



**Science and Information Technology Department**

European Tourist Perspective on Destination Satisfaction:  
A Business Analytics approach

Lisa Maria Luís Vieira

Dissertation submitted as partial fulfilment of the requirement for the  
degree of

Master in Computer Science and Business Management

Supervisor:

Prof. Raul M. S. Laureano, Assistant Professor, ISCTE Business School,  
Department of Quantitative Methods for Management and Economics

September, 2016



## **Acknowledgments**

After the completion of this project, it is important to thank everyone who, directly or indirectly, led to the final results of this research work.

Firstly, I would like to express my sincere gratitude to my advisor Professor and Supervisor Raul Laureano for the advices, guidance, availability, help, quick assistance to all of my requests and all the data mining knowledge conveyed to me. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my project.

Besides my advisor, I would like to thank my parents and family for supporting me throughout writing this thesis, and especially to my boyfriend who supported me and encouraged me to complete this project. I also extend my gratitude to everyone who believed in my effort and my devotion.



## **Abstract**

For many years that tourism information has been collected and stored, allowing increased interest in the data mining (DM) areas. This leads to a need of research and discovery of new patterns to develop automated procedures to improve the tourism knowledge management.

The relationship between the tourist characteristics and preferences and the tourist satisfaction was never studied in order to provide useful knowledge to the tourism industry. Therefore, there was the need to investigate the explanatory factors of the tourist satisfaction with the destination to allow the tourism companies to define the correct assumptions about a certain travel.

This dissertation used the data from Flash Eurobarometer 414 “Preferences of Europeans towards tourism 2015” with data from the 28 countries of the European Union (EU).

A predictive model was obtained for the tourist satisfaction, through the discovery of existing patterns in the process of the tourist travel, using DM techniques on the data referred above. The definition of an explanatory model allowed to obtain useful knowledge for tourism agencies, enabling the development of marketing strategies according to the tourist profile and ensuring development of promotional messages for products and experiences, ensuring that correct assumptions are made about their customers.

**Keywords:** Data Mining, Business Intelligence, Tourism, Tourist, Satisfaction, CRISP-DM.

## Resumo

Desde há muito tempo que é recolhida e armazenada informação sobre turismo, permitindo captar o interesse das áreas de *data mining* (DM). Consequentemente, surgiu a necessidade de pesquisa e descoberta de novos padrões para desenvolver procedimentos automatizados, de forma a melhorar a gestão deste tipo de informação.

A relação entre as características do turista, as suas preferências e a satisfação nunca foram estudadas extensivamente de forma a criar conhecimento útil para a indústria do turismo. Desta forma, havia a necessidade de investigar e estudar os fatores explicativos da satisfação do turista com o destino, para que seja possível às empresas de turismo traçar o perfil de turista adequado e transmitir as campanhas de *marketing* de forma assertiva e eficiente.

Nesta dissertação foram utilizados os dados do Flash Eurobarometer 414 “Preferences of Europeans towards tourism 2015”, que contém dados dos 28 países da União Europeia.

Através da descoberta de padrões existentes no processo de viagem do turista, utilizando técnicas de DM sobre os dados acima referidos, foi possível obter um modelo preditivo para a satisfação do turista. A definição de um modelo explicativo permitiu obter conhecimento útil para as empresas de turismo, facilitando o desenvolvimento de estratégias de marketing de acordo com o perfil do turista e de mensagens promocionais para produtos e experiências, garantindo que são definidos pressupostos adequados para os seus clientes.

**Palavras-chave:** *Data Mining*, *Business Intelligence*, Turismo, Satisfação, Turista, *CRISP-DM*.

## List of Abbreviations

AUC	Area under the ROC Curve
CART	Classification and Regression Tree
CHAID	Chi-squared Automatic Interaction Detector
CRISP-DM	CRoss-Industry Standard Process for Data Mining
DM	Data Mining
DT	Decision Tree
EU	European Union
eWOM	Electronic word-of-mouth
FN	False negative tuple
FP	False positive tuple
HF	Highest frequency
KMO	Kaiser-Meyer-Olkin
LF	Lowest frequency
LOS	Length of stay
M	Mean
MR	Multiple Regression
PCA	Principal Component Analysis
PMML	Predictive Model Markup Language
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SEMMA	Sample, Explore, Modify, Model, Assess
TN	True negative tuple
TP	True positive tuple
UNWTO	United National World Tourism Organization





# Index

Acknowledgments .....	iii
Abstract .....	v
Resumo .....	vi
List of Abbreviations .....	vii
List of Figures .....	xi
List of Tables .....	xiii
1 Introduction .....	1
1.1 Theoretical background and motivation .....	1
1.2 Objectives .....	2
1.3 Methodology .....	4
1.4 Structure .....	4
2 Literature review .....	5
2.1 Tourism: concept and importance .....	5
2.2 Travel phases .....	6
2.2.1 Pre-travel .....	6
2.2.2 On-travel .....	17
2.2.3 Post-Travel .....	21
2.3 Satisfaction .....	23
2.3.1 Destination attributes .....	23
2.3.2 Personal characteristics .....	26
2.3.3 Complaints .....	27
2.4 Overview of tourist satisfaction studies .....	30
3 The CRISP-DM methodology .....	35
3.1 Business understanding .....	36
3.2 Data understanding .....	37
3.3 Data preparation .....	46
3.4 Modelling .....	54
3.4.1 Cluster Analysis: Tourist satisfaction assessment .....	54
3.4.2 Classification Analysis: Predictors of tourist satisfaction .....	55
3.5 Evaluation .....	62
3.5.1 Models to perform .....	64
3.6 Deployment .....	64
4 Results and discussion .....	65
4.1 Tourist satisfaction assessment .....	65
4.2 Predictors of Tourist satisfaction .....	72
4.2.1 Sensitivity Analysis .....	85
4.3 Discussion .....	89
5 Conclusion .....	91
5.1 Synthesis .....	91
5.2 Contributions .....	91
5.3 Limitations of the Research .....	93
5.4 Further Research .....	93
6 References .....	95
7 Appendix .....	103
A. Tourist characterisation and satisfaction cluster .....	103
B. Tourist characterisation and formal complaint .....	104
C. Predictors of tourist satisfaction .....	108
D. Decision trees rules .....	112
I. Highest Accuracy .....	112
II. Less Satisfied .....	112
III. Very Satisfied .....	116



## List of Figures

Figure 1: Satisfaction with the destination.....	2
Figure 2: Satisfaction with the accommodation.....	3
Figure 3: Explanatory factors of satisfaction with the destination and accommodation.....	3
Figure 4: Determinants of destination image .....	10
Figure 5: Relation between travel phases and travel experience .....	12
Figure 6: Predictors of loyalty.....	34
Figure 7: CRISP-DM phases.....	36
Figure 8: Satisfaction assessment: dendrogram from the hierarchical cluster analysis .....	54
Figure 9: ROC Curve representation.....	63
Figure 10: Gains chart representation.....	64
Figure 11: IBM SPSS Modeler stream for the social demographics characteristics for tourist characterisation and on-travel variables .....	72
Figure 12: IBM SPSS Modeler stream for the pre-travel variables.....	72
Figure 13: ROC Curve for the model of the social demographics characteristics and tourist behaviour characterisation variables.....	74
Figure 14: Gains chart for the model of the social demographics characteristics and tourist behaviour characterisation variables.....	74
Figure 15: ROC Curve for the model of the pre-travel variables.....	75
Figure 16: Gains chart for the model of the pre-travel variables.....	75
Figure 17: ROC Curve for the model of the on-travel variables.....	76
Figure 18: Gains chart for the model of the pre-travel variables.....	76
Figure 19: ROC Curve for the model with the highest accuracy.....	77
Figure 20: Gains chart for the model with the highest accuracy.....	77
Figure 21: Explanatory attributes of the satisfaction with destination and accommodation .....	79
Figure 22: ROC Curve for the model with the highest percentage of less satisfied tourists correctly classified .....	80
Figure 23: Gains chart for the model with the highest percentage of less satisfied tourists correctly classified .....	80
Figure 24: Explanatory attributes of the satisfaction with destination and accommodation .....	82
Figure 25: ROC Curve for the model with the highest percentage of very satisfied tourists correctly classified .....	83
Figure 26: Gains chart for the model with the highest percentage of very satisfied tourists correctly classified .....	83
Figure 27: Explanatory attributes of the satisfaction with destination and accommodation .....	84
Figure 28: Sensitivity analysis for the model with highest accuracy (variable D3_1) .....	86
Figure 29: Sensitivity analysis for the model with highest accuracy (variable D3) .....	86
Figure 30: Sensitivity analysis for the model with highest accuracy (variable OT1.1_2) .....	86
Figure 31: Sensitivity analysis for the model with highest accuracy (variable OT1.1) .....	87
Figure 32: Sensitivity analysis for the model with highest accuracy (variable BT1_MT) .....	87
Figure 33: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D3) .....	88
Figure 34: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D6) .....	88
Figure 35: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D2) .....	89
Figure 36: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D1) .....	89
Figure 37: IBM SPSS Modeler 10-fold stream .....	108



## List of Tables

Table 1: Pre travel studies resume – Non-Europe .....	15
Table 2: Pre-travel studies resume .....	16
Table 3: On-travel studies resume .....	20
Table 4: Post-travel studies resume .....	22
Table 5: Destination attributes studied and their relationship with satisfaction .....	25
Table 6: Personal Characteristics of tourists studied and their relationship with satisfaction .....	27
Table 7: Demographic characteristics as predictors of satisfaction .....	30
Table 8: Pre-travel behaviours as predictors of satisfaction .....	31
Table 9: On-travel behaviours as predictors of satisfaction .....	32
Table 10: Post-travel behaviours as satisfaction predictors .....	33
Table 11: Description of the variables of the social demographics characteristics for tourist characterisation .....	38
Table 12: Descriptive statistics of the social demographics characteristics for tourist characterisation .....	39
Table 13: Descriptive statistics of the historical tourist behaviour for tourist characterisation .....	40
Table 14: Description of the pre-travel variables .....	41
Table 15: Descriptive statistics to the pre-travel variables .....	42
Table 16: Description of the on-travel variables .....	43
Table 17: Descriptive statistics to the on-travel variables .....	43
Table 18: Description of the post-travel variables .....	44
Table 19: Descriptive statistics to the post-travel variables .....	45
Table 20: Description of the variables created of the social demographics characteristics for tourist characterisation .....	47
Table 21: Descriptive statistics to the variables created of the social demographics characteristics for tourist characterisation .....	47
Table 22: Description of the pre-travel variables created .....	48
Table 23: Descriptive statistics to the variables created for the pre-travel phase .....	49
Table 24: Description of the variables created for the on-travel phase .....	50
Table 25: Descriptive statistics to the variables created for the on-travel phase .....	50
Table 26: Description for the variables created for tourist the post-travel phase .....	51
Table 27: Descriptive statistics to the variables created for the post-travel phase .....	51
Table 28: Varimax rotated component matrix and communalities .....	53
Table 29: Statistics of the tourist satisfaction .....	53
Table 30: Social demographic characteristics and tourist behaviour: variables description .....	57
Table 31: Pre-travel: variables description .....	57
Table 32: On-travel: variables description .....	57
Table 33: Post-travel: variables description .....	57
Table 34: Decision trees parametrization .....	58
Table 35: Confusion matrix .....	62
Table 36: Satisfaction assessment: clusters descriptive statistics and validation .....	65
Table 37: Registered formal complaints by satisfaction cluster .....	66
Table 38: The social demographics characteristics by satisfaction cluster .....	68
Table 39: Tourist behaviour characterisation by satisfaction cluster .....	70
Table 40: Confusion matrix for the model of the social demographics characteristics and tourist behaviour characterisation variables .....	73
Table 41: Confusion matrix for the model of the pre-travel variables .....	73
Table 42: Confusion matrix for the model of the on-travel variables .....	73
Table 43: Confusion matrix for the model with the highest accuracy .....	77
Table 44: Confusion matrix for the model with the highest percentage of less satisfied tourists correctly classified .....	79
Table 45: Confusion matrix for the model with the highest percentage of very satisfied tourists correctly classified .....	82
Table 46: Tourist country of nationality by satisfaction cluster .....	103
Table 47: The social demographics characteristics by formal complaint .....	105

Table 48: Tourist country of nationality by formal complaint .....	106
Table 49: Tourist behaviour characterisation by formal complaint.....	107
Table 50: Results of the models for the social demographics characteristics for tourist characterisation variables.....	109
Table 51: Results of the models for the pre-travel variables.....	109
Table 52: Results of the models for the on-travel variables.....	109
Table 53: Results of the models for the group with all variables.....	110
Table 54: Results for the boosting and bagging models performed .....	111

# 1 Introduction

## 1.1 Theoretical background and motivation

In the past few years, the use of the word “tourism” and the events and conferences related to this subject are more frequent. Tourism is one of the industries with more economic growth in the world and is one of the most income sources for several countries, but also a source of employment and private sector growth (Khan, 2014; UNWTO, 2016).

This industry includes many services, such as accommodation, entertainment, food, tickets (airplane or others) and culture. This creates a huge amount of money, which is a way of creating competition between countries, which leads to a higher investment in this industry (Bose, 2009).

Moreover, tourism represents about 10% of the worldwide gross domestic product, which is the proof of the tourism success. In 2015, arrivals from foreign tourist to the Europe increased 5% in the first semester of the year, comparing to the same period in 2014 (UNWTO, 2016).

In addition to the tourism growth, the evolution from Web 1.0 to the Web 2.0 brought new ways of communication and methods of searching for information, where tourism information is included. The sources of information to search for accommodation, travel tickets, prices and destinations had changed and tourism companies need to be aware and keep up with this changes in order to not lose customers. For example, before these customers, complaints were registered by the traditional way, talking to the staff or manager of the service. Now, more people start to complaint online in order to show their dissatisfaction with the service or with the staff of a specified place (Hernández-Méndez et al., 2013).

With the evolution of the Web, the amount of tourism information available is getting bigger. There is a lot of new ways to get information from users, which requires different methods to analyse tourism data. Data can be obtained through different ways, e.g. data stored by tourism operators (actors), travel agencies, hotels, flight companies or by massive surveys to operators or tourists. Data obtained by survey is always more difficult as people are not always available to fulfil (long) questionnaires.

Moreover, not always data collected through questionnaires is well explored and analysed, that is why some organizations make data available in public repositories in order to improve research about certain topics. For tourism industry we have Eurobarometer (European Commission, 2015a, b) since 2009. Eurobarometer annual report just shows descriptive results and does not explore complex relationships among tourists’ characteristics and travel behaviours.

In this context and with large samples, data mining techniques are often recommended to discover new and valuable knowledge, being the process of automatically discovering patterns from large amounts of data in a way to provide important and useful information to companies, in order to help them to pay attention to what is essential (Bose, 2009; Ioniță, 2015).

Data analysis provided by data mining techniques can be very useful to discover new information in order to identify tourists’ behaviour patterns and their preferences. The knowledge and understanding of the tourists’ profiles and travel patterns is important to tourism organizations, because

it allows a better match between the destinations and the tourists (Aghdam et al., 2013; Juwattanasamran et al., 2013).

According to Bose (2009), there is two flows of information. One is from the tourism companies to the tourists, which is information about the services that the tourists use and the other flow is about the tourist behaviour in response to the companies' services. In this project, the focus will be in study the tourist behaviour in response to the companies' services, i.e., study the satisfaction with the destination and accommodation services, in order to improve the information that tourism companies give to their customers.

## 1.2 Objectives

As mentioned above, the aim of this dissertation is to study the behaviour and habits of European tourists and its relation with satisfaction. In detail, the main objectives are:

1. Evaluate what factors create satisfaction for the tourist with both the destination and accommodation. These two important factors are essential in determining what can influence sustainability and help predict the future growth of the tourist industry. For example, a customer's loyalty with a particular destination and word-of-mouth advertising are important factors when identifying tourism growth.

The aim of this research project is to collect and analyse data from specific segments of the tourist industry and to see how they influence a customer's overall perception and how likely they are to return and/or recommend the destination to others. The areas to be analysed are as follows: i) nature; ii) activity/services quality; iii) cultural attractions; iv) prices; v) the reception by host community; vi) accessibility for people with special needs (Figure 1); and vii) satisfaction with accommodation, namely, quality and safety issues (Figure 2). Additionally, it will be studied the relation between satisfaction and formal complaints.

**Figure 1: Satisfaction with the destination**





**Figure 2: Satisfaction with the accommodation**



2. To identify what triggers a customer's feeling of satisfaction with both the destination and the accommodation. In this point, the goal is to identify the predictors that influence satisfaction with the destination and with the accommodation.

Among them we can find demographic characteristics, destination features, travel duration, information sources and booking method. To each of these factors, there are elements to be studied, which are described at the Figure 3.

**Figure 3: Explanatory factors of satisfaction with the destination and accommodation**



The results of the above analysis allow to do the following:

1. Identify profiles (segments) of European tourists who have similar behavioural patterns. This data can then be actively used by the tourism industry for appropriate marketing strategies.
2. This research delivers new statistical information about the behaviour and habits of European tourists. Once analysed this will provide valuable information for the tourist sector and also provide other students with data for further analysis and projects and this in turn may again be useful information for the tourist sector.
3. To test the effectiveness of data mining techniques where currently there is minimal available literature about its application.
4. To publish the findings to help the tourist industry link customers with destinations of their liking.

## 1.3 Methodology

This research used the CRISP-DM (CRoss-Industry Standard Process for Data Mining) methodology, which is composed of six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment (Chapman et al., 2000).

In the business understanding phase, the goal is to evaluate tourist satisfaction with the destination and accommodation, but also define the good satisfaction predictors.

These goals can also be defined as data mining goals, with a segmentation problem to identify tourist's profiles with different levels of satisfaction with the destination (environment, quality of activities and associated services, prices, cultural attractions, accessibility for people with special needs, reception by the host community, accommodation complaints, location of accommodation to main attractions and demographic characteristics) and with a classification problem to predict satisfaction with a low prediction error rate and also to identify which predictors are more important.

Given the expensive and time consumer activity of data collection and also that the data from Eurobarometer (European Commission, 2015a, b) is available and not appropriately studied, this study uses the data collected from tourism for 2015 in the 28 countries in the European Union, regarding main holiday in 2014.

In the second phase of CRISP-DM – data comprehension – 630 attributes collected from 30.101 European tourists were analysed and included in the Excel database. This size of data needs to be prepared and reduced, appealing to PCA (Principal Component Analysis) and other transforming procedures.

In the modelling phase, each goal uses two different algorithms. To the first goal, involving a segmentation problem, clustering analysis, namely, k-means algorithm was performed. To the second goal, consisting in a classification problem, decision trees were used with three different algorithms (CART, CHAID and C5.0).

In the model evaluation phase, as the database contains a large amount of data, it will be divided using a training sample (with 70% of the cases) and a test sample (with 30% of the cases). The model that best fits the goals will be chosen, based on the testing sample, taking in account the holdout and cross-validation (10-fold) methods results.

In the last phase, deployment, the intention is to present this dissertation and its results to the tourism industry by presenting at conferences and also publishing in scientific papers.

## 1.4 Structure

This dissertation follows a traditional structure. In addition to this introduction, it contains a further four sections. In Section 2 – Literature Review – an overview of tourism, including tourist behaviour, pre, during and post-travel behaviour, satisfaction with destination and with accommodation and a resume of the studies found. In Section 3 – Methodology – the CRISP-DM methodology will be described, with emphasis on the comprehension, preparation and evaluation phases. In Section 4 – Results – the findings are presented and discussed. In Section 5 – Conclusions – the conclusions are described focusing on contributions and study limitations.

## 2 Literature review

This chapter contains all the theory and studies about tourism and related issues that support this research. First it gives the definition of tourism and evidences its importance. Then, the three travel phases are described focusing on the tourists' behaviours on each one. In the third subchapter, the satisfaction explanatory factors are described and resumed.

### 2.1 Tourism: concept and importance

There are several tourism definitions, some more detailed and others more general. Since 1937 that exists a definition for tourism, given by the League of Nations (Sharpley & Telfer, 2002). Over time, tourism definitions have been changing and updated according to the decade and culture evolution. The most usual definition in the tourism industry and used for the quantitative measure of tourist traffic (Sharpley & Telfer, 2002) was created by World Tourism Organization (UNWTO, 1994:4-5): "Tourism comprises the activities of persons traveling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business, and other purposes" and it "refers to all activities of visitors, including both tourists (over-night visitors) and same-day visitors".

The term usual environment does not include trips within the area of usual residence, frequent and regular trips between the domicile and the workplace, and other community trips of a routine character, from the moment tourists leave until they return. Tourism is a worldwide industry, which includes hotels, transportation and the other services related to tourists and their needs and desires (UNWTO, 1994; Goeldner & Ritchie, 2009).

There are also definitions for traveller and tourist, which are different: a traveller is "any person on a trip between two or more countries or between two or more localities within his/her country of usual residence" (Goeldner & Ritchie, 2009:8) and all types off travellers are described as visitors. A tourist is in the category of visitors, such as same-day visitors. So, tourists are "visitors who stay in the country visited for at least one night" and same-day visitors are "visitors who do not spend the night in a collective or private accommodation in the country visited" (Goeldner & Ritchie, 2009: 8).

Tourists' activities have an impact in consumption of products and experiences when travelling. Tourists are looking for new experiences, either physical or emotional to satisfy their desires and needs. Their personality will define the choice of destinations and the attractions chosen (Goeldner & Ritchie, 2009; Khan, 2014).

Tourism is a source of economic development for most countries. It is important to several groups of people in a country from government to local community, because it helps the employment and allows the balance of payments. So, tourism is important to: i) the tourist; ii) the shops and providers of goods and services; iii) the government; iv) the politicians and; v) the destination community. As a result, countries want to invest in tourism to help economic growth and development on the local and regional level (Goeldner & Ritchie, 2009; Kachniewska, 2013).

The shops, providers of goods and services and the government of the destination take advantage of tourism, because they see it as an opportunity to make profits from selling to tourists. Politicians also take benefits from tourism, because it allows to collect tax receipts from tourism expenditures. In conclusion, tourism is a source of employment to the destination community and it is

seen by residents to be beneficial and like a positive factor to the destination. (Goeldner & Ritchie, 2009; Garau-Vadell et al., 2014).

## **2.2 Travel phases**

Travelling includes several stages, from the choice of the destination to the post-travel satisfaction or dissatisfaction and loyalty. The choice of the destination, the search of information and all the moments related to the travel that are done before travelling, are included at the pre-travel stage. Everything that is done during the travel such as the tourist behaviour at the moment of travel and travel duration are included at the on-travel phase. At last, the satisfaction, intention to recommend, intention to return and all the aspects after the trip are related to post-travel moment.

### **2.2.1 Pre-travel**

At the pre-travel stage, the main subjects to study are the sources of information and tools to search information about the destination, the motivations to choose a destination (push-pull approach and destination image), the sources of information to buy services and the accommodation and travel packages chosen. According to Akhoondnejad (2015), tourism managers must understand what factors may affect the choice of a destination.

#### **2.2.1.1 Sources of Information**

From the evolution of the Internet from Web 1.0 to Web 2.0, travel 2.0 tourism model emerged. This model allows users to “share their views and travel experiences with others” (Hernández-Mendéz et al., 2013:1001), which is an obstacle to tourism agencies, as information is viewed and disseminated by tourism agencies, but also Internet users. Information is now easily shared between blogs, social networks and all of Web 2.0 applications, making it difficult to control information.

Consumers have become independent, since they can search, collect and share information from the several information sources that Internet provides without the recurring of tourism agencies. This is a problem for tourism companies, because customers can buy and search travel information without using their services. To turn around this issue, tourism companies should also provide services through Travel 2.0 applications so that customers use these applications (Hernández-Mendéz et al., 2013).

The focus of tourism companies on these applications is relevant, as it allows to have direct relationships with users, but also to know their opinion about the company and what they need or want, which can be a great help to determine which products to sell.

From all the sources of information that can be used to search for travel information, the most relevant are: i) internet; ii) recommendations from friends and family and; iii) previous experiences. From the results of a study made by Hernández-Mendéz et al. (2013), when a tourist is choosing a destination or hotel, recommendations of friends and family have priority to recommendations from electronic users. In other hand, for Llodrà-Riera et al. (2015), Internet has a bigger influence than recommendations of family or friends.

Another important source of information is previous travel experience that was added by Kim et al. (2015), where this source took the second place, with Internet coming at first and recommendations of family or friends on third. The travel experience was studied by Kim et al. (2015) using four different generations, namely Silent Generation (1920s to the early 1940s), Baby Boomers (1946 to 1964), Generation X (early 1960s to 1970s) and Generation Y (1977 to 1994) as stated.

The studies concluded that the factor previous travel experience registered about 20% decrease from Silent Generation to Generation Y. Also, the youngest generations (Generations X and Y) search more about things to do at the destination, dining and entertainment, stores and places to shop. More than 50% of the Generation Y search for ideas of destinations to visit, which means this generation is open to choose a destination and do not have fix ideas. This can be explained by the lower travel experience of youngest generations (Kim et al., 2015).

Moreover, travel agencies are followed from friends and family and are mostly used by silent generation and diminish over the generations, as it happens with travel guidebooks. Additionally, older generations print and request more travel information and online brochures. In addition to these sources of information, travel agents, tourist offices, travel programs and advertising are the least used sources of information (Kim et al., 2015).

At least 20% of the respondents, belonging mostly from young generations, only use the Internet to search information for trip planning. Younger generations (Generation X and Generation Y) are more connected to information posted by other users, because they look at the comments that other travellers have posted, read travel blogs, watch and download videos and listen to travel audio files. They also are the generations that most print out coupons, which means that they are more concerned about saving money through promotions and discounts (Kim et al., 2015).

However, the youngest generations also use other types of sources, being the generations that most use travel documentaries, television and radio. Therefore, young people are divided between the ones who just use the Internet and who use a lot of other types of information (Kim et al., 2015).

Tourists who only use the Internet will not use traditional sources of information such as magazines, tourism companies, television and travel guidebooks. As they will only use the Internet, this is important information to define marketing strategies, since they will not respond to other type of advertising (Kim et al., 2015). About online sources of information, Hernández-Mendéz et al. (2013) concluded that the most important is official websites, i.e., official destination and official accommodation websites, but Llodrà-Riera et al. (2015) have put official websites next to search engines (e.g. Google) and Maps (e.g. Google Maps) as the most important, followed by websites with evaluations by users such as TripAdvisor and social networks.

The main travel website with user reviews used around the world is TripAdvisor which is an online application that allow users to plan their services such as flights, hotels and restaurants. In this platform, people can also comment and post their opinions about a certain service.

Kim et al. (2015) concluded that the information sources most used for trip planning are online travel agencies websites, tourism supplier websites such as airlines and hotels websites, search engines and destination websites. The older generations (Silent Generation and Baby Boomers) use more frequently the supplier websites and destination sites. The younger generations (Generation X and

Generation Y) prefer to use more frequently the search engines, destination websites and travel websites (e.g. personal blogs or social networks). Younger generations use more sources with indirect interaction with the supplier and tend to use “consumer generated content sites” and “social networking sites” (Kim et al., 2015:282), while older generations prefer direct interaction with the supplier. The most searched information by tourists is about a destination, hotel prices and availability, airline fares and schedule, which are the main services to choose when travelling (Kim et al. (2015).

Two-thirds of the respondents (66%) have never shared their experiences on a destination website, personal blogs or social networks and only 34% of Internet users have posted content on their blogs or websites (Hernandez-Mendéz et al., 2013). Although most women and men do not use travel website, with 60% of men and 69% of women never using it, women share less experiences online than men. The study revealed that when women’s experience of using Internet is low, just a few of them (9.3%) share their experiences online. As experience increases, more online experiences are shared. Despite the low usage of this type of information source, the information from these travel websites can be used by tourism companies to get to know what tourists think of the destination or tourism service, and use this information to improve their service and attract more clients (Hernandez-Mendéz et al., 2013).

SAß (2011) studied the use of Internet in Faro (Portugal) and concluded that 93% of the respondents of his survey use Google Search Engine as the first source of information. Despite search engines, travel blogs are more used than travel social networks. 38% of the respondents use always travel social networks such as TripAdvisor and 39% use them sometimes as a source of information (SAß, 2011; Llodrà-Riera et al., 2015). However, to Hernández-Mendéz et al. (2013), travel blogs or travel social networks are a second choice when compared to official websites.

The choice of official websites happens because they offer more detailed information, like the ways to get to the hotel, schedules and also allow customers to post comments and share opinions and experiences. Customers see official websites like a trustable source of information, without the need to search in another website (Hernandez-Mendéz et al., 2013). The reason why tourists use travel blogs is similar to official websites, as they can get more detailed information, “travel trivia”, “ideas” or “tips” (Hernandez-Mendéz et al., 2013: 1014), but also because they can read comments and reviews from other users that already had been there.

Travel social networks are one of the least used source of information, since they have essentially comments and feedbacks. This is a significant disadvantage, because tourists will have the need to search additional information about the destination or accommodation on other websites, which are probably official websites or travel blogs. However, the comments made online have influenced the majority of the respondents of the study made by Hernandez-Mendéz et al. (2013). They were classified according to their incomes and the results are that people with lower incomes are more influenced by electronic worth-of-mouth (Hernandez-Mendéz et al., 2013).

In conclusion, recommendations of friends and family and Internet are more effective to make someone chose a destination, but the priority differs between authors. Hernandez-Mendéz et al. (2013) concluded that friends and family are the most important source of information, but Llodrà-Riera et al. (2015) concluded that it is the Internet.

Differences between the conclusions of these authors can be explained by the country where the study was made. The study of Hernández-Mendéz et al. (2013) and Llodrà-Riera et al. (2015) was made in Spain and the study of Kim et al. (2015) was made in the United States of America, which are locations with very different cultures. Also, the study made in the United States of America (USA) was made after the others mentioned, which can lead to new variables (e.g. previous travel experience, maps and search engines) and therefore, different conclusions.

### **2.2.1.2 Motivations about the destination**

The motivations about the destination are the characteristics of the destination or any other reason inherent to the tourist that can lead the tourist to choose that destination. This section has two sub-sections: push-pull approach and destination image.

#### **2.2.1.2.1 Push-pull approach**

One commonly approach about motivations is the push–pull approach. Yoon and Uysal (2005) explained that tourists are pushed by their biogenic and emotional needs to travel and pulled by destination attributes.

The push factors, which are also known as motives, are emotions felt by the tourist. The most relevant push factors are novelty, knowledge-seeking, ego enhancement and rest and relaxation (Goossens, 2000; Dolinting et al., 2015; Yousefi & Marzuki, 2015).

On the other hand, pull factors explain the attractions or attributes of destination that match the interest of a person, such as climate, culture, among others. In particular, a combination of push and pull information and hedonic responses will motivate tourists to plan a trip (Goossens, 2000; Dolinting et al., 2015).

The most important pull factors resultant from the study of Yousefi and Marzuki (2015) are: i) environment and safety, ii) cultural and historical attractions, and iv) tourism facilities. In conclusion, from both push and pull factors, the most influencing to make a decision about a travel are novelty, knowledge-seeking and cultural and historical attractions (Yousefi & Marzuki, 2015). For instance, to the Baby Boomers generation, fun and enjoyment, relieve from stress and tension, escapism, relaxation, change and novelty are the main push factors, while the pull factors are attractiveness of the physical environment and better health (Naidoo et al., 2015).

The main motives to travel, which are a combination of both push and pull factors, are socialization with family/friends, discovery of new places, culture and historical attractions, shopping, relaxation, status and prestige, meet new people, influence of family or children, to break away from routine and pressure, education, among others. These motives can be divided into five different groups of motives, which are: i) recreational activities, ii) resting and relaxation, iii) personal values, iv) social experiences and, v) enriching and learning experiences (Vuuren & Slabbert, 2011; Hosany & Prayag, 2013).

According to Vuuren and Slabbert (2011) and Xu and Chan (2016), from those motives for travelling, resting and relaxation are the most important for tourists. The second and third motives to

travel are education and adventure. Personal values, such as spending time with family or friends is the motive with less value to the tourist (Xu & Chan, 2016).

Yuan and McDonald (1990) found that all people try to satisfy the same needs and the same desires. They also concluded that the attractions to choose a destination differ between countries and the importance given to that different attractions are not the same either.

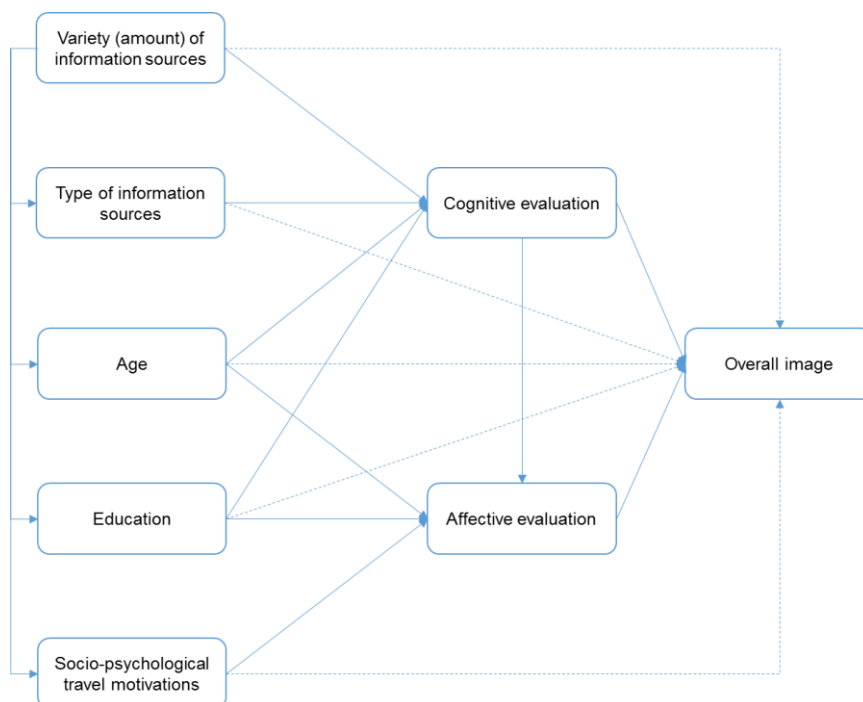
### 2.2.1.2.2 Destination image and the importance of information sources

Destination image is important in terms of understanding the travel behaviour and designing effective tourism marketing strategies, since it has a large impact on both supply and demand factors of marketing. This triggers the need to develop methodologies to comprehensively and accurately measure this concept in successful tourism development and destination marketing (Echtner & Ritchie, 2003; Tasci & Gartner, 2007).

The concept of destination image can be defined as “the sum of beliefs, ideas and impressions that a person has of a destination” (Crompton, 1979: 18).

Destination image was divided within two components by Gartner (1993): cognitive and affective. The cognitive component refers to the “beliefs and knowledge about an object” (Baloglu & McCleary, 1999:872) and the affective component refers to feelings about it. Both cognitive and affective evaluations form the overall image of a destination are presented in Figure 4 (Baloglu & McCleary, 1999).

**Figure 4: Determinants of destination image**



Source: Adapted from Baloglu and McCleary (1999:871).

The overall destination image consists of the cognitive and affective evaluation. The first one is composed by information sources, where is included the type and amount, but also by individuals age and education (Baloglu & McCleary, 1999). The affective evaluation is formed by age, education and tourist motivations (personal factors), when there is no previous experience. Information sources and



previous experience are designated stimulus factors, because they become from external factors and physical objects (Baloglu & McCleary, 1999; Frías et al., 2011).

The development of destination image was defined by Gunn (1988) in seven phases of the travel experience:

1. Accumulation of mental images about vacation experiences
2. Modification of those images by further information
3. Decision to take a vacation trip
4. Travel to the destination
5. Participation at the destination
6. Return home
7. Modification of images based on the vacation experience.

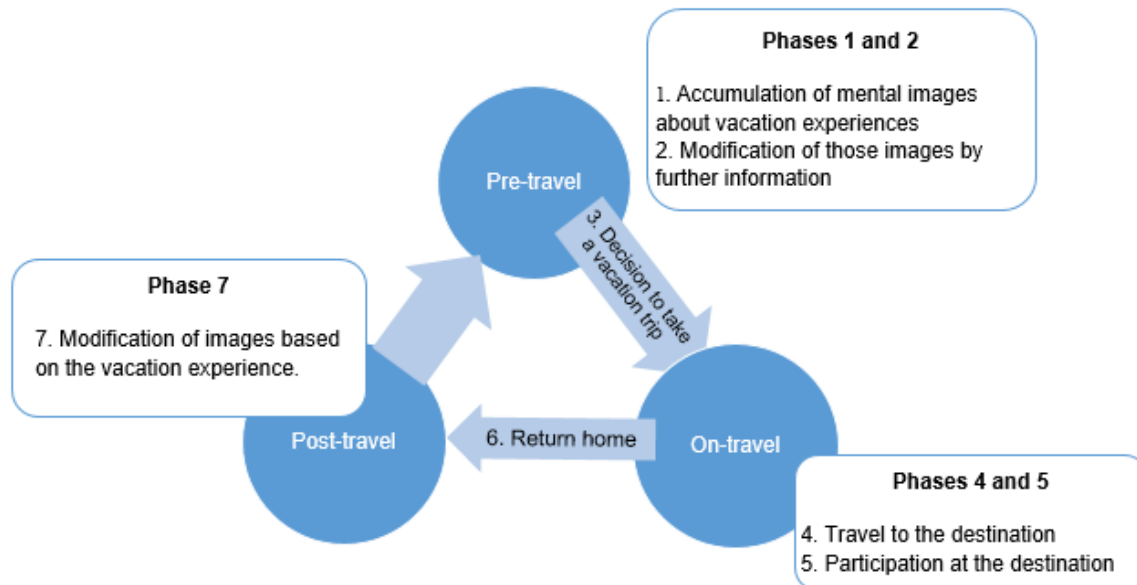
In phases 1 and 2, the reflection of secondary sources of information forms the destination image. These phases use several information sources and types of information to form images of the destination, which is an important factor to determine destination image (Echtner & Ritchie, 2003; Hanlan & Kelly, 2005). In phase 1, images are formed from media such as television and magazines, education and information obtained from friends and family. In phase 2, the image produced in phase 1 is modified by different kinds of information that is received by the tourist. This new image is referred to as "induced image" (Echtner & Ritchie, 2003:38). This information comes from travel sources, such as travel brochures, travel companies and travel guidebooks. The images developed from these sources should be the more realistic possible, since that a non-realistic destination image will lead to a negative word-of-mouth, resulting from the differences between this image and the reality (Beerli & Martin, 2004).

The phase 3 is the moment when the tourist decides to go on vacation, making this the last pre-travel phase. The phases 4 and 5 are on-travel phases, since phase 4 is the travel to the destination and 5 is the period when the tourist is at the destination.

In the phases 6 and 7, the tourist returns home and the destination images are modified based on travel experience. Consequently, destination image can change during the three big moments of travelling (pre-travel, on-travel and post-travel) (Echtner & Ritchie, 2003; Hanlan & Kelly, 2005).

The relationship between the seven different phases of travel experience and travel phases (pre-travel, on-travel and post-travel) are presented in Figure 5.

*Figure 5: Relation between travel phases and travel experience*



Source: Adapted from Echtner and Ritchie (2003) and Hanlan and Kelly (2005).

The process of destination image formation has two important parts. The first one is that people can have an image of a destination, even if they did not had contact with commercial information of that place. The best and worst factors of these images should be used to define marketing strategies that would be properly adjusted to the right target.

The separation of images of the people who have already been on the place and those who have not is also suggested by Echtner and Ritchie (2003), as someone that has already been in a certain destination has already a post-travel change of destination image, and on the other hand, someone that has not been in the destination, has not passed through this phase yet.

Ramseook-Munhurrun et al. (2015) concluded that destination image can be analysed through several aspects, namely, travel environment, attractions, events, infrastructure and sport. In addition to these aspects, destination image is influenced by the level of uncertainty avoidance of their national cultures (Frías et al., 2011).

Tourists from countries with high aversion to uncertainty cultures, such as France, Belgium or Italy, which only use the travel agency have a more favourable destination images. Tourists that use both the travel agency and Internet have a worst destination image, because travel agencies are in the bounds of their culture preferences and the risk associated to Internet is not. Destination image from these tourists is so negative, that it is similar to destination image of tourists from countries with low uncertainty avoidance cultures to the same sources of information (Frías et al., 2011).

In contrast, tourists from countries with low uncertainty avoidance cultures, for example people from United Kingdom, have no difference between using only travel agencies or both travel agencies and Internet. This happens because information sources have no impact to destination image of tourists from these type of countries. Tourists from countries with a high uncertainty avoidance cultures have a better destination image than tourists from countries with a low uncertainty avoidance culture (Frías et al., 2011; Abodeeb, 2014). A different culture can promote people to have different behaviour and “potentially quite contrasting perceptions about the same destination” (Abodeeb, 2014:31).

Dewar et al. (2007) concluded that previous experience and culture are a strong influence in the destination image. The differences between influences of Internet to tourists with different cultures result in differences in the destination image, because the influence of destination image depends from the congruency, incongruence or irrelevance of information sources, with regard to uncertainty avoidance (Dewar et al., 2007; Frías et al., 2011).

If the information is incongruent, the perception of people will be different about that destination depending from their uncertainty avoidance cultures. Incongruent information will not be remembered in the same way than congruent information will be, because congruent information is retained more accurately and will influence attitudes in a positive way. So, information from travel agencies is more correct and does not have inconsistencies, which leads to a better destination image than information from online sources (Frías et al., 2011).

### **2.2.1.3 Travel Booking**

Travel booking methods or ways to purchase travel products can be done using traditional ways, such as travel agencies or using the Internet, through supplier websites, online travel agent websites or portals for private houses (e.g. Airbnb) or hotel reservation (e.g. Booking.com).

The study of Kim et al. (2015) concluded that the most of the respondents use only the Internet, having a percentage of 35% tourists booked more than 75% of travel products or services on the Internet and having 25% of the tourists purchased more than 75% of travel products or services on the Internet. Tourism companies should be aware of this reality, because with most of the people booking or purchasing their services on Internet, companies need a strong online presence, services and a good reputation. The existence of the company services and products is becoming more important online than on physical agencies.

Travel reservations on supplier websites (e.g. airlines, hotels and rental cars) are higher for older generations (Silent Generation and Baby Boomers), but online travel agent websites (e.g. Travelocity, Expedia, eDreams) and tickets for plays, shows, concerts, festivals and events are searched mostly by younger generations (Kim et al., 2015).

This may be explained as older generations prefer direct interaction with the company, which is a result of the trust that older generations have in official companies and do not have in intermediaries and also that younger generations have a higher natural interest in events such as shows and concerts.

For Kim et al. (2015), from the several travel services available, the most booked travel services using the Internet are airline tickets and accommodation, having 60% of the respondents chosen these services. The fact that airline tickets are one of the most booked online travel services is easily explained by the trust that people have on them. This trust comes from the fact that most of the airline companies are known around the world and are big companies with credibility, which makes people trust them and have no concerns about purchasing airline tickets online. Referring to accommodation, people have more choices and different kind of places to stay if they search and book online, which is an understandable advantage when comparing to the traditional booking. For example, in a tourism company, a tourist cannot choose an apartment or a room as the ones that exist at Airbnb.com, for example.

However, according to Llodrà-Riera et al. (2015), websites related to accommodation such as portals for tourist accommodation letting (e.g. Ownersdirect) and portals for letting private houses or rooms (e.g. Airbnb) are not very used, with the exception of portals for hotel reservation, such as Booking.com.

#### **2.2.1.4 Accommodation and travel packages**

Accommodation refers to the types of accommodation used by tourists. The types of accommodation mostly used by tourists are: hotels, bed and breakfast, hostels, with family or friends, self-catering accommodation, in a camping site or in a guest house (Nash et al., 2006).

Travel services, where accommodation is included, can be purchased independently from each provider or together in a travel package. The effects of travel packages on tourist satisfaction differ between authors.

Differences between younger and older generations in the acquisition of travel packages were found by Kim et al. (2015). Younger people (Generation X and Generation Y) prefer travel packages, while the older generations (Silent Generation and Baby Boomers) usually buy travel services from different travel companies. This can be explained by the less time that young people have available, which they do not want to spend searching in several companies. Travel packages can be more advantaging due to time efficiency and convenience than buying independent tourism services (Naylor & Frank, 2001; Kim et al., 2015).

However, Chen et al. (2015) concluded that tourists that purchased travel packages had a low level of satisfaction and a lower intention to return.

Table 1 and Table 2 contain a resume of the results obtained from the studies described in the Pre-travel phase. Table 1 refers to the studies made in non-European countries, while Table 2 represents the studies made in Europe. Both tables include the country where the study was made and the main conclusions.

**Table 1: Pre travel studies resume – Non-Europe**

<b>Study</b>	<b>Conclusions</b>
Gunn (1988)	Creation of the destination image phases: accumulation of mental images about vacation experiences, modification of those images by further information, decision to take a vacation trip, travel to the destination, participation at the destination, return home and modification of images based on the vacation experience.
Echtner & Ritchie (2003)	Explanation of destination image phases.
Nash et al. (2006) Scotland	The types of accommodation used by tourists are: hotels, bed and breakfast, hostels, with family or friends, self-catering accommodation, in a camping site or in a guest house.
Hosany & Prayag (2013) United Kingdom (South east of England)	Five groups of tourists were defined according to the relationship between their emotional responses, satisfaction and behavioural intention.
Dolinting et al. (2015) Malaysia (Sabah)	Push-pull approach: tourists are pushed by their biogenic and emotional needs to travel and pulled by destination attributes.
Kim et al. (2015) United States of America (USA)	Information sources with most influence on tourists' choices are Internet, recommendations from friends and family and travel guides. The least important ones are suppliers, advertising and intermediaries.
	Websites used for travel planning are:
	- Tourism supplier websites such as airlines, hotels and rental cars (mostly used by Silent Generation and Baby Boomers);
	- Online travel agency sites (mostly used by Generation X and Generation Y);
	- Search engines (mostly used by Generation X and Generation Y);
	- Destination websites (mostly used by Silent Generation and Baby Boomers);
	- Social media (mostly used by Generation X and Generation Y).
	Younger generations use more intermediaries than older generations, who prefer direct interaction to build trust.
	Information searched is mainly about specific destination, hotel prices and availability and airline fares and schedule. More than 50% of the Generation Y search for ideas of destinations to visit.
	Younger generations search things to do at the destination.
Silent Generation, Generation X and Baby Boomers print or request more travel information or online brochures.	
Purchasing products	
- Most people (35%) buy travel products only on the Internet;	
- 25% of all respondents booked or purchased more than 75% of travel products or services on the Internet;	
- Travel reservations on supplier websites is higher in older generations (Silent Generation and Baby Boomers).	
Airline tickets and accommodation are the most booked travel services using the Internet, with 60% of the respondents.	
Portals for hotel reservation such as Booking.com are the most used for accommodation.	
Young generations prefer travel packages instead of buying the different services from different tourism companies or suppliers. Old generations prefer booking the services separately.	

**Table 2: Pre-travel studies resume – Europe**

<b>Study</b>	<b>Conclusions</b>
Yoon & Uysal (2005)	Push-pull approach: tourists are pushed by their biogenic and emotional needs to travel and pulled by destination attributes.
Frías et al. (2011) Spain (Andalusia)	<p>The level of uncertainty avoidance of tourists' cultures influences their destination images.</p> <p>Countries with high aversion to uncertainty cultures:            - That only used the Internet have more favourable destination images;            - That used travel agency and Internet have a negative image.</p> <p>Countries with low aversion to uncertainty cultures have no difference between the utilisation of travel agencies only or both, the travel agencies and Internet.</p> <p>Tourists from countries with a high uncertainty avoidance cultures have a more positive destination image.</p>
SAß (2011) Portugal (Faro)	<p>93% of the respondents use Google Search Engine as the first source of travel information.</p> <p>38% of the respondents use always travel social networks such as TripAdvisor and 39% use them occasionally.</p>
Hernández-Mendéz et al. (2013) Spain	<p>The information sources that most influence a tourist destination choice are recommendations from friends and family and electronic word-of-mouth (eWOM).</p> <p>The most used online applications/websites are official websites, travel blogs and travel social networks.</p> <p>People with lower incomes are more influenced by electronic word-of-mouth.</p> <p>The majority of the respondents have never published their travel experiences on an online source. Only 34% of Internet users have posted content on their own blogs or websites.</p> <p>People with higher expertise in the use of travel websites share more travel experiences.</p>
Llodrà-Riera et al. (2015) Spain (Majorca)	<p>Sources of information that most influence a tourist destination choice are the Internet, Travel Experience (to Generation X and Generation Y) and recommendations from friends and family.</p> <p>Traditional sources such as travel companies and travel guidebooks are more used by Silent Generation and Baby Boomers.</p> <p>Younger generations use more different kinds of sources, but are also the generation that use more only the Internet.</p> <p>Younger generations use with a higher frequency television, documentaries, movies and radio.</p> <p>Web platforms ordered by importance:</p> <ol style="list-style-type: none"> <li>1. Search Engines</li> <li>2. Maps</li> <li>3. Official websites</li> <li>4. Web pages with assessments by users (TripAdvisor)</li> <li>5. Social networks</li> <li>6. Web pages of intermediaries</li> <li>7. Web pages of suppliers</li> </ol>

## 2.2.2 On-travel

This chapter contains travel information about the frequency, duration or length of stay (LOS), behaviour and trip costs and expenditures. There are studies about the influence of these variables in tourist satisfaction.

### 2.2.2.1 Frequency

Frequency in this context is the amount of travels that a person does in a year. The amount of travels per year have been increasing due to the low cost flights and work trips. The low cost flights created a new travel concept: short-trips, which are trips of two or three days and allow to travel more times a year (Barros & Machado, 2010). The values of frequency found in the literature are usually between 1 and 30 times a year and the usual value for travel frequency for senior group is about 3 times (Losada et al., 2016).

Some demographic characteristics (age, gender, employment status, household size, number of financially dependent members in the household, household type), self-perceived health, economic status and time available were studied among seniors' tourists in order to determine their influence on travel frequency (Losada et al., 2016).

Gender, self-perceived economic status and self-perceived time available are the characteristics with a significant relationship to travel frequency. The study of gender revealed that female seniors travel more frequently than male seniors, which can be a result of the growth of female economic dependency in the last decades and also the higher number of women in the senior age. Self-perceived economic status has a positive relationship with travel frequency, given that people with more money have more possibilities to travel, while self-perceived time available has a negative relationship (Losada et al., 2016).

Therefore, the trips made by seniors are lower due to the decreasing of the perception of free time to travel, which increases with age. This is caused for the job absence and the time spend dealing with family issues (Cooper et al., 2007).

### 2.2.2.2 Length of stay

Duration or Length of stay (LOS) is the time (in days) that tourists stay travelling, i.e., staying in a place different than their homes. Peypoch et al. (2012) studied the influence of nationality (France, Austria, Canada and Italy), socio-demographic characteristics (age, education, gender and income), travel costs and destination attributes (gastronomy, nature, climate, sun and sea, security, physical appearance of population and life style) in length of stay (LOS). Authors concluded that nationality, socio-demographic characteristics and destination attributes have influence in length of stay. Specifically, age influences positively LOS, but also all the destination attributes, except gastronomy and lifestyle. Travel costs also influence LOS, but in a negative way, which can be easily explained by the fact that travelling with higher travel costs leads to a lower LOS.

### 2.2.2.3 Behaviour

The tourist behaviour is the influence of the tourist personality and characteristics when buying or experiencing travel products or experiences. It was studied by Gazley and Watling (2015) and the results of their studies suggested that there is a difference between a person when is not travelling and the same person when is travelling. People while travelling make different choices, such as, trying new food, meeting different people, going to an unusual kind of events and normally enjoying to know and experience the host culture.

The way that a person expresses feelings and ideas is a person's self-expression. Self-expression is different to each person and can influence their behaviour. Travel behaviour is influenced by self-expression, but also consumption of products and experiences during the travel. People who are more self-expressive want to show what they are achieving at the travel to themselves and others. Thus, their consumption of products and services is higher than people with less self-expression and is done to show that they have done that experience or got that product. Tourists who are less self-expressive will travel more and will not be concerned about acquiring products or experiences just to show to someone (Gazley & Watling, 2015).

The consumption of products increases with the growth of self-expression level, but the consumption of experiences does not. When people have a high level of self-expression, they want to show what they have done at the travel and products are something tangible that allow an easy display to other people. Experiences are submitted to judgments and are not tangible objects that the tourist can keep. Experiences have a more symbolic meaning than products. Moreover, consumption of products is also influenced by the push factors from a tourist country of origin and by the pull factors from the tourist destination. When both push and pull factors increase, the expression of a tourist's consumption is more expressive (Gazley & Watling, 2015).

The image that a tourist has of a product or experience and the real representation of it will also influence consumer behaviour. The level that a person sees a matching between his personality and a product or experience is named as self-congruity (Sirgy, 1982; Beerli et al., 2007; Hosany & Martin, 2012).

Beerli et al. (2007) added that if tourists are more involved with the travel, their personality, their self-expression and their self-congruity will be higher than tourists that do not have much involvement. When tourists are not involved with the travel, their self-concept is not correlated with the travel, which results in a lower involvement and self-congruity.

Gazley and Watling (2015) also concluded that people that are in high stimulation frequently, which means people that travel a lot and are constantly having different experiences, are less interested about the meaning of products and experiences they consume. There is a difference between tourist and traveller and this makes people have different behaviours and consume different kind of things. A tourist is more concerned about acquiring travel products and physical objects to take home and showing that they have "been there, done that, got the t-shirt" (Gazley & Watling, 2015: 651), while a traveller prefers consumption of experiences and enjoys the intangible nature of the destination.



According to Gazley and Watling (2015), if people travel with the purpose to relax and enjoyment, they are not concerned about consuming products, because they do not need to do it to facilitate or improve their self-esteem.

#### **2.2.2.4 Trip Costs and Trip Expenditures**

A study was made about trip costs and expenditures in order to identify their determinants, with students at college, with mean age at 22 years old. The variables used were: LOS, age, destination (travel inside or outside the country), repeat visitor and accommodation booking (Thrane, 2016).

Thrane (2016) concluded that the average of LOS is ten days and with a 10% increasing of length, trip costs increase 5%. This is just applicable until a certain number of days. For big values of LOS, this increase in costs disappears. Age is also related to trip costs and expenditures. Older students spend more money, with a 5% growth in trip costs and expenditures just by aging one year.

Moreover, the author concluded that going to big cities or foreign countries will increase costs. The trip costs from a foreign destination are much higher than domestic trip costs. However, when a person returns to the same place, the costs and expenditures decrease. This can be explained by the fact that people will not spend so much money in things that have already experienced (e.g. museums). Additionally, the tourist already knows the best places to stay, the restaurants and places to go.

From the results of the study made by Thrane (2016), with the increase in time, accommodation costs will also increase. However, for Kim et al. (2009), the prices of travel products decrease as the time to the booking is closer. Most of people book accommodation approximately 7.89 weeks before the travel. The study suggested that half of the respondents stayed in commercial accommodation, but people staying in private accommodation have 32% less costs than people staying in commercial accommodation (Thrane, 2016). Tourists who purchase a package to online travel companies have less expenses than who acquire independent services (Kim et al., 2009; Chen et al., 2015).

### 2.2.2.5 On-travel studies: resume

Table 3 contains a resume of the results obtained from the studies described at the On-travel phase. This table has the author, country and conclusions of the study.

**Table 3: On-travel studies resume**

Study	Conclusions
Sirgy (1982)	Created the self-congruity term, which is the level that a person sees a matching between her personality and a product or experience.
Beerli et al. (2007) Spain (Gran Canaria)	Self-congruity influences consumer behaviour. The involvement of tourists with the travel has a positive relationship with their self-expression and self-congruity.
Cooper et al. (2007)	With the age increase, the perception of free time to travel decreases.
Barros & Machado (2010) Portugal (Madeira)	The trips frequency per year is increasing due to low-cost flights and work trips.
Hosany & Martin (2012) Singapore	Self-congruity influences consumer behaviour.
Gazley & Watling (2015) United Kingdom and New Zealand	There is a difference between a person when is not travelling and the same person when is travelling. Self-expression influences travel behaviour, but also consumption of products and experiences during the travel. High levels of self-expression: high consumption of products, but not experiences. Low levels of self-expression: higher consumption of experiences, but not products. Push and pull factors increasing: the expression of a tourist's consumption is more expressive. High involvement with the travel: personality, self-expression and their congruity will be higher. Tourists that have stimulation frequently are less interested about the meaning of products and experiences they consume. Differences between a tourist and a traveller: 1. Tourist: has a higher consumption of products; 2. Traveller: want to enjoy the intangible nature of the destination.
Losada et al. (2016) Spain	Variables that have influence on travel frequency: 1. Gender: female seniors travel more frequently than male seniors. Self-perceived economic status: positive relationship with travel frequency, which means that people that have more financial capabilities travel more frequently; 2. Self-perceived time available: negative relationship with travel frequency, which means that when time available arises, travel frequency of senior tourists decreases. Travel frequency mean for this age group is about 3, with values from 1 to 30, which means that everyone that answered the survey did at least one travel on the year of 2014.
Thrane (2016) Norway	The average of LOS is ten days, but with a 10% increasing of length, trip costs increase 5%, but for big values of LOS, the increase in costs disappears. Age is related to trip costs and expenditures: as older students spend more money, with a 5% growth in trip costs and expenditures just by aging one year. Foreign countries as a destination lead to higher costs than domestic destinations. Accommodation costs grows up with the time. People staying in private accommodation have 32% less costs than people staying in commercial accommodation.

## 2.2.3 Post-Travel

This chapter contains the information about the influence of perceived value and negative factors of the destination in satisfaction, loyalty and intention to return.

### 2.2.3.1 Satisfaction, Loyalty and Intention to return

Some researchers used to examine the perceptions about the product, but according to Cohen et al. (2014), now they have moved away from this to the relationship between tourists and places as a determinant of satisfaction.

Overall satisfaction is “the extent of overall pleasure or contentment felt by the visitor, resulting from the ability of the trip experience to fulfil the visitor’s desires, expectations and needs in relation to the trip” (Chen & Tsai, 2007:1116).

Alegre and Garau (2010) concluded that the tourists’ judgment about the different attributes of the destination will determine a part of tourist satisfaction and intention to return. Dissatisfaction (negative) attributes have influence on tourists’ satisfaction, but their influence is lower than satisfaction (positive) attributes. For sun and sand destinations, negative factors such as overdevelopment, tourism congestion and environmental degradation are the most important factors which will influence the tourist not to return (Alegre & Garau, 2010).

Negative experiences at the destination do not define overall satisfaction, but make the destination less attractive, which reduces the probability of a return visit and probably intention to recommend. Both negative and positive factors define the overall post-travel image of a destination. So, how they affect a tourist is important to know. The study of negative factors is very important, because they can disadvantage the destination, in comparison with other destinations.

Chen and Tsai (2007) defined perceived value and behavioural intention (also called loyalty). Perceived value is the final value that the tourist retains; it is the result of what the tourist has received and what has been given. Behavioural intention is the willingness that a tourist has to recommend or revisit a certain destination. In this way, perceived value plays an important role in defining the level of satisfaction and behavioural intentions of tourists.

Perceived value, behavioural intentions and trust are related to satisfaction, being trust defined as “a willingness to rely on an exchange partner in whom one has confidence” (Moorman et al., 1992: 82). Kim et al. (2011) concluded that satisfaction has a significant positive effect on trust and trust and satisfaction have a significant positive effect on loyalty.

Loureiro and Kastenholz (2011) studied the influence of delight, arousal, positive affect and corporate reputation in tourist satisfaction and loyalty in a rural accommodation unit in Portugal. The conclusions are that perceived quality, unfulfilled expectations and corporate reputation have an impact on satisfaction, while positive affect does not. The authors also concluded that loyalty is explained by satisfaction (19%), reputation (23.6%), delight (10.4%) and perceived quality (6.9%). Moreover, the relationship between reputation and loyalty is higher than the other variables, being stronger than the relationship between delight and satisfaction with loyalty. However, the impact of corporate reputation on loyalty is higher than satisfaction, but the impact on loyalty is not enough to convince tourists to return

to the same accommodation unit despite a pleasant experience. The strongest determinant of satisfaction is perceived quality, while the strongest determinant of delight is arousal (Loureiro & Kastenholz, 2011).

A better destination image improves the tourist expectations, the propensity to perceive a higher quality, intention to return and recommend, and the destination perceived value. A better perceived value leads to a better level of satisfaction. So, the path “Destination image – Perceived value – Satisfaction” must be kept in mind of the tourism companies (Bigné et al., 2001; Xia et al., 2009; Ramseook-Munhurrun et al., 2015; Tsai, 2015).

A positive overall image of a place is a “prerequisite for successful tourism” (Akhoondnejad, 2015: 9). Destination image and satisfaction are the two most important variables to influence visitor’s behavioural intentions (Chen & Tsai, 2007).

Akhoondnejad (2015) made a study about Iran image and concluded that the image of that country before the visit was different to the image of the destination after the visit. Therefore, tourists do not stay with the first impression of the destination forever, because the emotions experienced at the visit will influence the image they have of the destination.

### 2.2.3.2 Post-travel studies: resume

Table 4 contains a resume of the results obtained from the studies described at the Post-travel phase. This table has the author, country and conclusions of the study.

*Table 4: Post-travel studies resume*

<b>Bibliography</b>	<b>Conclusions</b>
Bigné et al. (2001) Spain (Valencia)	Destination image has a positive effect on behavioural intentions.
	An improvement in the overall image of a place improves a tourist’s intention to return and to recommend.
Chen & Tsai (2007) Taiwan	Perceived value, behavioural intentions and trust are related to satisfaction.
	Destination image and satisfaction are the two most important variables to influence visitor’s behavioural intentions.
Alegre & Garau (2010) Spain (Majorca)	Judgment about the different attributes of the destination influences on tourist satisfaction.
	Overall satisfaction is the combination of the positive and negative attributes.
	Dissatisfaction attributes of a destination have an impact on the satisfaction of the tourist, but not as much as the satisfaction attributes of the place.
	Negative factors in a sun and beach destination that can make a person not to return are: <ol style="list-style-type: none"> <li>1. Overdevelopment;</li> <li>2. Tourism congestion;</li> <li>3. Environmental degradation.</li> </ol>
Kim et al. (2011) United States of America	Satisfaction has a significant positive effect on trust and trust and satisfaction have a significant positive effect on loyalty.
Akhoondnejad (2015) Iran (Isfahan)	Destination image can change during the visit to the destination, which means that the post-image is different from the pre-image of the destination.
	Satisfaction is influenced negatively by the duration of tourists’ stay.

## 2.3 Satisfaction

In all industries, it is important to measure satisfaction with every aspect of the product or service delivered. Almost everything can be measured and used to improve a service. This is also relevant in the tourism industry, where satisfaction can be measured using different aspects of the travel, such as destination attributes tourist personal characteristics and complaints.

### 2.3.1 Destination attributes

Destination attributes are the aspects related to the destination, such as place, the services offered there, people and environment. There are positive destination attributes, which are the positive factors of the destination, and negative destination attributes, which are the negative aspects. Each factor of the destination can provide a higher or lower level of satisfaction or not influence satisfaction at all. For example, negative destination attributes, such as construction intensity and water purity influence satisfaction in a negative way (Jarvis et al., 2016).

Destination attributes were studied by Sarra et al. (2015) in Lisbon and the results showed that safety, accommodation, urban land, architecture, atmosphere, local people, food and wine, geographic position and LOS make perceived quality higher, which leads to higher satisfaction, when compared to prices (the baseline item). Related to LOS, in the study of Jaafar and Khoshkam (2014), satisfaction is influenced negatively by the time tourists stay travelling, as people who travelled for more than 10 days were less satisfied than those who travelled for less days.

However, when the attributes are traffic, cleanliness, night life, handcrafts and traditions, the perceived quality is low, which leads to a lower satisfaction. These are important factors to take into account when trying to improve the tourism industry in Lisbon, because the positive factors can be used to sell the city and the negative ones should be used to improve themselves (Sarra et al., 2015).

Ragavan et al. (2014) concluded that tourists' perceptions of accommodation and food, climate, convenience, culture and people have a positive influence on satisfaction, while attractions, commodities and price have a negative influence. There are some differences between the dimensions of destination attributes studied by Ragavan et al. (2014) and Sarra et al. (2015), which are due to the different countries where the studies were made, Malaysia and Portugal (Lisbon), respectively. Nevertheless, the most important conclusion is that both studies agree that most of the destination attributes studied by them are related to satisfaction.

According to Chi and Qu (2009), the destination attributes that have a higher relationship with satisfaction are accommodation, attractions, environment and dining. The more positive these attributes, the higher the satisfaction. Some attributes are not significant in determining satisfaction, such as activity and events, shopping and accessibility.

Arasli and Baradarani (2014) studied the satisfaction of European tourists with some destination attributes in Jordan (Amman). The conclusions were that shopping, tourist attractions, environment, safety and food have a direct effect on satisfaction, while accommodation, restaurants, and transportation facilities do not have an influence on satisfaction.

Vajčnerová et al. (2014) studied the influence of several destination attributes in Brno (Czech Republic). Variables such as experiential activities, local transport, product packages, nature attractions and quality of available information do not have an impact on tourist satisfaction. On the other hand, the destination attributes that most influence tourist satisfaction are a friendly welcome and acceptance by the local residents, uniqueness of the destination, image of the destination, perceived safety, price levels of services and goods, quality of dining facilities, roads to the destination, transport accessibility, cultural and social attractions and quality of accommodation.

Xu and Li (2016) studied the customer satisfaction among different types of accommodation: i) limited-service hotels; ii) suite hotels with food and beverage; and iii) suite hotels without food and beverage. The conclusions were that the factors which influence the satisfaction are basically the same between different hotels types: i) location; ii) staff performance; and iii) room quality. Still, for each type of hotel, the factors had a different importance in customer satisfaction: i) good restaurant for full-service hotels; ii) good value for limited-service hotels and suite hotels with food and beverage and; iii) good complimentary breakfast for suite hotels without food and beverage.

Most authors who studied the influence of destination attributes on satisfaction studied very similar attributes. Chen et al. (2011) studied this relationship in Kinmen (Taiwan) and the satisfaction with the attributes were Warfare reserves (66%), historical relics (64%), beautiful scenery (60.6%), traveling security (57%), local hospitality (46.2%), local specialties (40.1%), entertainment facilities (26.1%) and low tour fee (16.5%). Warfare reserves cannot be compared with attributes from other countries, because they are a major attraction in Kinmen.

An analysis was made about the relationship between destination attributes and willingness to revisit Kinmen. The results were very interesting, because although tourists who wanted to revisit the place had a higher level of satisfaction with each attribute, tourists who did not have a willingness to revisit chose the same top attributes with higher satisfaction than the tourists who had a willingness to revisit. However, only 47.1% of the tourists showed willingness to revisit. Due to this low value, the satisfaction of tourists who travelled to different destinations (Japan, USA, Europe, Thailand, Singapore, China, Hong Kong, Australia, Canada and Malaysia) was studied to find the competitiveness between countries. The results showed that the destination's competitiveness is not correlated to the tourist satisfaction (Chen et al., 2011).

From the study made by Vajčnerová et al. (2014), the average value of overall satisfaction is 6.95, but just a few satisfaction components have higher values than this. The factors that have a lower value of satisfaction are price level of services and goods, uniqueness of the destination, perceived safety, image of the destination and friendly welcome and acceptance by the local residents. Vajčnerová et al. (2014) also concluded that probably these are the factors where the expectations were lower due to a lack of promotion or to some unusual attributes of the destination.

As it was referred above, the most studied attributes are similar to all the authors mentioned. These similarities and differences can be observed in Table 5, which resumes the destination factors and their influence on satisfaction. The destination factors are distributed in three columns, according to their negative, positive or no influence on satisfaction.

**Table 5: Destination attributes studied and their relationship with satisfaction**

<b>Bibliography</b>	<b>Positive influence on satisfaction</b>	<b>Negative influence on satisfaction</b>	<b>No relation</b>
Chi & Qu (2009) United States of America	Accommodation, attractions, environment and dining.		Activity and events, shopping and accessibility
Chen et al. (2011) Kinmen	Warfare reserves (66%), historical relics (64%), beautiful scenery (60.6%), traveling security (57%), local hospitality (46.2%), local specialities (40.1%), entertainment facilities (26.1%) and low tour fee (16.5%).		
Araslı & Baradarani (2014) Jordan (Amman)	Shopping, tourist attractions, environment, safety and food.		Accommodation, restaurants, and transportation facilities.
Jaafar and Khoshkam (2014) Malaysia (Langkawi)		LOS	
Ragavan et al. (2014) Malaysia	Perceptions of accommodation and food, climate, convenience, culture and people.	Attractions, commodities and prices.	
Vajčnerová et al. (2014) Czech Republic (Brno)	Host community, uniqueness of the destinations, image of the destination, perceived safety, price levels of services and goods, quality of dining facilities, roads to the destination, transport accessibility, cultural and social attractions and quality of accommodation.		Experiential activities, local transport, product packages, nature attractions and quality availability of information
Sarra et al. (2015) Portugal (Lisbon)	Safety, accommodation, urban land, architecture, atmosphere, local people, food and wine, geographic position and LOS.	Traffic, cleanliness, night life, handcraft and traditions.	
Jarvis et al. (2016) Australia (Queensland)		Construction intensity and water purity	

### 2.3.2 Personal characteristics

The personal characteristics of the tourist, such as place of residence, age, income, level of education, personal motivation to travel, previous visit experiences, gender and marital status are important factors to determine satisfaction (Ragavan et al., 2014; Sarra et al., 2015; Lu et al., 2015).

Lu et al. (2015) concluded that place of residence (nationality) influences satisfaction, but also concluded that it causes variances in the pre-travel and on-travel phases.

Jarvis et al. (2016) concluded that income has a positive influence in tourist satisfaction. However, Lu et al. (2015) concluded that income has a negative influence in travel motivations and tourist satisfaction, which means that a higher income leads to a lower level of travel motivations and satisfaction. Travel motivations have a positive correlation with shopping behaviour and satisfaction, which means that a tourist, motivated to travel, will have a higher consumption of products/experiences and a higher satisfaction too (Lu et al., 2015).

From a study made by Sarra et al. (2015) to evaluate tourist satisfaction in Lisbon, the results showed that factors such as previous visit experience and respondents' age do not affect tourist satisfaction. This does not match with Lu et al. (2015), who concluded that age is negatively correlated with satisfaction and that younger people with a wider education tend to have higher satisfaction levels than older people. Ragavan et al. (2014) also concluded that age has implications on satisfaction, because there are different degrees of satisfaction and differences in age leads to different levels of satisfaction. For example, for tourists over 40 years old, the perception of culture has a positive effect, whilst for young tourists there is no relation.

There is a relationship between marital status and satisfaction, because for single tourists, climate was important to determine satisfaction, whilst for married tourists, climate has no influence on satisfaction. Also, for married tourists, the host people are more important to determine satisfaction than for single people. Being a female or a male is also related to satisfaction. Females pay more attention to accommodation and food, climate and price. For females, the perception of convenience has a negative effect on satisfaction, but for men there is no significant relationship. Men place more importance on people, whilst for females there is no influence on satisfaction (Ragavan et al., 2014).

The expectations about a destination can influence perceived quality and the tourist satisfaction, but not perceived value. When the tourist expectation is fulfilled, the destination attributes might not have a strong influence in satisfaction. However, if the tourist expectations were not fulfilled or did not exist, the destination attributes influence satisfaction significantly (Xia et al., 2009; Vajčnerová et al., 2014). In addition to expectations, also emotional place attachment, which is an "affective bond or link between an individual and special places" (Hidalgo & Hernandez, 2001:274), has a positive influence on tourist satisfaction (Tsai, 2015).

Table 6 demonstrates the relationship between the tourist personal characteristics and satisfaction. The personal characteristics are distributed in four columns, according to their positive or negative relation, or relation and no relation on satisfaction.



**Table 6: Personal Characteristics of tourists studied and their relationship with satisfaction**

<b>Bibliography</b>	<b>Positive influence on satisfaction</b>	<b>Negative influence on satisfaction</b>	<b>Relation</b>	<b>No relation</b>
Xia et al. (2009) China (Guilin)	Tourist expectations			
Ragavan et al. (2014) Malaysia	Age		Marital status Gender	
Lu et al. (2015) Taiwan	Perceived quality Travel motivations	Personal income Age Level of education	Place of residence	
Sarra et al. (2015) Portugal (Lisbon)				Previous visit experience Age
Tsai (2015) Taiwan	Emotional place attachment			
Jarvis et al. (2016) Australia (Queensland)	Personal income Perceptions of personal safety			

### 2.3.3 Complaints

Complaints are important for measuring satisfaction in any industry, as they demonstrate where a service or product may need to be improved or changed. The effective management of complaints leads to customer satisfaction, retention and loyalty. The importance of complaints for analysing satisfaction and loyalty differs from one author to another.

When a tourist complains, something is not working to meet their expectations, therefore action should be taken to resolve this. Thus, complaints are negatively correlated with satisfaction. With an increase in satisfaction, complaints will decrease (Xia et al., 2009; Song et al., 2011).

Song et al. (2011) found that complaints have little effect on loyalty, which is explained by Zhan et al. (2015), that found that in average the number of complaints is small, which can be a reason for complaints to have little or no effect on customer loyalty.

The only activity where Song et al. (2011) found a significant relationship between complaints and loyalty was with retail shops, hotels and tour operators. However, Xia et al. (2009) concluded that complaints with the features of the destination have a significant and negative relationship with loyalty. The difference between the conclusions of these two authors can be because Song et al. (2011) was measuring three specific services at the destination and Xia et al. (2009) was measuring the general destination attributes.

Complaining does not come easily to everyone. Some people do not like to complain, because they feel embarrassed, whilst others do not experience the same problem. Song et al. (2011) concluded that most of the respondents did not want to make a public complaint, for the above reasons.

Customers make online complaints because others can learn of their bad experiences and perhaps be persuaded not to go there themselves. Poor handling of complaints at a hotel can be another

reason for on-line complaints and hotel management would be well advised to take note of such complaints in order to maintain a good reputation (Sparks & Browning, 2010).

### **2.3.3.1 Predictors of complaints**

Personal characteristics such as nationality, level of education, age and income act like drivers to complain. Ngai et al. (2007) studied the differences between Asian and non-Asian people concerning private complaint behaviour and they found that Asian people make more private complaints, pointing to nationality making a difference to the complaint behaviour. Asian people complain less to the hotel because they are less familiar with the process of complaining and are afraid of “losing face” (Ngai et al., 2007:1387). This makes access to complaint channels of greater importance to an Asian complainant. However, they make more private complaints than non-Asian people. This means they will spread negative opinions by word of mouth about their bad experiences at the hotel more than non-Asian people. Both nationalities, Asian and non-Asian, indicated a preference for having their complaint dealt with immediately instead of other more formal and lengthy methods for resolution, such as compensation.

Concerning age, different results were found by Lam and Tang (2003) and Ngai et al. (2007). The first authors concluded that younger groups (24 years old and below) have a greater tendency to engage in public complaint behaviour than the older groups, whilst the second authors concluded that people in the medium age group (between 31 and 50 years old) are the ones most likely to engage in public complaint behaviour.

Level of education is also related to complaint behaviour. Ngai et al. (2007) concluded that people with a higher level of education are less likely to make public complaints, but Lam and Tang (2003) concluded the opposite, which means that people with a higher level of education are more likely to engage in different methods of complaining, including uncomplimentary word of mouth, complaints to the management and the Internet. The differences between the conclusions of the two studies is possibly because the first one is more recent than the second one. Although both studies were made in Hong Kong, the study of Ngai et al. (2007) was undertaken in the passenger lounge of the Hong Kong International Airport and the study of Lam and Tang (2003) was undertaken in hotel restaurants in Hong Kong.

Income is also important for measuring complaint behaviour. Complaining to management is related with income, which is greater when income is higher. Uncomplimentary word of mouth, and the Internet are not related with income. The relationship between gender and complaints was also studied and no significant relationship was found between them meaning that the probability of a woman or a man complaining is the same (Lam & Tang, 2003; Sari et al., 2013).

### 2.3.3.2 Reasons to complaint

A total of two hundred complaints from TripAdvisor were studied by Sparks and Browning (2010). The reason for the complaints were classified as: features internal to the hotel (e.g. room features, customer service, public areas, star reference, food or beverage and ambience) and features external to the hotel (e.g. location relative to main attractions). Of the internal features of the hotel, the main complaints were about room features, consumer service and public areas.

Mainly complaints about the room related to the size of the room, cleanliness, condition of the furniture and the equipment in the room. Concerning customer service, complaints were mainly about the behaviour of the employees towards the guests and the conduct of the owners or management of the hotel. These were characterised as unhelpful, unfriendly, uncooperative, aggressive, rude, incompetent and failure to apologise for the problems experienced by the guests. The reasons for complaining about public spaces were general conditions of the hotel (being old, shabby, dirty, smelly, outdated décor and poor facilities, such as swimming pools, restaurants, parking and entertainment) (Sparks & Browning, 2010).

Other complaints of less frequency are also important. Guests commented that their expectations had not been met, pointing to misleading hotel marketing with the experience. Also, some guests made reference to the price that travel agencies charge to make the travel arrangements (Sparks & Browning, 2010).

An analysis was made of complaints in the accommodation sector reported website of the accommodation sector in Turkey. The main problems found were with catering, guestroom, cleanliness and employee rudeness. The complaints about noise were subject to few complaints, corresponding to only 0.4% of the issues raised. It was found that people who experienced problems with cleanliness, also had problems with food and beverage services, guestrooms and with unresponsive staff of the hotels. Problems with catering are linked with guestroom issues and lack of communication from employees (Sari et al., 2013).

The problems above are related with the main services provided in a hotel, which suggests that accommodation providers are not sufficiently attentive of the quality of services they provide and to the complaints made by customers. From a total of 397 online complaints, only thirteen were resolved. The number of complaints shows that the customers studied were of the opinion that complaining would make a difference and help companies to improve their services (Sari et al., 2013).

## 2.4 Overview of tourist satisfaction studies

This chapter describes the most important studies and their corresponding conclusions. All tables give the author(s) and the main goals of the study, sample characterisation, methods used and results. The first table (Table 7) contains the demographic studies, and the conclusions from the studies related to demographic characteristics of the tourist, such as culture, age and gender.

**Table 7: Demographic characteristics as predictors of satisfaction**

Bibliography			
	Frías et al. (2011)	Hosany & Prayag (2013)	Xu & Li (2016)
Goals	Study the influence of culture as a moderating variable in pre-visit tourist destination image formation, through the information sources utilized by tourists in the selection of a holiday destination.	Investigate relationships between tourists' emotional profiles and their post-consumption evaluations of satisfaction and intention to recommend.	Study about the influence of different types of accommodation in satisfaction.
Sample	Valid responses: 371 Local: Andalusia, Spain Questionnaire to tourists of different nationalities (United Kingdom, France, Italy, and Belgium) Year: non-defined	Valid responses: 520 Local: South East of England, UK Questionnaires in postage envelopes Year: non-defined	Valid responses: 3480 Local: 100 largest U.S. cities The data was collected from the third-party website <a href="http://www.booking.com">www.booking.com</a> . Year: non-defined
Methods	Hypothesis tests (ANOVA) Quasi-experimental designs	Clusters (K-means) Hypotheses tests (Chi-square and one-way ANOVA)	Term Frequency Matrix Transformation A term frequency-inverse document frequency (TF-IDF) weighting method Singular Value Decomposition
Results	<ul style="list-style-type: none"> <li>➤ The level of uncertainty avoidance of tourists' cultures influence their destination images.</li> <li>➤ Tourists from countries with a high uncertainty avoidance cultures had a more positive destination image.</li> <li>➤ The results show that the formation of a destination's pre-visit image among tourists, based on the information sources they use, is moderated by the level of uncertainty-avoidance of their national cultures.</li> </ul>	<p>Tourists' emotional responses correspond to satisfaction and behavioural intention. Five groups of tourists were found:</p> <ul style="list-style-type: none"> <li>➤ Delighted: highest levels of satisfaction and intention to recommend.</li> <li>➤ Negatives: lowest levels.</li> <li>➤ Passionate: second highest mean scores.</li> <li>➤ Mixed: relatively high levels.</li> <li>➤ Unemotionals: above average for both.</li> </ul>	<ul style="list-style-type: none"> <li>➤ The factors which influence the satisfaction are basically the same between different hotels types (limited-service hotels, suite hotels with food and beverage, suite hotels without food and beverage): i) location, ii) staff performance and iii) room quality.</li> <li>➤ For each type of hotel the factors had a different importance in customer satisfaction: i) good restaurant for full-service hotels, ii) good value for limited-service hotels and suite hotels with food and beverage and iii) good complimentary breakfast for suite hotels without food and beverage.</li> </ul>

Table 8 concerns the pre-travel studies, including references to the preferred sources of information and of booking sources, the tourism information searched, the influence of the destination image in tourist satisfaction and behavioural intentions (i.e. intention to return).

**Table 8: Pre-travel behaviours as predictors of satisfaction**

<b>Bibliography</b>					
	Chen & Tsai (2007)	SAß (2011)	Hernandez-Mendéz et al. (2013)	Kim et al. (2015)	Llodrà-Riera et al. (2015)
<b>Goals</b>	Impact of destination image on behavioural intentions.	Discover the degree of Internet usage of travellers and tourists as well as their preferences in online advertising.	Confirm how Travel 2.0 applications have influenced tourist decision-making behaviour.	Identify several aspects of Internet use among four generational groups (Silent Generation, Baby Boomers, Generation X, and Generation Y) over a six-year period.	To identify the most popular sources of information used by tourists.
<b>Sample</b>	Valid responses: 393 Local: Kengtin, Taiwan Surveys to individuals over the age of 18 years. Year: 2004	Valid responses: 248 Local: Faro, Portugal. Questionnaire Year: non-defined	Valid responses: 616 Local: Spain Personal online survey to Spain Internet users with ages 16–64 Year: 2010	Valid responses: 9691 Local: United States of America (USA) An Internet-based survey was conducted each January from 2007 to 2012	Valid responses: 541 Local: Mallorca, Spain Online questionnaire
<b>Methods</b>	Principal component analysis (PCA). Structural equation modelling (SEM)	Hypothesis tests	Hypothesis tests Clustering Decision trees	Cross-tabulation and chi-square test.	Principal Components Analysis
<b>Results</b>	The results show that destination image has both direct and indirect effects on behavioural intentions. Destination image → trip quality perceived → value satisfaction → behavioural intentions. ➤ Destination image and satisfaction are the two most important variables to influence visitor's behavioural intentions.	➤ 93% of the respondents use Google Search Engine as the first source of information ➤ 38% of the respondents use always travel social networks such as TripAdvisor and 39% use them sometimes.	Main sources of information: ➤ Friends and family ➤ Internet ➤ Official Websites ➤ Travel blogs ➤ Travel social networks ➤ People with lower incomes are more influenced by eWOM. ➤ People with greater expertise in the use of travel websites share more travel experiences.	Main sources of information: ➤ Internet ➤ Travel Experience ➤ WOM. Information searched: a specific destination, hotel prices and availability and Airline fares and schedule. Purchasing products: ➤ 25% of all respondents booked or purchased more than 75% of travel products or services on the Internet; ➤ Airline tickets and accommodation are the most booked travel services using the Internet. Portals for hotel reservation are the most used for accommodation.	Main sources of information: ➤ Internet ➤ Search engines ➤ Official websites ➤ Previous travel experience ➤ Recommendations from friends and family.

Table 9 describes the on-travel studies, which includes variables that can influence travel frequency and also the effects of the level of self-expression in travel behaviour and consumption of products and experiences while travelling.

**Table 9: On-travel behaviours as predictors of satisfaction**

		<b>Bibliography</b>		
		Gazley & Watling (2015)	Losada et al. (2016)	Thrane (2016)
<b>Goals</b>	Discover what benefits a tourist has through the symbolic consumption of tourist products and define what influences some variables (self-concept, self-congruity, motivation) have in symbolic consumption behaviour within a tourism setting.	Identify variables that influence travel frequency among Spanish senior tourists	Examines the factors that influence length of stay at summer vacation destinations	
<b>Sample</b>	Valid responses: 410 Local: Leeds University, United Kingdom (UK), and Victoria University of Wellington, New Zealand. Online survey to young adults who had recently completed an overseas experience (OE). Year: non-defined	Valid responses: 358 Local: Spain Telephone survey to seniors over 55 years of age. Year: 2012	Valid responses: 436 Local: Norway Questionnaires in a Norwegian university college Year: 2014	
<b>Methods</b>	Multiple regression analysis	Neg-Bin model, based on a Negative Binomial distribution.	OLS regression model, Weibull survival model and Zero-truncated negative binominal regression model	
<b>Results</b>	<ul style="list-style-type: none"> <li>➤ Self-expression influences travel behaviour, but also consumption of products and experiences during the travel.</li> <li>➤ High levels of self-expression: high consumption of products, but not experiences;</li> <li>➤ Low levels of self-expression: higher consumption of experiences, but not products.</li> <li>➤ Push and pull factors increasing: the expression of a tourist's consumption is more expressive.</li> <li>➤ High involvement with the travel: personality, self-expression and their congruity will be higher.</li> </ul> Differences between a tourist and a traveller: <ul style="list-style-type: none"> <li>➤ Tourist: have a higher consumption of products;</li> <li>➤ Traveller: intangible nature of the destination.</li> </ul>	Variables that have influence on travel frequency: <ul style="list-style-type: none"> <li>➤ Gender: female seniors travel more frequently than male seniors.</li> <li>➤ Self-perceived economic status: positive relationship with travel frequency.</li> <li>➤ Self-perceived time available: negative relationship with travel frequency.</li> <li>➤ Travel frequency mean for this age group is about 3.</li> </ul>	<ul style="list-style-type: none"> <li>➤ The average of LOS is ten days, but with a 10% increasing of length, trip costs increase 5%, but for big values of LOS, the increase in costs disappears.</li> <li>➤ Age is related to trip costs and expenditures: as older students spend more money, with a 5% growth in trip costs and expenditures just by aging one year.</li> <li>➤ Foreign countries as a destination lead to higher costs than domestic destinations.</li> <li>➤ Accommodation costs grows up with the time. People staying in private accommodation have 32% less costs than people staying in commercial accommodation.</li> </ul>	

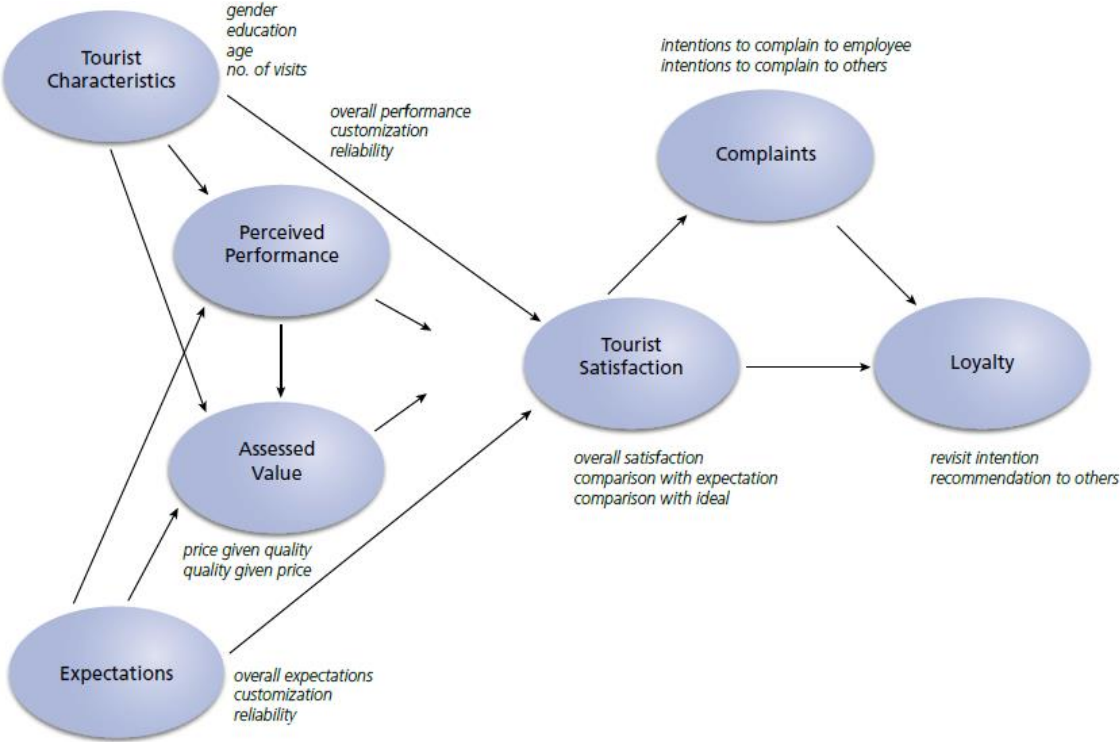
Table 10 contains studies of the variables related to the effects of the destination image of a country in satisfaction and intention to return and recommend, the features that can influence tourist satisfaction and the relationship between satisfaction and loyalty to the destination.

**Table 10: Post-travel behaviours as satisfaction predictors**

Bibliography				
	Bigné et al. (2001)	Alegre & Garau (2010)	Akhoodnejad (2015)	Kim et al. (2015)
Goals	Clarify the interrelationships among destination image, perceived quality, satisfaction, intention to return and willingness to recommend the destination.	Compare positive and negative attributes influence on tourist satisfaction for a certain destination.	Understand the perceptions of foreign tourists traveling to Iran of the image of Iran (both pre- and post-travel), trip value, satisfaction, intention to revisit Iran, and likelihood to recommend Iran as a travel destination to others.	Identify several aspects of Internet use among four generational groups (Silent Generation, Baby Boomers, Generation X, and Generation Y) over a six-year period.
Sample	Valid responses: 251 in Peníscola and 263 in Torrevieja Local: Peníscola and Torrevieja, Valencia Year: 2004	Valid responses: 2423 Local: Majorca (Balearic Islands). Survey with 24 dissatisfaction attributes and 13 for satisfaction about their summer holidays in 2004, 2005 and 2006.	Valid responses: 298 Local: Isfahan, Iran. Survey of international tourists, who travelled to Iran for the first time. Year: summer and autumn of 2013 as well as in early winter of 2014.	Valid responses: 9691 Local: United States of America An Internet-based survey was conducted each January from 2007 to 2012
Methods	Structural Equation Model Chi-square statistic Analysis of variance and Spearman coefficient. Path analysis.	Principal Component Analysis	Pearson correlation coefficient Sign Test analysis Structural equation modelling (SEM).	Cross-tabulation (and chi-square) analysis.
Results	<ul style="list-style-type: none"> <li>➤ An improvement in the overall image of a place improves a tourist intention to return and to recommend. This causal relationships were found:</li> <li>➤ Image → Quality</li> <li>➤ Quality Satisfaction →</li> <li>➤ Image → Satisfaction</li> <li>➤ Image → Satisfaction Quality</li> <li>➤ Return</li> <li>➤ Image → Recommendation</li> <li>➤ Image → Return</li> <li>➤ Satisfaction → Recommendation</li> </ul>	<ul style="list-style-type: none"> <li>➤ Overall satisfaction is the combination of negative and positive attributes.</li> <li>➤ Dissatisfaction attributes from a destination have impact in the satisfaction of the tourist, but not so much like the satisfactory attributes of the place.</li> <li>Negative factors in a sun and beach destination that can make a person not to return are: <ul style="list-style-type: none"> <li>➤ Overdevelopment.</li> <li>➤ Tourism congestion.</li> <li>➤ Environmental degradation.</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>➤ Destination image can change during the visit to the destination, which means that the post-image is different from the pre-image of the destination.</li> <li>➤ Post-travel image directly influences trip value and tourist satisfaction.</li> <li>➤ Trip value and tourist satisfaction influences revisit and recommend intentions.</li> <li>➤ Satisfaction is influenced negatively by the time tourists stay travelling.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Satisfaction has a significant positive effect on trust and trust and satisfaction have a significant positive effect on loyalty.</li> </ul>

Figure 6 is a resume, containing all the main variables that influence tourist loyalty: perceived performance or perceived value, tourist characteristics, expectations, assessed value, satisfaction and complaints. Therefore, loyalty, which is divided into revisit intention or intention to recommend, is influenced by several factors that have yet to be studied.

**Figure 6: Predictors of loyalty**



Source: TSI (2016).



### 3 The CRISP-DM methodology

The most important data mining methodologies are PMML (Predictive Model Markup Language) developed by Grossman et al. (1999), SEMMA (Sample, Explore, Modify, Model, and Assess) referred by Santos and Azevedo (2005) and CRISP-DM (CRoss-Industry Standard Process for Data Mining) introduced by Chapman et al. (2000). In this study it will be used the CRISP-DM model to analyse data, since this model is business oriented and the other models are not. Below it is presented a description of the CRISP-DM framework.

CRISP-DM is focused on industry. It is easy to understand and helps to manage a data mining project from novices to data mining experts. The data mining process is divided in six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment (Chapman et al., 2000).

The business understanding phase is the definition of the problem in data mining terms, i.e., understand the business and business problems in order to translate these problems into a data mining problem. A plan to achieve the objectives of the project is defined in this phase.

The data understanding phase starts with data gathering. After this, the analyst tries to discover data problems and find out patterns hidden in the data. This phase involves four steps: collection of initial data, description of data, exploration of data and verification of data quality.

At the data preparation phase, data is prepared to go to data mining tools to be treated. A process of selecting, cleaning, construction, integration and formatting is applied to data in this phase.

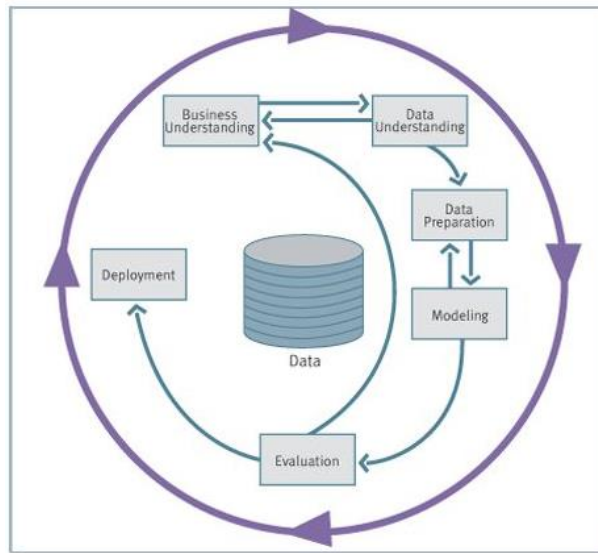
At the modelling phase, the selection and parametrization of modelling techniques is done. Some techniques need some specific data requirements to use data. The process of meeting these requirements are made in this phase, but sometimes stepping back to the preparation phase is necessary to meet these requirements. Therefore, modelling steps are: the selection of the modelling technique, generation of test design, creation of models, and assessment of models.

At the evaluation phase is made an evaluation to the model and its construction, to make sure that every step is correct and that nothing is missing in the model design. When this is guaranteed, the project leader has to decide what to do with the results that will be produced.

At the deployment phase the results are organized in order to the customer can use it. Normally data is given to the customer through a model implemented in the organization's decision-making processes. "The deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise" (Chapman et al., 2000:18). In conclusion, this phase includes plan deployment, plan monitoring and maintenance, the production of the final report, and review of the project.

Figure 7 illustrates the CRISP-DM methodology, with emphasis on the interactivity and constant presence of the manager. This methodology has the advantage of being able to progress and return to earlier stages.

**Figure 7: CRISP-DM phases**



Source: Chapman et al. (2000:13).

### **3.1 Business understanding**

In this phase the aim is to understand the business goals and requirements, reflecting the knowledge to the development of a DM problem, plan and define the criteria for a successful achievement of the objectives (Chapman et al., 2000).

The tourism industry has so many available data from surveys, but only descriptive studies are developed. Gazley and Watling (2015) wrote that tourism marketers should pay attention in the development of promotional messages for products and experiences, ensuring that correct assumptions are made about their customers.

As mentioned in the literature review, the tourist satisfaction with the destination is a determinant factor to their loyalty with the destination. Then, it is extremely important to the tourism companies to have this knowledge and identify their explanatory factors. In addition, the globalization and the increased competitiveness among the destination countries suggest the need to study the tourists from a country, but also from a wide group of countries, having been selected for this study tourists from EU countries.

Therefore, the data studied is from Flash Eurobarometer 414 “Preferences of Europeans towards tourism 2015”. Moreover, only the tourists from EU who travelled at least once in 2014 and stayed abroad for a minimum of four nights were considered in this study, as the tourists that stayed less time did not answer the adequate amount of questions.

Regarding the business goals, the first one is to evaluate the tourist satisfaction with the destination and accommodation services. The second objective is to identify explanatory factors of satisfaction with the destination and accommodation. These two objectives allow managers to better comprehend the travel preferences of the tourists and to develop better marketing campaigns promoting the destination.

In order to identify profiles of European tourists based on their satisfaction with destination and accommodation, a cluster analysis was used as it is an accurate method for generating groups with

different levels of satisfaction. To illustrate the differences between groups (to validate the clusters<sup>1</sup>) decision (classification) trees were used because of the transparency of the algorithm and the interpretability of results (Larose, 2005). This technique allows to identify the most important predictors of the satisfaction and to classify the tourists in one of the clusters created. Moreover, the decision tree should allow to classify 60% of the tourists into the right clusters.

The resources to be used are the IBM SPSS Statistics (v. 23) and the IBM SPSS Modeler (v. 18), where the manipulation and creation of decision trees is easily understood and the outputs are highly satisfactory.

### 3.2 Data understanding

This phase begins with the gathering, description, exploration of data and verification of its quality (Chapman et al., 2000).

As mentioned before, the data used in this study was collected from the Flash Eurobarometer 414 "Preferences of Europeans towards tourism 2015" in the 28 countries of the European Union, conducted by TNS Political & Social at the request of the European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (European Commission, 2015a, b). The interviews were conducted by phone in the respective national language. The analysis already made to this questionnaire is very simple and descriptive and do not explore complex relationships among variables.

The data available consists of 30.101 tourists and 629 variables, which need to be observed and cleaned in order to meet this study goals. For this study, only the tourists from EU who travelled at least once in 2014 for a minimum of four nights are relevant, because only those individuals answered the majority of questions, including the satisfaction question and the study of satisfaction regarding the main holiday in 2014 is the main goal of this study. Then, from the initial data of 30.101 tourists, only 13.853 travelled at least once for a minimum of four nights, which are the ones to be studied. Moreover, despite the EU nationality restriction, the holiday's destination country can be either European (own country or abroad) or non-European.

From all the 630 variables included in the data file retrieved, only the ones that can be related to tourist satisfaction were maintained, resulting in 37 final variables to be studied. These variables were divided into four different subgroups: tourist characterisation (including social demographics characteristics and also historical behaviour as tourist), pre-travel, on-travel and post-travel.

The tourist nationality was distributed among 28 variables, corresponding to the number of countries in the EU. These variables were transformed in one single variable, to facilitate the analysis. Some tourists had multiple citizenship, therefore were not used in the analysis.

Table 11 to Table 19 show the description of each variable, namely ID, nature (qualitative nominal, qualitative ordinal or quantitative, either discrete or continuous), studies that support them and the main goal. Also, the descriptive statistics to all these variables and the number of missing values are presented in those tables.

---

<sup>1</sup> Hypothesis tests (Oneway ANOVA) were also performed to groups' evaluation.

Table 11, Table 12 and Table 13 present the social demographic characteristics (D1 to D8), and also the historical tourist behaviour while tourist, such as, the accommodation and package types that usually are chosen and the number of travels per year (D9 to D13).

**Table 11: Description of the variables of the social demographics characteristics for tourist characterisation**

Variable ID and label		Description	Nature	Literature Review	Goal
<b>Social demographics</b>					
D1	Age	Tourist age	Ordinal	Ragavan et al. (2014); Lu et al. (2015); Sarra et al. (2015); Jarvis et al. (2016)	2
D2	Gender	Tourist gender	Nominal		2
D3	Nationality	Tourist nationality	Nominal		2
D4	Age education	Age when the tourist stop full-time education	Nominal		2
D5	Occupation	Occupation of the tourist	Nominal		2
D6	Household size	People aged 15 years or more who live in the tourist household	Ordinal		2
D7	Phone available	Tourist phone available (phone/mobile)	Nominal		2
D8	Type of community	Community where the tourist lives (e.g. rural area, small or middle size town or large town)	Ordinal		2
<b>Historical tourist behaviour</b>					
Accommodation – considering the personal travels in 2014 how many times the tourist stayed in the following types of accommodation					
D9.1	Paid accommodation with less than 20 guests	Number of times a tourist chose a certain type of accommodation (in 2014)	Ordinal	Nash et al. (2006); Xu & Li (2016)	2
D9.2	Paid accommodation with more than 20 guests		Ordinal		2
D9.3	Own property/second home		Ordinal		2
D9.4	Friends or relatives		Ordinal		2
D9.5	Camping site (tent, motorhome, caravan, etc.)		Ordinal		2
D9.6	Other accommodation		Ordinal		2
Paid accommodation – type of accommodation chosen in 2014					
D10.1	Always in paid accommodation	Type of accommodation chosen by the tourist	Nominal	Nash et al. (2006); Xu & Li (2016)	2
D10.2	Never in paid accommodation		Nominal		2
D10.3	Paid and unpaid accommodation		Nominal		2
D11	Frequency	Number of trips (more than 1 day) made by a tourist in 2014	Ordinal	Barros & Machado (2010); Losada et al. (2016)	2
Duration – type of holiday taken in 2014					
D12.1	Holidays with more than 13 consecutive nights away	Travel duration in 2014	Nominal	Peypoch et al. (2012); Jaafar & Khoshkam (2014); Sarra et al. (2015); Jarvis et al. (2016);	2
D12.2	Holidays between 4 and 13 consecutive nights away		Nominal		2
Package type – type of travel package chosen in 2014					
D13.1	Always 'all-inclusive holidays'	Type of travel package chose by the tourist in 2014	Nominal	Nash et al. (2006); Xu & Li (2016)	2
D13.2	Never 'all-inclusive holidays'		Nominal		2
D13.3	Combination of types of holiday		Nominal		2
D13.4	Don't know		Nominal		2

**Table 12: Descriptive statistics of the social demographics characteristics for tourist characterisation**

Variable ID and label		# Values	Descriptive Statistics				# Missings	Next-phase CRISP-DM
<b>Social demographic</b>								
<b>D1</b>	Age	4	Count		13853			Modelling
			HF	55 years and older	6226	44.9%		
			LF	14 - 24 years	1067	7.7%		
<b>D2</b>	Gender	2	Count		13853			Modelling
			HF	Female	7916	57.1%		
			LF	Male	5937	42.9%		
<b>D3</b>	Nationality	28	Count		12700		1153 (8.3%)	Data preparation
			HF	United Kingdom	926	7.3%		
			LF	Lithuania	144	1.1%		
<b>D4</b>	Age education	4	Count		13853		154 (1.1%)	Modelling
			HF	20 years and older	7109	51.9%		
			LF	No full-time education	64	0.5%		
<b>D5</b>	Occupation	4	Count		13807		46 (0.3%)	Modelling
			HF	Not working	5978	43.2%		
			LF	Manual workers	916	6.6%		
<b>D6</b>	Household size	4	Count		13853	100.0%	63 (0.5%)	Modelling
			HF	2	6702	48.6%		
			LF	4+	2055	14.9%		
<b>D7</b>	Phone available	3	Count		13649		204 (1.5%)	Modelling
			HF	Mobile and landline	10528	77.1%		
			LF	Landline only	563	4.1%		
<b>D8</b>	Type of community	3	Count		13770		83 (0.6%)	Modelling
			HF	Small or middle sized town	5375	39.0%		
			LF	Rural area or village	3783	27.5%		

Note: HF – Highest frequency; LF – Lowest Frequency

**Table 13: Descriptive statistics of the historical tourist behaviour for tourist characterisation**

Variable ID and label		# Values	Descriptive Statistics				# Missings	Next-phase CRISP-DM
<b>Historical tourist behaviour</b>								
Accommodation								
D9.1	Paid accommodation with less than 20 guests	6	Count		13740		113 (0.8%)	Modelling
			HF	None/ zero	7374	53.7%		
			LF	4	380	2.8%		
D9.2	Paid accommodation with more than 20 guests	6	Count		13730		123 (0.9%)	Modelling
			HF	None/ zero	6100	44.4%		
			LF	4	503	3.7%		
D9.3	Own property/second home	6	Count		13753		100 (0.7%)	Modelling
			HF	None/ zero	10053	73.1%		
			LF	4	249	1.8%		
D9.4	Friends or relatives	6	Count		13797		56 (9.4%)	Modelling
			HF	None/ zero	7519	54.5%		
			LF	4	482	3.5%		
D9.5	Camping site (tent, motorhome, caravan, etc.)	6	Count		13816		37 (0.3%)	Modelling
			HF	None/ zero	11907	86.2%		
			LF	4	77	0.6%		
D9.6	Other accommodation	6	Count		13726		127 (0.9%)	Excluded
			HF	None/ zero	11507	83.8%		
			LF	4	102	0.7%		
Paid accommodation								
D10.1	Always in paid accommodation	2	Count	Yes	3943	28.5%		Data Preparation
D10.2	Never in paid accommodation	2	Count	Yes	3192	23.0%		Data Preparation
D10.3	Paid and unpaid accommodation	2	Count	Yes	6714	48.5%		Data Preparation
D11	Frequency	6	Count		13853			Modelling
			HF	4 or 5 times	2733	19.7%		
			LF	More than 10 times	1965	14.2%		
Duration								
D12.1	Holidays with more than 13 consecutive nights away	2	Count	Yes	5002	36.1%		Data Preparation
D12.2	Holidays between 4 and 13 consecutive nights away	2	Count	Yes	11105	80.2%		Data Preparation
Package type								
D13.1	Always 'all-inclusive holidays'	2	Count	Yes	1745	12.6%		Data Preparation
D13.2	Never 'all-inclusive holidays'	2	Count	Yes	9906	71.5%		Data Preparation
D13.3	Combination of types of holiday	2	Count	Yes	2190	15.8%		Data Preparation
D13.4	Don't know package	2	Count	Yes	12	0.1%		Excluded

Note: HF – Highest frequency; LF – Lowest Frequency

Table 14 and Table 15 concern the pre-travel variables and refer to all the moments before the travel. These variables are the motivations to travel (BT1.1 to BT1.10), information (BT2.1 to BT2.10) and organising method (BT3.1 to BT3.9).

**Table 14: Description of the pre-travel variables**

Variable ID and label		Description	Nature	Literature Review	Goal
Motivations to travel - reasons for going on holiday					
<b>BT1</b>	Main reason for going on holiday	Main reason for going on holiday (e.g. sun, wellness, city trips, etc.)	Nominal	Naidoo et al. (2015); Yousefi & Marzuki (2015); Xu & Chan (2016)	2
<b>BT1.1</b>	Sun/Beach		Nominal		2
<b>BT1.2</b>	Wellness/Spa/health treatment		Nominal		2
<b>BT1.3</b>	City trips		Nominal		2
<b>BT1.4</b>	Sport-related activities		Nominal		2
<b>BT1.5</b>	Nature		Nominal		2
<b>BT1.6</b>	Culture		Nominal		2
<b>BT1.7</b>	Visiting family/ friends / relatives		Nominal		2
<b>BT1.8</b>	Specific events (sporting events/festivals/clubbing)		Nominal		2
<b>BT1.9</b>	Other motivation		Nominal		2
<b>BT1.10</b>	Don't know		Nominal		2
Information source – most important information source when making a decision about travel plans					
<b>BT2.1</b>	IS - personal experience	Main information source to make a decision about travel plans	Nominal	Hernández-Méndez et al. (2013); Kim et al. (2015); Llodrà-Riera et al. (2015)	2
<b>BT2.2</b>	IS - recommendations of friends, colleagues or relatives		Nominal		2
<b>BT2.3</b>	IS - paid for guidebooks and magazines		Nominal		2
<b>BT2.4</b>	IS - free catalogues, brochures		Nominal		2
<b>BT2.5</b>	IS - internet websites		Nominal		2
<b>BT2.6</b>	IS - social media sites		Nominal		2
<b>BT2.7</b>	IS - travel agencies / tourism offices		Nominal		2
<b>BT2.8</b>	IS - newspaper, radio, TV		Nominal		2
<b>BT2.9</b>	IS - other		Nominal		2
<b>BT2.10</b>	IS - don't know		Nominal		2
Organising method – searching for information, looking for prices and booking services					
<b>BT3.1</b>	Organising method – internet	Organising method used (searching for information, looking for prices and booking services)	Nominal	Kim et al. (2015); Llodrà-Riera et al. (2015)	2
<b>BT3.2</b>	Organising method - over the phone		Nominal		2
<b>BT3.3</b>	Organising method - by post		Nominal		2
<b>BT3.4</b>	Organising method - over the counter at a travel agency		Nominal		2
<b>BT3.5</b>	Organising method - through someone you know		Nominal		2
<b>BT3.6</b>	Organising method - over the counter of a transportation company		Nominal		2
<b>BT3.7</b>	Organising method - on-site (place of holidays)		Nominal		2
<b>BT3.8</b>	Organising method – other		Nominal		2
<b>BT3.9</b>	Organising method - don't know		Nominal		2

**Table 15: Descriptive statistics to the pre-travel variables**

Variable ID and label		# Values	Descriptive Statistics				# Missings	Next-phase CRISP-DM
Motivations to travel								
BT1	Main reason for going on holiday	9	Count		13777		76 (0.5%)	Modelling
			HF	Sun/beach	3990	29.0%		
			LF	Specific events	369	2.7%		
BT1.1	Sun/Beach	2	Count	Yes	6212	44.8%		Data preparation
BT1.2	Wellness/Spa/health treatment	2	Count	Yes	2140	15.4%		Data preparation
BT1.3	City trips	2	Count	Yes	3408	24.6%		Data preparation
BT1.4	Sport-related activities	2	Count	Yes	1674	12.1%		Data preparation
BT1.5	Nature	2	Count	Yes	4616	33.3%		Data preparation
BT1.6	Culture	2	Count	Yes	3823	27.6%		Data preparation
BT1.7	Visiting family/ friends / relatives	2	Count	Yes	4835	34.9%		Data preparation
BT1.8	Specific events (sporting events/festivals/clubbing)	2	Count	Yes	1040	7.5%		Data preparation
BT1.9	Other motivation	2	Count	Yes	1770	12.8%		Data preparation
BT1.10	Don't know motivation	2	Count	Yes	76	0.5%		Excluded
Information Source								
BT2.1	IS - personal experience	2	Count	Yes	4792	34.6%		Data preparation
BT2.2	IS - recommendations of friends, colleagues or relatives	2	Count	Yes	7647	55.2%		Data preparation
BT2.3	IS - paid for guidebooks and magazines	2	Count	Yes	1063	7.7%		Data preparation
BT2.4	IS - free catalogues, brochures	2	Count	Yes	1757	12.7%		Data preparation
BT2.5	IS - internet websites	2	Count	Yes	7564	54.6%		Data preparation
BT2.6	IS - social media sites	2	Count	Yes	1110	8.0%		Data preparation
BT2.7	IS - travel agencies / tourism offices	2	Count	Yes	2467	17.8%		Data preparation
BT2.8	IS - newspaper, radio, TV	2	Count	Yes	970	7.0%		Data preparation
BT2.9	IS - other	2	Count	Yes	233	1.7%		Data preparation
BT2.10	IS - don't know	2	Count	Yes	192	1.4%		Excluded
Organising method								
BT3.1	Organising method – internet	2	Count	Yes	8498	61.3%		Data preparation
BT3.2	Organising method - over the phone	2	Count	Yes	2098	15.1%		Data preparation
BT3.3	Organising method - by post	2	Count	Yes	302	2.2%		Data preparation
BT3.4	Organising method - over the counter at a travel agency	2	Count	Yes	2722	19.6%		Data preparation
BT3.5	Organising method - through someone you know	2	Count	Yes	2887	20.8%		Data preparation
BT3.6	Organising method - over the counter of a transportation company	2	Count	Yes	940	6.8%		Data preparation
BT3.7	Organising method - on-site (place of holidays)	2	Count	Yes	1566	11.3%		Data preparation
BT3.8	Organising method – other	2	Count	Yes	940	6.8%		Data preparation
BT3.9	Organising method - don't know	2	Count	Yes	577	4.2%		Excluded

Note: HF – Highest frequency; LF – Lowest Frequency



Table 16 and Table 17 are related to the on-travel variables. These variables concern the destination country in the last holidays (OT1.1 and OT1.2) and the party experiences while using paid accommodation (OT2.1 to OT2.8).

**Table 16: Description of the on-travel variables**

Variable ID and label		Description	Nature	Literature Review	Goal
Destination country in last holidays					
OT1.1	Main holiday destination country	Destination country for the main holiday in 2014. It is divided into "Own country", when the destination country is in the own country or abroad in another country from the EU.	Nominal	Akhoondnejad (2015); Thrane (2016)	2
OT1.2	Own country as destination country	The destination country is the own country (not abroad).	Nominal		2
Party experience – type of party experience while using paid accommodation during the main holiday in 2014					
OT2.1	Slip/Trip/Fall within the accommodation (with serious consequences)	Type of party experience while using paid accommodation in 2014	Nominal	Sari et al. (2013)	1, 2
OT2.2	Food poisoning or food-related sickness		Nominal		1, 2
OT2.3	Swimming pool incident (drowning/near drowning)		Nominal		1, 2
OT2.4	Fire-related emergency situation		Nominal		1, 2
OT2.5	Balcony fall or near fall		Nominal		1, 2
OT2.6	Incident involving glass doors or windows		Nominal		1, 2
OT2.7	Other safety incident		Nominal		1, 2
OT2.8	No safety incident		Nominal		1, 2
OT2.9	Don't know safety incident		Nominal		1, 2

**Table 17: Descriptive statistics to the on-travel variables**

Variable ID and label		# Values	Descriptive Statistics				# Missings	Next-phase CRISP-DM
Destination country in last holidays								
OT1.1	Main holiday destination country	33	Count		12664		1189 (8.6%)	Data preparation
			HF	In own country	4935	35.6%		
			LF	Lithuania	20	0.1%		
OT1.2	Own country as destination country	32	Count		13774		79 (0.6%)	Data preparation
			HF	Spain	1538	11.2%		
			LF	Luxembourg	13	0.1%		
Party experience safety issues while using paid accommodation								
OT2.1	Slip/Trip/Fall within the accommodation (with serious consequences)	2	Count	Yes	154	1.4%		Modelling
OT2.2	Food poisoning or food-related sickness	2	Count	Yes	205	1.9%		Modelling
OT2.3	Swimming pool incident (drowning/near drowning)	2	Count	Yes	55	0.5%		Modelling
OT2.4	Fire-related emergency situation	2	Count	Yes	43	0.4%		Modelling
OT2.5	Balcony fall or near fall	2	Count	Yes	19	0.2%		Modelling
OT2.6	Incident involving glass doors or windows	2	Count	Yes	55	0.5%		Modelling
OT2.7	Other safety incident	2	Count	Yes	190	1.8%		Data preparation
OT2.8	No safety incident	2	Count	Yes	9841	92.3%		Excluded
OT2.9	Don't know safety incident	2	Count	Yes	154	1.40%		Excluded

Note: HF – Highest frequency; LF – Lowest Frequency

Table 18 and Table 19 present the post-travel variables. They consist of variables to measure the satisfaction with the destination (PT1.1 to PT1.7), the existence of formal complaints (PT3.1) and the services to which a formal complaint was made (PT3.1 to PT3.7).

**Table 18: Description of the post-travel variables**

Variable ID and label		Description	Nature	Literature Review	Goal
Satisfaction – satisfaction with the main holiday in 2014					
<b>PT1.1</b>	Satisfaction with the quality of the accommodation	Satisfaction with several aspects at the destination for the main holiday in 2014	Ordinal	Chi & Qu (2009); Chen et al. (2011); Araslı & Baradarani (2014); Ragavan et al. (2014); Vajčnerová et al. (2014); Sarra et al. (2015)	1
<b>PT1.2</b>	Satisfaction with the safety of the accommodation		Ordinal		1
<b>PT1.3</b>	Satisfaction with the natural features		Ordinal		1
<b>PT1.4</b>	Satisfaction with the general level of prices		Ordinal		1
<b>PT1.5</b>	Satisfaction with how tourists are welcomed		Ordinal		1
<b>PT1.6</b>	Satisfaction with the quality of activities/services available		Ordinal		1
<b>PT1.7</b>	Satisfaction with the accessible facilities for people with special needs		Ordinal		1
Formal complaint – existence or not of a formal complaint about any service during the main holiday in 2014					
<b>PT2</b>	Formal complaint	Existence of formal complaint at the destination with any service during the main holiday in 2014	Nominal	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	1
Services complaints - service to which the tourist made a complaint during the main holiday in 2014					
<b>PT3.1</b>	Accommodation	Service to which the tourist made a complaint	Nominal	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	1 and 2
<b>PT3.2</b>	Transport		Nominal		1 and 2
<b>PT3.3</b>	Restaurants		Nominal		1 and 2
<b>PT3.4</b>	Leisure activities		Nominal		1 and 2
<b>PT3.5</b>	Other		Nominal		1 and 2
<b>PT3.6</b>	None		Nominal		1 and 2
<b>PT3.7</b>	Don't know		Nominal		1 and 2

**Table 19: Descriptive statistics to the post-travel variables**

Variable ID and label		# Values	Descriptive Statistics				# Missings	Next-phase CRISP-DM
Satisfaction								
PT1.1	Satisfaction with the quality of the accommodation	4	Count		13473		380 (2.7%)	Data preparation
			HF	Very satisfied	8551	63.5%		
			LF	Not at all satisfied	84	0.6%		
PT1.2	Satisfaction with the safety of the accommodation	4	Count		13404		449 (3.2%)	Data preparation
			HF	Very satisfied	9383	70.0%		
			LF	Not at all satisfied	67	0.5%		
PT1.3	Satisfaction with the natural features	4	Count		13579		274 (2.0%)	Data preparation
			HF	Very satisfied	9844	72.5%		
			LF	Not at all satisfied	97	0.7%		
PT1.4	Satisfaction with the general level of prices	4	Count		13268		585 (4.2%)	Data preparation
			HF	Fairly satisfied	6758	48.8%		
			LF	Not at all satisfied	178	1.3%		
PT1.5	Satisfaction with how tourists are welcomed	4	Count		12611		1242 (9.0%)	Data preparation
			HF	Very satisfied	7255	57.5%		
			LF	Not at all satisfied	130	1.0%		
PT1.6	Satisfaction with the quality of activities/services available	4	Count		12969		884 (6.4%)	Data preparation
			HF	Very satisfied	6835	52.7%		
			LF	Not at all satisfied	124	1.0%		
PT1.7	Satisfaction with the accessible facilities for people with special needs	4	Count		7937		5916 (42.7%)	Excluded
			HF	Fairly satisfied	3441	43.4%		
			LF	Not at all satisfied	414	5.2%		
Formal complaint								
PT2	Formal complaint	2	Count		13757		96 (0.7%)	Data preparation
			HF	No	13169	95.7%		
			LF	Yes	588	4.3%		
Services complaints								
PT3.1	Accommodation	2	Count	Yes	264	44.9%		Data preparation
PT3.2	Transport	2	Count	Yes	140	23.8%		Data preparation
PT3.3	Restaurants	2	Count	Yes	72	12.2%		Data preparation
PT3.4	Leisure activities	2	Count	Yes	39	6.6%		Data preparation
PT3.5	Other formal complaint	2	Count	Yes	124	21.1%		Data preparation
PT3.6	No formal complaint	2	Count	Yes	3	0.5%		Excluded
PT3.7	Don't know formal complaint	2	Count	Yes	2	0.3%		Excluded

Note: HF – Highest frequency; LF – Lowest Frequency

Some variables from the previous tables were excluded due to the low level of interest to the study, such as the variables where the answer is “don't know”. These variables do not allow to conclude or relate with the tourist satisfaction, because of their ambiguous meaning. These variables are:

- D13.4: Don't know package;
- BT2.10: IS – Don't know;
- BT3.9: Organising method – Don't know;
- OT2.8: No safety incident;
- OT2.9: Don't know safety incident;
- PT3.6: No formal complaint
- PT3.7: Don't know formal complaint.

Moreover, variable PT1.7 – “Satisfaction with the accessible facilities for people with special needs” was excluded due to the high level of missing values (42.7%).

### 3.3 Data preparation

Data preparation phase consists in the data selection, cleaning and formatting for data modelling (Chapman et al., 2000).

#### *Recodification of the existing variables*

Table 20 and Table 21 are related to the tourist characterization and social demographic new variables.

The variable D3 (nationality) was used to create a new variable D3\_1 (area of the European Union that the country of nationality belongs) to split the nationality countries into the European demographic divisions and also to create the variable D3\_2 (currency of the nationality country) to assign the value “Euro” or “Non-euro” to each country.

The country division was made using the following characterization (Simielli, 2012):

- North European countries: Denmark, Finland and Sweden.
- Central European countries: Austria, Belgium, France, Germany, Ireland, Italy, Luxembourg, Malta, Netherlands and United Kingdom.
- Countries of the Iberian Peninsula: Portugal and Spain.
- Eastern European countries: Czech Republic, Hungary, Poland, Rep. of Cyprus and Romania.
- Balkan countries: Bulgaria, Greece, Latvia, Republic of Croatia, Slovakia and Slovenia.
- Baltics: Estonia and Lithuania.

The creation of the variable D3\_2 was made using the following information (UE, 2016):

- Euro: Germany, Austria, Belgium, Cyprus, Slovakia, Slovenia, Spain, Estonia, Finland, France, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands and Portugal.
- Non euro: Bulgaria, Croatia, Hungary, Poland, Czech Republic, Romania, Sweden, Denmark and United Kingdom.

The variables D10.1\_3 (type of accommodation (paid or unpaid) in personal travels), D12.1\_2 (duration of personal travels taken) and D13.1\_3 (type of holiday (package)) were created to reduce the number of variables to use in the modelling phase. These variables replace the mutually exclusive variables D10 (D10.1, D10.2 and D10.3), D12 (D12.1 and D12.2) and D13 (D13.1, D13.2 and D13.3), respectively and will be excluded, due to the creation of the new variables, which contain all the information.

**Table 20: Description of the variables created of the social demographics characteristics for tourist characterisation**

Variable ID and label		Description	Nature	Literature Review	Goal
D3_1	Area of the European Union that the country of nationality belongs	Country area of the tourist nationality (e.g. Baltics, Central European countries)	Nominal	Ragavan et al. (2014); Lu et al. (2015); Sarra et al. (2015); Jarvis et al. (2016)	2
D3_2	Currency of the nationality country	Currency of the tourist nationality country (euro or non-euro)	Nominal	Nash et al. (2006); Ragavan et al. (2014); Vajčnerová et al. (2014); Xu & Li (2016)	2
D10.1_3	Type of accommodation (paid or unpaid)	Type of accommodation chosen by the tourist (e.g. paid, unpaid or both)	Nominal	Sarra et al. (2015); Jarvis et al. (2016)	2
D12.1_2	Duration of personal travels taken in 2014	Nights away that a tourist stayed when travelling (e.g. more than 13 consecutive nights away, between 4 and 13 consecutive nights away) in 2014	Nominal	Thrane (2016)	2
D13.1_3	Type of holiday (package) in 2014	Type of holiday package (e.g. always all-inclusive, never all-inclusive, a combination of both) experienced by the tourist in 2014	Nominal		2

**Table 21: Descriptive statistics to the variables created of the social demographics characteristics for tourist characterisation**

Variable ID and label		# Values	Descriptive Statistics				# Missings
D3_1	Area of the European Union that the country of nationality belongs	6	Count		12700		1153 (8.3%)
			HF	Central European countries	6178	48.6%	
			LF	Baltics	326	2.6%	
D3_2	Currency of the nationality country	2	Count		12700		1153 (8.3%)
			HF	Euro	8371	65.9%	
			LF	Non euro	4329	34.1%	
D10.1_3	Type of holiday (package) in 2014	3	Count		13841		12 (0.1%)
			HF	Never 'all-inclusive holidays'	9906	71.6%	
			LF	Always 'all-inclusive holidays'	1745	12.6%	
D12.1_2	Duration of personal travels taken in 2014	3	Count		13853		
			HF	Holidays between 4 and 13 consecutive nights away	8851	63.9%	
			LF	Both types of duration	2254	16.3%	
D13.1_3	Type of accommodation (paid or unpaid)	3	Count		13849		4 (0.03%)
			HF	Paid and unpaid accommodation	6714	48.5%	
			LF	Always in paid accommodation	3943	28.5%	

Note: HF – Highest frequency; LF – Lowest Frequency

Table 22 and Table 23 contain the description and descriptive statistics of the pre-travel new variables. These are related to the number of motives to travel (BT1.1\_8), number of information sources used (BT2.1\_8) and number of booking sources (BT3.1\_7).

The variable BT1.1\_9 (number of motives to travel with the number of travel motivations) answered by each tourist was created using the variables BT1.1 to BT1.9 related to the travel motivations.

The variables BT2.1 to BT2.9 related to information sources and BT3.1 to BT3.8 related to the organising method were also used to define two new variables: BT2.1\_9 (number of information sources used by each tourist) and BT3.1\_8 (number of booking sources used by each tourist), respectively.

Variable BT4 (the nationality and destination currency are the same) was created from the variables D3 and OT1.2. This new variable is presented as a motivation to travel.

Moreover, the variables BT1\_MT, BT1.1\_9\_MT, BT2.1\_9\_MT and BT3.1\_8\_MT were created, assuming the same values of the original variables (BT1, BT1.1\_9, BT2.1\_9 and BT3.1\_8), but just for the cases when the tourist had only one trip. When the tourist had travel more than one time, the variables assume the value of “Unknown”, because of the uncertainty of the tourist behaviour related to the main trip.

**Table 22: Description of the pre-travel variables created**

Variable ID and label		Description	Nature	Literature Review	Goal
<b>BT1_MT</b>	Main reason for going on main trip	Main reason that made each tourist chose a destination (e.g. sun/beach, wellness, city trips, sport-related activities, nature, etc.)	Nominal	Naidoo et al. (2015); Yousefi & Marzuki (2015); Xu & Chan (2016)	2
<b>BT1.1_9</b>	Number of reasons to travel	Number of reasons to travel chosen by each tourist	Discrete quantitative	Naidoo et al. (2015); Yousefi & Marzuki (2015); Xu & Chan (2016)	2
<b>BT1.1_9_MT</b>	Number of reasons to main trip	Number of reasons used by each tourist to choose a destination	Nominal	Naidoo et al. (2015); Yousefi & Marzuki (2015); Xu & Chan (2016)	2
<b>BT2.1_9</b>	Number of information sources	Number of information sources chosen by each tourist to make a decision about travel plans	Discrete quantitative	Hernández-Méndez et al. (2013); Kim et al. (2015); Llodrà-Riera et al. (2015)	2
<b>BT2.1_9_MT</b>	Number of information sources main trip	Number of information sources chosen by each tourist to make a decision about travel plans to the main trip	Nominal	Hernández-Méndez et al. (2013); Kim et al. (2015); Llodrà-Riera et al. (2015)	2
<b>BT3.1_8</b>	Number of organising methods	Number of organising methods chosen by each tourist to organise holidays (searching, looking for prices and booking)	Discrete quantitative	Kim et al. (2015); Llodrà-Riera et al. (2015)	2
<b>BT3.1.8_MT</b>	Number of organising methods used for main trip	Number of organising methods chosen by each tourist to organise holidays (searching, looking for prices and booking) to the main trip	Nominal	Kim et al. (2015); Llodrà-Riera et al. (2015)	2
<b>BT4</b>	The nationality and destination currency are the same	The nationality and destination currency are analysed to define if they have the same or different currencies	Nominal	Naidoo et al. (2015); Yousefi & Marzuki (2015); Xu & Chan (2016)	2

**Table 23: Descriptive statistics to the variables created for the pre-travel phase**

Variable ID and label		# Values	Descriptive Statistics				# Missings
<b>BT1_MT</b>	Main reason for going on main trip	10	Count		13777		76 (0.5%)
			HF	Sun/beach	806	5.9%	
			LF	Sports-related activities	54	0.3%	
			Frequency	Unknown	11143	80.9%	
<b>BT1.1_9</b>	Number of reasons to travel with the number of travel motivations answered by each tourist	5	Count		13853		
			Mean		2,00		
			Std. Deviation		0,946		
			Mode		2		
<b>BT1.1_9_MT</b>	Number of reasons to main trip	9	Count		13853		
			HF	2	1270	9.2%	
			LF	0	22	0.2%	
			Frequency	Unknown	11197	80.8%	
<b>BT2.1_9</b>	Number of information sources used by each tourist	4	Count		13853		
			Mean		1,98		
			Std. Deviation		0,858		
			Mode		2		
<b>BT2.1_9_MT</b>	Number of information sources main trip	4	Count		13853		
			HF	1	1161	8.4%	
			LF	0	58	0.4%	
			Frequency	Unknown	11197	80.8%	
<b>BT3.1_8</b>	Number of organising methods used by each tourist	8	Count		13853		
			Mean		1,44		
			Std. Deviation		0,846		
			Mode		1		
<b>BT3.1.8_MT</b>	Number of organising methods used for main trip	8	Count		13853		
			HF	1	1960	14.1%	
			LF	0	58	0.04%	
<b>BT4</b>	The nationality and destination currency are the same	2	Count		11709		2144 (15.5%)
			HF	Same currency	8571	73.2%	
			LF	Different currency	3138	26.8%	

Note: HF – Highest frequency; LF – Lowest Frequency

Table 24 and Table 25 consist of the description and descriptive statistics of the on-travel new variables. These are about the area of the EU that the destination country belongs (OT1.1\_1), currency of the destination country (OT1.1\_2), party experience any safety issues while using paid accommodation (OT2), number safety issues while using paid accommodation (OT2.1\_7), type of accommodation (paid or unpaid) (OT3), duration (OT4) and the type of holiday package (OT5).

A similar process to the creation of the variables D3.1\_1 and D3.1\_2 was used to create the variables OT1.1\_1 (area of the European Union that the destination country belongs) and OT1.1\_2 (currency of the destination country). These variables were created from the variable OT1.1 (main holiday destination country).

The variable OT1.1\_1 was created through the attribution of a destination country to a European area. The variable OT1.1\_2 was created from the association of a country to a currency. When the destination country was “own country”, there was the need to use the variable D3 (nationality) to find the nationality and the destination country.

The variable OT2 (party experience any safety issues while using paid accommodation) was created from the variables OT2.1 to OT2.7 to verify the tourists that experienced a new party experience and the tourists who did not.

The variable OT2.1\_7 (number of party experiences any safety issues) was created from variables OT2.1 to OT2.7 in order to count the number of party experiences and safety issues of each tourist while using paid accommodation during their main holiday in 2014. The phases to create this variable were: i) count the number of safety issues of each tourist (the answer can be zero – no answer or seven – all answers); and ii) the tourists with zero answers were defined as “No” safety issue and the others were defined as “Yes”.

**Table 24: Description of the variables created for the on-travel phase**

Variable ID and label	Description	Nature	Literature Review	Goal	
OT1.1_1	Area of the European Union that the destination country belongs	Country area of the tourist destination (e.g. Baltics, Central European countries)	Nominal	Ragavan et al. (2014); Lu et al. (2015); Sarra et al. (2015); Jarvis et al. (2016)	2
OT1.1_2	Currency of the destination country	Currency of the tourist destination country (euro or non-euro)	Nominal	Ragavan et al. (2014); Lu et al. (2015); Sarra et al. (2015); Jarvis et al. (2016)	2
OT2	Party experience any safety issues while using paid accommodation	Party experience any safety issues while using paid accommodation	Nominal	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	2
OT2.1_7	Number of party experiences any safety issues	Number of safety issues while using paid accommodation during their main holiday in 2014	Discrete quantitative	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	2
OT3	Type of accommodation main trip	Type of accommodation package (paid, unpaid or both) used in the main holiday in 2014	Nominal	Nash et al., (2007); Ragavan et al. (2014); Vajčnerová et al. (2014); Xu & Li (2015)	2
OT4	Duration of the main trip	Nights away that a tourist stayed when travelling (e.g. more than 13 consecutive nights away, between 4 and 13 consecutive nights away) during his main trip in 2014.	Nominal	Sarra et al. (2015); Jarvis et al. (2016)	2
OT5	Type of holiday (package) main trip	Type of holiday (all-inclusive, never all-inclusive or a combination) used in the main holiday in 2014	Nominal	Thrane (2016)	2

**Table 25: Descriptive statistics to the variables created for the on-travel phase**

Variable ID and label	# Values	Descriptive Statistics			# Missings
OT1.1_1	8	Count		12664	1189 (8.6%)
		HF	Own country	4935 39.0%	
		LF	Baltics	54 0.4%	
OT1.1_2	2	Count		12280	1573 (11.4%)
		HF	Euro	7604 61.9%	
		LF	Non euro	4676 38.1%	
OT2	2	Count		13853	
		HF	No	13191 95.2%	
		LF	Yes	662 4.8%	
OT2.1_7	5	Count		13853	
		Mean		0.05	
		Std. Deviation		0.224	
		Mode		0	
OT_3	4	Count		13849	4 (0.03%)
		HF	Paid accommodation	3943 28.5%	
		LF	Paid and unpaid accommodation	545 3.9%	
		Frequency	Unknown	6169 44.5%	
OT4	3	Count		13795	58 (0.4%)
		HF	Holidays between 4 and 13 consecutive nights away	1797 13.0%	
		LF	Holidays with more than 13 consecutive nights away	801 5.8%	
		Frequency	Unknown	11197 81.2%	
OT5	4	Count		13841	12 (0.1%)
		HF	Not 'all-inclusive holidays'	1923 13.9%	
		LF	Combination of types of holiday (package)	157 1.1%	
		Frequency	Unknown type of holiday (package)	11188 80.8%	

Note: HF – Highest frequency; LF – Lowest Frequency



Table 26 and Table 27 are the description and the descriptive statistics of the created variables to the post-travel behaviours of tourists. At the variables related to satisfaction, PT1.1 to PT1.6, the scale was redefined, in order to define a growing satisfaction scale from 1 (not very satisfied) to 4 (very satisfied).

The variables PT3.1 to PT3.5 related to complaints were used to define a new variable PT3.1\_5 (number of complaints made). The creation of this variable was through the count of the number of complaints made to each service (accommodation, transport, restaurants, leisure activities and others).

In addition, variables PT2\_MT and PT3.1\_5\_MT were created to the main trip complaints. In the cases that the tourist had only one trip, the variables' values stayed the same as the original variable and if the tourist did more than one travel, the variable was defined as "Unknown".

**Table 26: Description for the variables created for tourist the post-travel phase**

Variable ID and label		Description	Nature	Literature Review	Goal
<b>PT2_MT</b>	Formal complaint main trip	Existence of any formal complaint with any service during the main holiday in 2014 by each tourist	Nominal	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	2
<b>PT3.1_5</b>	Number of complaints made	Number of complaints made to the holidays in 2014 by each tourist	Discrete quantitative	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	2
<b>PT3.1_5_MT</b>	Number of complaints made main trip	Number of complaints made to the main holiday in 2014 by each tourist	Nominal	Lam & Tang (2003); Ngai et al. (2007); Sparks & Browning (2010); Chen et al. (2011); Sari et al. (2013); Araslı & Baradarani (2014); Vajčnerová et al. (2014); Zhan et al. (2015)	2

**Table 27: Descriptive statistics to the variables created for the post-travel phase**

Variable ID and label		# Values	Descriptive Statistics		
<b>PT2_MT</b>	Formal complaint main trip	3	Count		13853
			HF	No	2576 18.6%
			LF	Yes	80 0.6%
			Frequency	Unknown	11197 80.8%
<b>PT3.1_5</b>	Number of complaints made	5	Count		13853
			Mean		0.05
			Std. Deviation		0.23
			Mode		0
<b>PT3.1_5_MT</b>	Number of complaints made main trip	4	Count		13853
			HF	0	2576 18.6%
			LF	2	6 0.04%
			Frequency	Unknown	11197 80.8%

Note: HF – Highest frequency; LF – Lowest Frequency

## *Principal Component Analysis: Dimension reduction*

To accomplish Objective 1, a PCA analysis was performed to reduce the dimensionality. To resolve this problem of high dimensionality, the PCA method defines new attributes to the data (components). The aim of PCA is to define the more suitable variables to the data. The reduction of the dimensionality is only possible when the variables are correlated and when is defined a component with these correlated variables (Dunteman, 1989; Luukka, 2009).

The suitability of the data size was validated using both KMO (Kaiser–Meyer–Olkin Measure of Sampling Adequacy) Index and Bartlett's Test of Sphericity.

The KMO index represents the ratio of the squared correlation between variables to the squared partial correlation between variables. The values of KMO range between 0 and 1. Values close to 1 indicates a high correlation, and therefore the analysis will result in distinct and reliable components. The sample of the data set in this study is proper for use in factor analysis, since the KMO value is 0.789 which indicates a high correlation (Kaiser, 1974).

Bartlett's Test of Sphericity evaluates whether the population correlation matrix resembles an identity matrix and it should be significant at  $p \leq 0.05$  (or  $p \leq 0.1$ ). If there is an identity matrix, every variable has a low correlation with the other variables, which means they have a higher independence level between them (Field, 2005). Bartlett's test demonstrated that it is significant,  $p < 0.001$  (Table 28). Therefore, this result indicated the existence of correlations between the variables and also corroborates the adequacy of the data to the technique.

The PCA was adopted in order to identify and validate the underlying dimensions of the tourist satisfaction (PT1.1 to PT1.7)<sup>2</sup>. PCA decomposes data into a set of linear components within the data and indicates how much a variable contributes to that component. To begin, a matrix representing the relationships between the six satisfaction variables is employed (Hair et al., 1998; Luukka, 2009).

The variable "Satisfaction with the facilities available" was excluded from this analysis, because only 7.219 tourists answered to this question, which would reduce the sample to almost half of the respondents. In this way, to begin, a matrix representing the relationships between the six satisfaction variables is employed (Luukka, 2009).

The PCA solution identified three components of satisfaction after a Varimax rotation. The components explain 83% of the total variance and have easy interpretation: satisfaction with the accommodation, satisfaction with the reception, satisfaction with the natural features and satisfaction with the general level of prices (Table 28).

In order to operationalize the components, two indexes were created through the average of the two variables with higher loadings ( $> 0.5$ ) to each of the first two components. The internal consistency of these indexes was confirmed through the Cronbach's Alpha, showing values above 0.6 as can be observed in Table 28. As components 3 and 4 include just one variable, we used the two

---

<sup>2</sup> PCA was also applied to other groups of variables, namely motivation to travel (BT1.1 to BT1.9), information Sources (BT2.1 to BT2.9) and booking sources (BT3.1 to BT3.8). Although, only with satisfaction the results were satisfactory, due to lack of correlations among pre-travel variables.

original variables corresponding to these components. The summary statistics of the 4 variables used for the tourist satisfaction assessment are presented in Table 29.

**Table 28: Varimax rotated component matrix and communalities**

Satisfaction at the destination	Communalities	Components and loadings			
		Satisfaction with the accommodation	Satisfaction with the reception	Satisfaction with the natural features	Satisfaction with the general level of prices
Satisfaction with the quality of the accommodation	0.756	0.826			
Satisfaction with the safety of the accommodation	0.783	0.859			
Satisfaction with the way tourists are welcomed	0.711		0.797		
Satisfaction with the quality of activities	0.744		0.830		
Satisfaction with the natural features	0.999			0.973	
Satisfaction with the general level of prices	0.995				0.968
% total variance explained after rotation		25.58%	24.25%	16.66%	16.64%
Initial Eigenvalue		2.594	0.861	0.796	0.737
Cronbach's Alpha		0.703	0.622		

Rotation Method: Varimax with Kaiser Normalization.  
 KMO=0,789;  
 Bartlett(15)=12849,168;p<0,001;  
 n=11697

**Table 29: Statistics of the tourist satisfaction**

Dimensions of satisfaction <sup>(1)</sup>	Mean <sup>(2)</sup>	Standard Deviation	Minimum	Quartile 1	Median	Quartile 3	Maximum
<b>Satisfaction with accommodation</b>	<b>3.6</b>	<b>0.5</b>	<b>1.0</b>	<b>3.5</b>	<b>4.0</b>	<b>4.0</b>	<b>4.0</b>
Quality of the accommodation	3.6	0.6	1.0	3.0	4.0	4.0	4.0
Safety of the accommodation	3.7	0.5	1.0	3.0	4.0	4.0	4.0
<b>Satisfaction with the reception at the destination</b>	<b>3.5</b>	<b>0.5</b>	<b>1.0</b>	<b>3.0</b>	<b>3.5</b>	<b>4.0</b>	<b>4.0</b>
How tourists are welcomed (e.g. services for children, customer care, pets-welcomed)	3.5	0.6	1.0	3.0	4.0	4.0	4.0
Quality of activities/services available (e.g. transport, restaurants, leisure activities)	3.5	0.6	1.0	3.0	4.0	4.0	4.0
<b>Satisfaction with natural features (e.g. landscape , weather conditions)</b>	<b>3.7</b>	<b>0.6</b>	<b>1.0</b>	<b>3.0</b>	<b>4.0</b>	<b>4.0</b>	<b>4.0</b>
<b>Satisfaction with general level of prices</b>	<b>3.3</b>	<b>0.7</b>	<b>1.0</b>	<b>3.0</b>	<b>3.0</b>	<b>4.0</b>	<b>4.0</b>

Notes: (1) scale from 1 (not at all satisfied) to 4 (very satisfied);  
 (2) number of cases: n=11697

## 3.4 Modelling

Modelling is the selection and parametrization of modelling techniques, such as the cluster analysis and the decision tree algorithms (Chapman et al., 2000). In this study, it was used the classification approach, with clusters analysis and different decision trees algorithms.

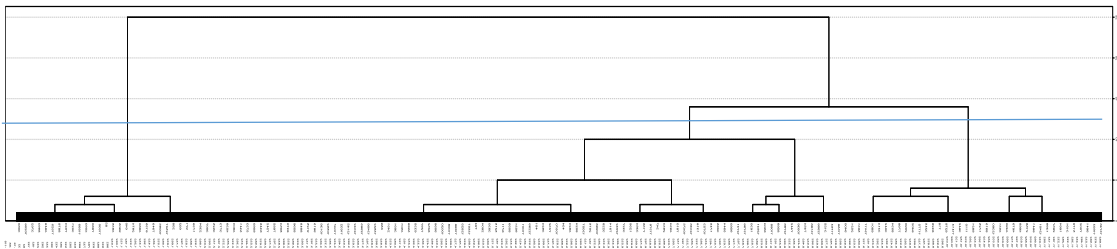
### 3.4.1 Cluster Analysis: Tourist satisfaction assessment

Regarding the Objective 1, a cluster analysis was performed in order to group the tourists according to their level of satisfaction. Two different methods were used: the hierarchical agglomerative and k-means. The hierarchical agglomerative method defines each entity as a group and these groups are then merged until there is only a group with all the entities. The clusters construction in this method can be visualized with a dendrogram.

One of the hierarchical agglomerative algorithms most used is the Ward method (Ward, 1963). It merges the two clusters that minimize the within-cluster variance at each iteration. This variance is measured as a weighted sum of squares, taking into account the cardinality of each cluster (Amorim et al., 2016).

By using hierarchical methods with Squared Euclidean distance and the Ward's method was possible to conclude from the dendrogram that one of the possible solutions was a number of clusters of 3 (Figure 8)<sup>3</sup>. Then, the 3 groups were defined using the k-means method (MacQueen, 1967), but also a 2 and 4 clusters solutions, also suggested by the dendrogram, were analysed.

**Figure 8: Satisfaction assessment: dendrogram from the hierarchical cluster analysis**



The k-means method is a partition algorithm where each entity is assigned to a different group until the k clusters are defined according to their similarity. This method minimizes the clustering error and does not depend on any initial conditions or adjustable parameters, which are significant advantages over all other clustering algorithms (MacQueen, 1967; Likas et al., 2003; Amorim et al., 2016).

The result of k-means method was 815 tourists in cluster 1 (not very satisfied), 6286 tourists in cluster 2 (very satisfied) and 4596 tourists in cluster 3 (satisfied). The clusters number was re-defined to order the groups by satisfaction level, which means the cluster 1 corresponds to the lower level of satisfaction and the cluster 3 to the higher level of satisfaction.

<sup>3</sup> Other hierarchical methods such as the furthest neighbour, between-groups linkage, nearest neighbour and centroid clustering were performed and the results were all similar. These methods were also performed using the standardized variables, but the differences were not relevant.

An F test (Oneway ANOVA) was performed to the clusters validation, in order to examine differences between them. Although, due to the violation of the assumption of the homogeneity of variances (according to the Levene test), it was used the Welch test (a robust test of equality of means) and the Scheffé post-hoc test to identify significant differences between each combination of clusters. In addition, the Eta<sup>2</sup> (Cohen's, 1988) values were computed. The Eta<sup>2</sup> is the measure of how much an independent variable (satisfaction dimension) has affected the dependent variable (cluster satisfaction level).

The significance of the relationship between the clusters and the variables related to tourist social-demographics (D1 to D8), tourist behaviour characterisation (D9 to D13) and the tourist complaints (PT2 and PT3.1\_5) was measured through the chi-squared test of independence and to complement an association measure (Cramer's V or Eta) was computed (Acock & Stavig, 1979). These methods were performed to determine the strength of the relationship between the cluster and the variables. In all the cases the independence between the variables were rejected (Acock & Stavig, 1979; Valle et al., 2012).

The same procedures were used to analyse the relationship between the formal complaints and the variables related to tourist social-demographics (D1 to D8) and tourist behaviour characterisation (D9 to D13).

### **3.4.2 Classification Analysis: Predictors of tourist satisfaction**

To accomplish Objective 2, decision tree algorithms were applied using the clusters defined in Objective 1 as dependent variable (target variable). A decision tree is an algorithm that tests a condition in every step of an analysis. "A Decision tree is a tree in which each branch node represents a choice between a number of alternatives and each leaf node represents a classification or decision" (Ayoubloo et al., 2011:10115).

Decision trees are easy to understand, require little data preparation, allow numerical and categorical data, are highly efficient with a large amount of data and perform a great way of results visualization and their relationships. There are several algorithms to build decision trees: Classification and Regression Tree (CART), Chi-squared Automatic Interaction Detector (CHAID), C5.0, among others (Delen et al., 2013).

CART (Breiman et al., 1984) is a binary decision tree that can predict continuous dependent variables (regression analysis) and categorical predictor variables (classification analysis) (Brezigar-Masten & Masten, 2012). This algorithm starts by splitting the initial data into two subsets and then these two subsets are partitioned. A subset is always more homogeneous than the previous ones. This process is applied to each subset until it is not possible to make more partitions. The subset will be partitioned only if it can produce the greatest accuracy improvement (Li et al., 2010; Delen et al., 2013). The criteria used to reduce the impurity in splitting for classification was the Gini Index, which is the recommended by Breiman et al. (1984).

This algorithm is easily understood by human beings (Li et al., 2010), is capable of "modelling complex relationships between independent and dependent features in the task without strong model assumptions" (Li et al., 2010:5896), handling missing values by "using surrogate splitting to make the

best use of the data” (Delen et al., 2013:3976) and “identifying significant independent features by itself”(Li et al., 2010:5896).

The CHAID (Kass, 1980) is a decision tree technique based on adjusted significance testing (Chi-squared test of independence). It is used mainly for segmentation or tree growing and it is very useful for demographic or behavioural data sets (Delen et al., 2013; Legohérel et al., 2015).

The Chi-squared tests of independence compare the square deviations between observed and expected frequencies. When the chi-square is high, there is a large difference between these two distributions, which means they are dependent. The variable with the highest chi-square value is chosen to split the subset (Valle et al., 2012).

This algorithm follows specific phases. It starts by finding the best partition for each predictor, compares the predictors, splits the data accordingly to it and at last “each of these subgroups are re-analysed independently to produce further subdivisions for analysis” (Kass, 1980:120).

C5.0 (Quinlan, 1993) is a decision tree algorithm, an improved version of C4.5 algorithm. It is capable of handling with any combination of nominal and numeric attributes, manage training data with missing attribute values, robust to noisy data, capable of learning disjunctive expressions and it has a process to solve the problem of the over-fitting (Polat & Gunes 2009; Mantas & Abellán, 2014). As in the C4.5, when the initial decision tree is constructed, the pruning procedure is initiated in order to decrease the overall tree size and the estimated error rate of the tree (Quinlan, 1993).

In the absence of an algorithm which is considered the most efficient and appropriate, several algorithms must be tried in order to get the model (the decision tree) more appropriate to the defined goal.

Therefore, the three most popular algorithms (CART, CHAID and C5.0) were chosen, with the satisfaction cluster as the dependent variable (which evaluates the satisfaction with the main trip) and with the social-demographic characterisation (D1 to D8), the historic behaviour as a tourist (D9 to D13), the pre-travel (BT1 to BT4) and the on-travel behaviour (OT1 to OT5) as the independent variables, which allow to explain and predict the tourist satisfaction.

The variables used in the decision trees algorithms are described in Table 30 (social demographic characteristics and tourist behaviour), Table 31 (pre-travel), Table 32 (on-travel) and Table 33 (post-travel).

**Table 30: Social demographic characteristics and tourist behaviour: variables description**

ID	Description
D1	Age
D2	Gender
D3	Nationality
D3_1	Area of the European Union that the country of nationality belongs
D3_2	Currency of the nationality country
D4	Age education
D5	Occupation
D6	Household size
D7	Phone available
D8	Type of community
D9.1	Paid accommodation with less than 20 guests
D9.2	Paid accommodation with more than 20 guests
D9.3	Own property/second home
D9.4	Friends or relatives
D9.5	Camping site (tent, motorhome, caravan, etc.)
D10.1_3	Type of accommodation (paid or unpaid) historical used on personal travels
D11	Frequency: number of trips
D12.1_2	Duration of personal travels taken
D13.1_3	Type of holiday (package) historical used on personal travels

**Table 31: Pre-travel: variables description**

ID	Description
BT1	Main reason for going on holiday
BT1_MT	Main reason for going on holiday main trip
BT1.1_9	Number of reasons to travel
BT1.1_9_MT	Number of reasons to main trip
BT2.1_9	Number of information sources used by each tourist
BT2.1_9_MT	Number of information sources main trip
BT3.1_8	Number of organising methods
BT3.1_8_MT	Number of organising methods used for main trip
BT4	The nationality and destination currency are the same

**Table 32: On-travel: variables description**

ID	Description
OT1.1	Main holiday destination country
OT1.1_1	Area of the European Union that the destination country belongs
OT1.1_2	Currency of the destination country
OT1.2	Own country as destination country
OT2	Party experience any safety issues while using paid accommodation
OT2.1_7	Number of safety issues while using paid accommodation
OT3	Type of accommodation main trip (paid or unpaid)
OT4	Duration of the main trip
OT5	Type of holiday (package) main trip

**Table 33: Post-travel: variables description**

ID	Description
Satisfaction	Satisfaction (cluster)

Decision trees were performed using the dimensions of each predictor in order to better understand their importance to the tourist satisfaction. At last, a decision tree was created with the majority of the predictors.

Since each algorithm allows several parameterizations, decision trees were performed using and experiencing different parameters, accordingly to Table 34, and appropriately to the data. The parametrizations also allow to avoid the overfitting of the models. Overfitting occurs when the learning algorithm keeps trying to find the best fit to the training data, which will include in the data some peculiarities of the training data instead of finding a general predictive rule and will increase the test set error and reduce the training set error (Dietterich, 1995).

**Table 34: Decision trees parametrization**

Algorithm	Parameters	Default	A	B	C	D	E
<b>CART</b>	Minimum Cases in Parent Node	100	2	2	2	2	2
	Minimum Cases in Child Node	50	1	1	1	1	1
	Improvement Depth	0.0001 5	0.0001 6	0.0001 4	0.0001 3	0.0001 3	0.0001 3
	Impurity measure for categorical targets	Gini	Gini	Gini	Gini	Gini	Gini
	Improvement method	-	-	-	-	Boosting	Bagging
	<b>CHAID</b>	Minimum Cases in Parent Node	100	2	2	2	2
Minimum Cases in Child Node		50	1	2	1	1	1
Depth		3	6	4	5	5	5
Significance level for splitting		0.05	0.05	0.05	0.05	0.05	0.05
Significance level for merging		0.05	0.05	0.05	0.05	0.05	0.05
Chi-square for categorical targets		Pearson	Pearson	Pearson	Pearson	Pearson	Pearson
Minimum change in expected cell frequencies		0.001	0.001	0.001	0.001	0.001	0.001
Maximum iterations for convergence		100	100	100	100	100	100
Improvement method		-	-	-	-	Boosting	Bagging
<b>C5.0</b>	Mode	Simple					
	Favor	Accuracy					
	Expected noise (%)	0					
	Improvement method	Boosting					



It is also important to emphasize about the dependent variable, which represents three categories, corresponding to three different levels of tourist satisfaction with the destination. However, the target was redefined after consulting a hospitality professional, Cristina Petrenko, that said “I think that the person will only return to the destination when very satisfied”<sup>4</sup>. Thereby, only two categories were considered: Less satisfied, which includes the 815 tourists from cluster 1 and the 4596 from cluster 2; and Very satisfied, which contains the 6286 tourists from cluster 3.

This target redefinition can be explained by the competitiveness of a destination and for the destination loyalty, which is mostly important to the tourists very satisfied with the destination (accommodation, reception, natural features and general level of prices). In this way, it is crucial to identify the less satisfied tourists’ profiles and their pre-travel and on-travel behaviours, which can lead to a satisfaction decrease.

### 3.4.2.1 Model quality and validity

In order to produce a higher accuracy in the analysis of the data set, there are techniques used to make the partition of the data into a training and a test set to evaluate the behaviour of the model in different data samples (Li et al., 2010). These techniques are holdout, cross-validation and bootstrap, and they allow to know if the models created work in different data sets.

The holdout method consists on splitting the original data set into a training and a test set of data. A certain amount of data from the data set is used for testing and the remaining data is used for training (Kohavi, 1995). For this study, two thirds of the data were used for training and one third for testing. This partition method has a problem: the subgroups should be split in order that all of them are representative. The method to diminish this problem is called stratification, which has the aim to split the data guarantee that each class is “properly represented in both training and test sets” (Witten & Frank, 2005:152). In the holdout method, the training and test sets are swapped (training the system on the test data and testing it on the training data) and the average of the two results is made to reduce the problem of an unrepresentative data set. The problem here is that the data is split in two equal parts and it is better to use more data to training than for testing (Kohavi, 1995; Witten & Frank, 2005).

Holdout has a variant procedure, which is named repeated holdout. In this variant, the process of training and testing the data is repeated with different data samples. In each time of the process, a certain amount of random data is selected for training and the other is used for testing. The error rate to this procedure is the average of the error rate to each time the process occurred to obtain an overall error rate (Witten & Frank, 2005).

Cross validation or k-fold is a simple variant of the Holdout method. In the cross validation procedure, the original data is partitioned in  $k$  different subsets with the same size. An amount of data (1 subset) is used to test and the other data ( $k-1$  subsets) is used for training. The data is used in turns for testing and training, which means this is a circular process in which all the subsets are used one time for testing (Kohavi, 1995). This method is used when the amount of data is limited. The standard subsets

---

<sup>4</sup> Free text translation from “acho que realmente a pessoa só irá voltar ao destino, quando ficar muito satisfeita.”

used in this method is ten and use stratification, which is called 10-fold cross-validation. Ten subsets is the best partition to obtain the most correct error rate (Witten & Frank, 2005).

Bootstrap was described for the first time in Efron and Tibshirani (1993). It is based on the statistical procedure of sampling with replacement and it is the best method to estimate error in small datasets. In the Bootstrap method, the training set will be a sample with replacement, which means that it can have repeated data elements (Kohavi, 1995).

The variant method - 0.632 bootstrap – has a dataset with  $n$  instances, that is sampled  $n$  times, with replacement, to give another dataset of  $n$  instances. The replacement means that one time that an element was chosen to the new dataset (the training set), that same element will continue in the original dataset (and in the training set also). Then, some elements will be repeated in the training set, but some elements in the original dataset have not been chosen. These elements are the ones that form the test set (Witten & Frank, 2005).

For this method, the test set is defined by approximately 36.8% of the instances and the training set is defined by approximately 63.2% of them. The error rate of a sample like this training set is not really exact when comparing with the 90% of the data used for the training set used in the 10-fold cross-validation. Therefore, a mix of the test set error rate and the resubstitution error to the training set is used to define a better error estimate (Kohavi, 1995; Witten & Frank, 2005):

$$e = 0.632 \times e_{test\ instances} + 0.368 \times e_{training\ instances}$$

The Bootstrap method is repeated with different replacement samples for the training set. Then, the results of the error estimate are averaged (Witten & Frank, 2005).

### 3.4.2.1.1 Bagging and Boosting

Bagging (Breiman, 1996) and boosting (Freund & Schapire, 1996) are methods used in the training data to generate different classifiers (models). The aim of these methods is to create a strong classifier in order to make accurate predictions (Quinlan, 2006).

Bagging uses bootstrap data samples (with replacement) to produce replicate training sets. It creates multiple models and combines them, in order to obtain more reliable predictions. Bagging requires that the learning system is unstable, for a small change to the training set be able to produce different classifiers (Quinlan, 2006; IBM, 2016a).

Boosting uses all the training data at each repetition, but it assigns a weight for each instance in the data that is its importance. This weight gives more importance to some instances, which allows the learner to focus on different instances. If an instance was misclassified by previous learners, the weight assigned to it should be higher, in order to the next learners give more attention to it during training. The learner will focus on different instances and form different classifiers. Boosting allows an increase in accuracy, but it leads to a deterioration on some data sets. Still, boosting does not exclude learning systems that generate poor predictors, since their error do not exceed 50% (Quinlan, 2006).

In both methods, the resulting classifiers are combined by voting to produce a composite classifier. Although, in bagging, all the classifiers have the same vote, while in boosting, different voting strengths are attributed to the classifiers, depending on their accuracy (Quinlan, 2006).

In the Bagging method, predicted values for categorical targets can be combined using voting, highest probability wins, or highest mean probability. Voting selects the category that has the highest probability most often across the base models. Highest probability wins selects the category that achieves the single highest probability across all base models. Highest mean probability selects the category with the highest value when the category probabilities are averaged across base models (IBM, 2016a).

In this study, the boosting method was performed to all the algorithms tested (CART, CHAID and C5.0). However, the bagging method was only performed to CART and CHAID, as in IBM SPSS Modeler software, does not exist the option to perform it to the C5.0. The results of these methods are in the Table 54 (Appendix C). Although these models should produce higher accuracies, the measures are similar to the models without these methods. The best models to each group of variables are described in the Results chapter (4.2. Predictors of Tourist satisfaction subchapter).

#### **3.4.2.1.2 Other procedures to increase accuracy**

In the cases that the data is unbalanced (the number of data representing the class of interest is much lower than the data of the other classes), the majority class (category with more data) has the highest interest from the model and the prediction of the minority class (category with less data) is not correctly developed. This induces a higher misclassification rate for the minority class instances (Lopez et al., 2013).

To avoid this problem, it is usual to balance the data sample in order to define a model with the same weight of each target categories in the models development. Resampling techniques can be categorized into (He & Garcia, 2009; Lopez et al., 2013):

1. Undersampling method, which create a subset of the original dataset by removing data (usually majority class instances) in a random way.
2. Oversampling method, which add a superset of the original dataset by replicating some instances from the minority instances or creating new instances from existing ones.

As the oversampling method produces tied instances, the classification performance on the unseen testing data is far worse than in the undersampling method. The problem of the undersampling is more objective and perceptible: some concepts from the majority class can be missed. Then, the undersampling method is better to avoid overfitting.

Another approach to solve this problem is the recurrence to misclassification costs. Misclassification costs are weights applied to specific output values, which means that they allow to define the relative importance of prediction errors. As a result, these weights are factored into the model and can change the prediction in order to protect from costly mistakes.

All the samples used for modelling were randomly selected, which guarantees (theoretically) the similar structure to the training and test sets. The undersampling method was used to build the model with the pre-travel variables, because, otherwise, only the very satisfied cases were being correctly predicted. Related to the misclassification costs, these were not used, due to the good results without using them and the difficulty to estimate the appropriate costs to use.

### 3.5 Evaluation

In this phase occurs an evaluation to the model and model construction, to make sure that every step is correct and that nothing is missing in the model design. When this is guaranteed, the project leader has to decide what to do with the results that were produced (Chapman et al., 2000).

The confusion matrix (or classification matrix) is a great tool to evaluate the model performance for classification models. A confusion matrix has two entries: the predicted and the actual classes from the model. The ideal scenario is a filled diagonal and zeros in the other rows (Delen et al., 2013; IBM, 2016a).

An example of a confusion matrix with two classes is represented at Table 35. The true positives (TP) are the positive tuples that were correctly classified by the classifier, while the true negatives (TN) are the negative tuples that were correctly classified. The false positives (FP) are the negative tuples that were incorrectly labelled and the false negatives (FN) are the positive tuples that were incorrectly labelled (IBM, 2016a).

**Table 35: Confusion matrix**

		Predicted class	
		Successful	Unsuccessful
Actual class	Successful	True positives (TP)	False negatives (FN)
	Unsuccessful	False positives (FP)	True negatives (TN)

Source: Adapted from Delen et al. (2013).

Regarding the importance of predicting the less satisfied tourists this category was considered the successful class and the very satisfied was considered the unsuccessful class.

For purposes of this study, the performance measures used were: i) overall accuracy; ii) specificity; iii) precision and; iv) AUC (Area Under ROC curve). Each model of the study was evaluated using these measures (Delen et al., 2013).

The overall accuracy is the percentage of data that is correctly classified by the model, that is, the ratio of correctly predicted cases to the total number of cases (Delen et al., 2013; IBM, 2016a).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy is a function of sensitivity and specificity. Sensitivity, true positive recognition rate or recall is the proportion of positive tuples that are correctly identified, that is, the ratio of the true positive to the sum of the true positive and false negative. (Delen et al., 2013; He & Garcia, 2009).

$$Sensitivity (\% \text{ cases well classified} - \text{less satisfied} / \text{total well classified}) = \frac{TP}{TP + FN}$$

Specificity or true negative rate (TN) is the proportion of negative tuples that are correctly identified, which means it is the ratio of the number of true negative to the sum of true negative and false positive (Delen et al., 2013).

$$Specificity (\% \text{ cases well classified} - \text{very satisfied}) = \frac{TN}{TN + FP}$$

Precision is the ratio of true positive correctly predicted to the sum of the true positive and false positive (He & Garcia, 2009; Delen et al., 2013).

$$Precision (\% \text{ cases well classified} - \text{less satisfied}) = \frac{TP}{TP + FN}$$

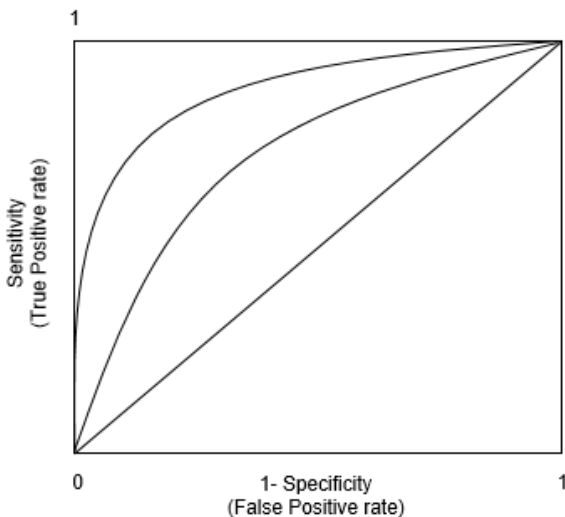
The F-measure is the mean of the precision and sensitivity performance measures (Witten & Frank, 2005).

$$F - \text{measure} = 2 \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

Moreover, the accuracy of a model can be observed using a ROC (Receiver Operation Characteristic) curve in Figure 9. It is a plot showing two probabilities, namely sensitivity and 1-specificity (Pundir & Amala, 2015).

ROC curves allow the visualization of the trade-off between the rate at which the model can accurately recognize the positive tuples versus the rate at which it identifies the negative tuples as positive for different portions of the test set. AUC is a measure of the accuracy of the model and an area of 1.0 represents a model with perfect accuracy (Pundir & Amala, 2015; IBM, 2016a).

**Figure 9: ROC Curve representation**



Source: Sayad (2016).

The Gain measure can be represented through a table or a chart. It measures the difference between the mean or proportion at a given node and the overall mean. The more far the mean or proportion differs from the overall mean or proportion, the more useful the tree model will be as a tool for making decisions (IBM, 2016a). The gains values are defined as:

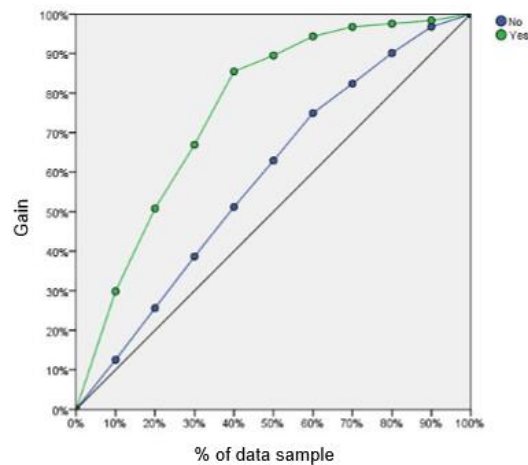
$$Gains = \frac{Hits \text{ in increment}}{Total \text{ number of hits}} \times 100$$

The chart (Figure 10) illustrates how widely it is necessary to cast the net to capture a given percentage of all the hits in the tree. It measures the effectiveness of a classification model calculated as the ratio between the results obtained with and without the model (Sayad, 2016).

From the chart analysis, it is possible to observe lines with two different formats (straight and curve). The straight diagonal line represents the expected response for all the data sample, if the model

was not used. The curved line represents how much the response can be improved if only those who rank in the higher percentages were included (IBM, 2016a).

**Figure 10: Gains chart representation**



Source: IBM (2016b).

The gains chart analyses the model performance only in a portion (node) of the data sample, while the confusion matrix evaluates on the whole data sample (Sayad, 2016; IBM, 2016a).

### 3.5.1 Models to perform

Therefore, there were created and evaluated six different models for each group of variables (social demographics characteristics and tourist behaviour characterisation, pre-travel and on-travel) and algorithm (CART, C5.0 and CHAID). The models evaluation was based not only in the main measures (i.e. the overall accuracy and AUC), when applied to test data, but also in the meaning of the models to the tourism industry.

For this purpose, the relative importance of each predictor was measured through IBM SPSS Modeler (v.18). The predictor importance chart indicates the relative importance of each variable for a particular model and the sum of the values for all predictors on the display is 1.0. However, the feature selection displays each variable based on the strength of its relationship to the specific target, independent of other variables. Predictor importance has no relationship with the model accuracy, because it “just relates to the importance of each predictor in making a prediction” (IBM, 2016a:42), not if it is accurate (IBM, 2016a).

Also a sensitivity analysis was performed (Cortez & Embrechts, 2013) which represents the relationship between each of the main predictors and the predicted class (cluster – satisfaction level).

### 3.6 Deployment

In the framework of this project would be important to develop and implement a system that could predict the most appropriate destination according to the tourist profile using data mining models. The deployment of this project is the elaboration of this report and the analysis developed. This document is also an added value for all tourism companies, to help understand the explanatory factors of the tourist satisfaction with the destination.

## 4 Results and discussion

The aim of this chapter is the analysis and evaluation of the developed methods in the CRISP-DM modelling phase. Firstly, using cluster analysis, the tourist satisfaction is assessed, then through the different decision tree models, the goal is to determine which attributes most explain the tourist satisfaction with the destination and accommodation and also which the best model to predict satisfaction is.

### 4.1 Tourist satisfaction assessment

As described, a k-means cluster analysis was made to obtain satisfaction clusters (groups) of European tourists, being identified three clusters: 815 tourists in cluster 1 (not very satisfied), 4596 tourists in cluster 2 (satisfied) and 6286 tourists in cluster 3 (very satisfied).

Table 36 allows to characterize the 3 clusters of tourists regarding their satisfaction with the destination. Cluster 1 is formed by the less satisfied tourists, showing mean values between 2.16 for satisfaction with the general level of prices and 3.19 for satisfaction with accommodation. In the opposite site, cluster 3 shows all the three mean values above 3.65 revealing that it represents the very satisfied tourists. In the middle, cluster 2 shows mean values between 2.96 for satisfaction with the general level of prices and 3.59 for satisfaction with the natural features.

These results allow to conclude that in general terms tourists are satisfied with the destination choice for their main holiday, and the differences found respect to the (three) levels of satisfaction. Moreover, the Eta<sup>2</sup> coefficients are all above 30%, allowing to conclude that all the four dimensions of satisfaction contribute to differentiate the clusters, being the satisfaction with the general level of prices (Eta<sup>2</sup>=45.7%) the one that most explains the tourist satisfaction. Evidently, all Welch tests reveal significant differences among the mean values of the three clusters.

**Table 36: Satisfaction assessment: clusters descriptive statistics and validation**

Satisfaction at the destination	Cluster 1		Cluster 2		Cluster 3		Welch-test (W) Eta <sup>2</sup>
	Not very satisfied		Satisfied		Very satisfied		
	(n=815)		(n=4596)		(n=6286)		
	M	SD	M	SD	M	SD	
Satisfaction with accommodation	3.19	0.68	3.36	0.47	3.89	0.26	$W_{(2;1986.3)}=2705.6;p<0.001;$ Eta <sup>2</sup> = 33.6%
Satisfaction with the reception at the destination	2.98	0.73	3.14	0.45	3.81	0.30	$W_{(2;2016.9)}=4128.5;p<0.001;$ Eta <sup>2</sup> = 42.0%
Satisfaction with the natural features	2.60	0.79	3.59	0.51	3.89	0.32	$W_{(2;2006.2)}=1596.1;p<0.001;$ Eta <sup>2</sup> =35.2%
Satisfaction with the general level of prices	2.16	0.75	2.96	0.46	3.66	0.47	$W_{(2;2109.0)}=3973.9;p<0.001;$ Eta <sup>2</sup> = 45.7%

M – Mean; SD – Standard Deviation.

Notes: Scale from 1 (not at all satisfied) to 4 (very satisfied);

Scheffé test shows that all differences are significant ( $p<0.001$ )

The proportions of tourists who register a formal complaint differ among clusters (Table 37). The results show that only 4.2% of tourists registered a complaint (in one of the following services – accommodation, transport, restaurants and leisure activities) and that the proportion of tourists who registered a complaint is higher in cluster 1 (6.4%) and then in cluster 2 (4.7%). In both cases this proportion is higher than the overall proportion. As expected, the proportion of complaints is lower in the cluster 3 (3.6%), which contains the very satisfied tourists.

The same conclusion is obtained when is made the analysis about the relationship between the clusters and the number of registered complaints. Tourists with zero complaints are more present in cluster 3 (96.4%) than in cluster 1 (93.6%), showing a relationship between the number of complaints and the tourist satisfaction.

The tourists who made a higher number of complaints (four) are zero in cluster 2 and cluster 3, which are the satisfied and very satisfied tourists, respectively. On the other hand, in cluster 1, there is 0.12% of tourists who made four complaints. Although it is a small percentage, it is possible to verify that the number of complaints is lower when the tourists are very satisfied (cluster 3).

To conclude, despite the relation between the cluster and the formal complaints is significant (independence chi-squared tests show an associated p-value lower than 0.001), this relation is weak (association measures lower than 0.1).

**Table 37: Registered formal complaints by satisfaction cluster**

Characteristics	Cluster satisfaction			Total	Chi-squared test <sup>a</sup> Measure of association
	Cluster 1 Not very satisfied	Cluster 2 Satisfied	Cluster 3 Very satisfied		
	(n=815) %	(n=4596) %	(n=6286) %		
<b>Formal complaint</b>					
No	93.6	95.3	96.4	95.8	$\chi^2_{(2)}=17.740; p<0.001^a$ Cramer's V = 0.039
Yes	6.4	4.7	3.6	4.2	
<b>Number of formal complaints</b>					
0	93.62	95.30	96.39	95.76	$\chi^2_{(4)}=46.569; p<0.001^a$ Eta = 0.047
1	5.52	4.31	3.42	3.92	
2	0.49	0.37	0.19	0.28	
3	0.25	0.02	0.00	0.03	
4	0.12	0.00	0.00	0.01	

Note: a) Monte Carlo p-value.

The differences among the clusters were also studied for the social demographic characteristics variables (Table 38) and for the tourist behaviour characterisation (Table 39).

Related to the social demographic characteristics, the variables in analysis are: i) gender; ii) age group; iii) area of the European Union; iv) currency of the nationality country; v) age education; vi) occupation; vii) household size; viii) phone available; and ix) type of community.

Referring to gender, it was found that in cluster 1 (44.7%) and in cluster 2 (47.4%), the values are higher than the male overall percentage (43.2%), having a proportion higher than in cluster 3 (40.0%).

In the age group, the not very satisfied tourists (cluster 1) have a higher proportion of younger tourists than in the clusters with higher levels of satisfaction. There are 12.8% tourists with age between 15-24 years in cluster 1, 9.6% in cluster 2 and 7.0% in cluster 3. This also happens to the ranges of 25-39 years and 40-54 years tourists, which means that younger tourists are less satisfied than the older tourists (55 or more years). The proportion of the older tourists (55 or more years) increases with the satisfaction (clusters). Cluster 1 has 32.5%, cluster 2 has 37.7% and cluster 3 has 46.0% of this group of tourists.



Concerning the area of the EU that the nationality country belongs, clusters 1 and 2 have a higher proportion of tourists from the countries of Iberian Peninsula, eastern European, Balkans and Baltics, than the overall percentage, being the less satisfied tourists. There is 10.8% of tourists from Iberian Peninsula in cluster 1, 11.4% in cluster 2 and 4.9% in cluster 3, meaning that they are mostly in clusters 1 and 2. In cluster 3, the North and Central European countries have a higher proportion than the overall percentage, being the most satisfied tourists. The North European countries have 7.4% of tourists in cluster 1, 12.2% in cluster 2 and 13.9% in cluster 3.

The results from the currency of the nationality country analysis show that tourists who use euro have a higher proportion in clusters 1 and 2, with 69.5% in cluster 1 and 68.9% in cluster 2, than in cluster 3 (64.5%) and the overall percentage, being the less satisfied tourists. Cluster 3 has a higher percentage of non-euro tourists (35.5%) than in cluster 1 and 2, being the most satisfied tourists.

Referring to the age education, it was observed that tourists who are still studying have a higher proportion in clusters 1 and 2, which means these tourists are majority less satisfied.

Regarding the occupation, the self-employed and employees are in higher proportion in clusters 1 and 2, with percentages above the overall percentage (10.1% and 42.1%, respectively), being the less satisfied tourists. Cluster 1 has 10.6% of self-employed tourists, cluster 2 has 11.3% and cluster 3 has 9.1%. However, the most of both self-employed and employees are in cluster 2 (satisfied). Manual workers are mostly in cluster 1 (7.4%) and tourists who do not work are mostly in cluster 3 (43.0%) and have a higher percentage than the overall, being the most satisfied tourists.

In what concerns the household size, tourists with a household size of three, and four or more people are in higher proportion in clusters 1 (three – 20.7%, four or more - 18.6%) and 2 (three – 18.6%, four or more - 16.6%), having percentages above the overall percentage, being the less satisfied tourists. Cluster 3 has the highest percentage of tourists with a small household size (one or two people), with 19.4% of tourists with one person and 50.1% with two people, being the most satisfied tourists.

Tourists only with landline (4.1%) or with landline and mobile phone (77.7%) in cluster 3 are in higher proportions, having percentages above the overall percentage, being the most satisfied tourists. Tourists who use mobile phone only are in higher proportions in cluster 1 (22.4%) and 2 (21.0%), having percentages above the overall percentage, being the less satisfied tourists.

The type of community analysis allows to conclude that tourists from the large towns are in higher proportion in clusters 1 (34.4%) and 2 (33.8%). This means that tourists from the larger towns slightly tend to have a lower level of satisfaction.

In conclusion, there are significant relations between the variables and the clusters, which can be evaluated by the chi-squared independent test.

**Table 38: The social demographics characteristics by satisfaction cluster**

Characteristics	Cluster satisfaction			Total	Chi-squared test <sup>a</sup> Measure of association
	Cluster 1	Cluster 2	Cluster 3		
	Not very satisfied	Satisfied	Very satisfied		
	(n=815)	(n=4596)	(n=6286)		
	%	%	%		
<b>Gender</b>					
Male	44.7	47.4	40.0	43.2	$\chi^2_{(2)}=60.777; p<0.001^a$ Cramer's V = 0.072
Female	55.3	52.6	60.0	56.8	
<b>Age group</b>					
15-24 years	12.8	9.6	7.0	8.4	$\chi^2_{(6)}=142.094; p<0.001^a$ Cramer's V = 0.078
25-39 years	24.3	22.1	17.3	19.7	
40-54 years	30.4	30.6	29.7	30.1	
55 or more years	32.5	37.7	46.0	41.8	
<b>Area of the European Union (nationality)</b>					
North European countries	7.4	12.2	13.9	12.8	$\chi^2_{(10)}=307.609; p<0.001^a$
Central European countries	43.3	43.0	54.7	49.5	
Countries of the Iberian Peninsula	10.8	11.4	4.9	7.8	Cramer's V = 0.120
Eastern European countries	16.2	17.1	12.8	14.7	If computed for nationality <sup>5</sup>
Balkan countries	19.7	14.1	11.5	13.0	Cramer's V = 0.201
Baltics	2.6	2.2	2.2	2.2	
<b>Currency of the nationality country</b>					
Euro	69.5	68.9	64.5	66.6	$\chi^2_{(2)}=24.111; p<0.001^a$ Cramer's V = 0.047
Non-euro	30.5	31.1	35.5	33.4	
<b>Age education</b>					
Up to 15	7.6	6.1	8.2	7.3	$\chi^2_{(6)}=104.215; p<0.001^a$ Cramer's V = 0.067
16-19	35.5	30.5	36.6	34.1	
20 years and older	48.0	55.6	49.7	51.9	
Still Studying	8.3	7.6	5.0	6.3	
No full-time education	0.6	0.2	0.5	0.4	
<b>Occupation</b>					
Self-employed	10.6	11.3	9.1	10.1	$\chi^2_{(6)}=31.920; p<0.001^a$ Cramer's V = 0.037
Employees	42.6	43.6	41.0	42.1	
Manual workers	7.4	6.7	6.9	6.9	
Not working	39.4	38.4	43.0	40.9	
<b>Household size</b>					
1	16.1	16.7	19.4	18.1	$\chi^2_{(6)}=46.362; p<0.001^a$ Cramer's V = 0.045
2	44.6	48.2	50.1	49.0	
3	20.7	18.6	15.7	17.2	
4+	18.6	16.6	14.8	15.7	
<b>Phone available</b>					
Mobile only	22.4	21.0	18.2	19.6	$\chi^2_{(4)}=23.561; p<0.001^a$ Cramer's V = 0.032
Landline only	2.5	3.3	4.1	3.7	
Mobile and landline	75.1	75.7	77.7	76.7	
<b>Type of community</b>					
Rural area or village	28.6	26.2	29.2	28.0	$\chi^2_{(4)}=14.693; p=0.006^a$ Cramer's V = 0.025
Small or middle sized town	37.0	40.0	39.2	39.3	
Large town	34.4	33.8	31.6	32.7	

Note: a) Monte Carlo p-value.

<sup>5</sup> View Table 46 in the Appendix A for the nationality country analysis.

The results from the relationship between the tourist behaviour characterisation variables and tourist satisfaction can be observed in Table 39. These variables are: i) paid accommodation with less than 20 guests; ii) paid accommodation with more than 20 guests; iii) own property/second home; iv) friends or relatives; v) camping site; vi) type of accommodation (paid or unpaid) in personal travels; vii) frequency: number of trips; viii) duration of personal travels taken; and ix) type of holiday (package).

Related to paid accommodation with less than 20 guests, tourists who went one (28.2% for cluster 1 and 25.5% for cluster 2) and five or more times (5.3% for cluster 1 and 5.4% for cluster 2) are in higher proportions in clusters 1 and 2, having percentages above the overall percentage, being the less satisfied tourists. Cluster 3 has a higher proportion of tourists who did not go to this type of accommodation (53.4%) and tourists who went 3 (5.2%), being the most satisfied tourists.

Tourists who went more times (two or more) to paid accommodations with more than 20 guests are the ones with more proportion in cluster 3. Cluster 3 is characterised by 14.8% of tourists who went two times to this kind of accommodation, 7.2% who went three times, 4.3% who went four and 7.1% who went five times or more. Clusters 1 and 2 have mainly tourists who had never go to this accommodation (46.6% for cluster 1 and 42.2% for cluster 2). This indicates that people who went more times to this type of accommodation are more satisfied.

From the results observed, the duration of personal travels taken is related with the tourist satisfaction. Tourists who normally travel between 4 and 13 nights away are in higher proportions in clusters 1 (67.9%) and 2 (68.0%), having percentages above the overall percentage, being the less satisfied tourists. Tourists who had both type of duration in their travel are in a higher proportion in cluster 3 (18.2%) than the overall percentage, being the most satisfied tourists.

The other variables (own property/second home, friends or relatives, camping site, type of accommodation (paid or unpaid) in personal travels, frequency and type of holiday (package)) have a Cramer's V far below 0.04. Thus, they are not related to the tourists' satisfaction.

**Table 39: Tourist behaviour characterisation by satisfaction cluster**

Characteristics	Cluster satisfaction			Total	Chi-squared test <sup>a</sup> Measure of association
	Cluster 1	Cluster 2	Cluster 3		
	Not very satisfied	Satisfied	Very satisfied		
	(n=815)	(n=4596)	(n=6286)		
	%	%	%		
<b>Paid accommodation with less than 20 guests</b>					
0	49.9	48.9	53.4	51.4	$\chi^2_{(10)}=39.831; p<0.001^a$ Cramer's V = 0.041
1	28.2	25.5	22.6	24.1	
2	10.7	12.0	11.5	11.6	
3	3.2	5.3	5.2	5.1	
4	2.7	2.9	2.9	2.9	
5 or more times	5.3	5.4	4.4	4.9	
<b>Paid accommodation with more than 20 guests</b>					
0	46.6	42.2	39.9	41.4	$\chi^2_{(10)}=35.440; p<0.001^a$ Cramer's V = 0.039
1	27.1	28.2	26.7	27.3	
2	10.2	13.1	14.8	13.8	
3	5.8	6.5	7.2	6.8	
4	3.7	3.2	4.3	3.8	
5 or more times	6.6	6.8	7.1	6.9	
<b>Own property /second home</b>					
0	72.1	72.2	74.9	73.6	$\chi^2_{(10)}=17.683; p=0.059^a$ Cramer's V = 0.028
1	8.3	9.4	8.3	8.7	
2	4.8	4.3	4.1	4.3	
3	2.3	2.9	2.4	2.6	
4	1.6	1.6	1.8	1.7	
5 or more times	10.9	9.6	8.5	9.1	
<b>Friends or relatives</b>					
0	55.6	54.7	56.8	56.0	$\chi^2_{(10)}=18.927; p<0.040^a$ Cramer's V = 0.028
1	16.3	16.9	17.2	17.0	
2	10.5	9.6	9.0	9.3	
3	5.4	5.3	4.6	4.9	
4	2.6	4.0	2.9	3.3	
5 or more times	9.6	9.5	9.5	9.5	
<b>Camping site (tent, motorhome, caravan, etc.)</b>					
0	85.5	85.4	86.0	85.7	$\chi^2_{(10)}=5.011; p=0.895^a$ Cramer's V = 0.015
1	8.6	8.4	8.0	8.2	
2	2.1	2.4	2.2	2.2	
3	0.6	1.2	1.1	1.1	
4	0.5	0.5	0.6	0.6	
5 or more times	2.7	2.1	2.1	2.2	
<b>Type of accommodation (paid or unpaid) in personal travels</b>					
Always in paid accommodation	30.8	28.9	30.9	30.1	$\chi^2_{(4)}=11.596; p=0.020^a$ Cramer's V = 0.022
Never in paid accommodation	21.1	18.6	19.4	19.2	
Paid and unpaid accommodation	48.1	52.5	49.7	50.7	
<b>Frequency: Number of trips</b>					
Once	24.0	17.8	18.9	18.9	$\chi^2_{(10)}=35.310; p<0.001^a$ Cramer's V = 0.039
Twice	16.8	16.1	17.0	16.7	
3 times	13.4	15.1	15.3	15.1	
4 or 5 times	15.7	20.7	19.8	19.9	
6 to 10 times	14.4	15.2	15.8	15.5	
More than 10 times	15.7	15.1	13.2	14.1	
<b>Duration of personal travels taken</b>					
Between 4 and 13 nights away	67.9	68.0	62.5	65.0	$\chi^2_{(4)}=61.867; p<0.001^a$ Cramer's V = 0.051
More than 13 nights away	20.6	18.4	19.2	19.0	
Both types of duration	11.5	13.6	18.2	16.0	
<b>Type of holiday (package)</b>					
Always 'all-inclusive holidays'	16.8	11.7	14.1	13.4	$\chi^2_{(4)}=28.585; p<0.001^a$ Cramer's V = 0.035
Never 'all-inclusive holidays'	69.1	71.4	68.2	69.5	
Combination of types of holiday	14.1	16.9	17.7	17.1	

Note: a) Monte Carlo p-value.

In conclusion, cluster 1 (with 815 tourists) is characterized by the tourists with the lowest satisfaction, which represent the highest percentage of formal complaints with a service (6.4%). The majority of the tourists are women (55.3%) with 55 or more years old (32.5%) or between 40 and 54 years old (30.4%) from the central European countries (43.3%), where the currency is euro (69.5%), with 20 years or more of full-time education (48.0%), work as employees (42.6%), have a household of two people (44.6%) in a small or middle sized town (37.0%) and have mobile phone and landline (75.1%). Related to the tourist behaviour characterisation, cluster 1 is defined by people who had never stayed in a paid accommodation with less than 20 guests (49.9%). This cluster has the highest proportion of tourists that never go to paid accommodation with more than 20 guests (46.6%). Tourists from cluster 1 stay in a combination of paid and unpaid accommodation (48.1%), never with all-inclusive holidays (69.1%) and travel between 4 and 13 nights away (67.9%), only travelled once in 2014 (24.0%).

Cluster 2 (with 4596 tourists) is characterized by satisfied tourists, which are in the middle of the other clusters about the execution of a formal complaints with some service (4.7%). The majority of the tourists are women (52.6%) with 55 or more years (37.7%) from the central European countries (43.0%), where the currency is euro (68.9%). This cluster has the highest percentage of tourists with 20 years or more of full-time education (55.6%), who work as employee (43.6%) and also the highest percentage of tourists from a small or middle sized town (40.0%). They have a household of two people (48.2%) and also mobile phone and landline (75.7%). Related to the tourist characterisation, cluster 2 has tourists that had never stayed in a paid accommodation with less than 20 guests (48.9%), who chose a combination of both paid and unpaid accommodation (52.5%) and also who never went to all-inclusive holidays (69.1%). They normally go for holidays 4 or 5 times a year (20.7%) and they are who more travel between 4 and 13 nights (68.0%).

Cluster 3 (with 6286 tourists) is characterized by very satisfied tourists, which represent the tourists who did not formal complaints with some service (96.4%), which tend to be women (60.0%), with 55 or more years (46.0%) and from the central European countries (54.7%). Also, the tourists are from countries which currency is the euro (64.5%). This cluster is also defined by tourists who do not work (43.0%), live in a small or middle sized town (39.2%) and who have a household of two people (50.1%) and have mobile phone and landline (77.7%). Related to the tourist behaviour characterisation, cluster 3 tend to have tourists who had never stayed in a paid accommodation with less than 20 guests (53.4%) and who never stay in all-inclusive holidays and travel between 4 and 13 nights away (62.5%). They mostly stay in a combination of paid and unpaid accommodation (49.7%) and travel between four or five times (19.8%)<sup>6</sup>.

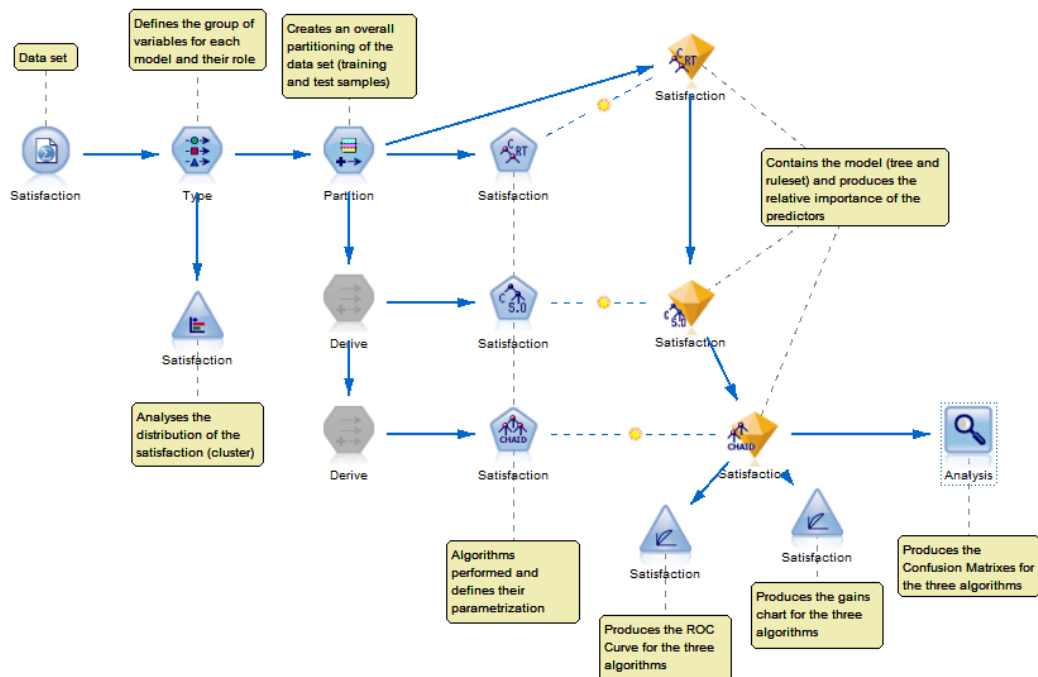
---

<sup>6</sup> In addition to the previous analysis, it was studied the differences between the existence of a formal complaint (no or yes) and the social-demographic characteristics variables (Table 47– Appendix B) and the tourist behaviour characterisation (Table 49– B).

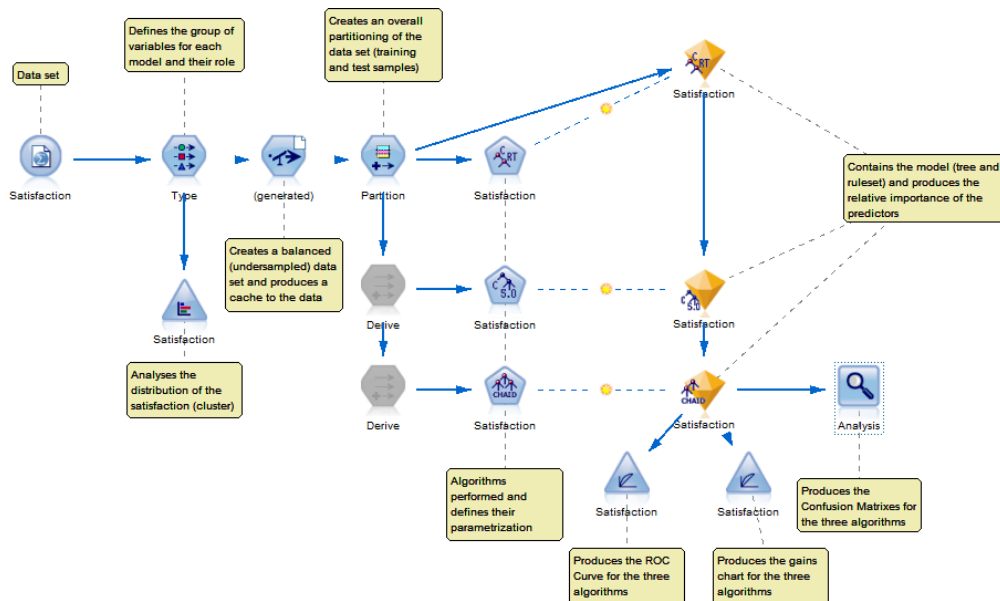
## 4.2 Predictors of Tourist satisfaction

The decision tree models were obtained through IBM SPSS Modeler (v. 18). To generate the decision trees, there were used four groups of variables: i) only social demographics and tourists behaviour; ii) only pre-travel; iii) only on-travel and; iv) all variables. The same stream was used for the social demographics, on-travel and all variables, which is illustrated in Figure 11. To the group of all variables, the 10-fold method was also performed, where the stream is demonstrated in Figure 37 (Appendix C). However, for the pre-travel variables, the (undersampling) balance of the data sample was performed, due to the inexistent prediction of the less satisfied tourists. Then, a second stream was performed (Figure 12).

**Figure 11: IBM SPSS Modeler stream for the social demographics characteristics for tourist characterisation and on-travel variables**



**Figure 12: IBM SPSS Modeler stream for the pre-travel variables**



Through the previous streams, some analysis were made about the resulting models. The results of the most common evaluation measures in classification problems are described in Appendix C: Table 50 (social demographics characteristics and tourist behaviour characterisation), Table 51 (pre-travel variables), Table 52 (on-travel variables) and Table 53 (all variables). The measures presented are the most common in this type of analysis and the measures not calculated, such as sensitivity, would be just a complement to the calculated ones.

In order to analyse the models performance and accuracy for the test sample, the confusion matrix is analysed for the best model of each group, namely in the Table 40 (social demographics characteristics and tourist behaviour characterisation), Table 41 (pre-travel variables) and Table 42 (on-travel variables).

**Table 40: Confusion matrix for the model of the social demographics characteristics and tourist behaviour characterisation variables**

		Predicted class	
		CHAID (Default)	Less satisfied
Actual class	Less satisfied	58.92%	41.08%
	Very satisfied	29.69%	70.31%

Note: Accuracy = 65.04%

**Table 41: Confusion matrix for the model of the pre-travel variables**

		Predicted class	
		CHAID (Default)	Less satisfied
Actual class	Less satisfied	63.52%	36.48%
	Very satisfied	56.58%	43.42%

Note: Accuracy = 53.58%

**Table 42: Confusion matrix for the model of the on-travel variables**

		Predicted class	
		CHAID (E Bagging)	Less satisfied
Actual class	Less satisfied	32.23%	67.77%
	Very satisfied	16.30%	83.70%

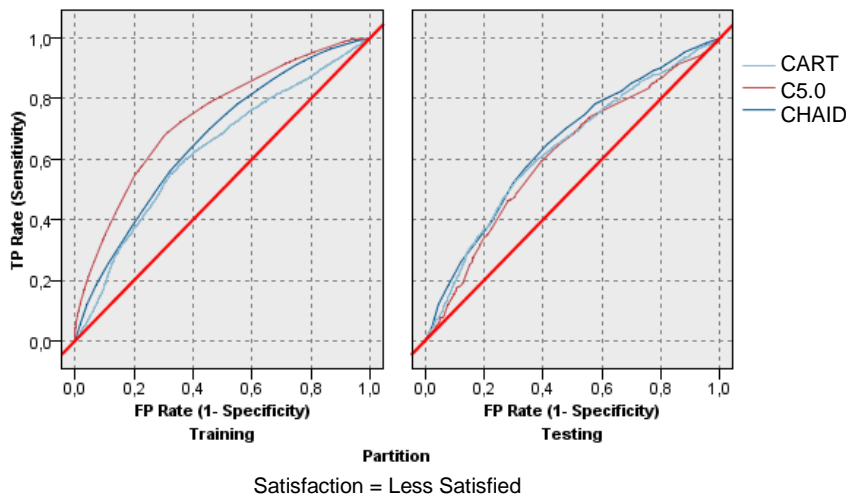
Note: Accuracy = 59.29%

The best algorithm for the social demographics, pre-travel and the group with all the variables is the CHAID algorithm. The model of the social demographics (Table 40) classified correctly 58.92% of the less satisfied tourists, i.e. it predicted 58.92% of the cases as less satisfied, when they are really less satisfied. In other hand, the model classified correctly 70.31% of the very satisfied tourists. In this way, the model predicted correctly 65.04% of the tourists' satisfaction, which makes it an efficient and accurate model. Related to the importance of the social demographics characteristics and historical

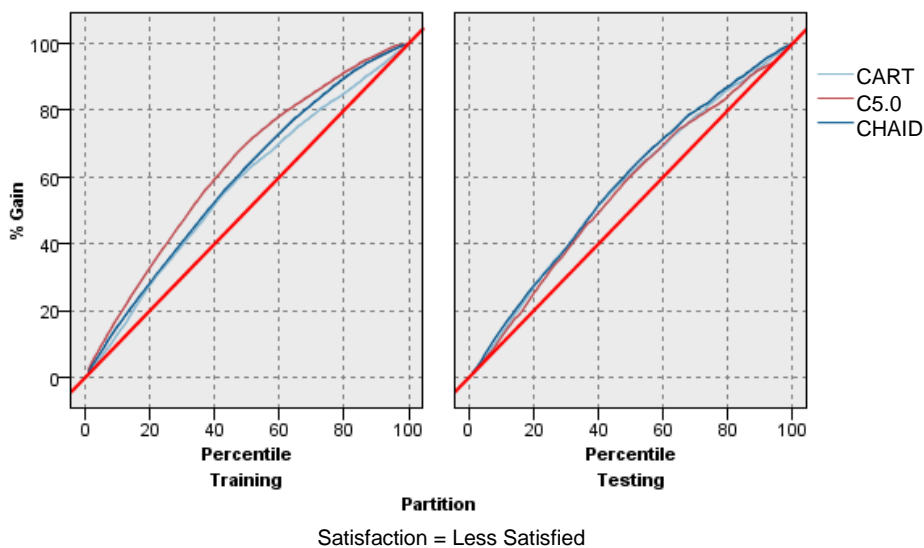
behaviour variables, the predictor importance to the variables with a higher importance are D3 (nationality) with 0.42, D1 (age) with 0.11 and D6 (household size) with 0.11.

The ROC curve (Figure 13) analyses the precision of the model predictions of the social demographics characteristics and tourist behaviour characterisation variables. Ideally, the closer the upper left corner point (0,1), better is the classifier, because it presents a higher True Positive rate and a lower False Positive rate. The AUC (Area Under the ROC Curve) is a measure of the accuracy of the model. The AUC is 0.631 for the CART algorithm, 0.612 for the C5.0 and 0.648 for the CHAID. Therefore, as referred before, the CHAID algorithm is the best one to predict the satisfaction with the demographic variables. The gains chart (Figure 14) measures the effectiveness of a model calculated as the ratio between the results obtained with and without the model. As it can be observed, the results are better with the model, since the model curves are apart from the baseline.

**Figure 13: ROC Curve for the model of the social demographics characteristics and tourist behaviour characterisation variables**



**Figure 14: Gains chart for the model of the social demographics characteristics and tourist behaviour characterisation variables**

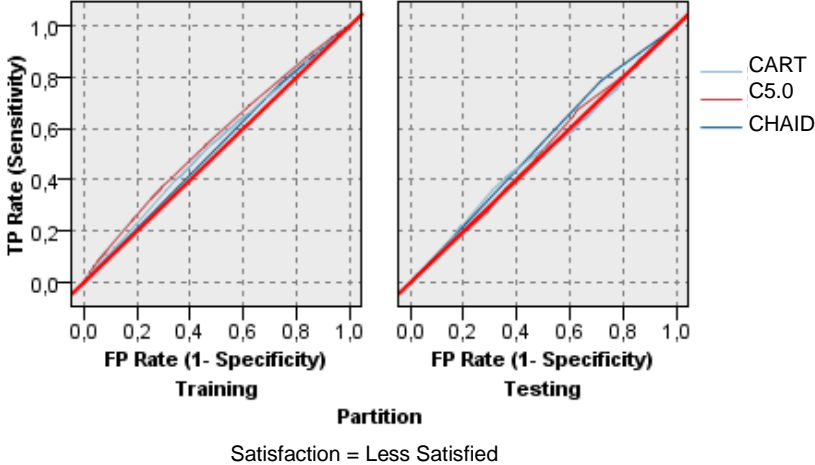




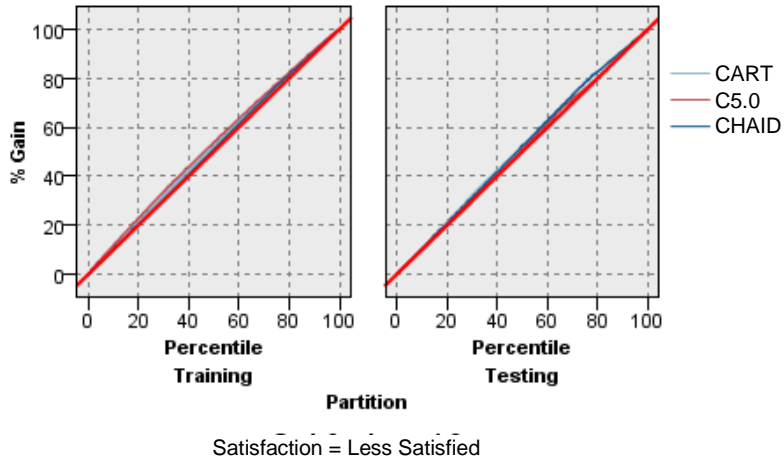
For the pre-travel group of variables (Table 41), the model classified correctly 63.52% of the less satisfied tourists. In other hand, the model predicted 43.42% of the very satisfied tourists. Then, the model predicted correctly the satisfaction of 53.58% of tourists. The models of this group of variables have a lower accuracy and a higher percentage in the correct classification of the less satisfied (precision), due to the data sample (undersample) balance, which reduces the number of very satisfied cases from the sample and then, the overall accuracy. Related to the importance of the pre-travel variables, the ones with a higher importance are BT4 (the nationality and destination currencies are the same) with a predictor importance of 0.54 and BT2.1\_9 (number of information sources used by each) with a predictor importance of 0.46.

The ROC curve (Figure 15) analyses the precision of the model predictions of the pre-travel variables. The AUC (Area Under the ROC Curve) is a measure of the accuracy of the model. The AUC is 0.535 for the CART algorithm, 0.522 for the C5.0 and 0.534 for the CHAID. Therefore, the CHAID algorithm is the best one to predict the satisfaction with the demographic variables. From the gains chart (Figure 16) it is possible to conclude that the model for the pre-travel variables is not very useful, since the model curves are very close to the baseline, as already evidenced by the ROC Curve, where the model curves are also very close to the baseline.

**Figure 15: ROC Curve for the model of the pre-travel variables**



**Figure 16: Gains chart for the model of the pre-travel variables**

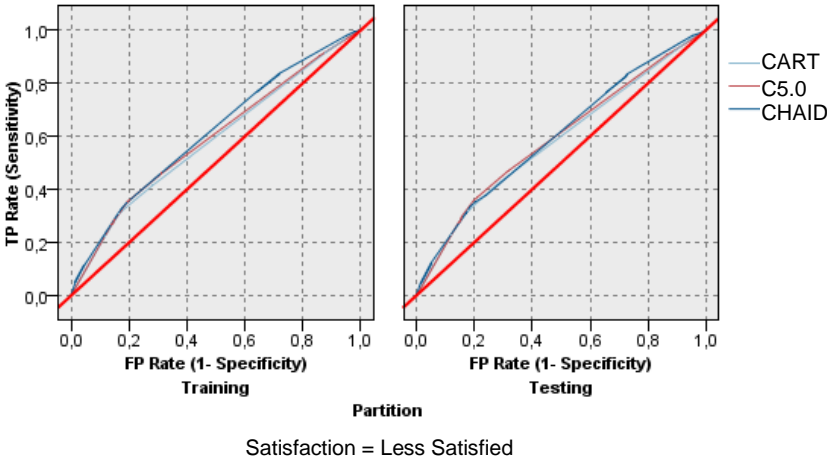


For the on-travel group of variables (Table 42), the model classified correctly 23.23% of the less satisfied tourists. In other hand, the model predicted 83.70 % of the very satisfied tourists. Then, the

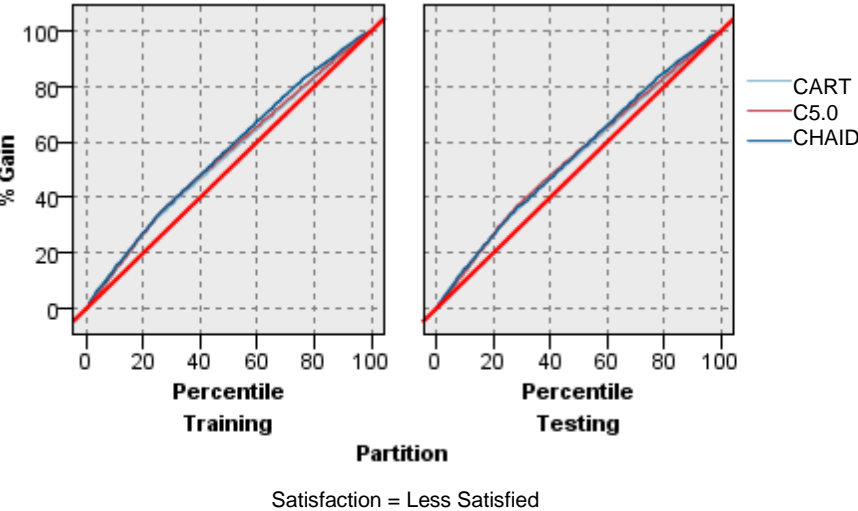
model predicted correctly the satisfaction of 59.29% of tourists, which makes this an efficient and accurate model related to the predictions of the relative importance of the on-travel variables. The variables with a higher relative importance for this group are OT1.2 (own country as destination country) OT1.1 (main holiday destination country) and OT1.1\_1 (area of the European Union that the destination country belongs) with 0.29.

The ROC curve (Figure 17) analyses the precision of the model predictions of the on-travel variables. The AUC (Area Under the ROC Curve) is a measure of the accuracy of the model. The AUC is 0.582 for the CART algorithm, 0.591 for the C5.0 and 0.596 for the CHAID. Then, the CHAID algorithm is the best one to predict the satisfaction with the demographic variables. From the gains chart (Figure 18) is possible to conclude that results are better with the model, since the model curves are apart from the baseline.

**Figure 17: ROC Curve for the model of the on-travel variables**



**Figure 18: Gains chart for the model of the on-travel variables**



Regarding the model with all the variables, the 10-fold validation method was performed. Three models were described: i) model with the highest accuracy; ii) model with the highest percentage of less satisfied tourists correctly classified and; iii) model with the highest percentage of very satisfied tourists correctly classified.

### Model with the highest overall accuracy

Table 43 shows the confusion matrix for the model with the highest overall accuracy. It classified correctly 56.25% of the less satisfied tourists and 71.47% of the very satisfied tourists. Then, the model predicts correctly the satisfaction of 64.30% of the tourists.

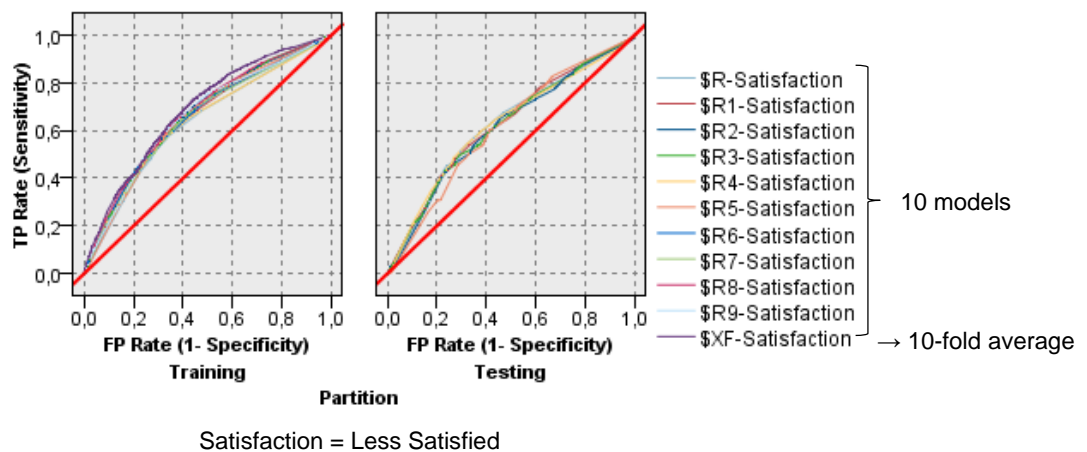
**Table 43: Confusion matrix for the model with the highest accuracy**

		Predicted class	
		CART (A)	Less satisfied
Actual class	Less satisfied	56.25%	43.75%
	Very satisfied	28.52%	71.47%

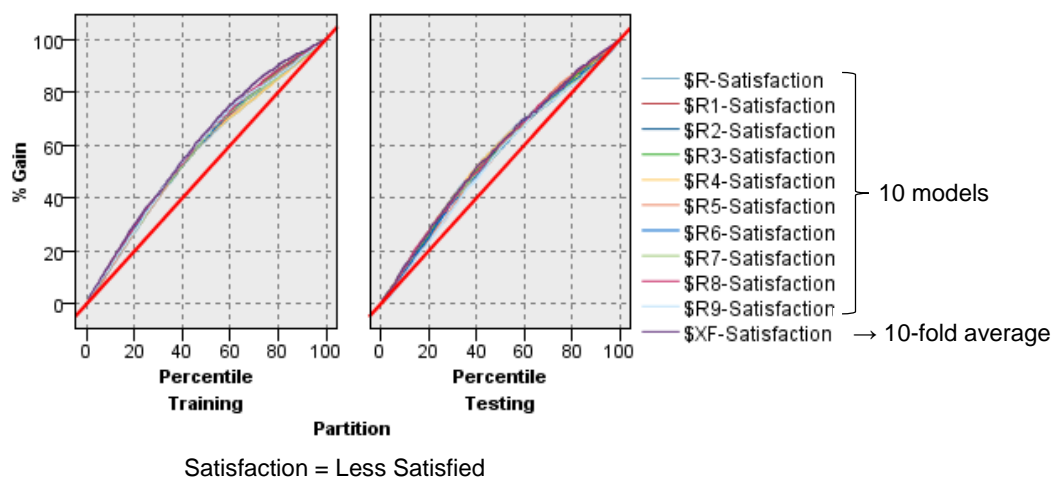
Note: Accuracy = 64.30%

The ROC Curve (Figure 19) analyses the precision of the model with the highest overall accuracy. The AUC (Area Under the ROC Curve) is a measure of the accuracy of the model and it is 0.632 for this model, which is a good value. The gain chart (Figure 20) measures the effectiveness of a model calculated as the ratio between the results obtained with and without the model. As the model curves are apart from the baseline, it means that it is a good model to predict the tourist satisfaction.

**Figure 19: ROC Curve for the model with the highest accuracy**



**Figure 20: Gains chart for the model with the highest accuracy**



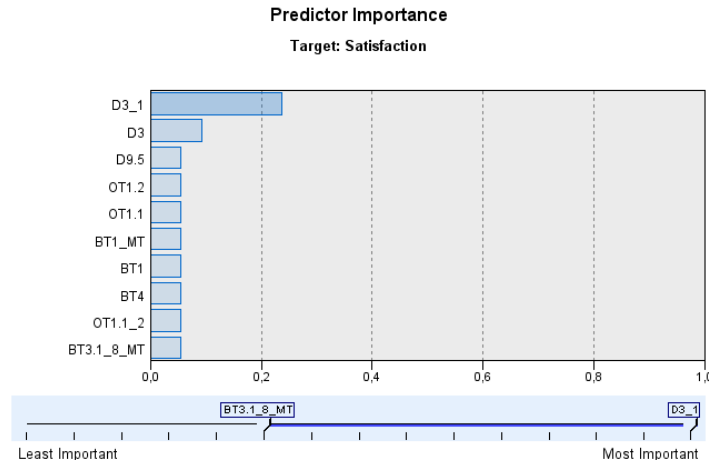
The decision tree (see Appendix chapter: I - Highest Accuracy) for all the variables allows to characterise different tourist profiles for each level of satisfaction (cluster). One of the 10 models built with the 10-fold validation was chosen and some nodes (profiles) were described. For instance:

1. Profiles for less satisfied tourists:
  - Node 6: if the tourist area of the European Union that the country of nationality belongs is the Iberian Peninsula, the Eastern European, the Balkans or the Baltics and the country of nationality is Bulgaria, Greece, Latvia, Lithuania, Poland, Portugal, Cyprus, Croatia, Romania or Spain, the tourist will be less satisfied with a confidence of 62.05% and a support of 23.10% (n=1523);
  - Node 8: if the tourist area of the European Union that the country of nationality belongs is the North European countries or the Central European countries and the country of nationality is Finland, France or Italy and the area of their destination country is the Central E0uropean countries or the own country, the tourist will be less satisfied with a confidence of 60.64% and a support of 9.98% (n=658);
2. Profiles for very satisfied tourists:
  - Node 4: if the tourist area of the European Union that the country of nationality belongs is the North European countries or the Central European countries and the country of nationality is Austria, Belgium, Denmark, Germany, Ireland, Luxembourg, Malta, Netherlands, Sweden or United Kingdom, the tourist will be very satisfied with a confidence of 63.48% and a support of 48.43% (n=3193);
  - Node 5: if the tourist area of the European Union that the country of nationality belongs is the Iberian Peninsula, the Eastern European, the Balkans or the Baltics and the country of nationality is Czech Republic, Estonia, Hungary, Slovakia or Slovenia, the tourist will be very satisfied with a confidence of 57.78% and a support of 13.26% (n=874);
  - Node 15: if the tourist area of the European Union that the country of nationality belongs is the North European countries or the Central European countries and the country of nationality is Finland, France or Italy and the area of the European Union that the destination country belongs is the North European, Iberian Peninsula, Eastern European, Balkans, Baltics or not in the EU and the main reason for going on holiday is wellness/spa/heath treatment, culture or specific events, the tourist will be very satisfied with a confidence of 72.22% and a support of 1.36% (n=90);
3. Profiles for a weak association to one of the two levels of tourist satisfaction:
  - Node 16: if the tourist area of the European Union that the country of nationality belongs is the North European countries or the Central European countries and the country of nationality is Finland, France or Italy and the area of the European Union that the destination country belongs is the North European, Iberian Peninsula, Eastern European, Balkans, Baltics or not in the EU and the main reason for going on holiday is sun/beach, city trips, sport-related activities, nature or visiting family/friends/relatives, the tourist will be less satisfied with a confidence

of 53.33% or very satisfied with a confidence of 46.67% and a support of 3.87% (n=255).

The results of the tree model allowed to identify the attributes that the model considers the most important to predict the tourists' level of satisfaction with the destination and accommodation. These attributes are showed in Figure 21, where the variable D3\_1 (area of the European Union that the country of nationality belongs) is the most important.

**Figure 21: Explanatory attributes of the satisfaction with destination and accommodation**



**Model with the highest percentage of less satisfied tourists correctly classified**

Table 44 represents the confusion matrix for the model with the highest percentage of less satisfied tourists correctly classified. The matrix classified correctly 56.00% of the less satisfied tourists and 70.00% of the very satisfied tourists. Then, the model explains correctly 63.42% of the data.

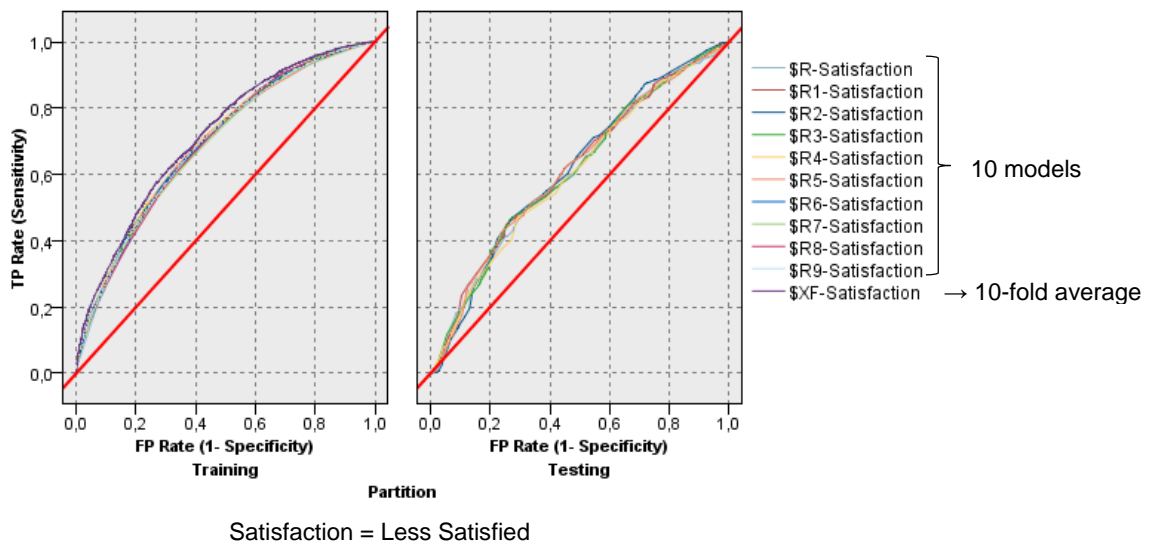
**Table 44: Confusion matrix for the model with the highest percentage of less satisfied tourists correctly classified**

		Predicted class	
		Less satisfied	Very satisfied
Actual class	CHAID		
	Less satisfied	56.00%	44.00%
Very satisfied	30.00%	70.00%	

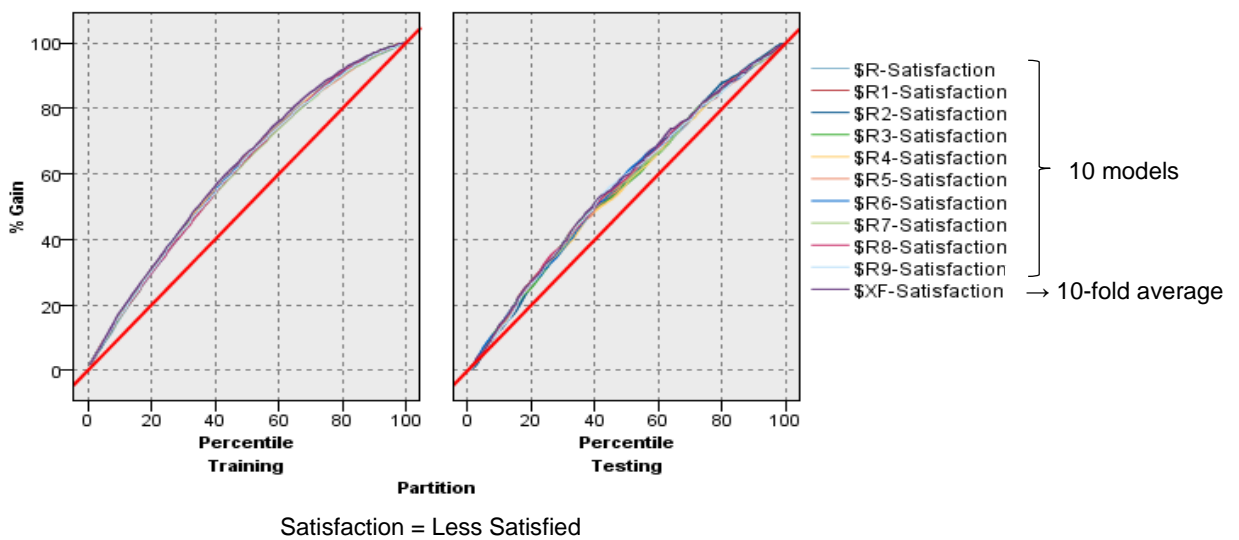
Note: Accuracy = 63.42%

The ROC curve (Figure 22) analyses the precision of the model with the highest percentage of less satisfied tourists correctly classified. The AUC is a measure of the accuracy of the model and it is 0.630 for this model, which is a good value. The gain chart (Figure 23) measures the effectiveness of a model calculated as the ratio between the results obtained with and without the model. As the model curves are apart from the baseline, this is a good model to predict the tourist satisfaction.

**Figure 22: ROC Curve for the model with the highest percentage of less satisfied tourists correctly classified**



**Figure 23: Gains chart for the model with the highest percentage of less satisfied tourists correctly classified**



The decision tree (see Appendix chapter: II - Less Satisfied) for all the variables allows to characterise different tourist profiles for each level of satisfaction (cluster). One of the 10 models built with the 10-fold validation was chosen and some nodes (profiles) were described. For instance:

1. Profiles for less satisfied tourists:

- Node 116: if a tourist's country of nationality is Bulgaria, Finland, France, Italy, or is missing and the currency is Euro or is missing and the type of accommodation (paid or unpaid) is always in paid accommodation or paid and unpaid accommodation and the household size is 2 or 4 or more people and the number of organising methods is one or none, the tourist will be less satisfied with a confidence of 59.32% and a support of 5.95% (n=563);
- Node 70: if a tourist's country of nationality is Bulgaria, Finland, France, Italy, or is missing and the currency is Euro or is missing and the type of accommodation

(paid or unpaid) is never in paid accommodation and is a male, the tourist will be less satisfied with a confidence of 57.07% and a support of 2.17% (n=205);

➤ Node 82: if a tourist's country of nationality is Latvia, Portugal, Romania or Spain, is a female, works as self-employed, employee or is not working and did not have a party experience for any safety issues while using paid accommodation, the tourist will be less satisfied with a confidence of 60.09% and a support of 4.93% (n=466).

2. Profiles for very satisfied tourists:

➤ Node 102: if a tourist's country of nationality is Austria, Denmark, Estonia, Ireland or Slovenia, has more than 40-years, never goes to paid accommodation with less than 20 guests, but goes 0, 1, 2, 4, 5 times or more or is missing to paid accommodation with more than 20 guests and goes to a camping site (tent, motorhome, caravan, etc.) for 0, 1 or 2 times or is missing, the tourist will be very satisfied with a confidence of 83.37% and a support of 4.45% (n=421);

➤ Node 112: if a tourist's country of nationality is Belgium, Czech Republic, Hungary, Luxembourg, Slovakia or Sweden, is a female with more than 25-39 years, used one or none information sources for their main trip and the duration of the personal travels taken in 2014 is between 4 and 13 consecutive nights or more than 13 consecutive nights with a confidence of 60.71% and a support of 6.22% (n=588);

➤ Node 38: if a tourist's country of nationality is Germany, Malta or Netherlands, and the duration of the personal travels taken in 2014 is between 4 and 13 consecutive nights and goes 0, 2, 3 or 5 times or more to paid accommodation with more than 20 guests or is missing, the tourist will be very satisfied with a confidence of 61.78% and a support of 4.62% (n=437);

3. Profiles for a weak association to one of the two levels of tourist satisfaction:

➤ Node 30: if a tourist's country of nationality is Belgium, Czech Republic, Hungary, Luxembourg, Slovakia or Sweden, is a male and did not have a party experience for any safety issues while using paid accommodation, the tourist will be very satisfied with a confidence 51.00% and a support of 7.96% (n=753);

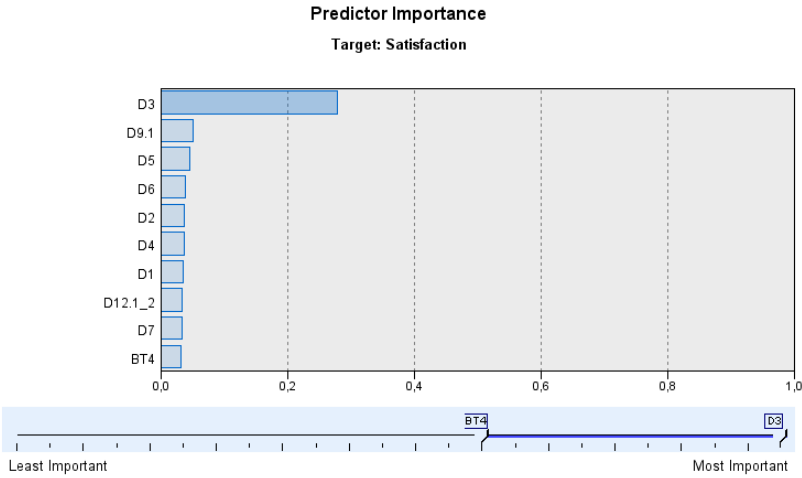
➤ Node 37: if a tourist's country of nationality is Bulgaria, Finland, France, Italy, or is missing and the currency is non-Euro and has more than 25-39 years, the tourist will be very satisfied with a confidence of 57.73% and a support of 4.10% (n=388);

➤ Node 88: if a tourist's country of nationality is Lithuania, Poland, Cyprus or Croatia, stays 0, 1, 2, 3 or 4 times in own property or second home at the destination or is missing, is a female and did not have a party experience for any safety issues while using paid accommodation, the tourist will be less satisfied with a confidence of 51.77% and a support of 4.19% (n=396).

The results of the tree model allowed to identify the attributes that the model considers the most important to predict the tourists' level of satisfaction with the destination and accommodation. These attributes are described in Figure 24, where the variable D3 (nationality) is the most important to

determine the tourists level of satisfaction with the destination and accommodation. The social demographics characteristics and tourist behaviour characterisation variables are the most important ones.

**Figure 24: Explanatory attributes of the satisfaction with destination and accommodation**



**Model with the highest percentage of very satisfied tourists correctly classified**

Table 45 is the confusion matrix for the model with the highest percentage of very satisfied tourists correctly classified. It classified correctly 48.61% of the less satisfied tourists and 72.61% of the very satisfied tourists. Then, the model explains correctly 61.34% of the data.

**Table 45: Confusion matrix for the model with the highest percentage of very satisfied tourists correctly classified**

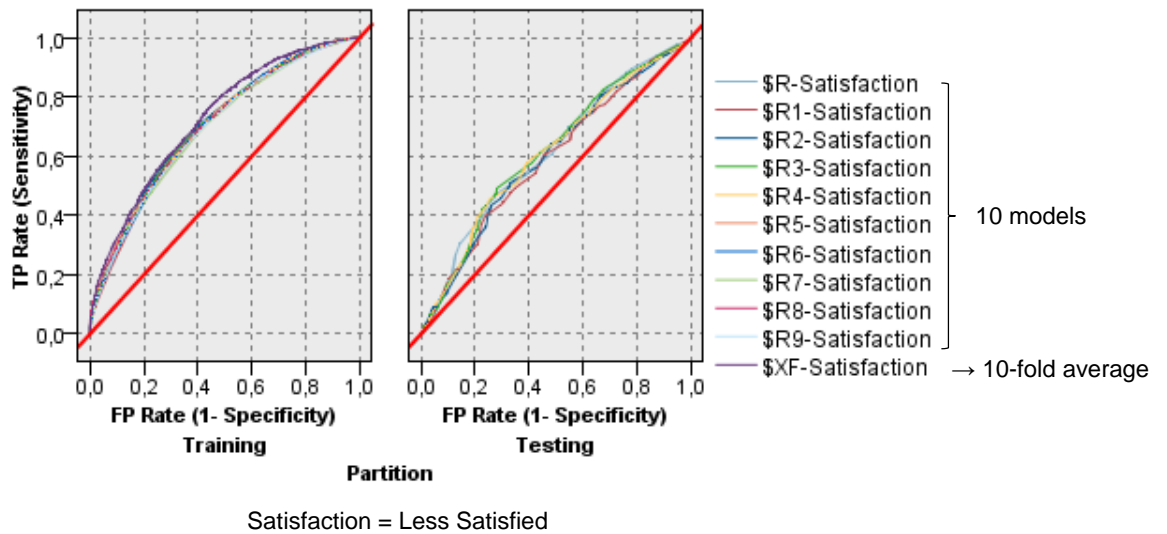
		Predicted class	
		Less satisfied	Very satisfied
Actual class	CART	48.61%	51.39%
	Less satisfied	27.38%	72.61%

Note: Accuracy = 61.34%

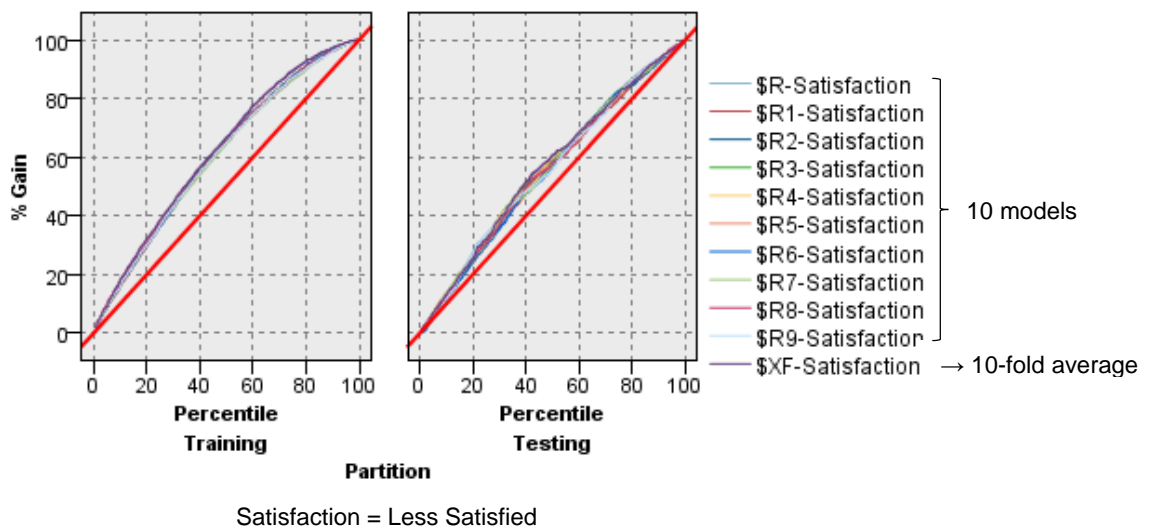
The ROC curve (Figure 25) analyses the precision of the model with the highest percentage of very satisfied tourists correctly classified. The AUC is a measure of the accuracy of the model and it is 0.623 for this model, which is a good value. The gain chart (Figure 26) measures the effectiveness of a model calculated as the ratio between the results obtained with and without the model. As the model curves are apart from the baseline, this is a good model to predict the tourist satisfaction.



**Figure 25: ROC Curve for the model with the highest percentage of very satisfied tourists correctly classified**



**Figure 26: Gains chart for the model with the highest percentage of very satisfied tourists correctly classified**



The decision tree for all the variables allows to characterise different tourist profiles for each level of satisfaction (cluster). One of the 10 models built with the 10-fold validation was chosen and some nodes (profiles) were described. For instance (III - Very Satisfied):

1. Profiles for less satisfied tourists:
  - Node 81: if a tourist's country of nationality is France, Italy, Lithuania, Poland, Cyprus or Croatia and the area of the destination country is the North European countries, the Central European countries or the own country and the type of holiday (package) main trip is not 'all-inclusive' holidays or unknown and the tourist is a male, the tourist will be less satisfied with a confidence of 65.55% and a support of 5.37% (n=508);
  - Node 82: if a tourist's country of nationality is France, Italy, Lithuania, Poland, Cyprus or Croatia and the area of the destination country is the North European countries, the Central European countries or the own country and the type of

holiday (package) main trip is not 'all-inclusive' holidays or unknown and the tourist is a female, the tourist will be less satisfied with a confidence of 59.50% and a support of 7.68% (n=726);

➤ Node 123: if a tourist's country of nationality is Latvia, Portugal, Romania or Spain, is a female, did not had a party experience to any safety issue while using paid accommodation, the type of holiday (package) in 2014 was a combination of types of holiday and never 'all-inclusive holidays' and is a self-employed, employee or not working, the tourist will be less satisfied with a confidence of 62.25% and a support of 4.31% (n=408);

2. Profiles for very satisfied tourists:

➤ Node 27: if a tourist's country of nationality is Austria, Estonia or Slovenia and did not had a party experience to any safety issue while using paid accommodation and is a female, the tourist will be very satisfied with a confidence of 73.75% and a support of 4.43% (n=419);

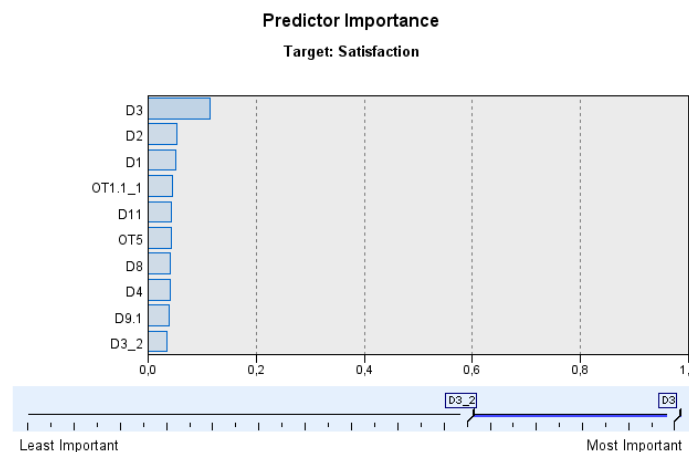
➤ Node 48: if a tourist's country of nationality is Germany, Malta or Netherlands and the holidays are between 4 and 13 consecutive nights, the tourist will be very satisfied with a confidence of 58.00% and a support of 6.67% (n=631);

3. Profiles for a weak association to one of the two levels of tourist satisfaction:

➤ Node 30: if a tourist's country of nationality is Bulgaria, Czech Republic, Hungary, Luxembourg, Slovakia or Sweden and is a male and did not have a party experience to any safety issue while using paid accommodation, the tourist will be very satisfied with a confidence of 52.21% and less satisfied with a confidence of 47.79% and a support of 7.90% (n=747);

The results of the tree allowed to define the attributes that the model considers the most explanatory factors for the tourists' level of satisfaction. These attributes are described in Figure 27, where the variable D3 (nationality) is the most important to determine the tourists level of satisfaction with the destination and accommodation.

**Figure 27: Explanatory attributes of the satisfaction with destination and accommodation**



From the previous analysis, it was possible to conclude that the variables related to the nationality country and area of the European Union that the country of nationality belongs are the ones that most explain the tourist satisfaction with the destination.

#### 4.2.1 Sensitivity Analysis

Sensitivity analysis is the evaluation of the models' main variables. The aim is to analyse which values of the variables predict each type of tourists (Cortez & Embrechts, 2013). The best models are the models with the highest overall accuracy and the models with the highest percentage of less satisfied tourists correctly classified.

For the first model, the predictors with more importance and with significant differences between the values are the D3\_1 (area of the European Union that the country of nationality belongs), D3 (nationality), OT1.1\_2 (currency of the destination country), OT1.1 (main holiday destination country) and BT1\_MT (main reason for going on holiday main trip).

For the area of the European Union that the country of nationality belongs (D3\_1) (Figure 28), the model predicts that the less satisfied tourists are from the countries of the Iberian Peninsula. Also, the Balkans and Eastern European countries have more than 50% of the tourists less satisfied. The North European countries have the highest percentage of very satisfied tourists, followed by the Central European countries and the Baltics.

Regarding the nationality (D3) (Figure 29), tourists from Portugal, Romania, Spain, Latvia, Greece, Lithuania, Croatia, Cyprus, Poland, Bulgaria, Italy, France and Finland tend to be predicted as the less satisfied. Tourists from Austria, Denmark, Estonia, Ireland, Malta, Netherlands, United Kingdom, Hungary, Slovenia and Slovakia tend to be classified as very satisfied.

For the currency of the destination country (OT1.1\_2) (Figure 30), when the currency of the destination country is Euro, the tourists tend to be predicted as less satisfied. For the main holiday destination country (OT1.1) (Figure 31), when the main holiday destination country is in the own country, the tourists tend to be predicted as less satisfied.

For the main reason for going on holiday main trip (BT1\_MT) (Figure 32), when the main reasons for going on holiday (main trip) are city trips, visiting family/friends/relatives or wellness/spa/health treatment, the tourists tend to be predicted as less satisfied. When the main reasons for going on holiday main trip are specific events (sporting events/festivals/clubbing), sport-related activities and culture, the tourists tend to be predicted as very satisfied.

Figure 28: Sensitivity analysis for the model with highest accuracy (variable D3\_1)

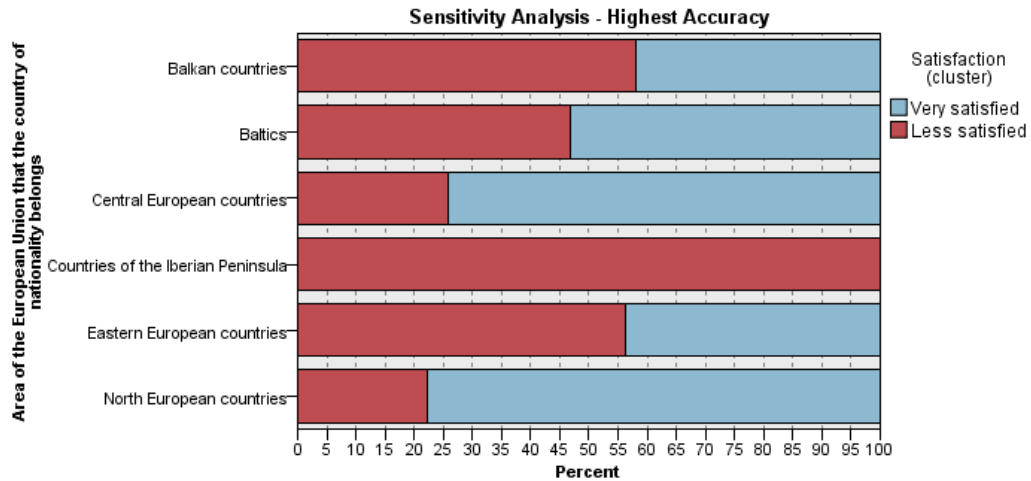


Figure 29: Sensitivity analysis for the model with highest accuracy (variable D3)

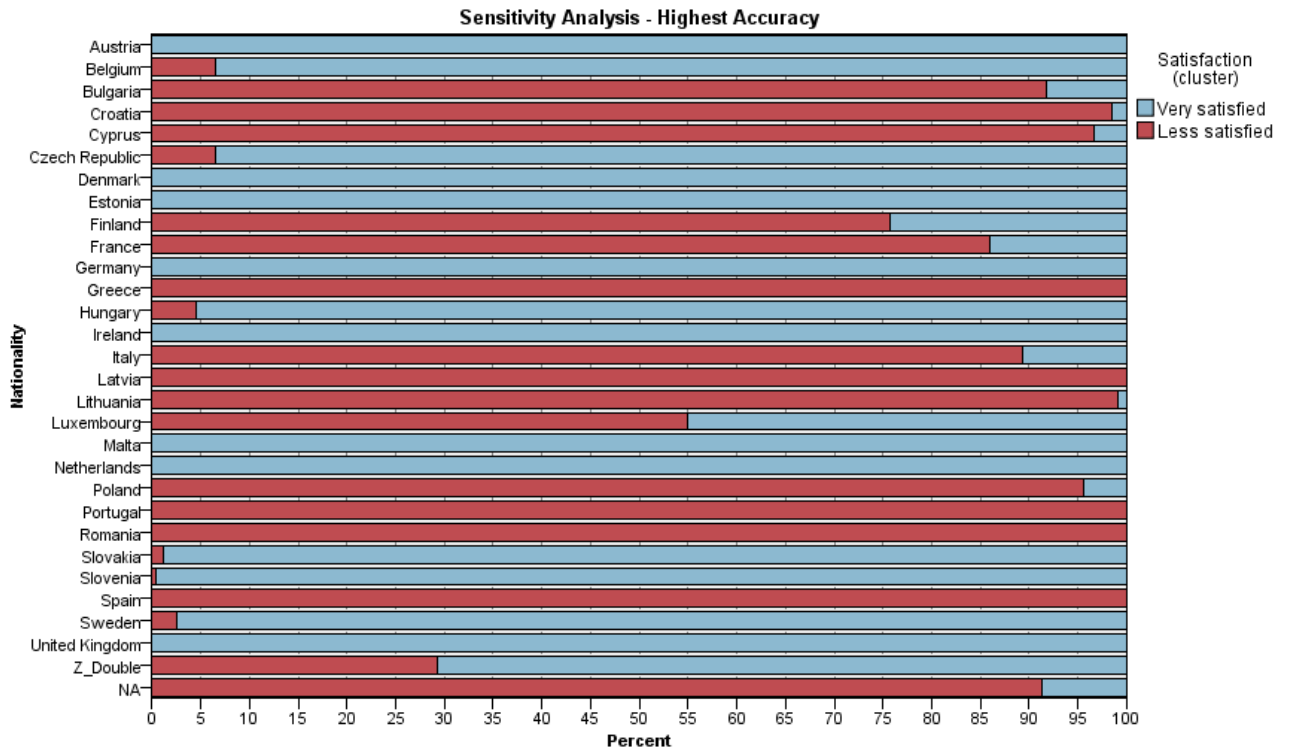
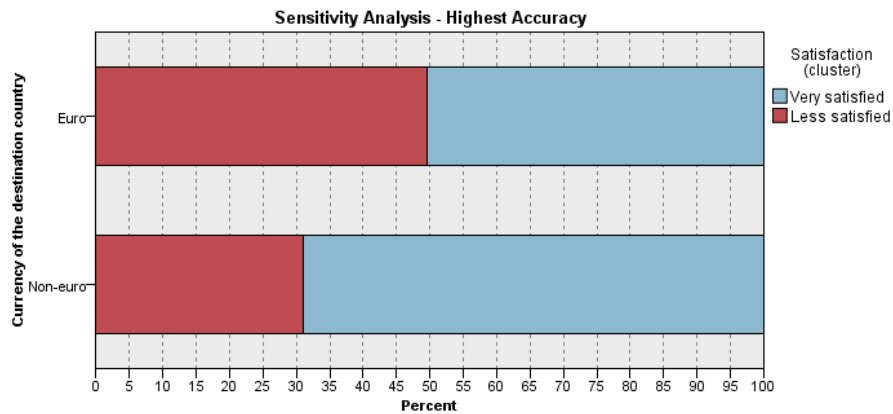
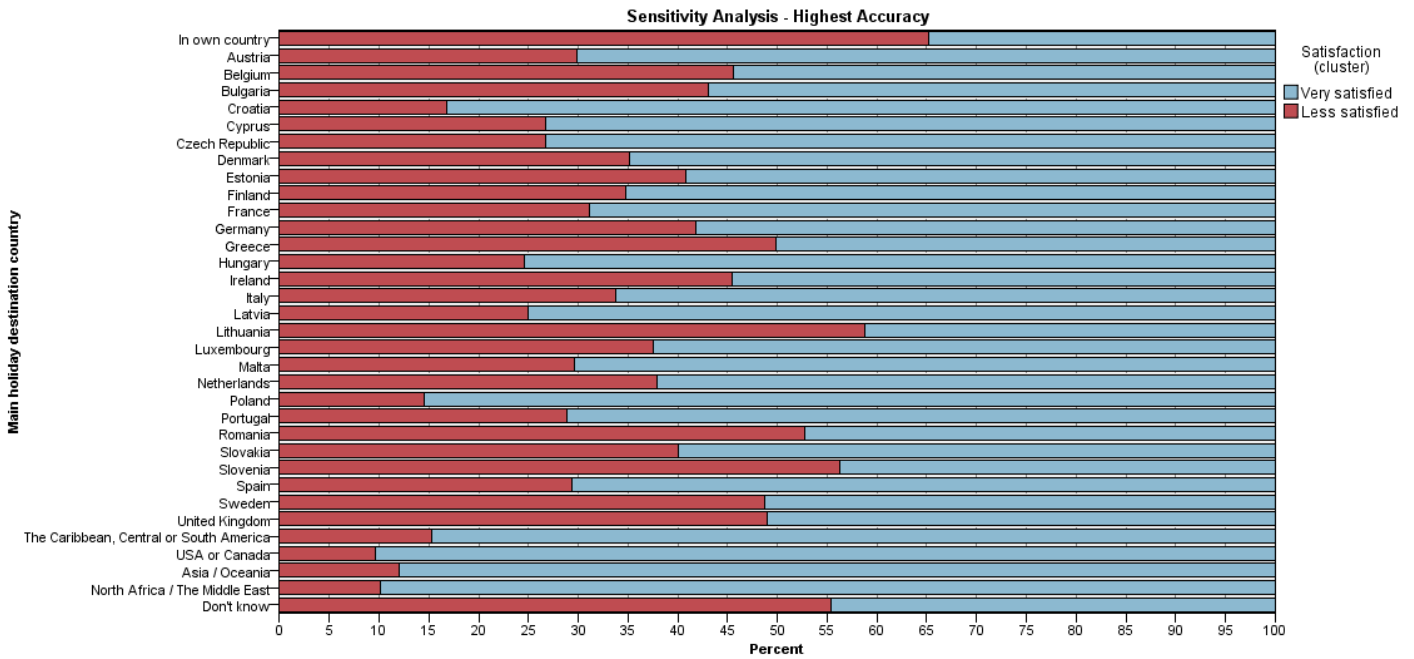


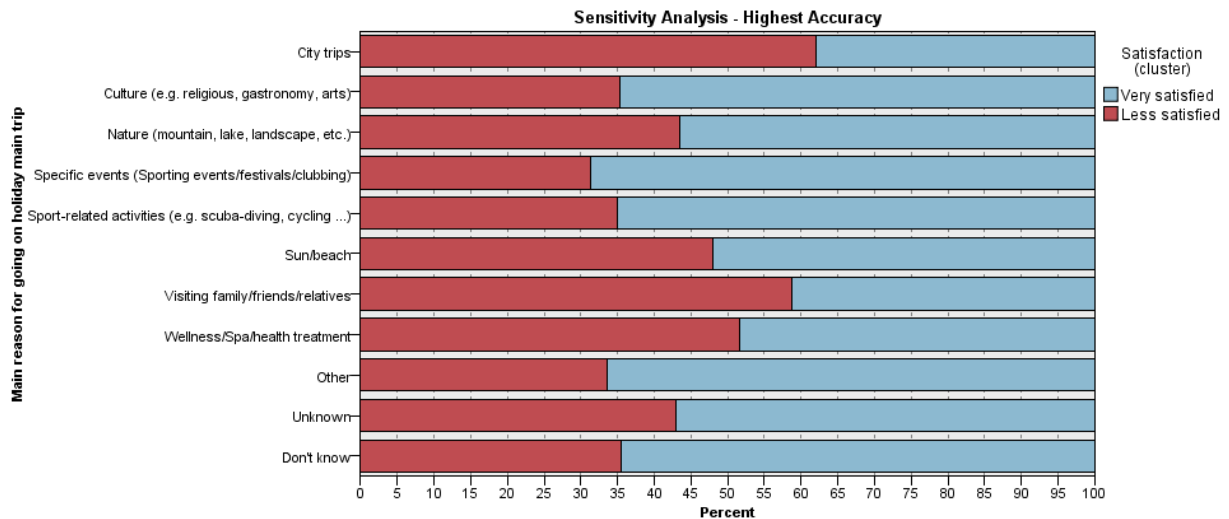
Figure 30: Sensitivity analysis for the model with highest accuracy (variable OT1.1\_2)



**Figure 31: Sensitivity analysis for the model with highest accuracy (variable OT1.1)**



**Figure 32: Sensitivity analysis for the model with highest accuracy (variable BT1\_MT)**



For the second model, the predictors with more importance and with significant differences between the values are the D3 (nationality), D6 (household size), D2 (gender) and D1 (age).

Regarding the nationality (D3) (Figure 33), the model predicted the similar countries of the previous model to the less satisfied tourists, where the differences are only about the percentage of satisfaction. Tourists from Latvia, Greece, Lithuania, Spain, Portugal, Croatia, Bulgaria, Romania, Poland, Cyprus, Italy, France and Finland tend to be predicted as less satisfied. Tourists from Malta, Slovenia, Ireland, Netherlands, Germany, Estonia, Denmark and United Kingdom tend to be predicted as very satisfied.

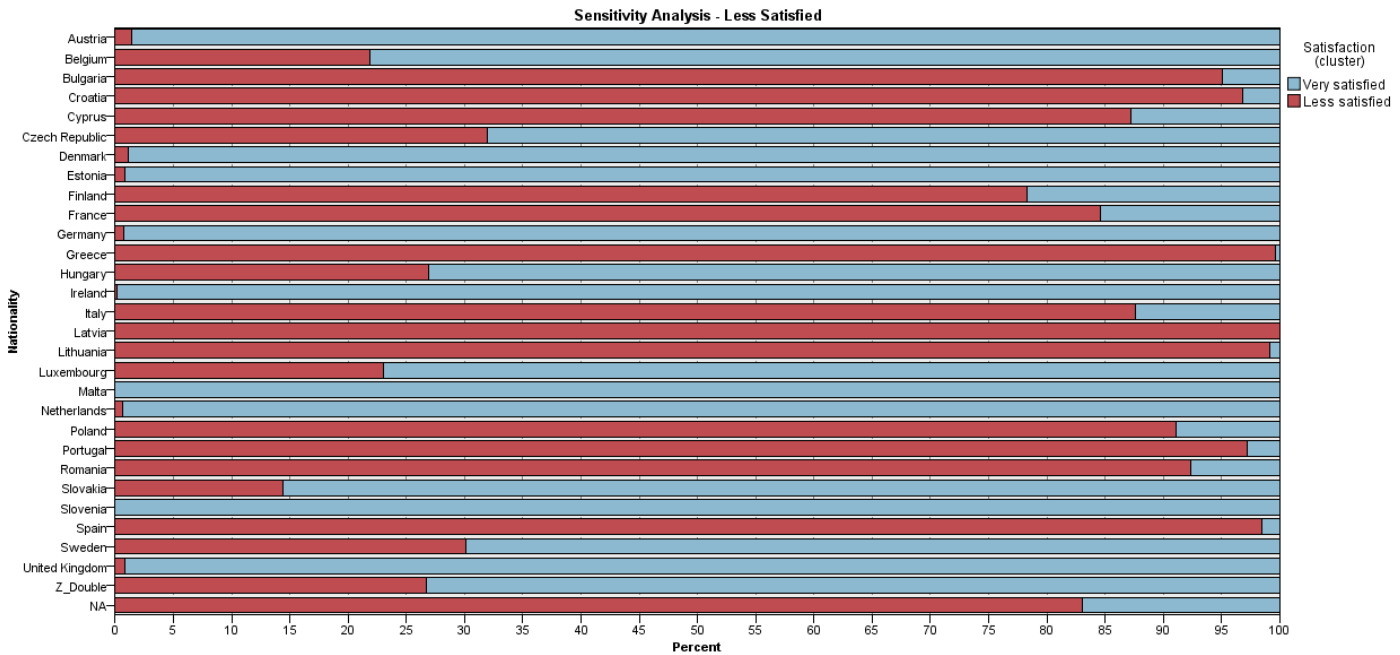
For the household size (D6) (Figure 34), as higher the household size, higher the percentage of less satisfied tourists is, i.e., the model predicted that tourists with a higher household size tend to be

predicted as less satisfied and the tourists with a small household size tend to be predicted as very satisfied.

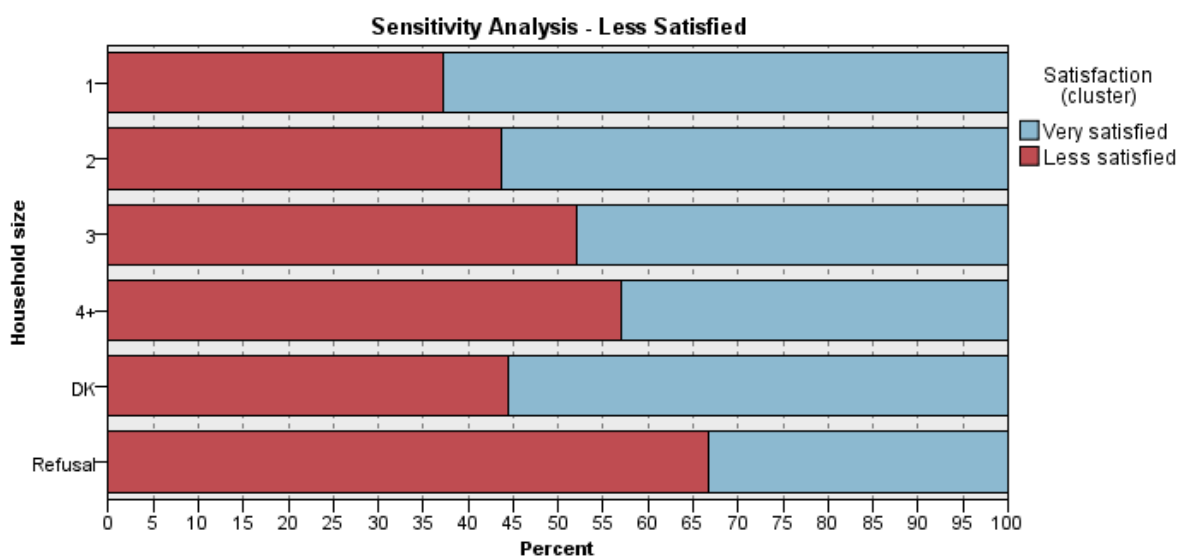
For the gender (D2) (Figure 35), the male tourists tend to be predicted as less satisfied and female tourists tend to be predicted as very satisfied.

At last, for the Age (D1) (Figure 36), as the age increases, the satisfaction also increases. In other words, older tourists tend to be predicted as very satisfied and younger tourists tend to be predicted as less satisfied.

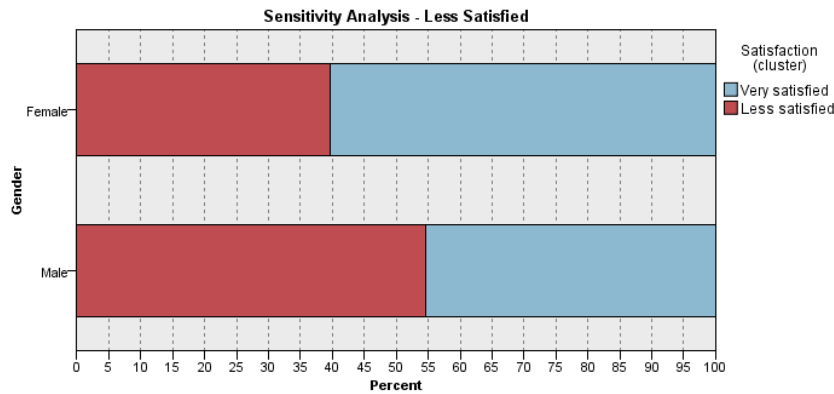
**Figure 33: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D3)**



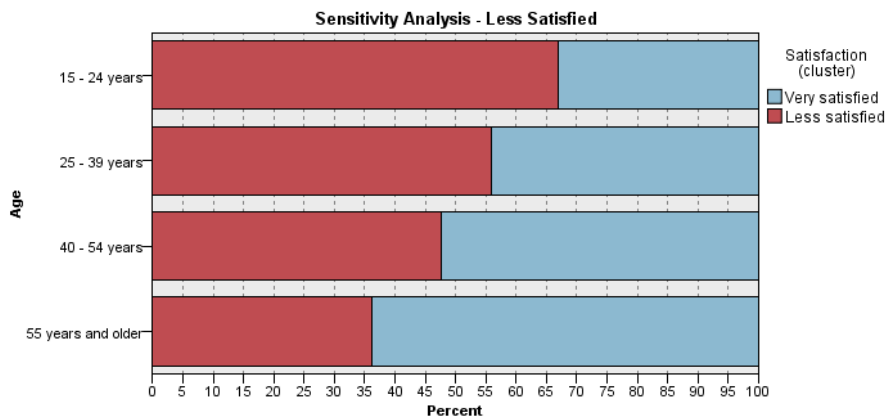
**Figure 34: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D6)**



**Figure 35: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D2)**



**Figure 36: Sensitivity analysis for the model with the highest percentage of less satisfied tourists correctly classified (variable D1)**



### 4.3 Discussion

The results obtained from the models allowed to conclude that the demographic characteristics of the tourist are the most important predictors of satisfaction. This happens because the inherent characteristics of the tourist are the factors that allow to choose the travel services (pre-travel) and experiences (on-travel), i.e., the tourist characteristics will influence in the pre-travel and on-travel choices, which will influence satisfaction. In other words, the nationality/culture, age, gender will lead to different choices and different ways to appreciate and enjoy the travel. Comparing two people from different countries, who have different cultures, with all the other variables being the same, one tourist can be very satisfied and the other be less satisfied, just because of the different nationalities.

As concluded by Frías et al. (2011), tourists from countries with a high uncertainty avoidance culture, such as France, Belgium and Italy, have a better destination image and then, a higher level of satisfaction. Countries where the level of uncertainty avoidance is lower, such as United Kingdom, the destination image is worse and then, the satisfaction level is lower. However, these results are majority in disagreement with this study conclusions, as France and Italy are countries with more than 85% of less satisfied tourists and United Kingdom is a country with almost all the tourists classified as very satisfied tourists, which is in discordance with the results of the study. Belgium is a country with the most of tourists very satisfied, which is in accordance with the study results, being the only country that this

happens. Lu et al. (2015) also concluded that there was variances in satisfaction, caused by the place of residence (nationality), which is also reinforced by the results of this study.

The results of this study suggest that there is a positive relation between age and satisfaction, which reinforces the conclusions of Ragavan et al. (2014), although it is not in accordance with the conclusions of Lu et al. (2015), that suggests that there is a negative relation, and Sarra et al. (2015), that suggests that there is no relation between both variables.

In terms of relation between gender and satisfaction, Ragavan et al. (2014) only concluded that there was a relation, but the results of this study can lead to a conclusion that males are less satisfied than females.

In short, the results of this study are important to discuss the conclusions of various authors mentioned throughout the chapters.



## 5 Conclusion

### 5.1 Synthesis

The objectives of this research were to evaluate the level of satisfaction with the destination and accommodation and identify the explanatory factors of the satisfaction with the destination and accommodation.

These goals were defined as data mining goals, with a segmentation problem to identify tourists' profiles with different levels of satisfaction with the destination and with a classification problem to predict satisfaction with a low prediction error rate and also to identify which predictors are more important. The methodology used was the CRISP-DM, since it is focused on the business approach.

In order to perform the models, the data was prepared, including the exclusion, recoding and creation of variables. After that, the k-means algorithm was performed to achieve the first goal. From this analysis, three clusters of tourists were defined: not very satisfied, satisfied and very satisfied tourists. Then, the demographic and behaviour profile of each group of tourists was identified.

Afterwards, given that to the tourism industry the most important is to differentiate between the less satisfied and the very satisfied tourists, to reach the second goal, predictive models of satisfaction were created based on the social demographic characteristics and tourist behaviour variables, but also on the pre-travel and on-travel behaviours. The two groups of tourists (less satisfied and the very satisfied) were used to perform decision tree algorithms, namely, CART, C5.0 and CHAID. The data set was divided using the holdout and the 10-fold methods. To both methods, the resulting models were evaluated through several metrics, such as, overall accuracy, precision, and specificity. Also, the ROC curve and Gains chart were performed. Through the ROC curve was possible to conclude that the algorithm with the highest accuracy is the CHAID algorithm, followed by the CART and C5.0 algorithms. Moreover, to increase accuracy the bagging, boosting and balanced samples methods were performed in some models. The bagging and boosting methods had not demonstrated a significant improvement in the models' accuracy. However, the balanced method allowed to predict the less satisfied tourists, which was not possible without this method.

The most efficient model, i.e., with the highest accuracy, allowed to classify correctly more than 60% of the tourists, which is slightly above the data mining goal defined, and allowed to identify the most important variables to predict satisfaction, which were the demographic and tourist behaviour variables (e.g. tourist nationality), followed by the on-travel and the pre-travel variables. In this way, the objectives were achieved successfully.

### 5.2 Contributions

Tourism is a growing industry and it is one of the most income sources for several countries. However, there is the need to focus on the data available from the tourism industry, in order to provide knowledge and develop systems that can help tourism companies to grow and be able to promote efficient marketing campaigns to the correct target and improve tourists' satisfaction and loyalty.

This research intends to increase the attention and interest of the tourism professionals to the importance of the study of the factors that can influence tourist satisfaction and consequently increase the loyalty, which benefits the tourism at the destination.

Therefore, this dissertation, by doing an extensive literature review related to the tourists' satisfaction with the destination and the factors that may explain it, provides to the investigators and professionals the acquisition of knowledge of the principal studies related to this subject.

In other hand, by verifying the availability of data related with tourists' behaviour and satisfaction in European Union that was not properly analysed and explored in the attempt to identify patterns in the data, this investigation illustrates to the academics and professionals that the application of DM techniques in the tourism sector can generate useful knowledge.

For the academics, this study demonstrates that different DM algorithms for classification problems, generate different outcomes and by that, enforces the need for the analysis to be done with the support of different DM techniques and algorithms.

For the professionals, this study demonstrates that in a sample of more than 10.000 European tourists, there is a high percentage of tourists that is not very satisfied with the destination of their trips, and that, in between more than 100 potential predictors, the variables that most influence the satisfaction, are mainly the demographic characteristics, highlighting the nationality.

By including all the different behaviour of tourists that they have from the preparation of the travel (pre-travel phase) to the return to their home (post-travel phase), makes this investigation one of the few (if it not the first, to my best knowledge), that evaluates attributes from the three phases of the travel in the same satisfaction analysis, allowing to compare the relative weight of each one of them in the satisfaction.

In more specific terms, this investigation presented the results of various algorithms of classification with different parameterizations, having included in some of them, techniques still not very used in studies, which happens due to lack of knowledge of investigators of these techniques, and by other hand, due to the lack of integration of these techniques in a lot of Analytics software, especially older versions. Therefore, this investigation results in a major contribution for investigators in the field of Business Analytics, and also for students, possibly turning out to be a consultation tool for them, in the course of their studies.

Finally, the good results achieved, in an issue with a particular complexity as it is the prediction of a tourist's satisfaction, may contribute for public and private organizations to consider making available its data to the academic institutes, so that these with their know-how, could work on the data and, by that, generate useful knowledge for the organizations.

### 5.3 Limitations of the Research

It is important to notice that there are some limitations in this research (as usual in all researches), namely:

- Although the sample size is not small, the data available can be insufficient to identify all the patterns in the tourists' behaviours;
- The data was available from a survey, which has some inherent problems. Also, the questions were limited and some tests may not have been made, but it was the best way to have a large sample and with a diversified geography;
- Some of the attributes proposed by the state of the art were not used due to the lack of them in the available data;
- Since the software chosen was the IBM SPSS Modeler, one of the most used software in the world to data modelling, the analysis are a result of the potential of this tool for each technique.

### 5.4 Further Research

This investigation provides several possibilities for further research. Although it was possible to define the major variables that have influence in the tourist satisfaction, it is important to distinguish further researches in the tourism industry area, and also with the same data, such as:

- An improvement in the DM techniques with the implementation of new parameters and the application of other techniques, such as, self-organising maps or two-step clustering for the first research goal, and logistic regression, neuronal networks and support vector machines for the second one;
- Find data with the variables not studied in this project, which can also be important to predict tourist satisfaction;
- Predict the number of complaints made through regression models;
- And, predict loyalty (repeat destination) with satisfaction and all variables of the three phases of the travel.



## 6 References

- Abodeeb, J. A. (2014). *Destination image from an Arab perspective: the influence of culture on perceptions of and motivations for visiting the Gold Coast, Australia* (PhD thesis, Southern Cross University, Australia). Retrieved September 13, 2016, from <http://epubs.scu.edu.au/theses/373/>.
- Acock, A. C., & Stavig, G. R. (1979). A Measure of Association for Nonparametric Statistics. *Social Forces*. 57 (4). pp. 1381-1386. doi: 10.1093/sf/57.4.1381.
- Aghdam, A. R., Kamalpour, M., Chen, D., & Sim, A. T. H. (2013). Finding Interesting Places at Malaysia: A Data Mining Perspective. In R. Neck (Ed.), *Mathematics and Computers in Contemporary Science* (pp. 89-93). Location: WSEAS Press.
- Akhoondnejad, A. (2015). Analyzing the Pre-Travel, On-Travel, and Post-Travel Behaviors of Iran's First-Time Visitors. *Journal of Travel & Tourism Marketing*. 32 (2). pp. 1023-1033. doi: 10.1080/10548408.2014.957796.
- Alegre, J. & Garau, J. (2010). Tourist Satisfaction and Dissatisfaction. *Annals of Tourism Research*. 37 (1). pp. 52-73. doi:10.1016/j.annals.2009.07.001.
- Amorim, R. C., Makarenkov, V., & Mirkin, B. (2016). A-Wardpβ: Effective hierarchical clustering using the Minkowski metric and a fast k -means initialisation. *Information Sciences*. 370-371. pp. 343-354. doi: 10.1016/j.ins.2016.07.076.
- Araslı, H. & Baradarani, S. (2014). European Tourist Perspective on Destination Satisfaction in Jordan's industries. *Procedia - Social and Behavioral Sciences*. 109. pp. 1416-1425. doi: 10.1016/j.sbspro.2013.12.644.
- Ayoubloo, M. K., Azamathulla, H. Md., Jabbari, E. & Zanganeh, M. (2011). Predictive model-based for the critical submergence of horizontal intakes in open channel flows with different clearance bottoms using CART, ANN and linear regression approaches. *Expert Systems with Applications*. 38 (6). pp. 10114-10123. doi: 10.1016/j.eswa.2011.02.073.
- Baloglu, S. & McCleary, K. W. (1999). A model of destination image formation. *Annals of Tourism Research*. 26 (4). pp. 868-897. doi: 10.1016/S0160-7383(99)00030-4.
- Barros, C. P. & Machado, L. P. (2010). The length of stay in tourism. *Annals of Tourism Research*. 37 (3). pp. 692-706. doi: 10.1016/j.annals.2009.12.005.
- Beerli, A. & Martín, J.D. (2004). Factors influencing destination image. *Annals of Tourism Research*. 31 (3). pp. 623-636. doi: 10.1016/j.annals.2004.01.010.
- Beerli, A., Meneses, G. D. & Gil, S. M. (2007). Self-congruity and destination choice. *Annals of Tourism Research*. 34 (3). pp. 571-587. doi:10.1016/j.annals.2007.01.005.
- Bigné, J. E., Sánchez, M. I. & Sánchez, J. (2001). Tourism image, evaluation variables and after purchase behaviour: inter-relationship. *Tourism Management*. 22 (6). pp. 607-616. doi: 10.1016/S0261-5177(01)00035-8.
- Bose, I. (2009). Data Mining in Tourism. In *Encyclopedia of Information Science and Technology*. (2nd ed., pp. 936-940). Hong Kong: IGI Global.
- Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. (1984). *Classification and Regression Trees*. Wadsworth: Taylor & Francis.
- Breiman, L. Bagging Predictors. (1996). *Machine Learning*. 24 (2). pp. 123-140. doi: 10.1023/A:1018054314350.
- Brezigar-Masten, A. & Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications*. 39 (11). pp. 10153-10159. doi:10.1016/j.eswa.2012.02.125.

- Chapman, P. & Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*. USA: SPSS Inc.
- Chen, C. & Tsai, D. (2007). How destination image and evaluative factors affect behavioral intentions? *Tourism Management*. 28 (4). pp. 1115-1122. doi: 10.1016/j.tourman.2006.07.007.
- Chen, C., Chen, S. H. & Lee, H. T. (2011). The destination competitiveness of Kinmen's tourism industry: exploring the interrelationships between tourist perceptions, service performance, customer satisfaction and sustainable tourism. *Journal of Sustainable Tourism*. 19 (2). pp. 247-264. doi: 10.1080/09669582.2010.517315.
- Chen, Y., Schuckert, M., Song, H. & Chon, K. (2015). Why Can Package Tours Hurt Tourists? Evidence from China's Tourism Demand in Hong Kong. *Journal of Travel Research*. 55 (4). pp. 427-439. doi: 10.1177/0047287515612597.
- Chi, C. G. & Qu, H. (2009). Examining the Relationship Between Tourists' Attribute Satisfaction and Overall Satisfaction. *Journal of Hospitality Marketing & Management*. 18 (1). pp. 4-25. doi: 10.1080/19368620801988891.
- Cohen, D. (1988). Teaching Practice: Plus Ça Change. In Jackson, P. W. (Ed.), *Contributing to Educational Change: Perspectives on Research and Practice*. National Society for the Study of Education Series on Contemporary Issues. Berkeley, CA: McCutchan.
- Cohen, S. A., Prayag, G. & Moital, M. (2014) Consumer behaviour in tourism: Concepts, influences and opportunities. *Current Issues in Tourism*. 17 (10). pp. 872-909. doi: 10.1080/13683500.2013.850064.
- Cooper, C., Fletcher, J., Gilbert, D. & Wanhill, S. (2007). *El turismo: Teoría y practica*. Madrid: Editorial Síntesis, S.A.
- Cortez, P., & Embrechts, M. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*. 225. pp. 1-17. doi: 10.1016/j.ins.2012.10.039.
- Crompton, J.L. (1979). Motivations for pleasure vacation. *Annals of tourism Research*. 6 (4). pp. 408-424. doi: 10.1016/0160-7383(79)90004-5.
- Delen, D., Kuzey, C. & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*. 40 (10). pp. 3970-3983. doi: 10.1016/j.eswa.2013.01.012.
- Dewar, K., Li, W. M. & Davis, C. H. (2007). Photographic Images, Culture, and Perception in Tourism Advertising: A Q Methodology Study of Canadian and Chinese University Students. *Journal of Travel & Tourism Marketing*. 22 (2). Pp. 35-44. doi: 10.1300/J073v22n02-03.
- Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys*. 27 (3). pp. 326-327. doi: 10.1145/212094.212114.
- Dolinting, E. E., Yusof, A. & Soon, C. C. (2015). Application of Push & Pull Theory in Island Sport Tourism: A Study of Sipadan Island, Sabah. *Journal of Physical Education and Sport*. 15 (2). 295-304.
- Dunteman, G. H. (1989). *Principal components analysis*. 69. Sage.
- Echtner, C. M. & Ritchie, J. R. B. (2003). The Meaning and Measurement of Destination Image. *The Journal of Tourism Studies*. 14 (1). pp. 37-46.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- European Commission (2015a). *Flash Eurobarometer 414* (Preferences of Europeans towards Tourism, 2015). Brussels: European Union. doi: 10.2873/632880.

- European Commission (2015b). *Flash Eurobarometer 414* (Preferences of Europeans towards Tourism, 2015). TNS Political & Social [producer]. GESIS Data Archive, Cologne. ZA6285 Data file Version 1.0.0. doi:10.4232/1.12344.
- Field, A. P. (2005). *Discovering Statistics using SPSS* (2nd ed.). London: Sage.
- Freund, Y. & Schapire, R. (1996) Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*. pp. 1-9.
- Frías et al., D. M., Rodríguez, M. A., Castañeda, J. A., Sabiote, C. M. & Buhalis, D. (2011). The Formation of a Tourist Destination's Image via Information Sources: the Moderating Effect of Culture. *International Journal of Tourism Research*. 14 (5). pp. 437-450. doi: 10.1002/jtr.870.
- Garau-Vadell, J. B, Diaz-Armas, R. & Gutierrez-Tano, D. (2014). Residents' Perceptions of Tourism Impacts on Island Destinations: A Comparative Analysis. *International Journal of Tourism Research*. 16 (6). pp. 578-585 doi: 10.1002/jtr.1951.
- Gartner, W.C. (1993). Image formation process, *Journal of Travel and Tourism Marketing*. 2 (2/3). pp. 191-215. doi: 10.1300/J073v02n02\_12.
- Gazley, A. & Watling, L. (2015). Me, My Tourist-Self, and I: The Symbolic Consumption of Travel. *Journal of Travel & Tourism Marketing*. 32 (6). pp. 639-655. doi:10.1080/10548408.2014.954690.
- Goeldner, C.H. & Ritchie, J.R.B. (2009). *Tourism: Principles, Practices and Philosophies* (11th ed.). New Jersey: John Wiley & Sons, Inc.
- Goossens, C. (2000). Tourism information and Pleasure Motivation. *Annals of Tourism Research*. 27 (2). pp. 301-321. doi: 10.1016/S0160-7383(99)00067-5.
- Grossman, R., Bailey, S., Ramu, A., Malhi, B., Hallstrom, P., Pulleyn, I., & Qin, X. (1999). The management and mining of multiple predictive models using the predictive modeling markup language. *Information and Software Technology*. 41 (9). pp. 589-595. doi:10.1016/S0950-5849(99)00022-1.
- Gunn, C. A. (1988). *Vacationscape: Designing tourist regions*. New York: Van Nostrand Reinhold.
- Hair, J, Anderson, R., Tatham, R. & Black, W. (1998). *Multivariate Data Analysis*. Upper Sadle River, NJ: Prentice Hall.
- Hanlan, J., & Kelly, S. (2005). Image Formation, Information Sources and an Iconic Australian Tourist Destination. *Journal of Vacation Marketing*. 11 (2). 163-177. doi: 10.1177/1356766705052573.
- He, H. & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*. 21 (9). pp. 1263-1284.
- Hernández-Méndez, J., Muñoz-Leiva, F. & Sánchez-Fernández, J. (2013). The influence of e-word-of-mouth on travel decision making: consumer profiles. *Current Issues in Tourism*. 18 (11). pp. 1001-1021. doi: 10.1080/13683500.2013.802764.
- Hidalgo, M. C. & Hernandez, B. (2001). Place attachment: Conceptual and empirical questions. *Journal of Environmental Psychology*. 21 (3). pp. 273-281. doi: 10.1006/jevp.2001.0221.
- Hosany, S. & Martin, D. (2012). Self-image congruence in consumer behaviour. *Journal of Business Research*. 65 (5). pp. 685-691. doi:10.1016/j.jbusres.2011.03.015.
- Hosany, S. & Prayag, G. (2013). Patterns of tourists' emotional responses, satisfaction, and intention to recommend. *Journal of Business Research*. 66 (6). pp. 730-737. doi: 10.1016/j.jbusres.2011.09.011.
- IBM. (2016a). *IBM SPSS Modeler 17 Modeling Nodes*. Retrieved July 26, 2016, from IBM Web site: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.0/en/ModelerModelingNodes.pdf>.

- IBM. (2016b). *Cumulative Gains and Lift Charts*. Retrieved September 11, 2016, from IBM Web site: [http://www.ibm.com/support/knowledgecenter/SSLVMB\\_24.0.0/spss/tutorials/mlp\\_bankloan\\_outputtype\\_02.html](http://www.ibm.com/support/knowledgecenter/SSLVMB_24.0.0/spss/tutorials/mlp_bankloan_outputtype_02.html).
- Ioniță, I. (2015). Data mining for predicting the military career choice. *Revista academiei fortelor terestre*. 20 (3). pp. 297-306.
- Jaafar, M. & Khoshkam, M. (2014). *Length of Stay and Tourism Facility Assessment: The Viewpoint of Malaysian Tourists at Langkawi*. Proceedings of the SHS Web of Conferences 12. pp. 1-9. doi: 10.1051/shsconf/20141201031.
- Jarvis, D., Stoeckl, N. & Liu, H. (2016). The impact of economic, social and environmental factors on trip satisfaction and the likelihood of visitors returning. *Tourism Management*. 52. pp. 1-8. doi: 10.1016/j.tourman.2015.06.003.
- Juwattanasamran, P, Supattranuwong, S. & Sinthupinyo, S. (2013). Applying Data Mining to Analyze Travel Pattern in Searching Travel Destination Choices. *The International Journal Of Engineering And Science*. 2 (4). pp. 38-44.
- Kachniewska, M. (2013). Towards the Definition of a Tourism Cluster. *Journal of Entrepreneurship Management and Innovation (JEMI)*. 9 (1). pp. 33-56.
- Kaiser, H.F. (1974). An index of factorial simplicity. *Psychometrika*. 39. pp. 31-36. doi: 10.1007/BF02291575.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*. 29 (2). pp. 119-127.
- Khan, M. A. (2014). A broad view of prospects of tourism industry with reference to India. *Pranjana: The Journal of Management Awareness*. 17 (2). pp. 56-63. doi: 10.5958/0974-0945.2014.00005.3.
- Kim, H., Xiang, Z. & Fesenmaier, D. R. (2015). Use of The Internet for Trip Planning: A Generational Analysis. *Journal of Travel & Tourism Marketing*. 32 (3). pp. 276-289. doi: 10.1080/10548408.2014.896765.
- Kim, J., Bojanic, D. C. & Warnick R. B. (2009). Price Bundling and Travel Product Pricing Practices Used by Online Channels of Distribution. *Journal of Travel Research*. 47 (4). pp. 403-412. doi: 10.1177/0047287508328658.
- Kim, M., Chung, N. & Lee, C. (2011). The effect of perceived trust on electronic commerce: Shopping online for tourism products and services in South Korea. *Tourism Management*. 32 (2). pp. 256-265. doi: 10.1016/j.tourman.2010.01.011.
- Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Proceedings of the 14th international joint conference on Artificial intelligence. 2. pp. 1137-1143.
- Lam, T. & Tang, V. (2003). Recognizing Customer Complaint Behavior. *Journal of Travel & Tourism Marketing*. 14 (1). pp. 69-86. doi: 10.1300/J073v14n01\_05.
- Larose, D. T. (2005). *Discovering knowledge in data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Legohérel, P. Hsu, C. H.C. & Daucé, B. (2015). Variety-seeking: Using the CHAID segmentation approach in analyzing the international traveler market. *Tourism Management*. 46. pp. 359-366.
- Li, B., & Yu, Q. (2008). Classification of functional data: A segmentation approach. *Computational Statistics and Data Analysis*. 52 (10). pp. 4790-4800. doi: 10.1016/j.csda.2008.03.024
- Li, H., Sun, J. & Wu, J. (2010). Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications*. 37. pp.5895-5904. doi:10.1016/j.eswa.2010.02.016.



- Likas, A., Vlassis, N. & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition*. 36 (2). pp. 451-461. doi: 10.1016/S0031-3203(02)00060-2.
- Llodrà-Riera, I., Martínez-Ruiz, M. P., Jiménez-Zarco, A. I. & Izquierdo-Yusta, A. (2015). Assessing the influence of social media on tourists' motivations and image formation of a destination. *International Journal of Quality and Service Sciences*. 7 (4). pp.458 - 482. doi: 10.1108/IJQSS-03-2014-0022.
- Lopez, V. Fernandez, A, Garcia, S., Palade, V. & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*. 250. pp. 113-141. doi: 10.1016/j.ins.2013.07.007.
- Losada, N., Alén, E., Domínguez, T. & Nicolau, J. L. (2016). Travel frequency of seniors tourists. *Tourism Management*. 53. pp. 88-95. doi: 10.1016/j.tourman.2015.09.013.
- Loureiro, S. M. C. & Kastenholz, E. (2011). Corporate reputation, satisfaction, delight, and loyalty towards rural accommodation units in Portugal. *International Journal of Hospitality Management*. 30 (3). pp. 575-583. doi: 10.1016/j.ijhm.2010.10.007.
- Lu, C., Hsu, Y., Lu, Y. & Lin, W. (2015). Measuring tourist satisfaction by motivation, travel behavior and shopping behavior: the case of lake scenic area in Taiwan. *International Journal of Organizational Innovation*. 8 (1). pp. 117-132.
- Luukka, P. (2009). Classification based on fuzzy robust PCA algorithms and similarity classifier. *Expert Systems with Applications*. 36 (4). pp. 7463-7468. doi: 10.1016/j.eswa.2008.09.015.
- MacQueen, J. B. (1967). *Some Methods for classification and Analysis of Multivariate Observations*. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. pp. 281-297.
- Mantas, C. J. & Abellán, J. (2014). Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Systems with Applications*. 41 (10). pp. 4625-4637. doi: 10.1016/j.eswa.2014.01.017.
- Moorman, C., Deshpande & R., Zaitnan, G. (1992). Factors Affecting Trust in Market Research Relationships. *Journal of Marketing*. 57 (1). pp. 81-101.
- Naidoo P., Ramseook-Munhurrin P., Seebaluck N. V., Janvier S. (2015). Investigating the Motivation of Baby Boomers for Adventure Tourism. *Procedia - Social and Behavioral Sciences*. 175. pp. 244-251. doi:10.1016/j.sbspro.2015.01.1197.
- Nash, R., Thyne, M. & Davies, S. (2006). An investigation into customer satisfaction levels in the budget accommodation sector in Scotland: a case study of backpacker tourists and the Scottish Youth Hostels Association. *Tourism Management*. 27. pp. 525-532. doi: 10.1016/j.tourman.2005.01.001.
- Naylor, G. & Frank, K. E. (2001). The effect of price bundling on consumer perceptions of value. *Journal of Services Marketing*. 15 (4). pp.270 - 281. doi: 10.1108/EUM0000000005506.
- Ngai, E. W. T., Heung, V. C. S., Wong, Y. H. & Chan, F. K. Y. (2007). Consumer complaint behaviour of Asians and non-Asians about hotel services. *European Journal of Marketing*. 41. pp. 1375-1391. doi: 10.1108/03090560710821224.
- Peypoch, N., Randriamboarison R., Rasoamananjara, F. & Solonandrasana, B. (2012). The length of stay of tourists in Madagascar. *Tourism Management*. 33 (5). pp. 1230-1235. doi: 10.1016/j.tourman.2011.11.003.
- Polat, K. & Gunes, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*. 36 (2). pp. 1587-1592. doi: 10.1016/j.eswa.2007.11.051.

- Pundir, S. & Amala, R. (2015). Evaluation of Biomarker using Two Parameter Bi-exponential ROC Curve. *Pakistan Journal of Statistics & Operation Research*. 11 (4). pp. 481-496. doi: 10.18187/pjsor.v11i4.992.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufman Publishers Inc.
- Quinlan, J. R. (2006). *Bagging, Boosting, and C4.5*. Proceedings of the thirteenth national conference on Artificial intelligence. 1. pp. 725-730.
- Ragavan, N. A., Subramonian, H. & Sharif, S. P. (2014). Tourists' perceptions of destination travel attributes: An application to International tourists to Kuala Lumpur. *Procedia - Social and Behavioral Sciences*. 144. pp. 403-411. doi: 10.1016/j.sbspro.2014.07.309.
- Ramseook-Munhurrun, P., Seebaluck, V. N. & Naidoo, P. (2015). Examining the Structural Relationships of Destination Image, Perceived Value, Tourist Satisfaction and Loyalty: Case of Mauritius. *Procedia - Social and Behavioral Sciences*. 175. pp. 252-259. doi: 10.1016/j.sbspro.2015.01.1198.
- SAß, J. (2011). Online Advertising in the Tourism Industry and its Impact on Consumers: A Study to Investigate Online Advertising Tools, the Degree of Usage and Customer Preferences. *Tourism & Management Studies*. pp. 101-107.
- Santos, M. F., & Azevedo, C. S. (2005). *Data Mining – Descoberta de Conhecimento em Bases de Dados*. Portugal: FCA Editores.
- Sari, F.O., Alikilic, O. & Onat, F. E. (2013). *Complaining: Analysis of Accommodation Customers' eComplaints from a Turkish Internet Website*. Proceedings of the 2013 International Conference on Information, Business and Education Technology. 26. pp. 561-565. doi:10.2991/icibet.2013.183.
- Sarra, A., Di Zio., S. & Cappucci, M. (2015). A quantitative valuation of tourist experience in Lisbon. *Annals of Tourism Research*. 53. pp. 1-16. doi: 10.1016/j.annals.2015.04.003.
- Sayad, S. (2016). *Model Evaluation - Classification*. Retrieved September 16, 2016, from Saed Sayad Web site: [http://www.saedsayad.com/model\\_evaluation\\_c.htm](http://www.saedsayad.com/model_evaluation_c.htm).
- Sharpley, R. & Telfer, D. J. (2002). *Tourism and Development: Concepts and Issues*. Clevedon: Channel View Publications.
- Simielli, M. E. R. (2012). *Atlas Geográfico Escolar* (36th ed.). 1. pp. 48. São Paulo: Ática.
- Sirgy, J. M. (1982). Self-Concept in Consumer Behavior: A Critical Review. *Journal of Consumer Research*. 9 (3). pp. 287-300.
- Song, H., Li, G., Van Der Veen, R. & Chen J.L. (2011). Assessing Mainland Chinese Tourists' Satisfaction with Hong Kong Using Tourist Satisfaction Index. *International Journal of Tourism Research*. 13 (1). pp. 82-96. doi: 10.1002/jtr.801.
- Sparks, B. A. & Browning, V. (2010). Complaining in Cyberspace: The Motives and Forms of Hotel Guests' Complaints Online. *Journal of Hospitality Marketing & Management*. 19 (7). pp. 797-818. doi: 10.1080/19368623.2010.508010.
- Tasci, A. D. A. & Gartner, W. C. (2007). Destination Image and Its Functional Relationships. *Journal of Travel Research*. 45 (4). pp. 413-425. doi: 10.1177/0047287507299569
- Thrane, C. (2016). Students' summer tourism: Determinants of length of stay (LOS). *Tourism Management*. 54. pp. 178-184. doi: 10.1016/j.tourman.2015.11.010.
- Tsai, C. (2015). The Relationships among Destination Image, Perceived Quality, Emotional Place Attachment, Tourist Satisfaction, and Post-visiting Behavior Intentions. *Marketing Review (Xing Xiao Ping Lun)*. 12 (4). pp. 455-479.

- TSI (2016). *Overall Tourist Satisfaction Index*. Retrieved September 20, 2016, from The Hong Kong Polytechnic University Web site: <http://hotelschool.shtm.polyu.edu.hk/tsi/common/methodology.jsp>.
- UE, (2016). *A zona do euro*. Retrieved September 14, 2016, from European Union Web site: [https://europa.eu/european-union/sites/europaeu/files/docs/body/map\\_ecfin\\_one\\_currency\\_pt.pdf](https://europa.eu/european-union/sites/europaeu/files/docs/body/map_ecfin_one_currency_pt.pdf).
- UNWTO (2016). *Tourism Highlights - 2016 Edition*. Retrieved September 14, 2016, from World Tourism Organization UNWTO Web site: <http://www.e-unwto.org/doi/pdf/10.18111/9789284418145>.
- UNWTO. (1994). *Recommendations on Tourism Statistics*. Series M (83).
- Vajčnerová, I., Žiaran, P., Ryglová, K. & Andráško, I. (2014). Quality management of the tourist destination in the context of visitors' satisfaction. *Procedia Economics and Finance*. 12. pp. 718-724. doi: 10.1016/S2212-5671(14)00398-0.
- Valle, P. O., Pintassilgo, P., Matias, A. & André, F. (2012). Tourist attitudes towards an accommodation tax earmarked for environmental protection: A survey in the Algarve. *Tourism Management*. 33 (6). 1408-1416. 10.1016/j.tourman.2012.01.003.
- Vuuren, V. C. & Slabbert, E. (2011). Travel behaviour of tourists to a South African holiday resort African. *Journal for Physical, Health Education, Recreation and Dance*. 17 (4). pp. 694-707.
- Ward, J. H. & Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. 58 (301). pp. 236–244.
- Witten, I.H. & Frank, E. (2005). *Data Mining: Pratical Machine Learning Tools and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers.
- Xia, W., Jie, Z., Chaolin, G. & Feng, Z. (2009). Examining Antecedents and Consequences of Tourist Satisfaction: A Structural Modeling Approach. *Tsinghua Science and Technology*. 14 (3). pp. 397-406.
- Xu, J. B. & Chan, S. (2016). A new nature-based tourist motivation model: Testing the moderating effects of the push motivation. *Tourism Management Perspectives*. 18 (2). pp. 107-110. doi:10.1016/j.tmp.2016.01.001.
- Xu, X. & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*. 55. pp. 57-69. doi: 10.1016/j.ijhm.2016.03.003.
- Yoon, Y. & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management*. 26 (1). pp. 45-56. doi:10.1016/j.tourman.2003.08.016.
- Yousefi, M. & Marzuki, A. (2015). An Analysis of Push and Pull Motivational Factors of International Tourists to Penang, Malaysia. *International Journal of Hospitality & Tourism Administration*. 16 (1). pp. 40-56. doi: 10.1080/15256480.2015.991987
- Yuan, S., & McDonald, C. (1990). Motivational determinants of international pleasure time. *Journal of Travel Research*. 29 (1). pp. 42-44. doi: 10.1177/004728759002900109.
- Zhan, G., Wang, P., Xiang, H., Jiali, C., Wang, Y. & Huang, P. (2015). Tourist Satisfaction survey and analysis on the Dongqian Lake, Ningbo, China. *Herald Journal of Environmental and Health Science*. 2 (1). pp. 004-012.



## 7 Appendix

### A. Tourist characterisation and satisfaction cluster

Table 46: Tourist country of nationality by satisfaction cluster

Characteristics	Cluster satisfaction			Total
	Cluster 1	Cluster 2	Cluster 3	
	Not very satisfied	Satisfied	Very satisfied	
	(n=815)	(n=4596)	(n=6286)	
	%	%	%	
Austria	4.2%	3.0%	6.2%	4.8%
Belgium	2.9%	4.1%	4.5%	4.2%
Bulgaria	2.9%	2.7%	1.9%	2.3%
Czech Republic	4.7%	4.3%	4.4%	4.4%
Denmark	2.0%	2.7%	5.5%	4.2%
Estonia	0.9%	0.9%	1.4%	1.2%
Finland	2.9%	4.4%	3.2%	3.6%
France	8.5%	9.0%	5.5%	7.1%
Germany	4.5%	5.8%	7.6%	6.7%
Greece	6.1%	3.5%	1.3%	2.5%
Hungary	1.6%	2.1%	2.2%	2.1%
Ireland	1.8%	3.6%	6.9%	5.3%
Italy	10.4%	6.8%	4.7%	5.9%
Latvia	2.3%	1.7%	0.8%	1.2%
Lithuania	1.8%	1.3%	0.8%	1.0%
Luxembourg	0.9%	1.9%	1.9%	1.9%
Malta	2.2%	1.1%	1.9%	1.6%
Netherlands	2.6%	3.9%	5.5%	4.7%
Poland	4.5%	5.8%	3.6%	4.5%
Portugal	4.1%	4.2%	2.0%	3.0%
Cyprus	2.0%	2.2%	1.3%	1.7%
Croatia	3.6%	1.9%	1.4%	1.8%
Romania	3.4%	2.8%	1.2%	2.0%
Slovakia	3.4%	2.8%	3.5%	3.2%
Slovenia	1.3%	1.4%	2.6%	2.1%
Spain	6.7%	7.2%	2.9%	4.8%
Sweden	2.5%	5.1%	5.2%	5.0%
United Kingdom	5.3%	3.7%	10.0%	7.2%
Austria	4.2%	3.0%	6.2%	4.8%
Belgium	2.9%	4.1%	4.5%	4.2%
Bulgaria	2.9%	2.7%	1.9%	2.3%
Czech Republic	4.7%	4.3%	4.4%	4.4%
Denmark	2.0%	2.7%	5.5%	4.2%
Estonia	0.9%	0.9%	1.4%	1.2%
Finland	2.9%	4.4%	3.2%	3.6%
France	8.5%	9.0%	5.5%	7.1%
Germany	4.5%	5.8%	7.6%	6.7%
Greece	6.1%	3.5%	1.3%	2.5%
Hungary	1.6%	2.1%	2.2%	2.1%
Ireland	1.8%	3.6%	6.9%	5.3%
Italy	10.4%	6.8%	4.7%	5.9%
<i>Chi-squared test</i> <sup>a</sup>				$\chi^2_{(54)} = 865.095; p < 0.001^a$
Measure of association				Cramer's V = 0.201

Note: a) Monte Carlo p-value.

## **B. Tourist characterisation and formal complaint**

From the Table 47, the results show that tourists who do not complaint are from the central European countries (48.4%), have a household size of two people (48.5%), are not working (43.4%) and live in a small or middle sized town (39.2%).

Tourists who complaint are also from the central European countries (54.8%) and have a household size of two people (50.5%). However, this group have a higher percentage of these tourists than the group "No". The "yes" group have a higher proportion of non-working tourists (40.9%) and live in large towns (37.4%). The variables occupation and household size are a not good ones to determine formal complaints, since their values to both groups (yes and no) are very close.

To all the other variables (e.g. gender, age group, currency of the nationality country, age education and phone available), the p-value is above 0.1, which means they have no relationship with the existence of a formal complaint.

**Table 47: The social demographics characteristics by formal complaint**

Characteristics	Formal complaint		Total	Chi-squared test <sup>a</sup> Measure of association
	No	Yes		
	(n=11200) %	(n=497) %		
<b>Gender</b>				
Male	42.9	42.4	42.9	$\chi^2_{(1)}=0.060$ ; $p=0.831^a$ <i>Cramer's V = 0.002</i>
Female	57.1	57.6	57.1	
<b>Age group</b>				
15-24 years	7.8	6.0	7.7	$\chi^2_{(3)}=3.596$ ; $p=0.302a$ <i>Cramer's V = 0.016</i>
25-39 years	18.1	19.0	18.2	
40-54 years	29.2	27.8	29.2	
55 or more years	44.8	47.2	44.9	
<b>Area of the European Union (nationality)</b>				
North European countries	13.1	23.4	13.6	$\chi^2_{(5)}=99.354$ ; $p<0.001a$ <i>Cramer's V = 0.088</i> <i>If computed for nationality<sup>7</sup></i> <i>Cramer's V = 0.127</i>
Central European countries	48.4	54.8	48.6	
Countries of the Iberian Peninsula	7.7	8.4	7.7	
Eastern European countries	15.0	6.6	14.7	
Balkan countries	13.1	6.2	12.9	
Baltics	2.7	0.5	2.6	
<b>Currency of the nationality country</b>				
Euro	66.0	64.1	65.9	$\chi^2_{(1)}=0.833$ ; $p=0.386a$ <i>Cramer's V = 0.008</i>
Non-euro	34.0	35.9	34.1	
<b>Age education</b>				
Up to 15	8.0	8.3	8.0	$\chi^2_{(4)}=3.728$ ; $p=0.441a$ <i>Cramer's V = 0.016</i>
16-19	33.8	33.9	33.8	
20 years and older	51.9	52.8	51.9	
Still Studying	5.9	5.0	5.9	
No full-time education	0.5	0.0	0.5	
<b>Occupation</b>				
Self-employed	9.8	10.8	9.8	$\chi^2_{(3)}=8.046$ ; $p=0.045a$ <i>Cramer's V = 0.024</i>
Employees	40.3	39.0	40.2	
Manual workers	6.5	9.3	6.6	
Not working	43.4	40.9	43.3	
<b>Household size</b>				
1	19.5	23.4	19.7	$\chi^2_{(3)}=10.837$ ; $p=0.012a$ <i>Cramer's V = 0.028</i>
2	48.5	50.5	48.6	
3	16.9	14.4	16.8	
4+	15.0	11.8	14.9	
<b>Phone available</b>				
Mobile only	18.8	18.2	18.7	$\chi^2_{(2)}=2.463$ ; $p=0.288a$ <i>Cramer's V = 0.013</i>
Landline only	4.2	2.9	4.1	
Mobile and landline	77.1	78.9	77.1	
<b>Type of community</b>				
Rural area or village	27.4	28.1	27.5	$\chi^2_{(2)}=6.080$ ; $p=0.048a$ <i>Cramer's V = 0.021</i>
Small or middle sized town	39.2	34.5	39.0	
Large town	33.3	37.4	33.5	

Note: a) Monte Carlo p-value.

<sup>7</sup> View Table 48 for the nationality country analysis.

**Table 48: Tourist country of nationality by formal complaint**

Characteristics	Formal Complaint		Total
	No (n=11200)	Yes (n=497)	
	%	%	
Austria	4.6%	6.0%	4.7%
Belgium	4.1%	2.2%	4.0%
Bulgaria	2.3%	1.6%	2.3%
Czech Republic	4.2%	3.7%	4.2%
Denmark	4.7%	5.7%	4.7%
Estonia	1.5%	0.5%	1.4%
Finland	3.6%	4.6%	3.7%
France	7.1%	2.9%	7.0%
Germany	7.1%	7.3%	7.1%
Greece	2.8%	0.5%	2.7%
Hungary	2.2%	0.0%	2.1%
Ireland	4.7%	8.8%	4.9%
Italy	5.5%	5.7%	5.5%
Latvia	1.2%	1.1%	1.2%
Lithuania	1.2%	0.0%	1.1%
Luxembourg	1.8%	2.6%	1.8%
Malta	1.5%	2.6%	1.5%
Netherlands	4.8%	7.1%	4.9%
Poland	4.9%	1.3%	4.7%
Portugal	2.9%	1.6%	2.9%
Cyprus	1.8%	1.5%	1.8%
Croatia	1.8%	0.7%	1.8%
Romania	1.9%	0.2%	1.9%
Slovakia	3.0%	2.2%	3.0%
Slovenia	2.1%	0.0%	2.0%
Spain	4.7%	6.8%	4.8%
Sweden	4.8%	13.2%	5.2%
United Kingdom	7.2%	9.5%	7.3%
Austria	4.6%	6.0%	4.7%
Belgium	4.1%	2.2%	4.0%
Bulgaria	2.3%	1.6%	2.3%
Czech Republic	4.2%	3.7%	4.2%
Denmark	4.7%	5.7%	4.7%
Estonia	1.5%	0.5%	1.4%
Finland	3.6%	4.6%	3.7%
France	7.1%	2.9%	7.0%
Germany	7.1%	7.3%	7.1%
Greece	2.8%	0.5%	2.7%
Hungary	2.2%	0.0%	2.1%
Ireland	4.7%	8.8%	4.9%
Italy	5.5%	5.7%	5.5%
<i>Chi-squared test<sup>a</sup></i>			$\chi^2_{(27)} = 205.209; p < 0.001^a$
Measure of association			Cramer's V = 0.127

Note: a) Monte Carlo p-value.

From the Table 49, the results show that tourists who do not complaint never go to paid accommodation with more than 20 guests (45.0%), go to a combination of paid and unpaid accommodation (48.1%), never go to all-inclusive holidays (71.9%) and stay between 4 and 13 nights way (64.2%). This group have the highest proportion of tourists who never go to paid accommodation with more than 20 guests, stay between 4 and 13 nights away and that never go to all-inclusive holidays.

Tourists who complaint go to paid accommodation with more than 20 guests (31.2%), go to a combination of paid and unpaid accommodation (57.2%), never go to all-inclusive holidays (63.1%) and stay between 4 and 13 nights way (57.5%). This group have the highest proportion of tourists who go to paid and unpaid accommodation.



To all the other variables (e.g. paid accommodation with less than 20 guests, own property/second home, friends or relatives and frequency), the p-value is above 0.1, which means they have no significant relationship with the existence of a formal complaint.

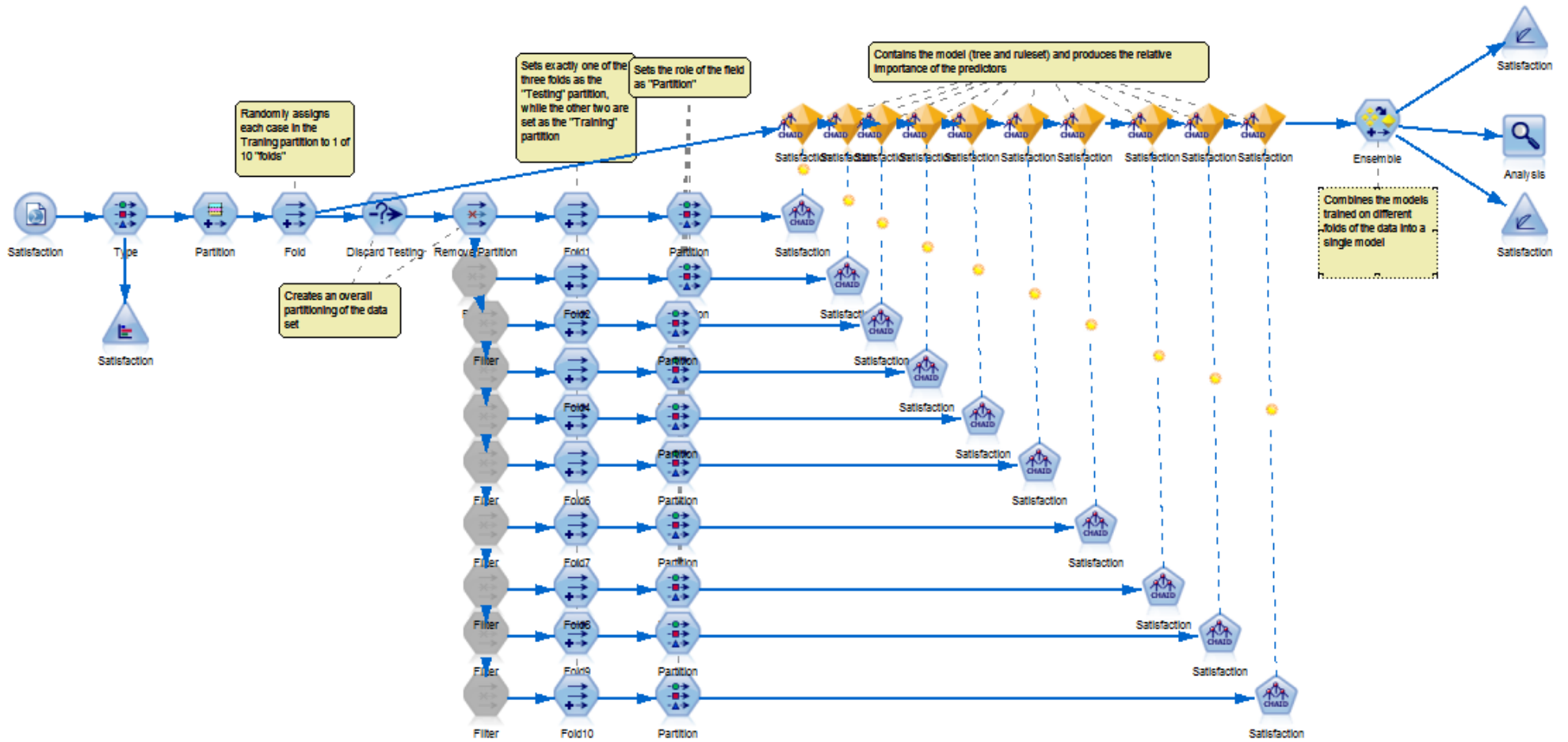
**Table 49: Tourist behaviour characterisation by formal complaint**

Characteristics	Formal complaint		Total	Chi-squared test <sup>a</sup> Measure of association
	No	Yes		
	(n=11200) %	(n=497) %		
<b>Paid accommodation with less than 20 guests</b>				
0	53.8	49.7	53.7	$\chi^2_{(5)}=13.614; p=0.018^a$ Cramer's V = 0.031
1	23.0	20.9	22.9	
2	10.9	12.8	11.0	
3	4.9	7.1	5.0	
4	2.7	3.5	2.8	
5 or more times	4.6	6.1	4.7	
<b>Paid accommodation with more than 20 guests</b>				
0	45.0	31.2	44.4	$\chi^2_{(5)}=67.241; p<0.001^a$ Cramer's V = 0.070
1	25.7	27.4	25.8	
2	13.0	15.0	13.1	
3	6.3	9.1	6.4	
4	3.5	7.6	3.7	
5 or more times	6.5	9.7	6.6	
<b>Own property /second home</b>				
0	73.1	73.6	73.1	$\chi^2_{(5)}=7.065; p=0.216^a$ Cramer's V = 0.023
1	8.9	7.6	8.9	
2	4.2	6.2	4.3	
3	2.6	2.2	2.6	
4	1.8	1.6	1.8	
5 or more times	9.3	8.8	9.3	
<b>Friends or relatives</b>				
0	54.4	55.8	54.5	$\chi^2_{(5)}=5.243; p=0.390^a$ Cramer's V = 0.019
1	17.7	17.6	17.7	
2	9.5	10.2	9.6	
3	5.2	4.6	5.1	
4	3.6	1.9	3.5	
5 or more times	9.6	10.0	9.6	
<b>Camping site (tent, motorhome, caravan, etc.)</b>				
0	86.2	84.9	86.2	$\chi^2_{(5)}=8.382; p=0.136^a$ Cramer's V = 0.025
1	7.8	7.2	7.8	
2	2.2	3.8	2.2	
3	1.1	1.0	1.1	
4	0.6	0.3	0.6	
5 or more times	2.1	2.7	2.1	
<b>Type of accommodation (paid or unpaid) in personal travels</b>				
Always in paid accommodation	28.4	29.9	28.5	$\chi^2_{(2)}=37.218; p<0.001^a$ Cramer's V = 0.052
Never in paid accommodation	23.5	12.9	23.0	
Paid and unpaid accommodation	48.1	57.2	48.5	
<b>Frequency: Number of trips</b>				
Once	19.4	13.7	19.2	$\chi^2_{(5)}=18.611; p=0.002^a$ Cramer's V = 0.037
Twice	16.8	14.8	16.7	
3 times	14.8	16.5	14.9	
4 or 5 times	19.6	22.8	19.7	
6 to 10 times	15.2	18.4	15.3	
More than 10 times	14.2	13.9	14.2	
<b>Duration of personal travels taken</b>				
Between 4 and 13 nights away	64.2	57.5	63.9	$\chi^2_{(2)}=19.953 p<0.001^a$ Cramer's V = 0.038
More than 13 nights away	19.8	19.7	19.8	
Both types of duration	16.0	22.8	16.3	
<b>Type of holiday (package)</b>				
Always 'all-inclusive holidays'	12.5	15.3	12.6	$\chi^2_{(2)}=22.376; p<0.001^a$ Cramer's V = 0.040
Never 'all-inclusive holidays'	71.9	63.1	71.6	
Combination of types of holiday	15.6	21.6	15.8	

Note: a) Monte Carlo p-value.

### C. Predictors of tourist satisfaction

Figure 37: IBM SPSS Modeler 10-fold stream



**Table 50: Results of the models for the social demographics characteristics for tourist characterisation variables**

Method		Default		A		B		C	
		Test	Training	Test	Training	Test	Training	Test	Training
CART	<b>Holdout</b>								
	Accuracy	59.97	61.15	60.39	61.41	59.97	61.15	60.08	60.98
	% cases well classified (less satisfied) – Precision	60.92	59.95	55.74	54.56	60.92	59.95	45.09	44.66
	% cases well classified (very satisfied) - Specificity	59.11	62.17	64.60	67.18	59.11	62.17	73.65	74.72
	AUC	0.627	0.643	0.612	0.739	0.627	0.643	0.627	0.631
C5.0	<b>Holdout</b>								
	Accuracy	59.43	68.42	-	-	-	-	-	-
	% cases well classified (less satisfied) - Precision	49.37	57.80	-	-	-	-	-	-
	% cases well classified (very satisfied) - Specificity	68.53	77.36	-	-	-	-	-	-
	AUC	0.612	0.739	-	-	-	-	-	-
CHAID	<b>Holdout</b>								
	Accuracy	65.04	70.66	61.21	63.00	61.21	62.97	61.21	63.00
	% cases well classified (less satisfied) - Precision	58.92	62.58	58.95	58.31	59.13	58.44	58.95	58.31
	% cases well classified (very satisfied) - Specificity	70.31	77.10	63.25	66.95	63.09	66.77	63.25	66.95
	AUC	0.648	0.667	0.644	0.680	0.644	0.678	0.645	0.679

**Table 51: Results of the models for the pre-travel variables**

Method		Default		A		B		C	
		Test	Training	Test	Training	Test	Training	Test	Training
CART	<b>Holdout</b>								
	Accuracy	52.57	51.83	52.51	51.77	52.41	52.05	52.44	51.94
	% cases well classified (less satisfied) - Precision	32.85	32.36	32.18	31.43	33.03	33.32	32.67	32.38
	% cases well classified (very satisfied) - Specificity	72.70	71.13	73.26	71.92	72.20	70.60	72.64	71.32
	AUC	0.532	0.526	0.533	0.520	0.531	0.530	0.532	0.523
C5.0	<b>Holdout</b>								
	Accuracy	51.10	53.61	-	-	-	-	-	-
	% cases well classified (less satisfied) - Precision	33.15	34.51	-	-	-	-	-	-
	% cases well classified (very satisfied) - Specificity	69.43	72.53	-	-	-	-	-	-
	AUC	0.522	0.551	-	-	-	-	-	-
CHAID	<b>Holdout</b>								
	Accuracy	53.58	51.61	53.58	51.61	53.58	51.61	53.58	51.61
	% cases well classified (less satisfied) - Precision	63.52	61.81	63.52	61.81	63.52	61.81	63.52	61.81
	% cases well classified (very satisfied) - Specificity	43.42	41.50	43.42	41.50	43.42	41.50	43.42	41.50
	AUC	0.534	0.518	0.534	0.518	0.534	0.518	0.534	0.518

**Table 52: Results of the models for the on-travel variables**

Method		Default		A		B		C	
		Test	Training	Test	Training	Test	Training	Test	Training
CART	<b>Holdout</b>								
	Accuracy	56.71	57.12	56.71	57.12	56.71	57.12	56.71	57.12
	% cases well classified (less satisfied) – Precision	45.47	44.93	45.47	44.93	45.47	44.93	45.57	44.93
	% cases well classified (very satisfied) – Specificity	66.81	67.38	66.81	67.38	66.81	67.38	66.81	67.38
	AUC	0.565	0.565	0.565	0.565	0.565	0.565	0.565	0.565
C5.0	<b>Holdout</b>								
	Accuracy	59.15	59.87	-	-	-	-	-	-
	% cases well classified (less satisfied) – Precision	35.93	34.75	-	-	-	-	-	-
	% cases well classified (very satisfied) – Specificity	80.17	81.02	-	-	-	-	-	-
	AUC	0.591	0.591	-	-	-	-	-	-
CHAID	<b>Holdout</b>								
	Accuracy	58.64	59.73	58.52	59.88	58.52	59.88	58.52	59.88
	% cases well classified (less satisfied) – Precision	36.11	35.20	35.28	34.64	35.28	34.64	35.28	34.64
	% cases well classified (very satisfied) – Specificity	79.04	80.38	35.28	34.64	79.58	81.13	79.58	81.13
	AUC	0.601	0.609	0.599	0.611	0.599	0.611	0.599	0.611

**Table 53: Results of the models for the group with all variables**

Method		Default		A		B		C	
		Test	Training	Test	Training	Test	Training	Test	Training
CART	<b>Holdout</b>								
	Accuracy	59.46	62.17	58.89	63.84	60.45	61.62	60.76	61.20
	% cases well classified (less satisfied) – Precision	55.68	56.62	52.23	55.47	54.49	53.67	52.41	50.38
	% cases well classified (very satisfied) – Specificity	62.88	66.84	64.92	70.88	65.84	68.31	68.32	70.32
	AUC	0.621	0.652	0.611	0.673	0.632	0.645	0.626	0.630
	<b>Cross-validation 10-fold</b>								
	Accuracy	62.92	67.64	64.30	71.67	63.58	65.99	62.77	64.23
	% cases well classified (less satisfied) - Precision	41.30	46.86	56.25	63.23	46.06	48.05	46.90	46.14
	% cases well classified (very satisfied) – Specificity	79.72	82.61	71.47	78.10	77.78	79.81	77.20	78.73
	AUC	0.636	0.660	0.623	0.726	0.625	0.639	0.632	0.645
C5.0	<b>Holdout</b>								
	Accuracy	60.02	68.68	-	-	-	-	-	-
	% cases well classified (less satisfied) – Precision	46.16	53.78	-	-	-	-	-	-
	% cases well classified (very satisfied) – Specificity	72.57	81.22	-	-	-	-	-	-
	AUC	0.611	0.741	-	-	-	-	-	-
	<b>Cross-validation 10-fold</b>								
	Accuracy	61.08	74.26	-	-	-	-	-	-
	% cases well classified (less satisfied) - Precision	38.53	53.93	-	-	-	-	-	-
	% cases well classified (very satisfied) - Specificity	80.11	90.45	-	-	-	-	-	-
	AUC	0.617	0.782	-	-	-	-	-	-
CHAID	<b>Holdout</b>								
	Accuracy	61.80	62.55	60.45	64.89	60.87	63.58	60.96	64.19
	% cases well classified (less satisfied) – Precision	59.61	58.61	46.94	50.27	54.79	55.52	49.49	51.13
	% cases well classified (very satisfied) – Specificity	63.79	65.87	72.68	77.20	66.38	70.36	71.34	75.19
	AUC	0.649	0.669	0.641	0.703	0.642	0.684	0.642	0.694
	<b>Cross-validation 10-fold</b>								
	Accuracy	62.28	66.48	61.34	73.20	63.10	70.52	63.42	72.04
	% cases well classified (less satisfied) - Precision	55.67	59.29	48.61	62.38	56.13	63.07	56.00	64.45
	% cases well classified (very satisfied) - Specificity	68.26	72.39	72.61	81.03	69.21	76.30	70.00	77.85
	AUC	0.625	0.678	0.623	0.726	0.630	0.703	0.630	0.716

Table 54: Results for the boosting and bagging models performed

Algorithm	Evaluation Measures	Demographics				Pre-travel				On-travel				All variables				
		Boosting		Bagging		Boosting		Bagging		Boosting		Bagging		Boosting		Bagging		
		Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	
CART	<b>Holdout</b>																	
	Accuracy	61.18	61.88	60.81	61.53	53.24	52.00	53.06	52.27	58.13	59.62	58.21	59.40	61.29	63.09	60.84	61.58	
	% cases well classified (less satisfied) - Precision	56.10	55.07	54.25	52.52	59.10	57.24	37.51	36.08	41.28	40.32	31.71	30.80	57.29	56.57	54.55	52.92	
	% cases well classified (very satisfied) - Specificity	65.79	67.61	66.76	69.12	47.25	46.82	68.93	68.31	73.38	75.87	82.22	83.48	64.92	68.58	66.54	68.87	
	AUC	0.631	0.628	0.632	0.639	0.535	0.525	0.542	0.533	0.586	0.594	0.582	0.582	0.631	0.637	0.631	0.638	
	<b>Cross-validation</b>																	
	Accuracy	-	-	-	-	-	-	-	-	-	-	-	-	-	64.19	68.82	61.48	62.87
	% cases well classified (less satisfied) - Precision	-	-	-	-	-	-	-	-	-	-	-	-	-	50.82	53.56	44.89	45.54
	% cases well classified (very satisfied) - Specificity	-	-	-	-	-	-	-	-	-	-	-	-	-	75.13	80.34	75.93	77.34
AUC	-	-	-	-	-	-	-	-	-	-	-	-	-	0.629	0.661	0.618	0.639	
C5.0	<b>Holdout</b>																	
	Accuracy	59.06	68.37	-	-	50.86	52.89	-	-	57.99	58.84	-	-	59.15	68.88	-	-	
	% cases well classified (less satisfied) - Precision	48.39	57.02	-	-	35.65	44.73	-	-	36.88	36.46	-	-	46.80	54.64	-	-	
	% cases well classified (very satisfied) - Specificity	69.10	78.15	-	-	66.44	63.43	-	-	77.10	77.67	-	-	71.60	81.18	-	-	
	AUC	0.605	0.713	-	-	0.530	0.560	-	-	0.588	0.596	-	-	0.616	0.745	-	-	
	<b>Cross-validation</b>																	
	Accuracy	-	-	-	-	-	-	-	-	-	-	-	-	-	63.74	86.41	-	-
	% cases well classified (less satisfied) - Precision	-	-	-	-	-	-	-	-	-	-	-	-	-	44.71	72.30	-	-
	% cases well classified (very satisfied) - Specificity	-	-	-	-	-	-	-	-	-	-	-	-	-	79.93	96.72	-	-
AUC	-	-	-	-	-	-	-	-	-	-	-	-	-	0.615	0.852	-	-	
CHAID	<b>Holdout</b>																	
	Accuracy	59.26	67.05	59.32	71.68	51.92	54.12	51.62	54.66	58.52	59.89	59.29	60.90	59.80	68.30	57.36	73.64	
	% cases well classified (less satisfied) - Precision	53.06	58.52	51.52	64.07	43.25	44.73	59.83	61.63	34.98	34.24	32.23	31.50	49.80	58.20	49.02	66.86	
	% cases well classified (very satisfied) - Specificity	64.87	74.22	66.38	78.08	60.78	63.43	43.24	47.75	79.85	81.49	83.70	84.89	68.86	76.80	64.92	79.35	
	AUC	0.618	0.731	0.589	0.763	0.512	0.560	0.511	0.567	0.601	0.613	0.596	0.632	0.613	0.752	0.564	0.779	
	<b>Cross-validation</b>																	
	Accuracy	-	-	-	-	-	-	-	-	-	-	-	-	-	59.26	67.05	63.96	72.34
	% cases well classified (less satisfied) - Precision	-	-	-	-	-	-	-	-	-	-	-	-	-	53.06	58.52	53.37	63.47
	% cases well classified (very satisfied) - Specificity	-	-	-	-	-	-	-	-	-	-	-	-	-	64.87	74.22	73.48	79.52
AUC	-	-	-	-	-	-	-	-	-	-	-	-	-	0.638	0.743	0.642	0.731	

## D. Decision trees rules

### I. Highest Accuracy

**Area of the European Union that the country of nationality belongs in [ North European countries Central European countries ] [ Mode: 0 ]**

**Nationality in [ Finland France Italy ] [ Mode: 1 ]**

**Area of the European Union that the destination country belongs in [ North European countries Countries of the Iberian Peninsula Eastern European countries**

**Balkan countries Baltics Not in the European Union ] [ Mode: 0 ]**

Main reason for going on holiday in [ Wellness/Spa/health treatment Culture (e.g. religious, gastronomy, arts) Specific events (Sporting events/festivals/clubbing) ] [

Mode: 0 ] => Very Satisfied

Main reason for going on holiday in [ Sun/beach City trips Sport-related activities (e.g. scuba-diving, cycling ...) Nature (mountain, lake, landscape, etc.) Visiting

family/friends/relatives Other ] [ Mode: 1 ] => Less Satisfied

**Area of the European Union that the destination country belongs in [ Central European countries Own country ] [ Mode: 1 ] => Less Satisfied**

**Nationality in [ Austria Belgium Denmark Germany Ireland Luxembourg Malta Netherlands Sweden United Kingdom ] [ Mode: 0 ] => Very Satisfied**

**Area of the European Union that the country of nationality belongs in [ Countries of the Iberian Peninsula Eastern European countries Balkan countries Baltics ] [ Mode: 1 ]**

**Nationality in [ Czech Republic Estonia Hungary Slovakia Slovenia ] [ Mode: 0 ] => Very Satisfied**

**Nationality in [ Bulgaria Greece Latvia Lithuania Poland Portugal Cyprus Croatia Romania Spain ] [ Mode: 1 ] => Less Satisfied**

### II. Less Satisfied

**Nationality = 1 or Nationality = 5 or Nationality = 6 or Nationality = 12 or Nationality = 25 [ Mode: 0 ]**

**Age <= 3 [ Mode: 0 ]**

**Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ]**

**Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 2 [ Mode: 0 ]**

Number of organising methods <= 2 [ Mode: 0 ] => Very Satisfied

Number of organising methods > 2 [ Mode: 1 ] => Less Satisfied

**Type of holiday (package) in 2014 = 3 [ Mode: 0 ] => Very Satisfied**

**Party experience any safety issues while using paid accommodation = 1 [ Mode: 0 ]**

**Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ]**

Duration of personal travels taken in 2014 = 1 or Duration of personal travels taken in 2014 = 2 [ Mode: 1 ] => Less Satisfied

Duration of personal travels taken in 2014 = 3 [ Mode: 0 ] => Very Satisfied

**Type of holiday (package) in 2014 = 2 [ Mode: 0 ]**

Paid accommodation with less than 20 guests = 0 or Paid accommodation with less than 20 guests = 2 or Paid accommodation with less than 20 guests = 3 or Paid accommodation with less than 20 guests = 5 [ Mode: 1 ] => Less Satisfied

Paid accommodation with less than 20 guests = 1 [ Mode: 0 ] => Very Satisfied

**Age > 3 [ Mode: 0 ]**

**Paid accommodation with less than 20 guests = 0 [ Mode: 0 ]**

**Paid accommodation with more than 20 guests = 0 or Paid accommodation with more than 20 guests = 1 or Paid accommodation with more than 20 guests = 2 or Paid accommodation with more than 20 guests = 4 or Paid accommodation with more than 20 guests = 5 or Paid accommodation with more than 20 guests IS MISSING [ Mode: 0 ]**

Camping site (tent, motorhome, caravan, etc.) = 0 or Camping site (tent, motorhome, caravan, etc.) = 1 or Camping site (tent, motorhome, caravan, etc.) = 2 or Camping site (tent, motorhome, caravan, etc.) IS MISSING [ Mode: 0 ] => Very Satisfied  
Camping site (tent, motorhome, caravan, etc.) = 3 or Camping site (tent, motorhome, caravan, etc.) = 4 or Camping site (tent, motorhome, caravan, etc.) = 5 [ Mode: 1 ] => Less Satisfied

**Paid accommodation with more than 20 guests = 3 [ Mode: 0 ]**

Frequency <= 5 [ Mode: 0 ] => Very Satisfied

Frequency > 5 [ Mode: 1 ] => Less Satisfied

**Paid accommodation with less than 20 guests = 1 or Paid accommodation with less than 20 guests = 2 or Paid accommodation with less than 20 guests = 3 or Paid accommodation with less than 20 guests = 4 or Paid accommodation with less than 20 guests = 5 or Paid accommodation with less than 20 guests IS MISSING [ Mode: 0 ]**

**Gender = 1 [ Mode: 0 ] => Very Satisfied**

**Gender = 2 [ Mode: 0 ]**

Occupation = 1 or Occupation = 2 or Occupation = 4 [ Mode: 0 ] => Very Satisfied

Occupation = 3 [ Mode: 1 ] => Less Satisfied

**Nationality = 2 or Nationality = 4 or Nationality = 11 or Nationality = 16 or Nationality = 24 or Nationality = 27 [ Mode: 0 ]**

**Gender = 1 [ Mode: 1 ]**

**Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ] => Very Satisfied**

**Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]**

**Frequency <= 4 [ Mode: 1 ] => Less Satisfied**

**Frequency > 4 and Frequency <= 5 [ Mode: 0 ]**

Phone available = 1 [ Mode: 1 ] => Less Satisfied

Phone available = 3 [ Mode: 0 ] => Very Satisfied

Frequency > 5 [ Mode: 1 ] => Less Satisfied

**Gender = 2 [ Mode: 0 ]**

**Age <= 2 [ Mode: 0 ]**

**Area of the European Union that the destination country belongs = 1 or Area of the European Union that the destination country belongs = 2 or Area of the European Union that the destination country belongs = 3 or Area of the European Union that the destination country belongs = 5 or Area of the European Union that the destination country belongs = 8 [ Mode: 1 ]**

Occupation = 1 or Occupation = 2 or Occupation = 3 [ Mode: 0 ] => Very Satisfied

Occupation = 4 [ Mode: 1 ] => Less Satisfied

**Area of the European Union that the destination country belongs = 4 or Area of the European Union that the destination country belongs = 7 or Area of the European Union that the destination country belongs IS MISSING [ Mode: 0 ] => Very Satisfied**

**Age > 2 [ Mode: 0 ]**

**Number of information sources main trip <= 1 [ Mode: 0 ] => Very Satisfied**

**Number of information sources main trip > 1 [ Mode: 0 ]**

Duration of personal travels taken in 2014 = 1 or Duration of personal travels taken in 2014 = 2 [ Mode: 0 ] => Very Satisfied

Duration of personal travels taken in 2014 = 3 [ Mode: 0 ] => Very Satisfied

**Nationality = 3 or Nationality = 7 or Nationality = 8 or Nationality = 13 or Nationality IS MISSING [ Mode: 1 ]**

**Currency of the destination country = 1 or Currency of the destination country IS MISSING [ Mode: 1 ]**

**Type of accommodation (paid or unpaid) = 1 or Type of accommodation (paid or unpaid) = 3 [ Mode: 1 ]**

**Household size = 1 [ Mode: 1 ]**

Gender = 1 [ Mode: 1 ] => Less Satisfied

Gender = 2 [ Mode: 0 ] => Very Satisfied

**Household size = 2 or Household size = 4 [ Mode: 1 ]**

Number of organising methods <= 1 [ Mode: 1 ] => Less Satisfied

Number of organising methods > 1 and Number of organising methods <= 2 [ Mode: 1 ] => Less Satisfied

Number of organising methods > 2 [ Mode: 1 ] => Less Satisfied

**Household size = 3 or Household size IS MISSING [ Mode: 1 ]**

Area of the European Union that the country of nationality belongs = 1 or Area of the European Union that the country of nationality belongs IS MISSING [ Mode: 1 ] => Less Satisfied

Area of the European Union that the country of nationality belongs = 2 or Area of the European Union that the country of nationality belongs = 5 [ Mode: 1 ] => Less Satisfied

**Type of accommodation (paid or unpaid) = 2 [ Mode: 1 ]**

**Gender = 1 [ Mode: 1 ] => Less Satisfied**

**Gender = 2 [ Mode: 0 ]**

Number of organising methods <= 0 [ Mode: 1 ] => Less Satisfied

Number of organising methods > 0 and Number of organising methods <= 1 [ Mode: 0 ] => Very Satisfied

Number of organising methods > 1 [ Mode: 1 ] => Less Satisfied

**Currency of the destination country = 2 [ Mode: 0 ]**

**Age <= 2 [ Mode: 1 ]**

**Currency of the nationality country = 1 or Currency of the nationality country IS MISSING [ Mode: 0 ]**

Age education <= 2 [ Mode: 0 ] => Very Satisfied

Age education > 2 [ Mode: 1 ] => Less Satisfied

**Currency of the nationality country = 2 [ Mode: 1 ] => Less Satisfied**

**Age > 2 [ Mode: 0 ] => Very Satisfied**

**Nationality = 9 or Nationality = 17 or Nationality = 18 [ Mode: 0 ]**

**Duration of personal travels taken in 2014 = 1 [ Mode: 0 ]**

**Paid accommodation with more than 20 guests = 0 or Paid accommodation with more than 20 guests = 2 or Paid accommodation with more than 20 guests = 3 or Paid accommodation with more than 20 guests = 5 or Paid accommodation with more than 20 guests IS MISSING [ Mode: 0 ] => Very Satisfied**

**Paid accommodation with more than 20 guests = 1 or Paid accommodation with more than 20 guests = 4 [ Mode: 1 ]**

**Phone available = 1 or Phone available = 2 [ Mode: 0 ] => Very Satisfied**

**Phone available = 3 [ Mode: 1 ]**

Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 2 [ Mode: 1 ] => Less Satisfied

Type of holiday (package) in 2014 = 3 [ Mode: 1 ] => Less Satisfied

**Duration of personal travels taken in 2014 = 2 or Duration of personal travels taken in 2014 = 3 [ Mode: 0 ]**

**Camping site (tent, motorhome, caravan, etc.) = 0 or Camping site (tent, motorhome, caravan, etc.) = 5 [ Mode: 0 ]**

**Age <= 3 [ Mode: 0 ]**

Paid accommodation with less than 20 guests = 0 or Paid accommodation with less than 20 guests = 4 [ Mode: 0 ] => Very Satisfied

Paid accommodation with less than 20 guests = 1 or Paid accommodation with less than 20 guests = 2 or Paid accommodation with less than 20 guests = 3 or Paid accommodation with less than 20 guests = 5 or Paid accommodation with less than 20 guests IS MISSING [ Mode: 0 ] => Very Satisfied

**Age > 3 [ Mode: 0 ]**

Paid accommodation with less than 20 guests = 0 or Paid accommodation with less than 20 guests = 4 [ Mode: 0 ] => Very Satisfied

Paid accommodation with less than 20 guests = 1 or Paid accommodation with less than 20 guests = 2 or Paid accommodation with less than 20 guests = 3 or Paid accommodation with less than 20 guests = 5 or Paid accommodation with less than 20 guests IS MISSING [ Mode: 0 ] => Very Satisfied

**Camping site (tent, motorhome, caravan, etc.) = 1 or Camping site (tent, motorhome, caravan, etc.) = 2 or Camping site (tent, motorhome, caravan, etc.) = 3 or Camping site (tent, motorhome, caravan, etc.) = 4 or Camping site (tent, motorhome, caravan, etc.) IS MISSING [ Mode: 0 ]**



Gender = 1 [ Mode: 1 ] => Less Satisfied  
 Gender = 2 [ Mode: 0 ] => Very Satisfied  
**Nationality = 10 [ Mode: 1 ]**  
 Phone available = 1 [ Mode: 0 ] => Very Satisfied  
 Phone available = 2 or Phone available = 3 [ Mode: 1 ] => Less Satisfied  
**Nationality = 14 or Nationality = 20 or Nationality = 23 or Nationality = 26 [ Mode: 1 ]**  
**Gender = 1 [ Mode: 1 ]**  
**Age education <= 1 [ Mode: 1 ] => Less Satisfied**  
**Age education > 1 or Age education IS MISSING [ Mode: 1 ]**  
**Paid accommodation with less than 20 guests = 0 or Paid accommodation with less than 20 guests = 3 or Paid accommodation with less than 20 guests = 5 [ Mode: 1 ]**  
 Currency of the destination country = 1 or Currency of the destination country IS MISSING [ Mode: 1 ] => Less Satisfied  
 Currency of the destination country = 2 [ Mode: 1 ] => Less Satisfied  
**Paid accommodation with less than 20 guests = 1 or Paid accommodation with less than 20 guests = 2 or Paid accommodation with less than 20 guests = 4 or Paid accommodation with less than 20 guests IS MISSING [ Mode: 1 ] => Less Satisfied**  
**Gender = 2 [ Mode: 1 ]**  
**Occupation = 1 or Occupation = 2 or Occupation = 4 [ Mode: 1 ]**  
 Party experience any safety issues while using paid accommodation = 0 [ Mode: 1 ] => Less Satisfied  
 Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ] => Less Satisfied  
**Occupation = 3 [ Mode: 0 ]**  
**Age <= 2 [ Mode: 1 ]**  
 Duration of personal travels taken in 2014 = 1 or Duration of personal travels taken in 2014 = 3 [ Mode: 1 ] => Less Satisfied  
 Duration of personal travels taken in 2014 = 2 [ Mode: 0 ] => Very Satisfied  
**Age > 2 [ Mode: 0 ]**  
 Number of organising methods used for main trip = 0 or Number of organising methods used for main trip = 1 or Number of organising methods used for main trip = 9 [ Mode: 0 ] => Very Satisfied  
 Number of organising methods used for main trip = 2 [ Mode: 1 ] => Less Satisfied  
**Nationality = 15 or Nationality = 19 or Nationality = 21 or Nationality = 22 [ Mode: 1 ]**  
 Own property/second home = 0 or Own property/second home = 1 or Own property/second home = 2 or Own property/second home = 3 or Own property/second home = 4 or **Own property/second home IS MISSING [ Mode: 1 ]**  
**Gender = 1 [ Mode: 1 ]**  
 Age <= 3 [ Mode: 1 ] => Less Satisfied  
 Age > 3 [ Mode: 0 ] => Very Satisfied  
**Gender = 2 [ Mode: 1 ]**  
**Party experience any safety issues while using paid accommodation = 0 [ Mode: 1 ] => Less Satisfied**  
**Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]**  
 Type of holiday (package) in 2014 = 1 [ Mode: 0 ] => Very Satisfied  
 Type of holiday (package) in 2014 = 2 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ] => Less Satisfied  
**Own property/second home = 5 [ Mode: 1 ]**  
**The nationality and destination currency are the same = 0 or The nationality and destination currency are the same = 1 [ Mode: 1 ] => Less Satisfied**  
**The nationality and destination currency are the same IS MISSING [ Mode: 0 ]**  
 Currency of the nationality country = 1 [ Mode: 1 ] => Less Satisfied  
 Currency of the nationality country = 2 [ Mode: 0 ] => Very Satisfied

**Nationality = 28 [ Mode: 0 ]**

**Phone available = 1 [ Mode: 1 ] => Less Satisfied**

**Phone available = 2 [ Mode: 0 ] => Very Satisfied**

**Phone available = 3 [ Mode: 0 ]**

**Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ]**

**Camping site (tent, motorhome, caravan, etc.) = 0 or Camping site (tent, motorhome, caravan, etc.) = 2 or Camping site (tent, motorhome, caravan, etc.) = 3 or Camping site (tent, motorhome, caravan, etc.) = 4 or Camping site (tent, motorhome, caravan, etc.) = 5 or Camping site (tent, motorhome, caravan, etc.) IS MISSING [ Mode: 0 ]**

Frequency <= 2 [ Mode: 0 ] => Very Satisfied

Frequency > 2 and Frequency <= 6 [ Mode: 0 ] => Very Satisfied

Frequency > 6 [ Mode: 0 ] => Very Satisfied

**Camping site (tent, motorhome, caravan, etc.) = 1 [ Mode: 0 ]**

Household size = 1 or Household size = 3 or Household size = 4 [ Mode: 0 ] => Very Satisfied

Household size = 2 [ Mode: 1 ] => Less Satisfied

**Party experience any safety issues while using paid accommodation = 1 [ Mode: 0 ]**

**Gender = 1 [ Mode: 0 ]**

Household size = 1 or Household size = 2 [ Mode: 0 ] => Very Satisfied

Household size = 4 [ Mode: 1 ] => Less Satisfied

**Gender = 2 [ Mode: 1 ]**

Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ] => Less Satisfied

Type of holiday (package) in 2014 = 2 [ Mode: 0 ] => Very Satisfied

### **III. Very Satisfied**

**Nationality = 1 or Nationality = 6 or Nationality = 25 [ Mode: 0 ]**

**Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ]**

**Gender = 1 [ Mode: 0 ]**

**Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 3 [ Mode: 0 ]**

Duration of the main trip = 1 or Duration of the main trip = 3 [ Mode: 0 ] => Very Satisfied

Duration of the main trip = 2 [ Mode: 1 ] => Less Satisfied

**Type of holiday (package) in 2014 = 2 [ Mode: 0 ] => Very Satisfied**

**Gender = 2 [ Mode: 0 ] => Very Satisfied**

**Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]**

**Type of accommodation (paid or unpaid) = 1 [ Mode: 1 ] => Less Satisfied**

**Type of accommodation (paid or unpaid) = 3 [ Mode: 0 ]**

**Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 2 [ Mode: 0 ]**

Type of community = 1 or Type of community = 3 [ Mode: 0 ] => Very Satisfied

Type of community = 2 [ Mode: 1 ] => Less Satisfied

**Type of holiday (package) in 2014 = 3 [ Mode: 1 ] => Less Satisfied**

**Nationality = 2 or Nationality = 4 or Nationality = 11 or Nationality = 16 or Nationality = 24 or Nationality = 27 [ Mode: 0 ]**

**Gender = 1 [ Mode: 0 ]**

**Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ] => Very Satisfied**

**Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]**  
**Frequency <= 4 [ Mode: 1 ] => Less Satisfied**  
**Frequency > 4 and Frequency <= 5 [ Mode: 0 ]**  
**Phone available = 1 [ Mode: 1 ]**  
 Age education <= 2 [ Mode: 0 ] => Very Satisfied  
 Age education > 2 [ Mode: 1 ] => Less Satisfied  
**Phone available = 3 [ Mode: 0 ]**  
 Paid accommodation with more than 20 guests = 0 or Paid accommodation with more than 20 guests = 1 or Paid accommodation with more than 20 guests = 2 or Paid accommodation with more than 20 guests = 3 [ Mode: 0 ] => Very Satisfied  
 Paid accommodation with more than 20 guests = 4 [ Mode: 1 ] => Less Satisfied  
**Frequency > 5 [ Mode: 1 ] => Less Satisfied**

**Gender = 2 [ Mode: 0 ]**  
**Age education <= 2 [ Mode: 0 ] => Very Satisfied**  
**Age education > 2 or Age education IS MISSING [ Mode: 0 ]**  
**Household size = 1 or Household size = 2 [ Mode: 0 ]**  
**Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 3 [ Mode: 0 ]**  
 Duration of the main trip = 1 or Duration of the main trip = 3 [ Mode: 0 ] => Very Satisfied  
 Duration of the main trip = 2 [ Mode: 1 ] => Less Satisfied  
**Type of holiday (package) in 2014 = 2 [ Mode: 0 ]**  
 Type of holiday (package) main trip = 2 [ Mode: 1 ] => Less Satisfied  
 Type of holiday (package) main trip = 4 [ Mode: 0 ] => Very Satisfied  
**Household size = 3 or Household size = 4 [ Mode: 1 ]**  
**Type of community = 1 or Type of community = 2 [ Mode: 1 ]**  
 Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ] => Less Satisfied  
 Type of holiday (package) in 2014 = 2 [ Mode: 1 ] => Less Satisfied  
**Type of community = 3 [ Mode: 0 ] => Very Satisfied**

**Nationality = 3 or Nationality = 7 or Nationality IS MISSING [ Mode: 1 ]**  
**Currency of the destination country = 1 or Currency of the destination country = 2 [ Mode: 0 ]**  
**Duration of personal travels taken in 2014 = 1 [ Mode: 1 ]**  
**Occupation = 1 or Occupation = 2 [ Mode: 1 ] => Less Satisfied**  
**Occupation = 3 or Occupation = 4 or Occupation IS MISSING [ Mode: 0 ]**  
**Phone available = 1 or Phone available = 2 or Phone available = 3 [ Mode: 0 ] => Very Satisfied**  
**Phone available IS MISSING [ Mode: 1 ]**  
 Number of information sources used by each tourist <= 2 [ Mode: 1 ] => Less Satisfied  
 Number of information sources used by each tourist > 2 [ Mode: 0 ] => Very Satisfied  
**Duration of personal travels taken in 2014 = 2 or Duration of personal travels taken in 2014 = 3 [ Mode: 0 ] => Very Satisfied**

**Currency of the destination country IS MISSING [ Mode: 1 ]**  
**Number of organising methods <= 0 [ Mode: 1 ]**  
 Type of holiday (package) in 2014 = 1 [ Mode: 0 ] => Very Satisfied  
 Type of holiday (package) in 2014 = 2 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ] => Less Satisfied  
**Number of organising methods > 0 and Number of organising methods <= 1 [ Mode: 1 ] => Less Satisfied**  
**Number of organising methods > 1 [ Mode: 1 ]**  
 Gender = 1 [ Mode: 1 ] => Less Satisfied

Gender = 2 [ Mode: 1 ] => Less Satisfied

**Nationality = 5 or Nationality = 12 or Nationality = 28 [ Mode: 0 ]**

Age <= 1 [ Mode: 0 ] => Very Satisfied

Age > 1 and Age <= 3 [ Mode: 0 ]

Area of the European Union that the destination country belongs = 1 or Area of the European Union that the destination country belongs = 7 [Mode: 0 ]

Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ] => Very Satisfied

Party experience any safety issues while using paid accommodation = 1 [ Mode: 0 ]

Type of holiday (package) in 2014 = 1 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ]

Main reason for going on holiday main trip = 1 [ Mode: 0 ] => Very Satisfied

Main reason for going on holiday main trip = 10 [ Mode: 1 ] => Less Satisfied

Type of holiday (package) in 2014 = 2 [ Mode: 0 ]

Camping site (tent, motorhome, caravan, etc.) = 0 or Camping site (tent, motorhome, caravan, etc.) = 1 [ Mode: 0 ] => Very Satisfied

Camping site (tent, motorhome, caravan, etc.) = 5 [ Mode: 1 ] => Less Satisfied

**Area of the European Union that the destination country belongs = 2 or Area of the European Union that the destination country belongs = 3 or Area of the European Union that the destination country belongs IS MISSING [ Mode: 0 ]**

Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ]

Friends or relatives = 0 or Friends or relatives = 1 or Friends or relatives = 2 or Friends or relatives = 3 or Friends or relatives = 5 or Friends or relatives IS MISSING [ Mode: 0 ] => Very Satisfied

Friends or relatives = 4 [ Mode: 1 ]

Camping site (tent, motorhome, caravan, etc.) = 0 or Camping site (tent, motorhome, caravan, etc.) = 1 [ Mode: 1 ] => Less Satisfied

Camping site (tent, motorhome, caravan, etc.) = 2 [ Mode: 0 ] => Very Satisfied

Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]

Currency of the nationality country = 1 [ Mode: 0 ] => Very Satisfied

Currency of the nationality country = 2 [ Mode: 1 ]

Number of organising methods <= 2 [ Mode: 1 ] => Less Satisfied

Number of organising methods > 2 [ Mode: 0 ] => Very Satisfied

**Area of the European Union that the destination country belongs = 4 or Area of the European Union that the destination country belongs = 5 or Area of the European Union that the destination country belongs = 6 or Area of the European Union that the destination country belongs = 8 [ Mode: 0 ]**

Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ]

Nationality = 5 or Nationality = 12 [ Mode: 1 ] => Less Satisfied

Nationality = 28 [ Mode: 0 ]

Gender = 1 [ Mode: 0 ] => Very Satisfied

Gender = 2 [ Mode: 0 ] => Very Satisfied

Party experience any safety issues while using paid accommodation = 1 [ Mode: 0 ]

Type of community = 1 or Type of community = 3 [ Mode: 0 ] => Very Satisfied

Type of community = 2 [ Mode: 1 ] => Less Satisfied

Age > 3 [ Mode: 0 ]

Paid accommodation with less than 20 guests = 0 [ Mode: 0 ]

Currency of the nationality country = 1 [ Mode: 0 ]

Number of organising methods <= 2 [ Mode: 0 ] => Very Satisfied

Number of organising methods > 2 [ Mode: 1 ]

The nationality and destination currency are the same = 0 [ Mode: 0 ] => Very Satisfied

The nationality and destination currency are the same = 1 [ Mode: 1 ] => Less Satisfied

**Currency of the nationality country = 2 [ Mode: 0 ] => Very Satisfied**  
**Paid accommodation with less than 20 guests = 1 or Paid accommodation with less than 20 guests = 2 or Paid accommodation with less than 20 guests = 3 or Paid accommodation with less than 20 guests = 4 or Paid accommodation with less than 20 guests = 5 or Paid accommodation with less than 20 guests IS MISSING [ Mode: 0 ]**  
 Number of organising methods used for main trip = 0 or Number of organising methods used for main trip = 2 [ Mode: 1 ] => Less Satisfied  
 Number of organising methods used for main trip = 1 or Number of organising methods used for main trip = 3 or  
 Number of organising methods used for main trip = 9 [ Mode: 0 ] => Very Satisfied  
**Nationality = 8 or Nationality = 13 or Nationality = 15 or Nationality = 19 or Nationality = 21 or Nationality = 22 [ Mode: 1 ]**  
**Area of the European Union that the destination country belongs = 1 or Area of the European Union that the destination country belongs = 2 or Area of the European Union that the destination country belongs = 8 [ Mode: 1 ]**  
**Type of holiday (package) main trip = 1 or Type of holiday (package) main trip = 3 [ Mode: 0 ] => Very Satisfied**  
**Type of holiday (package) main trip = 2 or Type of holiday (package) main trip = 4 [ Mode: 1 ]**  
 Gender = 1 [ Mode: 1 ] => Less Satisfied  
 Gender = 2 [ Mode: 1 ] => Less Satisfied  
**Area of the European Union that the destination country belongs = 3 or Area of the European Union that the destination country belongs = 4 or Area of the European Union that the destination country belongs = 5 or Area of the European Union that the destination country belongs = 6 or Area of the European Union that the destination country belongs = 7 or Area of the European Union that the destination country belongs IS MISSING [ Mode: 0 ]**  
 Gender = 1 [ Mode: 1 ] => Less Satisfied  
 Gender = 2 [ Mode: 0 ]  
**Party experience any safety issues while using paid accommodation = 0 [ Mode: 0 ] => Very Satisfied**  
**Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]**  
**Duration of personal travels taken in 2014 = 1 or Duration of personal travels taken in 2014 = 2 [ Mode: 1 ]**  
 Age <= 1 [ Mode: 0 ] => Very Satisfied  
 Age > 1 [ Mode: 1 ] => Less Satisfied  
**Duration of personal travels taken in 2014 = 3 [ Mode: 0 ]**  
 Phone available = 1 [ Mode: 1 ] => Less Satisfied  
 Phone available = 3 [ Mode: 0 ] => Very Satisfied  
**Nationality = 9 or Nationality = 17 or Nationality = 18 [ Mode: 0 ]**  
**Duration of personal travels taken in 2014 = 1 [ Mode: 0 ] => Very Satisfied**  
**Duration of personal travels taken in 2014 = 2 or Duration of personal travels taken in 2014 = 3 [ Mode: 0 ]**  
 Camping site (tent, motorhome, caravan, etc.) = 0 or Camping site (tent, motorhome, caravan, etc.) = 3 or Camping site (tent, motorhome, caravan, etc.) = 5 [ Mode: 0 ] => Very Satisfied  
 Camping site (tent, motorhome, caravan, etc.) = 1 or Camping site (tent, motorhome, caravan, etc.) = 2 or Camping site (tent, motorhome, caravan, etc.) = 4 or Camping site (tent, motorhome, caravan, etc.) IS MISSING [ Mode: 0 ] => Very Satisfied  
**Nationality = 10 [ Mode: 1 ]**  
**Phone available = 1 [ Mode: 0 ] => Very Satisfied**  
**Phone available = 2 or Phone available = 3 [ Mode: 1 ]**  
**Type of community = 1 or Type of community = 2 or Type of community IS MISSING [ Mode: 1 ]**  
**Number of reasons to travel <= 1 [ Mode: 0 ]**  
**Type of accommodation main trip (paid or unpaid) = 1 or Type of accommodation main trip (paid or unpaid) = 3 [ Mode: 0 ]**  
 Type of community = 1 [ Mode: 0 ] => Very Satisfied  
 Type of community = 2 [ Mode: 0 ] => Very Satisfied  
**Type of accommodation main trip (paid or unpaid) = 2 or Type of accommodation main trip (paid or unpaid) = 4 [ Mode: 1 ] => Less Satisfied**

**Number of reasons to travel > 1 [ Mode: 1 ]**  
 Gender = 1 [ Mode: 1 ] => Less Satisfied  
 Gender = 2 [ Mode: 1 ] => Less Satisfied  
**Type of community = 3 [ Mode: 1 ] => Less Satisfied**  
**Nationality = 14 or Nationality = 20 or Nationality = 23 or Nationality = 26 [ Mode: 1 ]**  
**Gender = 1 [ Mode: 1 ]**  
**Frequency <= 4 [ Mode: 1 ] => Less Satisfied**  
**Frequency > 4 [ Mode: 1 ]**  
 Currency of the destination country = 1 or Currency of the destination country IS MISSING [ Mode: 1 ] => Less Satisfied  
 Currency of the destination country = 2 [ Mode: 1 ] => Less Satisfied  
**Gender = 2 [ Mode: 1 ]**  
**Party experience any safety issues while using paid accommodation = 0 [ Mode: 1 ]**  
**Type of holiday (package) in 2014 = 1 [ Mode: 0 ] => Very Satisfied**  
**Type of holiday (package) in 2014 = 2 or Type of holiday (package) in 2014 = 3 [ Mode: 1 ]**  
**Occupation = 1 or Occupation = 2 or Occupation = 4 [ Mode: 1 ] => Less Satisfied**  
**Occupation = 3 [ Mode: 0 ]**  
 Number of information sources used by each tourist <= 2 [ Mode: 0 ] => Very Satisfied  
 Number of information sources used by each tourist > 2 [ Mode: 1 ] => Less Satisfied  
**Party experience any safety issues while using paid accommodation = 1 [ Mode: 1 ]**  
 Friends or relatives = 0 or Friends or relatives = 1 or Friends or relatives = 2 or Friends or relatives = 4 [ Mode: 1 ] => Less Satisfied  
 Friends or relatives = 5 [ Mode: 0 ] => Very Satisfied