



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

**O Impacto do *COVID-19* na Satisfação dos Clientes nos Hotéis: uma
Análise de *Text Mining***

Mariana Gonçalves Pontes Jacinto

Mestrado em Gestão

Orientador:

Professor Doutor João Ricardo Paulo Marques Guerreiro, Professor Auxiliar
ISCTE Business School

Julho 2021

Departamento de Marketing, Operações e Gestão Geral

**O Impacto do *COVID-19* na Satisfação dos Clientes nos Hotéis: uma
Análise de *Text Mining***

Mariana Gonçalves Pontes Jacinto

Mestrado em Gestão

Orientador:

Professor Doutor João Ricardo Paulo Marques Guerreiro, Professor Auxiliar
ISCTE Business School

Julho 2021

AGRADECIMENTOS

“É

bem certo que as palavras nunca estão à altura da grandeza dos momentos” (José Saramago).

A finalização da presente dissertação constitui um momento bastante marcante na minha vida, a concretização de um objetivo fulcral na minha formação académica. E, como tal, é crucial agradecer a todos os que contribuíram para que este momento fosse possível.

Em primeiro lugar não posso deixar de agradecer ao meu orientador, Professor João Guerreiro, por todo o apoio que me deu ao longo deste percurso. A disponibilidade, orientação e rápidos esclarecimentos a todas as minhas questões e ainda pela sua partilha de conhecimento contínua.

Aos meus pais, agradeço todas as oportunidades que me proporcionaram e o apoio incessante. À minha irmã, agradeço pelo exemplo de determinação e ambição profissional que representa na minha vida. À minha avó, obrigada pela lição de força.

Gostaria também de expressar um agradecimento especial aos amigos que “ganhei” no mestrado, nomeadamente à Ana Mendes, Carolina Sousa, Francisco Santa Bárbara e Maria Monteiro, pelo companheirismo, apoio e amizade ao longo desta caminhada.

O IMPACTO DO COVID-19 NA SATISFAÇÃO DOS CLIENTES NOS HOTÉIS: UMA ANÁLISE DE TEXT MINING

RESUMO

A pandemia *COVID-19* e, conseqüentemente a crise económica mundial é, atualmente, a temática mais debatida pelo seu impacto sem precedentes, nomeadamente na indústria hoteleira. Atendendo à importância deste setor ao nível económico e social, mas também à sua elevada competitividade, é crucial que os hotéis adotem uma atitude resiliente que vise o seu sucesso, através de clientes satisfeitos.

Esta investigação é elaborada no âmbito da crescente importância da *Web 2.0* enquanto fonte de inovação, que contribui para o desempenho e competitividade da indústria hoteleira. Como tal, visa analisar se a pandemia está a ter repercussões na satisfação dos clientes, refletida através das *online reviews*. É feita uma análise de *Text Mining*, adotando a metodologia *CRISP-DM*, em que são avaliadas e comparadas *online reviews* pertencentes a um intervalo de tempo no decorrer da pandemia com as referentes ao período homólogo, antecedente ao *COVID-19*. Neste estudo são assim comparados os fatores que geram um nível de satisfação mais elevado e aqueles que geram sentimentos mais positivos e negativos por parte dos *reviewers*.

Neste contexto, são analisadas *online reviews* presentes na *Booking.com*, relativas a 64 hotéis localizados em quatro dos principais destinos turísticos da Península Ibérica: Algarve, Costa del Sol, Lisboa e Barcelona. Os resultados demonstram que não existem diferenças significativas no conteúdo textual das *online reviews* nem nos atributos que originam um nível de satisfação superior, representado através de ratings superiores a 7, numa escala de 1 a 10. Contudo, é perceptível o impacto da pandemia através de uma redução expressiva na satisfação dos clientes.

Palavras-Chave: *COVID-19*; *CRISP-DM*; Indústria Hoteleira; *Online reviews*; Satisfação dos Clientes; *Text mining*.

Classificação JEL: L83 Turismo; M31 Marketing.

THE IMPACT OF COVID-19 ON CUSTOMER SATISFACTION IN HOTELS: TEXT MINING ANALYSIS

ABSTRACT

The *COVID-19* pandemic and, consequently, the global economic crisis is, nowadays, the most discussed topic for its unprecedented impact, namely on the hotel industry. Given the economic and social importance of this sector, but also its high competitiveness, it is crucial that hotels adopt a resilient attitude to their success, through satisfied customers.

This research is developed within the growing importance of *Web 2.0* as a source of innovation, which contributes to the performance and competitiveness of the hotel industry. As such, it aims to analyze whether the pandemic is having repercussions on customer satisfaction, reflected through online reviews. A *Text Mining* analysis is performed, adopting the *CRISP-DM* methodology in which online reviews belonging to a time period during the pandemic are evaluated and compared with those referring to the homologous period prior to *COVID-19*. In this study, the factors that generate a higher level of satisfaction and those that generate more positive and negative feelings from reviewers are compared.

In this context, we analyze online reviews from *Booking.com* for 64 hotels located in four of the main tourist destinations of the Iberian Peninsula: Algarve, Costa del Sol, Lisbon and Barcelona. The results show that there are no significant differences in the textual content of online reviews and in the attributes that give rise to a higher level of satisfaction, represented by ratings higher than 7, on a scale from 1 to 10. However, the impact of pandemic is visible through a significant reduction in customer satisfaction.

Keywords: *COVID-19*; *CRISP-DM*; Hotel Industry; Online reviews; Customer Satisfaction; Text mining.

JEL: L83 Tourism; M31 Marketing.

ÍNDICE GERAL

Principais Abreviaturas Utilizadas	xi
Capítulo 1 – Introdução	1
Capítulo 2 – Revisão de Literatura	5
2.1. <i>Web 2.0</i>	5
2.2. <i>User Generated Content</i>	6
2.3. <i>As Online Reviews</i>	7
2.3.1. Características das <i>Online Reviews</i>	9
2.3.2. Satisfação dos Clientes através das <i>Online Reviews</i>	10
2.4. Impacto da Pandemia <i>COVID-19</i> na Hotelaria.....	12
Capítulo 3 – Enquadramento Metodológico	15
3.1. Limitações Metodológicas de Estudos Existentes.....	16
3.2. <i>Text Mining</i>	19
3.3. Metodologia <i>CRISP-DM</i>	20
Capítulo 4 –Aplicação e Resultados.....	23
4.1. Compreensão do Problema.....	23
4.2. Compreensão dos Dados.....	23
4.3. Preparação dos Dados.....	35
4.4. Modelação	43
4.5 Avaliação.....	45
Capítulo 5 – Discussão e Conclusão.....	51
5.1. Principais Resultados	51
5.2. Discussão	53
5.3. Contributos, Limitações e Reflexões para Investigação Futura	54
Referências Bibliográficas	57

Anexo A – Teste para Igualdade de Médias do Rating por Localização.....	69
A.1. Teste para Igualdade de Médias do Rating dos Hotéis Localizados no Algarve.....	69
A.2. Teste para Igualdade de Médias do Rating dos Hotéis Localizados em Barcelona.....	69
A.3. Teste para Igualdade de Médias do Rating dos Hotéis Localizados na Costa del Sol.....	69
A.4. Teste para Igualdade de Médias do Rating dos Hotéis Localizados em Lisboa.....	69
Anexo B – Teste para Igualdade de Médias do Rating por Tipo de Hotel.....	71
B.1. Teste para Igualdade de Médias do Rating em Hotéis pertencentes a Cadeias Hoteleiras.....	71
B.2. Teste para Igualdade de Médias do Rating em Hotéis Independentes.....	71
Anexo C – Teste para Igualdade de Médias do Rating por Categoria de Hotel.....	71
C.1. Teste para Igualdade de Médias do Rating em Hotéis de 2 Estrelas.....	71
C.2. Teste para Igualdade de Médias do Rating em Hotéis de 3 Estrelas.....	71
C.3. Teste para Igualdade de Médias do Rating em Hotéis de 4 Estrelas.....	72
C.4. Teste para Igualdade de Médias do Rating em Hotéis de 5 Estrelas.....	72
Anexo D – Teste para Igualdade de Médias do Rating por Tipo de <i>Reviewer</i>	73
D.1. Teste para Igualdade de Médias do Rating: “ <i>Couple</i> ”.....	73
D.2. Teste para Igualdade de Médias do Rating: “ <i>Family</i> ”.....	73
D.3. Teste para Igualdade de Médias do Rating: “ <i>Group</i> ”.....	73
D.4. Teste para Igualdade de Médias do Rating: “ <i>Solo Traveller</i> ”.....	73
Anexo E – Teste para Igualdade de Médias do Rating por País de Origem	73
E.1. Teste para Igualdade de Médias do Rating: clientes “ <i>Eastern Europe</i> ”.....	73
E.2. Teste para Igualdade de Médias do Rating: clientes “ <i>Northen Europe</i> ”.....	74
E.3. Teste para Igualdade de Médias do Rating: clientes “ <i>Portugal</i> ”.....	74

E.4. Teste para Igualdade de Médias do Rating: clientes “ <i>South and Central America</i> ”.....	74
E.5. Teste para Igualdade de Médias do Rating: clientes “ <i>Southern Europe</i> ”...	74
E.6. Teste para Igualdade de Médias do Rating: clientes “ <i>Spain</i> ”.....	74
E.7. Teste para Igualdade de Médias do Rating: clientes “ <i>Western Europe</i> ”....	75
Anexo F – Teste Não Paramétrico Wilcoxon ao Rating: clientes “<i>North America</i>”	75
Anexo G – Teste para Igualdade de Médias do Rating por Dimensão de Hotel	75
G.1. Teste para Igualdade de Médias do Rating em hotéis com ≤ 50 quartos.....	75
G.2. Teste para Igualdade de Médias do Rating em hotéis com > 50 e ≤ 100 quartos.....	75
G.3. Teste para Igualdade de Médias do Rating em hotéis com > 100 e ≤ 200 quartos.....	76
G.4. Teste para Igualdade de Médias do Rating em hotéis com > 200 e ≤ 300 quartos.....	76
G.5. Teste para Igualdade de Médias do Rating em hotéis com > 300 quartos.....	76
Anexo H – Teste para Igualdade de Médias do Rating por Duração da Estadia.....	77
H.1. Teste para Igualdade de Médias do Rating para estadias de 1 noite.....	77
H.2. Teste para Igualdade de Médias do Rating para estadias de 2 noites.....	77
H.3. Teste para Igualdade de Médias do Rating para estadias de 3 noites.....	77
H.4. Teste para Igualdade de Médias do Rating para estadias de 4 noites.....	77
H.5. Teste para Igualdade de Médias do Rating para estadias de 5 noites.....	77
H.6. Teste para Igualdade de Médias do Rating para estadias de ≥ 6 e < 12 noites.....	78
Anexo I – Top 100: Conceitos Extraídos (2019).....	79
Anexo J – Top 100: Conceitos Extraídos (2020).....	81

Anexo K – “Types” Extraídos (2019 e 2020)..... 83

Anexo L – Exemplos de Conceitos Associados aos Principais “Types”..... 85

Anexo M – Principais Ligações Textuais (2019 e 2020) 87

ÍNDICE DE FIGURAS E TABELAS

FIGURAS

Figura 1 – Metodologia <i>CRISP-DM</i>	21
Figura 2 – Rating Médio das <i>Reviews</i> , por Destino.....	26
Figura 3 – Rating Médio das <i>Reviews</i> , por Tipo de Hotel.....	27
Figura 4 – Rating Médio das <i>Reviews</i> , por Categoria de Hotel.....	28
Figura 5 – <i>Heatmap</i> , por Localização e Categoria do Hotel.....	29
Figura 6 – Rating Médio das <i>Reviews</i> , por Tipo de <i>Reviewer</i>	30
Figura 7 – <i>Heatmap</i> , por Localização e Tipo de <i>Reviewer</i>	30
Figura 8 – Votos de Utilidade por Tamanho Médio das <i>Reviews</i>	31
Figura 9 – Rating Médio das <i>Reviews</i> , por Origem do <i>Reviewer</i>	32
Figura 10 – Rating Médio das <i>Reviews</i> , por Número de Quartos dos Hotéis.....	34
Figura 11 – Rating Médio das <i>Reviews</i> , por Duração da Estadia (em Noites).....	35
Figura 12 – Representação do Processo de Análise de Sentimentos no <i>Orange</i>	36
Figura 13 – Representação da Atribuição de Sentimentos às <i>Reviews</i>	37
Figura 14 – Representação do Processo de Análise de Sentimentos no <i>SPSS</i> <i>Modeler</i>	38
Figura 15 – Padrões de Ligação de Texto “ <i>Location</i> ” e “ <i>Positive</i> ”(2019).....	40
Figura 16 – Distribuição da Polaridade das <i>Reviews</i> (2019), por Categoria.....	42
Figura 17 – Distribuição da Polaridade das <i>Reviews</i> (2020), por Categoria.....	43
Figura 18 – Representação do Processo de Construção dos Modelos.....	45
Figura 19 – Variáveis Mais Importantes na Previsão do Rating (2019).....	46
Figura 20 – Variáveis Mais Importantes na Previsão do Rating (2020).....	47
Figura 21 – Árvore de Decisão C5.0 (2019).....	48
Figura 22 – Árvore de Decisão C5.0 (2020).....	50

TABELAS

Tabela 1 – Contributos para a Análise de <i>Reviews</i> na Indústria Hoteleira.....	17
Tabela 2 – Características dos Hotéis Utilizados na Extração de <i>Reviews</i>	24
Tabela 3 – Volume de <i>Reviews</i> Analisadas, por Ano.....	25
Tabela 4 – Distribuição das <i>Reviews</i> , por Destino.....	26
Tabela 5 – Distribuição das <i>Reviews</i> , por Tipo de Hotel.....	27
Tabela 6 – Distribuição das <i>Reviews</i> , por Categoria de Hotel.....	28
Tabela 7 – Distribuição das <i>Reviews</i> , por Tipo de <i>Reviewer</i>	29
Tabela 8 – Volume de Respostas às <i>Reviews</i>	31
Tabela 9 – Volume de Votos de Utilidade.....	31
Tabela 10 – Distribuição das <i>Reviews</i> , por Origem do <i>Reviewer</i>	32
Tabela 11 – Distribuição das <i>Reviews</i> , por Número de Quartos dos Hotéis.....	33
Tabela 12 – Distribuição das <i>Reviews</i> , por Duração da Estadia (em Noites).....	34
Tabela 13 – Distribuição das <i>Reviews</i> , por Rating.....	36
Tabela 14 – Carga Sentimental das <i>Reviews</i>	37
Tabela 15 – Teste para Igualdade de Médias da Carga Sentimental(2019 e 2020).....	38
Tabela 16 – Distribuição das <i>Reviews</i> , por Categoria.....	41
Tabela 17 – <i>Accuracy</i> dos Modelos Obtidos.....	45
Tabela 18 – Exemplos de Conceitos Associados às Variáveis Presentes na Árvore de Decisão (2019).....	49
Tabela 19 – Exemplos de Conceitos Associados às Variáveis Presentes na Árvore de Decisão (2020).....	50

PRINCIPAIS ABREVIATURAS UTILIZADAS

AHP	– Associação de Hotelaria de Portugal
COVID-19	– <i>Coronavirus Disease 2019</i> / Doença por Coronavírus 2019
CRISP-DM	– <i>Cross Industry Process for Data Mining</i>
E-WOM	– <i>Electronic Word of Mouth</i>
F&B	– <i>Food & Beverage</i>
OTA	– <i>Online Travel Agency</i> / Agência de Viagens Online
PIB	– Produto Interno Bruto
REVPAR	– <i>Revenue per Available Room</i> / Receita por Quarto Disponível
UNWTO	– <i>United Nations World Tourism Organization</i>
UGC	– <i>User Generated Content</i> / Conteúdo Gerado pelo Utilizador

CAPÍTULO 1

INTRODUÇÃO

O novo e atual surto do vírus *COVID-19* foi declarado como pandemia, pela Organização Mundial da Saúde, no passado dia 11 de Março de 2020 (WHO, 2020). Como consequência da pandemia, foram impostas pelos governos inúmeras restrições que forçaram a suspensão de atividade de muitas empresas na indústria do turismo e hotelaria em todo o mundo. A Organização Mundial do Turismo classifica o *COVID-19* como uma crise sanitária, social e económica a uma escala global, sem precedentes (UNWTO, 2020).

De acordo com a Organização para a Cooperação e Desenvolvimento Económico, estima-se que a pandemia tenha provocado uma queda de cerca de 60% a 80% na economia global do turismo (OCDE, 2020). Relativamente a Portugal, os dados da Associação de Hotelaria de Portugal (AHP) evidenciam uma quebra de 73% face a 2019, nas receitas dos hotéis (Simão, 2021). Neste contexto, é possível averiguar que a pandemia está a afetar os recursos financeiros das empresas hoteleiras, pelo que é previsível que os gestores hoteleiros procurem formas de reduzir os custos. Contudo, a redução de custos por si só não será suficiente para as empresas, estas necessitam de adotar uma atitude resiliente para sobreviverem aos desafios impostos pela pandemia.

Segundo Cristina Siza Vieira, CEO da AHP, o verão representou um “balão de oxigénio”, sendo que em Junho 52% dos hotéis ainda tinham a atividade suspensa, decrescendo para 30% em Julho e 28% em Agosto (Magalhães, 2020).

É então, neste contexto, que, conciliando o exposto com as “*MSI Research Priorities 2020-2022*” (Marketing Science Institute, 2020), surge a necessidade de investigar o impacto que a pandemia *COVID-19* está a ter no desempenho dos hotéis, mensurado através da experiência e satisfação dos clientes. Pela sua crescente importância e utilização por parte dos investigadores, e por serem consideradas como um indicador chave para a mensuração da satisfação (Xiang et al., 2015) o presente estudo incidirá na análise de *online reviews*. Com efeito, pretende-se que esta mensuração da satisfação seja, então, efetuada de um modo transparente e justo, focando-se na análise de *online reviews* presentes numa plataforma em que apenas os indivíduos que efetivamente pernottaram num dado hotel, através de uma reserva realizada na plataforma têm a oportunidade de escrever *reviews*.

As *online reviews* têm sido alvo de estudo nos últimos anos porque permitem estudar a performance dos hotéis e respetiva satisfação dos hóspedes, pelo que, a sua análise traduz-se numa oportunidade de melhoria (Lopes, 2015). Por exemplo, Xiang et al. (2015) procederam ao estudo de *online reviews* com o intuito de compreender o conteúdo e estrutura destas e como estão associadas à satisfação dos clientes, através dos ratings atribuídos. Autores como Li et al. (2020) evidenciam diferentes fatores de satisfação para clientes nacionais e internacionais enquanto que, Moro et al. (2020) a diferença entre diferentes tipos de cadeias hoteleiras. A título de exemplo importa ainda destacar a investigação realizada por Zhou et al. (2014) que identifica os atributos chave para a satisfação dos hóspedes e a diferença de ratings atribuídos consoante a classificação (estrelas) e tipo (cadeia ou independente) do hotel e a área geográfica de origem dos *reviewers*. Contudo, têm sido apontadas diversas limitações, no que concerne às investigações existentes relativas à análise de *online reviews* na indústria hoteleira. Estas consistem, sobretudo, na exploração de hotéis de um país ou cidade em exclusivo, maioritariamente em países como a China e os Estados Unidos da América (Lee et al., 2020; Li et al., 2020), a não atualidade das *online reviews* extraídas (Xiang et al., 2015 e Li et al., 2020), a contemplação de hotéis limitados a uma determinada categoria (Zhou et al., 2014) e a cadeias hoteleiras específicas (Moro et al., 2020). Importa ainda realçar que, apesar da crescente utilização das *online reviews* no âmbito do estudo da satisfação dos clientes nos hotéis, “a grande maioria dos estudos existentes baseou-se principalmente em técnicas de investigação convencionais, tais como inquéritos aos consumidores ou entrevistas para avaliar o que resulta na satisfação dos hóspedes” (Xiang et al. 2015, p.121).

Já no que concerne ao contexto pandémico, até à data, foram investigados os efeitos da pandemia no marketing hoteleiro e práticas de gestão, nomeadamente ao nível da inteligência artificial, higiene e cuidados de saúde (Jiang e Wen, 2020), a adoção de novas tecnologias como elemento crítico na recuperação do setor (Ivanov et al., 2020) e redução da perceção de risco dos consumidores (Shin e Kang, 2020). Outros exemplos de investigações já desenvolvidas evidenciam a importância do desenvolvimento de estratégias de preços e investimento nos mercados nacionais (Zhang et al., 2020) bem como, o estudo do comportamento do turista e tomada de decisão perante a pandemia (Kock et al., 2020) e os determinantes que contribuem para a resiliência e desempenho dos hotéis (Melián-Alzola et al., 2020).

Face ao exposto, com o intuito de colmatar esta lacuna, a presente dissertação tem como objetivo principal ***identificar o panorama das online reviews antes do surto do COVID-19 e durante, de modo a compreender se a satisfação dos clientes mudou com a transformação do serviço hoteleiro provocada pela pandemia.*** Adicionalmente, é possível especificar outros objetivos como: (1) identificar se os atributos essenciais na ótica do cliente são os mesmos; (2) constatar se os fatores preditivos do rating se alteraram; (3) verificar se existe alguma diferença na carga sentimental das *reviews*, averiguando, portanto, a capacidade de resiliência dos hotéis face aos desafios impostos.

Atendendo aos objetivos definidos, o método de investigação iniciar-se-á com uma revisão de literatura relativa à temática, no âmbito da *Web 2.0* e da indústria hoteleira. De seguida, será realizado um enquadramento metodológico da abordagem a utilizar na presente dissertação, que consiste numa análise de *text mining* através da metodologia *CRISP-DM*. A metodologia mencionada será aplicada na componente empírica da dissertação. Para tal, serão extraídas as *online reviews* e respetivos dados considerados pertinentes, através da plataforma de reservas *Booking.com*, tendo como finalidade a identificação e reunião de informação necessária que reflita o panorama das *online reviews* em hotéis de quatro dos principais destinos turísticos da Península Ibérica, antes e durante a pandemia *COVID-19*. Para o efeito, recorrer-se-á à utilização dos softwares *IBM SPSS Modeler*, *IBM SPSS Statistics* e *Orange*. Por fim, serão apresentadas as conclusões de toda a investigação.

Como resultado basilar, a dissertação pretende a elaboração de um estudo pioneiro, até à data, de *online reviews* dos hotéis localizados em quatro dos principais destinos turísticos da Península Ibérica, numa ótica comparativa entre dois períodos temporais distintos, o antecedente à pandemia *COVID-19* (2019) e o período correspondente ao contexto pandémico (2020). Deste modo, pretende-se que o estudo: (1) possibilite a estruturação da satisfação dos hóspedes para com os hotéis; (2) constitua um indicador de avaliação do impacto da pandemia na indústria hoteleira; (3) estabeleça uma ferramenta base da aplicação da metodologia *CRISP-DM* na indústria hoteleira, no apoio à gestão, mas também ao nível da investigação na área do Marketing, auxiliando e sustentando investigações futuras. A título complementar, e após a conclusão do estudo, ambiciona-se a publicação dos resultados obtidos em formato de artigo científico, numa revista científica.

Posto isto, a vigente dissertação é constituída por cinco capítulos, dos quais a presente introdução constitui o primeiro capítulo:

- O *Capítulo 2* apresenta a revisão da literatura, que tem como objetivo primordial a constituição do alicerce teórico do estudo em evidência. Para tal, primeiramente, são evidenciados conceitos chave relacionados com as *online reviews*, partindo de uma perspetiva do geral para o particular, na indústria hoteleira. Complementarmente, são evidenciados conceitos inerentes à temática da investigação, tais como o impacto da pandemia no setor.
- O *Capítulo 3* retrata o enquadramento metodológico que sustenta a vertente empírica da presente dissertação. Deste modo são apresentados alguns dos estudos existentes e respetivas metodologias, bem como, os principais conceitos relacionados com *Text Mining* e *CRISP-DM*.
- O *Capítulo 4* expõe a componente empírica desenvolvida, transparecendo a aplicação da metodologia descrita no capítulo antecedente. Assim, neste capítulo é pormenorizado todo o processo de aplicação *de CRISP-DM*.
- O *Capítulo 5*, por último, apresenta a conclusão geral da investigação, na qual são evidenciados os principais resultados inerentes ao estudo e, também uma discussão em que estes são relacionados com a literatura existente. E, por fim, são apresentados os principais contributos da investigação, limitações e sugestões para investigação futura.

CAPÍTULO 2

REVISÃO DE LITERATURA

“A Web 2.0 surge como uma importante fonte de inovação que contribui para o desempenho organizacional e a melhoria da competitividade da indústria do turismo” (Cristobal-Fransi et al., 2020, p.1). É, neste contexto que, a avaliação dos produtos/serviços hoteleiros através da partilha de opiniões tem auferido uma importância cada vez maior no desempenho organizacional e competitividade. Atendendo à atual pandemia, a indústria apresenta agora, mais do que nunca a necessidade de resiliência. No presente capítulo, serão apresentados fundamentos que apoiam a necessidade de analisar as *online reviews* dos hotéis, bem como a necessidade de adoção de resiliência por parte das empresas, perante o impacto que a presente conjuntura está a ter na indústria hoteleira.

2.1. *Web 2.0*

A elevada difusão da Internet a que se tem assistido na última década e os correspondentes avanços nas tecnologias de informação reforçaram o desenvolvimento da *Web 2.0* (Wang & Zhang, 2012). A *Web 2.0* consiste numa versão da *Web* que permite conectar comunidades com interesses semelhantes e funciona como facilitadora de interações bidirecionais (Filiari et al., 2015). Assim, os utilizadores deixam de ser meros destinatários de informação e passam a desempenhar o papel de criadores e participantes na criação de informação (Moro et al., 2019). Enquanto que a *Web 1.0* se caracteriza pelo seu teor estático, a *Web 2.0* é caracterizada por serviços que, no fundo, permitem a criação de *networks* que são cada vez mais populares na Internet, como por exemplo as de carácter social ou profissional, como é o caso do *Facebook* e do *LinkedIn* (Sheng, 2012).

Autores como Pateli et al. (2020), referem-se às tecnologias *Web 2.0* como uma ferramenta fundamental para a melhoria da pesquisa de informação e experiência do consumidor, bem como enriquecer a estratégia digital de uma empresa. Em concordância com o descrito por O'Connor (2010), a *Web 2.0* redefiniu a forma de os utilizadores criarem, partilharem e utilizarem informação *online*.

As aplicações *Web 2.0* alteraram de forma significativa o modo pelo qual as empresas turísticas, como por exemplo os hotéis, criam valor para os seus clientes ao providenciarem acesso *online* a informação relativa com as reservas, assim como a avaliação dos seus produtos/serviços turísticos, através da partilha de opiniões e experiências (Gonçalves et al., 2018). Deste modo, é possível constatar que a *Web 2.0*, enquanto importante fonte de inovação, contribui para a melhoria da competitividade da indústria hoteleira e respetivo desempenho organizacional (Cristobal-Fransi et al., 2020).

De acordo com Marine-Roig e Huertas (2020), a *Web 2.0* tem sido caracterizada por um vasto desenvolvimento de “*social networks*” e, por conseguinte, por um aumento exponencial de *user-generated content (UGC)*, dependendo deste para reter os seus utilizadores.

2.2. *User Generated Content*

O *UGC* diz respeito a qualquer conteúdo criado por utilizadores na Internet (Babić Rosario et al., 2020), podendo adotar vários formatos como por exemplo publicações no *social media* ou *online reviews* (Mariani et al., 2019).

De acordo com Bronner e De Hoog (2011), o *UGC* revolucionou a forma como os consumidores procuram informação, dado que se trata de uma das fontes de informação mais importantes (Del Chiappa et al., 2015). Assim, o *UGC* tem tido uma importância crescente na compreensão da experiência dos consumidores, atendendo ao facto de ser responsável pela geração de uma grande quantidade de dados (Shin et al., 2019; Xiang et al., 2015).

O *UGC* é bastante reconhecido como uma fonte de informação válida e credível tanto para os turistas como para os investigadores e gestores (Lu & Stepchenkova, 2015). No turismo e hotelaria, o *UGC* é utilizado pelos turistas para reduzir a incerteza e risco associados à dificuldade de acesso à qualidade de serviço proveniente do teor intangível dos produtos e/ou serviços (Casaló et al., 2015). Neste sentido, existe uma evidência de que o *UGC* tem influência na tomada de decisão dos turistas, de acordo com Kim et al. (2011) e Mauri e Minazzi (2013). No que diz respeito à utilização de *UGC* por parte dos investigadores, é de realçar a sua utilidade crescente para estudos relacionados com a satisfação dos hóspedes nos hotéis, destacando-se a título de exemplo os que foram desenvolvidos por Xiang et al. (2015) e Zhou et al. (2014).

As principais fontes de *UGC* relacionado com viagens, ao qual se atribui frequentemente a denominação de *TGC (travel generated content)*, são os websites relacionados com viagens como o *TripAdvisor*, plataformas *P2P* como o *Airbnb* e agências de viagem online (*OTA's*) como por exemplo a *Booking.com* (Martin-Fuentes et al., 2018).

É através do *UGC*, que os clientes expõem livremente o que experienciam e julgam acerca de um produto e/ ou serviço (Zhou et al., 2014), auxiliando outros turistas antes, durante e após a sua viagem (Mauri & Minazzi, 2013).

2.3. As Online Reviews

Na era atual, do *e-commerce*, são cada vez mais os clientes que publicam *online reviews* de produtos e /ou serviços posteriormente ao seu consumo (Xiang et al., 2015). As *online reviews*, segundo Xu (2019), funcionam então como um driver para o desenvolvimento do efeito *electronic word of mouth (eWOM)*. Pyle et al. (2021) consideram que o *eWOM* no contexto das *online reviews* pode ser visto como um sistema de interconexões que se divide em três níveis que são os sites, os *reviewers* e as *reviews* em si.

É possível definir as *online reviews* como “*um tipo de informação sobre um produto criada pelos próprios utilizadores com base na sua experiência pessoal de utilização*” (Chen & Xie, 2008, p.477), sendo que as suas duas principais finalidades se traduzem no fornecimento de informações acerca de um produto e/ou serviço e na elaboração de recomendações (Kwok et al., 2017). Portanto, de acordo com Wattanacharoensil et al. (2017) é importante que os *reviewers* procurem escrever da melhor forma que represente a sua experiência, em virtude de as *online reviews* serem habitualmente livres de regras que impõem tópicos obrigatórios.

Para Chen et al. (2019) as *online reviews* e respetivas respostas consistem numa comunicação entre os consumidores e os seus pares, mas também entre os gestores e os consumidores, podendo ocorrer em diferentes tipos de plataformas: *social media*, plataformas *third-party booking* e plataformas diretas (Xu & Lee, 2020). As *online reviews* dos hotéis são um dos mais importantes tipos de *UGC* (Xiang et al., 2015) e desempenham um papel crucial no e-commerce (de Gregorio et al., 2021). É neste sentido que, Mariani e Borghi (2018) apontam o desenvolvimento e consolidação das agências

de viagem online (*OTA's*), como a principal razão da difusão de *online reviews* relacionadas com o turismo e hotelaria, das quais se destacam a *Booking.com* e a *Expedia*. Assim, é possível afirmar que as *online reviews* constituem uma importante fonte de partilha de conhecimento no contexto da gestão hoteleira, quer para os consumidores quer para os hoteleiros (Xu, 2019).

Chakraborty e Biswal (2020) referem que a leitura de várias *online reviews* de um determinado hotel cria uma impressão de confiança e fiabilidade nos potenciais clientes e, conseqüentemente, tem um impacto significativo na sua decisão de compra. É então evidente a especial importância das *online reviews* na indústria hoteleira, em virtude de os consumidores serem, na maioria das vezes, confrontados com a necessidade de reservarem um hotel que não conhecem (Gonçalves et al., 2018). No fundo, os consumidores leem *reviews* para reduzir o risco e incerteza, dado que estas lhes permitem o acesso à experiência de se hospedarem num determinado hotel sem terem pernoitado na realidade (Browning et al., 2013). É de constatar o facto de o número de consumidores que confia neste tipo de informação no âmbito da preparação de uma viagem, ser cada vez mais elevado (Browning et al., 2013; Mauri & Minazzi, 2013; Serra Cantallops & Salvi, 2014), com o intuito de tomarem decisões de compra informadas (Fang et al., 2016). Importa também referir que as *reviews* têm uma influência superior nos consumidores que utilizam frequentemente a Internet, pois já compreendem os benefícios e fiabilidade da utilização da mesma (Zhu & Zhang, 2010).

As *online reviews* são bastante relevantes para os hoteleiros visto que permitem a aquisição de informação essencial para a compreensão da preferência dos seus clientes (Sanchez-Franco & Rondan-Cataluña, 2010) e respetiva melhoria operacional dos hotéis (Lopes, 2015). As *reviews* são também pertinentes para os gestores hoteleiros devido ao impacto duradouro que têm na reputação das empresas (Hennig-Thurau et al., 2004), na medida em que permanecem nos *websites* por um longo período de tempo. É, portanto, neste âmbito da gestão da reputação que, a resposta da gestão às *online reviews* tem sido apontada como uma das estratégias mais eficazes, transmitindo a ideia de que os gestores se esforçaram para melhorar ou corrigir os seus produtos e/ou serviços (Baka, 2016; Proserpio & Zervas, 2017; Xie et al., 2017). Torres et al. (2015) acrescentam ainda o facto de as *reviews* funcionarem como uma ferramenta para os proprietários dos hotéis responsabilizarem as equipas de gestão.

Importa também ressaltar que existem motivações distintas para diferentes consumidores (Xu & Lee, 2020). Bronner e De Hog (2011) destacam cinco categorias de

motivação para publicar *online reviews*, nomeadamente as pessoais (auto-direcionadas), a preocupação social (ajudar os outros), benefícios sociais, *empowerment* do consumidor e por último, ajudar as empresas. No entanto, para Gonçalves et al. (2018), são quatro as categorias de motivações dos consumidores para escrever *online reviews*, nomeadamente, motivações pessoais, preocupação social (para ajudar outros viajantes, benefícios sociais e *empowerment* dos consumidores. Maioritariamente, os consumidores escrevem *reviews* com o intuito de informarem os pares acerca da sua experiência de alojamento e para partilharem o seu nível de satisfação e insatisfação, a qualidade de serviço percebida, o desempenho dos atributos do hotel e a conveniência da plataforma (Park & Allen, 2013; Serra Cantallops & Salvi, 2014; Yen & Tang, 2019). Já Kudeshia e Kumar (2017), sustentam a ideia de que a atitude do consumidor em relação a uma marca age como um driver de motivação para escrever uma *review*.

Além das motivações já mencionadas, Mariani et al. (2019) acrescentam também o grau de envolvimento com a compra e características demográficas como exemplos de antecedentes de *online reviews*.

2.3.1. Características das *Online Reviews*

Os indicadores essenciais para compreender os efeitos provenientes do conteúdo gerado pelos utilizadores através das *online reviews* são o volume e a valência (Babić Rosario et al., 2016).

O volume das *reviews* refere-se normalmente ao número de *reviews* existentes, refletindo, por isso, a popularidade do produto (Zhang et al., 2013). Esta é uma característica importante relativamente às *reviews* dado que existe uma tendência para os consumidores preferirem hotéis com um grande volume de *reviews* (Gavilan et al., 2018), e, portanto, Ye et al. (2009) constatam que o volume de *reviews* pode auxiliar na previsão de vendas.

No que concerne à valência, esta diz respeito à média das classificações/ratings das *reviews* (Babić Rosario et al., 2016), caracterizando-se por ser óbvia e acessível é, geralmente, o indicador ao qual os viajantes tendem a dar mais importância quando são confrontados com um elevado volume de informação e procuram tomar decisões eficazes sem despende muito tempo (Browning et al., 2013). De acordo com Xie et al. (2014), a

valência das *reviews* diz respeito ao quão os consumidores gostaram da sua experiência com a empresa, podendo ser positiva, neutra ou negativa. Neste âmbito, é possível afirmar que as *reviews* negativas contribuem para uma redução no número de reservas efetuadas online (Ye *et al.*, 2009) e podem também permitir que os gestores hoteleiros direcionem a sua atenção para aspetos menos positivos mencionados pelos clientes (Torres *et al.*, 2015). Já no que diz respeito às *reviews* positivas, estas aumentam significativamente a intenção de reserva de um determinado hotel (Ye *et al.*, 2009) e, segundo os autores Bronner e De Hoog (2011), auxiliam os consumidores a evitar escolher um hotel que possa não corresponder às suas expectativas. Ainda relativamente à valência das *reviews* é de realçar o facto de Gao *et al.* (2018) argumentarem que os hotéis pertencentes a uma cadeia hoteleira terem geralmente ratings mais baixos do que os hotéis independentes, enquanto que Lee *et al.* (2020) revelam o oposto.

Além do volume e valência, a variância é também uma característica importante das *online reviews*, que permite averiguar a heterogeneidade das opiniões dos clientes (Jiménez & Mendoza, 2013). Uma elevada variabilidade dos ratings aumenta a incerteza relativamente à qualidade e resulta numa redução de vendas, enquanto que uma variabilidade inferior estabelece uma consistência nas avaliações dos clientes em relação a um determinado produto ou serviço (Babić Rosario *et al.*, 2016).

As *online reviews* devem ainda apresentar qualidade, que, segundo Cheung *et al.* (2009), esta pode ser medida através da sua utilidade, que é normalmente mensurada através de votos de utilidade nos *websites* que disponibilizam *online reviews*. Assim, com o fim de uma *review* ser percecionada como de alta qualidade e útil deve incluir detalhes acerca dos produtos e/ou serviços e das respetivas empresas que os providenciam (Park & Nicolau, 2015). Chatterjee (2020) acrescenta também que as *reviews* devem apresentar veracidade e atualidade.

2.3.2. Satisfação dos clientes através das *Online Reviews*

Com o intuito de avaliar a qualidade de serviço e das práticas de gestão, identificar os fatores que devem ser melhorados, obter uma maior taxa de retenção e, portanto, alcançar o sucesso, é crucial para as empresas a mensuração da satisfação dos clientes (Su, 2004).

Em função da elevada competitividade que caracteriza a hotelaria, Xiang *et al.* (2015) consideram que existe uma ambição crescente relativa aos hotéis se distinguirem

uns dos outros e, como tal, a satisfação do cliente é um dos indicadores chave da eficácia de um hotel, quando comparado com o seu *competitive set* (que se traduz no grupo dos seus principais concorrentes). Contudo, estes mesmos autores referem que, atendendo à complexidade da experiência do hóspede, medir e gerir a satisfação do cliente é uma tarefa desafiante. Assim, é neste sentido que tem existido um esforço crescente em usar *UGC*, nomeadamente através das *online reviews* (Rhee & Yang, 2015), para medir/avaliar as experiências turísticas e respetiva satisfação do hóspede (Xu et al., 2019).

De acordo com Xu e Li (2016), as *reviews* positivas expressam satisfação dos clientes relativamente à sua experiência, enquanto que as *reviews* negativas manifestam insatisfação. Posto isto, a principal finalidade da utilização das *online reviews* é a identificação de atributos dos hotéis essenciais na ótica do cliente (Sainaghi et al., 2018), permitindo assim uma melhoria na qualidade de serviço, aumento de satisfação, acréscimo na procura, desempenho financeiro superior e consequente obtenção de vantagem competitiva (Sparks & Browning, 2011; Verma & Chandra, 2018).

Xu e Lee (2020) argumentam que os atributos sobre os quais os clientes tendem a refletir a sua satisfação nas *online reviews* correspondem ao quarto, ao *staff* e serviço, à localização e acesso e, por último, às operações e instalações. Já Lee et al. (2020) consideram que o serviço, o quarto e o valor são os três atributos básicos que afetam a satisfação dos clientes, pelo que os hoteleiros devem assegurar que providenciam sempre um serviço de alta qualidade, que o quarto tem as comodidades adequadas e que o preço dos seus produtos é estabelecido num nível apropriado de forma a que o cliente percecionasse valor. A qualidade do departamento de *Food & Beverage* (Zhou et al., 2014) e a qualidade do sono (Mao et al., 2018) são também considerados antecedentes importantes da satisfação. Miles et al. (2012) referem ainda que, perceções de limpeza mais positivas estão associadas a perceções mais positivas de qualidade de serviço e elevados níveis de satisfação.

No que concerne aos determinantes da insatisfação, de acordo com Xu e Li (2016), estes estão principalmente associados a problemas das instalações do hotel, como por exemplo a idade e o funcionamento defeituoso ou inexistência de certos equipamentos; problemas operacionais como limpeza ineficaz, existência de ruído e estacionamento sem capacidade; e problemas comportamentais, tal como um *staff* pouco prestável. Visto que atualmente a *Internet* é uma necessidade básica, particularmente para os indivíduos que

viajam em trabalho, o mau funcionamento do *wi-fi* é também considerado como um antecedente da insatisfação (Lee et al., 2020).

Contudo, importa referir que os atributos dependem do tipo de hotel (Xu & Li, 2016; Moro et al., 2020) e têm pesos distintos (Oh & Jeong, 2010), assim como o facto de as diferenças no país de origem dos clientes terem um impacto na sua satisfação (Tse & Ho, 2009; Li et al., 2020).

Por conseguinte, atendendo à importância da satisfação dos clientes na hotelaria, surge a necessidade de averiguar se os níveis de satisfação foram ou não afetados pela pandemia, visto que “a atual pandemia pode constituir uma mudança de paradigma na investigação relacionada com o comportamento e tomada de decisões dos turistas. O que foi anteriormente dado como adquirido pode não ser mais na era *COVID-19*” (Kock et al., 2020, p.2).

2.4. Impacto da pandemia *COVID-19* na hotelaria

A pandemia *COVID-19* afetou severamente o tecido empresarial mundial (Eggers, 2020). É, portanto, nesse contexto que Duarte Alonso et al. (2020) referem-se à pandemia como a mais recente prova a numerosos líderes, empreendedores e trabalhadores de praticamente todas as indústrias, das quais se destaca a indústria hoteleira, que foi particularmente afetada (Nicola et al., 2020).

No contexto pandémico atual, as previsões demonstram que o PIB de muitos países irá diminuir bastante num futuro próximo (Cohen, 2020). Segundo Smeral (2010) a uma quebra no rendimento dos consumidores está normalmente associada uma queda no consumo de serviços turísticos, logo, é também possível prever que irá existir uma diminuição profunda no consumo destes serviços (Napierała et al., 2020). Contudo, não se trata meramente de uma previsão, pois o efeito adverso da pandemia no consumo turístico é visível através da diminuição das taxas de ocupação dos hotéis e respetivas receitas (Naumov et al., 2020). Neste contexto, de acordo com a UNWTO (2021) em Janeiro de 2021 verificou-se uma queda de 87% nas chegadas de turistas internacionais, comparativamente ao período homólogo, no ano antecedente. Importa referir que o *COVID-19* não só teve efeitos económicos como operacionais, dado que os governos impuseram restrições de viagens, por considerarem que existe uma relação entre o risco de infeção e o consumo turístico (Yang et al., 2020). Deste modo, foram inúmeras as

empresas relacionadas com a hotelaria e turismo ao nível mundial que foram forçadas a suspender a sua atividade como é o caso das companhias aéreas, hotéis e restaurantes, por exemplo (Ivanov et al., 2020).

Considerando a definição de desastre estabelecida por Faulkner (2001, p.136) “desastre pode ser definido como a situação em que uma empresa é confrontada com mudanças catastróficas imprevisíveis sobre as quais tem pouco controlo”, um surto de um vírus como o *SARS* em 2003 e por conseguinte, o atual *SARS-CoV-2*, pode ser classificado como um desastre na indústria hoteleira mundial (Lo et al., 2006).

O advento da pandemia *COVID-19* estabeleceu uma nova realidade para o setor do turismo dado que, os testes obrigatórios, as quarentenas e o encerramento de fronteiras continuam a vigorar em todas as regiões do mundo (UNWTO, 2021). O mesmo aplica-se também às empresas hoteleiras, visto que, embora o serviço *core* não tenha alterado, o método de entrega do serviço está a ser alvo de uma transformação (Seyitoğlu & Ivanov, 2020). O distanciamento físico, máscaras e acrílicos transparentes, por exemplo, passaram a caracterizar o serviço hoteleiro, em especial nos departamentos de *front-office* (Seyitoğlu & Ivanov, 2020; Sigala, 2020). Deste modo, as medidas anti-*COVID-19* adotadas pelos hotéis poderão constituir um atributo sobre o qual os clientes refletem a sua satisfação através das *online reviews*. A importância da higiene e limpeza dos hotéis tornou-se particularmente fundamental (Jiang & Wen, 2020; Naumov et al., 2020), em virtude de a *World Health Organization* (WHO, 2020) alegar que existe a possibilidade de o *COVID-19* se propagar diretamente entre as pessoas, mas também indiretamente através de objetos e superfícies contaminadas. Ou seja, se a higiene e limpeza dos hotéis já era uma questão crucial na obtenção de clientes satisfeitos, agora com o advento da pandemia será ainda mais tal como os autores referem e, conseqüentemente, poderá existir um aumento da expressão de satisfação ou insatisfação para com este atributo.

O *COVID-19* instaurou, assim, uma necessidade crescente de os hotéis aumentarem a sua resiliência, conceito ao qual estão inerentes inúmeras definições na literatura existente. Herbane (2019) define a resiliência como um indicador de preparação e capacidade para enfrentar uma crise que, todavia, não pode ser apontada como uma garantia de recuperação bem-sucedida. Já, no que diz respeito aos autores Duchek et al. (2020), a resiliência pode ser considerada como um resultado ou como um processo. Segundo estes autores, enquanto que a resiliência como um resultado consiste na capacidade de recuperação e só pode ser avaliada após acontecimentos disruptivos, a

resiliência como processo implica lidar com eventos adversos de forma eficaz, não só após a sua ocorrência, mas também no seu decorrer. Ou seja, uma organização melhora a sua resiliência, somente, se desenvolver capacidades adaptativas, flexibilidade e promover a inovação e autoeficácia (Brown et al., 2017).

A exposição a um elevado grau de incerteza e mudança da hotelaria resulta num crescente interesse relativo à resiliência na referida indústria (Filimonau & De Coteau, 2020; Prayag, 2018). É, portanto, neste contexto que, Melián-Alzola et al. (2020) consideram que os gestores hoteleiros devem estar previamente preparados de modo a enfrentar os desafios, como a atual pandemia, e maximizar as oportunidades, mediante uma situação de mudança inesperada ao invés de apenas reagirem quando estas ocorrem. De acordo com Brown et al. (2017), a resiliência de um hotel consiste na capacidade da organização e respetivos *stakeholders*, de se adaptarem, inovarem e superarem potenciais transtornos. No que diz respeito à resiliência perante um determinado acontecimento, os hotéis devem constituir atividades de preparação, mas também de redução de risco, que incluem análises estruturais e não estruturais que analisam os fatores operacionais e de serviço (Brown et al., 2017). Como tal, Melián-Alzola et al. (2020) argumentam que um hotel para ser resiliente carece de dois fatores, nomeadamente de uma estratégia, isto é, de um plano de negócios, mas também de mudança, visando uma melhoria contínua através das mudanças necessárias ao plano inicialmente definido. Estes autores consideram assim que, a resiliência tem uma influência positiva em diversos indicadores de desempenho hoteleiro nomeadamente ao nível da lealdade dos clientes, por exemplo, e, por isso, os hotéis podem obter vantagens competitivas através da sua capacidade de adaptação e agilidade para com a mudança.

Por conseguinte, perante o atual contexto pandémico, é imprescindível a adoção de resiliência através de novas práticas operacionais por parte dos hotéis, com o intuito de alcançarem uma recuperação eficaz (Jiang & Wen, 2020) e concederem uma sensação de segurança imediata aos seus clientes (Shin & Kang, 2020). É, neste âmbito, que têm sido desenvolvidas iniciativas por parte de instituições oficiais para motivarem a adoção de novas práticas e, simultaneamente, a conceção de uma sensação de segurança aos hóspedes. A título exemplificativo, é possível mencionar o selo “*Clean & Safe*” desenvolvido pelo Turismo de Portugal que confere aos estabelecimentos hoteleiros o reconhecimento de cumprimento das recomendações da Direção Geral de Saúde (DGS), para evitar a propagação do *COVID-19*. No entanto, Tsionas (2020) alerta para o facto de que não vão ser todas as empresas hoteleiras a conseguir superar os desafios impostos

pelo *COVID-19*, uma vez que os hotéis reabriram com taxas de ocupação inferiores, mas com os mesmos custos fixos. Presumivelmente, a quebra nas taxas de ocupação poderá ser evidenciada através da redução do volume de *online reviews*. É, portanto, nesta ótica que é substancial a adoção de resiliência, visando assim a sua competitividade e consequente sobrevivência (Lee et al., 2013).

Em oposição à maioria das pandemias anteriores, que foram de curto-prazo (Gössling et al., 2020), a previsão é de que o surto do *COVID-19* tenha efeitos de longo prazo na indústria hoteleira (Ying et al., 2020), estando dependente da velocidade da distribuição mundial das vacinas, entretanto desenvolvidas (UNWTO, 2021). É, portanto, neste contexto que existe um consenso global em considerar que o impacto da pandemia no turismo e hotelaria é inédito e sem precedentes (Kock et al., 2020; UNWTO, 2020).

Assim, tendo por base o presente capítulo é possível definir um conjunto proposições, no qual esta investigação se pretende basear: (1) Existem diferenças nos principais indicadores das *online reviews* (volume e valência), perante a pandemia, consoante as características dos hotéis em estudo (destino, tipo, categoria e dimensão), características demográficas dos clientes (país de origem e propósito de estadia), bem como a duração da estadia; (2) O conteúdo textual das *online reviews* nomeadamente ao nível dos atributos mais abordados pelos clientes e respetivos pesos, sofreu uma alteração perante o despoletar da pandemia *COVID-19*, enfatizando-se a importância da limpeza; (3) Diminuição da satisfação dos clientes, como consequência da pandemia; (4) Os fatores mais relevantes na previsão de ratings mais elevados sofreram alterações, perante o contexto pandémico.

CAPÍTULO 3

ENQUADRAMENTO METODOLÓGICO

O capítulo anterior salientou a utilidade do estudo das *online reviews* no contexto da indústria hoteleira, assim como uma contextualização do período temporal representado pelos dados que serão analisados na presente investigação. O setor hoteleiro, necessita mais do que nunca assumir uma posição resiliente. Apesar do impacto sem precedentes que a pandemia está a provocar ao nível económico, a indústria hoteleira assume uma grande importância a nível mundial. Neste sentido, é essencial, mais do que nunca,

investir em investigação pertinente com metodologias adequadas, de modo a facilitar a tomada de decisão nos hotéis. Assim, neste capítulo serão então abordadas as principais limitações ao nível metodológico de estudos existentes bem como as temáticas de *text mining* e *CRISP-DM*.

3.1. Limitações Metodológicas de Estudos Existentes

A investigação no âmbito das *online reviews* da indústria hoteleira está em constante evolução, pelo que, têm sido muitas as contribuições para um melhor entendimento da sua utilidade. A análise de *online reviews* tem sido efetuada de diversas formas, não existindo uma metodologia universalmente aceite. Nesse sentido, a *Tabela 1*, auxilia na compreensão do panorama geral de diversos estudos relacionados com a temática da presente dissertação.

Através da *Tabela 1* é possível apontar como principais limitações: (1) a área geográfica investigada; (2) o período temporal; (3) variabilidade do método utilizado. No que concerne à área geográfica é perceptível que a maioria dos estudos se centra em hotéis localizados na China e nos Estados Unidos da América. Grande parte dos estudos selecionados é recente, mas, no entanto, os dados extraídos já têm vários anos, pelo que podem já não representar o cenário atual. E, por último, apesar de existir uma tendência crescente na utilização de técnicas de *business analytics* nesta área de estudo, a maioria recorre a análises de regressão.

Assim sendo, a presente investigação é desenvolvida com o propósito de colmatar as principais limitações mencionadas, através da aplicação de uma análise de *text mining* adotando a metodologia *CRISP-DM*, com foco numa localização geográfica ainda pouco explorada e, simultaneamente com uma elevada recentidade.

AUTOR	AMOSTRA	MÉTODO	CONTRIBUIÇÃO	LIMITAÇÕES RECONHECIDAS PELOS AUTORES
Zhou et al. (2014)	<ul style="list-style-type: none"> • Fonte de extração: Agoda.com • Data: 2012 • Localização: China 	<p><i>Text Mining;</i> ANOVA e Scheffe</p> <p>Software: Não especificado</p>	<p>Identifica vinte e três atributos chave nas <i>reviews</i> para a satisfação dos clientes, a diferença de ratings atribuídos consoante a classificação(estrelas) e tipo (cadeia ou independente) do hotel e a área geográfica de origem dos hóspedes.</p>	<ul style="list-style-type: none"> • Contempla apenas os hotéis de quatro e cinco estrelas. • Limitação geográfica a uma cidade chinesa (Hangzhou). • Não considera as <i>reviews</i> em todos os idiomas relevantes (coreano e japonês).
Xiang et al. (2015)	<ul style="list-style-type: none"> • Fonte de extração: Expedia.com • Data: 2007 • Localização: Estados Unidos da América 	<p><i>Text Mining</i> Análise Factorial; ANOVA</p> <p>Software: Microsoft Access; Microsoft Office Excel</p>	<p>Investiga o conteúdo e estrutura das <i>reviews</i> e como estas estão associadas com a satisfação do hóspede. Considera o contexto da estadia, a avaliação verbal dos clientes e a expressão de possíveis futuras ações. Evidencia a alta dependência do fator higiene para com o nível de satisfação.</p>	<ul style="list-style-type: none"> • O estudo representa apenas um país. • Os atributos identificados refletem a percepção de aspetos relacionados com a localização dos hotéis. • Os dados foram recolhidos vários anos antes do estudo.
Xu e Li (2016)	<ul style="list-style-type: none"> • Fonte de extração: Booking.com • Data: Não especificada • Localização: Estados Unidos da América 	<p><i>Text Mining;</i> Análise Factorial</p> <p>Software: RapidMiner</p>	<p>Evidencia a diferença da importância dos fatores de satisfação e insatisfação consoante o tipo de hotel (<i>full service, limited service e suite</i>). Demonstra a importância da qualidade do quarto enquanto fator chave para a satisfação.</p>	<ul style="list-style-type: none"> • O estudo não considera os dados demográficos dos clientes.
Lee et al. (2020)	<ul style="list-style-type: none"> • Fonte de extração: TripAdvisor • Data: 2004-2015 	<p><i>Text Mining;</i> Regressão Linear Múltipla</p> <p>Software: LIWC2015; XLMiner</p>	<p>Analisa o serviço, quarto e valor, como os três atributos mais importantes que afetam a satisfação dos clientes.</p>	<ul style="list-style-type: none"> • Limitação geográfica a uma cidade (Nova Iorque). • Foco na identificação de fatores significantes que

	<ul style="list-style-type: none"> • Localização: Estados Unidos da América 		Demonstra a eficácia da utilização das técnicas de <i>Business Analytics</i> aplicadas ao setor hoteleiro.	afetam a satisfação dos clientes.
Li et al. (2020)	<ul style="list-style-type: none"> • Fonte de extração: TripAdvisor • Data: até 2014 • Localização: China 	<p><i>Text Mining;</i> Regressão Linear Múltipla Software: R</p>	<p>Analisa a influência que o país de origem dos hóspedes tem nas suas expectativas de desempenho dos hotéis. Diferencia diferentes fatores de satisfação para clientes nacionais e internacionais.</p>	<ul style="list-style-type: none"> • Limitação geográfica aos hotéis chineses. • O estudo foca-se em apenas 5 atributos dos hotéis. • Análise limitada aos dados do TripAdvisor até 2014.
Moro et al. (2020)	<ul style="list-style-type: none"> • Fonte de extração: TripAdvisor • Data: 2005-2018 • Localização: Amesterdão, Bruxelas, Frankfurt, Londres e Paris. 	<p><i>Text Mining;</i> Heat map Software: R</p>	<p>Evidencia a diferença entre cadeias hoteleiras “<i>high-end</i>” (representadas pela <i>Sheraton</i>) e “<i>low-end</i>” (representadas pela <i>Ibis</i>). Clarifica a importância da qualidade de serviço e os fatores mais importantes em hotéis de aeroporto.</p>	<ul style="list-style-type: none"> • Limita-se apenas a duas cadeias hoteleiras, não podendo ser generalizado. • Análise limitada aos dados disponíveis no TripAdvisor.
Xu e Lee (2020)	<ul style="list-style-type: none"> • Fonte de extração: TripAdvisor, Expedia.com e Marriot.com • Data: Não especificada • Localização: Estados Unidos da América 	<p><i>Text Mining;</i> Regressão Logística Software: SentiStrength; RapidMiner</p>	<p>Analisa o conteúdo e estilo linguístico das reviews em diferentes tipos de plataformas, evidenciando que os determinantes da satisfação não diferem.</p>	<ul style="list-style-type: none"> • O estudo foca-se apenas nas características linguísticas e atributos abordados nas respostas da gestão às <i>reviews</i>.

Tabela 1: Contributos para a Análise de *Reviews* na Indústria Hoteleira

Fonte: Elaboração Própria

3.2. *Text Mining*

O *data mining* é, cada vez mais, uma ferramenta imprescindível para transformar os dados em conhecimento (Berka & Rauch, 2010). Segundo os autores, o *data mining* intenta a aquisição de conhecimento subentendido, mediante a análise dos dados para desenvolver modelos de classificação e predição, como suporte às decisões.

De acordo com Fan et al. (2006) *text mining* é um tipo específico de *data mining* que consiste na análise de conteúdo textual inserido em documentos, comentários, *reviews* ou qualquer outro tipo de informação relacionada, de forma a transformá-lo em conhecimento ativo. Jain et al. (2013) afirmam que o *text mining* permite às empresas a compreensão das opiniões dos clientes a respeito da empresa em si e dos seus produtos e/ou serviços de modo a que possam melhorar a sua estratégia de marketing e respetivo posicionamento no mercado. O *text mining* pode ser aplicado para a gestão da reputação, relações públicas e previsão de tendências de mercado (Xiang et al., 2015).

O *text mining* trata-se, assim, de uma técnica que, no que concerne à análise de *online reviews*, ao contrário dos estudos que utilizam métodos qualitativos, viabiliza uma abordagem mais objetiva atendendo às suas características matemáticas (Xu & Li, 2016). É neste contexto, portanto, que tem atraído cada vez mais atenção académica (Goh et al., 2013) e sido utilizado em diversas investigações importantes associadas ao relacionamento dos clientes com as organizações, nomeadamente através das *online reviews* (Xu & Li, 2016).

Calheiros et al. (2017) consideram que o *text mining* implica, primeiramente, dois processos que dizem respeito à eliminação de palavras irrelevantes, como por exemplo advérbios e artigos, e *stemming*, com o intuito de reduzir as palavras derivadas para uma única palavra, exemplificando a redução da palavra “*feelings*” para “*feel*”.

Segundo os autores Liu et al. (2005) uma análise de *text mining* inclui principalmente tarefas como categorização, agrupamento de texto e análise de sentimentos, que é utilizada para identificar a polaridade dos dados classificando-os como positivos, negativos ou neutros. É neste sentido que, Xiang et al. (2015) referem-se à análise de sentimentos como um tipo específico de *text mining*, que tem como finalidade identificar afirmações subjetivas, opiniões e sentimentos, especialmente no conteúdo gerado pelos utilizadores, na Internet.

A quantidade de dados gerada diariamente ao redor da hotelaria é de elevada dimensão pelo que, o *text mining* se trata de uma ferramenta bastante útil (Godbole et al., 2010). Aplicada à indústria hoteleira, uma análise de *text mining a online reviews* pode auxiliar os gestores hoteleiros a compreenderem as características relacionadas com os serviços oferecidos nos seus hotéis que influenciam a satisfação dos clientes, facilitando a definição de estratégias de gestão de relacionamento com os clientes (Calheiros et al., 2017).

Em suma, *text mining* diz respeito ao processo de extrair informação útil e relevante de conteúdo textual não estruturado para lidar com a elevada dimensão de informação (Xu & Li, 2016). Como tal, existem várias metodologias para a aplicação do processo extração, das quais é possível destacar a metodologia *CRISP-DM*.

3.3. Metodologia *CRISP-DM*

A *CRISP-DM*, denominada de “*Cross-Industry Standard Process for Data Mining*”, trata-se de um procedimento que tem como objetivo providenciar as orientações para a execução de *data mining* e que se caracteriza por ser independente do setor da indústria e da tecnologia utilizada (Schröer et al., 2021). Huang e Hsieh (2020) referem-se à metodologia *CRISP-DM* como o processo standard mais utilizado por especialistas em *data mining*, sendo que este se baseia sobretudo na sua experiência prática, segundo Chapman et al. (2020).

Na presente investigação será adotada uma metodologia *CRISP-DM* visto que é a metodologia que se foca nas necessidades dos gestores e na resolução dos seus problemas de gestão (Laureano et al., 2014), mas também pelo seu teor flexível e facilmente adaptável (IBM, 2020).

Em conformidade com esta metodologia, segundo Chapman et al. (2000), o processo de *data mining* abrange seis fases distintas representadas na *Figura 1*, em que não existe a necessidade de serem sequenciais, ou seja, trata-se de um processo em que é possível avançar ou recuar nas fases, consoante o resultado de cada fase. As fases consistem então em: (1) Compreensão do Problema (“*Business Understanding*”); (2) Compreensão dos Dados (“*Data Understanding*”); (3) Preparação dos Dados (“*Data Preparation*”); (4) Modelação (“*Modeling*”); (5) Avaliação (“*Evaluation*”); (6) Implementação (“*Implementation*”).

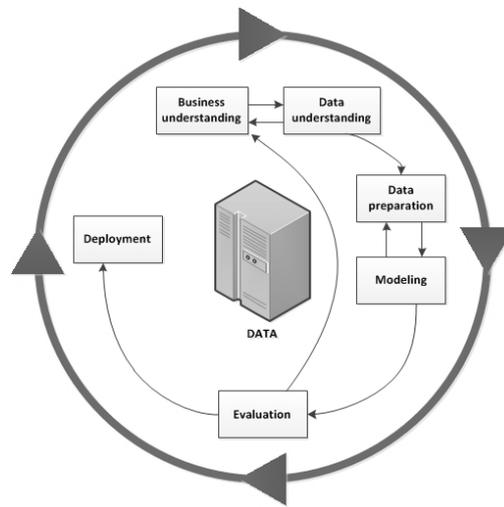


Figura 1: Metodologia CRISP-DM

Fonte: IBM

De acordo com os autores Schröer et al. (2021), a primeira fase consiste na obtenção de uma visão global dos recursos disponíveis e necessários, para o desenvolvimento do projeto. É neste sentido que Plotnikova et al. (2020) referem que esta primeira fase se foca na compreensão do contexto do problema, objetivos e necessidades do processo de *data mining*, tratando-se, portanto, do ponto de partida de um determinado projeto.

A segunda fase diz respeito à recolha de dados inicial, assim como a uma análise descritiva e exploratória dos mesmos, com o objetivo de verificar a qualidade dos dados e obter os primeiros insights (Chapman et al., 2020).

A preparação dos dados, que se refere à terceira fase consiste em selecionar os dados a serem utilizados na análise através da definição de critérios de inclusão e exclusão, com o objetivo de constituir o conjunto de dados final a ser utilizado nas fases subsequentes. Esta etapa, de preparação dos dados, engloba tarefas que podem ser executadas repetidamente (Plotnikova et al., 2020).

Segundo Chapman et al. (2000), na etapa seguinte (modelação), são selecionadas e aplicadas várias técnicas de modelação para testar e analisar os dados, com o intuito de obter um modelo e respetiva avaliação. Uma das técnicas de modelação mais notáveis diz respeito às árvores de decisão, que permitem prever os valores de uma variável dependente (*target*), em função dos valores das variáveis independentes, com base num conjunto de regras (IBM, 2020).

A quinta fase consiste em verificar se os modelos alcançam os objetivos inicialmente definidos. Nesta etapa (avaliação), os resultados devem ser alvos de

interpretação e deve-se proceder a uma revisão de todo o processo, no geral (Schröer et al., 2021).

E, por fim, a última fase traduz-se em planejar a implementação, monitorização e manutenção dos dados obtidos, de modo a suportarem decisões. De acordo com Chapman et al. (2000), a fase de implementação pode-se traduzir num mero relatório com as análises obtidas ou por uma repetição de todo o processo de *data mining*. Assim, na presente investigação a fase de implementação corresponde à elaboração do presente relatório e respetiva análise desenvolvida.

CAPÍTULO 4

APLICAÇÃO E RESULTADOS

O presente capítulo tem o intuito de explicitar componente empírica do estudo, que consiste numa análise de *text mining* através da aplicação da metodologia *CRISP-DM*. Através deste procedimento será possível executar uma análise de dados de forma estruturada. Neste sentido, serão apresentadas as cinco fases subjacentes: (1) Compreensão do Problema; (2) Compreensão dos Dados; (3) Preparação dos Dados; (4) Modelação; (5) Avaliação.

4.1. Compreensão do Problema

A presente investigação pretende, através da análise de *online reviews* de diversos hotéis localizados em vários destinos turísticos identificar o panorama das *online reviews* antes do surto do *COVID-19* e durante, de modo a compreender se a satisfação dos clientes mudou com a transformação do serviço hoteleiro provocada pela pandemia. Deste modo, surge a necessidade de obter informação acerca das diferenças nos principais indicadores das *online reviews* (volume e valência), do conteúdo textual das *reviews* nomeadamente ao nível dos atributos mais abordados pelos clientes e respetivos pesos, da carga sentimental das *reviews* e dos fatores relevantes na previsão de ratings mais elevados. Com esta informação ambiciona-se averiguar se os hotéis estão então a adotar uma atitude resiliente.

Neste sentido, é pretendido responder à questão de investigação: *Estará a pandemia COVID-19 a ter impacto na satisfação dos clientes nos hotéis?*

4.2. Compreensão dos Dados

Os dados utilizados foram obtidos através de extração manual de conteúdos presentes no site *Booking.com*. A escolha da plataforma prende-se ao facto de se tratar da agência de viagens online com maior volume de vendas a nível mundial (Xu & Li, 2016) e, também, pela sua transparência e fiabilidade, dado que apenas os indivíduos que efetuam

efetivamente uma compra num determinado estabelecimento têm a possibilidade de escrever uma *review*. Os dados recolhidos referem-se a *online reviews* de 64 hotéis (Tabela 2), datadas entre o dia 1 de Julho e o dia 31 de Agosto, em ambos os anos 2019 e 2020, nos três idiomas mais relevantes em termos quantitativos (inglês, espanhol e português). O intervalo de datas para a extração dos dados tem em consideração o facto de os meses de verão terem representado um “balão de oxigénio” para a indústria hoteleira (Magalhães, 2020). Importa, ainda, referir que para o processo de extração foram criados dois ficheiros *Excel* relativos aos dois períodos temporais em estudo.

De acordo com o Barómetro do Turismo Mundial o Sul da Europa e Mediterrâneo continua a ser o destino europeu com maior peso, representando cerca de 40,8% do total das chegadas de turistas internacionais e 40,7% das receitas de turismo da região europeia (UNWTO, 2020). Deste modo, no que concerne à localização dos hotéis a estudar, foram seleccionados quatro destinos turísticos, o Algarve, Costa del Sol, Lisboa e Barcelona. A região algarvia foi escolhida pelo facto de ser o principal destino turístico de Portugal com 33,1% das dormidas totais, ter a maior oferta hoteleira (número de hotéis) e, também, por ter o segundo valor de *REVPAR* mais elevado do país (INE, 2020). A Costa del Sol é o principal destino peninsular espanhol, no âmbito de “sol e praia”, tendo por base os principais indicadores: *REVPAR* e taxa de ocupação (Christie & Co, 2019). Lisboa é o segundo destino português mais expressivo, com cerca de 25,9% das dormidas totais e o *REVPAR* mais elevado do país (INE, 2020). Relativamente a Barcelona a sua escolha prende-se ao facto de ser o destino espanhol com maior oferta hoteleira (número de hotéis) e, também, a cidade com o *REVPAR* mais elevado e maior taxa de ocupação (Christie & Co, 2020). É, portanto, nesta ótica que, no presente estudo os dados referentes ao Algarve e Costa del Sol surgem como representantes dos hotéis de destinos de “sol e praia”, enquanto Barcelona e Lisboa de destinos citadinos, atendendo às características dos destinos.

Classificação	Destino							
	Algarve		Costa del Sol		Lisboa		Barcelona	
	NºHotéis	Dimensão média	NºHotéis	Dimensão média	NºHotéis	Dimensão média	NºHotéis	Dimensão média
2 Estrelas	4	57	4	84	4	74	4	101
3 Estrelas	4	131	4	178	4	149	4	84
4 Estrelas	4	139	4	211	4	167	4	115
5 Estrelas	4	260	4	188	4	198	4	256
Total	16	147	16	165	16	147	16	139

Tabela 2: Características dos Hotéis Utilizados na Extração de *Reviews*

Fonte: Elaboração Própria

Anteriormente, no Capítulo 2 do estudo, é mencionado que o volume e valência são os principais indicadores das *online reviews* (Babic Rosario et al., 2016), que existem diferenças nos níveis de satisfação expressos nas *reviews* consoante se tratar de um hotel independente ou pertencente a uma cadeia hoteleira (Gao et al., 2018 e Lee et al., 2020) e que as características demográficas, como por exemplo o país de origem dos clientes, têm impacto na sua satisfação (Tse e Ho, 2009 e Li et al., 2020). Além disso, a qualidade das *reviews* poder ser medida através da sua utilidade, nomeadamente através dos votos de utilidade (Cheung et al., 2009) e acresce-se ainda o facto de a resposta da gestão ser considerada como uma estratégia importante na gestão da reputação dos hotéis (Xie et al., 2017). Como tal, além das *reviews* propriamente ditas foram extraídas outras informações relativas ao hotel nomeadamente a classificação (estrelas), o tipo (independente ou cadeia hoteleira), e a dimensão (número de quartos). Para cada *review* foi também extraído o rating correspondente, atribuído pelos clientes, os votos de utilidade e a informação se existiu ou não uma resposta por parte do hotel. Por último, acrescentam-se ainda, informações como o número de noites pernoitadas, o tipo de viagem (casal, família, grupo e viajante individual) e o país de origem do hóspede.

No total foram analisadas 9387 *reviews*, das quais 61% representam o período antecedente à pandemia (entre 1 de Julho e 31 de Agosto de 2019) e 39% o seu decorrer (entre 1 de Julho e 31 de Agosto de 2020), tal como é possível constatar na *Tabela 3*. Deste modo é possível afirmar que o volume de *reviews* durante a pandemia diminuiu, comparativamente ao período homólogo.

2019		2020		Total	
N	%	N	%	N	%
5726	61%	3661	39%	9387	100%

Tabela 3: Volume de *Reviews* Analisadas, por Ano

Fonte: Elaboração Própria

Para uma melhor compreensão dos dados, recorreu-se ao software *IBM SPSS Modeler*, através da função “*Data Audit*”, que providencia uma primeira abordagem aos dados presentes nos ficheiros com dados estatísticos, que serão apresentados seguidamente.

Através da *Tabela 4* é possível constatar que existiu uma redução do volume de *reviews*, em todos os destinos. Contudo, em termos percentuais, importa referir que a representatividade dos hotéis de destinos urbanos (8% em Barcelona e 13% em Lisboa)

apresenta uma quebra face ao período pré-*COVID-19*, enquanto que, existe um aumento nos destinos de “sol e praia” (5% no Algarve e 15% na Costa del Sol).

2019			2020			Variação	
Value	%	N	Value	%	N	%	
Algarve	25.74	1474	Algarve	31.14	1140	5.4	
Barcelona	20.43	1170	Barcelona	12.76	467	(7.67)	
Costa del Sol	24.05	1377	Costa del Sol	39.09	1431	15.04	
Lisbon	29.78	1705	Lisbon	17.02	623	(12.76)	

Tabela 4: Distribuição das *Reviews*, por Destino

Fonte: Elaboração Própria

Ainda relativamente à localização dos hotéis, através da *Figura 2*, é possível verificar a média da valência das *reviews* por destino, sendo que esta é representada pelo *rating* associado às *reviews*, que, no ano de 2019 assume valores de 2,5 a 10 e em 2020 de 1 a 10. Assim, é visível que Lisboa se trata do destino com o *rating* médio mais elevado, em ambos os períodos temporais em estudo. É ainda possível constatar que existe uma diminuição dos *ratings* médios em todos os destinos, visto que em 2019 todos assumem valores superiores a 8 enquanto que, em 2020 apenas Lisboa possui um valor médio superior a 8.

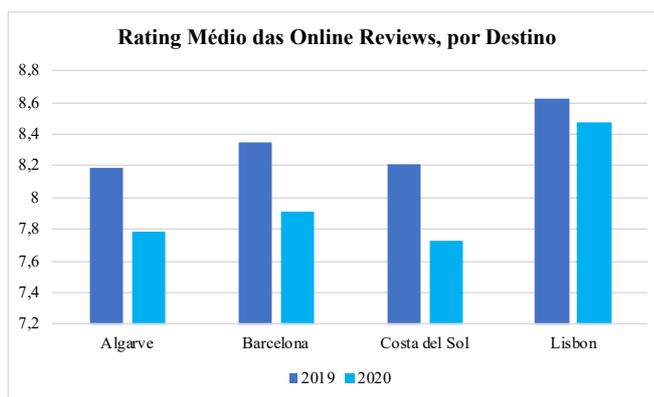


Figura 2: Rating Médio das *Reviews*, por Destino

Fonte: Elaboração Própria

Neste seguimento, através do *SPSS Statistics*, procedeu-se à realização de um *T-Test* para verificar se a diferença do *rating* médio representada anteriormente, em cada destino é significativa. Contudo, importa realçar que primeiramente foi assegurado o cumprimento dos pressupostos subjacentes a esta tipologia de teste estatístico. O *rating* médio das *online reviews* é significativamente menor durante o período pandémico do que antes da pandemia nas regiões do Algarve (média antes da pandemia= $8,19 \pm 1,67$; média durante a pandemia= $7,78 \pm 1,91$), $t(2270) = 5,69$, $p= 0,001$, em Barcelona (média antes da pandemia= $8,35 \pm 1,73$; média durante a pandemia= $7,91 \pm 2,12$), $t(726)=3,96$,

$p=0,001$, e na Costa del Sol(média antes da pandemia= $8,21 \pm 1,62$; média durante a pandemia= $7,73 \pm 2,09$), $t(2686)=6,78$, $p=0,001$. Em oposição, relativamente aos hotéis localizados em Lisboa, não existe uma diferença significativa no rating médio(média antes da pandemia= $8,63 \pm 1,5$; média durante a pandemia= $8,48 \pm 1,75$), $t(974)=1,89$, $p=0,059$. É assim possível afirmar que perante a pandemia, existe uma redução significativa no rating médio atribuído aos hotéis de todos os destinos em estudo, à exceção dos que se localizam em Lisboa (*Anexo A*).

Os hotéis pertencentes a uma cadeia hoteleira são predominantes, representando cerca de 70% e 73 % das *reviews* relativas a 2019 e 2020, respetivamente (*Tabela 5*).

2019			2020		
Value	%	N	Value	%	N
Hotel Chain	69.59	3985	Hotel Chain	73.37	2686
Independent	30.41	1741	Independent	26.63	975

Tabela 5: Distribuição das *Reviews*, por Tipo de Hotel

Fonte: Elaboração Própria

No que concerne à distinção do rating atribuído em função do tipo de hotel, em ambos os períodos temporais, são os hotéis independentes que possuem um rating médio mais elevado (*Figura 3*). Em 2020, contudo, existe uma intensificação na diferença da valência média entre os hotéis que pertencem a uma cadeia hoteleira e os que não pertencem.

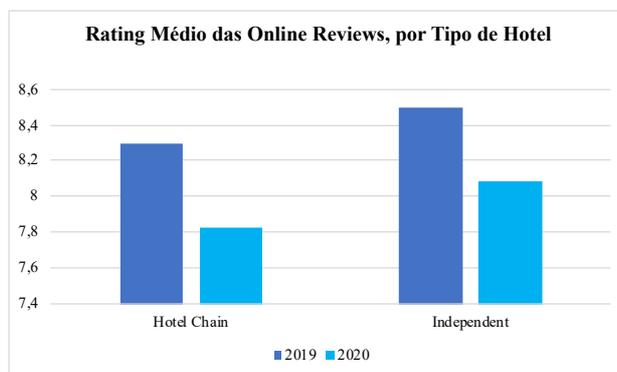


Figura 3: Rating Médio das *Reviews*, por Tipo de Hotel

Fonte: Elaboração Própria

Neste seguimento, através da realização de um *T-Test*, verificou-se que o rating médio das *online reviews* é significativamente menor durante o período pandémico do que antes da pandemia quer no que diz respeito aos hotéis que pertencem a uma cadeia hoteleira (média antes da pandemia= $8,3 \pm 1,69$; média durante a pandemia= $7,83 \pm 2$),

$t(5098,5)=9,98$, $p=0,001$, quer nos hotéis de gestão independente (média antes da pandemia= $8,49 \pm 1,48$; média durante a pandemia= $8,1 \pm 1,99$), $t(1582)=5,58$, $p=0,001$, (*Anexo B*).

No que diz respeito à classificação dos hotéis, tal como é visível na *Tabela 6*, os hotéis de três estrelas em ambos os intervalos de tempo analisados, são os que mais representam as *reviews* em estudo.

2019			2020		
Value	%	N	Value	%	N
2	23.54	1348	2	19.01	696
3	30.11	1724	3	34.31	1256
4	21.62	1238	4	27.94	1023
5	24.73	1416	5	18.74	686

Tabela 6: Distribuição das *Reviews*, por Categoria do Hotel

Fonte: Elaboração Própria

Através da *Figura 4* é possível averiguar que em 2019, o rating médio mais elevado está associado aos hotéis de quatro estrelas enquanto que, os hotéis de três estrelas são os que geram os ratings mais baixos, em termos médios. Relativamente ao ano de 2020, são os hotéis de cinco estrelas que possuem o rating médio mais elevado (cerca de 8,4 numa escala de 1 a 10). É também possível concluir que são os hotéis de quatro estrelas que possuem a quebra mais expressiva (0,6) no seu rating médio.

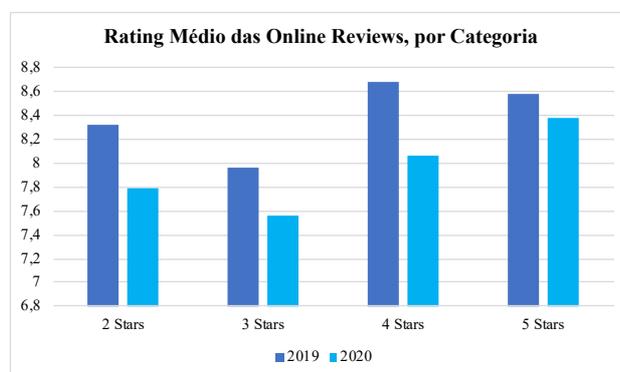


Figura 4: Rating Médio das *Reviews*, por Categoria de Hotel

Fonte: Elaboração Própria

Mediante a realização de um *T-Test* concluiu-se que a pandemia despoletou a existência de uma diminuição significativa no rating médio em todas as categorias de hotel nomeadamente os de 2 estrelas (média antes da pandemia= $8,32 \pm 1,51$; média durante a pandemia= $7,79 \pm 1,99$), $t(1117)=6,15$, $p=0,001$, os de 3 estrelas (média antes da pandemia= $7,97 \pm 1,75$; média durante a pandemia= $7,56 \pm 2,05$), $t(2441)=5,7$, $p=0,001$, os de 4 estrelas (média antes da pandemia= $8,69 \pm 1,46$; média durante a pandemia= $8,06 \pm 2$), $t(1824)=8,31$, $p=0,001$, e os de 5 estrelas (média antes da

pandemia= $8,58 \pm 1,64$; média durante a pandemia= $8,38 \pm 1,78$), $t(1256)=2,49$, $p=0,013$, tal como se constata no *Anexo C*.

Como complemento, de forma a visualizar os dados de uma outra perspectiva, foram ainda desenvolvidos *heatmaps* (recorrendo ao software *SPSS Modeler*), que permitem uma visualização do nível de rating médio atribuído através da intensidade da cor, com o intuito de traçar o perfil dos hotéis que possuem ratings mais elevados e ou inferiores em ambos os períodos temporais. Assim, a *Figura 5* permite-nos afirmar que no período antecedente à pandemia os hotéis de três estrelas localizados no Algarve são os que refletem um menor nível de satisfação, com base nas *online reviews* analisadas. Já no contexto pandémico o mesmo aplica-se aos hotéis de três estrelas localizados na Costa del Sol. Importa ainda realçar o facto de que os hotéis de cinco estrelas situados em Lisboa, em ambos os períodos temporais são os que apresentam os ratings mais elevados.



Figura 5: Heatmap, por Localização e Categoria do Hotel

Fonte: Elaboração Própria, através do SPSS Modeler

As pessoas que viajam em casal são as que mais explicam as *reviews* em estudo, com uma representatividade de 46% no ano de 2019 e cerca de 51% em 2020 (*Tabela 7*). A maior diferença refere-se às pessoas que viajam em família, que registou uma descida de cerca de 6%, relativamente ao período anterior à pandemia.

2019			2020		
Value	%	N	Value	%	N
Couple	46.26	2649	Couple	51.46	1884
Family	35.33	2023	Family	29.66	1086
Group	9.87	565	Group	8.14	298
Solo traveller	8.54	489	Solo traveller	10.73	393

Tabela 7: Distribuição das Reviews, por Tipo de Reviewer

Fonte: Elaboração Própria

Em ambos os períodos temporais, são as pessoas que viajam em grupo (“*Group*”) que, em média, atribuem ratings mais elevados (*Figura 6*). Os indivíduos que viajam em família (“*Family*”) e individualmente (“*Solo traveller*”) são os que apresentam uma

redução maior no que respeita aos ratings atribuídos, comparativamente ao período homólogo antecedente à pandemia.

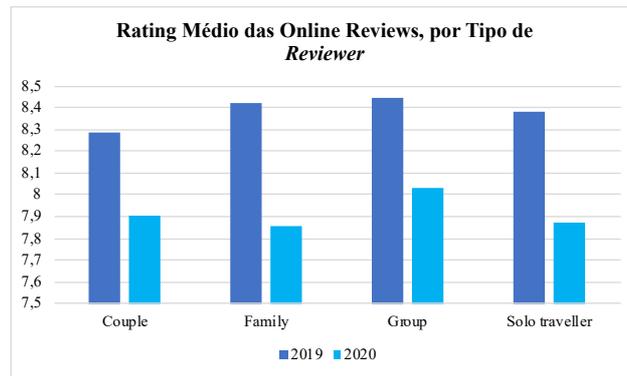


Figura 6: Rating Médio das Reviews, por Tipo de Reviewer

Fonte: Elaboração Própria

Através da realização de *T-Test*, verificou-se que o rating médio atribuído pelos clientes apresenta uma diminuição estatisticamente significativa nos clientes que viajam em casal (média antes da pandemia= $8,28 \pm 1,66$; média durante a pandemia= $7,91 \pm 2,04$), $t(3511)=6,618$, $p=0,001$), em família (média antes da pandemia= $8,42 \pm 1,61$; média durante a pandemia= $7,86 \pm 1,93$), $t(1904)=8,26$, $p=0,001$), em grupo (média antes da pandemia= $8,44 \pm 1,57$; média durante a pandemia= $8,03 \pm 2,05$), $t(484)=3$, $p=0,003$) e sozinhos (média antes da pandemia= $8,38 \pm 1,62$; média durante a pandemia= $7,87 \pm 1,93$), $t(765)=4,2$, $p=0,001$) tal como se pode verificar através do *Anexo D*.

Atendendo à localização dos hotéis e ao tipo de *reviewer* constata-se que os indivíduos que usufruíram de uma estadia no Algarve e Costa del Sol, em casal, são os mais insatisfeitos atendendo aos dados de 2019 (*Figura 7*). No que diz respeito ao período pandémico, o maior nível de satisfação provém dos clientes que pernотaram em casal, nos hotéis de Lisboa, enquanto que os clientes que viajaram sozinhos para um hotel da Costa del Sol são os menos satisfeitos.



Figura 7: Heatmap, por Localização e Tipo de Reviewer

Fonte: Elaboração Própria, através do SPSS Modeler

A *Tabela 8* demonstra que em 2019 apenas 42,8% das *reviews* obtiveram uma resposta por parte dos hotéis e em 2020 a percentagem aumentou para 52,7%, evidenciando assim um aumento de cerca de 10% nas respostas.

2019			2020		
Value	%	N	Value	%	N
No	57.2	3275	No	47.23	1729
Yes	42.8	2451	Yes	52.77	1932

Tabela 8: Volume de Respostas às *Reviews*

Fonte: Elaboração Própria

A *Tabela 9* demonstra a quase inutilização, por parte dos utilizadores da plataforma *Booking.com*, da ferramenta de voto que permite votar numa determinada *review* como sendo “útil” e ainda um retrocesso de 4% em 2020.

2019			2020		
Value	%	N	Value	%	N
0	86.06	4928	0	89.76	3286
1	9.43	540	1	7.76	284
>=2	4.51	258	2	1.88	284

Tabela 9: Volume de Votos de Utilidade

Fonte: Elaboração Própria

Ainda no que diz respeito aos votos de utilidade, existe uma tendência para a existência de votos em *reviews* que contêm entre 50 e 80 palavras, em ambos os períodos temporais em estudo (*Figura 8*).

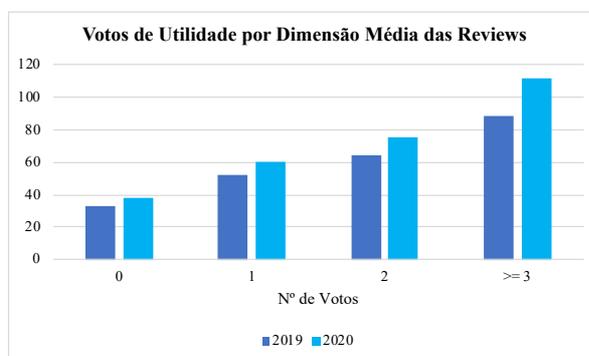


Figura 8: Votos de Utilidade por Tamanho Médio das *Reviews*

Fonte: Elaboração Própria

Inicialmente, em 2019 eram 123 os países de origem dos hóspedes e 84 em 2020, pelo que se decidiu agrupar os países em regiões, à exceção de Portugal e Espanha pela sua elevada representatividade. Através da *Tabela 10*, é visível a diminuição da expressão de todas as regiões em 2020, à exceção de Portugal, Espanha e Europa Ocidental (*Western*

Europe), o que evidencia claramente o efeito das restrições de circulação impostas, a nível internacional.

2019			2020		
Value	%	N	Value	%	N
Africa	1.9	109	Africa	0.27	10
Asia and Oceania	3.32	190	Asia and Oceania	0.27	10
Eastern Europe	2.1	120	Eastern Europe	1.42	52
Middle East	3.49	200	Middle East	0.38	14
North America	6.18	354	North America	0.68	25
Northern Europe	20.49	1173	Northern Europe	8.96	328
Not specified	0.12	7	Not specified	0.11	4
Portugal	12.73	729	Portugal	24.47	896
South and Central America	12.8	733	South and Central America	1.78	65
Southern Europe	1.17	67	Southern Europe	1.86	68
Spain	30.42	1742	Spain	53.87	1972
Western Europe	5.27	302	Western Europe	5.93	217

Tabela 10: Distribuição das *Reviews*, por Origem do *Reviewer*

Fonte: Elaboração Própria

Através da observação da *Figura 9* é possível averiguar que em 2020 existe um aumento da satisfação refletido através do rating, dos clientes provenientes de África, enquanto que houve uma redução em todos os outros à exceção de “*Not specified*”, que diz respeito aos *reviewers* que não revelam o seu país de origem. Num período pandémico, são os indivíduos oriundos de países da América do Norte, África, Europa Ocidental e Sul da Europa os que, em média, atribuem os ratings mais elevados.

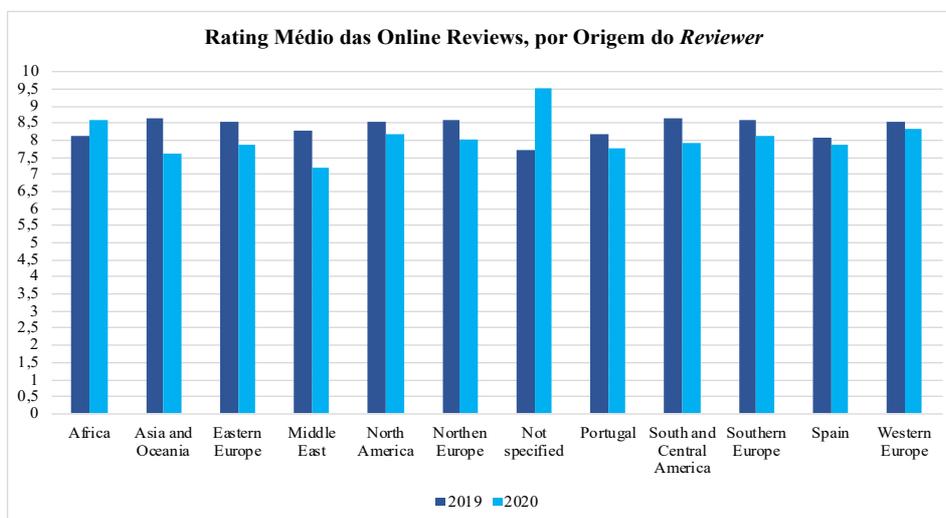


Figura 9: Rating Médio das *Reviews*, por Origem do *Reviewer*

Fonte: Elaboração Própria

Através da realização de um *T-Test* foi possível averiguar que o rating médio apresenta uma quebra significativa quando atribuído por clientes oriundos de Espanha (média antes da pandemia= $8,07 \pm 1,69$; média durante a pandemia= $7,89 \pm 2,01$), $t(3702)=3$, $p=0,002$), Portugal(média antes da pandemia= $8,19 \pm 1,56$; média durante a

pandemia= $7,75 \pm 1,93$), $t(1623)=5,09$, $p=0,001$, América do Sul e Central(média antes da pandemia= $8,62 \pm 1,51$; média durante a pandemia= $7,94 \pm 2,25$), $t(69)=2,38$, $p=0,020$ e Norte da Europa (média antes da pandemia= $8,59 \pm 1,55$; média durante a pandemia= $7,99 \pm 2,05$), $t(438)=4,92$, $p=0,001$ (*Anexo E*). No que concerne à região norte americana, por não se verificarem os pressupostos inerentes à realização de um *T-Test*, procedeu-se à realização de um teste não paramétrico *Wilcoxon*, que nos permite concluir que a pandemia não despoletou uma mudança estatisticamente significativa no rating atribuído pelos clientes oriundos desta região ($Z= -0,59$, $p=0,56$, *Anexo F*). Para as restantes regiões não foi possível realizar testes estatísticos dado o escasso número de observações existentes perante o contexto pandémico, relacionado ao não cumprimento dos pressupostos que a estes estão associados.

Relativamente à dimensão dos hotéis em estudo, por praticamente todos assumirem valores distintos, estabeleceram-se vários intervalos de dimensão, como é perceptível através da *Tabela 11*. Constatase então, que a maioria das *reviews* utilizadas no estudo, refere-se à avaliação de hotéis com uma dimensão compreendida entre os 201 e 300 quartos.

2019			2020		
Value	%	N	Value	%	N
<=50	14.15	810	<=50	11.94	437
>50 and <=100	27.79	1591	>50 and <=100	22.1	809
>100 and <=200	19.12	1095	>100 and <=200	22.64	829
>200 and <=300	31.96	1830	>200 and <=300	36.9	1351
>300	6.99	400	>300	6.42	235

Tabela 11: Distribuição das *Reviews*, por Número de Quartos dos Hotéis

Fonte: Elaboração Própria

No que concerne à relação entre os ratings atribuídos e a dimensão dos estabelecimentos hoteleiros, representada na *Figura 10*, é nos hotéis de dimensão inferior (com dimensão igual ou inferior a 50 quartos) que incidem os ratings médios mais elevados. Importa também realçar que enquanto que num período antecedente à pandemia eram os hotéis com dimensão compreendida entre os 101 e 200 quartos que representava o rating médio mais baixo, no contexto pandémico são os que detêm de 201 a 300 quartos.

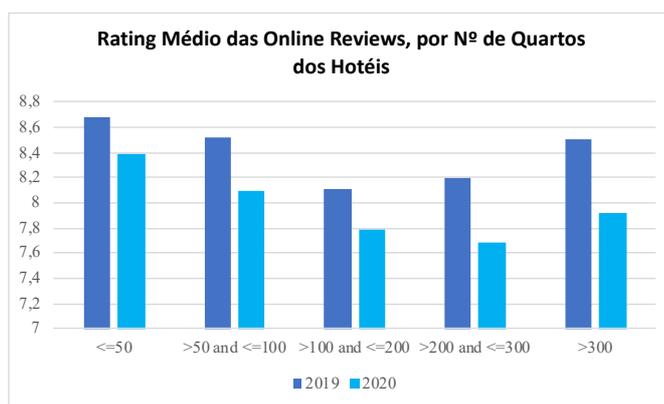


Figura 10: Rating Médio das *Reviews*, por Número de Quartos dos Hotéis

Fonte: Elaboração Própria

Complementarmente aos dados evidenciados na *Figura 10*, é possível afirmar que o rating médio é significativamente menor durante o período pandémico do que antes da pandemia nos hotéis de todas as dimensões (*Anexo G*).

O mesmo processo foi efetuado para o número de noites pernoitadas associadas à estadia dos *reviewers*. Deste modo, atendendo à *Tabela 12*, percebe-se que a maioria das *reviews* diz respeito a uma estadia curta, de 1 a 3 noites. Importa realçar o facto de que em 2020 o número de *reviews* associadas a uma estadia de 1 noite aumentou cerca de 9% e 4% a estadias de 2 noites, enquanto que todas as durações de estadia mais compridas diminuíram.

2019			2020		
Value	%	N	Value	%	N
1	23.98	1373	1	32.81	1201
2	20.92	1198	2	25.1	919
3	19.94	1142	3	14.91	546
4	13.64	781	4	9.4	344
5	7.95	455	5	5.95	218
>6 and <=12	13.01	745	>6 and <=12	11.47	420
>=12	0.56	32	>=12	0.36	13

Tabela 12: Distribuição das *Reviews*, por Duração da Estadia (em Noites)

Fonte: Elaboração Própria

A *Figura 11* evidencia que em 2019 existe uma discrepância entre os ratings atribuídos, atendendo ao número de noites associadas à estadia dos clientes que efetuaram as *reviews* em análise, destacando-se um rating médio superior proveniente de indivíduos que pernoitam apenas 1 noite. Já no que diz respeito aos dados relativos à pandemia, a única diferença mais expressiva nos ratings atribuídos consoante a duração da estadia remete-se às estadias prolongadas (iguais ou superiores a 12 noites).

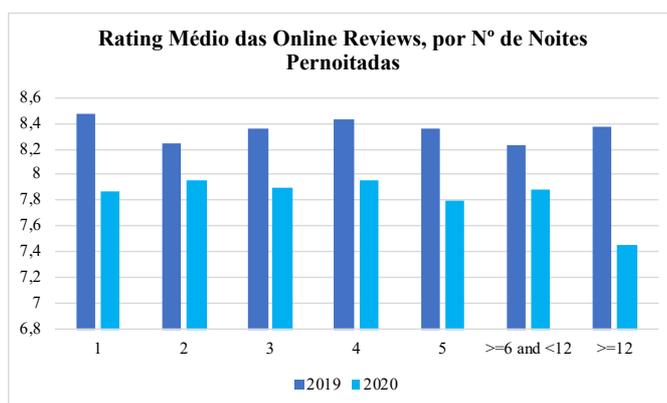


Figura 11: Rating Médio das *Reviews*, por Duração da Estadia (em Noites)

Fonte: Elaboração Própria

Através da realização de um *T-Test* verificou-se que o rating médio, perante o contexto pandémico apresenta uma quebra significativa quando associado a estadias inferiores a 12 noites (*Anexo H*). Já no que diz respeito a estadias iguais ou superiores a 12 noites não é possível aferir se a diminuição do rating é estatisticamente significativa, pois não são cumpridos os pressupostos inerentes à realização do teste estatístico.

4.3. Preparação dos Dados

Esta etapa de preparação dos dados visa a obtenção de informação relevante e alterações que considerem necessárias para a obtenção de um novo *dataset* a utilizar nas fases posteriores (modelação e avaliação).

Como tal, na preparação dos dados foram efetuadas algumas adaptações aos dados extraídos. Em primeiro lugar, o texto das *online reviews* foi traduzido na íntegra para inglês através da função de tradução do *Excel*. No decorrer da presente investigação, foram extraídas, no total, 16242 *online reviews*, contudo, nesta fase, foi tomada a decisão de filtrar as *reviews* de forma a manter apenas aquelas que contêm conteúdo informativo relevante. Assim, foram excluídas todas as *reviews* com menos de cinco palavras dado o seu carácter repetitivo e pouco informativo, o que resultou num total de 9387 *reviews*, originando assim uma redução de 6855 *reviews*, comparativamente às inicialmente extraídas. Importa também referir que, devido à existência de um elevado número de possibilidades de valores em algumas variáveis em estudo, tal como se verifica na secção anterior, nomeadamente o país de origem dos hóspedes, a dimensão dos hotéis os votos de utilidade, a duração da estadia e a dimensão média das *reviews*, surgiu ainda a

necessidade de os agrupar em categorias de modo a facilitar todo o processo. Esta fase do estudo careceu também de uma preparação ao nível da variável “Rating”, através da criação de três categorias (“*High*”, “*Medium*” e “*Low*”), visto tratar-se de uma variável contínua e, portanto, a possibilidade de assumir uma grande diversidade de valores (compreendidos entre 1 e 10) dificultar a análise pretendida nas etapas seguintes. Assim, a categoria “*High*” inclui todos os valores acima de 7 (inclusive), a categoria “*Medium*” engloba todos os valores superiores a 4 e inferiores a 7 e, por último, a categoria “*Low*” todos os valores inferiores a 4 (inclusive). Tal como é possível comprovar através da *Tabela 13*, de 2019 para 2020 existe um aumento 5% da percentagem de *reviews* com uma classificação baixa (“*Low*”) e, em contrapartida, uma diminuição das *reviews* com classificação média (“*Medium*”) e alta (“*High*”).

2019			2020		
Value	%	N	Value	%	N
High	83.16	4762	High	81.54	2985
Medium	14.81	848	Medium	11.28	413
Low	2.03	116	Low	7.18	263

Tabela 13: Distribuição das *Reviews*, por Rating

Fonte: Elaboração Própria

Nesta etapa do processo *CRISP-DM* recorreu-se ao *software Orange*, que possui uma ferramenta que permite identificar o sentimento contido no texto das *reviews*. Importa assim, referir que o sentimento é extraído pelo *software* através do texto das *reviews*, enquanto que o rating é atribuído pelos clientes, aquando a realização de uma *online review*. A *Figura 12* ilustra o processo desenvolvido no *Orange*, sendo que foi realizado de igual forma para ambos os períodos temporais.

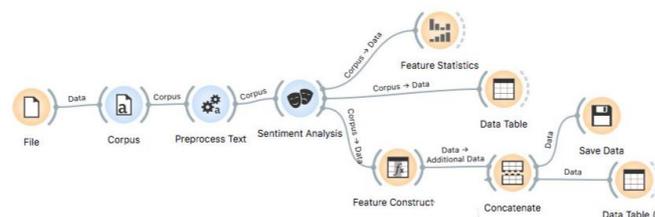


Figura 12: Representação do Processo de Análise de Sentimentos no *Orange*

Fonte: Elaboração Própria

Primeiramente foi então necessário recorrer à função de pré-processamento de texto do *software* que, no fundo, é definido como o processo que “divide o texto em unidades menores(*tokens*), filtra-as, executa a normalização (*stemming*, lematização), cria n-gramas e marca essas unidades com rótulos consoante a classe gramatical” (Orange, n.d.). Este pré-processamento envolveu então tarefas como a remoção de acentos,

tokenização (“*tokenization*”), que reparte o texto das *reviews* em palavras, a remoção de *stopwords* (“*and*”, por exemplo) e ainda *stemming* e lematização, com o intuito de reduzir as palavras derivadas e/ou relacionadas para uma única palavra (“*feelings*” para “*feel*”, por exemplo). De seguida, procedeu-se então à análise de sentimentos, que prevê o sentimento presente no texto das *reviews*, através do algoritmo “*Liu Hu*” disponibilizado pelo *software*, indicado para conteúdo na língua inglesa, que calcula uma pontuação única de sentimento que se trata da “diferença entre a soma de palavras positivas e a soma de palavras negativas, normalizada pelo comprimento do documento e multiplicada por 100” (Hu et al., 2004). A *Figura 13* retrata assim a pontuação mencionada previamente, através de um excerto da análise de sentimentos efetuada pelo *software*, em que quanto mais baixo o valor for, mais negativo é o sentimento e vice-versa.

	Review text	YEAR	sentiment
1	A stay very ...	2019	1.53846
2	Good. The ...	2019	8.69565
3	Experience t...	2019	-0.980392
4	Great value ...	2019	11.4286
5	Less bad. Th...	2019	4
6	I liked the ...	2019	22.2222
7	Good locatio...	2019	-3.33333

Figura 13: Representação da Atribuição de Sentimentos às *Reviews*

Fonte: Elaboração Própria, através do Orange

Assim, através dos outputs obtidos da análise realizada no *Orange* foi possível obter a carga sentimental de cada *review* (*Figura 13*) e, conseqüentemente, a carga sentimental relativa aos dois períodos temporais em estudo (*Tabela 22*). Averigua-se então que a média da carga sentimental das *reviews* diminuiu durante o período pandémico.

2019				2020			
Center	Dispersion	Min.	Max.	Center	Dispersion	Min.	Max.
9.38381	1.04977	-40	57.1429	8.31852	1.13619	-50	55.5556

Tabela 14: Carga Sentimental das *Reviews*

Fonte: Elaboração Própria, através do Orange

De seguida, procedeu-se à realização de um *T-Test* (através do *SPSS Statistics*), para testar a existência de diferenças significativas na carga sentimental média das *reviews* em função do período temporal, a 2019 (pré-*COVID-19*) e 2020 (período pandémico). Como tal, a presente investigação constatou que a carga sentimental média das *online reviews* é, de forma significativa, estatisticamente inferior ($8,32 \pm 9,45$) perante

a pandemia comparativamente com o período pré-*COVID-19* ($9,38 \pm 9,85$), tal como é possível averiguar através da *Tabela 15*.

2019			2020			Teste T para Igualdade de Médias	
N	Média	Desvio Padrão	N	Média	Desvio Padrão	Sig. (2-tailed)	Diferença Média
5726	9.38384	9.85169	3661	8.31934	9.45156	<.001	1.06449355

Tabela 15: Teste para Igualdade de Médias da Carga Sentimental (2019 e 2020)

Fonte: Elaboração Própria, através do SPSS Statistics

Após determinada a carga sentimental das *reviews*, procedeu-se à utilização do *SPSS Modeler Text Analytics*, enquanto ferramenta de análise de sentimentos (IBM, 2020), seguindo a metodologia utilizada por Guerreiro e Rita (2019). O processo desenvolvido através do nó de “*Text Mining*” disponibilizado pelo *SPSS Modeler* está representado na *Figura 14*, sendo que o mesmo se aplicou de igual forma aos dados referentes a 2019 e 2020. O procedimento de *Text Mining* efetuado no *SPSS Modeler* consiste num processo de extração em que todo o conteúdo textual é analisado (IBM, 2020). Aquando a extração, o conteúdo textual é analisado de forma a identificar palavras individuais que constituem “conceitos”. Por sua vez, o *software* procede ao agrupamento dos conceitos semelhantes em apenas um, constituindo os chamados “*types*” (IBM, 2020), ou seja, podem existir vários conceitos subjacentes a um “*type*”. Além disso, o *SPSS Modeler* permite também a criação de categorias que representam a agregação de conceitos e “*types*” de modo a obter uma análise mais resumida dos dados extraídos. Em suma, durante o processo de extração os termos são identificados e agrupados em conceitos, que por sua vez constituem “*types*”, existindo ainda a possibilidade de reunir ambos, através da criação de categorias.

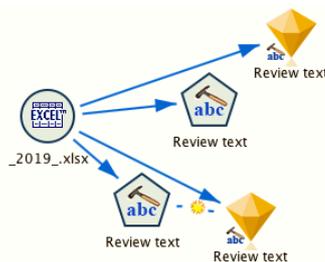


Figura 14: Representação do Processo de Análise de Sentimentos no SPSS Modeler

Fonte: Elaboração Própria

Além disso, importa também contextualizar que este *software* disponibiliza “*libraries*” de análise de texto que facilitam a categorização do texto, dado que contêm os recursos linguísticos necessários para codificar uma elevada quantia de dados de uma

forma rápida e fácil (IBM, 2020). As referidas “libraries” contêm três tipos de dicionários, nomeadamente os dicionários de “types”, de substituição e de exclusão. No que respeita ao dicionário de “types”, este é constituído por um conjunto de conceitos associado a um determinado “type”. Já o dicionário de substituição, tem como função o agrupamento de palavras (conceitos) consideradas como sinónimos. E, por último, o dicionário de exclusão permite a criação de um conjunto de termos e “types”, cuja sua apresentação nos resultados após a extração não seja pretendida.

No decorrer do processo representado na *Figura 14*, foram configurados diversos parâmetros como a seleção da library “Hotel Satisfaction (English)”, extração de unitermos, inclusão de erros de pontuação e extração de entidades não linguísticas. Neste âmbito importa referir que, na presente investigação, foram desenvolvidas algumas intervenções relativas à library pré-definida pelo software “Hotel Satisfaction (English)” desenvolvida especificamente para hotéis, nomeadamente ao nível do conteúdo relacionado com a temática da pandemia. No que concerne ao dicionário de “types” é de realçar a criação de um “type” específico denominado de “Covid-19” em que termos como “mask” e “social distancing” por exemplo foram aqui agrupados, suscitando a necessidade de adaptação do dicionário de substituição, sendo possível exemplificar com a congregação de termos como “coronavirus” e “SARS-CoV-2” a “Covid-19”.

Posto isto, de seguida serão então apresentadas as informações obtidas através do processo representado na *Figura 14*. No que concerne ao número de conceitos extraídos, o presente estudo constata que existe uma redução de 2019 para 2020, de 1095 para 7753 conceitos, respetivamente. Relativamente à sua frequência (*Anexos I e J*), em ambos os períodos, conceitos como “excellent”, “good”, “room”, “location”, “hotel”, “staff”, “breakfast”, “friendly”, “clean”, “pool”, “comfortable”, “bed”, são os que mais se destacam. No ano de 2020 surgem, entre os 100 conceitos mais frequentes, “covid-19” em 10% da totalidade das reviews, “measures” em 4% e “safe” em 2%.

De seguida, os dados foram agrupados em 56 “types” (*Anexo K*), de acordo com a informação semântica, como por exemplo: “Positive”, “Location”, “Hotel General”, “Room”, “Negative”, “Personnel”, “PositiveFeeling”, “RoomAmenities”, “NegativeFeeling”, “NegativeBudget”, “Suggestion”, “NegativeAttitude”, “Internet”, entre outros. Para a obtenção desta análise do conteúdo textual, o software extrai os conceitos, alocando-os a um determinado type, que consiste num conjunto de conceitos semelhantes ou relacionados (IBM, 2020). Assim, é possível constatar que não houve

alterações significativas no tipo de conteúdo mais abordado nas *online reviews*, comparando o período pré-*COVID-19* com o período pandémico. Em ambos os períodos temporais, é destacada a presença de conotação positiva (“*Positive*”) na maioria das *reviews* (93% em 2019 e 92% em 2020). A localização (“*Location*”), o hotel em si (“*Hotel General*”), o quarto (“*Room*”), o staff (“*Personnel*”), os sentimentos positivos (“*PositiveFeeling*”), a comodidades do quarto (“*RoomAmenities*”), o restaurante (“*Restaurant*”) e a presença de uma conotação negativa (“*Negative*”) é o que mais caracteriza o panorama das *reviews* em análise. Importa ainda referir que, em 2020, apesar de o contexto pandémico vigorar, apenas 15% da totalidade das *reviews* está alocada ao *type* “*Covid-19*”.

Neste seguimento importa exemplificar com alguns conceitos que constituem os “*types*” mais relevantes (*Anexo L*). Relativamente ao *type* “*Positive*”, estão associados conceitos como “*good*”, “*better than expected*”, “*excellent*” e, por oposição, ao “*Negative*” é possível destacar conceitos como “*worse*”, “*bad*” e “*ridiculous*”. Outros conceitos como “*love*”, “*beautiful*” e “*clean*” integram o *type* referente aos sentimentos positivos (“*PositiveFeeling*”). E, por último, a título de exemplo destaca-se também o *type* “*Covid-19*” que, contém conceitos como “*mask*”, “*safety distance*” e “*alcohol gel*”.

Após a extração de conceitos e *types*, o *SPSS Modeler* permite analisar ligações entre o texto, denominadas por “*Text Link Analysis*”, em que são definidos padrões de ligação entre os conceitos organizados por *types* (IBM, 2020). Para os dados de 2019 foram reconhecidos 1700 padrões e 1394 para 2020. Em ambos os períodos, a ligação mais relevante diz respeito ao tipo “*Location*” e “*Positive*”, com 2292 ligações em 2019 e 1327 em 2020. A *Figura 15*, exemplificativa do processo desenvolvido nesta fase, ilustra os padrões identificados pelo software no que diz respeito à ligação textual entre o tipo “*Location*” e “*Positive*”, no ano de 2019.

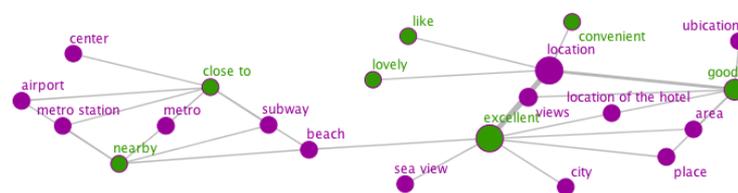


Figura 15: Padrões de Ligação de Texto “*Location*” e “*Positive*” (2019)

Fonte: *Elaboração Própria, através do SPSS Modeler*

Através deste tipo de análise é possível caracterizar os *types* anteriormente definidos consoante a sua ligação textual. No caso particular da *Figura 15* é perceptível

que a proximidade ao aeroporto, a uma estação de metro e à praia e a importância de uma boa vista são aspectos que provocam percepções positivas dos *reviewers* em relação à localização dos hotéis. Neste seguimento, foram obtidos inúmeros outputs semelhantes ao representado na *Figura 15*, para ambos os períodos temporais que se encontram sintetizados no *Anexo M*, onde é possível constatar que existe uma elevada semelhança entre o padrão das ligações textuais de ambos os períodos temporais.

No Capítulo 2 da presente investigação são referidos como atributos essenciais para a satisfação dos clientes o quarto, o staff, o serviço, a localização e as instalações (Xu e Lee, 2020); o preço (Lee et al., 2020); o departamento de *Food & Beverage* (Zhou et al., 2014) e a limpeza (Miles, Miles e Canon, 2012). Assim, através das associações textuais identificadas na “*Text Link Analysis*”, através do *SPSS Modeler* foram criadas as categorias com ambas as polaridades (positiva e negativa) “*Staff*”, “*Room*”, “*Service*”; “*Price*”; “*Location*”; “*Facilities*”, “*F&B*”, “*Cleaning*” e “*Covid-19*” (exclusivamente para os dados de 2020).

Na *Tabela 16*, está representada a proporção de cada categoria criada em relação ao número total de *reviews* analisadas (5726 referentes ao ano de 2019 e 3661 a 2020). Deste modo, é possível concluir que em ambos os períodos temporais a categoria “*Facilities*” é a mais frequente, estando presente em 45% e 49% das *reviews*, de 2019 e 2020 respetivamente. De uma perspetiva geral, o panorama das *reviews* representado através das categorias não difere muito. Contudo, a limpeza (“*Cleaning*”), demonstra uma crescente importância com o advento da pandemia na medida em que a sua presença nas *reviews* aumentou 18%. Importa também referir que a pandemia, de forma direta, representada através da categoria “*Covid-19*” se encontra em apenas 10% das *reviews* tratando-se, por isso, da categoria com menos representatividade no ano de 2020.

2019			2020		
Categoria	Reviews	%	Categoria	Reviews	%
Location	2295	40%	Location	1403	38%
Facilities	2581	45%	Facilities	1804	49%
Staff	1854	32%	Staff	1124	31%
F&B	1640	29%	F&B	954	26%
Room	2522	44%	Room	1539	42%
Cleaning	178	3%	Cleaning	774	21%
Price	929	16%	Price	557	15%
Service	617	11%	Service	487	13%
			Covid-19	382	10%

Tabela 16: Distribuição das *Reviews*, por *Categoria*

Fonte: Elaboração Própria

A *Figura 16* evidencia a polaridade das *reviews* por categoria, com o intuito de compreender a proporção de sentimentos positivos e negativos associada a cada categoria

criada. Assim, podemos verificar que em 2019 conotação positiva é predominante em todas as categorias. A localização do hotel (“*Location*”) é o atributo dos hotéis com maior percentagem de sentimentos positivos (88%) e, conseqüentemente, menor percentagem de sentimentos negativos (12%). Apesar de os sentimentos positivos associados ao preço (“*Price*”) e serviço (“*Service*”) prevalecerem, é nestas duas categorias que é visível uma maior sensibilidade por parte dos *reviewers*, visto que 46% das *reviews* em que estão presentes estão associadas a sentimentos negativos.

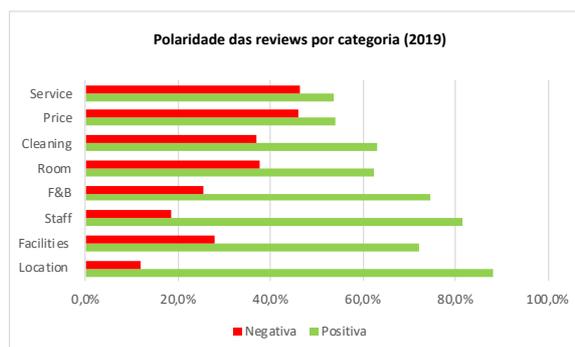


Figura 16: Distribuição da Polaridade das *Reviews* (2019), por *Categoria*

Fonte: Elaboração Própria

No que concerne ao período pandémico, todas as categorias apresentam mais sentimentos positivos associados do que negativos, à exceção do serviço (“*Service*”), tal como se pode verificar a partir da *Figura 17*. Enquanto que, em 2019 (*Figura 16*) existiam mais sentimentos positivos do que negativos associados ao “*Service*”, através da *Figura 17* é notável um impacto da pandemia, visto que a situação se inverteu, existindo, portanto, mais sentimentos negativos do que positivos relacionados com o serviço dos hotéis. Cerca de 86% das *reviews* categorizadas com “*staff*” e “*location*”, têm conotação positiva, o que demonstra que a maioria dos *reviewers* ao mencionar aspetos relativos à localização e recursos humanos dos hotéis, expressa sentimentos positivos. Tendo em conta a temática da presente investigação importa também realçar a categoria “*Covid-19*”, que apesar de pouco expressiva (*Tabela 24*), cerca de 64% das *reviews* associadas apresentam conotação positiva.

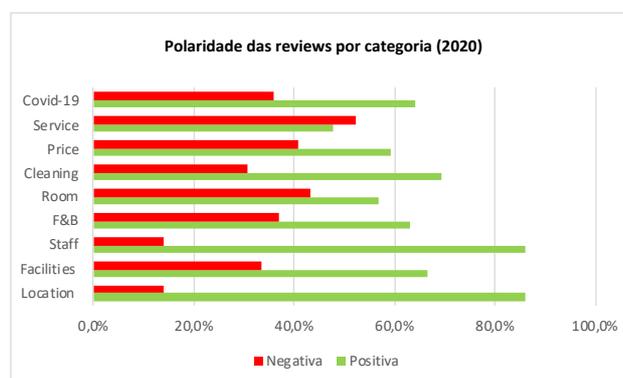


Figura 17: Distribuição da Polaridade das *Reviews* (2020), por *Categoria*

Fonte: Elaboração Própria

Em suma, comparando a polaridade associada a cada categoria, nos dois períodos temporais, é possível constatar que os sentimentos negativos associados às categorias “*Location*”, “*Facilities*”, “*F&B*”, “*Room*” e “*Service*” aumentaram durante a pandemia. Contudo, os *reviewers* no decorrer do período pandêmico, os *reviewers* demonstraram uma satisfação maior superior no que concerne ao preço, staff e à limpeza.

4.4. Modelação

Após a preparação dos dados mencionados acima é, entretanto, possível testar quais os modelos que se adequam melhor para com vista à compreensão dos fatores que originam ratings de satisfação mais elevados, num período pré-*COVID-19* e durante o período marcado pela presença do *COVID-19*. Neste âmbito, o *SPSS Modeler* disponibiliza múltiplos métodos de modelação com base em inteligência artificial, *machine learning* e estatísticas, que permitem obter novas informações a partir dos dados e desenvolver modelos preditivos (IBM, 2020). No referido *software* estes métodos de modelação dividem-se em três categorias distintas, nomeadamente de classificação, associação e segmentação. Dado que, os modelos de classificação têm como principal objetivo a explicação dos dados e a capacidade de previsão da variável dependente, através de uma ou mais variáveis independentes (IBM, 2020), o presente estudo recorre a esta tipologia de modelos. Por conseguinte, os referidos modelos de classificação subdividem-se em várias categorias, das quais é possível destacar as árvores de decisão, as redes neuronais, as listas de decisão e os modelos estatísticos, por exemplo.

De acordo com Berka e Rauch (2010), as árvores de decisão são os modelos mais conhecidos para tratar de tarefas de classificação. Contudo, existem estudos recentes (como por exemplo Xu e Lee, 2020) que se referem aos modelos estatísticos como uma metodologia eficaz no estudo de *online reviews*, nomeadamente o modelo de regressão logística. A essência das várias árvores de decisão disponíveis no *SPSS Modeler* é basicamente a mesma, ou seja, “*they examine all of the fields of your dataset to find the one that gives the best classification or prediction by splitting the data into subgroups*” (IBM, 2020). Já no que diz respeito aos modelos estatísticos fornecidos pelo *software* a dinâmica difere visto que estes se baseiam em equações matemáticas para a codificação da informação extraída dos dados (IBM, 2020).

Nesta etapa, procedeu-se à constituição de um novo conjunto de dados, agregando os dados resultantes da extração inicial a partir da *Booking.com* (rating e votos de utilidade atribuídos pelos clientes e informações relativas à estadia e aos hotéis), apresentados na etapa de “compreensão dos dados”, com os dados obtidos através da análise de sentimentos efetuada no decorrer da etapa de “preparação dos dados”, para cada um dos períodos temporais. Importa realçar que esta agregação foi elaborada através da conciliação dos outputs gerados pelo *SPSS Modeler* anteriormente, juntamente com a função de “*pivot tables*”, no *Microsoft Excel*.

Como tal, na presente investigação, o rating atribuído pelos *reviewers* constitui a variável *target* e as variáveis “*Hotel Type*”; “*Stars*”; “*Location*”, “*Number of rooms*”; “*Number of nights*”; “*Reviewer type*”; “*Guest country*”; “*Hotel response*”; “*Helpful votes*” e os “*types*” associados a cada *review* obtidos na análise de sentimentos (secção anterior), como *inputs*, ou por outras palavras, variáveis independentes. Assim, a variável *target* em estudo trata-se de uma variável categórica, visto que assume três categorias de valores: “*High*”, “*Medium*” e “*Low*”. Neste sentido, foram desenvolvidos três modelos distintos nomeadamente duas árvores de decisão (denominadas “*C&R*” e “*C5.0*”) e uma regressão logística. A árvore de decisão *C&R* “*uses recursive partitioning to split the training records into segments by minimizing the impurity at each step, where a node in the tree is considered “pure” if 100% of cases in the node fall into a specific category of the target field*” (IBM, 2020). Já no que diz respeito ao modelo *C5.0*, este divide os dados com base na variável que fornece o maior ganho de informação em cada nível (IBM, 2020). Importa referir que enquanto que no modelo *C&R* existe a possibilidade de quer as variáveis *target* quer as independentes serem numéricas ou categóricas, no modelo *C5.0* existe o pressuposto da variável *target* ter de ser categórica. Além destes dois

modelos, foi ainda desenvolvido um modelo de regressão logística (denominado “*logistic regression*”), que se trata de uma técnica estatística que classifica a ocorrência de um variável *target* com base nas relações existentes entre esta e os valores assumidos pelas variáveis independentes (inputs). A regressão logística consiste, portanto, num processo equivalente a uma regressão linear, mas com uma variável dependente categórica (como acontece com a variável *target* do presente estudo, ao invés de numérica (IBM, 2020).

Deste modo, dadas as características dos modelos mencionadas e dada a natureza dos dados em estudo, foi tomada a decisão de desenvolver cada um dos três modelos para cada um dos dois conjuntos de dados referentes a 2019 (período pré-*COVID-19*) e 2020 (período durante o *COVID-19*), cujo processo se encontra representado na *Figura 18*.

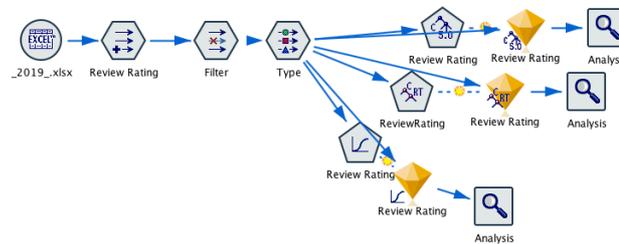


Figura 18: Representação do Processo de Construção dos Modelos

Fonte: *Elaboração Própria, através do SPSS Modeler*

4.5. Avaliação

Após o desenvolvimento do processo explícito na secção anterior, surge a necessidade de verificar o desempenho dos vários modelos desenvolvidos, que é obtido através da função “*Analysis*” do *SPSS Modeler*. Tal desempenho é representado através do valor de “*accuracy*”, em que quanto mais elevada a percentagem, melhor é o desempenho do modelo. Através da *Tabela 17* é perceptível que os modelos obtidos apresentam valores muito semelhantes de “*accuracy*”.

2019		2020	
Modelo	Accuracy	Modelo	Accuracy
C5.0	84,93%	C5.0	83,86%
C&R	84,04%	C&R	83,86%
Regressão Logística	85,92%	Regressão Logística	84,43%

Tabela 17: Accuracy dos Modelos Obtidos

Fonte: *Elaboração Própria, através do SPSS Modeler*

Atendendo ao facto de a diferença do desempenho entre os modelos não ser significativa, procedeu-se à análise e ponderação de cada um dos modelos obtidos para compreender o que melhor se adequava aos dados em estudo. Apesar de se tratar do modelo com maior “accuracy”, o modelo de regressão logística não foi selecionado por já ter sido utilizado em estudos recentes no âmbito de *online reviews* (como é o caso do desenvolvido por Xu e Lee, 2020), mas também devido à inexistência de resultados suscetíveis de uma interpretação que se adeque aos dados utilizados em estudo. Apesar dos modelos obtidos através dos algoritmos “C5.0” e “C&R” serem bastante semelhantes, o modelo nomeado para a apresentação dos resultados foi o “C5.0”. Tal escolha, baseia-se, sobretudo, no fator explicabilidade visto que o modelo C5.0 baseia-se na totalidade das *reviews* (5726 referentes a 2019 e 3661 ao ano de 2020) para a previsão do rating enquanto que o modelo “C&R” exclui 30% das *reviews* e, portanto, ao invés de explicar 100% dos dados, explica apenas 70%. Em suma, é através da análise do modelo “C5.0” que a investigação incide para averiguar quais os atributos associados às *reviews* que originam um rating mais elevado.

Por conseguinte, a *Figura 19* demonstra a importância das variáveis independentes mais relevantes na previsão do *rating*, relativamente ao período pré-*COVID-19*. Assim, é possível constatar que a presença de sentimentos negativos numa *online review* é o fator mais relevante na previsão do rating. Através da *Figura 19* é possível averiguar que os atributos mais expressivos dizem respeito ao conteúdo textual das *reviews*, à exceção do país de origem dos clientes (“*Guest country*”), que também se apresenta correlacionado com o rating.



Figura 19: Variáveis Mais Importantes na Previsão do Rating (2019)

Fonte: Elaboração Própria, através do SPSS Modeler

Relativamente ao período pandémico é a presença de conteúdo com conotação positiva numa *online review* o fator mais significativo para prever o rating das *reviews*.

Isto significa o “*type positive*” está altamente correlacionado com o rating. A *Figura 20* retrata as variáveis mais importantes na previsão do rating no período pandémico. Assim como no período pré-*COVID-19*, é o conteúdo textual das *reviews*, representado através dos “*types*” que tem mais expressão. Contudo, a localização do hotel (Costa del Sol, Barcelona, Lisboa ou Algarve), o tipo de hotel (independente ou de cadeia hoteleira), o tipo de *reviewer* (casal, família, viajante individual ou grupo) e a resposta dos hotéis às *reviews* passam a representar fatores que antecipam o rating.

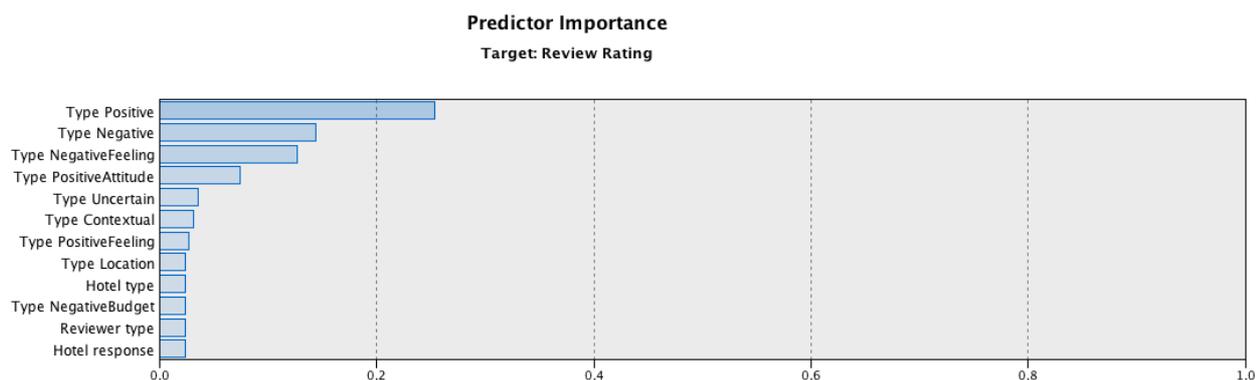


Figura 20: Variáveis Mais Importantes na Previsão do Rating (2020)

Fonte: *Elaboração Própria, através do SPSS Modeler*

Na *Figura 21*, é possível observar a árvore de decisão relativa a 2019, desenvolvida com o intuito de compreender os fatores que originam ratings mais elevados. O tipo “*NegativeRecommendation*” foi identificado como aquele que mais influencia o rating atribuído pelos clientes. Assim, uma *review* que contenha uma recomendação negativa possui uma elevada propensão (70%) para originar ratings inferiores a 7 (considerando uma escala de 2,5 a 10). Enquanto que, uma *review* que não envolva recomendações negativas é expectável (84%) que origine um rating elevado. A combinação do conteúdo expectável para originar resultados mais positivos, através de ratings mais elevados, está representada no Nó 7, que é a combinação de *reviews* que não contenham recomendações negativas, não aludam a atitudes negativas e que contenham uma conotação positiva (86,94%). É também perceptível que é a combinação que menos ratings classificados como “*Low*” e “*Medium*” (valores inferiores a 7, inclusive).

Existe uma propensão de 66% para os ratings médios, aqueles que assumem valores de 4 a 7, derivem de uma *review* que não contenha uma recomendação negativa (“*Type NegativeReccomendation*”) nem conteúdo positivo (“*Type Positive*”) mas que expresse um sentimento negativo (“*Type NegativeFeeling*”).

É, ainda, importante destacar o Nó 12 (*Figura 21*), em que a propensão de existirem ratings classificados como “Low” é nula. Assim, constata-se que na avaliação do desempenho dos hotéis na ótica dos clientes, é a presença de sentimentos positivos e simultânea inexistência de sentimentos negativos que permite a ausência de elevados níveis insatisfação (ratings iguais e inferiores a 4).

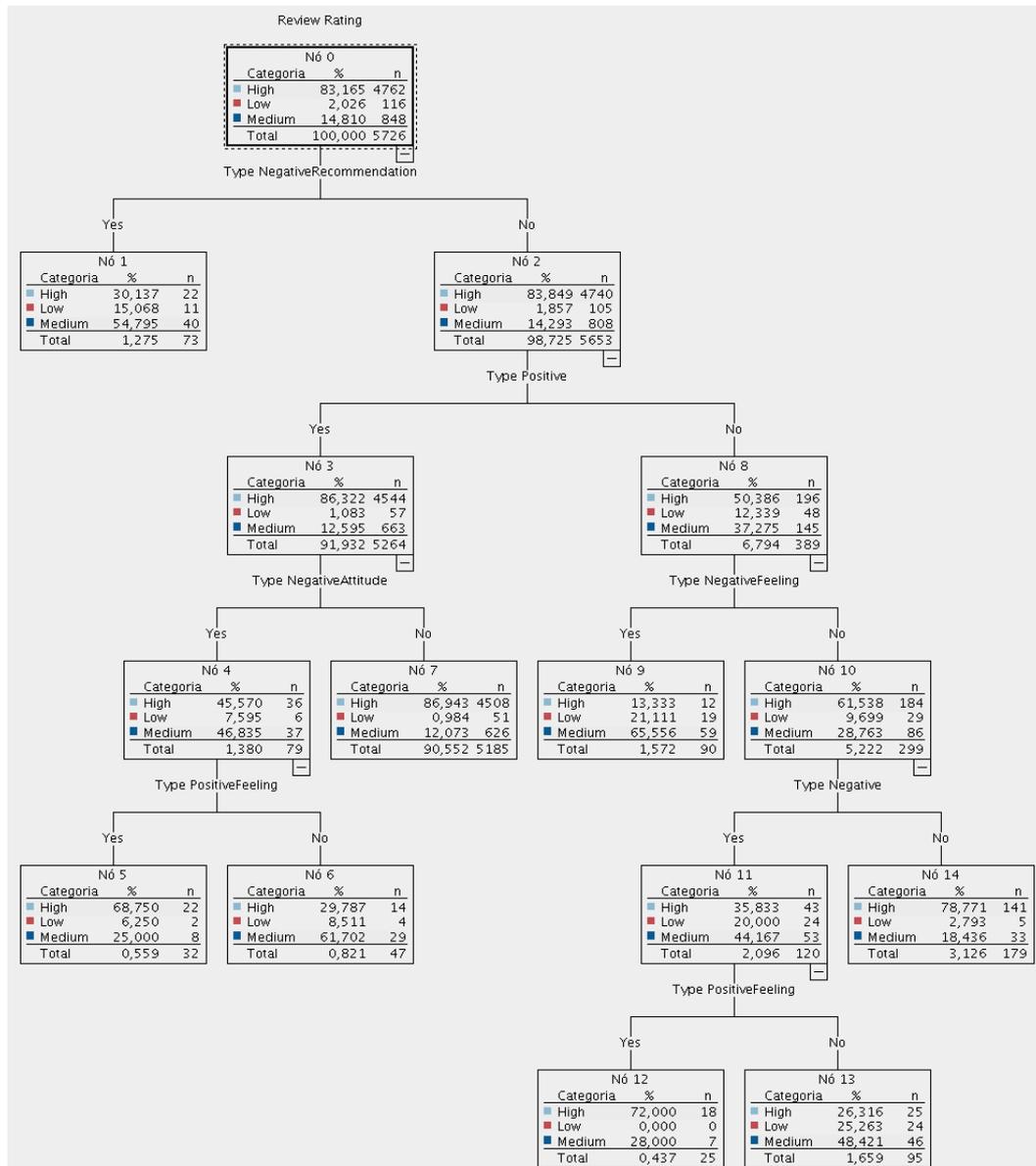


Figura 21: Árvore de Decisão C5.0 (2019)

Fonte: Elaboração Própria, através do SPSS Modeler

A Tabela 18 retrata alguns exemplos de conteúdo associado às variáveis que constituem a árvore de decisão referenciada anteriormente (*Figura 21*).

2019	
Type	Exemplos
Positive	excellent; like; good; normal; correct; satisfactory; no problem; adequate; positive; convenient; acceptable; right; well; interesting
Negative	bad; problem; wrong; slow; complaint; lack; negative; difficult not worth; limited; low-quality; mistake; miserable; worse; too old
PositiveFeeling	clean; comfortable; relaxing; beautiful; good-tasting; fresh; feeling at home; elegant; attractive; modern; quiet; love; stylish
NegativeFeeling	noisy; disappointing; dirty; ugly; smelly; not comfortable boring; nauseating; not attractive; not relaxing; not good-tasting
NegativeAttitude	not friendly; rude; not honest; not attentive; arrogant; not dedicated not professional; cruel; don't care; intrusive; nervous; abusive
NegativeReccomendation	intention to change; will not recommend; no intention to use

Tabela 18: Exemplos de Conceitos Associados às Variáveis Presentes na Árvore de Decisão (2019)

Fonte: Elaboração Própria, através do SPSS Modeler

No que concerne aos dados relativos a 2020, que representam o período pandémico, a árvore de decisão obtida, representada na *Figura 22*, permite-nos constatar que a expressão de um sentimento negativo numa *online review* é o fator que mais explica o seu rating. Deste modo, existe uma probabilidade de 84% de o rating ser elevado (“*High*”), no caso das *reviews* que não contêm sentimentos negativos e, pelo contrário, caso contenham existe uma predisposição de cerca de 66% para originarem ratings inferiores a 7 (“*Low*” e “*Medium*”). O Nó 5 corresponde à combinação do conteúdo textual com a maior propensão (87,4%) de obter ratings mais elevados, ou seja, uma *review* que não contenha qualquer sentimento negativo, mas pelo menos um conteúdo textual com conotação positiva, origina ratings superiores a 7 (considerando uma escala de 1 a 10). A maior probabilidade de ocorrência (26%) de ratings considerados médios verifica-se no Nó 2 que acomoda a combinação da existência de conteúdos negativos e conteúdo com conotação positiva (“*Type Positive*”).

Por outro lado, as *reviews* que contenham conteúdo negativo (“*Type Negative*”), mas não expressem sentimentos negativos nem termos positivos (“*Positive*”), possuem uma propensão de 59% para resultarem em ratings classificados como “*Low*” (iguais ou inferiores a 4).

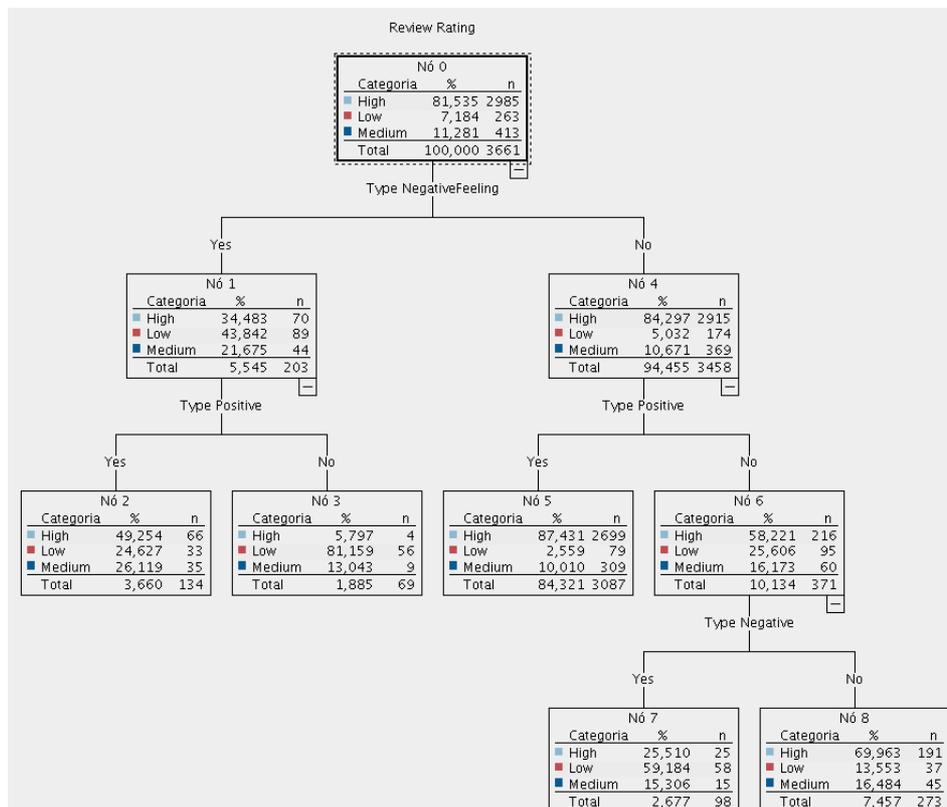


Figura 22: Árvore de Decisão C5.0 (2020)

Fonte: Elaboração Própria, através do SPSS Modeler

Na *Tabela 19* estão representados alguns dos conceitos que mais constituem as variáveis que influenciam o rating relativo às *online reviews* de 2020.

2020	
Type	Exemplos
Positive	excellent; good; safe; well; varied; cool; correct; acceptable; available organized; satisfactory; right; normal; interesting; proper; successful
Negative	bad; problem; wrong; poor; worse; limited; closed; difficult; slow crowded; excessive; dangerous; lack; not enough; chaotic; bad quality
NegativeFeeling	dirty; disappointing; noisy; not comfortable; dusty; boring; disturbed smelly; ugly; tasteless; not attractive; loud; old-fashioned; fear

Tabela 19: Exemplos de Conceitos Associados às Variáveis Presentes na Árvore de Decisão (2020)

Fonte: Elaboração Própria, através do SPSS Modeler

Posto isto, é possível afirmar que a pandemia não provocou alterações significativas na previsão do rating (Figuras 21 e 22), embora o nível de importância de cada variável (Figuras 19 e 20) difira ligeiramente.

CAPÍTULO 5

DISCUSSÃO E CONCLUSÃO

No presente capítulo, são apresentados os principais resultados e limitações da investigação, bem como uma breve discussão com o intuito de relacionar os resultados obtidos com a teoria apresentada no Capítulo 2. Além disso, serão ainda evidenciados os contributos teóricos e práticos resultantes do estudo, bem como sugestões para investigações futuras.

5.1. Principais Resultados

Como principais resultados, derivados da comparação entre os dois períodos temporais em estudo é possível evidenciar: (1) diferenças nos principais indicadores das *online reviews* (volume e valência), validando a 1ª proposição do estudo; (2) inexistência de diferenças expressivas no conteúdo textual das *reviews*, contrapondo-se à 2ª proposição; (3) diminuição da satisfação dos clientes, legitimando a 3ª proposição; (4) ausência de alterações significativas na previsão do rating das *online reviews*, refutando a 4ª proposição.

Atendendo à análise de sentimentos realizada através do *SPSS Modeler*, conclui-se que o *COVID-19* não teve impacto significativo no conteúdo textual das *online reviews*. À semelhança do que acontecia no período antecedente à pandemia, os conceitos que mais se destacam são “*excellent*”, “*good*”, “*room*”, “*location*”, “*hotel*”, “*staff*” e “*breakfast*”. A grande maioria das *reviews* (92%) contém pelo menos um termo com conotação positiva, tal como acontecia num período pré-*COVID-19* (93%).

De forma direta, a temática do *COVID-19* não tem uma expressão significativa na estrutura das *reviews*, dado que, apenas 10% das *reviews* analisadas possui conceitos diretamente relacionados. Contudo, importa referir que apesar da sua pouca representatividade no panorama geral das *reviews*, cerca de 64% dos sentimentos relativos ao *COVID-19* são positivos, o que transmite satisfação dos clientes em relação às medidas adotadas pelos hotéis. Em ambos os períodos temporais os atributos dos hotéis mais abordados dizem respeito às instalações do hotel, à localização, ao quarto e ao staff. No que diz respeito à polaridade dos sentimentos associados aos principais atributos dos

hotéis (preço, serviço, limpeza, quarto, departamento de *F&B*, staff, instalações e localização), são os sentimentos positivos que predominam antes e durante a pandemia, à exceção do serviço que num contexto pandémico são os sentimentos negativos que predominam. Deste modo, é possível constatar que a pandemia está a ter impacto na satisfação dos clientes ao nível do desempenho do serviço dos hotéis.

Os dados obtidos permitem-nos afirmar que para ambos os períodos temporais em estudo, é maioritariamente o conteúdo textual das *reviews* que mais explica o rating atribuído pelos *reviewers*. Todavia, variáveis como o país de origem dos hóspedes (no caso do período antecedente à pandemia) e o tipo de *reviewer* e hotel, no contexto pandémico, contribuem também para a previsão do rating. Por conseguinte, conclui-se que em plena pandemia, uma *review* que não contenha qualquer sentimento negativo, mas sim pelo menos um termo com conotação positiva traduz-se numa predisposição de cerca 87% para gerar ratings mais elevados (superiores a 7, numa escala de 1 a 10).

Além disso, com base nas *reviews* analisadas, a presente investigação conclui que a pandemia realçou a superioridade do nível de satisfação médio (representado pelo rating atribuído pelos *reviewers*) associado aos hotéis independentes, em detrimento daqueles que pertencem a uma cadeia hoteleira. Enquanto que, no período antecedente à pandemia são os hotéis de 4 estrelas que possuem os ratings médios mais elevados, no contexto pandémico são os hotéis de 5 estrelas e os de menor dimensão (constituídos por 50 ou menos quartos), que possuem um maior nível de satisfação.

Apesar do estudo não ter incidido na resposta dos hotéis às *online reviews*, é possível afirmar que existe uma preocupação crescente neste sentido dado que, existe um aumento 10% na frequência de respostas num contexto pandémico, comparando com o período antecedente ao *COVID-19*.

No que concerne ao nível de satisfação dos clientes é possível averiguar que existe uma desigualdade expressiva, comparando o período pandémico ao seu período homólogo, através dos ratings médios obtidos em função das várias variáveis em estudo, mas também da carga sentimental média das *reviews*.

Em suma, a presente investigação constata que a pandemia *COVID-19* está a ter um impacto significativo na satisfação dos clientes nos hotéis, apesar de não existirem diferenças significativas acerca daquilo que transmitem através das *reviews* ou nos fatores preditivos do rating, estes encontram-se mais insatisfeitos. Assim, será necessário que os hotéis aumentem a sua capacidade resiliência num curto prazo temporal, dado que a

pandemia perdurará e simultaneamente será necessário incrementar a satisfação dos clientes.

5.2. Discussão

A vigente pandemia *COVID-19* e, como repercussão, a crise económica mundial, têm sido as temáticas mais abordadas na atualidade. Deste modo, o setor do turismo do qual os estabelecimentos hoteleiros fazem parte, pela sua elevada importância na contribuição para o PIB de inúmeros países, poderá ser um fator chave no fomento da recuperação económica (Abbas et al., 2021). É, portanto, evidente a necessidade de resiliência por parte das empresas hoteleiras (Lee et al., 2013) que requer ações que visem o seu sucesso, no qual a satisfação dos clientes representa um indicador crucial (Xiang et al., 2015). Com o despoletar da pandemia máscaras, acrílicos e distanciamento social passaram a caracterizar o serviço hoteleiro (Seyitoğlu & Ivanov, 2020; Sigala, 2020). Neste sentido, os resultados obtidos no presente estudo constataam uma quebra na satisfação dos clientes em relação ao serviço, no geral, evidenciando um claro impacto da pandemia na prestação do serviço hoteleiro, embora a temática do *COVID-19* não seja praticamente abordada nas *reviews*, de forma direta.

Tal como o que é defendido por autores como Jiang e Wen (2020) e Naumov et al. (2020), os resultados obtidos realçam o aumento da importância da higiene e limpeza dos hotéis, perante um contexto pandémico. O presente estudo vem também reforçar os resultados obtidos por autores como Xu e Li (2016) e Li et al. (2020), por exemplo, evidenciando que as características dos hotéis têm influência na satisfação dos clientes. Evidencia, portanto, que os níveis de satisfação diferem consoante a categoria do hotel (estrelas), bem como a sua dimensão (número de quartos). Acresce-se ainda o facto de os hotéis independentes possuírem ratings mais elevados do que os hotéis que pertencem a uma cadeia hoteleira em ambos os períodos temporais, indo de encontro aos resultados obtidos no estudo de Gao et al. (2018)

Atendendo ao facto da resposta às *reviews* ser apontada como uma das estratégias mais eficazes na gestão da reputação (Baka, 2016; Proserpio & Zervas, 2017; Xie et al., 2017), o aumento na sua frequência evidenciado na presente investigação, reflete uma maior preocupação por parte dos gestores hoteleiros neste sentido.

Os resultados obtidos demonstram que o quarto, a localização, as instalações do hotel e o staff correspondem aos atributos sobre os quais os clientes tendem a refletir mais sobre a sua satisfação, tal como Xu e Lee (2020) defendem. Além disso, outro atributo que demonstra ser particularmente relevante é o departamento de F&B, tal como Zhou et al. (2014) referem.

Para Park e Nicolau (2015) para uma *review* ser considerada útil e de qualidade deve incluir detalhes acerca dos produtos e/ou serviços e das respetivas empresas que os providenciam, o que é salientado pela vigente dissertação, dado que, os resultados demonstram que são as *reviews* de maior dimensão que apresentam mais votos de utilidade.

Apesar de já existir uma tendência para o aumento da preocupação com a satisfação dos hóspedes nos hotéis, a sua mensuração deve ser considerada como uma questão desafiante (Xiang et al., 2015), dado que, atendendo à elevada quantidade de informação gerada diariamente (Xu e Li, 2016), as empresas precisam de adotar métodos que permitam o tratamento de uma elevada quantidade de dados, de forma eficaz. É neste sentido que se torna relevante a adoção de uma metodologia *CRISP-DM* como a utilizada na presente investigação, na qual é reforçada a ideia do seu carácter standard e replicável (Huang e Hsieh, 2020).

5.3. Contributos, Limitações e Reflexões para Investigação Futura

A presente investigação propõe uma nova abordagem na compreensão do impacto provocado pela pandemia *COVID-19* na indústria hoteleira. Neste sentido, a dissertação contribui para mais uma compreensão da satisfação dos clientes nos hotéis, expressa através de *online reviews*, de carácter inovadora. Ou seja, apesar de existir já um número elevado de estudos acerca da satisfação nos hotéis, este é único, pelo menos de conhecimento da autora até à data, no sentido em que trata esta temática perante um contexto único, numa ótica comparativa entre o período pandémico que vivenciamos atualmente e o antecedente. Além disso, destaca-se também a autenticidade da amostra explorada, pois considera-se que pela importância que os destinos turísticos em estudo auferem é necessário investir mais em investigação científica que seja útil para a comunidade científica, mas também para as entidades e empresas turísticas das regiões.

Existem vários contributos teóricos inerentes ao estudo dos quais é possível destacar o facto do país de origem dos clientes bem como a localização, categoria, tipo e dimensão do hotel influenciarem significativamente o nível médio de satisfação dos clientes (representado através do rating). Pelo contrário, nem a duração da estadia nem o tipo de viagem (família, casal, grupo, viajante individual) têm influência na satisfação expressa através do rating das *online reviews*. No entanto, no que diz respeito à análise preditiva dos ratings, esta é, sobretudo, executada com base no conteúdo textual das *reviews*. Perante os resultados da investigação sabe-se que perante uma pandemia, nomeadamente a *COVID-19*, a satisfação dos clientes nos hotéis sofre alterações. Estas alterações consistem na diminuição da carga sentimental das *reviews*, no decréscimo do rating médio, no aumento de sentimentos negativos de forma generalizada, mas principalmente ao nível do serviço.

Importa ainda realçar que a metodologia aplicada na presente dissertação demonstra a aplicabilidade da metodologia *CRISP-DM* a qualquer estabelecimento hoteleiro que tencione adquirir e compreender a informação contida nas *online reviews*, de modo a mensurar a satisfação dos seus clientes e visando o seu aumento. Além disso, pretende incentivar os investigadores à sua utilização, mas também sugerir a sua utilização por parte de entidades oficiais para os mais variados fins.

Contudo, a presente investigação não está isenta de limitações. O estudo incide sobre hotéis localizados em alguns dos destinos turísticos mais expressivos na Península Ibérica: Algarve, Costa del Sol, Lisboa e Barcelona, com o intuito de verificar se a pandemia *COVID-19* está a ter impacto na satisfação dos clientes, representada através do rating e sentimentos associados às *online reviews*. Trata-se, portanto, de uma amostra que se restringe a dois países (Portugal e Espanha), pelo que as suas conclusões não devem ser replicadas aos hotéis de outros destinos. O facto de o site *Booking.com* não ser compatível com o *scraper* inicialmente idealizado para utilização (*ParseHub*) com vista à extração automática das *reviews* representa também uma limitação. Ou seja, o volume de *reviews* utilizado é limitado, pois resulta de uma extração manual.

Além do já referido, constata-se como principal limitação, o facto de a presente dissertação incidir sobre um período temporal limitado e atual, visto que o contexto pandémico perdura.

Tendo por base o desenvolvimento da investigação exposta e respetivos métodos são várias as propostas para investigações futuras, nomeadamente (1) utilizar uma

metodologia distinta e/ou replicar o processo seguido na presente dissertação com hotéis de destinos diferentes dos usados, com o intuito de replicar os resultados obtidos ou alcançar outro tipo de resultados; (2) repetir o estudo, utilizando dados de um período temporal mais alargado e/ou para um segmento específico (por exemplo hotéis de uma determinada categoria ou destino) e, por último (3) aplicar a metodologia empregue a outros contextos, como por exemplo ao setor da restauração, que também foi bastante afetado pela pandemia.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abbas, J., Mubeen, R., Iorember, P. T., Raza, S., & Mamirkulova, G. (2021). Exploring the impact of COVID-19 on tourism: transformational potential and implications for a sustainable recovery of the travel and leisure industry. *Current Research in Behavioral Sciences*, 2, 100033.
- Babić Rosario, A., de Valck, K., & Sotgiu, F. (2020). Conceptualizing the electronic word-of-mouth process: What we know and need to know about eWOM creation, exposure, and evaluation. *Journal of the Academy of Marketing Science*, 48(3), 422–448.
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318.
- Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, 53, 148–162.
- Berka, P., & Rauch, J. (2010). Machine Learning and Association Rules. *19th International Conference on Computational Statistics COMPSTAT 2010*, 1–29.
- Booking.com (2020), *A maior seleção de hotéis, casas e alojamentos para férias, disponível online em: https://www.booking.com/index.pt-pt.html?label=gen173nr1BCAEoggI46AdIM1gEaLsBiAEBmAEfuAEHyAEP2AEB6AEBiAIBqAIDuAKos_qDBsACAdICJGJIZmZhZDZkLTU4NTEtNDM0Yi05NDNkLWRIZGM1YmE5ZjRkMdgCBeACAQ;sid=d29bcff64e04d0109a0429d975e6481a;keep_landing=1&sb_price_type=total* [Dezembro 2020].
- Bronner, F., & de Hoog, R. (2011). Vacationers and eWOM: Who posts, and why, where, and what? *Journal of Travel Research*, 50(1), 15–26.
- Brown, N., Rovins, J., Feldmann-Jensen, S., Orchiston, C., & Johnston, D. (2017). Exploring disaster resilience within the hotel sector: A systematic review of literature. *International Journal of Disaster Risk Reduction*, 22, 362–370.
- Browning, V., So, K., & Sparks, B. (2013). The Influence of Online Reviews on Consumers' Attributions of Service Quality and Control for Service Standards in Hotels. *Journal of Travel and Tourism Marketing*, 30(1–2), 23–40.

- Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675-693.
- Casaló, L., Flavián, C., Guinalíu, M., & Ekinici, Y. (2015). Do online hotel rating schemes influence booking behaviors? *International Journal of Hospitality Management*, 49, 28–36.
- Chakraborty, U., & Biswal, S. (2020). Impact of Online Reviews on Consumer's Hotel Booking Intentions: Does Brand Image Mediate? *Journal of Promotion Management*, 26(7), 943–963.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
- Chatterjee, S. (2020). Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. *International Journal of Hospitality Management*, 85, 102356.
- Chen, W., Gu, B., Ye, Q., & Zhu, K. X. (2019). Measuring and managing the externality of managerial responses to online customer reviews. *Information Systems Research*, 30(1), 81–96.
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477–491.
- Cheung, M., Luo, C., Sia, C., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce*, 13(4), 9–38.
- Christie & Co (2019). *Spanish Hotel Market Resort Destinations*, disponível online em: https://es.christie.com/christieMediaLibraries/christieSpain/PDFs-Publications/Mercado-Hotelero-en-Espana_Destinos-Vacacionales-SINGLE-PAGE.pdf?ext=.pdf [Dezembro 2020].
- Christie & Co (2020). *Destinos Urbanos en España 2019*, disponível online em: <https://es.christie.com/christieMediaLibraries/christieSpain/PDFs-Publications/Destinos-Urbanos-en-Espana-2019.pdf?ext=.pdf>
- Cohen, M. (2020). Does the COVID-19 outbreak mark the onset of a sustainable consumption transition? *Sustainability: Science, Practice, and Policy*, 16(1), 1–3.
- Cristobal-Fransi, E., Daries, N., Martin-Fuentes, E., & Montegut-Salla, Y. (2020). Industrial Heritage 2.0: Internet presence and development of the electronic commerce of industrial tourism. *Sustainability (Switzerland)*, 12(15).

- de Gregorio, F., Fox, A., & Yoon, H. (2021). Pseudo-reviews: Conceptualization and consumer effects of a new online phenomenon. *Computers in Human Behavior*, *114*, 106545.
- Del Chiappa, G., Lorenzo-Romero, C., & Alarcón-del-Amo, M. (2015). Profiling tourists based on their perceptions of the trustworthiness of different types of peer-to-peer applications. *Current Issues in Tourism*, *21*(3), 259–276.
- Duarte Alonso, A., Kok, S., Bressan, A., O’Shea, M., Sakellarios, N., Koresis, A., Buitrago Solis, M., & Santoni, L. (2020). COVID-19, aftermath, impacts, and hospitality firms: An international perspective. *International Journal of Hospitality Management*, *91*, 102654.
- Duchek, S., Raetze, S., & Scheuch, I. (2020). The role of diversity in organizational resilience: a theoretical framework. *Business Research*, *13*(2), 387–423.
- Eggers, F. (2020). Masters of disasters? Challenges and opportunities for SMEs in times of crisis. *Journal of Business Research*, *116*, 199–208.
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, *49*(9), 76-82.
- Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498–506.
- Faulkner, B. (2001). Towards a framework for tourism disaster management. *Tourism Management*, *22*(2), 135–147.
- Filieri, R., Algezau, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, *51*, 174–185.
- Filimonau, V., & De Coteau, D. (2020). Tourism resilience in the context of integrated destination and disaster management (DM2). *International Journal of Tourism Research*, *22*(2), 202–222.
- Gao, B., Li, X., Liu, S., & Fang, D. (2018). How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers’ travel experience. *Tourism Management*, *65*, 176–186.
- Gavilan, D., Avello, M., & Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, *66*, 53–61.

- Godbole, S., Bhattacharya, I., Gupta, A., and Verma, A. (2010). Building re-usable dictionary repositories for real-world text mining. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1189–1198). New York, NY, USA: ACM.
- Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research*, 24(1), 88–107.
- Gonçalves, H., Silva, G., & Martins, T. (2018). Motivations for posting online reviews in the hotel industry. *Psychology and Marketing*, 35(11), 807–817.
- Gössling, S., Scott, D., & Hall, C. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism*, 29(1), 1–20.
- Guerreiro, J., & Rita, P. (2020). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*, 43, 269–272.
- Hennig-Thurau, T., Gwinner, K., Walsh, G., & Gremler, D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Herbane, B. (2019). Rethinking organizational resilience and strategic renewal in SMEs. *Entrepreneurship and Regional Development*, 31(5–6), 476–495.
- Hu, Minqing and Bing Liu (2004). Mining opinion features in customer reviews. In *Proceedings of AAAI Conference on Artificial Intelligence*, vol. 4, pp. 755–760.
- Huang, C. H., & Hsieh, S. H. (2020). Predicting BIM labor cost with random forest and simple linear regression. *Automation in Construction*, 118(May), 103280. <https://doi.org/10.1016/j.autcon.2020.103280>
- IBM (2020), *IBM Docs*, disponível online em: <https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=help-about-spss-modeler-text-analytics> [Janeiro 2021].
- INE- Instituto Nacional de Estatística (2020), *Estatísticas do Turismo 2019*, disponível online em: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=133574&PUBLICACOESmodo=2 [Dezembro 2020].
- Ivanov, S., Webster, C., Stoilova, E., & Slobodskoy, D. (2020). Biosecurity, crisis management, automation technologies and economic performance of travel, tourism and hospitality companies – A conceptual framework. *Tourism Economics*, 1–24.

- Jain, M., Kumar, M., and Aggarwal, N. (2013). Web usage mining: An analysis. *Journal of Emerging Technologies in Web Intelligence*, 5(3), 240–246.
- Jiang, Y., & Wen, J. (2020). Effects of COVID-19 on hotel marketing and management: a perspective article. *International Journal of Contemporary Hospitality Management*, 32(8), 2563–2573.
- Jiménez, F., & Mendoza, N. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing*, 27(3), 226–235.
- Kim, E., Mattila, A., & Baloglu, S. (2011). Effects of gender and expertise on consumers' motivation to read online hotel reviews. *Cornell Hospitality Quarterly*, 52(4), 399–406.
- Kock, F., Nørfelt, A., Josiassen, A., Assaf, A., & Tsionas, M. (2020). Understanding the COVID-19 tourist psyche: The Evolutionary Tourism Paradigm. *Annals of Tourism Research*, 85, 103053.
- Kudeshia, C., & Kumar, A. (2017). Social eWOM: does it affect the brand attitude and purchase intention of brands? *Management Research Review*, 40(3), 310–330.
- Kwok, L., & Xie, K. (2016). Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role? *International Journal of Contemporary Hospitality Management*, 28(10), 2156–2177.
- Kwok, L., Xie, K., & Richards, T. (2017). Thematic framework of online review research. *International Journal of Contemporary Hospitality Management*, 29(1), 307–354.
- Langan, R., Besharat, A., & Varki, S. (2017). The effect of review valence and variance on product evaluations: An examination of intrinsic and extrinsic cues. *International Journal of Research in Marketing*, 34(2), 414–429.
- Laureano, R. M. S., Caetano, N., & Cortez, P. (2014). Previsão de tempos de internamento num hospital português: Aplicação da metodologia CRISP-DM. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, 13, 83–98.
- Lee, A., Vargo, J., & Seville, E. (2013). Developing a tool to measure and compare organizations' resilience. *Natural Hazards Review*, 14(1), 29–41.
- Lee, M., Cai, Y., DeFranco, A., & Lee, J. (2020). Exploring influential factors affecting guest satisfaction: Big data and business analytics in consumer-generated reviews. *Journal of Hospitality and Tourism Technology*, 11(1), 137–153.

- Li, C., Cui, G., & Peng, L. (2017). The signaling effect of management response in engaging customers: A study of the hotel industry. *Tourism Management*, 62, 42–53.
- Li, H., Liu, Y., Tan, C. W., & Hu, F. (2020). Comprehending customer satisfaction with hotels. *International Journal of Contemporary Hospitality Management*, 32(5), 1713–1735.
- Liu, B., Hu, M., & Cheng, J. 2005. Opinion observer. Proceedings of the 14th International Conference on World Wide Web - WWW '05, 342.
- Lo, A., Cheung, C., & Law, R. (2006). The survival of hotels during disaster: A case study of Hong Kong in 2003. *Asia Pacific Journal of Tourism Research*, 11(1), 65–80.
- Lopes, J. P. (2015, Outubro). *Reputação Online: Implicações na Gestão Hoteleira, Caso de Estudo: Pestana Group*, disponível online em: <https://repositorio.iscte-iul.pt/bitstream/10071/11520/1/Jos%C3%A9%20Pedro%20Lopes%20Outubro%202015.pdf> [Outubro 2020]
- Lu, W., & Stephenkova, S. (2015). User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. *Journal of Hospitality Marketing and Management*, 24(2), 119–154.
- Magalhães, L. (2020, Outubro 23). AHP: 25% dos hotéis estiveram fechados em Setembro. *opção turismo*. Disponível online em: <https://opcaoturismo.pt/wp/ahp-25-dos-hoteis-estiveram-fechados-em-setembro/> [Outubro 2020].
- Mao, Z., Yang, Y., & Wang, M. (2018). Sleepless nights in hotels? Understanding factors that influence hotel sleep quality. *International Journal of Hospitality Management*, 74, 189–201.
- Mariani, M., & Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management*, 66, 47–52.
- Mariani, M., Di Fatta, G., & Di Felice, M. (2019). Understanding Customer Satisfaction with Services by Leveraging Big Data: The Role of Services Attributes and Consumers' Cultural Background. *IEEE Access*, 7, 8195–8208.
- Marketing Science Institute. (2020, Maio 7), *MSI Research Priorities 2020–2022*, disponível online em: <http://www.msi.org/wp-content/uploads/2020/09/MSI-2020-22-Research-Priorities-final.pdf> [Setembro 2020]

- Marine-Roig, E., & Huertas, A. (2020). How safety affects destination image projected through online travel reviews. *Journal of Destination Marketing and Management*, 18, 100469.
- Martin-Fuentes, E., Fernandez, C., Mateu, C., & Marine-Roig, E. (2018). Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *International Journal of Hospitality Management*, 69, 75–83.
- Mauri, A., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, 34(1), 99–107.
- Melián-Alzola, L., Fernández-Monroy, M., & Hidalgo-Peñate, M. (2020). Hotels in contexts of uncertainty: Measuring organisational resilience. *Tourism Management Perspectives*, 36, 100747.
- Miles, P., Miles, G., & Cannon, A. (2012). Linking servicescape to customer satisfaction: Exploring the role of competitive strategy. *International Journal of Operations and Production Management*, 32(7), 772–795.
- Moro, S., Esmerado, J., Ramos, P. and Alturas, B. (2019), "Evaluating a guest satisfaction model through data mining", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 4, pp. 1523-1538.
- Moro, S., Lopes, R. J., Esmerado, J., & Botelho, M. (2020). Service quality in airport hotel chains through the lens of online reviewers. *Journal of Retailing and Consumer Services*, 56, 102193.
- Napierała, T., Leśniewska-Napierała, K., & Burski, R. (2020). Impact of geographic distribution of COVID-19 cases on hotels' performances: Case of Polish cities. *Sustainability (Switzerland)*, 12(11), 1–18.
- Naumov, N., Varadzhakova, D., & Naydenov, A. (2020). Sanitation and hygiene as factors for choosing a place to stay: perceptions of the Bulgarian tourists. *Anatolia*, 00(00), 1–4.
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193.
- O'Connor, P. (2010). Managing a hotel's image on Tripadvisor. *Journal of Hospitality Marketing and Management*, 19(7), 754–772.

- OECD- Organization for Economic Co-operation and Development (2020), *Tourism Policy Responses to the coronavirus (COVID-19)*, disponível online em: <https://www.oecd.org/coronavirus/policy-responses/tourism-policy-responses-to-the-coronavirus-covid-19-6466aa20/> [Outubro 2020].
- Oh, H., & Jeong, M. (2010). Evaluating stability of the performance-satisfaction relationship across selected lodging market segments. *International Journal of Contemporary Hospitality Management*, 22(7), 953–974.
- Orange (n.d.), *Orange Widget Catalog*, disponível online em: <https://orangedatamining.com/widget-catalog/> [Março 2021].
- Park, S., & Allen, J. (2013). Responding to Online Reviews: Problem Solving and Engagement in Hotels. *Cornell Hospitality Quarterly*, 54(1), 64–73.
- Park, S., & Nicolau, J. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
- Pateli, A., Mylonas, N., & Spyrou, A. (2020). Organizational adoption of social media in the hospitality industry: An integrated approach based on DIT and TOE frameworks. *Sustainability (Switzerland)*, 12(17).
- Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: A systematic literature review. *PeerJ Computer Science*, 6, 1–43.
- Prayag, G. (2018). Symbiotic relationship or not? Understanding resilience and crisis management in tourism. *Tourism Management Perspectives*, 25, 133–135.
- Proserpio, D., & Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5), 645–665.
- Pyle, M., Smith, A., & Chevtchouk, Y. (2021). In eWOM we trust: Using naïve theories to understand consumer trust in a complex eWOM marketplace. *Journal of Business Research*, 122, 145–158.
- Rhee, H., & Yang, S. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior*, 50, 576–587.
- Sainaghi, R., Baggio, R., Phillips, P., & Mauri, A. (2018). Hotel performance and research streams: a network cluster analysis. *International Journal of Contemporary Hospitality Management*, 32(2), 425–462.

- Sanchez-Franco, M., & Rondan-Cataluña, F. (2010). Virtual travel communities and customer loyalty: Customer purchase involvement and web site design. *Electronic Commerce Research and Applications*, 9(2), 171–182.
- Schröer, C., Kruse, F., Marx, J., Kruse, F., & Marx, J. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181(2019), 526–534.
- Serra Cantallops, A., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, 36, 41–51.
- Seyitoğlu, F., & Ivanov, S. (2020). A conceptual framework of the service delivery system design for hospitality firms in the (post-)viral world: The role of service robots. *International Journal of Hospitality Management*, 91.
- Sheng, M. (2012). The utilitarian and social dual presence in Web 2.0 services. *Total Quality Management and Business Excellence*, 23(7–8), 875–890.
- Shin, H., & Kang, J. (2020). International Journal of Hospitality Management Reducing perceived health risk to attract hotel customers in the COVID-19 pandemic era: Focused on technology innovation for social distancing and cleanliness. *International Journal of Hospitality Management*, 91, 102664.
- Shin, S., Chung, N., Xiang, Z., & Koo, C. (2019). Assessing the Impact of Textual Content Concreteness on Helpfulness in Online Travel Reviews. *Journal of Travel Research*, 58(4), 579–593.
- Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, 117, 312–321.
- Simão, R. (2021, Março 3). Receitas na hotelaria nacional registam quebra de 73% em 2020. *Publituris*. Disponível online em: <https://www.publituris.pt/2021/03/03/receitas-na-hotelaria-nacional-registam-quebra-de-73-em-2020/> [Março 2021].
- Smeral, E. (2010). Impacts of the world recession and economic crisis on tourism: Forecasts and potential risks. *Journal of Travel Research*, 49(1), 31–38.
- Sparks, B., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6), 1310–1323.

- Sparks, B., So, K., & Bradley, G. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management, 53*, 74–85.
- Su, A. (2004). Customer satisfaction measurement practice in Taiwan hotels. *International Journal of Hospitality Management, 23*(4), 397–408.
- Torres, E., Singh, D., & Robertson-Ring, A. (2015). Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. *International Journal of Hospitality Management, 50*, 77–83.
- Tse, C., & Ho, S. (2009). Service quality in the hotel industry: When cultural contexts matter. *Cornell Hospitality Quarterly, 50*(4), 460–474.
- Tsionas, M. (2020). COVID-19 and gradual adjustment in the tourism, hospitality, and related industries. *Tourism Economics*, 1–5.
- UNWTO (2020), *International Tourism Highlights*, disponível online em: <https://www.e-unwto.org/doi/epdf/10.18111/9789284422456> [Dezembro 2020].
- UNWTO (2021, Março 31), *Tourist Arrivals Down 87% in January 2021 as UNWTO calls for Stronger Coordination to Restart Tourism*, disponível online em: <https://www.unwto.org/taxonomy/term/347> [Abril 2021]
- Verma, V., & Chandra, B. (2018). Sustainability and customers' hotel choice behaviour: a choice-based conjoint analysis approach. *Environment, Development and Sustainability, 20*(3), 1347–1363.
- Wang, C., & Zhang, P. (2012). The evolution of social commerce: The people, management, technology, and information dimensions. *Communications of the Association for Information Systems, 31*(1), 105–127.
- Wattanacharoensil, W., Schuckert, M., Graham, A., & Dean, A. (2017). An analysis of the airport experience from an air traveler perspective. *Journal of Hospitality and Tourism Management, 32*, 124–135.
- WHO- World health Organization (2020), *How is COVID-19 transmitted?*, disponível online em: <https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted> [Outubro 2020].
- Xiang, Z., Schwartz, Z., Gerdes, J., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120–130.

- Xie, K., So, K., & Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management*, 62, 101–110.
- Xie, K., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1–12.
- Xu, F., La, L., Zhen, F., Lobsang, T., & Huang, C. (2019). A data-driven approach to guest experiences and satisfaction in sharing. *Journal of Travel and Tourism Marketing*, 36(4), 484–496.
- Xu, X. (2019). Examining the Relevance of Online Customer Textual Reviews on Hotels' Product and Service Attributes. *Journal of Hospitality and Tourism Research*, 43(1), 141–163.
- Xu, X., & Lee, C. (2020). Utilizing the platform economy effect through EWOM: Does the platform matter? *International Journal of Production Economics*, 227(71872200), 107663.
- 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, 57–69.
- Xu, Y., Zhang, Z., Nicolau, J., & Liu, X. (2020). How do hotel managers react to rating fluctuation? *International Journal of Hospitality Management*, 89, 102563.
- Yang, Y., Zhang, H., & Chen, X. (2020). Annals of Tourism Research Coronavirus pandemic and tourism: Dynamic stochastic general equilibrium modeling of infectious disease outbreak. *Annals of Tourism Research*, 83, 102913.
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180–182.
- Yen, C., & Tang, C. (2019). The effects of hotel attribute performance on electronic word-of-mouth (eWOM) behaviors. *International Journal of Hospitality Management*, 76, 9–18.
- Ying, T., Wang, K., Liu, X., Wen, J., & Goh, E. (2020). Rethinking game consumption in tourism: a case of the 2019 novel coronavirus pneumonia outbreak in China. *Tourism Recreation Research*, 0(0), 1–6.

- Zhang, K., Hou, Y., & Li, G. (2020). Threat of infectious disease during an outbreak: Influence on tourists' emotional responses to disadvantaged price inequality. *Annals of Tourism Research*, 84, 102993.
- Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694–700.
- Zhang, Z., Zhang, Z., Wang, F., Law, R., & Li, D. (2013). Factors influencing the effectiveness of online group buying in the restaurant industry. *International Journal of Hospitality Management*, 35, 237–245.
- Zhou, L., Ye, S., Pearce, P., & Wu, M. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1–10.
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on Sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133–148.

Anexo A.1 Teste para Igualdade de Médias do Rating dos hotéis localizados no Algarve

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1474	8.191	1.6677	1140	7.785	1.9069	5.689	2270.74	<.001	.4054

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo A.2 Teste para Igualdade de Médias do Rating em Barcelona (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1170	8.351	1.7287	467	7.913	2.1185	3.964	726.177	<.001	.4372

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo A.3 Teste para Igualdade de Médias do Rating na Costa del Sol (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1377	8.208	1.6201	1431	7.730	2.0877	6.783	2686.46	<.001	.4773

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo A.4 Teste para Igualdade de Médias do Rating em Lisboa (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1705	8.629	1.4975	623	8.480	1.7502	1.891	973.86	.059	.1493

Fonte: Elaboração Própria, através do SPSS Statistics

ANEXO B

Anexo B.1 Teste para Igualdade de Médias do Rating em Hotéis pertencentes a cadeias hoteleiras (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
3985	8.299	1.6898	2686	7.830	1.9985	9.985	5098.47	<.001	.4687

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo B.2 Teste para Igualdade de Médias do Rating em Hotéis Independentes (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1741	8.493	1.4775	975	8.086	1.9923	5.577	1582.398	<.001	.4070

Fonte: Elaboração Própria, através do SPSS Statistics

ANEXO C

Anexo C.1 Teste para Igualdade de Médias do Rating em Hotéis de 2 estrelas (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1348	8.317	1.5077	696	7.789	1.9910	6.149	1116.96	<.001	.5283

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo C.2 Teste para Igualdade de Médias do Rating em Hotéis de 3 estrelas (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1724	7.967	1.7483	1256	7.560	2.0465	5.707	2441.55	<.001	.4078

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo C.3 Teste para Igualdade de Médias do Rating em Hotéis de 4 estrelas (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1238	8.688	1.4583	1023	8.064	2.0035	8.308	1823.85	<.001	.6240

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo C.4 Teste para Igualdade de Médias do Rating em Hotéis de 5 estrelas (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1416	8.584	1.6366	686	8.382	1.7846	2.49	1255.97	.013	.2013

Fonte: Elaboração Própria, através do SPSS Statistics

ANEXO D

Anexo D.1 Teste para Igualdade de Médias do Rating: “*Couple*” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
2649	8.285	1.6583	1884	7.908	2.0447	6.618	3510.6	<.001	.3777

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo D.2 Teste para Igualdade de Médias do Rating: “*Family*” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
2023	8.423	1.6097	1086	7.856	1.9303	8.26	1903.91	<.001	.5673

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo D.3 Teste para Igualdade de Médias do Rating: “*Group*” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
565	8.445	1.5696	298	8.033	2.0557	3.030	483.65	.003	.4126

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo D.4 Teste para Igualdade de Médias do Rating: “*Solo Traveller*” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
489	8.381	1.6216	393	7.869	1.9290	4.197	765.39	<.001	.5114

Fonte: Elaboração Própria, através do SPSS Statistics

ANEXO E

Anexo E.1 Teste para Igualdade de Médias do Rating: clientes “*Eastern Europe*” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
120	8.538	1.5020	52	7.885	2.3149	1.870	70.301	.066	.6529

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo E.2 Teste para Igualdade de Médias do Rating: clientes “Northen Europe” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1173	8.595	1.5530	328	7.995	2.0469	4.921	437.588	<.001	.5993

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo E.3: Teste para Igualdade de Médias do Rating: clientes “Portugal” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
729	8.193	1.5589	896	7.752	1.9272	5.095	1622.95	<.001	.4406

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo E.4 Teste para Igualdade de Médias do Rating: clientes “South and Central America” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
733	8.619	1.5114	65	7.942	2.2508	2.381	69.211	.020	.6778

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo E.5 Teste para Igualdade de Médias do Rating: clientes “Southern Europe” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
67	8.600	1.4659	68	8.147	1.8947	1.555	25.946	.123	.4529

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo E.6 Teste para Igualdade de Médias do Rating: clientes “Spain” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1742	8.072	1.6882	1972	7.888	2.0144	3.032	3701.89	.002	.1843

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo E.7 Teste para Igualdade de Médias do Rating: clientes “Western Europe” (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
302	8.524	1.5341	217	8.321	1.8803	1.307	405.47	.192	.2028

Fonte: Elaboração Própria, através do SPSS Statistics

ANEXO F

Anexo F Teste Não Paramétrico Wilcoxon ao Rating: clientes “North America” (2019 e 2020)

2019			2020			Teste Wilcoxon		
N	Média	Desvio Padrão	N	Média	Desvio Padrão	Wilcoxon W	Z	Sig. (2-tailed)
354	8.511	1.8202	25	8.200	2.0616	4444	-.589	.556

Fonte: Elaboração Própria, através do SPSS Statistics

ANEXO G

Anexo G.1 Teste para Igualdade de Médias do Rating em hotéis com <= 50 quartos (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
810	8.676	1.3553	437	8.381	1.5302	3.368	805.383	<.001	.2941

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo G.2 Teste para Igualdade de Médias do Rating em hotéis com > 50 e <=100 quartos (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1591	8.520	1.5331	809	8.098	1.9483	5.365	1329.95	<.001	.4214

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo G.3 Teste para Igualdade de Médias do Rating em hotéis com > 100 e <=200 quartos (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1095	8.116	1.7503	829	7.792	2.0689	3.63	1610.54	<.001	.3239

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo G.4 Teste para Igualdade de Médias do Rating em hotéis com > 200 e <=300 quartos (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1830	8.190	1.6886	1351	7.684	2.0710	7.354	2547.44	<.001	.5059

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo G.5 Teste para Igualdade de Médias do Rating em hotéis com > 300 quartos (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
400	8.500	1.7225	235	7.917	2.0763	3.638	421.11	<.001	.5839

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo H.1 Teste para Igualdade de Médias do Rating para estadias de 1 noite (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1373	8.479	1.6253	1201	7.865	2.0781	8.267	2261.48	<.001	.6142

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo H.2 Teste para Igualdade de Médias do Rating para estadias de 2 noites (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1198	8.248	1.6714	919	7.955	2.0135	3.575	1766.44	<.001	.2936

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo H.3 Teste para Igualdade de Médias do Rating para estadias de 3 noites (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
1142	8.363	1.6477	546	7.904	1.8766	4.887	958.61	<.001	.4592

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo H.4 Teste para Igualdade de Médias do Rating para estadias de 4 noites (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
781	8.435	1.5676	344	7.952	1.9239	4.100	552.17	<.001	.4834

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo H.5 Teste para Igualdade de Médias do Rating para estadias de 5 noites (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
455	8.355	1.5894	218	7.793	2.0605	3.554	344.96	<.001	.5623

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo H.6 Teste para Igualdade de Médias do Rating para estadias de ≥ 6 e < 12 noites (2019 e 2020)

2019			2020			Teste T para Igualdade de Médias			
N	Média	Desvio Padrão	N	Média	Desvio Padrão	t	df	Sig. (2-tailed)	Diferença Média
745	8.222	1.6223	420	7.886	1.9351	3.007	750.48	.003	.3355

Fonte: Elaboração Própria, através do SPSS Statistics

Anexo I **Top 100: conceitos extraídos (2019)**

2019							
Concept	Global	Reviews	%	Concept	Global	Reviews	%
excellent	5885	3451	60%	cleanliness	173	172	3%
good	3887	2631	46%	coffee	171	157	3%
room	2547	1952	34%	night	168	149	3%
location	2119	1935	34%	area	166	154	3%
hotel	1885	1381	24%	attentive	163	159	3%
staff	1813	1646	29%	well-located	163	155	3%
breakfast	1509	1386	24%	better	162	153	3%
comfortable	991	918	16%	floor	162	143	2%
friendly	924	882	15%	problem	160	150	3%
no	887	738	13%	wifi	160	151	3%
clean	864	814	14%	refrigerator	159	153	3%
pool	806	702	12%	poor	157	139	2%
bed	728	668	12%	not comfortable	155	147	3%
bad	665	539	9%	drinks	152	143	2%
small	562	503	9%	noise	151	138	2%
stay	525	483	8%	pleasant	150	143	2%
helpful	493	469	8%	lack	146	135	2%
like	464	409	7%	que	142	120	2%
beach	440	390	7%	noisy	138	130	2%
bathroom	432	401	7%	space	138	124	2%
restaurant	394	364	6%	children	137	120	2%
service	378	347	6%	variety	135	133	2%
nearby	375	342	6%	choice	135	130	2%
food	353	326	6%	love	134	124	2%
nothing	348	341	6%	dislike	132	125	2%
views	325	300	5%	lisbon	131	122	2%
close to	313	300	5%	relaxing	129	121	2%
lovely	293	256	4%	buffet	129	122	2%
quiet	287	261	5%	towel	126	123	2%
parking	287	254	4%	disappointing	124	119	2%
facilities	280	272	5%	modem	124	114	2%
price	278	260	5%	water	123	110	2%
expensive	259	236	4%	available	121	112	2%
beautiful	247	215	4%	no problem	121	117	2%
bar	231	216	4%	time	119	113	2%
place	226	209	4%	acceptable	119	118	2%
old	225	203	4%	metro	115	112	2%
large	223	204	4%	car	115	98	2%
air conditioning	222	213	4%	door	114	105	2%
people	210	191	3%	check-in	113	107	2%
roomy	203	198	3%	balcony	113	109	2%
front desk	201	190	3%	family	112	105	2%
cleaning	195	186	3%	terrace	112	101	2%
reception	189	173	3%	center	111	108	2%
not working	184	164	3%	dinner	110	99	2%
day	184	168	3%	options	110	97	2%
free	184	172	3%	walk	109	99	2%
dirty	176	155	3%	attention	107	102	2%
shower	174	160	3%	general satisfaction	107	105	2%
quality	174	167	3%	airport	107	98	2%

Fonte: Elaboração Própria, através do SPSS Modeler

Anexo J Top 100: conceitos extraídos (2020)

2020							
Concept	Global	Reviews	%	Concept	Global	Reviews	%
excellent	3518	2166	59%	dirty	139	122	3%
good	2336	1626	44%	buffet	138	120	3%
room	1775	1293	35%	lack	137	128	3%
location	1284	1171	32%	day	134	122	3%
hotel	1277	936	26%	shower	134	126	3%
staff	1150	1029	28%	cleanliness	130	130	4%
breakfast	840	741	20%	terrace	130	124	3%
pool	751	649	18%	poor	126	114	3%
friendly	685	648	18%	water	123	109	3%
comfortable	644	585	16%	attention	119	112	3%
clean	528	483	13%	lovely	119	104	3%
beach	435	367	10%	front desk	118	113	3%
bed	425	383	10%	variety	118	113	3%
covid-19	424	382	10%	free	114	99	3%
small	387	337	9%	floor	114	96	3%
like	378	327	9%	acceptable	113	111	3%
bad	365	312	9%	expensive	113	100	3%
stay	362	329	9%	space	108	99	3%
views	318	291	8%	dislike	107	102	3%
perfect	297	268	7%	better	106	102	3%
service	272	245	7%	night	106	97	3%
restaurant	258	232	6%	well-located	103	101	3%
bathroom	254	234	6%	sea	100	88	2%
parking	252	225	6%	noise	99	89	2%
food	234	214	6%	served	98	90	2%
time	229	206	6%	children	98	84	2%
nothing	227	218	6%	lunch	97	88	2%
facilities	217	208	6%	pleasant	97	95	3%
price	206	187	5%	refrigerator	96	92	3%
old	202	176	5%	drinks	94	89	2%
air conditioning	201	186	5%	disappointing	94	85	2%
quiet	200	184	5%	available	94	88	2%
cleaning	197	191	5%	dinner	92	82	2%
people	194	168	5%	side	90	80	2%
helpful	193	184	5%	safe	90	85	2%
place	185	164	4%	general satisfaction	89	86	2%
beautiful	166	143	4%	meal	89	83	2%
bit	165	156	4%	open	86	79	2%
not working	157	139	4%	towel	85	76	2%
quality	156	147	4%	modern	82	80	2%
attentive	151	142	4%	experience	81	72	2%
large	151	137	4%	pillows	81	79	2%
nearby	147	136	4%	choice	81	77	2%
roomy	146	141	4%	wifi	80	76	2%
problem	145	122	3%	relaxing	79	79	2%
area	143	132	4%	customers	79	69	2%
bar	143	133	4%	recommend	79	77	2%
close to	142	138	4%	use	78	67	2%
reception	140	129	4%	elevator	78	73	2%
measures	139	130	4%	professional	78	75	2%

Fonte: Elaboração Própria, através do SPSS Modeler

Anexo K “Types” extraídos (2019 e 2020)

2019				2020			
Type	Global	Reviews	%	Type	Global	Reviews	%
Positive	15415	5337	93%	Positive	9396	3376	92%
Location	6198	3311	58%	Contextual	5301	2108	58%
Unknown	7344	3080	54%	Unknown	5157	1972	54%
Contextual	6286	2890	50%	Location	4084	2157	59%
Hotel General	4786	2706	47%	Hotel General	3641	1875	51%
Room	4292	2678	47%	Room	2848	1679	46%
Negative	4456	2374	41%	Negative	2796	1498	41%
Personnel	3115	2367	41%	RoomAmenities	2152	1206	33%
PositiveFeeling	3173	2116	37%	Personnel	2149	1542	42%
RoomAmenities	3258	1916	33%	PositiveFeeling	2039	1332	36%
Restaurant	2263	1839	32%	HotelAmenities	1684	1182	32%
HotelAmenities	2496	1700	30%	Restaurant	1467	1096	30%
TimePeriod	1941	1272	22%	TimePeriod	1411	899	25%
PositiveAttitude	1506	1251	22%	PositiveAttitude	1110	882	24%
SportsFacilities	1305	990	17%	SportsFacilities	1080	853	23%
Budget	1203	866	15%	Food	945	549	15%
Food	1370	816	14%	ActionDone	855	590	16%
Services	1013	813	14%	Covid-19	797	559	15%
NO	887	738	13%	Budget	749	573	16%
NegativeFeeling	885	701	12%	Services	659	518	14%
ActionDone	826	617	11%	NegativeFeeling	571	446	12%
PositiveCompetence	679	614	11%	Cleaning	496	442	12%
Drinks	777	538	9%	Drinks	478	337	9%
FoodPlaces	582	510	9%	FoodPlaces	408	341	9%
Uncertain	533	497	9%	Uncertain	384	354	10%
Cleaning	485	447	8%	NegativeFunctioning	375	313	9%
NegativeFunctioning	505	411	7%	PositiveCompetence	363	318	9%
NegativeBudget	439	369	6%	Time	220	165	5%
PositiveBudget	338	307	5%	NegativeBudget	208	182	5%
Suggestion	294	280	5%	PositiveBudget	203	176	5%
Time	309	240	4%	CheckInOut	176	144	4%
Internet	255	226	4%	Organization	174	149	4%
NegativeAttitude	258	220	4%	Suggestion	156	150	4%
CheckInOut	275	219	4%	NegativeAttitude	142	121	3%
PositiveRecommendation	194	183	3%	Internet	140	122	3%
Person	165	158	3%	Person	132	119	3%
Reservation	158	136	2%	Reservation	127	101	3%
Organization	150	128	2%	PositiveRecommendation	99	92	3%
NegativeCompetence	111	108	2%	PositiveFunctioning	92	87	2%
PositiveFunctioning	109	103	2%	Product	59	56	2%
Bill	107	98	2%	Membership	55	53	1%
Currency	98	84	1%	NegativeRecommendation	55	53	1%
NegativeRecommendation	94	83	1%	NegativeCompetence	42	41	1%
Product	62	60	1%	Bill	41	39	1%
Date	62	60	1%	Date	39	37	1%
Weights-Measures	62	57	1%	Currency	37	34	1%
PositiveFeeling_Emoticon	45	43	1%	PositiveFeeling_Emoticon	31	27	1%
FoodType	33	33	1%	Weights-Measures	30	24	1%
Age	33	33	1%	Percent	27	27	1%
Frequency	21	21	0,37%	ProblemResolution	23	22	1%
ProblemResolution	19	18	0,31%	Frequency	22	22	1%
Percent	16	15	0,26%	FoodType	16	16	0,44%
ActionNotDone	5	5	0,09%	Age	12	11	0,30%
NegativeFeeling_Emoticon	5	5	0,09%	ActionNotDone	6	6	0,16%
Membership	4	4	0,07%	NegativeFeeling_Emoticon	4	4	0,11%
Hashtag	2	2	0,03%	UserName	1	1	0,03%

Fonte: Elaboração Própria, através do SPSS Modeler

Anexo L Exemplos de conceitos associados aos principais “types”

Type	Exemplos de Conceitos
Positive	good; better than expected; right; exceptional; unique; well; excellent
Location	attractions; historic center; airport; views; metro; location of the hotel
Unknown	loungers; shade; luggage; flight; alarm; score; mosquitoes
Contextual	information; half; maximum; white; communication; effort; yellow; strong
Hotel General	entrance; 2-star hotel; 4-star hotel; boutique hotel; building
Room	room; bathroom; balcony; double room; bedroom
Negative	hard; a bit small; not worth; ridiculous; too old; lack; worse; bad; limited
Personnel	dining staff; team; staff; treatment of staff; cleaning staff; concierge; front desk;
Positive Feeling	gorgeous; beautiful; love; stylish; good-tasting; comfortable; quiet; clean
Room Amenities	hair conditioner; tv; minibar; sheet; towel;
Restaurant	buffet dinner; breakfast buffet; menu; buffet; restaurant
Hotel Amenities	hall; rooftop; roof; grounds; sky bar; parking; kids club; sun loungers; facilities
Time Period	next morning; current; weekend; next day; period; night
Positive Attitude	proactive; motivated; friendly; pay attention; not intrusive; willing to help; kind;
Sports Facilities	pool; tennis court; gym; pool area;
Budget	charge; commission; budget; credit; fee; ratio; price; discount; deal
Food	croissant; chicken; cakes; cereal; food; salad; bread; cheese; fruit; eggs
Services	service; room service; entertainment; activities
NO	no
Negative Feeling	loud; not relaxing; disappointing; smelly; noisy; dirty; not fresh
Action Done	provided; replaced; carried out; allocated; informed; prepared; explained
Positive Competence	knowledgeable; efficient; helpful; professional; educated
Drinks	tea; water; coffee; juice; lemonade; drinks; wine; gin; beer; coffee; cocktails
Food Places	italian restaurant; coffee bar; snack bar; restaurant; bistro
Uncertain	not sure; don't know; no preference; no experience; nothing
Cleaning	cleaning; overall cleanliness; cleaning noises; hygiene of the hotel

NegativeFunctioning	not working; stuck; broken; not functional; needs to be replaced;
NegativeBudget	expensive; pay extra; not free; have to pay; billed twice; additional cost
PositiveBudget	free; cheap; good value for money; economical; affordable; included in the price
Suggestion	needs; would have; would be; recommendable
Time	hours; 09:00 a.m.; 08:00 a.m.
Internet	wifi; network; wifi coverage; internet; access to wifi
NegativeAttitude	selfish; not careful; sarcastic; intrusive; rude; not friendly; not attentive
CheckInOut	early arrival; queue to check-in; check-out; early check-out; late check-out
PositiveRecommendation	recommend; would recommend
Reservation	booking; reservation request; reservation; booking site
Organization	queues; organization; line;
NegativeCompetence	inexperienced; not organized; not fluent; not qualified
PositiveFunctioning	able to use; functional; functioning
Bill	deposit; taxes; supplement; safety deposit; extra
Currency	USD; EUR; GBP
NegativeRecommendation	will not recommend; no intention to use;
Date	August; October; July; June; mid-august
Weights-Measures	100m; 50m; 1.5 km
FoodType	seafood; fast food; junk food; vegetarian; local food
Frequency	once a week; twice per day; 3 times; numerous times
Covid-19	rules; restrictions; covid-19; mask; safety distance; alcohol gel

Fonte: Elaboração Própria, através do SPSS Modeler

Anexo M Principais Ligações Textuais (2019 e 2020)

LIGAÇÃO	ANO	EXEMPLOS	ANO	EXEMPLOS
“LOCATION” + “POSITIVE”	2019	Excellent: sea view; city; views; place; beach Close to: metro; metro station; airport; subway; beach Good: ubication, area, place, location of the hotel Location: convenient; like; lovely; excellent	2020	Excellent: situation; site; sea; location of the hotel; views Close to: airport; beach Good: ubication; area; location of the hotel; beach; place Exciting: views
“LOCATION” + “NEGATIVE”	2019	Bad: area; views, beach; location Not central: location Poor: location Busy: street Long: walk	2020	Bad: views; beach; access to the beach Difficult: access to the beach Poor: beach Busy: area
“HOTEL GENERAL” + “POSITIVE”	2019	Hotel: large; right; simple; well-located; like; lovely Stay: pleasant Excellent: place to stay; property; resort; terrace Good: experience; stay; quality	2020	Hotel: like, lovely; well-located; good Excellent: hotel; attention; experience; quality; floor; stay; terrace Good: experience, stay, quality Stay: pleasant; like

<p>“HOTEL GENERAL” + “NEGATIVE”</p>	2019	<p>Hotel: worse; busy; lack; negative Experience: bad Walls: thin Quality: poor</p>	2020	<p>Hotel: worse; too old; bad; problem Experience: bad Too much: noise Walls: thin Quality: poor Décor: out of date</p>
<p>“ROOM” + “POSITIVE”</p>	2019	<p>Room: upgraded; ready; lovely; like; available; safe; tidy; superior; large; excellent; spotless Good: balcony; size of the room Accommodations: excellent Bathroom: good; large</p>	2020	<p>Room: available; large; excellent; simple; spotless; superior; lovely; like Excellent: view of the room; balcony Bathroom: excellent; good</p>
<p>“ROOM” + “NEGATIVE”</p>	2019	<p>Bathroom: narrow; out of date; small Room: dark; too small; a bit small; bad; not cleaned; too old; not ready; wrong</p>	2020	<p>Room: too old; too small; too hot; dislike; not large; not ready; dark; poor; bad; a bit small; poorly equipped; out of date Bathroom: tires; dislike; out of date Soundproofing of the room: poor</p>
<p>“PERSONNEL” + “POSITIVE”</p>	2019	<p>Staff: like; available; excellent; lovely; correct; fast; pleasant Good: staff; front desk; reception Excellent: treatment of staff; front desk; reception; staff service</p>	2020	<p>Staff: available; lovely; pleasant; like; excellent Excellent: restaurant staff; hospitality; team; reception; treatment of all staff; hotel staff; member of staff Good: staff; treatment of staff</p>
<p>“PERSONNEL” + “NEGATIVE”</p>	2019	<p>Bad: attitude; reception; management; staff; front desk Staff: dislike; not enough Busy: front desk Reception: problem; complained; complaint</p>	2020	<p>Staff: not sufficient; bad; lack; problem Slow: reception; front desk</p>

<p>“RESTAURANT” + “POSITIVE”</p>	2019	<p>Breakfast: lovely; correct; memorable; love; simple; worth; like; rich; normal; high quality; good; varied Excellent: meal; dinner; breakfast buffet; buffet</p>	2020	<p>Breakfast: simple; reasonable; correct; rich; varied; abundant; like; well-organized; plentiful; excellent; good Excellent: buffet breakfast; dinner; meal</p>
<p>“RESTAURANT” + “NEGATIVE”</p>	2019	<p>Bad: dinner; breakfast buffet; buffet; meal; breakfast Breakfast: not change; weak; limited; dislike; chaotic; out of stock; not varied; crowded; poor; not variety</p>	2020	<p>Breakfast: limited; bad; poor; not varied; weak; doesn't meet expectation Meal: not variety; repetitive</p>
<p>“HOTEL AMENITIES” + “POSITIVE”</p>	2019	<p>Parking: easy; available; nearby; good Good: rooftop; Rooftop bar: lovely; excellent; good Excellent: sky bar; swimming pool; rooftop pool; bar; facilities</p>	2020	<p>Parking: nearby; available; easy; good Good: garden; hotel facilities Excellent: bar; roof; rooftop; rooftop pool Facilities: like; excellent; good</p>
<p>“HOTEL AMENITIES” + “NEGATIVE”</p>	2019	<p>Parking: limited; lack; narrow; difficult to find; difficult; bad Facilities: too old; bad Elevator: slow</p>	2020	<p>Park: difficult; not easy; not possible; not nearby Parking: difficult; problem; lack; difficult to find; not available Elevator: bad, slow Bar: closed Facilities: out of date</p>
<p>“SPORTS FACILITIES” + “POSITIVE”</p>	2019	<p>Excellent: sauna; pool terrace; Pool: cool; like; good; love; large; lovely Spa: good; excellent Gym: good; excellent</p>	2020	<p>Excellent: views from the pool; pool area Pool: pleasant; large; close to; good; well; lovely; like; excellent</p>
<p>“SPORTS FACILITIES” + “NEGATIVE”</p>	2019	<p>Pool: bad; crowded; closed; not use; confusing Pool area: not large; crowded; bad</p>	2020	<p>Pool: unable to use; crowded; bad; closed; dislike; not use; too small; wrong Closed: spa, gym</p>

<p>“SERVICES” + “POSITIVE”</p>	2019	<p>Excellent: room service; customer service; activities Room service: fast; good Service: like; fast; good; pleasant; excellent</p>	2020	<p>Excellent: customer care; customer service; room service; housekeeping Room service: good; excellent Animation: excellent; improved; good Entertainment: good Service: like; excellent; good</p>
<p>“SERVICES” + “NEGATIVE”</p>	2019	<p>Service: poor; bad; slow Bad: entertainment; room service; customer service Lack: entertainment</p>	2020	<p>Service: bad; poor; lack; limited; slow Animation: dislike Waiting: too long</p>
<p>“BUDGET” + “POSITIVE”</p>	2019	<p>Price: high-quality; reasonable; great value; fair; excellent; good Good: cost; value for money; money; value; deal; quality price relation Excellent: budget; value; deal; value for money</p>	2020	<p>Price: right; worth; great value; excellent; good Good: price/quality ratio; quality price relation; deal; value for money; value; money; performance ratio Excellent: value for money; deal; value; price</p>
<p>“BUDGET” + “NEGATIVE”</p>	2019	<p>Not worth: charge; money Value for money: poor Compensation: not convenient Price: bad; dislike</p>	2020	<p>Value for money: bad; poor</p>
<p>“CLEANING” + “POSITIVE”</p>	2019	<p>Cleaning: good; excellent Cleanliness: like; good; excellent</p>	2020	<p>Excellent: hygiene; cleaning; cleanliness Good: hygiene; cleaning; cleanliness</p>
<p>“CLEANING” + “NEGATIVE”</p>	2019	<p>Cleaning: doesn't meet expectation; bad; poor; dislike Cleanliness: poor; doesn't meet expectation</p>	2020	<p>Cleaning: poor; limited; bad; doesn't meet expectation Cleanliness: lack</p>

<p>“PERSONNEL” + “POSITIVE ATTITUDE”</p>	2019	<p>Friendly: hotel staff; staff; front desk; treatment of staff; reception; workers Staff: courteous; kind; attentive; willing to help; cheerful; welcoming; responsive; accommodating</p>	2020	<p>Friendly: cleaning staff; reception; hotel staff; front desk; treatment of staff; workers Staff: attentive; kind; courteous; welcoming</p>
<p>“PERSONNEL” + “NEGATIVE ATTITUDE”</p>	2019	<p>Staff: not welcoming; not attentive; not friendly Rude: front desk; manager; staff</p>	2020	<p>Staff: not attentive; not friendly; rude</p>
<p>“ROOM AMENITIES” + “POSITIVE FEELING”</p>	2019	<p>Bed: beautiful; luxurious; clean; comfortable Soft: linen; mattress Quiet: air conditioning Pillows: comfortable</p>	2020	<p>Bed: comfortable; clean Mattress: comfortable; soft Furniture: beautiful Clean: towel; shower</p>
<p>“ROOM AMENITIES” + “NEGATIVE FEELING”</p>	2019	<p>Dirty: toilet; carpet; sheet; shower; towel; bath towels Not comfortable: pillows; sofa bed; bed; mattress Noisy: air conditioning; refrigerator</p>	2020	<p>Air conditioning: not refreshed; noisy; loud Dirty: curtains; sheet; furniture; pillows; towel Not comfortable: sofa bed; bed; mattress; pillows</p>
<p>“ROOM” + “POSITIVE FEELING”</p>	2019	<p>Room: quiet; well-kept; stunning; well-decorated; well-appointed; well-equipped; clean; beautiful; modern; newly renovated</p>	2020	<p>Room: modern; beautiful; quiet; well-equipped; stylish; clean; comfortable</p>
<p>“ROOM” + “NEGATIVE FEELING”</p>	2019	<p>Room: noisy; disappointed; needs remodeling; dirty; smelly; cockroach; smelt; older; not comfortable Bathroom: needs remodeling; dirty; not comfortable</p>	2020	<p>Room: needs remodeling; noisy; not comfortable; smelly; dirty; dusty Bathroom: dusty; dirty Door: slamming</p>

“COVID-19” + “POSITIVE”	NA	NA	2020	Covid-19: well applied; good; understandable; safe; excellent Social distancing: good
“COVID-19” + “NEGATIVE”	NA	NA	2020	Covid-19: limited; not control; problem

Fonte: Elaboração Própria, através do SPSS Modeler