INSTITUTO
UNIVERSITÁRIO
DE LISBOA

The Geography of Taste: Analytical Insights for Sabores do Ti Pereira's Expansion

Bruna Filipa Mendes Pereira

Master in Business Analytics

Supervisor:
Professor Nuno Duarte Fialho Sanches Borges dos Santos, Guest Assistant Professor, Iscte Business School

October, 2025

BUSINESS
SCHOOL

Department of Quantitative Methods for Management and Economics

The Geography of Taste: Analytical Insights for Sabores do Ti Pereira's Expansion

Bruna Filipa Mendes Pereira

Master in Business Analytics

Supervisor:
Professor Nuno Duarte Fialho Sanches Borges dos Santos, Guest Assistant Professor, Iscte Business School

October, 2025

Acknowledgments

To my parents who, without ever questioning my choices, gave me the freedom to dream and the courage to pursue my own path.

To my brothers and sisters-in-law, for their constant encouragement and for reminding me that determination and perseverance can turn any goal into reality.

To my nephews, whose laughter and energy brought lightness and joy to the longest days.

To my friends, who stood by me through every stage of this journey, listening to endless conversations about my thesis, offering words of reassurance, and believing in me even when I doubted myself.

To my supervisor, Professor Nuno, whose guidance, patience, and optimism made this challenge not only possible but deeply rewarding. Thank you for your trust and for showing me that every obstacle can be turned into motivation.

Finally, to everyone who shared this path with me. Thank you.

Abstract

This study analyses the territorial and socioeconomic characteristics that define potential locations for the expansion of *Sabores do Ti Pereira, Lda*, a Portuguese company operating in the bakery and pastry sector. The company, headquartered in Abrantes, currently manages seventeen stores across seven municipalities in central Portugal. To support its continued growth, the research aimed to identify statistically suitable areas for new store openings.

The methodological framework combined quantitative and spatial approaches, integrating both company and external data. Internal information was provided by the company, while external data were obtained from the Instituto Nacional de Estatística (INE) and OpenStreetMap (OSM), covering demographic, socioeconomic, and infrastructural indicators. The datasets were processed and analysed using Principal Component Analysis (PCA) to reduce dimensionality and identify territorial patterns, followed by Cluster Analyses to group statistical sections with similar characteristics.

The results revealed distinct territorial profiles across Portugal and led to the identification of a top five civil parishes that represent the most promising areas for future expansion. These findings provide practical insights into data-driven retail location strategies and contribute to the understanding of spatial dynamics and small and medium-sized enterprises (SMEs).

Keywords: Expansion, Retail Industry, Bakery, Pastry, Location

JEL Classification: M11, R11

Resumo

Este estudo analisa as características territoriais e socioeconómicas que definem potenciais localizações para a expansão da *Sabores do Ti Pereira, Lda*, uma empresa portuguesa que atua no setor da pastelaria e panificação. A empresa, sediada em Abrantes, gere atualmente dezassete lojas distribuídas por sete municípios da região centro de Portugal. Com o objetivo de apoiar o seu crescimento contínuo, esta investigação procurou identificar áreas estatísticas e espacialmente adequadas para a abertura de novas lojas.

O enquadramento metodológico combinou abordagens quantitativas e espaciais, integrando dados internos e externos. A informação interna foi fornecida pela empresa, enquanto os dados externos foram obtidos a partir do Instituto Nacional de Estatística (INE) e do OpenStreetMap (OSM), abrangendo indicadores demográficos, socioeconómicas e infraestruturais. Os conjuntos de dados foram processados e analisados através da Análise de Componentes Principais (ACP), de forma a reduzir a dimensionalidade e identificar padrões territoriais, seguida de Análise de Clusters para agrupar secções estatísticas com características semelhantes.

Os resultados revelaram perfis territoriais distintos em Portugal e permitiram identificar um top 5 de freguesias que representam as áreas mais promissoras para futura expansão da empresa. Estas conclusões fornecem insights práticos para estratégias de localização comercial baseadas em dados e contribuem para a compreensão das dinâmicas espaciais no contexto das Pequenas e Médias Empresas (PME).

Palavras-Chave: Expansão, Indústria do Retalho, Pastelaria, Padaria, Localização

Classificação JEL: M11, R11

List of Acronyms and Abbreviations

API: Affinity-based Popularity Inference

BFS: Bundesamt für Statistik (Swiss Federal Statistical Office)

BPS: British Population Survey

CANSIM: Canadian Socio-Economic Information Management System

CZSO: Czech Statistical Office

DC-CNN: Double-Channel Convolutional Neural Network

e.g.: Exempli gratia (for exemple)

EA: Environics Analytics

EPOS: Electronic Point of Sale

ERS: Economic Research Service

et al: Et alii (and others)

EUROSTAT: Statistical Office of the European Union

FMCG: Fast-Moving Consumer Goods

GIS: Geographic Information System

GWR: Geographically Weighted Regression

HESA: Higher Education Statistics Agency

IAUIDF: Institut d'Aménagement et d'Urbanisme d'Île-de-France (Ile de France Region Planning Agency)

INE: Instituto Nacional de Estatística

INSEE: French National Institute of Statistics

KNN: K-Nearest Neighbours

LBS: Location-Based Service

Lda: Limitada

LISA: Local Indicators of Spatial Association

LOJIC: Lousville and Jefferson County Information Consortium

LZs: Logistics Zones

MNL: Multinomial Logit

NAVER: The largest portal search in Korea

NETS: National Establishment Time-Series

OECD: Organisation for Economic Co-operation and Development

OLS: Ordinary Least Squares

OMB: Office of Management and Budget

ONS: Office for National Statistics

OSM: OpenStreetMap

PME: Pequenas e Médias Empresas

POIs: Points of Interest

SHS: Survey of Household Spending

SIM: Spatial Interaction Model

SMEs: Small and Medium Enterprises

SNS: Social Network Services

UK: United Kingdom

US: United States

USDA: U.S. Department of Agriculture

Vs: Versus

Index

Acknowledgments ..... **i**

Abstract ..... **iii**

Resumo ..... **v**

List of Acronyms and Abbreviations ..... **vi**

1 Introduction **1**

1.1 Theme and its Importance 1

1.2 Research Problem and Research Questions 2

1.3 Objectives and contributions 3

1.4 Methodological approach 3

1.5 Dissertation structure 4

***2* Systematic Literature Review** **5**

2.1 Systematic Literatura Review Protocol 5

2.1.1 Systematic Analysis 10

2.2 Quality Assessment & Discussion 11

2.2.1 Scope of the Studies 15

2.2.2 Methods Used for Store Location 21

2.2.3 Location Factors in Retail Site Location 27

2.2.4 Study Evaluation 28

2.3 Summary and Implications of Literature 35

3 Methodology **37**

3.1 Methodological framework 37

3.1.1 Case study: Company overview 38

3.2 Data Sources and Variables 40

3.2.1 Company Data 40

3.2.2 External Data 40

3.2.2.1 Sources of External Data 40

3.2.2.2 Variables Groups 41

3.2.3 Section-Level Context Dataset 42

3.3 Data Dictionary 43

3.4 Data Preparation and Processing 43

3.5 Data Analysis Framework 45

4 Results and Discussion **47**

4.1 Descriptive Statistics 47

4.1.1 Company Data 47

4.1.2 External Data 51

4.2 Principal Component Analysis (PCA) 53

4.3 Cluster Analysis 59

4.4 Selection of Priority Clusters and Identification of Potential Locations 62
5 Conclusion 67
5.1 Research Overview and Purpose 67
5.2 Summary of Key Findings 67
5.3 Contributions 68
5.4 Limitations and Future Research Recommendations 69
6 References 71
7 Appendix 75

Index of Tables

Table 2.1: Inclusion and exclusion criteria 6
Table 2.2: Articles included in Systematic Literature Review 8
Table 2.3: Quality criteria for article evaluation 10
Table 2.4: Assessment of Article Quality 13
Table 2.5: Scope of the studies 16
Table 2.6: Techniques and Variables Used 23
Table 2.7: Characterisation of Data, Sources and Models 24
Table 2.8: Contributions, Future recommendations, Limitations and Suggestions 30
Table 4.1: Analysis of Revenue and Issued Documents 48
Table 4.2: Descriptive Analysis of Revenue 48
Table 4.3: Characterisation of Portuguese Territory 53
Table 4.4: Analysis of Civil Parishes Selected 63
Table 4.5: Analysis of Civil Parishes where Company Operates 64
Table 7.1: Store Level Variables from Company, INE, and OpenStreetMap Data 75
Table 7.2: Statistical Section Level Variables from INE and OpenStreetMap Data 83
Table 7.3: Derived Analytical Variables 89
Table 7.4: Total Variance Explained 90
Table 7.5: Rotated Component Matrix 91
Table 7.6: Evaluation of Clustering Solutions Using TSS, WSS and BSS 91
Table 7.7: Cluster Centres of Territorial Profile Analysis 92
Table 7.8: Cluster Centres of Proximity Analysis 92
Table 7.9: Selection of Statistical Sections 93
Table 7.10: Selected Statistical Sections 93

Index of Figures

Figure 2.1: Diagram of article selection process 7
Figure 3.1: Diagram of Methodological Process 37
Figure 3.2: Map of Portuguese Territory 39
Figure 3.3: Map of Store Locations 39
Figure 3.4: Map of Store Locations in the Abrantes Region 39
Figure 4.1: Issued Documents vs Revenue 2023 (All Stores) 50
Figure 4.2: Issued Documents vs Revenue 2023 50
Figure 4.3: Issued Documents vs Revenue 2024 51
Figure 4.4: Quintile Map of Urban Density and Infrastructure Dimension 55
Figure 4.5: Quintile Map of Socioeconomic and Demographic Vitality Dimension 56
Figure 4.6: Quintile Map of Commercial and Service Concentration Dimension 57
Figure 4.7: Map of Territory Profile Cluster Analysis 60
Figure 4.8: Map of Proximity Cluster Analysis 61
Figure 4.9: Geographical Distribution of the Civil Parishes Selected 64

1 Introduction

The present study is an exploratory study that aims to identify and explore potential expansion opportunities and new locations for a Portuguese local company (*Sabores do Ti Pereira, Lda*) from the bakery and pastry sector.

This section explores the significance of the chosen theme and its importance to the companies around the world. Furthermore, it also introduces the research problem, outlines the objectives and potential contributions, explains the methodological approach and presents the structure of the dissertation.

1.1 Theme and its Importance

Nowadays, with the ongoing expansion of e-commerce, retailers face a challenging environment (Formánek & Sokol, 2022). Despite the rapid growth of technology and digital communication on the retail sector, physical stores remain essential, especially to the category of Fast-Moving Consumer Goods (FMCG), where traditional sales channels are essential (Ouyang et al., 2020; Formánek & Sokol, 2022).

The company *Sabores do Ti Pereira, Lda* used to plan the opening of a new store without conducting any formal study. Accordingly, the purpose of this research is to support the selection of new store locations by analysing the performance of existing stores and the characteristics of statistical sections in Portugal.

Store location is a determinant key of success, as it can impact consumer service quality, pricing strategies, and product variety (Suárez-Vega et al., 2015; F. Wang et al., 2014). However, medium-sized chains normally rely on the judgment of senior executives rather than data-driven analysis and small retailers use to choose store location based on personal preferences (Formánek & Sokol, 2022). As it involves significant financial costs and operational risks, planning without any research can be risky (Formánek & Sokol, 2022).

As a result, the effective location analysis should be based on sales data performance insights combined with geographical and sociodemographic factors to ensure informed decision-making (Formánek & Sokol, 2022).

In some cases, customers from more distant locations prefer to compare multiple products during a single visit (Oppewal & Holyoake, 2004), as it improves the attractiveness of the place (Teller & Thomson, 2012; Teller et al., 2016). Under these circumstances, agglomerations of stores can be beneficial as they attract this type of customers (Oppewal & Holyoake, 2004) and retailers can take advantage from locating stores next to other retailers (Teller et al., 2016). However, retailers might aim to reduce competition by strategically positioning stores far away from competitors or opt to adjust their marketing strategies (Thomadsen, 2007).

The location of commercial stores has been guided by a demand-driven location theory, which focused on the distribution and accessibility of consumers within the physical space (Hotelling, 1929; Lösch, 1954). Consumers used to choose easily accessible stores, which are the ones along main streets and close to major public transportation network (Lee et al., 2020).

Therefore, studying retail site selection in the bakery and pastry sector represents an important contribution to society, especially for retailers, as choosing the right location can result in high profitability and better market performance (Ouyang et al., 2020). Nonetheless, the topic is also crucial for the academics as the sector of pastry and bakery remains underexplored in the literature. This study may provide a valuable basis for future investigations.

1.2 Research Problem and Research Questions

The current research pretends to answer the following research problem: *Sabores do Ti Pereira, Lda* faces the challenge of determining viable strategies for geographic and market expansion in order to sustain its competitive growth. However, the factors influencing optimal location selection and expansion strategy are not yet clearly defined.

To address the research problem, this study should answer the following research questions:

(i.) What are the key factors influencing location decisions for business expansion?

(ii.) Which analytical models can best evaluate and compare potential expansion locations based on these factors?

(iii.) How can these insights be integrated into coherent expansion strategy for *Sabores do Ti Pereira, Lda*?

Normally, location decisions are based on sites that are easily accessible to consumers, ensuring reduced commuting time (Shim & Eastlick, 1998; Yan & Eckman, 2009). As the

competitiveness intensified, the store location acquired even more relevance (Formánek & Sokol, 2022).

To select the store site, retailers must plan and consider existing stores, the future stores planned to open in the area, and the improvement of performance and profitability of other retail locations (Ghosh & Craig, 1983). The purpose of this research is to analyse potential future locations based on the characteristics of statistical sections in Portugal. Using this information, spatial analysis models will be developed to address the research questions above.

1.3 Objectives and contributions

Considering the research question previously mentioned, the primary objective of this research is to identify, through spatial analysis models, the different expansion alternatives available to the company.

Consequently, to fulfil the overall objective, three specific objectives were defined: (1) Identify the specific endogenous and exogenous factors influencing store performance; (2) Segment the territory based on the critical performance factors; and (3) Identify priority territorial segments for expansion.

By addressing the objectives mentioned before and answering the research questions, this research will contribute to help the company finding the optimal location for its store.

1.4 Methodological approach

This study adopts a quantitative methodological approach designed to identify the territorial and socioeconomic characteristics that define potential locations for the expansion of *Sabores do Ti Pereira, Lda*. The research follows a multivariate analytical design, integrating company and external data to ensure a comprehensive understanding of both business performance and territorial context. Data from the company were combined with open-access sources, namely the Instituto Nacional de Estatística (INE) and OpenStreetMap (OSM). A descriptive statistical analysis was first performed to characterise the datasets, followed by a Principal Component Analysis (PCA) to reduce dimensionality and identify the most relevant territorial dimensions. Finally, two Cluster Analyses were conducted to classify statistical sections according to their socioeconomic, and spatial characteristics, supporting the identification of suitable areas for future store expansion.

1.5 Dissertation structure

To assist in the different strategic expansion alternatives available to the company, this research has been structured into five chapters, each addressing a specific aspect of the study.

The first chapter introduces the theme and its relevance, the research problem, the guiding research question, the primary and specific objectives and the study's contribution for society. Furthermore, the chapter presents the methodological approach, detailing the methods employed to address the research question and achieve the objectives.

The second chapter is dedicated to a systematic literature review. This review examines the critical role of location in business success and explores the application of various spatial analysis models in previous research.

The third chapter provides a detailed description of the research methodology, encompassing all aspects of data collection and analysis, while the fourth chapter presents the research findings and a comprehensive discussion of their implications.

Finally, the concluding chapter presents the research findings, discusses the study's limitations, and highlights its potential contributions to academical and business community.

2 Systematic Literature Review

Throughout the years, managers adopted a lot of methods to choose the location for a new retail store. The methods turned more sophisticated and nowadays it is possible to develop models that improves decision-making for site planning by incorporating temporal variation in customer flows (T. Waddington et al., 2019) and help managers predict when and where demand will be highest, using venue type, time patterns, and location data (Hsieh et al., 2019).

According to Formánek & Sokol (2022) and Horton (1968), the location is a key variable in understanding store success. As a result, this chapter aims to conduct a comprehensive analysis of existing literature on retail site selection. A significant number of articles reviewed build their methodologies upon prior studies conducted in similar contexts. This practice highlights the importance of understanding previous research before proposing new approaches. Furthermore, many of them are centred on evaluating how spatial and locational characteristics influence performance. Accordingly, the following questions were developed to explore how different strategic alternatives for a new retail location can be studied.

Q.1: What are the most important location factors?

Q.2: Which models estimate alternative locations based on location factors?

Q.3: How can business expansion strategy be defined?

2.1 Systematic Literatura Review Protocol

The present study aims to find the answer to the problem:

(i.) "What are the different expansion alternatives available for the next store of the company *Sabores do Ti Pereira, Lda*?".

In addition, the research must also identify the sectors in which location studies have already been conducted, the key factors influencing site selection, and the models used to assist companies in determining the alternative locations.

Firstly, to conduct the present research, the Scopus database was used to identify relevant documents containing valuable information. To ensure a comprehensive selection of pertinent studies based on title, abstract, and keywords, a query was formulated to facilitate the retrieval of essential articles for this research.

- ("new store*" OR "new location*" OR "location" OR "expansion*" OR "new branch" OR "business expansion" OR "site selection") AND ("predict" OR "forecast" OR "market potential" OR "location factor*") AND ("space" OR "spatial" OR "spatio*" OR "territory" OR "geography" OR "map") AND ("bakery" OR "pastry" OR "coffee shop" OR "restaurant" OR "convenience store" OR "grocery store" OR "retail" OR "fmcg" OR "beverages")

From the previous query, 149 documents were obtained. After collecting the articles, each abstract was read and analysed to determine their relevance to the study. As a result, 26 articles were selected based on inclusion and exclusion criteria, as presented in table **2.1** and explained in the Figure **2.1.**

Table 2.1: Inclusion and exclusion criteria

Inclusion criteria	Articles that discuss business expansion models
	Articles that analyse business location factors
	Articles that use models to determine a new location
	Articles written in English
	Articles focused on the fields of Social Sciences, Business, Management and Accounting, Computer Science, Economics, Econometrics and Finance, Decision Sciences, Multidisciplinary and Mathematics
Exclusion criteria	Articles that do not provide any contribution to the study
	Not available for download

Source: Own elaboration

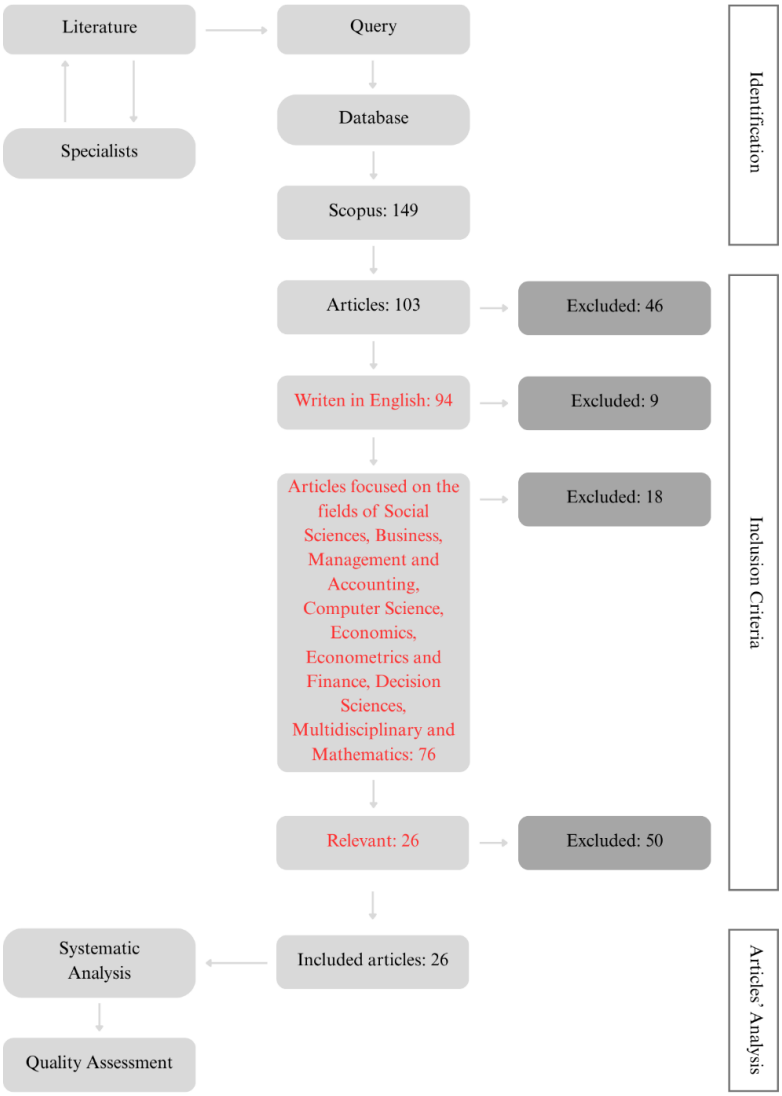


Figure 2.1: Diagram of article selection process

Table 2.2 presents the selected articles, ordered from the most recent to the oldest. Upon analysing the table, it can be concluded that the majority of journals have published only one article, with exceptions of *Applied Geography*, *Geographical Analysis*, *ISPRS International Journal of Geo-Information*, *Journal of Retailing and Consumer Services* and *Journal of Transport Geography*, each of which has published two. It is also important to note that a considerable number of journals are from the field of Geography (n=7) and the field of Management and Marketing (n=5).

Table 2.2: Articles included in Systematic Literature Review

ID	Authors	Year	Title	Source title
1	Hunneman A.; Bijmolt T.H.A.; Elhorst J.P.	2023	Evaluating store location and department composition based on spatial heterogeneity in sales potential	Journal of Retailing and Consumer Services
2	Formánek T.; Sokol O.	2022	Location effects: Geo-spatial and sociodemographic determinants of sales dynamics in brick-and-mortar retail stores	Journal of Retailing and Consumer Services
3	Lee J.-H.; Jo J.; Kim J.W.; Lee K.; Choi M.Y.	2022	Spatial distributions of restaurants emerging from pedestrian behavior and online information sharing	Physica A: Statistical Mechanics and its
4	Beckers J.; Birkin M.; Clarke G.; Hood N.; Newing A.; Urquhart R.	2022	Incorporating E-commerce into Retail Location Models	Geographical Analysis
5	Low S.A.; Bass M.; Thilmany D.; Castillo M.	2021	Local Foods Go Downstream: Exploring the Spatial Factors Driving U.S. Food Manufacturing	Applied Economic Perspectives and Policy
6	Ouyang J.; Fan H.; Wang L.; Yang M.; Ma Y.	2020	Site selection improvement of retailers based on spatial competition strategy and a double-channel convolutional neural network	ISPRS International Journal of Geo-
7	Khare A.	2020	Location and agglomeration factors predicting retailers' preference for Indian malls	Journal of Marketing Analytics
8	Sakai T.; Beziat A.; Heitz A.	2020	Location factors for logistics facilities: Location choice modeling considering activity categories	Journal of Transport Geography
9	Guidon S.; Reck D.J.; Axhausen K.	2020	Expanding a(n) (electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests	Journal of Transport Geography
10	Wojtyra B.; Kossowski T.M.; Březinová M.; Savov R.; Lančarič D.	2020	Geography of craft breweries in Central Europe: Location factors and the spatial dependence effect	Applied Geography
11	Hsieh H.-P.; Lin F.; Li C.-T.; Yen I.E.-H.; Chen H.-Y.	2019	Temporal popularity prediction of locations for geographical placement of retail stores	Knowledge and Information Systems
12	Waddington T.; Clarke G.; Clarke M.C.; Hood N.; Newing A.	2019	Accounting for Temporal Demand Variations in Retail Location Models	Geographical Analysis
13	Glaeser C.K.; Fisher M.; Su X.	2019	Optimal retail location: Empirical methodology and application to practice	Manufacturing and Service Operations

ID	Authors	Year	Title	Source title
14	Newing A.; Clarke G.; Clarke M.	2018	Applied spatial modelling for retail planning in tourist resorts	International Journal of Retail and Distribution
15	Robinson D.T.; Balulescu A.M.	2018	Comparison of methods for quantifying consumer spending on retail using publicly available data	International Journal of Geographical
16	Wang L.; Fan H.; Wang Y.	2018	Site selection of retail shops based on spatial accessibility and hybrid BP neural network	ISPRS International Journal of Geo-
17	Waddington T.B.P.; Clarke G.P.; Clarke M.; Newing A.	2018	Open all hours: spatiotemporal fluctuations in U.K. grocery store sales and catchment area demand	International Review of Retail, Distribution and
18	Cró S.; Martins A.M.	2018	Hotel and hostel location in Lisbon: looking for their determinants	Tourism Geographies
19	D'Silva K.; Noulas A.; Musolesi M.; Mascolo C.; Sklar M.	2018	Predicting the temporal activity patterns of new venues	EPJ Data Science
20	Dock J.P.; Song W.; Lu J.	2015	Evaluation of dine-in restaurant location and competitiveness: Applications of gravity modeling in Jefferson County, Kentucky	Applied Geography
21	Finazzi F.	2013	Geostatistical modeling in the presence of interaction between the measuring instruments, with an application to the estimation of spatial market potentials	Annals of Applied Statistics
22	Schuetz J.; Kolko J.; Meltzer R.	2012	Are poor neighborhoods retail deserts?	Regional Science and Urban Economics
23	Zhu T.; Singh V.	2009	Spatial competition with endogenous location choices: An application to discount retailing	Quantitative Marketing and Economics
24	Mendes A.B.; Themido I.H.	2004	Multi-outlet retail site location assessment	International Transactions in
25	Pillsbury R.	1987	From hamburger alley to hedgerose heights: Toward a model of restaurant location dynamics	Professional Geographer
26	Horton F.K.	1968	Location factors as determinants of CONSUMER ATTRACTION TO RETAIL FIRMS	Annals of the Association of American

Source: Own elaboration

Following a comprehensive analysis of selected articles, it is crucial to assess their quality and relevance to the study. This evaluation will identify the most valuable articles, those that make the most significant contribution to achieving the study's objectives. To guide this evaluation, several quality-items criteria were established based on three key questions: (i) What are the scope of the study? (ii) What methods are used for store location? (iii) How is the study evaluation conducted?

For the evaluation process, a scoring system was implemented. If an article provides a comprehensive answer to a question, it receives a score of 1. If the response is partial, the article is given a score of 0.5. If the article fails to address the question, it receives a score of 0.

2.1.1 Systematic Analysis

To answer the research question "What are the different expansion alternatives available for the next store of the company *Sabores do Ti Pereira, Lda*?", it is necessary to make a critical review analysis to understand what have been studied in the existing literature and understand the gaps that still exists. Therefore, the quality criteria mentioned in Table 2.3 were created.

Table 2.3: Quality criteria for article evaluation

Scope of the studies	**Q1.1:** Are the objectives of the study clearly stated?
	Q1.2: Is the industry sector to which it applies mentioned?
	Q1.3: Is the sector under study the pastry and bakery industry?
	Q1.4: Is the sector under study the food retail sector or the food service (restaurant) sector?
	Q1.5: Is the study focused on small and medium-sized enterprises (SMEs)?
	Q1.6: Is the study focused on a rural area?
	Q1.7: Is the country under study identified?
	Q1.8: Does the study clearly define the store location factors?
	Q1.9: Are the period of study mentioned?
Methods Used for Store Location	**Q2.1:** Are the techniques/models described and justified?
	Q2.2: Are the variables clearly identified?
	Q2.3: Is the type of data clearly specified?
	Q2.4: Are the data sources clearly identified?
	Q2.5: Does the model take spatial factors into account?
	Q2.6: Does the model predict the performance of new locations?
	Q2.7: Does the model incorporate exogenous factors?
	Q2.8: Is the forecasting horizon clearly defined (e.g., short-term, medium to long-term)?
Study Evaluation	**Q3.1:** Does the study provide theoretical and practical contributions to decision-making in the opening of new stores?

Q3.2: Is the economic and financial impact of adopting certain models clearly presented?
Q3.3: Are suggestions for the future research indicated?
Q3.4: Were the models implemented or tested?
Q3.5: Are the limitations of the models explicitly discussed?
Q3.6: Does the study provide suggestions for potential store locations or expansion strategies?

Source: Own elaboration

To develop this analysis, the study focuses on three components: Scope of the studies, Methods Used for Store Location and Study Evaluation.

2.2 Quality Assessment & Discussion

Table 2.4 presents the scores of each article based on the quality assessment criteria. As previously mentioned, each article is assigned a score of 0, 0.5 or 1, depending on how effectively it addresses the research questions.

Among the 26 articles selected for the Systematic Literature Review, none reached the maximum possible score of 23 points, although the highest score obtained was 20.5 (T. Waddington et al., 2019). The criteria with the highest score were “Q1.1: Are the objectives of the study clearly stated?”, ”Q1.2: Is the industry sector to which it applies mentioned?” and “Q1.7: Is the country under study identified?”. These findings indicate that nearly all articles clearly state their research objectives, specify the industry sector under investigation and identify geographical scope.

The best-performing analytical dimension relates to the methods used for selecting a store location, having achieved a score of 23.19 out of 26. This indicates that most of the articles provide a clear description and justification of the models and techniques used, as well as specify the variables, type of data and data sources used. It also suggests that several articles take spatial factors into account, predict the performance of new locations and incorporate exogenous factors. The forecasting horizon is also specified in almost every article.

Conversely, the dimension related to the overall scope of the studies scored the lowest. Although this dimension includes criteria with the highest individual quality ratings, it also records the lowest scores overall. Particularly, criteria Q1.3 (“Is the sector under study the pastry and bakery industry?”) received the lowest score of 0, which suggests an underexplored research niche that may be studied in a future investigation.

Concluding, for the dimension regarding the scope of the studies, the articles by Lee et al. (2022) and Robinson & Balulescu (2018) stand out as the most relevant. In contrast, for both the methodological dimension related to store location modelling and the dimension assessing practical contributions and implementations, several articles should be considered.

Table 2.4: Assessment of Article Quality

ID	Scope of the studies									Methods Used for Store Location									Study Evaluation					Total
	Q1.1	Q1.2	Q1.3	Q1.4	Q1.5	Q1.6	Q1.7	Q1.8	Q1.9	Q2.1	Q2.2	Q2.3	Q2.4	Q2.5	Q2.6	Q2.7	Q2.8	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6	
1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
2	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0.5	1	1	1	1	19
3	1	1	0	1	0	0	1	1	1	1	1	1	1	1	0.5	1	1	1	0.5	1	1	1	1	20
4	1	1	0	1	0	0	1	1	0.5	1	1	1	1	0.5	0	1	1	1	0.5	1	0.5	1	1	17
5	1	1	0	0.5	1	0.5	1	0.5	1	1	1	1	1	1	0	1	1	1	0	1	0.5	1	0.5	17.5
6	1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0.5	1	18.5
7	1	1	0	0	0	0	1	1	0	0.5	0.5	0.5	0.5	1	0	0.5	0	0.5	0	0.5	0	1	0.5	10
8	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0.5	1	1	1	1	17.5
9	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0.5	1	1	1	1	19
10	1	1	0	0	1	0.5	1	1	1	1	1	1	1	1	0	1	1	0.5	0.5	0.5	0	0.5	0.5	16
11	1	1	0	0.5	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0.5	1	1	1	1	19
12	1	1	0	1	0.5	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20.5
13	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
14	1	1	0	1	0	0	1	1	1	1	1	1	1	1	0.5	1	1	1	0.5	1	1	1	1	19.5
15	1	1	0	1	0	0	1	1	1	1	1	1	1	0.5	0	1	1	1	0.5	1	1	1	0.5	18.5
16	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
17	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
18	1	1	0	0.5	0	0	1	1	0.5	0.5	0.5	0.5	1	1	0	1	1	1	0.5	1	0	0.5	1	14.5
19	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0.5	1	1	1	1	18.5
20	1	1	0	1	0	0	1	1	1	1	1	1	0.5	1	1	1	1	1	1	1	1	1	1	19.5
21	1	1	0	0	0	0	1	0.5	0	1	1	1	1	1	1	1	1	1	0.5	1	1	1	1	17

ID	Scope of the studies									Methods Used for Store Location									Study Evaluation						Total
	Q1.1	Q1.2	Q1.3	Q1.4	Q1.5	Q1.6	Q1.7	Q1.8	Q1.9	Q2.1	Q2.2	Q2.3	Q2.4	Q2.5	Q2.6	Q2.7	Q2.8	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q3.6		
22	1	1	0	0	0.5	0	1	1	1	1	1	1	1	1	0	1	1	1	0.5	1	1	1	1	18	
23	1	1	0	0	0	0	1	1	1	1	1	1	0.5	1	1	1	1	1	1	1	1	1	1	18.5	
24	1	1	0	1	0.5	0	1	1	0	1	1	1	0.5	1	1	1	1	1	1	1	1	1	1	19.5	
25	1	1	0	1	0.5	0	1	0.5	1	0.5	0.5	0.5	0.5	1	0	0.5	0	0.5	0	1	0	0	1	12.5	
26	1	1	0	0	0	0	1	1	1	0.5	0.5	0.5	0.5	1	1	0.5	1	0.5	0	1	0	0.5	1	13.5	
Total	26	26	0	11.5	6	1	26	24.5	19	24	24	24	23	25	17	24.5	24	24	15.5	25	20	23	24		

Source: Own elaboration

2.2.1 Scope of the Studies

According to table 2.5, all the reviewed articles share a common objective: all seek to analyse the local factors of a specific sector and support managerial decision-making regarding optimal store placement.

By analysing the results presented in the table, it is possible to state that articles address different sectors. Many of the articles study the retail sector or the foodservice sector. However, while some studies do not belong to the retail sector at all (Lee et al., 2022; Low et al., 2021; Sakai et al., 2020;Guidon et al., 2020; Wojtyra et al., 2020; Cró & Martins, 2018; Dock et al., 2015;Pillsbury et al., 1987), others are part of the retail domain but focus on areas that are not comparable to the bakery and pastry sector (e.g., Hunneman et al., 2023). Nevertheless, the articles remain essential to understand the methods they employ to analyse the location factors.

Only four articles explicitly focus on Small and Medium-sized Enterprises (SMEs) (Low et al., 2021; Ouyang et al., 2020; Wojtyra et al., 2020; Wang et al., 2018), while another four include SMEs only partially, as they are not the main focus of the analysis (Waddington et al., 2019; Schuetz et al., 2012; Mendes & Themido, 2004; Pillsbury, 1987).

Most of the articles focus on urban areas. Only two studies explicitly examine both urban and rural contexts (Low et al., 2021; Wojtyra et al., 2020), while two others are primarily urban in scope but refer to or include rural characteristics in their analysis (Hunneman et al., 2023; Formánek & Sokol, 2022).

Geographically, most of the studies (n=12) are in Europe, followed by those from America (n=9), and only four studies focusing on Asia countries.

As shown in the following table, the analysis of the location factors reveals that the most frequently cited elements include accessibility (n = 10), competition (n = 9), public transportation (n = 8), demographic characteristics (n = 8), socioeconomic characteristics (n = 7), and travel distance (n = 7).

Finally, while many studies do not specify the period of analysis, the most recent data used in the studies extends up to the year of 2020. These findings highlight the importance of conducting more studies to draw timely and relevant conclusions that can effectively support business growth.

Table 2.5: Scope of the studies

ID	Study Objectives	Sector	Focus on SMEs (Yes/Partially/No)	Urban or Rural Area	Country	Store Location Factors	Period of Study
1	Examine and evaluate the importance of store location as well as the size of its sections.	Clothing Retail	No	Urban but it can have some rural areas	The Netherlands	Competition; Customer travel costs; Regional customer preferences; Marketing Mix	ND
2	Analyse the influence of store location and size on retail performance.	Brick-and-mortar Retail	No	Urban but it mentions some rural areas	Czech Republic	Accessibility; Accessibility to transportation; Presence of competitors; Attractiveness; Distance between customers and stores; Store agglomeration	2006–2015
3	Analyse restaurant distribution based on customer behaviour, online review, and prices and the impact of online reviews.	Restaurant Industry	No	Urban	South Korea	Distance to public transportation	2010-2012
4	Examine the impact of location on e-commerce in the retail industry, considering delivery logistics, technological innovation, and demand drivers.	Grocery sector, focused on e-commerce	No	Urban	United Kingdom	Accessibility; Socioeconomic characteristics; Travel time-based distance	ND
5	Examine the spatial factors influencing food manufacturing start-ups in the United States of America.	Food and Beverage Manufacturing Industry	Yes	Both	United States of America	Infrastructure; Agglomeration; Product and input markets; Labor market; Entrepreneurial activity; Financial capital availability; Strength of the local food sector; Direct and organic food sales; Agricultural production diversity; Agricultural land area and value	2004–2015

ID	Study Objectives	Sector	Focus on SMEs (Yes/Partially/No)	Urban or Rural Area	Country	Store Location Factors	Period of Study
6	Improve the location selection for retailers with high market demand potential.	Retail Sector (small supermarkets and convenience stores)	Yes	Urban	China	Migration cost and location stability; Sales potential and competition; Customer benefits and proximity; Store area; Driving time	ND
7	Analyse the decision-making process regarding the location of malls.	Retail Sector, specifically shopping malls	No	Urban	India	Catchment area; Transport and communication access; Accessibility; Distance between malls; Tenant Mix; Population structure; Retail saturation; Retail competition; Economic conditions; Presence of anchor stores; Transports costs; Product variety; Agglomeration	ND
8	Analyse location choices for the development of logistics facilities.	Logistics	No	Urban	France	Zoning regulations; Accessibility to workplaces; Land price; Accessibility to the resident population and production industries; Accessibilities to potential customers plays a limited role	ND
9	Predict the demand for bicycle-sharing systems before expanding to a new city.	Transportation (Bicycle-sharing)	No	Urban	Switzerland	Distance to POIs; Elevation; Proximity to public transports; Urban Density	2018–2020
10	Explain the spatial distribution of craft breweries.	Craft Breweries	Yes	Both	Czech Republic, Poland and Slovakia	Access to materials; Market proximity and costs; Population density and income; Infrastructures, education and quality of life; Proximity to other breweries;	2017

ID	Study Objectives	Sector	Focus on SMEs (Yes/Partially/No)	Urban or Rural Area	Country	Store Location Factors	Period of Study
						Entrepreneurial environment and presence of millennials	
11	Predict the popularity of locations of the new retail stores using check-in data.	Retail Sector	No	Urban	United States of America	Population; Spatial density; Traffic; Competition; Interaction of people with different categories of locations	2011–2013
12	Improve retail GIS by adding demand layers based on the working population and students to capture more consumer behaviour variations.	Grocery sector	Partially	Urban	United Kingdom	Seasonality of consumption	ND
13	Optimize location choice and schedule decisions using machine learning and econometric methods.	Retail sector	No	Urban	United States of America	Locations of competing physical stores; Salary level of the area; Market size; Competitive strength of the firm	ND
14	Evaluate a spatial model's predictive capacity for seasonal sales and its applications in retail location research and impact assessment.	Grocery sector	No	Urban	United Kingdom	Accessibility; Seasonality of tourist demand; Resident population; Local competition; Proximity to tourist attractions	ND
15	Analyse consumer spending patterns and their relation to publicly available sales data.	Retail sector	No	Urban	Canada	Ideal store size; Timing of entry; Pricing mechanisms; Market share; Feasibility of adding a new store in a region; Consumer spending; Population density; Regional accessibility	2019
16	Develop a two-step model for selecting optimal sites for small shops based on	Retail sector	Yes	Urban	China	Economic, environmental and social factors; Socioeconomic factors;	ND

ID	Study Objectives	Sector	Focus on SMEs (Yes/Partially/No)	Urban or Rural Area	Country	Store Location Factors	Period of Study
	spatial accessibility and market potential.					Competition; Traffic; Travel time; Nearby points of interest; Active consumers	
17	Improve sales forecasting models to optimize supermarket location planning.	Grocery sector	No	Urban	United Kingdom	Consumer behaviour; Store type; Spatiotemporal characteristics of demand; Consumers' time availability; Proximity of residential and commercial areas	2018
18	Explore the spatial factors influencing hotel location strategies.	Hospitality sector	No	Urban	Portugal	Traffic and geographical conditions; Hotel characteristics; Public infrastructure and transport accessibility; Proximity to tourist attractions; Agglomeration benefits; Influence of urban policies, crime and nightlife	ND
19	Help business owners make informed decisions on staffing, supply, and resources through location-based analysis.	Retail and services sector	No	Urban	United Kingdom	Mobility data; Temporal patterns; Urban regions; Locality similarity; Proximity to other similar venues; Neighbourhood characteristics	2011–2013
20	Develop an effective method for site selection using gravitational modelling applications to assess restaurant location and competitiveness.	Restaurant Industry	No	Urban	United States of America	Market competitiveness; Attractiveness	2002
21	Estimate the spatial market potential using sales data, spatial characteristics and store interactions.	Media and Publishing distribution	No	Urban	Italy	Geographic distance; Transportation cost; Transport time; Consumer willingness to travel to a specific location; Store attributes	ND

ID	Study Objectives	Sector	Focus on SMEs (Yes/Partially/No)	Urban or Rural Area	Country	Store Location Factors	Period of Study
22	Analyse the relationship between neighbourhood income and retail presence using various estimation strategies.	Retail sector	Partially	Urban	United States of America	Store density; Fixed costs of stores; Transportation costs	ND
23	Analyse factors influencing business location choices and competitive interactions.	Discount Retail	No	Urban	United States of America	Population proximity and demographics; Household income and education levels; Retail costs and nearby competition; Proximity to company headquarters and distribution centres; Demand conditions, logistics, and consumer base size	2006
24	Review techniques and models used in location and sales evaluation for multiple retail points.	Retail (multi-outlet)	Partially	Urban	Portugal	Accessibility; Positive environment; Socioeconomic factors	ND
25	Develop a model explaining restaurant location dynamics, considering a combination of social, economic, and behavioural factors.	Restaurant Industry	Partially	Urban	United States of America	Accessibility; Traffic density; Demographics; Environment and customer experience	1964-1965 (observational)
26	Understand how location-specific factors, competition, and investment influence consumer attraction to retail establishments.	Retail sector	No	Urban	United States of America	Proximity to traffic generators; Population density; Household income level; Competition location; Contiguity effects	1970s

Source: Own elaboration

2.2.2 Methods Used for Store Location

After analysing the scope of the selected articles, it is crucial to understand methodologies and variables employed to analyse retail site selection and related spatial decision-making processes. This analysis enables the identification of the most appropriate methodological approach and variables for the present research.

As presented in the table 2.6, regression models emerge as the most frequently used technique. Many articles apply linear regression models (e.g., Formánek & Sokol, 2022) while others employ more advanced approaches, such as spatial regression (e.g., Guidon et al., 2020), multinomial logistic regression (e.g., Sakai et al., 2020), or negative binomial regression (e.g., Low et al., 2021).

In addition, spatial modelling is widely adopted. Methods such as Geographically Weighted Regression (GWR) (e.g., Hunneman et al., 2023) and gravity-based models (e.g., Dock et al., 2015) are used to capture spatial dependencies and estimate consumer attraction across geographic space. Some articles combine these methods with exploratory spatial analysis or clustering techniques to better understand geographic patterns (e.g., Newing et al., 2018; Schuetz et al., 2012).

More recent studies also incorporate machine learning methods, including Random Forests, K-Nearest Neighbours (KNN) and Neural Networks, particularly in contexts involving large datasets and predictive models (e.g., Ouyang et al., 2020; Wang et al., 2018). In a different approach, Zhu & Singh (2009) use game theory modelling to simulate competitive behaviour and endogenous location decisions among retail chains.

About variables used, most of the studies employ a combination of spatial, demographic, economic, and behavioural data. Commonly used variables include store performance indicators, population density, income levels, education, employment, distance to public transportation, travel time, and proximity to competitors. Accessibility and visibility also appear frequently.

Some studies also include consumer behaviour data derived from online platforms, such as check-ins, temporal usage patterns, or footfall estimates (e.g., Lee et al., 2022; Hsieh et al., 2019). This reflects a growing trend towards integrating real-time and user-generated data.

According to Table 2.7, which reviews the methodological scope, data sources, spatial considerations, and forecasting orientation of the studies, most rely on spatial or cross-sectional spatial data. However, some incorporate temporal or longitudinal components (e.g., Hsieh et

al., 2019; T. B. P. Waddington et al., 2018; D'Silva et al., 2018), while a few are primarily qualitative or rely on field observations and interviews (e.g., Khare, 2020; Pillsbury, 1987).

With respect to data sources, retailer-internal datasets and official national or regional statistics, such as those from the ONS, EUROSTAT and Census Bureaus, are commonly employed. Furthermore, some studies incorporate open or crowdsourced data, such as OpenStreetMap, Twitter, Foursquare, and Sina Weibo, particularly for capturing mobility and behavioural insights. Qualitative research is often based on in-depth interviews, field research, or typology analysis (Khare, 2020; Pillsbury, 1987).

Regarding spatial considerations, all articles incorporate spatial factors into their models, highlighting a consistent focus on geographical context in retail research. In addition, exogenous factors, such as demographics, accessibility, competition, population density, or economic conditions, are included in every study, even in the qualitative or field-based analyses.

In terms of predictive capabilities, approximately 77% of the studies present models that forecast the performance of new store locations. In contrast, qualitative or exploratory studies (e.g., Lee et al., 2022; Beckers et al., 2022) generally do not include predictive components.

Concerning time horizons, a significant number of studies (n=12) focus on short-term horizons, reflecting their emphasis on operational and tactical decisions (e.g., Hunneman et al., 2023; Formánek & Sokol, 2022). Studies dealing with strategic planning, long-term market evolution, or infrastructure tend to have a medium-long term scope (e.g., Lee et al., 2022; Low et al., 2021). Finally, qualitative or exploratory works often report unspecified horizons (e.g., Khare, 2020; Robinson & Balulescu, 2018)

Table 2.6: Techniques and Variables Used

ID	Used Techniques	Variables Used
1	Geographically Weighted Regression (GWR); Spatial econometrics; Sales Surface Optimization	Store performance (sales), department composition, local demographic and economic variables, location coordinate
2	Linear Regression Models; Linear Mixed-Effects Models	Store-level sales data, income, education, population density, proximity measures
3	Kernel Density Estimation; Spatial Autocorrelation (Moran's I); Clustering; Check in Data Analysis	Foursquare check-ins, pedestrian flows, number and types of restaurants, spatial geographic coordinates
4	Modified Spatial Interaction Model (SIM)	E-commerce penetration, store attractiveness, customer accessibility, delivery range
5	Negative Binomial Regression Model; Location quotient analysis	Manufacturing presence, agriculture-related variables, employment density, rurality indicators
6	Spatial Competition Model; Spatial Competition Strategy Model; Double-Channel Convolutional Neural Network (DC-CNN)	POI categories, consumer behaviour data, competition indicators, accessibility metrics
7	Survey-based factor analysis; Descriptive statistics	Retailer preferences, mall features, parking, accessibility, catchment population
8	Multinomial Logit (MNL) Model; Maximum like hood estimation	Land use category, transport access, labour availability, zoning regulations
9	Linear Regression; Spatial Regression (SARAR model); Random Forests	Trip origin/destination density, land use type, population, POI, elevation, distance to transit
10	Ordinary Least Squares (OLS) Regression Models; Spatial Lag Regression Models; Local Indicators of Spatial Association (LISA); Spatial econometrics	Brewery counts, GDP, employment, population age groups, proximity to urban centres
11	Affinity-based Popularity Inference (API) Model; Comparison models; Linear Regression	Temporal activity patterns, Venue category, Check-in data, Geographic coordinates
12	Time weighted Spatial Interaction Model (modified Huff model)	Store size, Population density, Travel time, Temporal demand weights
13	Regression analysis; Site scoring model	Retail sales, Demographics, Competitor presence, Accessibility scores
14	GIS-based spatial analysis; Statistical modelling	Tourist flows, Retail type, Accessibility, Spatial clustering
15	Regression models, Data aggregation from public sources	Consumer spending, Store location, Population data, Transport access
16	BP Neural Network; Spatial accessibility modelling	Accessibility index, Population density, Land use type, Shop characteristics
17	Cluster analysis; Spatiotemporal demand modelling	Sales volume (hourly), Loyalty card data, Catchment areas, Time of day
18	Binary Logit Regression; GIS mapping	Hotel grade, Accessibility, Agglomeration effects, Urban rehabilitation, Crime levels
19	K-nearest neighbour; Gaussian Processes	Check-in data, Venue type, Geographic coordinates, Transitions between venues
20	Gravity model; Market potential model; Huff model	Income, Population, Distance, Seating capacity, Visibility, Parking, Menu price

21	Geostatistical modelling; Bayesian estimation; Kriging; Spatial interpolation	Sales potential, Market response, Spatial geographic coordinates, Measurement instrument parameters
22	Regression analysis; Spatial econometrics; Geographic accessibility modelling	Store accessibility, Income, Car ownership, Population density, Neighbourhood classification
23	Game Theory Modelling; Structural econometric estimation; Simulation	Profit, Market share, Distance to competitors, Store location choices, Cost structures
24	Gravity Models (Huff model); Regression analysis; Cluster analysis	Attractiveness index, Travel time, Store area, Competitive intensity, Demographics
25	Field Survey; Interview analysis; Spatial typology mapping	Restaurant type, Accessibility, Agglomeration, Socioeconomic traits, Traffic density
26	Regression modelling; Covariance analysis; Spatial trip generation modelling	Consumer trips, Store size, Employees, Population density, Proximity to competitors

Source: Own elaboration

Table 2.7: Characterisation of Data, Sources and Models

ID	Type of data	Data Sources	Model(s) take spatial factors into account	Model(s) predict the performance of new locations	Model(s) incorporate exogenous factors	Horizon (e.g., short-term, medium to long-term)
1	Spatial	Retailer Customer Database; Store Manager Surveys; Commercial Demographic Data	Yes	Yes	Yes	Short-term
2	Cross-sectional, Spatial	Internally generated data from a major drugstore chain operating in Czechia; Czech Statistical Office (CZSO); Eurostat; OpenStreetMap (OSM) database	Yes	Yes	Yes	Short-term
3	Spatial	Ministry of Public Administration and Security of South Corea; NAVER; Minister of Land, Infrastructure and Transport of South Corea; Open Government Data Portals; Social Network Services (SNS) Reviews	Yes	No	Yes	Medium to Long-term
4	Cross-sectional, Temporal	Office for National Statistics (ONS); British Population Survey (BPS) (2010); CACI estimates; Statista (2019); Retailer-Provided Data; GMAP Ltd (2016); Clarke, Thompson, and Birkin (2015); Living Costs and Food Survey (ONS)	Partially	No	Yes	Medium to Long-term
5	Cross-sectional, Spatial	National Establishment Time-Series (NETS); USDA Census Data of Agriculture (2012); Census Data Bureau; Bureau of Economic Analysis (2012); U.S. Patent and Trademark Office (1995-99);	Yes	No	Yes	Long-term

		Office of Management and Budget (2003); Federal Communications Commission, Form 477 (December 2003); Esri Data & Maps; National Centre for Education Statistics (2012); National Conference of State Legislatures (data prior to 2013); USDA ERS Natural Amenities Scale (McGranahan – 1999); FDIC Summary of Deposits (2000); Bureau of Labor Statistics (2012); Industry data based on 50 largest firms;				
6	Spatial	Retail sales and store location data; WorldPop; Sina Weibo Location-Based Service (LBS); Baidu Map API; National Catalogue for Geographic Information (China); OpenStreetMap (OSM)	Yes	Yes	Yes	Short-term
7	Qualitative	Qualitative data from forty in-depth interviews; Press and retail reports	Yes	No	Yes	NA
8	Spatial	Subset of the data of logistics facilities; Ile de France Region Planning Agency (IAUIDF); Data for Logistics Zones (LZs); French National Institute of Statistics (INSEE); Route 120 database (2015)	Yes	Yes	Yes	Long-term
9	Spatial	Smid System (Free-floating E-Bike-Sharing System); Swiss Federal Statistical Office (BFS); Swiss Federal Office of Transport; OpenStreetMap (OSM); Swiss Federal Office for Spatial Development; Swiss Open Data Portal	Yes	Yes	Yes	Short-term
10	Spatial	Official Associations of Brewery Companies; Central Statistical Offices; Eurostat (2016); The Brewers of Europe; Untappd; RateBeer; Google Maps; Official Websites and Fan Pages of Breweries; Interviews with People Connected with Breweries	Yes	No	Yes	Long-term
11	Longitudinal, Spatial	Twitter; Foursquare Check-In Data	Yes	Yes	Yes	Short-term
12	Temporal, Spatial	Major U.K. retailer; 2014 ONS Living Costs and Food Survey; Census Data (2011); Acxiom Ltd; Office for National Statistics (ONS)	Yes	Yes	Yes	Short-term
13	Cross-sectional	Partner Retailer's; U.S. Census Data; OpenStreetMap (OSM)	Yes	Yes	Yes	Medium to Long-term
14	Cross-sectional, Spatial	Partner U.K. Retailer; Geocoded databases of visitor accommodation; Census Data; Local and National Surveys; Insight from Store Loyalty Cards provided by a partner retailer; Office for National Statistics (ONS) Living Costs and Food Survey; Specific Reports and Studies	Yes	Yes	Yes	Short-term
15	Cross-sectional	Statistics Canada National Surveys; National Household Surveys; Survey of Household Spending (SHS); Canadian Socio-Economic Information Management System (CANSIM); Census Data; North	Partially	No	Yes	Long-term

		American Industry Classification System (NAICS); IHS Global Insight (2013) Report; Environics Analytics (EA)				
16	Spatial	Users; Local Company Website or Internal Data; OpenStreetMap (OSM)	Yes	Yes	Yes	Short-term
17	Temporal, Spatial	Electronic Point of Sale (EPOS) System from a Major U.K. Grocery Retailer; Loyalty Card Scheme; 2011 Census Data; Higher Education Statistics Agency (HESA); Annual Survey of visits to visitor attractions by VisitEngland	Yes	Yes	Yes	Short-term
18	Cross-sectional, Spatial	Turismo de Lisboa Website; Booking Website; www.lisbonlux.com/crime; Ferreira (Dissertação de Doutoramento); GIS tools	Yes	No	Yes	Medium to Long-term
19	Longitudinal, Spatial	Dataset from a collaboration with Foursquare Check-in Data; Office for National Statistics (ONS); Greater London Authority (GLA) LSOA Atlas	Yes	Yes	Yes	Short-term
20	Cross-sectional, Spatial	USA.com; American FactFinder on the U.S. Census Data Bureau website; Individual reports from CoStar, Commercial Real-Estate Information Company Website Or Internal Data; Field research conducted at each restaurant; Lousville and Jefferson County Information Consortium (LOJIC)	Yes	Yes	Yes	Short-term
21	Spatial	Data provided by Alberto Saccardi Da Nunatac Srl	Yes	Yes	Yes	Medium to Long-term
22	Cross-sectional, Spatial	National Establishment Time-Series (NETS) database (Wall and Associates / Dun & Bradstreet); GeoLytics normalized U.S. Census Data (1990 and 2000); Office of Management and Budget (OMB); MABLE/Geocorr engine	Yes	No	Yes	Medium to Long-term
23	Spatial	Original Dataset compiled by the Authors (store-level data for Aldi and Lidl); German Municipal and Regional Data Sources	Yes	Yes	Yes	Long-term
24	Cross-sectional, Spatial	Extensive databases used by a large Portuguese Retail Group; EUROSTAT; In-depth interviews	Yes	Yes	Yes	Medium to Long-term
25	Cross-sectional	Atlanta Telephone Directory Yellow Pages; Census Data; Random interviews with both patrons and restaurant personnel; In-depth interviews of local restaurant personnel	Yes	No	Yes	NA
26	Cross-sectional, Spatial	Origin and Destination Study conducted in Texas Highway Department; Land Use Survey conducted by the Waco Urban Planning Department	Yes	Yes	Yes	Short-term

Source: Own elaboration

2.2.3 Location Factors in Retail Site Location

The choice of location represents a long-term strategic decision that significantly influences the performance and survival of retail stores (Hunneman et al., 2023; Formánek & Sokol, 2022). Recent studies demonstrate that an optimal site maximizes sales potential while minimizing both mobility and competitive costs (Ouyang et al., 2020; Glaeser et al., 2019). The literature converges on the idea that territorial, socioeconomic and behavioural dimensions must be analysed within an integrated and multidimensional framework (Beckers et al., 2022; L. Wang et al., 2018).

Based on the 26 studied analysed, the main location factors identified in the literature can be grouped into five overarching dimensions, each capturing a distinct set of determinants that shape retail spatial dynamics.

The demographic and socioeconomic dimension includes factors such as population density, average income, education level and age structure, which determine the size and profile of local demand (Wojtyra et al., 2020; Zhu & Singh, 2009). Proximity to populations with higher purchasing power further increases the attractiveness of an area (Schuetz et al., 2012). Moreover, several urban studies emphasise the importance of local economic structure and employment levels in defining the potential of retail markets (Low et al., 2021).

The accessibility and infrastructure dimension highlights that the quality of road network, public transport availability, parking facilities and overall mobility are key determinants of store attractiveness (Formánek & Sokol, 2022; Sakai et al., 2020). Furthermore, in convenience and foodservice sectors, factors such as visibility and pedestrian flow play a particularly relevant role (Dock et al., 2015; Lee et al., 2022). In addition, GIS-based approaches and gravity models reinforce that the distance between consumers and stores remains one of the main predictors of store choice (Hsieh et al., 2019; L. Wang et al., 2018).

The competition and agglomeration dimension shows that the presence of competing or complementary stores can generate positive agglomeration effects (Khare, 2020; Oppewal & Holyoake, 2004). However, recent literature highlights the importance of balancing competitive proximity with synergistic attraction between retailers (Teller et al., 2016; Ouyang et al., 2020). Empirical evidence also indicates that stores located within commercial clusters tend to achieve higher performance due to increased consumer flow (Hunneman et al., 2023; Formánek & Sokol, 2022).

From a spatial perspective, centrality, land use, and urban density shape market potential and retail catchment areas (Newing et al., 2018; Cró & Martins, 2018). Areas combining residential and commercial functions are especially suitable for everyday retail sectors such as bakery and pastry shops (T. B. P. Waddington et al., 2018). The broader urban context also influences perceptions of safety and convenience, which indirectly affect location attractiveness (Hernández & Bennison, 2000; Pillsbury, 1987).

Finally, the economic and contextual dimensions emphasises that land and rental prices, access to financing, and market stability are recurrent determinants in location decision-making (Sakai et al., 2020; Low et al., 2021). Furthermore, the environmental and sustainability dimensions have recently gained attention, particularly in studies linking location choice to local consumption and short supply chains (Low et al., 2021).

These five dimensions underscore the multifaceted nature of retail location choice and highlight the importance of considering a range of spatial, socioeconomic, and contextual factors in retail site selection analysis.

2.2.4 Study Evaluation

Table 2.8 outlines the contributions provided by the reviewed literature, the future investigations suggested by the articles or based on logical inferences, the limitations and the suggestions for store locations. This analysis is relevant for reviewing the findings and identifying the existing gaps that can be addressed through future research.

By analysing the contributions of all the articles, it is evident that they offer different types of insights. Some articles contribute with methodological frameworks and results that can be extended to other industries (e.g., Formánek & Sokol, 2022; Hsieh et al., 2019), although others provide conclusions on customer behaviour (e.g., Hunneman et al., 2023) and location factors (e.g., Khare, 2020; Wojtyra et al., 2020), presenting valuable insights for future research. In addition, some studies present relevant information about specific regions that can be generalised to others (e.g., Wang et al., 2018), while others offer useful perspectives for optimising store location strategies (Newing et al., 2018).

In terms of future research directions, a lot of the articles suggest considering more factors in retail site selection (e.g., Ouyang et al., 2020; Hsieh et al., 2019). Nevertheless, some also suggest expanding the analysis period (e.g., Low et al., 2021), use more recent data (e.g., Beckers et al.,

2022), consider multiple regions or cities (e.g., Guidon et al., 2020), develop more models (e.g., T. Waddington et al., 2019) and include temporal variations (e.g., Finazzi, 2013).

Concerning the analysis of limitations, there are some studies that are restricted to a single retailer, city or sector (e.g., Formánek & Sokol, 2022; Glaeser et al., 2019) and lack external validation or replication in other contexts (T. Waddington et al., 2019; Newing et al., 2018). Furthermore, some incorporate outdated data or rely on proxies (e.g., Beckers et al., 2022; Ouyang et al., 2020) and lack data on consumer behaviour or temporal dynamics (D'Silva et al., 2018; Horton, 1968).

Finally, regarding the suggestions for potential store location or expansion strategies, several studies recommend locating stores in areas with high population density or concentrated residential presence (Hunneman et al., 2023;Formánek & Sokol, 2022; T. B. P. Waddington et al., 2018), while others highlight the importance of accessibility via transport nodes, mobility patterns and proximity to infrastructure (Lee et al., 2022; Sakai et al., 2020; Wang et al., 2018). In addition, some suggestions focus on locating near existing commercial hubs, leveraging brand clustering or co-location effects (Khare, 2020; D'Silva et al., 2018; Zhu & Singh, 2009) and targeting tourist hubs or seasonal demand zones appears as a specific strategy in some papers (Newing et al., 201); (Cró & Martins, 2018). Some studies also emphasize the value of expanding into underserved zones or areas with retail scarcity (Robinson & Balulescu, 2018; Schuetz et al., 2012; Horton, 1968) and guide physical store planning based on online demand or digital consumer behaviour (Beckers et al., 2022; T. Waddington et al., 2019). At last, there are also some studies that support the use of models to simulate network expansion, test new formats, or evaluate scenarios (Ouyang et al., 2020; Dock et al., 2015; Mendes & Themido, 2004).

Table 2.8: Contributions, Future recommendations, Limitations and Suggestions

ID	Contributions	Future Research	Limitations	Suggestions for potential store locations or expansion strategies
1	Customers tend to visit stores closer to their homes, and areas with higher populations generate more sales. The introduced model helps retailers predict sales and assess the impact of location and store layout on performance.	Consider more than one chain of stores.	Limited to a single retailer and geographic context; Lacks Generalizability to other sectors.	Focus on high-density residential areas with established customer bases. Optimize store layout for target demographics.
2	The methodology and results can be generalized to other business, whether in development or already established.	Methodology and results can be applied to other markets and products.	Does not address sector-specific variation; Results may not fully translate across industries.	Methodology could be applied to identify optimal locations in similar markets segments or other European urban contexts.
3	A logistic function explains restaurant distribution, predicts customer numbers, and guides new openings in cities with public transport influence.	Suggest applying the model to other cities. It also suggests that its model-driven approach can complement data-driven approaches.	Restricted to urban areas with developed public transport networks; Lack cross-regional validation.	Target restaurant openings in zones with pedestrian flow and proximity to transit hubs.
4	The model represents empirical variations in consumer trends in e-commerce within the food retail market and predicts the impact of e-commerce on consumer demand and physical sales.	Investigating the correlation between a store and a click-and-collect behaviour; Incorporating additional metrics like discounts and price incentives; Utilizing more recent data to improve the model; Developing estimated demand by disaggregating parameter y; Using optimization models to calibrate parameters simultaneously.	Data limited to earlier e-commerce trends; Model may underperform with recent consumer behaviours.	Use output to refine click-and-collect location planning in grocery retail. Align with online demand clusters.
5	Links agriculture (direct/organic sales) to food start-ups, emphasizing spatial factors and market innovations along rural-urban continuum.	Studying startup survival and growth over time; Analysing the impact of agriculture and entrepreneurial ecosystems on sector-specific startups; Expanding the analysis period and refining regional factors.	Lack of endogeneity controls beyond lagging variables; Data restricted to NETS and only up to 2015; Earlier period results (2004–06) not detailed.	Regions with supportive rural infrastructure and entrepreneurial policy environments show potential. Focus on counties with higher food manufacturing activity and related agri-businesses.
6	A model to identify locations for stores.	Integrating spatiotemporal big data from crowdsourcing; Considering factors like income, store size, policy, public safety, and local consumption structure in retail location selection.	Relies solely on spatial simulation and social media data; May not reflect actual purchasing behaviour	Combine predicted demand and social media engagement to identify underserved urban areas.

7	Highlights key factors influencing retail location decisions, including economic factors, infrastructure, clustering, brand presence, and connectivity. Has insights for further quantitative research.	Exploring the impact of location attributes, such as transports and nearby services, on shopping centre planning; Analyse consumer attitudes towards shopping centre locations and facilities; Investigating how insights can guide site development based on consumer needs.	Exploratory and qualitative in nature; Lacks quantitative validation of identified factors.	Inform site selection based on consumer preferences for anchor brands and service diversity.
8	Emphasizes the role of land-use regulations, traditional clusters, and highway access in the location of logistics facilities.	Investigating land price and accessibility indicators for logistics facilities; Analysing spatial autocorrelation and unobserved factors; Conducting case studies in various cities; Exploring logistics facility location patterns and land use policy implications.	Limited to one type of facility; May not generalize to other retail formats.	Logistics facilities could be located near highway nodes and existing industrial clusters to maximize accessibility and agglomeration benefits.
9	Spatial regression outperforms random forests in predicting demand for new cities, emphasizing the importance of factors like population and employment.	Expanding analysis to multiple cities; Examining the impact of bike availability on demand with comparable systems; Investigating the relationship between bike density and demand; Exploring alternative predictive models; Combining spatial regression and random forest models using conjoint factor analysis.	Predictive performance was limited due to non-transferable spatial patterns and city-specific factors. Centrality measures reduced model accuracy in new cities. Supply-side constraints and omitted variables were not modelled.	Recommended focusing on areas with high population, employment, and presence of bars/restaurants. Avoid over-reliance on centrality measures; Emphasize transferable land-use features.
10	Key factors influencing craft brewery locations include quality of life, education, economic activity, creativity, and demographic factors. It identifies regions with growth potential and suggests the Czech market as a model for Poland and Slovakia.	Extending the study to other regions, particularly individual countries; Leveraging the availability of more detailed data of further analysis.	Different spatial analysis methods across regions; Use of Eurostat and NUTS 3 data introduces modifiable areal unit problem; Fixed agglomeration thresholds inconsistent across countries.	Target high-GDP, urban-adjacent regions with strong millennial and middle-aged demographics. Hot spots include nearly all Czech Republic and Trnavský (Slovakia); Avoid eastern and northern Poland.
11	Method to predict the popularity of store locations and optimize marketing strategies based on visitor patterns. It can be applied across various regions to identify location-specific popularity predictors.	Incorporating diverse urban data into temporal popularity models; Exploring factors like traffic, weather, and social interactions.	Does not explicitly address challenges with sparsity in less visited areas; Long-term success predictions are outside scope.	Suggests identifying neighbourhoods with similar temporal visitation patterns to inform placement of new venues.
12	SIM model that improves revenue forecasting accuracy, optimizes store design, and accounts for spatiotemporal demand variations to enhance retail strategy.	Developing models that are applicable both academically and in industry settings.	No external validation reported; Applicability to sectors beyond UK grocery retail formats.	The model could be used to plan store formats in regions with fluctuating temporal demand.

13	Demonstrates how retailers can use big data by combining external data with internal sales and operations data to predict performance of potential locations. The model can be applied by other businesses, increasing profits by up to 86%.	Test the model in different sectors and validate predictive performance using independent retail datasets.	Study based on internal data from a single U.S. retailer; External validity in not guaranteed.	Use the model to guide site selection in other retail categories with large-format stores or chain operations.
14	The model explores store extensions, helping retailers test existing stores for future competitive environments. It demonstrates excellent predictive capacity for new store characteristics and proves useful for retail planning and the public sector in Looe.	Explore the model's applicability across other coastal cities or resort-based retail areas.	Case study limited to one town; Results not tested across diverse urban or retail contexts.	Model could support expansion in seasonal tourist destinations with similar demand profiles to Looe.
15	The methodology helps municipalities estimate retail sector spending capacity and supports better zoning decisions. It also identifies development pressures from large retailers.	Exploring the application of methods in a single retail sector for better location selection; Using independent sales data to enhance spatial mapping of surpluses and leakages for improved decision-making in retail locations.	Lacks direct integration of consumer behaviour or expenditure tracking; Relies on employment and zoning proxies.	The model could inform zoning strategies and investment planning in commercially underserved neighbourhoods.
16	New two-step model for retail site selection integrates spatial accessibility and market potential evaluation, using socioeconomic and social media data for more accurate analysis. The model's extensibility allows for its application in other cities and industries.	Investigating temporal changes in sales and the spatiotemporal relationship between socioeconomic data and market potential; Integrate factors like road connectivity, climate, and purchasing power of more accurate geographic results.	Relies on social media proxies. The model could be applied in cities where traditional datasets are available for benchmarking.	The model could be used to identify locations with strong accessibility and youth-oriented consumer density.
17	A clustering framework is proposed to optimize retail operations, predict demand patterns, and enhance resource allocation for new and existing stores.	Integrating temporal demand profiles into forecasting models; Combine sales data with time of the day information for better insights; Extend spatiotemporal demand analysis using refined spatial models.	Existing store classifications inadequate for grouping stores with similar temporal sales profiles; Limited understanding of temporal demand in convenience formats.	Improved store clustering based on hourly sales profiles enables better store planning and location decisions, especially for convenience store formats.
18	Results indicate that accessibility to transports, restaurants, and cultural attractions is not critical for hotel and hostel location, with lower-quality hotels actively seeking clustering effects. Additionally, high-end hotel location patterns shifted after Expo 98.	Developing a more comprehensive chronological analysis of the influences on spatial distribution; Incorporating a more detailed account of government intervention; and obtaining historical data for establishments prior to 1980.	Generalizability beyond Lisbon may be limited; Not all urban tourism factors fully explored.	Upscale hotels should consider rehabilitated urban areas; Lower-grade hotels benefit from agglomeration.
19	Characteristic temporal curves offer valuable insights for optimizing store operations, supply	Investigate location with varying demand characteristics, not just the most popular	Challenges with data sparsity at fine spatial resolution and	Use neighbourhoods with similar temporal activity patterns and locality

	purchasing, and demand forecasting. They also aid in designing location-based technologies and understanding how a local influences its neighbourhood's temporal profile.	ones; Apply the models to other cities for broader analysis; Train a different model for each location, potentially adjusting k values and Gaussian Process hyperparameters.	generalizability beyond London.	for demand prediction and venue planning.
20	A gravity model with carefully selected variables effectively evaluates supply and demand interaction for restaurant locations. It improves predictive accuracy, helping restaurant owners select high-quality sites and optimize market share by incorporating factors like seating availability, parking, visibility, and menu pricing.	Improve predictive accuracy of gravitational model; Explore contrasting study areas with more dispersed patterns; Test new variables like signage, special pricing, and pedestrian traffic; Replace current configurations with peak-hour traffic, per capita income, or flexible menu pricing.	Model performance is sensitive to variable selection; Limited to a localized study area in Kentucky.	Refined gravity models could be used to consider income, accessibility, and visibility to select competitive restaurant sites.
21	The analysis concludes that daily sales volume can be significantly increased by focusing efforts on city areas with high market potential that are not adequately covered by the retail network.	Incorporate temporal variations and adjust for stores with different attractiveness levels.	Lacks detail on temporal shifts in demand; Assumes static attractiveness of locations.	Target expansion in urban zones identified as underserved based on spatial sales potential analysis.
22	Retail employment is lower in high-poverty areas due to fewer chain stores, while higher income increases employment and store area. Food desert classification depends on employment density, store size, and chain presence.	Explore the integration of consumer-side data to complement employment-based analysis and improve retail desert classification.	Focuses on employment outcomes; Does not include demand-side consumer data or sales performance.	Use findings to guide chain expansion strategies in higher-income, high-employment neighbourhoods.
23	Incorporating firm-specific asymmetries, the study enhances understanding of competitive interactions and market preferences in oligopolistic environments, offering a more precise view of firm behaviour.	Investigate strategic decision-making in dynamic games with ex-post regret; Include locally unobservable variables and utilize multi-agent dynamic games for market interactions.	The model applies only to discount retail chains in oligopolistic markets, limiting broader retail applicability; Less applicable to fragmented or small business environments.	The approach may inform competitive site planning for large-format retailers operating in concentrated markets.
24	Review of techniques and models for assessing location and revenue across multiple retail points.	Validate reviewed models through empirical application in various retail formats and countries.	Review does not assess effectiveness of techniques in real retail context.	Phased site selection methodology supports decisions about where to locate new outlets and what characteristics to prioritize in different regions.
25	It provides insights into the market positioning of different restaurant types and their location strategies, emphasizing the role of environment	Develop the temporal continuity of restaurant landscapes; Investigate relationships between restaurant clusters and their surroundings; Develop a predictive	Exploratory and descriptive; Lacks generalizability or formal modelling for replication or application.	Reveals typologies such as 'hamburger alleys' and 'ambient clusters' as patterns worth exploring

	and psychological factors in consumer choice beyond simple distance.	model for restaurant locations; Study influence of local factors in Atlanta on the results; Examine restaurant dynamics in rural and small-town areas.		when selecting new restaurant locations.
26	It highlights key location factors in consumer attraction and improves retail trip prediction accuracy through establishment-level analysis.	Explore the impact of psychographic factors on shopping choices; Integrate situational variables like peak times and traffic in predictive models; Examine competition dynamics in various regional markets and their urban planning implications.	Only accounts for trip frequency, without capturing visit purpose, loyalty, or frequency variation over time.	Leverage detailed trip-level data to refine predictions and improve urban retail zoning accuracy.

Source: Own elaboration

2.3 Summary and Implications of Literature

The Systematic Literature Review aimed to critically analyse a comprehensive selection of academic studies to identify the main research topics, methodological approaches, and gaps in existing knowledge. This process is essential to provide an informed basis for further research, as it highlights areas that have already been explore, and, crucially, those that remain underdeveloped. The review thus enabled the identification of specific research gaps that future studies on store location decisions can address, contributing to the advancement of knowledge in the field of retail location modelling.

While the retail sector has been widely examined in the existing literature, the food retail segment remains significantly underexplored. In particular, the bakery and pastry sector has received no academic attention, revealing a clear thematic gap for future research. Although a few studies have adopted broader geographical or organisational scopes (Hunneman et al., 2023; (Wojtyra et al., 2020), most existing studies focus on a single retail chain (Mendes & Themido, 2004), a single city (Cró & Martins, 2018) or a single country (Low et al., 2021), which limits the generalizability of their findings.

In addition, most studies focus exclusively on urban areas (e.g., Lee et al., 2022; Beckers et al., 2022; Ouyang et al., 2020) with only a few addressing both urban and rural contexts (e.g., Low et al., 2021;Wojtyra et al., 2020). Similarly, Small and Medium-sized Enterprises (SMEs) have received limited scholarly attention. As a result, both rural areas and SMEs remain largely underexplored, revealing another important gap in the literature.

The most recent data used in the reviewed studies dates to 2020 (Guidon et al., 2020). Since then, the retail landscape has been profoundly reshaped by the COVID-19 pandemic, which accelerated the digitalisation of consumption, boosted online retail sales, and altered the digitalisation and shopping behaviour (Organisation for Economic Co-operation and Development (OECD), 2020;Organisation for Economic Co-operation and Development (OECD), 2021). Consequently, more up-to-date research is needed to ensure the applicability of the findings to the current context.

The reviewed literature employs a wide range of modelling approaches, including regression analysis (e.g., Formánek & Sokol, 2022; Guidon et al., 2020) gravity models (e.g., T. Waddington

et al., 2019; Dock et al., 2015), neural networks (Ouyang et al., 2020; Wang et al., 2018), and game-theoretic frameworks (Zhu & Singh, 2009). In addition, spatial techniques, such as geostatistical modelling (Finazzi, 2013), spatial econometrics (e.g., Hunneman et al., 2023; Wojtyra et al., 2020) and spatial interaction models (Beckers et al., 2022; (T. Waddington et al., 2019) are frequently applied. Models based on spatial regression and interaction are highlighted as the most effective and integrate multiples sources of company and external data thereby increasing the robustness of their results. However, to enhance the predictive power and real-world applicability of these models, future research should incorporate a broader set of contextual variables, include indicators of consumer behaviour, and account for temporal variations.

3 Methodology

3.1 Methodological framework

This chapter presents the methodological framework adopted in this study, which aims to identify the territorial and socioeconomic characteristics of potential locations for the expansion of *Sabores do Ti Pereira, Lda*, a company operating in the bakery and pastry retail sector in Portugal. The research follows an exploratory and spatial approach, combining quantitative analysis with spatial data visualization.

This approach allows a better understanding of the business environment and company data, as well as the relationship between store performance and territorial context, while also producing insights that can be applied to similar retail businesses.

As presented in Figure 3.1, the methodological process comprised several stages: (i) literature review to identify location factors; (ii) collection of company and external data; (iii) data preparation; (iv) dimensionality reduction using Principal Component Analysis (PCA); (v) grouping of statistical sections through Cluster Analysis; and (vi) identification of the most suitable areas for new store locations.

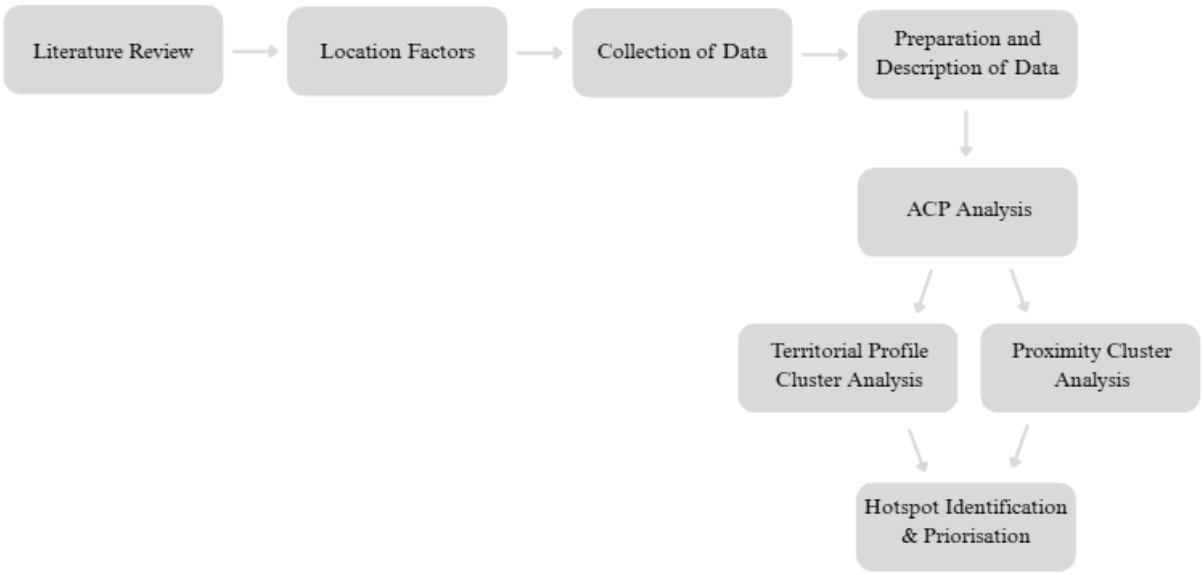


Figure 3.1: Diagram of Methodological Process

The analysis was conducted using Excel, GeoDa, QGIS and SPSS, which supported data processing, spatial analysis, and statistical modelling. The results were then interpreted considering the existing literature, focusing on location factors that may support store location decisions.

3.1.1 Case study: Company overview

This study focuses on *Sabores do Ti Pereira, Lda*, a Portuguese company operating in the bakery and pastry retail sector. With seventeen stores operating, the company represents a relevant case study for analysing the spatial dynamics and characteristics of local retail food businesses, and for understanding how geographical features can influence store performance and expansion decisions.

Sabores do Ti Pereira, Lda opened its first bakery in Abrantes at the end of the year 2001. Two years later, the company inaugurated its second unit, followed by a third in 2006, both also located in Abrantes. In 2008, a new bakery was established in Vale das Mós, which, although within the municipality of Abrantes, is geographically more distant from the city centre. In the same year, the company began its expansion beyond Abrantes by opening its first store in the municipality of Ponte de Sor. In 2010, another store was opened in Abrantes, and in 2011, two additional stores were inaugurated in Ponte de Sor, just one week apart. In 2012, the expansion continued with the opening of a store in Constância, marking a new municipality for the company. Between 2017 and 2019, three new stores were launched in Abrantes and one in Sardoal. The year 2023 recorded the highest number of openings, with four new stores inaugurated in different municipalities: Alter do Chão, Abrantes, Campo Maior and Gavião.

The municipality of Abrantes was the pioneer in this organization. Despite the company's subsequent expansion, it still has the highest number of stores, with a total of nine, as shown in the Figure 3.4. It is followed by the municipality of Ponte de Sor, where three stores are currently in operation. In addition, the company also maintains one store in each of the following municipalities: Alter do Chão, Constância, Campo Maior, Gavião and Sardoal.



Figure 3.2: Map of Portuguese Territory



Figure 3.3: Map of Store Locations

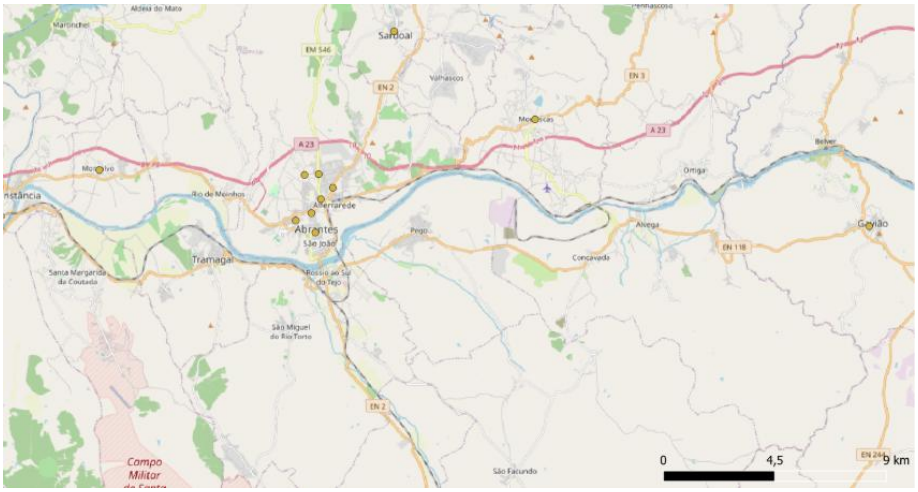


Figure 3.4: Map of Store Locations in the Abrantes Region

Given the company's continuous growth and expansion strategy, the study aims to support managerial decision-making by identifying the territorial and socioeconomic characteristics that define suitable locations for future stores.

3.2 Data Sources and Variables

At this stage, company data provided by the company and external data obtained from Instituto Nacional de Estatística (INE) and OpenStreetMap were collected, processed, and analysed to characterise the territorial and socioeconomic environment surrounding the company's stores.

During the preparation process, the company data was combined with the Census 2021 data available on the INE website and with Points of Interest (POIs) and road network layers extracted from OpenStreetMap. This integration was carried out using QGIS, which enabled the spatial linkage between the stores and the corresponding statistical sections.

In addition, external data was organised using QGIS and SPSS, which allowed data cleaning, standardisation, and the execution of the Principal Component Analysis (PCA) and two Cluster Analyses. These procedures were essential to reduce data dimensionality, identify spatial patterns, and distinguish different territorial profiles that could influence store performance and support future location decisions.

3.2.1 Company Data

To develop a reliable analytical database, the company provided a comprehensive database containing all variables necessary to describe and analyse the background of the case.

The dataset included detailed information on the location of each store, store areas, indoor and outdoor seating capacity, as well as total revenues and revenue by service type.

This company dataset, developed in close collaboration with the company, was subsequently complemented with external spatial and statistical data, as described in the following subsection.

3.2.2 External Data

3.2.2.1 Sources of External Data

To conduct a more comprehensive analysis, it was necessary to complement the company dataset with external information, and a dataset at the statistical section level was developed. Two main sources were used in this investigation:

- The Portuguese Census 2021 (INE), which provides statistical information at the section and subsection level, including demographic, housing and economic indicators.

- OpenStreetMap (OSM), which offers spatial data on the built environment, infrastructure, and points of interest, referring to Portuguese territory and processed using QGIS (data extracted in July of 2025).

These external sources are open-access and provide detailed information that characterises the areas surrounding each store and each statistical section, enabling a more accurate contextualisation of their operating environments. The selection of both sources was based on data availability and the potential for spatial comparability across territorial units, as noted by L. Wang et al. (2018).

3.2.2.2 Variables Groups

According to the sources previously presented, external data was organised into five groups of variables.

Sociodemographic data included indicators such as total population, age structure, gender distribution, level of education completed, and the proportion of economically active and inactive individuals, as well as employment by sector. These factors are widely recognised in retail location studies for capturing the demographic and economic characteristics of local demand and workforce composition (Formánek & Sokol, 2022; Glaeser et al., 2019; Schuetz et al., 2012). Variables such as population density, age distribution, education level, and employment structure influence purchasing power, consumer behaviour, and consequently the sales potential of retail stores (Low et al., 2021; Hunneman et al., 2023).

Housing and building information involved the number of residential and non-residential buildings, as well as the number of dwellings, classified as occupied, vacant, main and secondary residences. The pattern of occupancy provides insight into population stability and seasonality, which are important for assessing both regular and potential market demand (T. B. P. Waddington et al., 2018). Moreover, identifying areas with higher housing density and greater number of main residences reveal areas that are generally more favourable for proximity-based businesses such as bakeries and pastry shops (Cró & Martins, 2018).

Infrastructure and accessibility variables referred to the built environment, particularly to the length of different categories of roads, such as residential, primary, secondary, cycleways, and

others. These indicators are fundamental for store location decisions, as they determine consumer accessibility and reflect the degree of urbanisation of the surrounding area (Formánek & Sokol, 2022; Sakai et al., 2020). The type and length of road network indicate the level of urban development and influence both consumer flow and store visibility (L. Wang et al., 2018; Dock et al., 2015).

Points of interest identified the presence of different type of establishments such as bakeries, cafés, restaurants, and other services, which may influence the store activity. This information allowed the characterisation of the functional and commercial environment surrounding each store, identifying not only direct competitors but also complementary businesses (Hsieh et al., 2019; D.'Silva et al., 2018).

Competition variables were incorporated to analyse the competitive environment surrounding each store. These indicators account for the number and type of potential competitor located within the defined buffer zones. Furthermore, the distance to the nearest competitor was computed, providing a more precise measure of spatial proximity to rival establishments. These metrics enable the evaluation of the level of market saturation and the intensity of competitive pressure faced by each store, as highlighted in previous studies (Zhu & Singh, 2009; Hunneman et al., 2023).

These categories were identified as relevant, given their contribution to a more comprehensive characterisation of the areas surrounding each store, capturing demographic, physical, and competitive aspects of the local environment. By combining information on population, housing and buildings, infrastructure and accessibility, points of interest, and competition, it becomes possible to understand not only the heterogeneity of the environment in which the stores operate but also the potential relationships between these external factors and store performance. As a result, the inclusion of these factors enables a more robust and contextualised analysis of store performance.

3.2.3 Section-Level Context Dataset

To support the identification of potential new locations, an external dataset was constructed at the statistical section level, independent from existing locations. This dataset includes the same thematic groups of variables – sociodemographic, housing and building, infrastructure and

accessibility, points of interest, and competition – transformed to ensure a comparability and consistent interpretation across territorial units.

In addition to these variables, two new indicators were created to enrich the analysis. The first is a store presence flag, where a value of 1 identifies statistical sections that already contain a store of *Sabores do Ti Pereira, Lda* and 0 those without. The second variables measure the Euclidean distance (in metres) between each section centroid and company's central distribution facility, providing an indicator of logistical accessibility and potential delivery costs.

All variables were cleaned, standardised, and transformed as needed in QGIS and SPSS. This section-level dataset constitutes the input for the multivariate analysis (PCA and cluster analysis) applied across all sections in the study area, enabling a systematic screening of territories that exhibit profiles consistent with favourable store performance.

3.3 Data Dictionary

For a clearer understanding of the data used in this investigation, a data dictionary was developed to ensure transparency and facilitate the correct interpretation of each variable. As the groups of variables were outlined in the previous section, the dictionary provides a systematic summary including their definition, source, type (continuous, discrete, or categorical) and unit of measurement. The complete dictionary is presented in Appendix A, serving as a reference for future research and enabling replication or extension of the analysis.

3.4 Data Preparation and Processing

This subsection details the data preparation and processing procedures adopted prior to the statistical and spatial analyses. The process involved integrating internal business data with external territorial datasets, ensuring comparability and analytical readiness.

The company dataset was initially structured with the support of the company. Using QGIS and Google Maps, each store was georeferenced through the attribution and validation of geographic coordinates. This step enabled the integration of company information with spatial datasets and the subsequent execution of spatial analysis. All processing tasks were performed in QGIS, while Excel was used for data aggregation and preliminary organisation.

Regarding the sources presented in the previous section, the external variables were cleaned and prepared to ensure consistency and comparability across spatial units. Several indicators derived from the Census were converted into percentages to allow more accurate comparisons across stores in heterogeneous territorial contexts. In contrast, absolute measures such as the Number of Conventional Buildings, Number of Exclusively Residential Buildings, Total Number of Dwellings, Number of Family Dwellings and Total Population were preserved to provide a realistic representation of scale and density. Missing values and outliers were not removed, as they reflect genuine contextual differences rather than data errors.

Spatial information on road networks was obtained from OSM, from which a total length variable was generated for each road type to assess accessibility and the intensity of urban development. Points of Interest (POIs) were compiled from two different OSM datasets, one containing polygon features and the other point features. To construct a unified and reliable dataset, both layers were extracted from QGIS, merged in Excel, while duplicates were identified and removed through visual inspection of the maps in QGIS.

To characterise the environment surrounding each store, circular buffers of 200 metres and 400 metres were created in QGIS. These two scales were designed to capture both immediate proximity (200m) and a wider neighbourhood effect (400m), enabling a more nuanced spatial characterisation of each store's surroundings.

To ensure accurate measurements of road lengths and distances, all spatial layers were projected into the EPSG: 3763 (ETRS: 89 / PT-TM06) coordinate system. Datasets originally in different coordinate systems were converted to this projection to guarantee spatial accuracy and consistency across analysis.

In addition to the store-level database, a second dataset was developed at the statistical section level to enable a comprehensive spatial analysis of potential new store locations. This dataset, independent from the existing store network, integrates information from the 2021 Census with spatial data on POIs and road networks, providing a detailed territorial and socioeconomic characterisation of each statistical section.

To construct the dataset, information on buildings, population, POIs and roads was merged in QGIS. The resulting layers were transformed into percentages using Excel to enhance comparability between areas, while variables representing totals (e.g., number of buildings or total population) were retained as absolute values to preserve information on scale and density.

Using SPSS Statistics, several derived variables were created to facilitate interpretation and the detection of spatial patterns. These included density indicators (buildings, population, POIs, and roads), as well as specific measures such as the number of cafés, bakeries, malls, supermarkets, and competitors within each section. Additional variables were also generated to capture the presence of competitors (competitors_flag) and to normalise results by population, such as competitors per 1000 inhabitants, malls and supermarkets per 1000 inhabitants, and POIs per 1000 inhabitants.

To reduce data dimensionality and identify the most relevant territorial dimensions, a Principal Component Analysis (PCA) was conducted. The resulting components summarised the underlying spatial and socioeconomic variability across sections. Subsequently, a Cluster Analysis was performed in GeoDa to group statistical sections with similar characteristics and to support the visual identification of territorial typologies favourable to store expansion.

The overall purpose of this process was to ensure that both datasets were internally consistent, comparable across spatial units and suitable for subsequent statistical and spatial analysis, thereby reinforcing the robustness and validity of the results.

3.5 Data Analysis Framework

The aim of this section is to identify both spatial and statistical patterns related to the store performance and to the territorial context characteristics of their surrounding environments. This analytical process combines quantitative and spatial approaches to provide a comprehensive understanding of how internal and external factors influence store outcomes and to support data-driven location decisions.

The analytical workflow was designed as a sequential process comprising several stages. It began with an exploratory analysis of the datasets, during which data were cleaned, standardised, and spatially prepared using QGIS and Excel. This was followed by a Principal Component Analysis (PCA), conducted in SPSS Statistics, aimed at reducing data dimensionality and identifying the main components that summarise the most relevant territorial and socioeconomic dimensions.

To perform this analysis, a set of variables describing the demographic, economic, and structural characteristics of each statistical section was transformed into percentile rankings. These included indicators related to demography and human capital – such as the Share of Population

Aged 65 or Older, Share of Students, Share of Population with Higher Education Completed, and Share of Economically Active Population – as well as economic and employment structure, represented by the Share of Employed Population in the Tertiary Sector.

In addition, variables capturing the physical and spatial structure of the territory were incorporated, including Road_dens (Road Network Density), Total Roads Length, Number of Conventional Buildings, Building_dens (Building Density), and pop_ dens (Population Density). Finally, variables representing commercial dynamics and service intensity were included, namely Poi_dens (POI Density), competitors_1000inhab (Competitors per 1000 inhabitants), poi_1000inhab (POIs per 1000 inhabitants), mall_supermarket_1000inhab (Malls and Supermarkets per 1000 inhabitants) and Total Population.

This multidimensional set of variables was designed to capture the main territorial, socioeconomic, and competitive factors potentially influencing store performance and location suitability.

Subsequently, a K-Means Cluster Analysis was performed in GeoDa to group statistical sections with similar characteristics, thereby delineating distinct territorial profiles. The procedure aimed to identify homogeneous areas in terms of their economic, structural, and commercial characteristics, as summarised by the principal components.

Two complementary analyses were performed. The first, based on the percentile rankings of the three principal component scores obtained from the PCA, produced a five-clusters solution that classified statistical sections according to their overall territorial profiles. The second analysis considered the distance from each section to the company's distribution centre as an additional criterion, resulting in a ten-cluster solution that also accounted for spatial proximity.

The use of K-Means clustering was based on its proven effectiveness classifying spatial sections with similar socioeconomic and territorial characteristics (Ouyang et al., 2020;Guidon et al., 2020; Formánek & Sokol, 2022). The integration with PCA scores follows a widely applied analytical approach for identifying territorial typologies and urban profiles.

This combined approach, integrating multivariate and spatial methods, enables a systematic detection of spatial regularities, simplifies complex datasets into interpretable structure, and support the identification of areas with favourable conditions for future store expansion.

4 Results and Discussion

This chapter presents the results obtained from the applied methodology and interprets them considering study's objectives. The analysis provides essential insights that support the discussion of store performance and the identification of territorial characteristics suitable for future expansion.

4.1 Descriptive Statistics

According to Formánek & Sokol (2022), effective store location analysis requires combining sales performance data with the geographic and sociodemographic characteristics of the surrounding area. Following this approach, a descriptive analysis of both company and external datasets was conducted. This step was fundamental to identify the key variables influencing store performance and to provide a comprehensive understanding of their spatial, operational and socioeconomic context. The analysis was performed using SPSS Statistics to explore patterns and relationships that would guide the subsequent multivariate analyses.

4.1.1 Company Data

The company data provided by the company includes performance and operational indicators for each store, such as total revenue and number of transactions (tickets). These data were aggregated annually to allow comparison across stores.

In 2023, the company achieved a total of 4 024 129.12 €, corresponding to an average of 236 713.48 € per store. In 2024, the revenue increased by 1 179 426.74 €, reaching a total of 5 203 555.86 €, with each store generating an average of 306 091.52 €. Regarding the number of issued documents, stores produced a total of 1 316 516 in 2023 and 1 577 531 in 2024, representing an overall increase of 16.55%. On average, each store issued approximately more 15 354 documents when comparing both years.

The comparison presented in the previous paragraph did not consider that some stores only opened during 2023. A more refined analysis, considering only those that were operating throughout the entire year, reveals an average of 270 981.38€ and a total of 3 522 757.90 €. In terms of transactions, the total of issued documents was 1 197 973 with an average of approximately 92 000 per store.

Table 4.1: Analysis of Revenue and Issued Documents

	Average	Total (Year)
Revenue (2023)	236 713.48 €	4 024 129.12 €
Revenue (2023) - Only stores opened the entire year	270 981.38 €	3 522 757.90 €
Revenue (2024)	306 091.52 €	5 203 555.86 €
Variation	69 378.04 €	1 179 426.74 €
Number of Issued Documents (2023)	77 442.12	1 316 516
Number of Issued Documents (2023) - Only stores opened the entire year	92 151.77	1 197 973
Number of Issued Documents (2024)	92 795.94	1 577 531
Variation	15 353.82	261 015

Source: Company records

When comparing stores by revenue, the highest value was recorded in Ponte Sor 2 (587 524.52€), while the lowest was observed in Encosta da Barata (67 841.73 €). Across all stores, the standard deviation was 144 272.90 € and the median stood at 201 816.63€. In 2024, the same stores registered the maximum and minimum revenues (608 218.21 € and 75 289.79€, respectively). The standard deviation decreased slightly to 141 819.37€, while the median increased to 307 715.04€.

Considering just the stores that were operating throughout the entire year of 2023, the minimum and maximum revenues remained unchanged. However, the median and standard deviation increased to 284 487.81 € and 148 038.56 €, respectively. This analysis reveals greater variability among the fully operating stores, suggesting that differences in performance become more evident once recent opened units are excluded.

Table 4.2: Descriptive Analysis of Revenue

	Median	Maximum	Minimum	Standard Deviation
Revenue 2023	201 816.63 €	587 524.52 €	67 841.73 €	144 272.90 €
Revenue 2023 – Stores open the entire year	284 487.81 €	587 524.52 €	67 841.73 €	148 038.56 €
Revenue 2024	307 715.04 €	608 218.21 €	75 289.79 €	141 819.37 €

Source: Company records

According to the revenues recorded in both years, all stores increased their sales, except the Euromilhões store, which registered a decrease of 1 517.23 €. However, this store also operates Jogos Santa Casa, another source of income that is not included in this analysis.

A scatterplot was created to allow a clearer analysis of relationship between the revenue and issued documents. To ensure a more accurate analysis, the axes of the graph in Figure 4.1 were defined using the median of each indicator. This approach was adopted due to the high heterogeneity among stores in both variables. Consequently, the use of median provides a more robust and representative comparison. Therefore, the graph is divided into four quadrants:

- Quadrant I (upper right): high revenue and high transaction volume, representing the strongest performers;
- Quadrant II (upper left): high revenue but low transaction volume, indicating stores with fewer transactions but a higher average ticket value;
- Quadrant III (lower left): low revenue and low transaction volume, reflecting underperforming stores.
- Quadrant IV (lower right): low revenue but high transaction volume, suggesting stores with many transactions but relatively low value per sale.

Regarding graph 4.1, in 2023, the analysis of issued documents and revenue across all seventeen stores highlights a clear heterogeneity in performance. Two stores, Abrantes (Lidl) and Ponte Sor 2, are not represented in the graph because they are outliers, with results far above overall distribution. Among the remaining fifteen stores, six stores (35%) recorded values below the central tendency for both variables, while one store (6%) had transaction volume around the median but lower revenue, suggesting a lower average transaction value. In contrast, four stores (24%) performed above the median in both indicators, representing the strongest units, and three (18%) were positioned around the median in both variables. Finally, one store combined high revenue with a mid-range number of issued documents, suggesting a profile based on higher-value transactions rather than volume.

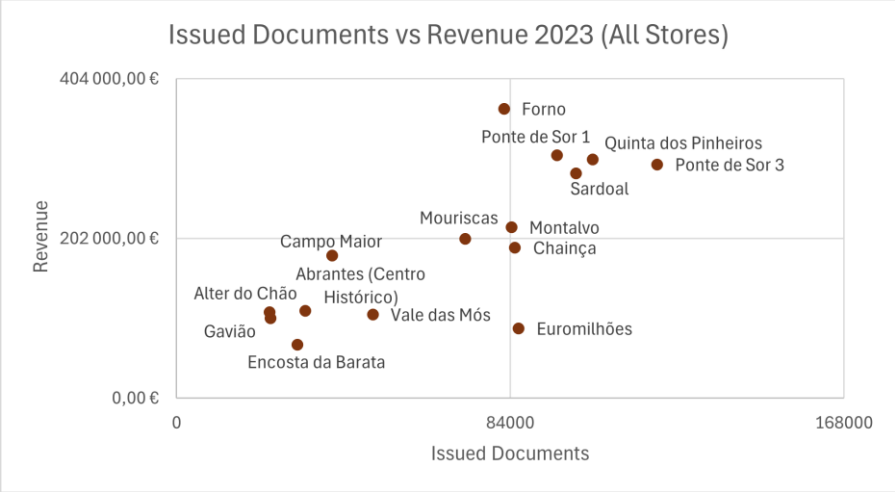


Figure 4.1: Issued Documents vs Revenue 2023 (All Stores)

When restricting the analysis to the stores that operated throughout the entire year, Ponte Sor 2 unit is not represented because it is an outlier. Among the rest of the stores, six (35%) were in quadrant III, with three of them really close to the threshold of issued documents. In contrast, five units (29%) were in quadrant I, with four stores lying just above the revenue median and one, Abrantes (Lidl), standing out clearly with both transaction volume and revenue at highest levels. Forno store is in quadrant II, combining revenue above the median with transactions volume close to the benchmark.

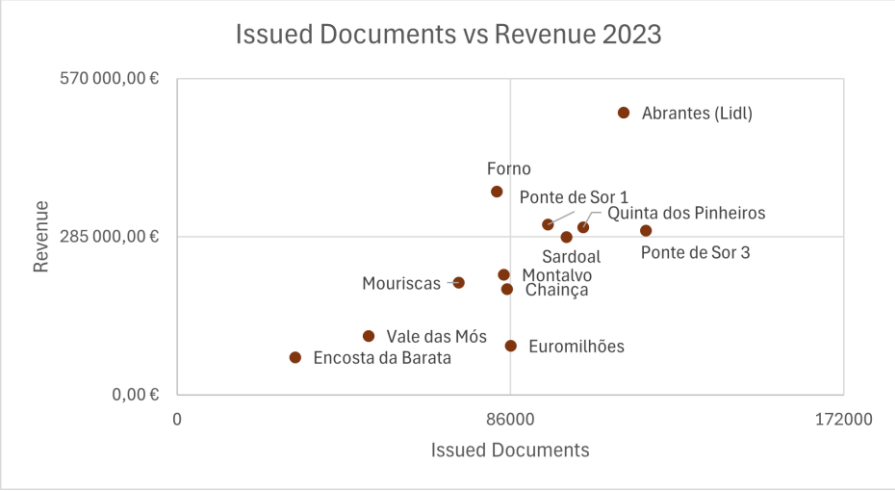


Figure 4.2: Issued Documents vs Revenue 2023

In 2024, the overall distribution remained consistent with the previous year. Ponte Sor 2 is clearly an outlier, with both revenue and transaction volume far above the rest of the network. Six stores (35%) were located in quadrant III, with both variables below the median, indicating weaker performs such as Encosta da Barata and Vale das Mós. A larger number of stores, including Abrantes (Lidl), Abrantes (Centro Histórico), Gavião and Quinta dos Pinheiros, were positioned in quadrant I, combining above median revenue and transactions volume. Three stores were located near the axes of quadrant I and IV, showing intermediate performance. Overall, although some minor adjustments occurred, the performance profiles of most stores were consistent when comparing both years.

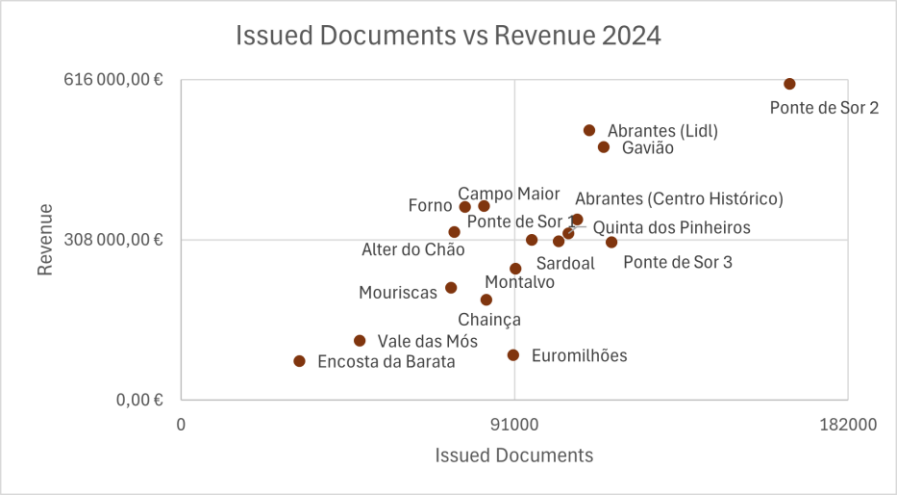


Figure 4.3: Issued Documents vs Revenue 2024

Overall, the findings demonstrate the diversity of store performance profiles, ranging from underperformers to top performing stores. More importantly, the comparison between 2023 and 2024 shows that these profiles remained largely stable over time. Even though newly opened units might still be adjusting to local demand, the relative stability over the years indicates that each store tends to keep its performance profile.

4.1.2 External Data

To understand the spatial structure and socioeconomic diversity of the Portuguese territory, data from OSM and INE were used to characterise the 9 929 statistical sections. The selected indicators reflect demographic, infrastructural, and competitive dimensions that have been consistently

recognised in retail location research as key determinants of market potential and accessibility (Formánek & Sokol, 2022; Ouyang et al., 2020; (Teller et al., 2016). Variables such as road and building density, population characteristics, and the presence of POIs and competitors provide complementary perspectives on accessibility, urbanisation intensity, and local market potential.

The table below presents a descriptive statistical summary of the indicators mentioned above. Overall, the 9 929 statistical sectors reveal strong territorial heterogeneity, as reflected by the high standard deviations observed for most variables. This variability highlights the marked contrasts between densely populated and well-connected urban areas and sparsely populated rural regions.

Regarding sociodemographic factors, the total population ranges from 8 to 2551 inhabitants per sector, with an average of 993, revealing an uneven population distribution and a clear concentration in urban areas. The average share of the population aged 65 or older (0.26) confirms a moderate level of demographic ageing, with extremes values ranging from 0 to 0.78, indicating sectors characterised by a predominantly elderly population. The average shares of the population in higher education completed (0.17) and economically active population (0.45) reflect moderate levels of education and labour market participation. In turn, the average share of students (0.06) highlights the limited presence of individuals still in education, while the share of employed population in the tertiary sector (0.29) reveals the predominance of service-related activities across several areas, consistent with the urban economic pattern of the country.

From a housing and building perspective, the average number of conventional buildings (340.62) and building density (549.44) show significant variance, suggesting the coexistence of dense urban zones and dispersed rural settlements. Higher building densities are typically concentrated in metropolitan and coastal areas, whereas rural regions tend to exhibit lower densities, reflecting the dual and polycentric urban structure of Portugal (Cró & Martins, 2018).

In terms of infrastructure and accessibility, the high amplitude between the minimum (701.06) and the maximum (215647) of the road density, illustrates substantial inequality between territories. However, the very high maximum value of both road density and the total road length (1107238.32) point to a strong concentration of road networks within urban areas.

Regarding the service and commercial environment, the standard deviation of POIs density (208.52) and POIs per 1000 inhabitants (135.96) present high values, showing a concentration of commercial and service establishments in a small number of urban sectors. The average number of competitors per 1000 inhabitants (1.9) and of malls and supermarkets per 1000 inhabitants

(0.45) suggest markets that are generally unsaturated, although some zones experience intense competitive pressure. This variation reveals the coexistence of areas with untapped market potential and others characterised by high retail concentration.

Table 4.3: Characterisation of Portuguese Territory

	Minimum	Maximum	Mean	Std. Deviation
Number of Conventional Buildings	3.00	1049.00	340.62	206.13
Total Population	8.00	2551.00	992.64	408.85
Share of Population Aged 65 or Older	0.00	0.78	0.26	0.10
Share of Population with Higher Education Completed	0.00	0.67	0.17	0.11
Share of Economically Active Population	0.06	0.72	0.45	0.08
Share of Students	0.00	0.26	0.06	0.02
Share of Employed Population in the Tertiary Sector	0.02	0.61	0.29	0.09
Total Roads Length	508.92	1107238.52	52144.70	71622.70
Population Density (Pop_dens)	0.89	65151.31	4395.18	7484.71
Building Density (Building_dens)	1.59	14558.01	549.44	807.06
Road Density (Road_dens)	701.06	215647.00	27744.38	25749.18
POIs Density (Poi_dens)	0.00	2744.60	87.38	208.52
Competitors per 1000 inhabitants (Competitors_1000inhab)	0.00	500.00	1.91	6.31
Malls and Supermarkets per 1000 inhabitants (Mall_supermarket_1000inhab)	0.00	125.00	0.45	1.73
POIs per 1000 inhabitants (Poi_1000inhab)	0.00	9750.00	37.93	135.96

Source: INE and OSM

The analysis of the external data highlights pronounced territorial disparities between dense, accessible, and economically dynamic urban areas and ageing rural regions with lower infrastructure and limited economic activity.

4.2 Principal Component Analysis (PCA)

As previously described in the methodology chapter, a Principal Component Analysis (PCA) was conducted to reduce dimensionality. Therefore, the variables analysed in the previous section were selected for this procedure as they capture the main territorial, socioeconomic, and infrastructural dimensions identified in retail location research as key determinants of market potential and accessibility (Formánek & Sokol, 2022; Ouyang et al., 2020).

Drawing on the literature, these variables capture complementary aspects of demographic structure, socioeconomic profile, accessibility and infrastructure, urban density, and competition and agglomeration effects, which together shape retail spatial dynamics. Population related indicators, including total population, age structure, education level, and economic activity, reflect the size and profile of local demand (Wojtyra et al., 2020;Zhu & Singh, 2009; Low et al., 2021).

Measures of accessibility and infrastructure, such as road length and road density, operationalise the role of mobility and connectivity in store attractiveness (Formánek & Sokol, 2022; Sakai et al., 2020; Hsieh et al., 2019).

Indicators of urban form, including population and building density, reflect spatial concentration and centrality effects that influence retail catchment areas (Newing et al., 2018; Cró & Martins, 2018; T. B. P. Waddington et al., 2018). Finally, variables capturing commercial intensity and competitive pressure proxy competition and agglomeration dynamics, which have been shown to generate both competitive and synergistic effects in retail location outcomes (Oppewal & Holyoake, 2004; Teller et al., 2016; Ouyang et al., 2020; Hunneman et al., 2023).

The initial dataset, collected from the INE and OSM, consisted exclusively of variables measured at absolute levels. To mitigate the strong spatial dependence inherent to territorial data, a data preparation process was implemented, including the relativisation and standardisation of variables, thereby reducing scale bias and improving the robustness of the PCA results. As a result, the fifteen variables were prepared and analysed.

As presented in Appendix B, the analysis extracted three principal components, which summarise the dominant territorial and socioeconomic dimensions of the study area. Component extraction was based on the Kaiser criterion (eigenvalue greater than 1), and the three retained components together explained approximately 68.1% of the total variance (Component 1 ≈ 35.98; Component 2 ≈ 17.97; Component 3 ≈ 14.15). This level of explained variance is considered satisfactory, as it captures most of the variability across statistical sections while avoiding redundancy among correlated variables.

To enhance the interpretability of the components, as orthogonal Varimax Rotation with Kaiser Normalisation was applied. This procedure facilitated a clearer factor structure by maximising the variance of factor loadings, thereby supporting the substantive interpretation of each component.

Given the exploratory nature of the study and the objective of improving the readability of quintile-based maps, the components were converted into percentile rankings (PFAC1_1, PFAC2_1, PFAC3_1). This transformation affects only the map legends and does not alter the underlying graphical representation of the results.

The first component (FAC1) presents high loadings on variables such as road density (0.880), population density (0.795), building density (0.758), and a strong negative loading for Number of

Conventional Buildings (-0.771). This component represents an urban density and infrastructure dimension, distinguishing compact, well-connected urban areas with high concentration of population, buildings, and services, from more dispersed rural territories.

Higher values on this component indicate territories characterised by greater accessibility, denser built environments, and stronger functional integration, typically corresponding to central and metropolitan regions such as Lisbon and Porto, as well as in some regional urban centres along the coast, while lower values dominate the rural municipalities, reflecting their population and infrastructural density, as shown in the figure 4.4.

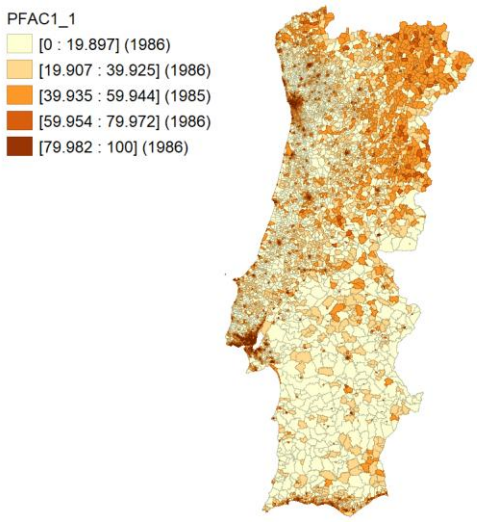


Figure 4.4: Quintile Map of Urban Density and Infrastructure Dimension

The second component (FAC2) is defined by high positive loadings on variables Share of Economically Active Population (0.767) and Share of Students (0.767), and a strong negative association with Share of Population Aged 65 or Older (-0.899). This dimension reflects socioeconomic and demographic vitality, contrasting territories with a predominantly active, younger, and more educated population against areas characterised by ageing and demographic decline.

High scores on this component indicates territories where the share of economically active individuals and students is significantly higher than inactive population, demonstrating stronger labour market participation and potential consumer dynamics. As shown in Figure 4.5, these areas are mainly concentrated in metropolitan and coastal regions, particularly around Lisbon, Porto, and the Algarve. Conversely, inland regions display lower scores, reflecting an ageing demographic structure and lower levels of economic and educational activity.

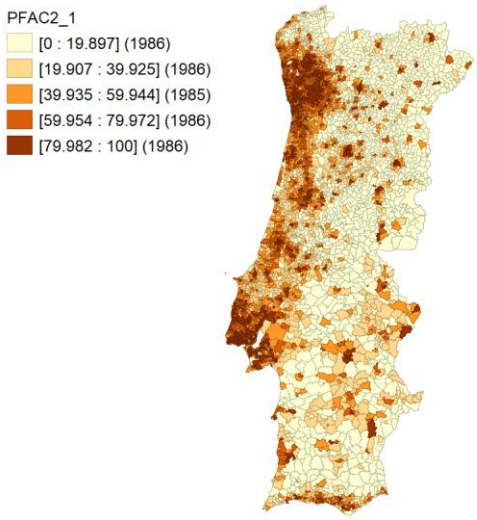


Figure 4.5: Quintile Map of Socioeconomic and Demographic Vitality Dimension

The third component (FAC3) is defined by very high loadings for competitors_1000inhab (0.913), poi_1000inhab (0.896), and mall_supermarket_1000inhab (0.858), describing a commercial and service concentration dimension. This component captures the intensity of retail and service activity within each statistical section, indicating areas where the concentration of commercial establishments and competitors is the highest.

As shown in Figure 4.6, higher scores on this component, represented by the darkest shades, are widely distributed across the Portuguese territory. These areas exhibit strong retail presence and a diversified service structure, reflecting higher levels of economic and commercial dynamism. In contrast, lower scores are mainly found in the interior and northern rural territories, where the service offer is more limited and competition is weaker, indicating less developed commercial networks.

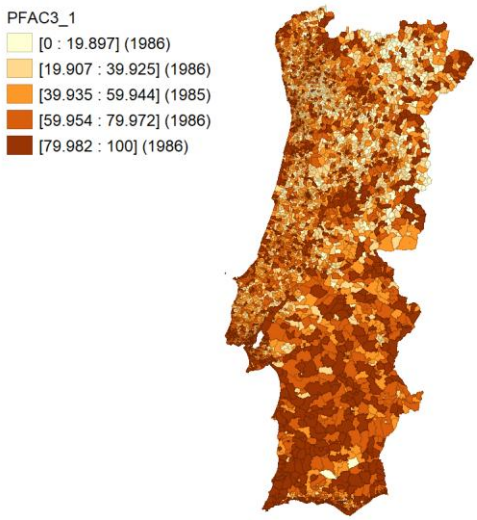


Figure 4.6: Quintile Map of Commercial and Service Concentration Dimension

The three extracted components are consistent with established evidence on retail location dynamics. The first component confirms that, as highlighted by Formánek & Sokol (2022) and Beckers et al. (2022), urban density and accessibility remain central factors for store success. Furthermore, Hunneman et al. (2023) demonstrated that both road and building density are also crucial determinants of commercial performance, which aligns with the results obtained in the PCA.

This pattern also aligns with Christaller's (1933) Central Place Theory, which postulates that higher population and infrastructure concentration increase the attractiveness and commercial reach of central places (Christaller, 1966). The strong positive loading of variables related to road, building, and population density, combined with the negative loading for the number of conventional buildings, confirm the internal consistency of the component and support its interpretation as an urban density and infrastructure dimension.

The second component aligns with the findings of Low et al. (2021), who demonstrated that regions with a higher proportion of active and educated population tend to exhibit stronger economic contexts and more dynamic retail activity. Their study also highlights that the presence of students and younger professionals increases the demand for convenience and food service establishments.

Similarly, Khare (2020) found that areas characterised by younger and economically active residents are more attractive to retailers due to the higher frequency and diversity of consumption, confirming that socioeconomic and demographic vitality is a key factor in location choice.

In contrast, Wojtyra et al. (2020) observed that zones with a predominance of elderly population tend to show lower mobility and reduced patronage of local businesses, which is consistent with the negative loading associated with this variable in the component.

Finally, Mahmud & Habib (2024) emphasises that, when developing business location choice models, socioeconomic characteristics, such as employment, education, and income, should be systematically integrated, as they are critical determinants of market attractiveness and viability. These findings collectively support the interpretation of the second component as a key determinant of territorial attractiveness.

The third component aligns with the findings of Zhu & Singh (2009), who demonstrated that the proximity between retailers create agglomeration economies, where stores benefit from shared consumer flows. However, the same authors highlighted that excessive competition could reduce the store performance, suggesting that high values of this component may represent both opportunities for commercial dynamics and risks of market saturation.

Dock et al. (2015) further found that commercial density and service diversity enhance territorial attractiveness, emphasising the importance of complementary retailers in forming local commercial centres. Their findings support the inclusion of variables related to POIs to capture these spatial dynamics.

Similarly, D'Silva et al. (2018) analysed clustering effects and concluded that the agglomeration of complementary activities increases the probability of location choice, reinforcing that bakeries and cafés benefit from proximity to other services. Khare (2020) also strengthened the concept of commercial synergy, showing that higher diversity of offerings increases the likelihood of attracting varied consumer profiles. The author demonstrated that zones with higher commercial and services density generate stronger attractiveness, especially among urban and young consumers.

More recently, Zhang et al. (2024) identified spatial clustering patterns in food retail and demonstrated that factors, such as population density, POI concentration and infrastructures accessibility are decisive in determining new store locations. Their results also showed that commercial concentration is positively associated with sustainability and network expansion, validating the interpretation of this component as a commercial and service concentration dimension.

4.3 Cluster Analysis

To further analyse and characterise the territorial context in Portugal, a cluster analysis was conducted. This technique was applied to identify spatial patterns and to group statistical sections with similar territorial and socioeconomic characteristics. Two analyses were performed: the first based on the percentile rankings of each principal component scores obtained from the PCA, and the second incorporating the distance from each section to the company's distribution centre, to assess spatial proximity and identify the most suitable areas for potential expansion.

In the first analysis, the Territorial Profile Cluster Analysis, a five-cluster solution was identified. This solution was selected based on both the interpretability and parsimony of the resulting territorial segments, as well as on an evolution of clustering solutions ranging from two to ten clusters using the ratio of between-cluster to total sum of squares (BSS/TSS).

The results show that the BSS/TSS ratio stabilises at five clusters, explaining approximately 65% of the total variance, as shown in the Appendix C, thereby supporting the choice of this clustering configuration. The K-Means algorithm was applied to the standardised component scores (Z-scores) to classify statistical sections into homogeneous territorial groups.

To guarantee the stability of the solution K-Means++ initialization method with 150 re-runs and Euclidean distance was used, ensuring a robust and stable clustering solution. The ratio of between to total sum of squares (0.648) indicates that approximately 65% of the total variance is explained by the differences between clusters, indicating a good level of separation and internal consistency across groups.

As shown in Appendix C, Cluster 1 represents areas of low urbanization (PFAC1_1 = 29.95) and low socioeconomic vitality (PFAC2_1 = 18.24), with only a modest commercial presence (PFAC3_1 = 46.48). Cluster 2 shows a high socioeconomic vitality and a moderate level of competition (PFAC2_1 = 67.27 and PFAC3_1 = 50.21) but low population and infrastructure density (PFAC1_1 = 19.46). Cluster 3 is characterised by high urban density (PFAC1_1 = 78.69) and an intermediate level of socioeconomic activity (PFAC2_1 = 51.22), yet a relatively weak commercial concentration (PFAC3_1 = 14.87). Cluster 4 reveals consistently high values across all components (PFAC1_1 = 65.06; PFAC2_1 = 82.79; PFAC3_1 = 68,08), representing developed, highly urbanised, and commercially active territories. Finally, Cluster 5 includes urban and commercially concentrated areas (PFAC1_1 = 71.97; PFAC3_1 = 83.12), but with lower socioeconomic vitality (PFAC2_1 = 30.22).

The spatial distribution of clusters, illustrated in Figure 4.7, reveals clear territorial patterns. Cluster 1 dominates the national territory, corresponding to low-density rural areas, whereas Clusters 4 and 5 are concentrated in major metropolitan and coastal regions, reflecting highly urbanised and commercially dynamic contexts. Intermediate clusters (C2 and C3) represent transition zones, combining moderate urban density with varying socioeconomic or commercial characteristics. Furthermore, Clusters 1 and 2 (2752.86 and 2442.07, respectively) exhibit greater internal heterogeneity due to their extensive spatial coverage. In contrast, Clusters 4 and 5 (1615.23 and 1429.58) are more homogeneous, representing well-defined urban and commercial profiles.

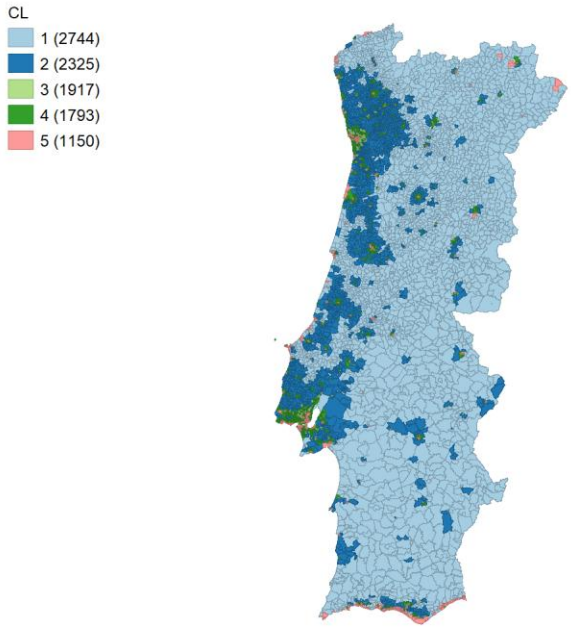


Figure 4.7: Map of Territory Profile Cluster Analysis

The second cluster analysis, the Proximity Cluster Analysis, followed the same methodological structure as the first one, but incorporated variables Flag_Store and Distance_C. Ten clusters were defined to represent ten proximity levels, capturing the distribution of statistical sections according to their proximity to the company's distribution centre. The main objective of this analysis was to delineate zones with similar logistical characteristics, thereby identifying the most accessible and operationally viable areas for potential expansion. This approach complements the previous analysis, shifting the focus from territorial and socioeconomic dimensions to spatial and operational accessibility.

The ratio of between to total sum of squares (0.9929) confirms the robustness of the solution, as shown in Appendix C, indicating an almost perfect spatial separation between the clusters. The

results reveal a clear concentric pattern centred on distribution hub. Cluster 10 corresponds to the sections that currently host stores, as identified by the Flag_Store variable, while Cluster 9 represents the nearest zones to the distribution centre without existing stores. From Cluster 1 to Cluster 10, the zones progressively approach the distribution centre, forming well-defined concentric rings that reflect gradual decrease in distance, as shown in Figure 4.8.

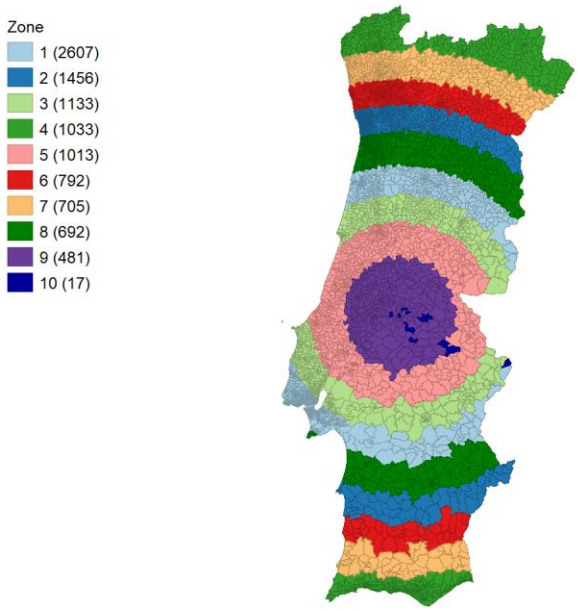


Figure 4.8: Map of Proximity Cluster Analysis

The first cluster analysis aligns with established evidence on spatial retail patterns (Beckers et al., 2022; Zhang et al., 2024), confirming that urban density, accessibility, and commercial concentration are key factors defining distinct territorial contexts for retail expansion. According to Formánek & Sokol (2022) and Beckers et al. (2022), this segmentation confirms that commercial density and urbanisation are determinant variables in territorial differentiation. Complementing the previous analysis with a distance-based cluster analysis between the centroid of each statistical section and the distribution centre is consistent with the principles of location-allocation models and competitive network optimisation (Mendes & Themido, 2004; Drezner & Eiselt, 2024), enabling the identification of logistically efficient expansion zones.

Both analyses provide a comprehensive understanding of Portugal's territorial dynamics. The first highlights market potential and the territorial diversity, while the second identifies

operationally feasible zones for network expansion. This dual perspective combines market opportunity with logistical efficiency, supporting a balanced approach to network growth.

4.4 Selection of Priority Clusters and Identification of Potential Locations

Following the analysis of both PCA and the Cluster Analysis results, integrating their findings is crucial to identify statistical sections that simultaneously combine favourable market conditions with logistical efficiency.

For the Territory Profile Cluster Analysis, Cluster 3 presents a balanced territorial profile, characterised by high urban density (PCA1), moderate socioeconomic vitality (PCA2) and low commercial concentration (PCA3). This combination reflects urban areas with well-developed infrastructure and population accessibility yet limited competitive saturation.

According to Khare (2020), urban zones with adequate infrastructure, but low commercial concentration, offer high growth potential while avoiding excessive competitive pressure. Similarly, Low et al. (2021) highlight that areas with strong accessibility and demographic vitality, even when less commercially agglomerated, tend to demonstrate greater long-term resilience. Hernández & Bennison (2000) also argue that under-served yet accessible urban markets represent strategic niches for store openings, balancing demand potential with reduced competition. Furthermore, Hunneman et al. (2023)confirm that population density and accessibility remain critical determinants of store performance, but excessive commercial concentration may dimmish profitability.

Although several authors (e.g., Beckers et al., 2022; Zhang et al., 2024) emphasises the advantages of locating a store in highly competitive and commercially dense urban environments, the choice of Cluster 3 represents a strategic deviation towards emerging urban markets.

For the Proximity Cluster Analysis, Cluster 9 corresponds to the statistical sections located near the company's distribution centre but without existing store, since Cluster 10 includes the sections where stores are already established. The objective is to identify locations that combine logistical efficiency with the opportunity for network expansion into new areas.

According to Mendes & Themido (2004)minimising the distance to the distribution centre is essential to reduce operational costs and increase overall network efficiency. Similarly, Drezner & Eiselt (2024) highlight that retail expansion strategies should prioritise areas not yet served by

existing stores to avoid market cannibalisation and ensure efficient coverage of the service area. This decision also aligns with the principles of Christaller's (1966) Central Place Theory, which postulates that proximity to the distribution centre enhances accessibility and operational performance.

Consequently, Territory Profile Cluster and Proximity Cluster merge as the most suitable group of statistical for expansion, combining favourable market characteristics with logistical efficiency.

To further refine the selection, the intersection of these clusters was analysed using Excel. A total of 37 statistical sections were identified as belonging simultaneously to both clusters.

To evaluate these sections, an analysis of their corresponding civil parishes was conducted considering three key indicators: total population, total number of competitors, area (Km^2) and population density. These variables were selected based on the literature, which consistently associates demographic vitality and competitive pressure with retail performance and market attractiveness.

From the selected statistical sections, fifteen civil parishes were analysed in detail, and a top 5 were identified based on their total population, as presented in Table 4.4. The geographical distribution of these top-ranking is illustrated in Figure 4.9.

Table 4.4: Analysis of Civil Parishes Selected

DTMNFR21	Total Population	Competitors	AreaKm2	Population Density	Civil Parish	Municipality
141633	7967	4	1.37	5821.97	União das Freguesias da Cidade de Santarém	Santarém
141002	6437	2	0.95	6774.55	Nossa Senhora de Fátima	Entroncamento
140301	6208	1	1.14	5453.95	Almeirim	Almeirim
141001	5459	3	3.59	1522.05	São João Baptista	Entroncamento
141921	2105	3	0.72	2908.18	União das Freguesias de Torres Novas (São Pedro), Lapas e Ribeira Branca	Torres Novas
140401	1916	1	0.89	2153.27	Alpiarça	Alpiarça
140303	1132	0	0.97	1168.52	Fazendas de Almeirim	Almeirim
141920	926	0	0.16	5682.20	União das Freguesias de Torres Novas (Santa Maria, Salvador e Santiago)	Torres Novas
101608	832	0	0.53	1568.60	Mira de Aire	Porto de Mós
141910	831	0	0.27	3065.67	Riachos	Torres Novas

DTMNFR21	Total Population	Competitors	AreaKm2	Population Density	Civil Parish	Municipality
140107	703	0	0.54	1299.69	Pego	Tomar
142115	681	0	0.09	7182.61	Nossa Senhora da Piedade	Ourém
141821	674	0	0.20	3363.67	União das Freguesias de São João do Batista e Santa Maria dos Olivais	Tomar
140124	618	0	0.75	827,47	União das Freguesias de São Miguel do Rio Torto e Rossio ao Sul do Tejo	Abrantes
051002	132	0	13.05	10,11	São João do Peso	Vila de Rei

Source: INE and OSM

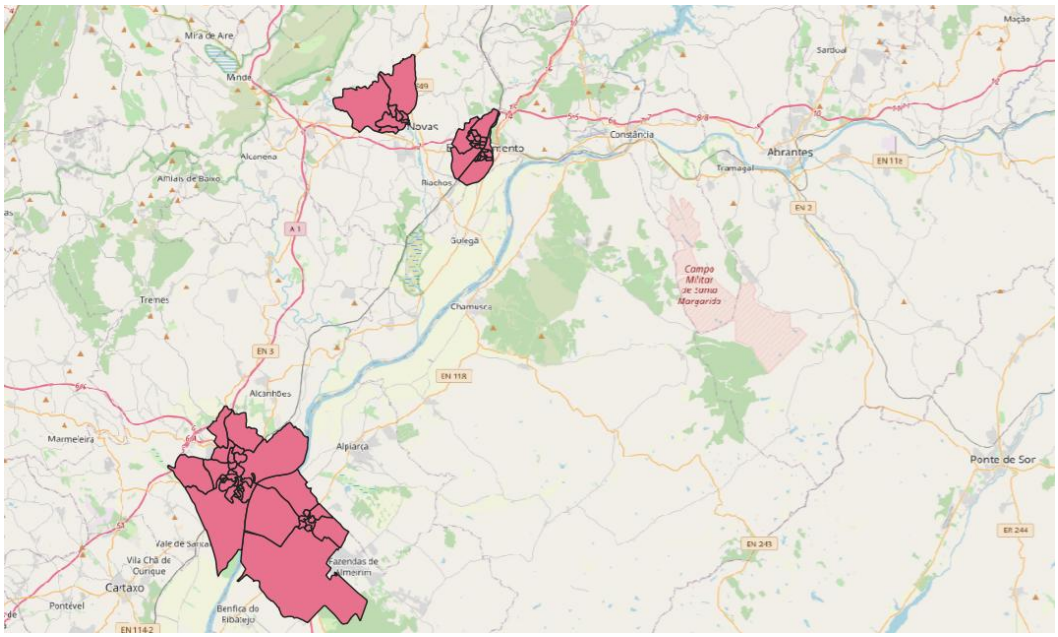


Figure 4.9: Geographical Distribution of the Civil Parishes Selected

To provide a clearer interpretation of these results, Table 4.5 presents the civil parishes where the seventeen stores of the company already operated.

Table 4.5: Analysis of Civil Parishes where Company Operates

DTMNFR21	Total Population	AreaKm2	Population Density	Company Revenue 2024	Civil Parish	Municipality
140120	7002	5.13	1364.61	1913122.41	União das Freguesias de Abrantes (São Vicente e São João)	Abrantes

121308	4411	24.91	177.05	1219201.51	União das Freguesias de Ponte de Sor, Tramaga e Vale de Açor	Ponte de Sor
120403	1818	36.81	49.38	373366.62	São João Batista	Campo Maior
140802	1239	12.81	96.69	252785.17	Montalvo	Constância
120906	990	60.14	16.46	486329.2	União das Freguesias de Gavião e Atalaia	Gavião
120101	883	130.53	6.76	323408.87	Alter do Chão	Alter do Chão
140123	702	45.90	15.30	114182.83	União das Freguesias de São Facundo e Vale das Mós	Abrantes
141703	647	9.92	65.23	305005.98	Sardoal	Sardoal
140106	610	6.44	94.73	216153.27	Mouriscas	Abrantes

Source: INE and OSM

The comparison between the identified potential locations and the existing store network shows that the selected parishes present a higher total population. Furthermore, the selected civil parishes have a balanced territorial condition. They are characterised by well-developed road networks, a moderate share of economically active population, and a potential commercial market with relatively low competitiveness.

5 Conclusion

5.1 Research Overview and Purpose

The main objective of this study was to identify the territorial and socioeconomic characteristics that define potential locations for the expansion of *Sabores do Ti Pereira, Lda*.

The research began with a systematic literature review, which made it possible to identify the key factors influencing retail location decisions and to understand how previous studies have approached this topic. Particular attention was given to the methodologies, findings, and limitations of prior investigations, which provided a theoretical foundation and informed the methodological design of the present study.

The empirical component combined quantitative and spatial analytical methods. The process started with a descriptive analysis of the company dataset, followed by a multivariate exploration of external data obtained from INE and OSM. PCA was applied to reduce data dimensionality and identify main territorial and socioeconomic dimensions, while Cluster Analysis was used to group statistical sections with similar characteristics.

Overall, this research contributes to a better understanding of how spatial and socioeconomic factors influence retail performance and provides practical insights for data-driven location planning. Moreover, it demonstrates how spatial analytical tolls can be applied to support strategic decision-making within small and medium-sized retail networks.

5.2 Summary of Key Findings

To determine the key territorial and socioeconomic characteristics influencing store location, a PCA and a Cluster Analysis were conducted.

The PCA identified three main components: the first (PCA1) related to urban density and infrastructure, the second (PCA2) to socioeconomic vitality, and the third (PCA3) to commercial and service concentration.

Following this analysis, two complementary cluster analyses were developed. The first defined five territorial profiles based on the percentile rankings of each principal component scores obtained from the PCA, while the second established ten spatial profiles based according to the distance from each statistical section to the company's distribution centre.

Based on the literature and the integration of both analyses, Cluster 3 (Territorial Profile Cluster) and Cluster 9 (Proximity Cluster) were selected as the most suitable profiles for expansion. A total of 37 statistical sections were identified as belonging simultaneously to both clusters, as presented in Appendix D. To further refine the selection, the corresponding civil parishes were evaluated based on their demographic characteristics and commercial agglomeration patterns, which let to the identification of five parishes as the most promising potential locations for future expansion.

The results confirmed the relevance of classical location theories (Christaller, 1966) as well as recent empirical evidence (Khare, 2020;Low et al., 2021;Zhang et al., 2024), demonstrating that accessibility, socioeconomic vitality, and balanced competition remain critical determinants of retail success.

5.3 Contributions

This research contributes to the theoretical and methodological understanding of retail location dynamics, while also providing practical and managerial implications that can inform real-world decision-making processes.

From a theoretical and scientific perspective, this study validates classical location theories, such as Christaller's (1966)Central Place Theory, by demonstrating that population density and accessibility remain key determinants of commercial attractiveness. Furthermore, the findings emphasise the relevance of socioeconomic vitality and balanced competition in the store location decision-making process. The analysis also revealed that the integration of social, demographic and infrastructural variables provides a more comprehensive understanding of the commercial environment and its spatial dynamics.

In addition, this research addresses an existing gap in literature by extending the study of retail geography to small and medium-sized enterprises (SMEs). It contextualises location theory within the Portuguese setting through the application of international analytical models to regional data, thereby offering empirical evidence that enhances the understanding of retail dynamics in Portugal, particularly within the bakery and pastry industry.

Regarding methodological contributions, this study demonstrated the potential of integrating quantitative and spatial analytical methods by combining PCA and Cluster Analysis with

Geographic Information Systems (GIS) tools such as GeoDa and QGIS. The use of the data from INE and OSM illustrates how open-access spatial datasets can effectively support strategic retail location decisions. Moreover, the development of this methodological framework represents a valuable contribution, as it can be replicated in other studies or adapted to different companies and contexts. The research also incorporated principles of logistical optimisation, highlighting the relevance of spatial proximity and distribution efficiency in retail expansion planning.

About practical and managerial contributions, the study provides the company with an evidence-based approach to decision-making, identifying seven priority statistical sections for potential expansion. These results reflect a strategic alignment between market opportunity and logistical efficiency, ensuring that future expansion decisions are both commercially and operationally viable. The methodological framework and analytical toolset developed in this research can be used for ongoing spatial planning, assisting not only *Sabores do Ti Pereira, Lda* but also other companies in selecting store locations. Furthermore, the study offers valuable insights for other Portuguese SMEs facing similar location challenges, promoting a more analytical and data-driven approach to retail expansion.

In conclusion, this research contributes to the theoretical, methodological and practical dimensions of retail location analysis, highlighting the importance of integrating company data and external data to support strategic, evidence-based decision-making.

5.4 Limitations and Future Research Recommendations

The present study provides relevant contributions to both society and the academic community. However, certain limitations must be acknowledged.

Regarding data, the research relied on information from the 2021 Census (INE) and OSM, which provide a static representation of territorial and socioeconomic conditions. This may limit the ability to capture recent demographic or infrastructural changes, Moreover, as Census data refer to 2021, some information may no longer fully reflect the current context. Similarly, OSM data, while highly valuable and accessible, may contain incomplete or outdated details concerning points of interest and road networks.

Additionally, the study did not include behavioural variables (such as consumer mobility or purchasing patterns) nor economic indicators (such as household income or expenditure levels),

which could provide a more comprehensive understanding of local market potential and demand dynamics.

From a methodological perspective, the reliance on an exploratory analytical approach constitutes a limitation, as the applied techniques, PCA and K-Means Cluster Analysis, allow the identification of spatial and statistical patterns but do not predict store performance or quantify causal relationships. Moreover, the number and composition of clusters may vary depending on the choice of parameter and distance measures, potentially affecting the stability and reproducibility of results. The research focused exclusively on quantitative indicators, without incorporating qualitative aspects such as consumer preferences or behavioural patterns, which could enrich the interpretation of spatial attractiveness and better reflect real decision-making processes.

From a contextual standpoint, the study was conducted within the Portuguese context, which may differ from other countries in demographic and infrastructural characteristics. Consequently, the results should not be generalised without careful consideration of regional specificities.

For future investigation, it is recommended to incorporate updated and diversified datasets, including economic and behavioural variables, to enhance the accuracy and explanatory power of model. The integration of predictive analytical techniques, such as spatial regression or machine learning models, could enable the forecasting of store performance and provide more reliable and actionable result. Furthermore, combining quantitative approaches with qualitative methods, such as surveys or interviews with consumers, would complement the statistical findings with insights into consumer perceptions and behaviours, offering a more comprehensive understanding of market attractiveness and brand positioning.

6 References

Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A., & Urquhart, R. (2022). Incorporating E-commerce into Retail Location Models. *Geographical Analysis*, *54*(2), 274–293. https://doi.org/10.1111/gean.12285

Christaller, W. (1966). *Central Places in Southern Germany* (C. W. Baskin, Trans.). Prentice-Hall, Inc.

Cró, S., & Martins, A. M. (2018). Hotel and hostel location in Lisbon: looking for their determinants. *Tourism Geographies*, *20*(3), 504–523. https://doi.org/10.1080/14616688.2017.1360386

Dock, J. P., Song, W., & Lu, J. (2015). Evaluation of dine-in restaurant location and competitiveness: Applications of gravity modeling in Jefferson County, Kentucky. *Applied Geography*, *60*, 204–209. https://doi.org/10.1016/j.apgeog.2014.11.008

Drezner, Z., & Eiselt, H. A. (2024). Competitive location models: A review. In *European Journal of Operational Research* (Vol. 316, Issue 1, pp. 5–18). Elsevier B.V. https://doi.org/10.1016/j.ejor.2023.10.030

D'Silva, K., Noulas, A., Musolesi, M., Mascolo, C., & Sklar, M. (2018). Predicting the temporal activity patterns of new venues. *EPJ Data Science*, *7*(1). https://doi.org/10.1140/epjds/s13688-018-0142-z

Finazzi, F. (2013). Geostatistical modeling in the presence of interaction between the measuring instruments, with an application to the estimation of spatial market potentials. *Annals of Applied Statistics*, *7*(1), 81–101. https://doi.org/10.1214/12-AOAS588

Formánek, T., & Sokol, O. (2022). Location effects: Geo-spatial and socio-demographic determinants of sales dynamics in brick-and-mortar retail stores. *Journal of Retailing and Consumer Services*, *66*. https://doi.org/10.1016/j.jretconser.2021.102902

Ghosh, A., & Craig, C. S. (1983). Formulating Retail Location Strategy in a Changing Environment. In *Source: Journal of Marketing* (Vol. 47, Issue 3). http://www.jstor.orgURL:http://www.jstor.org/stable/1251197http://www.jstor.org/stable/1251197?seq=1&cid=pdf-reference#references_tab_contents

Glaeser, C. K., Fisher, M., & Su, X. (2019). Optimal retail location: Empirical methodology and application to practice. *Manufacturing and Service Operations Management*, *21*(1), 86–102. https://doi.org/10.1287/msom.2018.0759

Guidon, S., Reck, D. J., & Axhausen, K. (2020). Expanding a(n) (electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests. *Journal of Transport Geography*, *84*. https://doi.org/10.1016/j.jtrangeo.2020.102692

Hernández, T., & Bennison, D. (2000). The art and science of retail location decisions. *International Journal of Retail & Distribution Management*, *28*(8), 357–367. https://doi.org/10.1108/09590550010337391

Horton, F. E. (1968). Location Factors as Determinants of Consumer Attraction to Retail Firms. In *Source: Annals of the Association of American Geographers* (Vol. 58, Issue 4).

Hotelling, H. (1929). Stability in Competition. In *Source: The Economic Journal* (Vol. 39, Issue 153). http://www.jstor.orgURL:http://www.jstor.org/stable/2224214

Hsieh, H. P., Lin, F., Li, C. Te, Yen, I. E. H., & Chen, H. Y. (2019). Temporal popularity prediction of locations for geographical placement of retail stores. *Knowledge and Information Systems*, *60*(1), 247–273. https://doi.org/10.1007/s10115-018-1311-x

Hunneman, A., Bijmolt, T. H. A., & Elhorst, J. P. (2023). Evaluating store location and department composition based on spatial heterogeneity in sales potential. *Journal of Retailing and Consumer Services*, *73*. https://doi.org/10.1016/j.jretconser.2023.103355

Khare, A. (2020). Location and agglomeration factors predicting retailers' preference for Indian malls. *Journal of Marketing Analytics*, *8*(4), 245–266. https://doi.org/10.1057/s41270-020-00075-7

Lee, J. H., Jo, J., Kim, J. W., Lee, K., & Choi, M. Y. (2022). Spatial distributions of restaurants emerging from pedestrian behavior and online information sharing. *Physica A: Statistical Mechanics and Its Applications*, *597*. https://doi.org/10.1016/j.physa.2022.127265

Low, S. A., Bass, M., Thilmany, D., & Castillo, M. (2021). Local Foods Go Downstream: Exploring the Spatial Factors Driving U.S. Food Manufacturing. *Applied Economic Perspectives and Policy*, *43*(3), 896–915. https://doi.org/10.1002/aepp.13046

Mahmud, N., & Habib, M. A. (2024). A Comprehensive Business Location Choice Model Leveraging Machine Learning in Systematic Choice Set. *Transportation Research Record*, *2678*(12), 1856–1871. https://doi.org/10.1177/03611981241253609

Mendes, A. B., & Themido, I. H. (2004). *Multi-outlet retail site location assessment 1*.

Newing, A., Clarke, G., & Clarke, M. (2018). Applied spatial modelling for retail planning in tourist resorts. *International Journal of Retail and Distribution Management*, *46*(11–12), 1117–1132. https://doi.org/10.1108/IJRDM-10-2017-0253

Oppewal, H., & Holyoake, B. (2004). Bundling and retail agglomeration effects on shopping behavior. In *Journal of Retailing and Consumer Services* (Vol. 11).

Organisation for Economic Co-operation and Development (OECD). (2020). *E-commerce in the times of COVID-19*.

Organisation for Economic Co-operation and Development (OECD). (2021). *The role of online platforms in weathering the COVID-19 shock*. www.cnbc.com/2020/09/04/disruptors-paypal-and-square-surpass-wall-

Ouyang, J., Fan, H., Wang, L., Yang, M., & Ma, Y. (2020). Site selection improvement of retailers based on spatial competition strategy and a double-channel convolutional neural network. *ISPRS International Journal of Geo-Information*, *9*(6). https://doi.org/10.3390/ijgi9060357

Pillsbury, R. (1987). THE PROFESSIONAL GEOGRAPHER FROM HAMBURGER ALLEY TO HEDGEROSE HEIGHTS: TOWARD A MODEL OF RESTAURANT LOCATION DYNAMICS. In *Professional Geographer* (Vol. 39, Issue 3).

Robinson, D. T., & Balulescu, A. M. (2018). Comparison of methods for quantifying consumer spending on retail using publicly available data. *International Journal of Geographical Information Science*, *32*(6), 1061–1086. https://doi.org/10.1080/13658816.2018.1423684

Sakai, T., Beziat, A., & Heitz, A. (2020). Location factors for logistics facilities: Location choice modeling considering activity categories. *Journal of Transport Geography*, *85*. https://doi.org/10.1016/j.jtrangeo.2020.102710

Schuetz, J., Kolko, J., & Meltzer, R. (2012). Are poor neighborhoods retail deserts? *Regional Science and Urban Economics*, *42*(1–2), 269–285. https://doi.org/10.1016/j.regsciurbeco.2011.09.005

Shim, S., & Eastlick, M. A. (1998). *The Hierarchical Influence of Personal Values on Mall Shopping Attitute and Behavior*.

Suárez-Vega, R., Gutiérrez-Acuña, J. L., & Rodríguez-Díaz, M. (2015). Locating a supermarket using a locally calibrated Huff model. *International Journal of Geographical Information Science*, *29*(2), 217–233. https://doi.org/10.1080/13658816.2014.958154

Teller, C., Alexander, A., & Floh, A. (2016). The impact of competition and cooperation on the performance of a retail agglomeration and its stores. *Industrial Marketing Management*, *52*, 6–17. https://doi.org/10.1016/j.indmarman.2015.07.010

Teller, C., & Thomson, J. A. (2012). Gender differences of shoppers in the marketing and management of retail agglomerations. *Service Industries Journal*, *32*(6), 961–980. https://doi.org/10.1080/02642069.2011.559725

Thomadsen, R. (2007). Product positioning and competition: The role of location in the fast food industry. *Marketing Science*, *26*(6), 792–804. https://doi.org/10.1287/mksc.1070.0296

Waddington, T. B. P., Clarke, G. P., Clarke, M., & Newing, A. (2018). Open all hours: spatiotemporal fluctuations in U.K. grocery store sales and catchment area demand. *International Review of Retail, Distribution and Consumer Research*, *28*(1), 1–26. https://doi.org/10.1080/09593969.2017.1333966

Waddington, T., Clarke, G., Clarke, M. C., Hood, N., & Newing, A. (2019). Accounting for Temporal Demand Variations in Retail Location Models. *Geographical Analysis*, *51*(4), 426–447. https://doi.org/10.1111/gean.12179

Wang, F., Chen, C., Xiu, C., & Zhang, P. (2014). Location analysis of retail stores in Changchun, China: A street centrality perspective. *Cities*, *41*(PA), 54–63. https://doi.org/10.1016/j.cities.2014.05.005

Wang, L., Fan, H., & Wang, Y. (2018). Site selection of retail shops based on spatial accessibility and hybrid BP neural network. *ISPRS International Journal of Geo-Information*, *7*(6). https://doi.org/10.3390/ijgi7060202

Wojtyra, B., Kossowski, T. M., Březinová, M., Savov, R., & Lančarič, D. (2020). Geography of craft breweries in Central Europe: Location factors and the spatial dependence effect. *Applied Geography*, *124*. https://doi.org/10.1016/j.apgeog.2020.102325

Yan, R. N., & Eckman, M. (2009). Are lifestyle centres unique? Consumers' perceptions across locations. *International Journal of Retail and Distribution Management*, *37*(1), 24–42. https://doi.org/10.1108/09590550910927144

Zhang, E., Zhou, Y., Chen, G., & Wang, G. (2024). Classified Spatial Clustering and Influencing Factors of New Retail Stores: A Case Study of Freshippo in Shanghai. *Sustainability (Switzerland)*, *16*(15). https://doi.org/10.3390/su16156643

Zhu, T., & Singh, V. (2009). Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics*, *7*(1), 1–35. https://doi.org/10.1007/s11129-008-9048-6

7 Appendix

A: Data Dictionary

Table 7.1: Store Level Variables from Company, INE, and OpenStreetMap Data

Variable Name	Source	Type	Unit	Description
ID	Company Records	Discrete	-	Unique identifier for each store.
Store	Company Records	Categorical	-	Name or label of the store.
Opening Month	Company Records	Categorical	-	Month in which the store opened.
Opening Year	Company Records	Discrete	-	Year in which the store opened.
Address	Company Records	Text	-	Address of the store.
Postal Code	Company Records	Text	-	Postal code of the store location.
Municipality	Company Records	Categorical	-	Municipality where the store is located.
Latitude	Google Maps	Continuous	Decimal degrees	Latitude coordinate of the store.
Longitude	Google Maps	Continuous	Decimal degrees	Longitude coordinate of the store.
Number of Opening Hours per Week	Company Records	Discrete	Hours	Total number of opening hours per week.
Number of Opening Days per Week	Company Records	Discrete	Days	Total number of opening days per week.
Sales Area (m²)	Company Records	Continuous	Metres	Area of the sales section available to customers.
Storage Area (m²)	Company Records	Continuous	Metres	Area dedicated to storage.
Service Area (m²)	Company Records	Continuous	Metres	Area for customer service and interaction.
Production Area (m²)	Company Records	Continuous	Metres	Area used for production activities.
Kitchen Area (m²)	Company Records	Continuous	Metres	Area used for kitchen purposes.
Circulation Area (m²)	Company Records	Continuous	Metres	Space used for circulation within the store.
Toilet Area (m²)	Company Records	Continuous	Metres	Area of restrooms used by customers and staff.
Utility / Storage Room Area (m²)	Company Records	Continuous	Metres	Utility or auxiliary storage room area.
Counter Area (m²)	Company Records	Continuous	Metres	Conter or service desk area.
Average Number of Direct Employees per Day	Company Records	Discrete	-	Average number of direct employees working per day.

Variable Name	Source	Type	Unit	Description
Average Number of Indirect Employees per Day	Company Records	Discrete	-	Average number of indirect employees working per day.
Meals?	Company Records	Binary	Yes/No	Indicates whether the store serves meals.
Kitchen?	Company Records	Binary	Yes/No	Indicates whether the store has a kitchen.
Own Production?	Company Records	Binary	Yes/No	Indicates whether the store produces its own goods.
Parking?	Company Records	Binary	Yes/No	Indicates whether the store has parking available.
Renovated Space?	Company Records	Binary	Yes/No	Indicates whether the store was renovated.
Outdoor Seating	Company Records	Binary	Yes/No	Indicates whether the store has outdoor seating available.
Number of Indoor Tables	Company Records	Discrete	-	Total number of indoor tables.
Number of Indoor Chairs	Company Records	Discrete	-	Total number of indoor chairs.
Number of Indoor Sofas	Company Records	Discrete	-	Total number of indoor sofas.
Number of Outdoor Tables	Company Records	Discrete	-	Total number of outdoor tables.
Number of Outdoor Chairs	Company Records	Discrete	-	Total number of outdoor chairs
Revenue 2023	Company Records	Continuous	Euros	Total store revenue in 2023.
Revenue 2024	Company Records	Continuous	Euros	Total store revenue in 2024.
Revenue Variation	Company Records	Continuous	%	Percentage change in revenue between 2023 and 2024.
Revenue – Bakery (2023)	Company Records	Continuous	Euros	Revenue from bakery products in 2023.
Revenue - Small Pastry (2023)	Company Records	Continuous	Euros	Revenue from small pastry items in 2023.
Revenue - Large Pastry (2023)	Company Records	Continuous	Euros	Revenue from large pastry items in 2023.
Revenue - Beverages (2023)	Company Records	Continuous	Euros	Revenue from beverages in 2023.
Revenue - Coffee Shop (2023)	Company Records	Continuous	Euros	Revenue from coffee shop sales in 2023.
Revenue - Meals (2023)	Company Records	Continuous	Euros	Revenue from meals in 2023.
Revenue - Processed Products (2023)	Company Records	Continuous	Euros	Revenue from processed products in 2023.
Revenue - Bakery (2024)	Company Records	Continuous	Euros	Revenue from bakery products in 2024.
Revenue - Small Pastry (2024)	Company Records	Continuous	Euros	Revenue from small pastry items in 2024.
Revenue - Large Pastry (2024)	Company Records	Continuous	Euros	Revenue from large pastry items in 2024.
Revenue - Beverages (2024)	Company Records	Continuous	Euros	Revenue from beverages in 2024.
Revenue - Coffee Shop (2024)	Company Records	Continuous	Euros	Revenue from coffee shop sales in 2024.
Revenue - Meals (2024)	Company Records	Continuous	Euros	Revenue from meals in 2024.
Revenue - Processed Products (2024)	Company Records	Continuous	Euros	Revenue from processed products in 2024.
Number of Issued Documents (2023)	Company Records	Discrete	-	Number of issued sales documents in 2023.
Number of Issued Documents (2024)	Company Records	Discrete	-	Number of issued sales documents in 2024.

Variable Name	Source	Type	Unit	Description
Length of Residential Roads (200m)	OMS	Continuous	Metres	Total length of Residential Roads within the defined buffer.
Length of Primary Roads (200m)	OMS	Continuous	Metres	Total length of Primary Roads within the defined buffer.
Length of Footways (200m)	OMS	Continuous	Metres	Total length of Footways within the defined buffer.
Length of Pedestrian Roads (200m)	OMS	Continuous	Metres	Total length of Pedestrian Roads within the defined buffer.
Length of Tertiary Roads (200m)	OMS	Continuous	Metres	Total length of Tertiary Roads within the defined buffer.
Length of Living Streets (200m)	OMS	Continuous	Metres	Total length of Living Streets within the defined buffer.
Length of Secondary Roads (200m)	OMS	Continuous	Metres	Total length of Secondary Roads within the defined buffer.
Length of Cycleways (200m)	OMS	Continuous	Metres	Total length of Cycleways within the defined buffer.
Total Roads Length (200m)	OMS	Continuous	Metres	Sum of total length of all roads within the defined buffer.
Number of Conventional Buildings (200m)	INE	Discrete	-	Number of Conventional Buildings located within the buffer.
Number of Exclusively Residential Buildings	INE	Discrete	-	Number of Exclusively Residential Buildings located within the buffer.
Total Number of Dwellings (200m)	INE	Discrete	-	Number of Dwellings located within the buffer.
Number of Family Dwellings (200m)	INE	Discrete	-	Number of Family Dwellings located within the buffer.
Share of Main Residences (200m)	INE	Continuous	%	Proportion of Main Residences relative to the total Family Dwellings within the
Share of Vacant or Secondary Residences	INE	Continuous	%	Proportion of Vacant or Secondary Residences relative to the total Family
Total Population (200m)	INE	Discrete	-	Total population within the buffer.
Share of Male Population (200m)	INE	Continuous	%	Proportion of Male Population relative to the total within the buffer.
Share of Female Population (200m)	INE	Continuous	%	Proportion of Female Population relative to the total within the buffer.
Share of Population Aged 0–14 (200m)	INE	Continuous	%	Proportion of Population Aged 0–14 relative to the total within the buffer.
Share of Population Aged 15–24 (200m)	INE	Continuous	%	Proportion of Population Aged 15–24 relative to the total within the buffer.
Share of Population Aged 25–64 (200m)	INE	Continuous	%	Proportion of Population Aged 25–64 relative to the total within the buffer.
Share of Population Aged 65 or Older (200m)	INE	Continuous	%	Proportion of Population Aged 65 or Older relative to the total within the buffer.
Share of Population with No Educational Level	INE	Continuous	%	Proportion of Population with No Educational Level Completed relative to the
Share of Population with Primary Education	INE	Continuous	%	Proportion of Population with Primary Education Completed – 1st Cycle
Share of Population with Primary Education	INE	Continuous	%	Proportion of Population with Primary Education Completed – 2nd Cycle
Share of Population with Lower Secondary	INE	Continuous	%	Proportion of Population with Lower Secondary Education Completed – 3rd
Share of Population with Upper Secondary or	INE	Continuous	%	Proportion of Population with Upper Secondary or Post-secondary Education
Share of Population with Higher Education	INE	Continuous	%	Proportion of Population with Higher Education Completed relative to the total
Share of Economically Active Population	INE	Continuous	%	Proportion of Economically Active Population relative to the total within the
Share of Economically Inactive Population	INE	Continuous	%	Proportion of Economically Inactive Population relative to the total within the
Share of Employed Population (200m)	INE	Continuous	%	Proportion of Employed Population relative to the total within the buffer.

Variable Name	Source	Type	Unit	Description
Share of Students (200m)	INE	Continuous	%	Proportion of Students relative to the total within the buffer.
Share of Homemakers (200m)	INE	Continuous	%	Proportion of Homemakers relative to the total within the buffer.
Share of Retired Population (200m)	INE	Continuous	%	Proportion of Retired Population relative to the total within the buffer.
Share of Employed Population in the Primary	INE	Continuous	%	Proportion of Employed Population in the Primary Sector relative to the total
Share of Employed Population in the Secondary	INE	Continuous	%	Proportion of Employed Population in the Secondary Sector relative to the total
Share of Employed Population in the Tertiary	INE	Continuous	%	Proportion of Employed Population in the Tertiary Sector relative to the total
Share of Foreign Nationals (200m)	INE	Continuous	%	Proportion of Foreign Nationals relative to the total within the buffer.
Number of Supermarkets (200m)	OMS	Discrete	-	Number of Supermarkets located within the buffer.
Number of Pitches (200m)	OMS	Discrete	-	Number of Pitches located within the buffer.
Number of Playgrounds (200m)	OMS	Discrete	-	Number of Playgrounds located within the buffer.
Number of Fountains (200m)	OMS	Discrete	-	Number of Fountains located within the buffer.
Number of Schools (200m)	OMS	Discrete	-	Number of Schools located within the buffer.
Number of Veterinary Clinics (200m)	OMS	Discrete	-	Number of Veterinary Clinics located within the buffer.
Number of Stadiums (200m)	OMS	Discrete	-	Number of Stadiums located within the buffer.
Number of Libraries (200m)	OMS	Discrete	-	Number of Libraries located within the buffer.
Number of Cafés (200m)	OMS	Discrete	-	Number of Cafés located within the buffer.
Number of Museums (200m)	OMS	Discrete	-	Number of Museums located within the buffer.
Number of Bars (200m)	OMS	Discrete	-	Number of Bars located within the buffer.
Number of Arts Centres (200m)	OMS	Discrete	-	Number of Arts Centres located within the buffer.
Number of Convenience Stores (200m)	OMS	Discrete	-	Number of Convenience Stores located within the buffer.
Number of Restaurants (200m)	OMS	Discrete	-	Number of Restaurants located within the buffer.
Number of Pubs (200m)	OMS	Discrete	-	Number of Pubs located within the buffer.
Number of Car Dealerships (200m)	OMS	Discrete	-	Number of Car Dealerships located within the buffer.
Number of Beverage Shops (200m)	OMS	Discrete	-	Number of Beverage Shops located within the buffer.
Number of Kindergartens (200m)	OMS	Discrete	-	Number of Kindergartens located within the buffer.
Number of Parks (200m)	OMS	Discrete	-	Number of Parks located within the buffer.
Number of Sports Centres (200m)	OMS	Discrete	-	Number of Sports Centres located within the buffer.
Number of Clinics (200m)	OMS	Discrete	-	Number of Clinics located within the buffer.
Number of ATMs (200m)	OMS	Discrete	-	Number of ATMs located within the buffer.
Number of Bookshops (200m)	OMS	Discrete	-	Number of Bookshops located within the buffer.
Number of Fast Food Restaurants (200m)	OMS	Discrete	-	Number of Fast Food Restaurants located within the buffer.

Variable Name	Source	Type	Unit	Description
Number of Pharmacies (200m)	OMS	Discrete	-	Number of Pharmacies located within the buffer.
Number of Town Halls (200m)	OMS	Discrete	-	Number of Town Halls located within the buffer.
Number of Stationery Shops (200m)	OMS	Discrete	-	Number of Stationery Shops located within the buffer.
Number of DIY Stores (200m)	OMS	Discrete	-	Number of DIY Stores located within the buffer.
Number of Attractions (200m)	OMS	Discrete	-	Number of Attractions located within the buffer.
Number of Tourist Information Shops (200m)	OMS	Discrete	-	Number of Tourist Information Shops located within the buffer.
Number of Artworks (200m)	OMS	Discrete	-	Number of Artworks located within the buffer.
Number of Water Towers (200m)	OMS	Discrete	-	Number of Water Towers located within the buffer.
Number of Marketplaces (200m)	OMS	Discrete	-	Number of Marketplaces located within the buffer.
Number of Universities (200m)	OMS	Discrete	-	Number of Universities located within the buffer.
Number of Banks (200m)	OMS	Discrete	-	Number of Banks located within the buffer.
Number of Post Offices (200m)	OMS	Discrete	-	Number of Post Offices located within the buffer.
Number of General Stores (200m)	OMS	Discrete	-	Number of General Stores located within the buffer.
Number of Swimming Pools (200m)	OMS	Discrete	-	Number of Swimming Pools located within the buffer.
Number of Bakeries (200m)	OMS	Discrete	-	Number of Bakeries located within the buffer.
Number of Toilets (200m)	OMS	Discrete	-	Number of Toilets located within the buffer.
Number of Benches (200m)	OMS	Discrete	-	Number of Benches located within the buffer.
Number of Post Boxes (200m)	OMS	Discrete	-	Number of Post Boxes located within the buffer.
Number of Kiosks (200m)	OMS	Discrete	-	Number of Kiosks located within the buffer.
Length of Residential Roads (400m)	OMS	Continuous	Metres	Total length of Residential Roads within the defined buffer.
Length of Primary Roads (400m)	OMS	Continuous	Metres	Total length of Primary Roads within the defined buffer.
Length of Footways (400m)	OMS	Continuous	Metres	Total length of Footways within the defined buffer.
Length of Pedestrian Roads (400m)	OMS	Continuous	Metres	Total length of Pedestrian Roads within the defined buffer.
Length of Tertiary Roads (400m)	OMS	Continuous	Metres	Total length of Tertiary Roads within the defined buffer.
Length of Living Streets (400m)	OMS	Continuous	Metres	Total length of Living Streets within the defined buffer.
Length of Secondary Roads (400m)	OMS	Continuous	Metres	Total length of Secondary Roads within the defined buffer.
Length of Cycleways (400m)	OMS	Continuous	Metres	Total length of Cycleways within the defined buffer.
Total Roads Length (400m)	OMS	Continuous	Metres	Sum of total length of all roads within the defined buffer.
Number of Conventional Buildings (400m)	INE	Discrete	-	Number of Conventional Buildings located within the buffer.
Number of Exclusively Residential Buildings	INE	Discrete	-	Number of Exclusively Residential Buildings located within the buffer.
Total Number of Dwellings (400m)	INE	Discrete	-	Number of Dwellings located within the buffer.

Variable Name	Source	Type	Unit	Description
Number of Family Dwellings (400m)	INE	Discrete	-	Number of Family Dwellings located within the buffer.
Share of Main Residences (400m)	INE	Continuous	%	Proportion of Main Residences relative to the total of Family Dwellings within
Share of Vacant or Secondary Residences	INE	Continuous	%	Proportion of Vacant or Secondary Residences relative to the total of Family
Total Population (400m)	INE	Discrete	-	Total population within the buffer.
Share of Male Population (400m)	INE	Continuous	%	Proportion of Male Population relative to the total within the buffer.
Share of Female Population (400m)	INE	Continuous	%	Proportion of Female Population relative to the total within the buffer.
Share of Population Aged 0–14 (400m)	INE	Continuous	%	Proportion of Population Aged 0–14 relative to the total within the buffer.
Share of Population Aged 15–24 (400m)	INE	Continuous	%	Proportion of Population Aged 15–24 relative to the total within the buffer.
Share of Population Aged 25–64 (400m)	INE	Continuous	%	Proportion of Population Aged 25–64 relative to the total within the buffer.
Share of Population Aged 65 or Older (400m)	INE	Continuous	%	Proportion of Population Aged 65 or Older relative to the total within the buffer.
Share of Population with No Educational Level	INE	Continuous	%	Proportion of Population with No Educational Level Completed relative to the
Share of Population with Primary Education	INE	Continuous	%	Proportion of Population with Primary Education Completed – 1st Cycle
Share of Population with Primary Education	INE	Continuous	%	Proportion of Population with Primary Education Completed – 2nd Cycle
Share of Population with Lower Secondary	INE	Continuous	%	Proportion of Population with Lower Secondary Education Completed – 3rd
Share of Population with Upper Secondary or	INE	Continuous	%	Proportion of Population with Upper Secondary or Post-secondary Education
Share of Population with Higher Education	INE	Continuous	%	Proportion of Population with Higher Education Completed relative to the total
Share of Economically Active Population	INE	Continuous	%	Proportion of Economically Active Population relative to the total within the
Share of Economically Inactive Population %	INE	Continuous	%	Proportion of Economically Inactive Population relative to the total within the
Share of Employed Population (400m)	INE	Continuous	%	Proportion of Employed Population relative to the total within the buffer.
Share of Students (400m)	INE	Continuous	%	Proportion of Students relative to the total within the buffer.
Share of Homemakers (400m)	INE	Continuous	%	Proportion of Homemakers relative to the total within the buffer.
Share of Retired Population (400m)	INE	Continuous	%	Proportion of Retired Population relative to the total within the buffer.
Share of Employed Population in the Primary	INE	Continuous	%	Proportion of Employed Population in the Primary Sector relative to the total
Share of Employed Population in the Secondary	INE	Continuous	%	Proportion of Employed Population in the Secondary Sector relative to the total
Share of Employed Population in the Tertiary	INE	Continuous	%	Proportion of Employed Population in the Tertiary Sector relative to the total
Share of Foreign Nationals (400m)	INE	Continuous	%	Proportion of Foreign Nationals relative to the total within the buffer.
Number of Supermarkets (400m)	OSM	Discrete	-	Number of Supermarkets located within the buffer.
Number of Pitches (400m)	OSM	Discrete	-	Number of Pitches located within the buffer.
Number of Police Stations (400m)	OSM	Discrete	-	Number of Police Stations located within the buffer.
Number of Playgrounds (400m)	OSM	Discrete	-	Number of Playgrounds located within the buffer.
Number of DIY Stores (400m)	OSM	Discrete	-	Number of DIY Stores located within the buffer.

Variable Name	Source	Type	Unit	Description
Number of Department Stores (400m)	OSM	Discrete	-	Number of Department Stores located within the buffer.
Number of Fountains (400m)	OSM	Discrete	-	Number of Fountains located within the buffer.
Number of Fire Stations (400m)	OSM	Discrete	-	Number of Fire Stations located within the buffer.
Number of Kindergartens (400m)	OSM	Discrete	-	Number of Kindergartens located within the buffer.
Number of Schools (400m)	OSM	Discrete	-	Number of Schools located within the buffer.
Number of Pharmacies (400m)	OSM	Discrete	-	Number of Pharmacies located within the buffer.
Number of Car Dealerships (400m)	OSM	Discrete	-	Number of Car Dealerships located within the buffer.
Number of Restaurants (400m)	OSM	Discrete	-	Number of Restaurants located within the buffer.
Number of Veterinary Clinics (400m)	OSM	Discrete	-	Number of Veterinary Clinics located within the buffer.
Number of Clothes Shops (400m)	OSM	Discrete	-	Number of Clothes Shops located within the buffer.
Number of Sports Centres (400m)	OSM	Discrete	-	Number of Sports Centres located within the buffer.
Number of Town Halls (400m)	OSM	Discrete	-	Number of Town Halls located within the buffer.
Number of Attractions (400m)	OSM	Discrete	-	Number of Attractions located within the buffer.
Number of Stadiums (400m)	OSM	Discrete	-	Number of Stadiums located within the buffer.
Number of Libraries (400m)	OSM	Discrete	-	Number of Libraries located within the buffer.
Number of Parks (400m)	OSM	Discrete	-	Number of Parks located within the buffer.
Number of Picnic Sites (400m)	OSM	Discrete	-	Number of Picnic Sites located within the buffer.
Number of Tracks (400m)	OSM	Discrete	-	Number of Tracks located within the buffer.
Number of Caravan Sites (400m)	OSM	Discrete	-	Number of Caravan Sites located within the buffer.
Number of Bars (400m)	OSM	Discrete	-	Number of Bars located within the buffer.
Number of Cafés (400m)	OSM	Discrete	-	Number of Cafés located within the buffer.
Number of Hotels (400m)	OSM	Discrete	-	Number of Hotels located within the buffer.
Number of Museums (400m)	OSM	Discrete	-	Number of Museums located within the buffer.
Number of Pubs (400m)	OSM	Discrete	-	Number of Pubs located within the buffer.
Number of Arts Centres (400m)	OSM	Discrete	-	Number of Arts Centres located within the buffer.
Number of Fast Food Restaurants (400m)	OSM	Discrete	-	Number of Fast Food Restaurants located within the buffer.
Number of Public Buildings (400m)	OSM	Discrete	-	Number of Public Buildings located within the buffer.
Number of Convenience Stores (400m)	OSM	Discrete	-	Number of Convenience Stores located within the buffer.
Number of Beverage Shops (400m)	OSM	Discrete	-	Number of Beverage Shops located within the buffer.
Number of Clinics (400m)	OSM	Discrete	-	Number of Clinics located within the buffer.
Number of ATMs (400m)	OSM	Discrete	-	Number of ATMs located within the buffer.

Variable Name	Source	Type	Unit	Description
Number of Bookshops (400m)	OSM	Discrete	-	Number of Bookshops located within the buffer.
Number of Stationery Shops (400m)	OSM	Discrete	-	Number of Stationery Shops located within the buffer.
Number of Swimming Pools (400m)	OSM	Discrete	-	Number of Swimming Pools located within the buffer.
Number of Drinking Water Points (400m)	OSM	Discrete	-	Number of Drinking Water Points located within the buffer.
Number of Car Washes (400m)	OSM	Discrete	-	Number of Car Washes located within the buffer.
Number of Tourist Information Points (400m)	OSM	Discrete	-	Number of Tourist Information Points located within the buffer.
Number of Artworks (400m)	OSM	Discrete	-	Number of Artworks located within the buffer.
Number of Community Centres (400m)	OSM	Discrete	-	Number of Community Centres located within the buffer.
Number of Theatres (400m)	OSM	Discrete	-	Number of Theatres located within the buffer.
Number of Water Towers (400m)	OSM	Discrete	-	Number of Water Towers located within the buffer.
Number of Recycling Glass Points (400m)	OSM	Discrete	-	Number of Recycling Glass Points located within the buffer.
Number of Castles (400m)	OSM	Discrete	-	Number of Castles located within the buffer.
Number of Marketplaces (400m)	OSM	Discrete	-	Number of Marketplaces located within the buffer.
Number of Universities (400m)	OSM	Discrete	-	Number of Universities located within the buffer.
Number of Courthouses (400m)	OSM	Discrete	-	Number of Courthouses located within the buffer.
Number of Banks (400m)	OSM	Discrete	-	Number of Banks located within the buffer.
Number of Post Offices (400m)	OSM	Discrete	-	Number of Post Offices located within the buffer.
Number of General Shops (400m)	OSM	Discrete	-	Number of General Shops located within the buffer.
Number of Viewpoints (400m)	OSM	Discrete	-	Number of Viewpoints located within the buffer.
Number of Bakeries (400m)	OSM	Discrete	-	Number of Bakeries located within the buffer.
Number of Dentists (400m)	OSM	Discrete	-	Number of Dentists located within the buffer.
Number of Toilets (400m)	OSM	Discrete	-	Number of Toilets located within the buffer.
Number of Benches (400m)	OSM	Discrete	-	Number of Benches located within the buffer.
Number of Post Boxes (400m)	OSM	Discrete	-	Number of Post Boxes located within the buffer.
Number of Graveyards (400m)	OSM	Discrete	-	Number of Graveyards located within the buffer.
Number of Cinemas (400m)	OSM	Discrete	-	Number of Cinemas located within the buffer.
Number of Opticians (400m)	OSM	Discrete	-	Number of Opticians located within the buffer.
Number of Kiosks (400m)	OSM	Discrete	-	Number of Kiosks located within the buffer.
Number of Florists (400m)	OSM	Discrete	-	Number of Florists located within the buffer.
Number of Hairdressers (400m)	OSM	Discrete	-	Number of Hairdressers located within the buffer.
Number of Greengrocers (400m)	OSM	Discrete	-	Number of Greengrocers located within the buffer.

Variable Name	Source	Type	Unit	Description
Number of Jewellers (400m)	OSM	Discrete	-	Number of Jewellers located within the buffer.
Number of Garden Centres (400m)	OSM	Discrete	-	Number of Garden Centres located within the buffer.
Distance to the nearest competitor	OSM	Continuous	Metres	Distance from each store to the nearest competitor (bakery or café).

Table 7.2: Statistical Section Level Variables from INE and OpenStreetMap Data

Variable Name	Source	Type	Unit	Description
fid	INE	Discrete	-	Unique identifier for the statistical section.
DTMNFRSEC21	INE	String	-	Official INE code for the statistical section (Census 2021).
NumberofConventionalBuildings	INE	Discrete	-	Number of conventional buildings in the section.
ShareofBuildingswith1to2Floors	INE	Continuous	%	Derived indicator of local environmental or socioeconomic characteristics.
ShareofBuildingswith3to4Floors	INE	Continuous	%	Derived indicator of local environmental or socioeconomic characteristics.
ShareofBuildingswith5ormoreFloors	INE	Continuous	%	Derived indicator of local environmental or socioeconomic characteristics.
NumberofExclusivelyResidentialBuildings	INE	Discrete	-	Number of buildings used exclusively for residential purposes.
TotalNumberofDwellings	INE	Discrete	-	Number of dwellings in the section.
ShareofMainResidences	INE	Continuous	%	Proportion of dwellings occupied as main residences.
ShareofVacantorSecondaryResidences	INE	Continuous	%	Proportion of dwellings that are vacant or used as secondary residences.
TotalPopulation	INE	Discrete	-	Total resident population in the section.
ShareofMalePopulation	INE	Continuous	%	Proportion of male residents.
ShareofFemalePopulation	INE	Continuous	%	Proportion of female residents.
ShareofPopulationAged0-14	INE	Continuous	%	Proportion of residents in the specified age group.
ShareofPopulationAged15-24	INE	Continuous	%	Proportion of residents in the specified age group.
ShareofPopulationAged25-64	INE	Continuous	%	Proportion of residents in the specified age group.
ShareofPopulationAged65orOlder	INE	Continuous	%	Proportion of residents in the specified age group.
ShareofPopulationwithNoEducationalLevelCom	INE	Continuous	%	Proportion of residents with no formal education.
ShareofPopulationwithPrimaryEducationComp	INE	Continuous	%	Proportion of residents who completed primary education (1st cycle).
ShareofPopulationwithPrimaryEducationComp	INE	Continuous	%	Proportion of residents who completed primary education (2nd cycle).
ShareofPopulationwithLowerSecondaryEducati	INE	Continuous	%	Proportion of residents who completed lower secondary education (3rd cycle).
ShareofPopulationwithUpperSecondaryorPostse	INE	Continuous	%	Proportion of residents who completed upper secondary or post-secondary
ShareofPopulationwithHigherEducationComple	INE	Continuous	%	Proportion of residents who completed higher education.
ShareofEconomicallyActivePopulation	INE	Continuous	%	Proportion of economically active residents (employed + unemployed).

Variable Name	Source	Type	Unit	Description
ShareofEconomicallyInactivePopulation	INE	Continuous	%	Proportion of economically inactive residents.
ShareofEmployedPopulation	INE	Continuous	%	Proportion of residents currently employed.
ShareofStudents	INE	Continuous	%	Proportion of residents who are students.
ShareofHomemakers	INE	Continuous	%	Proportion of residents whose main status is homemaker.
ShareofRetiredPopulation	INE	Continuous	%	Proportion of residents who are retired.
ShareofEmployedPopulationinthePrimarySecto	INE	Continuous	%	Proportion of employed residents working in the primary sector.
ShareofEmployedPopulationintheSecondarySec	INE	Continuous	%	Proportion of employed residents working in the secondary sector.
ShareofEmployedPopulationintheTertiarySecto	INE	Continuous	%	Proportion of employed residents working in the tertiary sector.
ShareofForeignNationals	INE	Continuous	%	Proportion of residents with foreign nationality.
Shape_Length	QGIS	Continuous	Metres	Perimeter length of the section polygon (metres).
Shape_Area	QGIS	Continuous	Metres	Total area of the section polygon (square metres).
share_alpine_hut	OSM	Continuous	%	Proportion of points of interest classified as 'alpine hut' relative to all points of
share_archaeological	OSM	Continuous	%	Proportion of points of interest classified as 'archaeological' relative to all points
share_arts_centre	OSM	Continuous	%	Proportion of points of interest classified as 'arts centre' relative to all points of
share_artwork	OSM	Continuous	%	Proportion of points of interest classified as 'artwork' relative to all points of
share_atm	OSM	Continuous	%	Proportion of points of interest classified as 'atm' relative to all points of interest
share_attraction	OSM	Continuous	%	Proportion of points of interest classified as 'attraction' relative to all points of
share_bakery	OSM	Continuous	%	Proportion of points of interest classified as 'bakery' relative to all points of
share_bank	OSM	Continuous	%	Proportion of points of interest classified as 'bank' relative to all points of
share_bar	OSM	Continuous	%	Proportion of points of interest classified as 'bar' relative to all points of interest
share_battlefield	OSM	Continuous	%	Proportion of points of interest classified as 'battlefield' relative to all points of
share_beauty_shop	OSM	Continuous	%	Proportion of points of interest classified as 'beauty shop' relative to all points of
share_bench	OSM	Continuous	%	Proportion of points of interest classified as 'bench' relative to all points of
share_beverages	OSM	Continuous	%	Proportion of points of interest classified as 'beverages' relative to all points of
share_bicycle_rental	OSM	Continuous	%	Proportion of points of interest classified as 'bicycle rental' relative to all points
share_bicycle_shop	OSM	Continuous	%	Proportion of points of interest classified as 'bicycle shop' relative to all points
share_biergarten	OSM	Continuous	%	Proportion of points of interest classified as 'biergarten' relative to all points of
share_bookshop	OSM	Continuous	%	Proportion of points of interest classified as 'bookshop' relative to all points of
share_butcher	OSM	Continuous	%	Proportion of points of interest classified as 'butcher' relative to all points of
share_cafe	OSM	Continuous	%	Proportion of points of interest classified as 'cafe' relative to all points of interest
share_camera_surveillance	OSM	Continuous	%	Proportion of points of interest classified as 'camera surveillance' relative to all

Variable Name	Source	Type	Unit	Description
share_camp_site	OSM	Continuous	%	Proportion of points of interest classified as 'camp site' relative to all points of
share_car_dealership	OSM	Continuous	%	Proportion of points of interest classified as 'car dealership' relative to all points
share_car_rental	OSM	Continuous	%	Proportion of points of interest classified as 'car rental' relative to all points of
share_car_sharing	OSM	Continuous	%	Proportion of points of interest classified as 'car sharing' relative to all points of
share_car_wash	OSM	Continuous	%	Proportion of points of interest classified as 'car wash' relative to all points of
share_caravan_site	OSM	Continuous	%	Proportion of points of interest classified as 'caravan site' relative to all points of
share_castle	OSM	Continuous	%	Proportion of points of interest classified as 'castle' relative to all points of
share_chalet	OSM	Continuous	%	Proportion of points of interest classified as 'chalet' relative to all points of
share_chemist	OSM	Continuous	%	Proportion of points of interest classified as 'chemist' relative to all points of
share_cinema	OSM	Continuous	%	Proportion of points of interest classified as 'cinema' relative to all points of
share_clinic	OSM	Continuous	%	Proportion of points of interest classified as 'clinic' relative to all points of
share_clothes	OSM	Continuous	%	Proportion of points of interest classified as 'clothes' relative to all points of
share_college	OSM	Continuous	%	Proportion of points of interest classified as 'college' relative to all points of
share_comms_tower	OSM	Continuous	%	Proportion of points of interest classified as 'comms tower' relative to all points
share_community_centre	OSM	Continuous	%	Proportion of points of interest classified as 'community centre' relative to all
share_computer_shop	OSM	Continuous	%	Proportion of points of interest classified as 'computer shop' relative to all points
share_convenience	OSM	Continuous	%	Proportion of points of interest classified as 'convenience' relative to all points
share_courthouse	OSM	Continuous	%	Proportion of points of interest classified as 'courthouse' relative to all points of
share_dentist	OSM	Continuous	%	Proportion of points of interest classified as 'dentist' relative to all points of
share_department_store	OSM	Continuous	%	Proportion of points of interest classified as 'department store' relative to all
share_doctors	OSM	Continuous	%	Proportion of points of interest classified as 'doctors' relative to all points of
share_dog_park	OSM	Continuous	%	Proportion of points of interest classified as 'dog park' relative to all points of
share_doityourself	OSM	Continuous	%	Proportion of points of interest classified as 'doityourself' relative to all points of
share_drinking_water	OSM	Continuous	%	Proportion of points of interest classified as 'drinking water' relative to all points
share_embassy	OSM	Continuous	%	Proportion of points of interest classified as 'embassy' relative to all points of
share_fast_food	OSM	Continuous	%	Proportion of points of interest classified as 'fast food' relative to all points of
share_fire_station	OSM	Continuous	%	Proportion of points of interest classified as 'fire station' relative to all points of
share_florist	OSM	Continuous	%	Proportion of points of interest classified as 'florist' relative to all points of
share_food_court	OSM	Continuous	%	Proportion of points of interest classified as 'food court' relative to all points of
share_fort	OSM	Continuous	%	Proportion of points of interest classified as 'fort' relative to all points of interest
share_fountain	OSM	Continuous	%	Proportion of points of interest classified as 'fountain' relative to all points of

Variable Name	Source	Type	Unit	Description
share_furniture_shop	OSM	Continuous	%	Proportion of points of interest classified as 'furniture shop' relative to all points
share_garden_centre	OSM	Continuous	%	Proportion of points of interest classified as 'garden centre' relative to all points
share_general	OSM	Continuous	%	Proportion of points of interest classified as 'general' relative to all points of
share_gift_shop	OSM	Continuous	%	Proportion of points of interest classified as 'gift shop' relative to all points of
share_golf_course	OSM	Continuous	%	Proportion of points of interest classified as 'golf course' relative to all points of
share_graveyard	OSM	Continuous	%	Proportion of points of interest classified as 'graveyard' relative to all points of
share_greengrocer	OSM	Continuous	%	Proportion of points of interest classified as 'greengrocer' relative to all points of
share_guesthouse	OSM	Continuous	%	Proportion of points of interest classified as 'guesthouse' relative to all points of
share_hairdresser	OSM	Continuous	%	Proportion of points of interest classified as 'hairdresser' relative to all points of
share_hospital	OSM	Continuous	%	Proportion of points of interest classified as 'hospital' relative to all points of
share_hostel	OSM	Continuous	%	Proportion of points of interest classified as 'hostel' relative to all points of
share_hotel	OSM	Continuous	%	Proportion of points of interest classified as 'hotel' relative to all points of
share_hunting_stand	OSM	Continuous	%	Proportion of points of interest classified as 'hunting stand' relative to all points
share_ice_rink	OSM	Continuous	%	Proportion of points of interest classified as 'ice rink' relative to all points of
share_jeweller	OSM	Continuous	%	Proportion of points of interest classified as 'jeweller' relative to all points of
share_kindergarten	OSM	Continuous	%	Proportion of points of interest classified as 'kindergarten' relative to all points
share_kiosk	OSM	Continuous	%	Proportion of points of interest classified as 'kiosk' relative to all points of
share_laundry	OSM	Continuous	%	Proportion of points of interest classified as 'laundry' relative to all points of
share_library	OSM	Continuous	%	Proportion of points of interest classified as 'library' relative to all points of
share_lighthouse	OSM	Continuous	%	Proportion of points of interest classified as 'lighthouse' relative to all points of
share_mall	OSM	Continuous	%	Proportion of points of interest classified as 'mall' relative to all points of
share_market_place	OSM	Continuous	%	Proportion of points of interest classified as 'market place' relative to all points
share_memorial	OSM	Continuous	%	Proportion of points of interest classified as 'memorial' relative to all points of
share_mobile_phone_shop	OSM	Continuous	%	Proportion of points of interest classified as 'mobile phone shop' relative to all
share_monument	OSM	Continuous	%	Proportion of points of interest classified as 'monument' relative to all points of
share_motel	OSM	Continuous	%	Proportion of points of interest classified as 'motel' relative to all points of
share_museum	OSM	Continuous	%	Proportion of points of interest classified as 'museum' relative to all points of
share_newsagent	OSM	Continuous	%	Proportion of points of interest classified as 'newsagent' relative to all points of
share_nightclub	OSM	Continuous	%	Proportion of points of interest classified as 'nightclub' relative to all points of
share_nursing_home	OSM	Continuous	%	Proportion of points of interest classified as 'nursing home' relative to all points
share_observation_tower	OSM	Continuous	%	Proportion of points of interest classified as 'observation tower' relative to all

Variable Name	Source	Type	Unit	Description
share_optician	OSM	Continuous	%	Proportion of points of interest classified as 'optician' relative to all points of
share_outdoor_shop	OSM	Continuous	%	Proportion of points of interest classified as 'outdoor shop' relative to all points
share_park	OSM	Continuous	%	Proportion of points of interest classified as 'park' relative to all points of
share_pharmacy	OSM	Continuous	%	Proportion of points of interest classified as 'pharmacy' relative to all points of
share_picnic_site	OSM	Continuous	%	Proportion of points of interest classified as 'picnic site' relative to all points of
share_pitch	OSM	Continuous	%	Proportion of points of interest classified as 'pitch' relative to all points of
share_playground	OSM	Continuous	%	Proportion of points of interest classified as 'playground' relative to all points of
share_police	OSM	Continuous	%	Proportion of points of interest classified as 'police' relative to all points of
share_post_box	OSM	Continuous	%	Proportion of points of interest classified as 'post box' relative to all points of
share_post_office	OSM	Continuous	%	Proportion of points of interest classified as 'post office' relative to all points of
share_prison	OSM	Continuous	%	Proportion of points of interest classified as 'prison' relative to all points of
share_pub	OSM	Continuous	%	Proportion of points of interest classified as 'pub' relative to all points of interest
share_public_building	OSM	Continuous	%	Proportion of points of interest classified as 'public building' relative to all
share_recycling	OSM	Continuous	%	Proportion of points of interest classified as 'recycling' relative to all points of
share_recycling_clothes	OSM	Continuous	%	Proportion of points of interest classified as 'recycling clothes' relative to all
share_recycling_glass	OSM	Continuous	%	Proportion of points of interest classified as 'recycling glass' relative to all
share_recycling_metal	OSM	Continuous	%	Proportion of points of interest classified as 'recycling metal' relative to all
share_recycling_paper	OSM	Continuous	%	Proportion of points of interest classified as 'recycling paper' relative to all
share_restaurant	OSM	Continuous	%	Proportion of points of interest classified as 'restaurant' relative to all points of
share_ruins	OSM	Continuous	%	Proportion of points of interest classified as 'ruins' relative to all points of
share_school	OSM	Continuous	%	Proportion of points of interest classified as 'school' relative to all points of
share_shelter	OSM	Continuous	%	Proportion of points of interest classified as 'shelter' relative to all points of
share_shoe_shop	OSM	Continuous	%	Proportion of points of interest classified as 'shoe shop' relative to all points of
share_sports_centre	OSM	Continuous	%	Proportion of points of interest classified as 'sports centre' relative to all points
share_sports_shop	OSM	Continuous	%	Proportion of points of interest classified as 'sports shop' relative to all points of
share_stadium	OSM	Continuous	%	Proportion of points of interest classified as 'stadium' relative to all points of
share_stationery	OSM	Continuous	%	Proportion of points of interest classified as 'stationery' relative to all points of
share_supermarket	OSM	Continuous	%	Proportion of points of interest classified as 'supermarket' relative to all points
share_swimming_pool	OSM	Continuous	%	Proportion of points of interest classified as 'swimming pool' relative to all
share_telephone	OSM	Continuous	%	Proportion of points of interest classified as 'telephone' relative to all points of
share_theatre	OSM	Continuous	%	Proportion of points of interest classified as 'theatre' relative to all points of

Variable Name	Source	Type	Unit	Description
share_theme_park	OSM	Continuous	%	Proportion of points of interest classified as 'theme park' relative to all points of
share_toilet	OSM	Continuous	%	Proportion of points of interest classified as 'toilet' relative to all points of
share_tourist_info	OSM	Continuous	%	Proportion of points of interest classified as 'tourist info' relative to all points of
share_tower	OSM	Continuous	%	Proportion of points of interest classified as 'tower' relative to all points of
share_town_hall	OSM	Continuous	%	Proportion of points of interest classified as 'town hall' relative to all points of
share_toy_shop	OSM	Continuous	%	Proportion of points of interest classified as 'toy shop' relative to all points of
share_track	OSM	Continuous	%	Proportion of points of interest classified as 'track' relative to all points of
share_travel_agent	OSM	Continuous	%	Proportion of points of interest classified as 'travel agent' relative to all points of
share_university	OSM	Continuous	%	Proportion of points of interest classified as 'university' relative to all points of
share_vending_any	OSM	Continuous	%	Proportion of points of interest classified as 'vending any' relative to all points
share_vending_cigarette	OSM	Continuous	%	Proportion of points of interest classified as 'vending cigarette' relative to all
share_vending_machine	OSM	Continuous	%	Proportion of points of interest classified as 'vending machine' relative to all
share_vending_parking	OSM	Continuous	%	Proportion of points of interest classified as 'vending parking' relative to all
share_veterinary	OSM	Continuous	%	Proportion of points of interest classified as 'veterinary' relative to all points of
share_video_shop	OSM	Continuous	%	Proportion of points of interest classified as 'video shop' relative to all points of
share_viewpoint	OSM	Continuous	%	Proportion of points of interest classified as 'viewpoint' relative to all points of
share_waste_basket	OSM	Continuous	%	Proportion of points of interest classified as 'waste basket' relative to all points
share_wastewater_plant	OSM	Continuous	%	Proportion of points of interest classified as 'wastewater plant' relative to all
share_water_mill	OSM	Continuous	%	Proportion of points of interest classified as 'water mill' relative to all points of
share_water_tower	OSM	Continuous	%	Proportion of points of interest classified as 'water tower' relative to all points of
share_water_well	OSM	Continuous	%	Proportion of points of interest classified as 'water well' relative to all points of
share_water_works	OSM	Continuous	%	Proportion of points of interest classified as 'water works' relative to all points
share_wayside_cross	OSM	Continuous	%	Proportion of points of interest classified as 'wayside cross' relative to all points
share_wayside_shrine	OSM	Continuous	%	Proportion of points of interest classified as 'wayside shrine' relative to all
share_windmill	OSM	Continuous	%	Proportion of points of interest classified as 'windmill' relative to all points of
share_zoo	OSM	Continuous	%	Proportion of points of interest classified as 'zoo' relative to all points of interest
share_competitors	OSM	Continuous	%	Proportion of points of interest classified as 'competitors' relative to all points of
GrandTotal	OSM	Discrete	-	Number of points of interest counted within the section.
LengthofCycleways	OSM	Continuous	Metres	Total length of cycleways within the section.
LengthofFootways	OSM	Continuous	Metres	Total length of footways within the section.
LengthofLivingStreets	OSM	Continuous	Metres	Total length of living streets within the section.

Variable Name	Source	Type	Unit	Description
LengthofPedestrianRoads	OSM	Continuous	Metres	Total length of pedestrian roads within the section.
LengthofPrimaryRoads	OSM	Continuous	Metres	Total length of primary roads within the section.
LengthofResidentialRoads	OSM	Continuous	Metres	Total length of residential roads within the section.
LengthofSecondaryRoads	OSM	Continuous	Metres	Total length of secondary roads within the section.
LengthofTertiaryRoads	OSM	Continuous	Metres	Total length of tertiary roads within the section.
TotalRoadsLength	OSM	Continuous	Metres	Derived indicator of local environmental or socioeconomic characteristics.
section_area	SPSS	Continuous	Metres	Total area of the statistical section (from geometry).
bakery	SPSS	Discrete	Count	Number of bakery establishments within the section.
cafe	SPSS	Discrete	Count	Number of cafe establishments within the section.
mall	SPSS	Discrete	Count	Number of mall establishments within the section.
supermarket	SPSS	Discrete	Count	Number of supermarket establishments within the section.
competitors	SPSS	Discrete	Count	Number of competing establishments within/near the section.
mall_supermarket	SPSS	Discrete	Count	Combined count of shopping malls and supermarkets within the section.

Table 7.3: Derived Analytical Variables

Variable Name	Source	Type	Unit	Description
pop_dens	SPSS	Continuous	People/km²	Population density (people per square kilometre).
Building_dens	SPSS	Continuous	Buildings/	Building density (buildings per square kilometre).
Road_dens	SPSS	Continuous	m/m²	Road network density (metres of road per square metre).
Poi_dens	SPSS	Continuous	POIs/km²	Point of interest density (points of interest per square kilometre).
competitors_flag	SPSS	Categorical	-	Indicator variable taking the value 1 if true and 0 otherwise. Binary indicator of
competitors_1000inhab	SPSS	Continuous	-	Competitors per 1000 inhabitants within the section.
mall_supermarket_1000inhab	SPSS	Continuous	-	Malls and supermarkets per 1000 inhabitants within the section.
poi_1000inhab	SPSS	Continuous	-	Number of points of interest per 1000 inhabitants within the section.
FAC1_1	SPSS	Continuous	-	Principal Component score derived from PCA.
FAC2_1	SPSS	Continuous	-	Principal Component score derived from PCA.
FAC3_1	SPSS	Continuous	-	Principal Component score derived from PCA.
PFAC1_1	SPSS	Continuous	-	FAC1_1 Percentile Rank.
PFAC2_1	SPSS	Continuous	-	FAC2_1 Percentile Rank.
PFAC3_1	SPSS	Continuous	-	FAC3_1 Percentile Rank.
Flag_Store	QGIS	Categorical	-	Indicates whether a section already contains a store (1 = yes; 0 = no)

Variable Name	Source	Type	Unit	Description
Distance_C	QGIS	Continuous	Metres	Euclidean distance between the centroid of the section and the company's
Rank1_W	GeoDa	Continuous	-	Spatially weighted ranking of FAC1_1, computed using spatial weight matrix
Rank2_W	GeoDa	Continuous	-	Spatially weighted ranking of FAC2_1, computed using spatial weight matrix
Rank3_W	GeoDa	Continuous	-	Spatially weighted ranking of FAC3_1, computed using spatial weight matrix
CL	GeoDa	Continuous	-	Cluster identifier to each statistical section.
Zone	GeoDa	Continuous	-	Cluster identifier representing the zone classification of each section based on
Flag_Top5	QGIS	Categorical	-	Indicates whether a section is in the top5 civil parishes (1 = yes; 0 = no)

B: Results of PCA

Table 7.4: Total Variance Explained

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.397	35.982	35.982	5.397	35.982	35.982	3.975	26.503	26.503
2	2.695	17.969	53.951	2.695	17.969	53.951	3.735	24.897	51.400
3	2.122	14.146	68.096	2.122	14.146	68.096	2.504	16.697	68.096
4	0.896	5.974	74.071						
5	0.750	4.998	79.069						
6	0.576	3.843	82.912						
7	0.524	3.493	86.405						
8	0.446	2.972	89.377						
9	0.410	2.735	92.112						
10	0.340	2.268	94.380						
11	0.229	1.527	95.906						
12	0.212	1.413	97.320						
13	0.172	1.148	98.467						
14	0.146	0.970	99.438						
15	0.084	0.562	100.000						

Table 7.5: Rotated Component Matrix

Rotated Component Matrix

	Component		
	1	**2**	**3**
Road_dens	0.880	0.188	0.025
Pop_dens	0.795	0.129	-0.117
Number of Conventional Buildings	-0.771	-0.171	-0.022
Building_dens	0.758	-0.016	-0.021
Poi_dens	0.697	-0.022	0.254
Share of Population with Higher Education Completed	0.526	0.484	0.130
Total Roads Length	-0.518	-0.458	0.032
Share of Population Aged 65 or Older	-0.115	-0.899	0.008
Share of Economically Active Population	0.232	0.869	0.025
Share of Students	0.055	0.767	0.012
Total Population	-0.069	0.751	-0.176
Share of Employed Population in the Tertiary Sector	0.538	0.694	0.050
Competitors_1000inhab	0.076	-0.056	0.913
Poi_1000inhab	-0.010	-0.075	0.896
Mall_supermarket_1000inhab	0.027	0.057	0.858

Rotation Method: Varimax with Kaiser Normalization

C: Results of Clusters Analyse

Table 7.6: Evaluation of Clustering Solutions Using TSS, WSS and BSS

Clusters	TSS	WSS	BSS	BSS/TSS
1	29784	21725.9	0	0.000
2	29784	21725.9	8058.11	0.271
3	29784	16226.8	13557.2	0.455
4	29784	12511.9	17272.1	0.580
5	29784	10481.2	19302.8	0.648

6	29784	8716.02	21068	0.707
7	29784	7485.66	22298.3	0.749
8	29784	6653.86	23130.1	0.777
9	29784	6181.12	23602.9	0.792
10	29784	5770.67	24013.3	0.806

Table 7.7: Cluster Centres of Territorial Profile Analysis

	Cluster Centres		
	PFAC1_1	**PFAC2_1**	**PFAC3_1**
C1	29.9543	18.2385	46.4832
C2	19.4576	67.2719	50.2083
C3	78.693	51.2203	14.8707
C4	65.0626	82.7871	68.0809
C5	71.9653	30.2223	83.1219
The ratio of between to total sum of squares			0.648091

Table 7.8: Cluster Centres of Proximity Analysis

	Cluster Centres	
	Flag_Store	**Distance_C**
C1	0	124985
C2	0	188513
C3	0	103176
C4	0	267597
C5	0	71871.8
C6	0	212356

C7	0	237474
C8	0	159384
C9	0	35400
C10	1	19774.5
The ratio of between to total sum of squares		0.992291

D: Final Statistical Sections

Table 7.9: Selection of Statistical Sections

	1	**2**	**3**	**4**	**5**	**Total**
1	250	350	942	675	390	2607
2	216	327	433	320	160	1456
3	295	276	254	186	122	1133
4	314	126	135	116	342	1033
5	344	290	75	182	122	1013
6	156	375	75	129	57	792
7	266	206	49	117	67	705
8	265	236	46	66	79	692
9	259	99	37	37	49	481
10	4	6	1	4	2	17
Total	2369	2291	2047	1832	1390	9929

Table 7.10: Selected Statistical Sections

DTMNFR21	**DTMNFRSEC21**	**CL**	**Zone**
141821	141821005	3	9
141001	141001007	3	9
141001	141001001	3	9
141002	141002005	3	9
141002	141002003	3	9
141001	141001006	3	9

DTMNFR21	DTMNFRSEC21	CL	Zone
141001	141001005	3	9
141002	141002010	3	9
141001	141001003	3	9
141002	141002004	3	9
141002	141002009	3	9
141001	141001008	3	9
141002	141002006	3	9
051002	051002001	3	9
141910	141910003	3	9
141921	141921002	3	9
141920	141920005	3	9
141921	141921001	3	9
142115	142115001	3	9
140401	140401004	3	9
140401	140401003	3	9
140107	140107001	3	9
101608	101608001	3	9
140301	140301007	3	9
140301	140301011	3	9
140303	140303002	3	9
140301	140301004	3	9
140301	140301005	3	9
140301	140301002	3	9
141633	141633002	3	9
141633	141633015	3	9
141633	141633001	3	9
141633	141633009	3	9
141633	141633010	3	9
141633	141633018	3	9
141633	141633017	3	9
140124	140124001	3	9