FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Understanding demand and pricing behavior in the car rental business

Sara Isabel Serra Costa

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Beatriz Brito Oliveira

June 24, 2019

Resumo

Com a crescente importância de guardar e analisar dados, como é que uma empresa de rent-a-car Portuguesa pode extrair conhecimento dos seus dados de forma a entender os comportamentos da procura e dos preços no mercado? Como lidar com quantidades significativas de dados e obter informações úteis?

Face à falta de um sistema que permita às empresas de rent-a-car responder de forma eficiente a estas questões, uma solução personalizada foi desenvolvida para a empresa em análise neste trabalho, podendo também ser relevante para outras empresas. A solução desenvolvida foi projetar e implementar um sistema integrado que recolhe todos os dados relevantes de preço, frota e ocupação mantidos em diferentes ficheiros pela empresa e os armazena numa base de dados, para que posteriormente seja possível conectá-los a diferentes ferramentas de análise de dados.

Vários módulos de software foram desenvolvidos em diferentes formatos e plataformas. Além disso, algumas análises foram projetadas e implementadas usando diferentes técnicas de data mining, com ênfase na visualização e regressão.

Este sistema permitirá futuras análises de outros dados relevantes por outros investigadores e funcionários da empresa. As análises obtidas comprovam as funcionalidades e capacidades do sistema. Além disso, oferecem insights de gestão que a empresa pode utilizar para adotar e adaptar comportamentos e obter uma melhor compreensão do seu posicionamento no mercado de rent-a-car em Portugal.

Abstract

With the growing importance of saving and analyzing data, how can a Portuguese car rental company extract knowledge from its data to understand demand and market pricing behavior? How can a company tackle significant amounts of data and get useful information from it?

Car rental companies do not have such a system, thus a customized solution has been developed for the company analyzed in this work, which could also be relevant to other companies. The solution developed included the design and implementation of an integrated system that collects all the relevant price, fleet, and occupation data maintained in different files by the company and stores them in a database, so that later it is possible to connect them to different tools for data analysis.

Multiple modules of software were developed in different formats and platforms. Moreover, some analyses were designed and implemented using different data mining techniques, with an emphasis on visualization and regression.

This system will allow future analyses of other relevant data by other researchers and company employees. The analyses obtained are a proof-of-concept of the functionalities and capabilities of the system. Moreover, they offer managerial insights that the company can use to adopt and adapt behaviors and to gain a better understanding of its positioning in the Portuguese car rental market.

Acknowledgements

Firstly and above all, I would like to thank my supervisor, Prof. Beatriz Brito Oliveira for the possibility to be a part of this project. Her confidence in me and support was fundamental for my personal and professional growth throughout this semester.

I also want to thank my family, especially my mother and sister, for always being there to support and love me. I hope this work leaves my father, who is no longer present among us, proud. You have made me the person I am today, and I will continue to work hard to become an even better person and professional.

Sara Isabel Serra Costa

Contents

1 Introduction **1**
1.1 Purpose . . . 1
1.2 Structure . . . 2

2 Literature Review **3**
2.1 Application: the car rental business . . . 3
2.2 Methodology . . . 4
2.2.1 Data mining . . . 4
2.2.2 Economics . . . 7

3 Problem Description **13**
3.1 Company description . . . 13
3.2 Pricing and reservation procedures of the studied company . . . 14
3.3 Main objectives . . . 15

4 Proposed Solution **17**
4.1 Data Analysis Support System design . . . 17
4.2 Data handling . . . 18
4.2.1 Location and structure of files folders . . . 18
4.2.2 Data contained in the files . . . 18
4.3 Modular methodological approaches . . . 20
4.3.1 Database schema creation . . . 21
4.3.2 Extracting the data . . . 23
4.3.3 Connection to a Data Visualization tool . . . 25
4.3.4 Extracting prices and occupation values . . . 26
4.3.5 Calculating elasticities . . . 30
4.3.6 Connection to the Data Mining tool . . . 32

5 Analyses **35**
5.1 Visualization . . . 35
5.1.1 High Season analyses . . . 35
5.1.2 Middle Season Analysis . . . 40
5.2 Elasticities . . . 45
5.2.1 Group C . . . 45
5.2.2 Group J . . . 46
5.2.3 Group MINI . . . 48
5.2.4 Group H . . . 48
5.3 Regressions . . . 49

5.3.1 Group C 49
5.3.2 Group J 49
5.3.3 Group MINI 50

6 Conclusions and future work **53**

References **55**

List of Figures

2.1 Market Equilibrium [1] 8

3.1 Car-Rent Vehicle Groups 16

4.1 Data Analysis Support System, icons made by [2, 3, 4, 5, 6, 7, 8, 9, 10, 11] . . . 17
4.2 Calculation of available fleet 20
4.3 SQL table Broker 21
4.4 SQL table Fleet 21
4.5 SQL table Occupation 22
4.6 SQL table Price 22
4.7 SQL table Rental 23
4.8 SQL table Vehicle 23
4.9 Scheme of the fleet and occupation table 25
4.10 Scheme of the price and rental table 26
4.11 Example of a 5-day rental beginning before the start day 27
4.12 Example of a 7-day rental beginning before the start day 29
4.13 Regression Process 33
4.14 Cross-Validation block 33

5.1 Occupation value for different brokers 36
5.2 Occupation values for vehicle group categories 37
5.3 Price values for Car-Rent 38
5.4 Price values of Car-Rent and price of its competitors for economic groups 39
5.5 Price values of Car-Rent and price of its competitors for luxury groups 40
5.6 Price values of Car-Rent and price of its competitors for SUV groups 41
5.7 Occupation value for different brokers 42
5.8 Occupation values for vehicle group categories 43
5.9 Price values for Car-Rent 44
5.10 Price values of Car-Rent and price of its competitors for economic groups 45
5.11 Price values of Car-Rent and price of its competitors for luxury groups 46
5.12 Price values of Car-Rent and price of its competitors for SUV groups 47
5.13 Price Elasticity of Demand group C 48
5.14 Price Elasticity of Demand group J 49
5.15 Price Elasticity of Demand group J 50
5.16 Price Elasticity of Demand group MINI 51
5.17 Regression result Group C 51
5.18 Regression result Group J 51
5.19 Regression result Group MINI 51

List of Tables

2.1 Values and meanings of price elasticity demand 10

Acronyms

CSV	Comma Separated Values
ERP	Enterprise Resource Planning
RM	Revenue Management

Chapter 1

Introduction

The car rental market has gained increasing importance in the transportation sector, due to several factors, but mainly due to the increase in tourism. In Portugal, six in every ten euros of revenues in this sector come from tourism [12]. In 2017 alone, the car rental sector recorded revenues of around 700 million euros, a growth of 45% over the year 2016 [12]. Moreover, in December 2018, the acquisition of 3,071 light passenger vehicles was registered, affirming the sector as the largest buyer of new vehicles in Portugal [13].

This business involves several areas such as logistics, management, and marketing, and its operational efficiency is increasingly fundamental. The market is highly competitive, and therefore it is gradually necessary to develop efficient and flexible systems to manage a significant amount of data on reservations, rentals, fleet management, prices, and customers, among others. This data is often disorganized and without potential use, mainly due to its quantity. There is still a lack of a data analysis framework to tackle the data and draw relevant insights from it. The application of this framework, and of the analyses it implies, has the potential to increase sales, improve customer relationships, and reduce risk.

1.1 Purpose

The purpose of this dissertation is to design and develop an integrated system for a Portuguese car rental company to collect pricing and fleet occupancy data and connect it to data analysis tools, using real data to obtain relevant information. The analyses performed in this dissertation are a proof-of-concept of the system and allow the company to draw relevant managerial insights. Building this system implied understanding the problems and daily operations of the company, mapping and structuring the available data, developing adequate software as well as data analysis techniques, and drawing relevant insights.

1.2 Structure

This work is divided into the following chapters: Literature Review, Problem Description, Proposed Solution, Analyses, and Conclusions. The Literature Review, chapter 2, presents the data analysis techniques and economic concepts applied in the analyses. The company, its procedures, and the main objective of the analyses will be explained in the Problem Description, chapter 3. In the Proposed Solution, chapter 4, the system design, data handling, and modular methodological approaches are described. The different analyses obtained are explained in the Analyses, chapter 5, and finally, the Conclusions, chapter 6, present the future work and final insights of this work.

Chapter 2

Literature Review

In this chapter, an overall description of relevant problems in the car rental industry tackled by the literature will be presented. In addition, a summary of some data analysis techniques and fundamental economic concepts that were crucial to the development of the system are discussed.

2.1 Application: the car rental business

The car rental is a growing business. From 2016 to 2021, the global car rental industry is expected to grow 5.6%, due to increasing tourism activities, the globalization of operations, and the global rise of income levels [14]. Costs are heavily correlated with the unoccupied fleet. Therefore, the revenue management problem not only deals with uncertain demand (highly dependent on the companies' positioning versus the other prices on the market) but it also deals with the need to maximize the occupation of the fleet for each day, ensuring the cars were booked at the highest possible price [15].

Demand is the most recognized and addressed problem by car rental companies [14]. However, other issues often addressed also include revenue management and fleet management. Internet sales have significantly impacted the car rental business, mainly due to new sales channels: broker websites that compare the prices of different competitors in the market. This impact is especially relevant for car rentals because of the lack of differentiation of their product. Since the vehicles and pick-up stations are the same, the clients can compare all the offers in the market with full transparency, making the price an even more decisive role in their decision [15].

The car rental fleet management problem embeds decisions that are traditionally framed within different strategic levels and studied by different research areas. Overall, the main decisions are related with clustering locations that share the same fleet, deciding on the fleet size and composition, distributing fleet amongst rental stations, deciding on prices, selecting which reservations to accept, and assigning these reservations to specific vehicles [14].

According to [16], the primary aim of any business company is to make a profit, and to achieve more profits increasing or improving sales is critical. If the company is capable of predicting the cost-output ratio, the income of the consumer, and the impact of advertising, the company may improve its sales. Therefore, to understand the behavior of sales and plan their growth, it is necessary to identify the factors that influence the demand and estimate their effect. To understand and analyze the demand behavior of the company, management should integrate economic concepts with business decision-making practices. In this process, the idea of elasticity plays a vital role, which helps to provide a quantitative value for the responsiveness of the change in quantity demanded to each factor that will influence demand [16].

As stated by [16], in the analysis of sales behavior, the insights of elasticity of demand can be used for forecasting change in the demand for the product due to the expected change in the demand determinants. To understand the behavior of sales it is necessary to be aware of the influence of factors under the control of the company and the factors beyond control. The factors under the control of the company include price, advertising, quantity, and quality of the product. The factors beyond control include tastes, preferences, and incomes of consumers, and competitors' prices. From the main literature available, it is possible to conclude that some factors that are relevant in the car rental business are the prices of the company and its competitors, as well as advertising and preferences and incomes of consumers.

2.2 Methodology

In this section, the data analysis techniques and key economic concepts used for the analyses are described.

2.2.1 Data mining

In this sub-section, the description of the data mining techniques used, Data Visualization and Regression, is based on [17, 18, 19, 20, 21]. Nowadays, significant amounts of data are collected daily, thus analyzing it has growing importance. Data mining can meet the need for scalable, flexible, and practical data analysis by providing tools to obtain knowledge and relevant information from data. Therefore, data mining can be referred to as knowledge discovery from data. The following tasks can define this process of discovery:

- Data cleaning: removing noise and inconsistent data;
- Data integration: combining multiple data sources, if necessary;
- Data selection: retrieving relevant data to the analysis from the database;
- Data transformation: transforming and consolidating data into appropriate forms for mining;
- Data mining: applying intelligent methods to extract data patterns;
- Pattern evaluation: identifying the interesting patterns;
- Knowledge presentation: using visualization and other representation techniques to present mined knowledge.

2.2.1.1 Data handling and preprocessing

The tasks of data cleaning, integration, selection, and transformation can be agglomerated in one big task of preprocessing. Before proceeding with data preprocessing, which is the first primary task of the data mining process, it is useful to know the data sets. Data sets are composed of data objects. A data object represents an entity (in a university database, for example, the objects may be students, professors, or courses), and attributes typically describe them. An attribute is a data field that represents a characteristic or feature of a data object. The attributes may be nominal (symbols or names of things), binary (only two categories: 0 or 1), ordinal (values that have an exact order), and numeric (measurable quantity).

Basic statistical descriptions provide the analytical foundation for data preprocessing. The basic statistical measures include mean, weighted mean, median, and mode for measuring the central tendency of data, and range, variance, and standard variation for measuring the dispersion of data. Graphical representations, such as boxplots, histograms, and scatter plots, provide an easier visual inspection of the data.

Data sets have quality if they satisfy the requirements of the intended use. Some factors used to assess data quality include accuracy, completeness, consistency, and interpretability. There are many reasons for inaccurate data, such as human or computer errors occurring at data entry, an error with the data collection instruments, or users that purposely submit incorrect data so as not to submit personal information. Incomplete data can occur due to equipment malfunctions, information not available (e.g., in personal information of users), not being relevant at the time of entry, or being inconsistent with other recorded data.

Focusing on the task of preprocessing, the first process is data cleaning. Data cleaning consists of filling in missing values, smoothing out noise while identifying outliers, and correcting inconsistencies in the data. Data integration, the process that follows, is the merging of data from multiple data sources. However, to avoid and reduce inconsistencies and redundancies, there are a few steps that can be performed, such as data conflict detection, correlation analysis, and resolution of semantic heterogeneity.

Other tasks include data reduction techniques, which are used to obtain a reduced representation of the data while minimizing the loss of information, or data transformation, which converts the data into appropriate forms for mining. Some of the transformation techniques are normalization and data discretization.

2.2.1.2 Data Visualization

Data Visualization aims to communicate data clearly and effectively through graphical representation, allowing the discovery of information that otherwise was not readily observable by looking at the raw data. A visualization should provide a clear answer to a question without irrelevant details, and also provide information that raises additional questions. Some approaches include pixel-oriented techniques, geometric projection techniques, icon-based techniques, and hierarchical and graph-based techniques.

Pixel-oriented visualization techniques

A simple way to visualize the value of a dimension is to use a pixel where the color of the pixel reflects the value of the dimension. For a data set of *m* dimensions, pixel-oriented techniques create *m* windows on the screen, one for each dimension. The *m* dimension values of a record are mapped to *m* pixels at the corresponding positions in the windows. The colors of the pixels reflect the corresponding values.

Geometric projection visualization techniques

A disadvantage of the pixel-oriented visualization techniques is that it does not help understanding the distribution of data in a multidimensional space. Geometric projection techniques help find engaging projections of multidimensional data sets. The main challenge of this technique is how to visualize a high-dimensional space on a 2-D display.

A scatter-plot displays 2-D data points using Cartesian coordinates. A third dimension can be added using different colors or shapes to represent different data points. For data sets with more than four dimensions, scatter-plots are generally ineffective, so a scatter-plot matrix is a better option. For an *n*-dimensional data set, a scatter-plot matrix is an *n* x *n* grid of 2-D scatter plots that provides a visualization of each dimension with every other dimension. However, as the dimensionality increases, the scatter-plot matrix becomes less effective.

Another technique that can handle higher dimensionality is the parallel coordinates. To visualize *n*-dimensional data points, the parallel coordinates technique draws n equally spaced axes, one for each dimension, parallel to one of the display axes. A significant limitation of this technique is that it cannot conclusively show a data set of many records.

Icon-based visualization techniques

Icon-based visualization techniques use small icons to represent multidimensional data values. Two popular icon-based techniques are Chernoff faces and stick figures.

Chernoff faces display multidimensional data of up to 18 variables (or dimensions) as a cartoon human face. This technique helps to reveal trends in the data. Components of the face, such as the eyes, ears, mouth, and nose, represent values of the dimensions by their shape, size, placement, and orientation.

The stick figure visualization technique maps multidimensional data to five-piece stick figures, where each figure has four limbs and a body. Two dimensions are mapped to the display axes, and the remaining dimensions are mapped to the angle and/or length of the limbs.

Hierarchical visualization techniques

The previous techniques focus on visualizing multiple dimensions simultaneously. However, for an extensive data set of high dimensionality, it would be difficult to visualize all dimensions at the same time. Hierarchical visualization techniques partition all dimensions into subsets. These subsets are visualized hierarchically.

In summary, visualization provides useful tools to explore data. It is crucial to choose the right technique depending on the information to provide and to choose the right elements to convey a clear answer (area, color, and volume are subjective and difficult to interpret if they are not correctly used). In addition to visualizing data, visualization can be used to represent the data mining process, such as the patterns obtained from a mining method.

2.2.1.3 Regressions

One of the most important types of data analysis that helps to make data-driven decisions is regression [22]. Regression analysis allows understanding which factors have the most impact on a variable, and how they interact with each other. The regression model is a relationship between two different types of variables: dependent and independent variables. The dependent variable is the main factor that we are trying to understand or predict. The independent variables are the factors we suspect have an impact on the dependent variable [22].

According to [22], most companies use regression analysis to explain a phenomenon they want to understand, predict things about the future, or to decide what to do. However, when working with regression analysis, it is important to consider that correlated factors are not necessarily connected by cause and effect. Correlation is not causation; therefore, the data should be used to guide more experiments, not to make conclusions about cause and effect [22]. To make a better analysis, there are several aspects to keep in mind: selecting the factors that potentially have an impact, not all possible factors; the relevance of the independent factors; using reliable data; and evaluate the correctness of the analysis' results.

2.2.2 Economics

For a better understanding of the results from the data analysis tools and of how they can benefit the company, it is crucial to have knowledge of some basic and key economic concepts. The description of these concepts, such as demand, supply, and elasticity, presented in this sub-section, is based on [23].

2.2.2.1 Demand, supply, balance

The market of any good, or service, is composed of the buyers and sellers, effective and potential, of this good. The interactions made between buyers and sellers characterize the prices and quantities of this good in the market, and can be described using supply and demand curves.

The supply curve indicates, for a given price, the total amount of a good that suppliers are willing to sell. In general, suppliers will be willing to sell if their return covers the initial cost of investment to provide the good. In addition, they are willing to sell more at higher prices (supply curves have a positive slope). Supply curves are primarily affected by factors such as technology, production prices, and in the case of agricultural products, weather conditions. For example, when there is a shortage of agricultural products due to sudden changes in weather conditions, there is a decrease in quantity and an increase in price, shifting the supply curve.

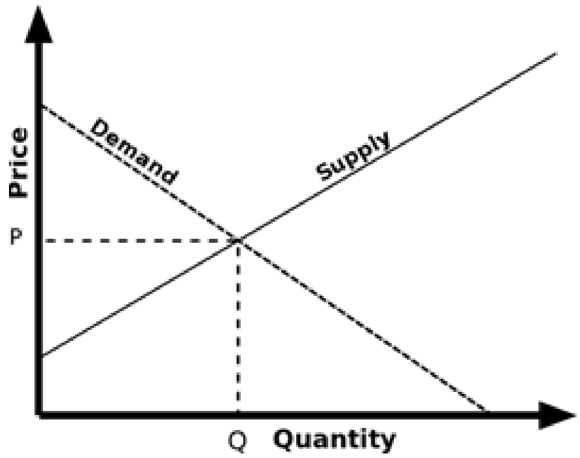

Figure 2.1: Market Equilibrium [1]

The demand curve indicates, for a given price, the total amount of a good that consumers desire to purchase. This curve has a negative slope in relation to the slope of the supply curve. Consumers are willing to buy less at higher prices. Income, tastes, and population are among some of the factors that affect demand curves. For example, when there is a shortage of a product due to a sudden interest of the population, there is an increase in quantity and price, shifting the demand curve.

For a balanced market, assuming a rational behavior, there is no reason for its constituents to change their behavior and therefore no tendency to change the prices or quantity of a good. Therefore, when there is a market equilibrium, all buyers and producers are "satisfied" in their respective quantities at the market prices. To find this balance, the demand and supply curves are used. The price at which a good will be sold (equilibrium price), and the quantity that will be sold (equilibrium quantity), are determined by the intersection of the two curves, supply and demand, as seen in Figure 2.1 from page 8.

Suppliers may have excess supply, which is the difference between the quantity offered and the quantity demanded, when the price exceeds the equilibrium price. They can also face excess demand, which is the difference between the quantity offered and the quantity demanded, when the price is set below the equilibrium price.

Typically, prices and quantities are driven to equilibrium by the actions of consumers and suppliers. If a price is initially high, resulting in oversupply, producers will lower prices in order to sell more. On the other hand, if initially a price is low, resulting in an excess of demand, competition among suppliers will drive the price up.

2.2.2.2 Elasticity

The elasticity is a measurement of how a variable responds to a change in another. In this work, we will address price elasticity of demand. However, there are other relevant concepts such as price elasticity of supply, cross-elasticity of demand, cross-elasticity of supply, and income elasticity of demand. The price elasticity of supply is the percentage change in the offered quantity of a good that results from a change in its price. The cross elasticity of demand is the percentage change in the demanded quantity of a good that results from a change in the price of another good. The cross elasticity of supply is the percentage change in the offered quantity of a good resulting from a change in the price of another good. The income elasticity of demand is the percentage changed in the demanded quantity resulting from a change in the income.

2.2.2.3 Price elasticity of demand

According to the economic law of demand, if no other factors are impacting the demand, generally consumers will purchase less of a good if the price of the good increases. As stated by [24], price elasticity of demand is one of the critical elements of pricing and shows precisely how sensitive customer demand is for a good based on its price. In other words, the price elasticity of demand measures how a percentage change in the price of a good will affect the quantity demanded of that good. As mentioned by [24], several factors affect elasticity, such as the product being sold, marketing, the income of the target consumers, the health of the economy, and the competitors' behavior. For example, when there are substitutes for a good, elasticity tends to become more elastic, as consumers can easily switch to other goods in response to price changes. If there is a high cost involved with switching between products, then demand may become relatively inelastic. Moreover, if the consumer's income decreases, consumers will tend to be more sensitive to price changes.

In the short run, demand for goods tends to be inelastic, as consumers do not immediately notice or react to price changes. However, this may vary depending on the product and the sale channel used. For the car rental business, new broker websites help consumers to compare price changes almost immediately, so the demand for these goods may be elastic in the short run. In the long run, demand for goods tends to be elastic, as consumers have the opportunity to change their spending habits. As a good becomes more necessary for a consumer, elasticity decreases, as the consumer will purchase the goods at nearly any price. Goods that are considered luxury items tend to be more elastic, as consumers can do without them. However, if they are in desperate need of a luxury good and there are no immediate substitute products, the demand may be more inelastic. In the car rental business, there are different types of rental vehicles' categories, such as luxury vehicles, possibly for high-end consumers, and standard vehicles, perhaps for consumers on vacation, and these goods may have different elasticities.

Depending on the value of price elasticity of demand, the demand of a good can be more elastic or inelastic. Table 2.1 shows the values and meanings of different price elasticities of demand, based on [24].

The elasticity can often be an inexact calculation, according to [24]. Since it is practically impossible to know what customers will do at every price point, it is also essential to understand consumer behavior. The goal is to understand the current price elasticity of demand and the factors that make it elastic or inelastic and how these factors will change over time.

Table 2.1: Values and meanings of price elasticity demand

Price Elasticity of Demand	
Value	Meaning
0	Perfectly inelastic: the quantity demanded does not change when the price changes.
$\varepsilon < 1, \varepsilon > -1$	Relatively inelastic: very large changes in price case small changes in the quantity demanded.
-1, 1	Unit elastic: the change in price of a good is equal to the change in the quantity demanded.
$\varepsilon > 1, \varepsilon < -1$	Relatively elastic: small changes in price cause large changes in the quantity demanded.
$\infty, -\infty$	Perfectly elastic: any very small change in price results in a very large change in the quantity demanded. The products considered perfectly elastic are considered mostly “pure commodities”.

Companies that stay relevant to consumers and differentiated from its competitors, as said by [24], can adjust their prices up or down to represent better the level of value provided to their customers. Their current price elasticity of demand is just one data point that helps to make future decisions. Nevertheless, it is essential for companies to have this information, and having significant data helps to obtain “average” elasticities that can be useful decision-making support information.

2.2.2.4 Types of markets

For different types of markets, specific price elasticities of demand may be expected. The perfect competition market is a market in which no single supplier has a significant influence on the market price of the product. The perfect competition is an ideal case, which only in the limit approaches the mode of operation of the real sectors of activity since consumers compare prices of the suppliers and mostly make their decisions based on the price and the existence of substitute products. The imperfect competition markets are:

- Monopoly: a single supplier of a single product, without close substitutes. According to [24], the perfectly inelastic price elasticity of demand is mostly seen in companies that have a monopoly on demand. Even if they change the price, the clients still have to buy the product.

- Oligopoly: only a few companies sell a particular product. For this market, the price elasticity of demand will tend to be relatively inelastic, since there are some other offers from other companies for the product.

- Monopolistic competition: typically constituted by many companies that sell the same product, slightly differentiated. Thus, some suppliers influence the market price of the product. In this market, the price elasticity of demand may be relatively elastic or even more elastic, depending on the product, due to the existence of several product substitutes.

The car rental business can be seen as a monopolistic competition market, especially with the use of broker websites that allow clients to compare prices of slightly differentiated products for multiple suppliers. In this business, the competitors' price plays a crucial role in the pricing process of companies, since it is vital to have a competitive price market. The price elasticity of demand will depend on the product and season, since the demand for this market has fluctuations throughout the year, and will probably tend to be elastic.

Chapter 3

Problem Description

The problem tackled in this work concerns a Portuguese car rental company. The company has a significant amount of raw data that they want to analyze with the purpose of understanding demand and pricing behavior, and obtain managerial insights. The problem that this work intends to solve will be explained with more detail in this chapter. In addition, the car rental company and its procedures will be described.

3.1 Company description

The problem described in this project concerns a Portuguese car rental company, which we will call Car-Rent, for simplicity. It has been present in the national market since the eighties and is the second company in fleet size operating in Portugal. The company has more than 50 rental stations throughout Portugal, including the archipelagos of Madeira and the Azores, where it is possible to collect and deliver the rented vehicles. During the peak season in 2018, it had 12,000 vehicles in fleet and more than 10,000 vehicles simultaneously on-rent. The company also represents other international brands in Portugal.

The year is divided into four different seasons, related to holidays and vacations, which are:

- Low season – January 2nd to February 22nd ,2018; February 23rd to March 22nd ,2018.
- Middle season – April 2nd to June 30th ,2018; October 1st to 31st ,2018.
- High season - December 16th ,2017 to January 1st ,2018 (Christmas); March 23rd to April 1st ,2018 (Easter); July 1st to 14th ,2018 (Summer); August 21st to September 30th ,2018 (Summer).
- Super high season – July 15th to August 20th ,2018 (Summer); November 1st to December 15th ,2018; December 16th ,2018 to January 1st ,2019 (Christmas).

The company has three customer segments:

- Corporate – corporate clients with whom inflexible and long-term contracts are established, with fixed services and prices, including a significant amount of insurance companies.

- International leisure – end-clients who use online brokers to compare the prices of competitors, and book vehicle reservations. These clients are the focus of this study.

- Direct – clients who rent the vehicle directly to the company.

Car-Rent's fleet is divided into two categories: special vehicles and free-sale vehicles, each consisting of several rental groups. Special vehicles belong to luxurious ranges or have distinct characteristics (such as minivans and off-road vehicles), thus existing a smaller number of vehicles available for each group that belongs to this category. Often this means that these vehicles are not located in the required stations leading to the need to perform empty transfers, moving the vehicle from the station where it is to the station where it will be needed. The free-sale vehicles exist in larger quantities and therefore are usually permanently available at each station.

The vehicles are divided into groups according to their characteristics. These groups are: Group C and MINI, the most economical; Group H, G, L and Z, the most expensive; Group N and O, the minivan; Group E, J, F, SUVC and SUVM, the SUV; and Group 02, 03, 04 and 05, the commercials, as seen in Figure 3.1, from page 16.

3.2 Pricing and reservation procedures of the studied company

The company calculates and provides prices for each different type of rental, which is characterized by start and end date, start and end location, and vehicle group required. These prices are also different for each broker and change with the distance to the start date, i.e. the antecedence. All other additions – such as insurance or rental of child seats – are indexed to these base prices.

The company currently is in the process of modifying its traditional pricing process, which is manual, and thus time-consuming and slow. The process begins by obtaining prices for different orders on the brokers' websites and in the fleet management system of the company and then analyzing the market conditions (i.e., the prices of competitors). Once prices have been formulated and approved, they are recorded in the Enterprise Resource Planning (ERP) of the brokers (a business management software used by companies to store, manage, and interpret data). This procedure has disadvantages, such as the difficulty of matching fluctuations in competition and the difficulty of including fleet analysis in the pricing decision.

The main objectives of the new pricing decision support system that the company is developing are the optimization and acceleration of pricing decisions. The aim is to maximize profit considering fleet availability, automate the pricing procedure, speed up the process to update prices quickly, improve fleet reposition decisions, and reduce costs. The solution developed for this new pricing process was a decision-support system to control the flow of information, and a Revenue Management (RM) tool, based on a priority schedule procedure and a price update heuristic. The decision-support system collects data from the web and the company's fleet management system. Then, the prices are collected from the web, updated by the RM tool, and fed back to the brokers' ERPs. The RM tool also considers up-to-date data on fleet occupation, as well as other relevant parameters. This dissertation appears in the context of this new pricing process. Its development has given rise to a significant amount of data, which the company now intends to analyze.

As far as the rental procedures are concerned, reservations can be made through Car-Rent's website, online brokers, mobile application, and call-center of the company. The customer, from the international leisure customer segment, provides the following information: the date and station at which he intends to collect the vehicle (pick-up or check-out), and the date and station at which he intends to deliver the vehicle (drop-off or check-in). Customers who have reservations or those that do not have reservations but expect to rent directly with the company ("walk-in" costumers) carry out the rental contract at the counter. For this work, only customers with a reservation will be analyzed. The company intends to give to the customer the vehicle group he or she has requested. However, this is not always possible. In these cases, it is common to offer the customer a vehicle for the same price but from a higher-valued group, that is, to offer an upgrade. If this is not possible, the customer is offered a vehicle from a lower-valued group than requested but at a discount, that is, a downgrade is offered.

In some instances, such as breakdown, accident or cleaning of the vehicle, on delivery of the same, a "preclose" situation is activated to ensure enough time to restore the vehicle to its original conditions. For these situations, the date the vehicle enters in "preclose" status and the date on which it ends are registered.

3.3 Main objectives

This project intends to build a system that tackles significant amounts of data that arose from the new pricing process the company Car-Rent is implementing, and that helps to understand the demand and pricing behavior of this Portuguese car rental company. The main objectives we aim to address are:

- How to process and analyze the data;
- How to obtain relevant information from the data;
- Understand the relationship between the prices charged by the company and the competitors;
- Understand how occupation levels evolve throughout the season/year;
- Through price elasticity of demand, understand demand;
- Understand the factors that impact the price elasticity of demand.

(a) Vehicle from Group C

(b) Vehicle from Group MINI

(c) Vehicle from Group H

(d) Vehicle from Group G

(e) Vehicle from Group L

(f) Vehicle from Group Z

(g) Vehicle from Group N

(h) Vehicle from Group O

(i) Vehicle from Group E

(j) Vehicle from Group J

(k) Vehicle from Group F

(l) Vehicle from Group SUVC

(m) Vehicle from Group SUVM

(n) Vehicle from Group 02

(o) Vehicle from Group 03

(p) Vehicle from Group 04

(q) Vehicle from Group 05

Figure 3.1: Car-Rent Vehicle Groups

Chapter 4

Proposed Solution

The company has a significant amount of raw data encompassing their prices for each type of rental, the prices applied by their competitors for the same type of rental, and the values for the available fleet and occupation. In this case, the company currently does not have the tools to analyze this data and obtain information that can improve decisions on pricing and fleet management, or information about the market and the position of the company in it. A flexible and scalable system is required that allows for various types of analysis according to the needs of the company.

4.1 Data Analysis Support System design

The solution proposed in this work is the design and implementation of a Data Analysis Support System. The system collects data from the diverse pricing, fleet, and occupation files of the company and loads them into a database. With this database, it is possible to establish a connection with several data analysis tools. The analyses carried out on these tools demonstrate the versatility and capabilities of the system, and obtain managerial insights as well, which the company can use to support their decisions. The Figure 4.1 from page 17, presents the Data Analysis Support System, as well as its inputs and outputs.

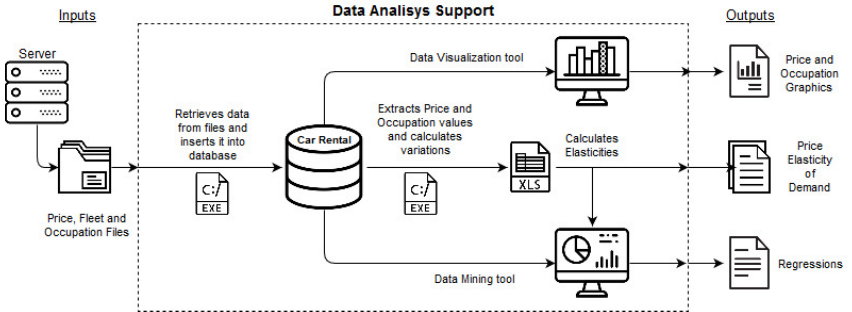

Figure 4.1: Data Analysis Support System, icons made by [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]

The methodology used in the development of this system will be described following its modular structure, comprehending database, data extraction, connection to the data analysis tools, price extraction and calculation of elasticities.

4.2 Data handling

In this section, the location and structure of the data will be described. Moreover, the attributes of the data contained in the files will be explained.

4.2.1 Location and structure of files folders

The text files containing the data are stored on the company server. Since there was a significant number of files, and their exhaustive analysis would have significant time implications, the files extracted to use as design validation and as inputs to the proof-of-concept analyses belong to four periods of the year 2018, derived from the four seasons in a year defined by the company. These periods are: February 19th to March 4th (low season), May 7th to 20th (medium season), July 2nd to 15th (high season) and July 30th to August 12th (high season). The files are distributed in a specific folder structure or repository. If the file is a price file, the repository is organized by pick-up zone of the rental vehicle, broker website from which prices have been withdrawn, duration of the rental in days, and start date of rental. For fleet files, the repository is organized by pick-up zone of the rented vehicle, and start date of rental.

4.2.2 Data contained in the files

All files containing relevant data are text files, specifically Comma Separated Values (CSV) files. These contain a list of data separated by commas and are useful to exchange data between different applications without significant compatibility issues.

4.2.2.1 Price files

Car-Rent represents several brands; one of them is the low-cost brand Car-Low-Cost. The system extracted the data concerning Car-Low-Cost, but no analysis was performed with it. Since the company intends to discontinue this brand, the aim was only to verify that it is possible to use the values of several brands in the system and not to pursue a specific analysis for them. For price files, the data list includes:

- The date and time at which the price was registered;
- The website or broker from which prices were withdrawn;
- The vehicle group;
- The pick-up zone;

- The start date of rental;
- The duration in days of the rental (only the durations of 1, 3, 5, 7, 15 and 28 days have differentiated prices);
- The price for Car-Rent's brand calculated with the new pricing system;
- The margin applied by the broker to Car-Rent;
- The price for Car-Low-Cost's brand calculated with the new pricing system;
- The margin applied by the broker to Car-Low-Cost;
- The date of the pricing decision (equal to the date of registration for all registers);
- The price for Car-Rent's brand calculated with the old pricing system;
- The price for Car-Low-Cost's brand calculated with the old pricing system;
- The minimum price applied by the competitors of Car-Rent (in the same circumstances);
- The minimum price applied by the competitors of Car-Low-Cost (in the same circumstances).

Due to the significant amount of price files, only the prices for a website broker, Auto Europe (France), were extracted. At the same time, the company believes this broker applies the contracted fixed margins while others have dynamic and variable margins, which hinders the accurate analysis of the effects of pricing decisions. Therefore, when analyzing the elasticities, only reservations that come from the Auto Europe website are considered for the occupation of the fleet.

4.2.2.2 Fleet and Occupation files

Before describing the contents of the fleet and occupation files, it is essential to explain the distinction between fleet (i.e. available fleet) and occupation. The fleet refers to vehicles available for reservation through the website of online brokers at the time in question. The company calculates the fleet value, as seen in Figure 4.2 from page 20, using accessible data on the total fleet available (excluding reserved or blocked vehicles) and a parameterization of vehicles that must be kept free for Corporate and Direct customers, as well as other international customers using different sales channels. These values may present some errors in specific dates due to issues on extraction from the fleet management system and changes in the parameterization. As for the occupation value, it refers to the number of vehicles already booked for reservations made by international clients through broker websites.

The fleet repository contains the fleet and occupation files, with one file for each vehicle group and date. The fleet files contain the following data:

- The zone to which the fleet belongs (and where the booked vehicle will be picked-up);

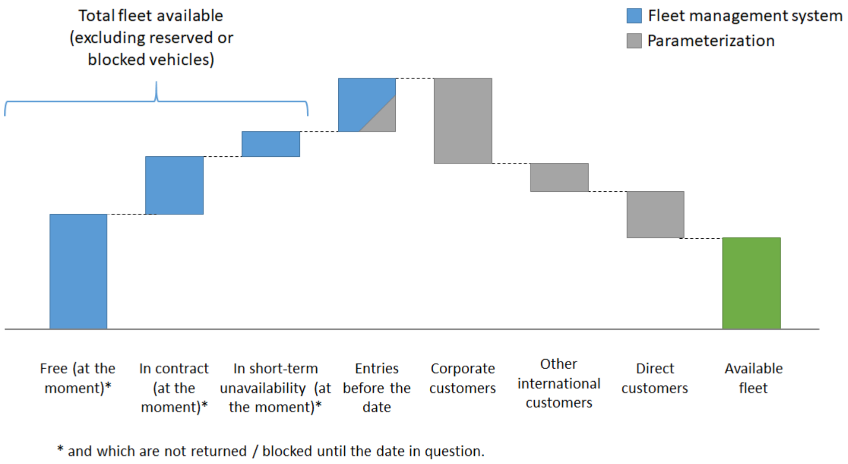

Figure 4.2: Calculation of available fleet

- The date for which the fleet value is being registered;
- The vehicle group (only one per file);
- The date on which the fleet value was registered;
- The calculated value.

The occupation files include:

- The zone where the booked vehicles will be picked-up;
- The date for which the occupation value is being registered;
- The vehicle group (only one per file);
- The date on which the occupation value was registered;
- The broker where the reservations were booked;
- The occupation value for the date in question.

The broker referred to in the occupation file is a general broker (*SuperBroker*), that is, it includes websites of the same broker from different countries (France, UK and Germany). However, the website broker referred to in the price file concerns a single country (France).

4.3 Modular methodological approaches

In this section, the methods and procedures used to assemble the system will be described as well as the possible analyses that can be made with the system. The section is organized following the different components or modules of the system: creation of a database schema, extraction from price and fleet files, connection to a visualization tool, extraction of price and occupation variation, calculation of elasticities and, for last, connection to a data mining tool.

4.3.1 Database schema creation

To more efficiently handle and access the data for analysis, which is found in several files spread over several folders, and allow the connection to various data analysis tools, a database schema was created. The database schema *car rental* was created in MySQL and consisted of six tables: *Broker*, *Fleet*, *Occupation*, *Price*, *Rental* and *Vehicle*, according to the data from the files. The *Vehicle* and *Broker* tables are auxiliary tables in the sense that they contain additional information about the groups of vehicles and brokers used.

The table *Broker*, in Figure 4.3 from page 21, consists of three columns: *ID*, *Superbroker* and *Broker*. This table contains the name of the brokers most used by international clients (*Superbroker*) and their divisions by country (*Broker*). The column *ID* is the primary key of the table.

Broker	
PK	**ID**
	Super Broker
	Broker

Figure 4.3: SQL table Broker

The table *Fleet*, in Figure 4.4 from page 21, consists of six columns: *ID_Fleet*, *ID_Vehicle*, *Zone_Fleet*, *Date_Fleet*, *Register_Fleet* and *Value_Fleet*. This table indicates the value of the company's fleet (*Value_Fleet*) for a specific rental zone (*Zone_Fleet*), date (*Date_Fleet*), and vehicle group (*ID_Vehicle*). The latter is a foreign key for the table *Vehicle*. The column *Register_Fleet* contains the date in which the fleet value was registered. The column *ID_Fleet* is the primary key of the table.

Fleet	
PK	**ID Fleet**
FK	**ID Vehicle**
	Zone Fleet
	Date Fleet
	Register Fleet
	Value Fleet

Figure 4.4: SQL table Fleet

The table *Occupation*, in Figure 4.5 from page 22, consists of seven columns: *ID_Occupation*, *ID_Group*, *Zone_Occupation*, *Date_Occupation*, *Register_Occupation*, *Value_Occupation* and *Superbroker*. This table indicates the value of the company's occupied fleet (*Value_Occupation*) for a specific rental zone (*Zone_Occupation*), vehicle group (*ID_Group*) and date (*Date_Occupation*). The latter is a foreign key for the table *Vehicle*. The column *Register_Occupation* contains the date in which the occupied fleet value was registered and the *Superbroker* column indicates through which broker the reservation was made. The column *ID_Occupation* is the primary key of the table.

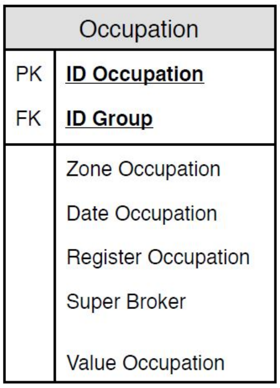

Figure 4.5: SQL table Occupation

The table *Price*, in Figure 4.6 from page 22, consists of seven columns: *ID_Price*, *ID_Rental*, *PriceCar-Rent*, *PriceCar-LowCost*, *PriceCar-RentComp*, *PriceCar-LowCostComp* and *Advance*. This tables indicates the prices calculated by the company (*PriceCar-Rent* and *PriceCar-LowCost*) for a specific rental type and the minimum prices of the company's competitors (*PriceCar-RentComp* and *PriceCar-LowCostComp*). The table has a foreign key (*ID_Rental*) for the table *Rental*, that characterizes the rental type. The column *Advance* indicates the advance of the price, i.e. the number of days missing to the start of the rental when the price was calculated (and made available to the end-customers). Finally, the column *ID_Price* is the primary key of the table.

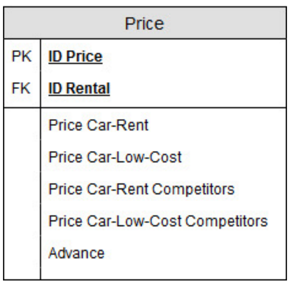

Figure 4.6: SQL table Price

The table *Rental*, in Figure 4.7 from page 23, consists of six columns: *ID*, *Broker*, *Group*, *Zone*, *Beginning* and *Length*. This tables characterizes the types of rentals by broker (*Broker*), vehicle group (*Group*), rental zone (*Zone*), start date of the rental (*Beginning*) and duration in days of the rental (*Length*). The columns *Broker* and *Group* are foreign keys for the table *Broker* and *Vehicle*, respectively. The primary key of the table is the column *ID*.

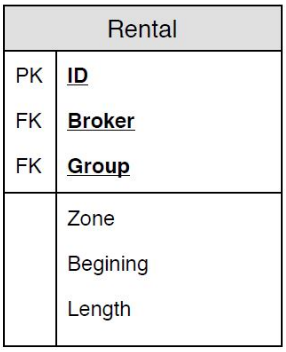

Figure 4.7: SQL table Rental

The table *Vehicle*, in Figure 4.8 from page 23, consist of two columns: *ID* and *Group*. The first column is the table's primary key, and the second column is the vehicle group. For future works, it is possible to add more columns to this table that will represent additional characteristics of the vehicles, such as the vehicles groups' hierarchy to analyze vehicles' upgrades.

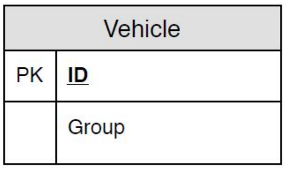

Figure 4.8: SQL table Vehicle

4.3.2 Extracting the data

A program in C++ was developed to extract the relevant data from the price and fleet text files, retrieved from the company server. The program searches for the folders of the corresponding fleet, occupation and price files, opens them, and reads and inserts the data in the adequate database table.

The folders are organized by pick-up zone (*Faro, Funchal, Lisboa, Porto*), broker (*Rental Cars, Auto Europe*, among others) and rental duration (for example, 1, 3 or 5 days) for the price files. For the fleet and occupation files, the folders are organized by the pick-up zone and start date of rental. Since the new pricing system is still in the test phase, the prices that will be extracted are the prices of the old system, both for Car-Rent and Car-Low-Cost, since these are the ones to which the end-users had access and that influence the behavior. The date used for the analyses was the decision date.

Two libraries were needed for this program: the MySQL library to connect and make queries to the database schema, and the Boost Filesystem library [25] that provides methods to manipulate files, directories, and their paths. The program has eight functions, to get the list of files for each directory, read the data, and insert it into the database tables.

The function *getALLFilesInDir* receives a path directory as an argument, either of a price folder or of a fleet folder, and returns a vector containing all the paths to the files in the given directory as well as its subdirectories. Two functions were created to retrieve the data from the files: *getFilesPrices* for price files, and *getFilesFleet* for fleet files. These functions receive a vector containing all the paths to their corresponding files and a text file. In the text file are written the names of the price and fleet files read, and the time it took to insert the data into the database tables for each.

In the *getFilesPrices* function, for each file, the function *getData* is called to read the file and return the data in a vector of vectors. This function iterates through each line of the file and splits its content using a delimiter received as an argument of the function. The delimiter is needed because the price and fleet files are CSV files, so the delimiter used in the function is a comma. After receiving the file data, for each line of the file that is read, the data is processed. The advance in relation to the start of rental with this price is calculated, and the format of the date is changed.

After processing the data, it is inserted into the database through two functions: *insertIntoRental* and *insertIntoPrice*. The first one receives as arguments the characteristics of a type of rental, the group, the broker, the pick-up zone, the start date, and the duration in days. The function begins by establishing a connection to the database. If the connection has been established, two queries are executed on the *Vehicle* and *Broker* table to know the vehicle group and broker *ID* to be used as foreign keys. Next, the existence of this type of rental in the database is verified, through a query to the table *Rental*. If it already exists then the *ID_Rental* is returned, if not then a new type of rental is inserted in the database and its *ID* is returned.

With the returned *ID_Rental*, a function to insert the prices in the database is called: *insertIntoPrice*. The function takes as arguments the four prices (Car-Rent, Car-Low-Cost and the minimum prices of the corresponding competitors), the *ID_Rental* as a foreign key, and the calculated advance in relation to the start of rental with these prices. This function is more straightforward than *insertIntoRental*: after establishing a connection and verifying that it was well established, a query is executed to the *Price* table, inserting the prices and the calculated advance, with a foreign key for the *Rental* table, which characterizes the type of rental to which the prices are linked.

In the *getFleetFiles* function, it is first checked whether the file is a fleet or occupation file. For each file, the function *getData* is called to read every line and separate the data by its delimiter, and then return it in a vector of vectors. After receiving and processing the data, the functions *insertIntoFleet* and *insertIntoOccupation* are called to insert the data into the corresponding database tables.

For the *insertIntoFleet* function, the following data is received as arguments: the pick-up zone, the start date, the group, the date when the fleet value was calculated, and the calculated fleet value. The function starts by establishing a connection to the database and verifying that it was well established. Then, a query is made to the *Vehicle* table to know the vehicle group *ID*. Afterwards, a query is made to the *Fleet* table, where the data received as an argument is inserted.

For the *insertIntoOccupation* function, the following data is received as arguments: the pick-up zone, the start date, the group, the date when the occupation value was extracted, the corresponding broker and the occupation value. The function establishes a connection to the database and verifies that it has been well established. Similarly, to the *insertIntoFleet* function, a query is made to the *Vehicle* table to know the vehicle group *ID* that will be used as a foreign key. Then, a query is made to the *Occupation* table to insert the data received as an argument.

4.3.3 Connection to a Data Visualization tool

Visualization allows understanding the data and what kind of information it may contain. It is, therefore, the first analysis to be carried out. The tool used for this analysis is Tableau, a software that helps people see and understand data, thus transforming the way people use data to solve problems [26]. Tableau saves time to connect and prepare data, and it can connect to various databases or file types. Data analysis with this tool is quick and intuitive. It is possible to create visualizations in the form of dashboards or worksheets, which can be easily understood by professionals of different levels.

After the database is established and completed, it is connected to the Tableau visualization tool. In establishing the connection, two tables are created for the extraction of the data for analysis: one for the fleet and occupation and other for the prices.

To analyze the values of the fleet and its occupation, the following database tables were merged: *Fleet*, *Occupation* and *Vehicle*, as seen in Figure 4.9 from page 25. The *Fleet* table has a one-to-many relationship with the *Occupation* table: for a certain zone and date, the fleet value corresponds to multiple occupation values (a value for each broker). Therefore, the values of the zone, date, and register are the same in both tables. These two tables have a one-to-one relationship with the *Vehicle* table, where the *ID_Vehicle* and *ID_Group* are foreign keys of the primary key of the *Vehicle* table.

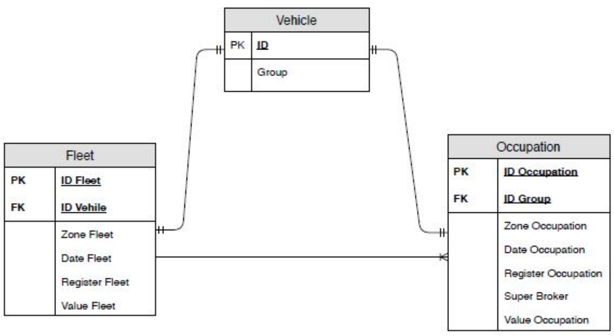

Figure 4.9: Scheme of the fleet and occupation table

To analyze prices, four tables were used: *Vehicle*, *Rental*, *Price* and *Broker*, as seen in Figure 4.10 from page 26. The *Rental* table has two foreign keys with the *Vehicle* and *Broker* tables, forming a one-to-one relationship with them. The *Price* table has a many-to-one relationship with the *Rental* table, that is, for a type of rental, there are several prices, and the *Price* table has a foreign key *ID_Rental*.

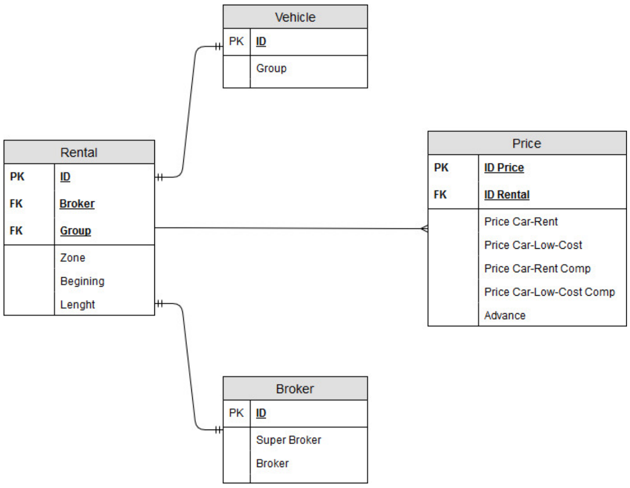

Figure 4.10: Scheme of the price and rental table

4.3.4 Extracting prices and occupation values

To calculate the elasticities, it was necessary to know the percentage variation of price and occupation. To do so, two programs were created, one to extract prices and calculate price variations and other to extract occupations and calculate occupation variations, from the database tables to an Excel file where the percentage variations will be calculated, one for each program. Since there was a constraint in the time available for the analyses, the values where extracted for four pick-up zones (*Lisboa, Porto, Faro and Funchal*), four vehicle groups (the H, J, C, and MINI groups were chosen) and for the period of July 2nd to 15th, 2018, which belongs to the High Season. The pick-up zone, vehicle group and period can be changed, making the programs versatile and allowing multiple analyses to be run.

The extraction contemplated the values of the last 30 days before the rental start date. Variations with a one-day interval were analyzed, so it was assumed that occupation changes and price changes corresponded to the same day, that is, customers observe the price for a particular day and make a reservation. However, a future analysis may also include the case of customers studying price changes and booking a vehicle only after weighing the price variation and strategically anticipating price changes. So, there may exist price and occupation variations with intervals longer than one day and considering trends.

The occupation value includes all rentals of different durations that begin on the start day, which we will call "occupation day", but also rentals that began on previous days but whose duration makes them last at least until the start day (i.e. end after the "occupation day"). Therefore, the various prices of different rentals (with different rental durations), which we will call "price days", will influence the occupation value.

The occupation values are extracted for the last 30 days before the rental start date, and the advance in relation to the occupation day with these occupation values is called occupation advance. For prices, their advance is in relation to the price day with these occupation values, and it is called price advance. Figure 4.11 from page 27 illustrates an example of a 5-day rental that begins on June 30th, 2018, which is the price day, and that influences the occupation day of July 3rd, 2018. On the 23rd of June, the occupation value for the occupation day (3rd of June) is extracted with ten days of advance. Nevertheless, the price for this rental that influences it is extracted with seven days of advance.

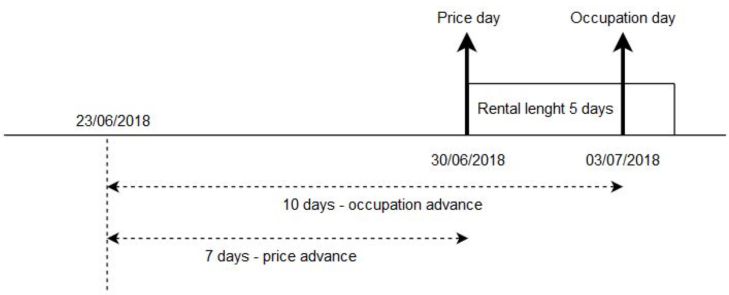

Figure 4.11: Example of a 5-day rental beginning before the start day

4.3.4.1 Extracting Occupation

The program begins with a function that creates an Excel file where the occupation values extracted and occupation variations calculated will be stored. The values are extracted for each pick-up zone, vehicle group, occupation day, and occupation advance (from 30 to 0 days), forming four cycles. At each iteration of the last cycle, the *extractOccupation* function is called to extract the occupation value, receiving as arguments the pick-up zone, the occupation day, the group, and the date on which the occupation value was registered (calculated with the occupation day and occupation advance).

The function *extractOccupation* starts by establishing a connection with the database. After verifying that the connection was successful, it executes a query to the table *Occupation* with the values received as arguments, returning the occupation value. The occupation value returned can be zero if no value is found in the database table, or if the value registered in the occupation file was in fact zero. Therefore, in this case, the occupation value is corrected, and it is assumed to be equal to the previous occupation value other than zero.

With the returned occupation value, the absolute occupation variation can be calculated by subtracting the occupation value to the previous occupation value. Then, a function is called that inserts the values into the Excel file having as arguments the occupation day, the occupation advance, the occupation value extracted, the occupation value corrected, and the occupation variation. For the first occupation advance (30 days) the variation is considered to be zero. Algorithm 1 from page 28 presents the pseudo-code for the algorithm of the occupation values extraction program.

```
create Excel file;
for each pick-up zone do
    for each vehicle group do
        for each occupation day do
            for each occupation advance do
                Extract occupation value;
                Calculate occupation variation;
                Write to Excel file;
            end
        end
    end
end
```

Algorithm 1: Extracting Occupation algorithm

4.3.4.2 Extracting Price

This program begins with a function that creates an Excel file where the price values extracted, and price variations calculated will be stored. The values are extracted for each pick-up zone, vehicle group, occupation day, occupation advance (from 30 to 0 days), rental duration (1, 3, 5, 7, 15 and 28 days) and price day, forming six cycles. The price days are calculated with the occupation day and rental duration and include the days from the occupation day minus the rental duration until the occupation day. For example, for a rental duration of three days, the price days include the period from three days before the occupation day until the occupation day.

The occupation advance is not the same for the price days, so the price advance needs to be calculated. The price advance is calculated by subtracting the difference in days of the occupation and price day to the occupation advance. For example, if the difference between the price and occupation days is three days, then the price advance is equal to the occupation advance minus three. However, if the occupation advance is inferior to the rental duration, the price advance would be negative for some price days. Therefore, for these cases, the price days can only include the days from the occupation day minus the occupation advance until the occupation day. Figure 4.12 from page 29 illustrates an example of a 7-day rental that begins on June 27th, 2018, where the occupation day is July 3rd, 2018, and the occupation advance is four days. In this example the price days are from June 29th to July 3rd, 2018, since the first days of the rental are past the occupation advance.

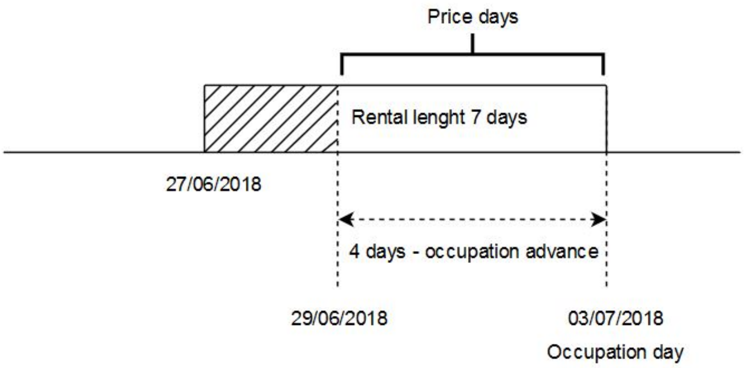

Figure 4.12: Example of a 7-day rental beginning before the start day

Before the last cycle, for each price day, the program verifies if the occupation advance is inferior to the rental duration. This will determine the price days for the last cycle. For each iteration in the last cycle, the function *extractPrice* is called to extract the current iteration price and the previous price from the database table and receives as arguments the pick-up zone, the price day, the group, the price advance, and the rental duration. The function establishes a connection to the database and verifies if the same was successfully established. Then, it executes a query to the *Rental* table to determine the *ID_Rental*, which corresponds to the type of rental that has the characteristics received as arguments. After knowing the *ID*, another query is executed on the *Price* table, which returns the price value.

With both the current iteration price and the previous price, returned by the function *extract-Price*, the price variation is calculated. Finally, a function is called to insert the values into the Excel file having as arguments the occupation day, the occupation advance, the price day, the price advance, the rental duration, the price value, and the price variation. Algorithm 2 from page 30 presents the pseudo-code for the algorithm of the price values extraction program.

```
create Excel file;
for each pick-up zone do
    for each vehicle group do
        for each occupation day do
            for each occupation advance do
                for each rental duration do
                    if occupation advance < rental duration then
                        for each price day do
                            Extract price values;
                            Calculate price variation;
                            Write to Excel file;
                        end
                    else
                        for each price day do
                            Extract price values;
                            Calculate price variation;
                            Write to Excel file;
                        end
                    end
                end
            end
        end
    end
end
```

Algorithm 2: Extracting Price algorithm

4.3.5 Calculating elasticities

The calculation of the elasticities allows obtaining managerial insights that the company can use to understand how increasing or decreasing their prices will affect their demand. The elasticity is calculated by dividing the percent change in occupation by the percent change in price. Through the price and occupation extraction program, absolute variations in occupation and price were calculated (*final occupation-initial occupation*, and *final price-initial price*), and it was considered a variation period of one day. Dividing these variations by the initial occupation and price values, it is possible to obtain the percent variation of occupation and price needed. The elasticities are calculated for each pick-up zone, start day, and advance. However, since there were multiple variations of price and occupation, it facilitates the analysis to calculate an average of the variations. The procedure followed to perform these elasticity calculations is presented in 4.1 from page 31.

$$price\ elasticity\ of\ demand = \frac{\%\ change\ in\ occupation}{\%\ change\ in\ price} = \frac{\frac{occupation\ variation}{initial\ occupation}}{\frac{price\ variation}{initial\ price}} \tag{4.1}$$

Following the law of demand, if no other factors affect the demand (which is a strong assumption in this context), then the price and demand would be inversely related. Therefore, the elasticities would always be negative. However, there can be goods that do not behave according to the law of demand. For example, when clients are in desperate need of a vehicle and scared of a shortage of supply, they continue to make reservations despite increases in price. Therefore, the analyses in this work will not assume a negative signal or module for elasticities, since some of the vehicle groups may not behave accordingly to the law of demand.

The occupation variation can show either an increase or a decrease in the occupied fleet, in which the increase represents a growth in the number of reservations, and the decrease represents a loss in reservations (i.e. cancellations). The elasticities allow an understanding of how a variation in price affects the occupation; nonetheless, a decrease in occupation is more likely to be related to other factors rather than price variation, such as "change of minds" of the clients. Therefore, an analysis was performed in which decreases of occupation were replaced by zero. With this price variations will only affect positive occupation variations.

As for the results of the elasticity calculations, three different analyses were performed to accommodate different interpretations of the elasticity results when the price variation is null. The elasticity intends to show how the variation of the quantity of occupied fleet responds to a variation in price. However, if there is not a variation in price, a variation in quantity will presumably not be related to the variation in price.

The difference between the three analyses is related with the treatment of cases where the occupation variation was other than zero (the numerator on the elasticity calculation) and the price variation was zero (the denominator). According to the definition of elasticity, if there is no price variation it cannot indeed cause the occupation variation. If these cases exist in a large number, the way they are mathematically considered in the calculation of the average may have a significant impact on the results. Therefore, three different ways were selected and compared. For all of the analyses, if both the percentage of occupation and price variation were null, the resulting elasticity was not considered when calculating the average. In the first analysis, in these cases, the elasticity was considered to be zero. The second analysis, contrarily, considers that the elasticity tends to infinity. To achieve this, the price variation was considered to be a marginal value so that the elasticity could be calculated as a very large value. The third analysis ignores these values when performing the elasticity average.

The limitations in the calculations are related to the use of the average since not all the values for different advances may have the same weight in the final elasticities. The same conclusion is valid for price variations: not all rental durations will have the same influence on the variation of occupation. Nevertheless, for now, there is a lack of data to verify what impacts and how it affects the elasticities.

4.3.6 Connection to the Data Mining tool

Considering the foregoing limitation, a regression analysis was performed to verify which variables may impact the elasticity. The RapidMiner software was used for these analyses. RapidMiner is a data analysis tool that provides an intuitive framework for several types of analysis using different data handling and data mining techniques, which may be significantly useful also for future work. It is capable of loading, accessing, and analyzing most types of data. Moreover, it is easy to use, builds efficient models, and accurately estimates the performance of the models [27].

The procedure to connect the other modules of the Data Analysis Support System to RapidMiner is as follows. The elasticities calculated in Excel are loaded into a table, and the day of the week and month of the start date are generated by the tool. The regression process, represented in Figure 4.13 from page 33, starts by loading the obtained table and replacing possible missing values of elasticities by the value - these missing elasticities values are the ones that were not considered when performing the average calculation in the analyses, the third analysis referred in 4.3.5, and will be filtered out during the regression process. To perform the linear regression, it is necessary to define the dependent variable - also called the “label” in RapidMiner - using the *Set Role* block. The label is the elasticity value since we want to understand which independent variables affect the elasticity.

The *Select Attributes* block selects the independent variables for the calculation: the day of the week, the day of the month, the advance, and the group. With the day of the week, we pretend to understand if there is a change in the elasticities during weekdays or weekend days. Moreover, since the month analyzed is July, the days closer to the middle of the month may impact the elasticity differently from the first days of the month. As for the advance, the days closer to the rental start may affect more the elasticity than the ones further away.

The vehicle groups analyzed have different characteristics: a Group H vehicle is a luxury vehicle, while those belonging to Group C are economical vehicles. Consequently, we expect elasticity to have a different behavior depending on the vehicle group. Therefore, the *Filter Examples* block filters the data by vehicle group as well as removing the elasticities equal to -1. For the regression, all values must be numerical, so the *Nominal to Numerical* block is used to convert these attributes through dummy coding.

The last block in the regression process is the *Cross-Validation* block. This block, presented in Figure 4.14 from page 33, is divided into a training phase, where the linear regression model is calculated, and a testing phase, where the linear regression model performance is evaluated by the absolute, relative, root mean squared and squared errors calculated by the *Performance block*.

In the *Linear Regression* block, the option "eliminate colinear features" was selected. With this option, the algorithm tries to eliminate colinear features during the regression, considering a minimum tolerance of 0.05, and therefore, not all the independent variables being tested could be presented in the linear regression results.

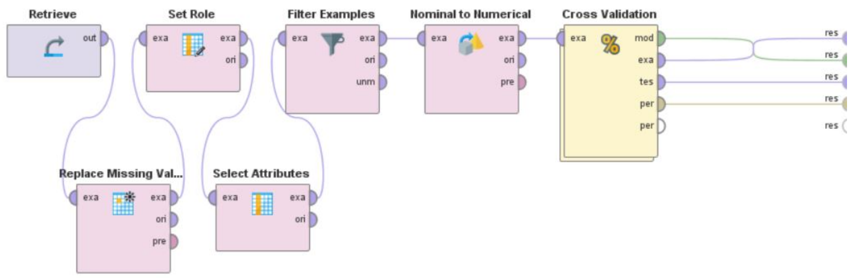

Figure 4.13: Regression Process

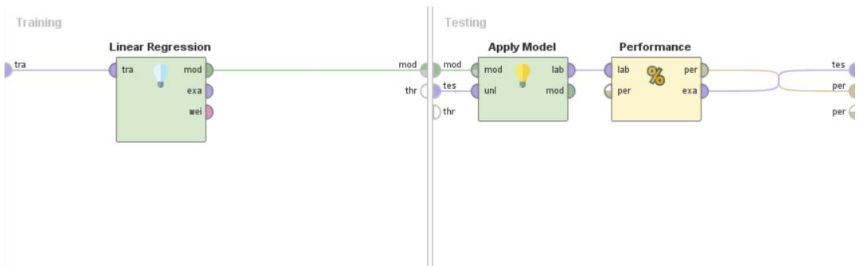

Figure 4.14: Cross-Validation block

Chapter 5

Analyses

This chapter describes the analyses performed as a proof-of-concept of the capabilities and functionalities of the Data Analysis Support System, which allowed drawing relevant managerial insights for the company. In each analysis, the objective, results, data used, limitations, and conclusions will be presented. This chapter is divided into three parts corresponding to the visualization, elasticity, and regression analyses.

5.1 Visualization

Focusing on the pick-up zone of Lisbon, the analyses compared two periods with higher and lower demand. The period with higher demand corresponds to a high season period, and the period with lower demand to a middle season period.

5.1.1 High Season analyses

The high season period analyzed is from July 2nd to 15th, 2018. First, the analyses of the occupied fleet for this period will be described, followed by the analyses of prices. The competitors' prices are the minimum price being applied in the market. Thus the prices can be from different competitors.

5.1.1.1 Occupation for each broker

To understand if the vehicles reserved are related to the broker website used for making the reservation and if the broker has a significant impact on how the occupation variations throughout the advance days, the following analysis was conducted.

The occupation value is distributed by advance (in days) and divided by super-broker and group. In Figure 5.1 from page 36, we can conclude that *Rental Cars* has the largest occupation values. This broker has around 100 more occupied vehicles in the maximum day than *Auto Europe* for Group C. Indeed, this group is the favorite among costumers regardless of the broker, since this is an economic group. Moreover, the fluctuations of occupation for each broker are very similar

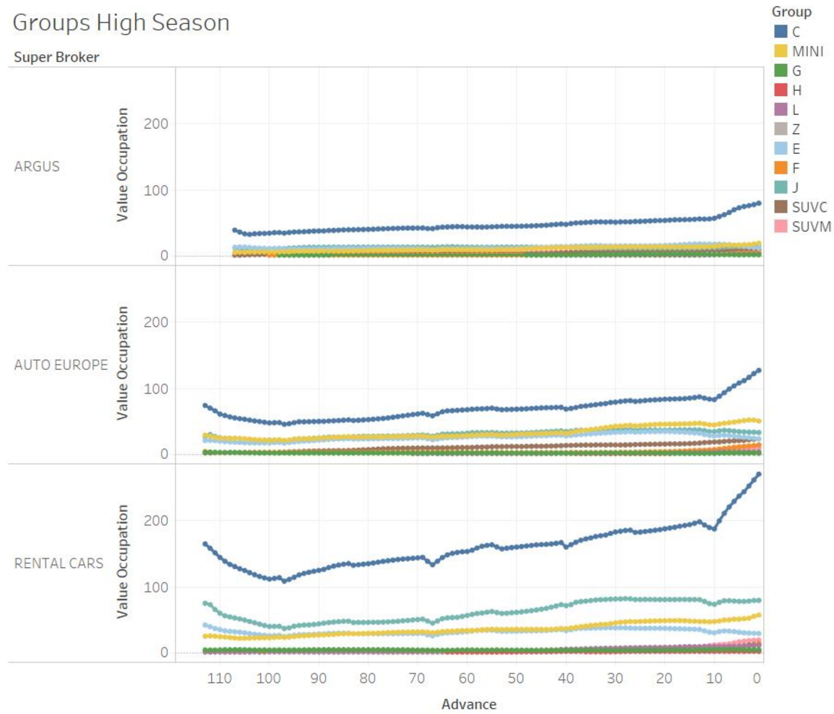

Figure 5.1: Occupation value for different brokers

between them, especially for group C, so the broker does not seem to have a significant impact on the percentage variation in occupation but has an impact on the absolute occupation value.

5.1.1.2 Occupation for different vehicle groups categories

To understand how occupation varies throughout the advance days for the three categories of the vehicle group (economic, SUV, and luxury), the following analysis was conducted.

To obtain the graphic presented in Figure 5.2 from page 37, the occupation values were distributed by advance days and divided by categories of the vehicle group. As expected, luxury groups have the lowest occupation values, since they are more expensive than the other groups. In fact, the cheapest vehicles have the largest occupation values. Groups C, J, and E have similar variations of occupation except when the date is closer (smaller antecedences) where C has an increase in reservations while the others have losses. These probably are not related to a change in price but are last-minute cancellations. In general, all groups gradually receive reservations, increasing their occupation values as the rental start date approaches. Around the last ten days, there is a more significant increase in occupation, corresponding to the last-minute reservations.

5.1.1.3 Price for each group

The following analysis intends to understand which groups have more price variations and how those are related to occupation values.

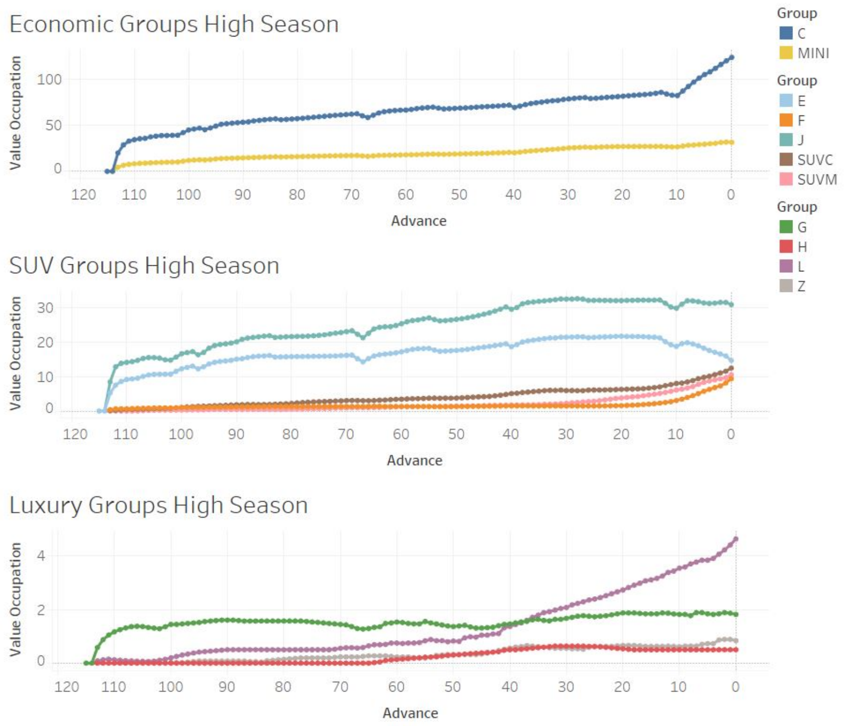

Figure 5.2: Occupation values for vehicle group categories

The graphic presented in Figure 5.3 from page 38 shows the Car-Rent price distributed by advance and divided by group. As a direct observation, the luxury groups have the highest price, and the economic groups have the lowest price, as expected and according to the occupation values observed, vehicles with a smaller price have higher occupation values. Moreover, for group H, the price starts very high, almost 350€ higher than the group below. In general, prices start to be constant for some days, and only around 75 days before the rental begins, they start to have variations. Price variations happen again in the last month before the rental starts and in the last few days as well, when this variation is generally an increase in price. The price increase in the last day may be the result of a strategy of the company, since there are always last minute clients that are in great need of a reservation, so they are willing to spend a little more.

The groups J and SUVC are the only groups that present a decrease in price in the few days before the rental begins. Nevertheless, their occupation does not have a higher increase in the corresponding days, compared with the other groups. This decrease could be a way to get more reservations, but the clients did not seem to have responded to it. Compared with other groups, the economic group C seems to have a smaller variation of price. However, besides having some occasionally decreases in occupation, its value gradually increases with advance. It is a possible scenario that clients reserve group C regardless of price changes because it is one of the cheapest groups.

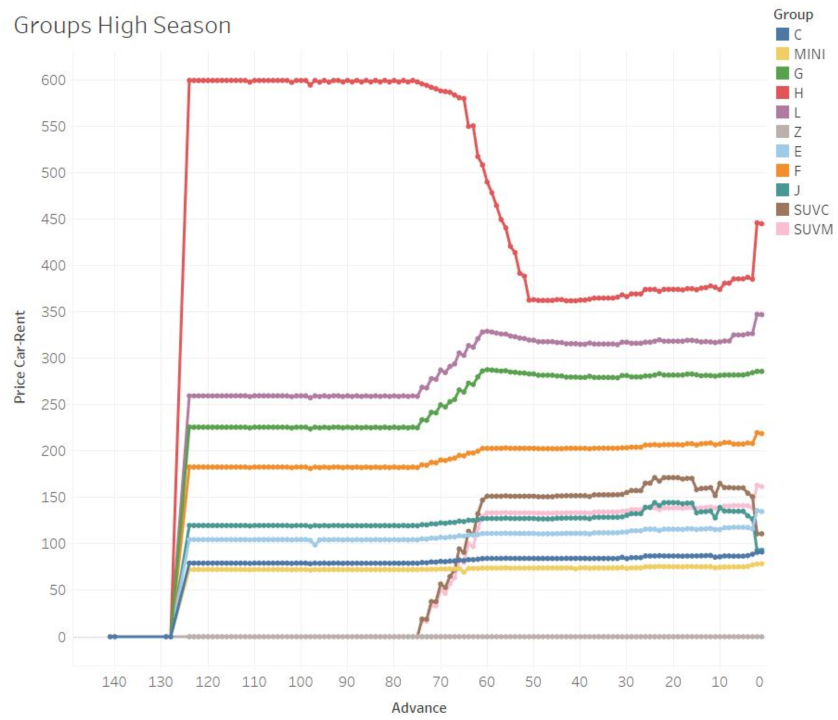

Figure 5.3: Price values for Car-Rent

5.1.1.4 Price Car-Rent and competitors for economic groups

Focusing on the economic groups, how the price variation for Car-Rent and its competitors behaves throughout the advance days was analyzed.

Figure 5.4 from page 39 shows the average price values for Car-Rent and its competitors for economic groups (C and MINI). The prices were distributed by advance and divided by economic group. In a darker color, the figure presents the price value for Car-Rent and in a lighter color the price for the competitors. Group C is represented in blue and group MINI in orange. Until around 75 days of advance, the prices for Car-Rent are quite constant and below the competitors' values. As the prices of the competitors start to decrease, Car-Rent increases theirs, mostly during the last month and days before the reservations start, when more reservations are expected. In the last few days, the competitors also increase their price, similarly to Car-Rent. It seems that both Car-Rent and its competitors are aware of market prices, tending to follow them, and having similar pricing behaviors. Car-Rent's pricing behavior, of increasing their prices when the competitors lower theirs, may be due to a starting under-valued price, which has had to be increased to compensate losses. Alternatively, it may be the result of a longer-term strategy, since group C is the most requested group and the closest to the due date gets, the more "desperate" and willing-to-pay clients become.

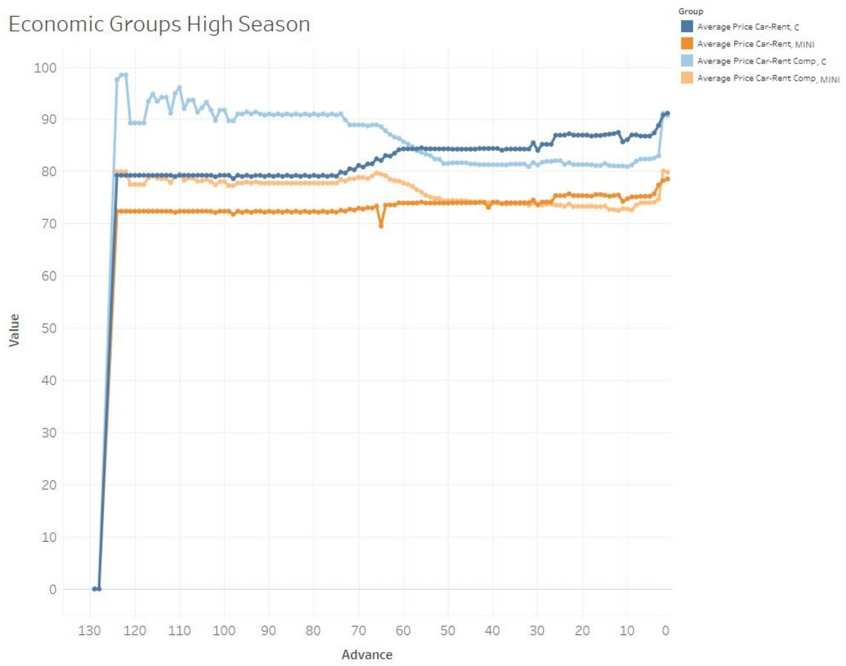

Figure 5.4: Price values of Car-Rent and price of its competitors for economic groups

5.1.1.5 Price Car-Rent and competitors for luxury groups

Focusing on the luxury groups, how the price variation for Car-Rent and its competitors behaves throughout the advance days was analyzed.

Figure 5.5 from page 40 shows the average price values for Car-Rent and its competitors for luxury groups (G, H, L and Z). The prices were distributed by advance and divided by luxury group. In a darker color, the figure presents the price value for Car-Rent and in a lighter color the price for the competitors. For group H, Car-Rent's prices start very high compared with other luxury groups and to the competitors' prices. Around 75 days before the start of the rental, prices begin to change. This group decreases its price, and so do the competitors, increasing it a few days before the beginning of the rental. However, Car-Rent's price is always higher than the competitors', even with the decrease, except for the last days, when it does not present a significantly competitive price.

For group G, both Car-Rent and the competitors have the same price until 75 days before the beginning of the reservation, when Car-Rent increases its price - the opposite behavior from the competitors - maintaining it consistently for the rest of the time period. Once more, this behavior may not allow for a significant price competition but may be the result of other strategies. Car-Rent's prices have a similar behavior to the competitors' but are more expensive.

5.1.1.6 Price Car-Rent and competitors for SUV groups

Focusing on the SUV groups, how the price variation for Car-Rent and its competitors behaves throughout the advance days was analyzed.

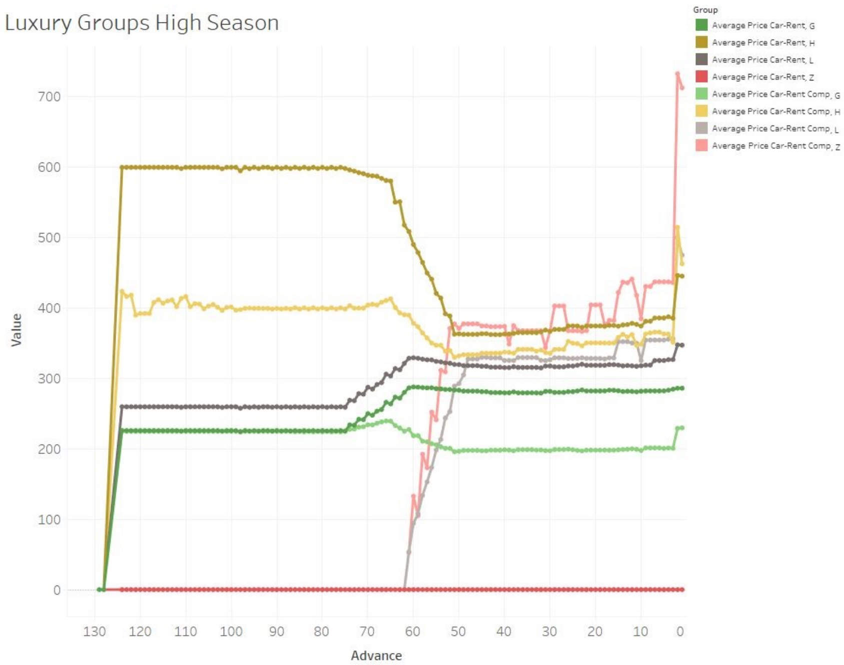

Figure 5.5: Price values of Car-Rent and price of its competitors for luxury groups

Figure 5.6 from page 41 shows the average price values for Car-Rent and its competitors for SUV groups (E, F, J, SUVC and SUVM). The prices were distributed by advance and divided by SUV group. In a darker color, the figure presents the price value for Car-Rent and in a lighter color the price for the competitors. Car-Rent's prices follow the same behavior seen before for the other groups. Until 75 days before the beginning of the reservation, prices are constant. Then, they start to increase while the competitors' prices decrease. A few days before the rentals start, the prices increase again for the last minute clients, except for group J and SUVC. However, for the group SUVC, the competitors' price is significantly higher than Car-Rent's price.

5.1.2 Middle Season Analysis

The middle season period analyzed is from 7th to 20th of May, 2018. First, the analyses of the occupied fleet for this period will be described, followed by the analyses of prices. The competitors' prices are the minimum price being applied in the market. Thus the prices can be from different competitors.

5.1.2.1 Occupation for each broker

To understand if the number of vehicles booked are related to the broker website used for the reservation, and if the broker has an impact on how the occupation varies throughout the advance days, the following analysis was conducted.

For Figure 5.7 from page 42, the occupation values were distributed by advance and divided by group and super-broker. As it was seen in the high season, the broker Rental Cars has higher

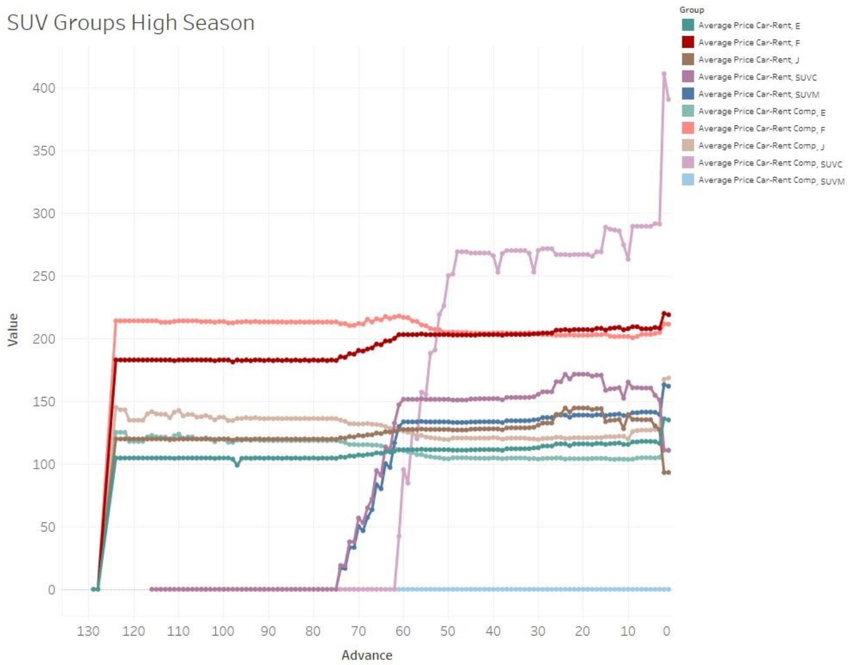

Figure 5.6: Price values of Car-Rent and price of its competitors for SUV groups

occupation values; therefore, this broker may be more popular among clients or have an effective advertisement. Since this is a season with lower demand values, the economic groups are the most booked and, once again, group C is the best-selling group. Moreover, the clients are not pressured to acquire their rentals so early, so the occupation values start appearing around 60 days before the beginning of the rentals, while in the high season these values appeared around 120 days before. The fluctuations of occupation are similar among brokers, so the percentage variation in occupation seems independent of the broker used for the reservation. However, the broker as a significant impact on the absolute occupation value.

5.1.2.2 Occupation for different vehicle groups categories

To understand how occupation varies throughout the advance days for the three categories of the vehicle group (economic, SUV, and luxury), the following analysis was conducted.

To obtain the graphic presented in Figure 5.8 from page 43, the occupation values were distributed by advance and divided by different vehicle group categories: economic, luxury, and SUV. The occupation values for the SUV and luxury groups are not very significant when compared with the values for the economic groups. In general, the occupation increases with the decrease of the advance day, having a higher increase in the last five days before the start of the reservation, half the days from the high demanded period.

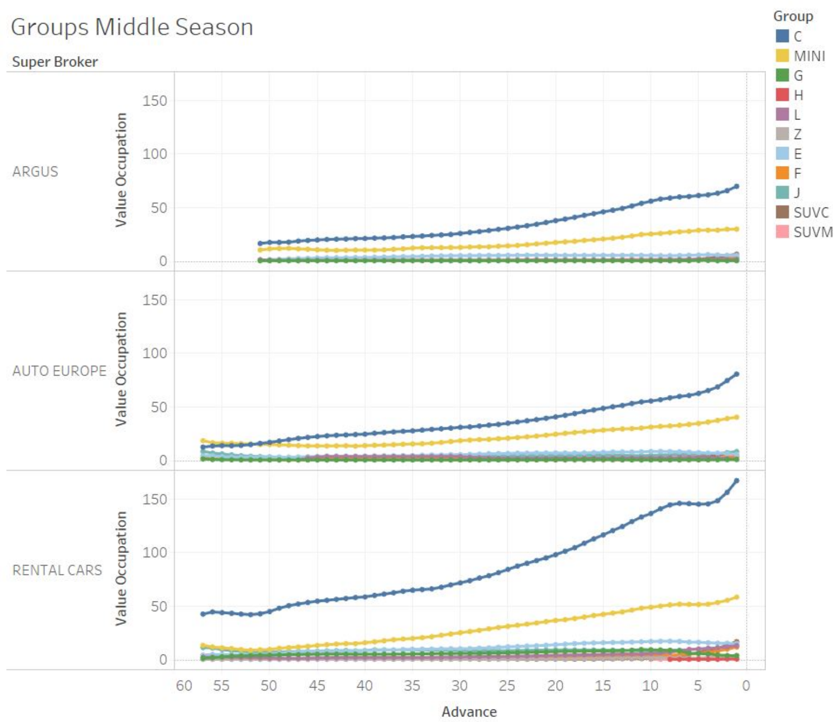

Figure 5.7: Occupation value for different brokers

5.1.2.3 Price for each group

The following analysis intends to understand which groups have more price variations and how they are related to occupation values.

The graphic presented in Figure 5.9 from page 44 shows the Car-Rent price distributed by advance and divided by group. The prices start to have values other than zero around 60 days before the beginning of the rental, which is 60 days less than in the high season period, as seen in Figure 5.3 from page 38. The prices are constant for most of the days until 20 days before the start of the reservation, when they increase or decrease. For all groups, the prices increase in the last five days, as happens with the occupation values. With this, we can conclude that being a high or medium demanded season, both prices and occupations increase in the last days before the beginning of the rental. For the middle season period, group H has a significant difference in price compared with the other prices, the same as the high season period, thus group H may be over-priced.

5.1.2.4 Price Car-Rent and competitors for economic groups

Focusing on the economic groups, how the price variation for Car-Rent and its competitors behaves throughout the advance days was analyzed.

Figure 5.10 from page 45 shows the average price values for Car-Rent and its competitors for economic groups (C and MINI). The prices were distributed by advance and divided by economic

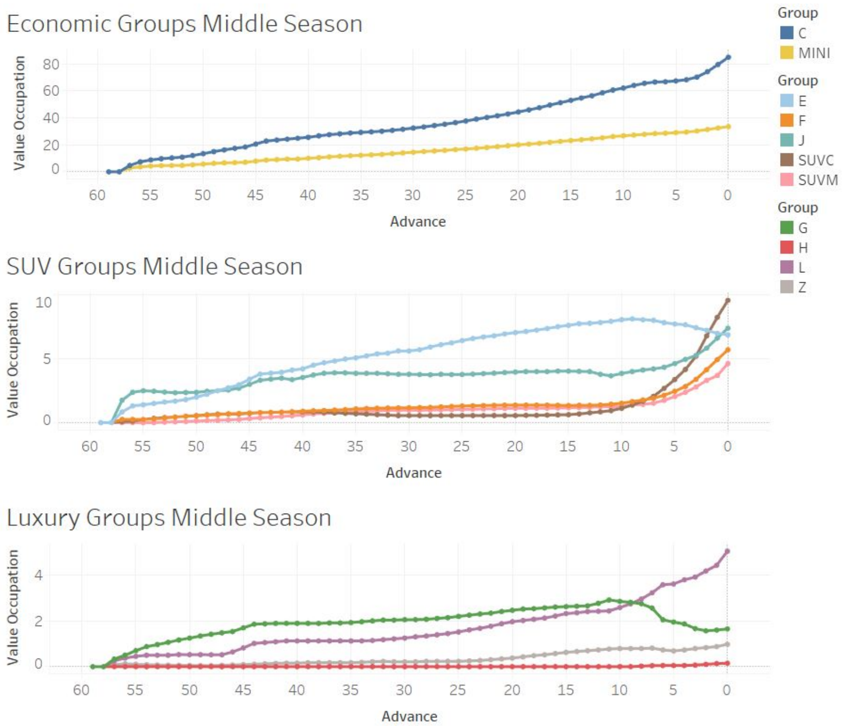

Figure 5.8: Occupation values for vehicle group categories

group. In a darker color, the figure presents the price value for Car-Rent and in a lighter color the price for the competitors. Until 20 days before the beginning of the rental, the prices are constant, and Car-Rent's prices are lower than the competitors. As the competitors decrease their price, Car-Rent does the same but maintaining the price lower. As the rental start approaches, the prices began to increase, and five days before the start of the reservation Car-Rent has now higher prices than the competitors. This behavior shows that Car-Rent is aware of its competitors' prices and follows their variation, or the competitors follow theirs. However, and differently from the high season period, they keep their prices lower than the competitors mostly for the entire season, having a much more competitive price. This change in behavior can be related to the lower demand of the season.

5.1.2.5 Price Car-Rent and competitors for luxury groups

Focusing on the luxury groups, how the price variation for Car-Rent and its competitors behaves throughout the advance days was analyzed.

Figure 5.11 from page 46 shows the average price values for Car-Rent and its competitors for luxury groups (G, H, L and Z). The prices were distributed by advance and divided by luxury group. In a darker color, the figure presents the price value for Car-Rent and in a lighter color the price for the competitors. The prices are constant until 20 days before the beginning of the reservation, when the prices increase or decrease. In the last five days, the prices increase for all groups. Car-Rent's price for group H is significantly higher than the competitors' is, as happens

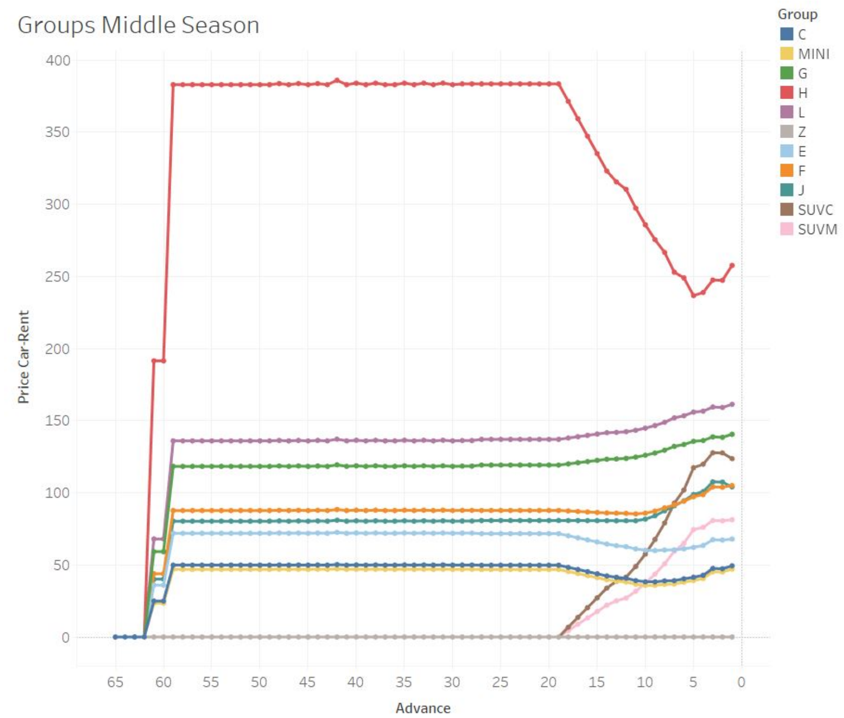

Figure 5.9: Price values for Car-Rent

in the high season. Only in the last ten days its price is lower than the competitors. For group G, Car-Rent's price is set lower than its competitors are until 20 days before the start of the rental. Then, it increases and reaches almost a difference of 50€ in the last days. Having a higher or lower price than the competitors, for these types of vehicles, does not seem to have a major effect since their occupation values for these groups are not significant.

5.1.2.6 Price Car-Rent and competitors for SUV groups

Focusing on the SUV groups, how the price variation for Car-Rent and its competitors behaves throughout the advance days was analyzed.

Figure 5.12 from page 47 shows the average price values for Car-Rent and its competitors for SUV groups (E, F, J, SUVC and SUVM). The prices were distributed by advance and divided by SUV group. In a darker color, the figure presents the price value for Car-Rent and in a lighter color the price for the competitors. Prices for groups E, F and J are constant until 20 days before the start of the reservation. Then, they start to decrease for a few days, similarly to the competitors' prices, and increase again. However, in the last days before the rental begins, they decrease a little since they are slightly higher than the competitors. Car-Rent's prices are again competitive and follow the behavior of the competitors. The SUVC and SUVM groups only start to have price values in the last 20 days, and their prices are significantly higher than the competitors' are.

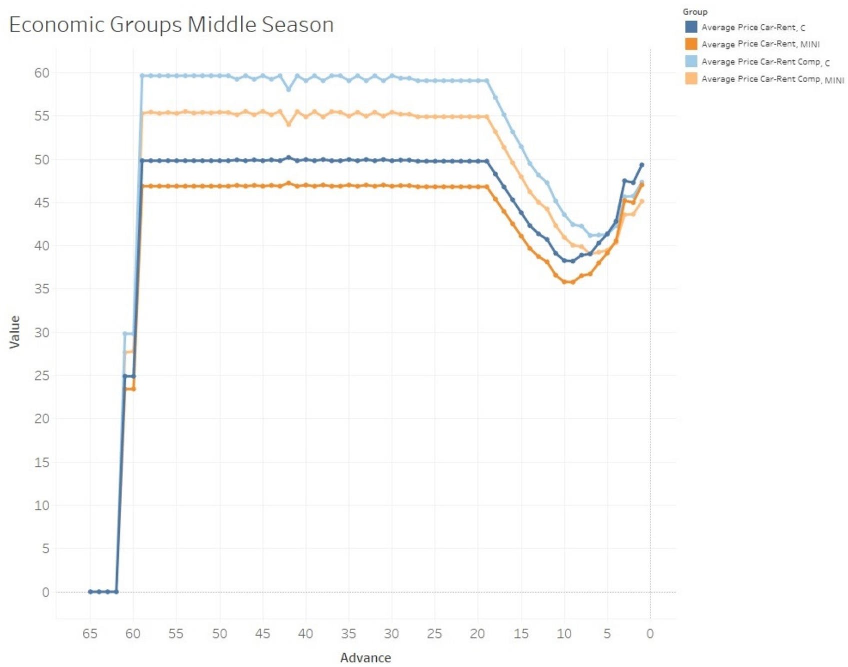

Figure 5.10: Price values of Car-Rent and price of its competitors for economic groups

5.2 Elasticities

Focusing on the pick-up zone of Lisbon and the high season period, the analyses of the elasticities differ on how to incorporate the cases where the price variation is equal to zero in the elasticity calculation . Moreover, an analysis was performed where the percentage of occupation variation is assumed as zero if it is negative, as explained in 4.3.5.

In preliminary tests, the results of the first two types of analyses showed that considering the elasticities as zero or as tending to infinity when the percentage of price variation is zero created a significant bias in the average calculation, limiting the final results, either decreasing or increasing in a disproportionate way the average elasticities values. Therefore, only the third type of analysis will be presented, where these cases were not included in the average calculation.

5.2.1 Group C

For Figure 5.13 from page 48, the price elasticities of demand were calculated, and then an average for each advance day was computed. The two yellow lines mark the area in which the demand is inelastic, according to the literature reviewed in 2.2.2.3 in Chapter 2 ($|\varepsilon| < 1$). That is to say, the area where the percent variation of price is higher than the percent variation of occupation. For the first days, the demand is inelastic, since the occupation variation has a small response to a price variation. For the graphics analyzed previously, the occupation values for group C gradually increase; however, the most significant increase occurs for the last days before

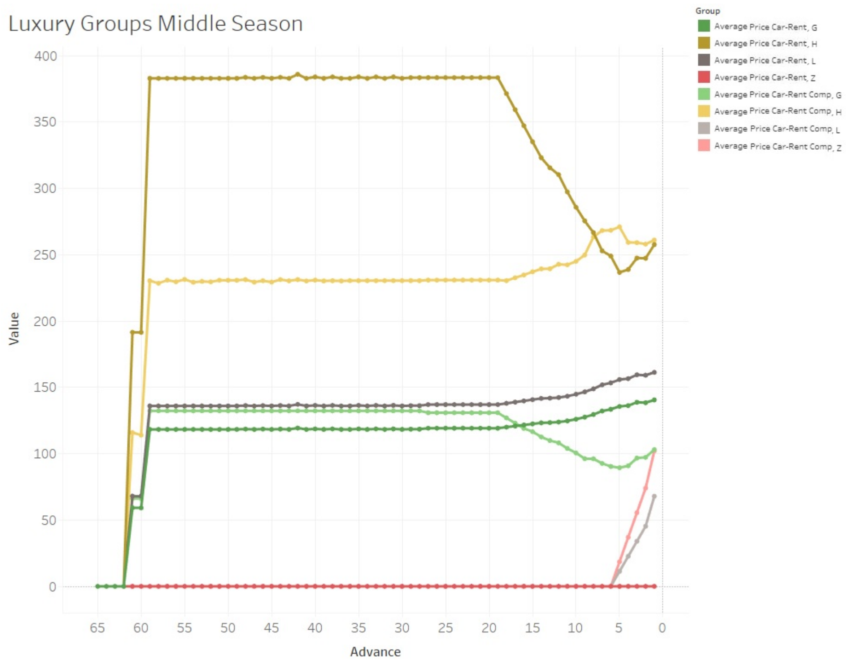

Figure 5.11: Price values of Car-Rent and price of its competitors for luxury groups

the beginning of the rental. With a month to go, clients probably want to compare more prices and do not want to book immediately.

From day 24 to 22, the demand is elastic, and the elasticity value is negative. Therefore, the occupation had a high increase for a decrease in price. The same happens between day 11 and 8, but with significantly higher elasticity. Starting from 10 days before the start of the reservation, there is a significant increase in occupation. On the other days, from day 21 to 12 and from day 7 to 1, the demand is elastic with positive elasticity, which means that the occupation had a higher increase for an increase in price, the clients may had been worried about further increases in price.

From 20 days forward, the demand is elastic, either with positive or negative elasticity values, so the variation in occupation is higher than the variation of price. The clients continue to reserve because of necessity, shortage of time to analyze prices, or because they are afraid of the prices getting higher. However, they have a higher response, the most significant elasticity, when the price decreases 9 days before the start of the reservation. Since group C did not have a decrease in occupation for the last 30 days before the beginning of the reservation during high season, the analysis not considering decreases in occupation is similar.

5.2.2 Group J

For Figure 5.14 from page 49, the price elasticities of demand were calculated, and then an average for each advance day was calculated. The yellow line marks the area in which the demand is inelastic, according to the literature reviewed in 2.2.2.3 in Chapter 2 ($|\varepsilon| < 1$). That is to say, the area where the percent variation of price is higher than the percent variation of occupation. For

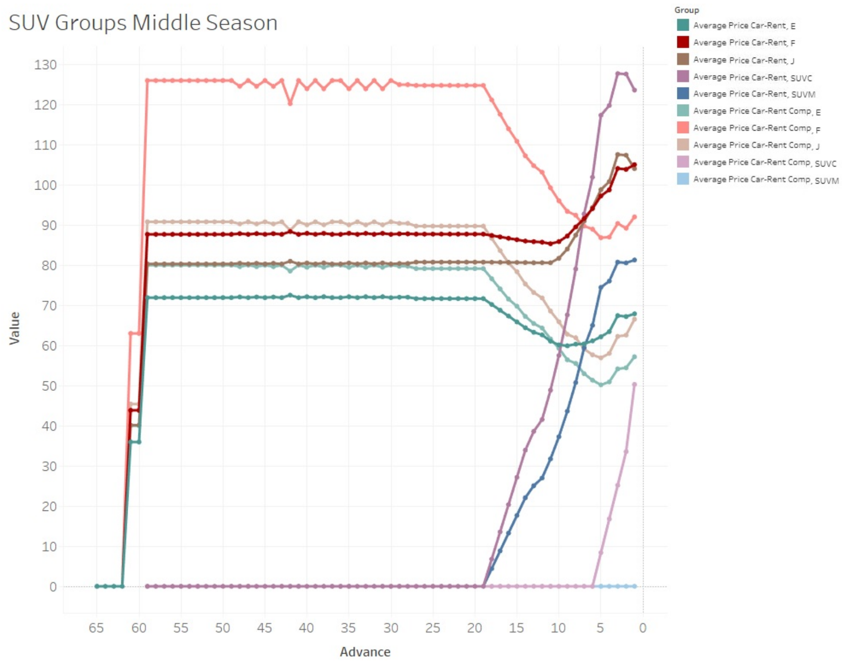

Figure 5.12: Price values of Car-Rent and price of its competitors for SUV groups

this group, the elasticities are all positive, if different from zero, meaning that occupation variation and price variation are not indirectly proportional. However, it is not known if any other factor is affecting the occupation, besides the price. Therefore, it cannot be concluded that this group behaves, or not, according to the law of demand.

The demand is mostly inelastic: the variation in occupation is smaller than the variation in price. For most of the days, clients book group J regardless of the price. These clients may have a higher income or specifically want this type of vehicle. From day 20 to 16, and from day 13 to 8, the demand is elastic and the elasticity value is positive, so for this period the variation of occupation is higher than the variation of price and directly proportional.

In the following analysis, negative occupation variations were considered to be null.

For Figure 5.15 from page 50, the percentage of occupation variation was considered to be only positive or zero. With that, the price elasticities of demand were calculated, and then an average for each advance day was calculated. The yellow line marks the area in which the demand is inelastic, according to the literature reviewed in 2.2.2.3 in Chapter 2 ($|\varepsilon| < 1$). That is to say, the area where the percent variation of price is higher than the percent variation of occupation. The only change from the previous analysis is that the demand of the period from day 13 to 8 is now inelastic. Therefore, for that period of days, there was a higher decrease in occupation for a smaller decrease in price. There are no apparent reasons why a decrease in price would decrease the occupation instead of increasing it.

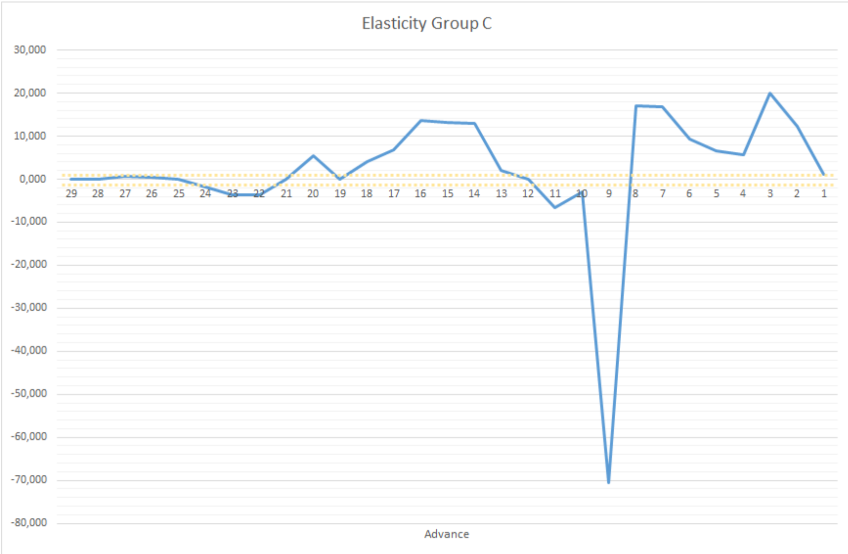

Figure 5.13: Price Elasticity of Demand group C

5.2.3 Group MINI

In Figure 5.16 from page 51, the price elasticities of demand were calculated, and then an average for each advance day was calculated. The yellow line marks the area in which the demand is inelastic, according to the literature reviewed in 2.2.2.3 in Chapter 2 ($|\varepsilon| < 1$). That is to say, the area where the percent variation of price is higher than the percent variation of occupation. The further from the rental start, the more an increase in price variation causes a decrease in occupation variation, even more if the competitors do not increase their prices. Since there is still time to make a reservation, clients seem more careful and aware of price changes.

Observing *Auto Europe* occupation values in Figure 5.1 from page 36, the occupation values for group MINI increase gradually through the advance days. In general, for the last days, a smaller increase in price has a higher increase in occupation, most likely because of the same reasons mention for group C: necessity, shortage of time to analyze prices, or because the clients are afraid of the prices getting higher.

5.2.4 Group H

For the high season period, this group has a constant occupation for the last 30 days before the rental start. Therefore, any change in price did not influence a change in occupation, so the elasticity could be considered perfectly inelastic. However, when observing the price difference between Car-Rent and its competitors, it may not be correct to claim that the demand is perfectly inelastic since it would be possible to propose a more competitive price and induce a change in the occupation variation.

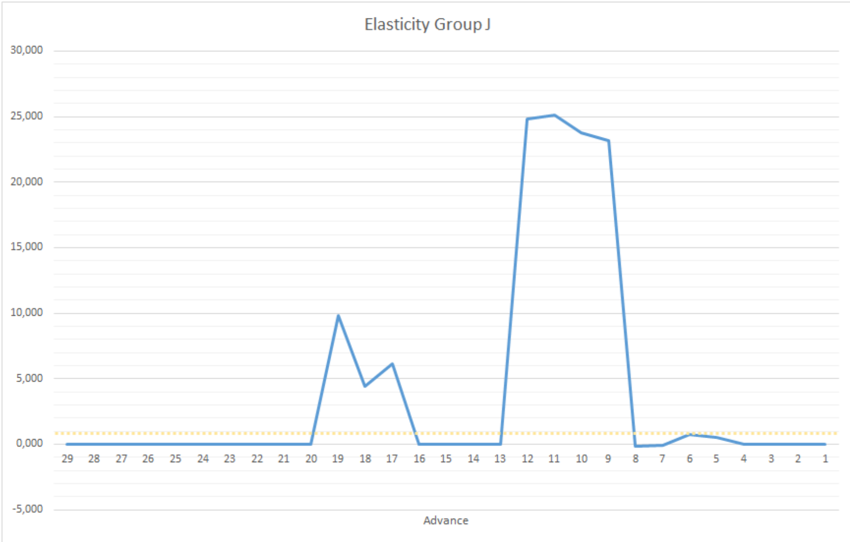

Figure 5.14: Price Elasticity of Demand group J

5.3 Regressions

Zooming in the pick-up zone of Lisbon, the regression analyses focused on trying to understand if the day of the week (Monday, Tuesday, ..., Sunday), day of the month, and advance (in days) have an impact on the elasticity calculated.

5.3.1 Group C

The elasticities used in the regression model were the ones from the third type of analysis: if the percentage variation of the price is zero, then the calculated elasticity is not considered. The elasticity for this group varied through the advance days, so it was expected that the advance days impacted the elasticities. However, observing the statistical significance (p-Value), in Figure 5.17 from page 51, the results show that only Sundays may have a (small) statistically significant impact on the elasticities, since it has the p-Value closer to zero (but still higher than 5%). The coefficient has a negative signal, which means that on Sundays, the elasticity may be slightly smaller.

For the high season period, the elasticity of group C may not be related to any of these factors. Since this is an economic group, and consequently the best-selling group, it will be continually booked regardless of the day.

5.3.2 Group J

For this regression model, the elasticities used were the ones from the third analysis as well. The regression results, in Figure 5.18 from page 51, indicate that, for this group, the day of the month and the advance (in days) have an impact on the elasticities since their statistical significance (p-Value) is closer to zero. The day of the month has a higher impact (p-Value smaller than 1%) and

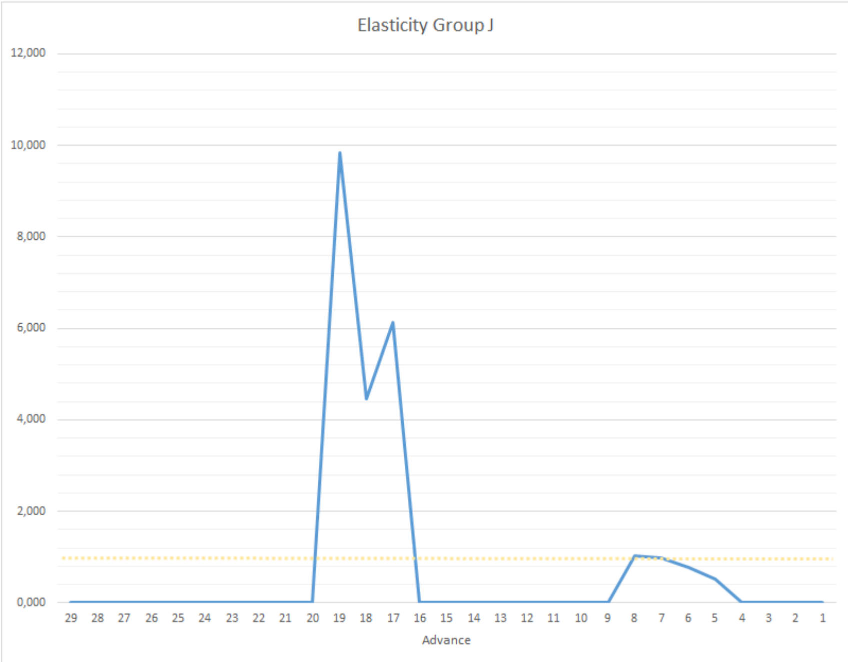

Figure 5.15: Price Elasticity of Demand group J

its coefficient is positive, which means that they are directly related, as the day in the high season period analyzed progresses, the elasticity is greater. For the advance (in days), the coefficient is negative, which means they are inversely related, so as the advance decreases, the elasticity has a higher value. Future analysis could determine which of the days and advances have an impact on the elasticity values.

5.3.3 Group MINI

In this regression model, the elasticities used were the ones from the third analysis as well. Considering the last regression results, it was not expected that these factors would impact the elasticities of group MINI, in Figure 5.19 from page 51, since this group had similar behavior as group C. As group C, for the high season period, group MINI is also booked gradually through the advance days. Therefore, these factors do not significantly impact the elasticity.

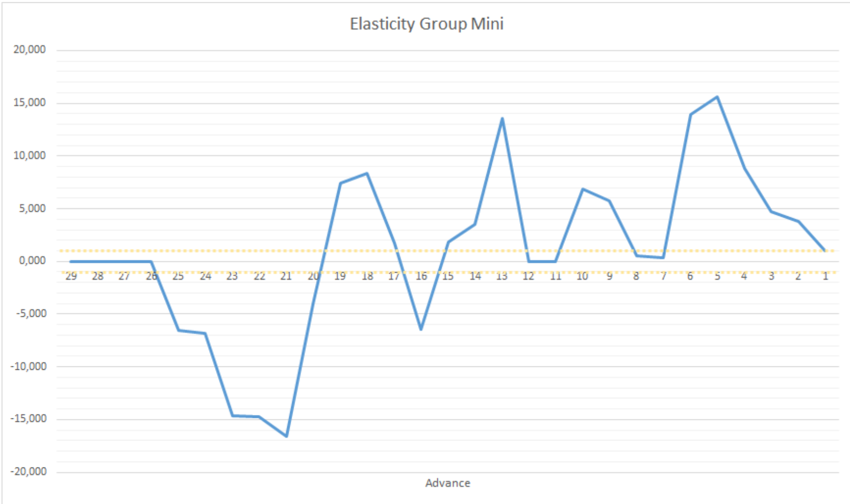

Figure 5.16: Price Elasticity of Demand group MINI

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
day of week = Mon	1.148	15.372	0.008	1.000	0.075	0.941	
day of week = Sun	-23.124	12.163	-0.204	0.989	-1.901	0.060	*
day of month	1.204	1.163	0.117	0.915	1.036	0.303	
ao	-0.336	0.642	-0.052	0.990	-0.523	0.602	
(Intercept)	1.788	13.372	?	?	0.134	0.894	

Figure 5.17: Regression result Group C

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
day of month	1.193	0.366	0.576	0.581	3.255	0.002	***
ao	-0.566	0.218	-0.460	0.581	-2.599	0.013	**
(Intercept)	4.477	2.702	?	?	1.657	0.105	

Figure 5.18: Regression result Group J

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
ao	-0.505	0.325	-0.155	1	-1.556	0.123	
(Intercept)	9.246	3.710	?	?	2.492	0.014	**

Figure 5.19: Regression result Group MINI

Chapter 6

Conclusions and future work

In order to tackle the significant amount of raw data, and be able to obtain managerial insights for a Portuguese car rental company, a Data Analysis Support System was designed and built from the start. This system receives the data to analyze, processes it, and connects the data to different data analysis tools where multiple analyses can be performed.

The system was successfully implemented, and some analyses were made as a proof-of-concept and to obtain managerial insights. The combination of the visualization, elasticity calculations and regression analyses allows for a better understanding of how prices and occupation (demand) behave and impact each other. As an example, the following managerial insights may be drawn for the pick-up zone of Lisbon:

- Prices and occupation values (demand) behave differently depending on the season. For the high season period, both prices and occupation values start to vary around 70 days before the rental start. For the middle season period, they start 50 days later, around 20 days before the rental start. Clients are aware of high-demanded seasons, and their consumer behaviors change according to it.

- Even though clients start to make reservations earlier in the high season, there are always last-minute clients. Moreover, occupation values increase the most when closest to the rental start for the high and middle season periods.

- Car rentals are aware of the last-minute clients, and so prices increase the most when closest to the rental start.

- The companies in the car rental market follow the same pricing behavior and observe the prices of each other.

- For a lower demand season, the middle season period, Car-Rent has a more competitive price to assure reservations. That pricing behavior changes for the high season period.

- Cheaper groups are booked gradually, regardless of increases in price, for the high season.

The system is scalable and adaptable, as it is possible to incorporate other functionalities, data, and data analysis tools. Also, it can be useful for other car rental companies and different types of data.

To process significant amounts of data, the software of the system needs to be efficient. As future work, the programs built can be improved to try reducing the data processing times and to develop a more user-friendly interface as well. There is still more data to be analyzed, so another stream of future work is to continue the analyses and try other points of view, building on the work already developed. New ways to calculate elasticities can be experimented, using an interval bigger than one day for the calculation of price and occupation variations, for example. Moreover, the case of upgrades and downgrades could be analyzed. Another line of future work could be combining the client's information with the reservation data. The client's age, nationality and profession can be related to the variation of occupation, and a pattern of consumption could be made. However, this analysis raises concerns about the protection of the clients' personal information, such as their permission to use the information.

References

[1] Wikipedia. Simple supply and demand graph., 2012. [image online] Available at: https://commons.wikimedia.org/wiki/File:Simple_supply_and_demand.svg [Accessed February 2019].

[2] Smashicons. Xls free icon. [image online] Available at: https://www.flaticon.com/free-icon/xls_136450 [Accessed June 2019].

[3] Smashicons. Database free icon. [image online] Available at: https://www.flaticon.com/free-icon/database_149206 [Accessed June 2019].

[4] Smashicons. Analytics free icon. [image online] Available at: https://www.flaticon.com/free-icon/analytics_238055 [Accessed June 2019].

[5] Smashicons. Analytics free icon. [image online] Available at: https://www.flaticon.com/free-icon/analytics_238057 [Accessed June 2019].

[6] Smashicons. Exe free icon. [image online] Available at: https://www.flaticon.com/free-icon/exe_136449 [Accessed June 2019].

[7] Smashicons. Folder free icon. [image online] Available at: https://www.flaticon.com/free-icon/folder_149344 [Accessed June 2019].

[8] Smashicons. Server free icon. [image online] Available at: https://www.flaticon.com/free-icon/server_149201 [Accessed June 2019].

[9] Smashicons. File free icon. [image online] Available at: https://www.flaticon.com/free-icon/file_149346 [Accessed June 2019].

[10] Good ware. Document free icon. [image online] Available at: https://www.flaticon.com/free-icon/document_702820 [Accessed June 2019].

[11] Good ware. Stats free icon. [image online] Available at: https://www.flaticon.com/free-icon/stats_702816 [Accessed June 2019].

[12] Diogo Ferreira Nunes. Rent-a-car vale 700 milhões de euros à boleia dos turistas. *Dinheiro Vivo,[online]*, 2018. (Last updated 6.00 AM on 14th February 2018). Available at: https://www.dinheirovivo.pt/empresas/rent-a-car-vale-700-milhoes-de-euros-a-boleia-dos-turistas/ [Accessed January 2019].

[13] Publituris. Aquisições no rent-a-car chegaram às 3.453 viaturas em dezembro. *Publituris, [online]*, 2019. (Last updated 10.22 AM on 10th January 2019). Available at: https://www.publituris.pt/2019/01/10/

aquisicoes-no-rent-a-car-chegaram-as-3-453-viaturas-em-dezembro/ [Accessed January 2019].

[14] Beatriz Brito Oliveira, Maria Antónia Carravilla, and José Fernando Oliveira. Fleet and revenue management in car rental companies: A literature review and an integrated conceptual framework. *Omega (United Kingdom)*, 71:11–26, September 2016. URL: https://www.sciencedirect.com/science/article/pii/S0305048316300342, doi:10.1016/j.omega.2016.08.011.

[15] Beatriz Brito Oliveira, Maria Antónia Carravilla, José Fernando Oliveira, Paula Raicar, Delfina Acácio, José Ferreira, and Paulo Araújo. Pricing for Internet Sales Channels in Car Rentals. *Springer International Publishing Switzerland*, 15:139–147, 2015.

[16] Purnachandra Rao. Insurmountable Business Problems and Optimal Managerial Decisions : Significance of Elasticity of Demand. *International Journal of Advances in Management and Economics*, 6(3):19–22, 2017.

[17] Kamber Micheline Han, Jiawei and Jian Pei. *Data Mining: Concepts and Techniques. 3rd ed.* [e-book]Amsterdam: Elsevier., 2012. Available through: Catálogo Universidade do Porto website https://catalogo.up.pt/F/?func=full-set-set&set_number=007614&set_entry=000001&format=999 [Accessed January 2019].

[18] José Luís Borges. Association Analysis, PRDEIG032 Extração de Conhecimento e Aprendizagem Computacional. 2018.

[19] José Luís Borges. Pre-Processing, PRDEIG032 Extração de Conhecimento e Aprendizagem Computacional. 2018.

[20] José Luís Borges. Data Visualization, PRDEIG032 Extração de Conhecimento e Aprendizagem Computacional. 2018.

[21] Rui Camacho and José Luís Borges. Classification, PRDEIG032 Extração de Conhecimento e Aprendizagem Computacional. 2018.

[22] Amy Gallo. A refresher on regression analysis. *Harvard Business Review, [online]*, 2015. (Last updated on 4th November 2015). Available at: https://hbr.org/2015/11/a-refresher-on-regression-analysis [Accessed June 2019].

[23] Robert Frank and Ben Bernanke. *Princípios de economia*. Lisbon: McGraw-Hill de Portugal, 2003. Translated from English by E. L. D'Oliveira, H. F. Martins, H. I. Coelho, J. P. Martins, L. B. de Oliveira M. J. Cotter.

[24] Amy Gallo. A refresher on price elasticity. *Harvard Business Review, [online]*, 2015. (Last updated on 21st August 2015). Available at: https://hbr.org/2015/08/a-refresher-on-price-elasticity [Accessed June 2019].

[25] Boost c++ libraries. [online] Available at: https://www.boost.org/ [Accessed March 2019].

[26] Tableau. [online] Available at: https://www.tableau.com/ [Accessed March 2019].

[27] Rapidminer. [online] Available at: https://rapidminer.com/ [Accessed March 2019].