

Mestrado em Gestão de Informação
Master Program in Information Management

HOW DATA VISUALIZATION TOOLS CAN IMPROVE DECISION MAKING PROCESSES BASED ON CUSTOMER SATISFACTION?

Project Work

Mariana Raquel da Silva Bento

Project Work presented as partial requirement for obtaining the Master's degree in Information Management with specialization in Knowledge Management and Business Intelligence

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**HOW CAN DATA VISUALIZATION TOOLS IMPROVE DECISION MAKING PROCESSES BASED
ON CUSTOMER SATISFACTION?**

by

Mariana Raquel da Silva Bento

Project work presented as partial requirement for obtaining the Master's degree in Information Management with a specialization in Knowledge Management and Business Intelligence.

Advisor: Diego Costa Pinto

December 2021

ACKNOWLEDGEMENTS

To my family, for believing in me since the beginning, for the unconditional love and for all the support and affection given, for having always believed in me and my dreams and for respecting my choices and achievements.

To João, for being my partner, my best friend, and my support always. For understanding me, respecting me, and helping me to accomplish this goal. For being my inspiration, professionally and academically, but above all as an incredible human being that he is.

To my friends, for the support and encouragement given, making me improve professionally and personally.

To my supervisor, professor Diego Costa Pinto, for his unconditional support throughout my master's degree and passage of the Nova Marketing Analytics Lab, for having given me the freedom to learn with him, the privilege of being able to work together and the humility of learning with me. Above all, for being a human being with a giant heart and an extraordinary teacher.

Finally, to the entire Nova Marketing Analytics Lab team, especially Carina, for the extraordinary work they do and for the generosity of teaching and helping me so much, thank you.

ABSTRACT

Companies seek to perform better using their own data to achieve sustainable competitive advantage. By enabling business intelligence tools and techniques companies can get ahead of the competition, however, even with these solutions, some questions still stand which this research aims to answer.

This study aims to answer the following research question: how data visualization tools can improve decision making processes based on customer feedback? To fill this gap, this research will explore the literature on data-driven marketing and its relationship with decision making processes, while also diving on data visualization and business analytics.

To improve the relevance of the study, a practical approach will also be used, using Business Intelligence tools and techniques, in particular Microsoft Power BI. The research uses real data from a major Portuguese hotel chain and aim to develop a new dashboard that can enhance the capability to improve data-driven marketing decision making, using customer data. The findings show that business intelligence and data visualization tools and techniques allow the development of dashboards with several KPI's and measures to improve strategic decision-making process, and it allows to still have a quick and objective views of business performance and productivity, exchanging company data in competitive advantages.

KEYWORDS

Keywords: Business Intelligence, Data Driven Marketing, Data Visualization, Marketing Analytics, Customer Feedback, Decision Making Process.

INDEX

1. INTRODUCTION	4
1.1. Relevance	4
1.2. Theoretical Gap.....	5
1.3. Research	6
2. LITERATURE REVIEW.....	9
2.1. Business Intelligence	10
2.1.1. Data Preparation	11
2.2. Data Visualization	13
2.2.1. DATA VISUALIZATION PIPELINE.....	13
2.2.2. DATA SCALES.....	14
2.2.3. Dashboards Design	15
2.3. Data-Driven Marketing.....	17
2.4. Customer Satisfaction.....	18
2.4.1. Definition of Customer Satisfaction	19
2.4.2. Importance of Customer Satisfaction.....	20
2.4.3. Customer Loyalty - Relationship with Customer Satisfaction	22
2.4.4. Marketing KPIs: Satisfaction and Loyalty	23
2.4.5. How to define customer profile in hotel sector	26
2.5. Sentiment Analysis.....	27
3. METHODOLOGY.....	30
4. Results and discussion	35
4.1. Preliminary Statistical Analysis	35
4.2. Power BI Data Visualization Model	37
4.2.1. Dashboard Customer Profile	37
4.2.2. Customer Satisfaction Dashboards	39
4.2.3. Key Influencers and Top Segments Dashboard	42
4.2.4. Text Mining and Sentiment Analysis Dashboard.....	47
5. Conclusions	50
5.1. Final Considerations	50
5.2. Study Limitations	53
5.3. Future Studies Recommendations	54
6. References.....	55
7. ANNEXES.....	59

LIST OF FIGURES

Figure 1 - Study Structure and Concepts Relationship.....	9
Figure 2 - Data Visualization Pipeline.....	14
Figure 3 - Aspects that Affects Loyalty	23
Figure 4 - Statistical Analysis Output Results	36
Figure 5 - Customer Profile Dashboard	37
Figure 6 - Customer Satisfaction Dashboard: Most Relevant Variables	40
Figure 7 - Customer Satisfaction Dashboard: Other Variables	41
Figure 8 - Key Influencers and Top Segments Dashboard	43
Figure 9 - Key Influencers and Top Segments Dashboard	45
Figure 10 - Text Mining and Sentiment Analysis Dashboard	47

LIST OF TABLES

Table 1 - Data Preparation Steps.....	11
Table 2 - Data Scales.....	15
Table 3 - Dashboards Key Criteria	16
Table 4 - Customer Satisfaction Definitions	20
Table 5 - Customer Satisfaction Score.....	24
Table 6 - Customer Profile Variables in Hotel Sector	26
Table 7 - Dictionary for the sentiment classification	28
Table 8 - Database Structure.....	32

1. INTRODUCTION

1.1. RELEVANCE

Many organizations nowadays are using business intelligence solutions to improve their strategic decision and is useful to create a sustainable competitive advantage. Business intelligence (BI) is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analysing data to help business users make better decisions (Watson, 2009). In general, BI has several advantages related with turning data into actionable information that allows companies to make better decisions and be able to predict trends and patterns.

BI tools and techniques enable us to ensure competitiveness and productivity due to making the process of analysing and interpreting data more efficient. Another great advantage of investing in business intelligence is the ability to turn data into insights about consumer behaviour patterns and trends, improving marketing and sales intelligence.

According to Mark Jeffery (2010) most marketing organizations do not use data driven marketing, meaning they do not have professional processes to manage marketing and most do not use marketing metrics in their day-to-day marketing activities. It is also particularly relevant to look at marketing organizations use of data:

- 57% do not use centralized database to track and analyse their marketing campaigns.
- 70% do not use an enterprise data warehouse (EDW) to track customer interaction with the company and with marketing campaigns.
- 71% percent do not use EDW and analytics to guide marketing campaigns selection.
- 80% not use an integrated data source to guide automated event-driven marketing.
- 82% never track and monitor marketing campaigns and assets using automated software such as marketing resource management (MRM).

This research will create new theoretical and practical knowledge about how business intelligence can drive marketing strategies and how data visualization tools can improve decision making processes. It aims to develop a data visualization dashboard with several KPI's and measures to improve strategic marketing decision making. To make this study more practical, one of the goals is to apply the research to a hospitality sector using real customer satisfaction survey data.

Finally, this project work will provide an interesting complement to the already existing approaches about data driven strategies, to give a practical solution to be applied in real life companies.

1.2. THEORETICAL GAP

There are many studies regarding customer feedback, published in top journals such as "International Journal of Production Research" and "Journal of Service Research", mainly about how to collect this type of data and on how to use this information to support the decisions related to products and services. These studies reveal that issues such as customer preference and customer satisfaction are rarely addressed and incorporated in the decision-making process.

There are also some papers about the process of analysing customer feedback data on the web, which is widely known as opinion mining, but the research scope is limited to text mining analyses that can be carried out using this type of data, while the extensive and strategic analysis that can be made using customer data, in specific customer feedback, are not explored.

One of the main problems regarding customer feedback is related to the fact that nowadays customers use different communication channels to give their feedback and this implies that companies have workflows that allow them to analyse and process this data from different sources and formats.

There are also several studies concerning data driven marketing and the importance of using customer data to support business strategies and decisions, leveraging on business intelligence processes and techniques to accomplish this. In that regard, there are several articles that explain how future marketers can use data to make their marketing decisions more relevant and to make a strategic difference. For this, there are a set of business intelligence tools and techniques to provide valuable real-time data that positively affects information quality.

However, none of literature actually apply all of these concepts to customer satisfaction, therefore there is a lack of research on how to use data visualization tools, data driven marketing and business intelligence techniques and tools to analyse, visualize and share this kind of data in order to make it accessible and accurate to support the decision-making process.

1.3. RESEARCH

The main objective of this study is to understand how business intelligence processes and techniques can help to improve marketing decisions and how data visualization tools can improve decision making processes based on customer feedback. In that sense, this study aims to develop a data visualization dashboard with several KPI's and measures associated to improve marketing decision making using customer data.

The database to be used corresponds to a Portuguese hotel chain and the aim is to use real business data to have the highest possible grip with the reality of the sector, while drawing a customer profile that repurchases.

Another goal is to organize all the information contained in the database on a dashboard using Power BI where it would be possible to see the information from different perspectives (date, type of hotel, location ...), thus it will be necessary to clean the data as well as making the appropriate transformations according to the type of data using the Power BI query editor.

Since we have comments from customers in the database, I will also complement the dashboard with word clouds allowing a text mining analysis.

As a final product, this study seeks to find a way to fill the gap that currently exists through an extensive dashboard aimed at supporting the marketing decision making process. To achieve this, the research will be supported by an in-depth study of the potential of data visualization techniques, the different approaches of data driven marketing using customer feedback data and finally the use of text mining analyses to provide a more clear and complete scenario.

1.4. CONTRIBUTION

This research contributes to create new theoretical and practical knowledge about how business intelligence can drive marketing strategies and how data visualization tools can improve decision making processes. It aims to develop a data visualization dashboard with several KPIs and measures associated to improve strategic marketing decision making. To make this study more practical, one of the goals is to apply the research to a specific sector using real data.

As a result, a data visualization model was developed and its great strength is that it can be applied in a real business context, as happened in this study, demonstrating the potential of using this type of methodologies in the decision-making process and how it can be applied in different contexts. It was also possible to generate several interesting academic and theoretical contributions through the interconnected exploration of different concepts that had been previously studied individually and, now with this study, were applied simultaneously and built upon each other, to develop an extensive and comprehensive literature review that connects all relevant aspects of business intelligence, data visualization, data driven marketing and customer satisfaction.

This research contributes to Diane Janvrin et al (2014) by building upon the literature to create a set of dashboards that put into practice the main concepts and recommendations provided to a specific sector, in this case the hospitality sector. It also adds to the literature on customer centric approaches based on customer satisfaction importance (Kotler et al, 2005; Shamsudin,

2020; Hardjono & San, 2017; Biesok & Wyród-Wróbel, 2011) by using customer data to better understand the client, introducing an analytical process based on Power BI, to then adapt the marketing strategies and initiatives to the customer needs.

Relating to sentiment analysis, this study contributes to the literature of Alaei et al (2019) and Chittipolu et al (2021), by applying the main principles of customer satisfaction and customer loyalty to develop an integrated solution, based on text mining, that allows to show the customer voice, generating relevant insights to support the decision-making process.

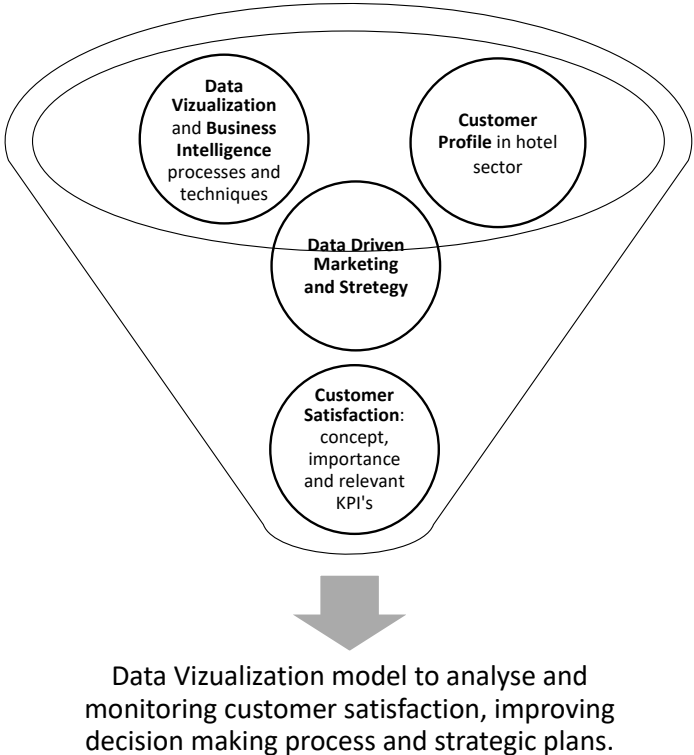
Finally, this project work adds to the field by providing an interesting complement to the already existing approaches about data driven strategies, through an integrated vision that complements the strategic aspects with analytical processes, complemented by a practical solution to be applied in real life companies, that is based on several different theoretical fundamentals from a comprehensive array of different authors and studies.

2. LITERATURE REVIEW

Business intelligence and data visualization tools can have a strong impact in the decision-making process, using analytical methods to minimize operating costs and to accurately forecast market trends (Rouhani et al, 2015). These tools help in collecting consumer data that give the capability to make relevant market offers according to consumers' interests and preferences, thus improving the performance of the company while also decreasing costs. Another idea is that marketing and strategic managers increasingly conduct strategies supported by business intelligence tools and the implementation of sophisticated database architectures supports data-driven decision processes, giving the business powerful competitive advantages. The applications of data visualization increase the level of understanding of data-driven performance measurement (Diane Janvrin et al, 2014), and it is also a powerful solution to make strategic decisions and anticipate business problems.

The following figure shows the relationship between the different concepts and approaches presented in the literature review and which are the structure and product of this study.

Figure 1 - Study Structure and Concepts Relationship



2.1. BUSINESS INTELLIGENCE

Business intelligence (BI) is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analysing data to help business users make better decisions. We can also define BI as a strategic approach to systematically target, track, communicate and transform relevant data into actionable information so that it can be used to support important decision-making processes (Kamel Rouibah & Samia Ould-ali, 2002). Through different points of view, BI can be extremely relevant in providing useful information to decision makers (Rouhani et al, 2015).

Among the main objectives of BI, it is possible to highlight as more relevant the fact of providing fast and accurate information allowing to make more effective and timely decisions, transforming corporate data into relevant information. And finally, it makes the decision-making process more transparent (Yohannes Kurniawan et al, 2005). BI increasingly assumes a decisive relevance in the data analysis process, mainly in the preparation of reports and analyses to support decision making.

The use of BI tools for data exploration and the construction of interactive ways of visualizing them by decision makers, requires a set of preliminary tasks of extracting data from different sources and transforming that same data into a format adapted to the type of analysis intended, which involves removing corrupted and duplicate data, verifying its integrity which involves correction and cleaning. Finally, and after all these steps it is time to applying joins and specific measures and calculations (Fikri et al, 2019). Thus, it is possible to group the necessary transformations to be carried out in the preparation of the analysis in two categories: **basic** and **advanced transformations**.

Basic transformations include cleaning, identify and removing duplication, format revision and key restructuring. While advanced transformations involved derivation, filtering, joining, splitting data, validation, summarization, aggregation and integration (Fikri et al, 2019).

2.1.1. Data Preparation

Data preparation is a fundamental step in business analytics process allowed manage huge datasets to ensure quality data. This is an extremely relevant process as it guarantees the integrity of the data so that it can be consumed for analysis at a later stage. Not getting this process is often the most time-consuming step when it comes to data analysis and the speed and efficiency of it automatically translates into the speed of discovering insights.

Due to its importance “data preparation is the focus of innovative software technology methods aimed at accelerating, if not automating, processes necessary to support business analytics” (Stodder et al, 2016).

The first and most important priority is to understand and assimilate the analysis objective, because only then can the data be treated and made as prepared as possible for the next reporting and analysis steps. In general terms the process initially integrates the ingestion of data sources, then the tools and methodologies are applied to adapt the data to the business objectives through transformations and enrichment, and also very important is the integration with governance and administration to ensure monitoring and improving data used for BI analysis (Stodder et al, 2016).

According to Patnaik & Vladicesco (2017) there are four fundamental steps in data preparation to ensure clean and quality data: **data cleaning, data integration, data transformation and data reduction.**

Table 1 - Data Preparation Steps

Steps	Definition	Tasks
Data Cleaning	Data cleaning is the process that includes all tasks and activities to detect and repair errors in the data (Chu & Ilyas, 2019).R	This step consists in fill in missing values, smooth noisy data, identify or remove outliers, and identify and solve inconsistencies.

Data Integration	Data integration aims to combine multiple sources like databases, data cubes or files into a single unified view (Patnaik & Vladicesco, 2017).	Data integration starts with the ingestion process that integrates a set of steps such as cleaning, ETL mapping and other transformations. Then the data is taken from its sources and finally is unified in a single database.
Data Transformation	The data transformation process is responsible for applying the necessary changes in order to ensure that the data is in a format suitable for the type of analysis desired. For this, it is necessary to emphasize the importance of understanding the objectives of the analysis and the business.	There are several transformations steps that could be applied, but the most used are cleaning (ex: convert NULL into 0 or Female into F), delete duplicate records, change format type (ex: date/time conversion) and filtering (choice only the columns and rows there are relevant for the analysis).
Data Reduction	Data reduction process allows to obtain a smaller sample of the data without compromising the quality of the analysis results and improve classification performance by eliminating redundant or irrelevant records (Rangarajan, 2010). This is an important process because larger datasets may not be supported by some tools capacity.	According to Microsoft data reduction techniques guidance (2019) there are some steps to apply in this process: remove unnecessary columns/rows, group by and summarize data, optimize columns data types and generate custom columns.

Source: Elaborated by the author

Depending on the type of analysis and tools used, some of these steps can be changed or aggregated, but the purpose inherent to each of them remains, but for BI analysis these are the most common.

The essential and most relevant goal of data preparation process is to understand the data and improve their quality and integrity, standardizing and structuring the information and developing a set of transformations in order to make the data useful for reports and analysis (Stodder et al, 2016), because only in this way it can be used as a support decision making.

2.2. DATA VISUALIZATION

Integrated data viewers facilitate the process of decision support to senior management by analysing large and complex data sets, while easily selecting different information types and formats that are most useful for decision making.

“Interactive data visualization” (IDV) refers to business intelligence technology and technique of gathering, processing, and analysis of large and complex amounts of data to a specific application of that technology, also known as “dashboards” (Diane Janvrin et al, 2014).

Interactive data visualization (IDV) simplifies the process of organizing and visually displaying data in an easy-to-use interface (Diane Janvrin et al, 2014) and increases the level of understanding of data-driven performance measurement.

Properly exploring the data visualization tools allows companies to have a deep understanding of the main results and to make important strategic and operational decisions easier (Jeanne Moore, 2017).

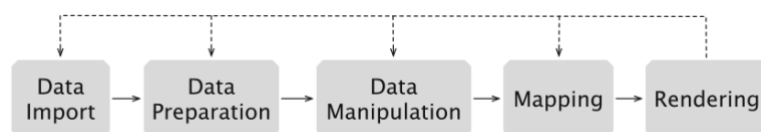
2.2.1. DATA VISUALIZATION PIPELINE

According to Hauser & Schumann (2009) “visualization pipeline is a general model for a typical structure of a visualization process”. To be able to achieve a correct and useful visualization of the data are necessary to a set of steps taken along the visualization pipeline including data import, preparation, manipulation, mapping and rendering (Qin et al, 2020).

All of these steps are absolutely essential and each of them is indispensable in the whole process:

- **Data import** is the process of extract the necessary data from a one or more data sources.
- **Data preparation**, according to a report from Yellowfin International (2016), “is the process by which data analysts transform and organize that data into new data sets suitable for exploration and analysis”. For data visualization this process is used to prepare the data that will be imported for visualization analysis.
- **Data manipulation** is a fundamental step for a consistent data visualization because choice and management the data that will be visualized. Generally, combines tasks like filter and sort data or merging and combine columns.
- **Mapping** is the central stage of the pipeline, in which the dimensions of the data are mapped so that they can constitute visual resources and structures. The most common procedures in the mapping stage are the ordering generation and attributing data dimension to visualization axes in different orders (Bertini et al, 2011). In essence, the different alternatives can be built using a set of visual features (colors, size, or shape).
- **Rendering** transforms visual structure into a visual representation through stipulating graphic features that transform these structures into pixels (Bertini et al, 2011).

Figure 2 - Data Visualization Pipeline



Source: Qin et al, 2020

2.2.2. DATA SCALES

In order to guarantee that the visualizations exactly fulfil the goals of the end user’s analysis it is necessary to make a correct survey of the data scales. The choice of a particular scale has the consequences of the mathematical operations that are allowed with the respective

variables, limits the transformations that can be made with the respective variables without loss or alteration of the information.

Data variables may have different scales, influencing which analyses and visual encodings can be used. It is possible to distinguish nominal, ordinal, interval, and ratio data based on the type of logical mathematical operations that are permissible.

Table 2 - Data Scales

Data Scale	Measure
Nominal	Mode
Ordinal	Median
Interval	Arithmetic mean
Ratio	Geometric mean

Source: Adapter from Börner et al, 2019.

2.2.3. Dashboards Design

Dashboards are representations of mostly visual information that are used to quickly monitor and implement various metrics and KPIs as well as current business conditions, thus facilitating their understanding through infographics or narrative visualizations (Sarıkaya et al, 2018).

Before starting dashboard design, it is important identify the audiences, their needs and business goals for dashboard analysis, outline key metrics and KPI’s and define the data that is relevant highlight. Other aspect that is fundamental state “*a priori*” is the dashboard interface and guarantee that is align with de analysis purposes.

Using the data to tell a story can be a valuable decision-making mechanism and, for that, it is important take into account the following aspects:

- Guarantee an easy, intuitive and dynamic visualization process.
- Promise an interactive experience.
- Allow a progressive data expose.

In order to properly build and evaluate dashboards, is must consider a set of criteria in this process, and we can classify them in seven categories involving user customization, knowledge discovery, security, information delivery, visual design, alerting and integration and system connectivity (d).

Table 3 - Dashboards Key Criteria

Key Criteria		
User Customization	Customizing Definitions	Goals
		Objectives
		Metrics
		Targets
		Calculations
		Correlation among metrics
	Categorization	Restricted access to specific metrics according to different users
		Assigning a group of users to a group of dashboards
	Feedback	Attach comments to metrics
Discussion forum among users		
Knowledge Discovery	Drill-down features	
	Dimensional modelling with hierarchies and levels	
	Dependency analysis	
	What-if analysis	
	Move from monitoring layer to analysis layer	
Security	Appropriate authentication and authorization methods	
	Backup and restore procedures	
	Version control dashboards	
	Protecting data from change	
	Defining role-based security	
	Automatic accessibility change by change in user roles or groups	
Information Delivery	Reasonable response time and latency	
	Customized layout of metrics for print	
	Exporting information in a format that is easily shared (spreadsheets, presentation slides, word or PDF)	
	Data filtering for selected reports	
	Sorting the report	
	Inserting/deleting columns	
	Scheduling automatic reports	

	Updating the reports
Visual Design	Visual intelligence to highlight areas and values
	Table and chart on same screen
	Toggling between tabular and chart views
	Resizing, maximize/minimize, re-ordering of zones
	Allowing different layouts
	Inclusion of metric definition and calculation
	Linking objectives with metrics
	Linking metrics together
	Having Metadata and help
Single screen with no scrolling	
Alerting	Defining and customizing the alerts
	Highlighting by colour coding for unexpected values
	Determining the timing of alerts
	Placing the alerts in context
	Delivering alerts through dashboard website, email and cell phone
	Showing the next step to do
System Connectivity and Integration	Explaining the problem using text
	Connectivity to a variety of data sources like OLAP cubes, databases, lists and spreadsheets
	Supporting different operating systems
	Integrating with portals and with other applications
	Recovering from software or hardware crash
Integrating with programmatic APIs for data and metadata	

Source: Adapted from Karami et al, 2017

2.3. DATA-DRIVEN MARKETING

It is possible to define Data Driven Marketing (DDM) as the process of gathering complex customer data through both online and offline channels, analysing them to find patterns of purchasing behaviour thus generating relevant insights for decision makers, in particular marketing teams as this allows to develop personalized strategies according to the preferences of the target audience (Balakrishna Grandhi et al, 2020).

With the increasing use of data on the part of companies to define strategies and anticipate customer needs, it has been verified that technology assumes an increasingly important role in the construction of predictive models which allows to anticipate problems, building customer-centered processes and making timely decisions. The use of customer data makes it possible to identify needs at each stage of the consumer's decision-making process (Balakrishna Grandhi et al, 2020).

The use of DDM techniques and principles improve the internal and external data analysis, integrating this information to enhance the development of new products and services or better the existing ones. This approach ensures greater productivity and enhances the acquisition of new clients and the trust of current ones, allowing to minimize costs and increase the company's productivity and efficiency (Balakrishna Grandhi et al, 2020).

2.4. CUSTOMER SATISFACTION

Collecting and analysing customer feedback data is an extremely relevant process as it allows a continuous learning centered on the customer to adapt the offer according to the customer's preferences and needs (Li et al, 2011).

It is increasingly evident that customers use different communication channels for feedback regarding their purchasing experiences, which implies complicated management by companies as they must develop efficient and effective processes to collect and analyse all information from different sources (Ordenes et al, 2014).

Technology has allowed companies to diversify the availability channels for collecting customer feedback, which, being structured, allows quantitative analyses in lieu of unstructured feedback, such as phone calls, emails and social networks, in which customers describe their experiences more freedom (Ordenes et al, 2014).

Despite efforts to find methodologies and processes to optimize the collection and analysis of customer data, several problems have persisted (Wang, 2013) such as:

- The need to incorporate customer satisfaction into the entire decision-making process.
- The characteristics of the product or services that are perceived as important by the customer vary depending on the individual and, therefore, a customer-oriented business approach is important to ensure that it is possible to identify and measure these characteristics.
- Not depending on subjective evaluations by specialists, but focusing on a market-oriented approach, it allows generating and evaluating concepts that are particularly relevant to identify new emerging market opportunities.

2.4.1. Definition of Customer Satisfaction

According to Shamsudin et al (2020) customer satisfaction is the general assessment that the customer makes after their purchase because of a comparison between their expectations and the perceived value of the product or service. Customer satisfaction is also a terminology used in marketing to measure how products or services match or exceed customer expectations (Razak & Nayan, 2020).

Several authors argue that customer satisfaction positively increases brand loyalty, the intention to repurchase and also allows to reduce marketing costs.

It's possible to divide customer satisfaction in two types, transaction-specific and cumulative. The first one indicates customer feedback regarding the latest transactional experience with the brand or company and the second is the satisfaction related it is the satisfaction associated with thousands of factors such as the product/service, quality-price ratio, brand trust or after sales services. Satisfaction is extremely relevant as it provides the company, particularly the marketing team, with a metric that allows it to implement a set of actions to improve the experience of the consumer and the business in general (Shamsudin et al, 2020).

Table 4 - Customer Satisfaction Definitions

Source	Definition
Oliver (2006)	“The consumer's fulfillment response. It is a judgment that a product or service feature, or the product or service itself, provided (or is providing) a pleasurable level of consumption-related fulfillment, including levels of under- or overfulfillment”
Philip Kotler et al (2005)	“The extent to which a product’s perceived performance matches a buyer’s expectations. If the product’s performance falls short of expectations, the buyer is dissatisfied. If performance matches or exceeds expectations, the buyer is satisfied or delighted.”
Hill et al (2007)	“Customer satisfaction, or dissatisfaction, is the feeling a customer has about the extent to which their expectations with an organization have met their needs.”
Falkowski & Tyszka (2009)	The positive emotional consumer’s reaction on product or service evaluation.
Farris et al (2016)	“The number of customers, or percentage of total customers, whose reported experience with a firm, its products, or its services (ratings) exceeds specified satisfaction goals. (...) and also indicates likelihood of repurchase.”
Shamsudin et al (2020)	It is the expectation perceived by both customer and consumer before purchasing and experiencing the product or services. The perceived value by customers and consumers before purchasing and experimenting the product or service.

Source: Elaborated by the author

2.4.2. Importance of Customer Satisfaction

Measuring customer satisfaction is an extremely relevant process for any company or business, above all because of the constant changes in the market.

Many companies and organizations use the measurement of customer satisfaction to (Biesok & Wyród-Wróbel, 2011):

- Understand the expectations and requirements of customers.
- Determine, how company and its competitors meet these expectations and requirements.
- Prepare services and products based on information obtained because of the survey.
- Find out, what current trends are in offer, to take an immediate action.
- Establish priorities, objectives, and standards, to assess how company has achieved it.
- Assess the impact of changes in politics, corporate strategy, products, and services.

It is possible to characterize the needs of consumers as unlimited and delighted and the objective is to make the most of the money spent on goods and services. In this sense, brands and companies must define their strategies and approaches to meet the needs and desires of their customers (Shamsudin et al, 2020).

Nowadays customers are seen as individuals looking for goods and services that are aligned with their values and are also more responsible spenders who want to maximize their money and are therefore more demanding consumers in terms of quality. The importance of customer satisfaction essentially depends on the fact that more satisfied customers tend to be more valuable customers as they tend to buy more (Shamsudin et al, 2020).

For companies that have a customer-centric approach, the purpose and one of the most important success factors is customer satisfaction. Highly satisfied customers bring a huge set of advantages for brands and companies: they are less sensitive to price and potentially long-term customers and therefore more valuable customers. These customers are often attentive to new products / services and their associated improvements, buying additional and complementary products over time and still recommending these products and services to friends and family (Kotler et al, 2005).

Companies and brands that adopt a customer-centric strategy focus on ensuring the maximization of their customers' satisfaction compared to the competition and this is always one of the biggest challenges. However, they are not always concerned with increasing satisfaction by decreasing the price or improving the product or service because it may imply increase costs and reduced profits for the company (Kotler et al, 2005).

There are also other important agents to consider for a company's strategy in addition to its customers, in this case we are talking about stakeholders, which generally include employees, resellers, suppliers and shareholders (Kotler et al, 2005).

For these partners, an increase in spending and possible decrease in profits to increase customer satisfaction may cause their funds and interests to have to be diverted for this purpose. The challenge is to find a balance in this dynamic to generate the maximum value for the customer in a profitable way for the company (Kotler et al, 2005).

2.4.3. Customer Loyalty - Relationship with Customer Satisfaction

First, it is important to distinguish the concepts of consumer satisfaction and loyalty as they are related but in essence differ from each other. According to Shamsudin et al (2020) "satisfaction is an attitude perceived by the customer while loyalty is a behavior of customers, it is the behavior that leads to repurchase and the disposition of the customer to acquire the organization's future product".

It is also possible to say that customer loyalty is a result of satisfaction since more satisfied customers tend to be more loyal to the brand, because they trust and therefore tend to spend more and are less sensitive to price, that is, small variations in price do not translate into a decreased demand for branded products or services (Shamsudin et al, 2020).

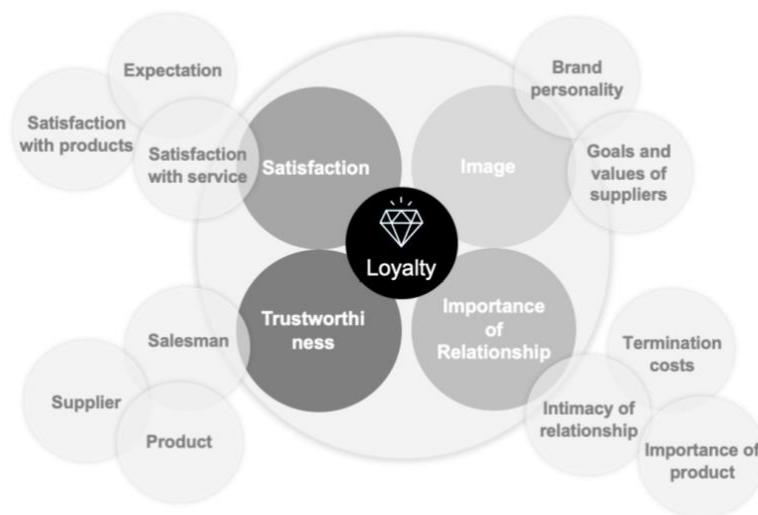
We can associate the concept of loyalty with relationship marketing whose focus is to guarantee long-term customer satisfaction, delivering the greatest possible added value in a long-term perspective of customer relationship management.

To ensure this dynamic it is necessary that this relationship be an objective of the whole company, requiring an articulation of work between all departments and the marketing team, involving all employees in an economic, social, technical, and legal effort to ensure high loyalty of the customer (Kotler et al, 2005).

For this study it is relevant to analyze the concept of customer loyalty in the context of the hotel sector, which will be applied in this study. The companies in this sector are characterized by strong competition with each other, it is a highly competitive sector and, therefore, customer loyalty gains even more relevance as it is the most effective mechanism that these companies must differentiate themselves from their competitors (Hardjono & San, 2017).

However, it is necessary to understand that "customer loyalty is a complex and multidimensional concept" (Majumdar, 2005), and this complexity makes it difficult to define this concept correctly and comprehensively. There are several aspects that affects customer loyalty, and it is possible to segment them as follows:

Figure 3 - Aspects that Affects Loyalty



Source: Adapted from Hardjono & San, 2017

2.4.4. Marketing KPIs: Satisfaction and Loyalty

Customer satisfaction score (CSAT) is the most usually satisfaction measure, collected on a 0-5 scale but can also be presented on a 10-point scale. Whatever the scale, the purpose is to measure customer satisfaction in relation to their experience with a product or service.

Table 5 - Customer Satisfaction Score

Very Dissatisfied	Somewhat Dissatisfied	Neither satisfied nor dissatisfied	Somewhat Satisfied	Very Satisfied
1	2	3	4	5

Source: Adapted from Farris et al, 2016

Another relevant aspect to perceive is whether expectations have been met or exceeded because they are an extremely relevant factor when measuring satisfaction. If the customer's expectations are very high and the reality does not correspond, we will have a dissatisfied customer (Farris et al, 2016).

Customer Satisfaction Score (CSAT) is a function of expectations and for that reason it is essential to be concerned with managing customer expectations, being as transparent as possible in relation to what the company is willing to offer. **This metric is subject to dubious responses, and it is important that it reflects the opinions of current customers and not lost customers** (Farris et al, 2016).

Another relevant metric related to satisfaction is **willingness to recommend** which indicates the percentage of surveyed consumers who would recommend the brand to family and friends. In this way it is possible to quantify more objectively how the customer is loyal to the brand (Farris et al, 2016) Loyal customers boost the cheapest and most effective type of marketing, positive word of mouth.

Associated with willingness to recommend is the **Net Promoter Score** that according to Farris et al (2016), "is a measure of the degree to which current customers will recommend a product, service, or company." This is a particularly relevant metric because it allows measuring more accurately customer satisfaction and loyalty. It is relevant to note that to obtain the net promoter score it is necessary to use a scale of responses from 0 to 10 (Farris et al 2016).

Net Promoter Score = Percentage of Promoters (%) – Percentage of Detractors (%)

According to Farris et al (2016), “loyalty itself is not a formal metric, but specific metrics measure aspects of this dynamic. New product entries may alter loyalty levels”. Hereupon, the main metrics to measure customer loyalty include **share of requirements, willingness to pay premium and willingness to search.**

Share of Requirements (SoR) corresponds to “brand purchases as a percentage of total category purchases by buyers of that brand” (Farris et al, 2016), and note that these purchases may be represented in units or revenue.

It is relevant to say that this metric has as purpose measure the level of commitment to the brand by its own customers.

SOR (%) = Brand Purchases (#) / Total Category Purchase by Brand Buyer (#)

About willingness to pay (WTP), this metric measure “both a consumer’s loyalty in the sense of attachment and differentiation, the fact that the product is very different and so there isn’t a close competitor that a consumer can purchase from instead” (Farris et al, 2016).

Finally, the willingness to search (WTS) represents the “percentage of customers willing to delay purchases, change stores, or reduce quantities to avoid switching brands” and also “indicates importance of distribution coverage”, however, it is not always easy to capture (Farris et al, 2016).

WTS (%) = Percentage of Customers Willing to Delay Purchases, Change Stores, or Reduce Purchase Quantities to Avoid Switching Brands

2.4.5. How to define customer profile in hotel sector

Defining the customer profile makes all the difference when it comes to directing marketing, sales, and service strategies. It is possible define that the consumer profile is a set of characteristics, behaviours and values observed in a specific audience. Perceiving the customer's profile allows to target the best market solutions to a specific target and make it easier to know where to find it and how to get in touch to try to sell a product or service that best suits their needs.

In terms of resource management, the definition of the customer's profile is very relevant because it allows for this allocation to be made more efficiently, reaching a specific audience that really corresponds to its target.

When it comes to the hotel sector, it is particularly important to define and understand the profile of the guests, in this case the target audience that a specific hotel or hotel group must reach. Usually, the trend is to try to please all guests in the same way, but each person has one has different needs, values different aspects, and has distinctive expectations. In this way, correctly identifying customer profiles and their motivations allows to increase recognition and satisfaction.

In the following table different variables are grouped according to their characteristics, and all of them allow to build a consistent host profile.

Table 6 - Customer Profile Variables in Hotel Sector

Group of Variables	Variables	Type	Example
Demographic	Age	Numerical	
	Gender	Categorical	Female, Male
	Civil Status	Categorical	Single, With Partner
	Country	Categorical	Spain, Germany
	State	Categorical	Berlin, Madrid

Behavioural	Stay Motivation How did the client know about the hotel chain?	Categorical Categorical	Business, Leisure, Golf Internet, friends
Generated by the hotel chain	Repeaters Cancelation Cluster X done by the hotel	Binary Binary Categorical	Yes or No Yes or No
Transactional Data	Reservation Source Entity is the last visit Type of price applied	Categorical Categorical Categorical	Phone, Website, Email Agency, B2B Promotion

Source: Adapted from Talón-Ballesteros et al, 2018

2.5. SENTIMENT ANALYSIS

It is increasingly evident the **importance of reviews when choosing a hotel**, being even a decisive factor for most consumers who are increasingly demanding and with high quality standards. In Portugal the demand for reviews from other customers, where **comments** are mainly included, is the **2nd most important factor** when choosing a hotel (Pimentel, 2015) and, therefore, more and more consumers are keen to let them write a review to share their experiences and opinions.

If this affects consumer choice so clearly, then it means that for companies the analysis of this information is an essential process in the strategic definition to acquire and retain customers, but also to increase levels of satisfaction and loyalty to the brand. For this reason, more and more companies are using **sentiment analysis** and **text mining** as an automated processes of examining semantic relationships and meaning in reviews (s).

According to Alaei et al (2019) “**sentiment analysis**, in general, aims to determine the overall contextual polarity of (..) a review, an opinion, or an emotion (..), whereby polarity can be positive, neutral, or negative”. The analysis of feelings related to customer evaluation, shows that the information obtained through text (comments) is mostly based on opinions, personal feelings, beliefs, and judgments about the customer's experience (Alaei et al, 2019).

Table 7 - Dictionary for the sentiment classification

Reduced term	Similar terms or from the same domain
Strong Positive	brilliant, excellent ,fantastic, phenomenal, wonderful, superb, beautiful, spectacular, high-quality, top notch, top, incredible, best, greatest, delightful, memorable, very fresh, very nice, real change, perfect, delicious, amazing, happier, perfectly, very much, fascinating, gorgeous, unforgettable, speechless, dreamful, dream, great, fabulous, paradise, enchantment, exceptional, delights, outstanding, surprising, gorgeous, :-) ,deeply touched, extraordinary, 5 stars, 5 star, stunning, surrender, breathtaking, marvel, exclusive, delighted, splendid, pampering ,maximum, wow, happiness, precious, exceptional, dream place, thumbs up ,impressive, remarkably, adorable, overlooking, stunning, tastefully, enchanting, touching, amazingly, genuinely, favourite, exciting, magnanimous, impeccably, knowledgeable, loved, inspirational, perfection, beautifully, magnificent, stupendous, inspiring, blow-away, marvellous, dazzled, impressed, high standard, honoured, super
Strong Negative	terrible, awful, stupid, horrible, unfortunately, ridiculous, difficulty, really hard, too long, weaknesses, very bad, too slow, nightmare, stupidity, food poisoning, unnecessary, extremely slow, nuisance, regrettable, not respect, disturb, discrimination, prejudice, worse, price disparity, discrimination policy, angry, cheated, not worth, not professional, lie, overpriced, unnoticed, envy, endless, criticism, not recommend, complaint, disappointed, unpleasant, nauseating
Ordinary Positive	cool, good, fashionable, better, cozy, modern, goodness, helpful, peaceful, beauty, quality, warm, respect, tasty, recommend, spacious, pleasure, elegant, sincere, joy, enjoy, enjoyed, smiles, serenity, worthy, comfortable, tasteful, excellence, come back, wealthy, harmony, sanctuary, happy, success, encouraging, luxury, sublime, affection, elegance, merit, diversity, fine, gratitude, inspiration, smiles, finely, picturesque, congratulations, congrats, thanks, thank you, remember, well-tasted, affection, surprise, healthy, large, fancy, interesting, grateful, courage, thankful, celebrate, cute, curious, delicacy, appreciate, return, like, liked
Ordinary Negative	bad, nervous, loss, aversion, sad, difficulty, quite small, little scattered, expensive, shame, unbalanced, spoiled, burst, tears, shy, back pain, missed, raw, blood, ignorance, bad time, mistakes, tough, failed, stressful, against, regret, doubly, stop, noise, rivals, hard, lacks, limited choice, less, dirty, blamed, concern, problems, trouble

Source: Calheiros, 2015

In table 5, there are a **sentiment classification dictionary** and the most common words assigned to each category (positive, negative, and neutral). It is through these ranking dictionaries that the predictive algorithms calculate the respective sentiment analysis score. The predicted scores vary between 0 and 1, with values less than 0,45 being negative, between 0,45 and 0,6 being neutral and values above 0,6 being considered positive.

While through **text mining** analysis it is possible to obtain the frequency of words obtained repeatedly in the comments and to classify these results as indicators of satisfaction or dissatisfaction (Chittiprolu et al, 2021). The most relevant output of this analysis is finding out the interesting patterns in text data. Using text mining techniques, it is possible to identify factors that are more determinant of customer satisfaction, their recommendations, customer standards and their complaints through the categorization of this information by frequency with which they are mentioned.

The processes of text mining and sentiment analysis allow transforming unstructured information into extremely important assets to **identify business opportunities and improve customer service**. This is because it becomes possible to identify the main reasons for customer satisfaction and dissatisfaction in a more detailed way that is not possible through the customer satisfaction score (CSAT).

3. METHODOLOGY

In terms of methodology, this is an empirical research with a practical part, where the literature review will support the data that is necessary to accomplish the practical aspect and assure the quality of the data.

The **data source** that will be used in this study was obtained in the scope of a project, which I was part of, at **NOVA Marketing Analytics Lab**, then being made available for academic research purposes. Nova Marketing Analytics Lab is a research laboratory at Nova IMS that develop projects in the areas of marketing analytics, data-driven marketing and neuromarketing.

This project consisted of using statistical methods and text mining to analyse an **extensive database of customer satisfaction questionnaires for one of the largest hotel chains in Portugal**, with sites located scattered through all countries, which ended up resulting in a set of very relevant insights for them, both in terms of the analysis of that information, as well as the collection and treatment of this data, which is fundamental for the results to be interesting from the point of view of improving the business. The insights gathered and the respective opportunities for future improvements were later presented to the CEO and information technologies director and the result was very positive since it was possible to understand the main challenges and constraints of the company in relation to the data collection, treatment, and analysis process of this type of information, so strategically important for the business in terms of positioning and investments.

Initially the data set was made over a period of 5 years (2015 to 2019), with 348 203 questionnaires was submitted where only 66 158 (19%) were fully answered, and the reason for this stay in data collection process.

The survey is composed of 17 questions, including open-ended questions, multiple choice questions and point scale ratings (1 to 5).

For this study it was necessary to proceed the data cleaning and appropriate transformations according to the type of data, to guarantee the integrity of the data, the veracity of the analysis and to keep the data relevant for the research questions at hand.

The tools that were used to make the data cleaning, transformations and processing were Microsoft Excel and Power BI, the last one was also used to develop any measures and KPI's necessary to the final dashboard and to all analysis performed.

The process of data cleaning and transformation took the following steps:

In Excel:

1. **Deleted columns that where not relevant to the analysis** (such as authorization for data collection or the open response field, keeping only the column referring to comments as it was the one that allowed a greater coherence of responses).
2. **Reduced the database** so that it can be read by Power BI without incurring problems of processing heavy data. For this, several transformations were applied, such as:
 - **Reducing the database only to the last 3 years (2017/18/19)**, reducing from 348.203 records to 241.429.
 - After, **inconsistent data was treated**, since some were answered on paper and the transcription for the excel was manual (country column with some errors, which makes it impossible to be read correctly by Power BI, among other).
 - Finally, **randomly eliminated 50% of the sample**, resulting in 112.676 responses from customers, which for the purposes of this study, is a number that allows analysis with a high level of credibility but without incurring processing problems.

In Query Editor (Power BI):

3. **Deleted all missing values:** deleted all the rows where there was a null in one column of the table, for this there was one constrain in the satisfaction fields, which should have been from 1 to 5, but some at a 0 as a non-response. To solve this, I replaced all the 0 in this column with a space, to then apply a formula to eliminate all null values contained in all columns.
4. **Rename columns** to facilitate interpretation for end users.

5. **Transformed the date format** to Date Only so it only showed the day, month, and year.
6. **Created new columns**, specifically, the custom column Stay Duration from the columns check-in date and check-out date, also created two new columns, the Year column and the Month column, by transforming the column Date.

With all this process of eliminating missing values and inconsistent data, we are left with a final sample of 21.637 records.

Also, some information was added to the system when the survey answers were being processed, such as the Repeater column (if the client repurchased or not). In the end of the transformation phase the database had the following structure.

Table 8 - Database Structure

Column Name	Data Scales	Description
Survey Code	Interval	Represent a simple unique identifier for survey.
Survey Type	Nominal	This column represents the type of survey (online or paper)
Hotel Name	Nominal	Corresponds to the name of the hotel where the customer was staying.
Check in Date	Interval	Date that indicates the day, month and year of customer's check in.
Check out Date	Interval	Date that indicates the day, month and year of customer's check out.
Stay Duration	Interval	This column corresponds to the measurement of the customer's stay duration, based on the difference in days between the check out and the check in dates.
Date	Interval	Date that indicates the day, month, and year of survey's response.
Year	Interval	Date that indicates the year in which the questionnaire was answered.
Location	Ordinal	This column indicates the degree of customer satisfaction in relation to the location, on a scale of 0 to 5.

Staff	Ordinal	This column indicates the degree of customer satisfaction in relation to the staff, on a scale of 0 to 5.
Cleaning	Ordinal	This column indicates the degree of customer satisfaction in relation to the cleaning, on a scale of 0 to 5.
Room	Ordinal	This column indicates the degree of customer satisfaction in relation to the room, on a scale of 0 to 5.
Comfort	Ordinal	This column indicates the degree of customer satisfaction in relation to the comfort, on a scale of 0 to 5.
Environmental Practices	Ordinal	This column indicates the degree of customer satisfaction in relation to the environmental practices, on a scale of 0 to 5.
Food	Ordinal	This column indicates the degree of customer satisfaction in relation to the food, on a scale of 0 to 5.
Sport and Animation	Ordinal	This column indicates the degree of customer satisfaction in relation to sport and animation, on a scale of 0 to 5.
Quality/Price Relationship	Ordinal	This column indicates the degree of customer satisfaction in relation to the quality/price relationship, on a scale of 0 to 5.
General Satisfaction	Ordinal	This column indicates the degree of overall customer satisfaction, on a scale of 0 to 5.
Recommend Hotel	Binary	This column indicates whether the customer recommends the hotel to family or friends. Recommend hotel is a binary variable, as the answers only assume the value 0 (no) or 1 (yes).
Source	Nominal	This is a nominal variable that show the reservation source, that can be email, website, phone, fax, walk in or other.
Market	Nominal	This is a nominal variable that show the type of trip that originated the reservation, that can be business, vacations, golf, congress or other.
Country	Nominal	This column indicates the client country.
Comments	Nominal	This column shows the clients comments related to their stay and experience.

Comment	Binary	This column indicates whether the customer has left any comments about their stay. Comment is a binary variable, as the answers only assume the value 0 (no) or 1 (yes).
Repeater	Binary	This column indicates whether the customer repurchased. Repeater is a binary variable, as the answers only assume the value 0 (no) or 1 (yes).

Source: Elaborated by the author

The data visualization approach and the transformations and analyzes carried out in this study seek to integrate the most valuable data analysis and reporting methodologies and technologies to manage the maximum added value and utility in the analyzes. Thus, it is possible to move to a level of superior analysis that allows more informed decision-making processes, oriented to results and in a timely manner.

4. RESULTS AND DISCUSSION

4.1. PRELIMINARY STATISTICAL ANALYSIS

In terms of results the preliminary statistical analysis, still within the scope of the **project of NOVA Marketing Analytics Lab**, it was possible to draw some conclusions regarding the correlation between the variables.

Through the application of several statistical tests performed using the SPSS tool, **linear regression, and binary logistic regression**, as it was possible to identify which variables, location, staff, cleaning, room, comfort, environmental practices, food, sport and animation and quality/price relationship (independent variables) had the greatest impact on dependent variable which is **general satisfaction (GS)**. In this process **several tests** were applied to understand the relationships between variables, such as **statistical significance, R-square** and **standardized beta**.

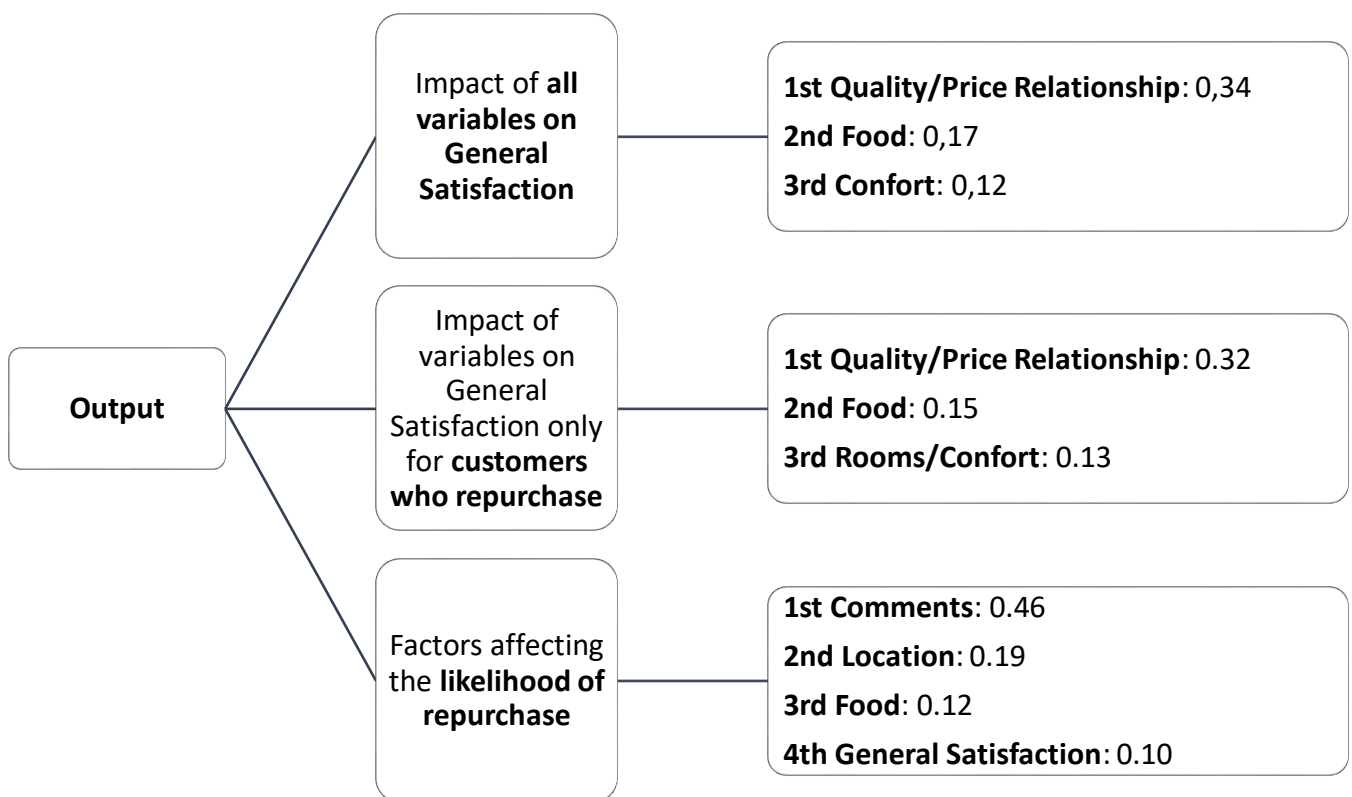
The **statistical significance** indicates if the independent variable impacts the dependent variable, and if the p-value is less than 0.005 we can conclude that the variable is significant, if is over than 0.005 is not significant and in this study, it was possible conclude that all variables fulfil this requirement. **The R-square** shows how independent variables explain general satisfaction, for example, if r-square is X means that independent variables explain in X% of general satisfaction. And finally, **standardized beta** indicates that a certain independent variable increases in percentage the general satisfaction.

From these preliminary statistical analyzes it was possible to draw some insight regarding the correlation between the variables:

- Regarding the **impact of all variables on general satisfaction**: variables explain 71.6% of general satisfaction, with the **quality/price relationship** standing out in first place, with a standardized beta of 0.34 which means that this variable explains in 34% general satisfaction. Followed by **food** and comfort which increases the general satisfaction by 17% and 12% respectively.

- About **customers who repurchase** the variables that most stand out are again the quality/price ratio (increased in 32% GS), followed by food (increased in 15% GS) and rooms/facilities (increased in 13% GS) with a technical draw.
- Through a binary logistic regression (introducing more variables to understand what affects the repurchase), it was possible to conclude that the **factors that most affect the repurchase** are first comments, following location and food and finally general satisfaction.

Figure 4 - Statistical Analysis Output Results



Source: Elaborated by the author

4.2. POWER BI DATA VISUALIZATION MODEL

As already mentioned, this study aims to go beyond simple statistical analysis and create a new approach to data analysis based on intuitive, interactive, and self-explanatory visuals with the power to present information in a consistent and objective way.

For this, several dashboards were developed using the Power BI tool, with the aim of analyzing the data in a segmented way and different perspectives. Altogether, 5 dashboards were elaborated each one with the purpose of extracting the maximum potential of the data, which are presented and discussed in the following sub-chapters.

4.2.1. Dashboard Customer Profile

The customer profile dashboard presents the general and complementary overview of the detailed **customer profile**, where four groups of variables (**demographic, behavioural, generated by the hotel chain and transactional data**) were analyzed and combined to build a consistent host profile.

Figure 5 - Customer Profile Dashboard



Source: Elaborated by the author

As can be seen above, this dashboard presents a complete and consistent set of variables that allow the construction of the customer's profile. Within these characteristics, the customer **country** (demographic), **repeaters** (generated by the hotel chain), the **stay motivation** such as leisure, golf, business or congresses (behavioural) and the reservation source that can be by email, website, fax, phone or walk in (transactional data).

This method of data representation and analysis using a data visualization dashboard allows to characterize the customer profile in generic terms as follows:

- The European and south American markets are the ones that represent the majority of the guests.
- The reason for the stay is in 94.7% of the cases destined to leisure.
- In accumulated terms, only 13.7% of guests return to repurchase.
- In the channels chosen to make the reservation, customers mostly prefer email (32%) and website (27.9%).
- In combined terms, the average length of stay for a guest is 4.96 days.
- Leisure stays are the ones that stand out in terms of longer stay with an average of 5.5 days.

Through this dashboard it is also possible to have an integrated view of the different variables for certain periods of time (months, quarters, years) or for each hotel, being able to analyse all variables according to any one of them, which allows the decision maker to have a complete view of the data and build analyses according to the strategic needs at each moment.

Based on this premise, through the **“Repeater” pie chart** it is possible to deepen the analysis by the type of customer that repurchases and that does not repurchase by analysing the other variables in terms of these two customer profiles, and the added value is to be able to perceive the behaviour of this type of host to adapt the offer to their needs.

The column chart **“Average of Stay Duration by Stay Type”** shows the relationship between the average length of stay according to the motivation of the stay, which allows to identify different clusters of customers according to the type of trip they take. This allows the hotel to

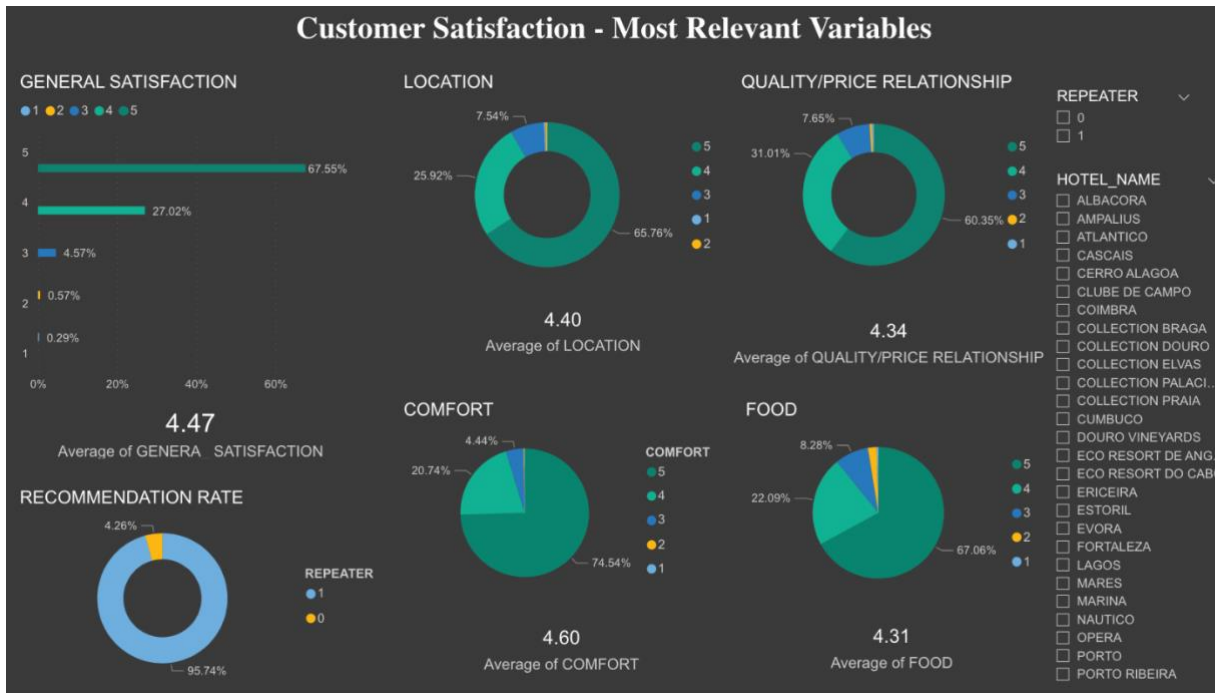
diversify its offer and respond in the best way to the expectations of these different types of customers.

Thus, with this first dashboard it's possible to have a set of relevant information and a very deep level of analysis of the customer's profile, which allows, before processing for any other type of analysis or strategic decisions, to have a clear and detailed definition of the audience. Being able to not only characterize the generic profile but also to identify the various types of customers is an extremely relevant exercise because each of them has different needs and expectations regarding their experience. As previously mentioned, customer satisfaction corresponds to a relationship between their expectations and emotional relationship with the experience someone had with a product or services (Shamsudin et al, 2020).

4.2.2. Customer Satisfaction Dashboards

The customer satisfaction dashboards are divided into "Most Relevant Variables" and "Other Variables" for clearer and more visually attractive data reading. These two dashboards have the purpose of giving visibility on the results of each one of the variables that determine the customer satisfaction, results that correspond to the answers of the customers through the surveys of satisfaction. In this way it is possible to define the aspects that are most valued and deserve more attention and those that are not relevant to the customer or that should be optimized by the hotel.

Figure 6 - Customer Satisfaction Dashboard: Most Relevant Variables



Source: Elaborated by the author

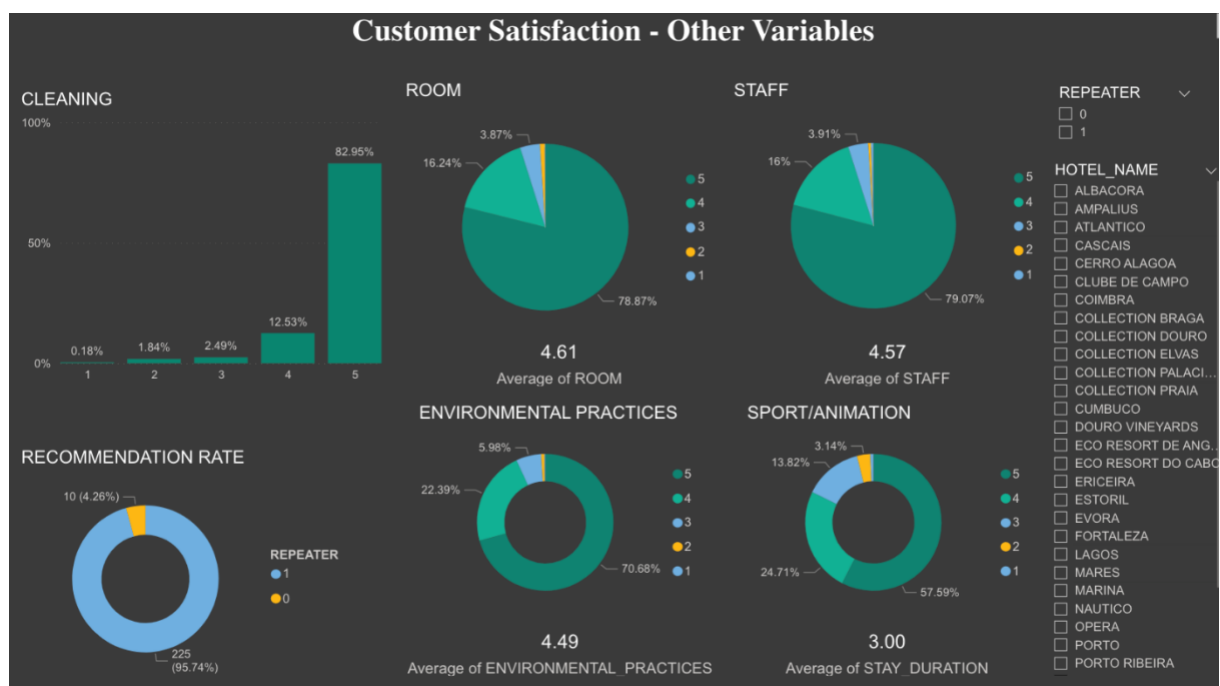
In the dashboard above, a detailed results analysis of the variables that most impact customer satisfaction is represented, being them “location”, “quality/price relationship”, “comfort”, “food” and “general satisfaction”. These results are presented on a scale of 1 to 5 through several graphical representations that allow us to have a view of the level of customer satisfaction (in percentage) in relation to each of the questions, with the scale varying from very dissatisfied to very satisfied.

In each of the five graphs relating to the questions in the survey show the percentage of responses for each level of the scale. For example, the “General Satisfaction” bar chart shows that more than half of the customers (59.49%) attributed the maximum score revealing that in general they were very satisfied with their stay. The remaining four graphs follow exactly the same approach and for each of them it is still possible to see the calculation of the average score from 1 to 5.

In this way, it is possible to perceive, in a very direct way, which aspects are contributing positively or negatively to the experience of the guest and thus planning and adapting the business based on feedback from the customers themselves. As in the “Customer Profile Dashboard”, all of this information can be filtered by hotel and by type of customer who repurchases or does not repurchase.

Another important highlight in this dashboard is the presentation of the recommendation rate, which is represented by a pie chart in which 1 corresponds to the percentage of customers who recommend the hotel and 0 exactly the opposite. Thus, being able to build the consumer profile that recommends the hotel to family or friends, that is a customer certainly with a good experience and is loyal to the brand.

Figure 7 - Customer Satisfaction Dashboard: Other Variables



Source: Elaborated by the author

As it was said, the dashboard above is an extension of the previous one and follows exactly the same approach and the same filter options but shows the remaining variables, that in this context do not have as much impact on customer satisfaction. This segmentation of variables into relevant results not only resulted from theoretical research related to this sector but also from preliminary statistical analyzes to this study that were applied to precisely the same context and database as it possible to see in “Preliminary Statistical Analysis” section of this study.

In short, the dashboards related to customer satisfaction variables introduce particularly relevant concepts like the **recommendation rate**, which assesses the customer loyalty, reflecting their satisfaction with the brand.

These dashboards allow a very intuitive reading and understanding of the data and provide a comparison analysis indispensable for decision making process.

4.2.3. Key Influencers and Top Segments Dashboard

The following dashboard, “**Key Influencers and Top Segments**” was built with the purpose of identifying **which variables affect the most the general customer satisfaction**. Through this visual representation that Power BI makes available, it is also possible to have access to a more detailed analysis about the **relative importance of these variables**.

With this data, it is possible to have an overview of which factors are most determinant of customer satisfaction and, at the same time, it indicates in a quantified way how an increase or decrease in the score of each of the variables translates into the score of the general satisfaction. At the end of this analysis, one of the objectives is, among others, **to compare the results of the preliminary statistical analysis** made in SPSS, with the **analysis produced in this dashboard** and to see if they match.

Figure 8 - Key Influencers and Top Segments Dashboard



Source: Elaborated by the author

Starting with the key influencers, we can choose one variable and see the variation impact in the general satisfaction (dependent variable). The idea is, by choosing one variable and locking the remaining it is possible clearly see how much a change in the independent variables affect the dependent variable.

Analysing the main results, it's possible to draw the following conclusions:

- The “**Location**”, “**Quality/Price Relationship**” and “**Food**” variables are the ones that have the **most impact on general satisfaction** which means that a variation in their value significantly affects the overall satisfaction score.
- When **location** is more than 4, the average of general satisfaction score undergoes an increase of 0,63. On the other hand, when the score is this question is 3 or less the average of general satisfaction decreases by 0,82.
- **Location** is undoubtedly the variable that most affects the general satisfaction score, especially when compared with the remaining independent variables. A variation in this variable significantly affects the dependent variable (general satisfaction) and, therefore, should be considered in strategic decisions, being clearly a point that is valued by the guests and that influences their opinion considerably.
- The “**Quality/Price Relationship**” is the variable following the location whose variation translates into a greater increase in the general satisfaction average, followed by “**Food**”. The analyses applied to these variables follow the same approach that was used in the localization.
- In the bar chart example on the left, **the average line** is calculated for all possible values of general satisfaction, except for location (which is the selected influencer).

In this way it is possible to conclude that the results of the relationship between the variables made through this data visualization tool match with the results obtained in the preliminary statistical analysis made previously using SPSS tool. However, and as already mentioned, the performance of these analysis using Power BI or other data visualization tool allows a much more intuitive and clear reading of the data and results, making them accessible to different types of end users and stakeholders.

Figure 9 - Key Influencers and Top Segments Dashboard



Source: Elaborated by the author

Another of the analysis that can be done in this dashboard has to do with the definition of top segments according to the sample size, segmenting the different groups of average general satisfaction scores.

This analysis allows the identification of clusters that contribute to the increase or decrease in the result of general satisfaction measured in different dimensions of the sample. This segmentation is particularly relevant because it allows for the identification of response profiles and the understanding the sample size that responds to each type of result.

For example, when selecting Segment 1, we can conclude that of the customers whose responses with the highest overall satisfaction score (4.97) represent 24.4% of the sample. The overall satisfaction of these customers averaged 4.97 since the variables "Comfort", "Food" and "Quality/Price Relationship" are evaluated with a score equal to or higher than 4, what makes this segment overall satisfaction 0,61 units higher than the average.

This reasoning is also applied to the segments with a lower score in general satisfaction and can also analyse the behaviour of this value according to the different scores in each answer and the respective sample size.

In brief, through this dashboard segmented into two types of different but complementary analysis (key influencers and top segments) it is possible to have a complete view of which variables most affect overall satisfaction and how their variation (positive or negative) translates into an increase or decrease in overall satisfaction (the dependent variable).

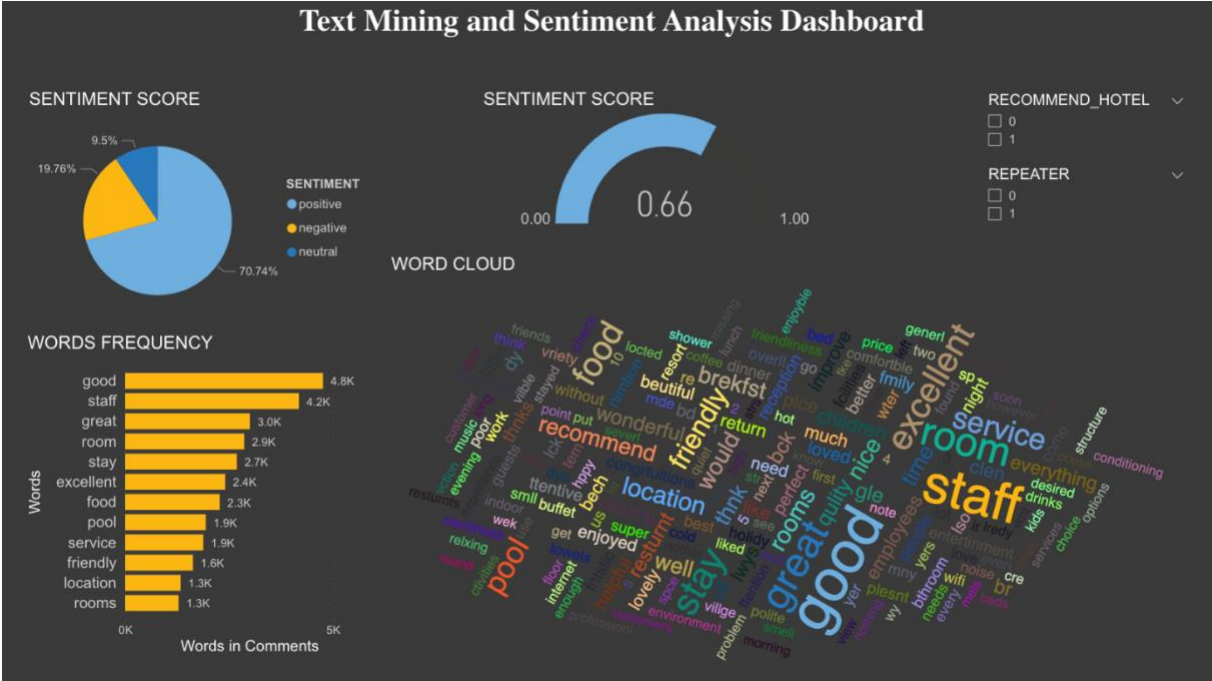
On the other hand, the identification of the top segments allows to define groups of scores of different sample dimensions and to perceive within each one of which variables and which score of them allows to reach a determined result in the general satisfaction.

These two perspectives allow to strategically define which aspects are most valued by the customer and which are the ones that most positively or negatively mark their experience with the brand, to give greater strength to the action plans to improve the experience of the guest.

4.2.4. Text Mining and Sentiment Analysis Dashboard

As previously mentioned, it is increasingly evident the **importance of reviews when choosing a hotel**, being even a decisive factor for most consumers who are increasingly demanding and with high quality standards. The added value of text mining and sentiment analysis is to allow transforming unstructured data into extremely important assets to **identify business opportunities and improve customer service**.

Figure 10 - Text Mining and Sentiment Analysis Dashboard



Source: Elaborated by the author

Given the relevance of these analyses and considering that the database of this study contains customer comments, an exclusive dashboard was built for **text mining and sentiment analysis**. This dashboard was developed with the aim of identifying in detail which aspects are most valued and criticized by guests and to **understand which feelings, emotions and perceptions** reflect the customers' experience.

Starting by **customer sentiment analysis**, it allows assigning a score to each word/comment and identifying it as positive, negative, or neutral. The predicted scores vary between 0 and 1, with values less than 0,45 being negative, between 0,45 and 0,6 being neutral and values above 0,6 being considered positive.

Through the application of a predictive algorithm using Azure Machine Learning, it was possible to reach the following conclusions:

- The average score is 0.66, which means that comments are mostly associated with positive feelings and emotions.
- 70.7% of the comments are positive, 19.76% are associated with negative feelings and 9.5% are neutral.
- The approximately 20% of negative comments represent an opportunity for improvement and through the text mining analysis described below, it is possible to specifically identify the most determining aspects for guests and with this information define specific action plans to mitigate negative comments and increase levels of satisfaction.

Note: this information can be analysed by customer profile (recommend hotel and repeaters).

To better understand these scores, a **text mining analysis** was developed through two graphic representations, a **word cloud** and the respective **word frequency chart**. The word cloud is the graphical representation of a hierarchical list of words and each one has its size governed by the relevance with which it appears in the text.

In this dashboard we also have the "Word Frequency" bar chart that identifies the 12 words that appear most frequently in the comments and their respective count. These counts are represented in the word cloud and allows to complete the following:

- The most repeated word is the adjective "good" which is associated with positive feelings followed by other equally positive adjectives such as "great", "excellent" and "friendly", indicating that, in general, the customer's experience was evaluated positively.

- The most relevant aspects and, therefore, mentioned most often by customers were **staff, room, food, service, and location**. These results do not differ much from the analysis present in the "**Key Influencers and Top Segments**" dashboard, in which food, room and staff are also highlighted as factors that most positively affect general satisfaction.

In short, the customer sentiment analysis today assumes a powerful tool of key information to support decision-making by companies in the hotel sector and beyond, to improve the experience of their guests and ultimately increase the organizations profits and profitability.

However, there are **many advantages** of transforming comments into structured information because it involves listening to **the voice of the customer**, which founds a much more informed decision-making power. Among the different assets, the following stand out:

- The first and most evident is **to improve customer service** because the use of written feedback and not just scores, allows to generate relevant insights to optimize customer service and which inclusive can explain the numerical/score results.
- **Improves product and service development**, as it allows anticipating market trends, obtaining relevant insights on which markets are most attractive and profitable for the brand and adapting the existing and future offer based on this.
- **Optimizing marketing campaigns investments**, through deeper knowledge of what are the feelings and emotions arising from customer experiences, it is possible to define a more representative and precise audience, which results in more effective and profitable marketing actions.

Developing this text mining and sentiment analysis dashboards allows to easily and intuitively achieve the tools to accomplish a **stronger product/service, more satisfied customers, and bigger competitive power**.

5. CONCLUSIONS

5.1. FINAL CONSIDERATIONS

Business Intelligence tools and techniques that result in an approach to systematize, track, communicate and transform data into information, are a precious asset for decision making and strategic definition of any business. This approach and the implementation of sophisticated databases architectures supports data-driven decision processes, giving the business competitive advantages. The applications of data visualization tools increase the level of understanding of data-driven performance measurement, and it is also a powerful solution to make strategic decisions, anticipate business problems and using customer data to identify their needs at each stage of the sales funnel. In this sense, this study aims to prove that BI and Data Visualization approaches are extremely useful and relatively easy to build and implement, depending on the size and structure of the business.

This study was applied to the hospitality sector, with the purpose of applying these tools and techniques to customer satisfaction, a strategic area for the customer loyalty, acquisition and for the product/service improvement since customer reviews are the 2nd most important factor when choosing a hotel in Portugal. The customer satisfaction measurement is essentially used to understand the customers' expectations and needs, define how the company wants to position itself in relation to its competitors, improve the products/services based on customer data and have access to current trends to adept diversify offer.

As an answer to the research questions how business intelligence can drive marketing strategies and how data visualization tools can improve decision making processes, this study presents a new approach to data analysis based on intuitive, interactive, and self-explanatory visuals with the power to present information in a consistent and objective way. For this, a set of dashboards was developed using the power bi tool, to analyse the data in a segmented way and from different perspectives. In all, five dashboards were created each to extract the maximum potential from the data and add value to the decision-making process.

The preparation of these dashboards served to prove that a data visualization model, achieved through the implementation of different business intelligence tools and techniques, is a key asset for the optimization of the decision-making process and, consequently, the increase of competitive power by allowing in-depth knowledge of the customer, their expectations, needs and interests.

- The first dashboard presents a set of information and a very deep level of analysis of the customer profile, which allows, before processing for any other type of analysis or strategic decision, to have a clear and detailed definition of the target audience.
- The dashboards that analyse customer satisfaction variables introduce particularly relevant concepts such as the recommendation rate, an indicator that assesses customer loyalty, which is one of the biggest insights on the customer satisfaction towards the brand. These dashboards allowed us to obtain a very intuitive view and understanding of the data and are an indispensable means of comparison for decision making.
- Key influencers and top segments dashboard were built with the purpose of identifying which variables most affect general customer satisfaction. The performance of these analysis using Power BI or other data visualization tool allows a much more intuitive and clear reading of the data and results, making them accessible to different types of end users and stakeholders.
- Finally, to make this study as complete and interesting as possible, a dashboard was developed for sentiment analysis and text mining through the analysis of written comments given by customers. Through this last dashboard was possible to develop an easy and intuitive tool to accomplish a stronger product/service, more satisfied customers, and bigger competitive power.

In short, business intelligence and data visualization tools and techniques allow the development of dashboards with several KPI's and measures to improve marketing decisions, making use of customer data also enable companies and brands to have a greater attention to detail than in traditional methods as it shows a greater depth of data through filters or graphics, just zoom in on certain topics. It also has the characteristic of being easier to share, as an image to facilitate understanding, allowing the decision maker to share information with other stakeholders in an easy and understandable way being an excellent method to optimize communication.

Working with visual information, it is much easier to compare data, problems, trends, or business opportunities. These methodologies make it easy to interpret differences and find their causes quickly, intuitively and reduce the percentage of error. Another of the advantages that it was possible to identify in this study is related to the possibility of obtaining quick predictions by drawing conclusions based on past data, predicting patterns of repetition of these behaviours in the future.

Finally, all these aspects (such as the ability to predict trends, ease of comparisons and understanding) are decisive and determinant in the strategic decision-making process, as they allow the display of quick and objective views of performance and productivity, offering a better understanding of what is necessary to optimize. In this way, it is possible to make more fundamental decisions, with the most up-to-date information giving them a greater degree of confidence and effectiveness.

5.2. STUDY LIMITATIONS

The main limitations that were detected throughout this study are related to the structure and scale used to measure customer satisfaction through the questionnaire that served as the database for this study, as well as the lack of other more detailed information about the business, such as sales evolution in volume and value, fixed and variable costs, investment and returns.

The use of the Likert scale, which classifies scores from 1 to 5, has some limitations, such as the lack of qualitative depth, as several questions are usually asked that encourage the client to agree and/or disagree, and the reason for the which the customer disagreed or agreed with those statements.

There is still a difficulty in transforming the answers into objective metrics that are easy to be followed by decision makers, as the result obtained does not allow for an easy comparison with companies in the competition. Furthermore, it limits the type of metrics and KPI's that can be measured using this data.

However, for this study, the greatest constraint felt was the fact that the scale on the questionnaire made it impossible to calculate the Net Promoter Score (possible only through a scale from 0 to 10), one of the most important indicators for measuring customer satisfaction since it indicates how likely a customer is to recommend the product/service to a family member or friend. The result of this score by consequence is one of the strongest indicators to determine overall consumer satisfaction as it identifies customer loyalty to the brand.

Finally, there was a difficulty in performing deeper and quantitative analyses such as profitability analyses, sales evolution, comparisons with the market since the database used did not contain this type of data that would allow to carry out this type of analyses.

5.3. FUTURE STUDIES RECOMMENDATIONS

For future studies, it is recommended to conduct interviews with professionals of different levels and from different sectors within departments of strategy, BI, sales, marketing, or others relevant to the study in question. It would be interesting to understand with business stakeholders how they are adapting and using these data driven results, business intelligence and data visualization approaches and methodologies as an integral part of the decision-making process.

It is also relevant to complement the study presented with data and analysis of profitability (like sales, investments, operating costs, etc.) and understand the impact that these profitability data have on the data visualization model.

6. REFERENCES

Jeffery, M. (2010). Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know.

Janvrin, D., Raschke, R., Dilla, W. (2014). Making sense of complex data using interactive data visualization. *Journal of Accounting Education*, 32, 31–48.

Moore, J. (2017). Data visualization in support of executive decision making. *Interdisciplinary Journal of Information, Knowledge, and Management*, 12, 125-138.

Kurniawan, Y., Gunawan, Kurnia, A. S. (2005). Application of business intelligence to support marketing strategies: a case study approach. *Journal of Theoretical and Applied Information Technology*, 64 (1).

Rouhani, S., Ashrafi, A., Ravasan, A. Z., Afshari, S. (2015). The impact model of business intelligence on decision support and organizational benefits. *Journal of Enterprise Information Management*, 29 (1), 19-50.

Rouibah, K., Ould-ali, S. (2002). PUZZLE: a concept and prototype for linking business intelligence to business strategy. *Journal of Strategic Information Systems*, 11, 133-152.

Stodder, D., & Matters, W. D. P. (2016). Improving data preparation for business analytics. *Transforming Data With Intelligence*, 1(1), 41.

Grandhi, B., Patwa, N., Saleem, K. (2020). Data-driven marketing for growth and profitability. *EuroMed Journal of Business*, 1450-2194.

Li, S., Sun, B., & Montgomery, A. L. (2011). Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research*, 48(4), 683-700.

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., Zaki M. (2014). Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach. *Journal of Service Research*, 17(3).

Wang, C. H. (2013). Incorporating customer satisfaction into the decision-making process of product configuration: a fuzzy Kano perspective. *International Journal of Production Research*, 51(22), 6651–6662.

Hamzah, A., A., Shamsudin, M., F. (2020). Why customer satisfaction is important to business?. *Journal of Undergraduate*, 2 (1), 2710-6918.

Kotler, P., & Armstrong, G. (2010). *Principles of marketing*. Pearson education.

Hill, N., Roche, G., Allen, R. (2017). *Customer Satisfaction. The customer experience through the consumer's eyes*. The Leadership Factor.

Falkowski, A., Tyszka, T. (2009). *Psychology of consumer behavior*. Gdańsk: Gdańsk Psychological Publishing House.

Farris, P., Bendle, N., Pfeifer, P., Reibstein, D. (2016). *Marketing Metrics. The Manager 's Guide to Measuring Marketing Performance*.

Biesok, G., Wyród-Wróbel, J. (2011). Marketing and logistic problems in the management of organization (23-41).

Hardjono, B., San, L. P. (2017). Customer Relationship Management Implementation and its Implication to Customer Loyalty in Hospitality Industry. *Journal Dinamika Manajemen*, 8 (1), 92-107.

Patnaik, S., & Popentiu-Vladicescu, F. (2017). *Recent Developments in Intelligent Computing, Communication and Devices*. Springer Singapore.

Chu, X., & Ilyas, I. F. (2019). *Data Cleaning*. ACM/Association for Computing Machinery.

Fikri, N., Rida, M., Abghour, N., Moussaid, K., & El Omri, A. (2019). *An adaptive and real-time based architecture for financial data integration*. *Journal of Big Data*, 6(1), 1-25.

Rangarajan, L. (2010). *Bi-level dimensionality reduction methods using feature selection and feature extraction*. *International Journal of Computer Applications*, 4(2), 33-38.

Microsoft (2019). Optimization Guide for Power BI. Retrieved from <https://docs.microsoft.com/pt-pt/power-bi/guidance/power-bi-optimization>.

Hauser, H., Schumann, H., Liu, L., & Özsu, M. T. (2009). *Visualization Pipeline*. *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_1133

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). *Making data visualization more efficient and effective: a survey*. *The VLDB Journal*, 29(1), 93-117.

Bertini, E., Tatu, A., & Keim, D. (2011). *Quality metrics in high-dimensional data visualization: An overview and systematization*. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2203-2212.

Yellowfin (2016). 7 Steps to Virtualized Data Preparation. Retrieved from <https://www.yellowfinbi.com/campaign/virtualized-data-preparation>.

Börner, K., Bueckle, A., & Ginda, M. (2019). *Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments*. *Proceedings of the National Academy of Sciences*, 116(6), 1857-1864.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2018). What do we talk about when we talk about dashboards?. *IEEE transactions on visualization and computer graphics*, 25(1), 682-692.

Karami, M., Langarizadeh, M., & Fatehi, M. (2017). *Evaluation of effective dashboards: key concepts and criteria*. The open medical informatics journal, 11, 52.

Razak, M. A. A., & Nayan, S. M. (2020). *The price of customer satisfaction*. Journal of Undergraduate Social Science and Technology, 2(2).

Oliver, R. L. (2006). *Customer satisfaction research. The handbook of marketing research: Uses, misuses, and future advances, 1*.

Majumdar, A. (2005). *A model for customer loyalty for retail stores inside shopping malls—an Indian perspective*. Journal of Services Research, 27(1), 5-21.

Talón-Ballester, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., & Rojo-Álvarez, J. L. (2018). *Using big data from customer relationship management information systems to determine the client profile in the hotel sector*. Tourism Management, 68, 187-197.

Pimentel, D. M. M. A. D. O. (2015). *Os guest reviews no processo de decisão de reserva num hotel: o caso português* (Doctoral dissertation).

Alaei, A. R., Becken, S., & Stantic, B. (2019). *Sentiment analysis in tourism: capitalizing on big data*. Journal of Travel Research, 58(2), 175-191.

Calheiros, A. C. D. S. (2015). *Sentiment analysis in hospitality using text mining: the case of a Portuguese eco-hotel* (Doctoral dissertation).

Chittiprolu, V., Samala, N., & Bellamkonda, R. S. (2021). *Heritage hotels and customer experience: a text mining analysis of online reviews*. International Journal of Culture, Tourism and Hospitality Research.

