NOVA IMS Information Management School

MDSAA

Master's degree Program in

Data Science and Advanced Analytics

Smart Transactions, Customer Analytics and Market Tracking in Oeste Region

A Visual Analytics Solution for Smart Regions

Xavier Golaio Gonçalves

Project Work

Presented as partial requirement for obtaining the master's degree Program in Data Science and Advanced Analytics

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

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

A VISUAL ANALYTICS SOLUTION FOR SMART REGIONS: SMART TRANSACTIONS, CUSTOMER ANALYTICS AND MARKET TRACKING IN OESTE REGION

by

Xavier Golaio Gonçalves
M20201090

Project Work report presented as partial requirement for obtaining the master's degree in Advanced Analytics, with a Specialization in Business Analytics

Supervisor: Miguel de Castro Simões Ferreira Neto

Co-Supervisor*:* João Bruno Morais de Sousa Jardim

November 2022

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Assinado por: **Xavier Golaio Gonçalves**
Num. de Identificação: 14318592
Data: 2022.11.29 21:42:26+00'00'

Lisboa, November 25th, 2022

A Visual Analytics Solution for Smart Regions: Smart Transactions, Customer Analytics and Market Tracking in *Oeste* Region

Xavier Gonçalves[1], Bruno Jardim[1], Miguel de Castro Neto[1]

[1]Information Management School, Universidade NOVA de Lisboa, Lisboa, Portugal

m20201090@novaims.unl.pt

bjardim@novaims.unl.pt

mneto@novaims.unl.ptm

ABSTRACT

Numerous industries and sectors of the economy and society are using information management and Business Intelligence systems, given the multiple benefits for the business processes in an organization. One of the areas where it is possible to take advantage of Business Intelligence and Data Visualization is public administration, where gathering and analysing data generates a more efficient and productive solution, with the help of modern Business Intelligence tools. We are reporting the conduction of a set of activities whose procedures and results propose to improve the use of public funds and directing investment, namely in areas of superior value-added, thus creating a more effective public management standard. In this project we analyse a particular case, covering a region of Portugal, a territory composed of twelve municipalities comprising the Intermunicipal Community of the Oeste Region (CIMO). Through a transactional database, covering all the purchases paid by electronic card and the cash surveys conducted in this area, it was possible to analyse a set of consumer behaviours and recognize shopping patterns by region and profile. Data modelling was carried out and later a dashboard was built that presents the main results of the analysis conducted, and where it would be possible to take benefit from data visualization, creating powerful insights. As a result of this project, it is possible to deliver significant insights and tools to local public administrators with valuable information on the development and monitoring of card transactions in amounts and consumer profiles to help stakeholders in the process of making more informed and efficient decisions in the scope of business as recognized places of self-brand, promoting sales and the assertion of the municipal distinctiveness in the region. The following article was submitted to a peer-reviewed scientific journal entitled "IEEE Transactions on Visualization and Computer Graphics", published by the IEEE Computer Society, and which covers topics related to computer graphics and visualization systems, among other subjects.

KEYWORDS

Smart Transactions, Data Visualization, Business Intelligence, Customer Analytics, Market Tracking, Sector Analysis, Smart Cities

INDEX

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The growing trend of using Business Intelligence (BI) tools in the public (Adelakun, 2012; Judek et al., 2017) and private sectors (Purnamasari et al., 2019) has reduced inefficiencies and brought countless advantages. In support of what was previously based exclusively on human knowledge and experience, with precarious and insufficient analyses, inefficient decision-making was carried out with statistically lower quality results and greater probability of failure, as expressed by Mix (2016). Given the advances and the introduction of more powerful tools in information management, data is now better transformed and analysed, towards producing a synthetic, structured, and meaningful set of information, possibly made available for strategic decision making, considered a new definition of intelligence (Hopken et al., 2015). How to consolidate the information and how to access it are some of the main challenges faced (Burstein & Holsapple, 2008; Nanda et al., 2019), as making the information available for the analysis and production of advanced and cross reports remains as its main purpose (Kazi et al., 2018).

Using Data Visualization and Business Intelligence tools, the public sector is an undeniable beneficiary (Kimball & Ross, 2013) in numerous areas it controls and influences as an important stakeholder in the economy and society of a region or country (Piechorowski, 2018). The volume of data produced is vast (Vanegas et al., 2009) and challenging to interpret for the local stakeholders, such as planners, policy makers, which highlights the importance of data visualization in extracting insights and understanding them. Through systems such as those proposed here, it is possible to perceive many dynamics that occur in the territories. Realizing who is visiting, how much is spent, how long was the stay and their journey through the region, support a set of advantages which will result in an increase in productivity and in sales of local businesses (Hopken et al., 2015), but also establishing a better organization of environmental, human, and cultural resources (Sengupta & Cao, 2019). On BI platforms it is comparably possible to allocate and to reorganize available human resources more efficiently in the areas most directly impacted by the municipalities, such as administrative office, urban cleaning, construction, and gardening workers. Considering the importance of these analyses in a context in which more and more functions and powers are delegated to local administrations (Olszak, 2014), in different areas, a correct integrated management of human resources (Kimball & Ross, 2013) is fundamental and can be made efficient.

In this paper, we verify the usefulness of Business Intelligence tools through a data analysis solution mainly related to transactions that took place in a region of Portugal in a period of nearly 34 months. Conclusions are made on the analysis of data in the scope of goods and services trade in the twelve municipalities, considering their potential relevance to the local and regional economy, enterprises and to the public administration, towards understanding the social dynamics and economic flow from corporate databases in local commerce, and subsequently obtaining the necessary evidence for decision-making processes.

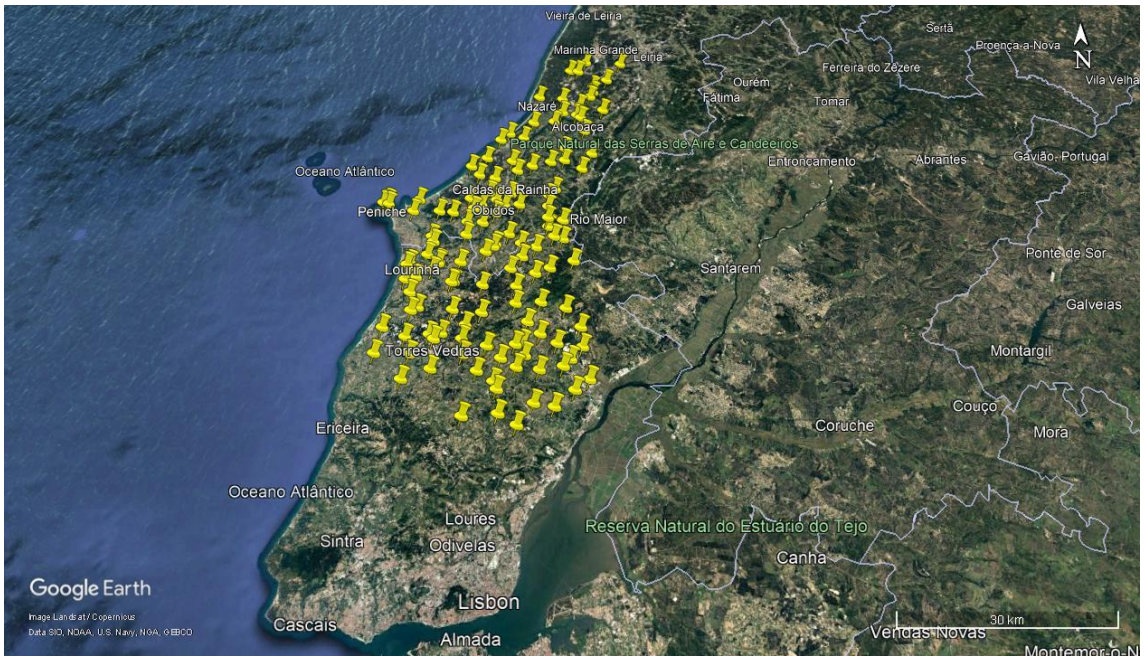

Figure 1 - Geographical Delimitation of the Oeste Region (Parishes)

The twelve municipalities of the region make up the *Intermunicipal Community of Oeste* (CIMO), namely: *Alcobaça*, *Alenquer*, *Arruda dos Vinhos*, *Bombarra*l, *Cadaval*, *Caldas da Rainha*, *Lourinhã*, *Nazaré*, *Óbidos*, *Peniche*, *Sobral de Monte Agraço* and *Torres Vedras* (CIMO, 2014).

2. Literature Review

Business Intelligence methods have been increasingly used by organizations (Chen & Siau, 2020; Richardson et al., 2020), including public sector (Popescu et al., 2020) and are responsible for better business and management decisions (Elena, 2010). A few years after development, its analytic component was emphasized, and the concepts of Business Intelligence and Data Analytics were substantially associated. Organizations, whether public or private, are collecting progressively new and ever more complex data. As reported by Kandel et al. (2012), Corporations depend on data analysis to fulfil tasks that that increase efficiency in costs, logistics and boost their sales, their brand in the market, or even accomplishing non-profit goals. By modelling customer engagement, streamline processes, organizations upgrade their business decisions, and combat fraud. Organizations, have been investing significant amounts in changes in their structures, according to Popescu et al. (2020), to process data and apply technologies in the scope of its collection, storage, access, analysis, and visualization (Nanda et al., 2019). Such information serves the purpose, in the scope of business intelligence, of supporting the processing in the decision-making activities. In this context, it has similarly gained position in local management, given its benefits and advantages (Olszak, 2014). Consequently, the relevance of BI depends on the relationship between the information delivery process, and all that individuals identify as the real needs of the organization (Teixeira et al., 2014).

Barbieri (2012) & Olszak (2014) explore the possible uses of Business Intelligence in local administrations, which are diverse, varying from water and electricity supply, waste management (Kimball & Ross, 2013), transportation systems (Shahini, 2020), urbanism (Mota, 2021), health structures and human resources management (Hartley, 2015). In this sense, its scopes stand in planning, task delivering and financial control, towards monitoring the performance of projects and activities, and subsequent evaluation of performance indicators, using more successful development analysis (Barbieri, 2012), and raising the main indicators of efficiency and cost effectiveness of projects (Burstein & Holsapple, 2008).

A different noteworthy application occurs in the fiscal area, with the potential of better integration of the information from land registry, place names, civil status, and property tax returns, to detect tax evasion and unauthorized building, crime or fraud (Barbieri, 2012). The performance assessment may also be applied to institutional affairs, assessing the productivity in case management through the analysis of the duration, detailed in stages, sectors, and progress (Judek et al., 2017).

The importance of BI tools regarding the potential for local commerce is associated to the way in which small businesses and commercial establishments may be capable of positioning their products and services in local markets (Khan et al., 2020). Public organizations may therefore make policies and conduct information to private economic agents (Lock & Gladstone, 2010), hence serving small and medium-sized companies to obtain relevant information about their consumer buying preferences, consumer purchases' time scales or income, and budget cycles. Therefore, customer targeting and business segmentation (Khan et al., 2020) are encouraged. In this way, local businesses in the region are capable of obtaining additional knowledge about market perspectives and thus be assisted in developing new (Purnamasari et al., 2019) and better products (Sengupta & Cao, 2019) and shift product portfolios.

In the context of Data Visualization, reports from instruments such as *Power BI*[1] or *Tableau*[2] are utilized to create interactive dashboards for target users (Piechorowski, 2018). Various visualization methods have demonstrated the efficacy of visual data exploration (Keim et al., 2007) in the proposal to find trends and patterns, in the visualization of diagrams and graphical statistics regarding the reported data. The proposed application of a dashboard is represented by the selections in its graphic layout and operational capabilities. Their purposes take place in the performance of dynamic analysis on a range of a topic information. In accordance with Sarikaya et al. (2019), the two main purposes of using dashboards are decision support, either in strategic, tactical or operational areas, but also as a communication and learning tool. On the one hand, dashboards will help an organization to plan and evaluate the impact of strategies or their results. On the other hand, the communication and learning factor emphasizes the production of decisions from an external perspective, with the purpose of teaching and educating the observer or analyst (Sarikaya et al., 2019).

In this context, the analysis of transactions that took place in a delimited area gives the user a general perspective and a variety of analyses on specific indicators in the area along with measures of analysis, as well as their chronological and geographic location (Dataflair, 2019). With the help of incorporated visualizations, data that were previously imperceptible, in large and complex volumes, can be interpreted (Kalaiarasan et al., 2020) and its organized analysis performed, i.e., to find patterns or trends, thus recognizing partnership and sponsorship opportunities (Chen & Siau, 2020). In economic terms, knowing where people are and consequently their consumption, might help to understand which are the geographic areas of greatest value (Kimball & Ross, 2013), of utmost interest, and through this information to be able to carry out a better allocation of resources with adequate urban policies. Illustrations exist when improving the pedestrian public space (Fangxiao, 2022), better transportation networks (Zhu et al., 2019), inside the domain or concerning connections with the outside.

However, there are challenges in the application of Business Intelligence systems (Teixeira et al., 2014) by organizations. When dealing with political management organizations at local level, there may be limitations regarding political and social matters, data privacy and local or regional law (Shahini, 2020). Other debate might consider the changes in existing technology (Olszak, 2014), the movement of data to cloud technology systems, the incorporation of data from multiple sources and with different origins (Khan et al., 2020), and data security and privacy (Gavrilov et al., 2018) are among the main challenges for these systems.

[1] https://powerbi.microsoft.com/
[2] https://www.tableau.com/

3. Data

The main source of data collected and managed in an OLTP environment is data collected by SIBS Analytics, an organization specialized in collecting and processing data obtained. The solutions provided offer aggregated consumption data through the multiple channels managed by SIBS, adapted to the idiosyncrasies of different realities, groups, companies and organizations.

At this point, *Oeste* region and its municipalities are encouraged to develop a more intelligent region from the perspective of its data usage and its purposes. The technical and dynamic analysis of the movements and transactions becomes an important pillar (Honcharenko et al., 2021). Data related to transfers imply the aggregation of a set of absolute values, relative values, and ratios, as well as their historic and location, with precision and granularity (Burstein & Holsapple, 2008).

The *Oeste* is a section of mainland Portugal made up of municipalities which are mainly located in the north of the Lisbon district and in the south of the Leiria district. It is a statistical territorial unit - NUTS III of the country (PORDATA, 2022). According to the censuses of 2021, it had around 355 thousand inhabitants (PORDATA, 2021). The region has coastal influence, but some of the municipalities are inland and rural. Some of the most relevant economic activities are fruit and vegetable farming, as well as some industry and tourism, given its vast historical heritage and appealing beaches (CIMO).

The operated data for the analysis is generated by individuals who make payments at an automatic payment terminal and carry out cash withdrawals at any ATM in the region. The data is generated and has a weekly aggregation, by weekday or weekend, in a tabular view. The data is mainly originated from purchases of goods and services carried out in the *Oeste* region, with each entry referring to the sum of the amounts of transactions, number of cards, number of operations, average number of operations, average number of operations per card and average value per transaction that took place in one of the defined periods, in each parish, with an implicit geographical origin from the card. The focus of the analysis is the set of the mentioned indicators at the level of aggregation in absolute value, such as the total amounts, but also the displays of cards and average transactions that have taken place. Table 1 compiles the variables that represent the tabular views, as well as the remaining relevant variables.

Concept	Description
Card Origin (Origin ID / Origin DESC)	Municipality in which the card carried out a greater number of operations (purchases, withdrawals, and payments for services/purchases) at physical points (TPAs and CAs) in the last 12 months.
Country of the Card (Origin ID / Origin DESC)	Country of the card issuing bank.
Type of Day	Relates to the type of day on which the operation was accepted (weekdays or weekends).
Municipality of the Transaction (Local DESC)	Municipality of the establishment where the TPA in which the purchase was made is registered. In the case of operations in CA, it corresponds to the Municipality of the CA where the operation was carried out.
Parish of the Transaction (Local DESC)	Parish of the business (store) where the TPA is registered, in which the purchase was made. In the case of operations in CA, it corresponds to the Parish of the CA where the operation was carried out.

Activity Sector (Sector ID/ Sector DESC)	Sector classification based on the CAE (Classification of Economic Activity) of the shopper who is registered with SIBS. Detail of the sectors in the Excel file sent "CSA_SIBS".
Week / Week DESC	Week of operations, identified by the first day of the week. For example: "20220207", equivalent to the week starting on February 7, 2022
Type of Local	Municipality or Parish of the Oeste Region
Number of Cards	Corresponds to the number of payment cards used. This is a reasonable approximation to the total number of consumers.
Number of Operations	Corresponds to the number of Operations performed.
Value of Operations	Total value of transactions carried out, in euros.
Average Number of Operations per Card	Corresponds to the average number of operations performed by each payment card.
Average Value per Card	Average value used per debit card, in euros.
Average Number of Operations	Average value used per transaction, in euros.

Table 1 – Relevant Variables of the Model

In this context, the Business Intelligence solution materialized in the production of a dashboard in PowerBI makes the compilation of interactive and dynamic visualizations. Interactions are thus simple and insightful information is obtained regarding the dynamics and patterns of consumption, given the transactions that took place in this region during the period covered. The implementation of the BI solution makes a tool available to the *CIMO*, allowing the interaction and thus providing smarter decisions, within the scope of its competences as a public organization and composed of the public power of the municipalities. Hence, the creation of brands associated with locations and territories (Kavaratzis, 2009) could be a reality, incrementing economic development through the creation of a brand and a better e-commerce system. Great examples exist of decent use of these structures (Purnamasari et al., 2019), and each stakeholder is able to benefit from this promotion.

Municipalities and economic agents thus gain a tool for tracking and monitoring purchases and ATM surveys carried out and sectors and products where these amounts are applied, and therefore equally boosting sales in the territory. Another relevant gain is to understand who is passing through the territory, to understand how many people and what is their origin (Hopken et al., 2015). In this way, it may be more easily completed, or even complementing this BI solution, where who spends the most and who spends the most in average terms comes from, according to the same authors. The existing granularity at the level of the municipality of origin makes it possible to know with greater precision the location of the Portuguese consumer, since this is done in a more granular way than in the case of foreigners, where it exists only by country of origin. Even so, the limitation of the data is controlled, and its availability is sufficient to draw fundamental conclusions for the analyses.

4. Methodology

As stated by Khan et al. (2020), the most varied methods are used within the scope of Business Intelligence systems. One of the challenges of this type of structures is to keep up with the latest technologies, so that their final product is relevant to organizations and makes them capable of achieving the intended purposes, making their processes more effective (Kovacic et al., 2021). In this way, organizations will be able to increase their output, whether social, economic (Mix, 2016), or other. To do so, the first segment of the process schema to understand is the dynamics of data sources and storage. The developed transformation and cleansing procedures will increase the value and purpose of the data (Piechorowski, 2018). Agreeing to the author, procedures would use this architecture with the goal of increasing business purposes.

While the standard model of a Business Intelligence system relies on marginally distinct steps, the attributes of the organization's layout did not result in the creation of a Data Warehouse in the present Business Intelligence solution. Thus, the ERD model was designed in the context of the Power BI functionalities, and the Extraction, Transformation and Loading (ETL) process largely took place in the same environment, with infrequent changes prior to the moment when the information was loaded. The operational databases were sent by SIBS Analytics, under similar formats, from which the data were extracted, connecting with the Power BI model. Routine transformation tools were employed to extract the files, querying the weekly databases, headed to produce visualizations and the process of making relevant analysis available. Several applications might have been used within the scope of these analyses, such as reports, visual dashboards, data mining tools and models, or assessments on more individualized perspectives of the data.

Error! Reference source not found. Figure 2 represents a simplification of a Business Intelligence system architecture (Lavalle et al., 2019).

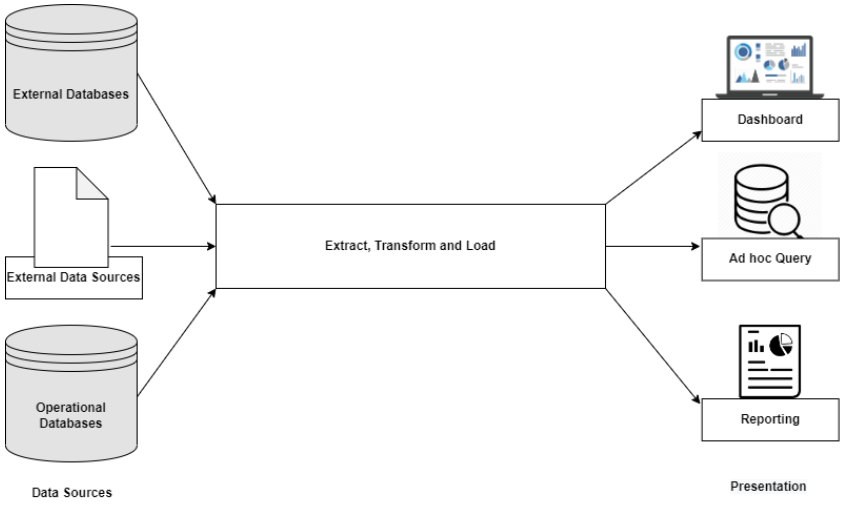

Figure 2 - Data Warehousing and Business Intelligence Architecture

Despite the power to find patterns or trends, there are other important advantages, such as the ability of more effective results and considering the needs of organizations, in forecasting and benchmarking (Richardson et al., 2020). The opportunity of easily discussing the main ideas with others (Kalaiarasan et al., 2020) is another important matter. On the other side of the scheme, visualization and reporting phases are crucial and exist as the main objective of the complete procedure.

Visualizations exist in the most varied structures (Richardson et al., 2020). From traditional reports, visualization in the form of Structured Query Language output, to a standard language for storing,

manipulating, and retrieving data in databases (Barbieri, 2012). Visualization of relevant data is based on capturing the attention of the audience and communicating findings in a short amount of time (Lavalle et al., 2019), making the audience access the performance of different insights to determine patterns and predict future years data (Africk & Levy, 2021). Visualizations must be accurate and easily understood by users (Lavalle et al., 2019), to obtain the correct analysis and capture users' viewing time and interest. The concept explored in this project has as its main way of assembling visualizations using a dashboard built in Power BI.

Thus, in the framework of analysing these data and creating a support tool that is easily updated and complemented, this project proposes the implementation of a Business Intelligence solution for the *CIMO* community. Evaluating the obtained data and the existing databases, a data model was created that efficiently schematizes the data.

Data Model

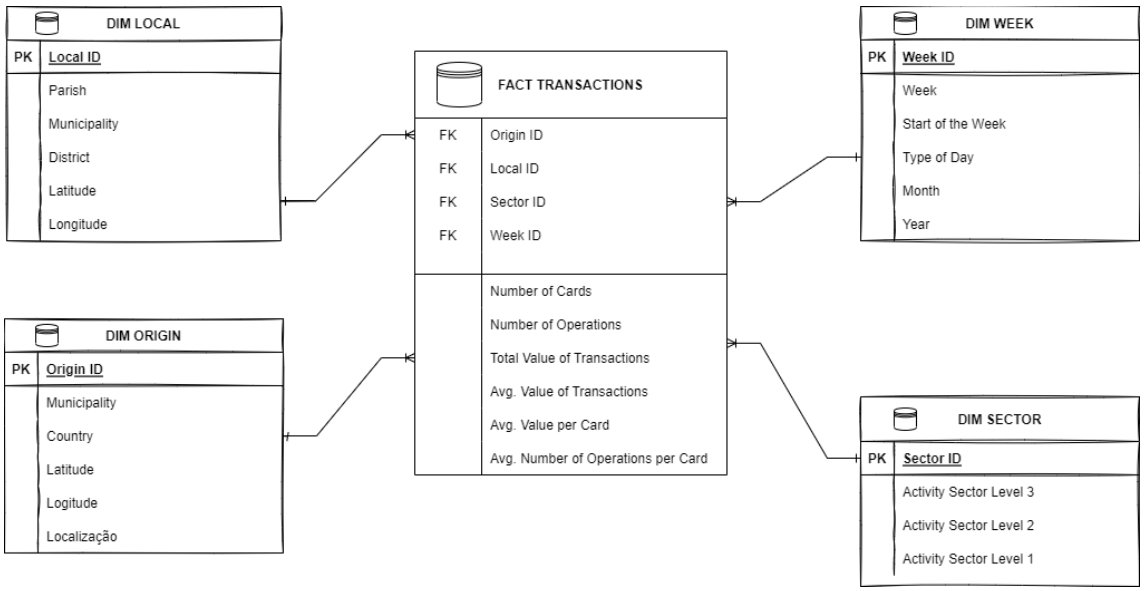

Figure 3 - Simplified Entity Relational Diagram (ERD) Proposed

In the context of Data Warehousing, the data model is represented as a set of relationships between populated tables, subject-oriented, integrated, normalized, time-variant and non-volatile data (Inmon, 2005), composed by data from one or more sources. Dimension Tables, which answer questions such as describing the facts and ranking them according to the questions "Who?", "What?", "When?", and the Fact Tables, which hold numerical measures to quantify and answer the question "How much?". Complying with the definition of functionality and other requirements that it must fulfil, the data model used must incur the desired functionality, which is called “Dimensional Model” (Judek et al., 2017; Kimball & Ross, 2013), a logical design method for communicating data in an efficient way (Dragicevic & Krneta, 2018). The model presents a Fact Table, the central dimension, where all other dimensions come together, in a Star Schema format, with the following dimensions:

- In the “*Dim Local*” dimension, the Surrogate Key (SK) “*Local ID*” is used, which locates the aggregation of transactions carried out, through the location code to which other attributes are associated, such as the name of the district, county and of the parishes that comprise it.
- “*Dim Origin*”, on the other hand, uses the field “*Origin ID*” as a Surrogate Key and specifies the origin of the card, which is complemented by a name of a municipality (in the case of Portugal)

or a country, for foreigners who have made purchases in the West, their latitude and longitude, as the main elements of this dimension.

- The "*Dim Sector*", which characterizes the consumption sectors. The attributes of this dimension are a subdivision of sectors into three levels. The "*Sector ID*" is the Surrogate Key and represents the code of the Activity Sector corresponding to the most granular level of the subdivision. The remaining levels are groupings of more granular sectors of activity, being valuable for integration into visualizations for more comprehensive view of a smaller set of areas where commerce is concentrated. For instance, "*Activity Sector Level 3"* indicates "Automobiles and Motorcycles Trade", while the second corresponding level stands as "Vehicles", and Activity Sector Level 1 is indicated as "Specialized Retail Trade".
- The "*Dim Week*" is another dimension, and it locates the purchases made in time, by the start date of the week. Its Surrogate Key, as in the other cases, which only indicates a date to the Fact Dimension is the "*Week ID*".

During the process and data transformation, various hierarchies were created in the dimension tables, necessary for some views to allow a drill-down and facilitate their handling.

- In the "*Dim Local*", the "*Local Hierarchy*" was created, comprising the attributes about the District, Municipality and Parish of the operation.
- In the "*Dim Origin*", the "*Origin Hierarchy*" was created, with information about the Country and Municipality fields.
- Lastly, In the "*Dim Week*", the "*Week Hierarchy*" was created, comprising information about the Year and Month (Number and three-letter identifier).

The central table in the schema is the Fact Table. At this point, Fact Tables are described and monitored (Barbieri, 2012) and store numerical indicators (Judek et al., 2017). In this case, information regards the transactions occurred. The facts in this dimension are the "*Number of Cards*", the "*Number of Operations*", the "*Total Value of Transactions*", representing the sum of the total of transactions, the "*Average Value of Transactions*", the "*Average Value per Card* ", and the "*Average Value of Operations per Card*", with all monetary values in the facts being represented in euros. The table contains information on all operations and cash withdrawals that took place in the region in the period under study, storing values in the respective financial value areas, consumer sectors, geographic regions, and timing in which they occurred.

On the one hand, in the initial information handling phase, to facilitate the creation of the model and allow further analysis, three auxiliary columns were built in the Power Query Editor that do not appear as visible in the final model, in the context of Dim Week, which made it possible to create a new Week ID with information about the type of day. Towards being able to handle the data according to the needs of the dashboard, a few measures and columns were created as to obtain an expected BI analysis. The filtering by measure of analysis, which chosen by the user, made it was necessary to generate measures that simulate five of the columns existing in the Fact Transactions table. On the other hand, two more measures were created to support the construction of breakdown perspectives of the average value of transactions per card used, conditional on the type of day of the week, and a third one to create the perspective of month in other visualizations. Auxiliary calculations made it possible to increase the range of views and articulate the model and existing visualizations, materialized in clearer and more effective results.

5. Results and Discussion

As a fundamental phase of the presented Business Intelligence solution, the results were obtained in the final stage of data visualization. The processing of the data thus made it feasible to have a conclusive purpose in obtaining several possible analyses and results, on the behaviour and patterns of consumption

in the region. In this context, we highlight analyses on areas with higher consumption in absolute values, average values, nationalities of consumers, its evolution through the period, among several other analyses that were carried out.

Regarding the Portuguese consumers, the relevant analysis is to understand where they come from and who they are, comparing, for instance, with foreign consumers, as well as to understand the dynamics that take place within the region itself. Among the twelve municipalities, there is free movement of people and goods, thus creating flows that must be understood for an indefinite economic and social decisions, realizing the purchasing power of citizens of different municipalities. In addition, local authorities could understand how the region is recovering from the pandemic that hit the world and that arrived in Portugal in March 2020.

5.1. Oeste: General Overview and Sector Analysis

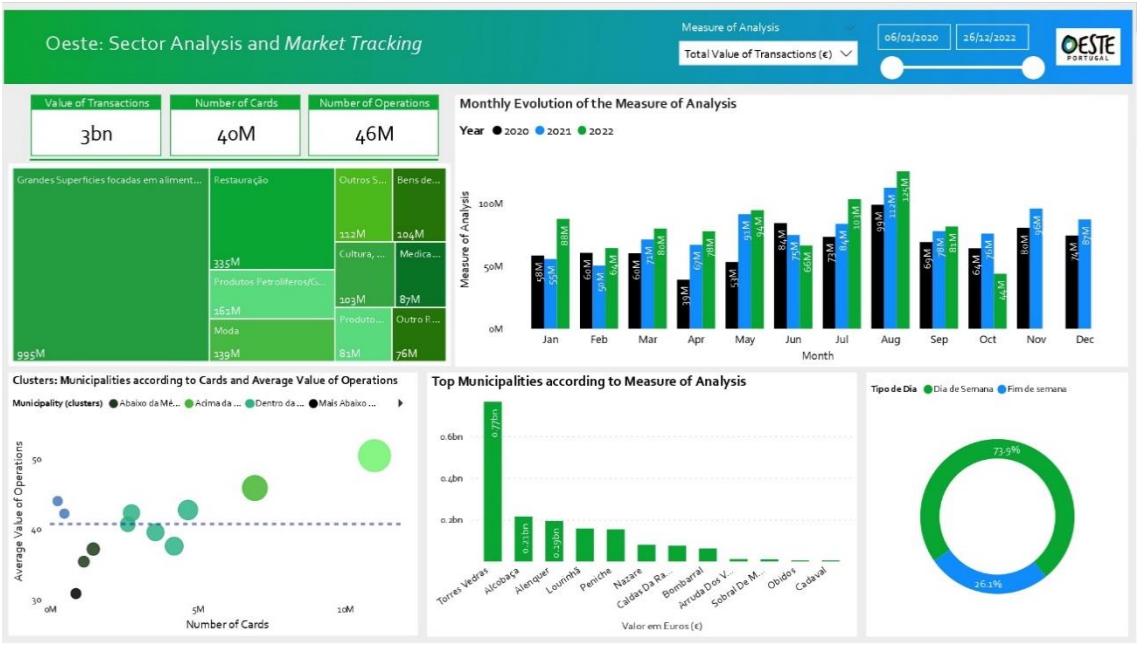

Figure 4 - Oeste: Sector Analysis and Market Tracking Tab

In the tab entitled *Oeste: Sector Analysis and Market Tracking* the user is invited to have a first outline of several key indicators in the region and an overview of the essential parameters. As such, and being a rule throughout the entire dashboard, the measure to be analysed is chosen through a filter, as well as its time horizon. These are just two basic filters that are available for this first stage, and it is possible to further interact with different graphs, such as selecting the part of the pie graph that concerns the measure chosen by type of day (weekday or weekend). In addition, graphs are available on the monthly evolution of the analysis measure, with annual and monthly perspectives, consumption sector, ranking of the region municipalities with the highest size of the chosen measure of analysis, as well as a plot chart that relates the number of cards and the average value of operations, which is a rigid graph and thus cannot be influenced by selecting the main filter on the sheet. At the top of the page, three main insights are highlighted: *The Total Value of Operations*, the *Number of Cards*, which represents a proxy to the number of consumers, as well as the *Number of Operations*.

Analysing the graph on the monthly/annual evolution of the analysis measure, and selecting "*Total Value of Transactions*", a phenomenon is seen across all years. The total value of sales has been increasing from year to year, and rises in the summer months every time, particularly if the two best months in the

observation are considered, which are July and August. The information for the year 2020 shows very low numbers of transactions for the months of April and May, months of intense lockdown in Portugal. A possible inference from this analysis would be to understand the impact that the pandemic had on the region and how it surpassed it, considering the data since 2020. In this analysis, we compare current periods with homologous periods towards understanding the changes that might have occurred.

If an analysis of the monthly evolution of the Average Value per Card is considered for 2022, March or April tends to be the months with the highest average values, and there is a trend across all three years for this value to fall during the summer months. It could, therefore, mean that during the warm months there is a greater volume of sales, but made by a much larger number of cards/consumers, who make more small purchases at a time.

On the other hand, the visualization on the left shows that there is clearly an advantage in the sectors of large food surfaces, supermarkets, and restaurants. Petroleum products (mostly car fuel) shortly occupy the third position. *Torres Vedras*, *Alcobaça* and *Alenquer* are in the top of municipalities with the highest volume of transactions, a label that essentially remains unchanged throughout the analysis in each of the three years covered. Torres Vedras is also the municipality highlighted as being very much above average, both in terms of number of cards (consumers) and in terms of *Average Value of Operations*. The city urban planner in these municipalities can improve on identifying trends and predicting future needs while changing behaviours. Understanding the relationship between the *Number of Cards* and the *Average Value per Card*, a possible approximation would be to characterize which type of tourist destination the region belongs to. This is pertinent whether agents would like to classify it as a low-cost destination or a premium destination. Most of the consumption takes place during weekdays (about 74% if we consider the entire data period since its beginning). The total amount of transactions in 2022 (until mid-October, around €824M in sales volume) has already exceeded the entire year of 2020 (€815M in the year of the beginning of the pandemic) and is relatively close to the total volume for the year of 2021 (about €942M).

5.2. Oeste: Market Tracking

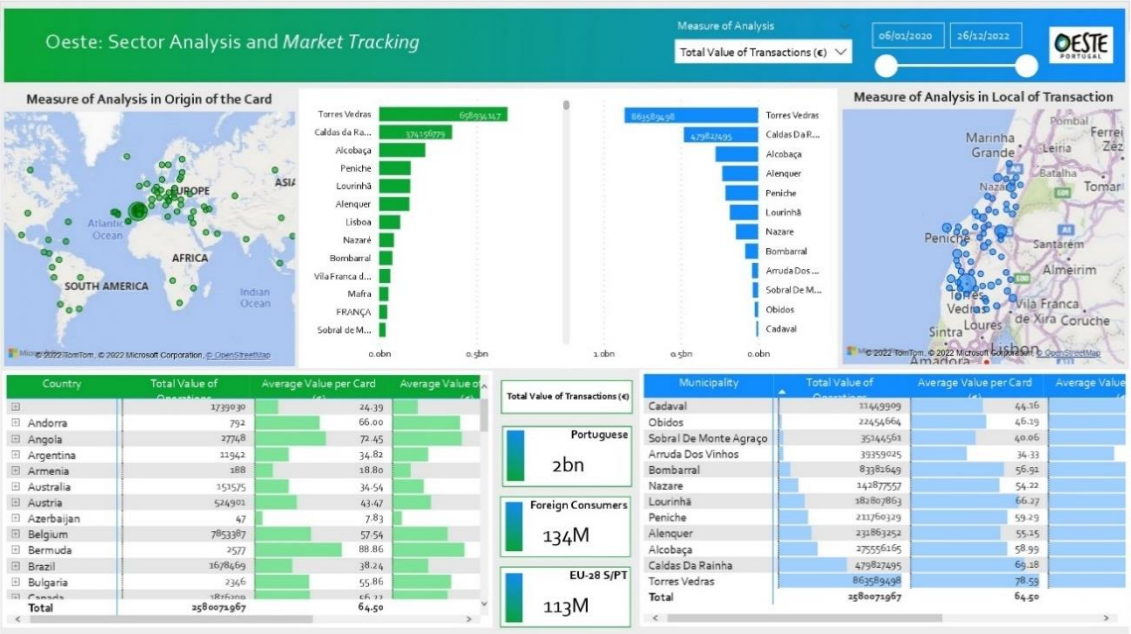

Figure 5 - Oeste: Sector Analysis and Market Tracking Tab

In the *Sectorial Analysis and Market Tracking* tab, the bipartite view displays the analysis measure chosen in the card origin filter on the left, and the same measure of analysis by location of transaction on the right, in the region defined by the twelve municipalities and their corresponding parishes. This tab makes the analysis of several dynamics related to transactions accessible, and thus designing a geographical view of the origin of consumers and the municipality or parish where they chose to generate wealth by consuming locally. The visual information is complemented by two tables that indicate the numbers respectively, and present bars supporting the interpretation of the amounts, as well as three main insights highlighted, which are the values of the chosen measure when consumption is made according to who is the target audience: One by the Portuguese, the other when it is accomplished by foreigners, considered all non-Portuguese, or in the third case when it is made by Europeans, disregarding Portugal in this last analysis.

Analysing the visualizations, the municipality of *Torres Vedras* is highlighted once more, considering the analysis on the right, if we contemplate the entire historical period, in all the measure of analysis chosen, due to its visual impact which confirms previous reporting. The analysis from the left includes more relevant insights. In almost all measures of analysis, the first location outside the west that appears in the ranking of origin of cards is Lisbon. There are, however, exceptions to this rule. When we consider the Average Value per Card, only two municipalities in the *Oeste* appear on the list, and all the others are, for instance, consumers from other municipalities in Portugal (*Mogadouro* with €153, *Oleiros* with €122, *Mealhada* with €116), as well as citizens from other countries (China with €131 and Bermudas with €89). On average, however, the Portuguese remain as the largest spenders per card, with around €66 against €50 for foreign consumers. The region received visits from around 3 million foreign cards in this period, which left over €134M in the area, of which €113M (84.3%) were spent by EU citizens. In terms of ranking, unsurprisingly, it was the French, the Spanish, the Germans, and the British who spent the most money in the West during this period.

Generating decisions about where to build infrastructures or maximizing the public utility of the existing ones could undertake the main objective of serving as a bridge between people and the private sector, businesses, enabling a better balance between local demand and supply. A relevant analysis would be to understand what the importance in values of the Portuguese emigrant community and its expenses are, its seasonality, and in a more complex analysis, how they differ from tourists coming from these same countries. In the same reasoning, it may become relevant for the public sector in the *Oeste* region to know the best markets to where local businesses are selling their local and tourist products. It is equally interesting to acknowledge the public domain perspective when understanding what is happening in real time at a location, or why consumers spend little or more money, and how much of that money is circulating in cash, as presented in a later analysis. More broadly, the specialization of parts of regions or municipalities may also be influenced by their ability of managing the fiscal situation of their territories, within the scope of local taxes, logistical support or with contributions of subsidies for regions, products, or brands in specific, to the extent of their capacities and the needs of the agents.

5.3. Customer Analytics: Municipality Level

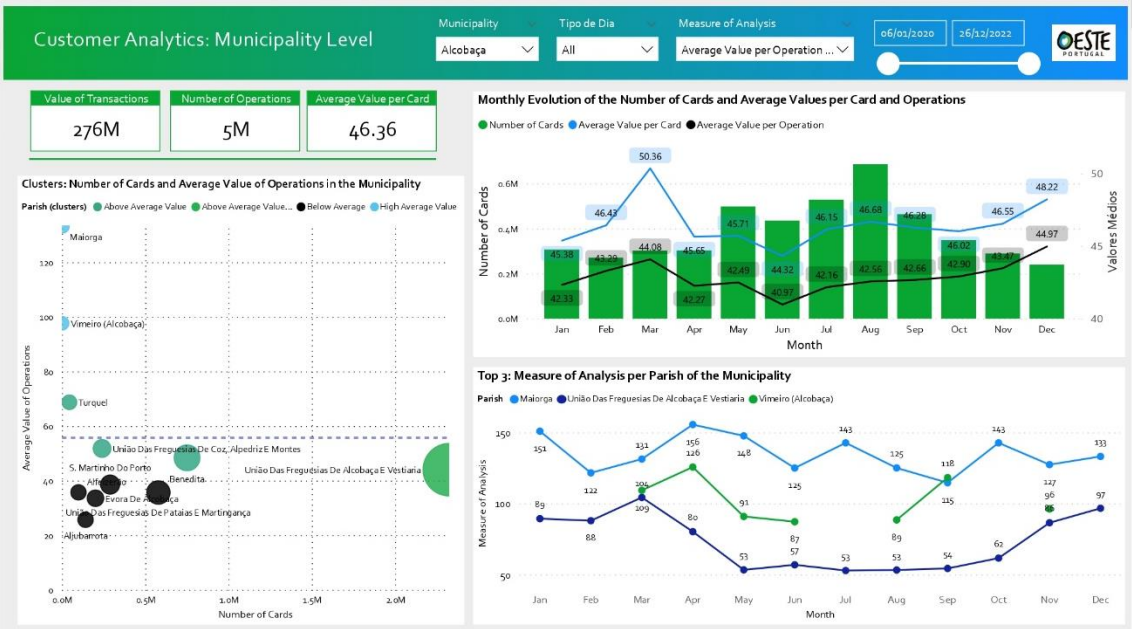

Figure 6 - Customer Analytics: Municipality Level Tab

In the next tab, the analysis is redirected to the county level. Here the users can select the measure of analysis, as in the previous tabs. Furthermore, the user may select one of the twelve municipalities and whether the analysis should focus on weekdays or weekends. There are several observations available at this point. The monthly evolution of the *Number of Cards*, *Average Value per Card,* and the *Average Value per Operation* is noteworthy. Another analysis is the scatter chart that holds the clusters obtained for the parishes according to the relationship between *Average Value of Operations* and *Number of Cards*, and, finally, the Top 3 of the parishes in the selected municipality with the highest measure of analysis chosen.

In this set of views, there is a single select in the municipality filter, a required unique selection for a more individualized analysis. The chart of the monthly evolution of the number of cards and the average values generally confirms the great trend seen initially. An analysis like the one encouraged here allows us to verify the parish level in relative terms, comparing with the chosen municipality in the general panorama, given that the clustering is done with all the parishes and only then is the municipality filtering applied. Thus, there are municipalities such as *Sobral de Monte Agraço* and *Óbidos* that have almost all their parishes below the average in the two parameters analysed (which may have a natural justification in the low population they have, at least in the parameter of number of cards), while that larger municipalities such as Torres Vedras or Alcobaça are much more representative of the general panorama of parishes in the *Oeste*. A parallel analysis allows us to pose another question: in some of the municipalities (i.e., *Arruda dos Vinhos* and *Bombarral*), there are sharp rises in the average parameters as we approach Christmas (November and December), but this does not happen in other areas such as in *Caldas da Rainha* or in *Torres Vedras*, with downward slopes or values nearly constant. This fact may suggest opportunities for improvement in local commerce and local Christmas marketing strategies.

5.4. Market Tracking: Municipality Level

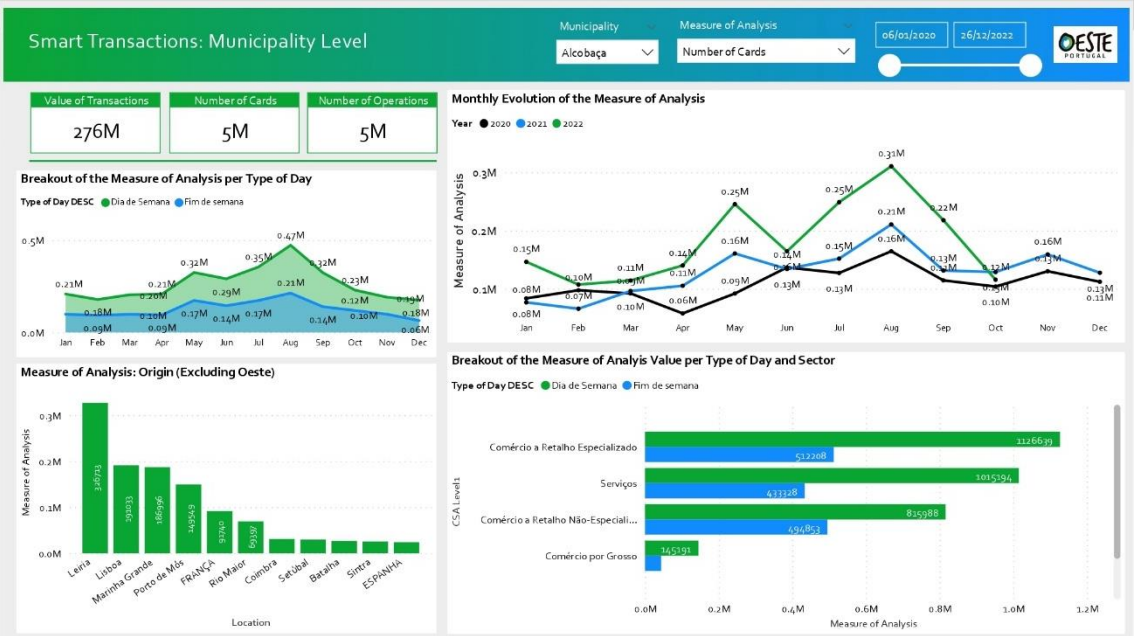

Figure 7 - Market Tracking: Municipality Level Tab

Then, the dashboard shows another set of possible analyses at the county level. The available filtering remains similar, with the choice of the possible measure of analysis, the intended analysis period, and the municipality. In part, it is expected to complement the previous tab at the same level. Regarding this fact, the analysis is complemented with information on the chosen measure during the week and at the weekend, contrasting both information. On the one hand, it is possible to analyse the measure chosen by week and by month, and then again using the ranking by sector (CSA Level). Through this analysis, it would also be possible to understand the dynamics between daily consumption on a weekday or else what happens on weekends, with the movement of people through the region, when tourism would be more expressive.

At this point, some analyses have common aspects comparing to those found previously for the entire region, when the objectives are currently understanding the breakout by type of day and by sector. For almost all municipalities, non-specialized retail trade or wholesale trade are the main segments where more money is spent in absolute values and in average parameters per card and per operation. However, in *Óbidos* there is an exception regarding the parameters associated with the segment of services, with a great lead over the remaining. In *Arruda dos Vinhos*, services are also leaders in some of the parameters, and in the others production and industry emerge as leaders (i.e., Average Value per Card and Average Value per Operation), inferring a relationship with the solid agri-food industry and others of the territory. In the same municipality, in February 2022, an unusual point was reached: the Average Value per Operation was higher on weekends, compared to weekdays, a difference of around €3.

The current assessment may allow the local economy to identify opportunities and plan accordingly, in the provision of products, logistics and services. Thus, the general communication between all sectors of the economy is improved, aligning the various stakeholders involved: the investors, the shoppers, and public administration, among others.

5.5. Customer Analytics: Parish Level

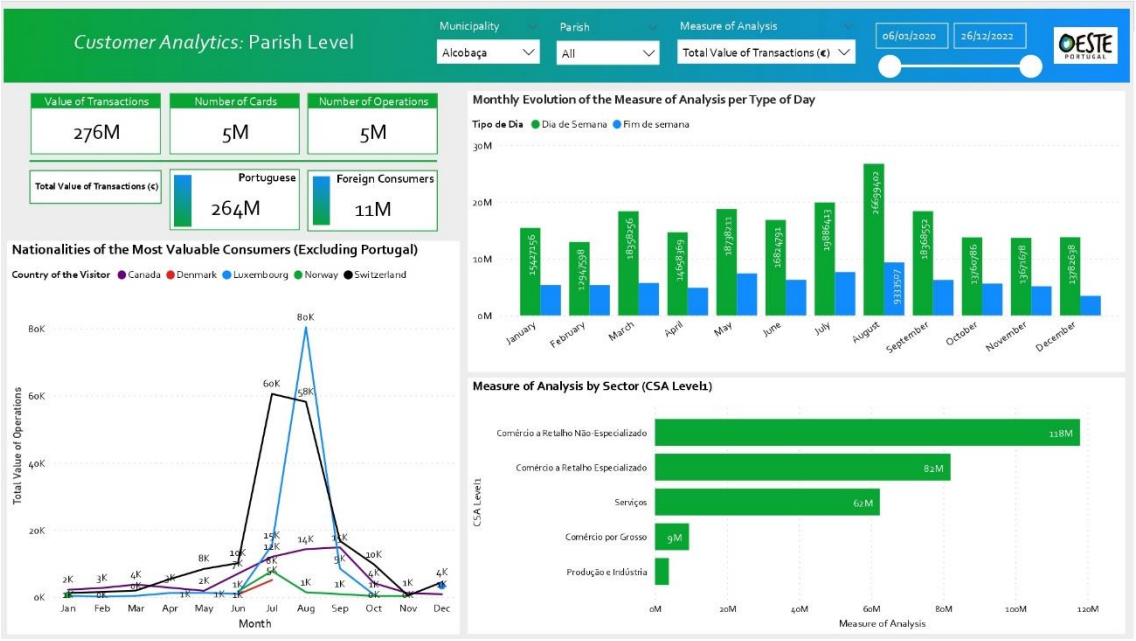

Figure 8 - Customer Analytics Overview: Parish Level Tab

Later, the study perspective specifies the parish. Comparable to the two sheets concerning the municipality level, the dashboard perspective moves to the parish level, allowing the user to select one or several. Some analytics are identical to what is found in previous tabs, with key total insights highlighted. Analysis by type of day, sector and nationality of the consumer form a representative perspective to be carried out at the most granular level of the geographic organization of the country. This set of analysis allows for a deeper and more detailed knowledge, for example, about the results of marketing strategies that have been applied, as well as the main performance indicators of the business and the local economy.

The analysis in the parish level confirms the general trend of the municipalities, in terms of monthly and annual evolution of the parameters. The main differentiating visualization in this area might be the graph that indicates the nationalities of the most valuable consumers, regarding the total amount spent in the respective parish. In general terms, the main countries that feature prominently in this analysis are central European countries where Portugal has a strong emigrant community, which manifests itself sharply in the summer months. Examples are Luxembourg and Switzerland, or France (which has a lot of expression in consumption in several municipalities such as *Caldas da Rainha*, *Bombarral*, or *Cadaval*, for instance), but there are some exceptions that should be highlighted. Thus, and only in its important parish of *União de Freguesias de Lourinhã e Atalaia*, *Lourinhã* generally attracts consumers from Canada, surpassing those from other countries such as Luxembourg. *Óbidos* also has a relevant feature, with the presence of a lot of consumption coming from Israel, ranked in the second place. In general, there are some other countries that manage to have some timid prominence, such as Mozambique, Norway, Angola, and the United States.

Analysing the clusters of parishes, it is possible to make a simplified and more intuitive analysis of the potential of each parish in their general context, driving a comparison and therefore generating opportunities for improvement of logistics chains, balancing financial support between similar zones and numerous further benefits.

5.6. Customer Analytics: Cash Withdrawals

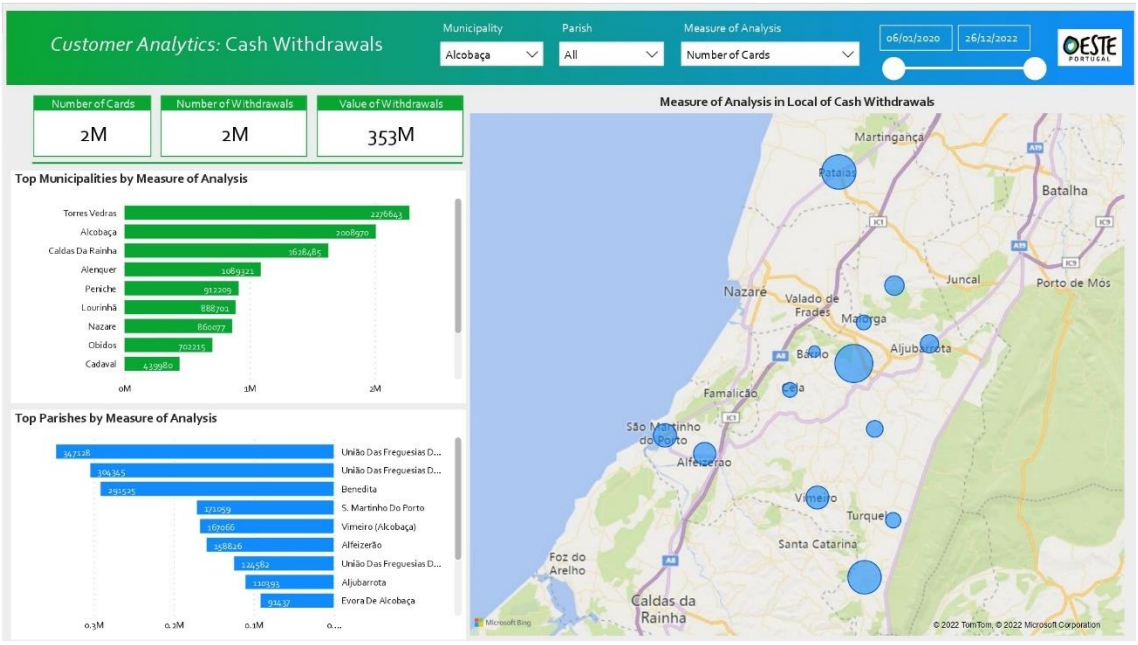

Figure 9 - Cash Withdrawals Overview Tab

In the last tab of the dashboard, the user is presented with a perspective on the dynamics of the cash surveys. It is possible to select municipality and parishes to understand the dimension of the chosen measure of analysis, but this time uniquely in terms of ATM surveys. The first bar graph indicates a top of the municipalities ordered by the chosen measure and is unalterable by the location filters, while, by selecting a town, its parishes will be ordered by the same measure and the map to the side will graphically display the parishes, where the size of the bubbles will be the dimension of the measure. Therefore, it is possible, for instance, to understand how many cards have withdrawn money, in which places citizens withdraw more money at a time and in total.

In this last analysis, the focus goes to the municipalities with the highest parameters in terms of cash surveys. For illustration, the *Average Value per Operation* could indicate that consumers are withdrawing a more money at a time because there are few ATMs and those that exist are far away, or the population is more dispersed. There should also be taken into consideration the possibility that there is still a greater predominance of small businesses location that does not have payments by terminal devices. In terms of average values collected each time, *Bombarral* (€189), *Torres Vedras* (€187) and *Caldas da Rainha* (€185) are the leaders, with the latter municipality occupying the first position if the average value collected is considered per card/consumer (235€), and *Alenquer* (221€) completes the podium that keeps the second position intact (with 226€ from *Torres Vedras*). Until October this year (2022) alone, around €709M had been withdrawn in the *Oeste*, with *Torres Vedras* being the leader in terms of value (€152M), equivalent to more than a fifth of the region's total.

These measures may be linked to factors of relevant analysis, such as the predominance of consumption in local commerce or large surfaces, assuming that there may be a greater predominance of cash payment in small or old shops and businesses, and never in large retails. Thus, parishes with higher average withdrawal amounts will be able to work on expanding ATM payment terminals in small local shops, or even increase the number of cash machines. By inference, the amounts collected per operation may tend to be higher if the distance to the nearest ATM is greater, for greater convenience and security.

6. Conclusion, Limitations and Future Work

This article presented a Business Intelligence solution applied to a region of the country, composed of the twelve municipalities in the *Oeste* region, in Portugal. This solution was designed to be as broad as possible so it can be reused throughout local government organizations, updating some changes to the data source, and revising processes and data to bring them up to date. With this solution, organizations in the region might benefit from a range of tools to support decision-making processes, in a context of high importance of information for local power organizations, and an asset for municipalities in policy making. with more direct impacts for citizens (Mix, 2016). In this way, the needs of the citizens residing in the *Oeste* may be more efficiently identified, their necessities met, and thus enabling the economic and social development of the region, through support for stakeholder agents, whether organizations or citizens. Leveraging the *Oeste* brand, the developments created may help municipalities to specialize in products in which they obtain higher returns or higher average returns, signalling products and sectors responsible for job creation in a region where there are several rural or small and medium-sized businesses, but whose geographical location enhances advantages in tourism and mobility.

The data obtained allowed the construction of a useful BI solution from which relevant information can be extracted, but for which there are limitations. As the main obstacles, there is, from the outset, a regulatory framework due to privacy rights issues which might turn harder the access to it. Afterwards, there are also limiting factors: the lack of granularity of the data in relation to the temporal dimension (only available for weekdays, or weekends, both as a whole), the number of cards that only serves as a general approximation (proxy) on the number of consumers, or the lack of knowledge about the breakdown of some of the consumer sectors represent some of the limitations.

Regarding possible developments and additions that can be made within the scope of this project and the continuity of the effort of obtaining, storing, and processing data by the *Oeste* region, limitations and challenges arise. Still within the scope of this Business Intelligence solution presented, an additional restriction of the model is the way in which the conclusions regarding the information on the measures and parameters analysed made with reference to the years 2020 and 2021 contain limitations. The information collected by SIBS Analytics for those years tends to be more imprecise and the comparison between years in some situations could not be harmonized, causing compatibility problems. In addition, the analysis for the year 2020 offers a high degree of bias in its analyses, given the occurrence of the first lockdowns within the scope of the pandemic, which may influence, for example, the *Number of Cards*, given that tend to have a smaller number of people made frequent purchases and the flows of consumers from outside the region decreased dramatically in this period. Thus, the complementarity of the data presented here with others from related areas, such as transport and population mobility, is of special relevance, to better understand displacement, the so called “Market Tracking”, the demographics of consumers, as well as other data. This provision includes information on tourism, which is relevant in numerous municipalities in the covered area, with the perspective of overnight stays, admissions to museums and galleries, as well as a clearer and more detailed view of two relevant sectors in the region, such as catering and agriculture, which are central and represent some of the best products the region has to offer.

REFERENCES

Adelakun, O. (2012). *The Role of Business Intelligence in Government: A Case Study of a Swedish Municipality Contact Center* [Master Thesis]. Hogskolan Vast.

Africk, E., & Levy, Y. (2021). An examination of historic data breach incidents: What cybersecurity big data visualization and analytics can tell us? *Online Journal of Applied Knowledge Management*, *9*(1).

Barbieri, D. (2012). *Business Intelligence and its Applications to the Public Administration*.

Burstein, F., & Holsapple, C. W. (Eds.). (2008). *Handbook on Decision Support Systems 2—Variations*. Springer-Verlag Berlin Heidelberg.

Chen, X., & Siau, K. (2020). Business Analytics/Business Intelligence and IT Infrastructure: Impact on Organizational Agility. *Journal of Organizational and End User Computing*, *32*(4), 24. https://doi.org/10.4018/JOEUC.2020100107

CIMO. (2014). *ESTRATÉGIA 2020 OESTE PORTUGAL: EIDT – Oeste*.

Data, F. (2019). Pros and Cons of Power BI – The Bright & the Dull side of visualization suite. *DataFlair*.

Dragicevic, M., & Krneta, D. (2018). *Business Intelligence in Power Distribution System*. 22–25. https://doi.org/DOI:10.20544/AIIT2018.P05

Elena, C. (2010). Business intelligence. *Journal of Knowledge Management, Economics, and Information Technology*.

Fangxiao, L. (2022). *A Walkability Simulator using a BI approach: A Lisbon Case Study* [Thesis]. Universidade Nova de Lisboa.

Gavrilov, G., Jakimovski, B., & Chorbev, I. (2018). *Cloud-Based Electronic Health Record for Health Data Exchange*. International Conference on Applied Internet and Information Technologiesv. https://doi.org/DOI:10.20544/AIIT2018.P03

Hartley, M. K. (2015). *An analysis of Business Intelligence for improved public service delivery* [Thesis]. Department of Information Systems, University of Cape Town.

Honcharenko, T., Kyivska, K., Mariia, L., Terentyev, O., Gorbatyuk, I., & Dolya, E. (2021). *Mathematical Modeling of Online Transaction Processing System for Design of Building Territory*. 552–556. https://doi.org/10.1109/UKRCON53503.2021.9575215

Hopken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2015). Business intelligence for cross-process knowledge extraction at tourism destinations. *Inf Technol Tourism (2015)*, *15*, 101–130. https://doi.org/DOI 10.1007/s40558-015-0023-2

Inmon, W. H. (2005). Building the Data Warehouse. In *Building the Data Warehouse* (Fourth Edition). Wiley Publishing, Inc.

Judek, M., Savic, Z., & Radovic, D. (2017). *Modeling a business intelligence system for school management on local government level*. https://doi.org/10.17559/TV-20160205121415

Kalaiarasan, T. R., Anandkumar, V., Nivetha, R., & Mathumitha, B. (2020). DATA ANALYTICS FOR RETAIL INDUSTRY USING QLIK. *Journal of Xi'an University of Architecture & Technology*, 1276–1279.

Kavaratzis, M. (2009). Cities and their brands: Lessons from corporate branding. *Place Branding and Public Diplomacy*, *5*, 26–37.

Kazi, Z., Kazi, Lj., Radosav, D., & Fenjcev, S. L. (2018). *Web Application for Art Gallery Management Support*. International Conference on Applied Internet and Information Technologies. https://doi.org/DOI:10.20544/AIIT2018.P01

Keim, D. A., Hao, M. C., Dayal, U., & Lyons, M. (2007). Value-Cell Bar Charts for Visualizing Large Transaction Data Sets. *IEEE Transactions on Visualization and Computer Graphics*, *13*(4), 822–833.

Khan, S., Redha Qader, M., Thirunavukkarasu, K., & Abimannan, S. (2020). Analysis of Business Intelligence Impact on Organizational Performance. *Way Towards a Sustainable Economy (ICDABI)*. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI). https://doi.org/10.1109/ICDABI51230.2020.9325610

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit* (Third Edition). Wiley.

Kovacic, lko, Schuetz, C. G., Neumayr, B., & Schrefl, M. (2021). OLAP Patterns: A pattern-based approach to multidimensional data analysis. *Data & Knowledge Engineering*.

Lavalle, A., Mate, A., Trujillo, J., & Rizzi, S. (2019). *Visualization Requirements for Business Intelligence Analytics: A Goal-Based, Iterative Framework*. 2019 IEEE 27th International Requirements Engineering Conference (RE). https://doi.org/10.1109/RE.2019.00022

Lock, M., & Gladstone, M. (2010). *Business Intelligence and its Applications to the Public Administration*.

Mix, G. (2016). *Best Practices for Local Government Business Intelligence Implementations* [Thesis]. University of Oregon.

Mota, A. R. de C. (2021). *Dashboards in smart city's sustainability performance measurement through business intelligence* [Thesis]. Universidade Nova de Lisboa.

Nanda, A., Gupta, S., & Vijrania, M. (2019). A Comprehensive Survey of OLAP: Recent Trends. *Proceedings of the Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019]*. Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019]. https://doi.org/10.1109/ICECA.2019.8822203

Olszak, C. M. (2014). Business Intelligence and Analytics in Organizations. *Advanced in ICT for Business, Industry and Public Sector*, 89–109. https://doi.org/10.1007/978-3-319-11328-9_6

Piechorowski, C. (2018). *The Extent of Autonomous Data Analysis for Non-IT Staff with Self-Service Business Intelligence Tools* [Master Thesis]. UNIVERSITY OF LJUBLJANA.

Popescu, G. H., Lazaroiu, G., Kovacova, M., Valaskova, K., & Majerova, J. (2020). URBAN SUSTAINABILITY ANALYTICS: HARNESSING BIG DATA FOR SMART CITY PLANNING AND DESIGN. *Theoretical and Empirical Researches in Urban Management*, *15*(2), 39–48.

PORDATA. (2021). Resultados dos Censos de 2021. *PORDATA*.

PORDATA. (2022). *O Que São NUTS?*

Purnamasari, A. M., Pah, C. E. A., Yoga, M. D. I., Girsang, A. S., & Isa, S. M. (2019). Business Intelligent in an E-Commerce Industry. *IOP Conference Series: Materials Science and Engineering*, *598 012085*. https://doi.org/10.1088/1757-899X/598/1/012085

Richardson, J., Sallam, R., Schlegel, K., Kronz, A., & Sun, J. (2020). *Magic Quadrant for Analytics and Business Intelligence Platforms*. Gartner Reprint.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What Do We Talk About When We Talk About Dashboards? *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 682–692. https://doi.org/10.1109/TVCG.2018.2864903

Sean, K., Andreas, P., Hellerstein, J. M., & Heer, J. (2012). Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2917–2926. https://doi.org/10.1109/TVCG.2012.219

Sengupta, A., & Cao, L. (2019). Value Creation of Business Intelligence and Analytics in Retailing. *Proceedings of the European Marketing Academy, 48th*. Annual EMAC Conference.

Shahini, R. (2020). *Business Intelligence in the Hotel Industry* [master's degree]. Linnaeus Univerity.

Teixeira, R., Afonso, F., Oliveira, B., & Santos, M. F. (2014). *Business Inteligence to improve the quality of Local Government Services—Case-Study in a Local Government Town Hall*. https://doi.org/10.5220/0005051601530160

Vanegas, C. A., Aliaga, D. G., Benes, B., & Waddell, P. (2009). Visualization of Simulated Urban Spaces: Inferring Parameterized Generation of Streets, Parcels, and Aerial Imagery. *IEEE Transactions on Visualization and Computer Graphics*, *15*(3), 424–435. https://doi.org/10.1109/TVCG.2008.193

Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *20*(1), 383–398. https://doi.org/10.1109/TITS.2018.2815678