NOVA IMS Information Management School

MGI

Master Degree Program in
Information Management

Understanding the External Factors Affecting Beer Sales: A Business Intelligence Approach

Beatriz Amorim Serrador

Project Work

presented as partial requirement for obtaining the Master Degree Program in Information Management

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

UNDERSTANDING THE EXTERNAL FACTORS AFFECTING BEER SALES: A BUSINESS INTELLIGENCE APPROACH

By

Beatriz Amorim Serrador

Master Thesis / Project Work presented as partial requirement for obtaining the Master’s degree in Information Management, with a specialization in Knowledge Management and Business Intelligence.

Supervisor: Gonçalo Baptista

November 2023

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.

Beatriz Serrador

Lisbon, 21/11/2023

ACKNOWLEDGEMENTS

I would like to thank all the wonderful women in my life. First and foremost, I want to give my deepest appreciation to my incredible girlfriend. Her endless encouragement, patience, and faith in me have been my guiding light. To my mom, her sacrifices and unwavering support have made this endeavor possible. I truly couldn’t be here without her. I am also blessed to have a circle of friends who have stood by me through thick and thin. Your words of encouragement have been priceless. Your friendship has made this journey memorable. And, of course, a special mention goes to my feline companions. Their comforting purrs, occasional distractions, and adorable adventures have provided moments of comfort and joy during this journey.

This work stands as a testament to the love and support that surround me, and I am profoundly thankful for your presence in my life.

ABSTRACT

New and modern data and analytical techniques allow organizations to leverage valuable insights to make informed decisions. Within this thesis, a new artifact was developed, supported in an innovative and comprehensive framework that employs a business intelligence approach to try to understand the dynamics of the Portuguese beer sales market. Several sales influencing factors were identified, each categorized within one of five pillars: microeconomic, macroeconomic, football, tourism, and climate. A list of relevant data sources was also used, including Nielsen, INE, Liga Portugal, and IPMA. The artifact includes an analytical model composed of data preparation, Feature Selection, and Ordinary Least Squares (OLS) regression phases, determining statistical significance among the indicators. Up until this point, according to the research carried out, earlier studies studying the factors impacting beer sales in Portugal are very limited. This emphasizes the uniqueness and significance of this research, which fills a gap in the existing literature and adds to a better understanding of the Portuguese beer industry. Besides the theoretical implications, the outcomes explain the impact of external factors on beer sales, unveiling insights that empower businesses to better prepare for market fluctuations, anticipate trends, and make informed decisions.

KEYWORDS

Business Intelligence, External Factors, Data-Driven Analysis, Beer Sales Volume, Regression Analysis

Sustainable Development Goals (SGD):

8 DECENT WORK AND ECONOMIC GROWTH

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

OLS	Ordinary Least Squares
DSR	Design Science Research
ETL	Extract, Transform, and Load
DW	Data Warehouse
OLAP	Online Analytic Processing
BI	Business Intelligence
BIA	Business Intelligence Analytics
CRISP-DM	Cross-Industry Standard Process for Data Mining
EDA	Exploratory Data Analytics
INE	Instituto Nacional de Estatística
CPI	Consumer Price Index
APT	Automatic Payment Terminals
US	United States
KPI	Key Performance Indicators
SEF	Serviço de Estrangeiros e Fronteiras

1. INTRODUCTION

1.1. Background and Problem Identification

In recent years, the field of Information Management, coupled with the growing importance of Business Intelligence, has witnessed significant advancements in understanding complex business phenomena through data analysis (Ragazou et al., 2023). With the increase in data sources and the arrival of sophisticated analytical techniques, organizations can now leverage valuable insights to make informed decisions, optimize strategies, and achieve competitive advantages (Bordeleau et al., 2018). In this context, this master's thesis aims to explore the influence of various external factors on beer sales in Portugal using a data-driven approach.

The principal spotlight of this thesis revolves around creating a dynamic, innovative, and integrated artifact to gather different data sources and to understand what the main drivers of the Portuguese beer market are. Specifically, to address the research question: "Are beer sales in Portugal affected by external factors?". These factors encompass a diverse set of hypotheses, including macroeconomic, microeconomic, tourism, climate, and sports indicators.

The external factors under consideration carry substantial significance, as they can play a vital role in modeling consumption patterns and economic behavior. Macroeconomic indicators, such as average monthly household income and consumer price index, are expected to shed light on how economic conditions impact beer sales in the country. As Valášková & Klieštik (2015) pointed out, consumption habits are heavily affected by the economic cycle. Throughout growth periods, consumers are motivated to spend more and, on the other hand, during economic depressions, they are reluctant to make purchases. Similarly, tourism indicators, including the number of guests and the number of nights they spend in the country, may offer insights into the impact of tourism on beer consumption. Furthermore, the influence of average monthly temperature on beer sales will be investigated, following previous studies (Colen & Swinnen, 2010). Understanding the complex relationships between these external factors and beer sales is crucial for businesses operating in the beer industry, as it can facilitate data-driven decision-making and marketing strategies tailored to the prevailing economic and environmental conditions.

To address the research question effectively, a methodological approach will be adopted that involves the collection, integration, and analysis of various data sources. Python was employed as the primary programming language to create a comprehensive artifact using a data frame that consolidates information from diverse sources, thereby enabling a unified analysis of the data. Data preparation will be a crucial step to ensure the reliability and accuracy of the findings. Feature Selection techniques will be applied to identify the most influential external factors affecting beer sales, streamlining the analytical process, and enhancing the interpretability of results. To model the relationships between beer sales and external factors, several regression techniques were implemented in the artifact, including Ordinary Least Squares (OLS), Lasso Regression, and Elastic Net Regression. By employing these regression models, underlying statistical patterns were identified and the impact of individual external factors on beer sales was quantified.

It is worth mentioning that earlier studies particularly studying the factors impacting beer sales in Portugal are very limited, according to the earlier literature assessment. This emphasizes the uniqueness and significance of this research, which fills a gap in the existing literature and adds to a better understanding of the Portuguese beer industry. The primary contribution of this master's thesis lies in the creation of an intelligent and innovative artifact capable of integrating, standardizing, and analyzing statistical relationships between beer sales and a multitude of external factors. This artifact will serve as a powerful tool for businesses and policymakers alike, enabling them to adapt their strategies based on real-time data and dynamic market conditions. The flexibility and on-demand updating capabilities of the developed solution will enhance its practical applicability across different organizations and industries. In summary, this master's thesis seeks to unveil critical insights into the relationship between beer sales in Portugal and external factors. By employing advanced data analytics and regression techniques, this study endeavors to contribute to the existing knowledge base, empower decision-makers, and advance the field of Business Intelligence.

1.2. Study Objectives

The first main objective of this master's thesis is to construct a dynamic artifact suited to enhance traditional analysis and to merge different non-normalized data sources under a single umbrella and contribute to the field of Information Management and Business Intelligence. This framework can integrate, standardize, and explore complex statistical relationships between variables. Not only that, but this dynamic tool can also offer the power of on-demand monitoring and updating of the discovered factors that affect sales. The second main objective is to use this artifact to uncover the underlying drivers that shape the Portuguese beer sales market.

In more detail:

1. Designing the framework:

Design and construct a robust framework – an artifact capable of gathering different non-normalized data sources. This framework stands as the embodiment of the integration of diverse external factors to analyze statistical relationships.

2. Creating an active data system:

By using advanced tools and methods strategically, a flexible structure should be built allowing the integration, standardization, and harmonization of different types of data together. This is vital for ensuring the artifact's durability, enabling the incorporation of new insights, and facilitating updates.

3. Exploring feature relationships:

The next step is to analyze the gathered data. Using the framework's abilities, the relationship between Portuguese beer sales and external factors is investigated by employing different models, namely Ordinary Least Squares, Lasso Regression, and Elastic Net Regression.

4. Giving the Industry the power of knowledge:

The last objective leads the research from analysis to foresight. Using the conclusions for the goal above, it will be possible to predict potential trends in the future of the Portuguese beer industry. By looking at the past it's possible to equip companies and breweries with the knowledge to make well-informed decisions, as well as anticipate market shifts and adapt to evolving dynamics.

1.3. METHODOLOGY

The methodology chosen for this research is Design Science Research (DSR), as introduced by Peffers et al. (2007). DSR involves the creation and evaluation of IT artifacts to solve identified organizational problems, aligning well with the goals of this study. The methodology consists of a systematic process to address the research problem, design a solution, and communicate the outcomes effectively to relevant stakeholders.

The DSR process, depicted in Figure 1, guides the development of the artifact, ensuring a structured and comprehensive approach to problem-solving. This approach is well-suited to the research question of understanding the impact of external factors on beer sales in Portugal.

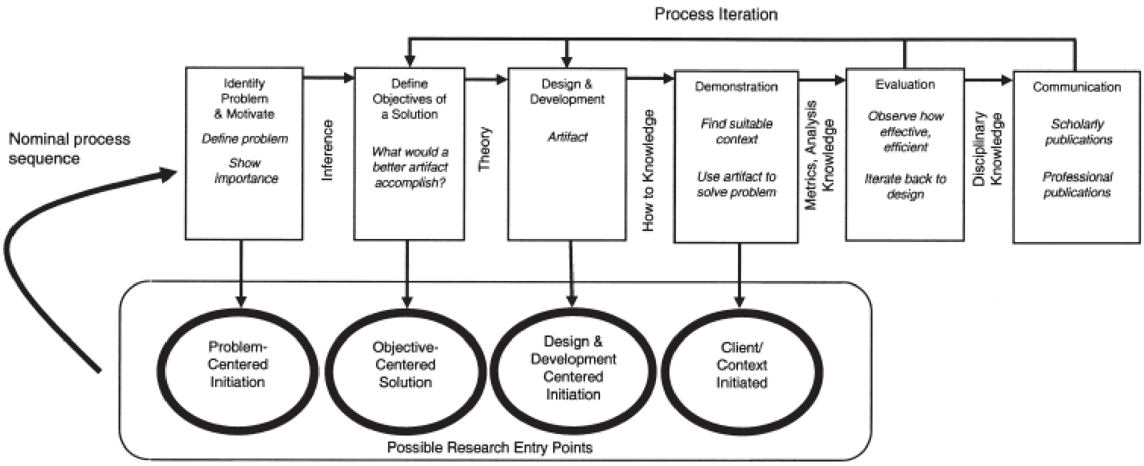

Figure 1 - DSR Process Model

Source: Peffers et al. (2007)

The DSR Process Model encompasses six key activities that will be meticulously applied in this research:

1. Problem identification and motivation:

The initial activity involves identifying the research problem of determining the external drivers of beer sales in Portugal and establishing the value that a data-driven solution can bring to the beer industry. A thorough review of existing literature and related studies provides insights into the state-of-the-art and helps frame the research objectives.

2. Definition of objectives:

Once the research problem is defined, the next step is to articulate clear and attainable objectives. This involves outlining the methods and tools necessary for the implementation of the artifact, ensuring a systematic and well-structured approach to achieving the research goals.

3. Design and development:

The heart of the DSR process is the design and development of the artifact itself. This phase involves gathering a diverse array of data, including sociodemographic indicators, beer sales-related information, and trends. Through this process, a sophisticated model was constructed capable of uncovering correlations between different variables, illuminating the crucial drivers shaping the Portuguese sales beer market.

4. Demonstration:

With the artifact created, the demonstration phase commences. The predefined criteria and analysis methods are applied to the dataset. This stage allows for the practical testing and validation of the artifact's effectiveness in capturing and analyzing the complex relationships between beer sales and external factors.

5. Evaluation:

The evaluation phase aims to measure the artifact's effectiveness and utility. By assessing the extent to which the defined objectives have been met, this phase ensures that the model aligns with the research goals and effectively contributes to the understanding of beer sales dynamics.

6. Communication:

The culmination of the DSR process is the communication of results and outcomes. Effective communication is essential to ensure that stakeholders, including those directly impacted by the initial problem and those who will work with the model, comprehend its functionality and potential contributions.

2. LITERARY REVIEW

2.1. BUSINESS INTELLIGENCE

By analyzing business intelligence-related academic and practitioner literature, Mashingaidze & Backhouse (2017) came up with the following definition: “BI is a set of integrated strategies, applications, technologies, architectures, processes and methodologies used to collect, store, retrieve and analyze data to support decision-making”. The expression has been around for over 20 years and more than the definition of BI, it is important to understand how its process works, as presented in Figure 2. Many of the concepts mentioned in the figure are often confused with the meaning of BI itself. The procedure starts with data, whether structured or unstructured, recovered from a database (Extract), making all the necessary changes until it’s in the desired format (Transform), and, finally, inserting it into the desired data warehouse (Load), which corresponds to the ETL process. A data warehouse (DW) is a “subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management’s decision-making process” (Inmon, 2005). When a company wishes to focus on a single business subject, like sales, finance, or logistics, it can resort to a data mart. Data marts are smaller than DW and, consequently, a more effective, faster, and cheaper solution. Enterprises deal with a vast quantity of data and business facts, so it’s important to perform multidimensional analysis and combinations, that’s where online analytics processing (OLAP) comes in. The final step in the BI process is to oversee the output information with visualization and dashboards, facilitating comprehension (Foley & Guillemette, 2010).

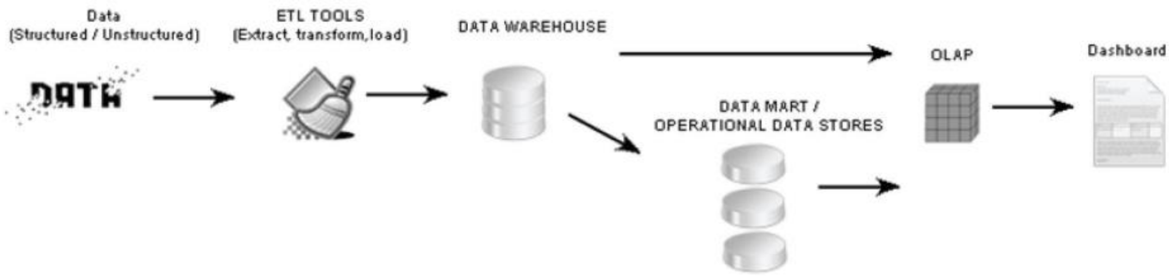

Figure 2 - Example of a BI technical architecture

Source: Foley & Guillemette (2010)

Business Intelligence (BI) has emerged as an essential component for making strategic decisions in organizations and governments worldwide. According to Tavera Romero et al., (2021), its significance lies in enabling businesses to thrive, establish robust partnerships, enhance counterintelligence measures, and achieve both short-term and long-term goals and objectives. Extensive research confirms the multitude of advantages derived from implementing BI, encompassing enhanced performance, efficiency, productivity, business expansion, effective resource planning, stronger supplier-buyer relationships, and cost reductions. These benefits, in turn, pave the way for gaining a competitive edge in the market.

Lim et al. (2013) argued that, even though business intelligence is based on the previously described process, the concept heavily relies on many data collection and analysis technologies, which can be referred to as business analytics. Therefore, the authors believe the two notions should be combined into business intelligence analytics (BIA).

2.1.1. Data Mining

In *Business Intelligence, Analytics, and Data Science: A Managerial Perspective,* Sharda et al. (2017), describe three levels of analytics: descriptive, predictive, and prescriptive. The first level, descriptive analytics, consists of understanding and finding patterns in the current state of a company. To do this, it is necessary to join different data sources to ensure significant outcomes by creating reports, queries, or dashboards, typically developing a data structure as part of a data warehouse. Predictive analytics, unlike descriptive analytics, aims to forecast the future using statistical methods. The final level – prescriptive analytics – is a combination of the previous analytics, trying to understand the present, as well as the future, to optimize decision-making.

Data mining is the process of obtaining patterns and trends from a large amount of data by using different approaches, such as clustering, classification, association, and regression (Gera & Goel, 2015). Since not all companies work the same or deal with the same data, it was necessary to create a process that was an industry-neutral standard – The Cross-Industry Standard Process for Data Mining (CRISP-DM) (Larose, 2015). This process argued that a data mining solution consists of six phases:

1. *Business Understanding*: define objectives according to the current business needs and convert them into a single data mining problem. Planning the necessary steps to reach the defined objectives is a crucial part of this phase.
2. *Data Understanding*: gather data and perform exploratory data analysis (EDA) to evaluate its quality and come up with preliminary insights that can lead the process in new directions.
3. *Data Preparation*: this is the most important and time-consuming part of the process. It involves cleaning the raw data, carrying out the necessary changes to variables, and selecting the most appropriate features.
4. *Modeling*: Like it’s mentioned above, there are several data mining tasks, and each of them offers a variety of different approaches or models. Having this in mind, this phase seeks to select the most appropriate methods to optimize the results.

5. *Evaluation*: the previous phase can output more than one solution, it's necessary to evaluate them for quality and effectiveness regarding the defined business problem.
6. *Deployment*: The final phase involves transmitting the results of the created model, whether through the conception of a report or an adjacent model.

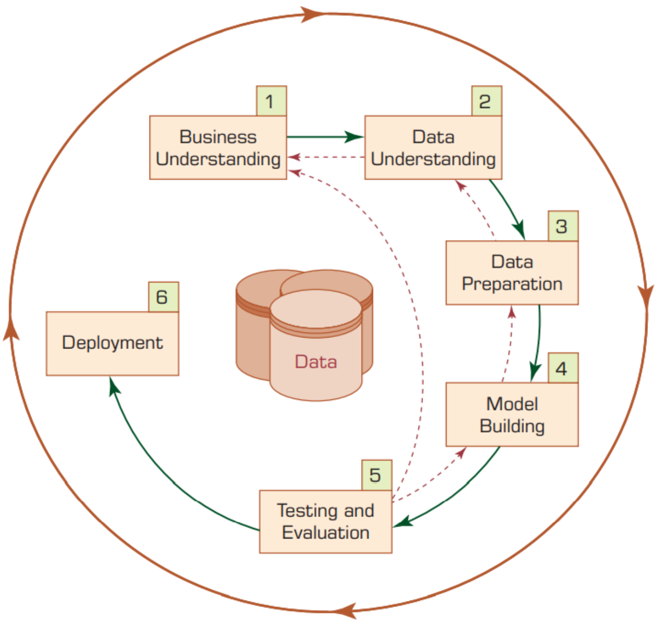

Figure 3 - The Six-Step CRISP-DM Data Mining Process

Source: Sharda et al. (2017)

2.1.2. Regression Analysis

In an overview of regression analysis and multivariate statistics, Duleba & Olive (1996), described Regression Analysis as a data mining technique that examines relationships between one or more variables Y (or dependent/ response) and one or more variables X (or independent/ predictor), where the intention is to forecast the dependent variable using the predictor features. Due to the easy interpretation and mathematical manipulation, the authors also pointed out that regression analysis can be used to understand the type and depth of the relationship among the variables.

Generally, there are three types of regression analysis:

1. Simple Regression Analysis: Describes the linear relationship between one dependent variable and one independent variable.
2. Multiple Regression Analysis: A natural extension of simple linear regression when more than one independent variable is used to define the dependent variable.

3. Multivariate Multiple Regression Analysis: When we are dealing with several independent variables, as well as several dependent variables.

2.1.3. Feature Selection

Feature Selection is one of the most popular and significant data preparation approaches, and it has grown to be an essential step in the machine-learning process (Kalousis et al., 2007). While reviewing the literature on the matter, Kumar (2014) stated that Feature Selection involves finding important features and eliminating unnecessary, redundant, or noisy data. This procedure accelerates data mining algorithms, enhances predicted accuracy, and makes data more understandable. Features that don't offer any helpful information are considered irrelevant, and redundant features don't offer any additional information above the features that are currently being used.

In general, there are three types of variable selection algorithms: filter, wrapper, and embedding techniques (Stijven et al., 2011). A filter approach will evaluate variable importance based on model assessment, but a wrapper method will solely use the properties of the data. The embedded technique incorporates Feature Selection into the model-building process such that the modeling process is guided by the relevance of the variables.

2.2. CONSUMER BEHAVIOR

In "In Times of Trouble", Campbell et al. (2020) mention that economic distress presents a threat to consumers' well-being and quality of life. Threats can be cataloged as economic, health, social, informational, and environmental. The first affects the consumer's economic condition, such as a recession or unemployment. As the name indicates, health threats can damage the consumer's health like getting sick or going through a pandemic. Social threats are hazards at a societal level, such as terrorist attacks, or something that pressures our role in society. The next category of threats is informational, which can affect the ability to learn and know – contradictory informational leaks or a newspaper closing, for example. Finally, environmental threats include risks to the environment that affect the consumer's safety. Threats are also prejudiced by severity (potential damage), scope (number of people affected), and psychological association (how close it is to the person comprehending the threat) (Campbell et al., 2020).

According to Valášková & Klieštik (2015), consumer behavior is "a decision-making process of searching, purchasing, using, evaluating, and disposing of products and services". These authors concluded that there were dependencies between changes in consumer behavior and certain demographic characteristics during the 2009 economic crisis. They found that changes in consumer behavior rested on disposable income and the length of the economic depression, however, it didn’t depend on the age of the population. These fluctuations happened quickly, consumers became more vigilant, more price-sensitive, and thoughtful, and bought private labels more often – a new consumer shopping behavior emerged.

2.3. Beer Consumption

The article "Beer Drinking Nations - The Determinants of Global Beer Consumption" by Liesbeth Colen and Johan Swinnen examined the determinants of beer consumption across countries and over time using panel data from 116 countries over the period 1960-2007. The study identifies several key factors influencing beer consumption, including income, temperature, religious beliefs, and culture. The article highlights the changing dynamics of beer consumption globally, with shifts in its share of total alcohol consumption between traditional beer-drinking countries and emerging economies. As beer consumption patterns continue to evolve, understanding these determinants is crucial for informed decision-making and policy formulation (Colen & Swinnen, 2010).

Kozak (2013) tried to explain the declining trend in beer consumption among Czech citizens, with each person averaging 134 liters in 2013, down from 160 liters in 1995. According to the Czech Beer and Malt Association's statistics, the drop could be due to external and inner factors. External factors contributing to this decline include an elevated excise tax on beer, the end of obligatory military service, alcohol checks for workers, a decrease in tourists, and changing beer-drinking patterns among the younger generation. Inner factors involve brewery pricing policies, profit maximization efforts, marketing focused on regular male pub-goers, individual beer imports, and the standardization of beer quality ("Euro-beer" production).

Betancur et al. (2020) discuss the complexity of factors influencing beer choice and consumption. It highlights that there are no single, clear-cut variables that directly predict beer choice; instead, it is the result of various factors interacting with each other. Product attributes and consumer variables alone do not independently affect beer choice, but rather they interact through psychological, socio-cultural, and biological mechanisms. The study also suggests the need for future studies that examine the relative contributions of these factors, considering data from consumers, products, and context. It emphasizes the importance of context, as beer choices can differ in shopping and consumer settings, depending on the environment.

Regarding alcohol consumption in the Portuguese population, Marques-Vidal & Dias (2005) studied trends and determinants of changes observed between 1995/1996 and 1998/1999. The paper suggested a changing pattern of alcohol consumption in Portugal, with a decreasing prevalence of drinkers and a shift among younger generations from wine to beer and spirits. Education level appears to be a significant factor influencing the choice of alcoholic beverage.

More recently, Silva et al. (2017) conducted a study to understand the characteristics of the average beer consumer in Portugal, who has a low monthly salary and primary education and is probably under 34 years old. The authors also concluded that beer is a relatively cheap alcohol segment, which might explain why it's attractive to younger consumers with less purchasing power. Moreover, they connected beer consumption with social gatherings such as parties, sports events, and music festivals, as well as pub culture and friendships, being a "flexible beverage", associated with meals and food. In the same study, according to 2011 data, 69% of beer consumption was out-of-home (On-Trade channel), including cafés, bars, and pubs, and 31% was in-home consumption (Off-Trade channel) through supermarkets.

3. METHODOLOGY

3.1. Conceptual Model

This framework begins by identifying and defining the key areas of study, taking into consideration the type of industry we are applying it to. Next, it is crucial to gather the necessary data sources to produce a rich repository, divide the different indicators into categories for better interpretation, and define the hypothesis to be tested to provide statistical significance.

The analytical model, crafted in Python, is the main component of this framework. It navigates through data preparation, feature selection, and Ordinary Least Squares (OLS) regression, revealing the statistical significance of each indicator. The outputted results from this model will be interpreted taking into consideration the research question, transforming data into knowledge and insights. This framework will equip businesses with the knowledge to adapt, anticipate market trends, and make informed decisions.

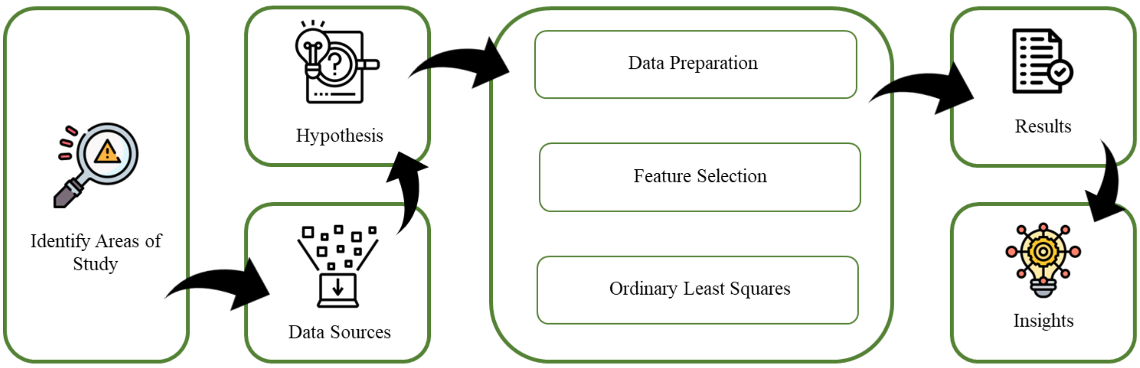

Figure 4 - Artifact Framework

Source: Own Elaboration

3.2. Data Collection

After gathering a thorough literature review, several data sources were searched to collect relevant information. One of those sources was INE (*Instituto Nacional de Estatística),* where it was possible to find relevant socioeconomic indicators such as the unemployment rate, consumer pricing index, degree of household savings, etc. Besides socioeconomic indicators, INE also provided a comprehensive view of the Portuguese tourism sector through the number of nights and guests in tourist accommodations.

The connection between beer consumption and sports, especially Football, is grounded in Portuguese culture. Therefore, information regarding the Portuguese 1st Football League, provided by Liga Portugal, was also incorporated, namely the number of games played.

In addition to public sources of information, data from Nielsen, a company specializing in collecting information on the sale of consumer goods, was also accessed. This data provided valuable insights into beer sales, including sales volume, and average price per liter.

After gathering the relevant data sources, the next step was to filter out the information based on the longest period window common to all sources, which was from 2018 to 2023, and to identify the optimal frequency, which was monthly. These were the chosen indicators:

Nº	*Data Source*	*Name*	*Beginning*	*Ending*	*Frequency*
1	INE	Perspective of the Country's Economic Situation in the Next 12 Months	1997	2023	Monthly
2	INE	Consumer Confidence Index	1997	2023	Monthly
3	INE	Degree of Household Savings	1997	2023	Monthly
4	INE	Unemployment Rate (Nº)	1998	2023	Monthly
5	INE	Activity Rate (%)	1998	2023	Monthly
6	INE	Average Monthly Gross Remuneration per Employee (€)	2014	2023	Monthly
7	INE	Consumer Price Index (%)	1948	2023	Monthly
8	INE	Purchases Through Automatic Payment Terminals using Foreign Card (€)	2010	2023	Monthly
9	INE	Nights Spent (No.) in Tourist Accommodation Establishments by National Guests	2017	2023	Monthly
10	INE	Nights Spent (No.) in Tourist Accommodation Establishments by Foreign Guests	2017	2023	Monthly
11	INE	National Guests (No.) at Tourist Accommodation Establishments	2017	2023	Monthly
12	INE	Foreign Guests (No.) at Tourist Accommodation Establishments	2017	2023	Monthly
13	Liga Portugal	Matches in 1st Football League (No.)	2009	2023	Monthly
14	IPMA	Average Air Temperature	2018	2023	Monthly
15	Nielsen	Beer Sales in Volume (Liter)	2011	2023	Monthly
16	Nielsen	Average Price per Liter (€/L)	2011	2013	Monthly

Table 1 - List of Variables

3.2.1. Variable Description

The **Degree of Household Savings**, in percentage, comes from INE, namely the Monthly Qualitative Consumer Survey. This variable is calculated through the difference between the percentage of positive answers ('increased', 'improved a lot', 'higher than normal', 'good', 'yes, absolutely sure', etc.) and the negative ones ('decreased', 'got a little worse', 'very unfavorable', 'probably not', etc.), when asked about the ability to save monthly income. Considering that beer is not an essential good, the degree to which households save can affect their sales – if households save less, they are likely to decrease their consumption. The Perspective on the **Economic Situation of the Country** in the next 12 months, in percentage, is also taken from INE's Qualitative Consumer Survey and calculated in the same way.

To encompass the indicators described above, the **Consumer Confidence Index** was also chosen, in percentage, which corresponds to the arithmetic mean of the balance of respondents to the following questions: "How has the financial situation of your household changed in the last 12 months? ", "How do you expect the financial situation of your household to evolve in the next 12 months?", "How do you think the general economic situation of the country will evolve in the next 12 months?", "Compared to the last 12 months, do you expect to spend more or less money on major purchases (such as furniture, appliances, computers, or other durable goods) in the next 12 months?". As this variable presents similarities to the different perspectives, they likely present a high correlation and, if this is the case, only one will be submitted to the model to avoid multicollinearity (Daoud, 2017).

Next, the **Consumer Price Index**, in percentage, represents the amount paid by households to purchase individual goods and services based on monetary transactions relative to the same period last year, and is a year-on-year rate. This amount corresponds to the amount the buyer pays at the time of purchase and includes all indirect taxes net of subsidies on products, rebates, and discounts provided they are generally applicable to consumers and exclude interest and other costs associated with buying on credit. In the case of this indicator, the Total CPI was chosen, because of the effect it has on consumers' wallets.

Still related to monetary issues that may affect the consumption of beer, the **Average Monthly Gross Remuneration per Employee**, in euros, refers to the gross amount paid to workers who are included in the concept of 'personnel at work', for the hours worked or for work performed during normal and overtime periods. It also includes payment for hours paid but not worked (holidays, holidays, and other paid absences) and allowances that are regular, such as allowances for food, function, accommodation, transport, etc. It is calculated by dividing the total gross remuneration paid by employers by the number of employees.

More related to the country's demographics, the **Activity Rate** is calculated by dividing the active population (aged 16 to 74) and the resident population of the same age. The active population is those of working age who are employed or unemployed, not including the inactive population, which are students, retirees, or those not looking for a job. The **Unemployment Rate** represents the percentage of the unemployed active population divided by the total active population.

Regarding tourism, we have both the number of **Guests** and the number of **Nights** spent in tourist accommodation establishments. A guest is an individual who spends at least one night in a tourist accommodation establishment, while a night is the stay of an individual in an establishment that provides accommodation. To understand the impact of both National and Foreign guests' behavior on beer consumption, this differentiation was added, resulting in four distinct variables – two for national guests and two for foreign guests.

Purchases through Automatic Payment Terminals (APT), in euros, are also a relevant indicator, encompassing the terminals existing in commercial establishments (points of sale) that allow the use of bank cards to make payments. Here, we have a variable for purchases with foreign cards for understanding tourists' purchasing behavior.

The number of **Matches in the 1st Football League**, provided by Liga Portugal, includes all games played per match week. The number of match weeks per month may vary according to the calendar stipulated by the competent authority, therefore it was necessary to sum the matches in each month, regardless of the match week.

Provided by *Instituto Português do Mar e da Atmosfera*, we have the **Average Monthly Air Temperature**. It has long been theorized that the heat influences people to seek a cool drink, often beer, for example at barbecues or parties. Thus, it was pertinent to have this variable present in the model to verify if it influences beer sales.

Finally, provided by Nielsen, the **Sales in Volume** is the total sales for the product expressed in volume, in this case, liters. The **Average Price per Liter** is the average price per liter, over time and is used to examine pricing trends for a category, segment, or trend. It's calculated by dividing the sales in value by the sales in volume.

3.3. Pillar Categorization

To obtain a more comprehensive analysis and simplify the hypothesis creation, this set of indicators was divided into distinct pillars. By organizing the indicators according to their natural characteristics it's easier to understand the multifaced external factors affecting beer sales in Portugal, but, at the same time, promotes the process of identifying and interpreting the relationship between the variables. Hopefully, this framework enhances the efficiency and accuracy of the results and contributes to more robust insights into the dynamics of the Portuguese beer market.

- **Macroeconomic** Pillar:
  - Degree of Household Savings
  - Perspective of the Country's Economic Situation in the Next 12 Months
  - Consumer Price Index (%)
  - Consumer Confidence Index
  - Activity Rate (%)
  - Unemployment Rate (Nº)
  - Average Price per Liter (€/L)

These variables focus on the macroeconomic conditions and sentiments within the country, including indicators related to savings, inflation expectations, economic outlook, consumer confidence, and labor market conditions. Identifying the impact of these indicators can provide insights into the economic health and effect on beer sales.

- **Microeconomic** Pillar:
  - Average Monthly Gross Remuneration per Employee (€)

Understanding the consumers' disposable income is pertinent since it can influence their purchasing behavior and intentions, including beer consumption. Therefore, this pillar represents microeconomic indicators linked to consumer spending power and income.

- **Tourism** Pillar:
  - Nights Spent (No.) in Tourist Accommodation Establishments (Foreign)
  - Nights Spent (No.) in Tourist Accommodation Establishments (National)
  - Guests (No.) at Tourist Accommodation Establishments (Foreign)
  - Guests (No.) at Tourist Accommodation Establishments (National)
  - Purchases Through Automatic Payment Terminals using Foreign Cards (€)

The variables present in this pillar are related to tourism activity by encompassing the number of nights spent and guests in touristic accommodations, as well as the amount of money spent in automatic payment terminals using foreign cards. Tourists often increase the demand for beverages in the country, including beer.

- **Climate** Pillar:
    - Average Air Temperature

The average air temperature variable represents the climate conditions, which can influence beverage consumption since higher temperatures could lead to an increase in beer consumption, especially during warm months and tourist seasons.

- **Football** Pillar:
    - Matches in 1st Football League (No.)

The Football Pillar centers on the Matches in 1st Football League (No.) indicator, acknowledging the potential impact of sports events, particularly football matches, on consumer behavior. By examining the influence of the number of football matches on beer sales, this pillar recognizes the significance of cultural and entertainment factors in shaping consumer choices.

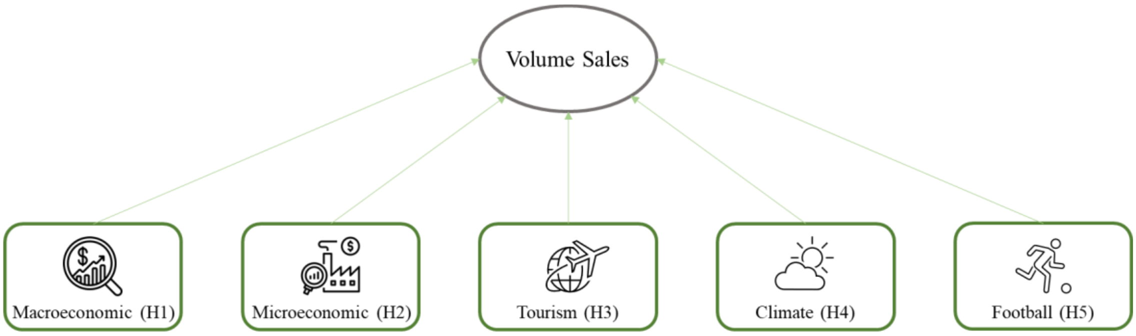

Figure 5 - Pillar Categorization Diagram

Source: Own Elaboration

3.4. Hypothesis Definition

In the process of understanding how external factors influence the Portuguese sales beer market, the chosen approach combines the principles of business intelligence with the research objectives. Therefore, as stated in the objectives section, this research will resort to statistics to test different hypotheses. This process starts with the definition of a null hypothesis (H0), which is tested by resorting to a certain test statistic. If the test statistic classifies the null hypothesis as implausible, then it's rejected, and, consequently, the alternative hypothesis (Ha) is accepted – the alternative hypothesis includes all alternatives to H0 (Duleba & Olive, 1996).

Hypotheses:

- **Null Hypothesis (H0)**: The Portuguese beer industry is not affected by any of the defined external factors, spanning macroeconomic, microeconomic, tourism, climate, and sports-related categories.
- **Alternative Hypothesis 1 (H1)**: The Portuguese beer industry is influenced by at least one of the indicators within the Macroeconomic Pillar.
- **Alternative Hypothesis 2 (H2)**: The Portuguese beer industry is influenced by at least one of the indicators within the Microeconomic Pillar.
- **Alternative Hypothesis 3 (H3)**: The Portuguese beer industry is influenced by at least one of the indicators within the Tourism Pillar.
- **Alternative Hypothesis 4 (H4)**: The Portuguese beer industry is influenced by at least one of the indicators within the Climate Pillar.
- **Alternative Hypothesis 5 (H5)**: The Portuguese beer industry is influenced by at least one of the indicators within the Football Pillar.

By formulating a single null hypothesis and five comprehensive alternative hypotheses, the aim is to discern whether external factors collectively play a role in shaping beer sales in Portugal. While it may not provide details on the exact contributors, it serves as a valuable starting point for understanding the potential implications of external drivers for the industry.

3.5. Data Analysis

After importing the data sources, it was necessary to perform data preprocessing techniques. The features in the dataset had different scales of magnitude. When modeling the data, some of these variables would have a much bigger impact on the results than others. All the features were scaled to not influence the output of the model. The chosen scaler was MinMax, because "Attributes are often normalized to lie in a fixed range — usually from zero to one — by dividing all values by the maximum value encountered or by subtracting the minimum value and dividing by the range between the maximum and minimum values" (Witten et al., 2011).

In order not to bias the analysis, observations were removed for the months of lockdown due to COVID-19, namely April and May 2020 and February and March 2021. In these months, although sales are recorded on the food channel, the Horeca channel (Hotels, Restaurants, and Cafés) reports very low sales, if it reports any sales at all.

To gain initial insights into the relationship between the variables, a correlation analysis was conducted. The correlation analysis aimed to identify the strength and direction of the association between pairs of variables. The results of the correlation analysis provided valuable information regarding the interdependencies among the variables. Based on the correlation coefficients, a few independent variables reported a strong correlation (over 0.7 or less than -0.7) between each other.

After gathering the previously mentioned data sources and variables, we are left with 16 columns and 62 observations – one for each month for five years, plus the first 6 months of 2023 and less the four months of lockdown. However, it is not feasible to create a model with that many columns, due to the curse of dimensionality. The curse of dimensionality is a common phenomenon observed in machine learning and data mining, it states that as the dimensionality (or number of features) increases, the number of data points (or observations) required for good performance of any machine learning algorithm increases exponentially (Hughes, 1968). Therefore, an initial feature selection framework was used. The premise of this framework was to compare the correlation between each independent variable and, if two variables reported a strong correlation, the correlation between those independent variables and the dependent variable was calculated. The variable with the lowest correlation with the dependent variable would be excluded from the model.

4. RESULTS

In a systematic exploration of the variables, it was observed that the "Consumer Confidence Index" and the "Perspective of the Country's Economic Situation" exhibited a perfect positive correlation, representing a coefficient equal to 1. Subsequently, when assessing the correlations of both these variables with the dependent variable "Sales in Volume," it was found that the "Consumer Confidence Index" emerged as the dominant feature. However, by continuing the analysis, it was recognized that the "Consumer Confidence Index" was significantly correlated with the "Degree of Household Savings" and was eventually excluded from the model.

A notable pattern emerged as the next set of variables was evaluated. The "Activity Rate" demonstrated a high correlation with the "Degree of Household Savings," with the latter exhibiting a comparatively lower correlation with the dependent variable. Furthermore, the "Consumer Price Index" showed a positive correlation with the "Activity Rate," but this relationship had limited influence on the dependent variable.

Among the tourism-related variables, an anticipated pattern was identified, with high correlations prevalent among these indicators. Ultimately, "National Guests" emerged as the variable showcasing the most substantial correlation with "Sales in Volume". Equally, the "Average Price" displayed a strong correlation with the "Activity Rate" and was subsequently excluded from the model. Finally, "Average Air Temperature" was found to have a significant correlation with "National Guests" and, since the tourism variable was revealed to be the most influential variable concerning "Sales in Volume", secured its place in the model.

Subsequently, the model was left with only five features: Unemployment Rate, Activity Rate, Average Monthly Gross Remuneration, Guests (National), and the number of Matches in 1st Football League. This study aimed to analyze the relationship between various independent variables and the target variable, utilizing three different regression models: Simple Ordinary Least Squares (OLS), OLS with Elastic Net, and OLS with Lasso Regression. These were the results from the models:

Pilar	Variables	Simple OLS	OLS w/ Elastic Net	OLS w/ Lasso
Macroeconomic	Unemployment Rate	-0,2370 *** (0,000)	-0,2553 *** (0,001)	-0,2553 *** (0,001)
	Activity Rate	-0.1436 *** (0,021)	-0.1650 *** (0,022)	-0,1650 *** (0,022)
Microeconomic	Average Monthly Gross Remuneration	0,1244 *** (0,017)	0,1028 * (0,101)	0,1028 * (0,101)

Pilar	Variables	Simple OLS	OLS w/ Elastic Net	OLS w/ Lasso
Tourism	Guests (National)	0,9334 *** (0,000)	0,9271 *** (0,000)	0,9271 *** (0,000)
Football	Matches in 1st Football League	-0,0595 (0,163)	-0,0425 (0,398)	-0,0425 (0,398)
	Constant	0.1489 *** (0.016)	0,1737 *** (0,019)	0,1737 *** (0,019)
	Observations	62	62	62
	R-squared	0,883	0,876	0,876

Table 2 - Regression Results

P-Values in parenthesis. *** p < 0,01, ** p < 0,05, * p< 0,1.

This table highlights the factors that have an impact on beer sales in volume in the Portuguese market and shows that all variables are statistically significant except for Matches in the 1st Football League. The Unemployment Rate variable has a negative coefficient of -0.2370, which indicates that as the unemployment rate increases, beer sales in volume tend to decrease, as well as the Activity Rate has a negative coefficient of -0,1436. As for the Average Monthly Gross Remuneration variable, the positive coefficient of 0.1244 suggests that as the average monthly gross pay of employees increases, beer sales tend to increase. The number of National Guests has a positive coefficient of 0.9334, implying that an increase in the number of national guests staying in tourist accommodation establishments is associated with higher beer sales. Despite not being statistically significant, the number of Matches in the 1st Football League has a negative coefficient of -0,0595.

Considering the pillar categorization and hypothesis definition, these results prove that Macroeconomic, Microeconomic, and Tourism indicators affect the Portuguese beer market, and, consequently, hypotheses H1, H2, and H3 are proven.

Evaluating the performance of the model, it is worth noting that the R-squared value is 88,3%. While this metric is traditionally regarded as a measure of model fit, it is essential to acknowledge that R-squared is not without its critics. Particularly, Quinino et al. (2013) highlight that R-squared lacks the foundation of a statistical test raises questions about its intuitive justification as a descriptive statistic, and even argues against reporting R-squared altogether. This sentiment is echoed by several prominent researchers, including Legates & Davis (1997), who posit that employing correlation-based metrics like R-squared for assessing model performance may not always be the most appropriate practice. In the context of social science research, an R-squared value falling within the range of 0.50 to 0.99 is generally deemed acceptable, especially when most of the explanatory variables demonstrate statistical significance. However, it is imperative to exercise caution, ensuring that the elevated R-squared is not a consequence of spurious causation or multi-collinearity among the explanatory variables (Ozili, 2022).

5. DISCUSSION

The regression results suggest that beer sales in volume are affected by **Macroeconomic indicators (H1)**, specifically, the Unemployment Rate and the Activity Rate. Cho and Ogwang (2018) found a positive relationship between beer sales and economic conditions in Canada, suggesting that beer sales increase during good economic times. Nelson (2010) found a pro-cyclical relationship between alcohol consumption and unemployment in an international panel of countries. However, Lester (1996) found a negative association between unemployment rates and alcohol consumption in eight nations studied. In this case, it seems that the relationship between beer consumption and unemployment is also pro-cyclical – this means that the behavior and actions of a measurable product or service move in tandem with the cyclical condition of the economy. Not only that but the results from this model seem to align more with Lester's study, since the Unemployment Rate has a negative coefficient.

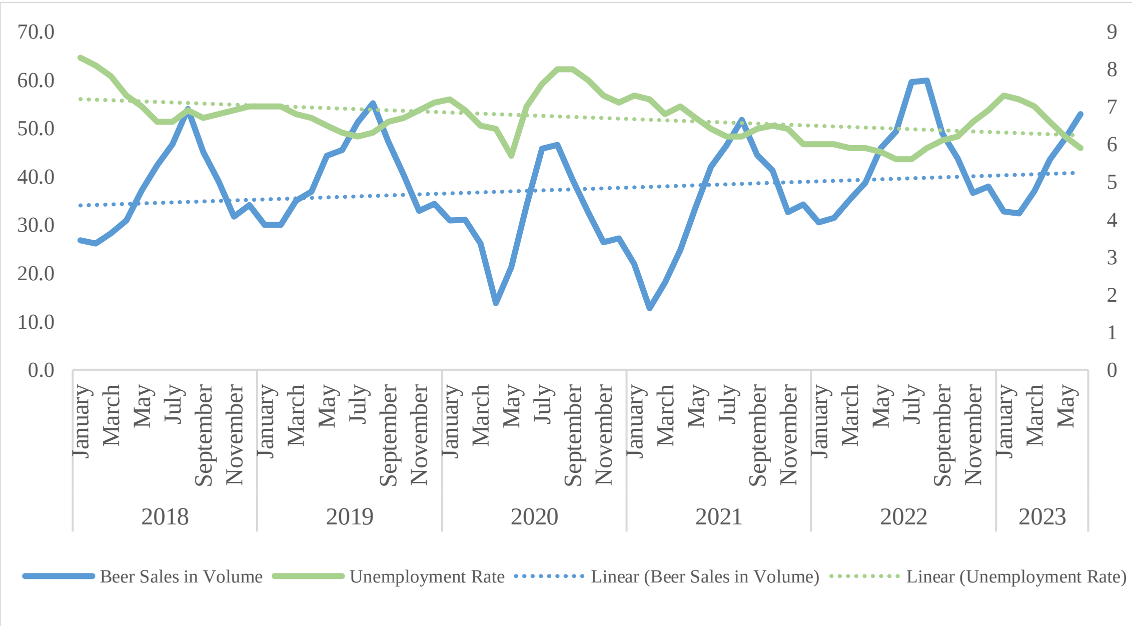

Figure 6 - Beer Sales in Volume and Unemployment Rate Line Graph

Regarding the activity rate, by analyzing a cross-country panel covering 169 nations and time-series observations across 52 years, Angerer et al. (2019) determined that an increase in beer consumption was positively related to the level of urbanization in a country and working age group (from 15 to 64 years). The activity rate represents the percentage of employed and unemployed people aged between 15 and 64. Beer is an alcoholic drink that usually begins to be consumed at the age of 18. Therefore, it can be theorized that an upsurge in the activity rate could lead to an increase in beer consumption for this reason. This growth in the activity rate could be due to an increase in immigration, especially from countries that consume a lot of beer. According to the 2022 Immigration, Borders, and Asylum Report drawn up by the Foreigners and Borders

Service (SEF), Brazil is the main foreign resident community in Portugal, followed by the United Kingdom and Cape Verde (Mota Lopes & Machado, 2022). In 2021, Portugal had a per capita beer consumption of 56.7 liters, lower than Brazil and the UK with 67.9 and 67.6 liters respectively (Kirin Holdings Company, 2022).

However, the results from this model suggest that when the activity rate increases, beer consumption in Portugal tends to decline. With a higher activity rate, there may be a general trend towards a healthier lifestyle. People engaged in more physical activities might be more conscious about their overall well-being, leading them to reduce their alcohol consumption, including beer, which is often associated with calorie intake. Preedy (2009) and Sohrabvandi et al. (2012) highlight the potential health benefits of moderate beer consumption, including nutritional benefits, cardioprotective effects, and immune system stimulation. Nevertheless, excessive consumption of beer can lead to health disorders such as obesity, increased risk of dementia, increased risk of cancer induction, and social misbehaviors.

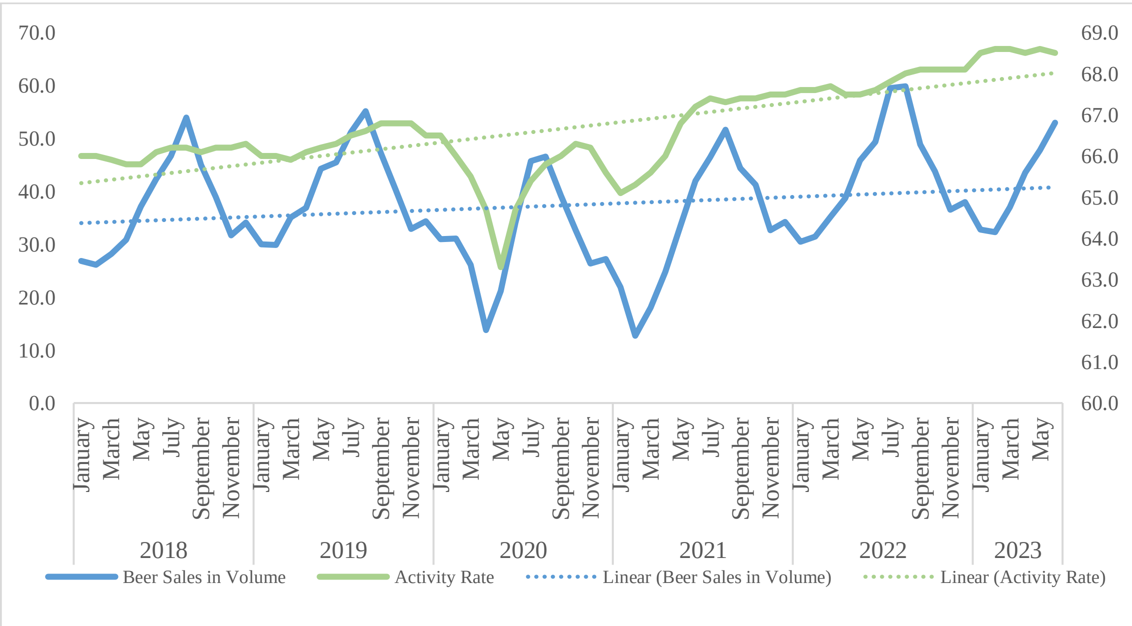

Figure 7 - Beer Sales in Volume and Activity Rate Line Graph

Besides the Macroeconomic Indicators, beer sales in volume also seem to be affected by **Microeconomic Indicators (H2)**, particularly the Average Monthly Gross Remuneration per Employee. Colen & Swinnen (2010) found an inverse U-shaped relationship between income and beer consumption, where beer consumption initially increases with rising income but decreases beyond a certain income level. In these results, it appears beer consumption increases when the average monthly gross remuneration also grows. By looking at the line graph of the data for these two variables, it is possible to see the cyclical behavior of the average monthly gross remuneration overlaps with the seasonality of beer consumption, particularly in the summer.

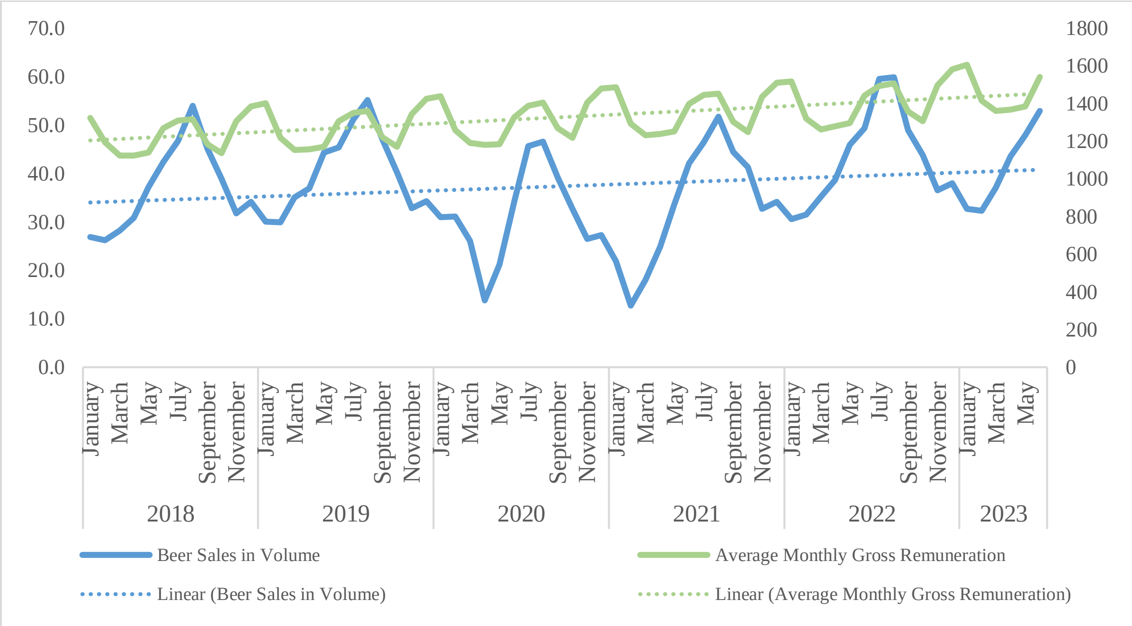

Figure 8 - Beer Sales in Volume and Average Monthly Gross Remuneration Line Graph

The model results also suggest that **Tourism (H3)** does have an impact on beer consumption. Örnberg & Room (2014) highlight that heavy drinking by tourists from high-consumption societies can increase alcohol consumption levels in low- and middle-income countries. What's interesting in this outcome is the fact that beer consumption is more related to national guests than foreign guests. Nonetheless, Stone et al. (2020) focus on beer tourists as a distinct market segment, noting that they participate in various tourism activities and spend more on food and beverages during their travels. Kraftchick et al. (2014) support the idea of beer tourism, identifying motivations such as craft brewery experience, enjoyment, socialization, and beer consumption. Through the line graph is possible to see how similar the two variables behave – both being affected by the summer seasonality.

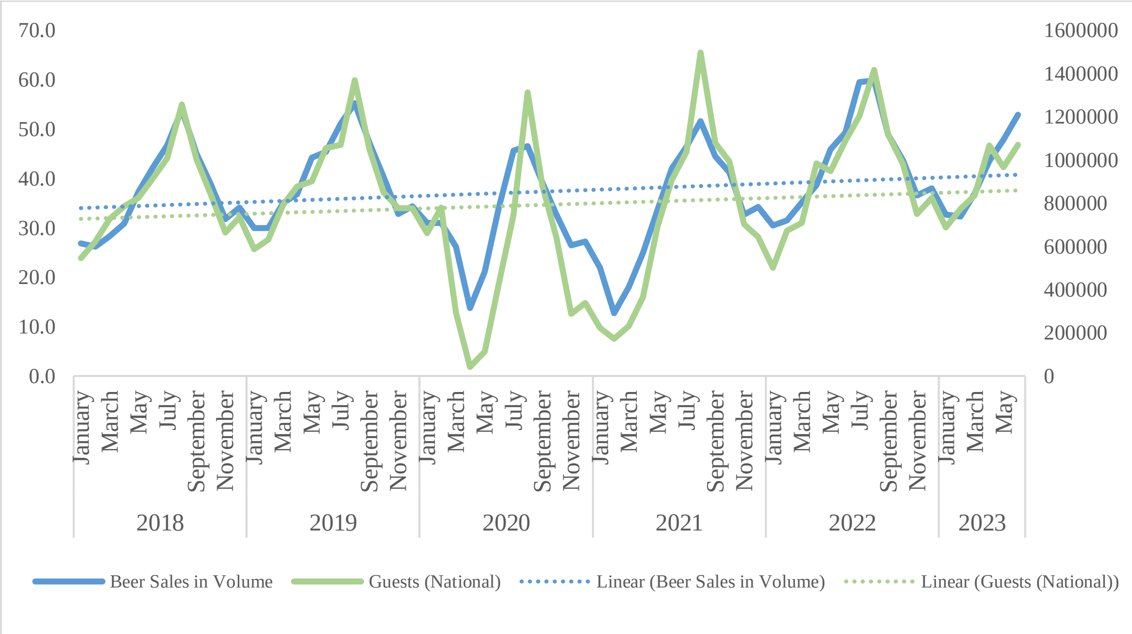

Figure 9 - Beer Sales in Volume and Guests (National) Line Graph

The number of Matches in the 1st Football League, which belonged to the **Football Pilar** (**H5**) was the only variable to be a part of the final model that proved not to be statistically significant. Estrada & Tryggvesson (2001) described the role of alcohol on Swedish football supporters and argued that beer has symbolic importance for football supporters, namely, helping them complete a moment of happiness, relaxation, and a sense of collectiveness. Taking this into consideration, this thesis theorized that the number of monthly football games in Portugal could affect beer consumption. However, it's necessary to keep in mind that the football season in Portugal ends in May and starts at the end of August. This means that there are no 1st league matches registered in the months where beer consumption is the highest, namely June, July, and August. Therefore, as it will be discussed in the limitations and future work section, it would be interesting to analyze other football moments in Portugal, such as Portuguese national team matches or European competitions.

Lastly, it is worth mentioning the **Climate Pilar (H4)**. This pillar, just like the previous one, only had one variable and, due to its high correlation with the Activity Rate variable, it was removed from the model. With that said, it is not possible to define the variable as not statistically significant since it wasn't in the model. Ventura-Cots et al. (2019) found an inverse correlation between colder weather and fewer sunlight hours with alcohol consumption and alcoholic cirrhosis worldwide. Hagström et al. (2019) conducted a study in Sweden and found a non-linear inverse association between temperature and alcohol consumption, suggesting that warmer temperature was associated with increased alcohol consumption. Overall, these papers present mixed findings, and further research is needed to understand the relationship between climate and alcohol consumption.

6. LIMITATIONS AND FUTURE WORK

To improve the robustness and applicability of this framework, some possibilities for future work are presented below. These suggestions include both the improvement of the data collection process, as well as and the expansion of the analytical setting. One opportunity for future work involves automating data collection through ETL (Extract, Transform, Load) processes. This would address the challenges caused by data sources such as the INE and Liga Portugal, which currently require manual extraction and updating.

Incorporating additional data sources could be an interesting path to follow. This includes expanding the climate pillar with data on precipitation, which can significantly impact consumer behavior. Moreover, data from Portuguese national team matches and European leagues could be added to the football pillar, offering a more comprehensive view of football-related dynamics. Recognizing the cultural and entertainment aspects that influence consumer choices, a potential option involves introducing a new pillar for music and culture. This would encompass factors such as shows, concerts, and music festivals. Considering the influence that wine consumption can have on beer consumption, it would be pertinent to understand the influence that this drink has on beer consumption, both in terms of consumption and in terms of price and supply because they are usually substitute goods – they are part of the same market as another product and compete for the same type of consumer.

To make the framework more accessible and interpretable for businesses, the development of an on-demand dashboard is a crucial opportunity. Such a dashboard would provide real-time access to results, enabling stakeholders to monitor and update Key Performance Indicators (KPIs) effortlessly.

Lastly, the application of this framework extends beyond just the beer industry. There is potential to broaden the scope and apply similar analyses to other product categories and markets. This expansion would allow businesses in various sectors to employ data-driven insights for informed decision-making.

7. CONCLUSION

In conclusion, an innovative artifact, and a framework to investigate the external factors that influence beer sales in Portugal was developed and supported in an OLS Regression model. Through this analysis, several key factors were identified that impact beer sales volume, namely the unemployment rate, the activity rate, the average monthly gross remuneration per employee, and the number of national guests. These results indicate convergences and divergences with earlier findings, confirming the unique characteristics of the region and the data sources used for the study. These findings provide valuable insights for the beer industry in understanding the drivers of beer sales and formulating effective strategies, improving both the Marketing and Sales departments' decision-making processes.

These results not only give practical insights for industry professionals but also add to academic knowledge by broadening awareness of the factors impacting beer sales in Portugal besides consumer behavior. It offers up new areas for research in this area and emphasizes the need to consider the country's economy and consumer's sociodemographic conditions. In terms of the work's objectives, they were fully accomplished, contributing to knowledge advancement.

8. BIBLIOGRAPHY

Angerer, M., Dünser, M., Kaiser, L., Peter, G., Stöckl, S., & Veress, A. (2019). What drives our Beer Consumption?---In Search of Nutrition Habits and Demographic Patterns. *Applied Economics*, *51*(41), 4539–4550. https://doi.org/10.1080/00036846.2019.1593938

Betancur, M. I., Motoki, K., Spence, C., & Velasco, C. (2020). Factors influencing the choice of beer: A review. *Food Research International*, *137*, 109367. https://doi.org/10.1016/j.foodres.2020.109367

Bordeleau, F.-È., Mosconi, E., & Santa-Eulalia, L. A. (2018). *Business Intelligence in Industry 4.0: State of the art and research opportunities*. https://doi.org/10.24251/HICSS.2018.495

Campbell, M. C., Inman, J. J., Kirmani, A., & Price, L. L. (2020). In Times of Trouble: A Framework for Understanding Consumers' Responses to Threats. *Journal of Consumer Research*, *47*(3), 311–326. https://doi.org/10.1093/jcr/ucaa036

Cho, D. I., & Ogwang, T. (2018). Investigating the Link between Beer Consumption and Economic Conditions in Canada: A Panel Data Analysis. *Journal of Alcoholism & Drug Dependence*, *06*(01). https://doi.org/10.4172/2329-6488.1000298

Colen, L., & Swinnen, J. F. M. (2010). *Beer Drinking Nations The Determinants of Global Beer Consumption*. http://www.econ.kuleuven.be/licos

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, *949*, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Duleba, A., & Olive, D. (1996). Regression Analysis and Multivariate Analysis. *Seminars in Reproductive Medicine*, *14*(02), 139–153. https://doi.org/10.1055/s-2007-1016322

Estrada, F., & Tryggvesson, K. (2001). A Part of the Game – alcohol, football fans and male comradeship. *Nordic Studies on Alcohol and Drugs*, *18*(3), 245–260. https://doi.org/10.1177/145507250101800312

Foley, É., & Guillemette, M. G. (2010). What is Business Intelligence? *International Journal of Business Intelligence Research*, *1*(4), 1–28. https://doi.org/10.4018/jbir.2010100101

Gera, M., & Goel, S. (2015). Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity. *International Journal of Computer Applications*, *113*(18), 22–29. https://doi.org/10.5120/19926-2042

Hagström, H., Widman, L., & von Seth, E. (2019). Association between temperature, sunlight hours and alcohol consumption. *PLOS ONE*, *14*(9), e0223312. https://doi.org/10.1371/journal.pone.0223312

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, *14*(1), 55–63. https://doi.org/10.1109/TIT.1968.1054102

Inmon, W. H. (2005). *Building the Data Warehouse* (4th edition). Wiley.

Kalousis, A., Prados, J., & Hilario, M. (2007). Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowledge and Information Systems*, *12*(1), 95–116. https://doi.org/10.1007/s10115-006-0040-8

Kirin Holdings Company. (2022). *Global Beer Consumption by Country in 2021*. https://www.kirinholdings.com/en/newsroom/release/2022/1223_01.html

Kozak, V. (2013). Analysis of reasons for beer consumption drop in the Czech Republic. *E+M Ekonomie a Management*, 130–138.

Kraftchick, J. F., Byrd, E. T., Canziani, B., & Gladwell, N. J. (2014). Understanding beer tourist motivation. *Tourism Management Perspectives*, *12*, 41–47. https://doi.org/10.1016/j.tmp.2014.07.001

Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, *4*(3). https://doi.org/10.6029/smartcr.2014.03.007

Larose, D. T. (2015). *Data Mining and Predictive Analytics*.

Legates, D. R., & Davis, R. E. (1997). The continuing search for an anthropogenic climate change signal: Limitations of correlation-based approaches. *Geophysical Research Letters*, *24*(18), 2319–2322. https://doi.org/10.1029/97GL02207

Lester, D. (1996). Unemployment and Alcohol Consumption. *Psychological Reports*, *79*(1), 150–150. https://doi.org/10.2466/pr0.1996.79.1.150

Lim, E. P., Chen, H., & Chen, G. (2013). Business intelligence and analytics: Research directions. In *ACM Transactions on Management Information Systems* (Vol. 3, Issue 4). https://doi.org/10.1145/2407740.2407741

Marques-Vidal, P., & Dias, C. M. (2005). Trends and Determinants of Alcohol Consumption in Portugal: Results From the National Health Surveys 1995 to 1996 and 1998 to 1999.

Alcoholism: Clinical & Experimental Research, *29*(1), 89–97. https://doi.org/10.1097/01.ALC.0000150001.31722.D1

Mashingaidze, K., & Backhouse, J. (2017). The relationships between definitions of big data, business intelligence and business analytics: a literature review. In *Int. J. Business Information Systems* (Vol. 26, Issue 4).

Mota Lopes, S., & Machado, R. (2022). *Relatório de Imigração, Fronteiras e Asilo 2022*.

Nelson, J. P. (2010). Alcohol, unemployment rates and advertising bans: international panel evidence, 1975–2000. *Journal of Public Affairs*, *10*(1–2), 74–87. https://doi.org/10.1002/pa.348

Örnberg, J. C., & Room, R. (2014). Impacts of Tourism on Drinking and Alcohol Policy in Low-And Middle-Income Countries: A Selective Thematic Review. *Contemporary Drug Problems*, *41*(2), 145–169. https://doi.org/10.1177/009145091404100202

Ozili, P. K. (2022). The Acceptable R-Square in Empirical Modelling for Social Science Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4128165

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Preedy, V. R. (2009). *Beer in Health and Disease Prevention*. Elsevier. https://doi.org/10.1016/B978-0-12-373891-2.X0001-6

Quinino, R. C., Reis, E. A., & Bessegato, L. F. (2013). Using the coefficient of determination R2 to test the significance of multiple linear regression. *Teaching Statistics*, *35*(2), 84–88. https://doi.org/10.1111/j.1467-9639.2012.00525.x

Ragazou, K., Passas, I., Garefalakis, A., Galariotis, E., & Zopounidis, C. (2023). Big Data Analytics Applications in Information Management Driving Operational Efficiencies and Decision-Making: Mapping the Field of Knowledge with Bibliometric Analysis Using R. *Big Data and Cognitive Computing*, *7*(1). https://doi.org/10.3390/bdcc7010013

Sharda, R., Delen, D., & Turban, E. (2017). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*.

Silva, A. P., Jager, G., Van Zyl, H., Voss, H.-P., Pintado, M., Hogg, T., & De Graaf, C. (2017). Cheers, proost, saúde: Cultural, contextual and psychological factors of wine and beer consumption in Portugal and in the Netherlands. *Critical Reviews in Food Science and Nutrition*, *57*(7), 1340–1349. https://doi.org/10.1080/10408398.2014.969396

Sohrabvandi, S., Mortazavian, A. M., & Rezaei, K. (2012). Health-Related Aspects of Beer: A Review. *International Journal of Food Properties*, *15*(2), 350–373. https://doi.org/10.1080/10942912.2010.487627

Stijven, S., Minnebo, W., & Vladislavleva, K. (2011). Separating the wheat from the chaff. *Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation*, 623–630. https://doi.org/10.1145/2001858.2002059

Stone, M. J., Garibaldi, R., & Pozzi, A. (2020). Motivation, Behaviors, and Travel Activities of Beer Tourists. *Tourism Review International*, *24*(2), 167–178. https://doi.org/10.3727/154427220X15912253254437

Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Prado, A. R. (2021). Business intelligence: business evolution after industry 4.0. In *Sustainability (Switzerland)* (Vol. 13, Issue 18). MDPI. https://doi.org/10.3390/su131810026

Valášková, K., & Klieštik, T. (2015). Behavioural reactions of consumers to economic recession. *Business: Theory and Practice*, *16*(3), 290–303. https://doi.org/10.3846/btp.2015.515

Ventura-Cots, M., Watts, A. E., Cruz-Lemini, M., Shah, N. D., Ndugga, N., McCann, P., Barritt, A. S., Jain, A., Ravi, S., Fernandez-Carrillo, C., Abraldes, J. G., Altamirano, J., & Bataller, R. (2019). Colder Weather and Fewer Sunlight Hours Increase Alcohol Consumption and Alcoholic Cirrhosis Worldwide. *Hepatology*, *69*(5), 1916–1930. https://doi.org/10.1002/hep.30315

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier. https://doi.org/10.1016/C2009-0-19715-5