SAVONIA

THESIS - **BACHELOR'S DEGREE PROGRAMME**
SOCIAL SCIENCES, BUSINESS AND ADMINISTRATION

DATA VISUALIZATION AND ITS CORRELATION TO THE BUSINESS DECISION-MAKING PROCESS.

APPLYING BUSINESS INTELLIGENCE TO A LATIN-AMERICAN SME.

Author: Karen Iliana Urbano Cuellar

SAVONIA UNIVERSITY OF APPLIED SCIENCES THESIS
Abstract

Field of Study Social Sciences, Business and Administration			
Degree Programme Degree Programme in Business Administration, International Business			
Author(s) Karen Iliana Urbano Cuéllar			
Title of Thesis Data Visualization and Its Correlation to the Business Decision-Making Process.			
Date	08-11-2020	Pages/Appendices	49/28
Supervisor(s) Mr. Abdelazim Hannoura, Mr. Leo Suomela			
Client Organisation /Partners Non-Disclosed			

Abstract

In 2020, Information Systems are revolutionizing business operations and within that context, the way businesses make their decisions based on their performance analysis. Determining if SMEs are undertaking non-grounded decisions is crucial to understand how business decisions are made and moreover if Information Systems are capable of improving the decision-making process within small and medium-sized companies.

Managers of SMEs, who are not aware of the importance of their business information and how it can affect their decision-making process, are more likely to experience poor managerial performance. This limits the level and quality of insights that these companies get from their overall performance. As a consequence, the process of Business Intelligence has been developed for the past few years and it has been widely adopted and incorporated as an analytical process in every large enterprise to support the decision-making process, yet not integrated into all the small and medium-sized enterprises.

The main purpose of this thesis was to analyze the impact of the Business Intelligence process (and data visualization as one of the main BI stages) as a completely new analytical tool introduced to a Peruvian SME that operates within the food industry and experiences several limitations such as their enterprise size, budget, and technological know-how.

The following academic research was carefully developed for four months and Business Intelligence instruments were developed to improve the company's decision-making process. Those BI instruments were the result of self-development work, and contained different BI tools such as interactive dashboards, databases, and data management processes. Data and observations were collected from top managers during every stage of this study. The company's unique performance, levels of satisfaction, and system needs were assessed on periodical one-to-one online interviews. Moreover, the last assessment evaluated the impact of the BI methodology in the company's operations and was measured through questionnaires. These assessed the usefulness and frequency of use of the BI instruments provided, as well as the end-user experience, and the impact of data visualization within the company.

The results obtained have demonstrated usage of only 58% for the BI instruments. However, the BI process brought up new information to 100% of the survey respondents, demonstrating the positive correlation between Business Intelligence and its contribution to the managerial decision-making process. Furthermore, the research has supported the hypothesis that companies that fulfill the aforementioned business-constraints criteria are more likely to experience limited performance insights that may happen unnoticed for top decision-makers. However, maybe the most important result this research pinpointed was the opposition to transitioning from more traditional management methodologies to technologically advanced based solutions due, once more, to the three main business limitations aforementioned.

In conclusion, this thesis project has been successful in demonstrating a positive correlation between data visualization and its contribution to the decision-making process within a SME, but it has also brought up the incognita if SMEs are prepared to upgrade their traditional managerial working and analytical methodology to the benefits provided by Business Intelligence.

Keywords
Data Visualization, Decision-Making, Strategic Management, Power BI, SQL, Business Intelligence, Databases, Data Cleansing, Dashboards

To my mother, father, and brother, for being my support during this journey.

To Ari Þór Vilhjálmsson, for being the reason of this journey, takk fyrir.

CONTENTS

1 INTRODUCTION 6
1.1 Abbreviations and Definitions 7
1.2 Partners and Copyright Holders or Other Parties 8
2 BUSINESS INTELLIGENCE 9
2.1 Data, Information & Knowledge 9
2.2 BI Advantages 10
2.3 The Business Intelligence Process 11
2.3.1 Databases 12
2.3.2 Data Modeling 13
2.3.3 Star Schemas 13
2.3.4 Data Visualization 14
2.4 Business Intelligence Tools 17
2.4.1 Excel for Business Intelligence 17
2.4.2 SQL for Business Intelligence 20
2.4.3 Power Query 21
2.4.4 MS Power BI 21
2.5 Business Intelligence Impact on Strategic Decision-Making 22
2.5.1 Strategic Management 23
2.5.2 Benchmarking in Strategic Management 24
2.5.3 Decision-making process 24
3 CASE COMPANY INTRODUCTION 27
3.1 Current Problems in the Case Company 27
4 DEVELOPMENT PLAN OF THE BUSINESS INTELLIGENCE PROCESS FOR THE CASE COMPANY 29
4.1 Step 1 of the Plan: Overall Company's Assessment 29
4.2 Step 2 of the Plan: Data Collection 30
4.3 Step 3 of the Plan: Recreating the Business Intelligence Process 30
4.4 Step 4 of the Plan: Questionnaire Assessment 31
4.5 Step 5 of the Plan: Questionnaires' Results Delivered to the User 31
5 BI DEVELOPED WORK USING EXCEL AS DATASOURCE 33
5.1 Stage 1 of the BI process: Data Cleansing. 33
5.2 Stage 2 of the BI process: Creating Relations. 36

5.3 Stage 3 of the BI process: Power Query. 37
5.3.1 Connecting Data to Power Query 37
5.3.2 Power Query Editor Interface 38
5.3.3 Editing Data with Power Query 38
5.4 Stage 4 of the BI process: Data Visualization. 38
5.4.1 MS Power BI Reporting Interface 38
5.4.2 Reporting with the Adequate Chart Type 39
5.4.3 Visualizing Data with Power BI 39
6 BI DEVELOPED WORK USING A DATABASE AS DATA SOURCE 41
6.1 Stage 1 of the BI process: Data Cleansing for a Database 41
6.1.1 Assigning Unique Identifiers to Support Data Cleansing for a Database 42
6.2 Stage 2 of the BI process: Relations and Tables 43
6.2.1 Creating Relations Across Tables 43
6.3 Stage 3 of the BI process: Creating a Database in SQL Server Management Studio 44
6.4 Stage 4 of the BI process: Connecting a Database to Power BI for Data Visualization 48
7 CONCLUSION 49
8 DISCUSSION 51
8.1 Data Cleansing 52
8.2 Creating Relations, Relations and Tables 52
8.3 Power Query 53
8.4 Creating a Database in SQL Server Management Studio 53
8.5 Data Visualization 53
REFERENCES AND SELF-PRODUCED MATERIALS 54
APPENDIX 1: BUSINESS INTELLIGENCE 58
APPENDIX 2: CASE COMPANY INTRODUCTION 59
APPENDIX 3: DEVELOPMENT PLAN OF THE BUSINESS INTELLIGENCE PROCESS FOR THE CASE COMPANY 60
APPENDIX 4: BI DEVELOPED WORK USING EXCEL AS DATASOURCE 63
APPENDIX 5: BI DEVELOPED WORK USING A DATABASE AS DATA SOURCE 73
APPENDIX 6: DISCUSSION 82

1 INTRODUCTION

The motivation behind this thesis project was the interest in determining if Small and Medium Enterprises are undertaking non-grounded business decisions and how information systems can enhance their decision-making processes. In 2020, Information Systems are revolutionizing business operations and how they base their decisions on their performance analysis. For that reason, it is important to understand how Information Systems improve the decision-making process of organizations that are used to implementing non-specialized software. An SME experiencing problems when making decisions, may be experiencing administrative deficiencies that assume a big stake in the main problem. However, these are not only the main causes of why small-sized companies are not making the best business decisions. Other reasons could be technical know-how deficiencies, misconceptions about technology implementation, and resources availability.

One of the objectives of this project was to measure the impact of advanced information systems on business operations that are usually only supported by more traditional software, e.g., MS Excel. Secondly, this project recreated the Business Intelligence process within a SME. Thirdly, was to demonstrate how the transition to a different technological process occurred, the complications discovered during the shift, and to present alternatives to the current strategic management approaches used in an SME.

Initially, the thesis project started with an evaluation of the Case Company's managerial performance, including their current software methodology and how their data were managed. Consequently, the company delivered the resources needed to start the BI procedure. The BI procedure included the development of BI tools such as data management processes, databases, and interactive dashboards. All the developed tools were successfully integrated into the company's daily operations for four months, raising awareness of the need of integrating the aforementioned processes into their daily business operations. The thesis project also provided a detailed analysis of the company's performance since the beginning of their operations in 2013. A monthly consultancy advice was constantly provided to the managers based on the findings discovered throughout the BI development. Finally, the user responded two questionnaires that evaluated the impact of the BI implementation in the company.

Despite of the successful BI implementation, the questionnaires responded by the final user, demonstrated that not all managers made use of the BI tools (e.g., BI reports). Only 58% of the respondents used the MS Power BI dashboards, however, 100% of the respondents agreed on have acquired new knowledge from the dashboards. Unfortunately, the Case Company, as part of the small and medium-size sector, suffers the scarcity of the three main limitations for companies to implement more complex data integration systems: business size, technological know-how, and budget. For that reason, SMEs such as the Case Company, will likely delay the upgrade of their current Information Systems. This is notoriously alarming when several errors exist and are result of the ongoing data management methodology and software.

1.1 Abbreviations and Definitions

SME = Small and Medium Enterprise

MNE = Multinational Enterprise

DAX = Data Analysis Expressions

SQL = Structured Query Language

Strategic management = development, utilization, and assessment of individual steps that each organization sets to reach common goals.

Synergy = the merge of the power of different objects that when working in conjunction, reach greater power rather than working individually.

SWOT analysis = strengths, weaknesses, opportunities, and threats analysis.

PEST analysis = political, economic, social, and technological analysis.

BI = Business Intelligence

DV = Data Visualization

.CSV = Comma-Separated Values

.xlsx = Microsoft Excel file

Power Query = data connection service that allows managing data sources as needed.

Tableau = reporting and analytics Business Intelligence software

DB = Database

Power BI = reporting and analytics Business Intelligence software

PB = 1 000 000 gigabytes

HR = Human Resources

CRM = Customer Relationship System

Granularity = the level of detail in the data

PDF = Portable Document Format

UI = User Interface

UX = User Experience

Relational database = a database that contains items related to each other

Integer = a number that has no fractions or decimals

String = a sequence that can contain characters such as letters, numbers, symbols, and punctuation marks.

Yape = Peruvian-origin app to make instant transfers without a bank account.

PK = Primary Key

FK = Foreign Key

SSMS = SQL Server Management Studio

View = (in SQL) is a table based on the result of another SQL statement.

AI = Artificial Intelligence

1.2 Partners and Copyright Holders or Other Parties

The following research has been carried out with the help of the different parties who supported the development of this project. Initially, Mr. Javier Gómez, Sr. Data Engineer and Business Intelligence instructor assisted in the search of a suitable company interested in participating in this thesis.

Consequently, I could establish communication with the final and end-user party of this project, a small-sized company based in Lima, Peru. The SME operates in the food sector and has been in the industry for seven years, since 2013. The company mainly distributes a variety of ice creams, healthy snacks, and dog treats to different districts alongside Lima, Peru. The main communication occurred between the author of the research and Mr. Rodrigo Gutiérrez, General Manager Assistant of the company.

The project was also partially supported by Mr. Christian García, Sr. Business Analyst who provided professional advice on Business Intelligence analytical tools and resources. Finally, more complex data integration methodology was supported by Mr. Omar Urbano, Sr. Software Developer, who assisted the academic research in different stages. To whom all the above, I am entirely thankful.

2 BUSINESS INTELLIGENCE

In an organization, historical data helps to detect trends that in cooperation with Information Systems, can predict important behaviors of the company. An organization that observes clear trends is likely to formulate a strategy to counteract or benefit from them while giving the business a purpose, e.g., increase/maintain sales across seasonal shifts. Perhaps this is an easy or less complicated task for MNEs rather than to SMEs as they establish more tailored strategies and set larger budgets for each of their departments, handle larger amounts of data, demand a larger storage capacity for their information, and also demand the automation of processes (Heang and Mohan s.a., 2.) Managers at MNEs have less time available to individually analyze data, managers at MNEs are also the main decision-makers for their own departments, and lastly, managers need to act fast and make quick decisions that may affect several business units (Heang et al s.a., 4). For these reasons, data needs to be treated in such a way that once handled, processed, and transformed, decision-makers will be able to easily spot, understand, and analyze new patterns and insights that support the decision-making process, and that is named Business Intelligence. The idea behind BI is to process organic data from sales, customers, employees, financial statements items, etc. to provide an overview of the business and support decisions (Sekhar 2013, 9.)

Today, large companies work with complex software and methodologies that support the BI development, and without those tools, MNEs would not be able to process the amount of data generated day by day nor perform optimal operations involving consumers, suppliers, etc. BI supportive tools do not only provide a business overview, but they also structure the way the information is stored and consumed, being this one of the main reasons why large enterprises opt to use more complex systems and methods to manage their data (Riahi 2017, 1.) The importance of having structured information is what can assure that the decisions made by managers are as reliable, understandable, and accurate as possible compared to how smaller companies tend to store their information in software that does not necessarily function as a database or analytics tool, e.g., spreadsheets, without discerning from their operational database (Riahi 2017, 3).

2.1 Data, Information & Knowledge

As aforementioned, the Business Intelligence implementation is a complex or non-intuitive process that – until some years ago – only large-sized companies had the opportunity to implement in order to achieve different purposes, such as saving costs. To understand the importance of using the BI methodology, managers and decision-makers must first engage with the qualitative and quantitative benefits of the process (Amoako 2013, 4.)

To gain an in-depth understanding of the BI process, it is needed to discern the differences between three important concepts involved: data, information, and knowledge. Plain business information collected due to the nature of the company and its operations, even when gathered and stored within the business systematic working methodology, cannot be immediately used for the decision-making process (Vercellis 2009, 6.)

Vercellis (2009, 6) supports that contrarily to how the company processes its data, in order to implement Business Intelligence, data must be first transformed into information and knowledge, and states the following definitions for recurrent BI concepts that are commonly used interchangeably:

- **Data**. Is the structured representation of single (or individual) data points that correspond to one or more entities, such as sales data or customers' data.
- **Information**. Refers to the results obtained from the activities applied to the data, e.g., extracting or processing. Data becomes meaningful once it has been manipulated and presented as information in a specific form.
- **Knowledge**. Refers to using information as the base for making decisions, developing strategies, and improving actions.

2.2 BI Advantages

At a first glance, the use of more common software to support SMEs administrative processes may appear inexpensive, easy to use, and effective, however, the problem appears when a business needs to make data-driven decisions and rely on accurate information that can be quickly and easily understood. For SMEs, it would be necessary to count with qualified staff capable to translate data analysis into knowledge as accessing highly detailed and structured data does not guarantee a better performance, but it only guarantees to display critical points that cannot be seen by the bare eye, and what a company does with that data is what guarantees success (Howson 2008, 3.)

Translated into benefits, the business intelligence process provides an overview of the company's performance, in most of the cases, in real-time, allowing data to be consumed in a more flexible and accessible fashion (Azvine, Cuid, & Nauck 2005, 1). Consuming a real-time business performance is supported by Data Visualization, which is one of the core stages of Business Intelligence, which allows users to rapidly understand if the company's metrics make sense within the overall picture. The level of interactivity reached throughout the BI process stands as a way to monitor how operative results behave in the organization, and to know whether if strategies are paying off and/or if milestones are being conquered (Howson 2008, 4.)

Supported by Imanuel (2020), there are three main advantages provided by the BI methodology being the **first benefit** how the intelligence in the process can assist SMEs with expertise-limitations to forecast their metrics, as different BI tools contain different in-app features that assist the user to process the data more easily and identify risks, trends, or patterns. A **second benefit** is how the information is being imperatively communicated across units as most of the BI platforms, if not all, are allocated on the cloud, making data to be more efficiently transmitted (or accessed) by the users who have to consume it. Lastly, the **third benefit** is the easiness to visually understand plain text and numbers that are structured and organized inside the environment of a BI reporting tool, which purpose is to ease the task of creating charts and graphs for Data Visualization.

2.3 The Business Intelligence Process

Supported by Vercellis (2009, 25) there are six major components in the BI system: *Data Sources, Data Warehouses and Data Marts, Data Exploration, Data Mining, Optimization,* and *Decisions*, as shown in FIGURE 1:

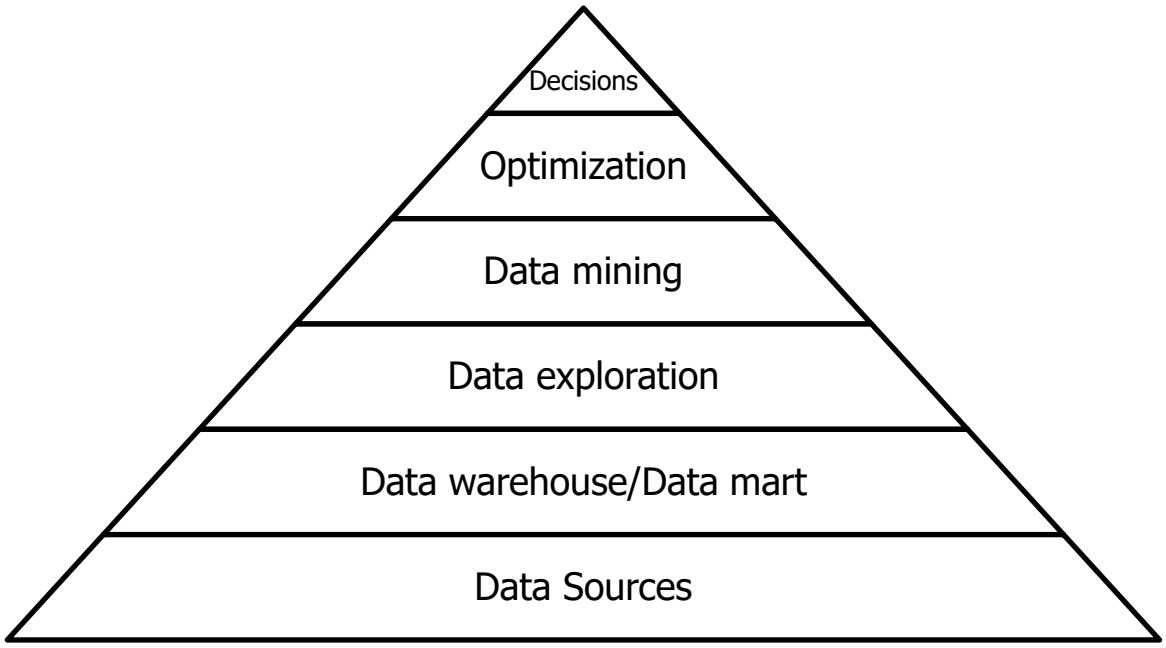


FIGURE 1. Six major components in the BI system (Vercellis, 2009).

Data Sources imply all the raw information that is stored in different services (e.g., cloud services, spreadsheets, emails, etc.) and later transposed to other integration systems. The conditional for data sources is that data points must comply with the same characteristics and create relations between each other, and this is what will turn data into information and allow BI reporting tools to effectively manage it. This is a stage that most of the new users of BI ignore and a conditional that is not always met. Shaping the information and unifying it with the same characteristics requires a major effort.

Data Warehouses and Data Marts are the second component of the BI process and refer to databases that allocate information in a structured and relational manner.

Data Exploration refers to the utilization of different tools to carry out analyses. These analyses may depend on queries, reporting tools, and mathematical/statistical methods.

Data Mining refers to the "extraction of information and knowledge from data".

Optimization refers to models that support the identification of the best solution amongst all available options.

Finally, the last component is **Decisions,** and it refers to choosing and accepting one alternative as a solution.

Referring to the six major components in the BI process, the final two stages are dedicated to the decision-making process, however, data exploration and data mining require tools that support visualization that could be either reached through an analytics tool such as Power BI or MS Excel (Vercellis 2009, 25; Havens Consulting 2020). A difference between spreadsheets software and the BI process relies heavily on the visualization scale (Havens Consulting 2020). FIGURE 2 displays how the development process in the BI methodology occurs, as well as the weight of each stage in the BI process.

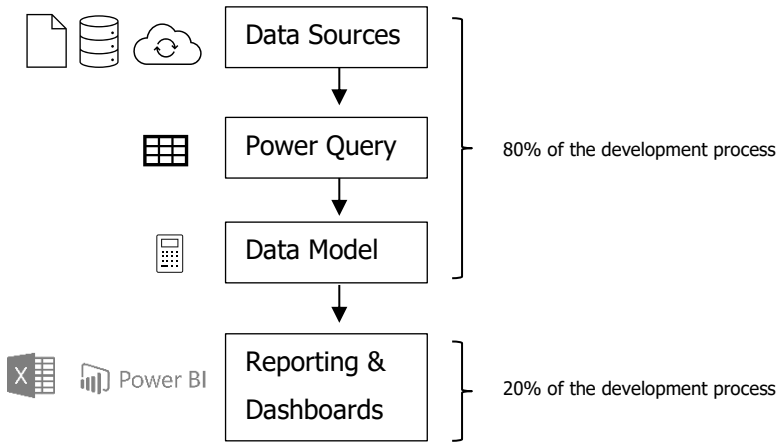


FIGURE 2. The Report Development Process (Havens Consulting, 2020).

2.3.1 Databases

As previously mentioned, the BI process does not only create charts that are visually appealing and provides a more accurate observation, but it also adds another level of structure within the organization. Employing database systems as part of the BI process has the purpose of gathering data and collecting it in a structured and organized manner where the main function of a database is to store and retrieve data in data warehouses and data marts as requested by the users (Alnoukari & Hanano 2017, 6.) Databases are then referred to different systems that accept third-party connections with DBs, such as MS Power BI where final users can develop specialized BI dashboards for their use in strategic management (Alnoukari et al. 2017, 8).

Amongst the several types of DBs, relational databases contain several tables and schemes that create relationships between data points across tables with the potential to store vast amount of qualitative information (Charis 2020). Databases can also collect quantitative information which in conjunction with qualitative information, is how the relations in databases are built (Pilon 2013).

2.3.1.1 Data Types

When creating a database, the most important aspect to consider is what are the data types we will be working with. The type of datum indicates what kind of data can be stored inside each column when creating a table (W3Schools s.a.). This functions as a guideline for DBs systems to understand what type of data is expected to be stored in the columns and helps the systems to identify how the database language will interact with the stored data (Microsoft 2017). Microsoft (2017) states the following six instances of data types:

`INT` = primary integer data type in SQL.

`VARCHAR` = this is a length string. It can store letters, numbers, and special characters.

`DECIMAL (size, d)` = An exact fixed-point number. *`Size`* specifies the integer number(s) and *`d`* the decimal number(s).

`DATETIME` = specifies a date and time combination in the format of YYYY-MM-DD hh:mm:ss.

`BIT` = an integer that can be only 0, 1, or `NULL`.

2.3.2 Data Modeling

A data model is the form in which data is structured in order to comply with the requirements of the language employed in a database system that relates data points and rules which are the key components for a database to operate (Mamčenko s.a.). To understand the relationships between data points, we must first need to understand the concept of granularity in the datasets. Granularity is the level of detail in the data and therefore, having more detailed information allows the user to perform more detailed analyses, however "one must have data at the right level of granularity" and the right level of granularity will depend on one's own needs. (Ferrari & Russo 2017b, 20.)

In order to create a more interesting and detailed data model, the table needs to be related to other tables with the help of relations. It is, therefore, the reason why the databases are so-called relational databases, as they are interconnected using unique identifiers that create that relationship between tables, a fact sustained by (Ferrari et al. 2017b).

2.3.3 Star Schemas

A company can present several assets that interact with each other and those assets can be products, categories, color, dimensions, price, brand, discounts, or customers' names and last names, addresses, age, purchasing amounts, phone number, etc. Stated by Ferrari et al. (2017b), these assets interact with each other and when this interaction occurs, events are created. If a customer interacts with a product the result might be a sale, but it could also be a return. According to Ferrari et al. (2017b), when the separation between assets and events occurs, the data modeling evolves to a star schema and the entities (assets) are separated into two categories: facts and dimensions.

A dimension refers to informative assets, as earlier mentioned, a product has different attributes such as brand, measures, category, etc. On the other hand, facts refer to numerical values that once related to a dimension provide valuable business insights, e.g., units sold, prices, costs, discounts, etc. (Ferrari et al. 2017b.) Another way to interpret a star schema is, for instance, to think that a sale could be linked to many dimensions, e.g., a sale contains information about the product sold, but also about the customer who bought it, date information when it was sold or the branch information where it was sold, etc., as seen in FIGURE 3.

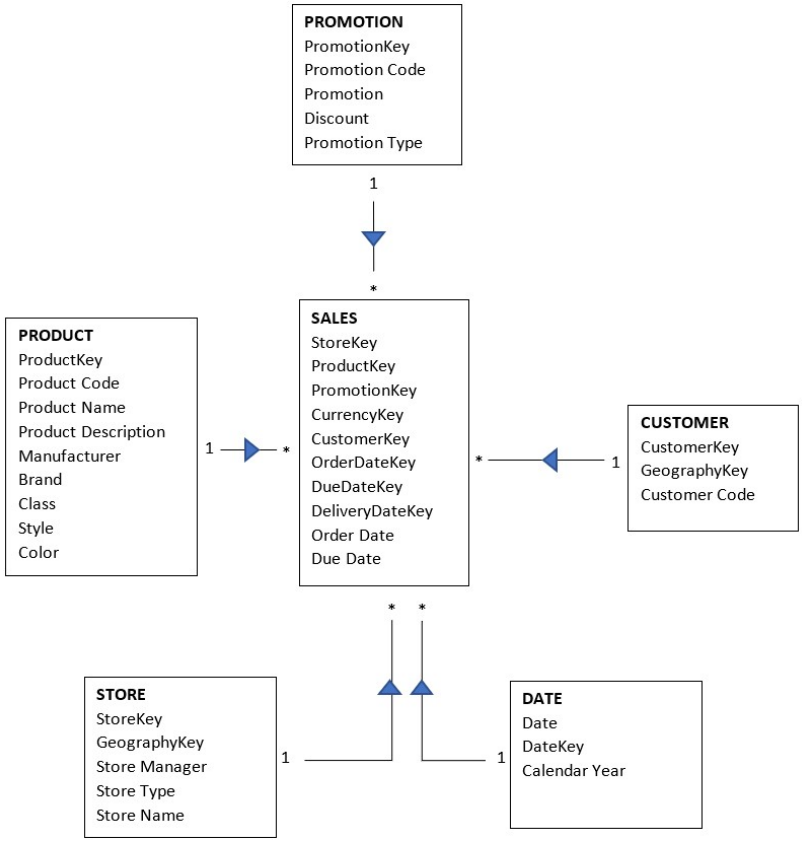


FIGURE 3. A star schema with the fact table in the center and dimension tables around. (Ferrari et al. 2017).

2.3.4 Data Visualization

Today, visualizing data has become more popular as there are more visualization tools available that allow users of any skill level to create graphics and transmit insights in a quicker, more powerful, and more effective form (Laumans s.a.).

As described by Tufte (2007, 7), "data visualization visually displays measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color". Therefore, data visualization can also be described as one of the last stages in the BI process where the information is presented to the final users in a visual representation using specific formats, such as shapes and colors that convert raw data into meaningful information that can be easily understood and analyzed (Tufte 2007, 7-8).

Supported by Zheng (2017), when visualizing business data, there are two main focuses, which are a) to visualize key metrics for an easier and faster comprehension that eases the process of decision-making, and b) to provide a visual and interactive form to exploring data. Perhaps the most common data visualization we are all used to the most is the weather forecast which presents interactive techniques to captivate the viewer's attention and clearly presents the information (Rautenhaus, Böttinger, Siemen, Hoffman, Kirby, Mirzargar, Röber, Westermann 2017). FIGURE 4 displays how a weather visualization looks like appealing to the easiness for the users to understand its meaning.

FIGURE 4. Example of a visualization of a weather forecast. (Yle, 2019).

Users or beneficiaries of data visualization make the minimal effort to identify and rapidly internalize the best knowledge out of the visual representation (Tableau s.a.). The weather visualization is a clear example of how data visualization works in a real-life context, and it can be perfectly extrapolated to a real business-case scenario.

2.3.4.1 Benefits of Data Visualization

As discussed before, DV enhances the form of how information is understood by final users and supports decision-making and problem-solving processes. Zheng (2017) suggests more benefits of data visualization:

- DV provides an overall picture of complex information and spots patterns, structures, relationships, and trends in an easier format.
- As DV transforms data into visual representations, it simplifies the cognitive process of loading and memorizing information (Zheng 2017; Borkin et al. 2013).
- DV contains elements that act as visual cues to draw attention to specific areas of interest.

2.3.4.2 Pre-Attentive Attributes

When a visual harness a specific element such as orientation, shape, color, etc., the user can distinguish key pieces of information even quicker, easier, and effortless (Rost 2018). Supported by the Corporate Finance Institute (2015), FIGURE 5 is an example of what one's sight does when visualization is presented in such a way that one's attention is drawn immediately to a specific part of the visual; pre-attentive attributes can make a significant difference once they are implemented in charts.

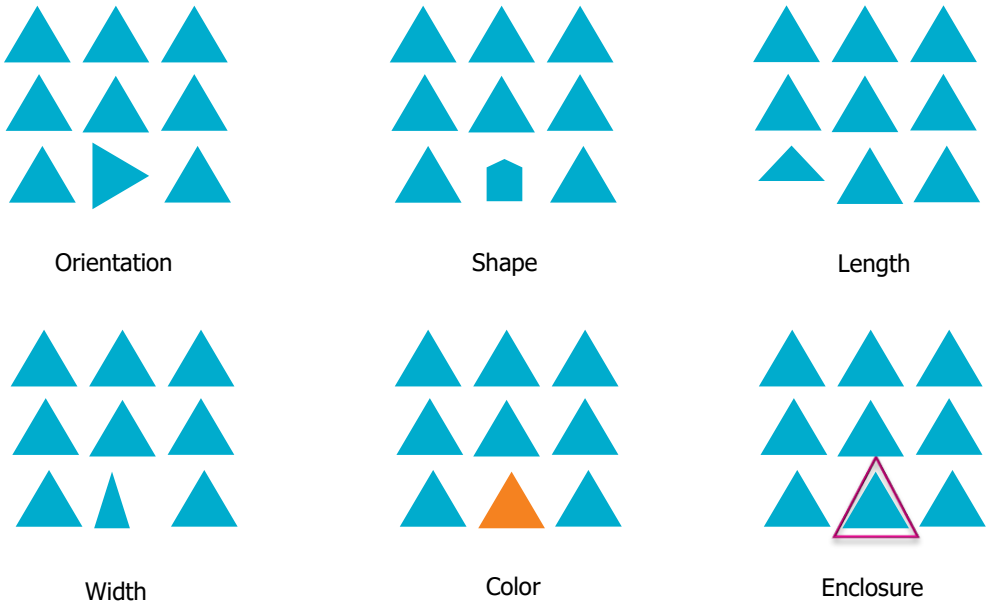


FIGURE 5. Examples of pre-attentive attributes. (Corporate Finance Institute, 2015).

As stated by the Corporate Finance Institute (2015), a chart using the attribute of **color** becomes intuitive and one's attention is immediately captivated by its attribute. In addition to the color attribute used in the second chart, it also uses plain text ("40") to emphasize the metric that could be the most important or relevant to highlight in the chart. The following FIGURE 6 below shows how the left bar chart is not deliberately using any pre-attentive attributes, and how the right chart uses the pre-attentive attribute of color in addition to plain text to highlight the most important datum.

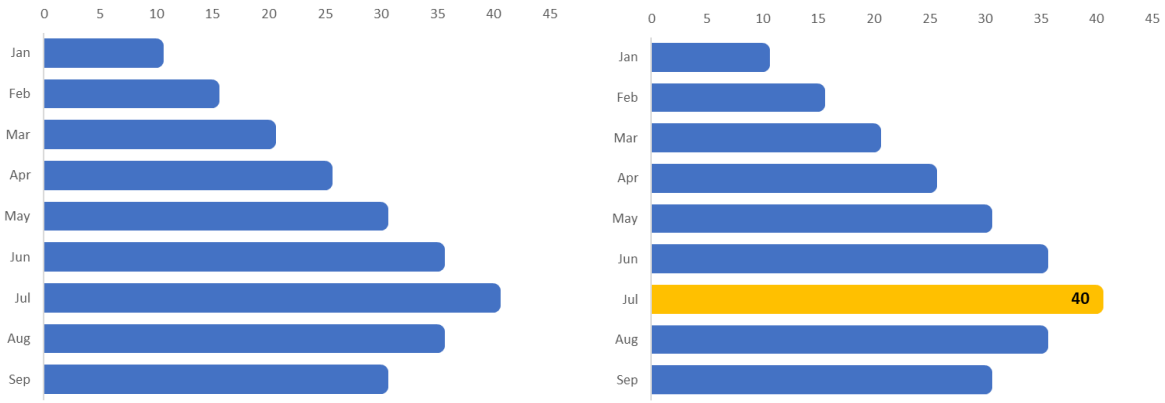


FIGURE 6. Incorporating the attributes with charts. (Corporate Finance Institute, 2015).

2.3.4.3 Multiple Types of Data Visualizations

BI results can be presented using reports, dashboards, and/or analytical tools and is common that BI results are presented in the shape of fixed reports that can be either in a PDF or printed format (Zheng, 2017). See Picture 1 in the Appendix 1.

Dashboards mainly include visualizations, but they could also contain plain text, images, interactive content, etc. in the form of static dashboards (Corporate Finance Institute 2015). See Picture 2 in the Appendix 1.

Likewise, today dashboards are a combination of data and interactivity that provides the user with a certain User Interface (UI) and a User Experience (UX) to control the way information is displayed, filtered, and manipulated. A visual interaction, differently than in MS Excel, happens when the user is able to control how charts behave and filter information respect to other charts and pages (Ferrari & Russo 2016a, 33). Contrarily to a static PDF report or printed report, MS Power BI dashboards display interactivity that provide the user the opportunity to not only read the data but to intuit the information by offering a user-friendly experience (Hsu 2020).

2.4 Business Intelligence Tools

Succeeding in the development of an accurate BI process entirely depends on the capabilities of the user to effectively apply the BI methodology combined with the right used of the Business Intelligence tools (Olszak & Zurada 2015). These software are designed to support the analysis, presentation, and mining of data that contribute to the overall BI process (Kolhe, Salkute, Patil 2011).

2.4.1 Excel for Business Intelligence

Traditionally, Excel is a tool that is most commonly used with accounting and financial purposes that however, in its 2013 edition has included two powerful add-ins: Power Pivot and Power View, including also the plug-ins availability of Power Query and Power Map that unravel a totally different capability for the software: the Business Intelligence purpose. With these additions, Microsoft has turned a commonly and widely well-known tool into a software that can leverage the advantages of already familiar Excel users in order to perform complex analyses, modeling, and visualizations (LeBlanc, Moss, Sarka, Ryan 2015, 14.)

The inclusion of Power Query and other tools for data analysis, revolutionizes MS Excel potential but despite the new add-ins, Microsoft have recognized the deficiencies that the software has in terms of scalability, transparency of analysis, and how data is presented. These weak points have led Microsoft to develop more tools that support the BI tasks that, however, still need third-party connections (Alexander, Decker, Wehbe 2014, 25.)

Third-party connections are inevitable when performing analytical tasks, and those needs surpass MS Excel despite its potential and as aforementioned, scalability, transparency of analyses, and data presentation are some of the limitations that push Analysts to experience beyond the always familiar MS Excel (Alexander et al. 2014, 26). Referring to the most important limitations:

- **Scalability**. Refers to the ability that a software has to meet the demand of growth and complex requirements that despite the amount of data, it can still perform the

tasks. Along the years, MS Excel has had several updates that have increased the amount of data that can be stored in the flat files that, however, does not solve the issues of performance and slowness that the .xlsx files present once they have grown big enough (Alexander et al. 2014, 26.)

- **Transparency of Analytical Processes**. When we think about MS Excel and its data analysis capabilities, we encounter how MS Excel is flexible at the moment of inputting data without considering the data types, e.g., a cell in MS Excel admits numbers, characters, formulas, etc., however, this is also one of its weaknesses: while databases follow rules to store data, MS Excel allows anything (Alexander et al. 2014, 26.)

- **Separation of Data and Presentation**. As aforementioned, MS Excel allows anything and when performing data analysis, the user tends to combine the data (as data source) with how the information is analyzed (the analysis) and presented (the visualization). For instance, when receiving a bank statement, what we see is not the data source, what we see is the visualization, and sometimes MS Excel users tend to combine all these aspects that when further and more complex analyses are required, it is impossible (or will require extra steps) to do it (Alexander et al. 2014, 27.)

2.4.1.1 Differences Between Databases and Spreadsheets

Storing information happens differently in MS Excel than in a database system. MS Excel allows adding any kind of formatting that the user considers helpful to better understand the information such as color-coding to easily spot an important datum, which is a valid practice until it becomes a habit, and the formatting practices overload the information (365 Data Science s.a.). Contrarily to spreadsheets, database systems do not allow unnecessary and/or unmeaningful formatting and in addition to that characteristic, DBs are a powerful tool that support a gigantic amount of data of 524 Petabytes (1 PB = 1 000 000 GB) of information, in its full version (Fernandes 2019). As a comparison MS Excel files' maximum capacity is 2GB however, the size limitations are not the only differences between DBs and spreadsheets (Microsoft 2020).

At a quick glance, comparing a database with a spreadsheet may seem similar as when the data is requested or called from a DB will also show up the information stored in columns and rows, just as in a spreadsheet but as aforementioned, the differences and advantages are vast and despite this can be achieved in a spreadsheet software like MS Excel it has several other limitations, for example, how the information is comprehended, related, and conditioned to the size limits of the spreadsheet (365 Data Science s.a.)

Another major difference is that MS Excel is not a BI tool but is capable of getting the task done, amongst several other kinds of tasks, although within the BI context, MS Excel offers a different level

of freedom to the spreadsheets' users but what happens in it is, that data can be tabulated, stored, designed, and formatted in any way, a non-tech user has no limitations to input information or creates charts and graphs (365 Data Science s.a.) However, it becomes a more intuitive solution to use instead of learning techniques and methodologies for each of the BI stages. For instance, a new user at an SME that has limited resources, e.g., as to time or budget, would rather acquire knowledge on MS Excel than specializing in databases languages such as SQL, Data Preparation such as Power Query, Data Modelling such as DAX, and reporting tools such as MS Power BI or Tableau (Phocas Software 2020.)

A strong difference between using a database or a spreadsheet, relies on the strategy of the company as when an SME has a vision and a plan for the future, the company will try to achieve greater results that will demand, and inevitably, bring more information to the business. In an effective strategic management context, analyzing information supports managers to understand the performance of a company, and when large amounts of data need to be processed, spreadsheets software like MS Excel have limitations: 1 048 576 rows and 16 384 columns that do not require from the user to understand data modeling (Ferrari & Russo 2017, 17.) Therefore, DBs can increase the value of the insights pulled out from information and allow this information to be connected across multiple systems, and use it for several purposes, including the BI process.

When an SME heavily relies on using merely spreadsheets for their management, it risks the possibility that files get ruined, deleted, or lost; compared to DBs which are stored in the cloud and in addition to supporting information related to the business operations they can also be part of different systems that support operative tasks, such as Inventory Control or Human Resources (Phocas Software 2020.). Managing HR through a database helps increase management efficiency while reducing costs and saving time. An HR system would, for example, administrate staff hours or payrolls, amongst others. The uses of databases are broad and today, databases can even take the form of a Customer Relationship Management or CRM (Lindblad 2019).

As aforementioned, MS Excel provides users with more freedom to edit, format, and manage the information as wanted as these users are already familiar with the software, however, it does not necessarily support an effective BI implementation as the software users are not necessarily aware of the methodology they need to follow in the Business Intelligence Process (Sekhar 2013, 60). Managers unaware of the BI methodology base their decisions entirely on a spreadsheet containing excessive formatting, deficient data organization, and a lack of business insights which difficults other users to understand key points and complicates the business performance as these pactices increment the size of the file used and complicate the performance of the company's devices, e.g. computers (Microsoft s.a.).

The difference between a BI chart and a chart that can result from a spreadsheet relies on which method provides more valuable insights in less time and less budget consuming. When a user wants to compare metrics, it would have to reshape how the information has been tabulated and create a different table containing only what the user wants to display; adding extra steps to a process that

demands providing quick and real-time business insights, does not bring value to managers who lead the company's strategies and need accurate, informative, and visually appealing analytical instruments (Linoff 2008, 45.)

2.4.2 SQL for Business Intelligence

SQL or Structured Query Language is the programming language that SQL Server Management Studio understands for performing managerial functions, programming, and manipulating datasets. On the other hand, SSMS is a tool that allows the user to perform database-related tasks such as developing and managing databases, developing SQL queries to retrieve and analyze data as required, and to import/export data from/to third-party software, amongst others (Alexander et al. 2014, 176; Microsoft 2019.)

Despite of being a language, SQL goes beyond the programming sequences and is in fact a whole bundle created by Microsoft that allows the tech user to perform an entire BI process in the SQL environment by accessing the following its own server anatomy (Withee 2010, 149). FIGURE 7 shows the SQL anatomy of its environment that supports the development of the BI process:

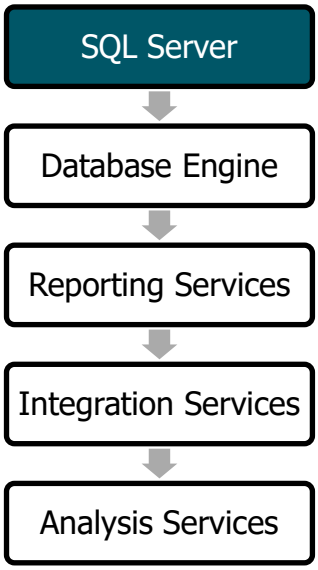


FIGURE 7. The SQL Server BI environment. (Withee 2010).

Solely, SQL is certainly not a BI process but becomes an environment that can perform the whole procedure by, and supported by Withee (2010, 150), offers the following BI platforms:

- **Database Engine.** Refers to the environment that allows the user to create relational databases, and it can be considered as the first step (the database creation) in the BI process.
- **Reporting Services.** These platforms offer the capability to generate somewhat static reports from the already processed and clean data.
- **Integration Services.** It refers to an additional tool that standardizes data in order to be used for an entire company.

- **Analysis Services.** The purpose of this service is to perform analytical tasks including data mining, allowing the user to generate knowledge and extract answers from queries.

2.4.3 Power Query

Power Query was first introduced by Microsoft, as previously mentioned, as an add-in capability that supports uploading third-party data to MS Excel and acts as a "self-service BI solution" that is translated as a tool that gives the user the capacity to generate his or her own dashboards and reports for decision-making purposes. In this way, analysts, and managers themselves have the chance to build their own and customized BI process (Webb 2014, 1.)

Power Query is part of a series that includes different components such as (Power Query), Power Pivot, Power View, and Power Map. These components form the MS Power BI or MS Excel BI environment, and are included in MS Power BI but need to be added to MS Excel as add-ins. Power Query thus is the element that allows the user to perform data cleansing, reshape, and filter data from different sources before it actually goes to integration services (MS Power BI in this case, Analysis Services for SQL) for final reporting or dashboarding (Webb 2014, 2.)

2.4.4 MS Power BI

MS Power BI works as an analytics software capable of analyzing extremely large datasets for different purposes such as extracting knowledge and sharing information across an organization. Developed and brought to the Information Systems world by Microsoft in 2014, it combines data analysis with data visualization tasks that can be display across several devices for the convenience of the end user (Cuddley 2016, 7.)

MS Power BI is intended to be used by, but not limited to, organizations that have the need to harness their internal information such as operative and financial metrics, as well as their external information such as competitors positioning in order to generate immediate strategies that could address any trends spotted and seize any opportunities mined. Concretely, the solution that MS Power BI offers to the organizations and users is the ability to create interactive reports in the form of data visualization dashboards that can held large amount of data and are able to stay always updated, contrarily to a flat data file (Cuddley 2016, 12.)

Supported by Cuddley (2016, 20-24), MS Power BI is part of an analytics suite that contains different tools that support the entire BI experience. These tools are presented as follows:

- **Power BI Desktop.** Is a desktop-based software that is mainly used to build visual analyses by connecting, shaping, visualizing, and sharing insights a wide range of data sources. Its platform is an intuitive environment with drag-and-drop features that ease the process of visualizing data.

- **Power BI Service.** Refers to a paid Microsoft cloud-based service that supports publishing reports either online, web-embedded, or internally across organizations under the same business domain.
- **Power BI Mobile.** This is Microsoft's app solution developed to access Power BI reports in almost any device, including Android and iOS smartphones and tablets.
- **Power BI Gateway.** Refers to a paid service that is installed in an organization's premises that refreshes the on-premises data sources linked to the Power BI cloud reports to always have it updated. Refreshes and updates can be scheduled as needed.

2.4.4.1 MS Power BI Perks

MS Power BI is undoubtedly not the only BI reporting service available in the market, but it does have certain specifications that comply well with businesses, for instance Cuddley (2016 8-11) suggests the following:

- **Free Desktop & Cloud Services.** Differently to other services, MS Power BI does offer a free version of its desktop service as well as its online service as long as the user signs up with an educative or organizational email address.
- **Accessibility.** MS Power BI offers a wide range of connections to almost all, if not all, the data sources in the market, including Excel, databases, Google Analytics. Thus, the connections can be locally stored or in the cloud.,
- **Real-Time Updates.** MS Power BI has dedicated features that can perform automatic, manually, and/or schedule updates to always maintain data sources and visualizations up to date.
- **AI-Based Answers.** An interesting feature in the software, is the attribute of asking questions using only plain text and the software will answer them by analyzing the data connected to the platform.
- **Confidentiality.** As one of the requirements to join the platform is to use an educative or business email address, all the content published in the cloud (depending on the subscription type) will only be shared within the same organization.

2.5 Business Intelligence Impact on Strategic Decision-Making

Despite Business Intelligence being a terminology from 1990s, its popularity has increased in the recent years and the systems that support this methodology entirely comply with the requirements of executives in organizations, these requirements can be translated into analytical, forecasting, and visualization needs. Decision-makers are in charge of the future of their organizations and making the most accurate decisions without organized data could lead to negative impacts, a solution for that problem is the BI methodology (Aghaei & Asadollahi 2013.)

Concretely, the impact of BI in an organization appears as an immediate result that assures the most accurate information and enhances the flexibility of managerial operations, the accuracy of the decisions made, and brings effectiveness and efficiency in the future plans of the company. Decision-makers must be truly impacted by reliable information as their actions in the organizations are the basis of the future success or failure for their companies (Aghaei et al. 2013).

There is a wide impact on BI on the strategic decision-making process that involves supporting decisions at any level in the organization, that is within the BI process, there are certain stages that allow the users to analyze the information in different ways that may include, for instance, a historical assessment of the company's data but also forecasting multiple scenarios able to predict future trends and patterns. Correctly implemented, BI provides a strong base for the exhausting role of leading the future of an organization that must look after different departments and their success (Mahmoudi & Mahdi 2009.) Business Intelligence brings a new perspective of the current company's performance and enhances it as well as it improves every organization's processes by simplifying data management processes where all the insights that matter for the company are located (Beikzadeh, Jafar, & Karim 2009; Golpaygani & Majid 2010).

2.5.1 Strategic Management

Despite its size and its industry, every organization must set and perform a series of steps in order to reach the company's overall goals, and every step involved in the strategy must be individually developed, carried out, and further assessed to function along with different departments and achieve the company's objectives. Thus, strategic management refers to the development, utilization, and assessment of the individual steps that each organization must set in order to reach common goals that every department of the organization must take part of in the implementation of the strategy, that is, departments such as Management, Finance, Marketing, Production or Sales, etc., must comply in a collective synergy towards the organization's success (David & David, 2016, 32.)

In highly competitive industries, e.g., retail, food, IT, the importance of effective strategic management is crucial at the moment of redirecting the company towards better industry-positioning, financial milestones, production efficiency, etc. (David et al., 2016, 33).

Supported by Harvey (2007), effective strategic management is the basis of creating a competitive advantage within organizations as managers and decision-makers at organizations must be aware of how the strategic-management process occurs.

Within the context of SMEs, the process of strategic management sometimes happens almost tacitly, and decisions made are not always strategically planned and have more of an informative purpose and in these cases, no strategic plans are made to address the future performance of the SME and contrarily to these practices, effective strategic management does provide a positive effect on the company's performance despite the business turn (Rajnoha, Stefko, Merkova & Dobrovic 2016).

2.5.2 Benchmarking in Strategic Management

When historical information is only treated as an informative asset, important trends in the company's performance may be ignored and those trends may be causing repetitive and negative behaviors over the years that do not support the company's growth strategies specially during economic crisis when inevitably companies decrease their performance (Rajnoha et al. 2016). In such cases, and with null strategic management implementation, one could say that the strategic management and decision-making processes of the organization have failed to address a particular issue and in fact, one could draw the conclusion of different factors inevitably and negatively affect the company year over year or at specific periods (Mintzberg, 1994.)

One of the most accurate methods to assure an increase in the company's performance and competitive advantages is benchmarking, which is a process that can be either implemented to carry out internal assessments in the company or to evaluate the overall competitors' scene (Rajnoha et al. 2016). Likewise, benchmarking can be further disclosed in:

- **Performance Benchmarking.** A process that compares the company's overall performance against competitors based on key performance indicators fundamental industry parameters.
- **Process Benchmarking.** Contrarily to performance benchmarking, this process analyzes the company's internal processes and compares them to other companies' processes. (Rajnoha et al. 2016.)

2.5.3 Decision-making process

The evaluation of how the decision-making process is implemented in an organization is a key piece to understand how the involvement of key managers or stakeholders reflects on the organization's performance. The optimal functionality of the process involves stakeholders, projects, objectives, environments, and analyses of the type of Business Intelligence that must be considered before making decisions (Barker, Bridges, Hunter, Johnson, Krupa, Murphy, & Sorenson 2001, 8.)

When evaluating a decision-making process, we must first define who are the decision-makers and stakeholders, or beneficiaries. Barker et al. (2001) suggest the following steps in the assessment:

- **Step 1 identifying the problem.** This step involves solving doubts about the problem, defining the problem, and submitting it for review to the decision-makers and stakeholders as, during this stage, the have the faculty to provide feedback. Corrections must be made before moving further in the evaluation process. It must be considered that step 1 involves the initial conditions of the problematic, that is, how the decision-making process is currently being carried out.

- **Step 2 in the evaluation must define what are the needs that any solution implemented must include to be considered feasible.** In this step, it is important to make an analysis of the current company's problem and try to identify the main cause(s). Finally, step 2 ends with a transformation of each cause, from individual problems to self-solutions. A powerful tool that supports the implementation of this methodology is the Problem and Objectives Tree.

- **Step 3 settles the goals to be achieved.** The goals must be considered as positive milestones and can be taken from the objectives tree.

- **Step 4 considers all the alternatives available to attain the milestones.** Any alternatives that are not suitable to meet the required criteria must be discarded.

- **Step 5.** Considers certain discriminating criteria that makes it possible to discard alternatives, if these do not meet the milestones or are unable to achieve the goals set.

- **Step 6.** Selecting a decision-making tool that will assist in the evaluation process.

- **Step 7 & step 8.** Selecting the most feasible alternatives and validate them against the initial requirements to make sure that the alternative(s) chosen truly solve(s) the company's initial problem.

In Figure 8, it can be seen a graphic flowchart defining every step in the analysis of an organization's decision-making process.

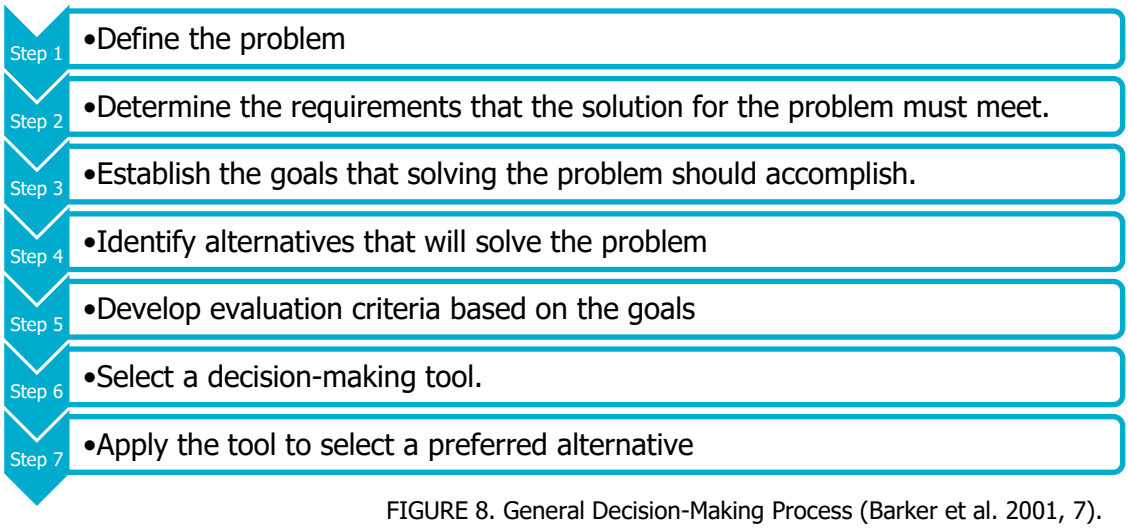


FIGURE 8. General Decision-Making Process (Barker et al. 2001, 7).

Decisions are the final outcome of the BI process and these are solely selected by decision-makers and the results of the intelligent model will be always subject to the final shape that managers decide

to give to the most feasible alternative once the qualified staff is able to operate the data accordingly to the requirements of the organization (Azvine 2005, 4).

3 CASE COMPANY INTRODUCTION

During the past seven years, the Case Company has been devoted to developing a home-made ice cream type of dairy product where their key strength is the use of natural fruits to provide a more unique and natural flavor to their products. The company started operations in 2013 in the city of Lima, Peru where the two main stakeholders, the CEO, and the Executive Chef, decided to offer to the Peruvian public, a different ice-cream option than those in the market.

Today, the company continues distributing its products only in the capital region of Peru and it has achieved different milestones along the way. An example of this is the company's headcount that has grown to up to 30-35 employees, allocating it in the range of the small and medium-sized companies.

The company has also grown its number of available products to up to 237 products in a year, however, this number varies. As a seasonal SME that appreciates during the summertime, their best-selling products are of the product line of ice milk products such as gelatos and sorbets. However, the company has successfully developed different kinds of products such as hot drinks, healthy snacks, pastries, and their most recent incorporation snacks for dogs.

3.1 Current Problems in the Case Company

Despite the potential behind the company and its vast products lines, there are five main problems that are affecting the company where three of them, have been constantly watching over the business in all its operational years, and one of them is an external problem that the managers successfully addressed. The four main issues of the company are as follows:

- **Strategy Issues:** The current managerial performance of the company has not been successful in how the company has addressed internal threatens. During the past seven years, the case company has experienced the same business trends year over year. It is expected, though, for a seasonal company to take off during its peak season and plummet otherwise, however, the strategic management process has not been optimized during the operative years.

- **Seasonal Sales Plummet:** During the last seven years, the company has experienced a constant plunge that happens from March to April in their yearly sales revenue. See Picture 3 in Appendix 2. Contrarily to the European seasons, summertime in Peru goes from January to March and therefore, such a trend is a predictable and natural behavior of a seasonal company. However, it is of the company's interest to lessen the consequences of an expected performance as the seasonal shift plunge highly affects the company's sales. Thus far, the company has not developed nor implemented any strategy to reverse the sales plummet and the company's performance during this specific period has remained the same during the past seven years.

- **Data Management:** the company currently works under the methodology of storing their data in spreadsheet software environments, such as MS Excel or Google Sheets. This practice becomes an issue when the data stored is manipulated in the wrong manner, for instance, the company uses excessive formatting that overlays the data and does not provide insightful meaning to the users. This practice does not only affect how sales and products data is consumed but it is also implemented for bookkeeping and accounting which leads to the accounting processes to become inaccurate and unreliable. In the same way, data is not treated under any Business Intelligence procedure and the company has experienced trouble when files have been corrupted, missed, or important information has been deleted.

- **COVID-19 Pandemic:** at the end of the first quarter of 2020, external factors affected the company's first semester of operations, which inevitably forced them to set new strategies. Concretely, the COVID-19 pandemic affected the Case Company's sales and in March 2020, managers opened a new sales channel via social media platforms to avoid the huge losses as their main and physical selling points were shut down during the outbreak. The impact of this new distribution channel named "Delivery Channel" was of the managers' interest to be analyzed and explained.

It is of the company's own interest to be aware of how the Management Department perceives external factors affecting the company. Strategies must be in constant change in response to internal and external factors. That is the reason why the usage of market analytical tools (e.g., SWOT or PEST analysis) is fundamental to revise the key drivers of a strategy, in addition to performance measurement, and taking corrective action when needed (David et al. 2016, 34.) In the same way the pandemic was addressed, it is of special interest for the company's managers to address the seasonal revenue plummet as well, however, managers do not consider this seasonal shift as an event the company can fight against and is considered to be something inevitable that will affect the sales year over year instead.

4 DEVELOPMENT PLAN OF THE BUSINESS INTELLIGENCE PROCESS FOR THE CASE COMPANY

From June 2020 to September 2020, the thesis project was developed in different stages that supported certain priorities for the company. It was of special interest for the managers to understand their data, concretely, the managers were interested in analyzing aspects of their sales numbers in both units and monetary facts. Additionally, aspects related to the improvement of the company's data management were proposed to the user and the company's overall data strategy was assessed as well. The stages of the developed work happened as follows:

1. **Overall Company's Assessment –** in the format of an online videoconference.
    a. Date: June 2020
2. **Data Collection** – sources collected monthly in the format of .xlsx files distributed by email.
    a. Date: June, July, August, September 2020
3. **Recreating the Business Intelligence Process**
    a. BI outputs monthly delivered to the user.
        i. Date: June, July, August, September 2020
    b. Monthly evaluations of the user's output experience in the format of videoconferences.
        i. Date: June, July, August, September 2020
4. **Questionnaires Assessment –** the impact of the BI process was evaluated twice in the format of online questionnaires.
    a. Date: July, September 2020
5. **Questionnaires Results Delivered to the User** – final results were communicated to the managers in the format of video conference.
    a. Date: September 2020

4.1 Step 1 of the Plan: Overall Company's Assessment

The first company's assessment was supported by the qualitative decision-making tool of Problems and Objectives Tree, which backs up the analysis of a decision-making process within an SME. During the first videoconference, the Problems and Objectives Trees supported the analysis of both positive and negative causes and consequences of one of the issues of the Case Company. See Picture 4 in Appendix 3.

In conjuction with the company's manager, the Problems Tree assisted in analyzing what was hiding underneath the company's main struggle which is the sales plummet after the seasonal shift. The Problems Tree also assisted in defining what hidden factors were causing the main issue. Lastly, this tool provided a scope of the impact of those causes when they hit the main problem. See Picture 4 in Appendix 3.

During the first company's assessment, the Objectives Tree provided a different perspective of the main problem, that was: a positive side beginning from the same starting point as in the Problem Tree. See Picture 5 in Appendix 3. The Objectives Tree provides an overview of how every problem (middle), cause (bottom), and consequence (top) is transformed into a solution. Solutions are connected step by step, from bottom to top and the purpose of those links is to achieve affordable relations that bring a final solution-outcome for the sales plummet after the seasonal shift issue, a fact communicated to the company's manager. See Picture 5 in Appendix 3.

During the first assessment, it was discussed how decisions can be split into two categories, those that are previously arranged and those that are not. The company's main problem of the revenue plunge after a certain period fits well into the category of arranged or scheduled decisions that can be set up, routinized, programmed, etc. prior to becoming a fact (Victoria 2019).

In addition to the aforementioned issues, several other company's details were assessed within the first video conference that included a revision of the company's strategy for the second half of the year 2020, which begun in August and consisted of the company's performance and business plan review. It was also agreed what was the most urgent data for the managers to analyze, being this the historical sales numbers but as a merely informative process only.

It was also planned that the company would use an analysis of how the external environment factors (e.g., COVID-19 pandemic) had positively/negatively affected the company's sales. Continuing with the agreement on what data to analyze, the company had a keen interest in knowing more details about their sales performance within different distribution channels, locations, and customers (e.g., wholesalers). Other areas of interest for the company were to know the sales behavior within marketing aspects such as sales by brands and type of customers. Lastly, during the video conference the current data management methodologies in the company were assessed as well as which practices were used to store information.

4.2 Step 2 of the Plan: Data Collection

The BI process developed for the Case Company collected the data from MS Excel. During the four months of the duration of the BI development process, different files were delivered by the company's General Director's assistant and approximately four to six batches of data were received. These batches included data about the company's yearly sales in units and currency, sales from the company's new sales channel named "Delivery", contribution margins, and invoices information. See Picture 6 in Appendix 3.

4.3 Step 3 of the Plan: Recreating the Business Intelligence Process

In this project, specialized software has been used to recreate the BI process from scratch. To begin with, as the information delivered by the case company was sent in the format of .xlsx files corresponding to MS Excel, this software was used to perform all the data cleansing processes. Later on,

a mockup database was generated using both SQL and the platform SQL Server Management Studio (SSMS). When developing databases, there were several SQL frameworks that could be used, e.g., MySQL. However, SQL was employed as it is the standard language used in the database development industry and the same approach this project followed.

To continue with the recreation of the BI process, after carrying out the data cleansing stage, data visualization continued. The aim of this process was to provide an interface that the user could actually see and interact with in order to analyze their data. In this area, there is a vast portfolio of specialized reporting and analytics tools available in the market, however, this project employed MS Power BI as its demand within companies increases all the time and has become a standard for performing BI tasks, such as decision-making.

After the four months of the recreation of the BI process, the MS Power BI Business Intelligence reports were combined with DAX-based formulas in order to perform mathematical operations that could provide specific calculations of interest for the company's managers.

Finally, a consultancy analysis of the results and a user-training were constantly provided to the final user during the whole project implementation.

4.4 Step 4 of the Plan: Questionnaire Assessment

Different aspects were evaluated after the development, delivery, and usage of the BI reports. The user had the task to evaluate, in the form of a questionnaire, the areas of data awareness, importance, usage & design, as well as future BI implementations. Two questionnaires were delivered during the project process and two reports were evaluated in each questionnaire which were responded by a sample population of six respondents from different areas of the company, who had to answer general questions about the BI implementation and specific questions for each of the four reports.

Each questionnaire had a structure of one brief introduction, 21 questions, 2 visuals referring to the reports being evaluated, and acknowledgment. Within the 21 questions, 1 question was of the type of multiple-choice, 17 questions were of the type of single selection, 1 question was the type of 1 to 4 Likert scale, and 2 questions were the type of 1 to 5 Likert scale.

The questionnaires were sent twice to the company during the four months of the project implementation.

4.5 Step 5 of the Plan: Questionnaires' Results Delivered to the User

After the implementation of the BI process and its evaluation, there was a gap of two weeks were the questionnaires' results were analyzed and further delivered to the user. See Picture 7, 8, and 9 in Appendix 1. The results included different assessments, such as:

- The operative role of the respondents.
- The percentage of the respondents who considered that the company's data is important in the decision-making process.
- The importance of data visualization amongst the departments.
- How much the respondents used the four reports. See Picture A in Appendix 1.

5 BI DEVELOPED WORK USING EXCEL AS DATASOURCE

Prior to this research, the Case Company had not implemented any BI systems or processes, therefore, using MS Excel was the most feasible solution to store and manage the company's data. However, the data management processes did not match the criteria of the BI methodology and as earlier stated, the data must meet the conditional of complying with the same characteristics and be related amongst each data point.

As found in the company's data source (MS Excel file) for the development of the first BI report, the company has tried to create relationships between the data points by classifying the information of their products according to their categories, brands, sizes, presentations, and product type. Concretely, the company has made the following classification for their products' information:

1) **Product Type**:
    a. Personal Containers (original: *envases personales*).
    b. Other Presentations (original: *otras presentaciones*).
    c. Distribution Products (original: *productos distribucion*).
    d. Discontinued Products (original: *productos descontinuados).*

However, the company's file presents a discrepancy at the moment of using colors to highlight specific data. Within the data source, the company has used the same colors to highlight information related to a brand as well as the type of container. The classification of the data has been mixed, which complicates how the user understands the data points. See Picture 10 in Appendix 4. The source of information does not contain the data in a uniform way and these data must be first processed before transposing it to other integrated systems.

5.1 Stage 1 of the BI process: Data Cleansing.

To begin transforming the data into a unified form, it is necessary to understand how the relationships within the individual data points work. Firstly, the data source provided by Case Company has been defined as to be: Monthly Sales of Products in Units. Recalling the definition of dimensions and facts, a dimension refers to informative assets, e.g., the attributes of a product; thus, a fact refers to a numerical value that is related to the dimension, e.g., sales. Therefore, the Monthly Sales will be treated as the fact and the Products will be treated as the dimension.

In order to clean and organize the data provided by the company for developing their first BI process, we need to start building relationships between the data points. Everything that acts as an attribute or characteristic of the product will be affecting it. The first dimension will be the product name itself and consequent dimensions, will be every characteristic that impacts the product. Every product in the company's Excel file will be taken separately despite of any present attributes, as applying other characteristics (dimensions) would increase the level of detail and no granularity needs to be applied yet.

The data cleansing process can be performed in a separate MS Excel worksheet where the transition of the raw data to the cleaned data happens in the extraction of the information from a POINT 11 to a POINT 12. See Picture 11 and 12 in Appendix 4. In this transition, individual data points need to be ordered by following a semantical structure, this means that contrarily to what the user originally did with the data, the data cleansing process will strictly collect only data points that are related, e.g., products' names only instead of combining products' names with brands' names. In total, the company has registers for 237 products and all those products must be separated into one single column of the worksheet while ignoring any secondary attributes previously given. See Picture 13 in Appendix 4.

To continue organizing the information, the user must consider what is the level of granularity required, in other words, how deep is the analysis required by the final user or decision-maker. It is a good practice to define the granularity, however, the level of granularity will not always be specified by the user and it must be imperatively but mandatorily mined.

During the first BI report developed for the Case Company in this project, the company's managers were not yet aware of what kind of information was the most important for them to analyze and therefore, the level of detail applied to this first BI report was mined from the data source.

As previously stated, the company attempted to classify their data by type of product, brand, presentation, and category, that is how every product, previously cleaned up, must be matched with its own attributes. This process takes now into consideration what are the secondary attributes affecting the dimension, in this case Products. See Picture 14 and 15 in Appendix 4. The most important step is to give a structure to the data and apply certain level of detail by classifying the data into their corresponding categories, e.g., a product name with its type of product, brand, presentation, and category. See Picture 14 and 15 in Appendix 4. Likewise, the same process must be done for all the products involved, 237 products in total. See Picture 16 in Appendix 4.

Despite the classification of the 237 products (see Picture 16 in Appendix 4), an extra column named "ID_PRODUCT" must be added. The function of this column is to assign a certain number to a specific product and that number cannot be repeated because each product should be unique, with its unique characteristics.

To continue with cleaning the data, the Sales fact data need to be organized in the same way as the Product dimension was done and certain details must be considered beforehand:

- The Case Company's information in the data source file covers the sales from the period of January 2018 to May 2020, without considering individual days.
    - Meaning that there are sales records for two years (2018 & 2019) plus the first five months of the third year (up to May 2020), in total **29** months.

- Each product (**237** in total) had a sale entry every month whether if the sale was >=$0.
    - Therefore, there are **6873** sales entries of each month by each product (**29** * **237**).

To organize the Sales fact, we must consider what are the elements affecting the sales:

1. the number of the sale whether if the sale was the first (#1), second (#2) or last sale (#6873).
2. The number of units sold in each sale.
3. The product sold in each sale.
4. Additionally, the quarter when the sale happened.
    - This extra criterion will add another level of detail to the analysis of the company's performance.

For instance, FIGURE 9 reflects how the elements affecting the sales are being identified from the original data source.

1. The first product (sale #1) was sold in January (original: *Ene*) 2018.
2. The first product sold was **Strawberry with Mango** (original: *fresa con mango*).
3. During January it sold 1782 units.
4. January is within the first quarter of the year.

Ventas Mensuales por Producto - UNIDADES							2,018									
	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	TOTAL	ENE	FEB	MA
ENVASES PERSONALES	13,759	9,609	14,710	8,408	8,029	4,992	6,203	4,743	8,404	8,198	9,464	9,804	106,323	17,804	14,438	11,2
PALETTI ENVASADAS	13,759	9,609	14,710	8,408	8,029	4,992	6,203	4,743	8,404	8,198	9,464	9,372	105,891	16,946	12,226	8,6
FRUTAS	2,036	1,727	2,346	1,266	1,164	606	856	590	1,096	1,070	1,438	1,238	15,623	2,854	2,000	1,2
FRESA CON MANGO	1,782	1,139	1,650	918	828	510	712	428	764	823	1,061	1,088	11,703	1,840	1,526	78
FRAMBUESA	372	426	396	216	228	54	78	102	208	169	197	150	2,596	516	330	24
KIWI CON FRAMBUESA Y MANDARINA JUMBO	72	162	300	132	108	42	66	60	124	78	180		1,324	498	144	18

FIGURE 9. Analysis of how to classify a product's sales within a period

Similarly, as with the cleansing process for the dimension, the data point must be separately organized. See Picture 17 in Appendix 4. Likewise, every sale entry (6873 in total) must be organized and simplified in a separate worksheet. See Picture 18 in Appendix 4.

5.2 Stage 2 of the BI process: Creating Relations.

One of the most time-consuming tasks in the BI process is the cleansing process, however, even though each sale entry has been organized and cleaned, the products and sales data would not function without a connection (Vercellis 2009, 6). In other words, products have been classified and sales haven organized but there is nothing that could tell what product corresponds to what sale, or the detail of each sale, and to solve this issue, we must apply the relationships between tables.

As earlier mentioned, the extra column added named ID_PRODUCT (or ID_PROD) is a unique identifier that cannot be repeated, each product is unique. On the other hand, a sale can be repeated for product A and product B, or products A, B, and C, etc., as a sale may contain many products.

To relate then the Facts with the Dimensions, there must exist a relationship. FIGURE 10 displays how the Sales fact table looks like and how the Product dimension table looks like without any relationship yet.

13	**SalesFact**		**ProductDim**
14	ID_SALE		ID_PROD
15	UNITS		NAME
16	PERIOD		TYPE
17			BRAND
18			PRESENTATION
19			CATEGORY

FIGURE 10. Example of a fact and dimension table

Now, each table is independent of each other but once we connect ID_PROD (ID_PRODUCT) from ProductDim table (Dim for dimension) to the SalesFact table, the relationship will be created, as shown in FIGURE 11.

	A	B	C
1	**SalesFact**		**ProductDim**
2	ID_SALE		ID_PROD
3	ID_PROD		NAME
4	UNITS		TYPE
5	PERIOD		BRAND
6			PRESENTATION
7			CATEGORY

FIGURE 11. Example of a relationship between a fact and a dimension table

In practice, what FIGURE 11 means is that each sale will be related to a product, that could be between 1 and 237. See Picture 19 in Appendix 4. In this way, we would have created a relational database that with following this practice, the data is simplified, and it is easier for a BI integration system, such as MS Power BI to understand it and process it. Additionally, it is important to keep every fact and dimension table in a separate worksheet. See Picture 20 in Appendix 4.

As aforementioned, data cleansing is a time-consuming process, and this is a fact that may concern and limit an SME from integrating BI processes within their daily operations however, this is not the

only way to maintain information organized, structured, and related but the process becomes necessary when the user mainly stores its information within multiple spreadsheets environments. It is, therefore, how creating a database is the most reliable form to maintain perfectly organized business data without the size-limit concerns, or without incurring the risk of broken and damaged MS Excel files. Nevertheless, the information cleaned, structured, and organized during this section is ready to be integrated into a BI system, in this case, MS Power BI.

5.3 Stage 3 of the BI process: Power Query.

Supported by Microsoft (s.a.), Power Query allows the user to connect, combine, refine, and manage data according to the user's needs. Power Query is one step before creating visualizations to support decision-making.

5.3.1 Connecting Data to Power Query

To connect data to the Power Query environment, the next process must be followed:

1. When initiating MS Power BI, a welcome screen will appear, and the user would have to click on Get Data. See Picture 21 in Appendix 4.

2. From the Get Data screen, there are several data sources that can be connected to MS Power BI, e.g., MS Excel files, plain text/CSV files, databases, PDF, etc. As in this case the data has been stored in an MS Excel file, click Excel>Connect. See Picture 22 in Appendix 4.

3. Once it has been chosen to connect to a MS Excel file type, we need to choose the file path.

4. Once the MS Excel file is connected, MS Power BI will display the Navigator screen where it is possible to choose which worksheets will be included in the report. See Picture 23 in Appendix 4.

5. Once the necessary worksheets have been selected, there are two options available for creating a visual report: 1) Load or 2) Transform Data. Loading the information will immediately put all the worksheets available to start visualizing the data. Transform Data would, on the other hand, open the Power Query Editor where the data can be further managed. For this BI report, the data need further transformation than what was done in MS Excel previously, therefore we will choose Transform Data. See Picture 23 in Appendix 4.

5.3.2 Power Query Editor Interface

Supported by Microsoft (2019), the Query Editor is distributed in four sections (see Picture 24 in Appendix 4):

- **Query Ribbon.** Contains five tabs Home, Transform, Add Column, View, and Help. Each of those tabs contain several buttons that allow the user to manage and interact with the data.
- **Queries Pane.** contains all the queries (or in this case worksheets) to interact with. See
- **Center Pane.** Is where the data is displayed.
- **Query Settings Pane.** Displays every step that has been applied to the data as well as its properties.

5.3.3 Editing Data with Power Query

As the data has been cleaned and organized in previous stages, only minor changes will be applied using the Query Editor. For instance, in the PRODUCTS (original: *products)* worksheet, the data in columns "NAME", "TYPE", "BRAND", "PRESENTATION", and "CATEGORY" should be transformed in order to show each word capitalized. This transition can be done by accessing in the Query Ribbon *Transform>Text Colum> Format>Capitalize Each Word*. See Picture 25 in Appendix 4.

After the formatting has been applied, each word would be capitalized in a question of only a few clicks. Similarly, there are different kinds of formatting, calculation, and data mining options that can be achieved by using the Power Query Editor.

To conclude the use of the Power Query Editor, click on Home>Close & Apply, to save all the changes done and continue to report the data.

5.4 Stage 4 of the BI process: Data Visualization.

Thus far, the data has been cleaned, structured, and organized in MS Excel and in Power Query, and the next step would be to start creating visualizations.

5.4.1 MS Power BI Reporting Interface

Supported by Microsoft (2020), there are six main sections (see Picture 26 in Appendix 4) in the reporting view of MS Power BI:

- **Section 1.** Displays a top ribbon with tasks that are associated with reporting and visualizing the data.
- **Section 2.** Is the middle canvas where all the visualizations are developed.
- **Section 3.** Allows the user to create or navigate pages.

- **Section 4.** Is where visualizations can be filtered.
- **Section 5.** Refers to the Visualization Pane where visuals can be added to the canvas, changed, modified, or customized.
- **Section 6: Fields pane.** Shows which are the available fields in the queries (in this case worksheets). Fields can be dragged and dropped onto the canvas to create or modify visualizations.

5.4.2 Reporting with the Adequate Chart Type

The core aim of visualizing data is to graphically represent information in such a way that the visualization methods chosen communicate the information effectively. In order to do that, the user must understand what questions need to be answered as is important to choose the right type of chart.

In the example of **Monthly Sales by Products in Units**, the user might be interested in knowing which products have sold the most. The user needs to consider which type of charts are the most adequate to display such data. For instance, displaying the sales by-products in units using a pie chart would display the information accurately but it becomes difficult to interpret. See Picture 27 in Appendix 4. This is due to the fact that pie charts are most effective for displaying a small number of categories (Lane, s.a.), and in this case, visualizing the sales of each product means visualizing the 237 products the company works with.

Contrarily, visualizing a large number of observations could be better achieved by using bar charts, which show a more intuitive distribution of the information compared to the pie chart. See Picture 28 in Appendix 4.

5.4.3 Visualizing Data with Power BI

For this first report, the questions that the user needed to answer were as follows:

1. User is interested in knowing the sales (in units) per year, quarter, and month.
2. Most/least selling products by brand.
3. What is the percentage of units sold by product presentation?
4. How many units were sold by product category and product type?

To answer the **first question**, we would need to combine the fact (numerical value) of units, with the dimension that must affect the units, in this case, the year. So, to start visualizing the data in the MS Power BI reporting canvas, it is needed to:

1. Select a Clustered Column Chart from the Visualization pane, a blank chart will be positioned onto the blank canvas. See Picture 29 in Appendix 4.

2. From the Fields pane, both the UNITS SOLD, and the YEAR (original: *año)* need to be dragged to the "Values" and "Axis" sections, respectively. See Picture 30 in Appendix 4. This will segment the units sold by the year and display it in a bar chart visualization, that can be later formatted (for color, shapes, etc.) by accessing the "paint roller" icon above the "Axis" section.

The **second question** involves the most/least selling products in units by brand, so it is needed to combine the UNITS fact and the BRAND dimension. As there are not many brands in these data, information could be visualized with either a bar or a pie chart and could be created in the same way as in the example in Picture 30 in Appendix 4.

The **third question** to solve is: "what is the percentage of units sold by product presentation?" This is approached similarly to the second question, and in this case, a pie chart is useful for visualizing percentages. See Picture 31 in Appendix 4.

Finally, the **fourth question** refers to "how many units were sold by product category and product type?" This is visualized by using a funnel chart. A funnel chart supports visualizing data in different stages and when is known that the first level of the funnel is larger than the rest of the levels. See Picture 31 in Appendix 4.

In this way, raw data has been transformed into useful visualizations that can immediately transmit information and convert it into knowledge.

6 BI DEVELOPED WORK USING A DATABASE AS DATA SOURCE

As earlier mentioned, cleaning and shaping information is the most tedious and time-consuming part of the BI process, however, when the information is been stored in a database, the BI process runs differently in such a way that the data only needs to be updated as needed but do not require any further cleansing processes anymore.

To develop a second BI report, the Case Company would like to move their following .xlsx dataset "Delivery Report for June" (*Original: "Reporte Delivery Junio"*) to a database. The file contains information about their new sales channel named "Delivery", sales channel that came into operations due to external factors that negatively impacted the company's traditional operations. The file contains raw data that need to go through the cleansing process but despite it, is a process that needs to be done only once even if future data needs to be added.

The differences between creating a visual report by connecting an MS Excel data file or a database to MS Power BI are that data entered into a database will always follow the same structure and avoids users from adding any extra formatting that could harm the original file and require, in the long-run, extra cleansing.

6.1 Stage 1 of the BI process: Data Cleansing for a Database

The process of cleaning the information happens similarly to the one previously done for an MS Excel-MS Power BI connection. In the raw data workbook "Delivery Report for June" (see Picture 32 in Appendix 5) there are three worksheets that the user has defined:

1. A worksheet for the clients' details.
2. A worksheet for the products' details
3. A worksheet for the billing information.

As the user has stored information in different worksheets, the user is aware that there is a need for some sort of datum able to link the data across worksheets. For that reason, in the **first worksheet**, the user has automatically created a unique identifier or ID for every customer, as every customer is a unique identity that cannot be repeated, in this case the unique identifier is "ID_NAME". See Picture 32 in Appendix 5. It is possible to relate the information across worksheets by creating a unique identifier in a specific worksheet that relates to another.

However, and despite the ID relationship, there are three fields within the clients' details worksheet that can be answered in a different format with "yes" (original: *si)* and "no" *(*original: *no)*. These questions are if the client has any disease (e.g., diabetes), if the client owns pets (as the company has a dedicated product line for pets), and if the client has children. See Picture 32 in Appendix 5. These questions, with exception of the disease condition question, are of the data type BIT, which can accept values of 1, 0, or null (Microsoft 2017).

In the **second worksheet**, the user has stored information about the products and is interesting to see how the user has intuitively related each client's name (ID_NAME) with their correspondent sale number (ID_SALE). The user does understand that without this relationship, it would be impossible to know what the clients bought, or which sale corresponds to which client. See Picture 33 in Appendix 5.

Moreover, the **third worksheet** contains information about the billing data where the user has once more, connected each sale number (ID_SALE) with the client's unique identifier (ID_NAME). The user has specified different payment methods:

- Transfer (original: *transferencia*).
- Yape (*mobile app for transfers*).
- Not identified.
- Cash (original: *efectivo*).
- Deposit (*deposito*).
- Payment against delivery (original: *contra entrega*).

In addition to the ID_SALE ↔ ID_NAME connection, the connections made by the user match the date when the sale took place, the amount (though not always tabulated as seen in cells E5, E7, etc.), and the payment method. See Picture 34 in Appendix 5.

6.1.1 Assigning Unique Identifiers to Support Data Cleansing for a Database

For the **first worksheet** and in the same way the user has given unique identifiers to the data points to clean the information, it is necessary to assign unique identifiers to the rest of the sections that affect the main dimension: products. For instance, unique identifiers need to be given to each payment method, from 1 to 6 for the six types of payment methods. These unique identifiers must be also set for all the number of Districts, Brands, and Presentations. See Picture 35 in Appendix 5.

Once each catalog has been identified with a unique ID (see Picture 35 in Appendix 5), the data points must be organized according to their categories and IDs. For instance, each client must be related with his/her own district already categorized by a unique number. See Picture 36 in Appendix 5.

As earlier mentioned, the first worksheet contains three open questions (see Picture 37 in Appendix 5), whether the client has an illness, pets, or children. The reason is that concretely, the user is interested in knowing if the client has an illness condition or not, and if yes, which one, e.g., diabetes (original: *diabetico)*. In case the client does not have any, the answer will be *NO*, in case the client does have an illness, the answer will be the name of the illness. Therefore, this data type cannot be understood as `BIT`, as there is still the need to input the name of the illness using plain text, which is a characteristic of the data type `VARCHAR`.

However, the situation for the two last questions about pets and children is a situation of simply answering 'Yes' or 'No'. This can be understood as using only two options to answer the question, 0 for 'No' and 1 for 'Yes', a characteristic of the `BIT` data type. (See Picture 37 in Appendix 5).

For the **second worksheet** each product, brand name, and presentation need to be classified uniquely, e.g., there cannot be "two **products** for **chocolate ice cream** of **brand Paletti** in the **presentation** of **473ml**" and therefore each of these categories must have a unique ID. (See Picture 38 in Appendix 5).

The second worksheet contains an extra column named "Sampling", where the user wants to identify those sales that included a free sample. (See Picture 38 in Appendix 5). This data type is a `BIT` type as the answer will be 0 for 'Yes' or 1 for 'No'. Also, within the second worksheet the user has failed to introduce the unit price for each individual product which will later create a problem of granularity. (See Picture 38 in Appendix 5).

Finally, in the **third worksheet**, the user has intuitively related each client's details with each sale's details by connecting IDs. (See Picture 39 in Appendix 5). It is noticeable too that some total amounts are missing.

As earlier mentioned, the second worksheet failed to display the unit price of each product and as seen in the third worksheet (see Picture 39 in Appendix 5), only total amounts are being shown, suggesting a low level of details, e.g., in the case the user would need to compare prices at a deeper level, it will not be possible as the unit prices are missing.

6.2 Stage 2 of the BI process: Relations and Tables

It would not be possible to create a database following the form the user has originally allocated the data. The reason for this is simple: a relational database must be linked through relationships across tables. As stated in previous chapters, there are facts and dimension tables and therefore, the information in this dataset must be stored following the same principles. Facts must contain their numeric values and their attributes connected to dimensions, that are thus connected to their own attributes. (See Picture 40 in Appendix 5). In other words, we must first create tables for each fact and dimension, in total ten tables will be created, and link them between each other.

6.2.1 Creating Relations Across Tables

For table 1 (see Picture 40 in Appendix 5), a catalog for the clients' DISTRICT has been created by storing the unique `ID` as an `INT` data type, (which will be always the case for the IDs). The name attribute has been defined as a `VARCHAR` data type, as the data will be stored as plain text, with the capacity to store as many characters as indicated between parentheses. The same logic applies to tables 2 to 6. (See Picture 40 in Appendix 5)

Table 7 (see Picture 40 in Appendix 5) uses the same logic with the exception that it also stores another ID (ID_BRAND) stored as an `INT`. Below ID_BRAND there is an "FK" abbreviation that stands for Foreign Key. A Foreign Key is indicating that the ID containing "FK" acts as the main ID of a foreign table and is also what connects the parent table (or Primary Key, PK) to the child table. Therefore, ID_BRAND in table 6 is referencing its parent table, table 6. Table 7 also contains the products' prices in `DECIMAL` format and the stock of the products as an `INT`.

In table 8 (see Picture 40 in Appendix 5), the elements NAME, ADDRESS, and TELEPHONE are stored as `VARCHAR` with a character-length indicated in their correspondent parentheses. ID_DISTRICT and ID_CONDITION are FKs referring to their parent tables' PKs. Table 8 also contains the `BIT` data type elements, PET and CHILDREN, that can only contain true or false values of 0 or 1.

Table 9 (see Picture 40 in Appendix 5) is a table that refers to or takes its information from elements using FKs to extract data from other tables, ID_DETAIL, ID_CLIENT, ID_PMTMTD (payment method). The only elements locally stored in table 9 are data for SUBTOTAL, TAX, and TOTAL.

Finally, table 10 (see Picture 40 in Appendix 5) follows the same logic as table 9 by collecting its information from foreign tables, and the only elements locally stored are the elements of QUANTITY as an `INT` and SAMPLING as a `BIT` data type.

6.3 Stage 3 of the BI process: Creating a Database in SQL Server Management Studio

It is possible to create a database when the user has understood the relationships between data points without the concern whether if the database would run or not. To create a database there are some lines of code needed.

We will start building the database by first creating the dimension tables with fewer elements, such as tables 1 to 6 (see Picture 40 in Appendix 5), as they all follow the same logic. On the SSMS canvas (see Picture 41 in Appendix 5) we will:

1. **Insert use the following line of code:**

```
CREATE TABLE DISTRICT
(
ID INT IDENTITY(1,1) PRIMARY KEY,
NAME VARCHAR(50) NOT NULL
)
```

Where `CREATE TABLE` is the command to create the dimension table with name the `DISTRICT` after the first parenthesis, `ID,` and `NAME` refer to the name of the elements inside the table.

`ID` is followed by its data type `INT` and `IDENTITY(1,1)` sets an automatic auto increasing ID count, e.g. ID 1, ID 2, ID 3, etc., where (1,1) refers to "start in number 1 and increment by 1". After that,

the sentence `PRIMARY KEY` indicates the type of key the ID represents, there can be only one PK in each table, but as many FKs as needed.

After the table `NAME` there is the data type `VARCHAR` followed by the length of the permitted characters of `50` characters per register, at last, the indicator of `NOT NULL` forces the user to enter a value for `NAME`, otherwise, the database will show an error.

Highlight the code `+ F5 (Execute)` to run it, now the table for `DISTRICT` has been created.

2. The attributes of each element need to be inputted.

There is no need to input each individual ID, as the command `IDENTITY(1,1)` will generate these automatically for each of the registers inputted. However, the name of the districts of the Case Company needs to be inputted.

The 21 districts will be inputted by using the following line of code:

```
INSERT INTO DISTRICT (NAME) VALUES
('SAN ISIDRO'),
('SURCO'),
('SURQUILLO'),
('MIRAFLORES'),
('BARRANCO'),
('LA MOLINA'),
('PUEBLO LIBRE'),
('SAN BORJA'),
('MAGDALENA'),
('LINCE'),
('JESUS MARIA'),
('SAN MIGUEL'),
('SAN LUIS'),
('CHORRILLOS'),
('LA VICTORIA'),
('ATE'),
('VILLA MARIA DEL TRIUNFO'),
('LOS OLIVOS'),
('SAN JUAN DE MIRAFLORES'),
('CALLAO'),
('SAN MARTIN DE PORRES')
```

Where `INSERT INTO` is the command that inserts into the table `DISTRICT`, under the element `NAME` the `VALUES` corresponding to the name of each district, following the syntaxis logic of the code above, highlight the code `+ F5 (Execute)` to run it.
After this, the table for the Districts had been successfully created and a form to confirm it is by writing the following code, (where `SELECT` commands to select elements and `*` picks all the elements from the table `DISTRICT`),

```
SELECT * FROM DISTRICT
```

highlighting it and pressing `F5` (or Execute) to visualize the table, (see Picture 42 in Appendix 5), where DISTRITO = *DISTRICT* and NOMBRE_DSTRT= *DISTRICT NAME*.

3. **The process to create tables 2 to 6 is exactly the same and the only difference will be the names of the tables.**

Tables 2 to 6 can be created in a separate SSMS sheet or in the same canvas used before.

4. **To create table 7, we will introduce the following code onto the canvas:**

```
CREATE TABLE PRODUCT
(
            ID INT IDENTITY(1,1) PRIMARY KEY,
            NAME VARCHAR(50) NOT NULL,
            ID_BRAND INT NOT NULL,
            STOCK INT NULL,
            PRICE DECIMAL(10,2) NULL
)
```

highlight the code `+ F5` (`Execute`) to run it.

Table 7 PRODUCT (see Picture 41 in Appendix 5) has five elements:

1) ID type `INT` with an auto increasing `IDENTITY (1,1)`, set as primary key.

2) `NAME` (of the product) as a `VARCHAR` type with capacity for 50 characters, which must contain a value (`NOT NULL`).

3) `ID_BRAND` which is a foreign key that refers to another table (parent table 6), is of type `INT` and cannot be null (`NOT NULL`).

4) `STOCK`, which is type `INT` and can be `NULL`, and

5) `PRICE` that is the type of `DECIMAL (10,2)` and can be `NULL`.

As `STOCK` and `PRICE` can be null, we will not input those values now, as those values are not delivered by the company.

5. **Connecting Tables.**

To relate tables, we need to connect foreign keys within their host tables, to their respective parent tables. This can be done by running the following code:

```
ALTER TABLE PRODUCT
ADD CONSTRAINT ID_BRAND_PD FOREIGN KEY (ID_BRAND_PD)
REFERENCES BRAND (ID)
```

After table 7 "PRODUCT" has been created, we need to `ALTER` the `TABLE PRODUCT` and `ADD` the `CONSTRAINT` of saying that `ID_BRAND_PD` (*ID product brand*) is a `FOREIGN KEY (ID_BRAND_PD)` that `REFERENCES` the parent table `BRAND` and connects to its `ID` element.

What we are doing with this, is to connect one foreign ID to its corresponding parent ID.

6. **Next, the values of the PRODUCT table need to be entered, by running the following code.**

Note that due to the extension of the values, this instance code will only introduce three records corresponding to the name of the product and the ID of the brand. Notice also that the price nor the stock values are being entered, as the data are missing, and both can be `NULL`.

```
INSERT INTO PRODUCT VALUES
('AFTER EIGHT MENTA',  3),
('CHOCOLATE BITTER',   3),
('CHOCOLATE LIGHT',    3)
```

Finally, entering, and highlighting `+ F5` (`Execute`) the following code, will return the table PRODUCTS with all the products' names that had just been created. (See Picture 43 in Appendix 5).

7. **Table 8 is created in the same way as table 7.**

It is important to connect every FK with its PK after creating the table, otherwise, the relations will not exist, and the DB will not function properly.

The only new element in table 8 is the `BIT` data type for both PET and CHILDREN elements. These `BIT` type elements can be inputted by introducing the next code within the rest of the `CREATE TABLE` line of code:

```
PET BIT NULL,
CHILDREN BIT NULL
```

8. **Table 9 SALES and 10 SALES_DETAILS can be created following the same guidelines as in table 8.**

   - When all the tables are created and related using FKs, the user can retrieve the table 9 SALES with the command line `SELECT * FROM SALES` and visualize it. (See Picture 44 in Appendix 5). Where:
     - ID_CLIENTE = *ID_CLIENT*, FECHA = *DATE*, ID_METODO_PAGO = *ID_PMTMTD (PAYMENT METHOD)*, and IMPUESTO = *TAX*.

   - In the same way, the user can visualize the table 10 SALES_DETAILS using the code `SELECT * FROM SALES_DETAILS`. (See Picture 45 in Appendix 5).
     - Where ID_VENTA = *ID_SALE*, ID_PRODUCTO = *ID_PRODUCT*, ID_PRESENTACION = *ID_PRESENTATION*, ID_DIMENSION = *ID_DIMENSION*, and CANTIDAD = *QUANTITY*.

As seen, (See Picture 44 & 45 in Appendix 5), the relationships are being created by numbers that, within their parent tables, match a description such as the product name, the district, the brand, etc., which is the purpose of the relational databases.

At first, it may look more complicated to create a DB than to handle information in Excel, but in the long-run, information stored in a database will be information that is structured and organized. Additionally, data can be allocated in the cloud, and with the help of lines of code, the user can access to only the data required without the need of retrieving the full data set. This is one of the advantages of using SQL within the BI process.

6.4 Stage 4 of the BI process: Connecting a Database to Power BI for Data Visualization

To connect a DB to MS Power BI, it is necessary to know the server name the data will be retrieved from.

1. Within the SSMS environment, we can find this name in the Object Explorer by pressing the Connect Object Explorer icon. See Picture 46 in Appendix 5.

2. Once clicking on the Connect Object Explorer, a SQL Server window will pop up indicating the server name, which will be later pasted to MS Power BI. See Picture 47 in Appendix 5.

3. The DB connects to MS Power BI by, on MS Power BI, clicking on Get Data > SQL Server. See Picture 48 in Appendix 5.

4. A window will pop up. Introduce the server name and click on OK. See Picture 49 in Appendix 5.

5. The Navigator pane will pop up and all the available DBs will show up. After selecting the right DB, click on either Load or Transform Data. See Picture 50 in Appendix 5.

6. Once the information is loaded, every table will be available to create visualizations, just as when using MS Excel. For instance, if the company would like to know how their products have sold since the introduction of their new sales channel "Delivery", we could create a clustered bar chart, (see Picture 51 in Appendix 5), and with this, the user could start visually analyzing the data having the information structured and organized, without size limitations or risks of damages/losses and with no need of future cleansing procedures.

7 CONCLUSION

During the development of this mockup BI process within a SME, it was observed that managers do undertake some non-grounded decisions. The uncertainty of those strategies relies on the absence of an in-depth analysis of the information collected.

It seems that leaders at SMEs do not give the right value to their organization's data which, as demonstrated in this paper, goes beyond functioning as just historical data. Data is complex and like an avalanche, the more data the business collects the bigger it becomes, turning it into an even more complicated asset to harness. For that reason, tabulating data into any-type of software should not be the issue SMEs should deal with but, collecting data should be an administrative task that supports more complex tasks such as decision-making.

A solution for that problem is Business Intelligence, which combines the theory and practice of different areas such as strategic management, decision-making, databases, and data visualization, amongst others. This is a process far from simple that however, goes beyond the task of simple data tabulation in spreadsheet environments. Advancing data management in businesses will lead SMEs to detect insights that cannot be seen by a bare eye that only inputs raw data to maintain records. These insights can either be human errata, non-provided information, miscalculations, missing or damaged files that harm the accuracy of the information, which does not support the idea of the effective decision-making process.

Business Intelligence requires two main elements to understand the process and perform it: knowledge, and skills. Moreover, by implementing the BI methodology, an SME would have to go into a deep understanding of sub-processes such as creating databases or generating visual reports in specialized software.

At first, all these individual processes of data cleansing, database development, data modeling, and data visualization may seem extensive and complicated but, these are procedures that act hand by hand and depend on each other to make the BI process more accurate and functional. Implementing the process for the first time could be the most complex task of the BI process itself, for instance, performing data cleansing consumes up to 80% of the time employed.

Once the BI process operates and runs, it decreases the time spent in generating strong tools and materials that are not only the base for future decisions but that provide knowledge of the company's performance. Moreover, tabulating data becomes minimal, connecting sources to create reports becomes effortless, and even fewer resources are needed.

Nonetheless and despite the benefits of the BI process, SMEs still need to deal with the biggest constrain behind the BI implementation: BI awareness. As demonstrated, 100% of the respondents acquired new and valuable information from the process that fulfills the objective of demonstrating

that Business Intelligence is valuable for the company's decision-making process. However, the percentage of usage of the developed tools was only 58%, supporting the conclusion that SMEs need to raise their awareness on the implementation of Business Intelligence to enhance operations.

In addition, overcoming the concern of transitioning to more complex working practices demands a vision beyond the three main limitations of a company when integrating complex operative systems, which are the business size, technological know-how, and budget.

The difference between managing data under limited methodologies and working under Business Intelligence relies on the strategy of the company and its vision and plans for the future. A company trying to orchestrate greater results will demand and bring in more data to the business that once effectively managed and analyzed, has the power to unravel the business potential based on an effective strategic management.

8 DISCUSSION

The overall objective of this project was to work closely with the Case Company for four months and recreate the Business Intelligence process to integrate it into the company's daily operations. Concretely, the aim was to develop four BI reports (see Pictures 51, 52, 53, and 54 in Appendix 6) using the platform MS Power BI and to create a database mockup using SQL Server. To this, the importance of developing the interactive dashboards and databases relied on improving the company's current data management processes. The user had a keen interest in implementing the BI methodology and support the company's decision-making process.

Several issues were encountered as soon as the project began with the assessment of the data management practices. Those issues were revealed throughout the entire processes and were mainly related to poor spreadsheet software practices with excessive formatting including excessive use of colors, uncategorized information, a mix of attributes, data mistakenly deleted, and missing formulas, to name a few. As the process advanced, more important issues were found specially when the consultancy advice was provided to the final user by implementing the BI dashboards, for instance and perhaps two of the most important issues encountered were:

- The company operated spreadsheets related to products and prices that contained outdated information and prices from years 2018, and in some cases, from year 2013, when the company began operations. As discussed with the final user, the Case Company is a seasonal entity that needs to fluctuate its prices and product lines, but this fact was not reflected in the spreadsheets they were using. The final user was aware that at least in practice, the prices did fluctuate which created a discrepancy with how the information was used for business operations and how the information was analyzed for internal strategies.

- Secondly, in the spreadsheets provided by the user, was observed how some products, including names and prices, were deleted with no justified reasons. It was mined that previous users deleted certain items as those items were discontinued by the company, however, this practice affected the analysis reflected in the BI dashboards as the total amounts in dashboards and spreadsheets did not match. The final user was not aware of such faulty practice, but these observations brought new insights to other managers who did not realize before that these phenomena were happening.

As discussed with the user, the reason behind the aforementioned issues is due to faulty spreadsheet practices result of a limited know-how that beyond implementing the BI methodology, must be revised, and corrected in order to rely on accurate data that could be further analyzed to support the internal decision-making process.

8.1 Data Cleansing

Spreadsheet software provides the user with unlimited tools, design options, and formatting alternatives that can be used to perform a specific task, e.g., highlight an important datum. Such a powerful and complex software as MS Excel could be, could also harm how the information is stored and consumed, specially when the user overlays the data with unnecessary practices that could be e.g. adding extra colors, adding more worksheets than needed, or mixing attributes. At the end of the day, these practices will only slow down how the company analyzes and harness their data but most important, the user will get used to these practices which will later become in an impediment to upgrade from spreadsheet environments to the BI methodology.

In order to fix the problem of unorganized, unstructured, and unshaped raw files, the data cleansing process must be performed, as done with the Case Company. The files provided by the company contained all the aforementioned constraints and each of the files was cleaned and all the data organized and unstructured.

The data cleansing process is perhaps the most time-consuming task in the BI process when is first implemented. Cleaning, organizing, and structuring the data of the company took up to 80% of the resources used in the project. Data cleansing for the case company required up to four daily hours in order to process the more than 150,000 data points and be able to develop the databases and BI dashboards.

Undoubtedly, data cleansing is the more demanding task, but it is also the main insights provider as at the moment of performing such process, the story behind the performance of the company starts arising and errors and miscalculations start appearing.

8.2 Creating Relations, Relations and Tables

After reorganizing the Case Company's datasets, the question of how to relate the information to make it work within the BI systems arises. It was demonstrated how the user related its own information as the user does understand that without it, the information would simply not connect. Linking, connecting, or relating information is the most important step before connecting the datasets to any BI environment including a database service such as SQL or even Power Pivot in Excel.

In every dataset there are facts and dimensions attributes that need to be categorized prior to attempting any connection between data points. Categorizing each of them needs an understanding of what a fact and a dimension refers to. In the case of the Case Company, a dataset about *monthly products' sales by units* would contain information about the products sold and their characteristics, but it would also contain information about the sales and how many units of each product were sold. A dimension will, therefore, contain information that gives a characteristic about something, e.g., the size, the color, the shape, etc., whilst a fact will contain information that affects the dimension, e.g., the price, the amount, the discount, etc.

When the user has understood the characteristics of both fact and dimension tables, it is needed to link each table by adding a unique identifier that could be found in both of the tables. This will automatically create a link between a fact and a dimension, and any reporting system or Excel Pivot Tables would understand.

8.3 Power Query

Similarly to MS Excel when performing Data Cleansing, Power Query is a MS Power BI in-app tool that allows the user to reshape and restructure information. For the Case Company, Power Query was used to perform smaller formatting tasks as well as calculations but the entirely cleansing process could not have been shaped using Power Query as the environment accepts connections to data sets that are already pre-cleaned.

8.4 Creating a Database in SQL Server Management Studio

As discussed with the final user, the employment of databases assures a better management of the information, however and as demonstrated in this project, the user must understand the code commands that are needed to create a database, its tables, and its relations. Transitioning from spreadsheets to databases will require to reallocate resources in order to acquire the technological expertise and it would imply in the short run, more expenses for the company.

Particularly in this case, the user does not perceive the need of implementing more complex data management processes than whatsoever MS Excel provides. However, it was discussed with the user the long-run complications that spreadsheets may have and how those complications would end up consuming extra resources from the company.

8.5 Data Visualization

Data visualization is the last stage of the practical methodology of the BI process, and it can only be carried out once the information is cleaned, organized, and structured. During this stage, there was constant communication between the final user as developing the interface and user experience of the BI dashboards was a process that entirely considered the needs of the user. The design of every dashboard was developed thanks to the constant feedback from the user, for instance, if the final user was a visually limited senior age person, the selection of colors, font sizes, images, etc., was carefully selected.

Every reporting software has a different interface, and it will depend on the user which one is the best suitable for the company, however, MS Power BI is one of the most demanded platforms in Business and therefore, was chosen for this project as it offers a wide range of different BI capabilities.

After developing each report, these were successfully shared with the final user in the form of a URL where the user could interact with the data prior to the final consultancy advice with each report.

REFERENCES AND SELF-PRODUCED MATERIALS

365 DATA SCIENCE s.a. Database vs Spreadsheet. 365 Data Science. [Accessed: 2020-10-18.] Available: https://365datascience.com/explainer-video/database-vs-spreadsheet/

AGHAEI, M., ASADOLLAHI, A. 2013. Analysis of Business Intelligence on Strategic Decision Making. International Journal of Scientific Management and Development [electronic journal] Vol. 2 No. 1. [Accessed 2020-11-28.] Available: https://www.researchgate.net/publication/303631722_Analysis_of_Business_Intelligence_on_Strategic_Decision_Making

ALEXANDER, M., DECKER, J., WEHBE, B. 2014. Microsoft Business Intelligence Tools for Excel Analysts. New Jersey: John Wiley & Sons, Inc.

ALNOUKARI, M. & HANANO, A. 2017. Integration of business intelligence with corporate strategic management. Journal of Intelligence Studies in Business [electronic journal] Vol. 7 No. 2. [Accessed 2020-10-03.]

AMOAKO, T. B. 2013. THE IMPORTANCE OF BUSINESS INTELLIGENCE AS A DECISION-MAKING TOOL: CASE STUDY ELECTRICITY COMPANY OF GHANA (E.C.G.). UNIVERSITY OF BORAS. SCHOOL OF BUSINESS AND IT. [Accessed: 2020-10-18.] Available: http://www.diva-portal.org/smash/get/diva2:1309561/FULLTEXT01.pdf

AZVINE, B., CUI, Z. & NAUCK, D. 2005. Towards real-time business intelligence. BT Technology Journal [electronic journal] Vol. 23 No. 3. [Accessed 2020-10-03.] Available: https://schep.home.xs4all.nl/bi/paper%203%20Towards%20real-time%20business%20intelligence.pdf

BARKER, Dennis, BRIDGES, Donald, HUNTER, Regina, JOHNSON, Gregory, KRUPA, Joseph, MURPHY, James, & SORENSON, Ken, 2001. Department of Energy. Guidebook to Decision-Making Methods. Location: U.S.: Department of Energy. Available: http://everyspec.com/DOE/DOE-PUBS/download.php?spec=WSRC-IM-2002-00002.036284.pdf

BEIKZADEH, J., ESKANDARI, K. 2009. Competitive Intelligence as a tool to support managers in decision-making processes. Journal of Management [electronic journal] Vol. 1 No. 20. [Accessed 2020-11-28.]

BORKIN, M. A., VO, A. A., BYLINKSKII, Z., ISOLA, P., SUNKAVALLI, S., OLIVA, A. & PFISTER, H. 2013. What Makes a Visualization Memorable? SEMANTIC SCHOLAR. [Article] 1. [Accessed 2020-08-30.] Available: https://www.semanticscholar.org/paper/What-Makes-a-Visualization-Memorable-Borkin-Vo/1febc55512d2f33686a626402145e602de2bcfdf

CHARIS, Danni 2020-10-10. WHY DATABASES ARE IMPORTANT TO BUSINESS. JCount. [Accessed: 2020-10-17.] Available: https://www.jcount.com/why-databases-are-important-to-business/

CUDDLEY, M. O. 2016. Introduction to Microsoft Power BI. Bring Your data to Life! [Accessed: 2020-11-28.] Available:
http://projanco.com/Library/INTRODUCTION%20TO%20MICROSOFT%20POWER%20BI.pdf

DAVID, F. R. & DAVID, F. R. 2017. Strategic Management. A Competitive Advantage Approach, Concepts and Cases. Malaysia: Pearson Education.

FERNANDES, Luis 2019. Size Limitation Database Size for Microsoft SQL Server versions from 2000 to 2016. GFI Software. [Accessed: 2020-10-18.] Available: https://support.gfi.com/hc/en-us/articles/360012842094-Size-Limitation-Database-Size-for-Microsoft-SQL-Server-versions-from-2000-to-2016

FERRARI, A. & RUSSO, M. 2017. Analyzing Data with Microsoft Power BI and Power Pivot for Excel. Washington: Microsoft Press.

HARVEY, Jasmin, and TECHNICAL, Information Service 2007. Effective decision making. CIMA. Topic Gateway Series [electronic journal] 40. [Accessed 2020-10-03.] Available: http://www.cimaglobal.com/documents/importeddocuments/40_effective_decision_making.pdf

HEANG, R. & MOHAN, R. s.a. LITERATURE REVIEW OF BUSINESS INTELLIGENCE. School of Business and Engineering. Halmstad University, Sweden. [Accessed: 2020-10-18.] Available: https://www.diva-portal.org/smash/get/diva2:1080911/FULLTEXT01.pdf

HOWSON, Cindi 2008. Successful BUSINESS INTELLIGENCE. Secrets to Making BI a Killer APP. USA: McGrow-Hill.

HSU, J. s.a. Take your first step into data visualization with Microsoft's Power BI. Tech Soup Canada. [Accessed: 2020-10-18.] Available: https://www.techsoupcanada.ca/en/community/blog/power-BI-101

IMANUEL 2020-01-01. WHY SMALL BUSINESSES NEED BUSINESS INTELLIGENCE SOFTWARE. PAT RESEARCH. [Accessed: 2020-09-29.] Available: https://www.predictiveanalyticstoday.com/why-small-businesses-need-business-intelligence-software/

KOLHE, S. R., SALKUTE, V. R., & PATIL, M. P. 2011. Business Intelligence (BI) Tools and Techniques: Indian Scenario. International Research Publication House [electronic journal] Vol. 2 No. 1. [Accessed 2020-11-28.] Available: http://www.irphouse.com/cdai/cdaiv2n1_4.pdf

LANE, David s.a. Online Statistics Education: An Interactive Multimedia Course of Study. Rice University. Available: http://onlinestatbook.com/

LAUMANS s.a. an introduction to VISUALIZING DATA. [Accessed: 2020-10-18.] Available: http://courses.haigarmen.com/intd319/files/2018/01/visualizingdata.pdf

LEBLANC, P., MOSS, J., SARKA, D., RYAN, D. 2015. Applied Microsoft Business Intelligence. Indianapolis: John Wiley & Sons, Inc.

LINDBLAD, Marilyn 2019-01-25. Database Uses in Business. Chron. [Accessed: 2020-10-18.] Available: https://smallbusiness.chron.com/database-uses-business-64298.html

LINOFF, Gordon 2008. Data Analysis Using SQL and Excel®. Indianapolis: Wiley Publishing, Inc.

MAHMOUDI, M. 2009. Approach to the construction and implementation of business intelligence systems. Tadbir Journal [electronic journal] Vol. 1 No. 201. [Accessed 2020-11-28.]

MAMCENKO, J. s.a. Introduction to Data Modeling and MS Access. Academia. [Article] 1. [Accessed 2020-06-30.] Available: https://www.academia.edu/6872508/Introduction_to_Data_Modeling

MICROSOFT 2017-09-08. int, bigint, smallint, and tinyint (Transact-SQL). [Accessed: 2020-10-18.] Available: https://docs.microsoft.com/en-us/sql/t-sql/data-types/int-bigint-smallint-and-tinyint-transact-sql?view=sql-server-

ver15#:~:text=The%20int%20data%20type%20is,by%20the%20int%20data%20type.&text=SQL%20Server%20does%20not%20automatically,%2C%20and%20int)%20to%20bigint.

MICROSOFT 2019-11-09. What is SQL Server Management Studio (SSMS)?. Microsoft. [Accessed: 2020-10-18.] Available: https://docs.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?redirectedfrom=MSDN&view=sql-server-ver15

MICROSOFT 2019-11-26. Quickstart: Using Power Query in Power BI Desktop. [Accessed: 2020-10-18.] Available: https://docs.microsoft.com/en-us/power-query/power-query-quickstart-using-power-bi

MICROSOFT 2020-08-19. Get started with Power BI Desktop. [Accessed: 2020-10-18.] Available: https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started

MICROSOFT s.a. Business Intelligence VS Spreadsheets in the Long Run. Clean excess cell formatting on a worksheet. Microsoft. [Accessed: 2020-10-18.] Available: https://support.microsoft.com/en-us/office/clean-excess-cell-formatting-on-a-worksheet-e744c248-6925-4e77-9d49-4874f7474738

MICROSOFT s.a. Maximum Memory or File Size Exceeded. Excel 2013. Microsoft. [Accessed: 2020-10-18.] Available: https://support.microsoft.com/en-us/office/maximum-memory-or-file-size-exceeded-2b43e34d-bb20-49e1-b168-7866206f3669#:~:text=In%20the%2032%2Dbit%20version,the%20workbook%20cannot%20be%20saved.

MICROSOFT s.a. Power Query – Overview and Learning. [Accessed: 2020-10-18.] Available: https://support.microsoft.com/en-us/office/power-query-overview-and-learning-ed614c81-4b00-4291-bd3a-55d80767f81d#:~:text=Power%20Query%20is%20a%20data,follows%20a%20few%20common%20steps.

MINTZBERG, Henry 1994. The Fall and Rise of Strategic Planning. Harvard Business Publishing. Harvard Business Review. [electronic journal] 1994 issue. [Accessed 2020-10-03.] Available: https://hbr.org/1994/01/the-fall-and-rise-of-strategic-planning

OLSZAK, C. M. & ZURADA, J. 2015. INFORMATION TECHNOLOGY TOOLS FOR BUSINESS INTELLIGENCE DEVELOPMENT IN ORGANIZATIONS. POLISH JOURNAL OF MANAGEMENT STUDIES [electronic journal] Vol. 12 No. 1. [Accessed 2020-11-28.] Available: https://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-073b1f7d-d4f1-40e3-ad7f-26b6a0d8890e/c/Olszak.pdf

PHOCAS Sotware 2020-11-26. Business intelligence vs. spreadsheets: What does your business need? Business Intelligence Blog. Phocas Software. [Accessed: 2020-10-18.] Available: https://www.phocassoftware.com/business-intelligence-blog/business-intelligence-vs-excel-spreadsheets-what-does-your-business-need

PILON, Annie 2013-06-02. What is a Database?. Small Business TRENDS. [Accessed: 2020-10-18.] Available: https://smallbiztrends.com/2013/06/what-is-a-database.html#:~:text=2-,What%20is%20a%20database%3F,about%20your%20customers%20and%20clients.

RAJNOHA, R., STEFKO, R., MERKOVA, M., & DOBROVIC, J. 2016. BUSINESS INTELLIGENCE AS A KEY INFORMATION AND KNOWLEDGE TOOL FOR STRATEGIC BUSINESS PERFORMANCE MANAGEMENT. INFORMATION MANAGEMENT. [Accessed: 2020-10-18.] Available: https://core.ac.uk/download/pdf/43641275.pdf

RASTILAV R., ROBERT, Š., MARTINA, M., JÁN, D. 2016. BUSINESS INTELLIGENCE AS A KEY INFORMATION AND KNOWLEDGE TOOL FOR STRATEGIC BUSINESS PERFORMANCE MANAGEMENT. Information Management. [Article] 1. [Accessed 2020-11-29.] Available: https://core.ac.uk/download/pdf/43641275.pdf

RAUTENHAUS, M., BÖTTINGER, M., SIEMEN, S., HOFFFMAN, R., KIRBY, R., MIRZARGAR, M. RÖBER, N., & WESTERMANN, R. 2017. Visualization in Meteorology – A Survey of Techniques and Tools for Data Analysis Tasks. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS. [Article] 1. [Accessed 2020-06-30.] Available: http://www.sci.utah.edu/publications/Rau2017a/286c33cb985b759cef6df8cd5568f430a49d.pdf

RIAHI, Y. 2017. Business Intelligence: A Strategy for Business Development. Faculty of Informatics. International University of Rabat. [Accessed: 2020-10-18.] Available: https://www.researchgate.net/publication/321942128_Business_Intelligence_A_Strategy_for_Business_Development

ROST, L. 2018-08-22. What to Consider When Choosing Colors for Data Visualization. Dataquest.io. [Accessed: 2020-10-18.] Available: https://www.dataquest.io/blog/what-to-consider-when-choosing-colors-for-data-visualization/

SEKHAR, S. G. V. 2013. Business Intelligence. DCAP606. Lovely Professional University. New Delhi: Excel Books Private Limited.

TABLEAU s.a. Data visualization beginner's guide: a definition, examples, and learning resources. Tableau. [Accessed: 2020-10-18.] Available: https://www.tableau.com/learn/articles/data-visualization

TUFTE, Edward R 2001. The Visual Display of Quantitative Information. 2nd ed. Connecticut: Graphics Press.

VERCELLIS, Carlo 2009. Business Intelligence. Data Mining and Optimization for Decision Making. Milano: John Wiley & Sons, Ltd.

VICTORIA, Artur 2019. Taking a Decision. Research Gate [electronic journal]. [Accessed 2020-10-03.] Available: https://www.researchgate.net/publication/331929637_Taking_a_Decision

W3SCHOOLS s.a. SQL Data Types for MySQL, SQL Server, and MS Access. [Accessed: 2020-10-18.] Available: https://www.w3schools.com/sql/sql_datatypes.asp

WEBB, C. 2014. Power Query for Power BI and Excel. New York: Springer Science+Business Media New York

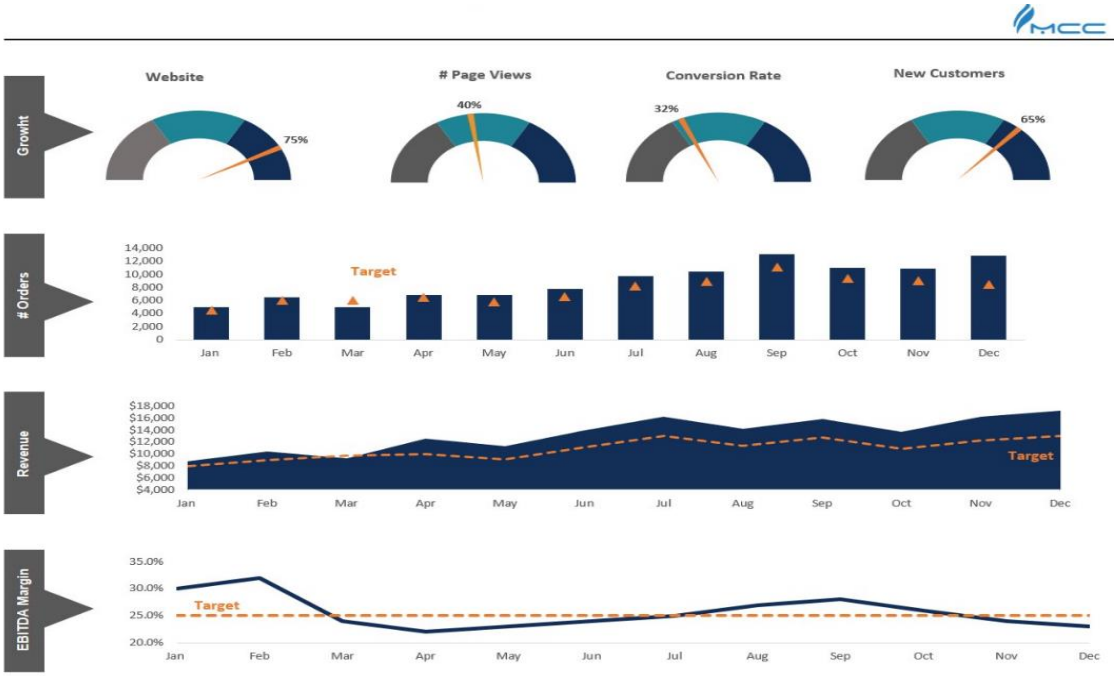
WITHEE, K. 2010. Microsoft® Business Intelligence For Dummies®. Indianapolis: Wiley Publishing, Inc.

ZHENG, Jack G. 2017. Data Visualization for Business Intelligence. Research Gate. [Article] 1. [Accessed 2020-06-30.] Available: https://www.researchgate.net/publication/321804138_Data_Visualization_for_Business_Intelligence

APPENDIX 1: BUSINESS INTELLIGENCE

Multiple Types of Data Visualizations

PICTURE 1. Investment Banking pitchbook as an example of a static dashboard (Corporate Finance Institute 2015)



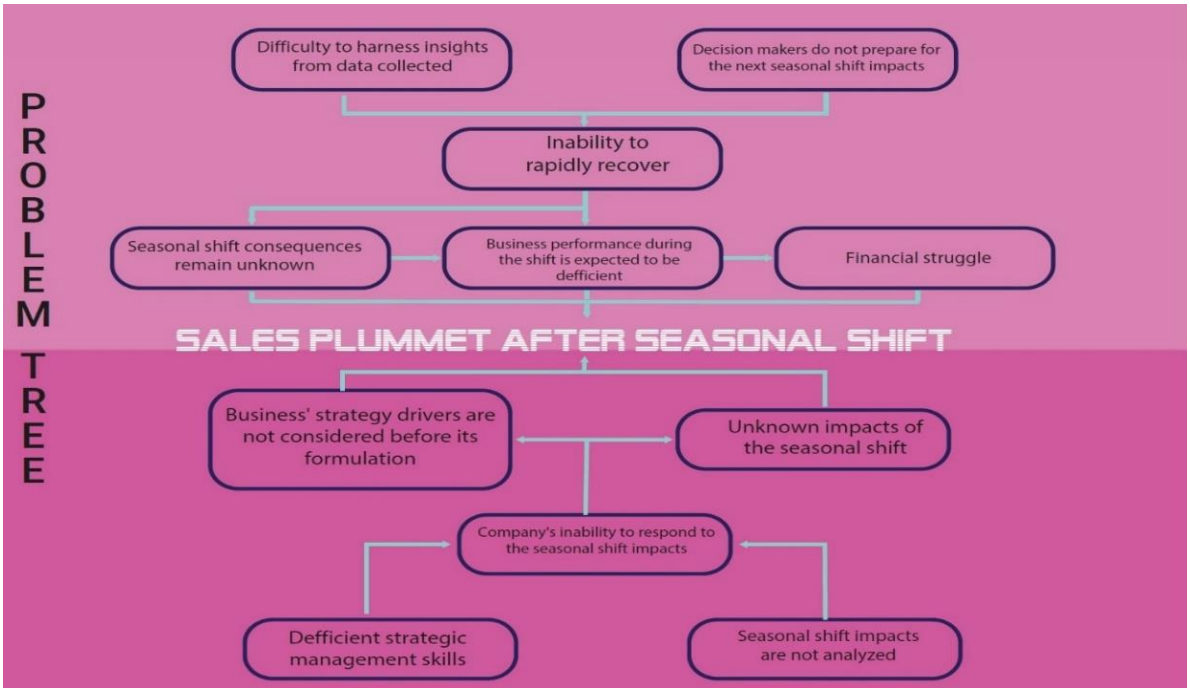
PICTURE 2. A dashboard containing plain text and visual but static content. (Corporate Finance Institute 2015)

APPENDIX 2: CASE COMPANY INTRODUCTION

PICTURE 3. The Case Company experiences a sales plunge from March to April (Cuéllar 2020-07-30).

APPENDIX 3: DEVELOPMENT PLAN OF THE BUSINESS INTELLIGENCE PROCESS FOR THE CASE COMPANY

Step 1 of the Plan: Overall Company's Assessment



PICTURE 4. Problem Tree. Sales Plummet After Seasonal Shift. (Cuéllar 2020-07-13).

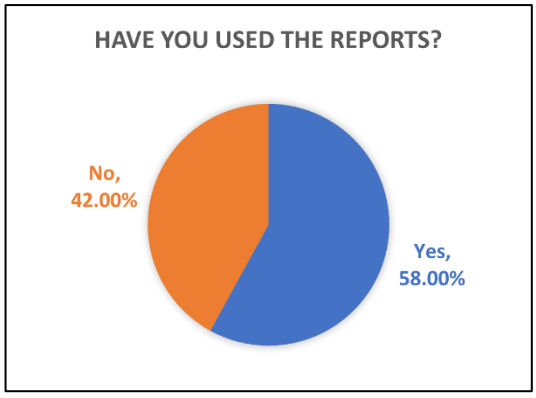
PICTURE 5. Objectives Tree. Sales Improvements After Seasonal Shift. (Cuéllar 2020-07-13).

Step 2 of the Plan: Data Collection

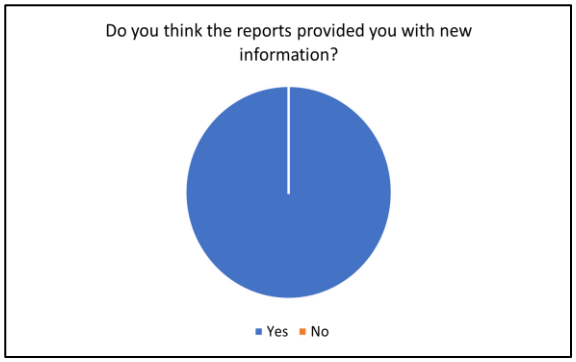
Ventas Mensuales por Producto - UNIDADES	2,018															
	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	TOTAL	ENE	FEB	MA
ENVASES PERSONALES	13,759	9,609	14,710	8,408	8,029	4,992	6,203	4,743	8,404	8,198	9,464	9,804	106,323	17,804	14,438	11.
PALETTI ENVASADAS	13,759	9,609	14,710	8,408	8,029	4,992	6,203	4,743	8,404	8,198	9,464	9,372	105,891	16,946	12,226	8,6
FRUTAS	2,226	1,727	2,346	1,266	1,164	606	856	590	1,096	1,070	1,438	1,238	15,623	2,854	2,000	1,2
FRESA CON MANGO	1,782	1,139	1,650	918	828	510	712	428	764	823	1,061	1,088	11,703	1,840	1,526	76
FRAMBUESA	372	426	396	216	228	54	78	102	208	169	197	150	2,596	516	330	24
KIWI CON FRAMBUESA Y MANDARINA JUMBO	72	162	300	132	108	42	66	60	124	78	180		1,324	498	144	16
CREMA	3,432	2,329	3,280	1,998	1,686	1,278	1,530	1,038	2,104	1,854	2,154	2,034	24,717	3,858	2,714	1,7
LUCUMA +40% (DESDE MARZO 2019)	1,160	769	1,027	600	450	348	546	294	347	600	672	858	7,671	858	846	53
QUESO HELADO +40% (DESDE MARZO 2019)	1,618	966	1,443	888	828	708	732	480	1,140	816	1,080	702	11,401	1,794	1,166	82
CHOCOLATE BITTER +40% (DESDE MARZO 2019)	522	384	498	354	282	174	186	180	365	306	246	390	3,887	690	462	30
LUCUMA Y CHOCOLATE JUMBO	132	210	312	156	126	48	66	84	252	132	156	84	1,758	516	240	10
YOGURT	330	252	588	270	294	102	156	162	390	324	402	222	3,492	930	490	33
YOGURT DE FRESA	216	144	342	186	204	78	102	120	240	216	156	180	2,184	432	210	16
YOGURT DE FRESA Y MARACUYA JUMBO	114	108	246	84	90	24	54	42	132	84	102	42	1,122	126	120	11
YOGURT NATURAL CON FRUTAS JUMBO	0	0	0	0	0	0	0	0	18	24	144		186	372	160	5
LIGHT	1,174	762	1,458	1,188	1,080	426	399	480	1,048	846	1,162	1,042	11,065	1,810	1,434	1,4
LUCUMA LIGHT	492	276	654	576	486	174	192	234	485	408	504	596	5,077	910	654	62
CHOCOLATE LIGHT	682	486	804	612	594	252	207	246	563	438	658	446	5,988	900	780	81
CREMAS DECORADAS	6,381	4,317	6,804	3,590	3,577	2,472	3,190	2,383	3,532	3,864	4,194	4,656	48,960	6,840	5,318	3,5
COOKIES AND CREAM +40% (DESDE MARZO 2019)	1,254	769	1,802	752	859	546	642	638	972	786	756	912	10,688	1,542	1,094	56
BANANA SPLIT +40% (DESDE MARZO 2019)	186	234	234	168	108	126	96	54	130	210	132	204	1,882	216	252	8
COCO CON MANJAR	0	0	0	0	0	0	0	0	12		84	192	288	234	234	14

PICTURE 6. Sales from the company's new sales channel named "Delivery" (Cuéllar 2020-07-13).

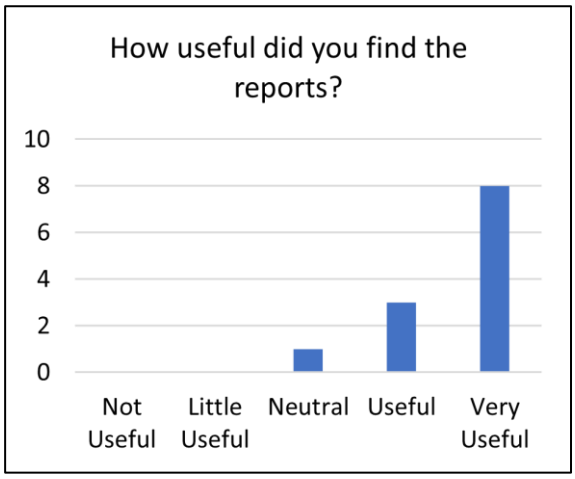
Step 5 of the Plan: Questionnaires' Results Delivered to the User



PICTURE 7. Have You Used the Report? Pie chart (Cuéllar 2020-09-30).

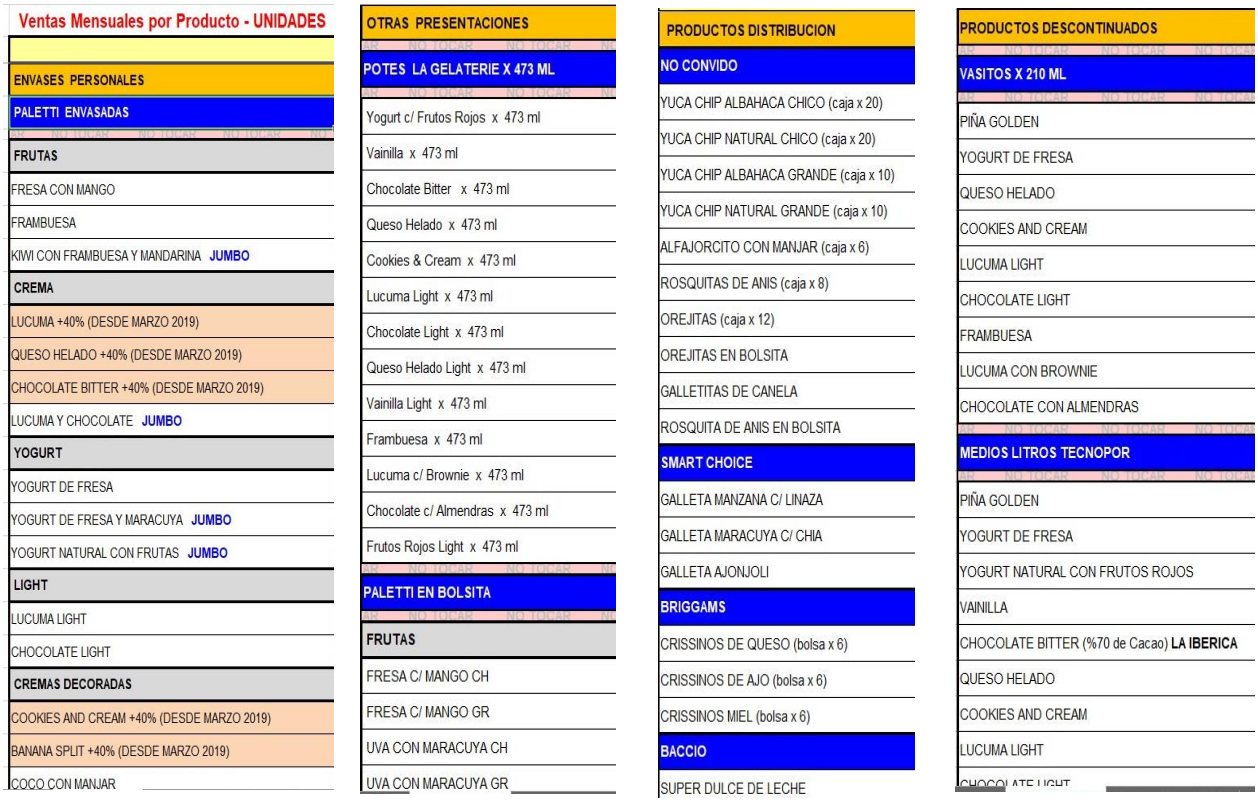


PICTURE 8. Do you think the reports provided you with new information? (Cuéllar 2020-09-30).



PICTURE 9. How useful did you find the reports? (Cuéllar 2020-09-30).

APPENDIX 4: BI DEVELOPED WORK USING EXCEL AS DATASOURCE



Ventas Mensuales por Producto - UNIDADES
ENVASES PERSONALES
PALETTI ENVASADAS
FRUTAS
FRESA CON MANGO
FRAMBUESA
KIWI CON FRAMBUESA Y MANDARINA JUMBO
CREMA
LUCUMA +40% (DESDE MARZO 2019)
QUESO HELADO +40% (DESDE MARZO 2019)
CHOCOLATE BITTER +40% (DESDE MARZO 2019)
LUCUMA Y CHOCOLATE JUMBO
YOGURT
YOGURT DE FRESA
YOGURT DE FRESA Y MARACUYA JUMBO
YOGURT NATURAL CON FRUTAS JUMBO
LIGHT
LUCUMA LIGHT
CHOCOLATE LIGHT
CREMAS DECORADAS
COOKIES AND CREAM +40% (DESDE MARZO 2019)
BANANA SPLIT +40% (DESDE MARZO 2019)
COCO CON MANJAR

OTRAS PRESENTACIONES
POTES LA GELATERIE X 473 ML
Yogurt c/ Frutos Rojos x 473 ml
Vainilla x 473 ml
Chocolate Bitter x 473 ml
Queso Helado x 473 ml
Cookies & Cream x 473 ml
Lucuma Light x 473 ml
Chocolate Light x 473 ml
Queso Helado Light x 473 ml
Vainilla Light x 473 ml
Frambuesa x 473 ml
Lucuma c/ Brownie x 473 ml
Chocolate c/ Almendras x 473 ml
Frutos Rojos Light x 473 ml
PALETTI EN BOLSITA
FRUTAS
FRESA C/ MANGO CH
FRESA C/ MANGO GR
UVA CON MARACUYA CH
UVA CON MARACUYA GR

PRODUCTOS DISTRIBUCION
NO CONVIDO
YUCA CHIP ALBAHACA CHICO (caja x 20)
YUCA CHIP NATURAL CHICO (caja x 20)
YUCA CHIP ALBAHACA GRANDE (caja x 10)
YUCA CHIP NATURAL GRANDE (caja x 10)
ALFAJORCITO CON MANJAR (caja x 6)
ROSQUITAS DE ANIS (caja x 8)
OREJITAS (caja x 12)
OREJITAS EN BOLSITA
GALLETITAS DE CANELA
ROSQUITA DE ANIS EN BOLSITA
SMART CHOICE
GALLETA MANZANA C/ LINAZA
GALLETA MARACUYA C/ CHIA
GALLETA AJONJOLI
BRIGGAMS
CRISSINOS DE QUESO (bolsa x 6)
CRISSINOS DE AJO (bolsa x 6)
CRISSINOS MIEL (bolsa x 6)
BACCIO
SUPER DULCE DE LECHE

PRODUCTOS DESCONTINUADOS
VASITOS X 210 ML
PIÑA GOLDEN
YOGURT DE FRESA
QUESO HELADO
COOKIES AND CREAM
LUCUMA LIGHT
CHOCOLATE LIGHT
FRAMBUESA
LUCUMA CON BROWNIE
CHOCOLATE CON ALMENDRAS
MEDIOS LITROS TECNOPOR
PIÑA GOLDEN
YOGURT DE FRESA
YOGURT NATURAL CON FRUTOS ROJOS
VAINILLA
CHOCOLATE BITTER (%70 de Cacao) LA IBERICA
QUESO HELADO
COOKIES AND CREAM
LUCUMA LIGHT
CHOCOLATE LIGHT

PICTURE 10. Monthly Sales by Product (Cuellar 2020-07-22).

Stage 1 of the BI process: Data Cleansing

Ventas Mensuales por Producto - UNIDADES
ENVASES PERSONALES
PALETTI ENVASADAS
FRUTAS
FRESA CON MANGO
FRAMBUESA
KIWI CON FRAMBUESA Y MANDARINA JUMBO

PICTURE 11. Point A. Unstructured Data About Products (Cuellar 2020-07-22).

A
FRESA CON MANGO
FRAMBUESA
KIWI CON FRAMBUESA Y MANDARINA

PICTURE 12. Point B. Products' Data Extracted (Cuéllar 2020-07-22).

	A
1	Product Name
2	Fresa con mango
3	FRAMBUESA
4	KIWI CON FRAMBUESA Y MANDARINA
5	LUCUMA
6	QUESO HELADO
7	CHOCOLATE BITTER
8	LUCUMA Y CHOCOLATE
9	YOGURT DE FRESA
10	YOGURT DE FRESA Y MARACUYA
11	YOGURT NATURAL CON FRUTAS
12	LUCUMA LIGHT
13	CHOCOLATE LIGHT
14	COOKIES AND CREAM
15	BANANA SPLIT
16	COCO CON MANJAR
17	LUCUMA CON BROWNIE
18	MENTA CON CHOCOLATE
19	STRACCIATELLA
20	VAINILLA CON BAÑO CROCANTE
21	DOBLE CHOCOLATE CON ALMENDRAS
22	PISCO SOUR
23	FRESA SOUR
24	MARACUYA SOUR
25	FRESAS CON LECHE CONDENSADA
26	MANGO CON CHOCOLATE
27	MARACUYA CON LECHE CONDENSADA
28	CHOCOLATE CON MARACUYA
29	LUCUMA CON BROWNIE
30	BROWNIE CON CHOCOLATE
31	CHOCOLATE & COOKIES
32	COCO CON MANJAR
33	YOGURT GRIEGO CON MIEL
34	FRESAS CON LECHE CONDENSADA
35	MARACUYA CON LECHE CONDENSADA
36	LUCUMA CON BROWNIE
37	CHOCOLATE & COOKIES

	A
202	FIOR DE PANNA(NATA) - MS MIX
203	VAINILLA
204	CHOCOLATE
205	FRESA
206	LÚCUMA
207	COCO
208	FIOR DE PANNA (NATA)
209	STRACCIATELA
210	MENTA CHIPS
211	TRICOLOR
212	TRICOLOR CON CHOCOLATE
213	PIÑA GOLDEN
214	YOGURT DE FRESA
215	QUESO HELADO
216	COOKIES AND CREAM
217	LUCUMA LIGHT
218	CHOCOLATE LIGHT
219	FRAMBUESA
220	LUCUMA CON BROWNIE
221	CHOCOLATE CON ALMENDRAS
222	PIÑA GOLDEN
223	YOGURT DE FRESA
224	YOGURT NATURAL CON FRUTOS ROJOS
225	VAINILLA
226	CHOCOLATE BITTER (%70 de Cacao) LA IBERICA
227	QUESO HELADO
228	COOKIES AND CREAM
229	LUCUMA LIGHT
230	CHOCOLATE LIGHT
231	FRAMBUESA
232	LUCUMA CON BROWNIE
233	CHOCOLATE CON ALMENDRAS
234	VAINILLA
235	CHOCOLATE BITTER (%70 de Cacao) LA IBERICA
236	QUESO HELADO
237	LUCUMA CON BROWNIE
238	CHOCOLATE CON ALMENDRAS

PICTURE 13. 237 products extracted from the unstructured data (Cuéllar 2020-07-22).

Ventas Mensuales por Producto - UNIDADES
ENVASES PERSONALES
PALETTI ENVASADAS
FRUTAS
FRESA CON MANGO
FRAMBUESA
KIWI CON FRAMBUESA Y MANDARINA JUMBO

	A	B	C	D	E
1	PRODUCT NAME	TYPE	BRAND	PRESENTATION	CATEGORY
2	FRESA CON MANGO	Frutas	Paletti	Envasadas	Envases personales
3	FRAMBUESA	Frutas	Paletti	Envasadas	Envases personales
4	KIWI CON FRAMBUESA Y MANDARINA	Frutas	Paletti	Envasadas	Envases personales

PICTURE 15. Data points after classification (Cuéllar 2020-07-22).

PICTURE 14. Data points before classification (Cuellar 2020-07-22).

	A	B	C	D	E	F
1	ID_PROD	NAME	TYPE	BRAND	PRESENTATION	CATEGORY
2	1	Fresa con mango	Frutas	Paletti	Envasadas	Envases personales
3	2	FRAMBUESA	Frutas	Paletti	Envasadas	Envases personales
4	3	KIWI CON FRAMBUESA Y MANDARINA	Frutas	Paletti	Envasadas	Envases personales
5	4	LUCUMA	crema	Paletti	Envasadas	Envases personales
6	5	QUESO HELADO	crema	Paletti	Envasadas	Envases personales
7	6	CHOCOLATE BITTER	crema	Paletti	Envasadas	Envases personales
8	7	LUCUMA Y CHOCOLATE	crema	Paletti	Envasadas	Envases personales
9	8	YOGURT DE FRESA	yogurt	Paletti	Envasadas	Envases personales
10	9	YOGURT DE FRESA Y MARACUYA	yogurt	Paletti	Envasadas	Envases personales
11	10	YOGURT NATURAL CON FRUTAS	yogurt	Paletti	Envasadas	Envases personales
12	11	LUCUMA LIGHT	light	Paletti	Envasadas	Envases personales
13	12	CHOCOLATE LIGHT	light	Paletti	Envasadas	Envases personales
14	13	COOKIES AND CREAM	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
15	14	BANANA SPLIT	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
16	15	COCO CON MANJAR	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
17	16	LUCUMA CON BROWNIE	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
18	17	MENTA CON CHOCOLATE	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
19	18	STRACCIATELLA	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
20	19	VAINILLA CON BAÑO CROCANTE	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
21	20	DOBLE CHOCOLATE CON ALMENDRAS	CREMAS DECORADAS	Paletti	Envasadas	Envases personales
22	21	PISCO SOUR	LICOR	Paletti	Envasadas	Envases personales
23	22	FRESA SOUR	LICOR	Paletti	Envasadas	Envases personales
24	23	MARACUYA SOUR	LICOR	Paletti	Envasadas	Envases personales
25	24	FRESAS CON LECHE CONDENSADA	grande	La calaca	Envasadas	Envases personales
26	25	MANGO CON CHOCOLATE	grande	La calaca	Envasadas	Envases personales
27	26	MARACUYA CON LECHE CONDENSADA	grande	La calaca	Envasadas	Envases personales
28	27	CHOCOLATE CON MARACUYA	grande	La calaca	Envasadas	Envases personales
29	28	LUCUMA CON BROWNIE	grande	La calaca	Envasadas	Envases personales
30	29	BROWNIE CON CHOCOLATE	grande	La calaca	Envasadas	Envases personales
31	30	CHOCOLATE & COOKIES	grande	La calaca	Envasadas	Envases personales
32	31	COCO CON MANJAR	grande	La calaca	Envasadas	Envases personales
33	32	YOGURT GRIEGO CON MIEL	grande	La calaca	Envasadas	Envases personales
34	33	FRESAS CON LECHE CONDENSADA	chica	La calaca	Envasadas	Envases personales
35	34	MARACUYA CON LECHE CONDENSADA	chica	La calaca	Envasadas	Envases personales
36	35	LUCUMA CON BROWNIE	chica	La calaca	Envasadas	Envases personales

PICTURE 16. Final classification of all the 237 products with their own attributes. (Cuéllar 2020-07-22).

A	B	C	E
ID_SALE	PERIOD	UNITS	ID_QUARTER
1	01/01/2018	1782	1

PICTURE 17. Data about a product's sales after being classified. (Cuéllar 2020-07-22).

ID_SALE	PERIOD	UNITS	ID_QUART
1	01/01/2018	1782	1
2	01/01/2018	372	1
3	01/01/2018	72	1
4	01/01/2018	1160	1
5	01/01/2018	1618	1
6	01/01/2018	522	1
7	01/01/2018	132	1
8	01/01/2018	216	1
9	01/01/2018	114	1
10	01/01/2018	0	1
11	01/01/2018	492	1
12	01/01/2018	682	1
13	01/01/2018	1254	1
14	01/01/2018	186	1
15	01/01/2018	0	1
16	01/01/2018	1338	1
17	01/01/2018	324	1
18	01/01/2018	1248	1
19	01/01/2018	522	1
20	01/01/2018	1509	1
21	01/01/2018	96	1
22	01/01/2018	42	1
23	01/01/2018	78	1
24	01/01/2018	0	1
25	01/01/2018	0	1
26	01/01/2018	0	1
27	01/01/2018	0	1
28	01/01/2018	0	1
29	01/01/2018	0	1
30	01/01/2018	0	1
31	01/01/2018	0	1
32	01/01/2018	0	1
33	01/01/2018	0	1
34	01/01/2018	0	1
35	01/01/2018	0	1
36	01/01/2018	0	1
37	01/01/2018	0	1

6836	01/05/2020	0	2
6837	01/05/2020	0	2
6838	01/05/2020	0	2
6839	01/05/2020	0	2
6840	01/05/2020	0	2
6841	01/05/2020	0	2
6842	01/05/2020	0	2
6843	01/05/2020	0	2
6844	01/05/2020	0	2
6845	01/05/2020	0	2
6846	01/05/2020	0	2
6847	01/05/2020	0	2
6848	01/05/2020	0	2
6849	01/05/2020	0	2
6850	01/05/2020	0	2
6851	01/05/2020	0	2
6852	01/05/2020	0	2
6853	01/05/2020	0	2
6854	01/05/2020	0	2
6855	01/05/2020	0	2
6856	01/05/2020	0	2
6857	01/05/2020	0	2
6858	01/05/2020	0	2
6859	01/05/2020	0	2
6860	01/05/2020	0	2
6861	01/05/2020	0	2
6862	01/05/2020	0	2
6863	01/05/2020	0	2
6864	01/05/2020	0	2
6865	01/05/2020	0	2
6866	01/05/2020	0	2
6867	01/05/2020	0	2
6868	01/05/2020	0	2
6869	01/05/2020	0	2
6870	01/05/2020	0	2
6871	01/05/2020	0	2
6872	01/05/2020	0	2
6873	01/05/2020	0	2

PICTURE 18. All the data about products' sales have been classified. (Cuéllar 2020-07-22).

Stage 2 of the BI process: Creating Relations

ID_SALE	PERIOD	UNITS	ID_PROD	ID_QUARTER
1	01/01/2018	1782	1	1
2	01/01/2018	372	2	1
3	01/01/2018	72	3	1
4	01/01/2018	1160	4	1
5	01/01/2018	1618	5	1
6	01/01/2018	522	6	1
7	01/01/2018	132	7	1
8	01/01/2018	216	8	1
9	01/01/2018	114	9	1
10	01/01/2018	0	10	1
11	01/01/2018	492	11	1
12	01/01/2018	682	12	1
13	01/01/2018	1254	13	1
14	01/01/2018	186	14	1
15	01/01/2018	0	15	1
16	01/01/2018	1338	16	1
17	01/01/2018	324	17	1
18	01/01/2018	1248	18	1
19	01/01/2018	522	19	1
20	01/01/2018	1509	20	1
21	01/01/2018	96	21	1
22	01/01/2018	42	22	1
23	01/01/2018	78	23	1
24	01/01/2018	0	24	1
25	01/01/2018	0	25	1
26	01/01/2018	0	26	1
27	01/01/2018	0	27	1
28	01/01/2018	0	28	1
29	01/01/2018	0	29	1
30	01/01/2018	0	30	1
31	01/01/2018	0	31	1
32	01/01/2018	0	32	1
33	01/01/2018	0	33	1
34	01/01/2018	0	34	1
35	01/01/2018	0	35	1
36	01/01/2018	0	36	1
37	01/01/2018	0	37	1

PICTURE 19. Example of relationships between ID_SALE and ID_PROD. (Cuéllar 2020-07-22).

ID_SALE	PERIOD	UNITS	ID_PROD	ID_QUARTER
1	01/01/2018	1782	1	1
2	01/01/2018	372	2	1
3	01/01/2018	72	3	1
4	01/01/2018	1160	4	1
5	01/01/2018	1618	5	1
6	01/01/2018	522	6	1
7	01/01/2018	132	7	1
8	01/01/2018	216	8	1
9	01/01/2018	114	9	1
10	01/01/2018	0	10	1
11	01/01/2018	492	11	1
12	01/01/2018	682	12	1
13	01/01/2018	1254	13	1
14	01/01/2018	186	14	1
15	01/01/2018	0	15	1
16	01/01/2018	1338	16	1
17	01/01/2018	324	17	1
18	01/01/2018	1248	18	1
19	01/01/2018	522	19	1
20	01/01/2018	1509	20	1
21	01/01/2018	96	21	1
22	01/01/2018	42	22	1
23	01/01/2018	78	23	1
24	01/01/2018	0	24	1
25	01/01/2018	0	25	1
26	01/01/2018	0	26	1
27	01/01/2018	0	27	1
28	01/01/2018	0	28	1
29	01/01/2018	0	29	1
30	01/01/2018	0	30	1
31	01/01/2018	0	31	1
32	01/01/2018	0	32	1
33	01/01/2018	0	33	1
34	01/01/2018	0	34	1
35	01/01/2018	0	35	1
36	01/01/2018	0	36	1
37	01/01/2018	0	37	1

SALES PRODUCTS Sheet1

PICTURE 20. Using separate sheets for each table to connect to Power BI. (Cuéllar 2020-07-22).

Stage 3 of the BI process: Power Query - Connecting Data to Power Query

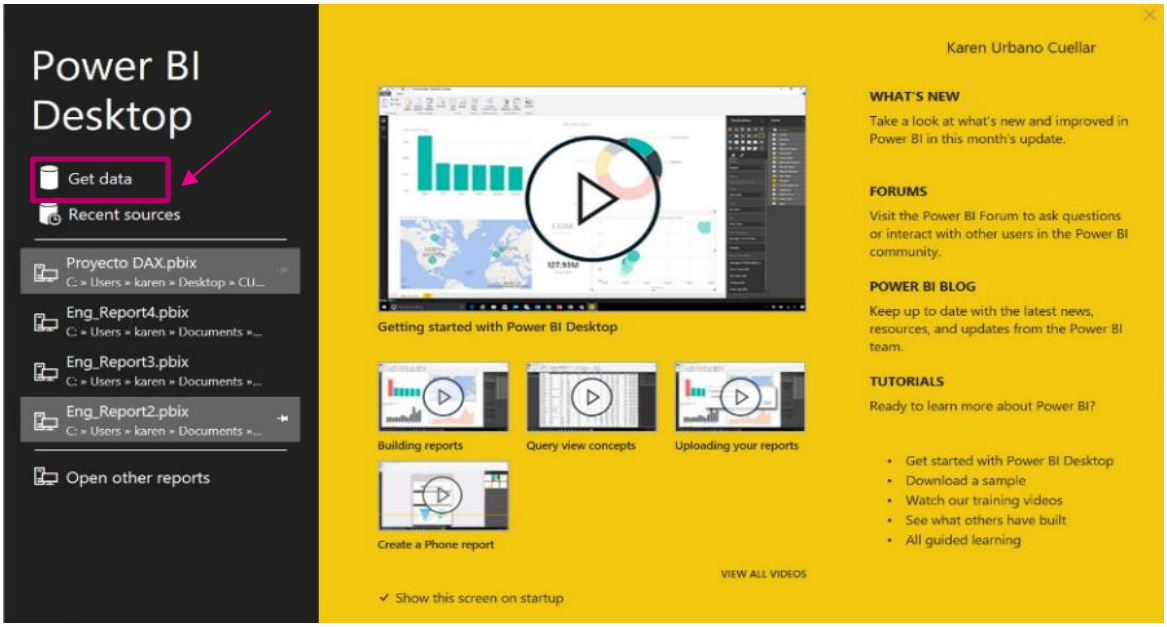


FIGURE 21. Power BI welcome screen with options to connect data, such as Get Data. (Cuéllar 2020-07-13).

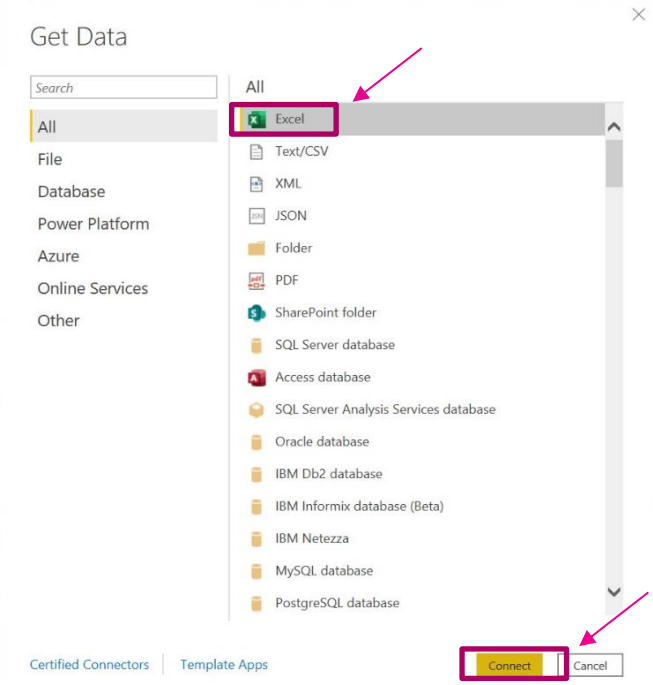


FIGURE 22. Connecting data to an Excel file. (Cuéllar 2020-08-23).

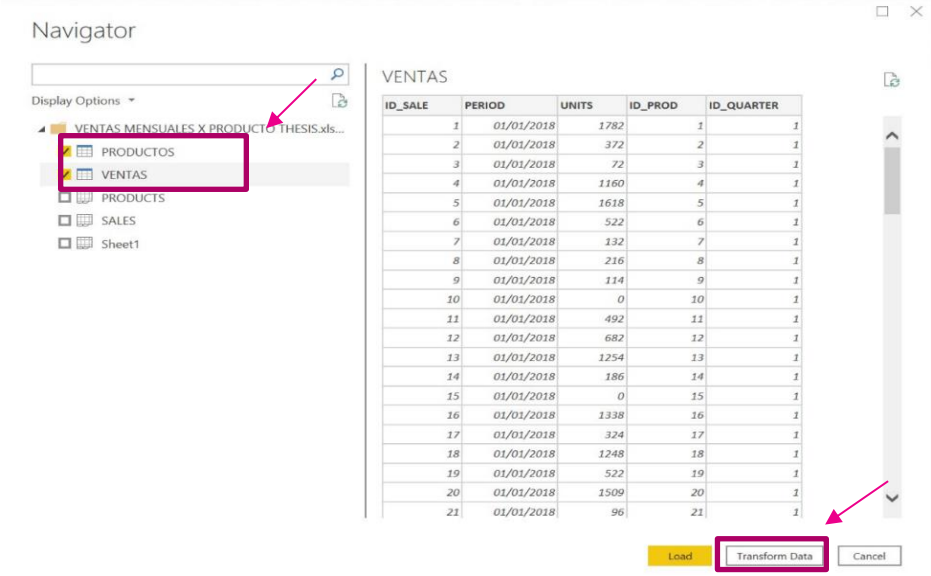


FIGURE 23. Worksheets available in the Excel file & options to create a visual report. (Cuéllar 2020-07-13).

Stage 3 of the BI process: Power Query - Power Query Editor Interface

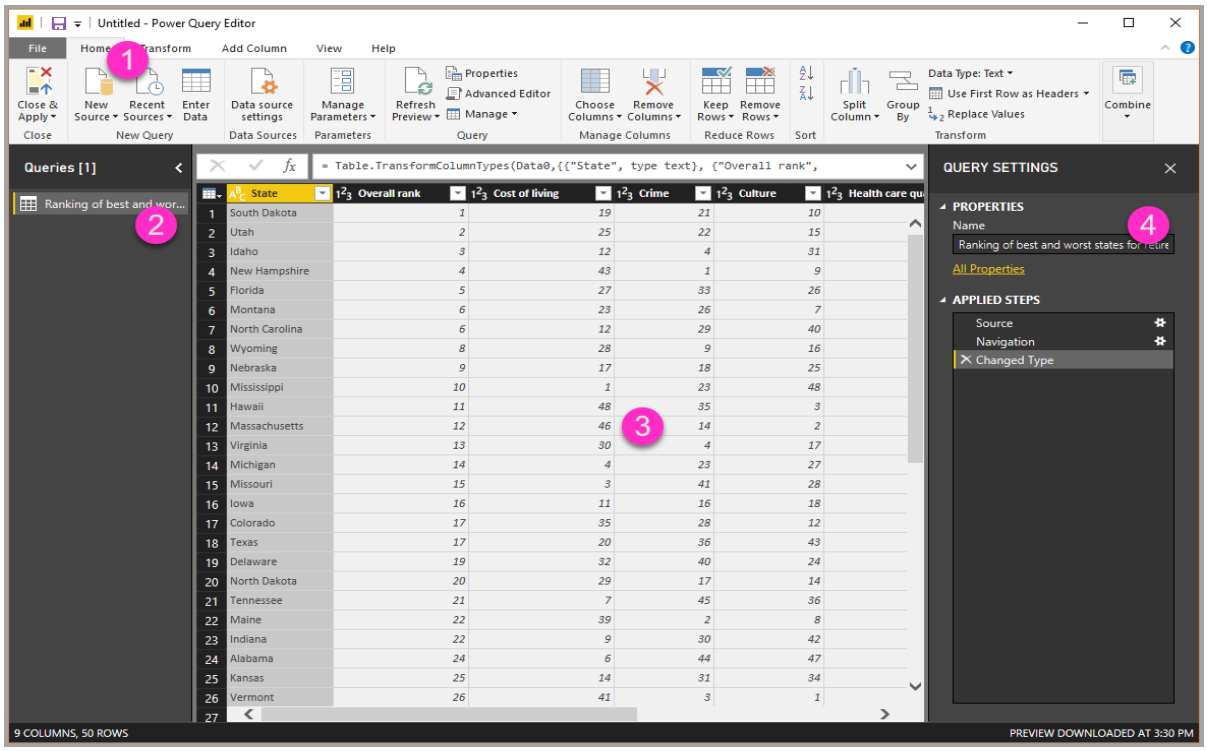


FIGURE 24. Power Query Editor interface sections. (Microsoft, 2019).

Stage 3 of the BI process: Power Query - Editing Data with Power Query

	NAME	TYPE	BRAND	PRESENTATION	CATEGORY
1	Fresa con mango	Frutas	Paletti	Envasadas	Envases personale
2	FRAMBUESA	Frutas	Paletti	Envasadas	Envases personale
3	KIWI CON FRAMBUESA Y MANDARI...	Frutas	Paletti	Envasadas	Envases personale
4	LUCUMA	crema	Paletti	Envasadas	Envases personale
5	QUESO HELADO	crema	Paletti	Envasadas	Envases personale
6	CHOCOLATE BITTER	crema	Paletti	Envasadas	Envases personale
7	LUCUMA Y CHOCOLATE	crema	Paletti	Envasadas	Envases personale
8	YOGURT DE FRESA	yogurt	Paletti	Envasadas	Envases personale
9	YOGURT DE FRESA Y MARACUYA	yogurt	Paletti	Envasadas	Envases personale
10	YOGURT NATURAL CON FRUTAS	yogurt	Paletti	Envasadas	Envases personale
11	LUCUMA LIGHT	light	Paletti	Envasadas	Envases personale
12	CHOCOLATE LIGHT	light	Paletti	Envasadas	Envases personale
13	COOKIES AND CREAM	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
14	BANANA SPLIT	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
15	COCO CON MANJAR	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
16	LUCUMA CON BROWNIE	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
17	MENTA CON CHOCOLATE	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
18	STRACCIATELLA	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
19	VAINILLA CON BAÑO CROCANTE	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
20	DOBLE CHOCOLATE CON ALMENDR...	CREMAS DECORADAS	Paletti	Envasadas	Envases personale
21	PISCO SOUR	LICOR	Paletti	Envasadas	Envases personale
22	FRESA SOUR	LICOR	Paletti	Envasadas	Envases personale
23	MARACUYA SOUR	LICOR	Paletti	Envasadas	Envases personale

	NAME	TYPE	BRAND	PRESENTATION	CATEGORY
1	Fresa Con Mango	Frutas	Paletti	Envasadas	Envases Personales
2	Frambuesa	Frutas	Paletti	Envasadas	Envases Personales
3	Kiwi Con Frambuesa Y Mandarina	Frutas	Paletti	Envasadas	Envases Personales
4	Lucuma	Crema	Paletti	Envasadas	Envases Personales
5	Queso Helado	Crema	Paletti	Envasadas	Envases Personales
6	Chocolate Bitter	Crema	Paletti	Envasadas	Envases Personales
7	Lucuma Y Chocolate	Crema	Paletti	Envasadas	Envases Personales
8	Yogurt De Fresa	Yogurt	Paletti	Envasadas	Envases Personales
9	Yogurt De Fresa Y Maracuya	Yogurt	Paletti	Envasadas	Envases Personales
10	Yogurt Natural Con Frutas	Yogurt	Paletti	Envasadas	Envases Personales
11	Lucuma Light	Light	Paletti	Envasadas	Envases Personales
12	Chocolate Light	Light	Paletti	Envasadas	Envases Personales
13	Cookies And Cream	Cremas Decoradas	Paletti	Envasadas	Envases Personales
14	Banana Split	Cremas Decoradas	Paletti	Envasadas	Envases Personales
15	Coco Con Manjar	Cremas Decoradas	Paletti	Envasadas	Envases Personales
16	Lucuma Con Brownie	Cremas Decoradas	Paletti	Envasadas	Envases Personales
17	Menta Con Chocolate	Cremas Decoradas	Paletti	Envasadas	Envases Personales
18	Stracciatella	Cremas Decoradas	Paletti	Envasadas	Envases Personales
19	Vainilla Con Baño Crocante	Cremas Decoradas	Paletti	Envasadas	Envases Personales
20	Doble Chocolate Con Almendras	Cremas Decoradas	Paletti	Envasadas	Envases Personales
21	Pisco Sour	Licor	Paletti	Envasadas	Envases Personales
22	Fresa Sour	Licor	Paletti	Envasadas	Envases Personales
23	Maracuya Sour	Licor	Paletti	Envasadas	Envases Personales

FIGURE 25. Data points before and after capitalization formatting. (Cuéllar 2020-08-23).

Stage 4 of the BI process: Data Visualization - Power BI Reporting Interface

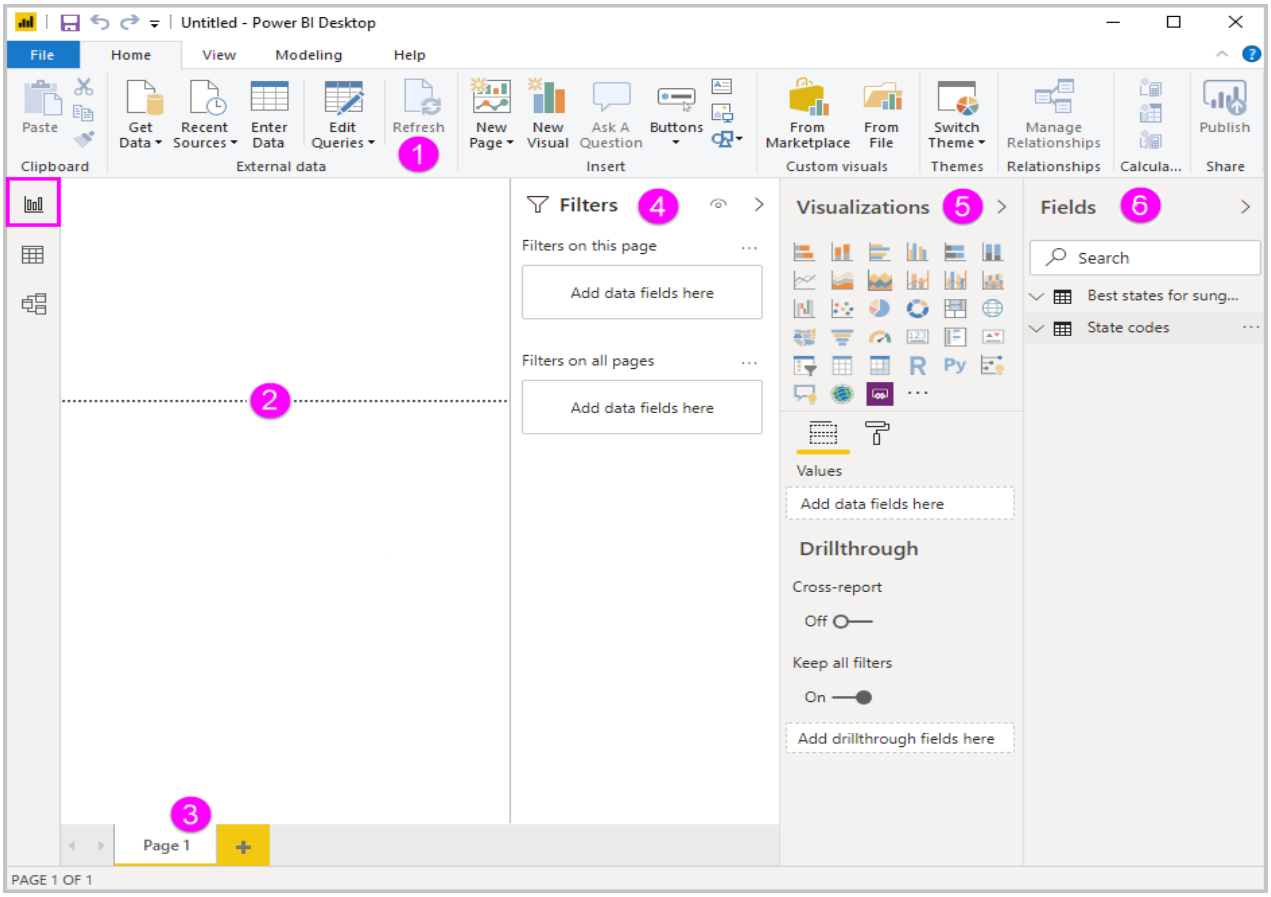


FIGURE 26. Power BI Reporting interface sections. (Microsoft, 2020).

Stage 4 of the BI process: Data Visualization - Charts Types

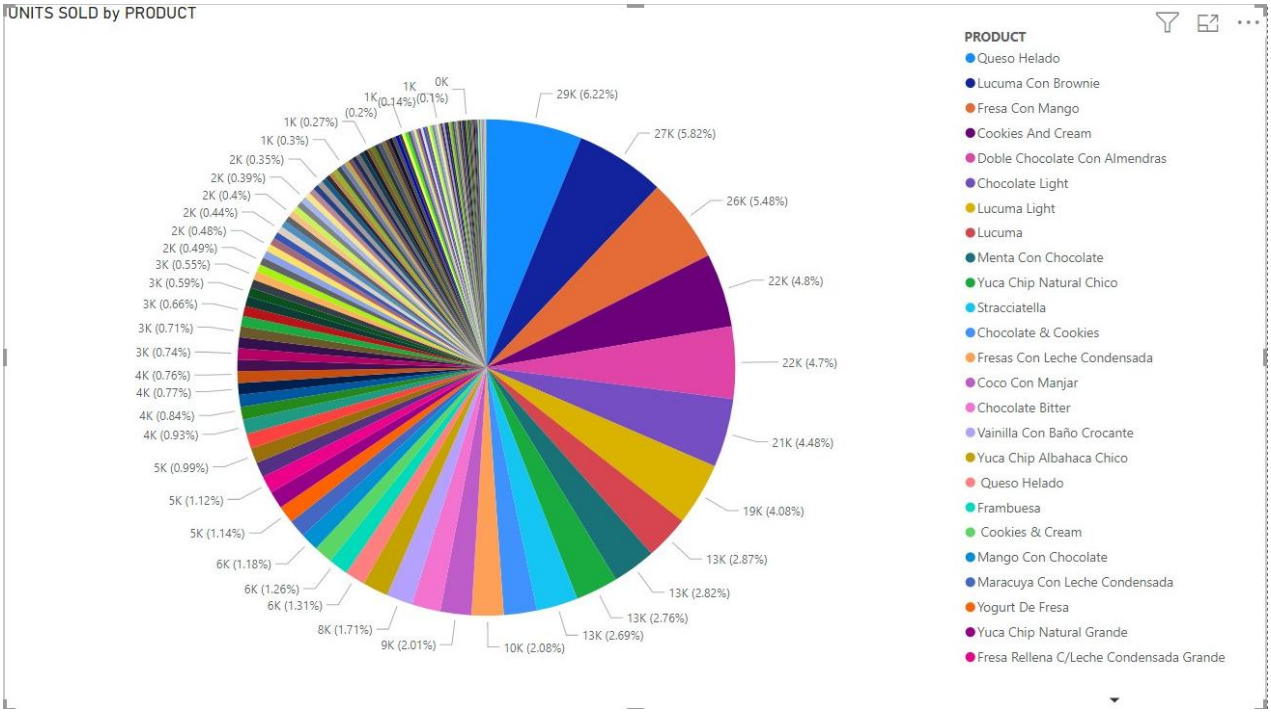


FIGURE 27. Units Sold by Product pie chart with an overload of information. (Cuéllar 2020-08-24).

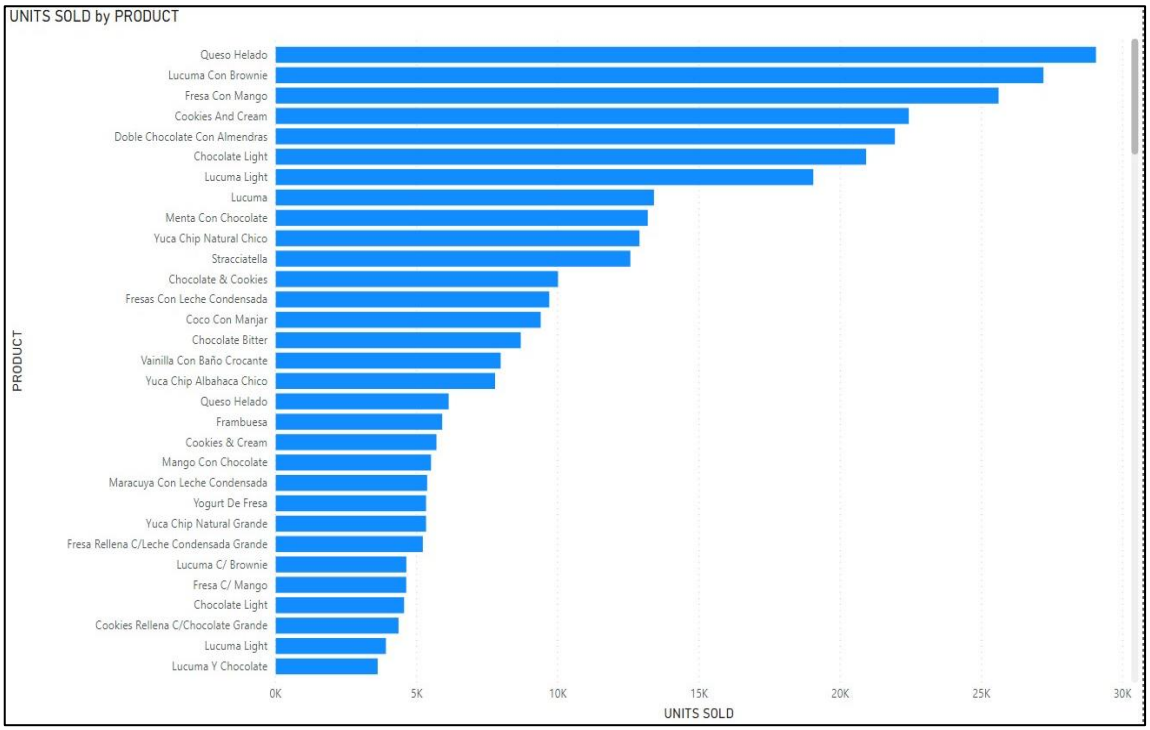


FIGURE 28. Units Sold by Product bar chart. Intuitive distribution. (Cuéllar 2020-08-24).

Stage 4 of the BI process: Data Visualization - Visualizing Data with Power BI

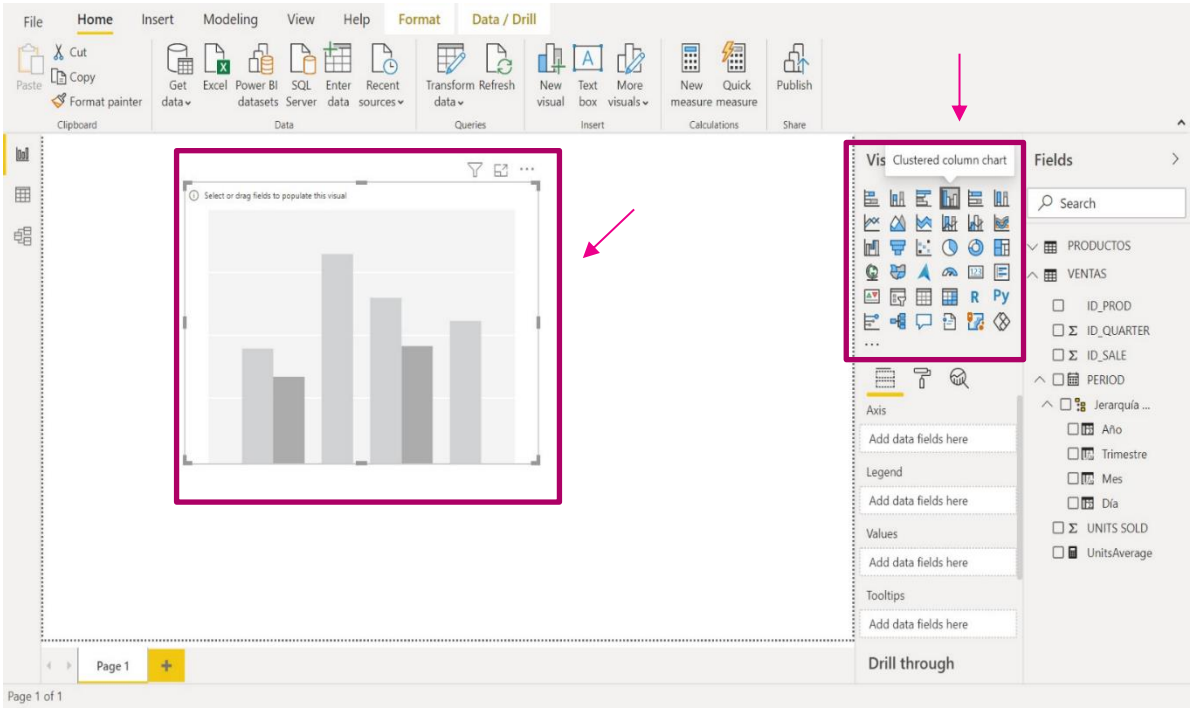


FIGURE 29. Selecting a visualization. Visual appears on the blank canvas. (Cuéllar 2020-08-24).

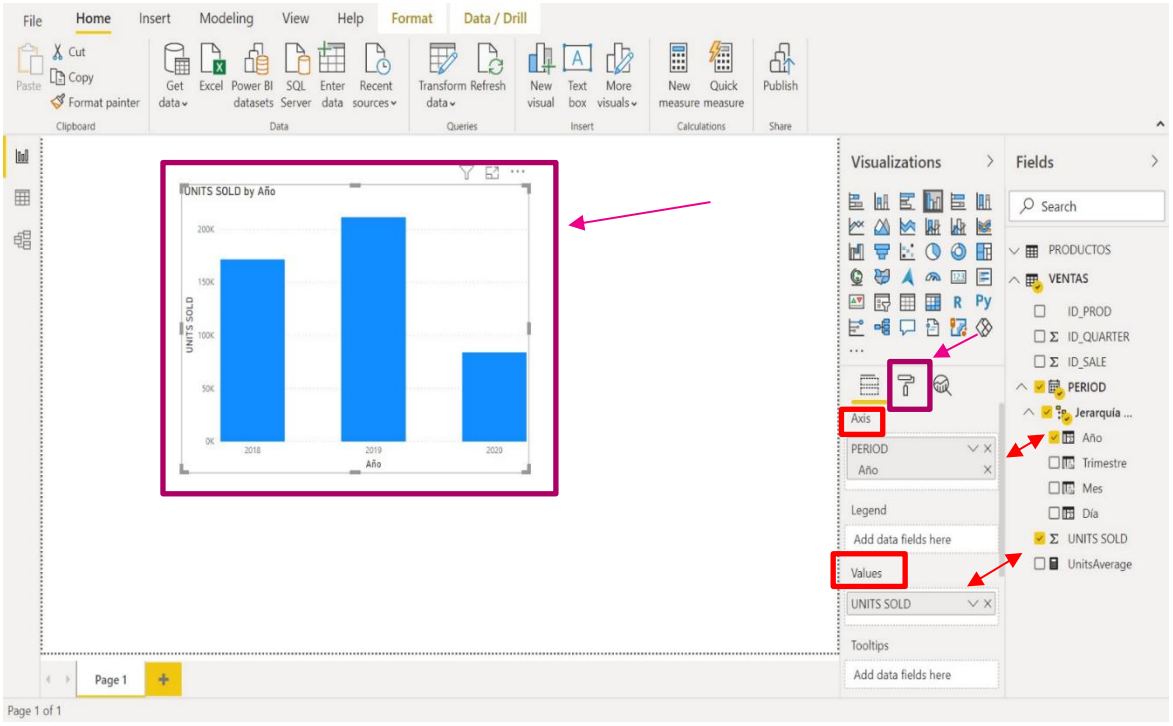


FIGURE 30. Dragging fields into Values and Axis sections to populate the bar chart. (Cuéllar 2020-08-24).

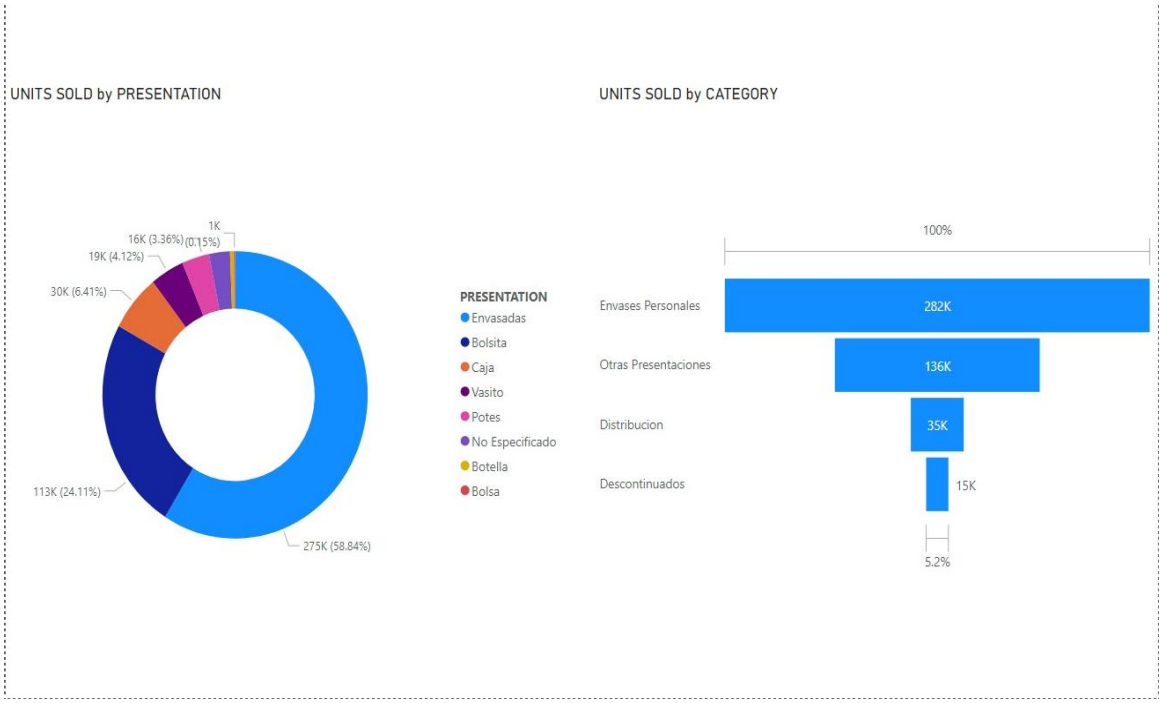


FIGURE 31. Pie and funnel charts. (Cuéllar 2020-08-24).

APPENDIX 5: BI DEVELOPED WORK USING A DATABASE AS DATA SOURCE

Stage 1 of the BI process: Data Cleansing for a Database

ID_NAME	NAME	ADDRESS	DISTRICT	TEL	ILLNESS?	PET?	CHILDREN
1	Jhoyma Perez	STREET1	San Isidro	111111111	No	NO	NO
2	Maura Boord	STREET2	Surco	111111112	No	NO	NO
3	Luis Velarde Aspillaga	STREET3	San Isidro	111111113	DIABETICO	NO	NO
4	Fiorella Carrano	STREET4	Surco	111111114	No	SI	NO
5	Magy Bertrand	STREET5	Surquillo	111111115	No	SI	NO
6	Cynthia Munailla	STREET6	Miraflores	111111116	No	NO	NO
7	Carlo Betetta	STREET7	Miraflores	111111117	No	NO	NO
8	Gabriela Vega	STREET8	Barranco	111111118	No	NO	NO
9	Inés Zegarra	STREET9	La Molina	111111119	No	NO	NO
10	Sigrid Bazán	STREET10	San Isidro	111111120	No	NO	NO
11	Liliana Trujillo	STREET11	Pueblo Libre	111111121	No	NO	NO
12	Christian Rodriguez	STREET12	La Molina	111111122	No	NO	NO
13	Alexandra Molina	STREET13	San Borja	111111123	No	NO	NO
14	Alexandra Molina	STREET14	Surco	111111124	No	NO	NO
15	Karla Arana	STREET15	Magdalena	111111125	No	SI	NO
16	Fiorella Klatich	STREET16	Miraflores	111111126	No	NO	NO
17	Cary Rutte	STREET17	Pueblo Libre	111111127	No	NO	NO
18	Roxana Delgado	STREET18	Magdalena	111111128	No	NO	NO
19	Tanya Valerio Saravia	STREET19	Surco	111111129	No	NO	NO
21	Jaghaira Enrriquez Torres	STREET20	Miraflores	111111130	No	SI	NO
22	María de Carmen Reátegui	STREET21	Lince	111111131	No	No	SI
23	Giuliana Zafra Ortiz	STREET22	Jesús María	111111132	No	No	NO
24	María Luisa Carrera	STREET23	San Borja	111111133	No	No	SI
25	Roberto Torpoco	STREET24	La Molina	111111134	No	No	NO
26	Mariel Yuyes	STREET25	Jesús María	111111135	No	No	NO
27	Fatmir Perez	STREET26	San Borja	111111136	No	No	NO
28	Dassia Castañeda	STREET27	Miraflores	111111137	No	No	NO
29	Tusy Arenas	STREET28	Surco	111111138	No	No	NO
30	Karina Bertello	STREET29	Miraflores	111111139	No	No	NO
31	Paola Pinto	STREET30	Jesús María	111111140	No	No	NO
32	Catherine Pacheco	STREET31	San Isidro	111111141	No	No	NO
33	Ana Cecilia Rosas	STREET32	San Borja	111111142	No	No	NO
34	Miguel Lazarte	STREET33	San Miguel	111111143	No	No	NO
35	Silvia Paz Macedo	STREET34	Surquillo	111111144	No	No	NO
36	Dennis Uribe	STREET35	Surco	111111145	No	No	NO
37	Diana Calderón	STREET36	Miraflores	111111146	No	No	NO
38	Adriana de La Flor Llona	STREET37	San Isidro	111111147	No	No	NO
39	Nérida Triveño	STREET38	San Borja	111111148	No	No	NO
40	Helena Vidal	STREET39	Surco	111111149	No	No	NO
41	Rosario Salazar	STREET40	Miraflores	111111150	No	No	NO
42	Rosa Valle Cortez	STREET41	Surco	111111151	No	No	NO
43	Patricia Meza	STREET42	La Molina	111111152	No	No	NO
44	Nel Boord	STREET43	San Borja	111111153	No	No	NO

PICTURE 32. Sheet 1. Delivery report for June. Clients' details. (Cuellar 2020-07-29).

VENTA DETALLE							
ID_NAME	ID_SALE	PRODUCT SOLD	QTY	BRAND	PRESENTATION	Sampling	PRICE
1	1	WHISKY CON CHOCOLATE	1	La Calaca	Bolsita		
1	1	RON CON PASAS	1	La Calaca	Bolsita		
1	1	FRESA RELLENA C/LECHE CONDENSADA	1	La Calaca	Bolsita		
1	1	LÚCUMA RELLENA C/CHOCOLATE	1	La Calaca	Bolsita		
1	1	CHOCOLATE RELLENA NUTELLA	1	La Calaca	Bolsita		
1	1	LÚCUMA CON BROWNIE RELLENA CHOCOLATE	1	La Calaca	Bolsita		
1	1	MENTA CON CHOCOLATE	1	La Calaca	Bolsita		
1	1	Chocolate c/ Almendras x 473 ml	1	La Gelaterie	473 ml		
1	1	Lúcuma c/ Brownie x 473 ml	1	La Gelaterie	473 ml		
2	2	LÚCUMA LIGHT CH	6	Paletti	Bolsita		
2	2	AFTER EIGHT (MENTA) CH	7	Paletti	Bolsita		
2	2	FRESA C/ MANGO CH	2	Paletti	Bolsita		
2	2	CHOCOLATE BITTER CH	1	Paletti	Bolsita		
2	2	CHOCOLATE RELLENA NUTELLA	1	La Calaca	Bolsita		
2	2	LÚCUMA CON BROWNIE RELLENA CHOCOLATE GRANDE	1	La Calaca	Bolsita		
2	2	Frambuesa x 473 ml	1	La Gelaterie	473 ml		
3	3	CHOCOLATE LIGHT CH	6	Paletti	Bolsita		
3	3	Chocolate Light GRANDE	6	La Gelaterie	110 ml		
3	3	Lúcuma Light	6	La Gelaterie	110 ml		
4	4	CHOCOLATE LIGHT	6	La Calaca	Bolsita		
4	4	LÚCUMA LIGHT	4	La Calaca	Bolsita		
4	4	FRESA CON MANGO 4 COLORES	4	La Calaca	Bolsita		
4	4	PALETA PERRO ZANAHORIA	1	La Calaca	Bolsita		
4	4	PALETA PERRO BANANA	1	La Calaca	Bolsita		
5	5	CHOCOLATE LIGHT	2	La Calaca	Bolsita		
5	5	LÚCUMA LIGHT	2	La Calaca	Bolsita		
5	5	DULCE DE LECHE RELLENA MANJAR	2	La Calaca	Bolsita		
5	5	COOKIES RELLENA C/CHOCOLATE GRANDE	2	La Calaca	Bolsita		
5	5	CHOCOLATE RELLENA C/PASTA DE MANI	2	La Calaca	Bolsita		
5	5	PALETA PERRO BANANA	1	La Calaca	Bolsita		
6	6	DULCE DE LECHE RELLENA MANJAR	3	La Calaca	Bolsita		
6	6	COOKIES RELLENA C/CHOCOLATE GRANDE	3	La Calaca	Bolsita		
6	6	CHOCOLATE RELLENA C/PASTA DE MANI	3	La Calaca	Bolsita		
7	7	LÚCUMA RELLENA C/CHOCOLATE	2	La Calaca	Bolsita		

PICTURE 33. Sheet 2. Delivery report for June. Products' details. (Cuellar 2020-07-29).

	A	B	C	D	E	F
1						
2		ID_NAME	ID_SALE	DATE	AMOUNT	PAYMENT METHOD
3	1	3	3	07/04/2020	169.24	Transferencia
4	2	4	4	14/04/2020	256.8	Transferencia
5	3	5	5	15/04/2020		Transferencia
6	4	7	7	15/04/2020	103.2	Transferencia
7	5	6	6	16/04/2020		Transferencia
8	6	8	8	16/04/2020	122.8	Transferencia
9	7	1	1	21/04/2020	115.8	Transferencia
10	8	2	2	23/04/2020	239.32	Transferencia
11	9	9	9	23/04/2020		Transferencia
12	10	10	10	23/04/2020	79.7	Transferencia
13	11	11	11	23/04/2020	156.4	Transferencia
14	12	12	12	23/04/2020	96.8	Transferencia
15	13	13	13	23/04/2020	166	Yape
16	14	14	14	23/04/2020		Yape
17	15	15	15	23/04/2020	76.5	Transferencia
18	16	16	16	24/04/2020	85.7	Transferencia
19	17	17	17	24/04/2020	87	Transferencia
20	18	18	18	24/04/2020	79.7	Transferencia
21	19	19	19	24/04/2020	93	Transferencia
22	20	4	20	24/04/2020	106.65	Transferencia
23	21	21	21	24/04/2020	97	Transferencia
24	22	22	22	24/04/2020	81.1	Transferencia
25	23	23	23	24/04/2020	107.6	Transferencia
26	24	24	24	24/04/2020	136.06	Transferencia
27	25	25	25	24/04/2020	78.9	Transferencia
28	26	26	26	24/04/2020	86	Transferencia
29	27	27	27	24/04/2020	113.6	Transferencia
30	28	28	28	24/04/2020	86.8	Transferencia
31	29	29	29	24/04/2020		NO IDENTIFICADO
32	30	30	30	24/04/2020	140	Transferencia
33	31	31	31	24/04/2020	130.4	Transferencia
34	32	32	32	24/04/2020	80.7	Transferencia

PICTURE 34. Sheet 3. Delivery report for June. Billing information. (Cuellar 2020-07-29).

Stage 1 of the BI process: Assigning Unique Identifiers to Support Data Cleansing for a Database

TRANSFERENCIA	1
YAPE	2
EFECTIVO	3
CONTRAENTREGA	4
DEPOSITO	5
NO IDENTIFICADO	6

	DISTRICT
1	San Isidro
2	Surco
3	Surquillo
4	Miraflores
5	Barranco
6	La Molina
7	Pueblo Libre
8	San Borja
9	Magdalena
10	Lince
11	Jesús María
12	San Miguel
13	San Luis
14	Chorrillos
15	La Victoria
16	Ate
17	Villa María del Triunfo
18	Los Olivos
19	San Juan de Miraflores
20	Callao
21	San Martin de Porres

BRAND	
LA CALACA	1
LA GELATERIE	2
PALETTI	3
SMART CHOICE	4
RELLEPAN	5
NO CONVIDO	6

PRESENTATION	
BOLSITA	1
CAJA	2
ENVASADO	3
VASITO	4

PICTURE 35. Unique identifiers set for Districts, Brand, and Presentation. (Cuéllar 2020-07-

ID_NAME	NAME	ADDRESS	DISTRICT	ID_DISTRICT	TEL
1	Jhoyma Perez	STREET1	San Isidro	1	111111111
2	Maura Boord	STREET2	Surco	2	111111112
3	Luis Velarde Aspillaga	STREET3	San Isidro	1	111111113
4	Fiorella Carrano	STREET4	Surco	2	111111114
5	Magy Bertrand	STREET5	Surquillo	3	111111115
6	Cynthia Munailla	STREET6	Miraflores	4	111111116
7	Carlo Betetta	STREET7	Miraflores	4	111111117
8	Gabriela Vega	STREET8	Barranco	5	111111118
9	Inés Zegarra	STREET9	La Molina	6	111111119
10	Sigrid Bazán	STREET10	San Isidro	1	111111120
11	Liliana Trujillo	STREET11	Pueblo Libre	7	111111121
12	Christian Rodriguez	STREET12	La Molina	6	111111122
13	Alexandra Molina	STREET13	San Borja	8	111111123
14	Alexandra Molina	STREET14	Surco	2	111111124
15	Karla Arana	STREET15	Magdalena	9	111111125
16	Fiorella Klatich	STREET16	Miraflores	4	111111126
17	Cary Rutte	STREET17	Pueblo Libre	7	111111127
18	Roxana Delgado	STREET18	Magdalena	9	111111128
19	Tanya Valerio Saravia	STREET19	Surco	2	111111129
21	Jaghaira Enriquez Torres	STREET20	Miraflores	4	111111130
22	María de Carmen Reátegui	STREET21	Lince	10	111111131
23	Giuliana Zafra Ortiz	STREET22	Jesús María	11	111111132
24	Maria Luisa Carrera	STREET23	San Borja	8	111111133
25	Roberto Torpoco	STREET24	La Molina	6	111111134
26	Mariel Yuyes	STREET25	Jesús María	11	111111135
27	Fatmir Perez	STREET26	San Borja	8	111111136
28	Dassia Castañeda	STREET27	Miraflores	4	111111137
29	Tusy Arenas	STREET28	Surco	2	111111138
30	Karina Bertello	STREET29	Miraflores	4	111111139
31	Paola Pinto	STREET30	Jesús María	11	111111140
32	Catherine Pacheco	STREET31	San Isidro	1	111111141
33	Ana Cecilia Rosas	STREET32	San Borja	8	111111142
34	Miguel Lazarte	STREET33	San Miguel	12	111111143

PICTURE 36. FIRST WORKSHEET, clients' data with each district categorized by a unique number. (Cuéllar 2020-

ID_CONDICION	ILLNESS?	PET 0 OR 1	PET?	CHILDREN 0 OR 1	CHILDREN?
1	No	1	NO	1	NO
1	No	1	NO	1	NO
2	DIABETICO	1	NO	1	NO
1	No	0	SI	1	NO
1	No	0	SI	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	0	SI	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	1	NO	1	NO
1	No	0	SI	1	NO
1	No	1	No	0	SI
1	No	1	No	1	NO
1	No	1	No	0	SI
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO
1	No	1	No	1	NO

PICTURE 37. FIRST WORKSHEET, three last questions of the clients' data using the corresponding data types. (Cuéllar

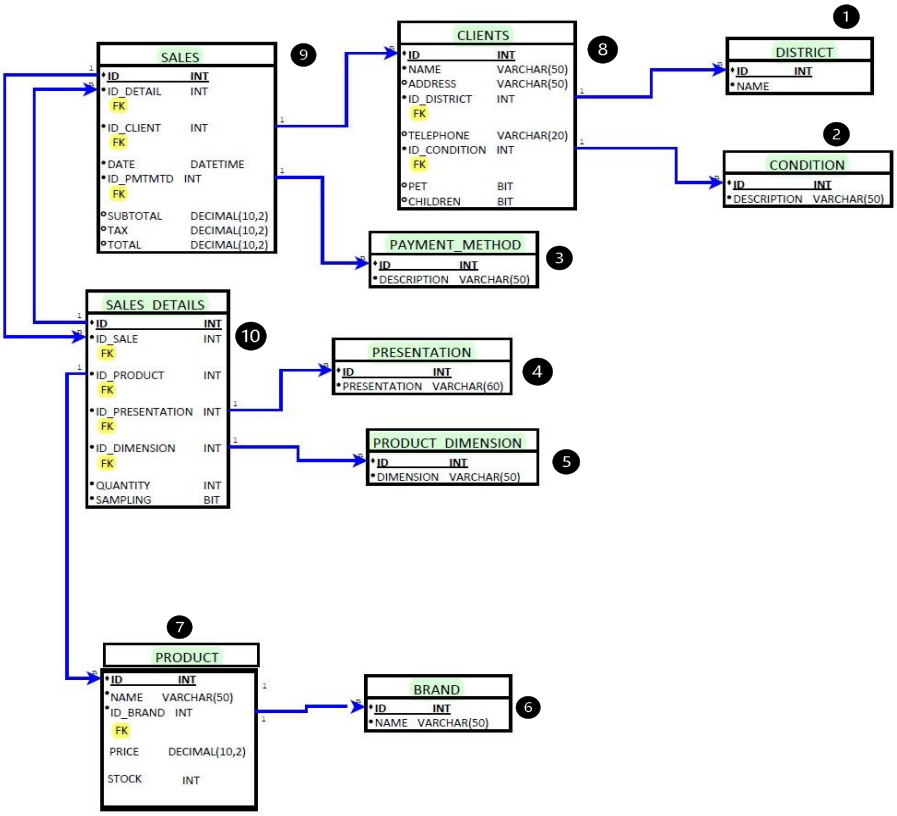
Stage 2 of the BI process: Relations and Tables

ID_NAME	ID_SALE	ID_PRODUCTO	PRODUCT SOLD	QTY	BRAND	ID_MARCA	PRESENTATION	ID_PRESENTACION	Sampling	SAMPLIN	PRICE
1	1	59	WHISKY CON CHOCOLATE	1	La Calaca	1	Bolsita	1		0	
1	1	53	RON CON PASAS	1	La Calaca	1	Bolsita	1		0	
1	1	40	FRESA RELLENA C/LECHE CONDENSADA	1	La Calaca	1	Bolsita	1		0	
1	1	15	LÚCUMA RELLENA C/CHOCOLATE	1	La Calaca	1	Bolsita	1		0	
1	1	34	CHOCOLATE RELLENA NUTELLA	1	La Calaca	1	Bolsita	1		0	
1	1	41	LÚCUMA CON BROWNIE RELLENA CHOCOLATE	1	La Calaca	1	Bolsita	1		0	
1	1	46	MENTA CON CHOCOLATE	1	La Calaca	1	Bolsita	1		0	
1	1	32	Chocolate c/ Almendras x 473 ml	1	La Gelaterie	2	473 ml	2		0	
1	1	13	Lúcuma c/ Brownie x 473 ml	1	La Gelaterie	2	473 ml	2		0	
2	2	14	LÚCUMA LIGHT CH	6	Paletti	3	Bolsita	1		0	
2	2	1	AFTER EIGHT (MENTA) CH	7	Paletti	3	Bolsita	1		0	
2	2	8	FRESA C/ MANGO CH	2	Paletti	3	Bolsita	1		0	
2	2	2	CHOCOLATE BITTER CH	1	Paletti	3	Bolsita	1		0	
2	2	34	CHOCOLATE RELLENA NUTELLA	1	La Calaca	1	Bolsita	1		0	
2	2	41	LÚCUMA CON BROWNIE RELLENA CHOCOLATE GRANDE	1	La Calaca	1	Bolsita	1		0	
2	2	7	Frambuesa x 473 ml	1	La Gelaterie	2	473 ml	2		0	
3	3	3	CHOCOLATE LIGHT CH	6	Paletti	3	Bolsita	1		0	
3	3	3	Chocolate Light GRANDE	6	La Gelaterie	2	110 ml	3		0	
3	3	14	Lúcuma Light	6	La Gelaterie	2	110 ml	3		0	
4	4	3	CHOCOLATE LIGHT	6	La Calaca	1	Bolsita	1		0	
4	4	14	LÚCUMA LIGHT	4	La Calaca	1	Bolsita	1		0	
4	4	39	FRESA CON MANGO 4 COLORES	4	La Calaca	1	Bolsita	1		0	
4	4	48	PALETA PERRO ZANAHORIA	1	La Calaca	1	Bolsita	1		0	
4	4	47	PALETA PERRO BANANA	1	La Calaca	1	Bolsita	1		0	
5	5	3	CHOCOLATE LIGHT	2	La Calaca	1	Bolsita	1		0	
5	5	14	LÚCUMA LIGHT	2	La Calaca	1	Bolsita	1		0	
5	5	38	DULCE DE LECHE RELLENA MANJAR	2	La Calaca	1	Bolsita	1		0	
5	5	37	COOKIES RELLENA C/CHOCOLATE GRANDE	2	La Calaca	1	Bolsita	1		0	
5	5	33	CHOCOLATE RELLENA C/PASTA DE MANI	2	La Calaca	1	Bolsita	1		0	
5	5	47	PALETA PERRO BANANA	1	La Calaca	1	Bolsita	1		0	
6	6	38	DULCE DE LECHE RELLENA MANJAR	3	La Calaca	1	Bolsita	1		0	
6	6	37	COOKIES RELLENA C/CHOCOLATE GRANDE	3	La Calaca	1	Bolsita	1		0	
6	6	33	CHOCOLATE RELLENA C/PASTA DE MANI	3	La Calaca	1	Bolsita	1		0	
7	7	15	LÚCUMA RELLENA C/CHOCOLATE	2	La Calaca	1	Bolsita	1		0	
7	7	41	LÚCUMA CON BROWNIE RELLENA CHOCOLATE GRANDE	2	La Calaca	1	Bolsita	1		0	
7	7	37	COOKIES RELLENA C/CHOCOLATE GRANDE	2	La Calaca	1	Bolsita	1		0	
7	7	8	FRESA C/ MANGO CH	2	Paletti	3	Bolsita	1		0	
7	7	11	KIWI FRAMBUESA Y MANDARINA GR	2	Paletti	3	Bolsita	1		0	
7	7	6	DOBLE CHOCOLATE CH	2	Paletti	3	Bolsita	1		0	
8	8	39	FRESA CON MANGO 4 COLORES	2	La Calaca	1	Bolsita	1		0	
8	8	40	FRESA RELLENA C/LECHE CONDENSADA GRANDE	2	La Calaca	1	Bolsita	1		0	
8	8	15	LÚCUMA RELLENA C/CHOCOLATE	2	La Calaca	1	Bolsita	1		0	
8	8	33	CHOCOLATE RELLENA C/PASTA DE MANI	2	La Calaca	1	Bolsita	1		0	

PICTURE 38. SECOND WORKSHEET, sampling and unit prices are missing. (Cuéllar 2020-07-29).

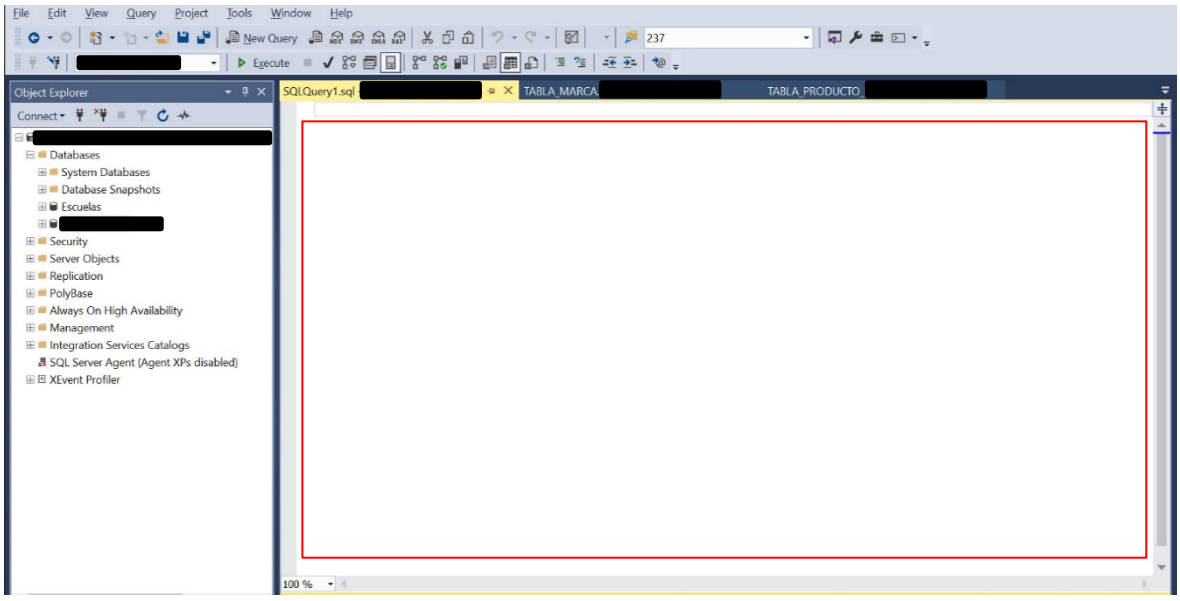
ID_NAME	ID_SALE	DATE	AMOUNT	PAYMENT METHOD	ID_PAYMENT
3	3	07/04/2020	169.24	Transferencia	1
4	4	14/04/2020	256.8	Transferencia	1
5	5	15/04/2020		Transferencia	1
7	7	15/04/2020	103.2	Transferencia	1
6	6	16/04/2020		Transferencia	1
8	8	16/04/2020	122.8	Transferencia	1
1	1	21/04/2020	115.8	Transferencia	1
2	2	23/04/2020	239.32	Transferencia	1
9	9	23/04/2020		Transferencia	1
10	10	23/04/2020	79.7	Transferencia	1
11	11	23/04/2020	156.4	Transferencia	1
12	12	23/04/2020	96.8	Transferencia	1
13	13	23/04/2020	166	Yape	2
14	14	23/04/2020		Yape	2
15	15	23/04/2020	76.5	Transferencia	1
16	16	24/04/2020	85.7	Transferencia	1
17	17	24/04/2020	87	Transferencia	1
18	18	24/04/2020	79.7	Transferencia	1
19	19	24/04/2020	93	Transferencia	1
4	20	24/04/2020	106.65	Transferencia	1
21	21	24/04/2020	97	Transferencia	1
22	22	24/04/2020	81.1	Transferencia	1
23	23	24/04/2020	107.6	Transferencia	1
24	24	24/04/2020	136.06	Transferencia	1
25	25	24/04/2020	78.9	Transferencia	1
26	26	24/04/2020	86	Transferencia	1
27	27	24/04/2020	113.6	Transferencia	1
28	28	24/04/2020	86.8	Transferencia	1

PICTURE 39. THIRD WORKSHEET, client's details with each sale's details by connecting IDs. (Cuéllar 2020-07-

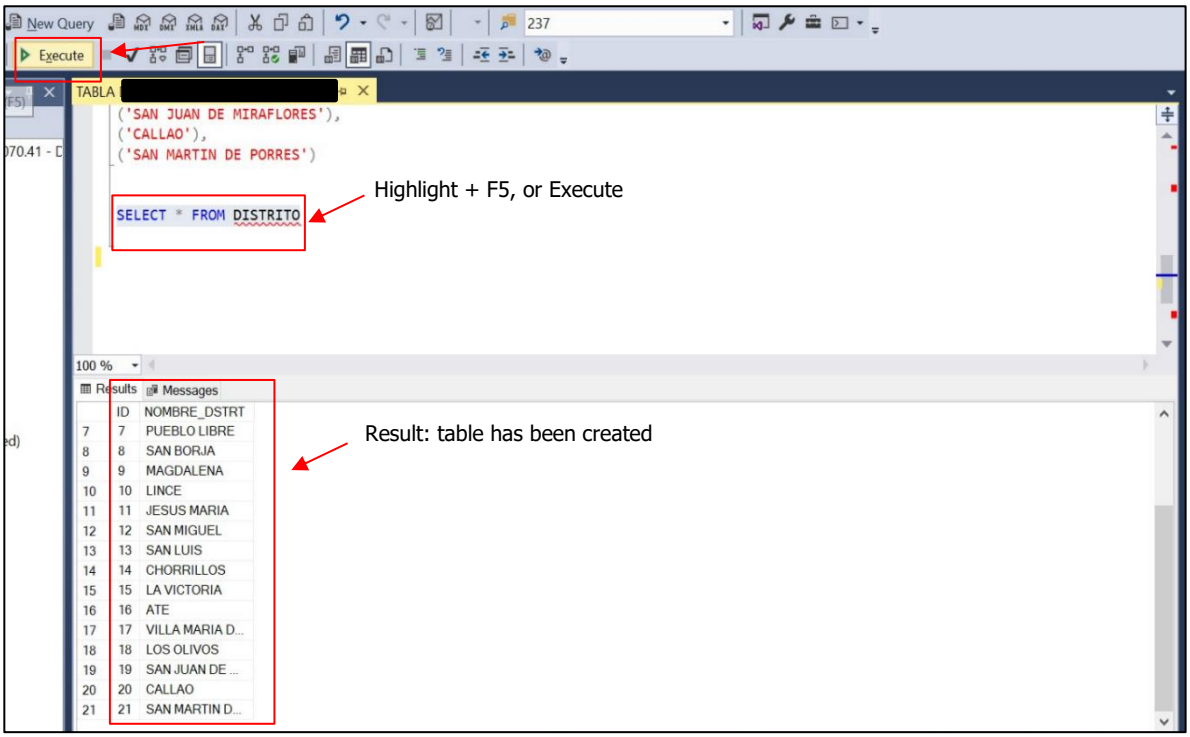


PICTURE 40. How the relationships from the dataset have been created. (Cuéllar 2020-07-29).

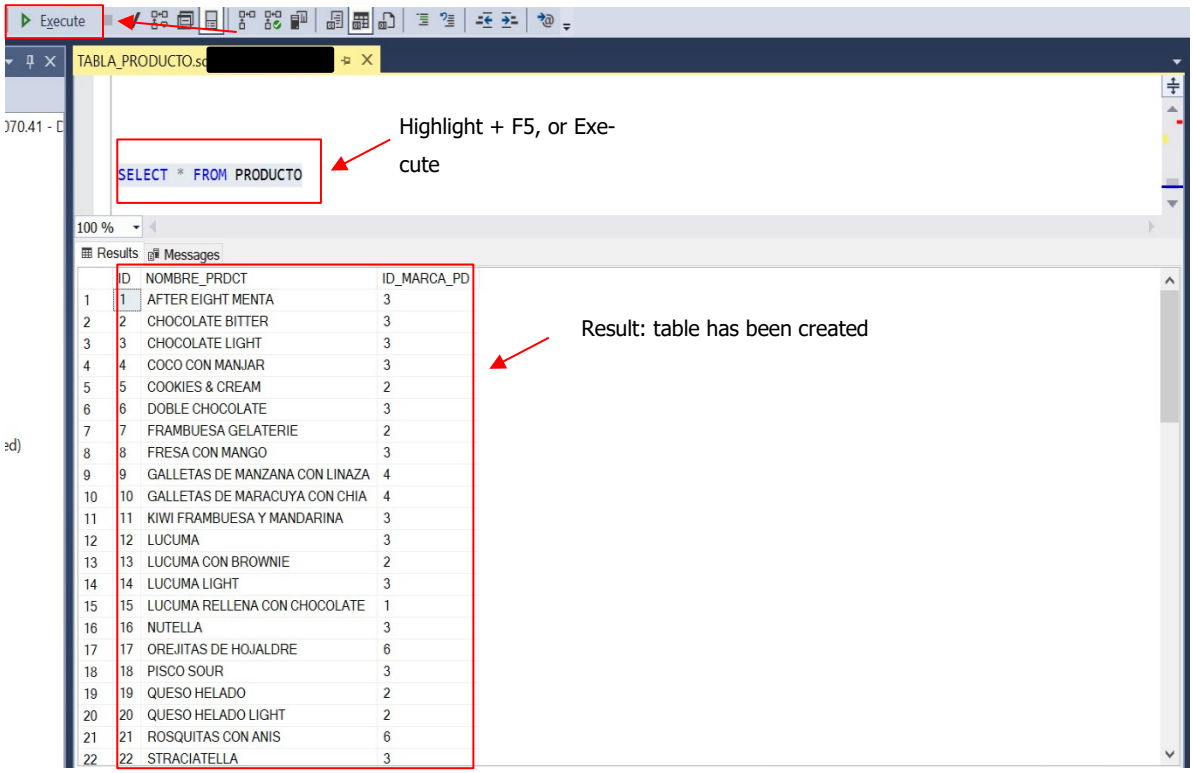
Stage 3 of the BI process: Creating a Database in SQL Server Management Studio



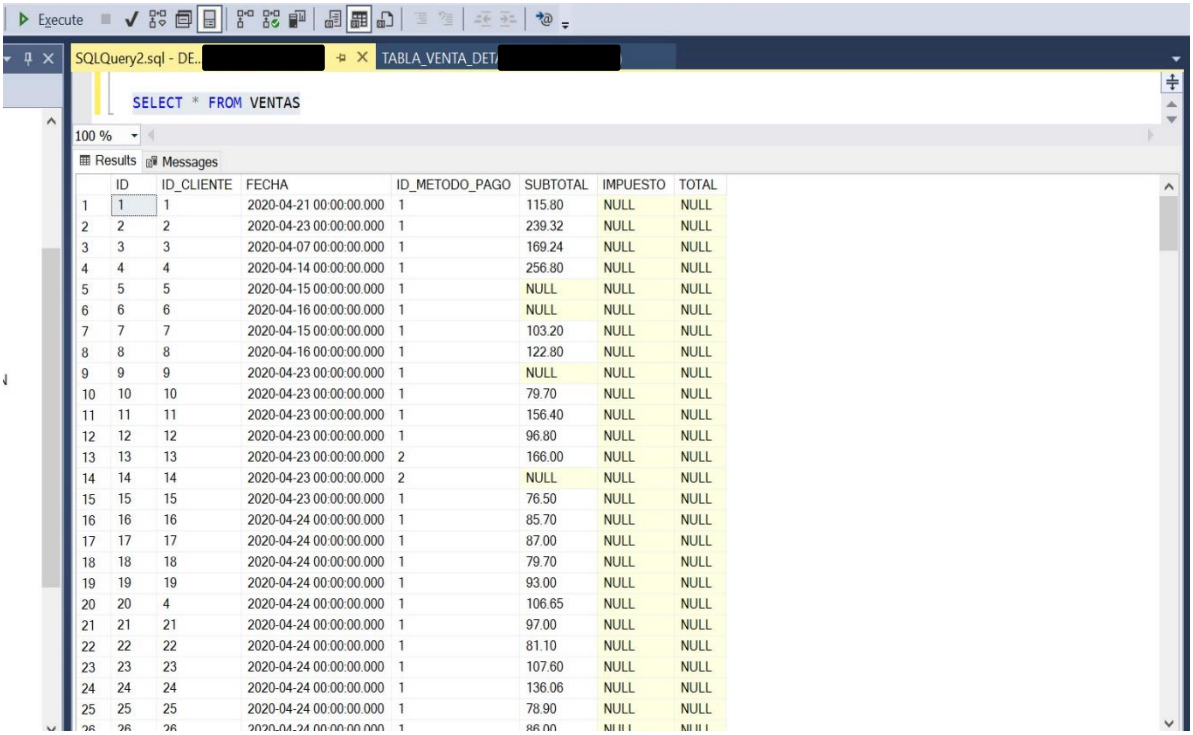
PICTURE 41. SQL Server Management Studio Canvas. (Cuéllar 2020-07-29).



PICTURE 42. Executing a code and its result shown as a table. (Cuéllar 2020-07-29).



PICTURE 43. The table for PRODUCTS has been created. (Cuéllar 2020-07-29).



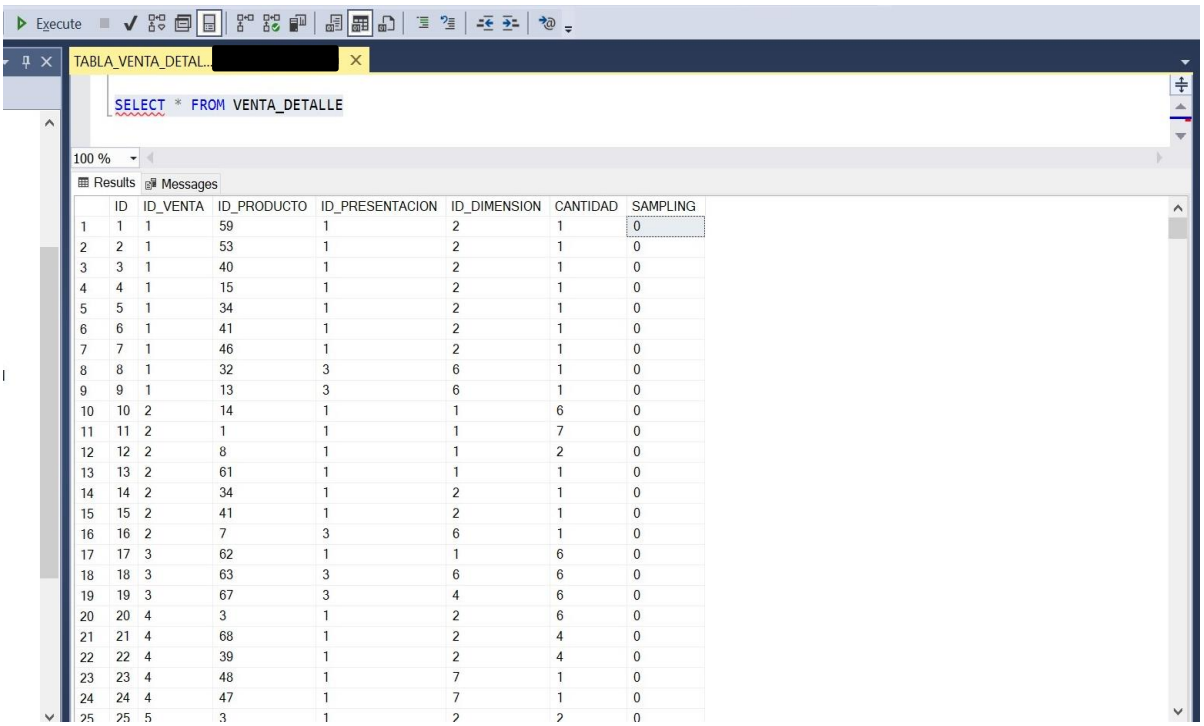
```
SELECT * FROM VENTAS
```

100 %

Results | Messages

	ID	ID_CLIENTE	FECHA	ID_METODO_PAGO	SUBTOTAL	IMPUESTO	TOTAL
1	1	1	2020-04-21 00:00:00.000	1	115.80	NULL	NULL
2	2	2	2020-04-23 00:00:00.000	1	239.32	NULL	NULL
3	3	3	2020-04-07 00:00:00.000	1	169.24	NULL	NULL
4	4	4	2020-04-14 00:00:00.000	1	256.80	NULL	NULL
5	5	5	2020-04-15 00:00:00.000	1	NULL	NULL	NULL
6	6	6	2020-04-16 00:00:00.000	1	NULL	NULL	NULL
7	7	7	2020-04-15 00:00:00.000	1	103.20	NULL	NULL
8	8	8	2020-04-16 00:00:00.000	1	122.80	NULL	NULL
9	9	9	2020-04-23 00:00:00.000	1	NULL	NULL	NULL
10	10	10	2020-04-23 00:00:00.000	1	79.70	NULL	NULL
11	11	11	2020-04-23 00:00:00.000	1	156.40	NULL	NULL
12	12	12	2020-04-23 00:00:00.000	1	96.80	NULL	NULL
13	13	13	2020-04-23 00:00:00.000	2	166.00	NULL	NULL
14	14	14	2020-04-23 00:00:00.000	2	NULL	NULL	NULL
15	15	15	2020-04-23 00:00:00.000	1	76.50	NULL	NULL
16	16	16	2020-04-24 00:00:00.000	1	85.70	NULL	NULL
17	17	17	2020-04-24 00:00:00.000	1	87.00	NULL	NULL
18	18	18	2020-04-24 00:00:00.000	1	79.70	NULL	NULL
19	19	19	2020-04-24 00:00:00.000	1	93.00	NULL	NULL
20	20	4	2020-04-24 00:00:00.000	1	106.65	NULL	NULL
21	21	21	2020-04-24 00:00:00.000	1	97.00	NULL	NULL
22	22	22	2020-04-24 00:00:00.000	1	81.10	NULL	NULL
23	23	23	2020-04-24 00:00:00.000	1	107.60	NULL	NULL
24	24	24	2020-04-24 00:00:00.000	1	136.06	NULL	NULL
25	25	25	2020-04-24 00:00:00.000	1	78.90	NULL	NULL

PICTURE 44. Visualizing or calling the table VENTAS (SALES). (Cuéllar 2020-07-29).



```
SELECT * FROM VENTA_DETALLE
```

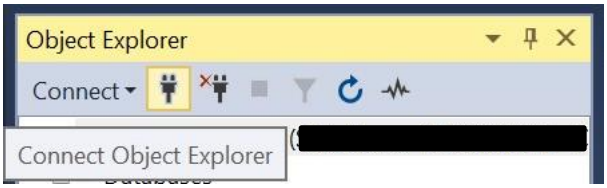
100 %

Results | Messages

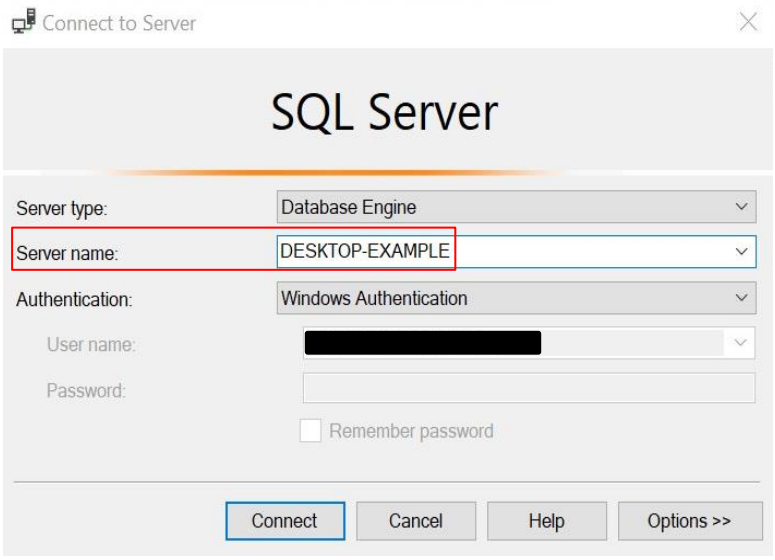
	ID	ID_VENTA	ID_PRODUCTO	ID_PRESENTACION	ID_DIMENSION	CANTIDAD	SAMPLING
1	1	1	59	1	2	1	0
2	2	1	53	1	2	1	0
3	3	1	40	1	2	1	0
4	4	1	15	1	2	1	0
5	5	1	34	1	2	1	0
6	6	1	41	1	2	1	0
7	7	1	46	1	2	1	0
8	8	1	32	3	6	1	0
9	9	1	13	3	6	1	0
10	10	2	14	1	1	6	0
11	11	2	1	1	1	7	0
12	12	2	8	1	1	2	0
13	13	2	61	1	1	1	0
14	14	2	34	1	2	1	0
15	15	2	41	1	2	1	0
16	16	2	7	3	6	1	0
17	17	3	62	1	1	6	0
18	18	3	63	3	6	6	0
19	19	3	67	3	4	6	0
20	20	4	3	1	2	6	0
21	21	4	68	1	2	4	0
22	22	4	39	1	2	4	0
23	23	4	48	1	7	1	0
24	24	4	47	1	7	1	0
25	25	5	3	1	2	2	0

PICTURE 45. Visualizing or calling the table VENTA DETALLE (SALES_DETAILS). (Cuéllar 2020-07-29).

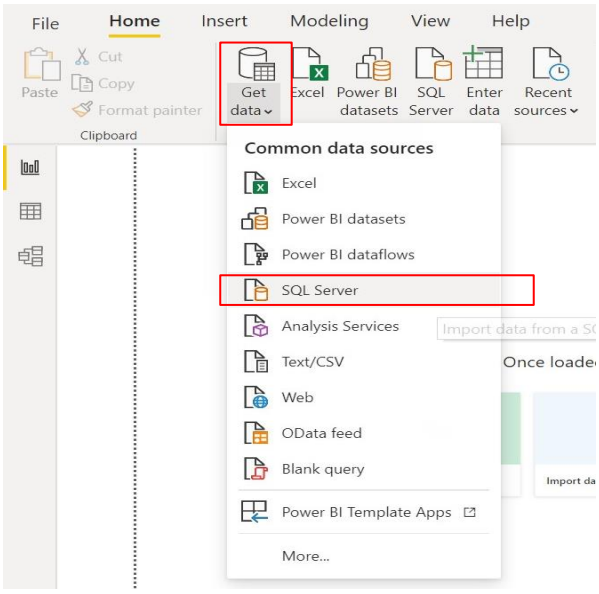
Stage 4 of the BI process: Connecting a Database to Power BI for Data Visualization



PICTURE 46. Object Explorer section in SQL Server Management Studio. (Cuéllar 2020-07-29).

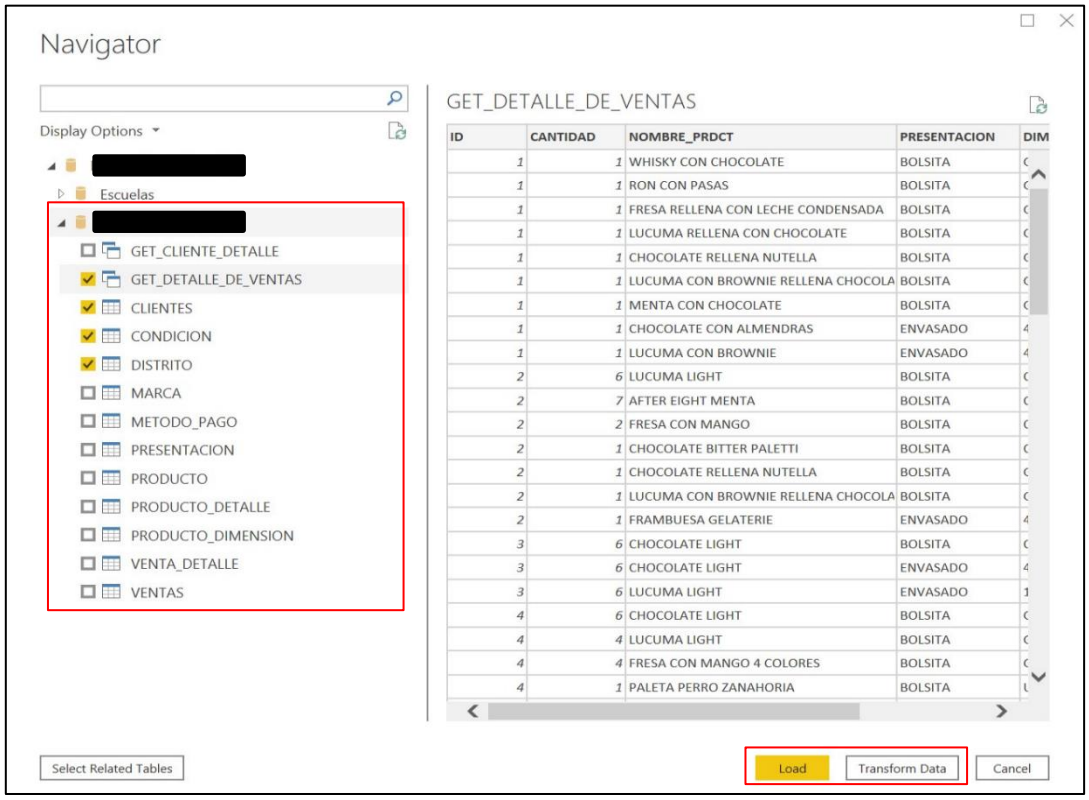


PICTURE 47. Connect to Server window. Displays server name. (Cuéllar 2020-07-29).

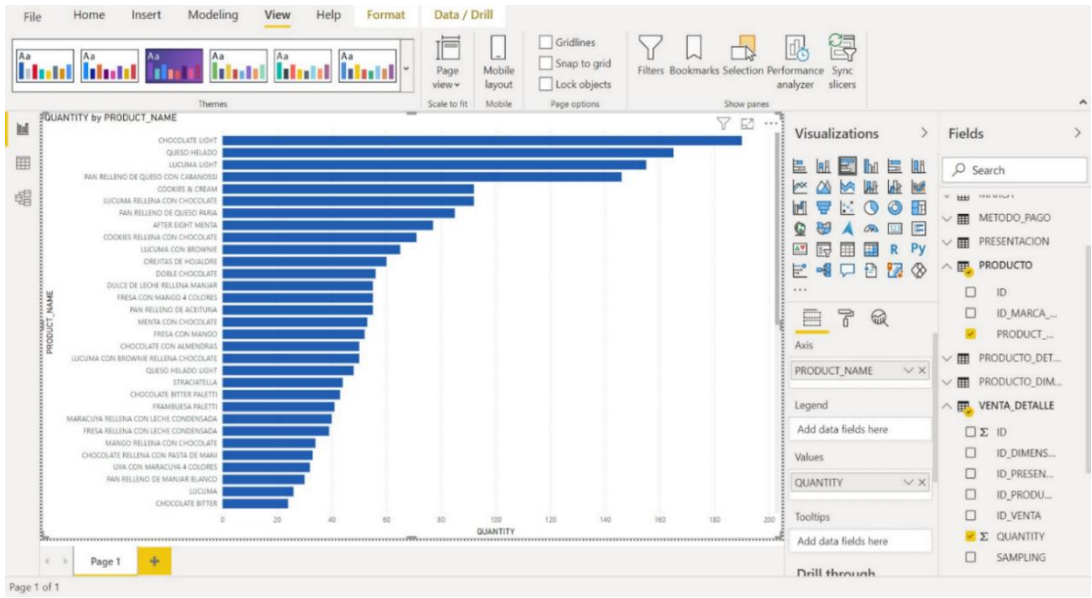


PICTURE 48. Connecting a database to Power BI. Choosing source type. (Cuéllar 2020-07-29).

PICTURE 49. Connecting a database to Power BI. Introducing server name. (Cuéllar 2020-07-29).

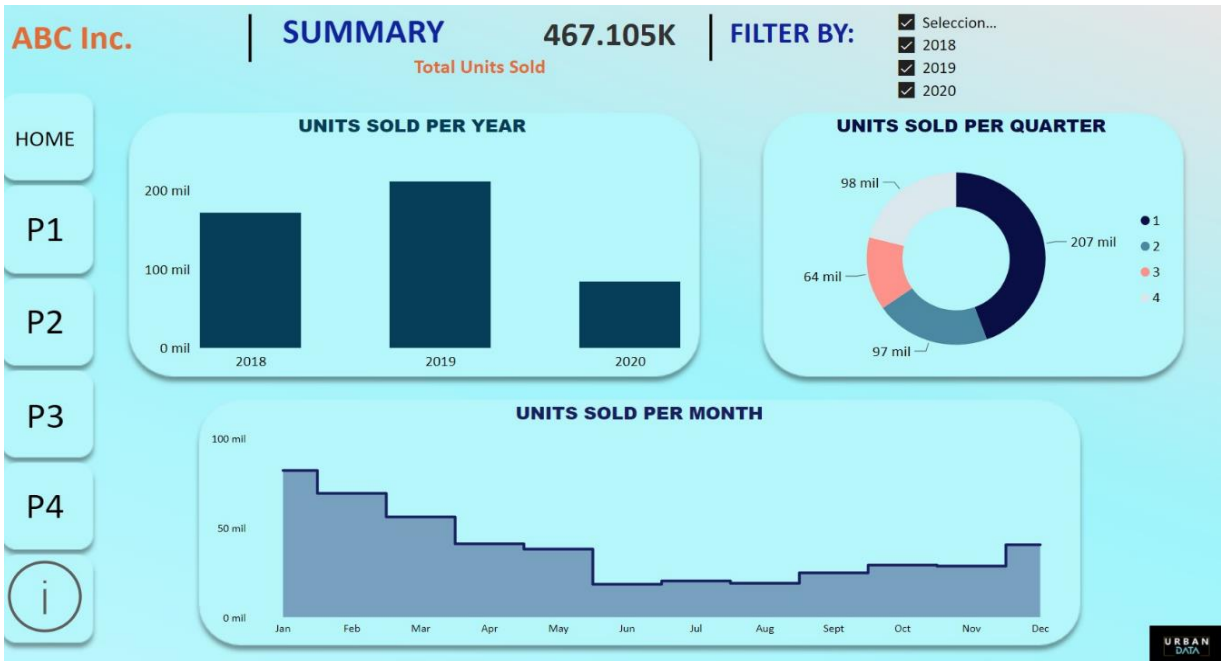


PICTURE 50. Selecting the tables to use from the database. (Cuéllar 2020-07-29).



PICTURE 51. Clustered bar chart to analyze the sales of the sales channel "Delivery". Quantity by Product Name. (Cuéllar 2020-07-29).

APPENDIX 6: DISCUSSION



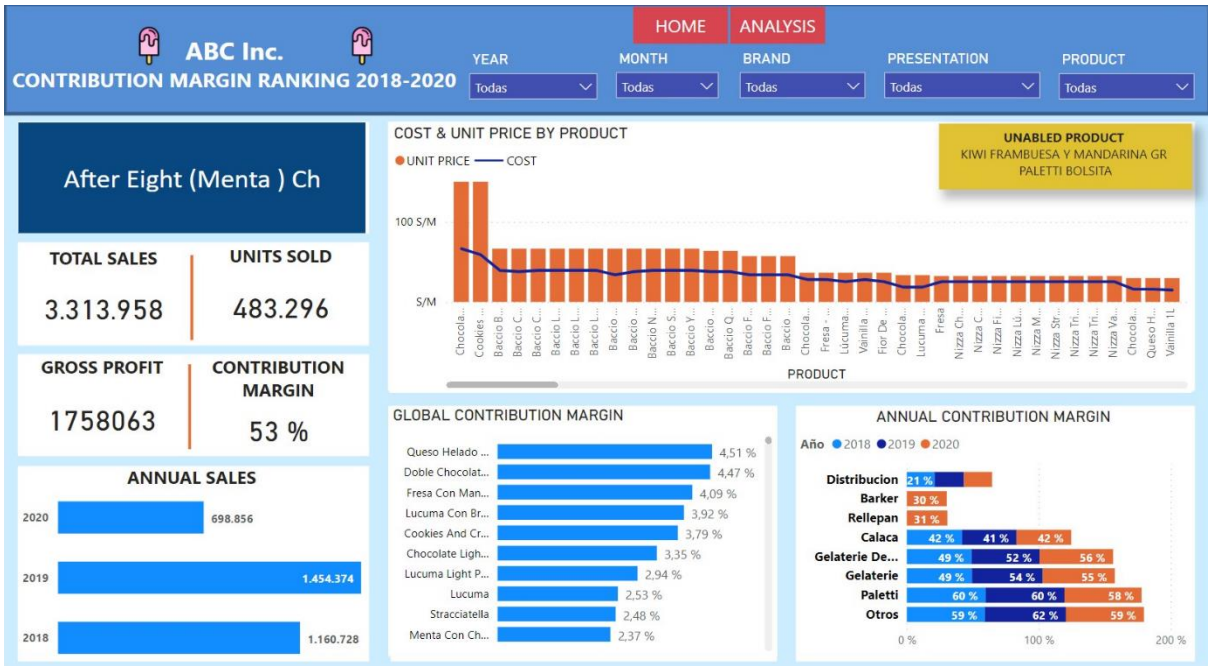
PICTURE 51. BI REPORT #1. (Cuéllar 2020-07-29).

PICTURE 52A. BI report #2, page 1/3. (Cuéllar 2020-08-15).

PICTURE 52B. BI report #2, page 2/3. (Cuéllar 2020-08-15).

PICTURE 52C. BI report #2, page 3/3. (Cuéllar 2020-08-15).

PICTURE 53A. BI report #3, page 1/3. (Cuéllar 2020-09-01).



PICTURE 53B. BI report #3, page 2/3. (Cuéllar 2020-09-01).

PICTURE 53C. BI report #3, page 3/3. (Cuéllar 2020-09-01).

PICTURE 54A. BI report #4, page 1/2. (Cuéllar 2020-09-15).

PICTURE 54B. BI report #4, page 2/2. (Cuéllar 2020-09-15).