Msc. Data Science and Advanced Analytics
Catarina Candeias, m20200656
Catarina Urbano, m20200607
João Henriques, m20200670
Margarida Pereira, m20201038
Rita Ferreira, m20200661

2021

BUSINESS INTELLIGENCE PROJECT

Land of Cars

INDEX

INTRODUCTION

As in the world of cars, there are thousands of pieces available, we thought that it would be wise to find a dataset for this topic as it would provide a large dataset with enough attributes to work with. After a huge search, we obtained a good dataset of an Italian store that sells for each car model the wanted parts.

Having the proper vehicle makes driving even more enjoyable. Some car parts are mandatory, others are only optional, depending on the driver's purpose.

This Italian manufacturer has an in-depth selection of superior car bulbs, brake pads, spark plugs, and other automotive parts designed to keep every car running at its absolute best.

This report explains in detail our procedure and how the business objective was achieved.

We are going to approach different techniques taught in the *Business Intelligence* Course, making use of the *Power BI*, especially (*Power*) *Dax, Power Query*, and *Power Editor*, and we will try to use domain and research knowledge in the mechanics' field.

PRESENTATION OF BUSINESS

Land of Cars is a B2B (Business to Business) company that sells car parts. Its main goal is to serve large car companies (such as Mercedes, Jaguar, and Ferrari) as well as local mechanical workshops – produces from the simplest screws to top-of-the-range engines, with the purpose of being used in the construction and reparation of automobiles. It is known for its high quality since its inception, 1920, as also for its huge sales volumes.

Although the company has existed since the beginning of the automotive business, it only started to collect data in 2005, and the available dataset presents data only about sales between 2005 and 2011.

These sales can be explained by the variety of its produced parts and its quality – 3464 car parts, more specifically.

These parts are produced in two countries – USA and Italy. Each country has 5 different factories (*Production Plants*), where different products are produced with different costs. Moreover, each product can also have 3 different series - P1, P2, or/and P3 - depending on your evolution level.

Although produced in Italy or the USA, the products are shipped to most of the world and in several different segments, such as the *Automotive, Distributor, CEMS, Consumer, Industrial, Lighting*, and *Power*. Anyway, the Industrial segment represents the vast majority, comprising approximately 37% of its clients.

Beyond that, it is possible to extract some insights regarding the salesmen, as it was provided information regarding their *Gender, Marital status, Age, Hire date,* and *PM*, *People Manager*.

Land of Cars, the freedom of travelling

BUSINESS NEEDS

The main objective of this Business Intelligence system is to improve the business processes and enable data-driven decision-making, always keeping in mind that it is important to have it aligned with the business, mainly its objectives and goals (business needs), vision, problems, among others.

No. 01 —Business Needs

- Increase Profit Margin;
- Streamline management decision-making process;
- Revise Financial projections;
- Understand the customers' preferences and behavior;
- Get insights into the products' catalog and their performance;
- To be able to write effective business reports;
- Reduce operational costs;
- Optimize the use of resources.

This company started off as a small business that would only offer a small variety of products. With the high increase in demand, the company was forced to keep up with it having to invest more in personnel and machinery, which later led to an increase in both product variety and assortment, accompanied by the creation of several teams and departments. However, its information systems were not able to sustain this growth: the traditional means were still used to manage massive volumes of data (ex: *static spreadsheets* and *PowerPoint* presentations), basically, data was not seen as an asset but rather as something costly to maintain and that is dispersed through a number of spreadsheets and databases from which is hard to measure the business performance and report it in real-time.

No. 02 — Business Problems

- Data is spread across multiple sources;
- The company is using information from old reports to understand market behavior;
- It is too time-consuming to report results. The company's ability to collect data is not enough to succeed, it is also required to turn it into valuable information for the business. This takes a significant amount of time to do, meaning that much likely by the time we get good insights, they will be already outdated;
- The company is not able to monitor its day-to-day activity in an efficient way which makes it difficult to determine its level of success and to keep up with the competition;
- In the current situation, to get even small insights of the data, it is required high IT involvement, which is a step back when managing productivity and operations;
- It is very difficult for the company to know for sure what are the most profitable and demanded products;
- There is disparate information at different levels of the company.

DESCRIPTION OF THE SOURCE DATA AND DISCOVERY PROCESS

As we know, the quality of the data is a crucial factor in order to obtain great insights and conduct a good analysis according to the stipulated business goals. For this reason, the two sources of data used to develop our research were in two different formats, *Excel* and *CSV*, whose labels are *AUTOMOTIVE* and *DATES*, respectively.

Regarding the first source of data mentioned previously, *AUTOMOTIVE*, refers to an Italian automobile company, whose products vary from some simple nails to mobile vehicle arrangements tools. Nonetheless, it is possible to understand how internationally their services are required and the costs that are supported by producing each manufactured good. Moreover, the quantities of the products ordered by the customers and the unit price of each are also information we can have access to.

In a more detailed description, *AUTOMOTIVE* provides five excel sheets which are focused on the records extracted from the *Customers, Fact Sales, Salesmen, Product, Production Plant* of the company.

Regarding the Fact Sales excel sheet, the following variables are represented:

- *Order Type* – Indicates which type of document we are analysing. In this case, it only presents one value which is “Invoiced Shipments”, which we conclude is not a relevant variable to take in further analyses.
- *Cust ID* – Corresponds to the customer ID to identify who has purchased something. As our entity is a B2B, the first four digits are the same as the *Corp Cust No*, represented in the *Customers* sheet. Moreover, the ID includes numbers, but also letters in some cases.
- *Product* – Represents the product purchased by each customer. These products are encoded in order to maintain the integrity of the data. However, we figured out that the first two digits of each product correspond to its product series.
- *Production Plant* – Demonstrates which factory is manufacturing the product in question. This entity has 10 different factories.
- *Order Entry Date* – Corresponds to the date when the customer solicited the purchase. The date format is *dd/month/yy*, which means it is only required the two final digits of the year and the month is displayed as a string abbreviation with only the initial three letters of the month’s denomination - for example, mar stands for March. It is also very relevant to notice the oldest date is the 26th of May 2005 and the latest one is the 31st of December 2010.

- *Scheduled Ship Date* – Indicates the date when it was supposed to send the product purchased to the customer, in other words, the expected shipment delivery date. The format of this variable is different from the previous one because it is displayed as *dd/mm/yyyy*, which means it is requested all the digits of the year, month, and day. Nonetheless, the ranges of dates presented here are from 22nd of October 2006 to 25th of January 2011. This way, it is possible to establish provisory deadlines.
- *Actual Ship Date* – Perceives the date when it happened the product's shipment delivery. The format of this variable is the same as the previous one and the interval of time assessed is between 24th of November 2006 and 31st of December 2010.
- *Invoice Date* –Refers to the date when the invoice was finally made regarding the purchase requested. The format is the same as in the *Order Entry Date* and the interval of dates in question is between the 1st of January 2007 and 31st of December 2010.
- *Payment Date* - As the name points out, this variable represents the date of payment of the purchase by the customer. The format is the same as in the *Order Entry Date*. The payments were effectuated between 24th of November 2006 and 5th of January 2011.
- *Quantity* – Expresses the units of each product that were indeed bought, which values range between 1 and 866 250.
- *Product Series* – Shows which serie each product belongs to, which can be *P1, P2,* or *P3*.
- *Salesman* – Indicates, encoded, the salesman who was in charge of the sale in question.

Secondly, concerning the Product excel sheet, the variables presented are the following ones:

- *Product* and *Production Plant* that were already explained before when describing the *Fact Sales*.
- *Material Cost* – Indicates the monetary amount spent on equipment, tools, or component needed to manufacture such product. The values range between 0$ and 667,5247$.
- *Labour Cost* – Point outs the monetary amount spent in the workforce for producing each unit. The values vary from 0,0012$ to 217,9963$.
- *Variable Costs* – Shows the monetary amount spent on not-fixed costs, such as electricity, water, among others; in other words, represents the costs that vary depending on the company's production volume. So, when the company does not produce, it also does not incur any cost. The values range between 0,0005$ and 51,5344$.
- *Unit Price* – Pinpoints the price of one product individually, in order to know what is the value of each unit, whose values range between 0,008 and 1182,92779. As it is possible to observe, it is not specified the currency of this variable; however, as it is expected to be consistent in the further analysis, it was assumed this variable would have the same currency as the previous ones, so in dollars.
- *Supplier* – Expresses encoded the supplier who provided the materials to conduct the business in question. This entity has 15 Suppliers, each one distributes a different product to *Land of Cars*.

Thirdly, respecting the Customers excel sheet, it is composed of eight features which can be described as followed:

- *Cust ID* and *Salesman* that were already explained previously, when analysing the *Fact Sales*.
- *Cust name* – Corresponds to the name of the customer, which is encoded due to the *General Data Protection Regulation, GDPR*. However, an important aspect to highlight is the fact that the three first digits refer to the *Corp Cust Name*.
- *Corp Cust No* – Reveals the ID, encoded, of each corporation that is our client.
- *Corp Cust Name* – Demonstrates the name, encoded, of each company. It is possible to verify there are twenty different enterprises who are our clients.
- *Salesarea* – Identifies which continent has required the product. Overall, it was only required in three distinct continents: *Europe, America*, and *Asia*.
- *Region* – Pinpoints the region area where the product was solicited.
- *Segment* - This corresponds to the segment the corporation operates, which can be classified as *Power, Lighting, Industrial, Automotive, Continuous emissions monitoring system (CEMS), Consumer,* and *Distributor*.
- *Country* – Refers to the location, country in this case, of the customer's entity.

Fourthly, analysing the Production Plant excel sheet, it is noticeable the following features:

- *Production Plant* that was already explained previously, when analysing the *Fact Sales*.
- *Factory's Country* – Represents the country where the factories are allocated. The only possible locations are Italy and the United States of America, USA.
- *Rent* – Pinpoints the value of each factory rent, whose values range between 2700 and 120000$. As there is no indication of the currency, we assume it is in the same currency as in the other tables, so in dollars.
- *Equipments* – Refers to the amount of equipment each factory possesses.
- *Area* – Shows the area of each factory, in square meters, which is between 3045 and 9461.
- *Number of employees* – Indicates, as the feature's name points out, the number of employees each factory has, whose values range between 91 and 300 people.

Fifthly, the Salesmen excel sheet is the next to be explained, which includes all the following variables:

- *Salesman* that was already explained previously, when analysing the *Fact Sales*.
- *Gender* – Identifies straightforward the salesman's gender, which can be female, *F*, or male, *M*.
- *Age* – Indicates the salesman's age, whose values range between 28 and 55 years.
- *Marital Status* – Pinpoints the marital status, as the variable name points out, which can take four distinct values: *Single, Married, Divorced*, or *Widower*.
- *Hire Date* – Corresponds to the year the salesman was hired, whose values range between 1978 and 2005.
- *PM* – Represents the abbreviation of "People Manager", which are five in total, from *PM1* to *PM5*.

Concerning the DATES file

Concerning the *DATES* file, in *CSV* format, its content focuses basically on a general sheet, where it is possible to extract all the information regarding the dates between 1998 and 2030. This source aims to identify in a quick way all the days, weeks, months, quarters, and years of each date. Obviously, the purpose of this source is to extract more insights regarding our business, evaluating it according to the data available in the *Fact Sales* excel sheet.

The Discovery Process

One of our biggest challenges was to obtain a good dataset, in order to meet the requirements asked for, which was accomplished when solicited it to an automobile entity, which wants to remain anonymous, due to data protection rules, *GDPR*. As it is possible to understand, the data taken into consideration is based on a real-life business corporation, which makes us more engaged and thrilled to develop some insights and to extract value to enhance it.

Furthermore, as we know, usually the raw data has some inconsistencies and needs some data preparation to have clean or manageable data, and ours was not the exception. From the different datatypes to be fixed, until the comprehension of the metadata, as it is encoded, and also the understandable of the primary keys and foreign keys, and the creation also of the surrogate keys, it was indeed a time-consuming process, as we had to be careful regarding the implementation of these concepts. Additionally, the decision of implementing a star-schema or a snow-flake model was also one of our main concerns, as it would influence the approaches to take into consideration in further analysis.

PERSPECTIVES OF ANALYSIS

No. 01 — Sales and Operational Perspective

As our company's activity is based on production, and in order to sustain our operational costs, through the revenue generated by sales, the first perspective analysed is concerning these two main matters: operational and sales, as it is possible to extract interesting insights from our data. We can explore more about each *Production Plant* and compare their productivity, their profitability, the types of products they produced, and for example the delays caused by the difference between the actual shipment dates and the scheduled ones. Furthermore, the possibility of understanding the distinct salesmen and their segments, in terms of years of experience and the sales amount generated, are also crucial aspects to evaluate.

Nonetheless, further analysis can be taken by studying their demographic aspects and interconnect them with the production's segment or even with each customer to understand who is more valuable for the company depending on the region they are located. Moreover, analyzing the sales of our products, comparing the quantities of each one usually sold and the correspondent prices are also one important aspect to explore. From here we will be able to find the profitability of each product from the period of time we have the data, and even check if there is any seasonality, for example.

No. 02 — Marketing Perspective

Then, to improve the success of our sales, a Marketing analysis is going to be performed. Here we can explore our different clients, their buying patterns, for example, as well as which products and product series each client tends to acquire in more quantity and frequency. Another interesting topic is the region from where the products are bought.

These are the two perspectives we are going to analyze in our project. We obviously may find other interesting topics to address along the way, as we get different insights from our data on *Power BI*, but these are the topics inside each perspective that caught more attention from our group.

DESIGN OF THE DIMENSIONAL MODEL

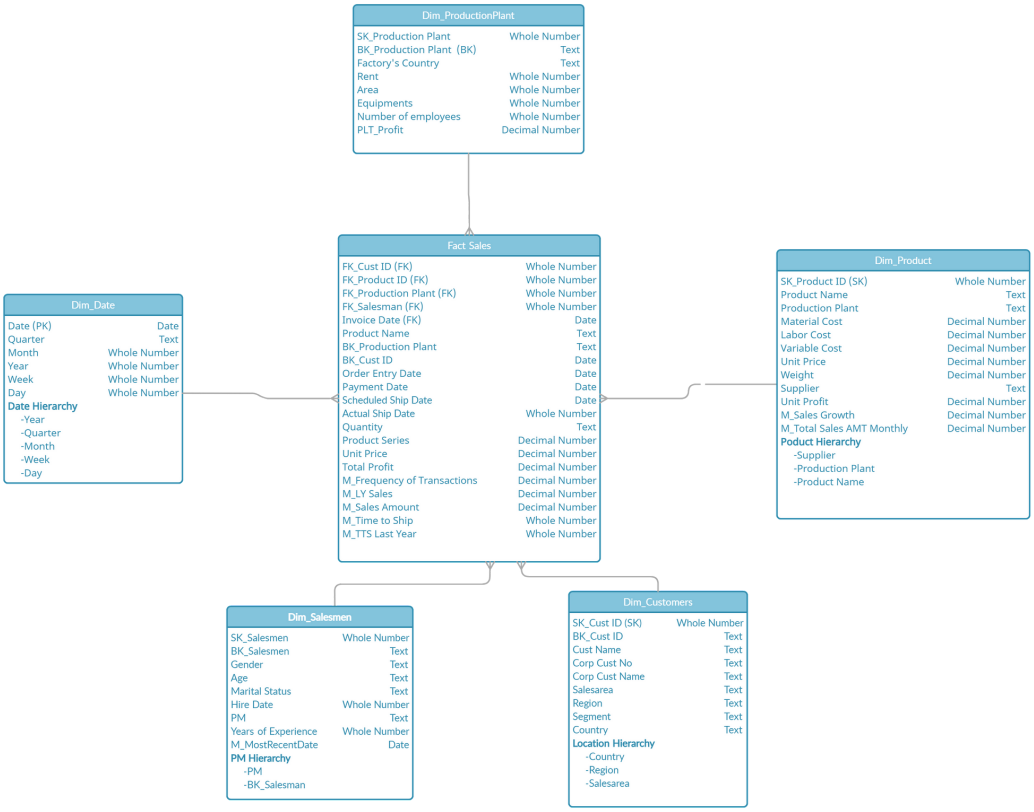


Fig. 1: Dimensional Model

Our Dimensional Model consists of a *Star Schema*. At the center we have our fact table with five dimensions connected as in a typical 5-dimension star schema would be. Moreover, in order to implement the design of the model in question, we have followed the Kimball methodology since it is a business-process-oriented one that meets the characteristics of *Land of Cars* (B2B company). Furthermore, the *Kimball* methodology considers the facts and dimensions using dimensional modeling, which is indeed what happens here, as it follows a star schema model and pursues a data warehouse bus architecture. Finally, the *Kimball* approach starts by identifying the key business processes and the key business questions that the data warehouse needs to answer, which was the first procedure taken at the beginning of this project.

Regarding the hierarchies created, we analyse four different ones:

- Firstly, the *Product Hierarchy*. As we can see in Figure 2, we are relating the *Supplier*, the factories each supplier provides the product, and the *Product* itself.
- Secondly, the *Date Hierarchy*, which represents the labels of the relationship between five variables: *Year, Quarter, Month*, *Week,* and *Day* – Figure 3.
- Thirdly, the *Location Hierarchy* explores the customer *Sales Area*, the *Region* and the *Country*, Figure 4.
- Finally, the *PM Hierarchy*, which is constituted by *PM* and *Salesman*, Figure 5.

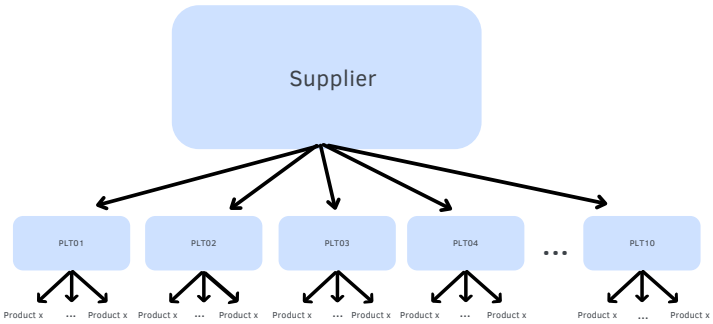


Fig. 2: Product Hierarchy

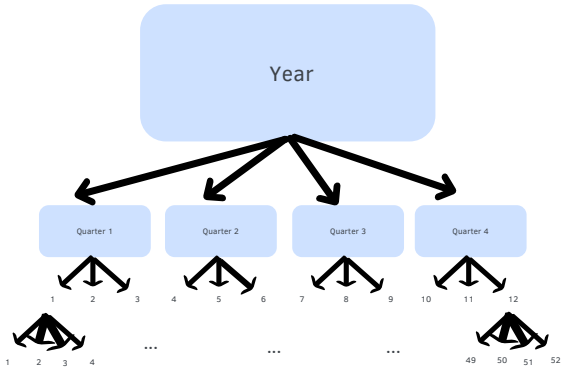


Fig. 3: Date Hierarchy

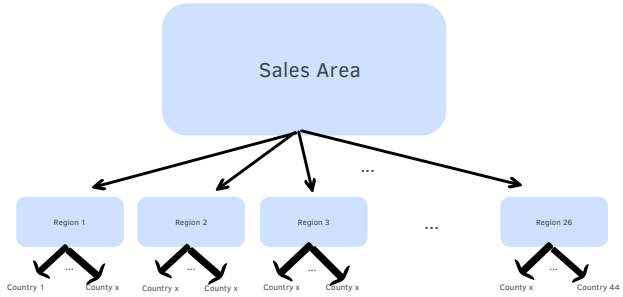


Fig. 4: Location Hierarchy

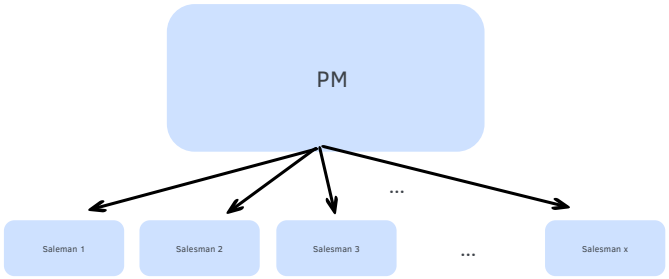


Fig. 5: PM Hierarchy

DATA INTEGRATION, TRANSFORMATION AND MODELLING

Data Integration

As mentioned previously, the data collected is from two different formats, *Excel* and *CSV*. Since the data integration is the process of combining data from many distinct sources, the only steps taken in *Power BI* at this stage were to select the respective excel and the correct option, *AUTOMOTIVE.xlsx*, and the correspondent *CSV*/ text option for the *DATES* file to import to this tool. Therefore, for the *AUTOMOTIVE*.xlsx file, we filtered the tables necessary to perform the desired report and to extract insights from the data, and then the data was loaded. This way, after the proper data loading process, the data sources were integrated into a single Dimensional Model, thus being able to access all the data we need to perform data transformations and the posterior analysis.

Data Transformation

As it is important to uniformize the data and have it managed to deal in the further analysis, the following step was the data transformation to a meaningful and more appropriate one. For this reason, a slight number of actions were performed, since this is crucial for the good performance of the upcoming report and dashboard. Most of the operations made were based on the Power Query Editor, as it was the most appropriate one, except the ones that are explicitly indicated to be performed in another tool.

Regarding the *Fact_Sales* table, we have performed the following transformations:

- Firstly, we have changed the datatypes that were not coherent with the data introduced. This includes the following variables: *Order Type, Cust ID, Product, Production Plant, Product Series* and *Salesman*, where it was indicated to be a mix of numbers and text, which we have modified to be only text datatype. Moreover, the dates were not in the correct format and it was not also identified as a date datatype. For this reason, it was also altered in order to be consistent with the information it provides. This includes the following variables: *Payment Date, Order Entry Date, Invoice Date, Scheduled Ship Date,* and *Actual Ship Date*, where the format chosen was *dd/mm/yyyy*, to be more concrete and better visualize it.
- Secondly and surprisingly, we have noticed some columns that were also imported to *Power BI*, which were completely fulfilled with *Null* values, which does not add any relevant information for our analysis. So, those columns were removed. Additionally, the fact that the *Order Type* column presents only a single value, “Invoiced Shipments”, was also eliminated, as it does not reveal any relevant insight regarding the business itself. So, we have at least removed two columns so far.

- Thirdly, to develop our model and to build one consistent and accordingly to the rules of its implementation phase, now we are focusing on the coherency of the names of the columns. Most of the columns presented are *Business keys* ones, so we have defined the name of them by beginning with “BK_” plus the corresponding column’s name. The features that have suffered this transformation were the following: *Cust ID, Product* and *Production Plant*.
- Finally, we have performed the split between the text and the digits from the *BK_Salesman* and from the *BK_Production Plant* columns, due to the fact we can instantly create the columns that are going to link the respective *Dimension tables*. Moreover, we have also changed the datatypes of each to a whole number and the name of each column, being “SK_” plus the name of the column in question.

Concerning the *Dim_Customers* table, the actions taken were the followings:

- Firstly, we have modified the datatypes that were not consistent with the columns’ data. This includes the following variables: *Cust ID*, *Cust name*, *Corp Cust Name, Salesarea, Region*, *Country,* and *Segment*, where it was expressed the fact that the datatypes were simultaneously in two distinct datatypes, text and number. For this reason, we have changed it to be only of a unique datatype, in a text format.
- Moreover, the *Corp Cust No* was not in the correct datatype neither, so the decision taken was to change the datatype to a whole number, as it was the more appropriate one.
- Lastly, in order to link this Dimension Table with the Fact Sales table, we have performed a Surrogate Key, where we have renamed it, being “SK_” plus “Cust ID”, and correctly defined its datatype, whole number.

Respecting the *Dim_Dates* table, the transformations made were concerning mainly the datatypes and the removal of unnecessary information to our analysis. Having said that, all the columns provided were initially in text datatype, which then was changed to the correct ones: this means that the *Date* column, for instance, was converted to date datatype, in the format *dd/mm/yyyy*, and the following features for whole number datatype, except the *Quarter* and *Period* columns, which remained with the initial format, text one.

Relatively to the *Dim_Product* table, some few and crucial operations were performed, in order to obtain concise data. In detail, the change of datatype was also the main transformation here, as we have converted all the mixed datatype to the adequate one: the *Product, Production Plant* and *Supplier* columns to the text format; the *Weight* column to decimal number and the remaining ones to a fixed decimal number, as we are in presence of monetary values. Then, we have also specified the currency type in the Formatting section in the Data View, to English (United States), so in dollars. Furthermore, in order to create the *Surrogate key*, we have converted also it to the whole number datatype and then we have renamed it as “SK_Product”, as it is the most appropriate nomenclature to distinguish the *Business key* and the *Surrogate key*.

Finally, regarding the *Dim_Salespeople* table, the main transformation was again the change of the datatypes, as the *Gender, Marital Status,* and *PM’s* columns were converted to a text format and the remaining ones to a whole number format. Nonetheless, in order to link this dimension table with the fact one, we have created a *Surrogate key*, where we have split the text characters and the digits from the column *Salesman*, and retained only the digits, converting it to a whole number format – *SK_Salesman* column. Then, we have also renamed the *Salesman* column to *BK_Salesman*, as we are now in presence of a business key one.

Data Modelling

Data Modelling is one crucial aspect we need to take into consideration in data warehouse development. For this reason, it is necessary to ensure that the final model accomplishes all its requirements. With the help of *Power BI Model View*, it provided a viable way to represent data in order to support end-users in understanding and being able to query them.

After all the previous treatment of data and cleaning made, the implementation carried out was to assure the model was indeed perfectly able to be manageable and correctly linked between the *Primary keys*, allocated in the *Dimension* tables, and the *Foreign keys*, built on the *Fact table*. Nonetheless, specifically oriented to our data and as explained above, it was indeed necessary to create *Surrogate keys* in the *Dimension* tables, due to the fact that we were in presence of *Business keys*, in a text format, and it is a good practice to present the keys in a whole number format. Afterwards, it was also performed the creation of *Foreign keys*, as the main objective of this phase is to structure the model to perform the necessary queries to extract the precise and correct insights of the data. In order to develop most of the *Surrogate keys* presented on the *Dimension tables*, basically, it was created an index for each observation, separately in a different column, through the “Index Column” option – the *Surrogate keys* that have a different procedure were already explained above in the *Data Transformation section*, which includes the *Dim_SalesPeople* table and the *Dim_ProductionPlant* table.

Having said that, the connections between each *Dimension* table and the *Fact table* are through the *Surrogate keys* performed in the *Dimension* tables and the *Foreign keys* developed in the *Fact table*. To ensure an adequate procedure, the use of the *DAX* function *LOOKUPVALUE* in the *Power BI Data View* tool was crucial for this task, as we could allocate the correspondent values of the *Surrogate keys* in the right position/row. However, there was an exception when performing this step with the *Dim_Product*. As it is possible to observe in the *Power BI Data View*, there are several Products that have different *Production Plants*, and, for this reason, it is not reasonable to perform a *LOOKUPVALUE* based only on a single feature – the most common thought would be to restrict only for the *Product* column, but with this implementation, the outcome would result in a misleading interpretation of the data. For this reason, it was applied first the *CONCATENATE()* function, creating the *Merge Column* that is hidden (due to the fact it is not useful for the Report generated through *Power BI*), between these two columns, the *Product_Name* and the *Production Plant*, and then we renamed it as our *Business key* column for this *Dimension* table. Regarding the *Fact table*, it was followed the same process for this specific case, where when applying the *LOOKUPVALUE() DAX* function we have indirectly concatenate the information regarding the *Product_Name* and the *BK_Production Plant*, in order to access the *Surrogate key* in the *Dim_Product* table and then implement the *Foreigner Key* in the *Fact Sales* table.

Furthermore, each *Dimension* table has a *One-to-Many* relationship with rows in the central *Fact table*. As we know, the *Dimension* table presents the attributes that characterized the data retained within the *Fact table* and is the core of multidimensional databases. Additionally, these are the ones who present the hierarchies with different levels, which we have already explained previously and that embraces at least three distinct levels, as requested – the *Date hierarchy*, the *Location hierarchy*, the *Product hierarchy,* and the *PM hierarchy*.

After all these implementations, it was reasonable to develop two different perspectives for the *Land of Cars* Report, to perceive the data provided and to enhance its business.

DESCRIPTION OF PBI REPORT FOR EACH PERSPECTIVE

The *Power BI Report View* is the most important tool to, in an efficient and perceptible mode, visualize the insights that can be extracted from the data and understand what are the main causes of the problems or situations that the *Land of Car's* business is dealing with.

As previously explained, we developed two different perspectives: the *Sales* and *Operational*, whose representations are more oriented to the intrinsic business itself, in other words, the processes of the organizations, the salesman data, the profits, and the optimization of the costs, among others; and the *Marketing perspective*, where the main objective is to perceive our customers and their preferences of products or whether the location influences the amount of purchases or the revenue generated, for example.

For this purpose, it was necessary to create *KPIs,* measures, and calculated columns. As we know, the measures represent the fact property that the users pretend to investigate and optimize, existing three distinct types: the *additive*, the *semi-addictive*, and the *non-addictive*. So, we needed to apply the theoretical concepts regarding the measures and the calculated columns to distinguish its differences, as we needed to implement them to compare distinct aspects that, without these resources, would be impossible.

On our initial page, there are two different *bookmarks*, being that these two redirects us to the distinct perspectives, allowing the user to choose which one he/she prefers to assess. Besides that, on each page, there are some icons that allow the user to navigate between report' pages or to return to the home page.

Sales and Operational

In order to do a more in-depth analysis, we created three different tabs to represent the Sales and Operational Report: the *Sales and Operational – Salespeople*, the *Sales and Operational – Sales Amount*, and the *Sales and Operational – Profit*.

With the business needs in mind, when it comes to sales, we started by creating a measure that returns the value of sales amount, according to the variable we wanted to analyze. This measure was calculated by multiplying the quantity of value, defined in *Fact_Sales*, by the unit price of each product, which is defined in *Dim_Product.*

In order to be able to compare different values for the year, we have the *LY Sales*, where it is calculated the sales amount for the previous year, in the same period of time, and *TTS Last Year*, which has the same line of thought as the previous one, but it is focusing on the time between the *Actual ship date* and the *Scheduled ship date*.

When trying to define what could be analyzed and what would be the most useful for *Land of Cars*, we easily realized that it would be important to define groups of *salespeople* and what characteristics have the people who are able to bring greater profits to the company. To help us with this analysis, the second measure created is related to the last ship date (the *actual ship date* or *scheduled ship date*, dependent on which is the most recent) for each salesman. This measure also helped us to create a new column – *Years of Experience*, which presents the years of experience of each salesman – making the difference between this measure and their *Hire date*. These two features made in *Power BI* will help the company to understand if any salesman has not made sales in a long time and if the number of years in the company will contribute or hinder its success as salespeople.

To perform this analysis, we segmented the salesman using an *Advanced Visual* – the *Scatter chart*, where the salespeople were divided into different clusters according to their *Years of Experience*, the number of invoices carried out, and the total sales amount achieved. In order to analyze the different clusters, *Parallel coordinates* were developed, where the values are representing the average of *Age, Years of Experience,* and *Sales Amount* of each cluster.

In the same tab, we used a *Tornado* to easily understand how the total sales amount change according to the *Marital Status* and the *Gender* of each Salesman. In spite of that, we used a *Data bar KPI* so that, if a specific characteristic is selected (be it *Marital Status, Gender,* or each *Cluster*), it will be possible to analyze the evolution of *Sales Amount* over year, targeting the value of sales amount from the previous year. Also in this tab, a slicer was placed with the different *People Manager*, to be easy to analyze the variables explained above by *PM*, and study which *PM* gets more sales and what characteristics of their salespeople allow that to happen.

More focused on the production and different market segments, we used a *Sankey* to show how much *Sales Amount* comes from each segment, as well as in which country these products are manufactured. With the same intention, a *treemap* was made to analyze the value of *Sales Amount* by *Region* – adding a *tooltip*, using a *Donut Chart*, to understand which *Salespeople* are responsible for these values in each *Region*. Still, on sales, a *Line Chart* was made with the new *Sales Amount* measure and the year of each Invoice, to analyze the variation of this value over the years and to predict how it will behave in the future years.

In more operational terms, it made sense to see which products have the most value for the company, so the *Unit Profit* was created, using the *Unit Price* and the 3 costs associated with each product (*Labour Cost, Material Cost,* and *Variable Cost*). After that, in the *Fact_Sales* we created a column with the *Total Profit*, using not only the *Unit Profit* previously defined in *Dim_product* but also the quantity of each invoice.

We also noticed that some deliveries took much longer than others, considering the entry date and the *actual ship date*. To have a more accurate analysis, we created the measure *Time to Ship*, to understand which products took the longest to produce and aim to decrease this production time. Besides that, we added a *Q&A* to the *Land of Cars*' Manager to be able to question whatever he/she wants to analyse and that is not present in the current report.

With this in mind, the last tab, *Sales and Operational – profit*, displays the profit by year - using a *KPI chart* - the *Unit Price*, average *Unit Profit,* and *Profit Margin %* of each product (in a *Matrix*), having the possibility to analyse the average of these values per *Production Plant* and per *Supplier*, since the *Product Hierarchy* is used on this table. To finish, a *KPI* is presented to show how the average of the *Time to Ship* is changing and whether it is achieving the goal or not - notice that for some selected products the *KPI* is empty, and that is because some products have not been sold in the last years of analysis, so it is not possible to analyse this *KPI* in those years. All of these visualizations can be updated according to the *Production Plant* we want to study.

Concerning the technical part of the report, each visualization was chosen over others, with proper reasons, being them the following:

- The *Scatter Chart* was chosen since allows us to analyse 3 different variables in the same category and use the Clustering option according to that variables.
- The *Tornado* by the characteristic of we can analyse 2 categorical variables in relation to a numerical one, in which one of the categorical only has 2 categories.
- The *Data Bar KPI* was chosen since not only allows us to compare different categories (in this case, different years) but also to define a goal and analyse how distant of the goal the current data are.
- Since we only wanted to inform the *Land of Cars* about the centroid of each cluster, the *Parallel Coordinates* seemed the most suitable and easy to read, because we easily perceive that variables have lower or higher average values and how far away are the clusters from each other.
- The *Sankey* diagram is a graphic illustration of flows like energy, material, or money flows. With this in mind, we have chosen this visualization to see the sales amount flow from the factory's country to each market segment.
- To see the evolution of *Sales Amount* by year, the *Line chart* seemed to be the best option since it not only is the best approach for time analysis but also enables to do a future valuation, using the forecast.
- We chose the *Treemap* for the *Sales Amount* by *Region* since it visualizes the attribute by size and color coding and shows ratios of each part to the whole.
- The *KPI* shows the evolution of the *Time to Ship* over a year and allows us to understand if this value has been decreasing, increasing, and if it is reaching the goal of not being higher than in previous years, so it seemed to us the most suitable visualization for this value.
- The *Matrix* was chosen over others since we have a big amount of products and it is the only one that allows us to see them all.

Marketing

Regarding the second perspective of analysis, the Marketing Perspective, contains two report pages, one more focused on the general insights and the other more oriented on the customer.

For the first page of the report, we started by implementing a *Slicer* that will grant us the possibility to do a filtration over the visualizations of this page - *KPI*, *Table*, and a *Multiple-Axis Chart*, namely – by month, where each month encompasses every year. Subsequently, for the creation of the *KPI*, two new measures needed to be built, both in the *DIM_product* table, the first one was the *Total Sales AMT Month Before*, being this measure designed only for the purpose of calculating the main measure, the S*ales Growth*. The measure *Total Sales AMT Month Before* calculates the total sales amount based on the previous month, being that one-time intelligence measure was used in this formula, the dateadd, that is, this function performs the calculation for every month and that's why we use the intelligent function, which we can understand whether it is a *Year*, a *Month*, or a *Day* that we want, and as we always have a month, we have to subtract one because we only want to keep the previous one. Concerning the second measure, the *Sales Growth*, it varies monthly, being this measure calculated by taking the difference between two measures, the *Sales Amount* and *Total Sales AMT Month Before*. In order to have it in relation to the previous month, we then divide this difference by the *Total Sales AMT* 1 *Month Before*. Therefore, this *KPI* will inform us whether or not there was an improvement or a decrease in the company's sales over a particular period of time (for all months or only for the selected ones), given as a percentage. Then, we have a table, which shows for each product, the quantity purchased, and the number of times the product was acquired. In this table, we can filter by ascending or descending order for the column *Quantity* and *Frequency of Transactions*, and also, we can search for a specific product by writing its *ID*. The last visual of the marketing report is the *Multi-Axis Chart*, for this one we specified two axes, one for the *Quantity* and the other for the *Frequency of transactions*, this way we could compare these two variables taking into account the *Product Series*. From its analysis, we can realize which product series have the higher number of quantities sold and also the one that has the higher frequency of transactions.

Taking now into consideration the second page of the marketing report, with the objective of making the report more interactive and appealing, we have also started by introducing a *Slicer*, where the user can make a filtration by the *Data Hierarchy*, that is, we can choose if we want to filter by *Year, Quarter, Month*, or *Day*. Afterwards, the best visualizations were chosen, that is, the visuals that better explain and represent our data for this perspective of analysis, being this page composed of the following set of visuals: a *Map*, a *Stacked Bar Chart*, and a *WordCloud*. For the construction of the *Map*, a new measure was created in the *Fact Sales* that will provide us very useful and meaningful insights. This measure called *Frequency of Transactions* counts the number of times each product was bought. Hence, the map represents the number of transactions carried out depending on where the customer lives, considering his *Country, Region*, and *Salesarea (Location Hierarchy*). In addition, to improve the quality of this graph we decided to use a *tooltip* over it, which shows through a stacked bar chart the most frequent customers of the company and the *Number of Transactions* displayed by a *Card* visualization. Considering the *Stacked bar chart*, it simply exhibits which segments give the most revenue in each *Segment*. For the development of this graph, we used a measure that was created and explained in the *Perspective Sales and Operational*. The last visual of this first page, the *WordCloud*, demonstrates which corporations buy more in terms of quantity of products, since our company is a *B2B* it makes perfect sense to have this information shown. With this first page, we can easily understand which regions our company should pay more attention and effort to, and also in which segments and corporations should we invest more.

Finally, concerning the technical part of the report, each visualization was chosen over others, with proper reasons, being them the following:

- The *Map* shows the *Frequency of Transactions* for the geographic location of the client, which serves our objectives;
- The *WordCloud* displays the companies that buy in more and less quantity our products, being them differentiated by the size of the font (bigger size buy more often, and vice-versa);
- The *Stacked bar chart* was chosen instead of lollipop because *lollipop* was a paid version;
- The *KPI* Indicator shows the percentage of the *Sales Growth* and the value for each month (indicated by a *Line chart*), being the *KPI* that better fits our purposes to show what we want,
- The *Multi-Axis chart* was the better visualization since it has the possibility to incorporate two different axes, thus, giving us the opportunity to compare two distinct variables;
- The *Table* was chosen over others since we have a big amount of products and it is the only one that allows us to see them all, as was doing in the last perspective.

Summary Tables

Measure	*Formula*
Sales Amount	Sales Amount = SUMX(ADDCOLUMNS('Fact Sales', "new_column", 'Fact Sales'[Quantity]* RELATED(Dim_Product[Unit Price])), [new_column])
TTS Last Year	TTS Last Year = CALCULATE([Time to Ship], SAMEPERIODLASTYEAR(Dim_DATES[Date]))
Time to Ship	Time to Ship = DATEDIFF(AVERAGE('Fact Sales'[Order Entry Date]),AVERAGE('Fact Sales'[Actual Ship Date]),DAY)
LY Sales	LY Sales = CALCULATE('Fact Sales'[Sales Amount], SAMEPERIODLASTYEAR(Dim_DATES[Date]))
Frequency of Transactions	Frequency of Transactions = CALCULATE(COUNTROWS('Fact Sales'))
Most Recent Date	MostRecentDate = YEAR(Max(MAX('Fact Sales'[Scheduled Ship Date]), Max('Fact Sales'[Actual Ship Date])))
Sales Growth	Sales Growth = IF(HASONEVALUE(Dim_DATES[Month]), DIVIDE('Fact Sales'[Sales Amount]- Dim_Product[Total Sales AMT Month Before],Dim_Product[Total Sales AMT Month Before]),BLANK())
Total Sales Amount Month Before	Total Sales AMT Month Before = IF('Fact Sales'[Sales Amount], CALCULATE('Fact Sales'[Sales Amount], DATEADD(Dim_DATES[Date],-1,MONTH)), BLANK())

Table 1 - Measures and Formulas

Calculated Column	Formula
Profit Margin %	Profit Margin % = (Dim_Product[Unit Price] - (Dim_Product[Labour Cost] + Dim_Product[Material Cost] + Dim_Product[Variable Cost])) / Dim_Product[Unit Price]
Merge	Merge = CONCATENATE([Product Name], [Production Plant])
Unit Profit	Unit Profit = Dim_Product[Unit Price] - (Dim_Product[Labour Cost] + Dim_Product[Material Cost] + Dim_Product[Variable Cost])
PLT_Profit	PLT_Profit = SUMX(VALUES('Fact Sales'[FK_Production Plant]),CALCULATE(SUM('Fact Sales'[Total Profit])))
Years of Experience	Years of Experience = [MostRecentDate] - [Hire Date]
FK_Cust ID	FK_Cust ID = LOOKUPVALUE(Dim_Customers[SK_Cust ID], Dim_Customers[BK_Cust ID], [BK_Cust ID])
FK_Product ID	FK_Product ID = LOOKUPVALUE(Dim_Product[SK_Product ID], Dim_Product[Merge], CONCATENATE([Product Name], [BK_Production Plant]))
Total Profit	Total Profit = RELATED(Dim_Product[Unit Profit]) * 'Fact Sales'[Quantity]
Unit Price	Unit Price = RELATED('Dim_Product'[Unit Price])

Table 2 - New columns and Formulas

ANALYSIS AND DISCUSSION

As previously explained, our approach since *Data Integration* until calculated measures had the goal of analyzing our data through two perspectives: *The Sales and Operational Perspective* and the *Marketing Perspective*. With the first one the main objective was to understand the *Land of Cars*' Salesmen and the company's sales amount through indicators and different variables as region, segment, and time as well as profitability and shipment times. Whereas in the second perspective, our focus was to study the company's customers, regions with more frequent sales, the different segments, and product preferences.

From this analysis, several conclusions about the data were reached as listed below.

Sales and Operational perspective

The clustering of *Salesmen* based on their years of experience, number, and sales amount of the transactions they were involved in, allowed us to distinguish three groups: *The Champions*, The *Experienced,* and *The Ordinary* ones. The *Experienced* ones are characterized for being older and having more years of experience, although with a slightly lower sales amount compared to the other two groups. The *Champions* group has the best sales amount value, comprehend middle-aged salesmen with approximately ten years of experience. Whereas the *Ordinary Salesmen* are the younger and with less experience, having a sales amount value in the middle of the two previous groups.

Additionally, it was observed that the *Married* salesmen are the ones who bring a higher sales amount ($10.9M, being $2.9M made by females and the remaining amount by males) whereas the widowed are the ones who bring less ($2M), although it's important to notice that this pattern also happens due to the fact that there are more *Married* people in the company in regards to the other *Marital Status*. Also, one group that positively distinguishes itself is the one composed of *Divorced* females (with sales of $4.9M).

Concerning the *Sales Amount*, the year 2007 was the one with better results ($6.9M). However, 2009 was the only one that outperformed the previous year (growth of 11.57%). Both in 2008 and 2010, the company underperformed its goal (regressed 11.98% and 4.84%, respectively). However, when analyzing the sales amount of each product manager (*PM*) individually, some differences were found, mainly: with *PM3*, 2008 was the only year without growth, contrarily to *PM2*, which only had growth in this year; with *PM5* there was a regression in the sales amount of 2009 (-6.68%) which contrasts with the growth it had in 2008 (14.93%) and in 2010 (3.21%).

Still, regarding the *Sales Amount*, it was possible to perceive that *Central Europe* was the region of origin of the customers where this amount was higher, with a total of $8.83M (being 22.24% of it made by *Salesman004* followed by *Salesman060* with 17.48%), followed by *China* with $3.72M and *Alpine/Eastern Europe* with $2.12M. Plus, the customers with ids 572100, 536700, and 537715 were the ones that most contributed to the values verified in these three regions, respectively (with the corresponding amounts of $1.2M, $0.57M, and $0.8M).

Additionally, using the sales amount observed between the year 2007 and 2010, it was possible to forecast with 95% of confidence the value for the year 2011. Although we don't have data for this year that enables us to evaluate the quality of this prediction, we estimate that the sales amount would be close to $6.65M, with the possibility of ranging between $5.8M and $7.5M.

Regarding the *Factory's country* and the destination of the product in terms of customer's segment, it was noticed that the transactions from the *USA* to the *Industrial* segment are the ones that most impact the sales with a total value of $5.1M. Overall, USA's factories are the ones that most contribute to the sales. When it comes to Italy's factories, the Automotive segment is the one with the most relevance, with an amount of $3.2M.

In terms of profit, 2007 was the year with higher *Total Profit* ($2.3M) whereas 2008 is the one with worst results ($1.98M). Looking at each product individually, the product P10162 is the one with higher unitary averaged (for the different *Production Plants*) profit ($5.5/unit). Nevertheless, it's important to take into account the performance of this product in terms of total profit. On that note, according to our data, this product was only sold in the year 2010 and accounted for an insignificant profit value ($868.5). So, the unitary profit margin is not always the best indicator of product performance.

Afterwards, the ten production plants were analyzed, which is important when studying the operations of the company. *PLT01* is the one with a higher average percentage of *Profit Margin* (61.96%). Also, for it, 2008 was the least profitable year ($19.1K) and 2007 the best one ($28.3K). For PLT02, 2009 was the worst in terms of profit ($55.5K) contrarily to 2008 ($82.7K). For PLT3, 2009 ($74.8K) contrasts with 2007 ($164K). The results, for the different production plants, can be seen in more detail in our *PBI* report. However, overall *PL04* seemed to have been the one with better results (with a profit of $0.34M in its worst year, 2008, and $0.53M in its best, 2007). Nonetheless, it's important to highlight that there is a high variability of the profit through time.

With respect to the shipment times and the goal of shipping at least as fast as in the previous year, 2009 was the year with better results, that is, with faster shipments, whereas 2010 was the one with worst results (with an average shipment time of 70 days, 2.94% more than in 2009). Looking at the production plants in particular, in 2010, the only one that outperformed the year of 2009 was *PLT10* (6.56% faster), while *PLT1* and *PLT6* remained the same, and the rest of the production plants underperformed, taking more time to ship the products.

Marketing perspective

Examining the product preferences, *P2* was the product series with both higher quantities and frequency of transactions (61.6M and 6.7K, respectively). Followed by *P3* if considering the total quantity and *P1* if the frequency. Apart from small variations, the results are similar through the different months. Additionally, *P34785* was the product with the most units sold in total (4.9M) and through several months (February, May, October, and November), and *P24520* the most frequently sold one in total (and in the months of March, May, June, and September). Other relevant products when studying each month individually are *P34768* (best in the months of March, April, and August, when looking at the quantities) and *P22516* (best in January, October, and November, when looking at the frequency).

Analyzing the sales growth by month, there is some variability, being possible to verify a great growth from June to July (+24.66%) and, in contrast, a decrease in growth between May and June (-20.45%).

Central Europe was the region with a higher frequency of transactions (2608), being 183 of it made by the customer with *ID 543600*, followed by the *479701* (with 176). Some regions that also have great relevance are: *China, Southeast Asia,* and *Southwest Europe*, with frequencies of 1907, 1456, and 995, respectively. Analyzing this aspect in the *Date Hierarchy*, these are also the regions that have higher frequencies of transactions. Additionally, studying the countries, in particular, some that stand out are: *Germany, Italy, the USA,* and *Hong Kong* with frequencies of 2543, 1487, 1433, and 1304, respectively. In terms of the *Sales area*, the one to highlight is *Euro* with 8202 transactions and the customer with *ID 573800* contributing to 243 of it.

Also, it was observed that *C04* was the client's corporation that bought higher quantities with 27.1M, followed by *C01, C02,* and *C08*, with 16.3M, 14.2M, and 13.6M items, respectively. Besides some minor changes, these products remain the top sold ones (in regards to quantity) over the different years, quarters and months.

Finally, studying the sales amounts of the different segments of *Land of Cars*' customers, *Distributor* proved to be the one with more relevance ($7.2M), followed by *Industrial* ($6.9M) and *Automotive* ($5.6M). Although, neither the ranking nor the ranges for the sales amount vary much from year to year, when analyzing each quarter individually it's observable that: in some quarters the *Automotive* segment is the one with better results (3rd quarter of 2007 with $0.48M and 1st quarter of 2008 with $0.38M), however, in most quarters either *Distributor* (e.g., $0.48M for the 1st quarter of 2007, $0.51M for the 3rd quarter of 2008, among others) or *Industrial* ($0.54M for the 2nd quarter of 2010, $0.61M for the 3rd quarter of 2010, among others) are the segments that are the most valuable for the company. Examining further the *Sales Amount* behavior by segment in the time hierarchy, the previously mentioned segments remain the best ones (monthly), yet when looking at each day individually there is higher variability of performance in the different segments, being this ranking not so clear, as it was expected.

CRITICAL ASSESSMENT

Taking a look into the first part of our report, we can compare how the final project came up versus how we initially thought it could be. It is obvious that the expectations are not 100% accurate when we have a project on sight, especially when that project is made on software we did not have experience with, but as students, we are always looking forward to learning new skills and abilities and in our case, we feel we have learned an extensive amount about *Power BI* during this project.

When starting the progress in our *Power BI* journey, we immediately faced some adversities we did not expect in the beginning. Our data did not have specific names for the products, for the customers, for the salesmen, and so on, and that created confusion regarding the meaning of business and surrogate keys. We decided to call those variables business keys as they also had text (for example *PLT01*, which referred to Production Plant 1) and created a surrogate key to face this issue. Apart from this, this also made it more difficult to interpret the data, especially regarding the products. We knew some of the characteristics of the products, such as the weight, unit price, and costs, but we did not know what they were, and we feel a lack of information could have brought more value to our analysis.

While working on the project we found adversity in the *Product* Dimension. This dimension has a *Product Name* that we found out was not unique. There were some rare cases of products that were produced in more than one production plant, so in those cases, there was more than one line. We obviously created a surrogate key that is an index for every different line, treating differently each product produced in each production plant.

Another difficulty we faced was with the column *Production Plant*. We had this column in the fact table, in the *Production Plant* dimension, and also in the *Product* dimension. This limited our creativity in terms of analysis, as we wanted to combine the unitary data we had for each product, multiplying it with the quantities sold in each month and then subtracting the monthly rent, so we could have got a better monthly insight of the profits each production plant had in each month.

But anyone would be very naïve to think there would not be any adversity. Overall, we believe we met most of the expectations we had before working on *Power BI* with this dataset. As aspiring Business Analysts, we had the capability to preview what were the aspects of our business we wanted to analyse and the insights we wanted to take, and we are very pleased with the final product.

If we take a look at the perspectives of analysis we wanted to analyse and consequently to our final report, we can conclude our outcome matched our desired objectives and with distinction. Throughout the project and with a closer look and relationship with the data we felt the need to explore further from what we had previously expected.

One of the most important aspects of a project is the achievement of a purpose. In this case, we feel we achieved the purpose desired. In the *Sales and Operational Perspective*, we managed to get great insights from the different products, specifically their profit margins, the time they normally take to ship since the moment the order is registered, and the regions they are mostly sold to, for example. Another important take is the segments of the products. As explained before, not having the name of the products limited our project, but here we were able to at least get an idea of the segment. We analysed the different segments and their distribution between our two countries where we produce: Italy and the USA.

Also, we managed to get more deeply into our *People Managers* and their correspondent Salesmen, making a demographic analysis that helped us better understand the disparity of results and outcomes of each other. This analysis can be very useful for the company's Human Resources to understand who their best salespeople are, how they should motivate them, whom they should dedicate more effort to motivate, etc.

The most crucial aspect in this perspective is obviously the sales. We managed to analyse the sales from the period of 2007 to 2010, and still making a forecast for the year of 2011. As we now know, those were difficult years of recovery from a huge financial crisis, so we felt a forecast added a huge value to our analysis. As well as the sales we compared the profits for the different periods of time. And here we did it also per *Production Plant* and per Product itself. Again, if we could have analysed the monthly profits with the rent of each production plant, we could have made more suggestions, but this does not make our analysis negative by any means. We believe a deep analysis of the most profitable products was very important, especially when presenting to a big company in this industry, where constant innovation needs to be in place.

Regarding our *Marketing Perspective*, we took a deep analysis of the different product series we have, especially regarding both the quantity sold and the frequency of transactions. The series of the products are very crucial to analyse, because in this sector the competition is very hard, and the manufacturers cannot risk being any step behind any competitor. It is important to know when to innovate a product and create a new series, as well as if there is not the need to rush on a new solution if the series is still competitive. This can bring more time for research and development to happen to have an even bigger advantage in the next product series. The insight we took was as expected, with the series 2 being the most bought both in quantity and infrequency of transactions.

Finally, to finish this perspective, we continued our analysis of the product segments, but this time comparing the sales amount each segment managed to get versus the quantities sold we had compared in the previous perspective. Also, we were interested to know what our best clients are. We know this is sensitive information and we understand the reasons behind the fact of not having the name of each, but this would bring more interest to the analysis. And apart from what our best clients are we also presented where they order us from with a wide world map view also showing the best customers from each region of the planet.

CONCLUSION

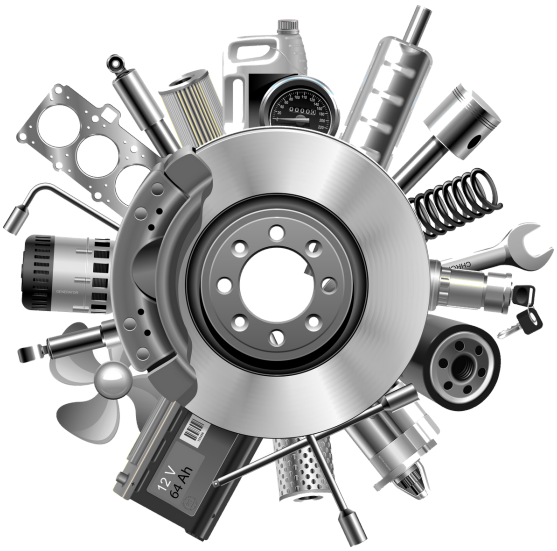
First of all, this project was contributed a lot to improve our problem-solving abilities since we learned and understand the functionalities and advantages of working with *Power BI* desktop/app to show different aspects of the business.

Important to say that one of our biggest obstacles and time spending was in the first phase of the project, where we had to search for a dataset that meets all the criteria imposed in the project description, that is, the searching for one dataset that meets all the needs for the development of our project.

With the realization of this project, we were able to present an interactive dashboard that exhibits the most important aspects of our *Power BI* report, which englobes two perspectives, the *Sales and Operational* and *Marketing*, respectively. The displayed perspectives allow the user to explore and have clear answers to the most critical and essential questions regarding our business, in other words, it helps to understand the pertinent information about the chosen departments.

To sum up, we can state that we have fulfilled all the objectives and requirements mentioned for carrying out the project and that despite the challenges encountered during the way, we managed to achieve a good final model (Data Warehouse design) that follows the *Kimball* methodology, which permitted us to have a wider vision of enterprise integration and consistency, capable of showing everything we had purposed at the beginning of the project, the business needs more precisely, with the support of the best interactive visuals. As we follow the extract, load, and transform (ELT) stage, and ending up with an application that shown the appropriate visualizations for the information we wanted to address, the company will have access to the necessary knowledge that will facilitate the process of decision-making for the further steps.

The *app* with the report and dashboard can be found here.



Land of Cars

Catarina Candeias, m20200656

Catarina Urbano, m20200607

João Henriques, m20200670

Margarida Pereira, m20201038

Rita Ferreira, m20200661

REFERENCES

Links

- *Enterprise Data Warehouse Bus Architecture - Kimball Group*
- *https://www.kimballgroup.com/2005/03/slowly-changing-dimensions-are-not-always-as-easy-as-1-2-3/*
- *The 16 Marketing KPIs You Should Be Measuring (vtldesign.com)*
- *Power BI documentation - Power BI | Microsoft Docs*