UNIVERSIDADE CATÓLICA PORTUGUESA

Monitoring Feasibility Studies:

Improvement of BI tools

Joana Filipa Gomes Areosa

Católica Porto Business School

2022

UNIVERSIDADE CATÓLICA PORTUGUESA

Monitoring Feasibility Studies:

Improvement of BI tools

Final Work in Organisational Context presented to Católica Portuguesa in order to obtain the master's Degree in Management with a Specialization in Business Analytics

by

Joana Filipa Gomes Areosa

under the guidance of

António Manuel Valente de Andrade

Católica Porto Business School, Universidade Católica Portuguesa

April, 2022

Acknowledgements

"The better you are at surrounding yourself with people of potential, the greater your chance of success" – John C. Maxwell

I would like to express my gratitude and appreciation to the following people, without whom I would not have made it through this research and, consequently my master's degree!

I am extremely grateful to my supervisor, professor António Andrade, for providing guidance and constructive criticism throughout this project.

I would also like to express my deepest recognition to Polisport Plásticos, SA, particularly the planning and management control department, for the warm welcome and willingness to impart their knowledge.

I am deeply indebted to my family, for providing me with the opportunity to study at a highly regarded university, for the profound belief in my abilities and for putting up with all my stress for the past few months.

I would like to extend my sincere thanks to my close friends, for all the meetings to work on our dissertations and for providing a sounding board when needed.

Abstract

At the beginning of the 21st century, the term Business Intelligence became a buzzword in the management world. Its importance has grown with the so-called Big Data, which arises from a digital world, implying, at the very least, the need to find patterns in the data and infer predictions. In a massive competitive period like today, information is one of the main sources of competitive advantage, turning the directors' focus to this discipline.

BI is a technology-driven process that analyses data in order to transform it into useful insight (information) for decision-making. Although the literature about this subject is still reduced, its growing relevance in business is raising learning opportunities, which inherently stimulates interest in this matter for further research.

This dissertation shows the restructuring of a report in Power BI of Polisport Plásticos, which objective was to improve the report in terms of data processing and data visualization. The methodology adopted was Action-Research structured in three designated cycles: Understand Data System Structure, Prepare Data and Report Design.

The main findings of this research conclude that using other tools associated with Power BI, understanding that data preparation and report design are phases that depend on each other, and that a good DW structure are fundamental for the report's efficiency.

Keywords: Business Intelligence, Power BI, Dashboard, Data Visualization

Words: 9911

Resumo

No início do século XXI, o termo *Business Intelligence* tornou-se um termo popular no mundo da gestão. A sua importância cresceu com o designado *Big Data* que resulta de um mundo mais digital implicando, no mínimo, a necessidade de encontrar padrões nos dados e de inferir previsões. Num período tão competitivo como o atual, a informação é uma das principais fontes de vantagem competitiva, ganhando a atenção dos diretores.

BI é um processo alimentado pela tecnologia que analisa dados e os transforma em informação útil para a tomada de decisão. Embora a literatura sobre esse assunto ainda seja reduzida, a sua crescente relevância nos negócios está a aumentar as oportunidades de ensino, o que, por natureza, estimula o interesse por esse assunto para novas pesquisas.

Esta dissertação mostra a reestruturação de um relatório em Power BI da Polisport Plásticos, cujo objetivo era melhorará-lo em termos de processamento e visualização de dados. A metodologia adotada foi a Investigação-Acção estruturada em três ciclos: Compreensão da Estrutura do Sistema de Dados, Preparação do Dados e Elaboração do Relatório.

As principais descobertas desta pesquisa concluem que a utilização de outras ferramentas associadas ao Power BI, o entendimento que a preparação dos dados e o desenho do relatório são fases que dependem uma da outra, e que uma boa estrutura do *Data Warehouse* são fundamentais para a eficiência do relatório.

Palavras-Chave: *Business Intelligence*, Power BI, *Dashboard*, Visualização de Dados

List of Abbreviations

BA – Business Analytics

BI – Business Intelligence

DW – Data Warehouse

ERP – Enterprise Resource Planning

IS – Information System

IT – Information Technology

Table of Content

List of Figures

List of Graphics

List of Tables

First Chapter

1. Introduction

This dissertation was developed under an internship at Polisport Plásticos, which proposed an improvement of a business report, taking into consideration upgrades in Power BI that improve the performance of the report, both in terms of data processing and visualization. The report's purpose is to keep track of the status of each investment already made. That is, the company wants to understand how the situation of a certain project is compared to the target made in the feasibility study, allowing the board of directors to decide about the continuity of a certain product and help commercials negotiate prices.

Polisport Plásticos, SA is a company specialized in Plastic Injection Molding, that focuses mainly on plastic products and accessories for two-wheeled vehicles. This company is part of Polisport Group, which was born in 1978 from the CEO's passion for the world of two wheels. The group already has a Management Planning and Control department, where the BA unit is included as a shared service for the five integrated group companies. Besides being a Data Analysts job, the project was coupled with the IT department, whose help was crucial for the project's success.

1.1. Report Restructuring Goals

Considering the fact that this dissertation relies on a report restructuring, it does not answer a proper research question. Instead, its main purpose is to reform one of the company's reports which resorts to the Power BI platform. Hereupon, the central objective is to improve the report efficiency and this is clarified in the following specific objectives:

- Improve the connection to data sources through Power BI
- Rule out tables' unnecessary information
- Formalize steps on Power Query
- Implement a star schema
- Optimize data visualization

There are large incentives for the development of this research, due to its relevance for business. Firstly, the matter of the report is important, since it allows the company to monitor their project investments and compare them to the feasibility study already made. By analysing the report, some conclusions about whether variables, for instance, costs or quantities, in the feasibility studies are too conservative or too optimistic can be taken, which is very useful for commercials. Besides, it also shows relevant financial information when deciding on some product continuance, such as profitability ratios and margins.

Secondly, since the work done consists of a review of an existing report in terms of upgrades in the Power BI platform, some of the changes made that increase efficiency, both in terms of data processing and data visualization, can be applied to the remaining reports.

1.2. Research Methodology

The practical component of this dissertation, the report restructuring, follows the Action-Research methodology. This methodology has a more practical strand since when analysing the results of some changes, new ideas for improvement can emerge.

The action-Research methodology consists of cycles in which within each one there is a phase of planning, action, observation and reflection that later gives rise to a new plan. Thus, for the restructuring of the report, the cycles are called Understand

Data System Structure, Prepare Data and Design Report, as they represent the main areas of analysis for improvement. However, in this situation, several turns were taken in each cycle for constant improvement in those areas and advancing to the next cycle does not invalidate the need to go back and formulate a new plan for the previous cycle, since each cycle depends on the previous ones.

1.3. Dissertation Structure

This dissertation is organized into five chapters. The ongoing chapter, Chapter 1, pinpoints the background in which this study was elaborated and how the restructuring itself emerged, an overview of the objectives to achieve, the methodology adopted and the structure of this research.

The second chapter provides a theoretical framework about what are the main knowledge hitherto known while defining the main concepts adjacent to Business Intelligence, Information Systems and Data Visualization required to understand the developed project.

The third chapter gives a glimpse of the context in which the internship was taken, the reasons why the report needs improvements and goes deep in explaining the goals. It also presents a theoretical description of the adopted methodology.

The fourth chapter shows all the initial analyses done to the base report along with its main problems. Consequently, presents the suggested changes and finalizes by showing the same analyses but this time done to the final report (after changes being applied) revealing indeed real evidence of improvement.

The last chapter, Chapter 5, discriminates the goals achieved, summarizes all the conclusions drawn from all the work carried out, indicates the major contributions of this work, both to analysts and end-users, and recommends analyses that can be done to conclude about the efficiency of other report's aspects.

Second Chapter

2. Literature Review: Business Intelligence World

"Data is what you need to do analytics. Information is what you need to do business" – John Owen

The current chapter provides a theoretical framework for this study. It approaches studies already taken about this discipline, central concepts and notions for understanding it.

In terms of architecture, this chapter is divided into three main sections: Business Intelligence, Information Systems and Data Visualization The main reasoning behind this structure is to firstly display BI history and show how it plays such an important role in today's business, then demonstrate how IS undergird the efficiency of BI, allowing easy access to information through tools for building analytical reports.

2.1. Business Intelligence

At a time when companies are facing rivalry at its highest level between industries, BI is the best source of differentiation (Davenport, 2006).

Besides its relatively recent interest among organizations, the concept reached its maximum worldwide popularity in 2005, according to Google Trends. In the last years, the term has been losing attractiveness (Graphic 1). However, its implementation has just started. In fact, Deloitte concluded, through its own survey, that only a minority of organizations are insight-driven and that the biggest drag factor is the culture. It states that getting access and knowledge to deal with analytical tools is not a difficult task, but instead changing behaviours and mindsets are. (Deloitte, 2019).

Graphic 1 - “Business Intelligence" word trend

Source: Google Trend (27.12.2021)

Over time, there were three milestones in the construction of the BI concept. The first is marked by Richard Miller Devens, who pioneered the use of the term discussed in 1865 in Cyclopaedia of Commercial and Business Anecdotes. The term described the knowledge that a banker, named Henry Furnese, had in some areas. Although he did not use that information for good, the seed was planted. However, in that era, the technological conditions were not suitable for the progress of this discipline. Thenceforward, the term was reintroduced by who is recognized as the “father of Business Intelligence”, Hans Peter Luhn. In an International Business Machines (IBM) article, Hans separately defines both business and intelligence. According to him, business is "a collection of activities carried on for whatever purpose" and intelligence is the “ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal” (Luhn, 1958). Therefore, it is possible to state that for him, BI were procedures done to gather information helpful to define the way forward regarding a goal. After almost a century, what leveraged this discipline was the IBM discovery of the computer disk storage system, that in turn, had a huge contribution to the development of other IT. In this sense, a more complex definition arises that describes BI as "concepts and methods to improve business decision making by fact-based support systems" (Dresner, 1989).

On that account, from the early 1990s until today, BI has gone through three generations, and there is no news that all of them are related to massive technological landmarks. The first generation (BI 1.0) embodies mainly structured data stored in a management database. It relies on ETL (extract, transform and load), OLAP[1] (online analytical processing), queries and reporting tools useful to integrate and analyse data, that give rise to dashboards for better visualization of the business performance. BI 2.0 ground heavily on both the internet and web, which boosted unstructured data offer by the now possible direct contact with customers. Besides, it also provided a better awareness of customer needs and preferences deriving from constant feedback. Ultimately, the Internet of Things is the fingerprint of this ongoing third generation. These devices generate cosmic amounts of location-aware, person-centred and context-relevant data (Chen et al, 2012).

2.2. Information Systems

As presented in the previous section, alongside the generations of BI, the amount of data available has burgeoned in volume, variety and velocity. This phenomenon led to the Big Data concept. According to McKinsey (2011), Big Data "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse".

Supporting this perspective, Big Data is far more than data with those characteristics. Instead, it merges the capacity of the IS to support such data. Hence, it is of utmost importance to implement a good IS.

[1] "OLAP is software for performing multidimensional analysis at high speeds on large volumes of data from a data warehouse, data mart, or some other unified, centralized data store." (IBM)

IS in its main core is a system responsible for collecting, transforming and storing information that is relevant to managerial decisions. However, nowadays, this notion is umbilically linked with the IT term, which according to Gartner includes "the entire spectrum of technologies for information processing". Therefore, it is currently seen as a subject whose purpose is to implement and maintain a technological information system. There are several stages throughout this system as shown in Figure 1.

Figure 1 - Information System Stages

Data can arrive from different sources, such as flat files, ERP and web services. The difference between these sources relies on their format. While flat files, for instance, Excel files, and ERP files are pleasantly organized, often by a matrix where rows represent records and columns represent attributes of each record, unstructured data that comes from web services does not follow this pattern or any other (Skiena, 2017). The following stage - Data Process - includes ETL tools which purpose is to extract data from data sources, manipulate it to meet certain standards (transform), and finally load this data. The process of loading the data initiates the Data Storage phase. According to specialists, despite the array of choices, data storage generally follows one of the two models: Bill Inmon's and Ralph Kimball's. Both experts were recognized for their seminal works *Building the Data Warehouse* (2002) and *The Data Warehouse Toolkit* (2000), respectively, but Inmon was the one titled as the "Father of Data Warehousing".

Mary Breslin, in an article for Business Intelligence Journal (2004), compared, in her words "The Giants" models. It is deduced that both models are similar on subjects such as the use of time-stamped data and ETL process since are imperative for the viability of the decision support system. However, only the first two phases of ETL are common. The conflict between these models strikes up with where transformed data is loaded. According to Inmon's perspective data is loaded to a unified DW whereas in Kimball's model data is loaded to smaller databases – Data Marts.

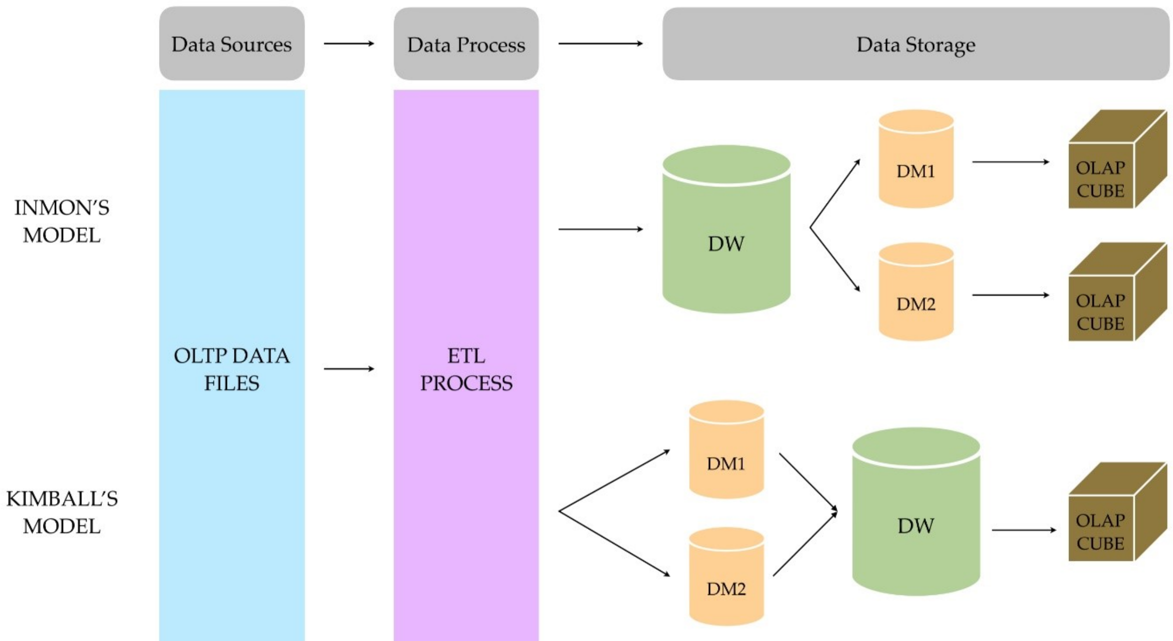


Figure 2 - Comparison between Inmon's and Kimball's models

The differences between the mentioned models are mostly related to their methodology, architecture, data modelling and philosophy. Focusing on the first two areas, Inmon's and Kimball's models have contrasting approaches, structures, and complexity – Table 1.

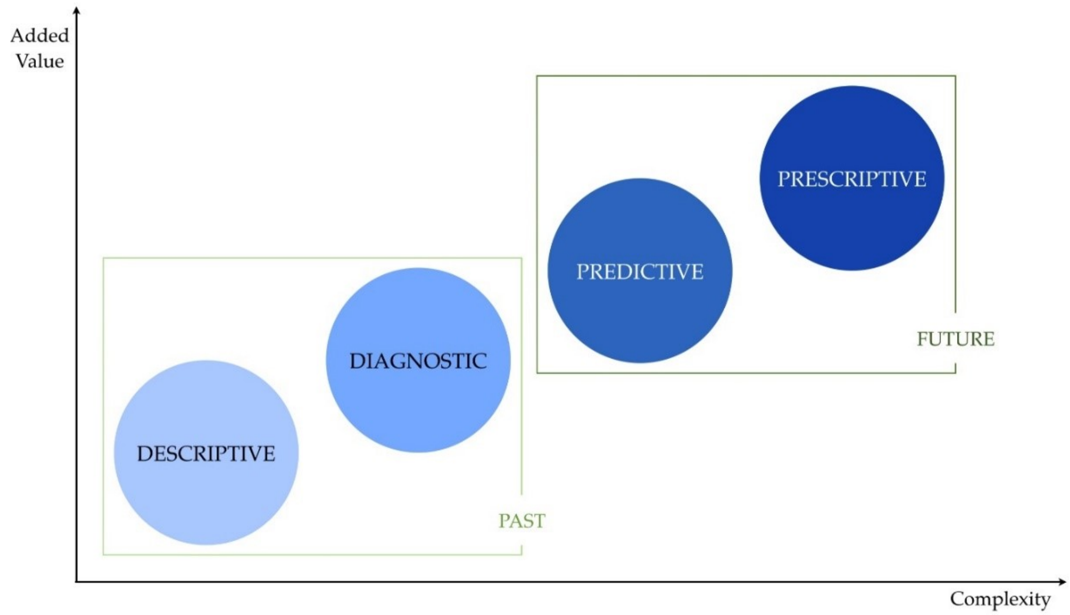
Inmon's model follows a top-down approach, meaning that data is loaded to an atomic DW, feeding all the organization's departments' databases. Whereas Kimball's adopts a bottom-up strategy, which means that data is firstly loaded to each department's Data Mart, that in turn creates the DW. Complexity is related to the number of OLAP Cubes created. A bottom-up approach, in its essence, involves fewer elements since cubes are originated from the DW instead of having a cube for each Data Mart (Inmon's model). It is also important to mention that Inmon philosophy has as its target audience IT users, with substantial knowledge on the matter. Oppositely, Kimball recognizes the end-user as an integral part of process development, making his model more user friendly.

	INMON	**KIMBALL**
Approach	Top-down	Bottom-up
Structure	Unified Data Warehouse supports departmental databases	Data Marts of each department feed the Data Warehouse
Complexity	High	Low
Audience	IT professionals	End users

Table 1 - Differences between Inmon's and Kimball's Models

Source: Adapted from Mary Breslin, 2004

Data Analysis, the following phase, is where BA takes action. As stated by Harvard Business School (2019), BA is "the process of using quantitative methods to derive meaning from data to make informed business decisions". To this definition can be added BA's purpose that, as reported by Gartner, is "to build analysis models and simulations to create scenarios, understand realities and predict future sales". In this sense, BA consists of four types of analysis – Descriptive, Diagnostic, Predictive and Prescriptive – all with different degrees of added value for the business itself and complexity as shown in Graphic 2. Obviously, these two characteristics are linear to each other, which means that more complex analyses add more value to the business and vice-versa.



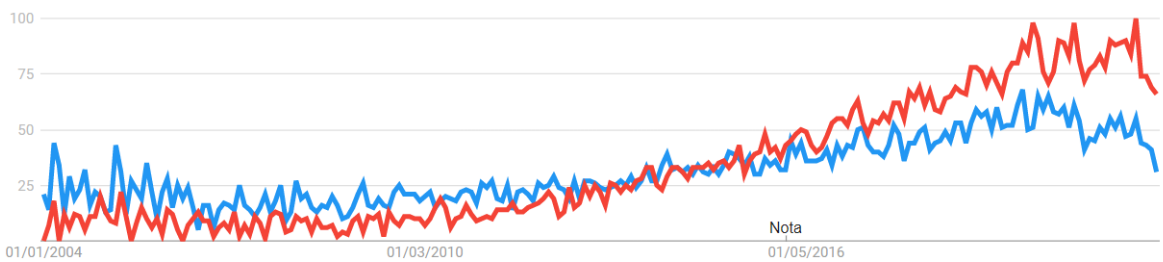
Graphic 2 - Types of Business Analytics

Gartner also defines each type of analysis. Descriptive Analytics is related to analyses performed on past data so that the question “What happened?” can be answered, and it is also characterised by traditional BI techniques and Data Visualization. The outcome of these analyses needs to be interpreted, which is done through Diagnostic Analytics, to answer “Why did it happen?”. Some examples of BI procedures used in this type of analytics are correlations, drill-downs and data discovery. While the first two types adopt a hindsight approach, the remaining ones ground a forecast analysis. Predictive Analytics is defined by more advanced techniques such as regressions and forecasting to explain “What is going to happen?”. Finally, Predictive Analytics represent the ultimate goal of knowing “What should be done?” or “What can be done to make that happen?”. Besides being the type of analysis that adds more value to a business, it also has the dichotomy of being the more complex one, consisting of simulations, heuristics and machine learning.

2.3. Data Visualization

Under the IBM perspective, Data Visualization is the process of translating large data sets and metrics into appealing visuals that allows an easier identification and sharing of trends, outliers and insights. Additionally, the technological corporation outstands its importance in seeing the big picture and in making better decisions.

Contrary to the BI term, and leveraged by BA (red line) popularity, Data Visualization (blue line) concept is also currently trending, as shown in Graphic 3. This phenomenon can be explained by the fact that organizations need to establish a robust IS throughout BI tools so that they can collect data to analyse and posteriorly develop credible dashboards and reports.



Graphic 3 - "Business Analytics" and "Data Visualization" words trend

Source: Google Trend (06.01.2022)

Stephen Few is known for his blog, Perceptual Edge, where he archives his works, to help people to communicate data more persuasively. One of the topics addressed in his book *Show me the Numbers* is what display to choose to represent quantitative data. He states that most of the time people tend to adopt the wrong one. Further, while showing real examples of how organizations present their data, he agrees with the saying "Keep it short and simple", explaining that even though people want to embellish their visuals, it often results in the distraction of the data reader from the pivotal point. Therefore, he identifies two fundamental challenges of data presentation, which are first determining the best way to tell the story (graph or table) and then designing its components for better message streaming.

On the other hand, in another of his books *Dashboard Design for at-a-glance monitoring* he addresses what analysts should avoid, identifying the following mistakes:

- Exceeding the boundaries of a single screen – the main point for avoiding the need to scroll down or over to see more data is that the ability to make comparisons and relationships and to see the big picture weakens.
- Supplying inadequate context for the data – with this mistake what Stephen wants to express is that data by itself does not tell much and always requires a context, such as a comparison to see if something is on track.
- Displaying excessive detail or precision – here the objective is to let the analyst know that too detailed data can delay the capture of information by the end-user. For example, rounding when displaying quantitative information.
- Expressing measures indirectly – the indication of what is being measured, and in which units must be clear to the data reader. He adds the following example: if the reader needs a difference between two values, there is no need to show both, but instead the dashboard must show the difference value directly.
- Choosing inappropriate display media/Introducing meaningless variety – a frequent mistake is to use the wrong type of graphic and sometimes people tend to choose different visuals to break the monotony of the dashboard, but it results in a wrong representation of the data.

- Using poorly designed display mechanisms – in addition to the choice of what visual to use, it is important to clearly communicate the data. For instance, if the purpose is to compare a real situation with its target one for different regions, using a bar chart it does not make sense to have the actual situation on the left and the target one on the right side of the chart. For a more intuitive understanding, the actual and target bar for each region should be placed together.
- Encoding quantitative data inaccurately – the alert made with this mistake is that people need to pay attention to graphics with more than two dimensions. When comparing revenues and costs over time is important to correctly define the scale of the first two variables, otherwise just looking at the graphic may be misleading.
- Arranging the data poorly/Ineffectively highlighting what is important – a dashboard should be organized in terms of importance, and the data must be presented in a way that fits the way it is used. The most prominent position is the top left corner and it should be reserved for the most important information. In this sense, there should not be a lot of information highlighted because if it is, it ends up not standing out anything.
- Cluttering it with useless decoration – as stated previously, Stephen is of the opinion of keeping the dashboard as clean as possible. For him, too much decoration can provide the idea that is needed to keep people entertained, when this is not the goal of a dashboard.
- Misusing or overusing colour – colour can help pass the right message; however, it must be used carefully. It is important to be aware that people tend to relate information in the same colour and eyes tend to roll to

anything that contracts the norm, even if it is not such relevant information.

- Designing an unappealing visual display – even though being attractive is not the main goal of a dashboard, an analyst does not want that people avert their eyes when looking at it, as it contains crucial information.

Few's perspective goes eye to eye with the multinational software corporation SAP one. According to SAP, ten rules should be followed when designing a dashboard. There are as follows:

PURPOSE	**RULE**
Ensure Appropriate Visualizations	Use containers that are best suited to showing and reading your type of data
	Compare data using charts appropriate for the type of comparison. Use truthful scale factors
	Show data and its relation as part of a whole
	Bind the data to where it makes sense geographically
	Use charts that plot the relationship between measures, and the distribution of values within a measure
Ensure Dashboard Optimization	Build using a grid to align and scale elements consistently
	Understand colour theory to use colour appropriately
	Contrast a single neutral colour for graphic elements with visible colours that have known semantic meaning for data highlights and alerts
	Keep the dashboard design uniform and coherent with the corporate style and visual standards
	Ensure everything in the dashboard is relevant and revealing so the user understands the data and is empowered to make a confident decision

Table 2 - SAP 10 Rules When Designing a Dashboard

Source: SAP PRESS

Tableau Software, a data visualization software company focused on BI, conducted a study to track a person's eyes while looking for the first time at a dashboard using a gaze opacity and heat map as control instruments. The results are shown in Figures 3 and 4. The gaze opacity methods consist of making the areas most focused on by peoples' eyes more transparent while in the heat map the areas of more focus are highlighted by warmer colours such as yellow, orange and red. Both methods foster the same conclusion, meaning that when overlapping them, the transparent areas tend to correspond to the hot areas.

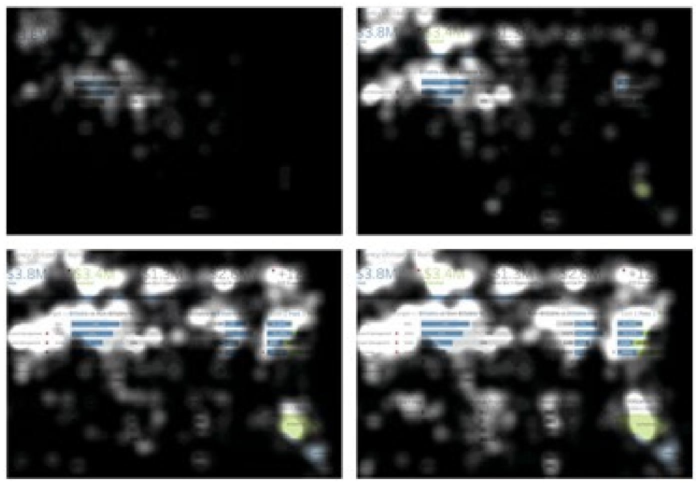


Figure 3 - Eye Tracking Map: Gaze Opacity

Source: Tableau

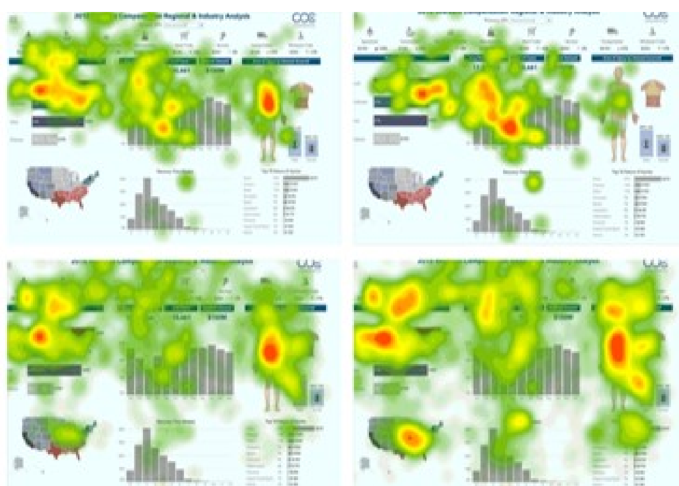


Figure 4 - Eye Tracking Map: Heat Map

Source: Tableau

These analyses help to understand the concept of Dashboard Real Estate. Few also addresses this concept. In his aforementioned book, he stated that researches indicate that different screen areas receive different levels of attention, which matches the Tableau study. Therefore, a page can be divided according to its prominence as follow:

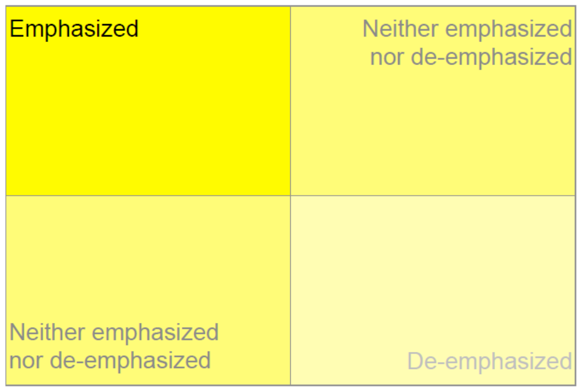
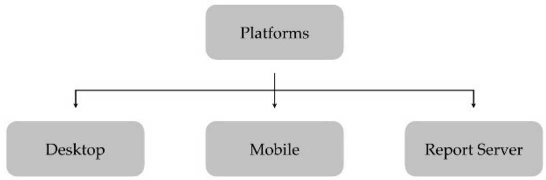


Figure 5 - Dashboard Real Estate

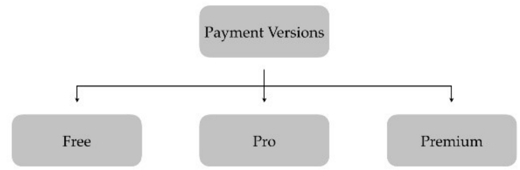
Source: Stephen Few, Perceptual Edge at Dashboard Design for at-a-glance monitoring

2.3.1. Power BI

One of the most acclaimed tools in terms of data visualization is Power BI from Microsoft launched in 2011 as a BI platform. It is defined by the organization as "a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights". There are three types both of platforms where Power BI can be used and of payment versions as presented in Graphics 4 and 5.



Graphic 4 - Power BI Platforms



Graphic 5 - Power BI Payment Versions

Power BI can be detached into two different sections depending on the adopted code. The section where the report is designed uses DAX which stands for Data Analysis eXpressions. This programming language lies on an internal engine called Tabular Models, which stores data in memory requiring, for that reason, a lot of memory and faster CPUs (Central Processing Units). It fits better columnar storage resulting in better data compression. The alternative model is Multidimensional, which is considered to be a more mature model but challenging to implement. On the other hand, the transform data section is based on M Code (M for Mash-up), which is the code behind Power Query. This is the section where the user can easily connect to different data sources and transform data according to his needs since the program automatically writes the code itself accordant to the action the user chooses.

This platform also distinguishes the concepts of dashboard and report. While Dashboard is more a daily view of Key Performance Indicators from different reports, providing a guiding point to the reports themselves, reports are a combination of related visuals that can be criteria filtered. However, the rules for designing a good dashboard also apply when building a report.

Third Chapter

3. Research Methodology

"No methodology can guarantee success. But a good methodology can provide a feedback loop for continual improvement and learning." – Ash Maurya

The previous chapter was crucial to provide a theoretical framework regarding important concepts related to the project on which this dissertation is based. Now, chapter three presents the methodology adopted along with the report restructuring.

This chapter is composed of three sections: project goals introduction in subchapter 3.1, research contextualization where the project was developed in subchapter 3.2 and finally, a theoretical description of the chosen method in subchapter 3.3.

3.1. Project Goals

As stated previously, the purpose of this dissertation is to present an improvement of one of the company's reports based on upgrades on BI tools, more specifically, Power BI from Microsoft. Even though the restructuring presented further ahead in chapter 4 was developed specifically for Polisport, the improvements made can be applied to any company from any business sector as long as they also resort on the used tool, as it focuses mainly on technical changes.

Since this dissertation is written in the form of an Internship Report, it does not answer a specific research question formulated based on literature gaps. Instead, so that the main goal, improving efficiency, be reached, this research has as its main purpose to achieve the following articulated objectives:

- Improve the connection to data sources through Data Marts and Data Flows
- Rule out tables' unnecessary information, by filtering and consequently not loading this information
- Formalize steps on Power Query, commenting the reason behind some steps
- Implement a star schema[2]
- Optimize data visualization, mostly according to the dashboard real state

The internship took place in a company called Polisport Plásticos, which will be presented in subchapter 3.2. Due to the amass of data, the dataset that undergirds the report was too heavy that was causing query problems, such as taking hours to update a query change, or in the worst-case scenario, not allowing the refresh of new data at all.

Therefore, Polisport proposed the topic "Improvement of BI tools" as the company believed it was time to upgrade the methods used in its reports to continue to withdraw knowledge from them. The report in question concerns the approved investment in projects. Its leading intention is to provide grounds to evaluate if the real situation of a certain project is in line with what was defined in the feasibility study and to forecast the consequences of changes in certain parameters, such as costs and quantities in cases where they may be too conservative or optimistic.

[2] Star Schema: is the result schema of crating joins between fact and dimension tables. There is a fact table surrounded by dimensional ones, creating the familiar form of a star (Ferrari and Russo, 2008).

3.2. Research Contextualization – Polisport

Polisport Plásticos, SA is a Portuguese company located in Oliveira de Azeméis, in the district of Aveiro, belonging to the Porto Metropolitan Area, in the North Region. The company is part of Polisport Group and is specialized in Plastic Injection Molding, which has its focus mainly on plastic products and accessories for two-wheeled vehicles.

The group was founded in 1978 by the CEO, Pedro Araújo, passion for the world of two wheels and with 44 years of history has a solid market presence in its core business, especially in the manufacture of baby carries, with a production that surpasses half a million units per years. It exports nearly 90% of its production to over than 70 countries worldwide and operates in both bicycle and off-road accessories. Nowadays, the group consists of five companies: Polisport Plásticos, Polinter Plásticos, Polisport Molds, Polipromotion and Polistar Brasil. All these companies combined are responsible for the creation of 530 workstations. Over time, the group has also been acquiring brands, creating its own portfolio, such as Bobike in 2013, which is a prestigious Dutch brand of baby carriers and Catlike in 2020, which is a Spanish brand known for its helmets.

The commitment of every single employee allowed the group to achieve several awards throughout its history. In 2008, Polisport won its first IMDA for Best Injection Molding with IPD. The next year, the group was again awarded more two IMDAs for Best Injection Mold Part and Best Durable Product IMD. In 2010 they were recognized as the Most Innovative Portuguese Company receiving the COTEC-BPI SME Innovation Award. Two years later, the bicycle child seat won the Taipei Cycle Award. Additional, in the same year, they also won two IF Design Awards, namely Product Design and Material Design. In 2015, Polisport was once again awarded an IF Design Award, this time with the Guppy Junior. All these awards and recognitions result from its passion, innovation and globalization. Polisport holds 33 patents, 10 trademarks and 26 design registrations with recognitions and awards worldwide.

Polisport Plásticos, in particular, is the company where the shared service centre is based. This company has an open space in its installations where the main departments are located. They are divided into support and operational. The first category integrates departments such as environmental quality and safety, internal audit, financial, facilities and equipment management, information technologies and systems, human resources and management planning and control. On the other hand, operational departments are more related to production and the market. It includes departments like commercial, purchases, engineering, innovation, logistics and marketing. Despite each department having its specific tasks, on daily basis there is a lot of communication between them, including periodic meetings, to ensure that the goals are aligned.

3.3. Method

Regarding the methodology chosen to achieve the restructuring goals proposed in subchapter 3.1, Action Research was adopted to inspire the project conduction as opposed to other alternatives. The choice had into consideration its practical strand as when analysing the results drives new ways of improving. That is, it looks for a change based on action, research and reflection of what has been done. Basically, this methodology chases action and research at the same time.

Despite there are no certainties about the emergence of the action research concept, the creation of the process is often attributed to Kurt Lewis due to his paper A*ction Research and Minority Problems* where he describes Action Research as "a comparative research on the conditions and effects of various forms of social action and research leading to social action" (Lewis, 1946). Even though some authors are defining the process differently, the majority agree on it not being a rigid definition, since it arises in different guises and have different applications. Some even there to state that "action research is a term which is applied to projects in which practitioners seek to effect transformations in their own practices…" (Brown and Dowling, 2001, p.152). Therefore, it is possible to conclude that this methodology is in its essence practical and reacts to the need of solving real-life problems. Kemmis and McTaggart (1989) explain Lewis' definition through the spiral of cycles presented in Figure 6. From it, it is possible to easily identify the four phases in Action Research: first, plan what is going to be done based on critical information; second, actually do what was planned – action; third, observe the output of what was done and finally, reflect about it to improve the plan, giving rise to a new revised plan. In each cycle of the spiral, the researcher's insights of the original problem increase.

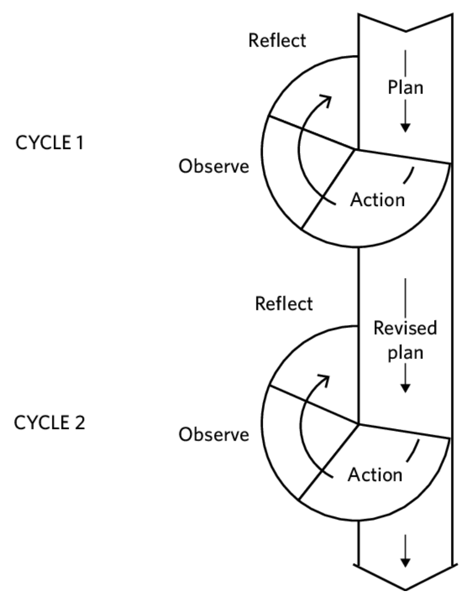


Figure 6 – Action Research: Spiral of Cycles

In the following chapter, chapter 4, the report restructuring took into consideration the adapted model presented in Figure 7. As can be seen, the spiral of cycles in the developed project consists of three cycles. All the phases detailed by Kemmis and McTaggart were repeated multiple times within the same cycle until the desirable result was achieved.

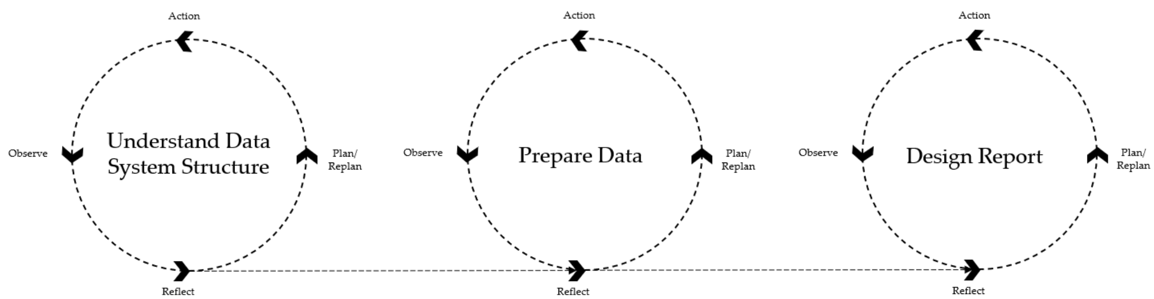


Figure 7 - Action Research: Adapted Model

The first cycle was named Understand Data Systems Structure. In this cycle the goal was to design a big picture of the IS and how information is stored, to see which model, Inmon's or Kimball's, the company was adopting. Additionally, and after finding out how the connections to the data were being made in the base report, a plan of how to allow an efficient flow of the data was traced.

The second cycle – Prepare Data – focuses on data analyses to see if it fits the needs of the report or not and on commenting on the reasoning behind the steps applied. For this reason, the plan was to rule out useless information, formalize the steps and implement an easily readable schema.

The third and last cycle, named Design Report, consists of understanding what information needs to be displayed. In accordance with that, both measures and the report's layout were evaluated. The plan was extremely connected to Few's approach on how to build a report, highlight the most important information and use appropriate visuals.

Fourth Chapter

4. Report Restructuring

“A project is complete when it starts working for you, rather than you working for it.” – Scott Allen

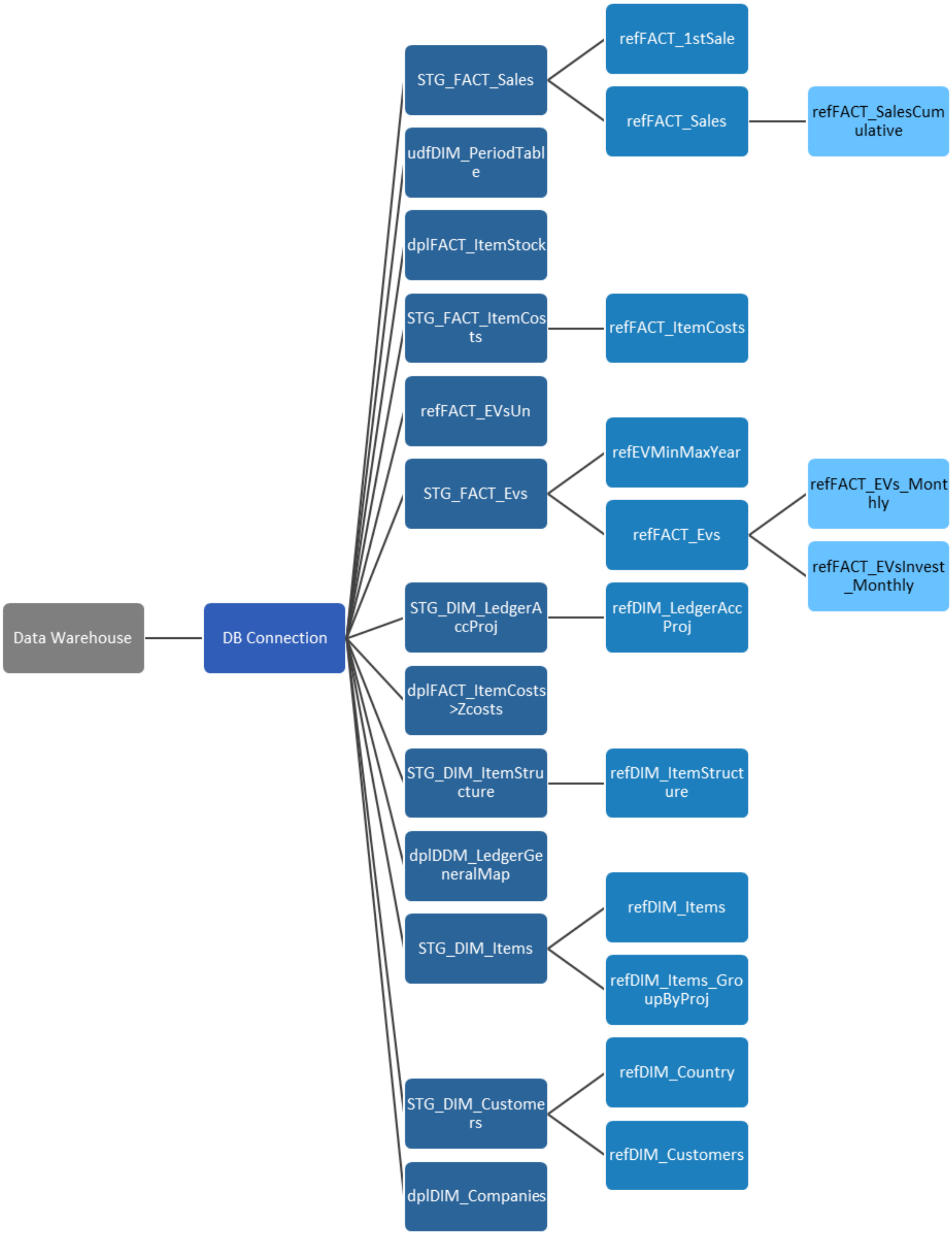
Whilst the third chapter provided a theoretical framework of the methodology used during the report restructuring at Polisport, this chapter will now present, in detail, the practical application that the Action Research method had on it. The subchapters contemplated here correspond to the three important stages of the restructuring. In this sense, subchapter 4.1 focuses on understanding what was done in the base report, analysing its performance, both in terms of processing and data interpretation, and concluding about aspects that can be improved; subchapter 4.2 explains what changes were made, their reasoning and how they were achieved, and lastly, subchapter 4.3, presents the final report, comparing it to the base one and showing true evidence of enhancement.

Throughout the subchapters the three cycles of Action Research – Understand Data System, Prepare Data and Design Report – will be addressed, once all these stages are worthy of analysis to discover potential improvements. It must be highlighted that more than one turn was given to the different cycles and that moving forward to the next cycle does not mean that there is no need to adjust something later in the previous one. In fact, going back a cycle was quite common when designing the report.

4.1. Base Report

Nowadays, Polisport is restructuring its DW, having as example Kimball's model. At the time the report was designed Polisport did not have that perspective into consideration and there were no Data Marts and Data Flows created in the Data System. In this sense, an analysis of the source of each table was conducted, resulting in the map of dependencies presented in Graphic 6. Polisport started to create a connection to the DW – *DB Connection* - and then all the needed tables were linked with that connection as the source. As can be seen from it, there were a total of 13 first-degree[3] tables, which in turn were the source of other tables.

[3] During this dissertation a first-degree table is a table that is directly connected to the Data System.



Graphic 6 - Map of Dependencies: Base Report

Besides these tables, connected directly or indirectly to the DW, in the report itself, a Calendar table was created from scratch using queries. These types of tables are referred to as Raw Tables in this study. Additionally, at that time, it was not possible to create folders inside tales to organize the measures, so five blank tables were also generated in the Power Query ribbon for this purpose. Their names and the type of measures for which they were created can be found in Table 3.

TABLE	MEASURES
Time Comparison	Time horizon measures (e.g., TTM)
Key Measures	Key measures for the report (e.g., sum of all sales value)
Cumulative	Cumulative measures (no filters applied)
Average	Averages of measures already created or average measures (e.g., Average Sale Price)
Allocation	Allocation measures (relate a measure to another or to an attribute)
Stata	Data mining related measures (e.g., correlations)

Table 3 - Blank Tables' Purpose

Regarding the linked tables, they were analysed both in terms of purpose and performance. The purpose analysis includes an understanding of what data each table stored that was relevant to compare the situation established in the feasibility study with the real situation of each approved project (Table 4). It is important to underline that a project can have more than one feasibility study. For instance, if the project relied on the manufacture of a handguard, each colour available represents a different feasibility study since the colours to dye the pieces have different costs.

TABLE	PURPOSE
LedgerGeneralMap	Supports the calculation of the real investment
Companies	Contains information about the companies of the group
dplFACT_ItemCosts>ZCosts	Supports the depreciation calculation
dplFACT_ItemStock	Contains stock information
refDIM_Items_GroupByProj	Contains information about how many items are associated with a certain project
refEVMinMaxYear	Gives information about the adjusted year range of the feasibility study
refFACT_1stSale	Gives the first date of real sales for each feasibility study
refFACT_EVsInvest_Monthly	Gives information about the predicted investment
refFACT_EVsUn	Gives the first date of sales for each project foreseen in the feasibility study
Customers	Contains information about all the clients
Items	Contains information about all the items
ItemStructure	Supports the depreciation calculation
LedgerAccProj	Contains information about all the projects
Evs	Contains information about the feasibility studies
ItemCosts	Supports the depreciation calculation
Sales	Contains information about the sales
Calendar	Contains date information
PeriodTable	Supports the dates division into different time horizons

Table 4 - Linked Tables' Purpose

The Power BI Desktop integrates some tools that help to gain knowledge about the data. One of those tools is called DAX Studio which can be found on the *External Tools* tab. This tool was critical to the performance analysis. When inside DAX Studio, there is a command called *View Metrics* that displays information and metrics about the model allowing a detailed analysis of it. An adapted result is shown in Figure 8. Despite not all the tables being loaded into the model, they interfere with the report's performance as queries continue to be run. These tables were called staging tables, being identified with a "STG", whereas the "DL", is for *Data Load*. In the base report, Polisport had as police to create a staging table whenever a new applied step to transform data broke the query folding[4]. To interpret the result is pivotal to apprehend what information each column is providing. Therefore, based on the definitions of the platform used, Attachment A was built.

In this sense, highlighted in red (Figure 8) are the cells with the highest numbers in each column, which represent the tables with greater weight in each variable. The table with more distinctive values is the factual *Item Cost*, which makes sense since it contains the cost of each item per day. As for columns and, consequently, table size the ones that stood out were the *Ledger General Map*, *Sales*, and, once again, the *Item Cost*. Following a logical reasoning, as these were the heavier tables, they also had to be the ones that represented a larger percentage of the database. In terms of the number of columns, there were many tables with a large number, namely the *Item* table with over than 50 columns. The existence of many columns suggests that the process of preparing and transforming data was not done having into consideration the needs and purpose of the report in advance. What reinforces this

[4] "Query folding is the ability for a Power Query to generate a single query statement to retrieve and transform source data. The Power Query mashup engine strives to achieve query folding whenever possible for reasons of efficiency" (Microsoft).

unfavourable situation is the fact that some of these tables were duplicated increasing the effect produced.

Table	STG/DL	Cardinality	Table Size	Columns Total Size	Data Size	Dictionary Size	Columns Hierarchies Size	Database Size %	Columns #
dplDDM_LedgerGeneralMap	DL	1 784 662	49 778 244	49 157 012	27 524 016	17 437 668	4 195 328	13,94%	39
dplDIM_Companies	DL	5	86 374	86 374	40	85 934	400	0,02%	5
dplFACT_ItemCosts>ZCosts	STG	4 626	191 913	191 913	22 176	128 289	41 448	0,05%	2
dplFACT_ItemStock	STG	23 692	1 107 173	1 107 173	54 328	857 309	195 536	0,31%	2
refDIM_Country	STG	101	40 020	40 020	192	38 148	1 680	0,01%	2
refDIM_Customers	DL	4 130	1 576 954	1 557 674	70 840	1 296 226	190 608	0,44%	37
refDIM_Items	DL	1 776	1 709 955	1 709 819	88 984	1 460 579	160 256	0,48%	68
refDIM_Items_GroupByProj	STG	165	22 648	22 648	320	20 784	1 544	0,01%	2
refDIM_ItemStructure	STG	4 047	310 234	310 234	12 968	232 386	64 880	0,09%	3
refDIM_LedgerAccProj	DL	79	131 368	131 368	616	126 192	4 560	0,04%	9
refEVMinMaxYear	STG	446	75 058	74 866	1 240	68 218	5 408	0,02%	5
refFACT_1stSale	STG	137	22 142	22 142	368	20 166	1 608	0,01%	2
refFACT_EVs	DL	2 172	773 902	773 670	148 728	479 854	145 088	0,22%	37
refFACT_EVs_Monthly	DL	22 800	2 389 106	2 388 874	738 928	1 433 930	216 016	0,67%	29
refFACT_EVsInvest_Monthly	DL	3 276	160 148	159 916	19 112	131 412	9 392	0,04%	8
refFACT_EVsUn	STG	184	27 584	27 584	592	24 864	2 128	0,01%	4
refFACT_ItemCosts	DL	16 555 145	78 605 247	78 582 431	71 472 464	5 950 255	1 159 712	22,01%	4
refFACT_Sales	DL	151 469	59 368 425	59 359 105	11 161 944	41 148 721	7 048 440	16,63%	46
refFACT_SalesCumulative	DL	28 186	1 568 666	1 568 666	272 896	945 354	350 416	0,44%	10
STG_DIM_Customers	STG	5 410	1 765 654	1 765 558	89 512	1 437 054	238 992	0,49%	37
STG_DIM_Items	STG	2 172	1 315 309	1 315 133	43 360	1 168 069	103 704	0,37%	51
STG_DIM_ItemStructure	STG	17 466	764 992	758 504	162 200	461 312	134 992	0,21%	12
STG_DIM_LedgerAccProj	STG	1 532	242 393	242 393	9 832	192 489	40 072	0,07%	5
STG_FACT_EVs	STG	3 849	1 899 641	1 899 449	110 904	1 457 889	330 656	0,53%	29
STG_FACT_ItemCosts	STG	22 227 565	108 442 964	108 417 396	96 672 120	10 302 076	1 443 200	30,37%	4
STG_FACT_Sales	STG	826 022	43 548 343	43 541 575	25 159 768	14 659 159	3 722 648	12,20%	26
udfDIM_Calendar	DL	6 938	694 950	694 950	90 304	538 502	66 144	0,19%	21
udfDIM_PeriodTable	DL	29 115	340 592	334 256	70 864	236 336	27 056	0,10%	11
Grand Total		**41 707 320**	**357 087 645**	**356 368 349**	**233 999 912**	**102 464 037**	**19 904 400**	**100,00%**	**527**

Figure 8 - DAX Studio - View Metrics Analysis: Base Report

In the data preparation part, many of the steps applied were not formalized, that is, there were not many comments on the code explaining the reasoning behind some more complex queries related to business. This does not appear to be a serious matter until someone who is not familiar with that report/data needs to make adjustments. Another situation that stood out was that some filters applied required manual updates if the situation changes. An example can be the company filter applied in almost every table as just the companies that have feasibility studies matter. However, if another group company decides to start predicting projects' situation, the filter must be updated one by one, which is definitely not efficient.

After finalizing the Power Query ribbon analyses, it was time to move forward with the understanding of the base report to the DAX ribbon, where there were three main areas of analysis: Schema, Measures and Layout.

The schema is what completes the data preparation and can be found on the *Model* tab. A schema is the basis of the entire report since this is where relationships between previously created or connected tables are established. Assessing the one on the base report it did not represent any theoretical model (Figure 9). Apart from containing mainly single directional relationships, which is the best practice, it also includes a bidirectional one between the *Calendar* and *PeriodTable*. The reason behind this type of relationship is to be able to filter both ways. Another factor that immediately stood out was the relationships between dimensions. The simplest form of schema – Star Schema – does not approve these relationships as it turns the design more complex. The one presented is more likely a snowflake, since fact tables are surrounded by dimensions tables, which are consequently surrounded by other dimensions tables.

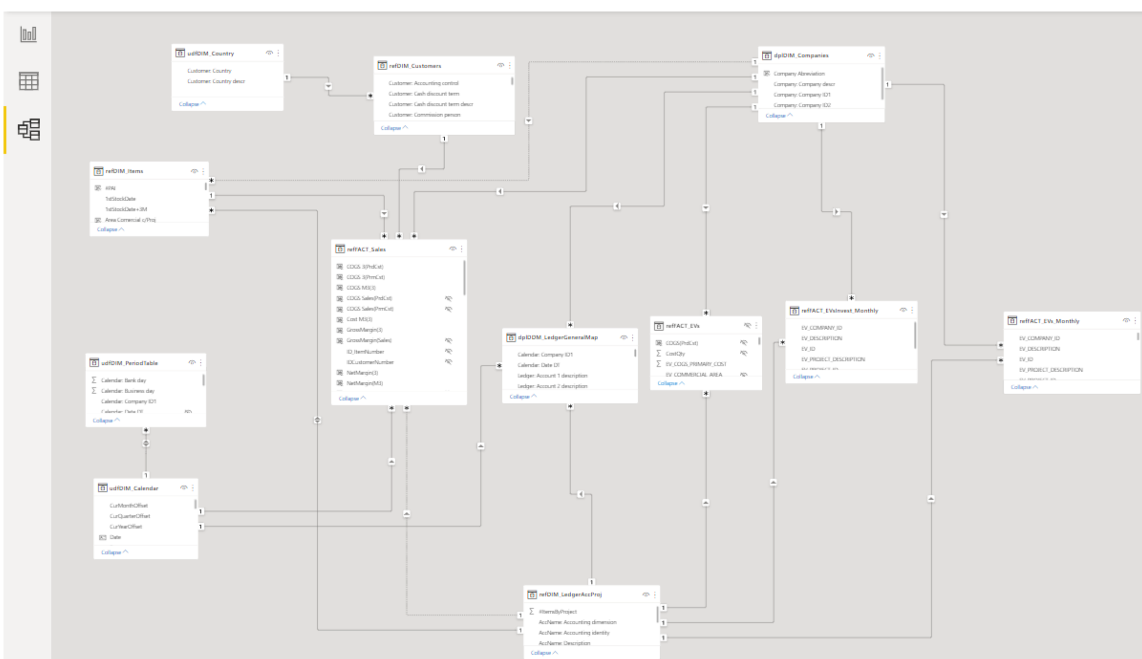


Figure 9 - Schema: Base Report

Moving forward to the measures, it was previously stated that were created blank tables to organize the measures created ahead. This was a strategic step due to the number of measures that were being created. All measures have a home table, where they can be found in the field list appearing with a calculator icon. This report contained 424 measures, in each over a hundred were time-related. That is, the report analysed the projects' situation in many time horizons, so some created measures overlapped each other being the only difference the analysed time. For an easier understanding, an example related to sales is presented in Appendix A From it, it is possible to see that the only addition to the main measure – *Sales* – is a date filter allowed by the *CALCULATE* formula.

When scrolling around the pages what immediately stood out from all the information were the slicers, which positions were not consistent across the report and, above that, some were even displayed on the most valuable area of the page as it is illustrated in Figure 10. Obviously, the slicers' importance is not being questioned, as they are required so that the user receives the information he wants. Nevertheless, they are not what the user wants to capture as an end goal.

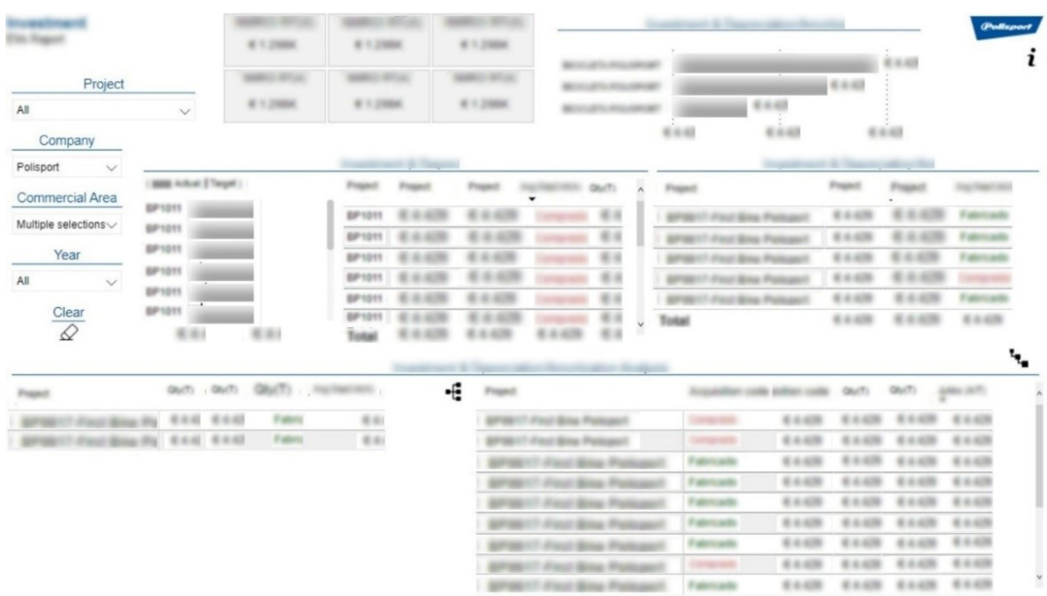


Figure 10 - Slicers' Positions: Base Report

The last aspect, this time not so noticeable, was the number of elements on sheets. For instance, one of the sheets contained 105 elements and each element is processed when the page opens. Additional, when displaying elements that overlap each other the analyst must pay special attention to the *Layer Order* available in the *View - Selection* tab.

4.2. Improvements

Nowadays already exists on the Data System a Data Mart containing tables with better-prepared data. Some of the base report tables were replaced by others already on the Data Mart, and time-related tables - *Calendar* and *PeriodTable* - were created and placed on the same Data Flow (Table 5). Additionally, some tables were no longer needed, such as the ones which purpose was to calculate an item's depreciation, because such information was incorporated on other tables already used, the ones duplicated for query folding reasons, and the blank ones, since it is now possible to organize measures in folders on a connected table. Table 6 shows the modifications made in this regard on connected tables.

TABLE	BASE SOURCE	TABLE NAME	NEW SOURCE	NEW TABLE
dplDIM_Companies	DW	Companies	DM	Division
dplFACT_ItemStock	DW	ItemStock	DW	Stock History
STG_DIM_Customers	DW	Customers	DM	Customer
STG_DIM_Items	DW	Item	DM	Materials
udfDIM_PeriodTable	DW	Calendar	DF	PeriodTable
refDIM_Country	DW	Customers	DW	Country_UnionBlock
udfDIM_Calendar	Raw Table		DF	Calendar

Table 5 - Source Changes

TABLE	REASON
STG_FACT_Sales + refFACT_Sales	Became just one DL table (query folding)
refFACT_SalesCumulative	Removed for repeated data
dplFACT_ItemCosts>ZCosts	Removed since it was no longer needed to calculate items' depreciation
STG_FACT_ItemCosts	Removed since it was no longer needed to calculate items' depreciation
refFACT_ItemCosts	Removed since it was no longer needed to calculate items' depreciation
STG_DIM_ItemStructure	Removed since it was no longer needed to calculate items' depreciation
refDIM_ItemStructure	Removed since it was no longer needed to calculate items' depreciation
STG_FACT_EVs + refFACT_EVs	Became just one STG table (query folding)
STG_DIM_LedgerAccProj + refDIM_LedgerAccProj	Became just one DL table (query folding)
STG_DIM_Items + refDIM_Items	Became just one DL table (query folding); replaced by *Materials*
STG_DIM_Customers + refDIM_Customers	Became just one DL table (query folding); replaced by the *Customer* table of the DM

Table 6 - Removed and Combined Tables

Beyond this, two lists were created to help filter the information from other tables. As the purpose of the report is to compare the real situation with the one in the feasibility studies all tables were filtered by the company *"010"*, which represents Polisport Plásticos, through the query shown below.

$$Table.SelectRows(dbo\_utbDAFACT\_Sales, each\ [\#"Sales{:}\ Company\ ID1"]\ =\ \text{"010"}\ )$$

For now, it is the only company with feasibility studies. However, the aim is that all of them start to foresee projects' situations over time. In this sense, it was created a *Company* list with the feasibility studies table as source. The reasoning behind this option was that if a company forecasted an item's situation the information of the feasibility study must appear on this table. The queries applied are presented in Figure 11.

```
let
    Source = Sql.Database("Hannibal", "DW"),
    dbo_utbDAFACT_EstudosViabilidade = Source{[Schema="dbo",Item="utbDAFACT_EstudosViabilidade"]}[Data],
    #"Filter.Company<>null" = Table.SelectRows(dbo_utbDAFACT_EstudosViabilidade, each [EV_COMPANY_ID] <> null),
    RemDup = Table.Distinct(#"Filter.Company<>null", {"EV_COMPANY_ID"}),
    List = RemDup[EV_COMPANY_ID]
in
    List
```

Figure 11 – M Code: Automation Company filter

Regarding the weight of each table that stayed and the excessive number of columns, the plan was to check if there was any possible way to reduce it. This improvement was very time consuming because there was not a direct way to discover what data could be removed from tables. Therefore, it was a process that was done alongside the creation of measures to see which columns and rows were needed. It is worth mentioning that replacing some tables with similar ones on the Data Mart really helped the process.

Analysing the schema, the best improvement solution goes through trying to transform it into a Start Schema. The report is mainly used once a month to update the board of directors on their approved projects, so the main reason behind this choice is that this type of schema is easier to understand, the queries run faster and it provides optimal disk usage. The reduction of the number of loaded tables definitively benefited the understanding of the model. The only snowflake kept was the one between *Calendar* and *PeriodTable*.

To decrease the number of measures which, consequently, leads to a decrease in complexity, the restructured report resort on another external tool incorporated on Power BI Desktop called *Tabular Editor*. This is the first tool that enables the creation of *Calculation Groups*, which is a feature shown in the fields lists as a table with a single column. However, that is not a typical column, but instead, it represents more than one calculation that can be applied to a certain measure. Herewith, a calculation group called *Time Intelligence* was created as shown in Figure 12.

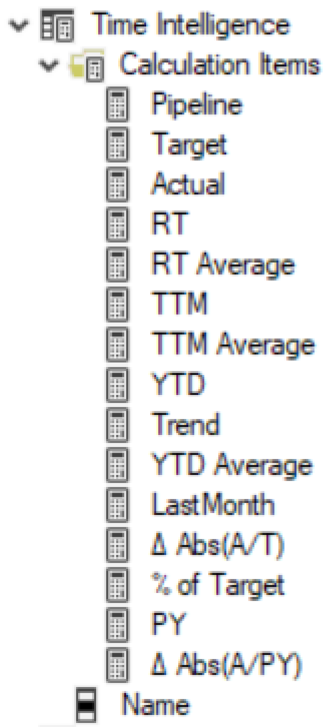


Figure 12 - Calculation Group: Calculation Items

The formula for each calculation is listed in Appendix B. The main reason behind the creation of this calculation group was to end with cases like the sales example presented in Appendix A, since it allows to achieve all those derived measures[5] creating only the main one. This calculation group was created apace from the report's layout to understand what would work better. For instance, blank calculations, namely *Pipeline* and *Trend*, were included so when displayed in a matrix form there was space to fit a graphic, as shown in Figure 13. However, for a matrix to show those columns it is necessary to activate the "*Show items with no data*" as demonstrated in Figure 14.

[5] Derived Measures refers to measures that depend on other measures previously created.

Figure 13 - Blank () Formula Purpose: Example

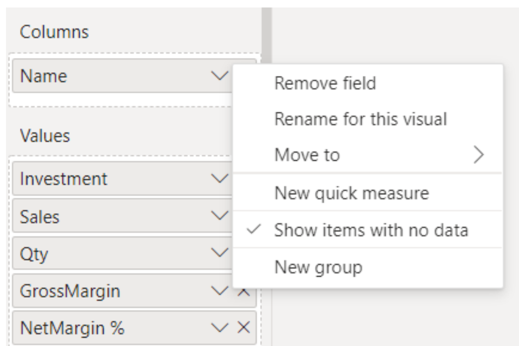


Figure 14 - Step to Show the Blank () Calculations

The best solution found to the slicers' problem was to resort to the concept of bookmarks[6], which allowed, in a way, to hide the slicers, increasing the free space on the report to display useful information. Along with the bookmark concept, the *Selection Pane* is also decisive for the solution's execution. On the *Selection* tab, those elements were separated into three groups:

- Group 1 – contains the slicers
- Group 2 – contains all the elements backing the visual of an open panel
- Group 3 – contains all the elements to be displayed with the panel closed

Figure 15 shows what elements each group contains. These groups must be placed in this specific order and are the first elements in the *Layer Order* so that when

[6] "A bookmark captures the state of a report page. This includes the settings you've made to filters, slicers, and visuals on that page. Simply select a bookmark, and Power BI takes you back to that view." (Microsoft)

the user wants to use a slicer, the set of elements that make up the bookmark upstages the page visuals. These groups must exist on all report pages. On the *Bookmarks* tab was added a bookmark to each page containing within it each of the three options – Clear, Slicers On and Slicers Off – as shown in Figure 16.

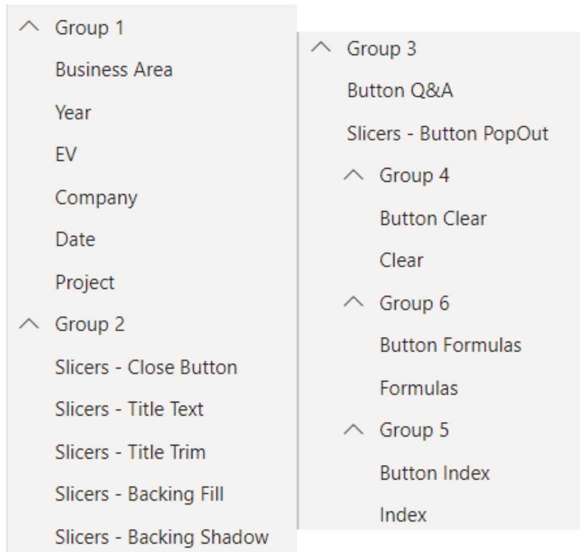


Figure 15 - Selection of Pane Groups' Elements

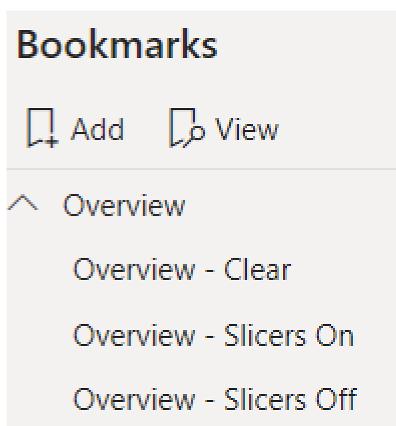


Figure 16 - Bookmark Page: Example

Table 7 tells if each group must be visible or not to achieve the goal of each bookmark. The cells highlighted in light grey mean that the group must be selected before updating the bookmark.

Bookmark/Group	Group 1	Group 2	Group 3
Clear	Hide	Hide	Visible
Slicers On	Visible	Visible	Hide
Slicers Off	Hide	Hide	Visible

Table 7 - How to Update the Bookmarks

Each of these bookmarks is activated by clicking on the element designated for that purpose, which are *Button Clear*, *Slicers – Button PopOut* and *Slicers – Close Button*, respectively. For this to happen, the last thing to do is to select each of these elements and add an action in which must be chosen the type *Bookmark* and then select the correspondent one, like the example shown in Figure 17.

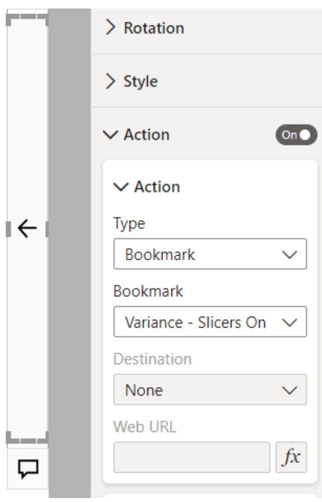


Figure 17 - Element/Bookmark Relation: Action Step

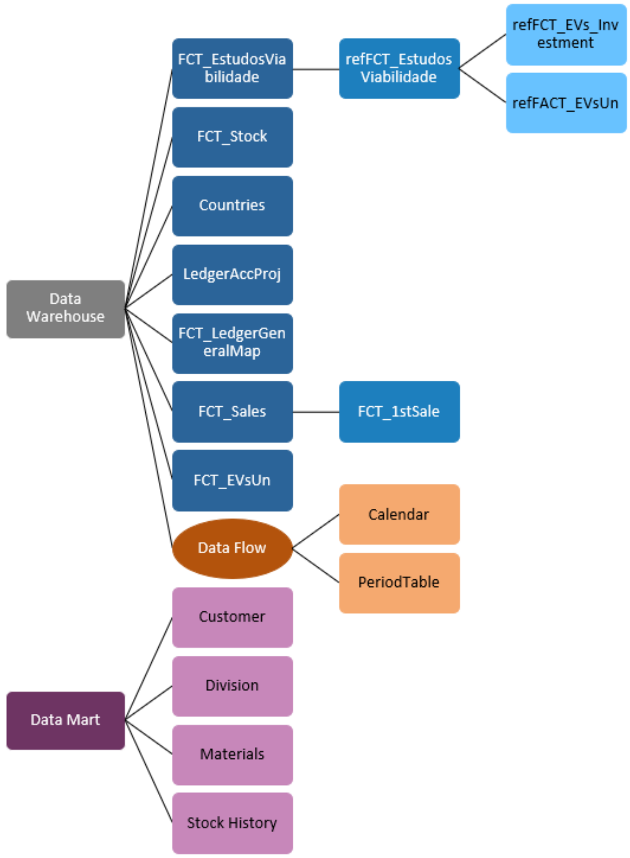
4.3. Final Report

To actually prove the existence of improvements in the processing and interpretation of the report, it is now necessary to analyse all the points mentioned in subchapter 4.1 but with the incorporation of the improvements described in subchapter 4.2.

Starting with the first cycle – Understand Data System Structure -, from the analyses made, it was possible to conclude the following aspects:

- Some connected tables could be replaced for others stored on the Data Mart
- The query folding suffered updates, allowing the reduction of the number of duplicate tables

That said, a new dependency map was developed as presented in Graphic 7 From it, it is possible to see that the number of tables on the *Power Query* ribbon has dramatically decreased, going from 28 (27 connected tables plus the calendar one which was a Raw table) to 17. No doubt understanding the dependencies got a lot easier. The circle – Data Flow - represent a new source, even though it still comes from the DW. The Data Mart is independent from the DW, representing a new completely new source. The ETL process behind the DW and the DM are also distinct.



Graphic 7 - Map of Dependencies: Final Report

Additionally, the nomenclatures "STG" and "DIM" on tables' names were ceased. The defence argument for such changes were, respectively, that the end-user will not see the staging tables (they are not loaded) and since the factual tables are already indicated with the "FCT, automatically it is noticeable that the remaining ones are dimensions. The organization method for the analyst is to allocate each table to the corresponding folder, as illustrated in Figure 18.

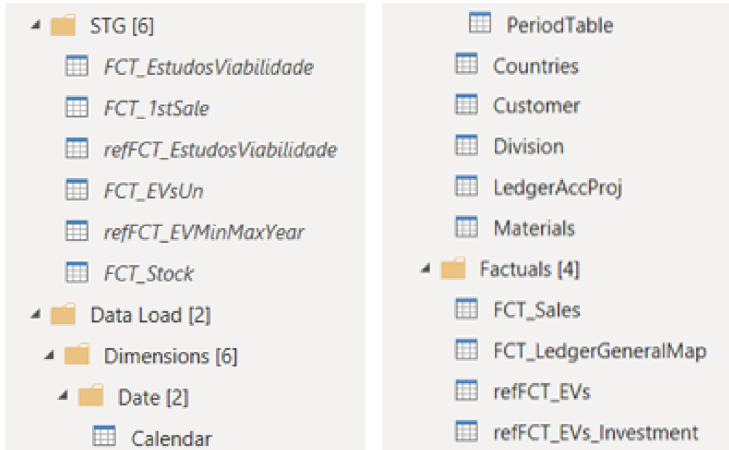


Figure 18 - Tables Organization: Power Query Ribbon

In the matter concerning the Data Preparation cycle, the majority of the complex steps were formalized, by leaving an explanatory comment on the code itself. The advantage of leaving a comment on the properties section of an applied step is that when hovering over the applied step, it automatically shows up as a kind of tooltip as displayed in Figure 19.

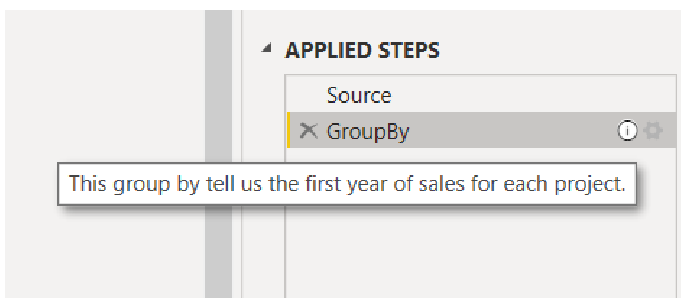


Figure 19 - Explanatory Comment: Example

The data stored in each table was also detailed and analysed. Having applied these changes, the *View Metrics* DAX Studio analysis was redone resulting in the information and metrics presented in Figure 20.

Table	STG/DL	Cardinality	Table Size	Columns Total Size	Data Size	Dictionary Size	Columns Hierarchies Size	Database Size %	Columns #
Calendar	DL	21 184	2 730 835	2 730 835	508 544	1 973 579	248 712	3,73%	37
Countries	DL	249	76 880	76 880	768	70 016	6 096	0,11%	3
Customer	DL	5 405	710 835	710 835	44 632	523 699	142 504	0,97%	6
Division	DL	5	36 030	36 030	24	35 766	240	0,05%	3
FCT_1stSale	STG	366	2 140 864	2 140 864	1 384	2 134 976	4 504	2,92%	3
FCT_EstudosViabilidade	STG	3 910	33 229 504	33 229 504	114 848	32 772 368	342 288	45,39%	32
FCT_EVsUn	STG	443	2 143 008	2 143 008	1 256	2 137 288	4 464	2,93%	4
FCT_LedgerGeneralMap	DL	57 293	852 967	852 399	479 544	272 991	99 864	1,17%	10
FCT_Sales	DL	181 325	12 529 170	12 521 106	6 304 440	5 091 786	1 124 880	17,12%	26
FCT_Stock	STG	77 439	551 536	551 536	217 472	20 560	313 504	0,75%	2
LedgerAccProj	DL	578	167 945	167 945	3 104	146 425	18 416	0,23%	4
Materials	DL	2 605	520 714	468 234	35 952	366 274	66 008	0,71%	16
PeriodTable	DL	27 428	325 870	319 158	65 424	225 270	28 464	0,45%	10
refFCT_EstudosViabilidade	STG	3 910	10 095 424	10 095 424	254 048	9 649 672	191 704	13,79%	31
refFCT_EVMinMaxYear	STG	445	3 213 248	3 213 248	1 264	3 206 608	5 376	4,39%	5
refFCT_EVs	DL	41 664	3 569 407	3 569 143	1 110 784	2 148 887	309 472	4,88%	25
refFCT_EVs_Investment	DL	5 328	199 179	198 923	35 352	149 827	13 744	0,27%	8
Grand Total		**429 650**	**73 199 046**	**73 130 702**	**9 178 968**	**61 030 382**	**2 921 352**	**100,00%**	**241**

Figure 20 - DAX Studio - View Metrics Analysis: Final Report

The first thing to have in mind while comparing the two analyses (Figures Figure 8 and Figure 20) is that they are more than six months apart in terms of data. The base report was lastly updated at the end of August since it started having updating problems due to excessive queries' weight. Therefore, when improvements occur, they are of even greater value.

As stated in the previous chapter, the *Item Cost* table was removed, which by itself already helped a lot in the report's processing. The tables that now stand out are the main two factual – *Estudos de Viabilidade* and *Sales*. Particularly on the *Estudos de Viabilidade* tables, it is possible to see that when we go down on the dependencies map, the tables' weight decreases, meaning that they are becoming more specific for their purpose. Regarding the percentage occupied in the database, it makes no sense to compare, since the sum must always be 100%, so when the tables that had the highest percentage were removed, others gained weight. In terms of the number of columns, in total, was reduced by more than half. The only exception is the *Calendar* table. As it was constructed a *Calendar* table in the DW, more information was added, namely denomination in both Portuguese and English language, and some offsets. However, it still represents a very small percentage of the database so it does not translate into processing problems.

The simplification at the tables' level becomes a requirement for a consequent simplification of the schema. All the relations between dimensions were eliminated, as best practices suggest, being the bidirectional relationship between the *Calendar* and the *PeriodTable* the only one that remained for the same reason it previously existed. Figure 21 shows the actual schema, where can be easily identified four Stars Schemas, each with a centre on a factual table – *FCT_Sales*, *FCT_LedgerGeneralMap*, *refFCT_EVs_Investment* and *refFCT_EVs*.

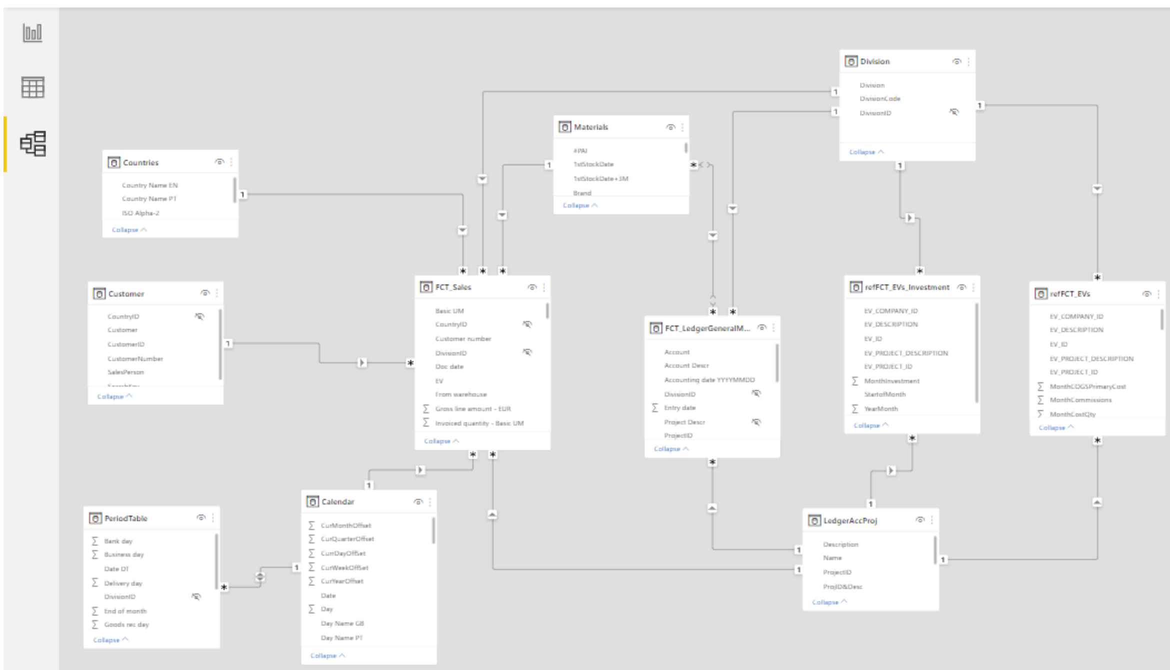


Figure 21 - Schema: Final Report

Reaching the third cycle - Design Report -, the analysis performed on DAX Studio, was also useful. From it also resulted a total of 192 measures. Remembering that the base report contained 424, this represents a reduction of approximately 54.2% in the number of measures created. Beyond those eliminated by the Calculation Group's potential, some of the ones created on the base report were not used in any visual being, therefore, worthless to the report's purpose. For this reason, also these measures were not replicated in the final report.

Last but not least, in what concerns the layout of the final report the main improvement, as explained in the previous subchapter, was the design of a slicers panel. The visual result of all the steps explained before is registered in Figures 22 and 23. As can be seen, it was also placed on an area with no great value, but the panel is large enough to be noticed by the end-user.

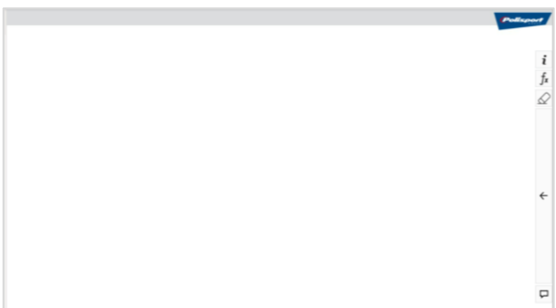


Figure 22 - Slicer Panel Off

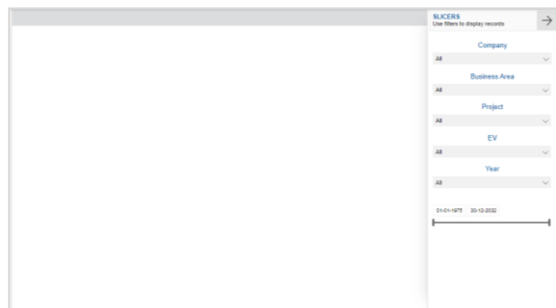


Figure 23 - Slicer Panel On

Fifth Chapter

5. Conclusion

This dissertation focuses on improving a report based on upgrades in the BI tool, Power BI. The report restructuring in question belongs to the company Polisport Plásticos, SA in which the internship that nourish this research took place. The first two subchapters identify the main conclusions and contributions of this work, respectively, and the last one presents some suggestions for future work related to the report. For that, various analyses were made to the base report to identify potential aspects that could be improved.

5.1. Final Conclusions

In what concerns the specific goals previously defined all of them were achieved and, consequently, the main objective of improving the efficiency of the report.

Furthermore, after the development of everything described in the previous chapter, several conclusions were drawn regarding both the construction of a report and the Power BI tool. They were as follows:

A good DW structure is essential to guarantee the quality and optimization of data when connected through BI tools so that they reach these platforms as appropriate as possible.

The vast knowledge of external tools, in this case, Tabular Editor and DAX Studio, associated with Power BI, when properly used, are essential to ensure good information processing.

There is no clear separation between both the data preparation and the report's layout design since there are cycles that are very interconnected with each other.

Ignoring this fact jeopardizes the efficiency of the report, by not removing unnecessary information from tables or by the exaggerated creation of measures.

5.2. Major Contribution

The major contributions of the practical work carried out during the internship and here explained can be divided into two points of view: analyst and end-user.

Table 8 discriminates the ones related to each point of view.

ANALYST	END-USER
Steps formalization	Placement of visuals according to the report's real estate
Automation of some manual fixes	Creation of "extra space" with the bookmark concept

Table 8 - Main Contributions: Analysts and End Users

Regarding the main contribution of this work for analysts, the ones that stand out are the formalization of the steps and the automation of some manual fixes applied in the Data Preparation cycle. The first one has to do with the fact that leaving comments in the code, related to the code itself or the business, makes it easier for anyone to understand what was done and why. This is crucial because the person who builds the report is not always the same person who monitors it, and even this can vary between people in the department depending on the needs. The second one has to do with the fact that in case of strategic or business-related changes, the report has to be updated manually, which requires a lot of communication between departments and much of the analysts' time. By making these updates automatic, analysts have more time for issues that really require their full attention.

As for the end-users, they benefit from the fact that their eyes immediately capture the most relevant information on each page, as the data importance/ dashboard real estate relationship has been respected. Additionally, the slicers pane allows a quick meeting of where these are organized, since it is common to all the pages, leaving more space for useful information.

Beyond these contributions, key technical changes – Calculation Groups and Slicer Panel - can be incorporated into other reports as they are not business or theme-oriented.

5.3. Recommendations for future work

In what concerns future work for this report restructuring at Polisport Plásticos, the following tasks are suggested:

- Run the *Diagnose Analysis*, as it tells the time that each step needs to be executed
- Run the *Performance Analyzer* for each page, as it tells the time each visual needs to be processed

Both these analyses, allow analysts to gain better knowledge about processing times regarding steps and visuals, which is fundamental when they are presented with more than one way of reaching the same outcome. For instance, this is useful when choosing the visual to display certain information since people tend to vary them thinking that it enriches the report, but as some are processed faster, they might indeed be decreasing reports' performance.

Bibliographic References

Breslin, M. (2004). *Data warehousing battle of the giants*. Business intelligence journal. 7, 6-20.

Brown, A., & Dowling, P. (2001). *Doing research/reading research: A mode of interrogation for teaching*. London: Routledge Falmer.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). *Business Intelligence and Analytics: From Big Data to Big Impact*. MIS - Quarterly, 36(4), 1165–1188.

Davenport, T. H. (2006, January). *Competing on Analytics*. Harvard Business Review.

Davenport, T. H., Guszcza, J., Smith, T., & Stiller, B. (2019). *Analytics and AI-driven enterprises thrive in the Age of With - The culture catalyst*. Deloitte Insights. 1–16. https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten*. Analytics Press. 2nd Edition. www.analyticspress.com

Few, S., & Edge, P. (2010). *Dashboard Design for at-a-glance monitoring*. Available in www.perceptualedge.com

Gartner Glossary. (2022). https://www.gartner.com/en/information-technology/glossary

Gavin, M. (2021). *Business Analytics: What it is & Why it's important*. Harvard Business School Online's Business Insights. https://doi.org/10.1021/cen-v078n032.p049

IBM. (2021). https://www.ibm.com/cloud/learn/data-visualization

Kemmis, S (1989). *Investigácion en la Accion*. Enciclopedia Internacional de la Educacion, Vol 6, pp. 3330-3337. Barcelona: Vicens-Vives/MEC

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (Third). John Wiley & Sons, Inc.

Lewin, K. (1946) *Action research and minority problems*. G.W. Lewin (Ed.) Resolving Social Conflicts. New York: Harper & Row (1948).

Luhn, H. P. (2010). *A Business Intelligence System*. IBM Journal of Research and Development, 2(4), 314–319. https://doi.org/10.1147/rd.24.0314

Manyika, J., Brown, M. C., Dobbs, R., Roxburgh, C., & Hung, A. B. (2011). *Big data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. 156. https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

Microsoft. (2022). Power BI. https://powerbi.microsoft.com/en-au/

Polisport. (2022). https://www.polisport.com/pt/

Russo, M., & Ferrari, A. (2008). *Introduction to the SQLBI Methodology*. p.21. www.sqlibi.com

Russo, M., & Ferrari, A. (2020). *Definitive Guide to DAX: Business intelligence with Microsoft Power BI, SQL Server Analysis Services, and Excel*. (Second, Vol. 560). Pearson Education, Inc.

SAP. (2021). *10 Golden Rules for Business Dashboard Design*. SAP Press. https://blog.sap-press.com/10-golden-rules-for-business-dashboard-design

Skiena, S. (2017). *Data science design*. Springer. Available in http://link.springer.com/10.1007/978-3-319-55444-0%0Ahttp://www.springer.com/series/3191%0Ahttp://link.springer.com/10.1007/978-3-319-55444-0%0Ahttp://link.springer.com/10.1007/978-3-319-55444-0%0Ahttp://www.springer.com/series/3191

Tableau. (2016). *Dashboard Design: A study on eye tracking*. https://www.tableau.com/resource/eye-tracking-study

Attachments

Attachment A

COLUMN	DESCRIPTION
Cardinality	The total number of distinct values in a column
Table Size	The total size of a table, including all columns, relationships and hierarchies
Col Size	The total size of the column = Data + Dictionary + Hierarchies
Data	The size of the data for the column
Dictionary	The size of the dictionary
Hier Size	The size of hierarchy structures
% DB	The space taken up as a percentage of the total size of the database
Columns #	The number of columns in the table

Attachment A - Meanings DAX Studio View Metrics Analysis

Source: DAX Studio

Appendix

Appendix A

MEASURE	DESCRIPTION
Sales	= SUM ('refFACT_Sales'[Sales: Net line amount - EUR])
Sales PY	= CALCULATE ('Key Measures'[Sales], SAMEPERIODLASTYEAR ('udfDIM_Calendar'[Date]))
Sales TTM	= CALCULATE ('Key Measures'[Sales], DATESBETWEEN ('udfDIM_Calendar'[Date], NEXTDAY (SAMEPERIODLASTYEAR (LASTDATE ('udfDIM_Calendar'[Date]))), LASTDATE ('udfDIM_Calendar'[Date])))
Sales RT	= CALCULATE ('Key Measures'[Sales], FILTER (ALL ('udfDIM_Calendar'[Date]), 'udfDIM_Calendar'[Date] <= MAX ('udfDIM_Calendar'[Date])))

Appendix A - Sales Measures Overlap

Appendix B

CALCULATION	FORMULA
Pipeline	Blank ()
Target	SWITCH (TRUE (), SELECTEDMEASURENAME () = "Investment(A)", [Investment(T)], SELECTEDMEASURENAME () = "Sales(A)", [Sales(T)], SELECTEDMEASURENAME () = "Qty(A)", [Qty(T)], SELECTEDMEASURENAME () = "GrossMargin(A)", [GrossMargin(T)], SELECTEDMEASURENAME () = "GrossMargin %(A)", [GrossMargin %(T)], SELECTEDMEASURENAME () = "NetMargin(A)", [NetMargin(T)], SELECTEDMEASURENAME () = "NetMargin %(A)", [NetMargin %(T)], SELECTEDMEASURENAME () = "COGSPrmCst(A)", [COGSPrmCst(T)], SELECTEDMEASURENAME () = "COGSPrdCst(A)", [COGSPrdCst(T)], SELECTEDMEASURENAME () = "SalePrice(A)", [SalePrice(T)], SELECTEDMEASURENAME () = "SalePrice(A)", [SalePrice(T)], SELECTEDMEASURENAME () = "Gross/NetMargin(A)", [Gross/NetMargin(T)], SELECTEDMEASURENAME () = "Gross/NetMargin%(A)", [Gross/NetMargin%(T)], SELECTEDMEASURENAME () = "NMROI(A)", [NMROI(T)], 0)
Actual	SELECTEDMEASURE ()
RT	CALCULATE (SELECTEDMEASURE (), FILTER (ALL ('Calendar'[Date]), 'Calendar'[Date] <= MAX ('Calendar'[Date])))

RT Average	var Running_Total = CALCULATE (SELECTEDMEASURE (), FILTER (ALL ('Calendar'[Date]), 'Calendar'[Date] <= MAX ('Calendar'[Date]))) RETURN DIVIDE (Running_Total, 'Calendar'[MonthsInYear])
TTM	CALCULATE (SELECTEDMEASURE (), DATESBETWEEN ('Calendar'[Date], NEXTDAY (SAMEPERIODLASTYEAR (LASTDATE ('Calendar'[Date]))), LASTDATE ('Calendar'[Date])))
TTM Average	var _TTM = CALCULATE (SELECTEDMEASURE (), DATESBETWEEN ('Calendar'[Date], NEXTDAY (SAMEPERIODLASTYEAR (LASTDATE ('Calendar'[Date]))), LASTDATE ('Calendar'[Date]))) RETURN DIVIDE (TTM, 'Calendar'[MonthsInYear])
YTD	TOTALYTD (SELECTEDMEASURE (), 'Calendar'[Date])
Trend	Blank ()
YTD Average	VAR _YTD = TOTALYTD (SELECTEDMEASURE (), 'Calendar'[Date]) RETURN DIVIDE (YTD, 'Calendar'[MonthsInYear])
LastMonth	CALCULATE (SELECTEDMEASURE (), Dateadd('Calendar'[Date], -1, month))

Δ Abs(A/T)	VAR _Target = SWITCH (TRUE (), SELECTEDMEASURENAME () = "Investment(A)", [Investment(T)], SELECTEDMEASURENAME () = "Sales(A)", [Sales(T)], SELECTEDMEASURENAME () = "Qty(A)", [Qty(T)], SELECTEDMEASURENAME () = "GrossMargin(A)", [GrossMargin(T)], SELECTEDMEASURENAME () = "GrossMargin %(A)", [GrossMargin %(T)], SELECTEDMEASURENAME () = "NetMargin(A)", [NetMargin(T)], SELECTEDMEASURENAME () = "NetMargin %(A)", [NetMargin %(T)], SELECTEDMEASURENAME () = "COGSPrmCst(A)", [COGSPrmCst(T)], SELECTEDMEASURENAME () = "COGSPrdCst(A)", [COGSPrdCst(T)], SELECTEDMEASURENAME () = "SalePrice(A)", [SalePrice(T)], SELECTEDMEASURENAME () = "SalePrice(A)", [SalePrice(T)], SELECTEDMEASURENAME () = "Gross/NetMargin(A)", [Gross/NetMargin(T)], SELECTEDMEASURENAME () = "Gross/NetMargin%(A)", [Gross/NetMargin%(T)], SELECTEDMEASURENAME () = "NMROI(A)", [NMROI(T)], 0) RETURN SELECTEDMEASURE () - '_Target'
% of Target	VAR _Target = SWITCH (TRUE (), SELECTEDMEASURENAME () = "Sales(A)", [Sales(T)], SELECTEDMEASURENAME () = "Qty(A)", [Qty(T)], SELECTEDMEASURENAME () = "NetMargin(A)", [NetMargin(T)], 0) RETURN DIVIDE (SELECTEDMEASURE(),'_Target')
PY	CALCULATE (SELECTEDMEASURE (), SAMEPERIODLASTYEAR ('Calendar'[Date]))
Δ Abs(A/PY)	VAR _PY = CALCULATE (SELECTEDMEASURE (), SAMEPERIODLASTYEAR ('Calendar'[Date])) RETURN SELECTEDMEASURE () - '_PY'

Appendix B - Calculation Items Formulas