

Performance Analysis of Product Portfolio and Customer Segmentation in the Pharmaceutical Sector

Duarte João Coutinho Bastos de Oliveira

Dissertação de Mestrado

Supervisor at FEUP: Dr^a Ana Maria Cunha Ribeiro dos Santos Ponces Camanho



Mestrado Integrado em Engenharia e Gestão Industrial

2021-06-28

“There’s a world out there to be created and I want to be part of it”

João Nuno Oliveira

Resumo

Num ambiente competitivo como o setor Farmacéutico, onde as equipes de vendas e negociação vivem para o melhor posicionamento de produto, detalhes de contrato e preço, é crucial não só que se extraia o máximo valor dos dados já existentes, como também tornar isso em informação valiosa para uma descoberta do comportamento do consumidor e definir as melhores formas de atuação.

Por um lado, as empresas devem ter uma estratégia bem definida para visualizar o seu desempenho de vendas, como também que cenários estratégicos definir e metas para atingir. Sem isto, podem futuramente perder velocidade de decisão e arriscar perder oportunidade de melhoria.

Pelo outro lado, as empresas precisam de tratar dos dados dos consumidores de forma a perceberem os seus padrões de consumo e desenharem táticas comerciais personalizadas para cada segmento a explorar.

Esta dissertação propõe duas novas ferramentas analíticas para uma empresa no setor farmacéutico: um novo dashboard e uma nova segmentação das Farmácias. O dashboard será capaz de analisar o desempenho dos produtos e das farmácias nos processos de vendas às Farmácias e vendas aos clientes finais, através das capacidades de drill-down, filtragens entre elementos gráficos e uma vasta experiência de objetivos orientada ao utilizador. A Segmentação de Clientes (Farmácias), que enriquece o dashboard, foi desenvolvida recorrendo ao conceito de Customer Lifetime Value das Farmácias, para que a empresa fosse capaz de elaborar ações de marketing e vendas mais personalizadas.

Finalmente, sugestões sobre o futuro de Business Intelligence e Segmentação de Clientes são referidas, para que a empresa possa avançar no futuro com estes tópicos.

Palavras-chave: Performance de Vendas, Objetivos, KPI, Experiência do Utilizador, Segmentação, K-Means, Decisão por Multi-Critério

Abstract

In competitive environment such as the Pharmaceutical sector, where sales and negotiations teams live for the best product placement, deal and price, it is crucial to both be able to extract the most from the available data and as well turn it into insightful customer discoveries to understand customer behavior and take action upon it.

On the one hand, companies must have a well-defined strategy on how to not only visualize their sales performance, but as well what strategic scenarios to look for and target to improve. Without doing so, they might be losing speed of decision and might risk losing sight of quick-wins.

On the other hand, companies need to be able to work around customers data to further understand their purchase behavior and design personalized commercial tactics for each of the segments explored.

This study proposes two new analytical tools for the company: a new dashboard capable of analyzing the performance of products and Pharmacies in the Sell-In and Sell-Out Processes, by delivering various drill-down, cross-filtering capabilities to the users and a broad target-oriented user experience. To enrich the business intelligence tool, a customer segmentation was developed, based on the Customer Lifetime Value of the Pharmacies, in order for the company to have more marketing and sales personalized approaches.

Finally, suggestions on the future of BI and Customer Segmentation are delivered, in order to further advance the topics in the company.

Keywords: Sales Performance, Dashboard, Targets, KPI, User Experience, Clustering, K-Means, Customer Segmentation, Multicriteria Decision-Making

Acknowledgements

I would like to start off by deeply thanking my colleagues Eduardo, Alberto and Maria, not only for their tactical and strategic advice on the work developed, but as well as for their joyful presence in every moment of the journey. May we find each other more times in the future. To Helena, Patrícia, Ana and Jorge for their unconditional support on the project and for my internal development both inside the company and in the sector as well.

Secondly, to my supervisor, Professor Ana Camanho, not only for having guided me throughout this project, but also for the tremendous support during my development as a student, professional and active citizen across my academic life.

On a more personal note, would like to thank my closest friends Pedro, Gustavo and João for their motivation and personal interest on my development as a person and as a professional during my academic life.

Lastly, to my girlfriend, Teresa, for always pushing and believing in me, even in the darkest moments. To my parents, whose love and wisdom allowed me to flourish and improve every day that goes by. To you, my deepest thanks.

Duarte João Oliveira

Contents

1	Introduction.....	1
1.1	The Company and the Retail Pharmaceutical Market	2
1.2	Project Challenge, Stakeholders and Main Scope	3
1.3	Project Methodology	5
1.4	Dissertation Structure.....	6
2	Literature Review	8
2.1	Product Portfolio Performance Analysis	8
2.2	Clustering and Customer Segmentation Analysis	10
3	Problem Description and Business Scenarios	20
3.1	Sell-In vs Sell-Out	20
3.2	The categories to take into account	22
3.3	The AS-IS Difficulties and Processes.....	23
3.4	Main Strategic Aspects currently analyzed by Key Accounts of XYZ for the Product Performance Analysis	23
3.5	Main Strategic Aspects currently Analyzed by Key Accounts of XYZ for the Pharmacy Segmentation	24
3.6	To-Be Analysis	25
4	Dashboard for the Product Performance Analysis of Supplier XYZ	28
4.1	Dashboard philosophy and main usage requirements	28
4.2	Performance Metrics Areas and Definition	29
4.3	Data Treatment and Tables Structure	32
4.4	Dashboard Usage Journey.....	33
4.5	Main UX/UI Elements	34
4.6	Tool Used and its limitations	36
4.7	Dashboard: the report pages and its interpretation	37
4.8	Main Conclusions.....	44
5	Pharmacy Segmentation – How do our clients perform and under which patterns for Supplier XYZ	46
5.1	Approach to be used throughout the Analytical Discovery	46
6	Conclusions and Future Work	59
7	References	61
8	Appendix A.....	64

Acronyms and Symbols

CLV	Customer Lifetime Value
DW	Data Warehouse
SI	Sell-In
SO	Sell-Out
POS	Point of Sale
AHP	Analytical Hierarchy Process
UML	Unified Modeling Language
KPI	Key Performance Indicator
KDD	Knowledge Discovery in Databases
CR	Consistency Ratio
ETL	Extraction, Transformation and Loading
BI	Business Intelligence
MNRSM	Medicamentos Não Sujeitos a Receita Médica
PSBE	Produtos de Saúde e Bem Estar

List of Figures

Figure 1.1: Project Methodology	6
Figure 3.1: Typical Supply Chain Concepts.....	20
Figure 3.2: Adapted Supply Chain steps for the Company Context	21
Figure 3.3: Decision-Tree made possible after the exploration of the Report Pages of the Dashboard.....	26
Figure 3.4: Thought Process after the Segmentation Process	27
Figure 4.1: Adapted UML for the SI Process from the Data Warehouse.....	33
Figure 4.2: User Journey when navigating through the Dashboard Report Pages	33
Figure 4.3: Slicers found across the Report Pages of the Dashboard.....	34
Figure 4.4: Examples of KPI Elements used across the Report pages in the Dashboard.....	35
Figure 4.5: Examples of Navigation Buttons used across the Report Pages of the Dashboard.....	35
Figure 4.6: Main User Discovery Elements in the Report Pages of the Dashboard	36
Figure 4.7: Executive View, where the user can see the most important KPIs.....	38
Figure 4.8: Menu Page for Dashboard Navigation.....	39
Figure 4.9: Global Analysis of the Business of SO	40
Figure 4.10: Global Analysis of the Pharmacy in SO	41
Figure 4.11: Market Penetration on SI	42
Figure 4.12: Internal Market Share of MNSRM and PSBE	43
Figure 4.13: SI and SO Alignment of the Whole Business and Pharmacy Specific	44
Figure 5.1: Boxplots of the 3 RFM Variables	50
Figure 5.2: Elbow Curve	51
Figure 5.3: Silhouette and Gap Statistic Method Plotted	51
Figure 5.4: 3D Representation of the Clusters and their respective Pharmacies.....	53
Figure 5.5: Decision-Tree for the RFM Analysis.....	54
Figure 5.6: Relationship between CLV and % of Purchases of Supplier of XYZ of each Cluster.....	55
Figure 5.7: Scatterplot of the Pharmacies belonging to C4-Dark Blue, by % of XYZ (X axis) and CLV (Y axis)	56
Figure 8.1: Top 20 Products in SI - Revenue and Quantity and the respective average	64
Figure 8.2: Scatterplots of Revenue and Growth of Revenue compared to last year by Pharmacy, Category, Sub-Category and Product	65
Figure 8.3: Top 20 Products in SI Revenue and their respective sum across time, both absolute and the percentage they represent in total revenue	65
Figure 8.4: Product Growth in Revenue of SI compared to Last Year and compared to average of the last 3 Months (both in Value and Quantity). Product and Category detail on the side for more exploration;.....	66

Figure 8.5: Count of Pharmacies with Purchase Levels of SI above a certain target, both for a specific month or across time	66
Figure 8.6: Top and Bottom 20 Pharmacies in terms of SI (Revenue and Quantity).....	67
Figure 8.7: Growth of a Category and Product comparing to the average of the last 3 Months	67
Figure 8.8: Stock Overview of the company, allowing to see which products should we pay attention to	68

List of Tables

Table 1.1: Company Internal Structure	2
Table 1.2: Market Share of the Company in important Segments	3
Table 2.1: AHP questionnaire for the pairwise comparison.....	18
Table 2.2: AHP scale, adapted from Shih and Liu (2003)	18
Table 4.1: Focus Areas and correspondent Metrics Developed during the creation of the Dashboard.....	29
Table 5.1: Sample of the Original Transaction Dataset.....	47
Table 5.2: Sample of the dataset to be used for the clustering step.....	48
Table 5.3: Statistical Summary of the RFM Variables.....	48
Table 5.4: Normalized Data for the Clustering Algorithms	49
Table 5.5: Matrix of Correlation between RFM Variables	50
Table 5.6: Sum of Square for each Cluster.....	52
Table 5.7: Normalized Results for each Cluster.....	52
Table 5.8: Average Values of Each Cluster	52
Table 5.9: CLV of each Cluster and % of XYZ in the Purchase Value	55
Table 5.10: Overview table of the Monetary focus XYZ Key Accounts and Operations Department could have	57

1 Introduction

When diving deep into the pharmaceutical sector, both worldwide and in Portugal, it's easy to grasp how competitive and fast-paced it is, requiring its players to be on the cutting-edge of process efficiency, costs implied in the overall business and, most importantly, every player needs to know how to effectively negotiate and make the best of each deal, since "price wars" and managing the clients sales trends are a must for everyone to be able to win, both in the short and long term. With this in mind, this is a sector that, among many others nowadays, and even more with the big amount of data it generates on a daily basis, needs both a structure and rationale behind the collection, treatment and management of this data, transforming it into actionable thoughts and decisions. Thus, it is easy to conclude that much of the future of the sector will involve a great amount of analytics and all the benefits it brings to businesses.

This project hereby written represents the final work/dissertation for the Masters Degree in Industrial Engineering and Management, in the Faculty of Engineering of University of Porto. This dissertation was done with the cooperation a Portuguese National group of pharmacies that encompasses many services, ranging from online retail of medicines, on-site diagnosis of recurrent diseases, personalized supervision of patients in their pharmacies across time and even the development of their own products. In its main vision, the company desires to make pharmacies a broader place for treatment and wealth for the patients, taking away much of the burden from the other health institutions.

In between many strategical points, the company realized the potential that the two following points would bring to the business: an overall day-to-day business intelligence solution allowing them to monitor and help on the decision-making process of every part of their product portfolio and an intelligent segmentation of their main POS (Points of Sale), which are the pharmacies, through the usage of clustering techniques. The first aspect came as a realization since the amount of data was already advanced to be managed only solely manually and without a proper strategy and the company was losing sight of many possible scenarios that could be hindering their price and product development strategy, making them vulnerable to the competition being better in these two points. As for the latter, the intelligent segmentation, it came as a natural conclusion that when comparing pharmacies in between them for analyzing the selling performance of the company towards their pharmacies, it was becoming increasingly difficult to make this a fruitful business exercise, since pharmacies have different profiles that aren't easily matched solely on their geographical positioning, thus, the whole grouping of them should be made on a more holistic perspective using clustering techniques.

Summing up, the project has the following motivation to happen: to deliver a turnkey business intelligence tool that can give a full picture of the current state of the product portfolio of the company (pricing, performance and stock usage) and an intelligent segmentation of the pharmacies, hoping to enable better benchmarking strategies withing the group to increase the sales performance of the company towards their pharmacies. The whole project is focused on the supplier XYZ, a supplier that works with the company.

1.1 The Company and the Retail Pharmaceutical Market

The company is historically one of the biggest buying group in Portugal, starting off its operations in 2007 (a buying group is a type of company that does not own their stores, in this case Pharmacies, but rather Pharmacies join its network and the responsibility of the company is to negotiate better product prices with the Suppliers, so that Pharmacies can buy the products with a higher discount). In 2008, the company entered in contract with its first pharmacy and in 2009, it makes even a higher jump on the market by introducing its first own brand product. In 2011, the company reached a historical record, by getting to its pharmacy portfolio 150 pharmacies, being 80 of them “Full Concept” (pharmacies that besides being able to get access to the companies’ products, they also commercialize the group services, such as diabetics and podiatry appointments). Following a great run on acquisitions, the company continued its quest to enter in contract more pharmacies to its name, having by that time more than 500 pharmacies, 130 of them full concept. Finally, today, the company encompasses around 250+ pharmacies in its portfolio. It is important to address that the company is part of a bigger multinational umbrella group, that helps on best practices and performance monitoring.

More than being a buying group, the company delivers more offers than only serving as an intermediary between the major Laboratories and the Pharmacies aiming to get the best price possible. The company is trying to create a first line of healthcare for every possible patient. Thus, in order to be able to provide services on so many aspects, in the Table 1.1, we can depict the major areas focus on to deliver better services both for the pharmacies and the final client as well:

Table 1.1: Company Internal Structure

Area of Focus	Mission of the area	Examples of Practices
Category Management	To make sure that Pharmacies know best how to place their products in the stores and its employees have the best guidelines and techniques on how to sell different categories of products.	<p>Establishing semestral objectives to achieve and monitoring them;</p> <p>Advising which categories and sub-categories to best cross-sell;</p> <p>Make sure the pharmacy is aware of its performance against the Top 10 Nationally;</p>
Operations	To be able to negotiate the best possible lower prices to its pharmacy network.	<p>Development of Promotional and Penetration product packs to include in the pharmacies in order to test a new product line;</p> <p>Continuous monitoring of the alignment between Sales and Purchases of the Pharmacies Network;</p> <p>Key Accounts Team responsible for</p>

		a wide portfolio of Suppliers
Online Services	To be able to reach a broader target group of customers and to provide quick assistance on common health questions from the patients.	Continuous Development of the Academy (online courses for the Pharmacies); Management of the Online Platform, where the patients can order medicines;
Projects and Services	Offering of specialized in-store services (appointments) and realization of many awareness campaigns	Nutrition appointments with specialized monitoring, across time, of the patient;
Pharmacies Consultancy	Overall monitoring of the Pharmacies performance across all the areas covered before	Development of monthly performance reports for the pharmacies;

On what the Pharmaceutical Market concerns, it is important to grasp some numbers to give context on the size and matter of the market. Thus, through research on multiple reports from [Research \(2020\)](#), it was found that in 2020, in Portugal, this Pharmaceutical Market made nearly 3.7 thousand million euros, where the research accounts for approximately 85% of the whole national pharmacy coverage (by coverage we mean that the research has information from 85% of the existing Pharmacies in Portugal on their sales performance). In this research, the company accounted for 5,36% of Market Share in that same year. If we look closely, the numbers for the different Market Shares in each of the relevant product segments can be found in Table 1.2:

Table 1.2: Market Share of the Company in important Segments

Segment Selected	Company's Market Share
Pharmaceutical	5,39%
Generic	5,34%
OTC	5,48%
Patient Care	5,29%
Personal Care	5,34%
Nutrition	5,28%
Supplements	5,08%

As seen above, the company has a steady Market Share across multiple segments, making it an established player on the market.

1.2 Project Challenge, Stakeholders and Main Scope

This whole project, both the business intelligence display and the pharmacies segmentation are tools developed for the Operations Department of the company. This department, among many duties, is mainly responsible for the overall strategy regarding the suppliers' portfolio, thus being the one that needs full overview of how well positioned their

products are and how well their negotiation process is being carried out, something that reflects directly on the selling process from the company to its pharmacy network.

Firstly, and going deeper now on each topic, the business intelligence tool will serve many internal customers, being thus important to list them and explain the direct usage they will give to the tool:

1. **Key Accounts Team:** for a long time the team had the need to fully visualize the overall performance of the suppliers, both across time and a head-to-head comparison between different categories and segments, being this on the side of sales and margins. The major goal is that the Key Accounts Team can understand which products, categories and sub-categories to invest the most, the ones to divest and where in time they struggled the most, being this supported with sales from pharmacies across the country to better explain such phenomenon's. For the context of this project, the team wanted to focus on a specific supplier, the supplier XYZ;
2. **Operations Head of Department:** the company as a buying group negotiates on a big scale, with big pharma Laboratories directly, being the main value proposition the great amount of pharmaceutical products purchased as a whole to be sold in its Pharmacies. However, across time, the Head of the Department has been noticing that on specific products, categories and segments, the prices and the packages are not competitive enough. This was noticed mainly due to the differences felt in selling to the Pharmacies and from the Pharmacies to the final consumer disparities that show up big improvement opportunities in some products to be better negotiated by the company. Thus, the business intelligence tool developed in this project aims as well to deliver, in a real-time perspective, to highlight the products of Supplier XYZ that could help the pharmacies that are not buying as much as expected from a specific segment and how does that relates to negotiation later on;
3. **Analytics and IT team:** finally, still on the business intelligence tool, this will be also be important for the Analytics and IT team. With the implementation of this real-time tool, this team will be able to supervise the quality of the data existent in the company and detect improvement points on both the databases that support the information and how the data is collected to support even better strategic decisions;

Thus, as we can depict, this BI tool will respond both to the daily tactical decisions of the company, but it will be a gateway for a more solid data strategy in the long run.

As for the second topic, the segmentation of the 160+ relevant pharmacies depending on their purchase profiles, it has as main goal a better grouping of these pharmacies, not only for benchmarking purposes, but as well as for better marketing decisions in the Operations Department. For quite some time, the company has been feeling that when addressing their pharmacies to make yearly reviews of their purchase performance, it has been doing it comparing many of the pharmacies with the Top 20 in this dimension or, even some times, Top 10 of pharmacies, leading these to the following problems:

1. Pharmacies with low purchase performance always feel demotivated since comparing them to the top performers make them feel incapable of ever reaching those results;

2. Many pharmacies, when receiving advice from the company Central teams related to real examples of the Top 20/10 performers on how they should improve their processes, feel completely disconnected from those advice, mainly due to the feeling of “that’s not my reality, it won’t work here, we have different cultures and priorities”, thus being difficult for the company to carry on change management initiatives and behavioral changes for that matter;

As for the company side, without this more delicate segmentation, the company is facing the following problem:

- When detecting poor performance in the purchase process for some products or segments and by following it up with the selling performance of the Pharmacies, the company is having difficulties explaining why that purchase behavior happens and if similar pharmacies are having the same purchasing behavior. Thus, it is difficult to draw a defined business strategy to improve many segments of products at the same time, thus being critical to better cluster pharmacies to detect better “customer” patterns;

So, all-in-all, the whole thesis project breaks-down into two main objectives: the creation of a ready-to-use Business Intelligence Dashboard that will allow better product management and negotiation decisions and the creation of clusters of pharmacies based on their purchase profiles, thus allowing the company to detect new patterns and therefore, design new strategies.

1.3 Project Methodology

Regarding the Methodology, the project/tools that gave the content to this dissertation went through many loops in order to be rightfully developed. Firstly, this project started off with an intensive introduction to the business structure and processes, where the newly recruited analytics team had several meetings with the Operations Department in order to develop basic data analysis and present them in two presentations. Secondly, discussions about the major business questions and challenges were conducted within the department to be able to know how to rightfully build the tools. Thirdly, on a more technical topic, requirements were listed in order to be able to develop both the dashboard and the segmentation of the pharmacies: KPIs to create and the built-in ones to use, filters necessary for the users, special report pages for special business questions and among other aspects. Forth, the dashboard creation, where various iteration rounds were conducted within the respective department in order to improve the user experience. Fifth, and with a more critical perspective driven by the dashboard analysis, the team further defined the business objectives with the Pharmacy Segmentation and clusters were created. Sixth, the final user feedback stage, where the tools (both BI and Segmentation) were published and ready to receive feedback as final products. Lastly, as a final step, recommendations on what to improve regarding the data collection of the company were given in order to improve the overall business strategy of the company, given in the format of an auxiliar Guideline. The methodology is illustrated in Figure 1.1:



Figure 1.1: Project Methodology

Going even further in the project methodology, as for checking the results and advancing stage by stage in the project, some basic requisites must be met. These requisites were the following:

1. Having beta-version of the dashboard fully functional from a user-experience perspective (bugs in some extra calculations where not crucial);
2. Well-documented all the functionalities of each report page of the dashboard;
3. Analytical and statistical summary of the Pharmacy Segmentation elaborated;
4. Usage Description and Strategy on the Pharmacy Segmentation created;

1.4 Dissertation Structure

Ending up the introduction, it is important to address and deliver a sneak peek about how this dissertation will evolve as the pages go by, thus presenting its structure. In this current chapter, an introduction to the main parties involved, main project goals and motivations and the main business problems to tackle were discussed and presented. Following, on the second chapter, a broad range of topics are presented as Literature Review in order to support the developments of the project, such as Dashboards and KPIs, Business Intelligence and Clustering Techniques. On the third chapter, we will look deeply into the details of the pharmaceutical business that are core to this project and the requirements asked for each tool and solution the team needs to be fulfilled. The fourth chapter will focus on how

the dashboard was developed and its results. On the fifth chapter, there will be the opportunity to uncover all the steps regarding the Pharmacies Segmentation and its implications. Finally, on the last chapter, the main conclusions will be presented, as well as main recommendations for future work.

2 Literature Review

This dissertation focuses on two major topics: Product Portfolio Performance Analysis and Customer Segmentation, and even though strategically they are connected regarding the overall direction and necessities of the companies for the Supplier XYZ, tactically and conceptually they are very different from one another, thus, being important two address both in different manners when it comes to the concepts supporting each other.

This chapter of the dissertation will deliver various insights on the main theories, concepts and rationales behind the main practical deliverables of the project: the Dashboard and the Pharmacy Segmentation. Firstly, the main concepts of Business Intelligence, Dashboards and KPIs will be discussed, where the reader will have the chance to understand the nuances of certain elements and stages of development of such tools. Secondly and lastly, the main techniques on clustering, data normalization, customer segmentation, and assessment of relative weights for the construction of the Pharmacies Clusters and respective importance will also be presented and discussed to enlighten the reader with the necessary concepts to further understand the decisions taken in Chapter 4.

2.1 Product Portfolio Performance Analysis

2.1.1 Dashboards and Visual Communication

According to [Pauwels et al. \(2009\)](#), dashboards are a powerful solution for organizations to look a deeper and more organized way to their information, delivering them their current state by centralizing the main analysis and KPIs into one place, and, consequently, allowing companies to deliver their first steps in measuring their improvements.

Through [Wind \(2005\)](#), dashboards, as a tool, deliver organizations concrete feedback on the outcomes achieved, allowing managers to work on the improvement of the drivers and processes. On a technical level, dashboards collect, summarize and showcase information from various sources such as BI and ERP softwares, delivering management teams a landscape of how KPIs are performing ([Yigitbasioglu and Velcu 2012](#)).

As for the main 4 factors that drive the urge to develop Dashboards, [LaPointe \(2005\)](#), has pointed out the following ones:

1. Lack of organized data in a relevant way;
2. Many biases found in the decision-making process for lack of treated information;
3. Great need of integration of performance reporting practices in a cross-departmental mindset;
4. Increasing demands of market accountability;

When it comes to the development of dashboards, [Pauwels et al. \(2009\)](#) delivers 5 key stages that need to be carefully taken into consideration, being them as follows:

1. *Selecting the Key Metrics*: here, there are two major ways of developing the Key Metrics – a general and a specific way. On to the general, it is relevant if the business needs to compare itself, on a high level, to other businesses, such as revenues and customer retention. The tailored approach is relevant for department specific necessities, requiring this a much higher level of communication between teams to understand each other's targets;
2. *Populating the Dashboard with Data*: the stage where multiple sources of data and levels of data need to come together to fill in the dashboard. One of the major difficulties is in the enormous amount of data with different periodicity;
3. *Establishing Relationships between the Dashboard Items*: this is the stage where the dashboard developer organizes the metrics and the information in order to deliver answers and understandings for the business;
4. *Forecasting and Scenarios*: the stage where the information displayed helps on the what-if situations allowing to predict possible increases or decreases in productivity;
5. *Connecting to Financial Consequences*: as the final stage, more on a managerial way, this stage is connection of the metrics displayed in the dashboard to the financial targets the company needs to fulfil;

When it comes to the more user experience perspective of a dashboard, again, through [Yigitbasioglu and Velcu \(2012\)](#), we can see that the design features of a dashboard can be divided into visual and functional features. Regarding the visual ones, these features refer to how data is visualized and elements used to do so, such as the usage of the screen pages and colors. When combined, all the elements help the user to improve cognition and interpretation. As for the functional features, here we refer to what the dashboard can provide as exploratory options to the user, such as flexibility, scenario analysis, drill-down capabilities and even automated alerts.

Going even deeper, it is relevant to address these two types of features even further, by exploring specific aspects of both. Again, through [Yigitbasioglu and Velcu \(2012\)](#), and starting off with the functional features, there is always the need to adapt the dashboards to the specific user and situation we are dealing with. These tools should be interactive and allow many drill-downs through a point and click interactivity. Because of this, it is recommended that the Dashboard is integrated into a data warehouse, so that users have full granularity on the data. Presentation flexibility allows business users to slice and filter information; Automated visual alerts identify areas that need attention and, finally, scenario analysis allow the teams to better plan. On to the visual features, through [Yigitbasioglu and Velcu \(2012\)](#), the main highlight is the space and colors usages. It is recommended that the data that is presented can fit and be arranged on a single screen if it belongs to the same topic. Furthermore, on the colors, they need to be applied carefully, as its abusive use can distract the business user and hinder its navigation experience.

Finally, coming back to [Wind \(2005\)](#) suggestions, dashboards should be simple and efficient tools that allow teams to understand their performance and to improve decision-making on the key metrics presented.

2.1.2 KPIs and Performance Measurement Concepts

According to [Bauer \(2004\)](#), even though all KPIs are metrics, not all metrics are KPIs, mostly due to the level of criticality the metrics may represent for the business. KPIs are metrics that can be quantified and that showcase the performance of an organization, always aligned with its goals and objectives. KPIs track the business status of the company and make sure that all individuals are going towards the same strategies and goals. Furthermore, they deliver a common focal point for enterprises to be collaborative, standardized and coordinated.

Trough [McNeeney \(2005\)](#), we can see that it is greatly important that metrics/KPIs are embedded into performance measurement systems so that individuals and groups understand how their activities fulfill the bigger corporate goals. Thus, according to [Weber and Thomas \(2005\)](#), each function in the organization must focus on the portion of the indicators that they influence, so that the performance can be improved and maintained. This is also something supported by [Neely \(1999\)](#), that underlines that the performance measures used in many organizations are usually not integrated with each other or aligned with a higher business process.

In order to have a more sequential perspective on how to build KPIs and the performance measurement systems, [Bourne et al. \(2000\)](#) presents us with its three main phases, as follows:

1. The creation of the Performance Measures and their interconnection with Key Objectives;
2. The implementation of the performance measures, that includes collection, collation sorting and distribution;
3. Finally, the usage of the performance measures to assess the implementation of the strategy and to challenge strategic assumptions;

2.2 Clustering and Customer Segmentation Analysis

Since the dissertation also comprises the usage of Data Science and Mining Techniques and segmentation of customers also using analytical methods, we will now uncover some of the key concepts used for that purpose, how to interpret them and the limitations associated with each of them.

2.2.1 Data Mining

Data Mining is the application of various data analysis steps to datasets in order to extract valuable information about dimensions of the dataset being studied, such as groups of dimensions, correlations between or even predict future outcomes.

According to [Goldschmidt and Passos \(2005\)](#), data mining is structured and contains three main phases: pre-processing, data task application and, finally, post-processing. Preprocessing relates to all the functions that are connected to data capturing, organization and preparation of the data. This is a critical phase, since it comprises actions such as selection, cleaning and preparation of data in order to be used further on. Data Task application refers to the effective search for useful knowledge within the already prepared dataset to be used for the task. Finally, post-processing concerns the treatment of the knowledge acquired in the previous data task application phase.

On a more extensive version of the data treatment and preparation, and according to [Fayyad et al. \(1996\)](#), we can name the process of having to work around data in order to

prepare it for further analysis ad KDD (Knowledge Discovery in Databases). According to the author, we can divide it into the following phases:

1. **“Learning the application domain:** includes relevant prior knowledge and the goals of the application”;
2. **“Creating a target dataset:** includes selecting a dataset or focusing on a subset of variables or data samples on which discovery is to be performed”;
3. **“Data cleaning and preprocessing:** includes basic operations, such as removing noise or outliers if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, and accounting for time sequence information and known changes, as well as deciding DBMS issues, such as data types, schema, and mapping of missing and unknown values”;
4. **“Data reduction and projection:** includes finding useful features to represent the data, depending on the goal of the task, and using dimensionality reduction or transformation methods to reduce the effective number of variables under consideration or to find invariant representations for the data”;
5. **“Choosing the function of data mining:** includes deciding the purpose of the model derived by the data mining algorithm (e.g., summarization, classification, regression, and clustering)”;
6. **“Choosing the data mining algorithm(s):** includes selecting method(s) to be used for searching for patterns in the data, such as deciding which models and parameters may be appropriate (e.g., models for categorical data are different from models on vectors over reals) and matching a particular data mining method with the overall criteria of the KDD process (e.g., the user may be more interested in understanding the model than in its predictive capabilities)”;
7. **“Data mining:** includes searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression, clustering, sequence modeling, dependency, and line analysis”;
8. **“Interpretation:** includes interpreting the discovered patterns and possibly returning to any of the previous steps, as well as possible visualization of the extracted patterns, removing redundant or irrelevant patterns, and translating the”;
9. **“Using discovered knowledge:** includes incorporating this knowledge into the performance system, taking actions based on the knowledge, or simply documenting it and reporting it to interested parties, as well as checking for and resolving potential conflicts with previously believed (or extracted) knowledge”;

If we now focus on the Task Application only part already mentioned, [Goldschmidt and Passos \(2005\)](#) go even further in what comprises the Task Application part. Amongst many tasks, the ones that follow are the ones more commonly used:

1. **Clustering:** it has the goal of creating new groups with the items being analyzed. Only after the groups are created, through its interpretation, the user can name them accordingly;
2. **Outlier Detection:** this task has the goal of identifying items whose characteristics are not inside the normal patterns of each of the dimensions of the dataset being used;
3. **Classification:** this task has the goal of classifying a new item based on previously learned categorical labels, through a supervised learning action. According to [Michie](#)

[Michie et al. \(1994\)](#), Neural Networks and Genetic Algorithms are examples of technologies able to make a classification task;

2.2.2 Clustering Theory

In order to further understand the clustering technique used in the context of this dissertation, it is important to address the rationale behind different ways of clustering a dataset. Thus, and according to [Tan et al. \(2013\)](#), we should be focusing on the following ones:

1. **Fuzzy versus Overlapping versus Exclusive:** with fuzzy clustering, each item in the dataset is attributed to all newly formed clusters with a certain degree of membership, ranging from 0% to 100%, meaning not belonging or fully belonging to the cluster, respectively, thus, being quite relevant if the user wants to decide after all to which cluster does the item belong; overlapping clustering, a point can be part of multiples clusters and, finally, in exclusive clustering, contrary to the previous two, the item can only belong to one cluster;
2. **Complete versus Partial:** with complete clustering, the user is forcing every item in the dataset to be allocated to one cluster, whereas in Partial, some items might not be allocated to a cluster;
3. **Hierarchical versus Partitional:** in partitional methods, each item is divided into a sole cluster that does not overlap with other ones. As for hierarchical methods, each object can belong to subsequent sub-clusters, nested in one another;

Assessing Clustering Tendency

Before the application of any sort of clustering techniques, it is important to analyze if the dataset has a non-random structure, in other words, if it is good enough to be used for a clustering task. According to [Tan et al. \(2013\)](#), if this step is not validated, we might be getting clusters without any type of meaning, completely random.

For this purpose, the Hopkins statistic is often used, which, according to [Lawson and Jurs \(1990\)](#), it is a statistics that assesses the randomness of the data. The expression underlying its calculation is shown in equation 2.1:

$$H = \frac{\sum_{i=1}^n dist(qi)}{\sum_{i=1}^n dist(pi) + \sum_{i=1}^n dist(qi)} \quad (2.1)$$

Here, we find the following elements:

1. D represents the dataset (not in the equation, but crucial for the following points);
2. $dist(pi)$ is the distance between n sample points pi from D and their nearest neighbours in D;
3. $dist(qi)$ is the distance between generated points qi uniformly distributed in the space of D and their nearest real neighbours in D;

In order to assess if the dataset has a cluster tendency, the value of H should tend to 0. The closer to 0, the better. With this, if $H > 0.5$, we reject the alternative hypothesis, which is that our dataset is tending to be uniformly distributed.

Clustering Techniques

In this sub-section, we will now explore the specific clustering techniques and the methods to evaluate the results coming out of those techniques.

From [Tan et al. \(2013\)](#), we can further understand the algorithm K-Means, which is a partitioning method that is able to separate the dataset in K number of clusters, specified by the user. Important to say that this method is un-nested, meaning, it only delivers one level of groups. Also from the same author, it is possible to point out two major drawbacks that show that K-Means is not the best for all possible projects: it is only suitable for quantitative data and it is affected by outliers. As a positive note, K-means is one of the most efficient methods in clustering large datasets ([Huang 1998](#)).

According to [Han et al. \(2011\)](#), we can find other partitioning methods for the same objective:

1. **K-Medoids:** in this method, each item is allocated to a cluster whose representative item is more similar to the item being evaluated. This method is better for the outlier sensitivity, as it uses a real item of the cluster;
2. **K-Median:** very similar to K-Means, the only difference is that instead of working with the mean of the cluster, we work with the median, thus reducing the outlier influence in the cluster creation;

To sum up, the main difference between K-means and the other two methods

Even though K-Median could have been a better choice scientifically, for the goal of this project K-Means was chosen due to its simplicity and easiness to communicate the results. Thus, by exploring even further this method, it is possible to describe the steps to apply it, including both the selection of the correct number of clusters (K) and the quality evaluation of its results.

Again, through [Tan et al. \(2013\)](#), we get the following algorithm that allows us to get the final cluster after applying it:

Algorithm 1 - K-Means Algorithm, inspired by [Tan et al. \(2013\)](#)

- 1 Select K Points as the initial centroids
 - 2 Repeat
 - 3 Form K clusters by attributing each item to the cluster whose centroid is closest to
 - 4 Re-Calculate the centroid of each cluster
 - 5 Until items do not change
-

Normalization of Variables

In order to be able to provide a useful clustering result, it is necessary to normalize the variables used for the task, as in most cases they are measured in different scales. To do it, it is possible to use a common method called min-max normalization. According to [Al Shalabi et al. \(2006\)](#), this technique performs a linear change to the original dataset. By analyzing the minimum and the maximum values for a certain attribute A, it maps a value v of A to v' in the range [new-min_a, new-max_a] by computing the equation 2.2:

$$v' = \frac{v - \min_a}{\max_a - \min_a} * (\text{newmax}_a - \text{newmin}_a) + \text{newmin}_a \quad (2.2)$$

In this equation, each element represents the following:

1. Newmax and Newmin represent the values chosen by the user for the minimum value and the maximum value to convert the attribute for. In this particular case, it is 1 for the maximum and 0 for the minimum;
2. Min_a and Max_a represent the minimum and maximum value on the original scale, respectively;
3. v represents the specific item in the attribute that will be converted to the new scale;

Deciding on the Number of K Clusters

Among several methods to find the most suitable K number of clusters, the first one we will use later in the project to establish our number will be the Elbow Method. According to [Nainggolan et al. \(2019\)](#), with the Elbow Method, the best number of clusters will be seen from the Sum of Square Error (SSE) value that has a significant and elbow-shaped decrease. In order to discover K, we need to plot the percentage of variance explained by the number of clusters and discover the point where the marginal gain is not significant enough.

From [Nainggolan et al. \(2019\)](#), we can as well base ourselves to provide the following algorithm 2 to discover the best K number of Clusters:

Algorithm 2 – Elbow Method adapted from [Nainggolan et al. \(2019\)](#)

- | | |
|---|---|
| 1 | Create the initial value of K |
| 2 | Repeat |
| 3 | Calculate the sum of square error for the current K Value and plot it |
| 4 | Until predefined values of K finish |
| 5 | Analyse the sum of square errors from all the K Values |
| 6 | Select the K that delivers the most drastic decrease in the plot |
-

In order to have a more robust opinion whether the choice of K was acceptable or not, it is relevant to analyze through other methods what could be the result of the number of K. For that, and in the context of this dissertation, two methods were used extra: the Average Silhouette Method and the Gap Statistic method.

Starting off with the Average Silhouette Method, it is important to address in simpler words what is the “Silhouette” measure: it is a measure of how similar an object is to its own cluster compared to other clusters. This value ranges from -1 to +1: the higher the value of a specific item in the dataset, the better it matches with the cluster it was allocated in. In terms of the algorithm, it computes the average silhouette of observations for various pre-defined values of K. The most suitable number of clusters K is the one that maximizes the average silhouette over a range of possible values for k ([Kaufman and Rousseeuw 2009](#)).

The algorithm to perform the Average Silhouette Method can be seen in algorithm 3:

Algorithm 3 – Average Silhouette Method adapted from [Kaufman and Rousseeuw \(2009\)](#)

- 1 Apply the clustering algorithm for each K number of Clusters pre-defined
 - 2 For each K number of Clusters, calculate the average silhouette of observations, $s(i)$
 - 3 Plot the $s(i)$ for all the number of K calculated in the previous steps
 - 4 Select the number of K Clusters that maximizes $s(i)$
-

Through [Kaufman and Rousseeuw \(2009\)](#), we can as well understand the equation 2.3 that lead us to the value of the average silhouette for each K number of Clusters used on its evaluation:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.3)$$

In the equation, we have:

1. i , that represents the number of clusters that are being plotted;
2. $s(i)$, that varies from -1 to 1. In order to assess if the data is well clustered, a value close to 1 is required and thus, $a(i) < b(i)$;
3. $a(i)$ shows the compactness of the cluster to which i belongs (measure of cohesion) – it gets more compact the smaller it gets;
4. $b(i)$ shows how much i is farther from other clusters (measure of separation) – it gets more separate the larger it gets;

As a second algorithm to cross-validate if the number of K Clusters was well selected, we have the Gap Statistic Method. According to [Tibshirani et al. \(2001\)](#), this method compares the total within intra-cluster variation for different values of K with their expected values under a null reference distribution of the data. The estimate of the optimal clusters will be the value that maximizes the gap statistic. This will mean that the structure of the clustering is far away from the random uniform distribution.

As an algorithm for the gap statistic, and also inspired by [Tibshirani et al. \(2001\)](#), we have algorithm 4:

Algorithm 4 – Gap Statistic Method adapted from [Tibshirani et al. \(2001\)](#)

- 1 Cluster the observed data, changing the number of clusters from $k = 1$ to $k = k_{\max}$ and compute the respective total within intra-cluster variation W_k
 - 2 Generate B reference datasets with a random uniform distribution. Then, cluster each of these reference datasets with varying number of clusters $k = 1$ to $k = k_{\max}$ and calculate the respective within intra-cluster variation W_{kb}
Calculate the estimated gap statistic as the deviation of the observed W_k value from its expected value W_{kb} under the null hypothesis:
 - 3
$$Gap(k) = \frac{1}{B} \sum_{b=1}^B \log(W_{kb}^*) - \log(W_k)$$

. Calculate also the standard deviation of the statistics;
 - 4 Select the number of clusters as the smallest value of k such that the gap statistic is within one standard deviation of the gap at $k+1$: $Gap(k) \geq Gap(k+1) - S_{k+1}$
-

Cluster Evaluation

Finally, after proceeding with the clustering techniques and deciding on the number of clusters, it is important to evaluate its results. The goal is to compare intra-cluster heterogeneity with inter-cluster homogeneity. As we were able to see across this chapter, Sum of Squared Error (SSE) is widely used and, in this case, it is used to analyze how compact each cluster is. According to [Han et al. \(2011\)](#), for each cluster C_i , the SSE between its centroid c_i and all the other items within that cluster, v (the items of the cluster being analysed), gives us an indication of its quality. In conclusion, in a dataset, the SSE for all its items is delivered by equation 2.4, where *dist* corresponds to the Euclidean Distance:

$$SSE = \sum_{i=1}^K \sum_{v \in C_i} dist^2(v, C_i) \quad (2.4)$$

2.2.3 Decision-Making Techniques and Customer Segmentation Models

Adding up to the clustering techniques, it is then important for all organizations to be able to use those structures from the algorithms and organize them in such way that is possible to deliver better decision-making processes. Thus, in this sub-section we will explore two main techniques used widely: multi-criteria decision-making techniques such as AHP and customer relationship management techniques, such as RFM Model to be able to characterize the behaviors of customers.

These two topics are highlighted here in the literature review as they will be fundamental to further understand the procedures and results of this dissertation.

Starting off with the customer behavior, through [Payne and Frow \(2005\)](#), we can see that CRM (Customer Relationship Management) plays a huge role in improving the relationship every organization has with its customers. In sum, CRM is the application of data mining techniques on the existing data from customers to better extract behavioral patterns. In the long run, this will improve customer acquisition, loyalty and profitability.

Furthermore, going even beyond the acknowledgement that CRM is an important aspect for every organization, it is important to quantify the value each customer has for us. For that, two important dimensions need to be discussed: what is the value of a customer and how do we calculate it. According to [Kotler and Armstrong \(2010\)](#), Customer Lifetime Value (CLV) is the key concept, defining it as “the value of an entire stream of purchases that the customer would make over a lifetime of patronage”. This shows us that the main objective is to measure the influence each customer has for the organization’s operations.

Thus, in order to quantify this CLV, among many possible analytical models, the RFM model is the basis for even more complex models we will address further here. According to [Hughes \(1996\)](#), RFM allows to detect different types of customers through the usage of three variables associated with their purchase history: Recency, Frequency, Monetary. They are as follows:

1. **Recency (R)** represents the length of time since the last purchase. On business environments, this variable increases its value the farthest the purchase was made;
2. **Frequency (F)** represents the number of purchases within a selected time period. On business environments, this variable increases its value if the customer makes many purchase orders;
3. **Monetary (M)** represents the amount of money spent in a selected time period. On business environments, this variable increases its value the more money the customer puts into the company through purchases;

Even though RFM variables are widely used, it is not the basic version that is the most useful to this project. Thus, a variation of it, WRFM (Weighted RFM) is the adapted version, where the business user can apply weights to each of the variables, so that it translates the organization strategy on what is most important to it. As an example of it, we have [Khajvand et al. \(2011\)](#), where this strategy was applied in a health and beauty organization. With this, the CLV of each cluster will now be able to be calculated as in equation 2.5:

$$CLV^i = C_R^i \times w_R + C_F^i \times w_F + C_M^i \times w_M \quad (2.5)$$

In this equation, the symbols represent the following:

1. C_R^i , C_F^i , C_M^i represent the normalized values obtained for the average values of each cluster i ;
2. w_R , w_F , w_M are the weights delivered for each variable, coming from the company’s priorities;

For the context of the dissertation, this will be the equation used later for Customer Segmentation of the Pharmacies.

Finally, to add up to this, it is necessary to have a proper methodology for assigning the correct weights to the equation formulated. For this, it is possible to use one of the most

popular techniques in this field: Analytical Hierarchy Process (AHP). Introduced by [Saaty \(2000\)](#), the author explains that AHP basis itself on the human capability to decide on the relative weight of various factors based on intuition, pairwise comparisons and experience.

In order to compute the relative weights, it is necessary to follow a methodology that will result in that precisely. Following what [Shih and Liu \(2003\)](#) propose, the steps to compute the weights of the variables are as follows:

1. The evaluators analyze and compare each pair of variables (Monetary with Frequency; Frequency with Recency; Monetary with Recency). To do this, we can use the scale proposed as well by [Shih and Liu \(2003\)](#) in Table 2.1:
2. After giving the relative importance to each of the pairs, the consistency ratio (CR) needs to be calculated to address the consistency of the evaluations made. Through [Saaty \(2000\)](#), if CR is below 10% then it is acceptable. If it surpasses, steps 1 and 2 must be repeated, as it is an indication that organization's assessment of the relationship between the variables is not coherent;
3. Finally, it is needed to calculate the relative weights of the variables (w_R , w_F , w_M) according to the pairwise comparison matrixes, through the eigenvector's theory of [Saaty \(1977\)](#).

Table 2.1: AHP questionnaire for the pairwise comparison

Criteria 1	Comparative Importance	Criteria 2
Monetary	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Frequency
Frequency	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Recency
Monetary	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	Recency

In the following table, we can see the AHP Scale adapted from [Shih and Liu \(2003\)](#), where it is possible to have a look at the scale presented to the managers of the organizations in Table 2.2 that allows them to evaluate the importance of a variable comparing to another one:

Table 2.2: AHP scale, adapted from [Shih and Liu \(2003\)](#)

Value Given	Description
1	Equally important
2	Between equal and weak
3	Weaker importance of one factor
4	Between weak and vital
5	Vital importance of one factor
6	Between strong and demonstrated
7	Demonstrated importance of one factor

8	Between demonstrated and absolute
9	Absolute importance of one factor

3 Problem Description and Business Scenarios

In the previous chapter, we were able to look at the main key scientific concepts related to business intelligence, sales and client segmentation. In this chapter, we will now focus on the core business intricacies and nuances that must be considered in order to better understand the final tools and results delivered. It is believed that by the end of this chapter, the reader can critically deliver opinions about how to analyze the business and bring value to the teams.

3.1 Sell-In vs Sell-Out

In the Buying Groups Market and in many other similar Markets, both in the Pharmaceutical and other sectors, it is crucial that the concepts of Sell-In (SI) and Sell-Out (SO) are well understood. On a theoretical level and for most of the businesses, there is an intermediary step, which is the Sell-Through. In order to better visualize the different steps, let us focus on Figure 3.1:

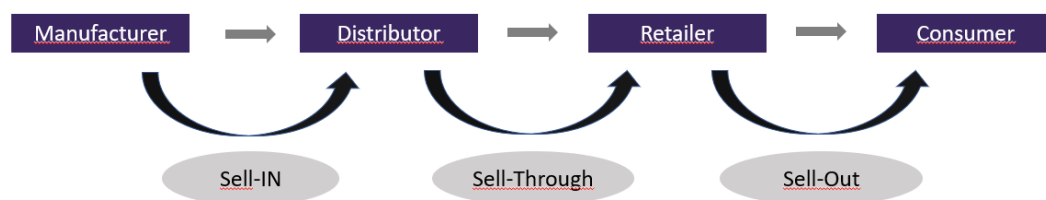


Figure 3.1: Typical Supply Chain Concepts

As seen above, SI usually concerns the products commercialized from the Manufacturer to the distributor, Sell-Through from the Distributor to the Retailer and SO the ones from the Retailer to the final Consumer. Thus, the main challenge in this process is to make sure that, in this supply chain, everything that goes to the next stage is sold as fast as possible to the next one, being that an indication of good performance and good forecast of the necessary quantity of products to be transacted.

For the context of the company and for this challenge in specific, Figure 3.1 becomes the Figure 3.2:

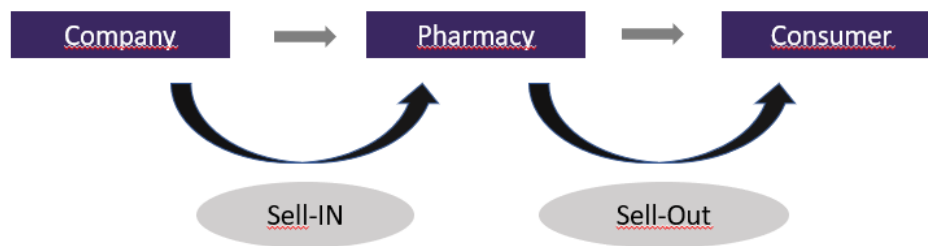


Figure 3.2: Adapted Supply Chain steps for the Company Context

As it is possible to analyze, for the day-to-day operations of the company, Sell-Through is left out of the equation and closer attention is left for SI (products commercialized from the company's commercial team to its Pharmacies) and SO (from the Pharmacies to the final Consumer).

3.1.1 SI vs SO: Pharmacies do not belong exclusively to the company and the problems associated with it

As counter intuitive as it seems, all Pharmacies that belong to the company's Network of Pharmacies are not owned by it, meaning, they don't have, for example, to exclusively buy and sell products that come from the company and they can leave as soon as they would like. This is a Portuguese regulation that does not allow for Pharmacy Networks to have their own Pharmacies, but solely associated ones.

With this, some major problems arrive that are useful to explore through analytics to detect the major deviations in the perfect scenario:

1. **Pharmacies can buy directly to the Laboratories (Manufacturers):** even though the company tries to negotiate large volumes of products with the Laboratories to execute the best possible deal for the products in its portfolio, it happens that Pharmacies, sometimes, negotiate directly with the Laboratories. This process makes the company to lose Market Share that could be concentrated in its portfolio and causes, in the long-run (2 to 3 years' time), that Pharmacies have less margin since the company can't improve its business practices to continuously decrease the price of each product unit. This rationale comes from a short-term desire from the Pharmacies of having the best possible margin right on the current month, making them go directly to the Laboratories as the prices per unit are lower (the company allocates an extra margin, thus making it higher for the Pharmacies);
2. **Pharmacies are autonomous in their product portfolio of choice to sell in their stores:** since Pharmacies do not have to strictly follow the strategic directions that come from the category management and the consultancy teams regarding which products should be commercialized, many times Pharmacies do not detect either margin opportunities or cross-selling opportunities. Thus, they end up buying products from the company that, again, delivers the maximum short-term margin for a specific

month, not thinking on the long-run returns, even if the company recommends the best ones for them;

3. **Pharmacies do not share their overall purchase data:** since Pharmacies are not owned legally by the company, they do not have to share the overall data of their purchases. Thus, this prevents the company from having direct access to missing opportunities or products from other Labs competing directly against the ones they propose to the Pharmacies, making it harder to have the full picture. This can be, nonetheless, turnaround by looking at the SI vs SO disparities and see which products are Selling-Out, but not In;

3.2 The categories to take into account

Across the whole business, the company operates on multiple categories of products, thus, it is crucial to take a closer look on the whole structure of products on which the company operates on and how do they divide themselves in different layers, as this knowledge is vital to take the best possible conclusions from the final dashboards and segmentation tools created. They are as follows:

1. **OTC (Over-The-Counter):** these are the medicines/products that do not need any type of prescription from a health professional (usually a doctor) in order to be sold to the final client, in a Pharmacy. For our context, these products are the ones the supplier XYZ focus on the most, so when presented the dashboard and the customer segmentation analysis through clustering, these are the products to be considered. Moreover, these products divide themselves in two further large categories very used in the pharmaceutical world:
 - a. **MNRSM (Medicamentos Não Sujeitos a Receita Médica) or P (Pharmacy):** these medicines/products can only be sold by a pharmacist in the Pharmacy, as they are not available for the client to pick them up physically in the store. As a curiosity, it is not possible to make a commercial campaign out of these products, for example. As examples we may find ibuprofen 200 mg and paracetamol 500 mg;
 - b. **PSBE (Produtos de Saúde e Bem Estar) or Consumer Healthcare Market:** these medicines/products are distinct from the previous ones described as they don't need a pharmacist to supervise and allow the purchase, the client can pick up the product by themselves. As examples of these we have body lotions and aspirins.
2. **MSRM (Medicamentos Sujeitos a Receita Médica) or POM (Prescription Only Medicines):** these are the medicines that need an authorization from a health professional (usually a doctor), to be able to be sold, since they can have harmful side effects if the patient has autonomy to administer them on its own behalf. As examples of these, we have paracetamol 1000 mg;

3.3 The AS-IS Difficulties and Processes

Up to date, the Key Accounts team did not have a fully functional dashboard, decision-making processes or other more advanced analytics tools put into practice, where it would be embedded in the daily culture of the company to be used.

As for the Business Intelligence tool requested, every team of the company, every time it was necessary to develop any kind of analysis, it was necessary to go through time-consuming ETL (Extraction, Transformation and Loading) processes from the Oracle database and multiple other data sources coming from the Pharmacies created for the company, leading that to the following problems:

1. **Not flexible filtering:** if the team forgot, for example, to add a specific Month for its analysis relevant for a near-future discussion, they have to go through the whole process again of ETL in order to capture and extract value from that missing Month;
2. **No cross-filtering:** without a proper BI tool, the team did not have enough data discovery capabilities, as classic tools such as Microsoft Excel do not deliver a nice GUI (Graphical User Interface) to make elements such as graphics to interact between each other with drill-downs and highlights capabilities;
3. **Different Treated Data Sources:** since every team developed self-made reports with different sources, many data validation scenarios were simply impractical to perform, as the teams had different results for the same phenomenon;
4. **Stability of the Data Used:**

When it comes to the Customer Segmentation, the company only had a segmentation done based on the subscription level of the Pharmacy inside the Company Group, not Purchase or Selling Performance oriented.

As for the subscription level, this represents the contract status of each Pharmacy towards the company, allowing the Pharmacy to have more or less benefits depending on that contract. This subscription is not an extra factor, but rather the main bond between the company and the Pharmacies. In the following list, we may find the levels more into detail:

1. **Base:** A Pharmacy that pays the minimum amount per month to have access to the the company's Portfolio and Discounts;
2. **Plus:** A Pharmacy that pays the intermediate amount per month to have access not only to the company's Product Portfolio, but as well as 3 of 7 services of choice;
3. **Full:** A Pharmacy that pays the maximum amount per month to have access to the company's Product Portfolio and all the 7 services provided by the company;

Even though this segmentation is relevant to know the level of service engagement each Pharmacy has with the group, it does not allow to better understand their buying behavior when it comes to specific products. Thus, the customer segmentation developed answers specific and necessary marketing questions about the Pharmacies, allowing the group to have new commercial initiatives with them.

3.4 Main Strategic Aspects currently analyzed by Key Accounts of XYZ for the Product Performance Analysis

Currently, the Key Accounts team has many strategic questions to address on monthly basis, mainly when they are making their monthly performance revision and planning the next

month(s) ahead. Thus, it is important to make a brief guide through on these main questions and what they represent for the business.

First things first, the team focuses mostly on trying to have as big Pharmacy penetration as possible with its products, meaning, trying to get as many Pharmacies to select their products as the main products for certain categories. This translates in how many Pharmacies each category, sub-category or product is present, delivering an indication of its attractiveness and knowledge present on the Pharmacy about it. Because of the conclusions taken from the previous point, the team tries to understand if new products should be launched or removed from the Portfolio or if current products should be rebranded alongside a commercial campaign, in order to revive their lifecycle.

Secondly, on a more strategic and broad perspective, the team has a high target oriented and growth mindset, always seeking to analyze their performance with specific months prior, prior to last year to against targets they define as “success” at the beginning of the year. With this, its fundamental to have overall monitoring of the business, monitoring of the performance and profile of each Pharmacy and how the different products and categories contribute to this matter. Because of these analysis, action plans and training sessions are designed for specific Pharmacies in order to increase their performance and bigger data exploration is performed to find potential problems not expected in commercial campaigns. Moreover, this perspective plays a crucial role in continuously reviewing the targets of the year, cumulative and monthly, with the Board Members.

Finally, the team also has necessities of monitoring the available stocks in warehouse and how much internal Market Share Supplier XYZ represent. As for the stocks, there’s the clear worrying of making sure the work is as much as possible in a Just in Time regime, meaning, you have the least possible in stock, since there are high costs associated with it and products can become commercially invalid. As for the Market Share, both in SI and SO, the team focuses on increasing as much as possible this number, as the Supplier XYZ is the most beneficial supplier both for the Pharmacy and the company for margin gains in specific cases

With this knowledge, it is going to be further easier to understand the rationale behind the tools created in this dissertation.

3.5 Main Strategic Aspects currently Analyzed by Key Accounts of XYZ for the Pharmacy Segmentation

On the level of the Pharmacy Segmentation, the Key Accounts have the main strategic concern of how they should group the Pharmacies and what are the best criteria to do so in terms of marketing actions. Even though, as explained previously, there was subscription-based segmentation (Base, Plus and Full), this was not based on marketing analysis neither on more sophisticated models leading to no possibility of more defined targeted actions regarding products market share in the Pharmacies and Pharmacy loyalty towards the specific supplier the Key Accounts are responsible for.

On a financial level, not having proper customer segmentation does not allow the company to evaluate precisely what is the current value of different groups of Pharmacies and their potential in the future, taking some prioritization away in the process regarding the commercial importance to the business.

On a benchmarking level, without the customer segmentation, the Key Accounts cannot have a precise awareness to which Pharmacies to compare specific ones during their Pharmacy Support, taking away more customized-approach opportunities.

Following up on the benchmarking level, in the operations and sales part, also without the customer segmentation, the team is limited in the priority ranking it delivers to each Pharmacy, as without it, the team can't rationally chose which group to improve firstly and, inside each group, which Pharmacy and why.

Finally, one of the key aspects of having customer segmentation is also the ability to monitor the evolution of one Pharmacy across the universe of segments and see if it went from a poor performance segment to a more competitive one across time, enabling the team to assess if their business improvement and sales initiatives worked as intended.

3.6 To-Be Analysis

With all the strategic questions highlighted above, which represent some of the strategic aspects both the dashboard and customer segmentation will bring, it is possible as well to define what are the concrete outcomes that are expected later on after deployment phase in both projects.

As for the dashboard, the team is expected to be able to:

1. Understand how attractive the categories and specific products are;
2. Understand which products to invest more, divest or make a rebrand;
3. Understand which Pharmacies are being the best performers in certain categories and periods of time;
4. Understand which Pharmacies are more aligned with the Supplier being analyzed when it comes to its offers;
5. Understand if both the levels of stock in the Pharmacies and centrally are inside the expected;

On Figure 3.3, it is possible to depict a possible exploration scenario coming from some reports pages of the built the dashboard. The goal is that the team can assess certain targets on the dashboards and have two main actions: be able to question the business around possible actions and behaviors took in the past and be able to also take action with the teams connected to it.

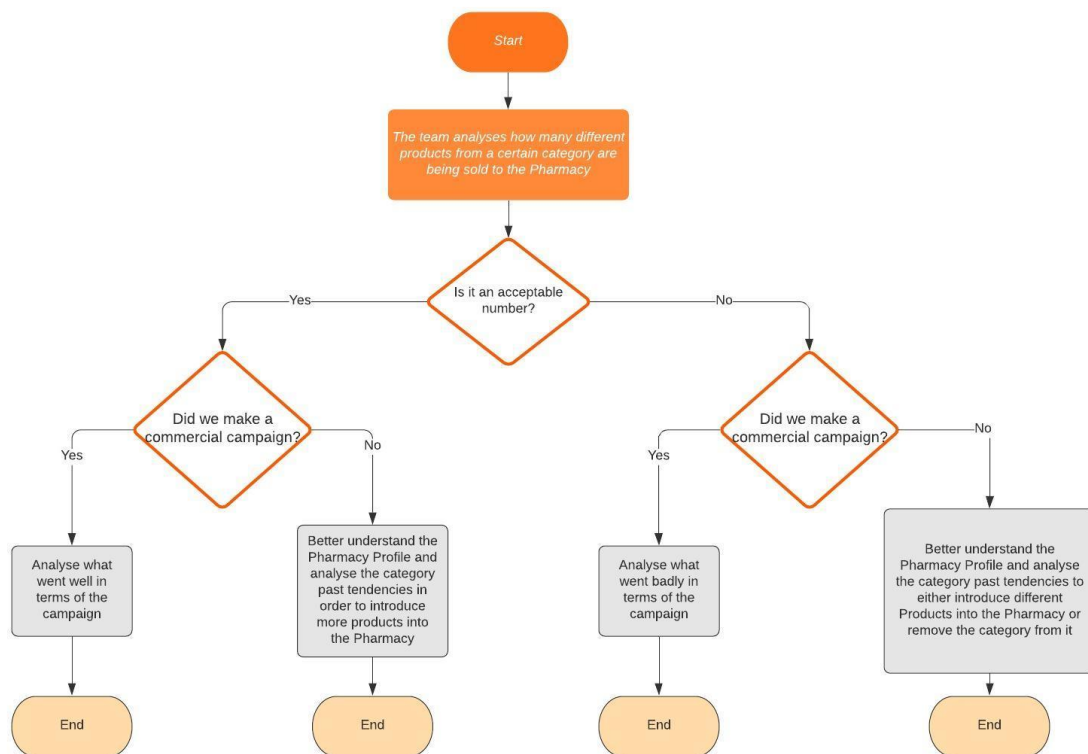


Figure 3.3: Decision-Tree made possible after the exploration of the Report Pages of the Dashboard

As for the Pharmacy Segmentation, the team is expected to be able to:

1. Prioritize which Pharmacies to tackle first and which to tackle later;
2. Understand the buying behavior of specific Pharmacies;
3. Understand and spot which Pharmacies deliver the most value to the specific supplier in analysis;
4. Define tactics to improve the purchase behavior of specific Pharmacies inside each segment;

Similarly to the dashboard, it is also important to display a possible thought process once the segmentation is complete, as we can see in Figure 3.4 :

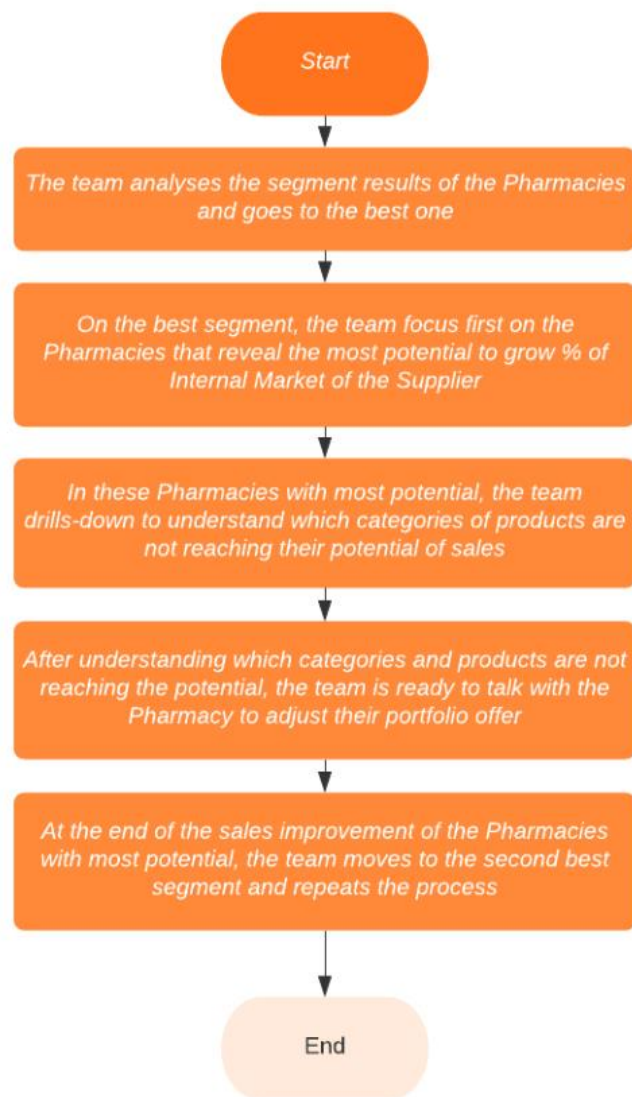


Figure 3.4: Thought Process after the Segmentation Process

4 Dashboard for the Product Performance Analysis of Supplier XYZ

After going through the main company strategic concerns, main usability and user experience requirements, one of the major dissertation outcomes was the development of a turnkey dashboard for the key account managers of Supplier XYZ. This tool that will enable them to have a more systematic way of comparing performance across time and have a better drill-down exploration on the main factors the team feels affect the business. Through this chapter, we will discover the main philosophy behind the dashboard, the key interactive features developed, the KPIs selected to convey context to the main questions of the business and, finally, some of the most important dashboard pages as well.

4.1 Dashboard philosophy and main usage requirements

Different teams and businesses need different types of dashboards. After discussions with the key account managers, the dashboard had to be constructed on the following premises:

1. **Exploration was the key:** the main goal was to create charts and elements that would motivate the team to further investigate on what went right or wrong with their prior business decisions. Thus, prescriptive aspects were not developed neither elements that would suggest any very direct and real action after dashboard interpretation;
2. **Extra graphics for drill-down:** even if in some dashboard pages it was not the main goal, the team came to the conclusion that it was beneficial to add, in most of the pages, graphics that had more detailed information about category and specific products performance, allowing the team to instantly understand why a certain KPI was behaving as such by looking at the specifics of the products;
3. **Metrics across time are vital:** also, as a mindset, the team agreed that, even though the main goal in some pages might not be to focus on evolution of the metrics across time, it's a crucial addition, since it enables the team to have eureka moments finding out increases or decreases, they might have not been expecting if they hadn't the time elements;
4. **Cross Filtering and Drill-Down:** as usual, but also reviewed and decided, the team felt that all the pages should have maximum possible interaction within its elements. If we are looking at a certain chart and we are interested in one aspect, we click and the whole dashboard adapts to that detail we are looking for;
5. **It should alarm towards the status of the business:** also, as a core mindset, the team felt the need for the dashboard to warn them about specific aspects that were either off track (negative towards a certain target) or totally great, that they should pay attention to how good of a performer a certain pharmacy or category/product was. For this,

many alarming colors were added conditionally to the dashboard to adapt with the users' selection;

4.2 Performance Metrics Areas and Definition

Before advancing to the building of the dashboard, the team presented many metrics relevant to the business, some of them quite complicated to perform on non-BI tools, thus the necessity of the dashboard. As for clarity of understanding, the metrics were divided into areas, not only for being easy to manage the timeline of the dashboard construction, but as well as for later reporting to be easier done in the company. Furthermore, it is important to notice that some of the metrics also came as the meetings went by, when new necessities were spotted and more complex analysis were defined to complement the initial requests, so, the developed metrics were divided into "Initial Request" (coming from past necessities already) and "Proposed" (developed alongside the team by critically assessing what was needed extra). The areas of the KPIs and the respective KPIs are as follows in Table 4.1:

Table 4.1: Focus Areas and correspondent Metrics Developed during the creation of the Dashboard

Focus Area	Performance Metric	Way of Development
General Business	Margin/Profit (Absolute) (€)	Initial Request
	Margin/Profit (%)	Initial Request
	Growth of Sales Compared to Last Year (%)	Initial Request
	Number of Pharmacies above X € Monthly of SI Purchases (#)	Initial Request
	Cumulative Sales (€)	Proposed
	Cumulative Sales Difference to Last Year (€)	Proposed
	YearToDate Sales (€)	Proposed
	MonthToDate Sales (€)	Proposed
Pharmacy Specific	Number of Pharmacies in between X and Y levels of euros in SI and SO	Initial Request
	Distinct Number of Products Sold/Bought	Proposed
	Ranking of Pharmacies considering their Growth in Sales to Last Year;	Initial Request
	Ranking of Pharmacies considering their Growth in Margin to Last year;	Initial Request
	Average Sales in Value and Quantity of the Top 20 Pharmacies in Sales	Proposed
Product Specific	Distinct Number of Products Sold	Proposed
	Distinct Number of Pharmacies that bought/sold a specific	Initial Request

	category/subcategory/product	
	Ranking of Products considering their growth in Sales to Last year;	Initial Request
	Ranking of Products considering the Margin to Last year;	Initial Request
	Growth of sales considering the average of Sales from Last 3 months;	Proposed
	% Representation in Total Sales of the Top 20 products in sales;	Proposed
	% Representation in Total Margin of the Top 20 Products in Sales	Proposed
	Average Sales in Value and Quantity of the Top 20 Products in Sales	Proposed
SI vs SO Alignment	Difference between products sold in SI and SO	Proposed
	Difference between Number of Pharmacies that buy in SI and SO	Proposed
	% of Product Quantity Sold Out from SI bought in the same time period	Proposed
Stock Specific	% of Stock in Quantity / % of SI in Quantity of a category or Product	Proposed
	% of Stock in Value / % of SI in Value of a category or Product	Proposed
	Number of Months still in Stock	Initial Request
	Participation in % of Stock of the Top 5 Products in Quantity in Stock	Proposed
	Participation in % of Stock of the Top 5 subcategories in Quantity in Stock	Proposed

- General Business:** this area and metrics are dedicated to deliver insights on the general state of the business, always thinking that most of these metrics are going to be not only relevant for the respective team of the supplier in question, but as well as the most condensed and crucial metrics that related, financially and strategically, to the metrics evaluated on the Board level every month and year. As a special note here, we should point out three important metrics that deliver powerful insights to the team. Firstly, “Growth of Sales Compared to Last Year (%)”: this is a classic metric, however, it delivers relevant knowledge on how the business was consistent and if any internal or external variable affected deeply the business; secondly, “Difference of Cumulative Sales Compared to Last Year” is a very powerful metric the team asked for, since it delivers how good the business is, at this moment in time, compared to last year, taking into consideration all the months you may be analyzing. As an example, a specific month it was not optimum compared to last year, but cumulatively you might be performing better overall, giving a better sense of long-term commitment to the team; finally, “Number of Pharmacies above X Euros in SI Bought

in a Month” represents the networks target of pharmacy alignment in that specific supplier we are analyzing. We can see, by order of explanation, the equations that lead to these metrics on the dashboards in 3.1 and 3.2:

$$\text{Revenue Growth to Last Year} = \frac{\text{Revenue Current Period Selection} - \text{Revenue homologous Period Selection}}{\text{Revenue homologous Period Selection}} \quad (3.1)$$

$$\text{Difference of Cumulative Revenue} = \text{Revenue Current Period Selection} - \text{Revenue Homologous Period Selection} \quad (3.2)$$

- **Pharmacy Specific:** this area refers to metrics that allow the team to extract the maximum number of information and detail about the Pharmacies performance and how its universe of work for the supplier. From here, one metric is relevant to point out and discuss about: “Distinct Number of products Sold/Bought” delivers how much loyal that Pharmacy is to the supplier products by showing the magnitude of acceptance of its portfolio of offer (this is calculated by counting the distinct products bought or sold, depending of SI or SO, for the Pharmacy or by the Pharmacy);
- **Product Specific:** this area refers to metrics relevant to study the universe of the products and to better understand which categories and products need the most attention from the team and which categories push the business even further. With this, link those conclusions to business actions that the team might have done in the past and will perform further in the future, such as marketing campaigns. Here, it’s relevant to point out two key metrics developed: firstly “Distinct Number of Pharmacies that bought/sold a specific category/subcategory/product” measures on how many different Pharmacies the product penetrated, meaning, how attractive the category/subcategory/product has been for the Pharmacy network, and it counts if it was sold/bought at least once per time frame we are analyzing (regardless of it was 100 units or 1 – analyzed by the SKU, as the “Pharmacy Specific Area”); secondly “% Representation in Total Sales of the Top 20 products in sales” that measures how much does the Top 20 products represent in terms of sales for the business, both in SI and SO, allowing the team to have a sense if they are too dependent or not in certain products, something that sparks adjustments in pushing or not more products in commercial campaigns or creating more products to dilute that percentage in case it is too high;

$$\% \text{ Representation of Top 20 Products in Revenue} = \frac{\text{Sum of the Revenue of the Top 20 Products in Revenue}}{\text{Total Product Revenue}} \quad (3.3)$$

- **SI vs SO alignment:** this area of metrics is a critical one to the business, because it allows to see how the products are flowing from the stage of SI to SO and detect the Pharmacy patterns in product accumulation tactics. Thus, one metric is relevant to point here: “% of Product Quantity Sold Out from SI bought in the same time period” displays how many products are being sold by a Pharmacy compared to how much are they buying (equation 3.4). This allows the team to understand what are the months the Pharmacy clearly accumulates stock and the others that are simply selling, delivering a sense of which Pharmacies concentrate efforts of purchase in specific months or if distributed along the time frame being analyzed;

$$\% \text{ Quantity Flow of SO/SI} = \frac{\text{Quantity of Items Sold in SO by the Pharmacy}}{\text{Quantity of Items Bought in SI by the Pharmacy}} \quad (3.4)$$

- **Stock specific:** this area refers to the analysis of the current stock of products referring to the supplier we are analyzing, delivering a great perspective to the team about the products that make up the most stock and that it does not represent a competitive advantage. Two metrics that need to be pointed out are: firstly, “Number of Months still in Stock” that is the division of the average sales in quantity of a specific category/product by the quantity existent in stock in that specific month, allowing the team to have a rough idea of how many months still of stock does it have to support the average from the previous months; secondly, and as an example, “% of Stock in Quantity / % of SI in Quantity of a category or Product” delivers the perception to the team on how much the representation of the quantity of stock a specific product has is aligned with that representation in SI. If a certain category/product is not aligned, it means that the stock should be further analyzed;

$$\text{Number of Months still in Stock} = \frac{\text{Average of SI Quantity of the Last X Months (per Month)}}{\text{Quantities in Stock Up to Date}} \quad (3.5)$$

$$\text{Stock Alignment with SI} = \frac{\% \text{ of Representation in Stock}}{\% \text{ of Representation in SI}} \quad (3.6)$$

4.3 Data Treatment and Tables Structure

As for the Data Treatment and the Tables Structure, not a lot of work was put into building it, but instead in acknowledging what was already built. For the project, the company already had established a Data Warehouse for all the teams to use in different scenarios with multiple tables already in place and connected. For the creation of the dashboard, there were some major tables that were used:

1. Fact Tables:

- Sell-In:** where all the transactions made to Pharmacies are registered;
- Sell-Out:** where all the transactions made from the Pharmacies to the final client are registered;
- Stocks:** where the current stock of each product in the warehouse is registered;

2. Dim Tables:

- Product:** each line represents a product present in the Portfolio of the company possible to be sold, depending only if it has stock;
- Geography:** each line containing data of the city, district and other details of a specific zip-code;
- Pharmacy:** each line contains information about one Pharmacy;
- Supplier:** each line contains information about one Supplier;

As for a quick overview of how the tables are connected, Figure 4.1 presents us a simplified UML version of it:

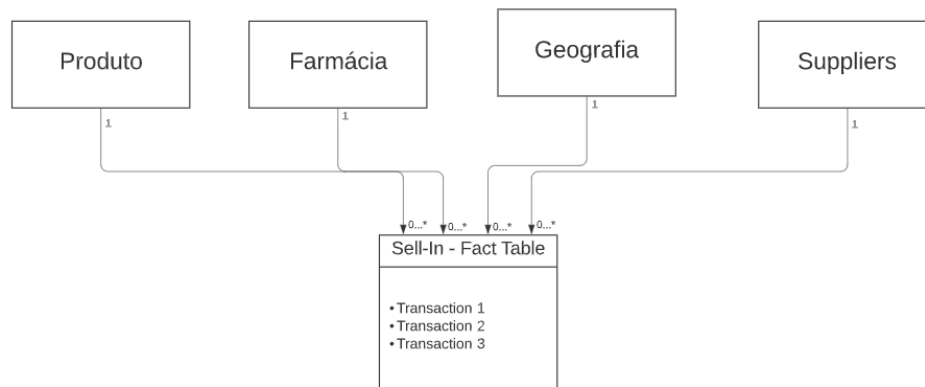


Figure 4.1: Adapted UML for the SI Process from the Data Warehouse

As for the remaining two Fact Tables, they are connected in the same way to the 4 Dimensions Tables.

When it comes to the data presented in the dashboard, all the values were changed in order to do not disclose any valuable information from the company and to provide, still, for the dissertation purpose, a nice experience by having data labels both on cards, graphics and tables.

The whole project was created by having a “Live Connection” established directly to the Dashboard, so, as stated in the beginning, no ETL processed was put into place not because of the lack of knowledge to do so, but due to the nature of the company, which was dealing with a Data Warehouse.

4.4 Dashboard Usage Journey

Before moving on to the potential interpretation of the most crucial pages of the dashboard, it is important to address how the dashboard was constructed and built-in terms of user experience. For this, it is important to address which pages types of pages are available and what is the user path in between them, so that we further take the most from the interpretation of it. For this, let us initially focus on the following diagram of Figure 4.2 that explains the flow from where the user would go through:

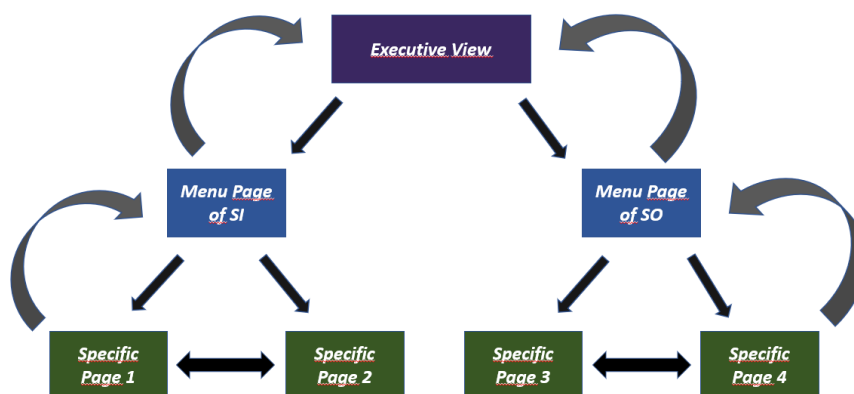


Figure 4.2: User Journey when navigating through the Dashboard Report Pages

Firstly, the user will face an “Executive View”, where the main KPIs of the high-level business associated with that supplier are displayed, with color codes in order to have a quick overview of what might be going as expected or not. Once first impressions are taken, in this “Executive View”, the user can choose to either analyze the SO metrics where they will be able to see the Pharmacies performance to the final client or they can chose to go to the SI pages. On the “Menu Page” level, the user can choose one of the various pages created to go to in order to analyze a specific business aspect, such as Pharmacies, Products, Stocks and so on. Now on the third level, the user will explore the main business questions and, if necessary, as shown in the diagram, go from one page to another inside each area (SI or SO) that have relationship for the team when it comes to the possible analyses to be done.

4.5 Main UX/UI Elements

During the usage of the Dashboard, the user/team will face many elements that help them to better navigate across pages and to take the most of each page, through advanced filtering and thoughtful display of the visual elements. Thus, in this section, we want to provide a small compilation of elements that are necessary to take into consideration.

4.5.1 The Slicers

As in any dashboard, and as well as described in the main rationale behind the dashboard, “Slicers” were included. This feature delivers the team flexibility to filter all the report pages only to see the performance of one product, or category or specific Pharmacy, one of the main goals with the project as well. By Figure 4.3, we can see some examples of it:

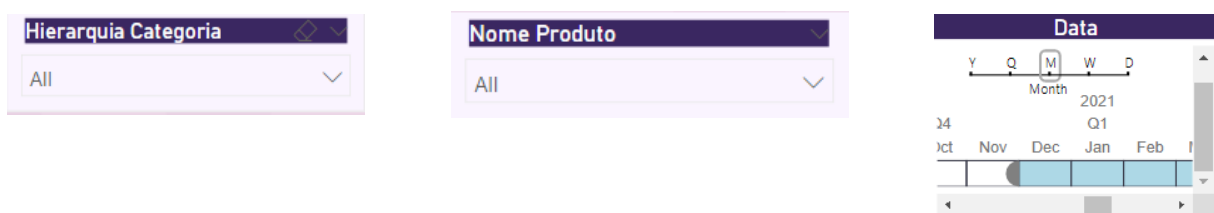


Figure 4.3: Slicers found across the Report Pages of the Dashboard

In every page, the team will have the opportunity to filter towards the month/range of date that is the most relevant to them. It was decided to use this more user-friendly date slicer as it allows drag selection for the user. All the slicer for products and categories were decided to be done in dropdown, because of the economy of space. Finally, all the slicers were put either on the left side or top of the dashboard, to not interfere with the main graphics usage.

4.5.2 The KPIs and Metrics Cards

In order to deliver a more target-oriented perspective to the dashboard, cards were added to various pages, both to condensate metrics in the graphics in one single elements (revenue of January, February and March combined in a single card) and to provide quick

color context, to indicate if a certain target was reached or not. Figure 4.4 delivers a nice example of the cards we may find:

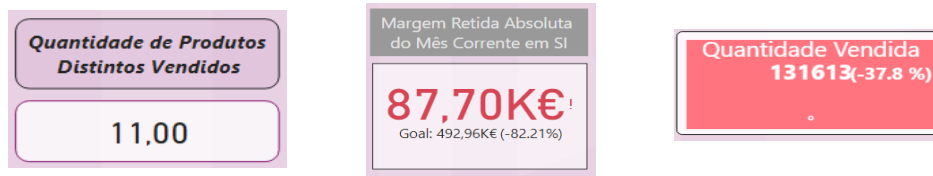


Figure 4.4: Examples of KPI Elements used across the Report pages in the Dashboard

As a decision, red was chosen always to represent “below target” and when it green is “above target”. Similar to the slicers, the KPIs and Cards were usually allocated on the sides or top of the Dashboard, to deliver a more organized experience and make the user focuses either on the graphics or the main KPIs.

4.5.3 Navigation Buttons

Since the dashboard has many different pages, it was crucial to deliver elements that would allow the users to go back and forth between levels of the dashboard or similar pages within the same level. Thus, with Figure 4.5, we can see some buttons added:



Figure 4.5: Examples of Navigation Buttons used across the Report Pages of the Dashboard

Both the arrow and the house symbol are the actual buttons for the user, making suggesting the actions of “going back” or “going home”, respectively.

4.5.4 Drilldowns, Points Selection and its effects on the whole page

In order to deliver as many exploration possibilities to the team as possible, the dashboard was built for the team to be able to physically interact with it. This means that by clicking on elements of the dashboard, the team will have a more refined perspective on the metrics being evaluated. For instance, if we click on a specific category in a bar chart, the whole report page will display the metrics only considering that category; if we click on a data point in a line chart across time, it will only display the metrics for that specific selected month. Finally, when with the drilldown feature (going one level down on the analysis), if we click on the graphic, it goes from showing all categories to showing all products of the selected category. Figure 4.6 illustrates what we described before:

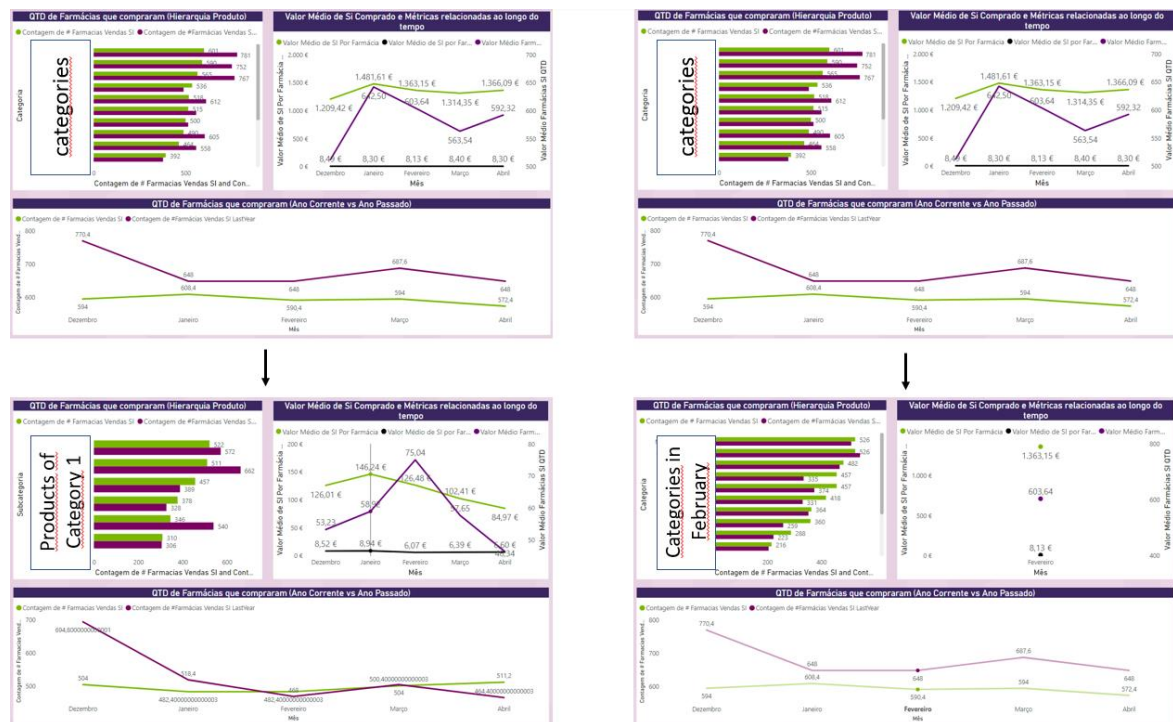


Figure 4.6: Main User Discovery Elements in the Report Pages of the Dashboard

4.6 Tool Used and its limitations

As for the Dashboard, the tool used was PowerBI Desktop, a BI software widely used for performance analysis. As shown above in section 4.3, the Dashboard was built on the premise to be a turnkey tool, meaning, once ready, the goal was to be used right away by the team and if the data used needed to be refreshed from one day to another, it would be possible to do so since the team was using a live connection to the Data Warehouse, also presented already in section 4.3.

PowerBI was the tool selected not only because of its ease to use but as well as its big interconnection with other Microsoft Tools, such as Power Apps, Power Automate, Azure and SQL Servers. This selection was also based on a long-term perspective of integration of the Dashboards created with some of these services to be able to further automate the discovery of results and the workflows within the company.

As a distinctive factor for this project, the whole dashboard was constructed using the programming language of PowerBI: DAX (Data Analysis Expression). This is the language in Microsoft PowerBI, Excel PowerPivot and Microsoft Analysis Services that allows the users to call for complex mathematical and filtering functions to create new metrics. As stated before, this was a necessity, since the whole dashboard was built on top of a complex already built data warehouse and not built through spreadsheets imports.

However, with every business choice, come its limitations. Thus, by having to work with live connection, the following business solutions and user experience aspects were not able to be implemented:

1. **Usage of Machine Learning charts:** with live connection, PowerBI does not allow the dashboard developers to take advantage of its machine learning charts. This is not

a limitation for descriptive analysis as the dashboard is built, but it removes possibilities such as root cause analysis with the “Key Influencers” Chart;

2. **Creation of auxiliar tables and columns:** with live connection, the dashboard developers cannot create new columns in the relationship model neither tables. As a result, it prevents the dashboard from having filters that allow to change from one analysis to another inside the same report page, taking out flexibility in one page, thus obliging the creation of multiple pages;

4.7 Dashboard: the report pages and its interpretation

With all the metrics, page flow, limitations and main targets of the construction of the dashboard, it is now time to dig deep into the pages themselves, look at the potential inside each page and why certain information was displayed in specific charts. As an initial disclaimer, it is important to notice that not all pages will be covered in the dashboard description, as it would be physically impossible to cover all of them in the dissertation core content. Thus, some of them that are repetition of the shown ones here (the main difference is that they are done for either SI or SO, what changes is the table we are working with), will be delivered as annexes.

4.7.1 Executive Page (Executive)

This page is the main page with all the high-level relevant KPIs to address in the business. It was fully built on the premise of delivering very quick insights on the performance of a specific month (by which you can change on the slicer on the top left “Mês Completo”), being used mainly for the current month the team may be present in. As we can see from Figure 4.7, the KPI card delivers us a quick sensation about whether or not the performance of the current month is below its homologous period (this last one as the target) and thus allows the team to have quick chats on what other metrics to explore associated with the 10 Key ones displayed. Visually, the goal was to deliver all the related KPIs in one line, so that SI and SO could be divided. By clicking on the respective buttons, we jump to the second layer of the dashboard, the “Menu Page”.

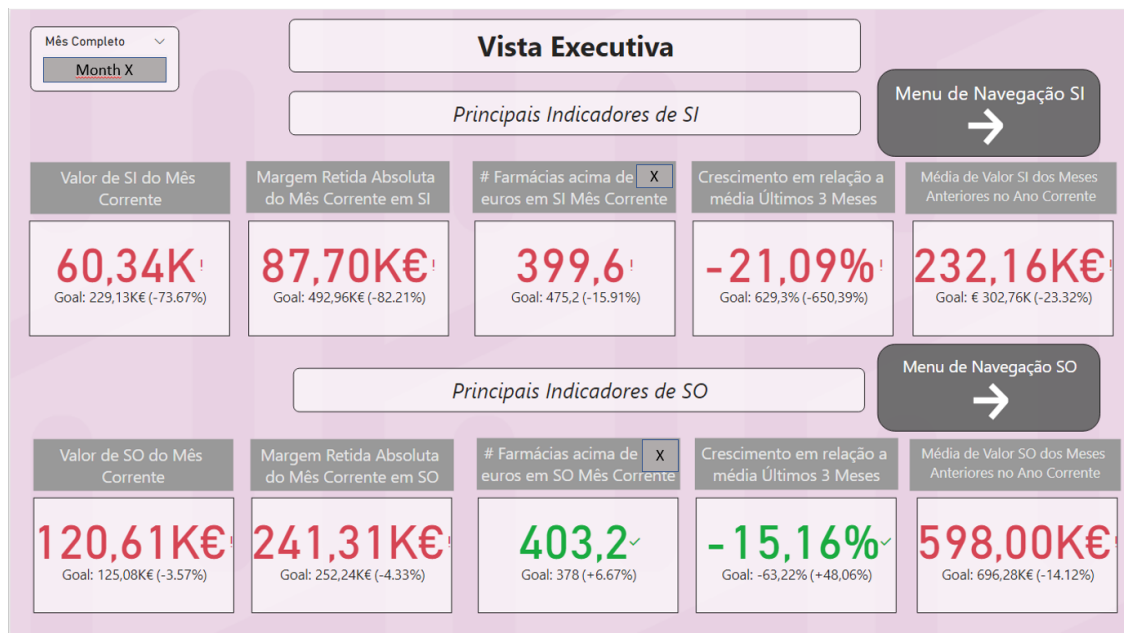


Figure 4.7: Executive Page, where the user can see the most important KPIs

4.7.2 Menu Page SO (Menus)

After having a quick glance at all the main business KPIs, the team can choose one of two menus: SI or SO. For demonstration purposes, we will focus on Sell-Out. The main purpose of this page was to deliver a full landscape of the options of pages the team had to select, each page corresponding to a technical/business question request. As we can see from Figure 4.8, the pages were clustered in different group types, being them the following: “Global Analysis”, where the team can look into broad details of the business across time, both on Pharmacy Level and the Business as a Whole; “Pharmacy Analysis”, the one on the left, allows the team to have dedicated pages on Pharmacy pattern discovery, in order to be able to compare Pharmacies between each other; “Product Analysis” allows the team to focus its attention on Product comparison and fine details of them; finally, “Extra Analysis” on the bottom will deliver the team important scatterplots that allow the user to detect how good of the performance of a certain product and category was in various dimensions at once. On the creation of this Menu Pages, it was paid attention to deliver information icons so that the user would have a brief explanation of what could be found before diving deeper.

Once the user knows that to look for, each title in the boxes represents a page to go through and it is at the same time a button, so the user just has to click on the title to go to the respective report page.

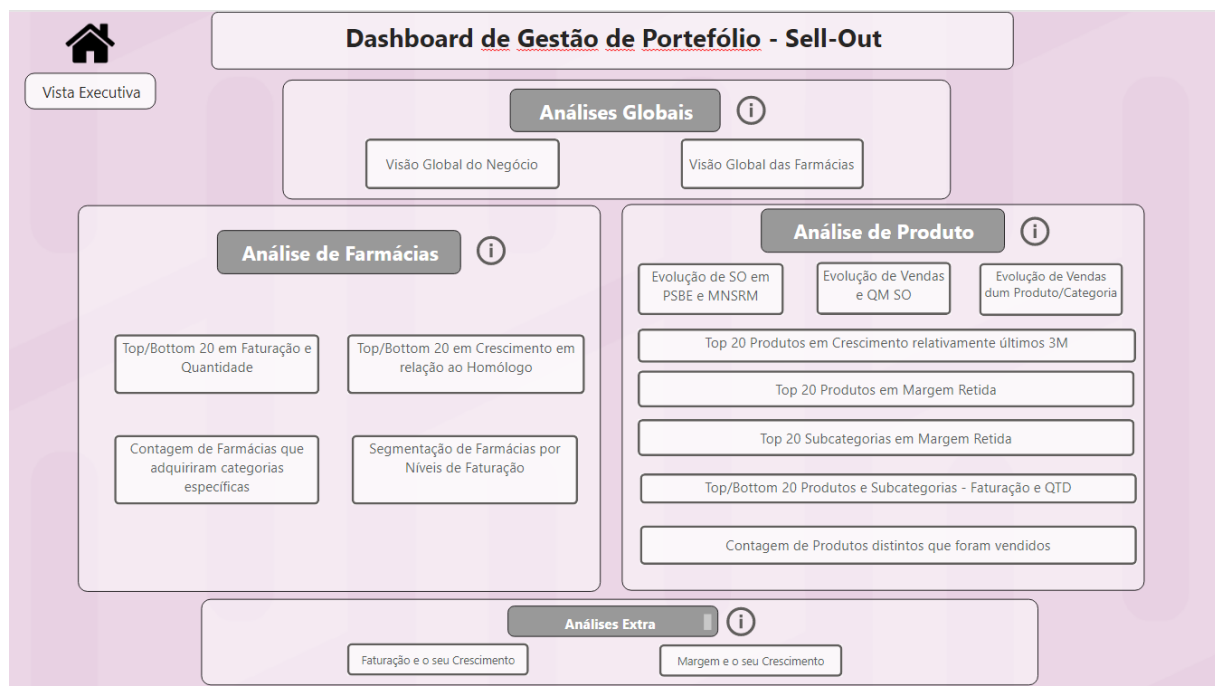


Figure 4.8: Menu Page for Dashboard Navigation

4.7.3 Global Analysis of the Business of SO (Global Business)

In the “Global Analysis of the Business of SO”, the team can look at the major KPIs that deliver a condensed idea on how the business is performing, but, unlike the “Executive Page”, it delivers it across time. As we can see in Figure 4.9, on the top of the page three line charts were aligned so that the team could visually focus only on half of the dashboard on Revenue, Absolute Margin and Percentage Margin, so that visual coherence and a better User Interface could be reached. Here, the team can look at specific months across time and discover that, for example, May 2021 had a better performance in Revenue and Absolute Margin, but not Percentage Margin than last year (green is current year and purple last year), meaning that even though, as a whole, more money was cashed into the business, per product it was made less profit, alerting the team to further analyze which products or categories caused this. For that detailed purpose, the team will have the Product pages to explore, however, as explained previously, we allocated bar charts in many similar pages with the performance of the products right on the side of the main requests, so that inside each page, the team could have exploration opportunities of products and categories. As a final note, we can also have a glance, across time, on how many Pharmacies we had above the critical target of SO alignment the team should be accomplishing for each Pharmacy, delivering the team a great assessment that, for example, May was better than last year’s one for this metric, giving a sense of continuous improvement. Finally, to deliver a broad target analysis (comparison to last year), it’s possible to see the cards applied on the left corner, where the team can see for the full time range selected how those metrics compare to last year’s one.



Figure 4.9: Global Analysis of the Business of SO

4.7.4 Global Analysis of the Pharmacy in SO (Global Business)

As for the team, this is the most complete Pharmacy focused analysis report page, where it is possible to have a full overview of the performance of a single Pharmacy. Without going for a complete description of the page, there are certain elements important to highlight that deliver a more immersive experience to the user. Firstly, the addition on the left side of the cumulative revenue (both current – green area – and the past year – purple area), where the team can analyze, at this moment in time, all the revenue aggregated, how is the Pharmacy performing and see which months delivered a great push to it. Furthermore, it was relevant to display the area chart below, the difference between this year and last year, so that more than the actual value, the team could have a very quick sense of competition vs the last year. Secondly, on the top of the page, the display of the 4 KPIs was a crucial addition for the team, mainly due to the fact that it delivers additional context for the graphics. For example, looking at the YTD (Year To Date) value, second and third KPI, we can see, in an aggregated way, the revenue and the margin, respectively, allowing the team to have a sense of the performance of the Pharmacy altogether. Thirdly, important to notice that this time, it was introduced bar charts for the direct comparisons of performance for the Product Revenues and the Product Quantities, allowing the team to have a more dissected perspective of which products evolved the most compared to last year. We can look at all of this in Figure 4.10:

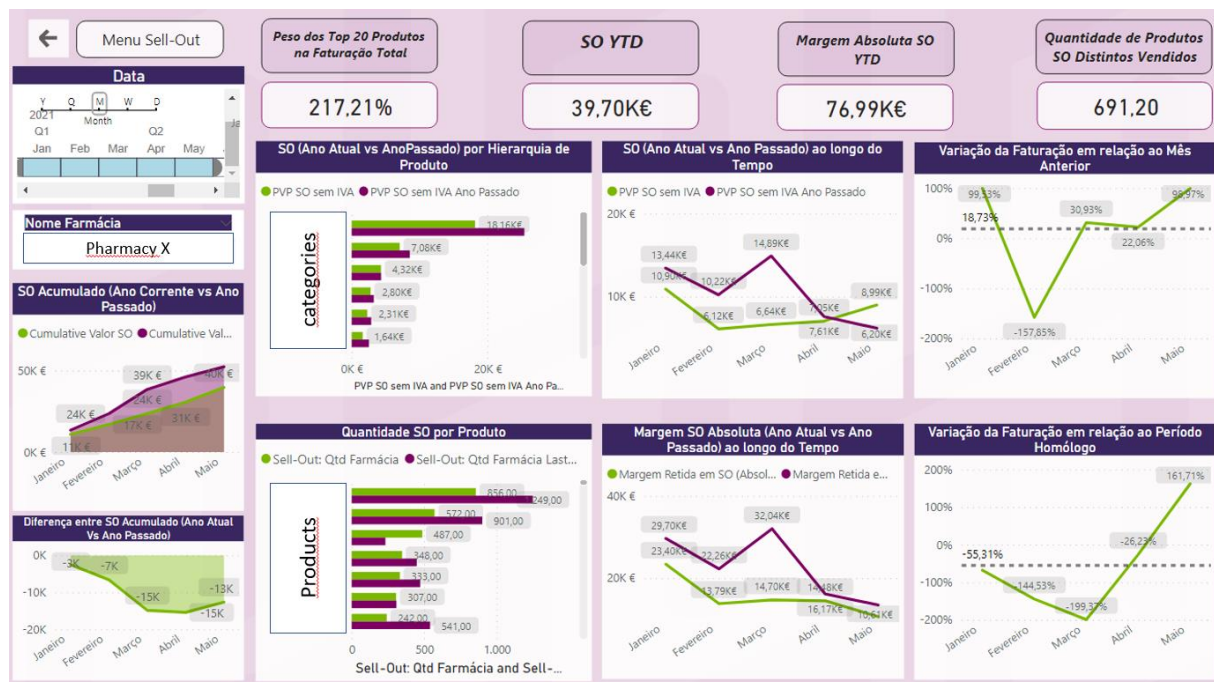


Figure 4.10: Global Analysis of the Pharmacy in SO

4.7.5 Pharmacy Penetration Analysis through SI (Pharmacy)

This page was fundamentally built for the team to analyze how attractive its portfolio is when it comes to acceptance of the categories/products of the supplier in question, by analyzing how many Pharmacies buy products. By analyzing Figure 4.11, the team has three main business opportunities to uncover: 1) How attractive each category and product is (1st bar chart) – here, it is possible to have a direct comparison between each category and product, not only between them, but with its last year (purple bar), giving the team a ranking perspective; 2) How valuable each product unit sold to the Pharmacy is, where across time (black line), we can analyze how much money per unit the Pharmacy invests in products of this supplier. Additionally, two lines in the same chart were added, so that the team could have a full perspective, across time, on how much the average money spent by each Pharmacy would vary (green line) and how many units of products on average per Pharmacy were bought (purple), allowing the team to see big drops or increases. 3) Finally, on the chart at the end, the line chart that goes across almost the whole page, delivers how many Pharmacies, in each month, bought products from this supplier, allowing the team to see which months were the best and the highest differences to last year (purple line).

Furthermore, in this page, one great user experience feature was added on top of it: each text box reacts to the selection being done in the page by showing the name of the category or product selected, both if we select a specific category or product on the slicer next to them or if we click directly on the graphics to further drill-down on the page information. This is useful for the team since it allows them to never get lost on what is being analyzed. Finally, the KPIs card were displayed on the side for the user not to get lost on the main discovery of the graphics.



Figure 4.11: Market Penetration on SI

4.7.6 Internal Market Share of PSBE and MNSRM from Supplier XYZ (Product)

In order to understand how competitive, inside the Company's Portfolio supplier XYZ was, a dedicated page for that purpose was built. By having a closer look to Figure 4.12 we can see many dimensions of the competitiveness displayed in various ways. Focusing on the top half of the page, the user can have a "Market Share" perspective, of the biggest clusters of products: MNSRM and PSBE. The main goal is to deliver the team the perspective of how much supplier XYZ represents in these two main clusters (green line current year, purple line last year) and allow them to monitor that percentage level, which is the one they use to create their targets. On the bottom half, instead of percentages, it was decided to deliver the full revenue provided by XYZ on the same clusters, so that the team would have a clear real perspective of how much money it represented for the business, which is, at the end of the day, the crucial aspect also for alignment with the Board members. As an interesting example of this is the month of January being displayed: even though, talking about Market share, this current year it has more in terms of percentage compared to last year, the total amount of money that represents is less (if we check the graphic below), allowing the team to see that it went, generally, worse than last year.

On a final user experience note, very relevant to the team, it was decided to add the two donut charts on the side. These donut charts show the user, for the money captured in revenue by the supplier XYZ, how much money comes from PSBE and MNSRM, in percentage, allowing them to detect if there was one month or year atypical for this distribution.

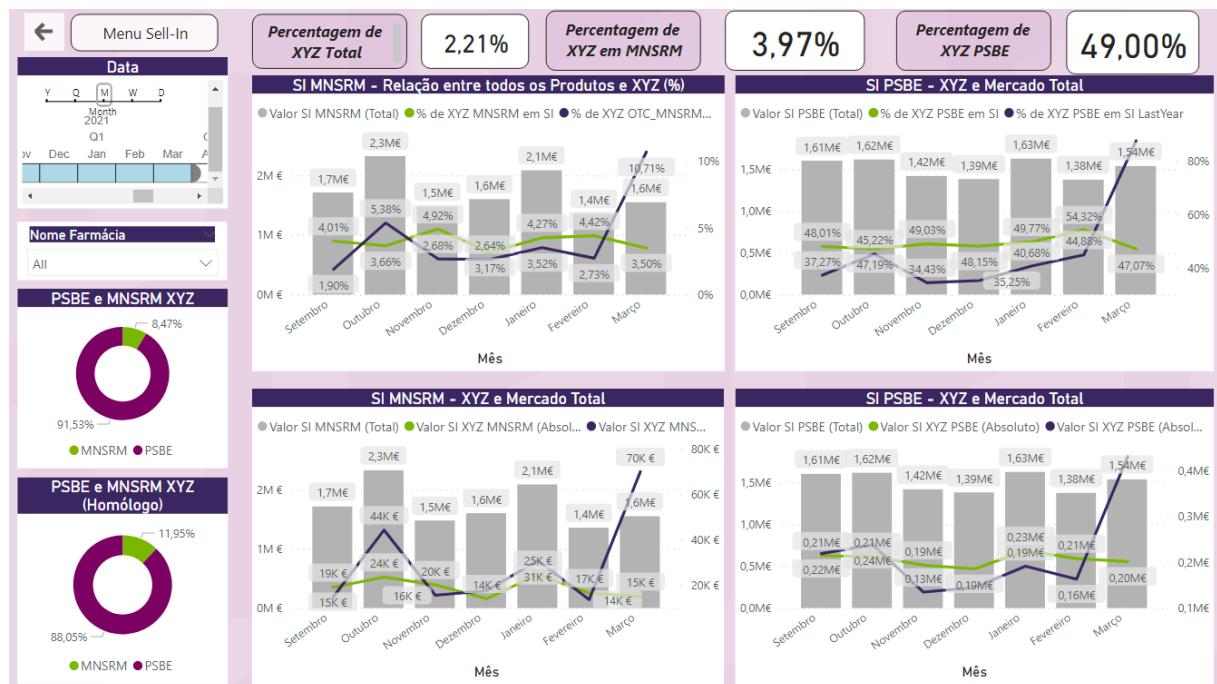


Figure 4.12: Internal Market Share of MNSRM and PSBE

4.7.7 SI vs SO Alignment (SI vs SO)

One of the most interesting pages for the team was the one created for understanding how SI numbers match or not SO numbers, so that conclusions could be drawn about the stock and Purchase Strategy of the Pharmacy. If we take a closer look at Figure 4.13, we can see the Alignment of SI and SO for the whole business, where the user can understand which periods were more critical that might need closer investigation: whether it was a forecasting error by the team or if it the team simply needs to better understand the purchase pattern of the Pharmacy.

Firstly, and addressing the KPI Cards on the top, we can depict right away that from December till April, the number of distinct Products (SKUs) sold in SO was much larger compared to SI, indicating this that Pharmacies might have acquired a lot of stock previously to this period. Moreover, if we head into the graphic below, the user can have a more detailed version of that, where it shows, for example, that the business is having higher numbers as SO than SI through the metric “Escoamento em QTD %” (SO QTD/SI QTD). Interestingly, the green line was added to the combo chart so that the team would have higher depth in understanding in which months there was a higher difference between SI and SO Quantities, in order to find possible outliers (positive numbers reveal more SI and SO quantities of products transacted). As a complement to the 1st graphic explained, on the top half of the page, the team can have a more detailed overview of the actual quantities sold in SI and SO Processes, in order to better investigate on the differences spotted on the first graphic.

On the bottom half, through the matrix created, the team can have a look now on the alignment per Pharmacy, where it is interesting to look at the metric “Meses de Stock Qtd” (Months of stock in Quantity). Through this metric, it is possible to analyze which Pharmacies have a lot of stock to sustain the next months.

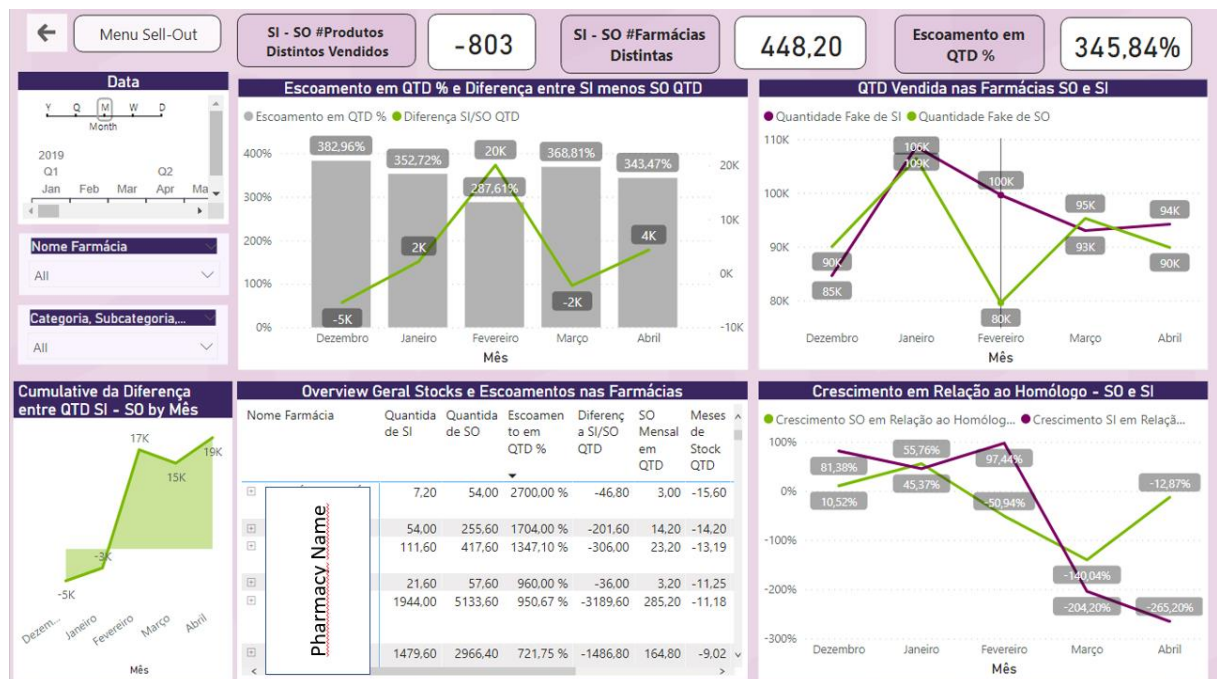


Figure 4.13: SI and SO Alignment of the Whole Business and Pharmacy Specific

4.8 Main Conclusions

Business Intelligence in the Pharmaceutical Industry is vital to be on the forefront of business alignment and monitoring of the best selling and purchasing strategies.

With this tool, the team felt a high increase in the drill-down potential of their Pharmacies and Product Categories, increase in the amount of different scenarios to analyze within the same page (facilitating the process discovery activities) and specially an increase in the filtering and cross-filtering potential for the business, where the user can easily click on a certain graphic and the whole page adapts to that request. Allow of these advantages brings a lot more richness and agility in team meetings, taking away enormous amount of time to prepare and present data that perhaps, could be wrongly done and the meeting was useless otherwise.

As a recap, the process started with the business understanding in terms of what metric were needed, what should be the mindset behind the dashboard and what should be the user experience when navigating through it. This was accomplished not only with multiple initial meetings, but as well as agile sprints of presenting minimum viable products of the Dashboard to the team and asking for improvement points from them. After the prototype was accomplished successfully, the dashboard was improved on its navigation and design, through the application of color codes, more contrast on its design and better placement of the main KPIs and Slicers to deliver a more standard and compact experience to the team.

Lastly, final rounds of feedback were taken and then launched on PBI Service (the online service of PBI enabling teams to use it collaboratively and as final user) so that the whole Operations Department was able to use it.

5 Pharmacy Segmentation - How do our clients perform and under which patterns for Supplier XYZ

With the performance analyzed on the Pharmacies individually, different products and their categories and as well as the business, for the Supplier XYZ, one question arises to complement its analysis and better understand how to exploit the hidden opportunities: how should we compare Pharmacies between them? Should we compare them to their Top 20 in different scenarios? What kind of groups, based on several factors, do we have that we can get together and discover new patterns for the pharmacies?

Thus, and as we could see from the previous chapters, the company across its years, had some Pharmacy Segmentation frameworks that they used to take advantage when doing their reports, however, none of these was based on more robust analysis and marketing theories in order to find relevant business patterns, always only using static targets based on empirical perception of the Pharmacies performances and the already mentioned subscription model in section 3.3.

Therefore, by investing in an analytical discovery, it is possible to detect buying behaviors that were never understood and analyzed profoundly. During this chapter, we will go through the considerations taken regarding the data used from the 172 relevant Pharmacies for the client segmentation, the processes and clustering techniques used, its main understanding and how can the company leverage its results for better on-site actions with both the Pharmacies and the commercial teams.

5.1 Approach to be used throughout the Analytical Discovery

In order to be able to develop a structured development of the customer segmentation initiative, it was decided to adopt the CRSIP-DM (Cross Industrial Standard Process for Data Mining) methodology proposed by [Shearer \(2000\)](#). Through the author, we can unveil the following 6 stages:

1. **Business Understanding:** understanding the objectives on the business perspective and the development of an action plan;
2. **Data Understanding:** focuses on the data collection, description and verification of its quality;
3. **Data Preparation:** focuses on the preparation of the specific datasets to be used correctly by the analytical models;
4. **Modelling:** in this stage, the team focuses on the selection of the data mining techniques for the problem in question. If these techniques require a different data treatment from the one provided in the 3rd stage, the team needs to go back to the “Data preparation” and build new datasets;

5. **Evaluation:** here, the focus is on understanding if the results obtained serve the purpose of the initial objectives;
6. **Deployment:** in this final stage, the team gets together to create a plan to integrate the results of the data mining initiative into the business day-to-day operations;

All these six stages were followed and used within this project.

5.1.1 Business Understanding

When it comes to the company, it was understood and detected that the way Pharmacies were being benchmarked in between them was not the most ideal, causing difficulties in the Key Accounts whenever they wanted to test out new policies for improving the Pharmacies performance or whenever a Pharmacy wanted to check how competitive they were towards others, similar to them. With this in mind, the company wanted to discover the buying patterns of its Pharmacies in order to match those patterns with their potential to buy even more of products from Supplier XYZ. As a result, the different segments found, together with the percentage of Products from XYZ that make up all the purchases value, will be key to define new actions both for the Key Accounts and as well the Operations Department in order to boost each Pharmacy's performance.

5.1.2 Data Understanding and Preparation

As for the dataset used, before being used for the clustering analysis, it went through many stages of preprocessing, necessary to make it viable for the final purpose. Regarding the content of the dataset, it relates to the Pharmacies information on Supplier XYZ Products Purchases from 1st of May 2020 till 30th of April 2021. This was the most accurate time to be analyzed, since if it did not encompass all the outliers existent in the months of March and April 2020 due to the outburst of the pandemic. Initially, the dataset was constructed in PowerBI Desktop, and had the following columns:

1. ANF Code of the Pharmacy: unique national code that identifies the Pharmacy;
2. Day of the Purchase: the day that a transaction/purchase took place;
3. Value of the Purchase: how much money did the Pharmacy spent on Supplier XYZ in that day;

As for this analysis, it was taken the decision that if a Pharmacy bought anything, regardless of if it was 1 product or 100 products with different purchase orders, if all of them were in the same day, it only counted as one transaction. The initial dataset contained 26503 transactions from 184 different Pharmacies and a sample of it can be found in the Table 5.1:

Table 5.1: Sample of the Original Transaction Dataset

ANF Code	Transaction Date	Total Value of Purchase
X1	03/07/2020	24,46 €
X2	15/05/2020	72,41 €
X3	21/08/2020	11,54 €

In order to perform the necessary RFM Analysis and the respective uncovering of even more purchase potential within each pharmacy, it was still needed to go through some more steps, as follows:

1. To eliminate the transactions that were 0 in Purchase Value (this was counted as an Informatic or Human Error in the POS in the Pharmacy);
2. To Filter the Pharmacies to only have the current and relevant ones – Pharmacies from specific not relevant places were excluded and past ones not active anymore as well, leaving us with 172 different Pharmacies;
3. To Match the ANF Code of each Pharmacy to its corresponding district and city, in case this information would be relevant later on;
4. To Match the ANF Code with each Pharmacy Name – not relevant for internal purposes for the teams (as they know the codes), but crucial for benchmarking presentations once in the Pharmacies;
5. Finally, to transform and incorporate the Frequency, Monetary and Recency columns necessary. Frequency is the number of days, among the 365 counted, that purchases were made (Number of Days); Monetary as the total purchase value made by the Pharmacy during the 365 days of the analysis (€); Recency as the number of days since the last purchase until 30/04/2021 (Number of Days). As an example, if a specific Pharmacy made an order on the 27/04/2021, the Recency value would be 3, for this specific case;

A sample of the final dataset can be found in the Table 5.2.

Table 5.2: Sample of the dataset to be used for the clustering step

ANF Code	Pharmacy Name	Frequency	Monetary Value	Recency	District	City
X1	Farmácia X1	20	530,20 €	9	Porto	Matosinhos
X2	Farmácia X2	337	7038,47 €	0	Braga	Braga
X3	Farmácia X3	131	4197,41 €	5	Algarve	Faro

Regarding the relevant dimensions for the specific task of clustering, it is possible to take out some conclusions right away only by looking at a brief summary of the dimensions Frequency, Monetary Value and Recency, in Table 5.3:

Table 5.3: Statistical Summary of the RFM Variables

Variable	Min	1 st Q	Median	Mean	3 rd Q	Max
Frequency (#Days)	2	86,75	145	147,31	209,50	349
Monetary Value (€)	3,1	1939,8	3722,6	5513	6594,1	53288,8
Recency (#Days)	0	0	1	9,884	3	266

As what stands out the most, we can depict that we will find key Pharmacies that display high numbers on Monetary Value (maximum value much higher than 3rd Quartile), being those ones clearly Pharmacies not to lose. Also looking at maximum values from Recency, much higher than the 3rd Quartile, we can already predict that some pharmacies are losing or already lost their interest in the products of supplier XYZ. However, on the frequency, interestingly we see an homogeneous distribution of the phenomenon, relevant later on to be discussed on the tactics the company will use with the results.

5.1.3 Data Exploration and Modelling

After preparing the data and having a glimpse of its characteristics, it is now time to go even further and provide more details, both visually and through more advanced statistics. Every analysis shown in this section was developed in R Studio. Even though in the literature review we cover more than one option to develop client segmentation using clustering techniques, the company felt comfortable in opting for the K-Means algorithm, so it will be the one displayed and used for the results.

Thus, and before showing the segmentation results, it is crucial that we have a sense of the “anatomy” of the dataset, so very briefly we will go through some outliers detected and their importance, the correlation between the variables, the cluster tendency and as well as the normalization of the data.

Variable Normalization

As a first step, the variables were normalized using the min-max normalization method described in section 2.2.2 by the equation 2.2, allowing us to have a better comparison in between them, given their high difference in amplitude between each other. Even though we normalized the variables, the data characteristics remain the same. The results are shown below, at Table 5.4:

Table 5.4: Normalized Data for the Clustering Algorithms

Variable	Min	1 st Q	Median	Mean	3 rd Q	Max
Frequency	0	0,2442	0,4121	0,4188	0,5980	1
Monetary Value (€)	0	0,0364	0,0698	0,1034	0,1237	1
Recency	0	0	0,0038	0,0372	0,0113	1

Correlation Evaluation

Next, after the data normalized, it was time to see how the data correlated, to be able to foresee any particularity from the data that could be useful in the results. Thus, the results were as follows in Table 5.5:

Table 5.5: Matrix of Correlation between RFM Variables

Variable	Frequency	Monetary	Recency
Frequency	1	0.5306	-0.4483
Monetary(€)	0.5306	1	-0.2162
Recency	-0.4483	-0.2162	1

From this correlation briefly, it's possible to observe that frequency is slightly correlated to Monetary and Frequency to Recency, but in a negative way, alarming us for the fact that Frequency will play a major role in segmenting with precision the groups we will form and use tactically. Going even further on the previous sentence, it is possible to see that there's a tendency that the more times a year a Pharmacy orders from the company, the higher the monetary value of it and the lower the time since the last order.

Outlier Detection

By the creation of boxplots of the RFM Variables, as seen above by the interquartile statistical data, it is possible to observe the existence of outliers that will originate relevant segments of Pharmacies later.

As it is possible to depict in Figure 5.1, there are outliers both in Recency and Monetary, allowing us to preview that the segmentation will deliver us Pharmacies that are either inactive/risk of becoming inactive and Pharmacies that represent extreme importance to the business of Supplier XYZ. After thorough analysis, these outliers were agreed to be kept in the analysis as they deliver a realistic overview of the business.

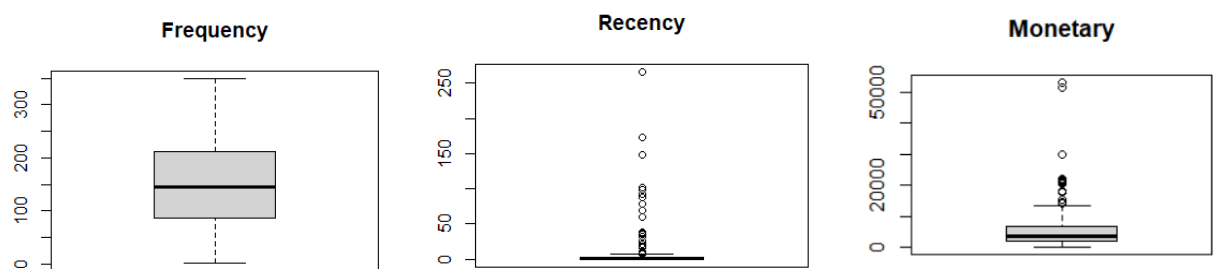


Figure 5.1: Boxplots of the 3 RFM Variables

Remarks on Cluster Tendency

Even though this step was not commercially crucial to continue the analysis or not, it was relevant to be done in order to further understand the data points distribution and assess if the dataset was suitable for clustering. Thus, the Hopkins statistic was calculated and delivered a value of 0,48972, meaning that the null hypothesis is rejected, and the data is slightly clusterable. This might be explained by the fact that the distance of the Pharmacies when it comes to the Frequency variable is not that much.

Development of the Clustering

After making brief analysis of the dataset, it is now time to decide how many clusters (or segments) should be defined. For this, the K-Means algorithm was used and, in order to choose the number of clusters, K, the Elbow Method was the one chosen. Thus, in Figure 5.2, we can see the Elbow curved displayed. This curve explains the percentage of variation as a function of the number of clusters potentially to be used, where we computed a maximum of 10. If we look at the graph, there was two possibilities: either 4 or 6. When deciding with the team, it was preferred to have a more fined tuned distinction of the Pharmacies patterns, specially to detect the more similar ones where “Frequency” was similar, so 6 was the initial intention.

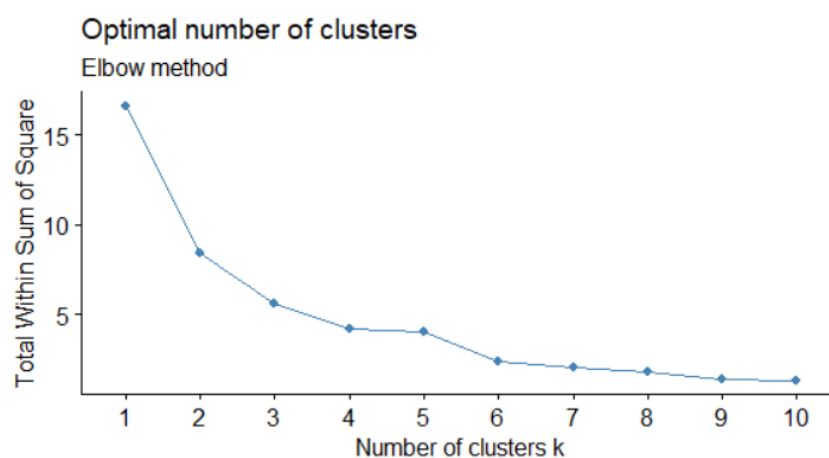


Figure 5.2: Elbow Curve

However, even though with K-Means seems reasonable to accept the 6 clusters as the right number, in order to have a broader feeling on what should be the right decision, the team decided to use two more methodologies to check if the number of clusters (six) was appropriate: the Average Silhouette Method, which aims to maximize the average total within sum of squares and analysis the quality of the cluster decision and the Gap Statistic method, where the estimate of the optimal clusters is the value that maximize the gap statistic, meaning that the clustering structure is far away from a random uniform distribution of points. By looking at Figure 5.3, we can be more critical towards the results of these two:

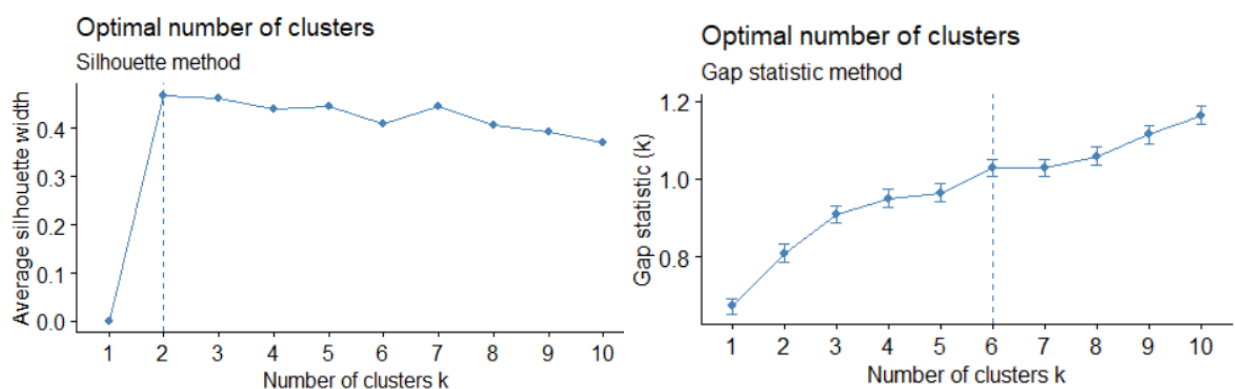


Figure 5.3: Silhouette and Gap Statistic Method Plotted

With Silhouette Method indicating 2 (the quality of 6 clusters is not considerably worse than 2) and Gap Statistic 6, the team felt confident, thus, to advance for a clustering with 6 segments. By conducting a K-Means with 6 clusters, that encompasses 86,1% of the total variance of the dataset. The results of the dispersion of each cluster are delivered in Table 5.6:

Table 5.6: Sum of Square for each Cluster

	C1	C2	C3	C4	C5	C6
Within-Cluster sum of squares	44,79	30,88	19,86	39,19	47,17	48,5
Between SS / Total_SS	86,1%					

The results of the final created clusters using normalized values can be seen in Table 5.7:

Table 5.7: Normalized Results for each Cluster

Cluster	%Pharmacies	#Pharmacies	Frequency Normalized	Monetary Normalized	Recency Normalized
1 – Black	5,23%	9	0.0122	0.0028	0.4666
2 - Red	19,19%	33	0.1154	0.0313	0.0493
3 - Green	1,74%	3	0.6820	0.8425	0.0000
4 – Dark Blue	12,79%	22	0.8478	0.2239	0.0007
5 – Light Blue	37,21%	64	0.3609	0.0732	0.0071
6 - Purple	23,84%	41	0.5930	0.1119	0.0023

However, in order to deliver more real context to the analysis and decisions upon each cluster, the mean values of each relevant dimension of the RFM Variables had to be computed. Table 5.8 and Figure 5.4 show the mean values of the clusters and the Pharmacies graphical distribution in a 3D perspective, respectively:

Table 5.8: Average Values of Each Cluster

Cluster	#Pharmacies	%Purchase Value Involved (SI)	Frequency	Monetary	Recency
1 – Black	9	0,14%	6,22	151,05	124,11
2 - Red	33	5,82%	42,03	1671,91	13,12
3 - Green	3	14,20%	238,67	44895,43	0,00
4 – Dark Blue	22	27,69%	296,18	11935,28	0,18
5 – Light Blue	64	26,33%	127,23	3901,70	1,89
6 - Purple	41	25,81%	207,78	5968,93	0,61

To facilitate the reading of the previous table, it means that, for example for C4 – Dark Blue, it is a cluster with 22 Pharmacies, the money in SI those 22 Pharmacies spend in products of Supplier XYZ represents 27,69% of the money in SI all the Pharmacies analysed spend in Supplier XYZ and, on average, each Pharmacy in this group buys something 296,18 days a year, spends 12000 euros a year and the last purchase was made 0,18 days ago (remember that the values are slightly changed).

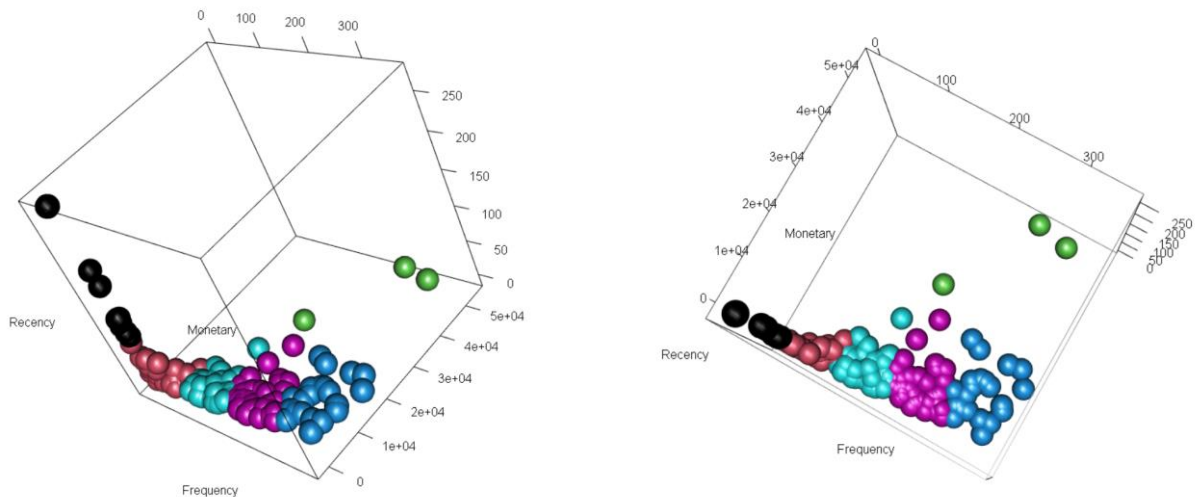


Figure 5.4: 3D Representation of the Clusters and their respective Pharmacies

By looking at the graphical results, it is interesting to observe how important Frequency is not only for the creation of the clusters, but as well as for the overall strategy purchase of the different Pharmacies we are dealing with. The results do not depict a normal RFM distribution of clients/customers, but instead they represent, also by strategic validation of the team, the strategy Pharmacies use to allocate their purchases, something dependent on the conditional rules of the ERP systems or the responsible of the Pharmacy. As an example, Pharmacies from the C4 – Dark Blue have highly automated purchase rules, that whenever a product goes below a certain stock level, it sends automatically a purchase order to the company's central teams, thus increasing not only the amount of times it purchases, but as well, in the long-term, the money invested in products, since it does not allow stock to go below a certain level. However, if our attention goes to C2 - Red, these are pharmacies that have a less rule-based purchase behavior, making them not only slack in the purchase orders, but as well as not having enough relevant stock, many times, to meet client demands, producing, thus, less money at the end and as well as being closer in Recency to the C1 - Black. Moreover, C2 - Red and C5 – Light Blue are characterized by purchasing more products at once as well, in bigger batches.

Focusing now our attention to a more visual perspective on the delimitations of the groups, we can see it through the created decision tree below, on Figure 5.5:

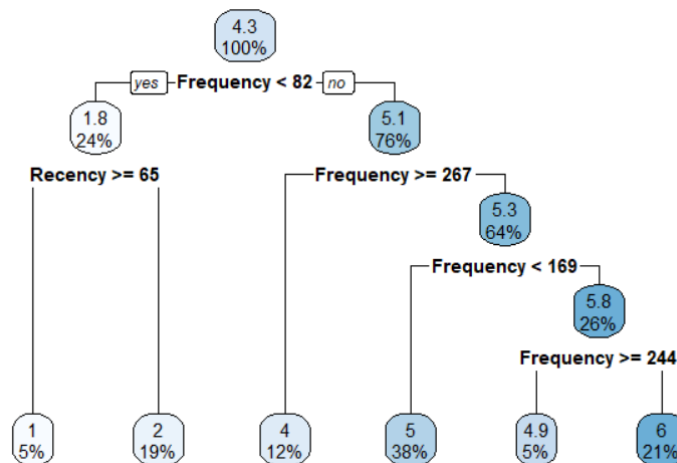


Figure 5.5: Decision-Tree for the RFM Analysis

As we can observe, through another algorithm, the main differentiators on the distinction of the clusters relies deeply on the Frequency patterns of the Pharmacies, which, in its origin, its related to the purchase strategy.

As a final note, we can also see that much of the money for Supplier XYZ products come from C4, C5 and C6 (Column “% Purchase Value Involved (SI)” in Table 5.8), which, by a fact, are the ones better on frequency.

There was a discussion upon whether there was the need to do sub-clusters, but the company felt confident with this one-stage process.

Further Points to enrich the Decision-Making upon the Pharmacies

After the creation of the clusters, questions were made upon what is the next step: “What extra data should we collect, treat or even generate in order to provide greater context when trying to decide and prioritize which Pharmacies to improve first?”

With this, the following stages were defined and brought to life into the next stages of the project:

1. Creation of a Performance Index using the RFM criteria in order to both make rankings of the clusters and the Pharmacies – the company felt this was a great method to find priorities within the list of Pharmacies;
2. Analysis of how much Percentage of Purchases (SI) from each Pharmacy corresponds to XYZ – the lower the Percentage and simultaneously the higher the already Purchase Value of XYZ, the more potential it has to increase its share of XYZ and unlock more Money for the company;
3. Analysis of how much money, in Sell-Out, does the Pharmacy does in XYZ, in order to analyze the potential in sales growth;

As for the CLV (the Performance Index mentioned just before) and being supported by section 2.2.3 of the literature review, the team felt comfortable in using the AHP methodology. With that, different weights were given to the RFM Variables. In order to do so, the team set the priorities in which their business should be run according to the RFM

variables. As a result of the discussion on the pairwise comparison, Monetary was considered of “vital importance” over Frequency (level 5 of the AHP scale), Frequency was considered in between “weaker importance” and “equally important” as Recency (level 2 of the AHP scale) and Monetary, with some error, was considered of “vital importance” over Recency (level 5 of the AHP scale). As a final validation step to make sure the analysis was correctly conducted, we got 4,67% of Consistency Ratio, below 10%, which is the required to proceed with the weights.

With this, the relative weights of the RFM Variables was defined as $w_M = 70,28\%$, $w_F = 18,22\%$ and finally $w_R = 11,49\%$.

In Table 5.9, it is possible to see the Clusters ranked from top to bottom by their CLV results and their respective % of Purchases of Supplier XYZ:

Table 5.9: CLV of each Cluster and % of XYZ in the Purchase Value

Cluster	CLV	% of Purchases of Supplier XYZ
3 - Green	0,83	23,59%
4 – Dark Blue	0,43	13,51%
6 - Purple	0,3	9,41%
5 – Light Blue	0,23	8,27%
2 – Red	0,15	4,82%
1 - Black	0,07	0,56%

Interestingly, it was very insightful for the team to realize that there is also a strong relationship between the % of Purchases of Supplier XYZ and the CLV the clusters get. At first, it seems basic, but it shows that a more penetrated Supplier XYZ is in the purchase profile of the Pharmacy, the higher the CLV, being this a good indication for the Key Accounts Team. We can see the scatterplot developed in Figure 5.6:

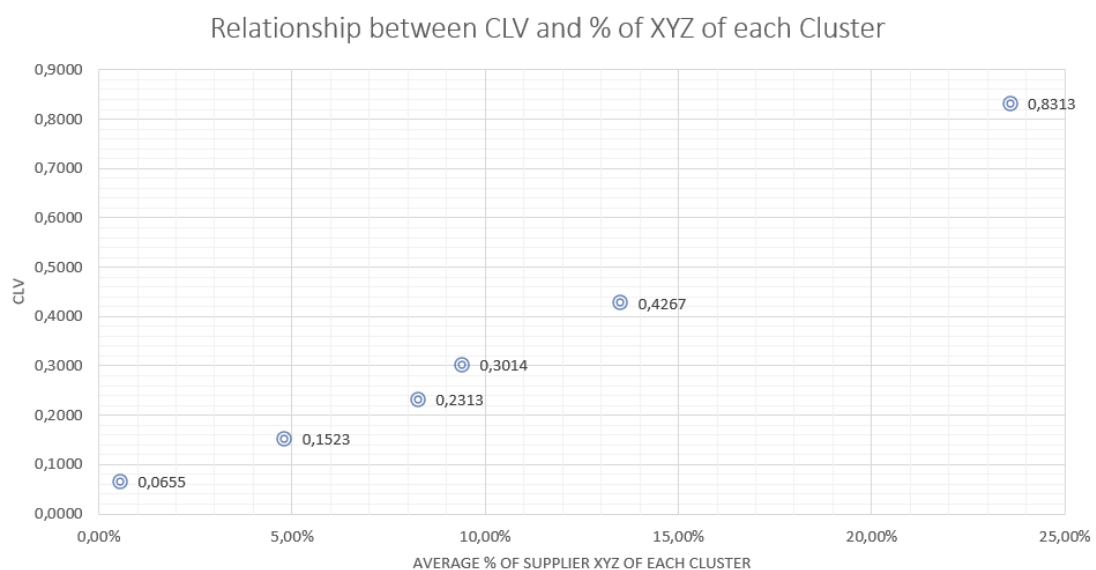


Figure 5.6: Relationship between CLV and % of Purchases of Supplier of XYZ of each Cluster

5.1.4 Evaluation and Deployment - Which Business Strategy to develop afterwards

Before addressing the tactics developed to address improvements in the purchase behavior of the Pharmacies, it is important to notice that there was no Marketing categorization ever done in the company, which means that for comparison terms, there was no previous historical model to compare with (already explored in section 3.3). Even without any possibility of benchmarking the team felt quite confident with the segmentation, as it allowed them, firstly, to identify the Frequency patterns of their Pharmacies.

On to the Business Tactics, there were certain questions that arise from all the information collected, such as:

1. How do we know which Pharmacies to tackle regarding their performance on Supplier XYZ?
2. Once we have the list of Pharmacies to tackle, which ones should we tackle first and how?

With this, 2 new elements were added to the discussion:

1. A scatterplot for each cluster, where we can see each Pharmacy displayed by its % of Purchases of Supplier XYZ in Sell-In (X Axis) by CLV (Y Axis);
2. A summary table displaying the potential money unlocked by following the newly created tactics, based on the RFM segmentation done previously;

As for the Scatterplot, which we can see in Figure 5.7, we need to separate the quadrants with a specification of the limits for each variable (we selected the mean for this purpose). With the mean of each dimension being displayed, it allows us to create 4 Quadrants that strategically deliver us different insights on the Pharmacies and as well as the priorities to address.

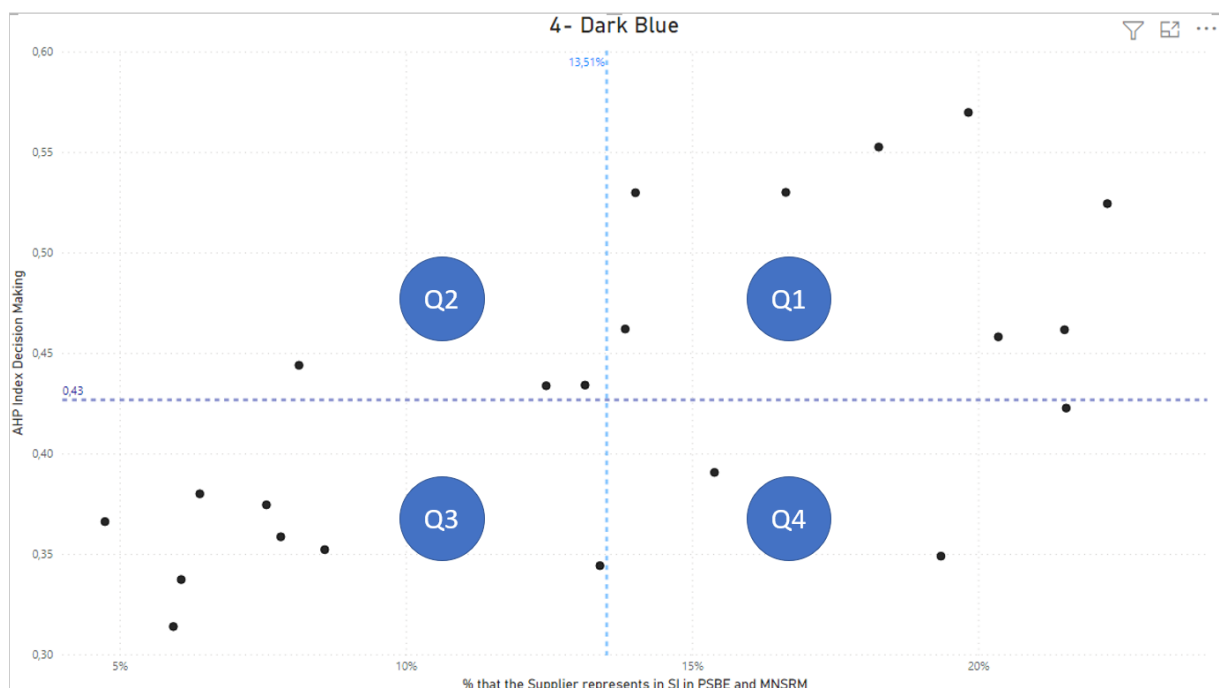


Figure 5.7: Scatterplot of the Pharmacies belonging to C4-Dark Blue, by % of XYZ (X axis) and CLV (Y axis)

By analyzing the quadrants, 4 different Categories of Pharmacies appear, each of them having its own information to deliver for the Business. The order with each of them should be tackled goes as follows:

1. Q2 – “Quick Wins – Question Marks”: Pharmacies with CLV above average, but below average on the % of Supplier XYZ. Thus, it means that the Key Accounts team should better understand which categories they should be promoting and exposing more to these Pharmacies in order to push them to Q1, the best group in the plot. These are called “Quick-Wins”, since they are already familiarized and happy with Supplier XYZ offers, as they have high CLV, but still with potential in some Supplier XYZ categories, being, thus, the easiest ones to further instill more purchases of this supplier;
2. Q3 – “Almost losing Pharmacies”: Pharmacies with below average on both dimensions of the plot, indicating either low knowledge on Supplier XYZ benefits or non-satisfaction with the prices worked with the company. These come in second place as priority, mostly due to their higher difficulty in improvement but great necessity to try to bring them back in terms of customer retention;
3. Q4 – “Possibly not Loyal”: Q4 is the most relevant for the entire Operations Department. These are Pharmacies that are below average on the CLV, but above average on % of Supplier XYZ. This means that they are good clients of Supplier XYZ considering their purchases, but overall they are not buying much from the company as a whole. This is a red flag for the whole department which needs to understand from which other suppliers these Pharmacies are buying;
4. Q1 – “The Best Ones - Stars”: Q1 is the study quadrant. The team felt these should be the last ones to be tackle when it comes to business improvement. They should serve as an example of how the selling process goes right and take conclusions for the others;

With this, the team got a deeper insight on the potential each Pharmacy holds for better sales and which priorities to set. As a Monetary overview of the potential growth in revenue these Pharmacies are capable of unlocking if looking at the Sell-Out, Table 5.10 shows exactly how many Pharmacies we have in each Quadrant of the Clusters and, based on that, how much money more the company could increase in these Pharmacies.

Table 5.10: Overview table of the Monetary focus XYZ Key Accounts and Operations Department could have

Quadrant	# Pharmacies	% Pharmacies	% of Total SI in XYZ	% of Total SO in XYZ
Q1	29	17%	16,13%	17,34%
Q2	45	26%	50,84%	47,66%
Q3	73	42%	21,83%	23,95%
Q4	25	15%	11,20%	11,05%

As we can see, by understanding the purchase behaviors of the Pharmacies, Key Accounts team is able to focus now on 37,96% of SI and 41,29% of SO (Q2 and Q3, most

crucial for XYZ, on both percentages). On a more supplier tactics perspective, Operations Department can focus on Q4, which concerns 11,20% and 11,05% on SI and SO respectively.

As a conclusion, the segmentation was very needed and very welcomed process and result for the company. It allowed the company to understand what were the segments that had more money related to it, what was the specific behavior of the Pharmacies in it and how to characterize each Pharmacy inside each cluster and with it, take action. The future of the usage of this segmentation is to be a living tool, where from time to time it is updated to have a renewed picture of the Pharmacies in each quadrant of its respective cluster.

6 Conclusions and Future Work

The Pharmaceutical Industry is quite competitive, leaving no room for negotiations errors or lack of awareness towards how suppliers, Pharmacies and final clients work and behave. Across the entire project, it was possible to observe how fine margins on portfolio decisions on various suppliers can have big impacts on the long-term revenues and margins of the company. Thus, a complete transparency of the sales performance and the purchase behaviors of the Pharmacies was necessary, not only to deliver new insights and opportunities, but as well as to bring a more consistent data-driven decision-making to the teams that benefitted from the tools created.

Regarding the Dashboard created, this was a tool that brought a lot more flexibility and speed to the Key Accounts Team when analyzing the performance of their products. By using it, the team can now easily spot the best Pharmacies in niche categories, analyze which categories of products drove the business the most and the ones to pay attention to, how did their marketing plans go on for the past months and whether their stock level is acceptable for the whole universe of products. Moreover, the dashboard allowed the team to have a more critical sense about the business, by driving the team to develop more specific targets for historical metrics. Finally, with the dashboard, it is now possible to have an alignment between the whole organization and continuous collaboration, as any member of the Operations Department can have access to it and propose more in-depth business exploration on various matters.

As for the Pharmacy Segmentation based on Customer Segmentation and Clustering Techniques, the team discovered new patterns and ways of classifying their Pharmacies. This not only enabled them to confirm important assumptions about the Pharmacies and their normal purchase behavior, but also allowed them to be able to make more defined business decisions on which Pharmacies to approach first and with which intention. In a more strategic way, it allowed the Operations Department to have a more marketing perspective towards how the data can be treated in order to bring more insights.

When it comes to future work perspectives, there were some opportunities found during the project development, both on the strategic and technical side.

Beginning with the technical side, it would have been relevant to be able to add the clusters created on R Studio to the DW of the company, so that the team would be able to filter their analysis on the created dashboard by the clusters, in order to analyze not only the variables used for the Pharmacy Segmentation, but as well in many others, such SI vs SO Alignment and main categories commercialized. Still on the DW field and Dashboard usage, the team felt the need to, prior to the request of business metrics, understand how much they would like the pages to be dynamic when it comes to Charts changes. This aspect comes since there is still room to reduce the number of pages created if, in the long run, there is a higher coordination with the IT Department for the creation of auxiliar tables in order to provide a more complete slicer and bookmark experience in PowerBI.

Now, on the strategic side, the team discovered many opportunities worth mentioning. Connected to the Dashboard created, it is possible, in the future, to make further product performance analysis based on the BCG Matrix, enabling the team to analyze which products to invest, which ones to discontinue and the ones that they need to maintain, as they represent their biggest and most consistent source of income. Moreover, as well in the dashboard, a lot of discussions went around forecasting various sorts of scenarios. Thus, as a recommendation, it should be possible to allocate efforts to have accurate forecasting methods for the SI and SO Sales and the stocks needed across the year. As for the business day to day life improvement, the team felt the need to map out, through the creation of more decision-trees, many exploration questions that come from the dashboard, similar to the ones shown on section 3.6 of this dissertation. If mapped out, the Key Accounts would have much more clarity on how to discuss each page and how to better navigate from one to another, leaving room as well for non-correct interpretations of the dashboards. Still on the day-to-day improvements, it is still possible to increase the usage of PowerBI Service (the online service that allows the dashboard to be published), where in the short-term there should be programmed alerts in the dashboards delivering direct emails to the teams regarding if a certain critical target was reached, so that immediate action could be taken.

On an extra technological note, and by looking at the whole AI intelligent tools portfolio (Tree Diagrams, Key Influencers, ...) delivered by PowerBI, the team agreed that there should be mapped out more narrowed down questions where the business would benefit in using these built-in solutions of PowerBI. By doing, so it would be possible to explore more root causes to good or bad performance, discover quickly big parts of the business responsible for big revenue slices and even be able to support even better the natural language interface with the dashboards (by the usage of the Q&A feature of PowerBI). However, the big effort to implement these features mentioned is that the company would need to both reflect on a long-term strategy to use the Live Connection to the DW and imported Datasets, as these AI features do not work with Live Connections setup.

Finally, at a Board level, there is the recommendation of backing up these Customer Segmentation and Dashboard Tools with Strategy Maps and Balanced Scorecards frameworks. If used, it would deliver a more sequential and connected mentally between the metrics and customer discoveries performed and the final ultimate financial goals and infrastructure of the company. By adding up a Strategy Map, for example, the company would be able to monitor not only the metrics, but the impact the initiatives had on the metrics and long-term financial results.

All in all, every project stakeholder got very accomplished and fulfilled with the results. The experience was high-opening for the company to bring an even-more data-driven culture inside to it and it allowed to discover many intricacies and technological opportunities within the pharmaceutical sector.

7 References

- (1) Al Shalabi, Luai, Zyad Shaaban, and Basel Kasasbeh. 2006. "Data mining: A preprocessing engine." *Journal of Computer Science* 2 (9):735-739.
- (2) Bauer, Kent. 2004. "KPIs-The metrics that drive performance management." *Information Management* 14 (9):63.
- (3) Bourne, Mike, John Mills, Mark Wilcox, Andy Neely, and Ken Platts. 2000. "Designing, implementing and updating performance measurement systems." *International journal of operations & production management*.
- (4) Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. "The KDD process for extracting useful knowledge from volumes of data." *Communications of the ACM* 39 (11):27-34.
- (5) Goldschmidt, Ronaldo, and Emmanuel Passos. 2005. *Data mining: um guia prático*: Gulf Professional Publishing.
- (6) Han, Jiawei, Micheline Kamber, and Jian Pei. 2011. "Data mining concepts and techniques third edition." *The Morgan Kaufmann Series in Data Management Systems* 5 (4):83-124.
- (7) Huang, Zhexue. 1998. "Extensions to the k-means algorithm for clustering large data sets with categorical values." *Data mining and knowledge discovery* 2 (3):283-304.
- (8) Hughes, Arthur M. 1996. "Boosting response with RFM." *Marketing Tools*:4-8.
- (9) Kaufman, Leonard, and Peter J Rousseeuw. 2009. *Finding groups in data: an introduction to cluster analysis*. Vol. 344: John Wiley & Sons.
- (10) Khajvand, Mahboubeh, Kiyana Zolfaghar, Sarah Ashoori, and Somayeh Alizadeh. 2011. "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study." *Procedia Computer Science* 3:57-63.
- (11) Kotler, Philip, and Gary Armstrong. 2010. *Principles of marketing*: Pearson education.
- (12) LaPointe, Patrick. 2005. *Marketing by the dashboard light: How to get more insight, foresight, and accountability from your marketing investments*: MarketingNPV.
- (13) Lawson, Richard G, and Peter C Jurs. 1990. "New index for clustering tendency and its application to chemical problems." *Journal of chemical information and computer sciences* 30 (1):36-41.
- (14) McNeeney, A. 2005. *Selecting the Right Key Performance Indicators*, Meridium.
- (15) Michie, Donald, David J Spiegelhalter, and Charles C Taylor. 1994. "Machine learning, neural and statistical classification."

- (16) Nainggolan, Rena, Resianta Perangin-angin, Emma Simarmata, and Astuti Feriani Tarigan. 2019. "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method." *Journal of Physics: Conference Series*.
- (17) Neely, Andy. 1999. "The performance measurement revolution: why now and what next?" *International journal of operations & production management*.
- (18) Pauwels, Koen, Tim Ambler, Bruce H Clark, Pat LaPointe, David Reibstein, Bernd Skiera, Berend Wierenga, and Thorsten Wiesel. 2009. "Dashboards as a service: why, what, how, and what research is needed?" *Journal of service research* 12 (2):175-189.
- (19) Payne, Adrian, and Pennie Frow. 2005. "A strategic framework for customer relationship management." *Journal of marketing* 69 (4):167-176.
- (20) Research, Health Market. 2020. Market Watch Portugal. Health Market Research.
- (21) Saaty, Thomas L. 1977. "A scaling method for priorities in hierarchical structures." *Journal of mathematical psychology* 15 (3):234-281.
- (22) Saaty, Thomas L. 2000. *Fundamentals of decision making and priority theory with the analytic hierarchy process*. Vol. 6: RWS publications.
- (23) Shearer, Colin. 2000. "The CRISP-DM model: the new blueprint for data mining." *Journal of data warehousing* 5 (4):13-22.
- (24) Shih, Ya-Yueh, and Chung-Yuan Liu. 2003. "A method for customer lifetime value ranking—Combining the analytic hierarchy process and clustering analysis." *Journal of Database Marketing & Customer Strategy Management* 11 (2):159-172.
- (25) Tan, PN, M Steinbach, and V Kumar. 2013. *Introduction to Data Mining: Pearson New International Edition (English Edition)*. Pearson Education Limited, Harlow, ESX, UK.
- (26) Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2):411-423.
- (27) Weber, Al, and Ron Thomas. 2005. "Key performance indicators: Measuring and managing the maintenance function." *Ivara Corporation*.
- (28) Wind, Yoram Jerry. 2005. "Marketing as an engine of business growth: a cross-functional perspective." *Journal of Business Research* 58 (7):863-873.
- (29) Yigitbasioglu, Ogan M, and Oana Velcu. 2012. "A review of dashboards in performance management: Implications for design and research." *International Journal of Accounting Information Systems* 13 (1):41-59.

8 Appendix A

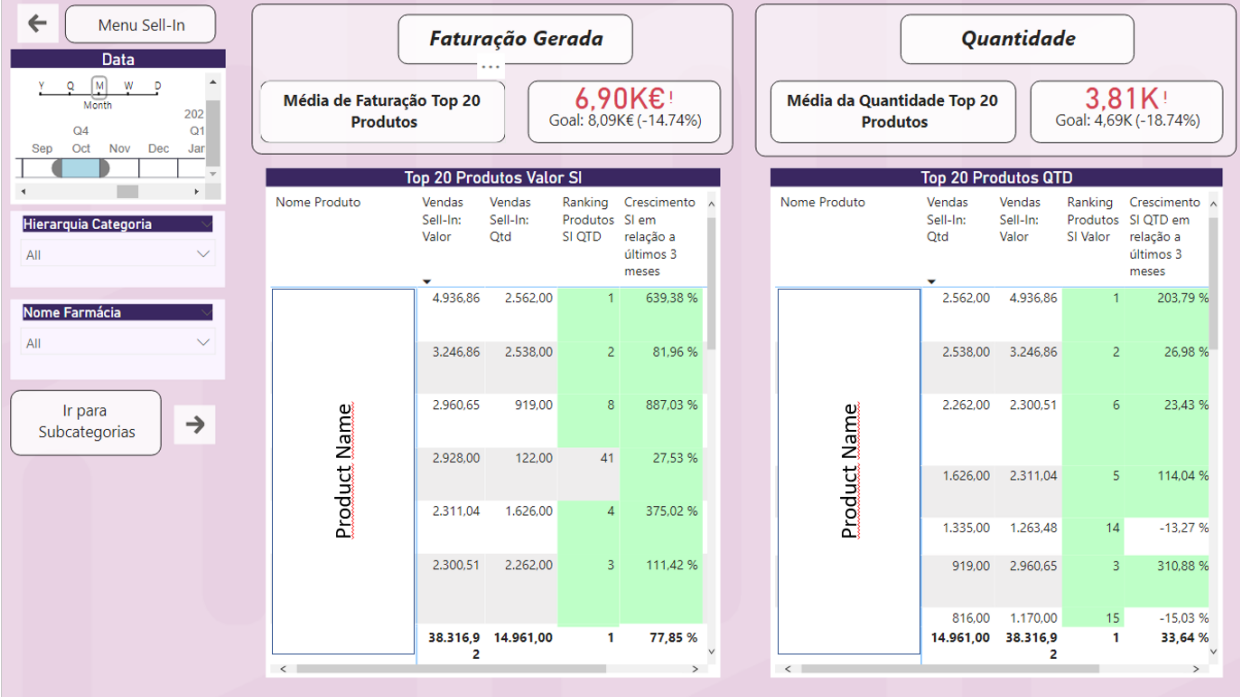


Figure 8.1: Top 20 Products in SI - Revenue and Quantity and the respective average



Figure 8.2: Scatterplots of Revenue and Growth of Revenue compared to last year by Pharmacy, Category, Sub-Category and Product

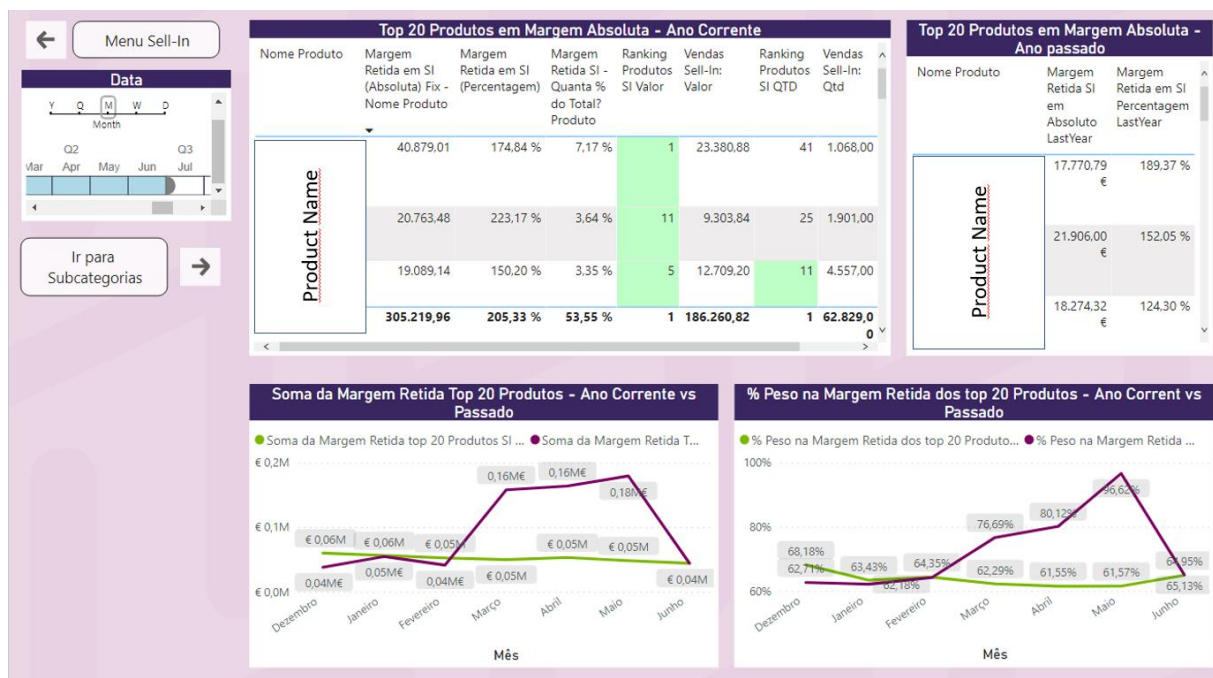


Figure 8.3: Top 20 Products in SI Revenue and their respective sum across time, both absolute and the percentage they represent in total revenue

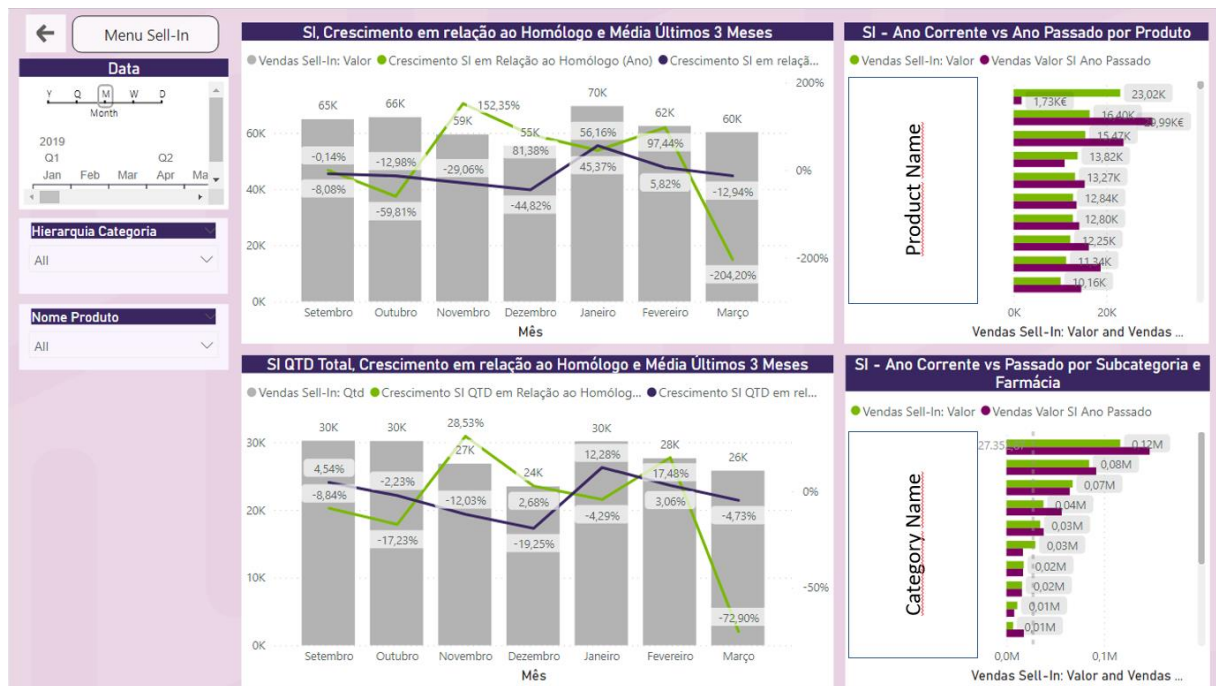


Figure 8.4: Product Growth in Revenue of SI compared to Last Year and compared to average of the last 3 Months (both in Value and Quantity). Product and Category detail on the side for more exploration;

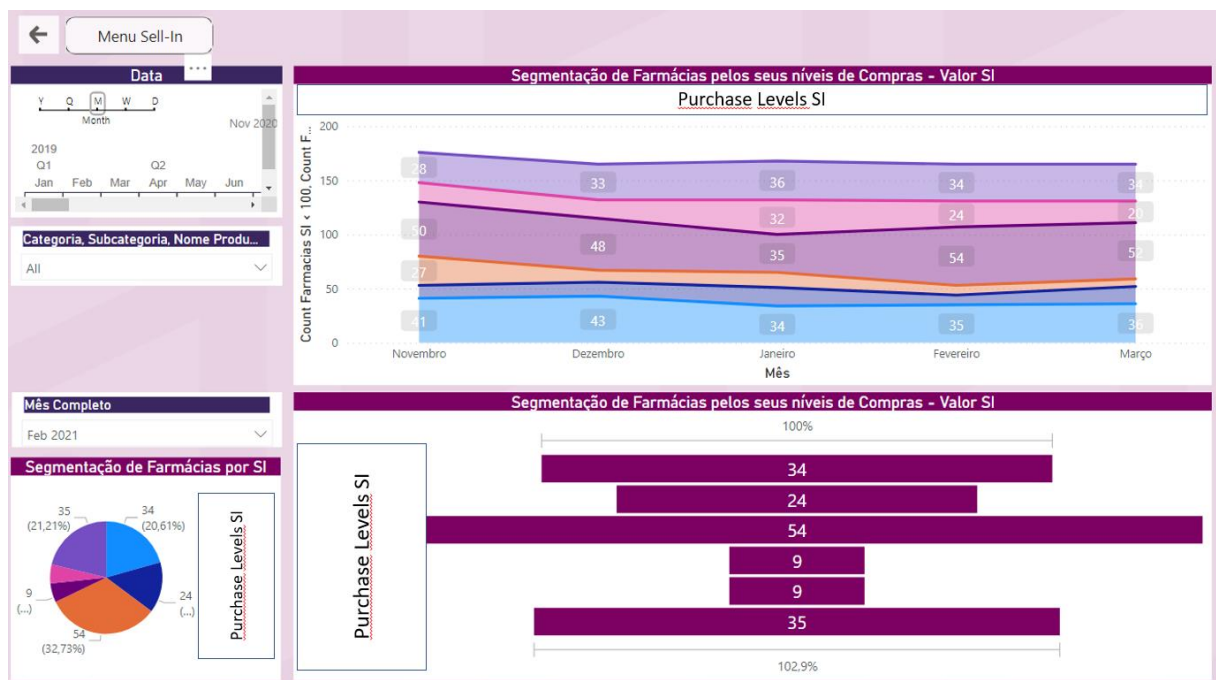


Figure 8.5: Count of Pharmacies with Purchase Levels of SI above a certain target, both for a specific month or across time

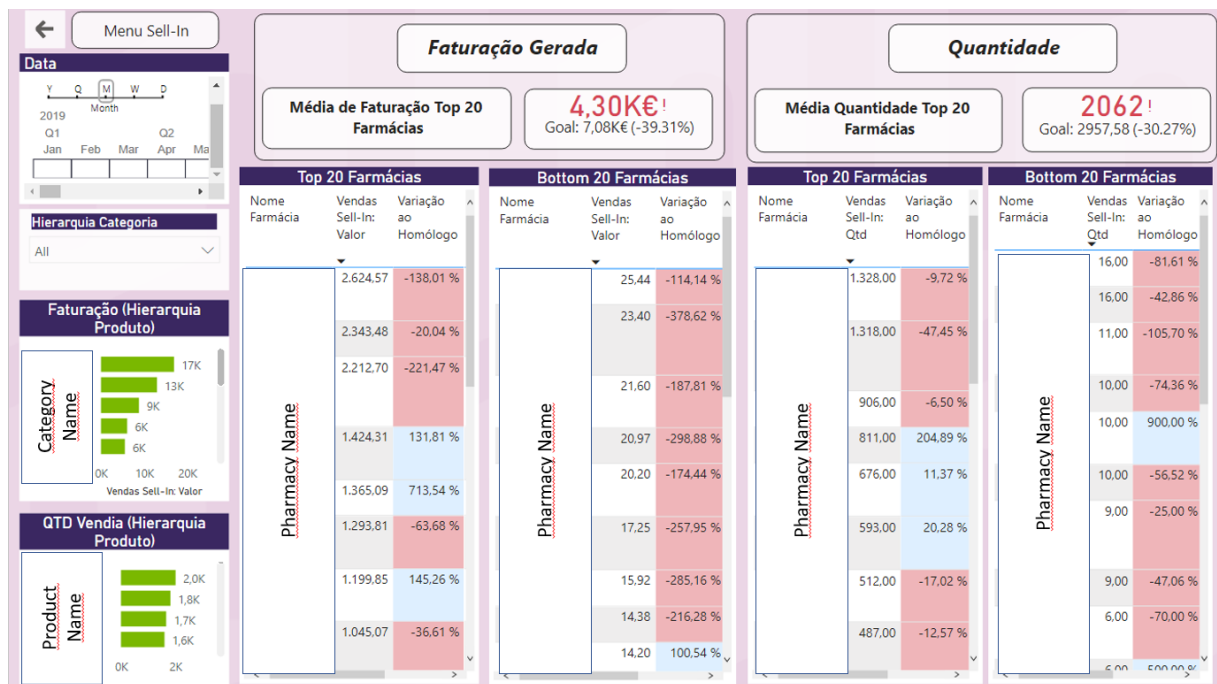


Figure 8.6: Top and Bottom 20 Pharmacies in terms of SI (Revenue and Quantity)



Figure 8.7: Growth of a Category and Product comparing to the average of the last 3 Months

Performance Analysis of Product Portfolio and Customer Segmentation - A Case Study on the Pharmaceutical Industry

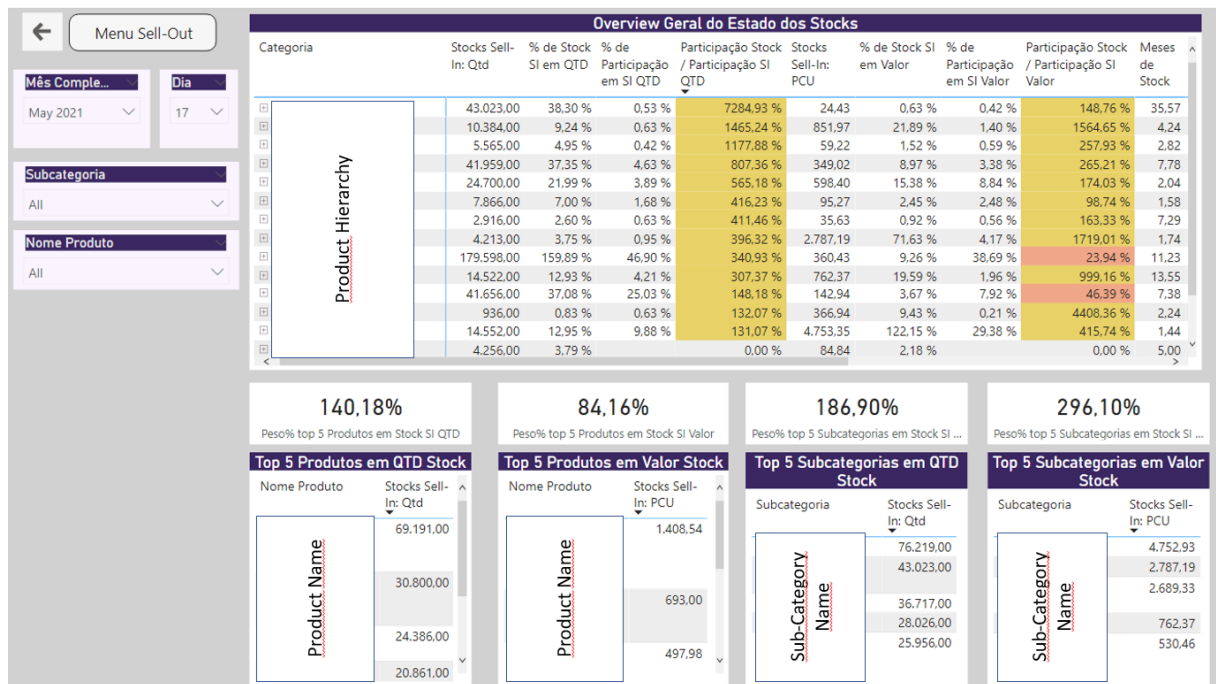


Figure 8.8: Stock Overview of the company, allowing to see which products should we pay attention to