NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

BUSINESS ANALYTICS SOLUTIONS' IMPLEMENTATION IN THE INSURANCE MARKET

Focus on Commercial Property Line of Business

Carolina Quintas Almeida Pina

Internship Report presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics, with a specialization in Business Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

BUSINESS ANALYTICS SOLUTIONS' IMPLEMENTATION IN THE INSURANCE MARKET

by

Carolina Quintas Almeida Pina

Internship report presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Business Analytics

Supervisor: Roberto Henriques, PhD

February 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Carolina Quintas Pina

Lisbon, February 27th, 2023

ACKNOWLEDGEMENTS

To my family and friends, for the constant support and encouragement to succeed in this important chapter of my life, and for always finding a way to comfort me.

To the PBA team, for the opportunity and knowledge shared with me throughout these months, and for welcoming and integrating me with such kindness.

To my mentor, João Oliveira, that never failed to care for and guide me in this project, showing me what real leadership and team-spirit means.

To my dear colleague, João André, for the long hours by my side, continuous teachings, endless patience, and most importantly for all the gentleness and laughs shared.

To my supervisor, Roberto Henriques, for all the help and guidance in this important period.

ABSTRACT

Business Intelligence solutions are becoming increasingly more and more important for organizations as they allow them to gain a deeper understanding of their operations and make more informed decisions. This report will explore the process of creating a Data Mart from start to finish is explored, as well as the consequent Technical Dashboard development for Ageas Seguros.

The Data Mart was designed to offer support, mainly to the company's Pricing and Business Analytics team, where the internship was developed, and occasionally to the Underwriting team. On the other hand, the Technical Dashboard serves a broader purpose, allowing several company departments to easily access and analyze the line of business intended. The benefits and challenges of implementing such a system are discussed, as well as its impact on the company's overall performance. This internship report aims to provide a detailed understanding of all the steps involved in creating and developing both tasks and can serve as a valuable resource for organizations looking to improve their data-driven internal processes and instruments of analysis.

The study finds that Business Intelligence solutions such as Data Marts and Technical Dashboards are two essential complementary tools for companies looking to expand their transversal data structures along with their decision-making processes, while staying competitive in today's business environment. They enable organizations to gain valuable insights from their data, ultimately leading to improved performance, increased efficiency, and overall company growth.

KEYWORDS

Insurance Business; Commercial Property; Data Mart; SAS Enterprise Guide; Business Intelligence; Data Visualization; Technical Dashboards; QlikSense

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BD	Big Data
BI	Business Intelligence
BI&A	Business Intelligence and Analytics
CP	Commercial Property
DM	Data Mart
DSR	Design Science Research
DW	Data Warehouse
KPI	Key Performance Indicators
LOB	Line of Business
NL	Non-Life
PBA	Pricing and Business Analytics
PH	Policy Holder
QS	QlikSense
SQL	Structured Query Language
TDB	Technical Dashboard
UML	Unified Modeling Language

1. INTRODUCTION

In today's fast-paced business environment, organizations constantly seek ways to improve their decision-making processes. One way to do this is by investing in organized and transversal data structures and more beneficial Business Intelligence (BI) tools and solutions. Solutions like Data Marts (DM) and Dashboards are becoming increasingly important for organizations as they provide a deeper understanding of operations and allow for more informed decisions. In addition, solutions like these enable more accurate and timely reporting, which is crucial in a business such as Insurance.

Overall, the focus of this internship was to provide a well-designed data management structure and implement appropriate tools and technologies to ensure that the company's data is properly collected, stored, and subsequently analyzed and delivered to the involved stakeholders.

This first chapter provides an introduction of the company, the team and the internship's objectives. Section 2 presents a Literature Review englobing all relevant topics associated with the project, and Section 3 showcases the framework used, with a higher focus on the methodology adopted. The fourth Section describes the conceptual approach used for developing the internship. Section 5 is focused directly on the DM and Technical Report development, describing in detail all phases and procedures, Chapter 6 displays the main results obtained from the project, and finally Section 7 focuses on the conclusion and main takeaways from the project developed.

1.1. COMPANY OVERVIEW

This internship was held for the multinational insurance group Ageas. The Group has been present in the market for almost 200 years. It operates in more than ten countries across Europe and Asia and provides Life and Non-life (NL) insurance solutions to millions of customers, from Personal to Retail and Business. Ageas headquarters are held in Belgium, where the company is known to be the national leader in the sector. As an insurance company, its primary business model is based on premium charging in return for insurance coverage to their clients. Generally, depending on how often the client activates each insurance policy, the company will produce more or less revenue.

Portugal is one of the countries where the Group operates and its several commercial brands represent the national market's top insurance companies. Its subsidiaries are the following: Ageas Seguros, Ocidental Seguros, Seguro Direto, Médis and Ageas Pensões. Considering the brands mentioned, the main difference lies in the type of product they cover and the type of client they insure. This project was specifically conducted for Ageas' NL insurance area, where different offers, company brands and Lines of Business (LOB) are integrated. The subsidiaries under study in his department are Ageas Seguros, Ocidental and Seguro Direto, where the most relevant LOBs are Motor insurance, Household insurance, Accidents at Work insurance, Commercial Property insurance and Personal Accidents insurance.

1.2. TEAM AND ACTIVITIES

This 12-month internship was developed under the guidance of the NL Operations' segment of the company and resided in the Pricing and Business Analytics (PBA) team, mainly responsible for the development of the pricing rules, adjustments, and value creation. Besides that, it focuses heavily on Business Analytics with tasks ranging from gathering, organizing, and analyzing information, to data

models implementation, development and execution of routines and had-hoc requests, Dashboards and also Technical Reports. Lastly, the PBA team also leads all Portfolio Management necessary for the company's NL LOB mentioned above.

In addition to these concreate tasks, this team also plays an extremely necessary role in providing any technical and/or analytical support to the several NL's strategic departments with their projects. As it is expected, a team like PBA is in constant symbiose with departments like Underwriting, Marketing or Financial.

1.3. INTERNSHIP GOALS

For this internship, the main intent was to study a particular LOB that until this point had not been analyzed in depth by the team, leaving the department lacking. Hence, the focus fell on Commercial Property (CP), a smaller LOB focused on companies' clients.

With this in consideration, the goal of this project was to first develop from beginning to end a data-storage solution to store all CP's existing information. The necessary data ranged from basic key and date variables, very relevant premiums values, and more particular policy variables, including object and coverage information, policy holder (PH) data, commercial structure information, and claims data. As this task had already been developed for other LOBs, the structure's intent was to remain similar, to ensure coherence across data structures. The proposed project had the challenge of including a new data set regarding policies' coverages. The team had never previously extracted nor incorporated this particular information in a data-storing solution, which allowed this project to innovate and provide a new, more complete drill-down analysis, where coverages' information would start to be studied and considered developing Technical Reports.

Once this complex data structure was developed, a second objective for this internship was defined: to create a Technical Dashboard (TDB) showcasing all this information, along with possible trends and business Key Performance Indicators (KPI). This dashboard can be seen as the end goal for this project, as it is the tool that provides the most dignificant global impact for the company. While the DM is a beneficial data storage solution for the PBA team in particular, the TDB is intended to reach several other teams and departments inside the NL Operations' segment of Ageas Portugal. With it, areas like Claims, Underwriting, Marketing and even top Management can access and interact with all CP's relevant information to improve their day-to-day work and consequently their business decisions. To guarantee the best use of the Technical Report as the BI tool that it is, the goal was not to have it display all the information stored in the DM. Instead, the focus here relies heavily on specific business KPIs, some policies' particularities and claims information.

Both phases were set to have a monthly update so that the data could be the most up-to date as possible.

2. LITERATURE REVIEW

This chapter describes a Literature Review including the main concepts, references and techniques applied as a theoretical framework for the project. This section aims to identify and showcase the best practices to develop the project at hands, as well as strengthen the decisions that will be made during its development to reach success.

2.1. INSURANCE BUSINESS

Insurance is a particular type of risk management and its fundamental role is to provide financial protection, offering a method of transferring the risk, in exchange for an insurance premium (David, 2015). Here an individual or organization performs regular premium payments to an insurance company in return for protection against financial losses due to unforeseen events or risks. The types of risks the insurance company assumes can vary from a wide range of options, including health, accidents, property damage, and others.

The main particularities of the Insurance business are, first, the fact that the ultimate/exact cost of an insurance policy is not known at the time of the sale, as most products are sold as a promise for future action if certain events take place during a specified time (Werner et al., 2016). Second, as a consequent response, not all risks are identified as equal; therefore, every individual or company insured will pay a correspondent premium or tariff depending on the presumed gravity of the risk presented (David, 2015), where typically, higher-risk individuals or entities will be paying higher premiums.

Insurance companies use actuarial science and statistic applications to determine the likelihood of a particular event occurring and its potential costs. Actuaries use various rate-making techniques given the particular circumstances, situation and product at hand (Werner et al., 2016). The major challenge here is to construct a fair tariff structure given that different insured risks have different premiums. The key idea behind risk classification is to split an insurance portfolio into classes that consist of risks with a similar profile and to design a fair tariff for each (Antonio & Beirlant, 2006). Also, a usual method to calculate the premium is to combine the conditional expectation of the claim frequency with the expected cost of claims, considering the observable risk characteristics (David, 2015).

Note that the most basic pricing model focuses on the idea that the decided policy price should reflect the costs associated with the product as well as incorporate an acceptable margin for profit, therefore, the initial price defined must be set so that the desired profit per unit of product will always be achieved (Werner et al., 2016).

2.1.1. Commercial Property

As mentioned previously, this project focuses on the Commercial Property insurance LOB. This type of property insurance has been imposed in the insurance and reinsurance markets to keep large multinational corporations within the traditional insurance market (Hobbs, 2020). This insurance provides coverage solutions to businesses and organizations for damage or loss of their physical assets, such as buildings, equipment, inventory, furniture, etc.

CP insurance policies can be tailored to meet the specific needs of different businesses, including small businesses, large corporations, and non-profit organizations. Premiums are typically based on the level

of risk associated with the business, such as the type of industry, the location of the property, and the amount of coverage needed.

For Ageas' solutions, the CP LOB can include two main types of secured objects – building and content – and these can be selected separately or in an aggregated format, where Ageas is responsible for securing both. The latter option is usually the most common. Each of these product's objects can be covered by a variety of possible coverages, given the various risk scenarios considered and insured. In Ageas, fixed and mandatory coverages are always included in both objects, which tend to be incapsulated in a 'Base Coverage' format. This is typically true for older products and basically, within the official systems of the company, the 'Base Coverage' is used in the policy creation phase as a single coverage that includes all correspondent coverages. These coverages vary given the product and the object at hand. All the additional coverages not included in the 'Base Coverage' need to be specifically required by the client to be added to the final policy offer.

2.1.2. Elementary Insurance Concepts

As a complementary segment, some of the most relevant basic concepts associated with the Insurance business are listed and described below (Werner et al., 2016):

- **Exposure:** the basic unit of risk that underlies the insurance premium.
- **Written Exposure:** the total Exposure developing from all policies issued in a particular period.
- **Earned Exposure:** the portion of the Written Exposure from which coverage has already been provided, given a certain point in time.
- **Premium:** the monetary amount the insured person pays for their insurance coverage.
- **Written Premium:** the total Premium associated with the policies that were issued during a particular period.
- **Earned Premium:** the portion of Written Premium from which coverage has already been provided, given a certain point in time.
- **Claim:** When the PH requests a compensation to the insurance company over a loss or damage covered under their insurance policy. After this request it is up to the insurance company to accept or deny it, in case it is determined that the policy does not cover the conditions of the incident. Only after it has been approved does it became an official Claim to the company.
- **Claim Cover[1]:** A single Claim can contain more than one Claim Cover. Suppose a particular event causes damage in more than one of the policy's coverages – an extreme raining incident can cause both flooding and machine breakdown, which can be both included into one policy – one specific Claim identifiers will include more than one Claim Covers identifiers.
- **Vigency[1]:** Vigencies are the measures used to identify an active period of a policy. By default, a Vigency corresponds to a full annuity (a complete year between the policy's start and end date); however, there are several cases where this period can be shortened. This usually occurs when a policy is cancelled before the end of its annuity, or also when the premium associated with the policy or relevant PH's information is altered mid-annuity. This concept developed by the company allows one policy's annuity to include more than one single vigency.

[1] Concept nomenclature developed particularly for Ageas.

2.2. BUSINESS INTELLIGENCE

BI has been a present and relevant topic in the last decades for any company. As data continues to be generated and stored at a rising volume, consequently creating a fast propagation of huge amounts of unfiltered information, organizations are now spending considerable resources on the tasks of managing and properly utilizing this data (Sidorova & Torres, n.d.). Given this Big Data (BD) reality, significant attention has been shed upon BI and its vast range of solutions regarding companies' management (Liang & Liu, 2018). Theoretically, BD is the term used for collecting of large and complex, organized, or unorganized, data from various sources that tend to be difficult to process using only traditional data management and processing applications (Huang et al., 2017).

In addition, according to the annual Society for Information Managements' IT Trend Study for 2021, Analytics/Business Intelligence/Data Mining/Forecasting/Big Data Information Technologies are in the Top 3 Largest IT investments verified within the sample examined. In addition, these are also identified as the number one technology category that should receive more investment (Kappelman, 2021). This study is developed annually, and these results have remained constant throughout the years.

2.2.1. Business Intelligence and Analytics

In the last couple of years, the new integrated concept of Business Intelligence and Analytics (BI&A) has emerged, and it is recognized as an important topic of Information Systems research with significant implications for businesses and practitioners. BI solutions that are able to apply Data Analytics to generate key information are inevitably a great and powerful resource in any company to support a data-driven decision-making processes (Liang & Liu, 2018).

BI is usually characterized as a process that incorporates the procedures and tools used to develop and retrieve information. This information is also identified as an element of BI and is the key factor when it comes to decision making. Generally, the most promoted definition for this concept is the "umbrella view", in which BI represents a collection of information-management technologies combined with information-seeking activities (Sidorova & Torres, n.d.). Following this view, Business Analytics is one of the several components of BI and it is seen as a tool for information's identification and development. Its main goal is to identify strategic business opportunities while improving business agility and gaining competitive advantage.

With this said, BI&A focuses on the development of technologies, systems, practices, and applications to analyze critical business data generating new and useful insights (Lim et al., 2013). These insights are designed to support reporting and decision-making, consequently also being used for attaining a better operational efficiency and improving services and products.

Globally, BI&A technologies facilitate data collection, analysis and information delivery and are designed to support decision-making processes while creating values for the organization. The four technological elements comprised in a BI&A application are Infrastructure, Data Management, Data Analysis, and Information Delivery. These are all integrated within each other, as the final and complete value of BI&A can only be effectively reached once all four elements work subsequently and in harmony. These building blocks are often referred to as the "architecture stack" of BI&A (Rikhardsson & Yigitbasioglu, 2018).

2.2.2. Business Intelligence Architecture

Architecture is the fundamental structure of any system. It tends to be embedded in the elements themselves, the relationship among the several elements, its design and evolution, and the visible properties such system can provide (Hightower & Shariat, 2007).

A solid architecture is a key factor for companies and organizations wishing to have better control over the implementation process and operation of their BI environment. If the architecture is not designed correctly, the chance of inconsistencies arising among the different components increases and the inability to share correct information, meet business requirements and provide good quality business performance (Ong et al., 2011). In these scenarios, companies may be unable to maximize their intended value production from their BI investments.

There exist several proposals of BI Architecture available in the literature; however, for this project the focus was on five main components that should always be considered in a BI environment: Data sources, ETL systems, Data Warehousing solution, End-user and lastly Metadata layers (Ong et al., 2011). The proposed architecture can be found in Figure 1 and below a description of its elements.

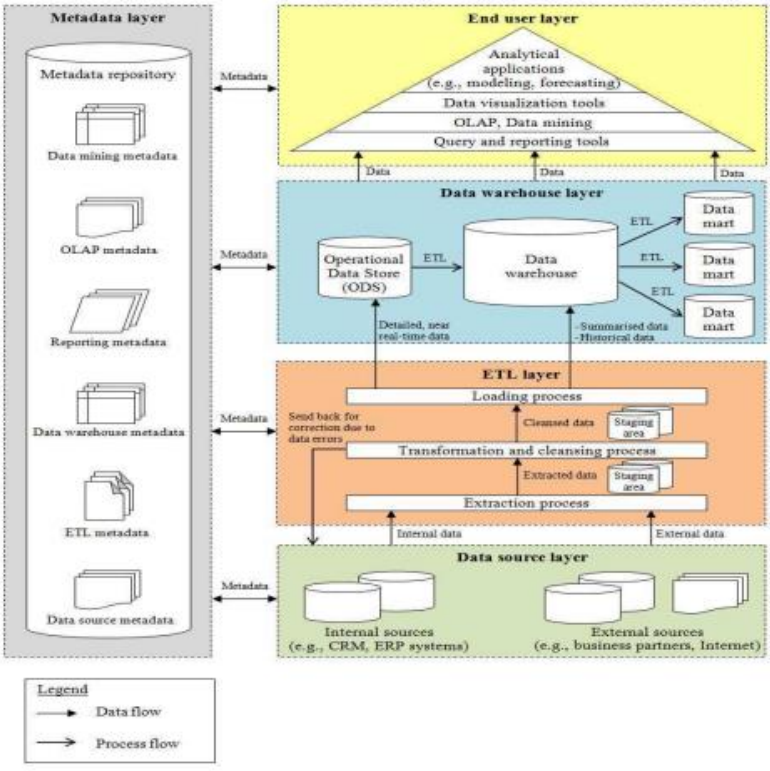

Figure 1 - Proposed BI Architecture (Ong et al., 2011)

I. Data Sources: The data available nowadays is not always structured and can come from internal or external sources. Internal data sources include data present in the company's operational systems, while external data can be collected from a wide range of possibilities outside the company's scope. The key here is to clearly identify which information is necessary and where to retrieve it from.

II. ETL: The process of extracting, transforming and loading data. The extraction of the raw data will vary depending on the type of sources available; its transformation will include fundamental

modifications as well as cleansing, aggregating, summarizing, integrating, and other forms of treatment; and lastly the load of the finalized data will be done into a data warehousing.

III. **Data Warehousing solution:** This layer can be composed of three different components – Operational Data Stores, Data Warehouses (DW) and/or DM. They each have their own particularities, unique purposes and benefits when used in a BI solution, and these should be considered when deciding the approach to follow.

IV. **End User:** This segment can be considered the final major component of a BI implementation. This layer consists of tools and BI applications that allow all the information to be displayed in different formats to different users through various analytical and reporting components. As seen in Figure 1 above, these can include data mining, predictive analytics, and data visualization solutions.

V. **Metadata layer:** This last layer is considered to be a complementary structural element that has recently been gaining appreciation within the BI architecture. Metadata is known to be data about the data, and its repositories and consequent management provide unique support to any BI solution, as it offers business users the chance to store and standardize relevant project's metadata across different systems. A metadata repository is used to store technical and business information about data as well as business rules and definitions. The tracking and monitoring of this information within the BI environment suggests better project management, a smaller development time and more straightforward maintenance of the entire architecture.

2.2.3. Data Warehouses and Data Marts

Data repositories have gained tremendous importance throughout the years with the emergence of large institutions and the fast growth of BD. It is now seen as a essential requirement to be able to find any data related to a specific subject in an organized data storage solution (Khalaf et al., 2021). As mentioned previously, these solutions can present different formats. In this segment, the most relevant concepts given the project developed – DW and DM – are described, focusing on their main particularities and key characteristics.

DW and DM are two concepts that are inevitably connected as they share the same base function and goals. A DW is typically associated with the baseline concept between the two, and it is characterized as a logical collection of non-volatile and integrated information, whose objective is to support management's strategic decision-making (Khalaf et al., 2021). The information is usually gathered from many different operational data sources, and it is treated to provide a coherent picture of the business conditions given a single point in time (Hightower & Shariat, 2007).

When studying a DM, is usually described as a subject-oriented decision support system generated from a pre-existing DW, that tends to have more limited use as it analyzes particular information from a specific area/segment of the business (Khalaf et al., 2021).

However, the relationship between the two concepts is not that straightforward. In reality, there are dependent and independent DM. If a DM is, in fact, a subset of the DW that is a relatively simpler and inexpensive platform, closer to the end user with periodical updates from the central DW, it is identified as a dependent DM. On the other hand, if the DM structure is created separately from the existence of a DW and involves a direct extraction from the sources of information, it is seen as an independent DM (Firestone, 1997). Typically, these share the same structural integrity as a DW and can be seen as a smaller DW that provides a dimensional view of a specific business function. In some cases, an official DW can be developed from a series of integrated DM's (Cabibbo & Torlone, 2004). Below you can find the visual representation of these two distinct concepts.

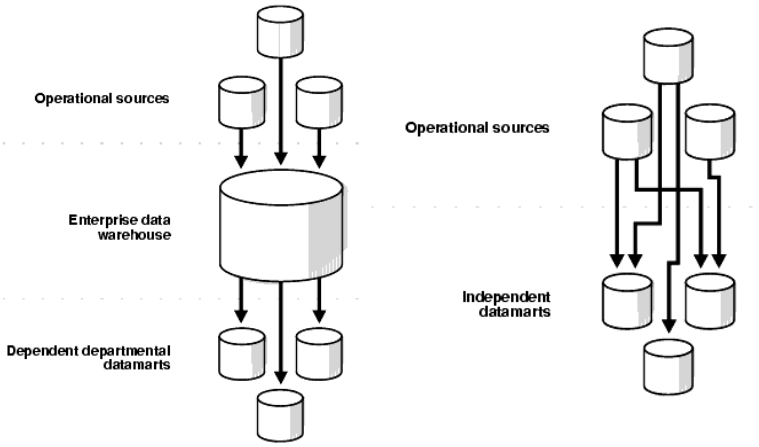

Figure 2 – Dependent vs Independent Data Mart structure

In addition, it became noticeable for many organizations that to directly implement a DW was a task that tended to be more complex, costly and with longer development periods (Firestone, 1997). Such findings also helped promote the development of independent DM.

2.2.4. Data Visualization and Reporting

In any company or organization, executive decisions are the core components affecting its growth. Because of this, it is mandatory that companies have an effective tool that measures and monitors the growth of the business and showcases all its trends and key performances. To be able to do so, the ideal proposal involves BI technologies, data mining techniques and data visualization technologies (Kumar & Belwal, 2018). In this segment, the focus is on the data visualization element.

Data Visualization is the graphical representation of information. The general process focuses on incorporating effective representation strategies to integrate, unify and standardize data from several different sources (Naidoo & Campbell, 2016). Its major objective is to provide the user/organization with a direct qualitative and easy understanding of the information content, usually in a dashboard format. In a complementary fashion, Data Visualization is used to aid analyses while also being an effective tool for communication (Evergreen & Metzner, 2013).

Because visuals and dashboards are both an increasingly relevant form of science communication, the need to properly deliver these elements, with a suitable understanding and interpretation of the data, is essential (Midway, 2020).

According to Shadan Malik there are five underlying elements that are essential for the success of any Enterprise Dashboard. These can be identified by the acronym SMART (Malik, 2005):

- **Synergetic**, meaning that the dashboard should be ergonomically and visually effective in the way that it contributes to the user, with the information displayed in each screen.
- **Monitor KPIs**, as it must showcase critical KPIs related with the business being studied, given that these are extremely effective and necessary for any decision-making process.
- **Accurate**, where the information being displayed must be entirely correct and validated, so as to ensure full confidence from the user in the dashboard being presented.
- **Responsive**, meaning that ideally it should respond to predefined needs, while also drawing attention to critical topics, through alerts, pagers, or other formats.
- **Timely**, as it must always present the most current information possible. The information must be in real-time, allowing for the most effective analysis possible.

In addition, Malik also states that a SMART dashboard is not sufficient to ensure effective organizational management. To complement this view, he suggests that the ideal dashboard should also include the following advanced elements: be interactive, provide more data history than only the present view, be personalized, be analytical, allow for collaboration amongst users and developers, and lastly, be trackable, allowing the user to customize the metrics they wish to follow (Malik, 2005).

It is worth noting that, despite the guidelines shared above, there is not a unique best version of a given graphic/figure. In reality, there may be several effective solutions for displaying a particular piece of information, and it is part of the developer's job to weigh the advantages and disadvantages of each, hopefully reaching the ideal solution given the scenario at hands (Midway, 2020).

3. PROJECT FRAMEWORK

This section focuses on the approach selected to guarantee the project's final success. First, the methodology followed will be described, along with its main structural components' listing and description. The main tools used to reach the internship's results will also be presented.

3.1. METHODOLOGY

The methodology used for this project was Design Science Research (DSR) for Information Systems. The main goal of this model is to enhance human knowledge by creating innovative and successful artifacts, providing solutions to real-world problems (vom Brocke et al., 2020).

DSR is an iterative process that involves developing an artifact based on experience and demonstration. It begins with the identification of a concrete, and typically complex, problem within a problem space (Carstensen & Bernhard, 2019; Goecks et al., 2021). In a cyclic manner, the DSR process generates, designs, and evaluates possible effective technology-oriented design products. These are typically presented in the form of models, methods, constructs or instantiations (Achampong, 2017). DSR is a process that requires several repetitions; therefore, the evaluation of the proposed solution cannot be the final step. To develop a successful project, five complementary phases need to be considered and incorporated: awareness of the problem, suggestion, development, evaluation, and conclusion (Mdletshe & Oliveira, 2020).

Bellow it is presented the schema for the DSR Process Model, and following the figure, each corresponding element of the process is explained with a description of its application in this project:

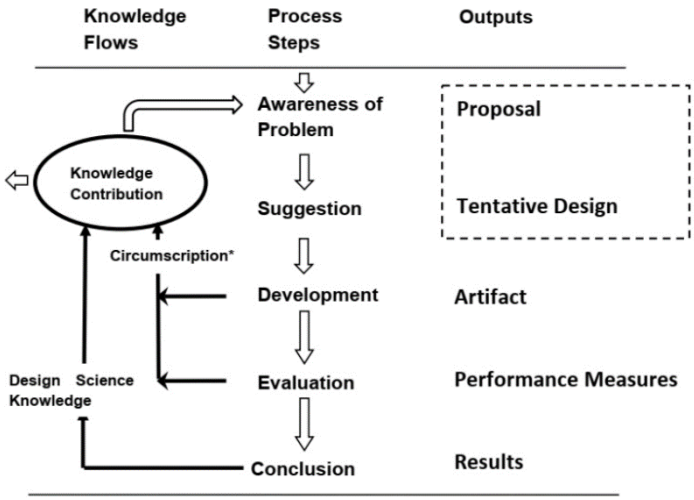

Figure 3 – DSR Process Model

I. Awareness of the problem: The amount of daily information produced in an insurance company is astronomical, so it is key to have an organized and well implemented data storage structure paired with meaningful reports that can provide knowledge to the various stakeholders. The motivation for this project was to develop and present such results to the company, focusing in the product area of CP, and in this way offer to the several teams involved the tools necessary to quantify and analyze the various business KPIs and analysis measures.

II. **Suggestion:** The proposed solution for the problem identified earlier was the creation of an independent DM and the consequent development of a TDB with complementary views regarding the LOB under analysis. With this suggestion, the team would be able to both have access to all the relevant and necessary information regarding CP by accessing the DM, while also be able to evaluate and retrieve more concrete business information from the Technical Report.

III. **Development:** It consists of two separate and sequential implementations, according to the proposed plan. The first phase consists of the transformation and centralization of all information regarding the company's portfolio for CP into one single data structure, whereas the second phase corresponds to the development of a final TDB, using an adapted summarized table from the DM previously created.

IV. **Evaluation:** This stage was recurrent throughout the development of the entire work, since constant validation was paramount to ensure the integrity of the information presented, which came from various sources. In addition, before the creation of the TDB, a more thorough evaluation took place to guarantee the data in the DM was aligned with other important and reliable company's sources, in particular financial sources. Finally, the final dashboard was also assessed to understand if the requirements and aims were fully met, while also measuring the added value such report brings to the team and department.

V. **Conclusion:** The last phase focuses on presenting the results to the teams involved with the LOB and collecting their feedback.

3.2. TOOLS AND TECHNOLOGY

For the proper development of the internship different software solutions were used, according to the project goals and the company's existing frameworks. Each technology was applied in a different phase of the project, as they are more appropriate for distinct, yet complementary, tasks. Below, a brief explanation of each software used can be found.

3.2.1. SAS Enterprise Guide

SAS Enterprise Guide (SAS EGuide) is a graphical point-and-click user interface that allows the user to retrieve, report and analyze statistical data without the requirement of knowing how to program in SAS (Meyers, Gamst, & Guarino, 2009). It can be connected to SAS on a local computer or via SAS Server.

This application provides an instrument to conduct analysis with a self-service environment specialized in workflow-based projects, offering a centralized system for managing access, while also enabling self-sufficient access to enterprise data sources for business analysts and programmers. Moreover, it allows the user to easily perform tasks like data preparation, data exploration and report generation.

For all these reasons, SAS EGuide is one of the main tools used transversely by the Group's analytical teams. The company uses a variety of SAS EGuide libraries to store all types of historical information, which varies from data extracted from the company's software solutions and systems to tables and projects developed within the application. Therefore, SAS EGuide was the main tool used in this internship, and it allowed to successfully reach the project's first goal of developing a CP's DM. This

new data storage structure needed to be updated monthly to feed the TDB developed in the second phase of this project.

Finally, it is relevant to mention that the main tool used in this application was 'Query Builder', a SAS component that by mandatorily retrieving data from one or more data sources, allows the querying of data. This querying includes, amongst other things, the specification of columns to consider in the analysis, the computation of new columns, the replacement of values and the introduction of conditional formats to the analysis. In addition to 'Query Builder', Structured Query Language (SQL) was also heavily used to query diverse programs and tables. Generally, and for this project in particular, SAS SQL procedure primarily enables the following procedures: retrieval and manipulation of stored data, creation of tables, development and use of SAS macro variables, insertion and removal of rows, and addition or modification of column's data values, among others.

3.2.2. QlikSense

While different departments and areas of the company use distinct BI visualization tools, the PBA team produces the majority of the TDB and Reports with QlikSense (QS). For this reason, this was the selected tool to develop the second phase of this internship.

QS is essentially a data visualization product that allows the user to create flexible, interactive visualizations that lead to meaningful and supported decisions, and offers an intuitive development interface that allows for free discovery and exploration of the data (Troyansky et al., n.d.). The base model of this product is their App, which consists of one or more sheets with a single or several visualizations. Each developer can create its own QS application, modify it, reuse it, and ultimately share it with other users at any time. Once the selected users have access to the applications, they are able to view the data, navigate through the visualizations and ultimately analyze the information adequately.

The biggest advantage extracted from QS is the effective monitorization of the monthly results and KPI's. With QS it is possible to always provide to the company a clear, objective and updated view of the data since any action can be triggered immediately upon the developer's request. This process not only saves time to the company as it delivers real-time insights by demand, but it also provides more cohesive and consistent analyses throughout the months, improving the decision-making process. Globally, with QS it is possible to accelerate business value as it allows the company to respond faster to events and trends in their own business.

4. CONCEPTUAL APPROACH

In this chapter, the Conceptual Model developed given the methodology mentioned previously is presented. The focus here is to explain the entire process flow that allowed this project to be completed successfully.

This includes the data retrieval process, the Conceptual Data Model in which the data was assembled, the DM envisioned structure and composition, the main KPIs and business measures considered for the project, and lastly the development of the data visualization solution.

4.1. DATA STRUCTURE

Since a major goal for this project was to collect, in one single data structure, all relevant information regarding the CP, this LOB's portfolio needed to be retrieved. Before this could be implemented, the first necessary step was to understand what information the company owned and wanted to have stored and presented for regular consultation, both in a table and dashboard format. Secondly, it was essential to identify its current structure within the company's existing infrastructures.

This was accomplished with the support of managers and future TDB's end users, those with the highest knowledge and sensibility regarding the LOB. Once this initial gathering was done, it was possible to determine that the necessary data fell under two main categories:

- The information generated by the company, already present within its systems, either in available databases and tables and/or as internal documents with complementary information, mainly in Microsoft Excel. Below you can find a descriptive mockup of the Data Flow scenario present in this particular project:

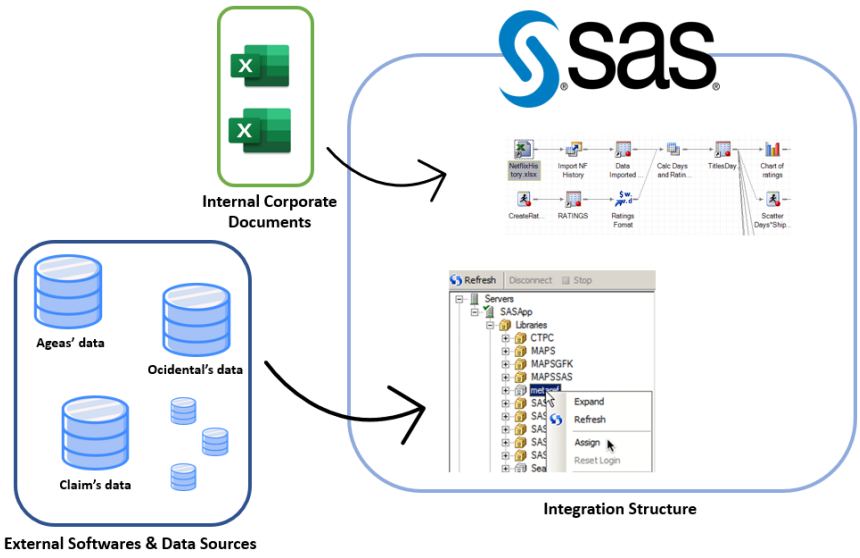

Figure 4 – Project's Data Flow Representation

- The information that needed to be produced for this project. This consisted of specific variables specially requested for this project and represented a significantly smaller percentage of the total data necessary to consider. In fact, most of the relevant information

originally intended to be stored in the new data structure is obtained and/or produced by the various systems and applications used by the company.

As seen in Figure 4, for the two brands that needed to be included in the study of this LOB – Ageas Seguros and Ocidental – there is a total of three different main applications collecting daily business data. Fortunately, the company already uses SAS EGuide and its libraries as a unifying structure to save all the dispersed information originated from the different software systems referred above, in table formats accessible to all employees with the appropriate accesses.

Although there was not a big variety of data sources' systems involved in the project, there were still more than fifteen different SAS tables and a few complementary Microsoft Excel files from which data needed to be retrieved, in order to successfully achieve the project's objectives. Having said this, and while one of the achievements of this project was to develop an independent DM with no previous assembled data structure, the fact that the company already maintained this data integration setup allowed the extraction of the data to be slightly more direct.

4.2. DATA MART STRUCTURE

When examining the DM structure, its two components ought to be mentioned: the Conceptual Data Model selected to model and design the structure of the DM; and the main variables that gave shape to the specific storage structure developed in this project.

4.2.1. Conceptual Data Model

For the context of this internship, given the complexity of the data sources and structure present, the Conceptual Modelling applied was the Unified Modeling Language (UML). It is worth noting that while UML is not specifically designed for DM development, it does represent a good solution when used to model and design the structure of one.

UML is a graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software system, that offers a standardized solution to developing the design of one. It includes conceptual notions such as business processes and system functions, as well as concrete ones such as programming language statements, database schemas, and reusable software components. (Luján S., 2005)

A UML data structure diagram is a partial graphical representation of a model under design or implementation. These are specifically used to represent the static structure of a system and its components, including classes, interfaces, and objects, and their relationships. It contains graphical elements which correspond to UML nodes connected with edges, also known as paths or flows, that represent the various elements in the designed model (Haji Ali et al., 2007).

From the different types of UML diagrams that exist, the most appropriate one for this project was the Class diagram, which is specifically used to represent the entities and their relationships in a DM. Class diagrams are a key artifact in the development of object-oriented systems as they lay the foundation for all subsequent design and implementation work (Genero & Piattini, n.d.). These diagrams provide a clear and concise way to understand and communicate the structure of the DM and are most useful when used in the design and development phases of the project. Below you can find the Class diagram established for the data structure of the DM developed in this project.

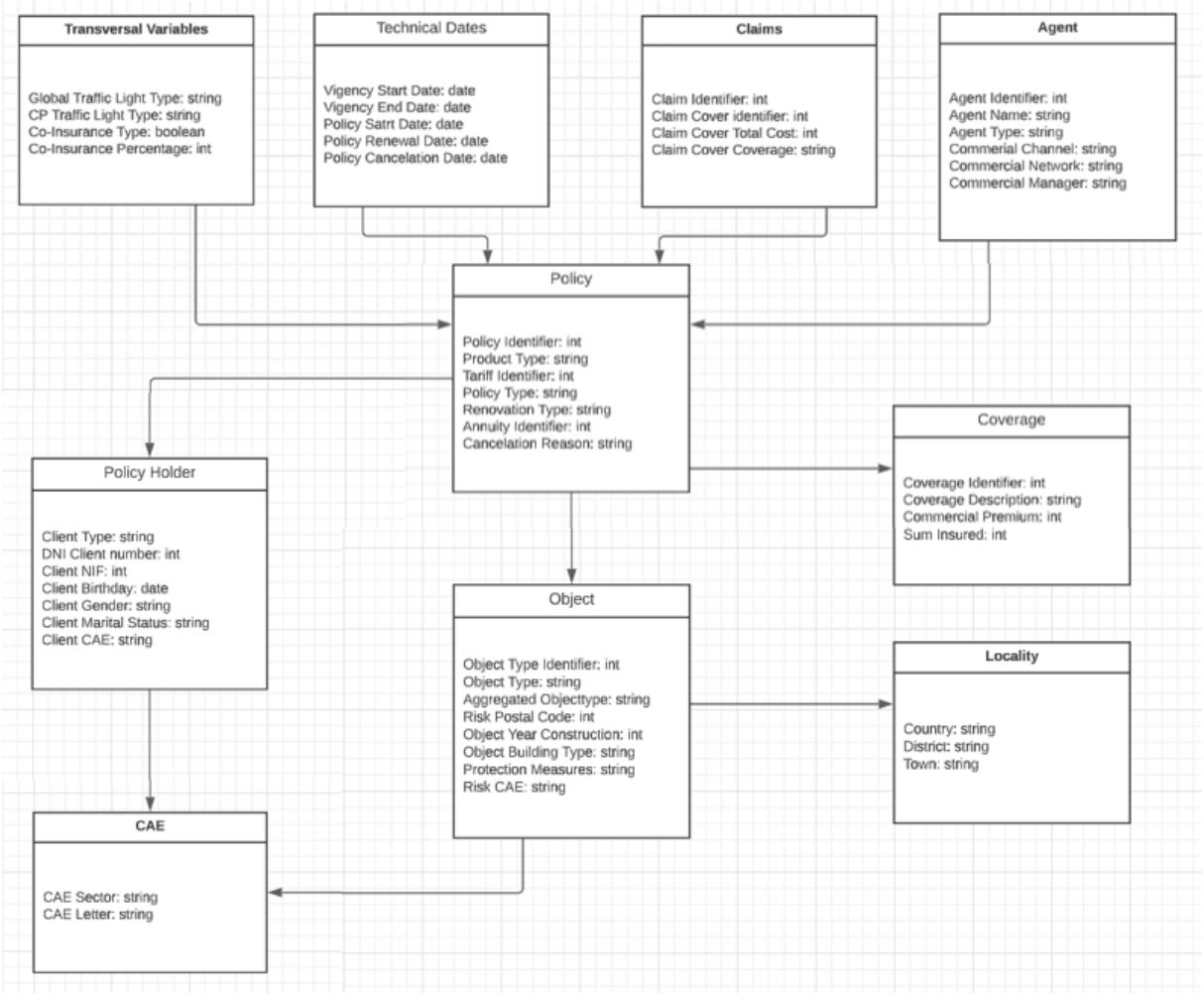

Figure 5 – UML Class Diagram developed for the Project

4.2.2. Data Mart Structural Composition

Regarding the DM technical structure, it is a subject-oriented relational database, where data is stored in a table format in SAS EGuide. Therefore, its structure is composed of rows and columns that are easy to access, organize and comprehend.

When deciding upon a row-oriented versus column-oriented structure for the database, it was decided to proceed with a classic row-analysis. The main reasoning behind this decision was the fact that column-stores are more efficient for read-only queries since they only have to read the attributes accessed by a query (Abadi et al., 2008). As the goal was, besides developing a TDB with this data, to be able to perform all types of necessary analyses on top of this DM table solution, the decision fell on the other option. In addition, the PBA team had previously developed projects with both approaches, and the consensus was that the row-oriented structure was more beneficial for the objective at hands.

In practical terms, the main unit of analysis in the DM were the individual policies, each represented as a record per row. A policy then had as many records as its number of vigencies, 'Accident months' in which it was active for those respective vigencies, number of objects within each one, and lastly given the number coverages included in each object. This organization was the best solution found

given all the necessary drill-downs and specifications studies and analysis intended to be performed afterwards.

Once it was known which information was needed for the development of this project, how its elements were connected amongst themselves and how they were going to be structured within the DM, a proper DM variables' composition was defined. As mentioned previously, this type of data storage infrastructure had already been developed inside the team for two other LOBs, Motor and Multi-risk Household, therefore the main arrangement of the DM was already pre-defined to some extent. In addition to this, a future goal for the team is to develop an integrated DW infrastructure, by aggregating all individual independent DM into one single structure, that would allow the team to have a client vision, instead of a LOB one. For this reason, the more uniform the structure between the LOBs' DMs, the easier it will be to combine them in the future.

Upon study and ponderation, it was decided that, while not diverging from the general structure of analysis implemented before, it was crucial for some additional information to be included in this project. The main characteristic present in this DM, comparing with the ones developed prior, was the introduction of a Coverage perspective.

Coverages are the smallest unit of granularity present in a policy and are one of the ideal elements to study and analyze within the Insurance business. Before, the structures developed only allowed policies to be explored up to their objects – a level of complexity above coverages. The addition of coverages' data also allowed the inclusion and study of 'Claim Covers', which as mentioned on chapter 2.1.2. are the smallest unit of analysis within a claim, to be established. This direct relationship between coverages and claim covers had never been created in a DM nor a TDB inside the company before, meaning that this project allowed for an entire new set of possible business analysis.

The DM structural segments proposed categorizes the data into the following seven different groups: key variables, Policy variables, PH variables, Commercial variables, Object variables, Coverage variables and Claim variables. Below you can find Table 1 related to the key variables segment and, on the appendix, Table 4 until Table 9 contain the composition of the remaining segments. Note that the listings already contain some variables and indicators that were created manually during the production of this DM. However, measures or purely auxiliar variables that were only developed to help the TDB's analysis were not included here and are explained in the 'Variable Creation' chapter.

Description	Example	Original vs Created
Source System Identifier	TECNISYS; COGEN	Original
Company Identifier	Ageas; Ocidental	Original
Line of Business	MRE	Created
Underwriting Period Identifier	202201; 202202; 202203	Created
Accident Period Identifier	202201; 202202; 202203	Created
Policy Identifier	0095254692XX; ME819657XX	Original
Vigency Start Date	2020-01-20	Original

Vigency End Date	2021-01-20	Original
Object Type Identifier	1; 2	Original
Coverage Identifier	Incendio; Tempestades	Original
Product Type	010; PNM	Original
Product Description	Commercialis; MULTIRISCOS EMPRESAS	Original
Tariff Identifier	010; M4	Original
Policy Type	N – Normal; R – Re-Insurance	Original

Table 1 – Key Variables Segment Table

4.3. MEASURES AND KPIS

Given the two main components of this project, a group of business analyses' measures and KPIs had to be considered and developed. This was done by identifying, with the members directly involved with the LOB, what were the critical success factors for CP and the best way to present and monitor them in this project. These were essential for the study of the LOB and served as the main strategic indicators when evaluating the business performance and results. Below you can find them listed, followed by a brief explanation.

- **Accident Period:** The smallest time unit used in both the DM and Technical Report references the month and year in which the analysis is being delivered. A policy will appear in as much accident periods as it is active with the company.
- **Underwriting Period:** The second time unit created, used to indicate the month and year in which a policy renewed, and it will be the same for every accident month impacted by that annuity. One underwriting period can have up to 13 accident periods, as the start month of the policy can correspond to the end month with a one-year difference.
- **Total Active/'In Force' Policies:** The total number of policies with an ongoing vigency in a particular time interval. It is analyzed as a total and also divided into 'New Business' and 'Renewed' policies.
- **Total Commercial Premium:** Total premium of a policy given its particular vigency. This premium is associated with the several individual coverages included in a policy, and their aggregation represents the total policy's premium.
- **Average Commercial Premium:** Total active policies' commercial premium divided by the total number of active policies.
- **Total Written Policies:** Total number of policies written in a particular period, considering all New or Renewed policies in that period.
- **Total Written Commercial Premium:** Total premium amount of the written policies in a particular period.
- **Exposure:** Total number of days that a single policy is considered to be active, within its vigency, given the time of data extraction. The smallest value of exposure corresponds to 1 day and the largest to 365 or 366, depending on the year.

- **Earned Premium:** Portion of the policy's commercial premium earned given its exposure by the time of data extraction. A policy's earned premium at the end of its vigency must be equal to the policies Commercial Premium for that specific vigency.
- **Average Earned Premium:** Total policies' earned premium divided by the total portfolio's exposure.
- **Sum Insured[2]:** Total amount of the sum insured for the active policies given a specific period.
- **Average Sum Insured:** Total active policies' sum insured divided by the total number of active policies.
- **Average Rate (%):** Total active policies' commercial premium divided by the total active policies' sum insured.
- **Total Number of Claims:** Total number of claims observed in a particular period, which can be disaggregated into 'Claim Covers' and studied.
- **Total Claims' Cost:** Value of the total cost associated with all claims occurring in a time period. Claims' costs can vary, however, if a claim presents a total cost of 0€ it will not be considered in the majority of the KPI's calculations.
- **Average Claims' Cost:** Total claim's cost divided by the total number of claims (with a cost higher than 0€).
- **Frequency (%):** Total number of claims (with a cost higher than 0€) divided by the total portfolio's exposure.
- **Loss Ratio (%):** Total claim's cost divided by the total portfolio's exposure.

4.4. Technical Report Structure

The final element produced in this internship is composed of several complementary dashboard pages, including the glossary and introductory sheet, all aligned with the good practices explained in the Literature Review above. Similarly to the scenario faced during the DM structuring, the PBA team had also previously developed some TDB for other LOBs. This led to the need to maintain a similar structure to the existing reports, in order to provide end users with a cohesive line of analysis across all LOB's studies delivered by the team.

One of the objectives of the report is to focus heavily on the following business KPIs, already explained previously, which are the most transversal and impactful business measures, thus being present in every dashboard page:

- Total number of Active policies
- Average Earned Premium
- Average Commercial Premium
- Average Claim Cost
- Frequency
- Loss Ratio

Complementing this information, there are different views and dimensions from which distinct results and conclusions can be reached. All sheets include seven fixed filters, illustrated below in Figure 6,

[2] Sum Insured translation to Portuguese is 'Capital'.

associated with key variables for the business, that allow the user to specify any of the analyses available. It is also important to mention that these filters can be used simultaneously, and more than one value can be selected at a time per filter.

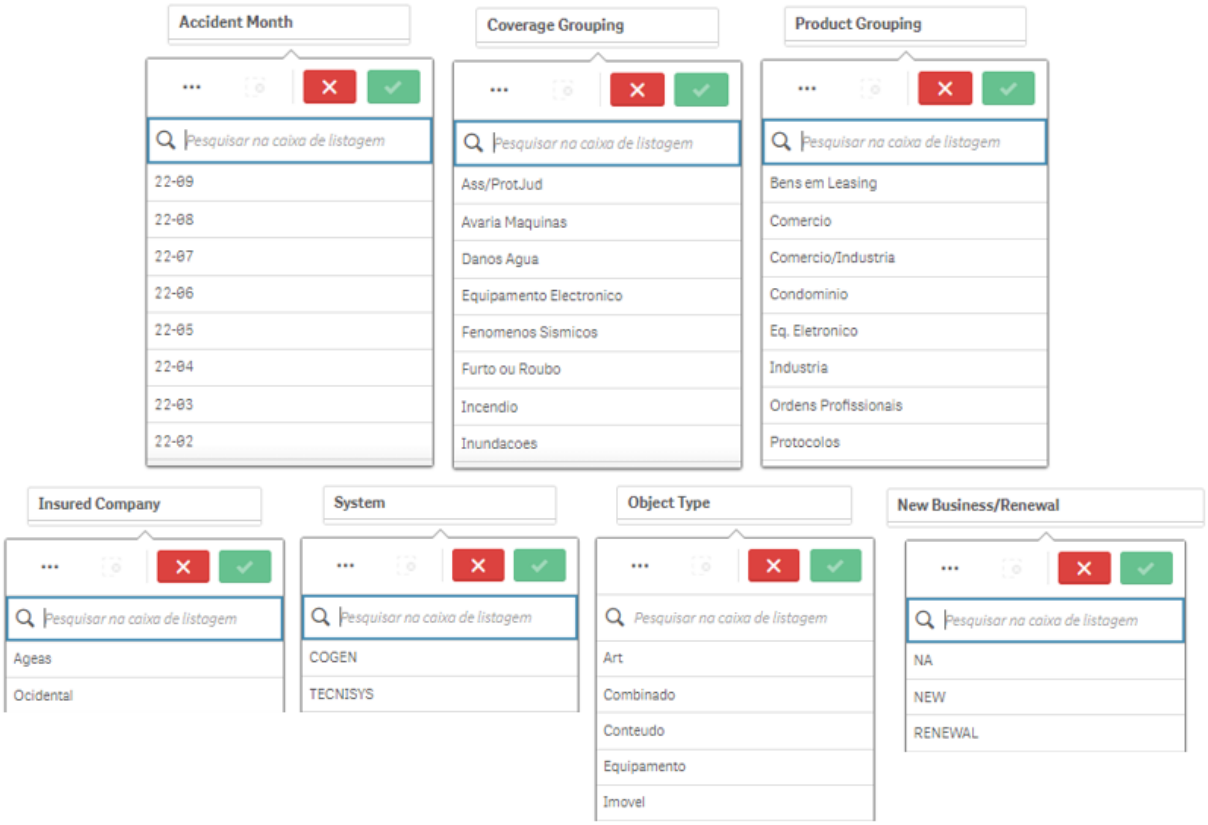

Figure 6 – Fixed transversal filters' description

Regarding the composition of the TDB's most relevant pages, the initial one contains the key business values, by default, for the most recent period of analysis - accident month -, along with their variation considering the homologous period of the previous year. In addition, bar charts showcasing some of the business metrics through a comparative analysis with the same time period in the last two years are also displayed.

The second page of the report is perhaps the most technical one, containing one single KPI's table with segmented values throughout different variables' sections. The user can select up to three distinct variables to combine at the same time, from more than ten different options, depending on the view in which they desire to explore the discriminated KPIs.

Lastly, for the most unique page of this report, the focus is on the individual coverages. The goal is to, again, study the most relevant KPIs segmented through the several coverages. Also in this sheet, there is a comparative visualization where both the average Commercial Premium and average Claim Cost of each Coverage are represented. This enables the user to better understand if any coverage is particularly more or less profitable than the remaining, also allowing to identify possible trends associated with this vital variable.

5. DEVELOPED WORK

The mission for this internship was to deliver two complementary components: a data storage solution that gathered all the historical portfolio's information and a resulting TDB that-provided an analysis of the LOB through a business perspective that would enable an improved decision-making process. To successfully achieve both these tasks a set of steps were performed. In this next section, the main actions are described and explained, as well as the internship's evaluation process.

5.1. DATA SOURCES AND COLLECTION

This project required the extraction of a significant amount of data, for which several SAS EGuide tables and some Microsoft Excel files were used. During this initial phase, the process was separated into brands, originating two separate processes of data collection: one for Ageas Seguros and another for Ocidental Seguros. With this said, since the goal was to extract the same information for both brands, the data collection process was very similar between the two.

As explained previously, the variables requested for this DM fell into seven main categories; however, these were located in more than twenty different tables, for both brands. For reasons of confidentiality the names of each table will not be disclosed but below, on Table 2, you can find the general structure of the data amongst the SAS EGuide libraries, whose names were altered for privacy reasons.

SAS EGuide Libraries	Nr Tables applied	Brands	General Information
OXX	9	Ageas Seguros	Transversal data
OXX_SEM	16	Ageas Seguros	Transversal data
SASPXX	11	Ageas Seguros & Ocidental	Claim data, Transversal data
PMCOP	2	Ocidental	Policy data
PMCOO	1	Ocidental	Object data
PMCOC	1	Ocidental	Coverage data
GPM2PXX	1	Ocidental	Renewals data
UNDUEXX	1	Ocidental	Transversal data
SEDCXX	1	Ocidental	Transversal data
FXX	2	Ocidental	Transversal data

Table 2 – General Structure of the data within SAS EGuide Libraries

As mentioned, a few Excel files providing specific information to complement the data already present were also imported. In these scenarios the data necessary was not described in any SAS table, and the documents were either official pre-existing company documentation or they were developed specifically for this project. The latter only occurred when dealing specifically with policy coverages. As stated previously, this was the team's first major project involving an analysis by coverage level, therefore there were some relevant perspectives and information missing.

5.1.1. ‘Base Coverage’ Disaggregation

As explained in chapter 2.1.1., most policies are composed of a ‘Base Coverage’ that includes a variety of mandatory coverages that are, by default, part of the product. Sometimes, that same policy can have additional coverages that are added by the client to complement the product. Until this point, because the analyses were being made only at a policy level, the only relevant values considered – mainly premium values, sum insured values and also claim information – were associated with the policy as an entirety.

For this transition to be made to each one of the coverages included in the policy, an additional study had to be developed, where a drill-down of the encapsuled ‘Base Coverages’ was built. The goal here was to understand which percentage of the total ‘Base Coverages’ commercial premium corresponded to each one of the singular coverages in it, while also extracting the sum insured value associated to each individual coverage – for the most part this was equal to the ‘Base Coverages’ total sum insured; however, for some coverages it presented its own unique values.

To be able to develop a document that would provide all this information, a variety of company’s ‘Product and Condition’ documentation was analyzed. This complex process had to consider the brand of the policy, the type of product secured and the insured object – building or content. Given that this scenario was present for both brands, and each brand had more than 10 available products, two main objects to consider and four different ‘Base Coverages’ identifiers, each to disaggregate, the possibilities for repartition were various.

The approach taken was to analyze the distribution of the coverages included in each ‘Base Coverage’ for the most recent LOB product’s portfolio – that did not have a ‘Base Coverage’ fixed coverage –, regarding commercial premiums and sum insureds, in order to understand their relevance and weight in both these variables. This task required a particular sensibility to the business that only experienced professionals could offer, so the analysis was developed with their constant orientation, which was essential to produce the most accurate and trustworthy distribution of the values.

5.2. Data Cleansing and Transformation

In this stage of the project, some adjustments and transformations were made to the data in SAS EGuide, using complementarily the traditional SAS point-and-click interface, SAS Code and SQL.

The main goal of this phase was to unify variables amongst the two brands, which was a complex process given the discrepancy in the data, due to its different sources. Most variables did not have the same nomenclature between the two brands, another great portion did not share the same structure when it came to the variable contents and how it was presented, and lastly there were some variables that were only present for one of the brands. All these factors led to a time-consuming variable analysis, followed by a proper data transformation. The objective was to unify the variables between the two brands, creating a homogenous data flow within the table. This process was done in parallel for each brand, and the information was only unified into one single structure once all the variables were aligned.

Regarding the cleansing of the variables’ values, as the goal was to store historical and official data, missing values and outliers were not handled so as not to disregard any record nor to alter its original value. However, all variables with a percentage of missing values higher than 95% are removed from

the table, since they do not bring enough relevant information to the study. Moreover, the data extracted was filtered to start in 2017, in order to include information regarding the last 50 months, and a few conditions were applied to guarantee that all the technical business assumptions were met. Below is a table with some of the most relevant assumptions considered during this DM development phase:

Conditional Statement	Output when TRUE
If 'End Vigency Date' is null and 'Cancellation Date' is not null	'End Vigency Date' = 'Cancellation Date'
If 'End Vigency Date' is not null and 'Start Vigency Date' is greater than 'End Vigency Date'	Delete record
If 'Cancellation Date' is not null and 'start vigency date' is greater than 'Cancellation Date'	Delete record
If 'Start Vigency Date' is not null and 'Start Vigency Date' is equal to 'Cancellation Date'	Delete record
If 'Start Vigency Date' is equal to 'End Vigency Date' and there is a premium alteration comparing to the previous vigency	Delete record
If 'Start Vigency Date' is equal to 'End Vigency Date' and there is not a premium alteration comparing to the previous vigency	Extend the previous vigency's 'End Vigency Date' to the current 'End Vigency Date' and delete the current record
If 'Acceptance Claim Date' is null and 'Claim Cost' is null	Delete Claim record

Table 3 – Business Technical Assumptions Table

Lastly, some variables were adapted to represent less storage space in the overall table. This was done by altering the original variables' content into codes instead of descriptions. To guarantee that this modification would not impact the end user experience when viewing and consulting the DM, a SAS EGuide formatting feature was used, that works similarly to a dictionary, called the Create Format. This simple feature enables improved understanding and grouping of data, as well as easy values conversion (Constable, 2010). It allowed the content of the variable to be adapted, while maintaining the showcased value as intended and defined by the developer, so the table would show the full description of the variable, while only using the storage space of a code. An example of this operation can be found on Figure 16, on the Appendix section.

5.3. VARIABLE CREATION

Following the Data Cleansing and Data Transformation stages, it was necessary to create additional variables that were missing in the analysis. In this phase, both Ageas Seguros and Ocidental

information was gathered into one single structure, therefore this process was carried out only once. As before, this process was developed using the traditional SAS point-and-click interface, SAS Code and SQL in SAS EGuide.

In total 40 new variables were created, including the 'Created' variables mentioned in chapter 4.2.2., but also some others more focused on facilitating the analytic view of the data. Hence, some of the variables created in this phase had the intent of assisting the future TDB, representing a more supporting role in the analysis itself, instead of being purely business informative variables.

Below you can find some of the most important variables created, divided by categories, along with their respective description:

- **Time Structure**

Defining a time structure is, perhaps, the most essential step when developing the DM as it ensures a uniform and organized data flow in the analysis. For this project 'Time' variables were created.

First, **Accident Period** was created to embody the concept of the month under analysis, and it is composed by the year and month being analyzed. The intent for this DM was for it to include the last 50 accident periods, meaning it only studied the CP's portfolio for the last 50 months. Given the monthly updates, these months are constantly being renewed. This variable represents the smallest level of analysis associated with the time structure, and it results in the creation of a record per month, given the active vicency periods of a policy.

The second and last variable created in this category was the **Underwriting Period**, referring to the month in which the policy renewed, also presented as a concatenation of a year and a month value. As it is expected, during a policy's entire annuity, the underwriting period will always remain the same, while the accident periods will vary given the vigency months in which the policies were active.

- **Policy Status**

To study the status of the policies in the CP portfolio throughout the last 50 months, two main identifier variables were developed. First a binary variable that indicates if, at a certain period, a policy was 'New' or 'Renewed' was created, according to the 'Policy Start Date' and 'Vicency Start Date'. If the two dates' year and month are the same, the policy is considered New, otherwise, it is Renewed.

Additionally, a complementary **Policy Situation** variable was created to tackle the need to distinguish between the types of cancellations. It takes one of the following values: 'Active', 'Renewal Cancelled' or 'Mid Term Cancelled', depending on the relation between the variables 'Policy Cancelation Date' and 'Policy Renewal Date'. Being able to distinguish between these types of cancellations is of crucial importance to understand the motives that may lead to the termination of a policy. For example, a Mid Term cancellation could indicate the client received a better offer from a different insurance company or even Ageas for the same product, while a cancellation upon renewal could simply indicate that the client no longer needed the services provided by the company.

- **Coverage Grouping**

As mentioned previously, this project included the analysis of coverages at a row-level. Since there were more than 50 different coverage identifiers available for each product of each brand – Ageas and

Ocidental – it was decided to group them into **Coverage Groupings** in order to reduce the number of records present in the DM and consequently save storage space.

From all the coverages available, the main and most impactful ones for the CP business were identified. This process originated 12 main groups and an additional 'Remaining' segment that englobed all the remaining coverages that did not get the chance to be analyzed separately, creating a variable with thirteen fixed values. As an example, the coverage of 'Theft or Robbery', between the two brands, had the following original identifiers: 002, 286, 287, BCF, FRB, FUR, R/F, ROP, ROU, FRI. These all fell under the category of 'Theft or Robbery'; however, each presented a particularity that did not allow them to originally belong to the same coverage ID.

When a policy contained more than one original coverage within a coverage grouping – something that occurred frequently with the 'Remaining' segment –, the procedure regarding the Commercial Premiums and Sum Insureds of that particular groupings was to summarize all the Premium into one total value per grouping and use the higher Sum Insured present as the grouping's final Sum Insured.

These coverage groupings were carried out throughout the entire project and became the official coverages identification adopted.

- **Earned Premium**

As described previously in Chapter 4.3., **Earned Premium** corresponds to the portion of the policy's Commercial Premium already earned given its exposure at a certain period. To create this variable and reach this value, at every monthly update of the DM the Earned Commercial Premium must be calculated for each coverage grouping given a particular policy vigency. This calculation is obtained by combining the Commercial Premium value, with the number of Exposure days verified until that period and a couple of other specific technical variables developed throughout the process.

Logically, for all the vigencies that were already concluded, the final Earned Commercial Premium verified in the vigency's last active month would be the same as that vigency's total Commercial Premium.

- **Auxiliar Variables**

Since the DM was developed at a row-level, with a variable drill-down from the policy down to the object, and from the object to the coverage groupings, it was necessary to guarantee that when the information was extracted in a summarized format and the TDB developed, there were no duplication of values. For example, if a policy is active in a particular accident month, all the records associated with that policy and period combination will be considered as 'Active' – these will represent as many rows as objects covered by the policy and how many Coverage Groupings contained in each object. Given this, if we were to summarize all the active policies in a selected accident period, we would have more records being counted than the exact number of distinct policies we initially intended to obtain. This same problem would occur if we decided to count the number of distinct objects or coverage groupings present.

To solve this issue and support the analysis, four binary (1,0), auxiliary variables were created to uniquely identify records and avoid duplicates among specific categories:

- **Ac_Policy:** uniquely identifies one single time a policy within an accident period.

- **Ac_Policy_Object:** uniquely identifies one single time an object within an accident period, so that if a policy has two different objects, they will both be separately identified once.
- **Ac_Policy_Object_Cov:** uniquely identifies a coverage grouping within an accident period. If a policy has two different objects, and a specific coverage grouping is present in both objects, it will be identified two times, once per object.
- **Ac_Cover:** uniquely identifies one single time a coverage grouping within an accident period. If a policy has two different objects, and a specific coverage grouping is present in both objects, it will only be identified one single time. For some analyses the previous variable would still provide duplicated values, therefore this was developed to aid that issue.

These four variables answer different necessities related to the TDB structure and analysis that was developed afterwards. Below you can find a practical example of these four variables:

	Id_Uw Period	Dt_VigStart Date	Dt_VigEnd Date	Id_Ac Period	Id_Policy	Cod_Product	Id_Object	Desc_CovGrouping	AC_Policy	AC_Policy_Object	AC_Policy_Object_Cov	Ac_Cover
1	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Ass/ProtJud	1	1	1	1
2	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Danos Agua	0	0	1	1
3	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Furto ou Roubo	0	0	1	1
4	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Incendio	0	0	1	1
5	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Inundacoes	0	0	1	1
6	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Pesquisa Avarias	0	0	1	1
7	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Quebra Vidros	0	0	1	1
8	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Remaining	0	0	1	1
9	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Responsabilidade Civil	0	0	1	1
10	202211	2022-11-06	2023-11-05	202212	0000000001	016	1	Tempestades	0	0	1	1
11	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Ass/ProtJud	0	1	1	0
12	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Avaria Maquinas	0	0	1	1
13	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Danos Agua	0	0	1	0
14	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Equipamento Electronic	0	0	1	1
15	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Furto ou Roubo	0	0	1	0
16	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Incendio	0	0	1	0
17	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Inundacoes	0	0	1	0
18	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Quebra Vidros	0	0	1	0
19	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Remaining	0	0	1	0
20	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Responsabilidade Civil	0	0	1	0
21	202211	2022-11-06	2023-11-05	202212	0000000001	016	2	Tempestades	0	0	1	0

Figure 7 – Data Mart sample for a policy in December of 2022

- **Claims Structure**

The last segment of developed variables focuses entirely on claims and claim covers. When the claims' data is extracted from the source it does not contain all the necessary nor wanted information to proceed with a complete analysis, so several adjustments and developments need to be made.

First the claims are separated into the following different technical categories in order to facilitate their study: Claims, Claim Covers, Closed Claim Covers, Ongoing Claim Covers, Large Claim Covers (with total cost over 50 million euros), Natural Events Claim Covers and Claim Covers with a total cost of 0€. Once these are identified, they are aggregated given a specific combined key – accident month, policy number, object identifier and coverage grouping. Upon aggregation, two complementary variables are created for each category: first the **Number of Claim Covers** is calculated in each segment; and then its correspondent **Total Claim Cover Cost** is summed.

In addition to these variables, a flag variable identified as **Claim Cover with No Exposure** is also created to identify claim covers that cannot be properly associated with a policy. This scenario tends to occur when either the period in which the claim cover occurred does not correspond to the accident months in which the correspondent policy was active, or when the object or coverage grouping associated with the claim cover, as registered by default in the system, does not found a match in the correspondent policy.

5.4. SUMMARIZED TABLE

As mentioned previously, the size of the data held in the DM was extremely large and would not be well supported in QS if integrated as is for the development of the TDB. For this reason, an intermediate summarized table was created, still in SAS EGuide, in order to adapt the data for the dashboards. The primary unit aggregated were the individual policies, resulting in the loss of the smallest granularity level, comparing with the DM. This step was already accounted for in the development of the DM, with the creation of appropriate variables that served as aggregating and counting variables – already described in Section 5.3..

In addition, the number of variables present in the Summarized Table was also inferior to the total number of variables in the DM. Here, only variables that would be included in the TDB were selected, which were the ones holding the most valuable information for the decision-making process and the business itself. With this selection, the number of variables considered reduced for almost half of its value.

5.5. DATA VISUALIZATION

The end goal for this project, and perhaps the internship's most relevant outcome for the company, was the creation of the TDB in QS. This tool fulfills the department's need to analyze and understand at a deeper level the CP LOB.

Here, the information was divided into 11 report pages, where each presented different yet complementary perspectives of the data. As mentioned previously in chapter 4.3., there are certain measures and KPIs essential to the business, that thus tend to be present in most pages, although showcased through distinct views. Below, a brief description of each page can be found:

Technical Sheet: First page explaining all KPIs, and relevant concepts present in the report.

User Guide: Complementary introduction regarding the report, with some important remarks and business disclaimers. Also includes a small guide describing each page's date analysis view.

Page 1 - Main KPIs: Initial view with business figures and the most relevant KPIs, intended to be analyzed specifically for a selected month, by default, the most recent one.

Page 2 - Summary Analysis Table: Complete table representation with various business values and metrics, filtered by up to three dimensions selected by the user.

Page 3 - Main KPIs Monthly View: Default comparative monthly evolutive analysis among main KPIs.

Page 4 - Main KPIs by Variable: Differs from the previous one by providing diverse variables from which data can be filtered by.

Page 5 - District Map Analysis: Geographic representation of Portugal, by district, where one of five KPIs can be selected.

Page 6 - Claims Analysis: Most relevant KPIs related to claims, filtered by a set of dimensions of choice.

Page 7 - Coverages Comparative Analysis: Coverages' comparisons of the most relevant KPIs. It also provides a direct comparison between Coverages and Claim Covers.

Page 8 - Large and Event Claims: Dedicated to Large Claims and Event Claims, both special type of claims that are typically associated with higher costs.

Page 9 - Annuity Claims: Understanding of policies' behaviors regarding their Claims' history in present and past annuities.

As mentioned in chapter 4.4., all visualization pages include seven fixed filters that help the user taper their analysis. In addition to the default filters, some pages also have complementary filters, associated with relevant variables for the business. All the variables used as filters were earlier identified by the main end users of the dashboard as necessary and value-adding when studying the CP LOB.

Regarding the type of analysis, in all pages it is possible to select the desired month or months under study, by manually selecting the preferred accident month(s). Note that with the right selection of months it is possible to obtain a year-to-date view of the data. In addition, in certain pages it is also possible to activate a 'Monthly' view filter, that offers a view of the exact accident month selected – if more than one month is selected, the analysis will be done only for the most recent one. Lastly, most pages also offer a '12-Month Rolling' filter option that allows the user to evaluate the behavior of the variables displayed over the last twelve months.

During the creation of this TDB, several measures and dimensions were developed in QS using QlikView script – the primary language used in QS. These elements were crucial when presenting all the required information, especially the KPIs and business measures throughout all time analysis views available. They were also extremely necessary to allow all the filters to work properly, as well as the different variables' variations presented.

Now, the most relevant pages will be partially showcased and briefly explained below. It is relevant to mention that because this is an official Ageas TDB currently in use, it is not possible to share the official values. For this reason, most of the figures shared in this Internship Report will have blurred values and/or graphics without legends or numbers.

The first page of the dashboard shows all the main KPIs, in this case, for the month of December of 2022. Globally, this page provides a direct view of the most important KPI values for the CP's LOB, while also offering a yearly comparative analysis, at the bottom, of four specific KPIs.

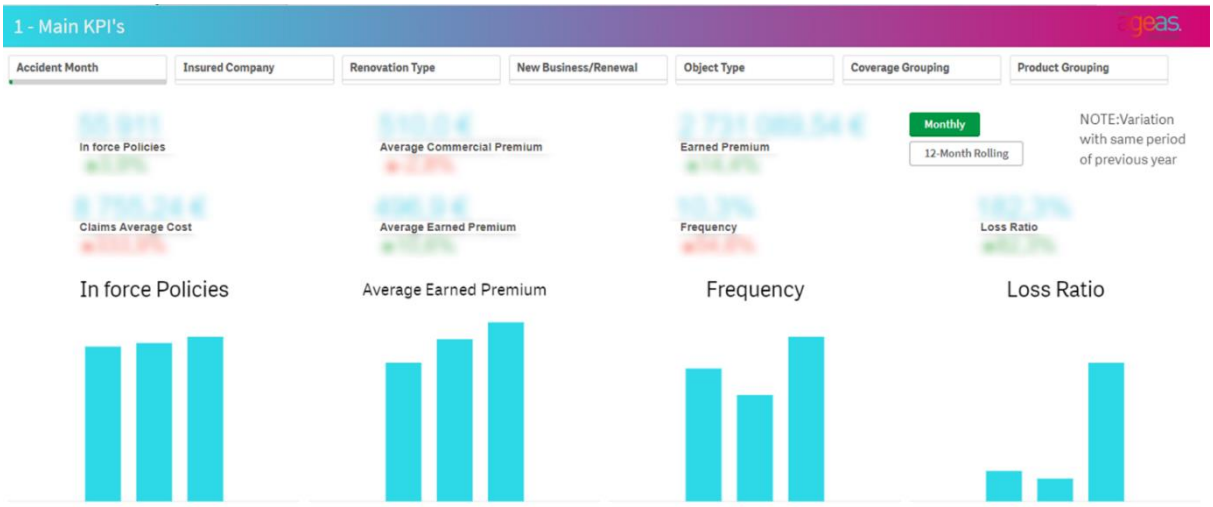

Figure 8 – Main KPIs page overview for December of 2022

Given the several filters available, it is possible to personalize all these values and re-direct any analysis, so that it is possible to:

- Understand if all accident months showcased similar values and trends and, if not, identify which were the outlier months that need further study.
- Visualize the portfolio's repartition between the two brands.
- Study only the 'Temporary' policies, that usually represent an extremely small percentage of the entire portfolio and can easily get lost in a more global analysis.
- Study the different impacts that 'New Business' can bring to the LOB.
- Specify the analysis to a particular type of object or coverage.
- Understand the different KPI values that distinct types of Products can have.
- Visualize the data through a 12-Month rolling view, that automatically offers a wider perspective of analysis.

Now describing the most technical sheet from the Report. Page 2 provides a Summary Table where each of its twelve columns represent a different business KPI, among: Exposure, Loss Ratio, Frequency, Average Claims Cost, Average Commercial Premium, Average Earned Premium, Earned Premium, Total Written Policies, Total Written Commercial Premium, Average Sum Insured, Average Rate.

At the same time, each row corresponds to a variable drill-down structure and can include information from up to three different variables selected by the user, by the order desired.

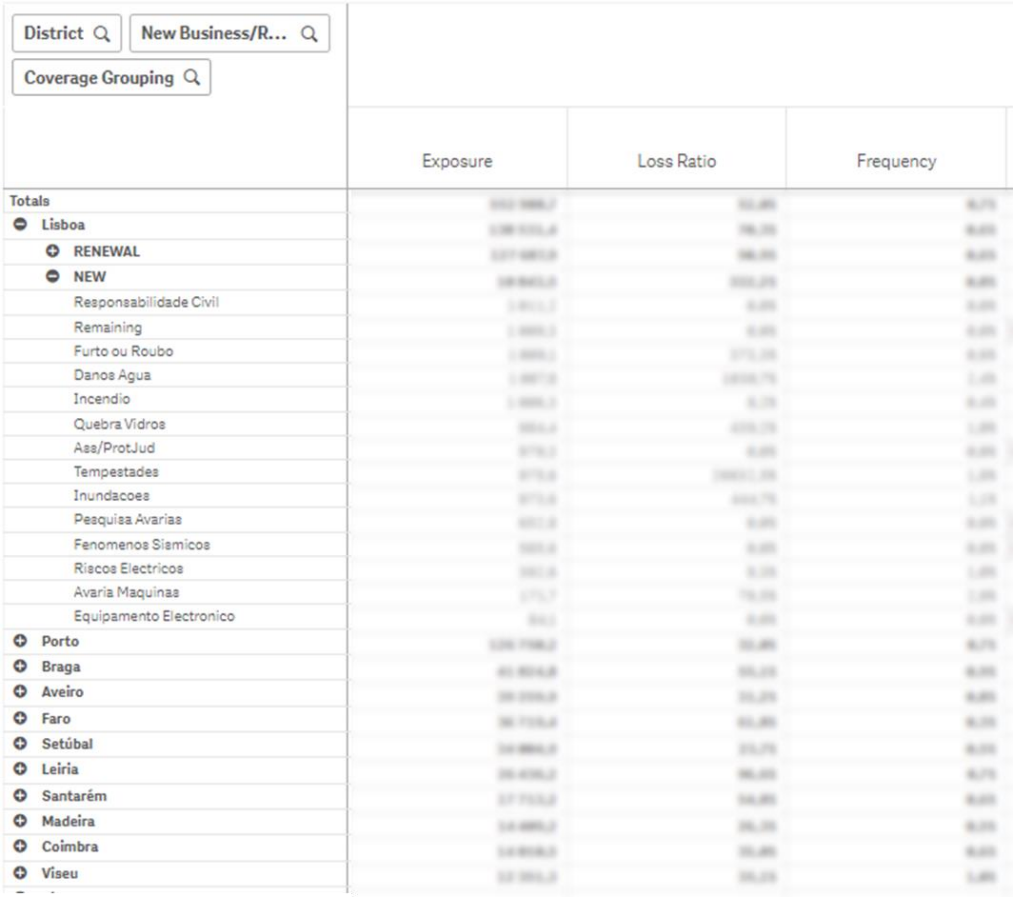

Figure 9 – Summary Analysis Table page sample for 2022

In the example showcased above in Figure 9, the represented values are firstly segmented by District, secondly by 'New Business' or 'Renewals', and lastly by type of Coverage Grouping. Each were selected from one of three 'Variable' options present in the dashboard. It is worth mentioning that the user is not obliged to always select three variables and perform their study following a drill-down structure. Below is the variable listing shared by the three filter options, from which the user can select what variable(s) to include in the analysis:

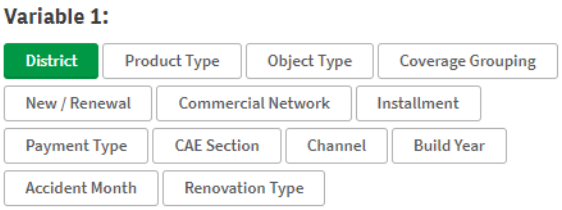

Figure 10 – Variable 1 Composition

Also in this sheet, it is also possible to filter the entire page to a specific 'Accident Year' using the correspondent filter available, instead of the standard 'Accident Month', as some KPIs provide better insights and are preferably analyzed in a Year-to-Date perspective.

Below, in Figure 11 the only geographical visualization of the TDB is represented, where the user can select which measure to study among the KPI filter, considering all Portuguese districts. The values will be presented in a cluster format with adapted value's intervals for each KPI.

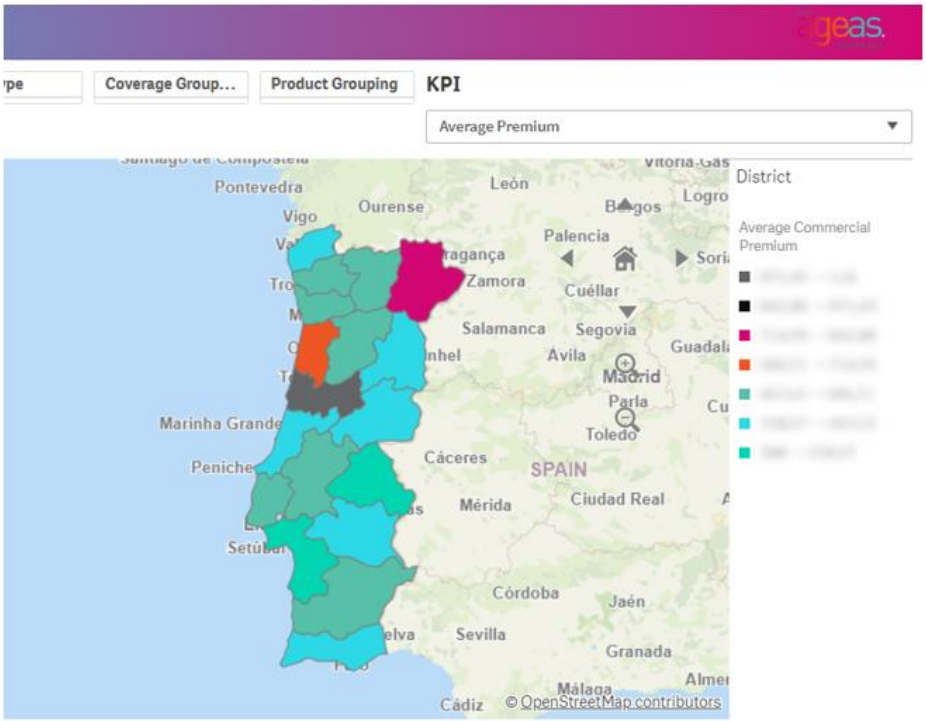

Figure 11 – District's Average Commercial Premium Analysis sample for December of 2022

Focusing on the coverage groupings, with the 'Coverage Comparative Analysis' page, the user can analyze each grouping individual's behavior. Here, for each coverage grouping, some key indicators are studied to ultimately identify relevant insights and remarks about each coverage. It also provides a unique perspective, where both the average commercial premium and average claim cost presented

by each coverage grouping, monthly, are directly connected and studied side by side. This is done with the help of an Indicators Table (Figure 17 on the Appendix), that provides five indicators for each coverage grouping, with the possibility of a drill-down by accident month, allowing a comparative and evolutive study. This type of analysis at a Coverage level had not previously been developed for any LOB, making this one of the main requests for this TDB. Moreover, this page also offers an additional monthly view of both Frequency and Loss Ratio values for the intended coverage grouping.

Lastly, Page 8 is relevant since, although it does not include any complex graphic or visualization, it focuses on Large and Event Claim Covers, which are relevant and unique types of claims. Both tend to have higher cost than the average claim, which sometimes can imbalance and mislead the global Claim analysis. By having a separate sheet like the one showed on Figure 12, the user can easily identify if, in any month, these claims had an effect in the overall portfolio values, and if so, the dimension and impact verified. The values studied are both the total number and total cost of each type of claim cost. As showed below, the visualizations are simple bar charts displayed throughout the several accident months, in which you can also see what were the types of coverages that were associated with each claim cover. On the left of each graphic, there is a "selection button" where the user can choose between Large Claim Covers or Natural Events Claim Covers.

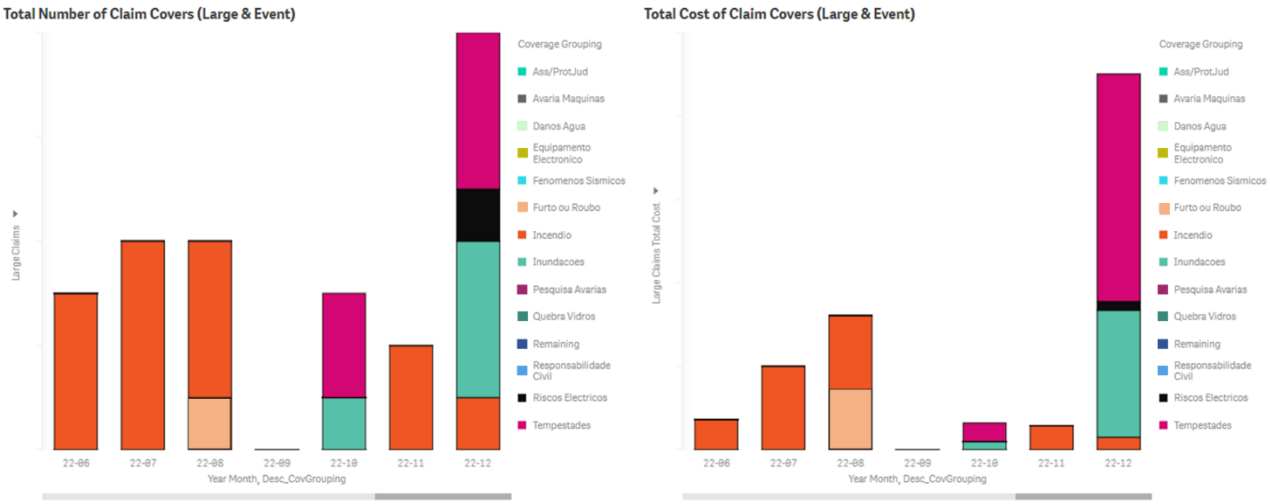

Figure 12 – Large Claim Covers Analysis

As it can be seen, in the last months, there were several large claim covers reported, particularly in December, which corresponded to a total cost significantly higher than the ones observed for the same year. Without this page, such increases in claim costs would be harder to comprehend and justify.

5.6. Process Evaluation

As part of the methodology used for this project, before the final product could be launched, it was submitted to an extensive validation process. This final step was divided into two stages, apart from the validation stages that took place throughout the entire development of both parts of the project.

First, the main portfolio values and KPIs were compared and studied against the company's financial values. PBA is a technical team that focuses on a more practical view and analysis of the business. This can lead to minor inconsistencies when comparing values and results with the official financial figures of the company, as some interpretations and assumptions are not equal between the two views and

approaches. With this said, to guarantee that the delivered report is reliable, and the analysis provided showcases truthful values, these possible variances needed to, firstly, be the smallest possible, and at the same time explained. To do so, a particular monthly financial document was used where the KPIs compared were the following:

- Total number of Active policies
- Total Earned Premium
- Average Earned Premium
- Average Commercial Premium
- Total portfolio's Exposure
- Total Written Premium
- Total number of Claims
- Total Claim Cost
- Average Claim Cost
- Frequency
- Loss Ratio

This evaluation originated, to some degree, a circular process where every time a metric would present a difference slightly larger than expected, a "deep dive" into the process would be performed to correct the possible concern. Most of the times there was no possible adjustment to be applied, as the values extracted from this project were completely in line with the technical view of the business held by the PBA team. In these cases, the goal was to identify why that difference was present, clarifying and registering it.

As an example, for this project it was considered that when a policy was cancelled, the effective date of cancelation was used as the final date for its vigency, which in several cases varied from the date when the cancellation was uploaded or shared to the company's systems. This means that a yearly policy could, for example, start on 01/01/2022 and on 01/01/2023 – the intended renovation date – the PH decided to cancel it applying its effect to 31/10/2022. Given the technical perspective in which the table was developed, any day after the 31st of October of 2022, the policy in question would not be considered active, therefore its exposure would not correspond to the full year. On the other hand, the accounting figures would register this policy's exposure to be the full year as the date when the cancellation was requested did not interject the previous vigency period of the policy. As showed in chapter 4.3., a policy's exposure also affects other important business KPIs like frequency and loss ratio.

After the first stage of validation was completed and the final values approved, the second phase begins, where managers and team members of the areas most related with the LOB were given a test version of the final TDB. This was done to guarantee that their experience and business know-how was also incorporated in the evaluation phase. They were asked to share final comments, requests, and possible concerns regarding particular KPI values or even visualizations, as a way of vetting and approving the dashboards before they were published within the company. This final phase did not involve a high number of adjustments as the test version presented was already considered the final ideal version. After both these stages were completed, the final product was successfully launched.

6. RESULTS

In the end, the two components of this internship were achieved with success, and it is possible to identify two clear and complementary results.

6.1. DATA MART

First, a new data storage infrastructure was built within SAS EGuide to provide a complete and up-to-date consultation and analysis solution for the PBA team, concerning the CP LOB. Until this point, all the data sources carrying its information were distributed and disorganized, not providing a clear pathway for analysis of any sorts. Ultimately, this led the company to slightly disregard this LOB, particularly when it came to its actual performance and business results.

With this DM, any member of the PBA team can now access, in one single structure, all the necessary information regarding CP. The selected structure allows any future studies, analysis and/or particular requests to be performed in an easier and faster way around the main key elements of the business, avoiding the need to perform any additional ad-hoc analysis. For the team, these main elements are: Accident Periods, Brand, Policies' information, Objects and Coverages. As mentioned previously, the DM composition includes portfolio's historical information regarding policies' current and past annuities, information regarding the contractual terms of each one, Commercial Premium and Sum Insured values, and lastly Claims data directly associated with the LOB.

The final DM, updated at the end of January of 2023, is composed of 135 different columns, and it has more than 38 million rows.

6.2. TECHNICAL DASHBOARD

The second accomplishment of this internship was the successful launch of the TDB for the same LOB. As mentioned previously, this was the main objective when developing the DM and the end goal for this project. Upon conclusion, it became possible to share with several company's department the most relevant analyses, trends and findings for CP under an interactive, direct and up-to-date BI solution in QS. Several teams work closely with this LOB and the introduction of this TDB was a clear necessity for the company. The insights obtained from the analyses of the TDB are varied and are mainly represented by values, charts, and tables. Since its conclusion in November of 2022, this TDB is updated once every month and it is available for consultation to several teams of Ageas's Operations NL.

To extract an initial and global view of the most relevant KPIs for the business, we can analyze the first page of the Report. The ideal analysis in this page is by selecting a particular month to study. Here, an assessment can be developed over the tendencies and global behaviors of the metrics, as well as their variations with the homologous period from the year before. These values can also be specific to particular business variables available as filters.

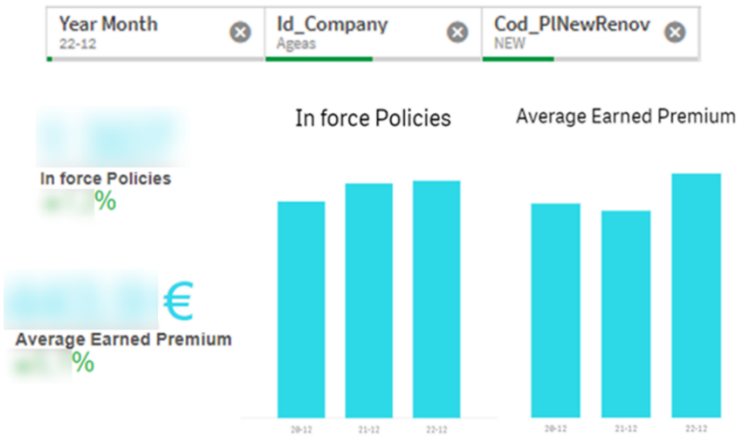

Figure 13 – Main KPIs page: Ageas Seguros' 'New' Business values

In Figure 13 it is possible to understand how many 'New' policies Ageas Seguros acquired in December of 2022, in Total Number of Active policies and also in Average Earned Commercial Premium generated for the company until that point, while also observing those values' correspondent percentual increases – variations – from the same period in the previous year, below. In addition, the evolution of the measures throughout the same months in the past three years can be observed.

If the purpose of the analysis is to extract a more complete and direct understanding of all measures and KPIs considered by the involved teams, there is a 'Summary Analysis Table' sheet that allows this. By selecting up to three variables, a decomposition view of all of them is generated, offering a greater level of detail and investigation ability. Here, the main advantage identified is the ability to comparatively analyze the different values and behaviors of the selected business dimensions/variables.

In this page it is also possible to obtain the exact value of each KPI, given several different time structures available upon selection. The selection of an 'Accident Year' allows for a year-to-date analysis of the measures, the selection of a 'Accident Month' offers a single month's analysis, and lastly, the activation of the '12-Month Rolling' button provides a more global and transversal view of the KPI values and behaviors.

In this TDB, there are also several visualization options that allow the user to understand the details of the data under analysis. For instance, it is possible to comprehend the impact and behavior of different variables' components within a monthly evolution analysis for the desired KPI.

Figure 14 below showcases the frequency and loss ratio of each type of policy object for Ocidental in 2022, along with the number of active polices for each object considered.

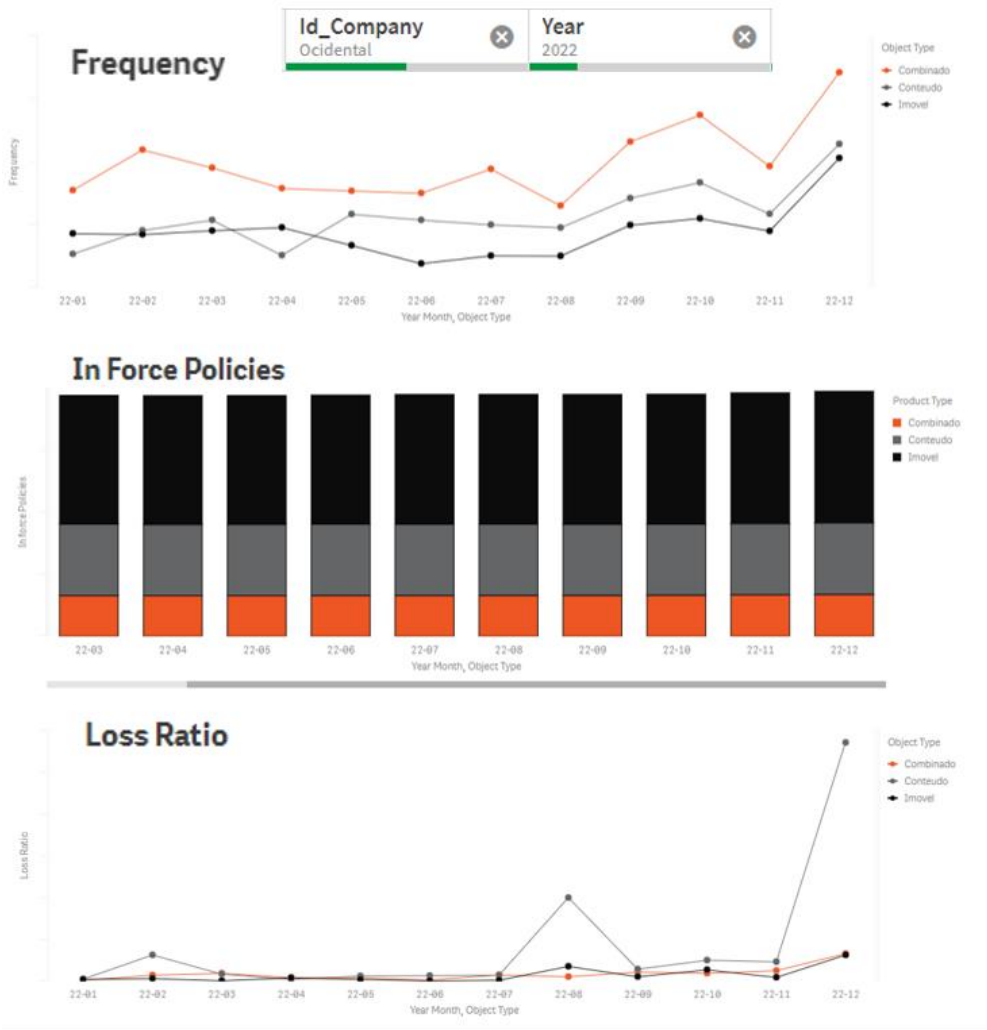

Figure 14 – Main KPIs by Variable page: Ocidental's Object Types Analysis for 2022

Knowing that the frequency represents the total number of claims divided by the total portfolio's exposure, and assuming a fairly even distribution of the exposure values by all object types, we can conclude that, for Ocidental, a policy that insures both Building[3] and Content[4] – object type Aggregated[5] – tends to present a higher number of claim incidents. Such conclusion is not necessarily unexpected, since if a policy is securing two different objects, instead of only one, the probability of suffering a claim is inevitably higher. With this said, such results can still be interesting and useful to acquire, especially when confirming that, from Ocidental's portfolio, this object type is not the one with the most active policies. In fact, the Aggregated object type shows a tendency for being the object with a significantly smaller number of active policies in the last year. In addition, it is possible to confirm that, as expected due to extreme weather conditions experienced, the winter months had generally more claim incidents than the remaining months.

[3] Building translation to Portuguese is 'Imovel', which is the variable name present in Figure 14.
[4] Content translation to Portuguese is 'Conteudo', which is the variable name present in Figure 14.
[5] Aggregated translation to Portuguese is 'Combinado', which is the variable name present in Figure 14.

Observing now the loss ratio analysis, also available in Figure 14, it can be confirmed that, particularly in December, the Content object type was the one that represented the highest increase in claim costs, as the loss ratio value escalated to almost five times the previous values for this object. It can also be observed that this specific object type, throughout the entire year, was the only one that showcased high picks in its loss ratio evolution. Such findings can suggest that the Content object type is more easily associated with higher claims total costs and tends to be more vulnerable to external disturbing factors related with claim incidents.

For a direct study between the policy's coverages and the claim clovers incidences associated with each one, there is a particular table that allows those conclusions to be made. This is found on the 'Coverage Comparative Analysis' sheet, already mentioned in this report, and it allows for a wider understanding of each individual coverage behavior, given a desired time period.

Here the user can obtain a more complete yet detailed study by following two main logical pathways. First, the analysis can be made to the specific indicators presented in the table, where all coverages' behaviors can be compared amongst each other, allowing the user to extract a wider perspective of all Coverages, considering the LOB's portfolio. Secondly, the user can also decide to initially focus on the frequency and loss ratio evolutive visualizations, and from here, identify and/or extract a specific period or coverage that needs further study. Knowing this, a complementary analysis can be performed using all the necessary measures present in the page's table.

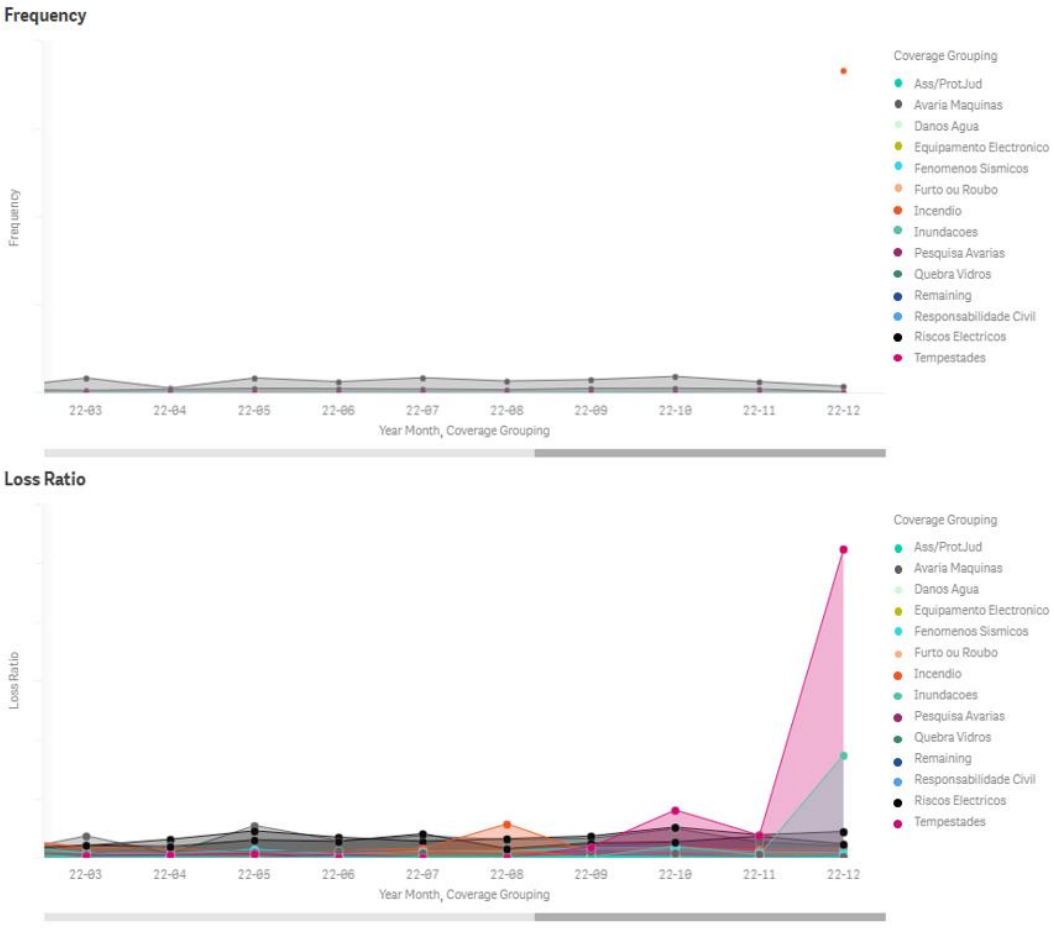

Figure 15 – Coverages Comparative Analysis page: Frequency and Loss Ratio Visualizations for Coverage Groupings

In the Figure 15 above, you can find an example of both these last visualizations, where clearly, for the last month of 2022, two separate coverage groupings stand out when comparing behaviors. Regarding frequency, it is clear that Fire[6] is the coverage category that reveals the highest frequency value. With the help of the table also available, any user will immediately be able to confirm that such values are completely justified by an extremely higher average claim covers cost verified for that month (note that these exact values are official company numbers, and therefore will not be displayed to visually support this statement). Regarding loss ratio, Storm[7] coverage is emphasized as presenting the highest ratio. Again, with the correct use of the elements in this page, the user is able to confirm what is the main cause for this outcome. As expected, it is due to the increase in claim incidents associated with this particular coverage in December.

The possibility to extract both these findings can be extremely useful to the users involved, as it allows them to develop a pattern and categorization for these important coverage groupings, and mainly conclude that both coverages do not behave in the same way. Although the extreme weather condition experienced in December of 2022 did not tend to influence an increase in Fire incidents, the ones that did occur demonstrated extremely high costs. On the other hand, and as expected, Storms became much more common, registering a notorious increase in claims, however, the damages caused by these incidents did not stand out as being infamously more expensive that the average of the remaining coverage groupings.

It is worth noting that, due to the outlier value presented by Fire in the frequency graph, the analysis of the remaining coverages' behaviors can become compromised. In these types of scenarios, the user can manually exclude this coverage grouping from the visualization in order to improve its readability properties.

As with every analysis offered in this TDB, this study can be highly enriched with the use of the available page filters, as the behavior of these coverage groupings and claim covers incidents can be done individually for a specific coverage grouping, or also significantly vary depending on the type of object or product that is being analyzed, for example.

Such analysis was highly requested by the teams and departments involved, as there is no doubt that it provides a much-needed perspective and vision over some of the most important business indicators in the Insurance business. The ability to study all these components side by side, will inevitably help the company reach more insightful conclusions regarding the CP LOB moving forward.

[6] Fire translation to Portuguese is 'Incendio', which is the variable name present in Figure 15.

[7] Storm translation to Portuguese is 'Tempestades', which is the variable name present in Figure 15.

7. CONCLUSION

This internship had two main focal points, and both surrounded the PBA team's and, consequently, the Company's need to better understand and evaluate its particular business area of Multi-risk CP and its portfolio. The goal was to help Ageas Operations' NL department improve its business decisions-making processes regarding this LOB, while also expanding its overall results.

To do so, the plan involved the development of a complete DM, given all the data available within the company and always considering the entire spectrum of relevant business KPIs and performance metrics used as decision making tools throughout the entire department. Such tool would not only be the main source of information for any analysis that may arise as a need within the involved teams, but also for the official LOB's TDB, which was identified as the second goal for this internship. The strategy for this dashboard was to understand what were the most relevant business insights that needed to be at an easy access and in constant display for all the departments and teams working with CP. Once these were identified, the dashboard was developed, evaluated for any possible issue, and successfully launched with the company.

7.1. LIMITATIONS AND ISSUES

Along with the development of this project a few limitations and particular issues were experienced.

As in any large company, the biggest challenge was the dimension of the data that needed to be considered. Although CP is not the company's LOB with the greatest portfolio, because this was a project that was meant to store historical data from every policy during the last 50 months, for two different brands, and with a new record being generated monthly for each policy, the size of the data was extremely large. This phenomenon led to two main occurrences. First, it is relevant to note that several teams within the company share the same SAS EGuide libraries to store different types of tables and projects. This fact, aligned with the high dimensions of the data used for developing the DM, resulted in a constant lack of space in the SAS common library, which led to frequent 'lack of space' errors while building and running the processes, thus delaying the project itself. Secondly, the constant awareness of the size of the data, also originated a higher caution when deciding which variables to include in the DM, conditioning the format and structure of the ones that were selected. A great example of this was the 'Coverage Grouping' variable creation, developed mainly to produce less records in the final table.

As an addition to this, it is relevant to mention the processing capacity of the data visualization tool. When faced with a higher amount of data as input, QS usually slows down its performance and originates higher waiting times, both in development and in production for the end user. This was the main reason why a summary table was created and imported to the application, instead of the complete DM.

Associated with the length of the data, the variety of source tables that needed to be connected to create the DM also represented a great issue. Not only were there several variables to consider, but there were two different brands to include in the project. As mentioned before, the company uses different tools and applications to retrieve data from its clients for years and, since the origin of the brands is not the same, the tools in which their information is collected is also different. The biggest issue associated with this was the heterogeneity in some of the information available. Most data that

was collected for both brands did not appear in the same formats, nor with the same basic definitions amongst them. Besides, there was also the case where some variables would be available for one brand, however, not for the other. The major challenge that resulted from this was the time spent identifying these cases while also consolidating and homogenizing the data into one cohesive table.

Finally, it is worth noting that, due to the business itself and its characteristic procedures, there is still a lot of information that is introduced to the systems manually. This is extremely common when registering a policy via an agent or representative, which still occurs a lot nowadays and can easily result in the input of several errors into the official company's systems. Throughout the project, these would sometimes be identified, however there was no possible correction that could be implemented to alter the original value.

7.2. Future Work

One of the most relevant projections for both the DM and the dashboard was their monthly update, where data regarding the most recent months could be constantly included and presented to the end user. Given the timeframe of the project, it was not possible to fully automate the DM into independently run and update the data every month. At this moment, and in particular for the DM, the monthly updates are still made by manually running the project corresponding SAS EGuide programs. An implementation of such mechanism would be highly beneficial for the team, as it would help to reduce greatly the execution period of the tasks.

Additionally, a regular and automatic data validation would also bring great value to the end-product, again focusing on the DM, that is responsible for presenting and storing all the CP portfolio information. This type of reassurance is crucial and could easily reduce the number of errors to the minimum possible, providing a higher security and confidence in the results presented from the PBA team to the rest of the company. Even though several validation procedures were carried out during the development of the DM and TDB, these were implemented as a test mainly for the deployment of both elements, and not as common or regular procedure.

Lastly, given that this was a project designed to assist and improve decision making within the company, it makes sense that neither of the project elements can be stagnant. Therefore, it is recommended that a rather regular evaluation and assessment of the utility of the information provided is implemented. In a constantly evolving business area like Insurance, there is always new information that can be considered to bring new value to an analysis.

7.3. Feedback

Considering the importance of the project already emphasized, the successful delivery of this project was extremely appreciated and valued, particularly by my manager and also the head of Operations of the department. They both reinforced the need for these tools inside the company. In addition, the effective introduction of policies' coverages into both analyses was also identified as a major accomplishment, as it was a key element of the business that was lacking, until this point, for the remaining LOBs.

8. BIBLIOGRAPHY

Antonio, K., & Beirlant, J. (2006). Risk Classification In Non-Life Insurance.

Abadi, D. J., Madden, S. R., & Hachem, N. (2008). Column-stores vs. row-stores: How different are they really? Proceedings of the ACM SIGMOD International Conference on Management of Data, 967–980. https://doi.org/10.1145/1376616.1376712

Achampong, E. (2017). Methodological Framework for Artefact Design and Development in Design Science Research General Medical Research View project Cloud computing and Electronic Health Records in Developing Countries View project. https://www.researchgate.net/publication/329775397

Cabibbo, L., & Torlone, R. (2004). On the integration of autonomous data marts. Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM, 16, 223–232. https://doi.org/10.1109/SSDM.2004.1311214

Carstensen, A. K., & Bernhard, J. (2019). Design science research–a powerful tool for improving methods in engineering education research. European Journal of Engineering Education, 44(1–2), 85–102. https://doi.org/10.1080/03043797.2018.1498459

David, M. (2015). Auto Insurance Premium Calculation Using Generalized Linear Models. Procedia Economics and Finance, 20, 147–156. https://doi.org/10.1016/S2212-5671(15)00059-3

Evergreen, S., & Metzner, C. (2013). Design Principles for Data Visualization in Evaluation. New Directions for Evaluation, 2013(140), 5–20. https://doi.org/10.1002/EV.20071

Firestone, J. M. (1997). Data Warehouses and Data Marts: A Dynamic View.

Genero, M., & Piattini, M. (n.d.). A Survey of Metrics for UML Class Diagrams. JOURNAL OF OBJECT TECHNOLOGY, 4(9), 59–92. Retrieved January 26, 2023, from http://www.jot.fm/issues/issue_2005_11/article1

Haji Ali, N., bin Idris, S., Shukur, Z., & Idris, S. (2007). Assessment System For UML Class Diagram Using Notations Extraction. IJCSNS International Journal of Computer Science and Network Security, 7(8). https://www.researchgate.net/publication/253243639

Hightower, R., & Shariat, M. (2007). Conceptualizing Business Intelligence Architecture. https://www.researchgate.net/publication/237013221

Hobbs, J. (2020). Insurance of Goods in Transit. Insurance Disputes, 581–600. https://doi.org/10.4324/9781003122906-22

Huang, S. C., McIntosh, S., Sobolevsky, S., & Hung, P. C. K. (2017). Big Data Analytics and Business Intelligence in Industry. Information Systems Frontiers, 19(6), 1229–1232. https://doi.org/10.1007/S10796-017-9804-9/METRICS

Kappelman, L. (2021). Issues, Investments, Concerns, & Practices of Organizations and their IT Executives 2022 Comprehensive Report: Results and Observations from the SIM IT Trends Study

IT TRENDS STUDY RESEARCH TEAM The 2021 SIM IT Trends Study The 2022 Comprehensive Report: Issues, Investments, Concerns and Practices of Organizations and their IT Executives. http://www.simnet.org/IT-Trends

Khalaf, A., Alnazer, M., Khalaf Hamoud, A., Kamil Hussein, M., Alhilfi, Z., & Hassan Sabr, R. (2021). Implementing data-driven decision support system based on independent educational data mart Image processing View project the optimum encryption method for image compressed by AES View project Implementing data-driven decision support system based on independent educational data mart. Article in International Journal of Electrical and Computer Engineering, 11(6), 5301–5314. https://doi.org/10.11591/ijece.v11i6.pp5301-5314

Kumar, S. M., & Belwal, M. (2018). Performance dashboard: Cutting-edge business intelligence and data visualization. Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017, 1201–1207. https://doi.org/10.1109/SMARTTECHCON.2017.8358558

Liang, T. P., & Liu, Y. H. (2018). Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. Expert Systems with Applications, 111, 2–10. https://doi.org/10.1016/J.ESWA.2018.05.018

Lim, E. P., Chen, H., & Chen, G. (2013). Business Intelligence and Analytics. ACM Transactions on Management Information Systems (TMIS), 3(4). https://doi.org/10.1145/2407740.2407741

Luján, S. (2005). Data Warehouse Design with UML. http://www.unlpam.edu.ar/

Malik, S. (2005). Enterprise Dashboards. In John Wiley and Sons Inc (Vol. 1). John Wiley and Sons.

Mdletshe, S., & Oliveira, M. (2020). The development of a computer-based teaching simulation tool to aid medical imaging educators in teaching pattern recognition. International Journal of Morphology, 38(5), 1258–1265. https://doi.org/10.4067/S0717-95022020000501258

Meyers, L. S., Gamst, G., & Guarino, A. J. (2009). Data analysis using SAS enterprise guide. Cambridge University Press.

Midway, S. R. (2020). Principles of Effective Data Visualization. Patterns, 1(9). https://doi.org/10.1016/J.PATTER.2020.100141

Naidoo, J., & Campbell, K. (2016). Extended abstract: Best practices for data visualization. IEEE International Professional Communication Conference, 2016-November. https://doi.org/10.1109/IPCC.2016.7740509

Ong, I. L., Siew, P. H., & Wong, S. F. (2011). A Five-Layered Business Intelligence Architecture.

Rikhardsson, P., & Yigitbasioglu, O. (2018). Business intelligence ampamp; analytics in management accounting research_ Status and future focus. https://doi.org/10.1016/j.accinf.2018.03.001

Sidorova, A., & Torres, R. R. (n.d.). Business Intelligence and Analytics: A Capabilities Dynamization View.

Troyansky, O., Gibson, T., & Leichtweis, C. (n.d.). QlikView your business : an expert guide to business discovery with QlikView and Qlik Sense. Retrieved January 26, 2023, from https://books.google.com/books/about/QlikView_Your_Business.html?hl=pt-PT&id=RpQvCgAAQBAJ

vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. 1–13. https://doi.org/10.1007/978-3-030-46781-4_1

Werner, G., Claudine Modlin, M., & Willis Towers Watson, M. (2016). BASIC RATEMAKING.

APPENDIX

Variable Description	Example	Original vs Created
Payment Plan Type	Annual; Monthly	Original
Payment Type	Direct Debit; Card	Original
Renovation Type	Temporary, One year and Beyond	Original
Traffic Light Identifier within the LOB	Green; Yellow; Red; Black	Original
Traffic Light Identifier within the company	Green; Yellow; Red; Black	Original
Co-Insurance Type (when applicable)	Leader; Non-Leader	Original
Co-Insurance Percentage (when applicable)	0.3; 0.6; 0.65	Original
Identifier Flag of Co-Insurance	1; 0	Created
Identifier Flag if Leader in Co-Insurance	1; 0	Created
Annuity Identifier	1; 2; 5	Original
Policy Start Date	10-03-1997	Original
Policy Renewal Date	12-12-2020	Original
Policy Situation	Active; Mid Term Cancelled; Renewal Cancelled	Created
Policy New or Renewal Status	New; Renewal	Created
Policy Cancelation Date	13-12-2021	Original
Policy Cancelation Reason	Lack of payment; Cancellation	Original

Table 4 – Policy Variables Segment Table

Variable Description	Example	Original vs Created
Client Type	Commercial Line; Personal Line	Original
DNI Client Number (National Identification Document)	999999999	Original
PH's NIF (Fiscal Identification Number)	999999999	Original
PH's Birthday	16-06-1971	Original
PH's Age	52	Created
PH's Gender	Male; Female	Original

PH's Marital Status	Married; Divorced	Original
PH's CAE (Economic Activity Classification)	86230; 68100	Original
PH's CAE Letter	G; Q; L	Original
PH's CAE Sector	862; 477	Original

Table 5 – PH Variables Segment Table

Variable Description	Example	Original vs Created
Agent Identifier	99999	Original
Agent Name	XXXXXXX	Original
Agent Type	Multi-Brand; Non-Exclusive	Original
Commercial Channel	Private; Multi-Brand Network	Original
Commercial Area	Center; North; Islands	Original
Commercial Network	RE; PR; CO	Original
Commercial Manager	XXXXXXX	Original

Table 6 – Commercial Structure Segment Table

Variable Description	Example	Original vs Created
Object Type	Content; Building	Original
Object Type (including 'Aggregated' if a policy has both 'Building' and 'Content)	Content; Aggregated	Original
Risk Unit's Postal Code	2615-164	Original
Risk Unit's 4 Digit Postal Code	2615	Created
Risk Unit's District	Porto; Lisboa	Original
Risk Unit's Town	Matosinhos; Lisboa	Original
Risk Unit's CAE	68100	Original
Risk Unit's CAE Letter	L	Original
Risk Unit's CAE Section	Actividades Imobiliárias	Original
Object's Year of Construction	1994	Original
Object's Age	29	Created

Object's Building Type	Mixed materials; Cement boards	Original
Protection Measures	Without Measures; Fire	Original

Table 7 – Object Variables Segment Table

Variable Description	Example	Original vs Created
Coverage's Commercial Premium	9999€	Original
Coverage's Sum Insured	9999€	Original
Coverage's Exposure	0.45	Created
Coverage's Earned Premium	9999€	Created

Table 8 – Coverage Variables Segment Table

Variable Description	Example	Original vs Created
Claim Identifier	18AXXX000420	Original
Claim Cover Identifier	18AXXX000420/001	Original
Claim Cover Total Cost	9999€	Original
Claim Total Cost	9999€	Created
Number of Claim Covers	1; 2; 3	Created
Identifier Flag of Claim with no Exposure	0; 1	Created

Table 9 – Claim Variables Segment Table

```
PROC FORMAT;
     VALUE $EstadoCivil
       'S'   = "S: Single"
       'C'   = "C: Married"
       'V'   = "V: Widower"
       'D'   = "D: Divorced"
       'O'   = "O: Other"
       'E'   = "E: Company"
       ''    = "NA";

RUN;
```

Figure 16 – 'EstadoCivil' Format Example

Coverage Grouping Analysis Table

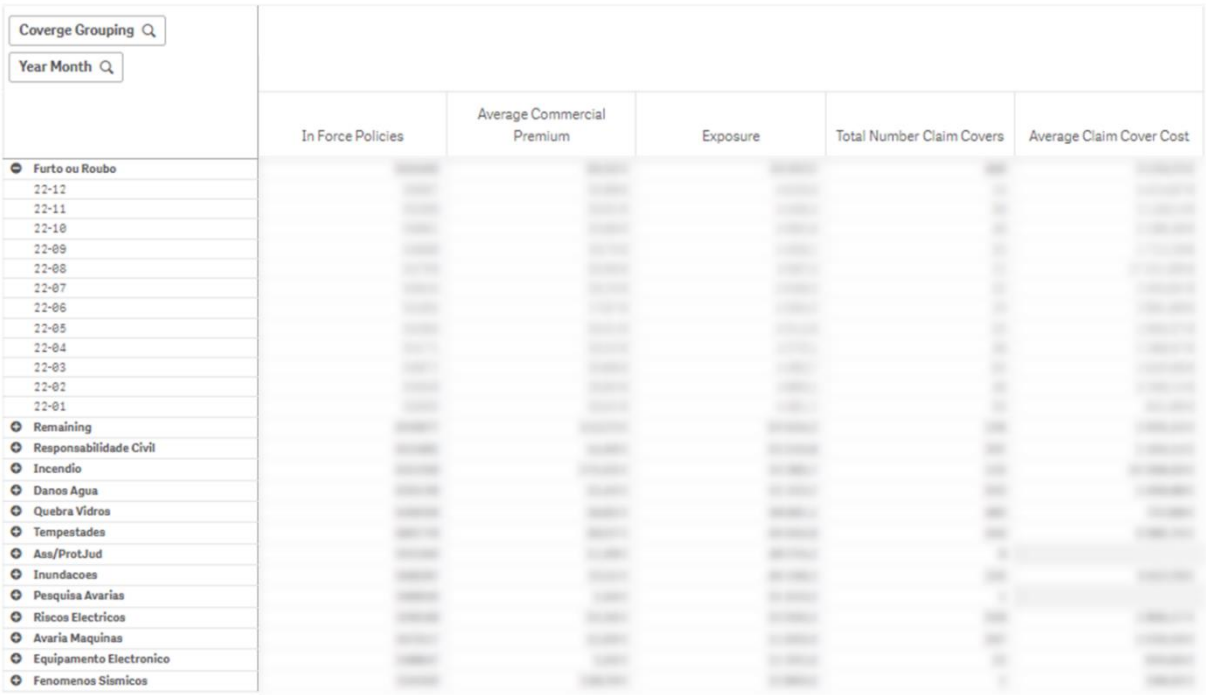

Figure 17 – Coverage Grouping Analysis Table for 2022

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa