University Institute of Lisbon

Department of Information Science and Technology

Self-Service Business Intelligence Data Analytics

Pedro Daniel Gonçalves Damásio

Dissertation in partial fulfillment of the requirements for the degree of

M.Sc. in Computer Science and Business Management

Supervisor

João Carlos Amaro Ferreira

ISCTE-IUL

October, 2019

Acknowledgements

I would like to thank my supervisor, Professor João Carlos Amaro Ferreira for the availability, guidance and patience shown.

Secondly, I would like to thank Tecmic for always showing availability and for helping me and supporting me throughout this work whenever it was needed.

Finally, a special thanks to my family, friends and colleagues for supporting me during this work.

Abstract

Information systems bring competitive advantages to companies when they know how to use them to their full potential. Business intelligence systems can garner large volumes of data and convert it into information but have the disadvantage of having associated high implementation costs both structurally and qualified people. Then comes Self Service Business Intelligence which provides companies with much lower costs of extracting information from their systems.

This work arises from Tecmic's need to extract information in the form of reports and dashboards from the incident system it has currently implemented. This system has limited information extraction and handling capabilities to help company managers make informed decisions about customer reported incidents.

In order to address this problem, it was proposed to implement the Self-Service Business Intelligence software from Microsoft, Power BI.

This work details all the preparatory steps to import the data into Power BI as well as an analysis of the results obtained. Additionally, through reports created in Power BI, it is explored what inefficiencies can be tackled in order to improve company performance and drive down the number of services created.

Keywords: Power BI, Self Service Business Intelligence, Incident Management, Data Analysis

Resumo

Os sistemas de informação trazem vantagens competitivas para as empresas que retiram máximo proveito deles. Os sistemas de *business intelligence* podem reunir grandes volumes de dados e convertê-los em informação, mas têm a desvantagem de ter elevados custos associados, estruturalmente e de trabalhadores qualificados. Surge então o *Self Service Business Intelligence*, que possibilita às empresas custos muito mais baixos de extrair informação de seus sistemas de informação.

Este trabalho surge da necessidade da Tecmic de extrair informação sob forma de relatórios e *dashboards* do sistema de incidentes que tem atualmente implementado. Este sistema tem capacidades limitadas no que diz respeito à extração e tratamento de informação o que dificulta os gestores da empresa de tomar decisões informadas sobre incidentes relatados pelos clientes.

Para resolver esse problema, foi proposto implementar o software de *Self-Service Business Intelligence* da *Microsoft*, o *Power BI.*

Este trabalho detalha todas as etapas preparatórias para importar os dados para o *Power BI*, bem como uma análise sobre os resultados obtidos. Adicionalmente, por meio de relatórios criados no *Power BI*, ssão exploradas as ineficiências que podem ser enfrentadas para melhorar o desempenho da empresa e diminuir o número de serviços criados.

Palavras-chave: Power BI, Self Service Business Intelligence, Gestão de Incidentes, Análise de dados

Contents

List of Figures

List of Tables

Abbreviations

BI: Business Intelligence
CRISP-DM: Cross-industry Standard for Data Mining
CRM: Customer Relationship Management
DM: Data Marts
DW: Data Warehouse
ERP: Enterprise Resource Planning
ETL: Extract-Transform-Load
IS: Information Systems
IT: Information Technology
ITS: Issue Tracking System
ODS: Operational Data Store
OLAP: Online Analytical Processing
PBI: Power BI
PQE: Power Query Editor
QA: Quality Assurance
SQL: Structured Query Language
SSBI: Self-Service Business Intelligence

Chapter 1

Introduction

Information Systems (IS) is one of the most relevant components of the current business environment. They offer great success opportunities for companies, given that they have the capability of collecting, processing, distributing, and sharing data in an integrated and timely manner. Furthermore, they help narrow geographical gaps, allowing employees to be more efficient, which is reflected in an improvement of the processes, administration, and the management of information, thus resulting in a positive impact on the productivity and competitiveness of companies [1] [19].

Over the past decades, Information Technology (IT) services have become an essential part of enterprise operations. Consequently, the efficient operation of these services represents a crucial component of corporate success, reflected in the vast IT support (or IT helpdesk, IT service desk) departments found at medium-sized and large companies. The purpose of service desk employees is to support other departments with problems arising from the introduction and use of hardware and enterprise software products [9].

Decision-making plays an important role in the competitive business environment as it ensures improved business performance. This has created the need to improve technologies to use the available data in order to improve decision-making [15]. One key technology is Business Intelligence (BI), which supports organizations in understanding large volumes of information, retrieving and analyzing information and making decisions. The information is usually in form of dashboards that organize and present information in an easy to read manner. The dashboards provide multiple visual components such as charts and tables on a single view so that information can be easly seen and monitored. [5]

1.1 Motivation

Tecmic is a Portuguese multinational company with sustained growth for over 25 years and has international recognition for their innovation and development capabilities. The company's activity is related to remote systems of monitoring and management of vehicle fleets, waste collection and urban public safety management, among others. It operates in several sectors such as public administration, waste collection and urban management, private security, distribution, logistics and freight transport, among others.

The company has developed a system of incidents tailored to its needs where it records all incidents related to the remote devices installed in the vehicles of its customers. All incidents are registered in a relational database in SQL (Structured Query Language) Server where they can be queried later. Querying information stored in the database is very limited through the application that manages the incidents because it does not allow the extraction and processing of information in order to perceive inefficiencies and problems in the service provided by the company. Because of this, the company intends to use a tool that operates on top of the SQL database and allows the creation of reports and dashboards in order to make better decisions and improve the service provided.

1.2 Objectives

The work carried out has two objectives. The first is to use a software that allows the extraction of information from the database that records incidents. This software extracts information and transforms it into graphics and dashboards, which simplifies the interpretation of the information presented. With this software, we intend to bridge the existing gap in Tecmic's incident system as well as to equip workers with information for decision-making.

The second objective of this work is to perform a data analysis of the data in the incident system using the software used in the first objective. This data analysis will be conducted through reports and dashboards and will aim to obtain more information of the data and its importance for the company's day-to-day operations and through this information, provide recommendations and withdraw conclusions on reported incidents.

1.3 Dissertation structure

This document is divided into structured chapters with the following information. In Chapter 1, an introduction is made to work carried out, and the motivation, objectives and research questions of the project are described. Chapter 2 includes the literature review where the structure and functionality of a system of incidents and business intelligence and self-service business intelligence are addressed. The methodology used in this work is explained in Chapter 3. This includes all the steps taken before creating the reports and dashboards. In chapter 4, the implementation is made, and an explanation of the various dashboards created contextualizing them in the company's business is given. Chapter 5 discusses data analysis, where data is explored in the incident system. Finally, in Chapter 6 the conclusions of the work carried out are drawn.

Chapter 2

Literature Review

This section of the document presents the literature review conducted to gain knowledge about issue tracking systems and incident management as well as the different types of BI available.

2.1 Issue Tracking System

Commercial software development is, by and large, an organization task. Project managers make requirements, programmers carry out functions, and Quality Assurance (QA) teams locate and check problems along with their connected fixes. Due to the high number and wide range of stakeholders involved with this form of group work, interaction, along with cooperation, develop into a fundamental piece of the process [13].

Besides demanding the initiatives of the many, often the development of high-quality software also calls for the mindful supervision of countless small related elements. Features and implementation responsibilities are generally born from requirements and flaws or bugs to tend to be unavoidably uncovered. These kinds of flaws are usually handled by making use of an Issue Tracking System (ITS). ITS assists software teams deal with issue reporting, assignment, monitoring, resolution, along with archiving through a trusted, distributed to-do checklist which also serves as a good archive of completed tasks [13].

2.1.1 What is an Issue Tracking System?

A good ITS is a software program that enables an organization to document and follow the development of each trouble or issue that the user of a computer system recognizes so that the issue is solved. ITS originated from manufacturing as being a paper-based reporting method, today the majority are web-based and related to Customer Relationship Management (CRM) environments, like call centres or even e-business websites, as well as with high-level technology environments including network operations centres [14].

By having an ITS, an issue, which may be nearly anything from a straightforward consumer issue to an in-depth technical report associated with an error or bug, is often followed through priority status, owner or some different personalized conditions. Issues that remain unsolved, maintain the “open ticket” in the work queue along with other issues of highest priority taking precedence in terms of work flow [14].

An ITS usually offers the user the possibility to record an issue, monitor progress toward its resolution, as well as understand who might be accountable for solving the issue. Furthermore, it enables the manager of the system to customize the tracking technologies to ensure unnecessary documentation on the part of the issue solvers will not turn into a waste of time. Numerous businesses use ITS software, including software developers, suppliers, IT help desks, along with service providers. Issue tracking offers some positive aspects like the capability to graph and or chart the improvement of people as well as the general resolution of the issue. Additionally, it provides the ability to monitor and keep end to end historical past with the issue to offer the people responsible for the issue, an extensive viewpoint of the measures taken to help solve the issue until that time. Another advantage of having issue history is usually the ability to recognize the kind of issues that the client has confronted in past times and the pattern in cases where any is available. Another benefit associated with issue tracking would be the capacity to create a report to the service provider to have a complete perspective regarding tickets and any filters that are being searched for. These kinds of reports also make it possible for the upper management to acquire a perspective of the work carried out and in case of any escalations from the consumer, it becomes an easy task to run through all these reports to acquire information on the case as well as properly manage the case [2].

2.1.2 Incident Management

An incident is an unplanned interruption or reduction in the quality of an IT service, and this leads to the creation of issues on an ITS that will be dealt with specific and skilled people. Incident Management (IM) is the process for dealing with all incidents and the primary goal is to restore normal service operation as quickly as possible and minimize the impact on business operations. IM can be triggered in a many different ways, the most common being through the user by phone, email or by using the incident logging application, however other methods can automatically identify incidents through IM tools. Much of the information used in IM comes from a tool adapted to it and this tool should implement at least incident recording and management capability [4].

2.1.3 Incident Management process flow

Figure 2.1 shows in detail the steps that should be followed in the IM process flow, from the incident identification and logging to closure [4].

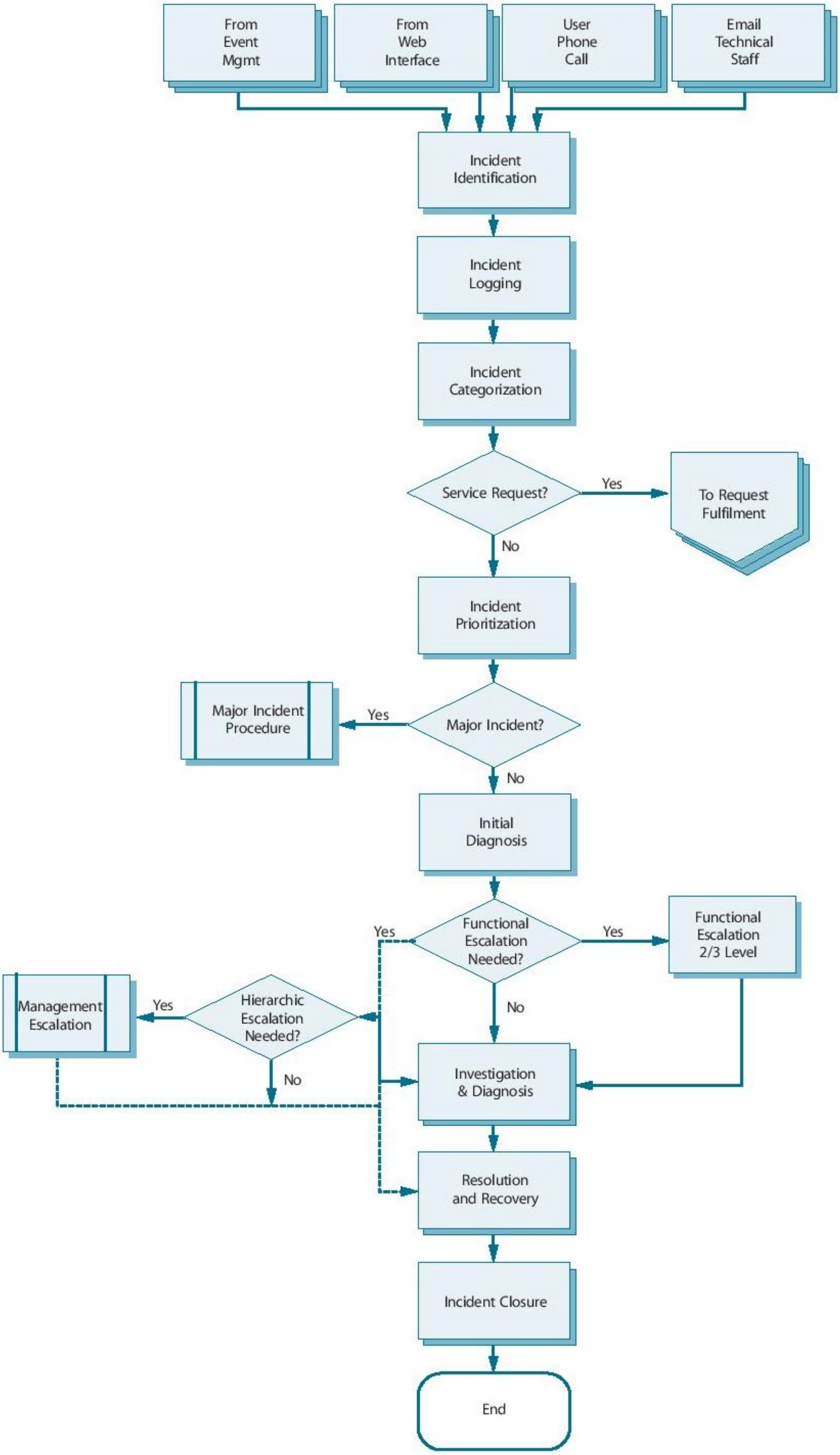

Figure. 2.1. Incident Management process flow [4].

Incident identification should be the first step, as work cannot start until the problem is identified. It is bad for the business to know about the incident from its customer's or Help Desk. It is much better to monitor all important components so that future problems can be avoided. The next step should be incident logging. This is when the system or person logs the date/time and important information about the current incident. It is important to have a full historical record of the incident because that way, it can be handled by other people. Some of the most important attributes of the incident are as follows: [4]

- Unique reference number
- Incident categorization
- Incident urgency
- Incident impact
- Incident prioritization
- Date/Time recorded
- Name/ID of the person and/or group recording the incident
- Method of notification (phone, automatic, e-mail, in person, etc.)
- Name/department/phone/location of user
- Call-back method (telephone, mail, etc.)
- Description of symptoms
- incident status (active, waiting, closed, etc.)
- Related CI
- Support group/person to which the incident is allocated
- Related problem/Known Error
- Activities are undertaken to resolve the incident
- Resolution date and time
- Closure category
- Closure date and time.

After logging the incident, a category should be assigned to it. The incident category is very important because later, when looking for incident types or frequencies, it is possible to establish trends and patterns. The categories used for an incident are different from organization to organization; however, a multi-level categorization with three to four levels of granularity should be used as shown for example purposes in Figure 2.2 [4].

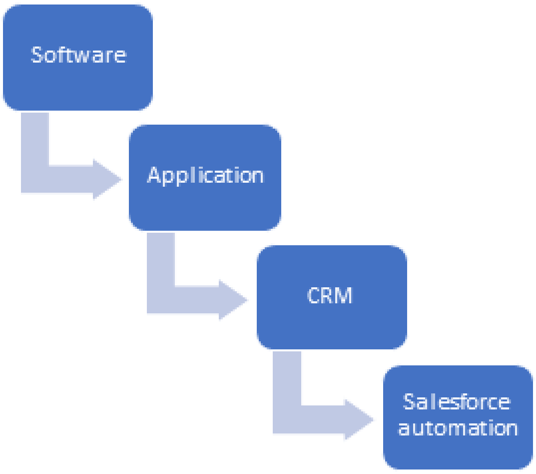

Figure. 2.2. Multi-level incident categorization, adapted from [4].

Another important step is to prioritize every incident as it will be necessary to allocate a person with the proper experience and expertise. The priority is usually determined by taking into account the urgency of the incident and the level of impact it is causing on the users. Sometimes the number of users affected by the incident might not be relevant as a single user with a great influence on the daily business operations can have a huge impact. A sample of what a priority system should be can be seen in Table 2.1 [4].

<table>
<tr><td colspan="5">Impact</td></tr>
<tr><td colspan="2"></td><td>High</td><td>Medium</td><td>Low</td></tr>
<tr><td rowspan="3">Urgency</td><td>High</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>Medium</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td>Low</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Priority Code</td><td colspan="2">Description</td><td colspan="2">Resolution Time</td></tr>
<tr><td>1</td><td colspan="2">Critical</td><td colspan="2">1 hour</td></tr>
<tr><td>2</td><td colspan="2">High</td><td colspan="2">8 hours</td></tr>
<tr><td>3</td><td colspan="2">Medium</td><td colspan="2">24 hours</td></tr>
<tr><td>4</td><td colspan="2">Low</td><td colspan="2">48 hours</td></tr>
<tr><td>5</td><td colspan="2">Planning</td><td colspan="2">Planned</td></tr>
</table>

TABLE 2.1: Priority coding system, adapted from [4].

Although the table presented can be consulted, there should be examples so that people dealing with problems can more easily determine levels of urgency and their impact. The fact that the priority is dynamic depending on the circumstances makes the team have to adapt in order to be able to solve problems with greater urgency or impact. The diagnostic is first done by phone if the incident is reported to the Help Desk with the user on the phone. If done this way, it is better to try to discover and take notes of every detail that caused this incident to understand exactly what went wrong and how to correct it. At this stage, it could be useful for the Help Desk to have a script for the most known types of issues and how to solve them so that a solution is given to the user on the phone. If the Help Desk cannot solve the incident, then the user should be informed and wait for further information. Investigation and diagnosis are the next steps taken, and this is when the support team gets involved with the incident. An incident history should always be maintained as this could be useful for future incidents of the same type. Some steps to solve the incident could be to try to investigate how the user identified the anomaly, identify the chronology of events until the identification of the anomaly, test (if applicable) the recurrence of the anomaly, and identify events or actions that may have triggered the anomaly. If it is not possible to identify the cause or find a solution, the incident should be immediately escalated to another support line with the necessary skills to solve it. Resolution and recovery are when a potential resolution has been identified and should be applied and tested. Even if the solution is found and tested, further testing should be done to prevent another failure and ensure the service is fully restored. The last step is to close the incident as this is usually done by the Help Desk. They should verify that the service has been restored and that the users are satisfied, and everyone agrees that the incident can be closed. The incident history should also be checked and updated if needed to ensure all information is correct so that it could be used to solve or prevent similar incidents. Even with all this care, an incident could recur, and in these special cases, it is recommended that very specific steps are defined and followed, and everyone involved agrees on the procedure to take to close the incident as fast as possible [4].

In the next chapter, it is given an overview of the traditional BI systems and their architecture. The next section also explains what Self-Service Business Intelligence (SSBI) is, what motivates organizations in adopting it and some of the leading tools available to the market.

2.2 Business Intelligence

Constant changes, as well as complexion in the business environment, pushed most business to make decisions quicker. All these actions associated with decision making demand a lot of relevant data, information knowledge, and wisdom. Procedures inside the framework need to be carried out successful and effective, and usually, these types of procedures require computerized support forcing businesses to need powerful decision making. As a solution, distributors began to develop enterprise-wide systems with visualization, notifications and performance measurement functions. These types of products or services are usually underneath the term BI [16].

As there is no uniform definition of BI, it can mean different things to different people. The following two definitions of BI are cited from Gartner and Forrester Research.

According to Gartner [7] "Business intelligence is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance."

Forrester Research says [6] that BI is "a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision-making.".

2.2.1 The architecture of a general Business Intelligence solution

There are several existing BI architectures, and they are different in their structures, such as layers, components, processes, and relationships. However, there are some common components among them like source systems, data storage, and reporting tools [12].

This paper will focus on the framework of five layers proposed by [12] as illustrated in Figure 2.3. The five layers are data source, Extract-Transform-Load (ETL), data warehouse (DW), end-user, and metadata and will be described in the next section.

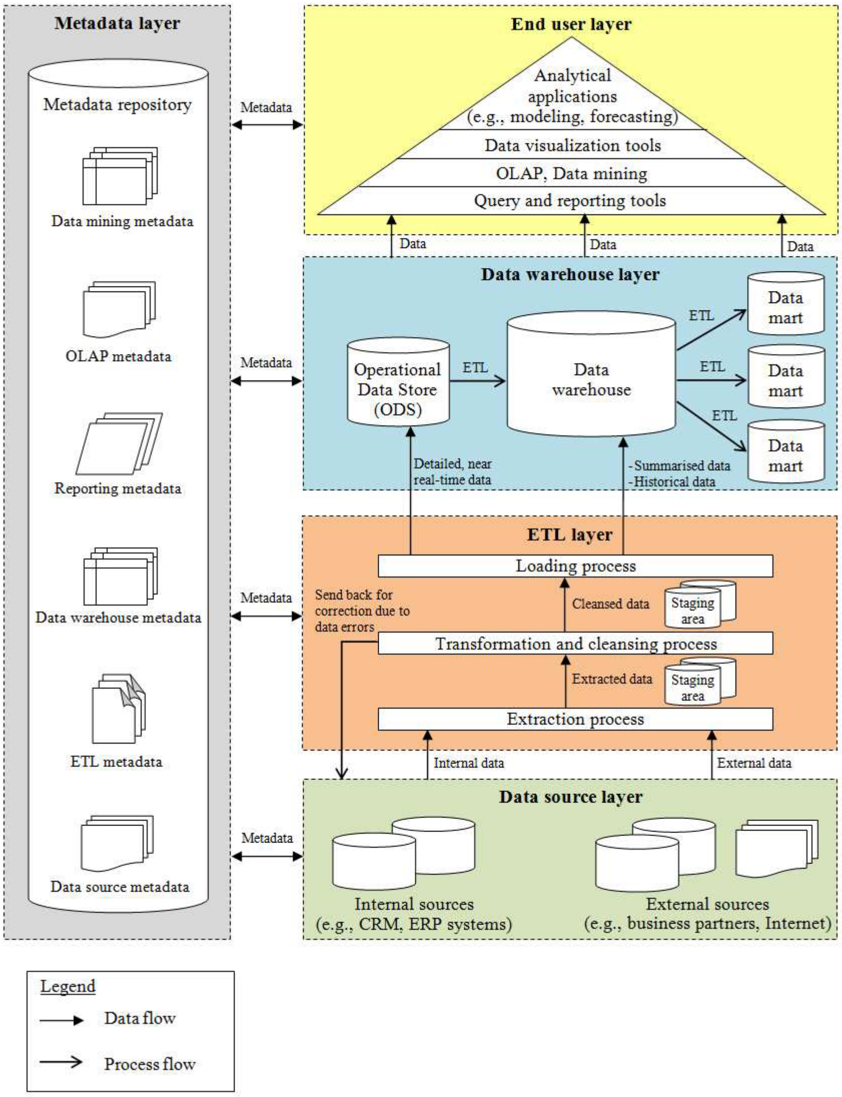

Figure. 2.3. Proposed BI Architecture [12].

2.2.2 BI architecture

The **data layer** used in BI applications can contain data that comes from internal or external sources. Internal sources are related to business operations and come from CRM and Enterprise Resource Planning (ERP) systems. These are operating systems and contain only current data that is used to support the daily business operations of an organization. External data is collected from sources outside of the organization such as business partners, syndicate data suppliers, the internet, governments, and market research organizations and are often related to competitors, market, environment, and technology. The **ETL layer** focuses on three main processes: extraction, transformation, and loading. Extraction is the process of identifying and collecting relevant information from different sources. Transformation is the process of converting data using a set of business rules into consistent formats for reporting and analysis. The transformation process includes defining business logic for data mapping and standardizing data definitions to ensure consistency across an organization. Loading is the last phase of the ETL process, where data in the staging area are loaded into a target repository. The **DW layer** is made up of three components, Operational Data Store (ODS), DW, and Data Marts (DM). The ODS integrates data from the ETL layer and into the DW. Since the data in the ODS is frequently updated, it does not store historical information, and so it is mainly used for short-term decision-making from the middle management of an organization. The second component is one of the most important, the DW. It collects information from external and internal sources and stores aggregated, summarised and huge amounts of historical data. Contrary to the ODS, the DW contains long term data as it contains historical and current data. This is a subject-oriented, integrated, time-variant and non-volatile variety of data to back up the management decision-making process. The last component is the DM, and it is like a very tiny DW as it contains data from only one topic to support one department or function of an organization and only stores data for a very limited time (60 to 90 days). The **metadata layer** stores details about data regarding different data such as source as well as where it is being used, the changes made to it and exactly how they relate to each other. Business rules, information and data information, definitions and technical information are all located on this layer. The BI architecture is supported by various kinds of metadata including data source, ETL, reporting, Online Analytical Processing (OLAP) and data mining. Data source consist of information regarding access mode, the structure of data sets and referential constraints. The ETL is a log which is made from ETL tools. As the data is incorporated into the DW, and the data is being extracted, this log maintains records of the changes made to the data factor to guarantee the quality of the data. Reporting is XML based, and they are used to store templates as well as reporting descriptions including report name, start date, and end date. OLAP metadata offers explanations related to structure of cubes, dimensions, hierarchies, levels, as well as the sort of drill paths being taken. Data mining metadata consist of information regarding algorithms and queries. All these metadata also incorporate information regarding structures of charts and queries. The **end-user layer** is a pyramid that consists of tools that can display information accordingly for the different levels of users inside an organization. The query and reporting tools are at the bottom of the pyramid and allow for quick information and reports to help decision making. OLAP and data mining are at the second level of the pyramid. OLAP allows users to analyze data fast and from different perspectives and compare it, as it shows this data in a user-friendly graphical way. The third level is data visualization tools that provide an overview of the business performance to managers and executives in the form of dashboards and scorecards [12].

2.3 Self-Service Business Intelligence

SSBI is a recent BI pattern, and it already grabbed huge attention from business, users and software vendors. What makes this type of BI so attractive to many people and what was the need for it? In the next sections, we are going to discuss this [10].

2.3.1 Definition of Self-Service Business Intelligence

The definition of SSBI is “The facilities within the BI environment that enable BI users to become more self-reliant and less dependent on the IT organization. These facilities focus on four main objectives: easier access to source data for reporting and analysis, easier and improved support for data analysis features, faster deployment options such as appliances and cloud computing, and simpler, customizable, and collaborative end-user interfaces.”. SSBI has four main objectives, easier to consume and enhance BI results, faster and easier management of DW, easier access to source data and easier to use BI tools as illustrated in Figure 2.4 [10].

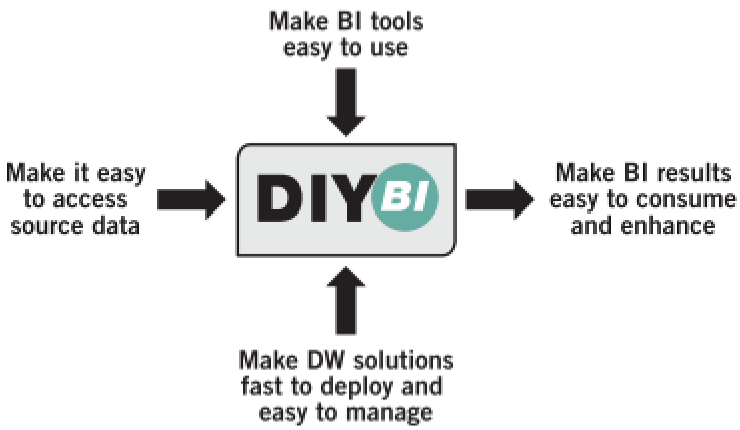

Figure. 2.4. The four objectives of SSBI [10].

To make BI results easy to consume, we must take into consideration that when users get information thrown at them, it makes it harder to understand, and that leads to poor decisions. An SSBI must be able to transform information, so it makes it quickly readable and understood by all users. The information is presented as a dashboard or reports, and it is highly customizable. This can be achieved with the simplistic and easy to use user interface of the software making workers self-sufficient [3] [10].

By making DW solutions fast to deploy and easy to manage, SSBI opens a whole new world to the users and organization as the solution provides good performance and scalability. From the organization perspective, when this type of BI is being accessed by the business community, applications are built that would not have been possible before and business can customize applications to their specific needs on their own timetable. From the user perspective, their satisfaction increases when they can explore, create, and manage reports and analytics [3] [10].

The final objective is to make the data sources easy to access. End users access information in a different manner than in a traditional BI environment, and this means that they don't require help from IT professionals to get this information. Although this does not mean that the IT department has no control over the information, they just support the end-user in whichever way necessary [3] [10].

2.3.2 The motivation for Self-Service Business Intelligence

A survey was done to understand what drives BI developers for a Self-Service platform for their users, as illustrated in Figure 2.5.

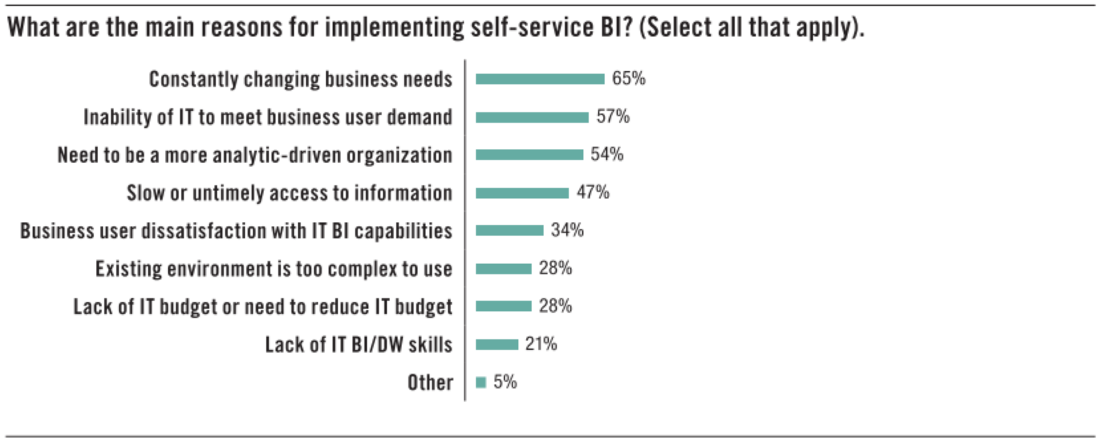

Figure. 2.5. Drivers for SSBI [10].

SSBI is the harmony between business needs and IT. Considering the dynamic changes in an organization environment, users would rather have access to real-time information to aid in decision making. IT professionals cannot fulfil these consistently shifting business needs making this the two top reasons for organizations to use an SSBI software. As the market gets more demanding and technology advances the process of decision-making is even more important for business, it must be fast and supported with the best information, if not, a decision made with poor or outdated information can cause enormous market loss, making businesses dependent on more data compared to before. With more data, businesses should react better as well as more intelligent to triumph competition as they need to count more on the analytical results than in the past. The traditional BI systems can be extremely difficult to use, as it requires skilled people and knowledge. These are the main reason organizations switch to an SSBI environment [10].

2.3.3 Overview of Self-Service Business Intelligence Tools

There is a lot of SSBI software in the market today, and in this section, we are going to have an overview of the three software leaders, Tableau, Qlik and Microsoft according to Gartner's Magic Quadrant as illustrated in Figure 2.6 [8].

Source: Gartner (February 2019)

Figure. 2.6. Magic Quadrant for Analytics and BI Platforms [8].

2.3.3.1 Tableau

Tableau provides an intuitive interactive visual-based exploration experience that permits organization users and just about any content creator to get into, put together, examine and present findings inside their information without having technical or coding abilities. Tableau provides three major products: Tableau Desktop, Tableau Server, and Tableau Online (its cloud offering). It has delivered a couple of enterprise functions which attracts IT customers such as licensed and recommended data sources for much better governance of large deployments, hybrid data support through the cloud, scheduling and alerting, and improved SDK and APIs, and extra collaboration, among other things. Tableau's primary product advantages keep on being its user-friendly interactive visualization and exploration, and analytic dashboarding functionality, for almost any repository, benefiting it's an in-depth group of data connectors, with both in-memory as well as direct query accessibility regarding bigger datasets. This combination, consisting of drag-and-drop advanced capabilities like forecasting, clustering, automated geocoding and assisted formula editing, permits users to try and do deep exploration and manipulation with their data easier and quicker compared to most competing platforms. Tableau may be deployed in the cloud, with Tableau Online, or on-premises. Its cloud deployment possibilities have advanced to provide pre-packaged virtual machines for Amazon Web Services and Microsoft Azure, to simplify deployment. In the past year, it has added in support for the Google Cloud Platform as well as hybrid data support to on-premises sources from the cloud. Although a wide variety of data source connectivity options are supported by Tableau, complex data modelling including multifactor table models should be developed inside a DW or through self-service data preparation partners. Additionally, bad efficiency regarding large in-memory extracts usually calls for modelling within an independent data repository which is directly queried through Tableau. Tableau is investing in, although not driving, the subsequent influx of disruptive innovation [8].

2.3.3.2 Qlik

QlikView is a BI data breakthrough solution for making guided analytics applications as well as dashboards tailor-made for business problems. The software allows the user to discover data insights as well as associations through numerous sources with QlikView's Associative Data Indexing Engine. This application reveals information which is not identified using query-based methods. QlikView also provides guided exploration, discovery as well as collaborative analytics intended for sharing information. Furthermore, this software permits customers to develop and even set up analytic applications without the need of specialized development expertise, assisting generate a quicker reaction to shifting organization demands, quicker time to value, and much more knowledge throughout an organization. QlikView features a trademarked in-memory technology which diminishes the need for IT individuals since the software possesses an inference engine to develop links with data automatically. The software consolidates information coming from numerous sources to deliver centralized information regarding advanced reporting. The user-friendly click through dashboards enables non-IT experts to recognize hidden tendencies [8].

2.3.3.3 Microsoft Power BI

Power BI (PBI) is a cloud-based SSBI solution. This means you may create and set up solution easily using data from the cloud as well as on-premises data sources, systems and application. All these are hosted in Microsoft's trust enterprise cloud. Every single Excel user could be the target of PBI, whether they have a need pertaining to SSBI features. Microsoft describes the target users of PBI in three groups: report creators like the data experts or even consultants, data stewards like data scientists, administrators or even IT specialists and report customers. PBI offers a reliable platform to share reports and analyses for a far better decision, which will stimulate the cooperation within the corporation, and everyone is able to become a report consumer by utilizing PBI [8].

BI seeks to find out foreseeable future tendencies according to previous findings regarding business process performance and predict the impact of business process changes. Through taking advantage of simulation, data mining, or optimization methods, BI concentrates on delivering longer-term predictions. To lower the delay involving a business event as well as the reaction to the event, real-time BI, also referred to as operational intelligence, have been suggested. Operational intelligence recommends real-time decision making and therefore brings out functionality for real-time, constant processes examination. Generally, operational intelligence procedures concentrate on the real-time prediction of aggregate indicators calculated for a group of process cases, for instance, key performance indicators which abstract from individual process performance measures. This implies, that operational intelligence techniques usually are unsuccessful in predicting the performance of individual process instances, and therefore simply provide constrained support to proactively tackle incidents associated with the execution of an individual process instance. Through taking into consideration process models which are mined through process logs, process mining enables dealing with circumstances where the actual process model is not predetermined or even where it is unknown [11].

Chapter 3

Methodology

The objective of this chapter is describe the methodology used to carry on the work done and to explain what was done for each step in the methodology. The methodology used is this work is an adaptation of the Cross-industry Standard for Data Mining (CRISP-DM) methodology as this is commonly used in solving analytical problems [17].

The CRISP-DM model proposes six steps that can be repeated if needed. The Figure 3.1 illustrates the model. An overview of what each step entails is given.

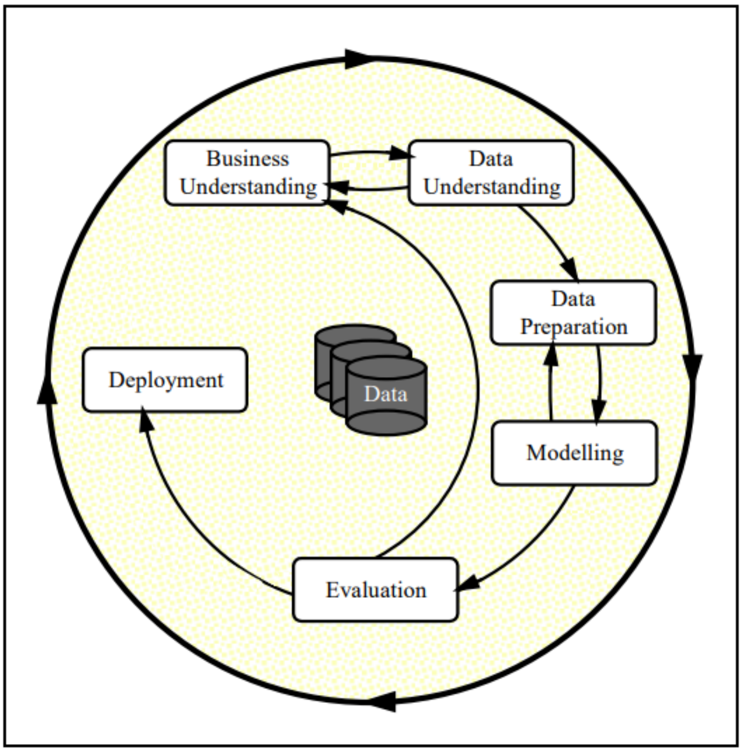

Figure. 3.1. CRISP-DM Process Model for Data Mining [18].

Business Understanding is the first phase where its primary function is to accomplish the business goals. It is an important step to take because this is where fundamental questions are asked to the organization and then find the right answers for the project. The lack of commitment at this stage can put at risk all future work. **Data Understanding** is where an understanding of all the data collected from the previous step is done. Some suggested methods to achieve this are describing the data and exploring it using tools and techniques for data visualization. There is a close link between this step and the previous one as they can be performed cyclic. It can become difficult to understand the data given and go back to the previous step to redo it. **Data preparation** consists on fixing data quality problems and transform it for the modelling phase. The objective of this phase is to obtain a final set of data based on the project goals. **Modelling** comes next and this is where a modelling technique is selected. Multiple can be used, and their parameters changed for better outcomes. Jumping back and forth between the modelling and data preparation is common because different techniques can give different results. **Evaluation** is as important as modelling. This is where the models are tested to see if they provide new information and can answer the problem identified. This step must be done thoroughly, review all the steps executed in the construction of the model to make sure it accomplishes the objectives of the business. **Deployment** is crucial for the project as this is where predictive analysis comes in and achieves the final goal of the business. During deployment it is important to monitor the results as incorrect use of the data provided can lead to various losses for the business [18].

As mentioned at the beginning of this chapter, the methodology used is an adaptation of CRISP-DM, because only three of the six steps will be performed with them being business understanding, data understanding and data preparation. This is due to the scope of the work done being focused on data analytics and extracting information of a database.

3.1 Business Understanding

At this stage it is intended to understand the objectives and requirements of the business in order to be able to reach a decision on which software to use. The goal of Tecmic is to extract information from its database where incidents reported by its customers are recorded. The need for information extraction arises because the incident system is limited in this regard. In order to define how it will meet the needs of the company, the author proposed to hold a meeting at Tecmic's premises. This meeting was attended by four people, three of them from Tecmic and the fourth person the author of this work. For reasons of anonymity the names of the people attending the meeting will not be disclosed. The three people from Tecmic are the administrator (person A), the director of the Electronics Development Unit (person B) and the third person from the Industrialization and Support Department (person C). Person A is the decision maker in the company, Person B is the one who has the technical knowledge of the incident system and the incident database model, and finally Person C is the person that uses the incident system on a daily basis. During the meeting requirements and restrictions were raised related to the software that should be used and these are as follows:

- Reporting capability in the form of graphics and dashboards
- Ease of use by operational staff
- Expandability to keep up with the incident tool growth
- Preference for Microsoft Tools
- On Premises solution

Acoording to the requirements and restrictions raised during the meeting and in agreement with those present, it was decided to use Microsoft SSBI software, PBI and thus concluding the first meeting.

The next step was to define what information should be presented in the dashboards to be created in the PBI and for this a new meeting with person B, person C and the author was scheduled. At this meeting person C was making the requests on what information needs to be present in PBI, and person B helped in guiding person C so that these requests could be acomplished with the information provided from the database and that the relationship model supports it. At the end of the meeting, the requests made from person C where translated into questions that could be answered through PBI and a total of nineteen questions are proposed and are presented below:

1. What is the percentage of services by technician?
2. What is the percentage of services by operator?
3. What is the percentage of services by state?

4. What is the percentage of services by type?
5. What are the causes of services?
6. What is the distribution of opened services by date?
7. What is the percentage of vehicles by type?
8. What is the percentage of vehicles by operator?
9. What is the percentage of vehicles by configuration?
10. What is the percentage of vehicles by registration date?
11. What is the percentage of vehicles by state?
12. What is the percentage of components by state?
13. What is the percentage of components by type?
14. What are the detail of components?
15. What is the percentage of tasks by type?
16. What is the percentage of tasks by result?
17. What is the percentage of tasks by operator type?
18. What is the percentage of tasks by operator?
19. What is the dristribution of tasks by date?

In the process of defining the questions presented, it was taken into consideration the data model where the incident system consumes and records information and this is the main reason why the questions are oriented to the main tables used by the system. In this meeting the author not only got the knowledge about the information to be extracted, but with the interventions of person B to orient the questions, it was possible to understand the relationships of the tables which leads us to the next step, data understanding.

3.2 Data Understanding

The process of understanding the data began in parallel with that of business understanding. This happened because at the meeting where the information to be reflected in the PBI were defined, it was also necessary to understand in which tables the necessary information is stored and how the tables relate to each other. All information used for this work comes from the incident system production database. It is a relational database that uses the Microsoft SQL Server Management system and contains 56 tables. Of these and with the questions defined and with the support of the knowledge of person B of the database, five main tables were identified, namely Services, Operators, Tasks, Vehicles and Components. In spite of this, it will be necessary to use other auxiliary tables, such as service types, vehicle status and others. Following is a brief contextualization of the five tables identified.

3.2.1 Services

The services table is where all incidents reported by Tecmic customers are recorded. This table has 22749 records and features a service across 36 attributes where most of them are Id's that relate to other tables.

3.2.2 Operator

The operator table contains personal information of the people involved in services. The most relevant information in this table is the typification given to an operator such as customer, supplier, installer and Tecmic. This table contains 321 records and characterizes an operator through 21 attributes.

3.2.3 Tasks

The task table is an extension of services because it provides additional information to a service. Most services have at least one associated task where it is typified, for example, maintenance, repair, and so on. The Task table also indicates the result of it by linking with the Task results table. In total, this table contains 22587 records, and a task is characterized by 18 attributes.

3.2.4 Vehicles

The table of vehicles similar to that of Operators as it contains information related to the characteristics of a particular vehicle, for example, license plate, colour, chassis and others. A row in this table represents a vehicle where a component has been or is installed. This table contains 8601 records and characterizes a vehicle through 40 attributes.

3.2.5 Components

This table contains information related to a given component, such as the model serial number and more. Usually, components are installed in vehicles and services performed are to carry out their installation, repair or maintenance. This table has 21823 records and features a component through 17 attributes.

A brief description of the five main tables identified for this work was given, but as mentioned earlier, it is necessary to use other auxiliary tables such as states or service typifications.

3.3 Data Preparation

Data preparation plays a key role in creating dashboards. In order to outline this step the author defined the process he used as shown in the Figure 3.2.

Figure. 3.2. Process detail for imported data into PBI.

For each of the steps defined in this process, a brief explanation of what each of them includes is given.

3.3.1 Connecting to the data source

In this step a link between the data source and the PBI file is made. The data source for this work was a copy of the company's production database to the local machine of the author held on September 9th, 2019. The Figure 3.3 shows an example of this process for the Services table.

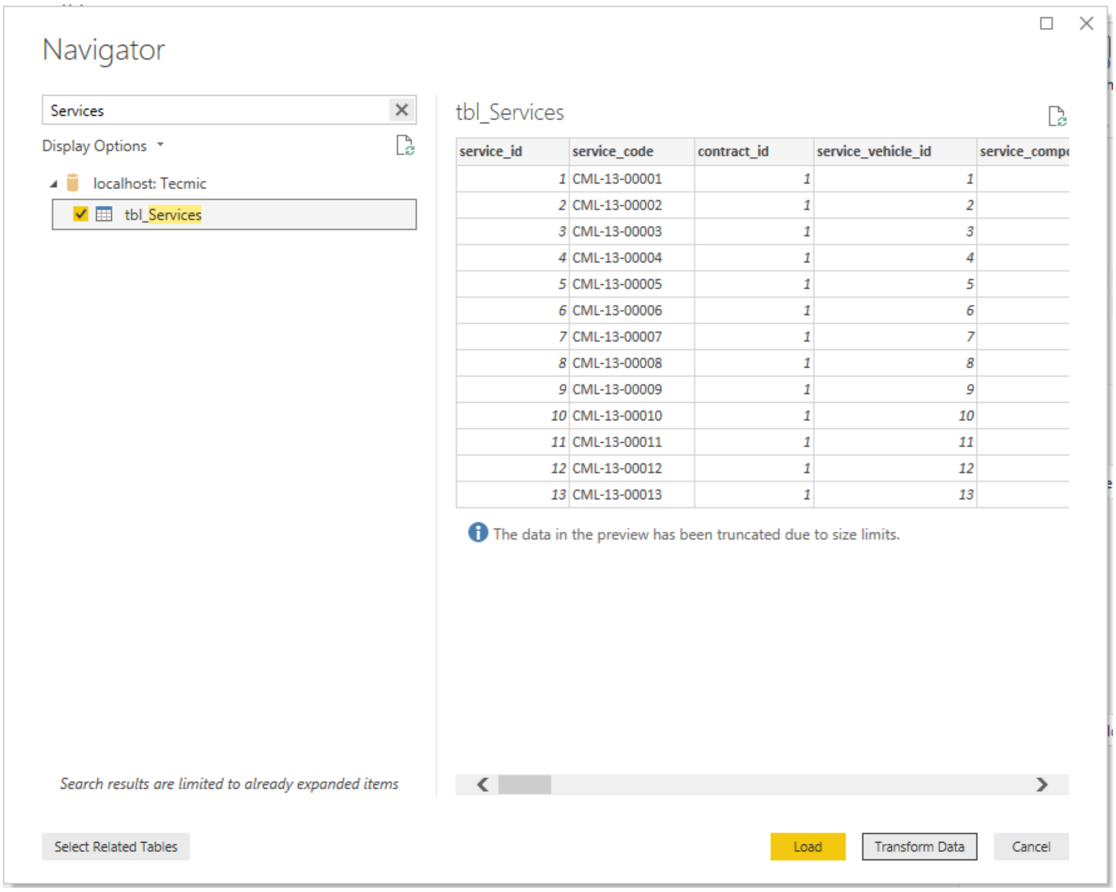

Figure. 3.3. Data transformation before importing into PBI.

3.3.2 Transforming data

The use of PBI allows the transformation of the data through the tool built into the software, Power Query Editor (PQE). To use this tool, a window appears when the table to import is selected. In this window, there is a button in the lower right corner, "Transform Data" that opens the PQE. The next steps to perform are data cleaning and data expansion and these were performed through PQE.

3.3.3 Cleaning Data

In the data cleaning process several columns from the main five tables have been removed. The criteria used for column removal are:

- Columns where all values are NULL
- Columns with descriptive content and inserted by a user such as incident description
- Metadata columns such as CreatedBy or Created In

The Table 3.1 summarizes the number of columns removed for each of the five main tables.

Table Name	Initial Number of Columns	Final Number of Columns
Services	36	21
Operator	18	5
Tasks	21	8
Components	17	8
Vehicles	40	10

TABLE 3.1: Comparison between initial and final number of columns in each main table.

3.3.4 Expanding Data

As previously mentioned, only 5 main tables that are the pillar for the work performed were identified. However, it is necessary to resort to auxiliary tables that contain information such as states and types of services, types of causes of services, and so on. As an example, the Services table has a relationship to a table called Service Type that contains all data for the types of services. After the merge is done between them, the Service table will have additional columns found in the Service Type table. If the reader is familiar with the SQL language, merging tables is the same as using a join between tables. The Table 3.2 summarizes the number of columns each table has after all the merges are done.

Table Name	Initial Number of Columns	Final Number of Columns
Services	21	48
Operator	5	7
Tasks	8	18
Components	8	12
Vehicles	10	36

TABLE 3.2: Comparison between initial and final number of columns in each main table after merging secondary tables.

Figure 3.4 shows a total of twenty-four tables imported into PBI. Of these, five are the principal ones previously identified, six are copies of the original but translated into English and are identified in the name by the word "Translation" and the remaining thirteen tables are auxiliary as they contain information such as types and states.

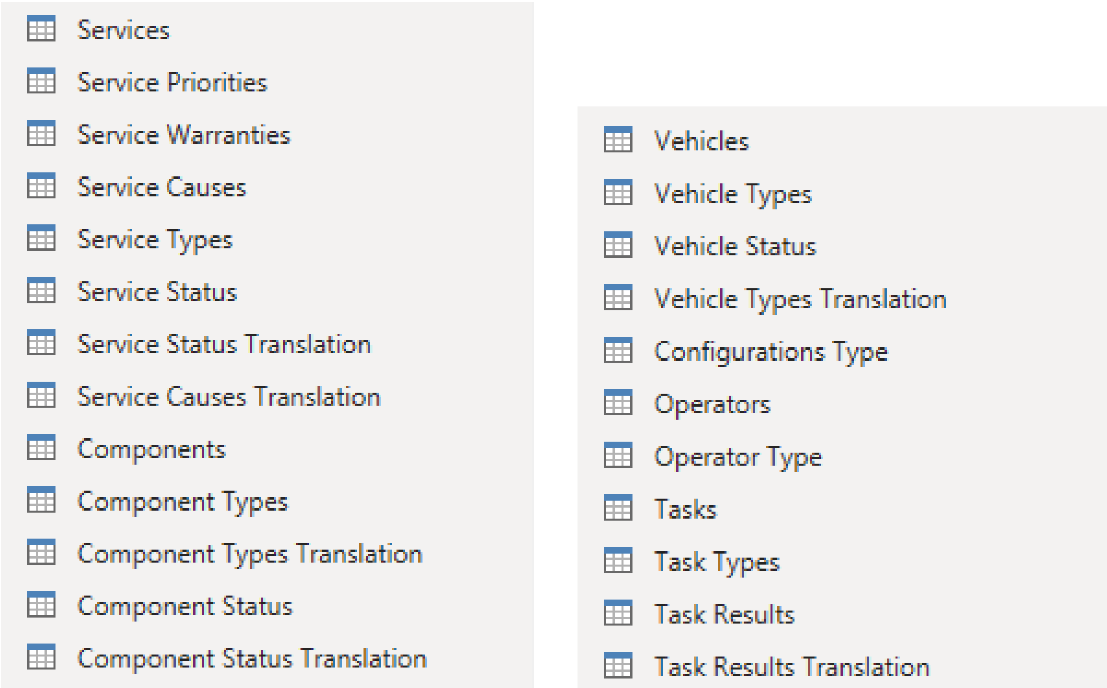

Figure. 3.4. Tables imported into PBI.

To get the data ready to work it was necessary to apply a series of transformations through the PQE. Figure 3.5 shows the transformations to the service table, Figure 3.6 to the components table, Figure 3.7 to the vehicles, Figure 3.8 to the operators, and finally Figure 3.9 to the tasks.

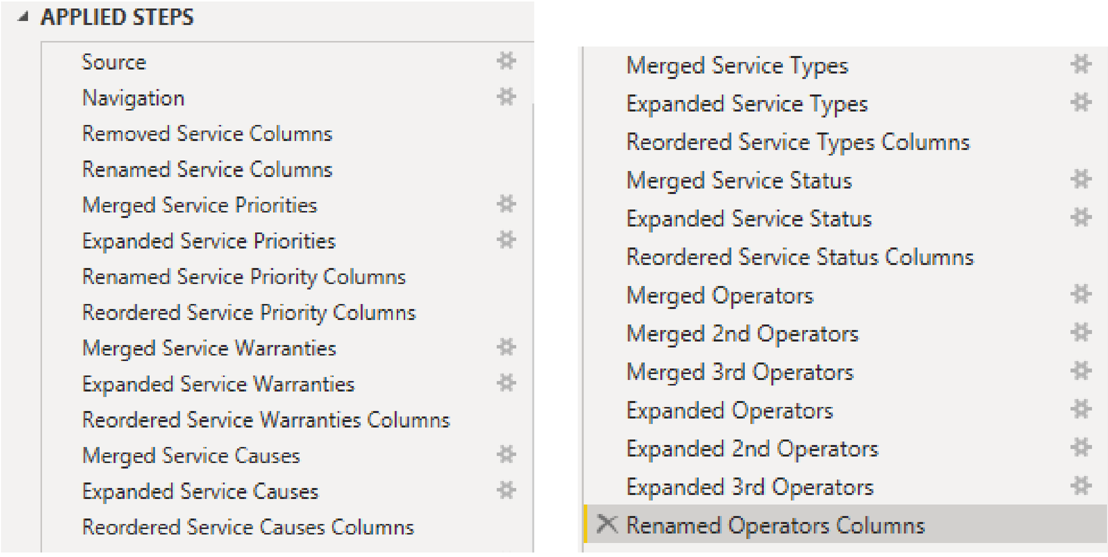

Figure. 3.5. Service table data transformations.

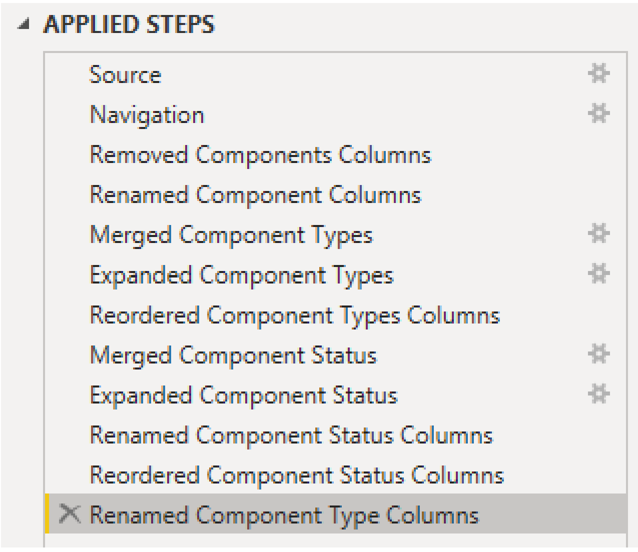

Figure. 3.6. Component table data transformations.

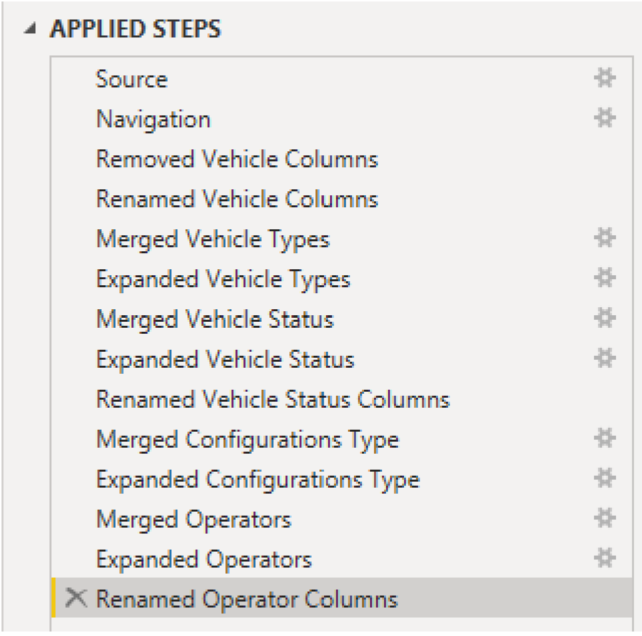

Figure. 3.7. Vehicles table data transformations.

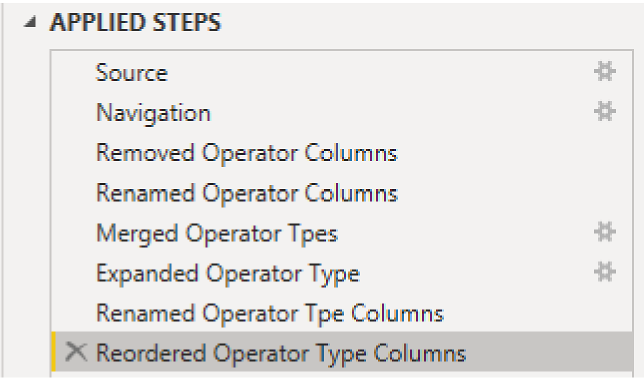

Figure. 3.8. Operator table data transformations.

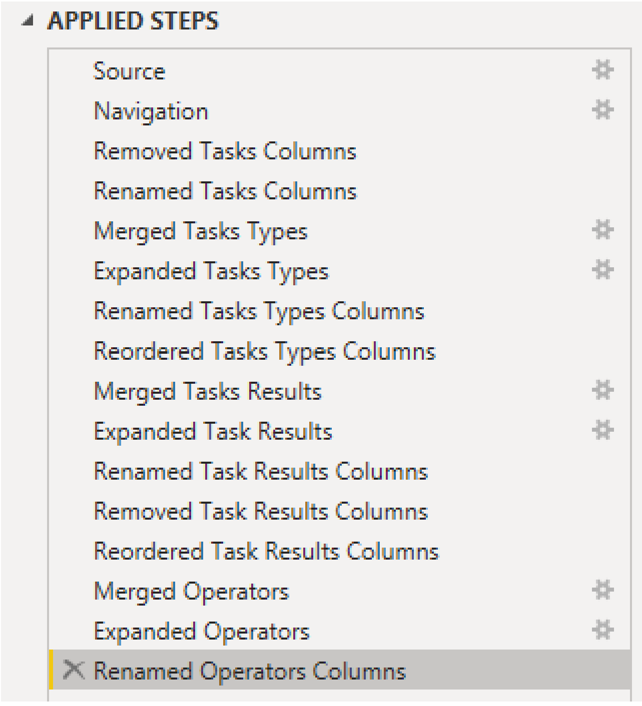

Figure. 3.9. Tasks table data transformations.

The final relationship model worked with is in Figure 3.10. Notice that although the Operators table is considered the main table, it is not shown in the model, this is because there is a limitation on what relationships between tables are active in PBI as it only allows one at a time. Because of this, the Operators table was merged onto the other 4 main tables.

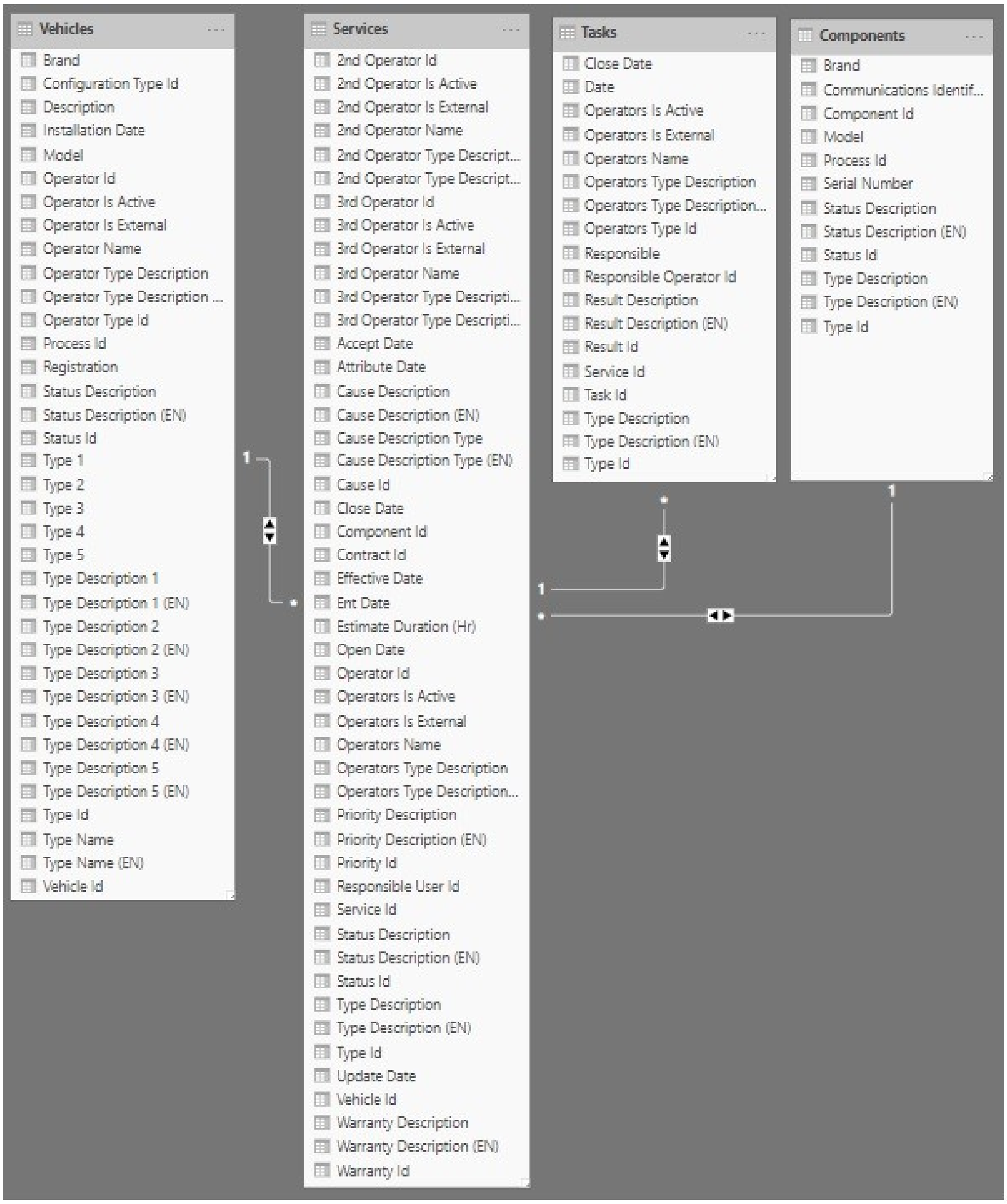

Figure. 3.10. PBI relationship model.

Chapter 4

Implementation

This chapter shows the reports and dashboards created according to the previously defined questions. In addition, a brief analysis is made in order to deepen the knowledge of the data and support the analysis made in the following chapter.

There are a total of 19 questions that need to be answered, and many of them are related to the same table. Because of this and in order to structure the dashboards created the author decided to form a group of questions based on what the main table is used to create a specific report. The groups created are represented in Table 4.1.

Question Number	Question	Question Group
1	What is the percentage of services by technician?	1
2	What is the percentage of services by operator?	1
3	What is the percentage of services by state?	1
4	What is the percentage of services by type?	1
5	What are the causes of services?	1
6	What is the distribution of opened services by date?	1
7	What is the percentage of vehicles by type?	2
8	What is the percentage of vehicles by operator?	2
9	What is the percentage of vehicles by configuration?	2
10	What is the percentage of vehicles by registration date?	2
11	What is the percentage of vehicles by state?	2
12	What is the percentage of components by state?	3
13	What is the percentage of components by type?	3
14	What are the detail of components?	3
15	What is the percentage of tasks by type?	4
16	What is the percentage of tasks by result?	4
17	What is the percentage of tasks by operator type?	4
18	What is the percentage of tasks by operator?	4
19	What is the distribution of tasks by date?	4

Table 4.1: Questions groups.

These groups create a structure for the final PBI file. Each dashboard created refers to a group of questions. This not only makes it easier for the people that are going to benefit from this work to navigate the reports, but it also provides a general guideline for what the next subchapters will cover.

With this said, the following sections will show the dashboards created for each group of questions as well as an analysis of the reports that were created. Because some dashboards have sensitive data, the attribute Operator Name and User Name are anonymized.

4.1 Services dashboard

The first dashboard created is shown in Figure 4.1 and contains the reports related to the services table. This dashboard provides an overview of general information regarding the services performed. A brief explanation of each report is given below.

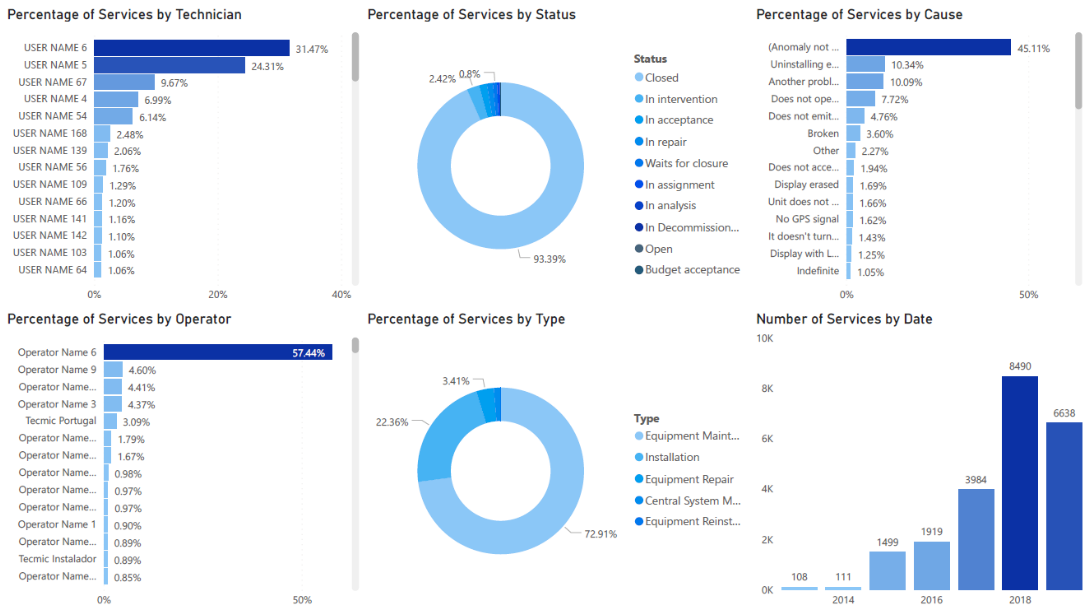

Figure. 4.1. Services dashboard.

This dashboard contains six reports and is based on the service table that has 22749 records. An overview of what each report represents and what information can be extracted from them is explained next.

The first report shows the percentage of services by the technician. A technician is a person responsible for the resolution of the service. With 73 technicians, only five stand out with two of them responsible for 55,78% of the total services done and the other 3 with 22,8%. The remaining 68 technicians have 21,42% of the total services done.

Bellow the technicians report there is the percentage of services done by the operator. An operator is a client that requested service. With 57,44% of the percentage of services, this is the main operator as the remaining 42,56% is spread over the 227 operators.

The next two reports allow us to understand the distribution of the services by status and type. Most of the services are closed, accounting for 93,39% of the total. Only 2,42% are In Intervention, and the remaining 4,19% is spread over the other statuses. The most service type done is equipment maintenance with 72,19% of the total services followed by Installation with 22,36% and Equipment Repair with 3,41%. Other types are Central System Maintenance with 1,31% and 0% for equipment reinstallation.

The rightmost report represents the percentage of services by cause. Unfortunately, the largest slice of services has the cause of "Unidentified anomaly" with a representation of 45,11%, which makes it more difficult to perceive in future analyses the causes for the existence of services. Almost tied with 10,34% and 10,09% are the causes "Uninstalling Equipment" and "Another problem" respectively followed by "Does not Open channel" still with a significant representation of 7,72% and "Does not emit voice" with 4,76%. These are the distributions of the top 5 causes with a representation of 78,02%, and the remaining 21,98% are distributed by the remaining 42 possible causes.

The last report shows how many services were opened by date. An increasing number of services are being opened as we move forward in time. The difference between 2015 and 2016 is just 420 services, an increase of 28%. More than double the number of services were opened in 2017 compared to 2016 and in 2018 compared to 2018, with increases of 107% and 113,1% respectively.

Although only simple metrics are represented in the graphs created, there is a great potential for exploring information through this dashboard. PBI cross-data and filtering capabilities allow people who make decisions to click on any bar or description of any graph by filtering the displayed information and affecting all displayed values. Questions such as knowing the volume of services per operator, date and the causes of the services performed are within a click of the operator's name. Another great advantage of the graph that shows the services by date is the ability to filter all the remaining charts by year and month simply by clicking on the desired year bar.

4.2 Vehicles dashboard

Similarly to the dashboard created for the services table, a dashboard for the vehicles was also created as shown in Figure 4.2. This dashboard provides an overview of general information regarding the clients vehicles. A brief explanation of each report is given below.

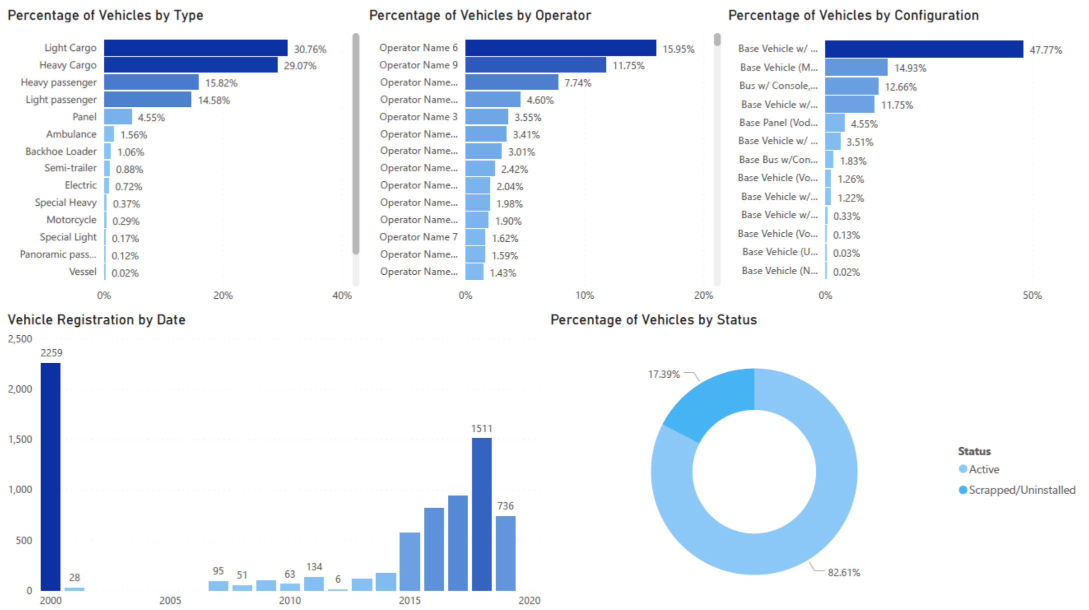

Figure. 4.2. Vehicles dashboard.

This dashboard contains five reports and is based on the vehicles table that has 8593 records. An overview of what each report represents and what information can be extracted from them is explained next.

The first report shows the percentage of vehicles by type. Cargo and passenger vehicles are the most common among the company clients, with 90,23% of the total vehicles. Of these, 59,83% are cargo vehicles and 30,4% are passenger vehicles.

The next report shows the percentage of vehicles by the operator. It can be said that although there are three operators who accumulate 35,44% of the vehicles, the remaining 273 operators occupy the remaining 64,56%. This report is quite useful when the information is crossed with the previous report. For example, the operator with more vehicles, 15,95% has mainly heavy passenger vehicles with 71,70%, where the second operator with 11,75% has 99,21% of heavy cargo vehicles.

The next report represents the percentage of vehicles per configuration. A configuration is a set of components installed in vehicles. The most common configuration represents 47,77% of all existing vehicle configurations, followed by three others with very similar percentages totalling 39,34%. Through this report, it is possible to perform analyses that aim to understand which have the most amount of services performed.

Finally, the last two reports represent the registration date of vehicles and their status. Regarding vehicle registration dates, there is a significant increase between

2015 and 2018, with a peak of 1511 vehicles in 2018. The increase in the number of vehicles means that there are more possibilities to install equipment than in turn provides more possibilities to do business. Finally, we have the report that shows the percentage between active and scrapped vehicles where over 80% are found to be active.

4.3 Components dashboard

The components dashboards shown in Figure 4.3 is simpler in the sense that it has a reduced number of reports against the previously presented dashboards. This dashboard features two charts and a table with the possibility to filter its results. A brief explanation of what each report represents is given below.

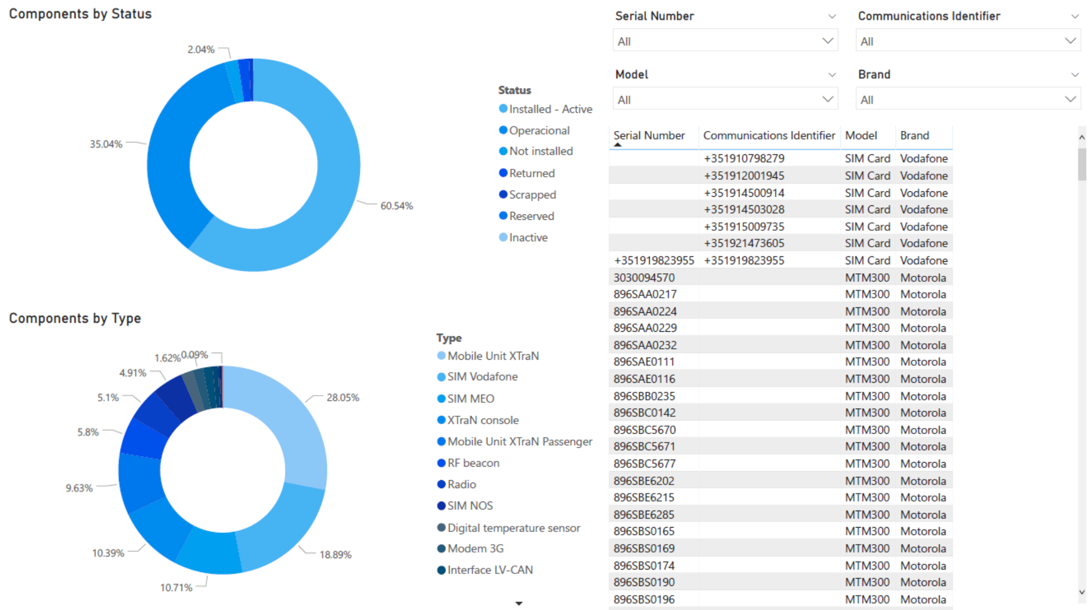

Figure. 4.3. Components dashboard.

The pie charts on the left represent components by state and components by type. Most of the components are installed and operational, with 95,58% of the total while the remaining 4,42% are distributed by the remaining statuses. In relation to the type of component, eleven types are observed where the five most common represent 68,84% of the total. On the right side is a table that contains detailed information regarding the components such as their serial number, communications identifier, model and brand. Above the table are the table filters where it is possible to combine them to perform a table search.

4.4 Tasks dashboard

Like the dashboard created for the services, one for the tasks was created as shown in Figure 4.4. This dashboard provides answers to the final group of questions.

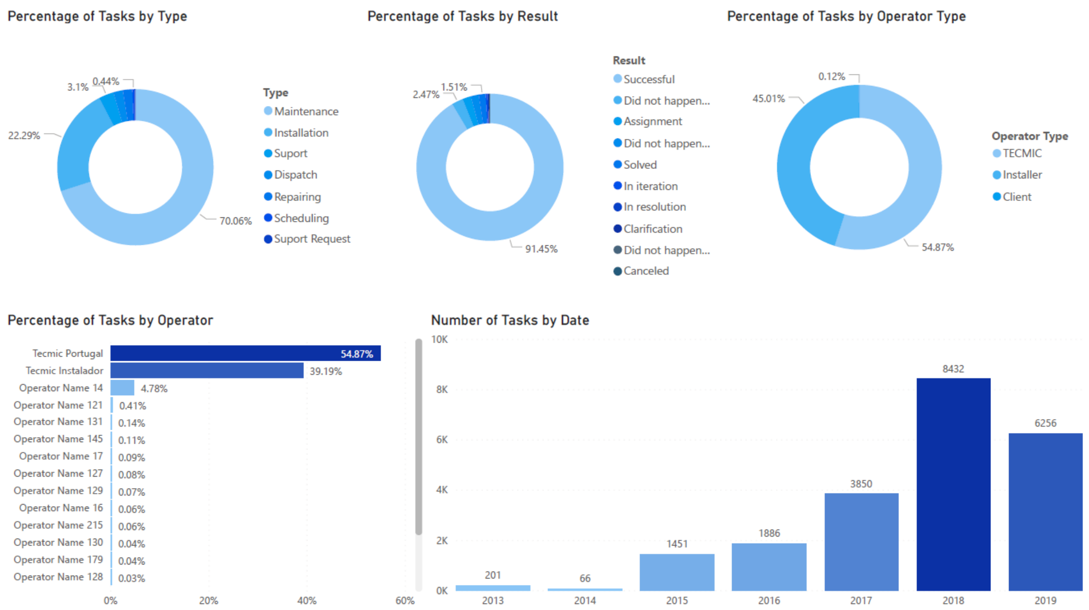

Figure. 4.4. Tasks dashboard.

The total number of tasks is 22586, which is a little less than the total services done with 22749. A task is created to provide additional information on a service. Some services require multiple tasks and others need only one. An overview of what each report represents and what information can be extracted from them is explained next.

The first report shows the percentage of tasks by type. There are seven types defined, and Maintenance is the most common accounting for 70,06% of the total tasks done with Installation as the second most common with 22,29%. The remaining 7,65% includes tasks such as Support, Dispatch, Repairing, Scheduling and Support Request.

Successful is the most common result of the tasks done, as shown in the next report, 91,45% of the total tasks. Some tasks did not happen due to technical reason with just 2,47%, and some tasks were solved with 1,51%. The remaining possible results account for the remaining 4,57%.

The next two reports show the percentage of tasks by operator type and percentage of tasks done by the operator. Two main types are prevalent, being Tecmic and

Installer totalling 99,88% of the tasks done, and the same pattern is seen in the operator with 94,06% of tasks done.

The last report shows the close date of tasks, and it is very similar to the report shown in the services dashboard.

With the creation of the presented dashboards, in addition to answering the questions posed at the beginning of this chapter, the reader has notions of how the information is distributed by the services, vehicles, components and tasks and how it is related.

All the information extracted from this chapter serves as a basis for support for the next chapter where the author performs a deeper analysis of the data enriching the PBI file where new reports and dashboards are created that allow providing recommendations on actions to be carried out.

Chapter 5

Data Analytics

This chapter aims to use the information extracted in the previous chapter and perform exploratory data analysis through reports created in PBI. At the end of each subchapter it should be possible to identify some aspects that can be improved to reduce the number of reported incidents or to improve Tecmic's performance. To achieve the objective described above, we propose to explore the data in order to be able to answer the following questions:

- What is the average service resolution times?
- How are component configurations distributed across vehicles?
- What configuration is the most frequent replaced?
- What is the average time between maintenance tasks for each configuration?

5.1 Services resolution time analysis

In this subchapter we intend to identify what are the average service resolution times and for this the report in the Figure 5.1 was created.

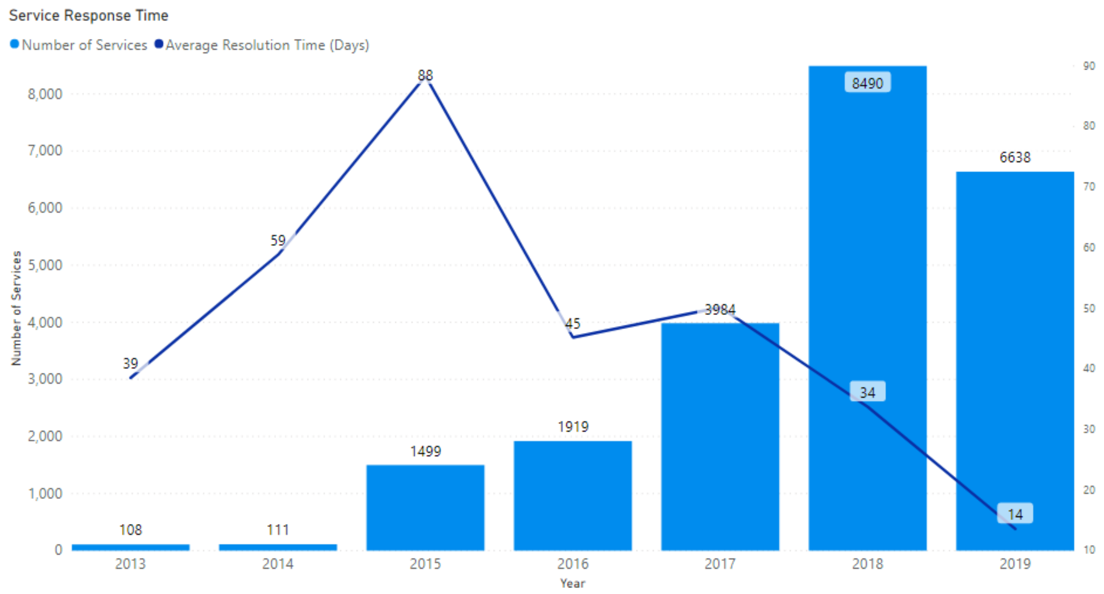

Figure. 5.1. Services response time.

In making this report several situations were identified where the average number of days exceeded 365, and for all these cases the value obtained was changed to the maximum value of 365. The displayed chart represents the number of services created and their average resolution time in days from 2013 to 2019. The resolution time is the number of days passed since the service is opened until all the tasks included in it are in the closed state. There is an increase in the average time between 2013 and 2015, where the difference in 2014 is approximately 1.5 times higher than the previous year, which is not accompanied by the proportional growth of the number of services that were almost identical with a difference of 3 services performed. In 2015, the average time increased again 1.5 times compared to 2014, but this time it is accompanied by a growth in services with an increase of approximately 13.5 times. From 2015 the performance of the company tends to get better as the average time decreases in proportion to the increase in the number of services. In 2016 a sharp decrease it observed in the average of days with a difference of almost half of the number of days needed on average to close services. This decrease is accompanied by an increase in the number of services by 1.28 times of the previous year. The next two years are what leads to believe the performance of the company is improving as relation between the number of days and number of services improves greatly. In 2017 although with an increase of almost 5 days, the number of services has more than doubled compared to 2016. With more than double the number of services in 2018, the average resolution time is decreasing by almost 30% less compared to 2017. The final reported year is not considered for this analysis because it only contains data until September but from what can be seen, the average resolution time is most likely to be even lower than 2018. In order to complement this analysis, other metrics were extracted as shown in the Figure 5.2.

Year	Average	Minimum	Maximum	Standard deviation	Variance	Median
2013	38.51	1	141	29.86	891.92	28
2014	18.57	0	82	15.61	243.80	11
2015	88.33	0	365	125.65	15789.01	30
2016	45.18	0	365	79.70	6352.56	18
2017	50.16	0	365	88.85	7894.66	15
2018	33.72	0	365	71.10	5055.02	5
2019	13.58	0	212	22.22	493.69	6
Total	36.01	0	365	74.32	5524.08	8

Figure. 5.2. Average resolution time analysis.

The image presents the average, minimum, maximum, standard deviation, variance and median by year of the time in days taken to close a service. On average, it takes 36 days to close service with 2015 as being the worst year with 88.33 days followed by 2017 with 50.16 days. The minimum is almost all identical among the years with 0 days with 2013 as an exception with 1 day. On the other hand, the maximum is 365 days observed in 4 of the 7 years analyzed with the lowest maximum being 82 days in 2014. Comparing the standard deviation to the mean by year, 2013, 2014 and 2019 have the closest values with a difference of less than 10 meaning that the observed average resolution times are very close to the average on the respective year. The same cannot be said to the values observed in the remaining years where the differences to the average are between 30 to 40, meaning that the averages on those years are more scattered. This leads to the variance analysis which confirms that the averages in 2015, 2016, 2017 and 2018 are scattered between 0 and 365 days with a lot a variability with 2015 with the highest value of 15789. Comparing the median to the average it can be said that again the same years of 2015,2016, 2017 and 2018 have skewed values in terms of the average values obtained as the difference between the observed medians and averages range from 28.72 to 58.33. The same is not observed on the other years as their median is very close to the average with a difference is less than 10.

In conclusion of this analysis, it can be said that the company has been improving its efficiency when it comes to resolution time of services as their average time has been decreasing over time even though the number of services is increasing. In the first few years it had some troubles since the time increased drastically, spiking in 2015 with 88 days on average but decreased in the following years. From the wide range of average values obtained it could be said that the services performed range from very simple when the resolution time is very short, or complex when the resolution time is high. Another consideration to the high-resolution time could be that some services include the installation of components in vehicles and these could be dependent on availability, scheduling among other external factors not controlled by the company which skews the average results as confirmed when comparing the median to the average values. Overall the resolution time of services is above 30 days with a high spread of averages observed from the high standard deviation value of 74.32 and extremely high variance of averages with a value of 5524.08.

In the next subchapter, vehicles configurations are explored based on the vehicle type and configuration type. By doing this, it provides knowledge that will be used when analyzing configurations breakdowns.

5.2 Vehicles Configuration analysis

This second subchapter answers the second question at the beginning of this chapter where two analysis are made. The first focuses on the typology of vehicles by their type and the second focuses on the number of configurations by theyr type.

In order to be able to perform these analyzes, it was necessary to extract information from the configuration table. A configuration represents a set of components installed in vehicles. The reports in Figure 5.3 and Figure 5.4 show the results obtained.

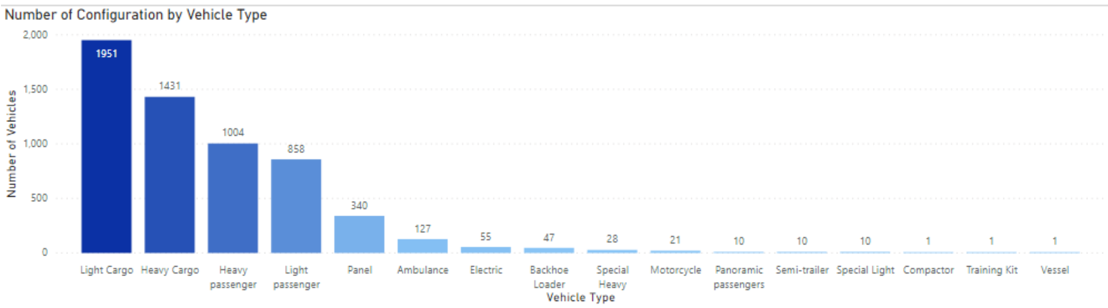

Figure. 5.3. Number of configurations by vehicle type.

There are 16 vehicle types where four of them stand out from the rest as they represent 72% of the total configurations by vehicle type. Of these 72%, light freight vehicles account for the largest share at approximately 33%, followed by 24% for heavy goods vehicles, 17% for passenger goods and 14.5% for passenger cars.

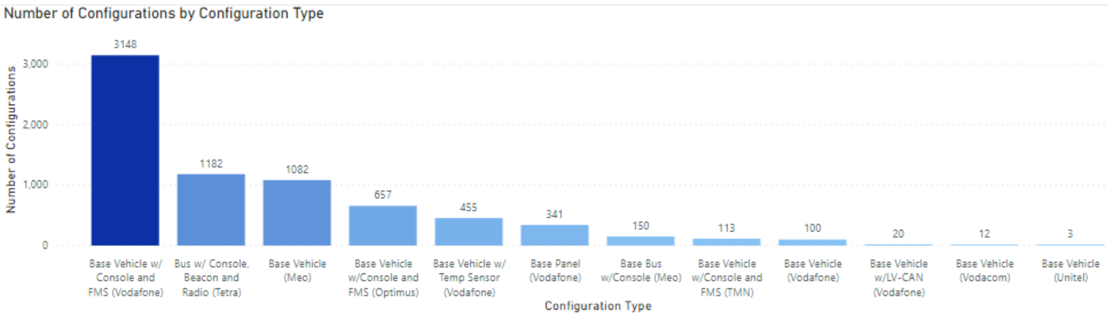

Figure. 5.4. Number of configurations by configuration type.

The most common configuration is on over 43% of the vehicles and the cargo and light passenger vehicles use it the most with percentages ranging from 11% to 16%. The second most common configuration is mostly used by heavy passenger vehicles, with 70% of them using it. Now that the most common configurations were identified and in which vehicles they are in, the next subchapter aims to explore components breakdowns. The information from both reports was filtered to see if there are types of vehicles where certain types of configuration prevail. For light cargo vehicles, 100% of the "Base Vehicle w / Temp Sensor" configurations and approximately 71% of the "Base Vehicle (Meo)" configurations are present in them. In heavy cargo vehicles, 100% of the "Base Vehicle w / Console and FMS (Optimus)" settings are found on this type of vehicle. Finally, heavy passenger vehicles also have a predominant configuration being "Bus w / Console, Beacon and Radio (Tetra)" with approximately 92% installed on this type of vehicles. The information extracted from the analyzes performed is useful because it will support the analysis perfomed in the following subchapter.

5.3 Components breakdown analysis

To conclude the analysis proposed at the beginning of this chapter, it is intended to understand in what kind of configurations the installation and maintenance services are performed. Additionally, it is also intended to analyze the average times between these services. To achieve the first proposed report, the necessary data was filtered to only include configurations that had installation or maintenance services performed, and 5863 vehicles meet that criteria. The results obtained can be seen in Figure 5.5

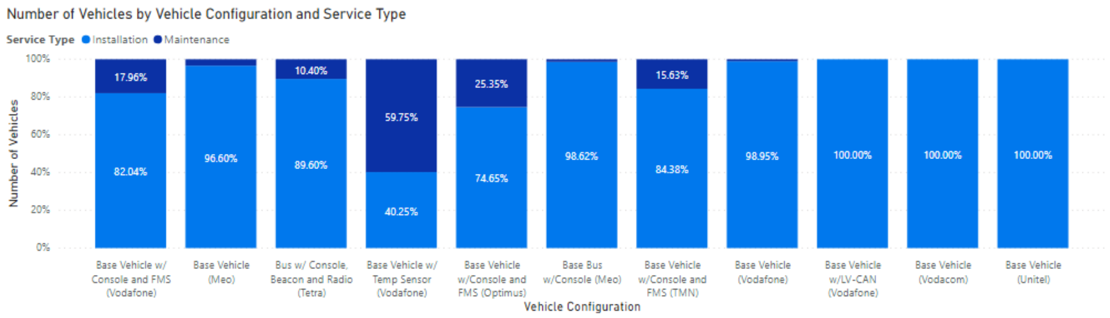

Figure. 5.5. Types of services performed based on configuration types.

Of the installation services there are 4753 vehicles representing 81% of vehicles with configurations leaving the remaining 19% for vehicles where maintenance services were performed. Of the twelve types of configuration, only five have a significant percentage with a maximum value of 59.75% affecting 423 vehicles for the fourth configuration shown in the graph. It should be noted that there are three configurations that have never been serviced and that three are also very close to 100%.

The report in Figure 5.6 shows the average time between maintenance services. Through the data presented it will be possible to identify which configurations and consequently the components, which fail more often.

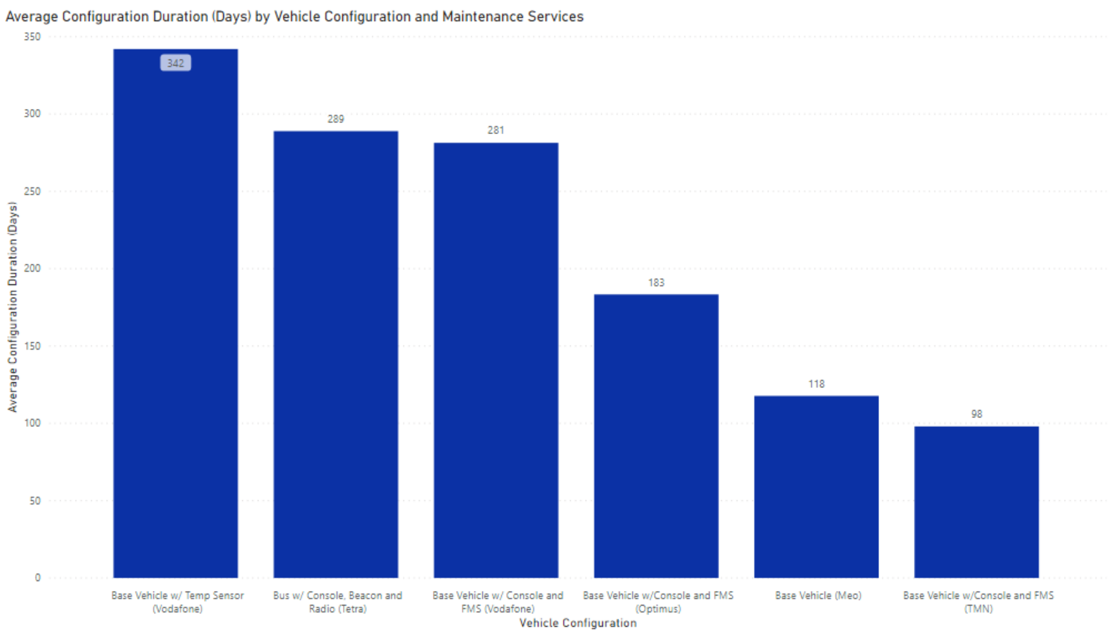

Figure. 5.6. Average component configuration in days.

Only configurations where maintenance services were performed are being analyzed. The average time between old and new component configurations by type were calculated. This means that a high average is a configuration that is less often replaced and the reverse for a low average. The observed averages show a very positive scenario where 4 of the 6 configurations present averages greater than 6 months. The highest average is very close to 1 year, with 342 days followed by two others with very close values of 289 and 281 days or roughly 10 months and another with 183 days which is approximately 6 months. Two very low averages are observed at 118 and 98 days, representing approximately 4 months.

By gathering the information from the reports generated in this chapter and the analyzes performed, conclusions can be drawn to help reduce the number of reported incidents.

- The "Base Vehicle w/ Temp Sensor (Vodafone)" configuration malfunctions after 1 year on average where 60% of the configurations of this type have already required maintenance services. This configuration only exists in light vehicles and 58,6% have already reported some form of malfunction.
- The configuration that is most often installed in vehicles, has only 16% of maintenance services performed which on average break down every 247 days and are installed in approximately 50% of light and heavy cargo vehicles and 78% in light passenger vehicles.
- The "Base Vehicle w / Console and FMS (TMN)" configuration malfunctions after 98 days on average. It is installed in 119 vehicles and of these, only 13% needed maintenance services, 94% of which were light cargo vehicles and the remaining 6% heavy cargo vehicles.

The points presented are just what are considered most important in order to make any recommendations, however there is plenty of possibilities to filter the reports in PBI and get even more information.

With the points presented it can be recommended to explore what goes wrong with the equipment installed in light cargo vehicles as these are have the greatest diversity of configurations being one of them that needed more maintenance services. It is recommended to check the components that are part of the "Base Vehicle W / Temp Sensor (Vodafone)" configuration as these have the most maintenance services performed even though they are only installed in 421 vehicles.

Chapter 6

Conclusions and Future Work

The first objective is to use PBI, Microsoft's SSBI software, and create reports and dashboards with information contained in the Tecmic Incident Database that enable the company's employees to make informed decisions about reported incidents. Additionally, an analysis was made of the reports created, which provides a solid knowledge base helping to understand the data explored to achieve the second objective.

The second proposed objective is to perform data analysis through new reports also created in PBI, which allows identifying points of improvement, inefficiencies in order to reduce the number of reported incidents.

Questions were raised by the incident system operational staff. These guide development to achieve the first goal. A methodology has been defined that includes all the necessary steps to perform before starting to import data into PBI. This methodology includes the steps of understanding the business and then cleaning and transforming them.

Once the data is ready to import, dashboards are created that focus on answering the previously stated questions. Associated with each dashboard is a data analysis that explains the importance and meaning of each report contained in it in order to provide a solid knowledge base to assist with the data analysis performed below. After the creation of the four dashboards, it was possible to achieve the first objective defined.

The following data analysis was also done using reports created in PBI. It is concluded at the end of this analysis that the company's performance in terms of average incident resolution time is on the right track because currently the value obtained is the lowest ever. Additionally it was noted that the large volume of incidents is concentrated in only one of the configurations of components installed in light vehicles.

With the questions answered in the form of dashboards and some aspects that are causing the creation of services, it is considered that the objectives proposed at the beginning of this work were achieved.

Future work aims to understand whether the information presented in PBI reports has impacted the day-to-day operations of the company and whether the number of services and their response times have improved. In addition, this work can be expanded by using Python or R language to perform predictive analytics.

Bibliography

[1] J. Yannis Bakos and Michael E. Treacy. Information Technology and Corporate Strategy: A Research Perspective. *MIS Quarterly: Management Information Systems*, 10(2):107, jun 1986.

[2] Stephen Blair. A Guide To Evaluating a Bug Tracking System White Paper. Technical report, 2004.

[3] Hsinchun Chen, Roger H.L. Chiang, and Veda C. Storey. Business intelligence and analytics: From big data to big impact. *MIS Quarterly: Management Information Systems*, 36(4):1165–1188, 2012.

[4] Office of Government Commerce. *ITIL Service Operation*. The Stationery Office, London, 2007.

[5] Justin Cutroni. *Google Analytics*. O'Reilly Media, Inc., 2007.

[6] Boris Evelson. Topic Overview: Business Intelligence, 2008.

[7] Gartner. Business Intelligence - BI - Gartner IT Glossary.

[8] Analysts Cindi Howson, James Richardson, Rita Sallam, and Austin Kronz. Magic Quadrant for Analytics and Business Intelligence Platforms. Technical report, Gartner, 2019.

[9] Jon Iden and Tom Roar Eikebrokk. Using the ITIL Process Reference Model for Realizing IT Governance: An Empirical Investigation. *Information Systems Management*, 31(1):37–58, jan 2014.

[10] Claudia Imhoff and Colin White. Self-Service Business intelligence TDWI best practices report. Technical report, The Data Warehousing Institute, Renton, 2011.

[11] Andreas Metzger, Philipp Leitner, Dragan Ivanovic, Eric Schmieders, Rod Franklin, Manuel Carro, Schahram Dustdar, and Klaus Pohl. Comparing and combining predictive business process monitoring techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(2):276–290, 2015.

[12] In Ong, Pei Siew, and Siew Wong. A Five-Layered Business Intelligence Architecture. *Communications of the International Business Information Management Association*, (May):1–11, 2011.

[13] Cr Reis and Rp De Mattos Fortes. An Overview of the Software Engineering Process and Tools in the Mozilla Project. *Proceedings of the Open Source Software Development Workshop*, pages 1–21, 2002.

[14] Margaret Rouse. What is trouble ticket (trouble report)?, 2007.

[15] Marten Schläfke, Riccardo Silvi, and Klaus Möller. A framework for business analytics in performance management. *International Journal of Productivity and Performance Management*, 62(1):110–122, jan 2013.

[16] Ramesh. Sharda, Dursun. Delen, and Efraim. Turban. *Business intelligence, analytics, and data science : a managerial perspective*. Pearson, 2017 edition, 2018.

[17] Sumana Sharma, Kweku Muata Osei-Bryson, and George M. Kasper. Evaluation of an integrated knowledge discovery and data mining process model. *Expert Systems with Applications*, 39(13):11335–11348, oct 2012.

[18] Rüdiger Wirth and Jochen Hipp. CRISP-DM: Towards a Standard Process Model for Data Mining. Technical report, 2000.

[19] Carmen Berenice Ynzunza Cortés and Juan Manuel Izar Landeta. Efecto de las estrategias competitivas y los recursos y capacidades orientados al mercado sobre el crecimiento de las organizaciones. *Contaduría y Administración*, 58(1):169–197, 2013.