NOVA IMS Information Management School

MDSAA

Master's degree Program in
Data Science and Advanced Analytics

Enhancing Analytical Insights: A Case Study of implementing a BI solution to a Multinational Company

Internship at Electrolux Group

Pauline RICHARD

Internship Report presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced 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

ENHANCING ANALYTICAL INSIGHTS: A CASE STUDY OF IMPLEMENTING A BI SOLUTION TO A MULTINATIONAL COMPANY

by

Pauline RICHARD

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

Supervisor: Mauro Castelli

Co Supervisor*:* Illya Bakurov

October 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.

15/11/2023, Krakow, Poland

ABSTRACT

In the current intricate business landscape, marked by the complexities of globalization and evolving environmental factors, the imperative for informed decision-making has led to a significant rise in the adoption of Business Intelligence (BI) systems. Despite possessing substantial data, many businesses struggle with the challenge of transforming this data into actionable insights. This often results in decisions being made based on intuition, leading to reduced productivity and flawed strategic choices.

This report delves into my role as a Business Data Analyst at Electrolux, a multinational corporation, focusing on the strategic implementation of a BI solution tailored to enhance data-driven decision-making processes. The initial scenario involved reliance on manual Excel-based analysis, a cumbersome process lacking visual clarity. Handling a lot of different variables and filters made calculations complicated, plus, without helpful visuals, managers found it hard to quickly understand the data trends. The report begins by discussing the importance of BI in today's complex business landscape and the challenges that businesses face in making data-driven decisions. It then describes the project at Electrolux, which aimed to implement a BI solution to improve the tracking of completed tasks and provide managers with a more visual and informative view of data trends. The report outlines the step-by-step process involved in implementing the BI solution, a comprehensive Power BI Report, from collecting data to pre-processing, analyzing, and visualizing the data. It also discusses the theoretical underpinnings of BI solutions, data warehouses, data analytics, and data visualizations. The report concludes by summarizing the key findings of the project and the benefits of using a BI solution to enhance data-driven decision-making.

KEYWORDS

Business Intelligence solutions, Data Modelling, Data Visualization, PowerBi

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

D&S	Documentation and Services
CDT	Content Development Team
BA Europe	Business Area Europe
ELUX	Electrolux Group
UM	User Manual
PBI	Power Bi
BI	Business Intelligence
DB	Database
ML	Machine Learning
DAX	Data Analysis Expressions

1. INTRODUCTION

1.1. BUSINESS CONTEXTUALIZATION

The Electrolux Group, a prominent multinational company, was founded in 1919. Throughout its rich history, the Electrolux Group has embodied values of innovation, quality, and consumer-centricity. Its widespread presence across 150 countries and a workforce of over 55,000 employees exemplify its status as a true multinational corporation. At the very core of the group's identity is its commitment to sustainability and responsible business practices. With a diverse portfolio spanning industries like home appliances, kitchen equipment, and cleaning solutions, the group enhances lives through ingenious design and functionality.

At the heart of the Electrolux Group's endeavors is its commitment to pushing the boundaries of innovation. Over the years, it has introduced a wide array of cutting-edge products and technologies that have revolutionized industries, elevated the group's reputation, but has also significantly impacted the lives of millions worldwide.

The group's reputation for quality and reliability is driven by meticulous attention to detail and stringent quality control processes. This relentless pursuit of excellence have set industry standards, making its brands such as Electrolux, AEG, and Frigidaire synonymous with durability and performance.

With a truly global presence, the Electrolux Group's focus on sustainability, from energy-efficient appliances to responsible sourcing, alongside an unwavering commitment to customer needs, the Electrolux Group has cultivated loyalty and maintained a competitive edge in this ever-evolving market. As the Electrolux Group advances into the future, it remains steadfast in its mission to innovate, inspire, and improve the lives of people through its diverse range of products and services, strengthening its position as a multinational force.

1.2. INTERNSHIP DESCRIPTION AND MOTIVATION

With a clear objective in mind and a well-defined plan, I joined the Global Electrolux Talent (GET) Program as a business data analyst in the Documentation and Service Department (D&S). The organization is playing an essential role in transforming the way Electrolux provides support to their consumers throughout the whole consumer journey. It cooperates with Product Lines, Research & Development, Marketing and Consumer Service functions to help in delivering consumer-focused product & ownership experiences across Taste, Care & Wellbeing product categories. From point of sales materials, through on-boarding guides, and user manuals, to service manuals and spare part catalogues, they aim to create outstanding consumer experiences by delivering multichannel product support and documentation services. The overall process has been described in Figure 1.

The department is composed of approximately 90 individuals divided across 7 core teams, they operate from Krakow (Poland), providing for the European market.

Figure 1 - D&S, Supporting Perfect Consumer Journey cycle. *Source: Electrolux website*

As Business Data Analyst in Documentation & Services, my primary role revolved around driving projects that innovated through data analytics and that will help transform the way Electrolux provides support to their consumers. In this dynamic and collaborative cross-functional environment, I was responsible for several key aspects, including identifying the right analysis methods and data sources, collecting and understanding data from various sources, analyzing and communicating results to relevant business members, and implementing Power BI to analyze and report on key figures.

Additionally, I worked closely with different teams across the business and the management team to understand their needs and find solutions. My role also involved spotting patterns and trends in data sets and defining new processes for collecting and analyzing data. This allowed me to explore data from Product Lines, Research & Development, Marketing, and Consumer Service to drive business performance and transformation at Electrolux.

The project which will be explained in this report, happened within the Content Development Team (CDT). This is the biggest team in the department. Their focus is on producing the User Manuals (abbreviated UMs) for all product ranges of Electrolux, by gathering various sources of information. While I cannot share confidential specifics, due to privacy matters, we will go more into details in Part 3, where the project is being described and the different variables introduced.

1.3. Problem Statement

Given the information provided above, the objective of the project was to assist the CDT team in analyzing the data stored in their database and helping them in making more data-driven decisions. When I started, the team was conducting basic analysis by exporting data to Excel and performing simple calculations manually. The challenge arose from dealing with numerous variables, factors, and filters, making calculations time-consuming and fastidious. Moreover, the absence of visual aids made it difficult for managers to quickly understand the context and trends shown by the numbers.

Another facet of the objective was to improve the tracking of completed tasks. This was particularly important since User Manuals are produced regularly, with some requiring substantial resources, both in terms of time and finances, while others involve minor updates. Keeping tabs on these activities became crucial for making critical business decisions, such as determining print quantities and translation needs. Given that these decisions directly influence costs, having accurate and up-to-date data is pivotal for achieving significant cost savings for the company, both in financial terms and in paper consumption.

To address these needs, a solution emerged in the form of introducing a Business Intelligence (BI) solution, more specifically, a Power BI dashboard. The aim was to introduce a fresh platform to the CDT Team, that will give them easy access to the data and will support their analysis and monitoring, on a monthly basis.

1.4. Report's Structure

Considering the bigger picture, the report will be split into two main parts. The first part will provide the theoretical background for the various concepts covered in this project. Moving on, the second part will demonstrate how these concepts were applied to the actual project at hand.

The report follows a chronological order. It's divided into different parts, each focusing on a specific aspect of the project.

Part 1 is already finished, covering the introduction to the company, explaining the internship scope, and giving context to the project.

Part 2 is the Literature Review. It will discuss important concepts like Data Mining, Business Intelligence solutions, Data Warehouses, and the significance of Data Analytics and Visualizations.

Part 3 goes through the actual project, the chosen methodology and details on its development and resolution.

Part 4 & 5 wraps up the project with the Results and a conclusion.

Part 6 outlines the project's limitations and potential future works.

2. LITERATURE REVIEW

Over the last fifteen years, the business landscape has become increasingly complex, and the utilization of Business Intelligence (BI) systems has correspondingly grown. In addition to globalization, other environmental factors are placing pressure on businesses to operate at a high level. However, due to limited access to those critical information, many business users often make decisions based on intuition rather than exploiting their substantial amounts of relevant and high-quality data. It results in a loss of productivity, reduced agility in the marketplace, and flawed decision-making (Boonsiritomachai et al., 2014). Consequently, recognizing the need for enhanced performance, companies are increasingly turning to decision-based systems. This trend is underscored by Olaniyi et al.'s study (2023), which observes that many Fortune 1000 companies are adopting innovative solutions to improve performance and customer experiences, requiring significant resources. However, he emphasizes that embracing big data analytics with business intelligence remains a compelling investment for them, enabling them to maintain a competitive edge.

The following focus on presenting a literature review regarding the technology required to deal with this knowledge gap, specifically focusing on Business Intelligence systems (BI). BI offers crucial insights by sorting and collecting the right data while Knowledge Management (KM) supports the company at making the best decision by organizing and presenting those results in a more comprehensive way, for the specific area of interest. Both BI and KM play pivotal roles in enhancing the quality and quantity of information available for decision-making processes and can mutually benefit from each other. However, we also observe a lack of consensus regarding the standards that support activities related to the development process of these systems, including ETL design, Datawarehouse design, OLAP cubes design, and others, which will also be explored in this paper.

The main process for the literature choosing method was by researching online, using google scholar and ScienceDirect, but also by having a closer look to the readings' recommendations from the Master's courses. The ones from our Business Intelligence classes, lead by professor Bruno Jardim and professor Miguel de Castro Neto, were particularly insightful and helpful.

2.1. Defining BI

Business intelligence (BI) is a subfield of Decision Support System (DSS) that focuses on providing tools and technologies for transforming data into information and knowledge, designed to support better decisions. Different interpretations of what business intelligence (BI) is, exist. These interpretations are influenced by a range of factors, including the individual's background and perspective.

The term "business intelligence" was first coined in the 1860s by Richard Miller Devens, a Massachusetts businessman. Devens used the term to describe the use of data to improve decision-making in business. The concept of business intelligence has evolved over time, but the core idea remains the same: business intelligence is about using data to make better decisions. Later on, in 1989, Howard Dresner popularizes the term "business intelligence" to describe it as the process of collecting, integrating, analysing, and visualizing data to support better decision-making (Gartner

Glossary of Terms). This quote provides a bit more detail about the different steps involved in the business intelligence process. It also highlights the importance of data visualization in helping businesses to understand their data and make better decisions.

2.2. History of BI

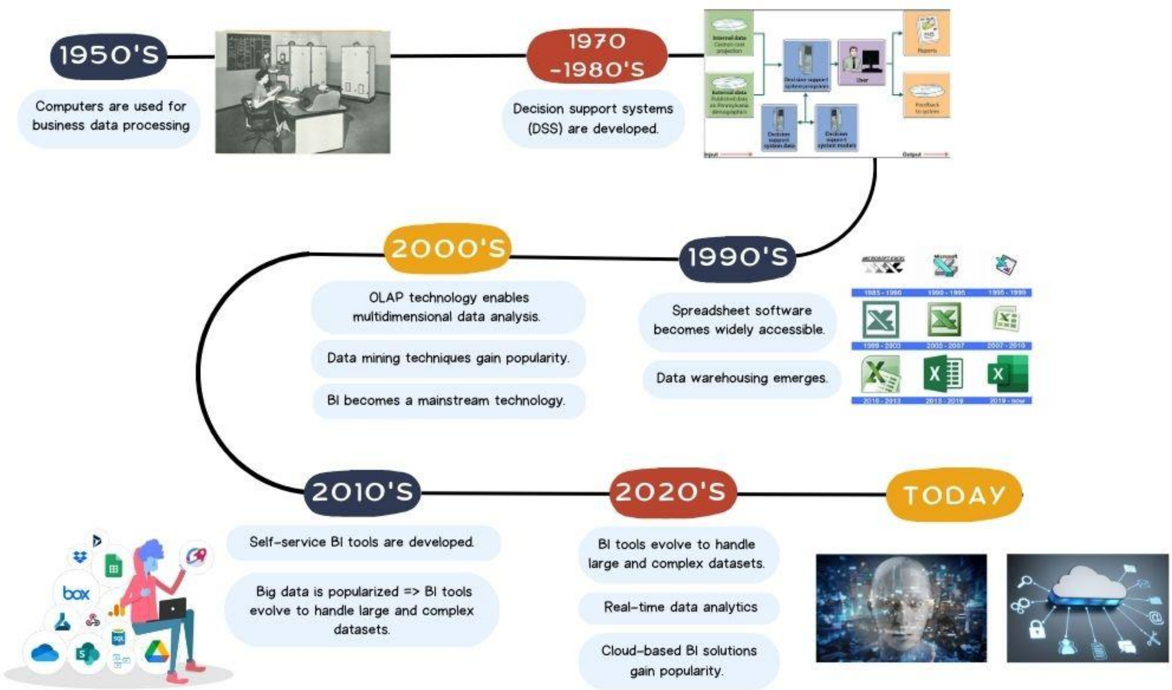


Figure 2 - Overview of the milestones in the evolution of Business Intelligence (BI). *Source: Created by me*

Business intelligence (BI) has undergone significant evolution since its inception in the 1950s. Driven by technological advancements and the ever-increasing volume of data available, BI has become an essential tool for businesses of all sizes. This timeline (Figure 2) highlights some of the key milestones in the development of BI, from its early beginnings to its current state-of-the-art capabilities.

1950s:

- ✓ Introduction of computers for business data processing.

1970s:

- ✓ Development of decision support systems (DSS) to aid decision-making in businesses.

1990s:

- ✓ Rise of the internet and e-commerce, leading to an increase in data.
- ✓ Accessibility of spreadsheet software like Microsoft Excel.
- ✓ Emergence of data warehousing for centralized data analysis and reporting.

2000s:

- ✓ Mainstream adoption of Business Intelligence.
- ✓ Introduction of Online Analytical Processing (OLAP) technology, enabling to analyze data from multiple dimensions and viewpoints.
- ✓ Popularization of data mining techniques for pattern discovery.

2010s:

- ✓ Development of self-service BI tools like Tableau, Power BI, and QlikView. Enabling non-technical users to create their own reports and dashboards, reducing dependency on IT departments.
- ✓ Popularization of the term "big data" and the rise of big data technologies.

2020s - Today:

- ✓ Integration of artificial intelligence and machine learning algorithms in BI tools.
- ✓ Increase in real-time data analytics capabilities.
- ✓ Rise of cloud-based BI solutions, providing scalability, flexibility, and accessibility.

More generally, those last years and the ones to come, we are witnessing a data-democratization in companies. The emphasis on data became stronger, aiming to make data and insights accessible to employees across all levels of the organization. That being said, the evolution of BI is a testament to the ever-growing importance of data in the modern world. As businesses strive to make better decisions, BI will continue to play a critical role in helping them to understand and leverage their data. The future of BI is likely to be even more data-driven, with new technologies emerging to help businesses get more value from their data.

2.3. Business Intelligence architecture and framework

A BI architecture is the design of the components and relationships that make up a BI system. It is composed of several layers that work together seamlessly to transform unprocessed data into meaningful information. Every layer is crucial to the overall data processing and analysis pipeline.

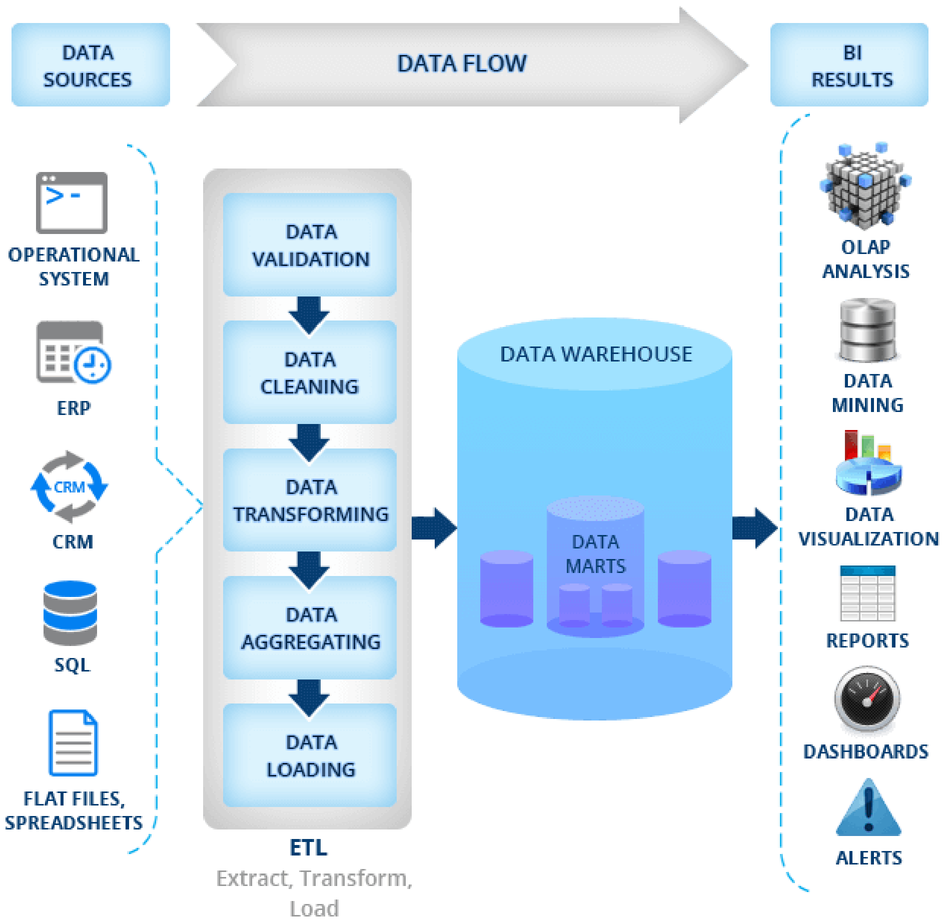


Figure 3 - Overview of the Business Intelligence process. *Source: Cognos Analytics IBM*

Figure 3 decomposed the Business Intelligence architecture into three high layers:

- **Data Sources** is about collecting data from various source systems within and outside the organization.
- **Data Flow** comes next. Once sourced, data needs to be integrated, cleaned, and transformed into a consistent format (Data Integration), before being stored (Data Warehousing) and modeled to create relationships and hierarchies (Data Modelling).
- **BI results** regroups access to the data with advanced analytics techniques such as data mining, statistics, predictive modeling, …etc, applied to uncover patterns, trends, and insights (Business Analytics) before generating reports and interactive dashboards for the end-user (Reporting and Dashboards)

Business Intelligence (BI) consist of various aspects and layers, some observable to users while others remain more hidden. Even models, such as the one depicted in Figure 3, have limitations; they do not fully capture the complexities of data preparation, including ETL processes and other essential steps. This complexity underscores that BI is more nuanced than it may seem, as it spans the entire process from data collection and storage (start) to decision-making and action (end). In the following sections, we will examine some of the components of the BI architecture in greater detail in order to gain a deeper understanding of some fundamental concepts in BI.

2.3.1 DATA SOURCE

Data sources are the cornerstone of a business intelligence architecture, constituting one of its most critical and challenging aspects. It provides the foundation upon which insights and informed decision-making are being built. As Kunnathuvalappil Hariharan (2018) aptly observes "gathering, preparing, and evaluating data are the most important tasks. The data must be of excellent quality (...). as it moves through multiple phases of informational metamorphosis. A BI organization fully uses data at every phase of the BI architecture.". The challenge arises from the scattered nature of data, often residing in various internal and external sources, each with its distinct format and varying level of quality. Internally, data emanates from systems like Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and Human Resources (HR). Externally, it extends to market research, social media, web analytics, and more. These sources provide the raw material indispensable for analysis, forming the foundation of any meaningful BI initiative.

However, the complexity of data arises not only from its scattered nature but also from its varied formats. Data can be unstructured, semi-structured, or well-organized structured data. Unstructured and semi-structured data, which includes emails, reports, articles, and social media content, poses a challenge due to its flexibility and the need for additional processing and normalization. This complexity is further underscored by the rapid growth in digital data. For instance, in 2012, 90 percent of information was reported to be unstructured, as per the IDC report *"The Digital Universe in 2012"*. This report highlights the arising challenge for Businesses to navigate this diverse data complexity and underscores the need for more effective data integration strategies, with a thorough understanding of the properties and formats of the data sources.

The complexity of data can be summarized in three dimensions:

- **Data volume:** Businesses are collecting and storing more data than ever before, due to factors such as the increasing use of sensors, the rise of social media, and the growth of e-commerce.
- **Data variety:** The types of data that businesses collect are also becoming more diverse, including structured data such as customer records and financial data, as well as unstructured data such as images and social media posts. This diversity adds layers of complexity to data analysis.
- **Data velocity:** The speed at which data is generated, especially in real-time processes such as online transactions, further intensifies the challenge of data analysis.

These complexities, if not effectively managed, limit businesses in making informed decisions, identifying trends, ensuring regulatory compliance, and seizing opportunities. However, they also present an opportunity. Businesses that are successfully navigating this intricate data landscape can gain a significant competitive edge. Data analytics, when harnessed effectively, can lead to product/service enhancements, innovation, and personalized customer experiences. Therefore, to navigate this complexity and minimize the time and cost associated with data collection, meticulous planning is essential. This includes identifying data sources, understanding their formats, and developing a robust strategy for cleaning and transforming the data to align with the BI system. It's like putting together a puzzle – collecting data from different places and making it fit together.

2.3.2 DATA FLOW

2.3.2.1. Data Transformation

As we just saw, data integration, is the first step before we can understand and use the data. A common mistake is to equalize the terms data and information, since they represent two different phenomena (Langit, 2007). Data is raw facts and figures that have not been organized or analyzed, while information is data that can be used by the end-user. In other words, data is the raw material, and to make it useful, we need to transform it into information.

Transforming data is the process of changing its shape or format so it can be used for analysis. Think of it as changing the shape or format of the puzzle pieces so they fit perfectly. There are tools and techniques for this, like the ETL (extract, transform, load) process, which involves collecting data, changing its format, and putting it where we can use it. Another method is ELT, which is similar, but the data changes shape after it's in the right place. Comparison is made in Figure 4.

Figure 4 - ETL vs ELT compared and explained. *Source: altexsoft*

We also have to make sure the data is good quality. We assess its accuracy, completeness, and other factors. If the data has errors, we fix them. This is called data cleansing. We might also pick out only the parts of the data that we need for a specific task. This is called data selection. Sometimes, we create new pieces for our puzzle by combining or transforming existing ones. This is called feature engineering. Lastly, there's something called data normalization, which is like organizing your puzzle pieces so you don't have duplicates. It helps us keep our data neat and efficient.

In summary, integrating data in BI is akin to solving a puzzle. We collect data from different sources, transform its shape to ensure a seamless fit, validate its quality, and organize it efficiently. To illustrate, consider a scenario where our raw puzzle/data represents a diverse set of pieces, each gathered from various places. First, a meticulous data quality assessment identifies missing pieces, akin to missing values in a dataset. Using data cleansing techniques, these gaps are filled, creating a complete set. Next, data selection acts like selecting specific puzzle pieces tailored for a particular model, optimizing its effectiveness. Additionally, through feature engineering, we craft new pieces, enhancing the puzzle's complexity and enabling deeper analysis. Each step is essential, transforming our jigsaw of data into a coherent and insightful picture ready for meaningful interpretation and decision-making.

In Figure 5 is illustrated the model inspired by Robert Taylor's influential concept, the 'Value Adding Spectrum,' introduced in 1982. This framework delineates the transformation of raw data into actionable insights through a process of continuous value addition. Taylor's model has been profoundly influential in shaping the landscape of data transformation models, providing a structured path for converting data from one form to another effectively.

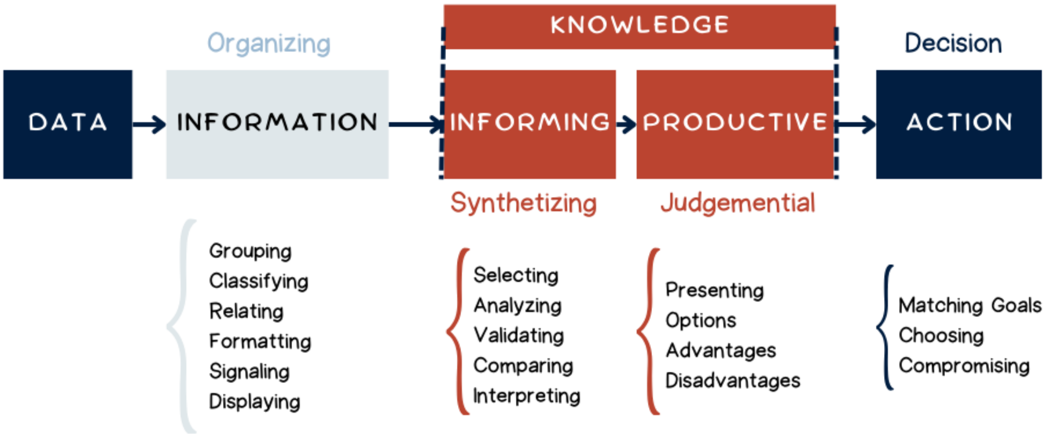


Figure 5 - Value adding spectrum (Taylor, 1982). *Source: Created by me*

The Value Adding Spectrum model is not merely theoretical; it emphasizes a fundamental truth – raw data, on its own, lacks utility until it is processed. In other words, data becomes meaningful information when it forms connections with other datasets, weaving together a coherent narrative. After information evolves to productive knowledge, it can be used to form decisions and later used to form actions (decisions) (Taylor, 1982.) Recognizing these connections within data presents a challenge in modern Business Intelligence (BI) as well. Hence, diverse methods to comprehend and prepare data are essential in BI practices. By understanding how data is transformed into information, knowledge, and action, BI practitioners can develop data transformation models and data preparation processes that produce the data needed to make informed decisions. The Value Adding Spectrum model highlights the critical journey from raw data to actionable insights, emphasizing the importance of meaningful data connections in driving effective decision-making and strategic actions in today's data-driven world.

2.3.2.2. Data Storage

After the Extract, Transform, Load (ETL) process has refined and prepared the data, data storage takes center stage, providing a robust foundation for meaningful analysis and strategic decision-making. And when it comes to storing data for a company's BI system, two primary methods come into play. A data warehouse serves as a comprehensive repository, storing a vast array of data for future utilization, while a data mart takes a more focused approach, storing data specific to particular areas like geographic regions, functional domains, or applications.

This careful storage and organization of data lay the groundwork for BI's analytical process. Data warehouses provide a holistic view, encompassing diverse datasets, while data marts offer nuanced insights tailored to specific business needs. The symbiotic relationship between these storage methods, as depicted in Figure 6, ensures that businesses possess precisely the data they need, ready for advanced analysis and well-informed decision-making.

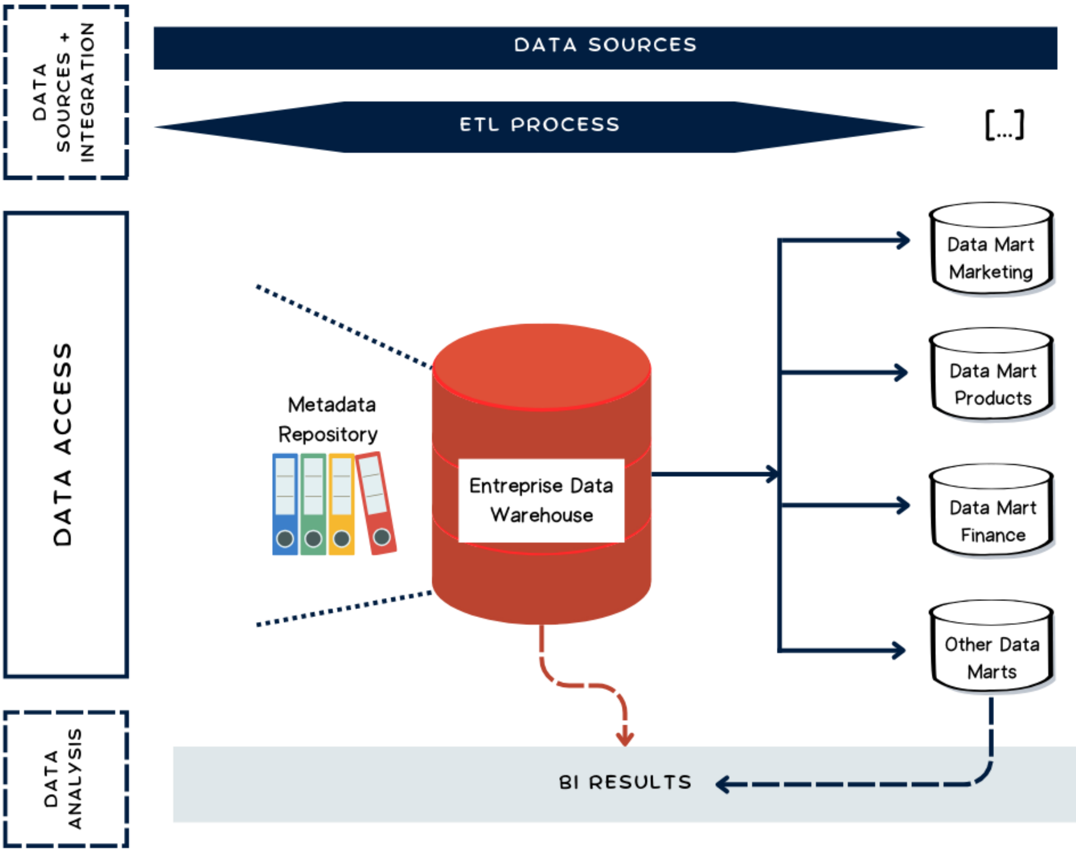


Figure 6 - Data Flow Diagram. *Source: Created by me*

In a world inundated with data, organizations must make strategic decisions about what to store to ensure system efficiency. Think of a data warehouse as a meticulously organized warehouse, each piece of data finding its designated spot, ensuring businesses rapid access to the information necessary for insightful analysis and strategic planning. Beyond operational tasks, the core purpose of a data warehouse lies in supporting organizational decision-making. Experts like Kimball, Corey & Abbey, and Inmon define data warehouses as subject-oriented, integrated, time-variant, and non-volatile collections of data. They enhance analytical processing by enabling efficient querying and reporting. By using techniques like indexing, segmentation, and data compression, data warehouses provide a historical context crucial for businesses to extract actionable intelligence from raw information.

This process is further streamlined by metadata, often termed as "data about data." Metadata acts as a bridge, offering essential information about the data warehousing system and its contents. It include details like formats, encoding/decoding methods, domain constraints, and data definitions, but also business-specific rules and assumptions. Through metadata browsers, users gain an intuitive and user-friendly view of the data warehouse, what information is available, how to access it, what it signifies, and when and how to use specific data sets. In essence, metadata acts as the connective tissue, linking users with stored data, ensuring they can seamlessly extract meaningful insights.

Additionally, in scenarios where data volumes escalate, especially in large companies, operational challenges may arise, leading to slowed functionality and extended search times for specific data. To mitigate this risk, companies can use data marts, which are similar to data warehouses but with a narrower focus. The key difference between the two reside in their scope. Data warehouses are designed to store and manage data from multiple sources across the entire enterprise. Data marts, on the other hand, are created by individual departments or divisions to support their own decision-making needs. Another difference is the type of data they store. Data warehouses typically contain structured data, while Data marts can contain structured or unstructured data, or a mix of both. Both data marts and data warehouses are important data storage solutions for BI systems. The best solution for a particular organization will depend on its specific needs. However, organizations working with too many different data marts in an overall corporate data warehouse strategy may face operational difficulties in using them because individual data marts may not be consistent with each other.

To conclude, the use of data warehouses and data marts is a common practice in Business intelligence, providing centralized hubs for integrated and structured data and enabling sophisticated analysis. However, the emergence of data lakes in the 2010s has challenged this paradigm, highlighting that not all data warehouses serve business intelligence purposes, and not all business intelligence applications require a data warehouse. Data lakes have emerged as a powerful alternative for storing and managing raw data, enabling its transformation and analysis. This shift is observed in a recent Procedia Computer Science article by A. Sukhobokov and E. Gapanyuk (2022), who state that the requirements for data organization and the availability of data from various sources for analytics have increased. And therefore, “to work effectively with large volumes of disparate data, the concept of a data lake began to be used (...) and has now become one of the most common.". Moreover, the role of data warehouses extends beyond traditional BI applications; they find utility in areas like data mining and machine learning, aiding in the development of predictive models for forecasting future trends. Consequently, the selection of data storage methods in the realm of BI is a nuanced decision, tailored to the specific analytical needs and scale of each application.

2.3.3 DIFFERENT TYPE OF BI RESULTS

Once data is meticulously sourced, cleansed, enriched, it can be standardized into "access-ready" formats, BI tools come to life, enabling businesses to extract valuable insights.

2.3.3.1. OLAP Analysis

OLAP, short for Online Analytical Processing, stands as a pivotal technology in the field of Business Intelligence. Designed for the complex and multidimensional data typically stored in data warehouses, OLAP systems enable businesses to swiftly and intuitively answer intricate analytical queries. By pre-aggregating data into multidimensional cubes, it is faster to generate reports and dashboards. Furthermore, what sets OLAP apart is its flexibility: not only can the cube be rotated for different perspectives and market trend predictions, but it also supports a range of operations (as illustrated in Figure 7). Common operations include:

- **Drill down**: analytical operation within OLAP that involves starting with a broad dataset and progressively breaking it down into smaller components, allowing users to delve into finer levels of data granularity. For example, individual products within overall sales figures.
- **Drill up**: also known as OLAP cube roll-up, is an analytical operation that involves aggregating all available data within one or more dimensions. To illustrate, consider a retail network with outlets in different cities. In this context, analysts could gather data about specific products from all locations and rolling it up to the company's central sales department.
- **Slice and Dice:** OLAP cube slice and dice operations involve two essential actions: first, extracting a specific dataset from a cube (referred to as "slicing"), and second, examining it from various perspectives or angles. Slice analysis allows users to dissect their data along a specific dimension, while dice analysis, permits users to explore data across two or more dimensions. This multi-dimensional approach provides a nuanced understanding of the data.

These operations, among others, contribute to OLAP's robust analytical capabilities, making it indispensable for businesses aiming to extract profound insights from their data. We could have also mentioned additional techniques and tools available to enhance visualization such as such as nesting, blending and OLAP cube reader.

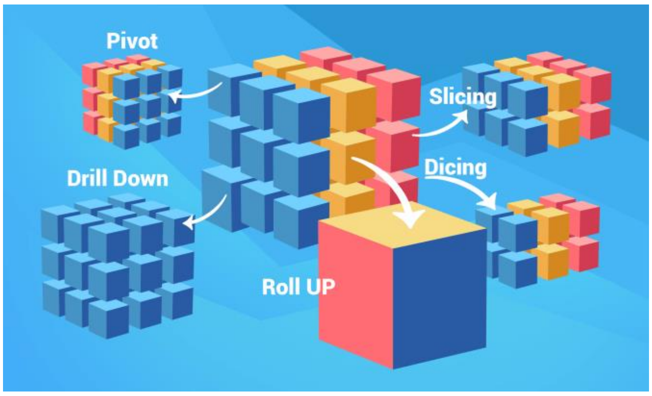


Figure 7 - Overview of OLAP Basic Operations. *Source:* [Galaktika Soft]

Both OLAP cubes and databases are used to store and organize data, however, there are some key differences between the two. Traditional databases are like flat files with rows and columns, limiting the way data can be analyzed. In contrast, OLAP cubes break down data barriers. Databases are typically used to store transactional data, such as customer orders or product inventory. OLAP cubes, on the other hand, are typically used to store analytical data, such as sales data or customer data. Another key difference is that OLAP cubes are pre-aggregated, which means that the data in the cube has already been calculated and summarized. Overall, OLAP cubes offer a number of advantages over traditional databases for analytical purposes, including faster performance and greater flexibility. Summary of those key differences can be found in Table 1.

Feature	*OLAP Analysis*	*OLAP Cube*
Purpose	Technique for analyzing multidimensional data	Database designed for OLAP analysis
Data structure	Any type of data	Multidimensional
Data aggregation	Not pre-aggregated	Pre-aggregated
Performance	Varies depending on the data	Fast
Flexibility	High	Low

Table 1 - Key differences between OLAP cubes and databases

2.3.3.2. Reports and Dashboards

In fulfilling their business intelligence requirements, companies utilize a combination of dashboards and reports to analyze and visualize their data. Reporting and Visualization Tools present a user-friendly interface for displaying analyzed data in an aesthetically pleasing and meaningful way. Through these tools, users can create tailored reports, interactive charts, and visually engaging dashboards, enhancing decision-making processes and enabling in-depth data exploration. It is crucial to note that dashboards and reports serve distinct purposes, and organizations employ both platforms for different functionalities. Reporting, when done effectively and strategically, can benefit organizations in a number of ways, as shown in Figure 8. The primary aim of BI reports is to offer comprehensive data that is easily accessible, understandable, and capable of generating actionable insights.

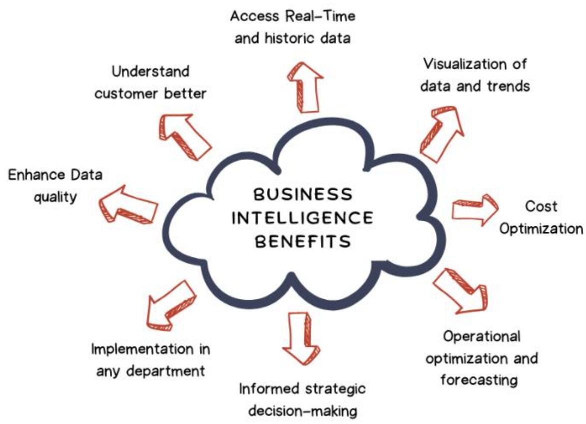


Figure 8 - Summary of benefits from BI reporting. *Source: Created by me*

Dashboards are carefully designed to display real-time or near-real-time data, seamlessly integrating with various data sources, such as databases and cloud services. They focus mainly on visual elements such as charts and graphs, dashboards provide a 360° view of reality rather than dwelling on complex details. By highlighting key performance indicators (KPIs), they offer a comprehensive view of business metrics, enabling users to quickly understand the overall picture. Confined to a single page, dashboards enable simultaneous viewing of multiple visualizations, facilitating a straightforward approach to monitoring overall company performance.

In contrast to dashboards, reports in business intelligence serve as official documents detailing specific areas of a company's performance at a given time. Typically, a business report incorporates statistical data, factual information, and research findings presented through graphs, charts, text, and tables. For example, a sales report might provide information about the performance of each sales region and employee, while a sales dashboard might only provide information about the sales of the entire company. Consequently, reports are visually and contextually richer, providing a deeper understanding to a more complex reality.

Reports are typically split across multiple pages or tabs so that users can navigate between them to explore specific aspects of the data. Although reports are more detail-oriented and accurate than dashboards, they still aim to provide an overview of the reality being explored through detailed and well-arranged information. In addition, reports are static in nature whereas dashboards are dynamic, they do not usually have an expiry date. And while reports are typically used to provide a comprehensive overview of data, dashboards are more commonly used to raise questions and identify areas for improvement. Key distinctions between reports and dashboards are summarized in Table 2 for easy reference.

Feature	*Reports*	*Dashboards*
Scope	Reports focus on one specific area and provide detailed data and information about that area	Dashboards provide a wide breadth of information/broad overview of data via several metrics to help examine the information at a glance
Interpretation	The person who prepares a report provides their own point of view on the data	Dashboards are essentially an objective collection of information
Purpose	To provide detailed information about a specific topic	To provide a high-level overview of key metrics and trends
View	Usually consists of multiple pages	Single-page or single-screen view
Format	Reports are static, so they aren't updated in real-time. You can print them or save them in digital format.	Dashboards are interactive and show real-time data. They are viewed on an electronic device and require a software interface like a website
Interactivity	May be interactive or non-interactive	Typically interactive, allowing users to drill down into data and explore different perspectives
Frequency	Typically generated on a scheduled basis, such as daily, weekly, or monthly	Typically refreshed in real time or on a regular basis

Table 2 - Reports Vs. Dashboards : What Are The Key Differences Between Reports and Dashboards?

2.3.3.3. Data Mining

Analyzing historical data can provide insights into past events, but it cannot predict future outcomes. To gain true foresight, historical data needs to be combined with real-time information and analyzed through advanced analytics. This is where data mining, also known as knowledge discovery in data (KDD), plays a pivotal role. Data mining involves extracting valuable patterns and information from large data sets. According to Chen et al. (2015), it is the process of "analyzing large amounts of raw data to identify patterns and extract useful information" (p. 1853). This method enables businesses to enhance their predictive abilities, make accurate forecasts, and refine classifications. Data mining integrates statistical, mathematical, and machine learning techniques, along with database systems. Companies leverage these insights to proactively address challenges, plan ahead, make informed decisions, manage risks, and seize new growth opportunities.

Data mining achieves to enhance organizational decision-making by analyzing through two main techniques: supervised learning and unsupervised learning. In supervised learning, algorithms learn from labeled data, discerning the relationship between input variables and the target variable. For instance, in a retail setting, it can predict customer purchasing behavior based on factors such as age, income, and past purchases. Linear regression predicts numerical outcomes, while classification algorithms like logistic regression and support vector machines categorize data into distinct classes. Unsupervised learning, conversely, handles unlabeled data. An important application is clustering, grouping similar data points together. For instance, it aids in customer segmentation, identifying distinct customer groups based on purchasing patterns without prior knowledge of customer segments. By employing powerful computers and algorithms, data mining uncovers hidden patterns, trends, and relationships in data that might be missed manually. Here is a summary of some common techniques used in data mining:

- **Association rule learning**
  Also called market basket analysis, association rule learning finds relationships between variables in a dataset that might not be immediately apparent, like identifying products often bought together.

- **Machine learning and deep learning algorithms**
  Machine learning is a broader concept that encompasses various algorithms and techniques that allow computers to learn from data and make predictions or decisions based on that data. Machine learning includes both traditional algorithms, such as decision trees (tree-like structure, predicting outcomes based on decisions), support vector machines, and k-nearest neighbors, as well as neural networks. Neural networks are inspired by the human brain structure and are the foundation of deep learning algorithms. These deep neural networks are capable of learning complex patterns and representations from large amounts of data because of their many layers (hence the term "deep"). Deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are particularly effective for tasks like image and speech recognition, language translation, and natural language processing (NLP).

- **Outlier detection and anomaly detection**
  Detecting outliers and anomalies is crucial for data accuracy and meaningful interpretation. Outliers can be describe as observations that are significantly distant from the distribution mean, not necessarily abnormal. While Anomalies are patterns unexplainable by the base distribution, considered impossibilities.

These techniques are employed to organize and filter data, bringing forth the most pertinent information, ranging from understanding user behaviors to identifying process bottlenecks (*Process Mining*) and even pinpointing security breaches (*Fraud detection*). Consequently, data mining finds applications in various sectors, including sales, marketing, product development, healthcare, and education (*EDM*). For instance, in the context of flood prediction, data mining methods have significantly impacted research, particularly in forecasting, classification, and clustering (Hakim et al., 2023). One of its most vital applications lies in early warning systems, where historical data is analyzed to create predictive models, enabling authorities to anticipate floods and provide affected regions with ample time for evacuation and necessary precautions.

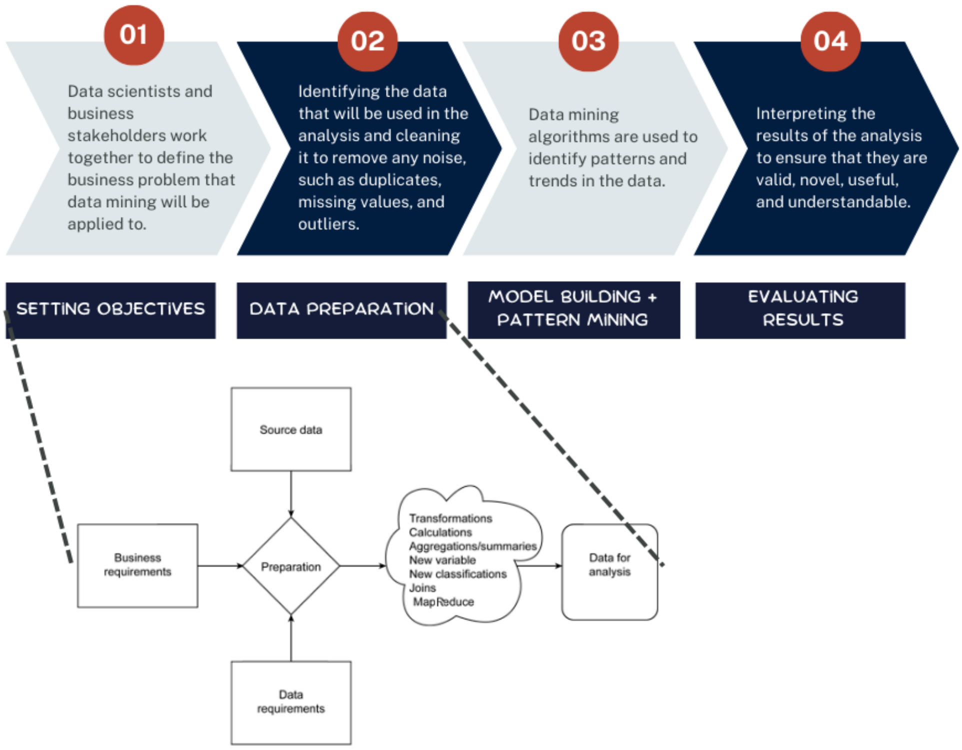


Figure 9 – Data Mining Process. *Source: Created by me*

As illustrated in Figure 9, the data mining process typically consists of four basic steps: defining objectives, gathering and preparing data, applying data mining algorithms, and evaluating outcomes. However, for a successful and reliable data mining process, data scientists adhere to a structured, repeatable methodology known as CRISP-DM, which stands for Cross Industry Standard Process for Data Mining. This standardized approach comprises six sequential phases:

1. **Business Understanding**: a comprehensive understanding of project parameters, including the current business situation, primary objectives, and and the criteria for success.
2. **Data Understanding and Collection**: Define the data that will be necessary to solve the problem effectively and gathering it from all available sources.
3. **Data Preparation:** Cleanse the data, address data quality issues such as missing or duplicate data, and transform the data into the appropriate format.
4. **Modeling**: Identify patterns within the data using algorithms and apply these patterns to create predictive models.
5. **Evaluation**: Assess whether the model's results align with the business goal. This phase often involves iterations to fine-tune the algorithm for optimal results.
6. **Deployment**: Run the analysis and make available the project's results to decision-makers.

Throughout this iterative process, collaboration between domain experts and data miners is crucial. This collaboration ensures a deep understanding of the data mining results in relation to the business question being answered. In summary, data mining proves valuable for businesses; however, it's essential to acknowledge that it's not a one-size-fits-all solution. Its effectiveness varies based on factors such as the available data, the specific business questions, and the desired outcomes. Determining the most suitable technique often involves a process of trial and error, where different methods are tested to identify the most effective approach for a given scenario.

2.3.3.4. Alerts

Alerts are essential for effective BI systems. They provide real-time notifications about critical changes or anomalies in business metrics, enabling proactive decision-making.These intelligent triggers notify users of specific data events, enabling proactive decision-making. For instance, low inventory triggers an alert to the line manager, or to the sales support staff and relevant management team members when a sales amount exceeds $1M. Custom alerts in modern BI solutions deliver timely critical information to designated recipients.

Static reports and alerting are also important ways for stakeholders to stay informed about their business and enabling prompt action. Effective BI software should enable users to effortlessly create and share static reports in common document formats, as well as establish real-time alerts triggered by specific thresholds.

2.3.4. DATA SECURITY

Data represents both a strategic advantage and a potential risk for any organization. Indeed, no organization is entirely secure; there's no foolproof method to guarantee 100% data protection. Nonetheless, failing to protect data can lead to significant costs — the average expense of a data breach has surged to a record high of US$ 4.45 million in 2023, marking a 2% increase from 2022 (IBM, 2023) - and most importantly can lead to market value declines, negative publicity, and erosion of trust. With the explosion of the amount of data captured, data-rich Business Intelligence and cloud systems are here to stay, and organizations need to find practical ways of ensuring their security.

Moreover, many companies routinely share customer data with partners, as part of their business model. That's why privacy regulations like CCPA (US version) and GDPR (European version) govern data storage, sharing and disclosure practices, challenging existing business models and data value transfer methods. While Data privacy is about protecting people's rights through regulatory compliance and business protocols, data security involves safeguarding all of an organization's data, including personal information.

When contemplating data security management, several important implications come into play.

- **Identify Vulnerabilities**

All data flows are vulnerable to attack by definition, so it is important to identify and mitigate all potential vulnerabilities. This includes both known and unknown vulnerabilities, as well as vulnerabilities that may be introduced by new technologies or processes. Merely accessing or copying data poses significant risks, allowing successful exploits to manipulate, delete, or create records.

- **Encrypt Communications**

Data encryption requires additional resources and expenses but proves to be a worthwhile investment. Business Intelligence tools are prime targets due to the competitive advantages they offer. Encrypting data adds a protective layer and distributing it across multiple servers with redundancy enhances security. The data should resemble a puzzle only solvable with a unique map.

- **Ensuring Device Security policies**

The growing trend of installing Business Intelligence tools on employee-owned devices (BYOD) poses a security risk, as these devices may be used in unsecured contexts such as public Wi-Fi networks. To mitigate this risk, organizations should enforce end-to-end encryption of all data.

- **Establish Permissions and Hierarchies**

Restrict knowledge by roles, establishing permission rules for each user or designated user groups. Users should have the minimum permissions necessary for them to perform their daily tasks. This will help to protect sensitive information from unauthorized access.

- **Segment and Distribute Data**

Segmenting and distributing data across multiple servers enhance data security by requiring attackers to compromise multiple servers to gain access, providing effective protection against central attacks.

- **Cloud Security**

If storing Business Intelligence data in the cloud, it is important to choose a reputable service provider with a strong security track record. Cloud providers such as Amazon, Google, and IBM have the necessary infrastructure, expertise, and financial resources to ensure the security of the data. While in-house server pose similar risks, achieving equivalent protection involves significantly higher costs. When sharing information with third-party cloud service providers, ensuring their trustworthiness and monitoring the workflow and data flow is essential. Additionally, having a plan in place to accommodate partner changes and uphold contractual agreements is crucial.

Security and privacy are crucial in big data scenarios, yet they often conflict. While security benefits from analyzing vast amounts of data in real-time, user privacy is compromised. Balancing these requirements is essential. Future business intelligence approaches must find a compromise between security, privacy protection, and the computational overhead posed by big data analysis techniques.

CONLUSION

It's important to note that the categorizations mentioned above represent various facets of Business Intelligence (BI). However, these categories are not rigidly compartmentalized. As articulated by Rick Sherman in his book 'Business Intelligence Guidebook : From Data Integration to Analytics' (2014), these categories overlap and interact, constituting an interactive process that transforms data into information and knowledge. He emphasizes that the components of a BI system are not isolated from each other; instead, they are interconnected and mutually reliant. For instance, a data warehouse serves not only as a storage repository but also as a data source for BI reports and dashboards. Similarly, BI reports and dashboards aren't static documents; they facilitate interaction with the data warehouse, enabling users to explore data and extract insights.

This interactive dynamic lies at the heart of BI, enabling the conversion of data into information and knowledge. By interacting with BI systems, users can gain a deeper understanding of the data and identify patterns and trends that would not be visible in the raw data. This enhanced understanding becomes the foundation for well-informed decision-making.

2.4. Data Visualization to enhance BI solutions

2.4.1. Defining Data Visualization

Data visualization is the graphical representation of data with the purpose of communicating insights and making information more accessible and understandable. Data visualization has been rising rapidly for the past few years in the BI and analytics industry, as part of the modern BI movement which emphasizes on self-service (Parenteau et al., 2016). The challenge of handling massive datasets and presenting them in an accessible manner has driven the adoption of data visualization tools. Data visualization serves as a vital conduit, translating complex information into clear and understandable visuals that empower decision-makers.

Data visualization and BI solutions are inextricably linked, each enhancing the other's capabilities. BI solutions provide the foundation for data visualization by collecting, cleaning, and transforming data into a format that can be easily visualized. Data visualization, in turn, enriches BI solutions by providing users with a powerful tool to explore, analyze, and communicate data insights.

In essence, visualization involves creating tangible, visually perceivable images in the human mind using visual elements like shapes, colors, and positions. Data visualization takes this a step further, allowing interactive exploration and graphical representation of data of varying sizes, types, and origins. This section investigates into the fundamental question that arises*: how, and in what form, data visualization contributes to the overall business intelligence process and system?* This question forms the core of our exploration, shedding light on the symbiotic relationship between data visualization and BI solutions and underscores the transformative impact of visually represented data in shaping business strategies and outcomes.

2.4.2. Purpose of Data Visualization

Data visualization involves meticulous selection and reinterpretation of data, guided by its intended application. This visual representation serves multifaceted purposes. By tracking trends over time, valuable insights can be gained, and starting with a baseline allows for observation of shifts in patterns to spot areas where adjustments are needed (cf. Line charts for example). Data visualization also helps monitor progress and identifies emerging trends, enabling strategic decisions that drive success. The significance of data visualization lies in its ability to uncover in data that are hard to discern just by looking at the numbers. Moreover, data visualization adds engagement and memorability to data, making it more compelling and easier to grasp.

Hence, knowing the purpose of the visualization and why it will be used, is crucial. Today, data visualizations typically fall into six distinct categories, each serving a specific purpose:

- **Comparison:** Focuses on showcasing changes in performance over time or across dimensions.
- **Composition:** Display the allocation of numerical data, often represented through pie or bar charts.
- **Distribution:** Illustrating how values are distributed over time or other dimensions, providing an overview of the data's dimensionality.
- **KPI (Key Performance Indicators):** Emphasizing the current status of metrics, offering a quick snapshot of essential performance indicators.
- **Relationships:** Displaying the connections and correlations between various metrics, often represented through scatter plots.
- **Location:** Mapping data on geographical maps (for example) providing spatial insights.

Therefore, creating impactful data visualizations requires careful planning and analysis. Understanding the insights to be shared and the audience that will use them is essential. Properly mapping out these details beforehand enables effective communication and ensuring maximum impact and usability for both the team and end-users (figure 10).

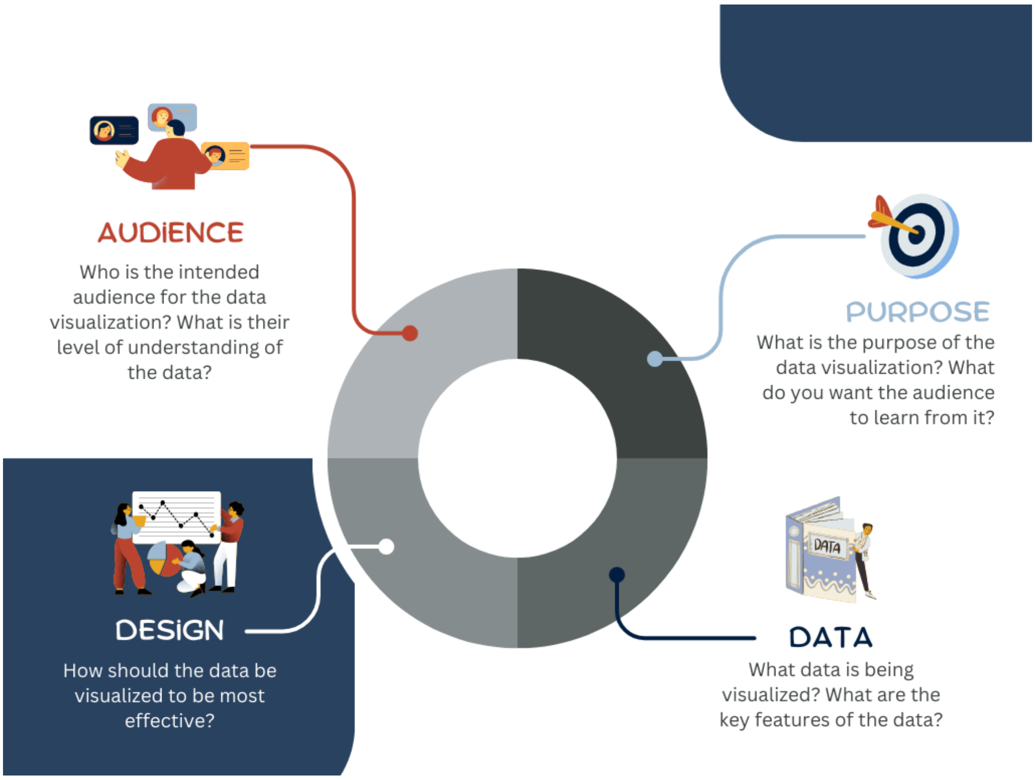


Figure 10 – Four factors to consider when designing data visualizations. *Source: Created by me*

2.4.3. Common types of data visualizations

Before designing a dashboard, it is essential to have a solid understanding of the different types of data visualization, so that you can choose the most appropriate ones to bring your data to life. Here's an overview of some commonly used visualization types, and example of their applications:

- **Sparklines**

A Sparkline is a condensed chart embedded within text, tables, images, or other information. It portrays general data patterns, such as variations and trends, in a simplified way. Unlike detailed charts, Sparklines lack interpretive elements like titles or legends. While miniature line charts are common, Sparklines can take various forms, including bar charts and bullet graphs. They are often used for quick insights in textual contexts.

Embedded visuals are graphical elements that are integrated into another presentation form, rather than being presented independently. Two common types of embedded visuals are those sparklines (or inline mini) and conditional formatting. Conditional formatting uses visual variables such as color and size to directly format or style text, numbers, shapes, and other content. Because it does not significantly alter the layout or flow of content, conditional formatting is less intrusive. Instead, it provides a decorative effect that reveals more meaning or highlights selected content from the data or text.

- **Heat Map**

A heat map employs colour to represent data magnitude, revealing patterns or clusters within two dimensions. By visualizing variations in colour intensity, heat maps are useful in understanding how phenomena cluster or vary across space, providing valuable insights into data trends.

- **Box Plot and Scatter Plot**

In general, box plots are a good choice for visualizing the distribution of a single variable, while scatter plots are a good choice for visualizing the relationship between two variables. Indeed, Box plots are used to show the distribution of a set of data. They show the median, quartiles, and outliers of the data. This can be helpful for understanding data variability, as well as identifying any outliers that may be skewing the results. On the other hand, Scatter plots employ X and Y axes to display data points, each representing specific qualities. By plotting objects on the graph, scatter plots reveal distributions and correlations between the two variables.

- **Charts and Diagrams**

Charts and diagrams are the two most common forms of data visualization. Charts are visual representations of quantitative data, using symbols such as points, lines, and areas, and visual variables such as colour, shape, and size. While diagrams are visual representations of data, which can be either quantitative or qualitative, and can be used to illustrate structures, relationships, sequences, and more. In the context of business data visualization, charts are often used to visualize business performance measures and indicators, and diagrams can be used to visualize qualitative information, such as customer journeys or employee workflows. Here is Table 3 that summarizes some basic types of charts and Table 4 summarizes the diagrams' ones:

Chart Type	*Description*	*When to Use*
Bar	Compares different categories of data.	When you want to compare the values of different categories of data, such as sales by product category or website traffic by source.
Pie	Shows the composition of a whole.	When you want to show how the parts of a whole make up the whole, such as the percentage of sales from each product category or the percentage of website traffic from each country.
Line	Shows how the value of a variable changes over time.	When you want to show how a value has changed over time, such as sales growth or website traffic over the past month.
Bubble	Shows the relationship between two variables, with the size of the bubble representing a third variable.	When you want to show and compare the relationships between different data points, such as the relationship between customer spending and customer satisfaction or the relationship between product price and product popularity.

Table 3 **-** Comparison of different common Chart's types

Type of Diagram	*Description*	*Purpose*
Tree	Shows hierarchical relationships between different elements in a nested tree structure.	Visualizing and understanding hierarchical data, such as taxonomic classification, family trees, or any other type of parent-child relationship.
Workflow	Illustrates the steps and sequences involved in completing a specific task or process.	Identifying bottlenecks and inefficiencies in workflow processes, as well as designing and implementing new workflow processes.
Use case	Models the behavior of a system from the perspective of its users.	Identifying the requirements for a system and designing the system so that it meets those requirements.

Table 4 **-** Comparison of Tree Diagrams, Workflow Diagrams, and Use Case Diagrams

- **Location-Based Visuals**

Location is a powerful dimension in business data analysis and decision-making. Many business activities, such as sales, marketing, and operations, are linked to locations. Location data is becoming increasingly available thanks to the widespread use of location sensors such as GPS. Location-based visuals, such as maps and heatmaps, can provide a familiar context for users and make location-related data more comprehensible and perceivable.

In conclusion, data visualizations can be used to communicate complex data in a clear and concise way. Each type of data visualization has a specific purpose, as shown in Annexe 1. Some visualizations are better suited for showing relationships between data points, while others are better suited for showing comparisons, composition, or distribution. There is no one-size-fits-all template for building a dashboard. The best visualizations to use will depend on the specific goals.

2.4.4. Popular data visualization tools

Data visualization tools are typically desktop applications with command dashboards that integrate with data sources via API. Once the data source is specified, the tool provides a canvas where visualizations can be placed and configured. Users can then combine different types of visualizations to create a report. Depending on the tool, reports can be further exported as CSV files or shared with other users.

Here are three example of popular data visualization BI tools:

- **Looker** offers a robust data analytics solution for businesses, featuring an intuitive drag-and-drop interface, customizable dashboards, and exportable reports.
- **Tableau** presents a comprehensive solution for all BI needs. Its user-friendly interface provides quick access to native integrations with various data sources, allowing transformation, cleaning, mapping, and sharing of analytics across multiple platforms.

- **QlikView** empowers teams to create customized software solutions tailored to their unique needs. Its freemium version offers an upgrade path to the full-featured Qlik Sense.
- …etc

These widely used data visualization tools automate data connections, allowing users to focus on creating reports without worrying about where the data comes from. Once the setup is done, it is simply needed to add the necessary information and then use the tools' user-friendly features to make engaging visuals and tell important stories. This makes data visualization and business intelligence more accessible and impactful for everyone.

CONLUSION

In conclusion, Data Visualization stands as a pivotal tool in the face of rapid growth in big data, which is expected to increase by 26.1% annually, as outlined in the International Data Corporation (IDC, 2023) report. This rapid expansion is estimated based on various factors such as the widespread adoption of cloud computing, making data storage and processing more accessible, the proliferation of data from sensors and devices through the Internet of Things (IoT), and advancements in artificial intelligence (AI). While this growth presents new opportunities for businesses, it also introduces challenges in harnessing this wealth of data effectively. Data Visualization emerges as a vital solution, swiftly extracting valuable insights from a sea of essential business data. Many businesses have recognized the importance of visual representation, placing it at the forefront of their strategies.

Visuals and user-friendly dashboards empower business users to delve into their datasets. These interactive illustrations make vast amounts of information comprehensible and engaging, providing management with real-time insights into their organization’s performance and growth. In essence, Data Visualization remains a critical facet of Business Intelligence. With the rise of advanced tools and technology, it transforms how companies interact and extract insights from their data.

2.5. What about Artificial Intelligence (AI)?

It is a common mistake to confuse BI and AI since those two terms are often used interchangeably, even though they have two different meaning. The distinction between the two is not always clear-cut but, in general, we call AI tools, tools that are focused on developing intelligent agents that can reason, learn, and act autonomously. BI tools, on the other hands, are focused on providing insights into data that can be used to inform decision-making. BI provides insights into past and present performance, it is about understanding what has happened and why, while AI is about predicting what will happen and what to do about it. BI is typically used by business users to make informed decisions, while AI is typically used by data scientists and machine learning engineers to automate tasks and make prediction (Figure 11)

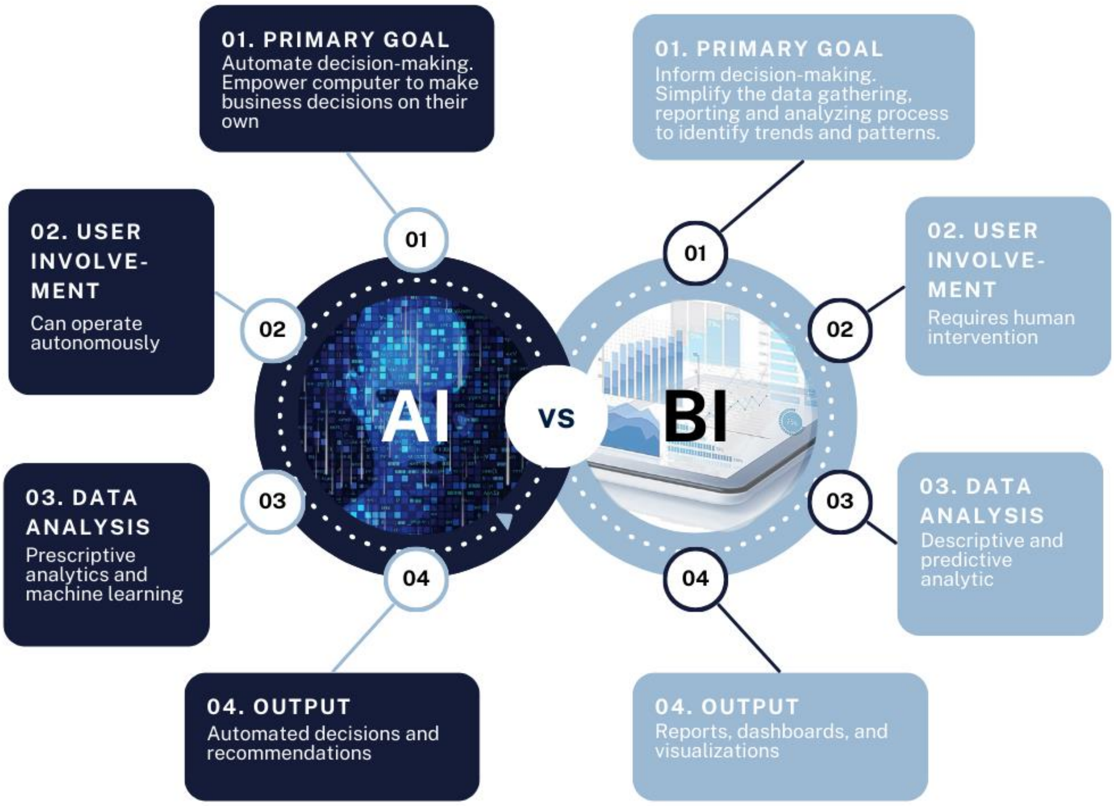


Figure 11 – AI vs BI. *Source: Created by me*

The use of AI in BI is still in its early stages, but it has the potential to revolutionize the way that businesses use data to make decisions. It is what is demonstrated in the paper research *Big Data and Predictive Analytics for Business Intelligence*, published in the MDPI Journal (2022): "With more and more data from the company's internal and external platforms, big data analysis has become the method with the most potential for BI insight extraction. In particular, with the advanced technology of artificial intelligence (AI) (…). Open-source analytics tools based on deep learning and machine learning are accessible and widely applied to the business decision-making processes, such as the tools of Microsoft Power BI, Google Analytics, SETLBI and the accessible models of GitHub Repositories."

As a result, BI and AI can be used together to provide a more complete picture of a business. AI is a rapidly growing field that is having a major impact on BI, and it is important to be aware of the ways in which AI is being used to enhance BI capabilities. It could automate certain BI tasks, such as data cleansing and preparation and be used to develop new BI tools and techniques, such as natural language processing (NLP) and machine learning (ML). For example, NLP can be used to extract insights from unstructured data, such as social media posts and customer reviews. ML can be used to develop predictive models that can be used to forecast future trends.

3. PROJECT DETAILS

3.1. Business context and Understanding

3.1.1. Business context

As introduced earlier, the project unfolds within the Content Development Team (CDT), responsible for producing User Manuals (UMs) covering the full range of Electrolux products. To provide further business context, all data and records related to the production of User Manuals are housed in an internal database known as ANDEX. The team relies on ANDEX to manage all information related to UM production and share it across departments. However, there's a challenge. ANDEX automatically generates a ton of records every day as UMs are created, approved, or updated. Unfortunately, there's no clear way to visualize or analyze this influx of data. This gap means the Content Development Team lacks the tools needed for effective decision-making and monitoring.

To address this gap, the decision was made to implement a Business Intelligence solution that would empower the team to extract valuable insights. The primary aim was straightforward: access the right information at the right time. This strategic move aimed not only at understanding and predicting demands but also at enhancing productivity and achieving cost savings.

3.1.2. Business Requirements

The Content Development Team is divided into four teams, each dedicated to a specific product area. While I can't provide exact team names due to privacy constraints, I've designated them as Team 1, 2, 3, and 4, for clarity.

To comprehend the full scope of the project, numerous vocabulary terms and variable definitions are essential. However, one crucial concept to understand before going into more details is the identification code assigned to each User Manual (UM), referred to as ANC. These ANC codes are unique identifiers, and there are two distinct types:

- Source ANC (English language): Assigned to the document when it's initially created.
- Target ANC (other languages): Used when the document is translated and created in languages other than English.

__Important Disclaimer:__ The screenshots and Figures that will follow in this section are taken from a "sample" version of the actual project where all sensitive information were replace such as names of the Teams, numbers, and other identifiers with generic or fictional information. While all numbers and analyses of this created version are then fictitious, it will effectively protect confidentiality while still allowing to illustrate the process of building the project from scratch and communicate on the key project elements.

3.1.3. Power BI

Considering these business requirements, the nature and volume of the data, the desired output, and the end users, implementing a Power BI dashboard for the team emerged as the most fitting solution.

Now, let's take a closer look at what Power BI is. Microsoft Power BI stands as a robust business intelligence platform, crafted by one of the industry leaders, offering unparalleled potential owing to its extensibility and scalability. Power BI boasts an intuitive interface, robust data visualization capabilities, and interactive features, democratizing analytics by making it accessible to users across all levels. Its support for query languages like DAX, R, Python, enables users to perform intricate calculations, including data querying, preparation, modelling, and forecasting.

Several aspects make Power BI particularly compelling comparing with other competitors:

- **Make Data-Driven Decisions:** Power BI simplifies self-service BI with features akin to Excel, making it adaptable for non-expert users like the Content Development Team (CDT).
- **Collaborate Across the Organization:** The platform facilitates quick sharing of reports and collaborative decision-making, eliminating the need to switch between different tools or exchange results via email. Power BI encourages collaboration with features such as team commenting, alerts, and content subscriptions (e.g., Teams).
- **Confidentiality and Privacy Settings:** As a prominent name in the software industry, Microsoft prioritizes security and protection for sensitive business information in Power BI. The platform enables adjustable permissions for publishing and sharing reports and dashboards, both internally and externally.

In conclusion, Power BI emerges as an excellent tool to facilitate data-driven decision-making, enhance collaboration within the organization, and ensure the confidentiality and privacy of the information. We can now move on with building our reports in Power BI, starting with the phases of Data Integration and Preprocessing.

3.2. Data Integration

In Power BI, steps before building the model typically involves shaping and transforming the data to create a structured and optimized dataset. This process often takes place within Power BI's Power Query Editor, where you can perform data cleansing, transformation, and modelling operations.

3.2.1. Loading the Source data

Data and variables descriptions

The raw data is housed in an Excel file, directly exported from the Andex database. Since there is no direct connection between Power BI and this internal database, the Excel file must be used as an intermediary. Loading data from an Excel file is a straightforward process. By navigating to the "Get Data" tab and selecting the desired file, Power BI provides options for data transformation through the Power Query Editor. This allows for tasks such as cleaning, filtering, and modifying the data before importing it into the Power BI model.

The Excel file contains 54,847 rows and 32 columns. The variables are of type integer, string, or object. The only numeric variables are the number of pages (NumberOfPage), the different date columns (AcceptanceDate, ApprovedBefore, CreationDate, LatestUpdate), and the unique code identifiers (ANC). All other variables are strings.

Import Mode vs DirectQuery Mode

By default, data is imported into the Power BI model, and a copy of the data is stored within the Power BI file (.pbix). In Import mode, data is loaded into the Power BI model, making it a self-contained dataset within the Power BI file. Consequently, it means that users are working with a snapshot of the data, and any changes to the original data source will not be automatically reflected in Power BI until a manual refresh is performed.

In contrast, DirectQuery mode allows Power BI to send queries directly to the data source at runtime. This means that Power BI maintains a live connection to the data source instead of importing it. This method was not chosen in this scenario because real-time data is not crucial and the team prioritized having more flexibility in terms of data transformation and modelling, as well as faster response times during analysis.

3.2.2. Data Exploring and Preprocessing

Before progressing further, once the data were loaded, the crucial step of data preprocessing was initiated to retain only the pertinent variables for our analysis. This initial phase consisted of reducing the size of our dataset by selecting and removing any variables that don't hold relevant information. This reduction involved removing columns related to the sales location of UMs, such as [TargetCountry], [SourceLangFlag] and [SoldtoC]. Several additional columns were removed due to their irrelevancy, such as the columns automatically generated by ANDEX to confirm document delivery to specific platforms ([SourceToM3], [SourceToUMR], [ReleasedStatus]).

Additionally, columns describing the UMs status, such as [Status], were removed as they provided excessive detail beyond the team's needs to track unapproved Ums ("Approved" but also "Derogated", "Draft" ...etc). To address the need of a clearer status, a calculated column with a concise distinction between "Approved" and "Not Approved" documents, was created during the modelling phase (explained in part 3.3.3).

Furthermore, certain columns contained relevant information for the team but were not mapped correctly due to being manually entered fields in the database without clear guidelines for data entry. This resulted in inconsistent or blank data for columns such as [Location] and [Platform]. To rectify this issue, a new custom column was created instead, in Power Query, using a mapping to populate the data accurately. This new column aimed to include subcategory details (referred to as Platform_) within the Team hierarchy. The manual mapping was possible because each ANC range belongs to a specific subcategory (e.g., ANC 100001 to ANC 200001 belongs to Platform 6). The M code for this added custom column is presented in Figure 12.

```
= Table.AddColumn(#"Changed Type", "Custom", each Table.SelectRows(Platform_,(x) => [ANC] >= x[Range
Min] and [ANC] <=x[Range Max] )[SubCategory]{0})
```

Figure 12 - M code for the subcategory custom column.

In Total, eighteen columns have been eliminated, reducing the total number of columns to fourteen.

The second phase involved assessing data quality and relevancy, and identifying anomalies to make the data more homogeneous. As we can see in Figure 13, some columns have empty rows, some of them have unique and distinct values.

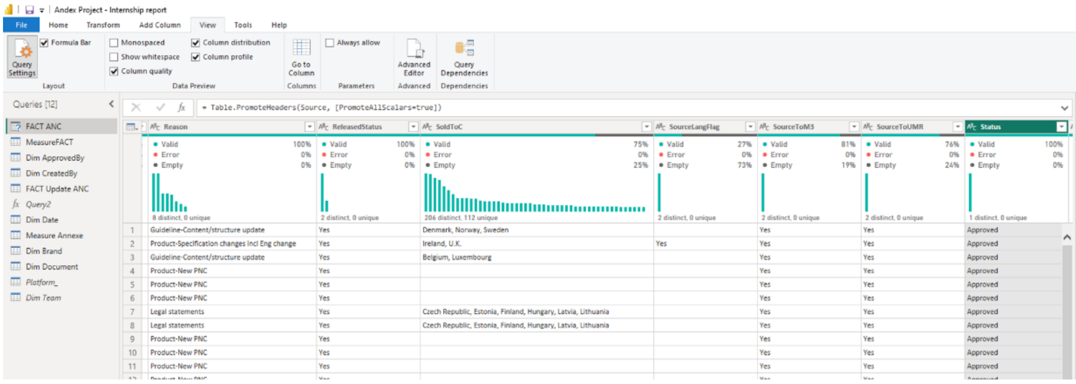


Figure 13 – Profiling of data quality columns in Power Query

Power Query played a pivotal role in cleaning and refining the data to ensure accuracy and homogeneity in our model. Here is a summary of some of the additional actions performed:

• **Checked for missing values and duplicates**

Examined the data for missing values and unusual characters (!, $, %, ?, *, +, _, @, €) that could indicate data entry errors.

• **Dealt with incoherent values**

Many of the values in the ANDEX database were manually typed by humans, so spelling mistakes were common. This could create redundancy, such as in the Brand column, where some people were typing "AEG" and others were typing "AEG-Electrolux," when both terms refer to the same thing. Power Query was used to clean up these inconsistencies.

• **Fixed wrong types**

For example, some of the date columns in the database were showing dates and times, while others were only showing times. Power Query was used to ensure that all date columns were in a consistent format.

• **Performed feature engineering to track updated ANCs**

One of the biggest challenges in preprocessing the data was extracting information about updated ANCs. Indeed, the database only shows the Approval Date of the latest version of the ANC, and the only way to know the historical versions is to go through the [HistoryComment] column, which contains unstructured data, as shown in Figure 14.

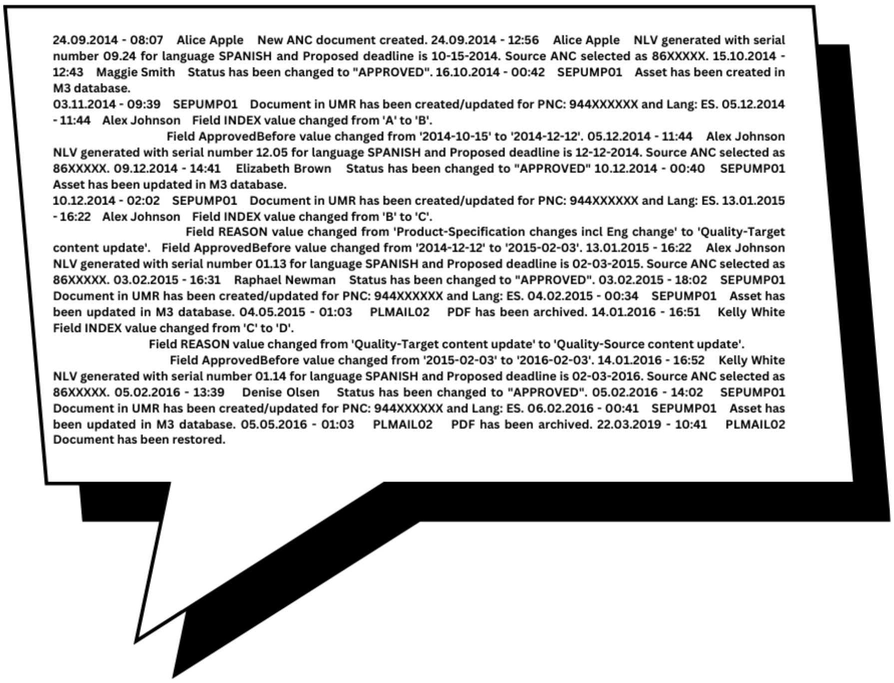


Figure 14 **-** Example of a unstructed data extracted from HistoryComment column

Figure 14 shows an example of those unstructured data for an ANC with four versions (= four updates). As we can see, information about the other updates is displayed in a single block and merged with other irrelevant information. To address this challenge, Power Query was used to split the [HistoryComment] column, using delimiters, and perform other steps to extract the desired data. Then, the date were grouped by rows with the same ANC but different versions (see Figure 15). This allowed to display four separate rows for the same ANC, one for each version, with the corresponding Approval Date.

```
= Table.Group(#"Removed Columns3", {"ANC"},
{{"AllRowsGroupes",
each Table.AddIndexColumn(
Table.Sort(_, {{"HistoricalCommentDate", Order.Ascending}}), "Row Rank",1, 1),
type table [ApprovedBy_ID=nullable number,
ANC=nullable number,
HistoricalCommentDate=nullable date,
HistoricalCommentTime=nullable text,
Row Rank=nullable number]}
}
)
```

Figure 15 **–** Code for the custom Power Query step

This information will be valuable later on for tracking ANC's updates frequency, in the visualization step.

3.3. Building the model

Data modelling is the process of creating a logical representation of data. The model in Power BI represents the relationships and structure of the data. This includes defining tables, establishing relationships between tables, and creating calculated columns or measures using Data Analysis Expressions (DAX). Building the model is crucial for organizing data in a way that makes it conducive to analysis and reporting within Power BI.

3.3.1. Kimball Methodology

The Kimball and Inmon methodologies are two different approaches to data warehousing. The Kimball methodology is a top-down approach that starts with the needs of the business users and then designs the data warehouse to meet those needs. This approach employs dimensional modelling, crafting star or snowflake schemas optimized for user-friendly reporting and analysis. The Inmon methodology is a bottom-up approach that starts with the raw data and then designs the data warehouse to store and manage that data.

The following table 5 summarizes the key differences between Kimball and Inmon methodologies:

Characteristic	*Kimball Methodology*	*Inmon Methodology*
Approach	Top-down	Bottom-up
Focus	Business user needs	Raw data
Data modeling	Dimensional modeling	Normalized modeling
Data warehouse	Encourages data marts	Encourages a single enterprise data warehouse

Strengths	Flexible, quick iterations, responsive to immediate business needs	Comprehensive, integrated view of enterprise data, long-term data management principles
Weaknesses	Can lead to silos of data	Can be complex and time-consuming to implement
Best suited for	Organizations that prioritize flexibility and responsiveness to immediate business needs, organizations with less complex data management requirements	Organizations that prioritize long-term data management principles and need a comprehensive, integrated view of enterprise data, organizations with more complex data management requirements

Table 5 **-** Kimball versus Inmon methodology approach

Considering the need for a modelling structure that supports fast and easy analysis, the Kimball methodology stands out as the most appropriate choice for our data warehouse project. Its simplified star schema structure, facilitates faster and more intuitive data analysis, enabling business users to quickly extract insights without extensive technical expertise. Moreover, no frequent changes in the varibales or business requirements are expected, so in this regards, the Kimball methodology's relative rigidity is not a major concern, as we prioritize a robust and efficient model over the Inmon's one flexibility.

Introduced in 1996, the Star Schema approach requires to, first, look at the requirements and needs of the business (done in part 3.1) and then, identify the fact table (or facts) and its attributes. Going from an Entity Relationships model to a Dimensional Model requires 3 steps, defined in 2002 by Moody and Kortnink:

1. Classify Entities
2. Identify Hierarchies
3. Produce the Dimensional Model

Starting with *Step 1: Classify Entities*, the entities of the ER Model need to be classified into different categories.

- The **Transaction entities** are the tables that contain details about events that happened in a moment m, with measurements. They are useful for decision-makers because they can be quantified and summarized into significant numbers. In this case, there is the ANC and details that contain all the information about UMs created by employees: languages, source or not, acceptance date, number of pages etc. There are also the Updated ANC details, created in power query that contains the information related to the historical updates of UMs.
- The **Component entities** are, as implied in their names, the detailed business components of the Transaction entities. They are directly linked using a One-to-Many relationship and will be the base of the Star Schema.
- Finally, there are the **Classification entities**. They are functionally dependent on the Component Entities, meaning each value of the Component entities is precisely associated with one value from the Classification entities. They are useful to create constraints between two sets of attributes but also contribute to the database's normalization, essential to reduce redundancy and improve its efficiency.

Continuing with *Step 2: Identify the Hierarchies*, Dimensions are organized hierarchically, utilizing a parent-child relationship. This arrangement allows for easy navigation between different levels and facilitates data manipulation for analysis. In this case, for Step 2, the goal was to identify the hierarchical relationships in the ER model. It is represented by any segment of entities linked with 1:n relationships and going in the same direction (Figure 16).

Step 3: Produce the Dimensional Model is the final step and involves transitioning from the Entity-Relationship Model to a Dimensional one, a Star Schema. This was achieved using two different operators:

Dimension tables

In Operator 1, the focus is on collapsing the Component entities established earlier into lower-level entities or classification tables to generate the Dimensions. The creation of Dimension tables is driven by the goal of providing a contextual framework for the fact table. Following this rationale, a Document entity was formed by collapsing all columns that provided contextual details of User manual properties, such as the ANC (unique identifier number), the number of pages, the document type, and so forth. Similarly, a Brand entity was created using the same approach, with all columns related to Brand details. Additionally, two separate entities were established to capture Approver names and Creator names details. These last two Dimension tables were not merged into a single "Employee" entity due to the importance of distinguishing between an approver and a creator for analysis purposes. It might happen that an employee could serve as both an approver and a creator, necessitating the creation of two distinct dimension tables with separate foreign keys on the fact table.

Moreover, a time dimension was needed. Time historical data is an important part of any analysis and, therefore, is a crucial component in any data warehouse. For this purpose, a dimension table, namely "Date," was built. Consequently, this process of collapsing hierarchies resulted in the creation of five Dimension tables.

Fact tables

Each of the transaction entities results in a Fact table. Coming along with building the Fact tables, the granularity had to be defined, meaning the lowest level of information that needed to be stored. It is known that records are being created daily in the ANDEX database, and the team need to look at those movements accordingly. Therefore, having ‘day’ as the lowest level of granularity in the date Dimension was what made more sense, so afterward, adequate analysis could be performed.

While a single fact table is often sufficient for uncomplicated data models, certain situations demand the utilization of multiple fact tables to enhance data organization, optimize query performance, and facilitate specialized analysis. In our specific scenario, the creation of two fact tables became imperative to accommodate intricate data relationships and meet specific analysis requirements. Both fact tables share the same contextual date ([ApprovalDate]), but this date should not be interpreted identically for both. The first table, {Fact Created ANC}, stores the actual records of the last version of ANCs approved and created. This table will be utilized in the analysis to monitor the number of pending ANCs (not approved). The second table, {Fact Update ANC}, is derived from data extracted from the [HistoryComment] column mentioned earlier. It stores historical records of approved ANCs and will be employed to monitor the updates movement. Creating two distinct fact tables was necessary because both tables handle the same types of data but with different time periods. Merging them into one would have limited our ability to effectively compare Approved ANCs versus Updated ANCs.

The role of Operator 2 is about aggregation attributes, produced to summarize data gathered and generated from aggregating different numerical attributes of the key dimensions. As can be seen in Figure 16, there are no aggregated facts.

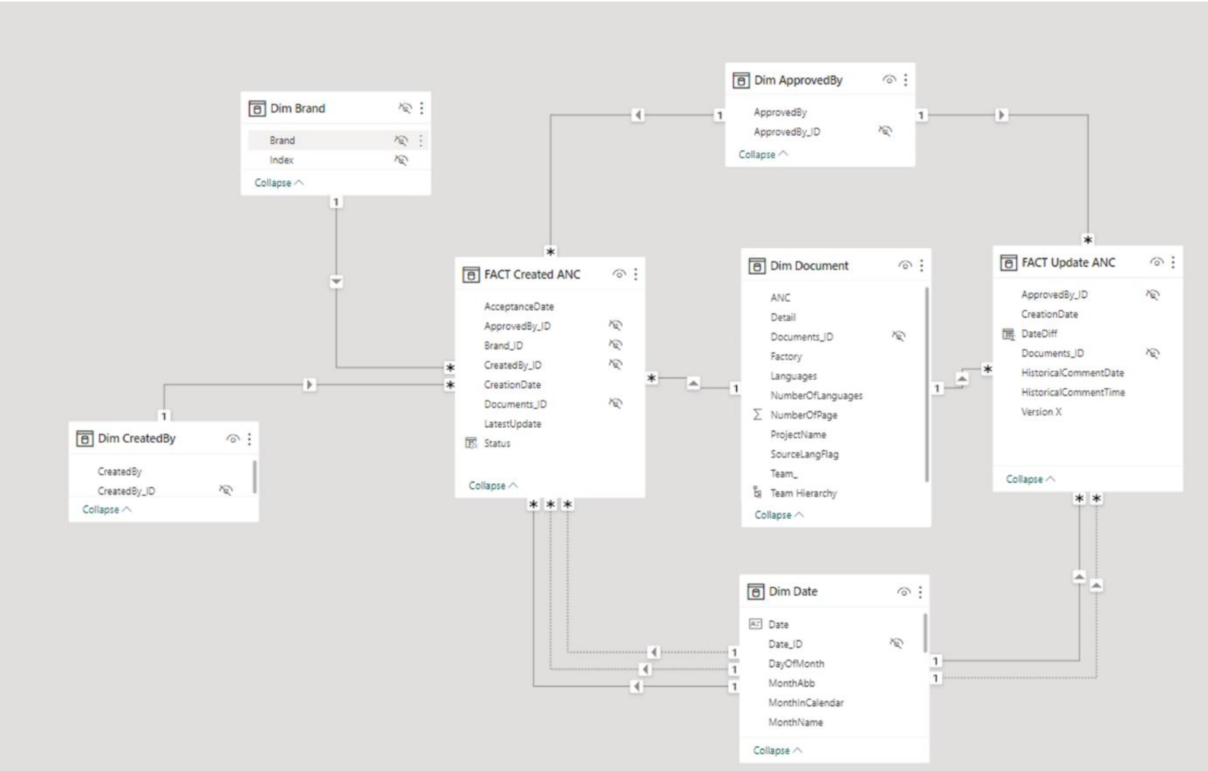


Figure 16 – Final Star schema

The final dimensional model is a combination of a two-star schema. It has a total of five dimensions and two fact tables. Notably, the two fact tables are not directly interconnected; rather, they share three common dimensions—Document, ApprovedBy, and Date. The CreatedBy and Brand Dimensions were intentionally not linked to the {FACT Update ANC} due to the absence of Creator names and brand details in the [HistoryComment] column, which served as the source of data for the Fact table. This set of star schemas, combined but not directly linked, is more commonly called a Galaxy schema.

It can be noticed that the dimension Date has a double connection with both fact tables and triple with the {FACT Created ANC}. This is due to the fact that the tables have more than one important date in them: AcceptanceDate, LatestUpdate, and Creation Date. However, only one of those connections is active at a time (cf solid lines).

3.3.2. Data Preparation

In this final step, the hierarchies are defined. Then, unnecessary information is minimized by hiding some columns in the data view from showing up in the report view. Firstly, all the primary keys and foreign keys were hidden. Secondly, any columns that already existed in a hierarchy were hidden as well since they will be displayed under the hierarchy in the report view. Finally, columns or tables that are not pertinent to the business objectives of the company were hidden. For instance, the Brand table was hidden because it is not used for now in the analysis but might be needed later on.

In this final step, the columns were changed to the proper data types, and the {Dim Date} table was marked as a date table to optimize the date-related visuals and to use time intelligence functions.

3.3.3. Measures and Calculated columns

The creation of calculated columns is necessary in Power BI when we require a static value or a value that remains constant for each row in the dataset. On the other hand, DAX measures are dynamic and recalculated for each visualization or interaction, making them suitable for aggregations or calculations based on user input. Depending on the nature of the calculation and the desired behavior, choosing between calculated columns and DAX measures is crucial to achieve accurate and efficient results in Power BI.

Several DAX measures were generated to calculate the necessary information and produce relevant visuals for the final dashboard. Examples of these measures are presented in table Figure 17.

DAX Measures	Formula

Number of Approvals	`A - Total # Approved = CALCULATE(`
	`COUNT('FACT Created ANC'[Documents_ID]),`
	`FILTER('FACT Created ANC', 'FACT Created ANC'[Status] = "Approved")`
	`)`

Average Workload in 1 day	`H - AVG of Approvals Daily = // Number of workload per working day (ANC created + updated)`
	`VAR NumberofWorkingDays = Calculate(`
	`DISTINCTCOUNT('FACT Update ANC'[HistoricalCommentDate]),`
	`USERELATIONSHIP('FACT Update ANC'[HistoricalCommentDate], 'Dim Date'[Date]))`
	`VAR Result =`
	`DIVIDE([H - Total # of Approval], NumberofWorkingDays)`
	
	`Return`
	`Result`

Average of days to approve a UM	`H - Avg of Days before Approvals = // Avg of days to approve a document` `AVERAGE('FACT Update ANC'[DateDiff])`
Number of Approvals Last Year	`H - LY # of Approvals =` `CALCULATE([H - Total # of Approval], SAMEPERIODLASTYEAR('Dim Date'[Date]))`

Number of ANC Created	`A - # of document Created = // Number of ANC created using CreationDate unactive relationship`
	`CALCULATE(`
	`COUNT('FACT Created ANC'[Documents_ID]),`
	`USERELATIONSHIP('FACT Created ANC'[CreationDate], 'Dim Date'[Date]))`

Number of ANC updated after less than 1 day	`H Support - # of Updates < 1 day =`
	`CALCULATE(`
	`COUNTROWS('FACT Update ANC'),`
	`FILTER('FACT Update ANC', 'FACT Update ANC'[Version X] > 1),`
	`FILTER('FACT Update ANC', 'FACT Update ANC'[DateDiff] <= 1))`

Figure 17 - Sample of some DAX Measures. *Source: Created by me*

In addition to the measures described above, two calculated columns were also required.

- The first calculated column, **[DateDiff],** was created in the {FACT Update ANC} table to improve performance. This column calculates the number of days between the creation date of each ANC and the update date for each version. This is a relatively complex calculation that will be performed frequently, so using a calculated column is the most efficient method possible. Additionally, with the [DateDiff] column, it is easier to filter the ANC table to display only those updated in the last 30 days (for example). This allows to quickly identify ANCs that need to be reviewed or updated.
- The {FACT Created ANC} table features a second calculated column, **[Status],** designed to determine the status of each ANC, classifying it as either "Approved" or "Not Approved" in instances where there is no Acceptance Date. This column was necessary to create Dax measures that would calculate the number of approvals (DAX detailed in Figure 16) and, additionally, would highlight pending ANCs—those that are categorized as "Not Approved" at present.

The logic behind both calculation is detailed in Figure 18.

<table>
<tr><th>Calculated column</th><th>Formula</th></tr>
<tr><td>Status</td><td>Status = IF('FACT Created ANC'[AcceptanceDate] = BLANK(), "Not Approved", "Approved")</td></tr>
<tr><td>DateDiff</td><td>DateDiff =
//FirstDate = It returns the Min Date for each group of ANC's update (ALLEXCEPT), that matches the first time an ANC have been approved
VAR FirstDateSub =
CALCULATE(
    MIN('FACT Update ANC'[HistoricalCommentDate]),
    ALLEXCEPT('FACT Update ANC', 'FACT Update ANC'[Documents_ID]))
//DemoteSub = It looks at the current row (EARLIER), per group of ANC, and check if there are smaller approval date. If there are, it returns next to the current row, the max of the smaller approval dates.
VAR DemoteSub =
CALCULATE(
    MAX('FACT Update ANC'[HistoricalCommentDate]),
    FILTER(
        ALLEXCEPT('FACT Update ANC', 'FACT Update ANC'[Documents_ID]),
        'FACT Update ANC'[Version X]<EARLIER('FACT Update ANC'[Version X])))
Return
IF(
    'FACT Update ANC'[HistoricalCommentDate]=FirstDateSub,
        IF(
            'FACT Update ANC'[HistoricalCommentDate]=FirstDateSub && DemoteSub,
            0,
            DATEDIFF('FACT Update ANC'[CreationDate],'FACT Update ANC'[HistoricalCommentDate],DAY)),
    DATEDIFF(DemoteSub, 'FACT Update ANC'[HistoricalCommentDate],DAY)
)</td></tr>
</table>

Figure 18 - Calculated columns code. *Source: Created by me*

3.4. Perspective of Analysis

3.4.1. Dashboard Perspectives

In formulating the analytical approach, the decision was made to strategically segment the final dashboard into three distinct perspectives tailored to meet the specific needs of the CDT team. A shared feature across these perspectives is the integration of slicers, empowering users to apply filters based on Team, Category, Year, and a specific selection for enabling a view of "*Source documents*" only or "*Target documents*" only. Given the array of filtering options available, users can also effortlessly remove all filtering by clinking on the eraser icon (linked to a bookmark to make this feature worked).

Each perspective in the dashboard showcases specific information. The Workload Overview (Figure 19) perspective presents workload-related metrics. In the Approval Repetitions perspective (Figure 22), the focus is on tracking the frequency of document updates and identifying unusually frequent occurrences within a specified time frame. The Paper saving Overview perspective (Figure 23) provides a comprehensive overview of paper-saving initiatives.

Workload Overview

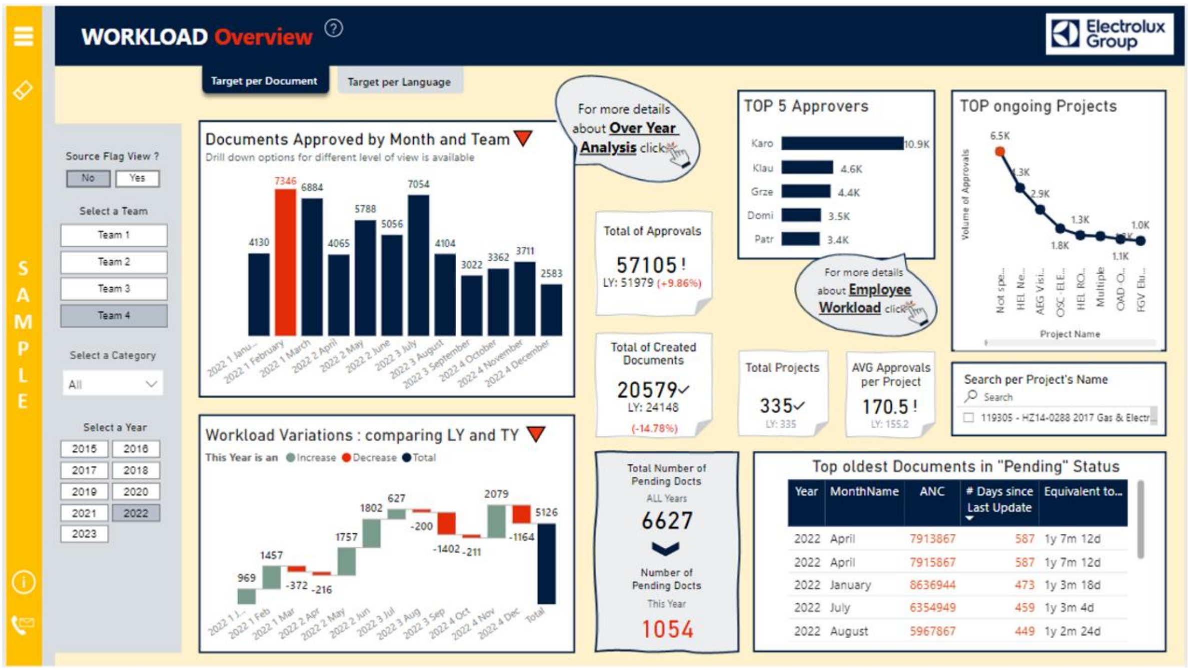


Figure 19 - Workload perspective, pbi project

This is the initial and main report, summarizing the monthly production of User Manuals. It provides a yearly view, allowing users to track trends over the year and compare the monthly volumes. with the previous six months. The report features several visual elements:

<u>Column Chart</u>: In the top left, a column chart displays the Volume of ANCs per month for the selected year. It is where the human eye is naturally first drawn, so it is where important information is being displayed for more efficiency. The visual's strength lies in its functionality to drill down, allowing

users to compare data other than just between months, using different hierarchies. This facilitates a deeper understanding of the workload distribution. For example, users can drill down the visual to get the Volume of ANCs not per month but per Team, to understand better who has the biggest workload.

<u>Waterfall Chart</u>: This chart serves to illustrate how the volume of ANCs has changed from the previous year. Red bars indicate a decrease, while green bars indicate an increase. This visualization provides a clear view of year-over-year trends in UMs production.

<u>Bar Chart</u>: This chart highlights the top 5 employees handling the most workload. By identifying the individuals with the highest contribution, team leaders can gain insights into resource allocation and potential areas for improvement.

<u>Table</u>: This table serves to flag ANCs that have not been approved for an extended period. A fast approval flow is crucial for the departmental efficiency, so tracking pending or overlooked items is essential. To achieve this, the table utilizes a DAX measure and the calculated column [DateDiff] introduced earlier. ANCs not approved in over a year are prominently displayed in red, drawing attention to items that require immediate attention.

<u>Cards and KPI Cards</u>: These cards provide a snapshot of key metrics, offering a quick overview of the business's state. Among these are two crucial KPI metrics: Total Approvals and Total Created Documents. The distinction between these metrics is important, as created documents typically demand more time and effort than updated ones. Understanding the percentage of Created ANCs out of the total Approvals is essential for balancing the workload and addressing potential bottlenecks. For instance, in Figure 18, we observe that over 82% of ANCs handled by Team 2 are newly created, while Team 4, with four times more Approvals, has only 36% of created documents. This information nuances the varying distribution of workload between the two teams.

For a more in-depth analysis, users can explore two sub-reports (Figure 20 and Figure 21) linked to this primary report:

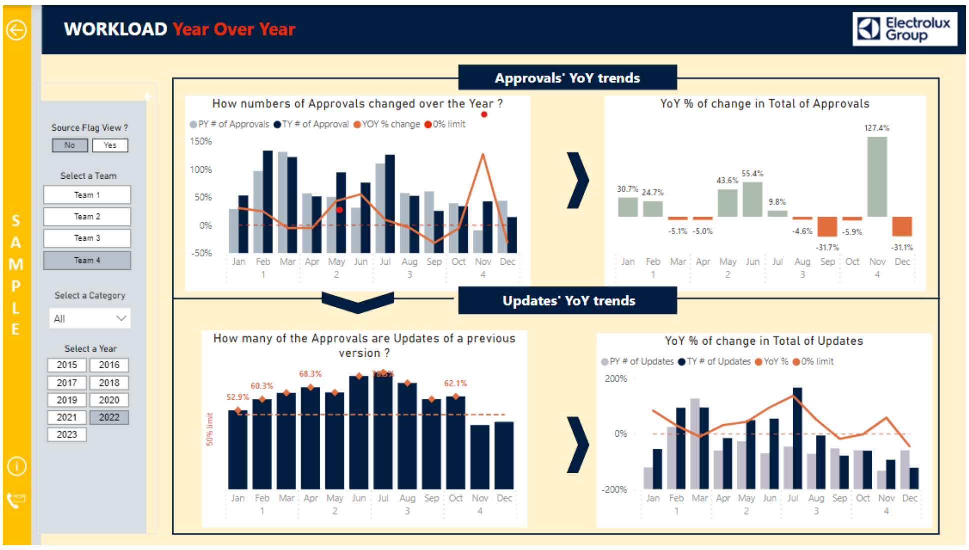


Figure 20 - Workload sub report focusing on Year over Year analysis

This first sub-report provides a detailed overview of how the workload varies over the year, with a focus on approved, created, and updated documents. It offers a more granular perspective on workload trends and helps identify areas for improvement.

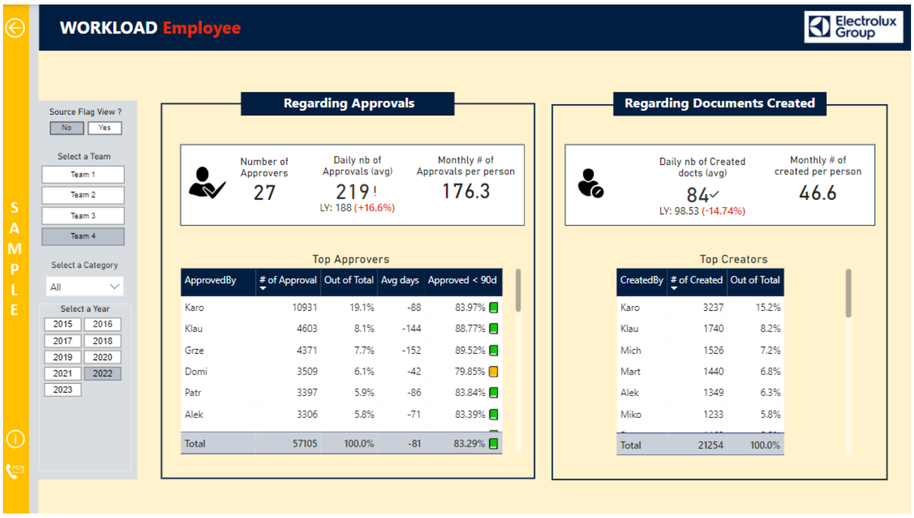


Figure 21 - Workload sub report focusing on Employees

This second sub-report focuses specifically on employee workload, providing insights into how much each employee is handling, who is approving documents more efficiently, and the average number of days taken to approve a document. It helps managers optimize resource allocation and address any potential bottlenecks in the approval process.

Approval Repetitions

Figure 22 - Approval Repetitions Perspective, pbi project

This Approval Repetitions' report focuses on updated ANCs, providing a comprehensive overview of their update patterns, version history, and potential anomalies. Its primary goal is to track the movement of updated ANCs, keeping tabs on their version count and release dates. Additionally, it identifies ANC updates with unusual patterns, such as those completed in a remarkably short time frame (less than a day). The development of this report was made possible by the transformation of the unstructured [HistoryComment] column into structured data using Power Query.

The report's top section provides a broader overview, while the lower visuals focus into individual ANC details.

Moreover, the date filter in this visual is set to a monthly time granularity. A monthly view is preferable because it offers a more granular perspective of what is happening on a month-to-month basis. A yearly view would be much larger and more difficult to interpret. This is because it would need to display data for all 12 months of the year, which would make it difficult to see any patterns or trends. A monthly view, on the other hand, is much more compact and easier to interpret.

<u>Clustered column chart</u>: This chart compares the number of updated ANCs with the total number of approved ANCs for the selected month. This comparison provides a relative understanding of update activity and shows how many of the approved ANCs have been updated. This visual also enables users to compare the Team's workload to the workload of other Teams.

<u>Cards</u>: These cards show important metrics, to be read from left to right. First one is the number of ANCs updated for the selected time period, then within the same time frame, how many were updated more than once, and finally the one updated in less than 1 day.

<u>Treemap</u>: Treemaps are particularly effective for showing the relationships between different levels of data and for identifying patterns and trends. In this report, the treemap is used to highlight ANCs with more than one approval in the selected time period.

<u>Line chart</u>: When an ANC is selected from the treemap, the line chart in the bottom right corner displays the update history for that ANC over the years. This allows users to see when the ANC was updated, who made the updates, and what the updates were. This information can help users understand why the ANC was updated, identify any potential issues, and know who to contact to solve them.

Similarly to first overview, this report offers a sub-report that dives deeper into the number of updates per factory, providing insights into their performance and update speed. This detailed view is particularly valuable for the CDT team, enabling them to compare the performance of different factories, identify areas for improvement, and communicate their needs more effectively. The sub-report helps them understand the capacity of each factory and ensures that their communication is aligned with their actual performance. (Figure 23)

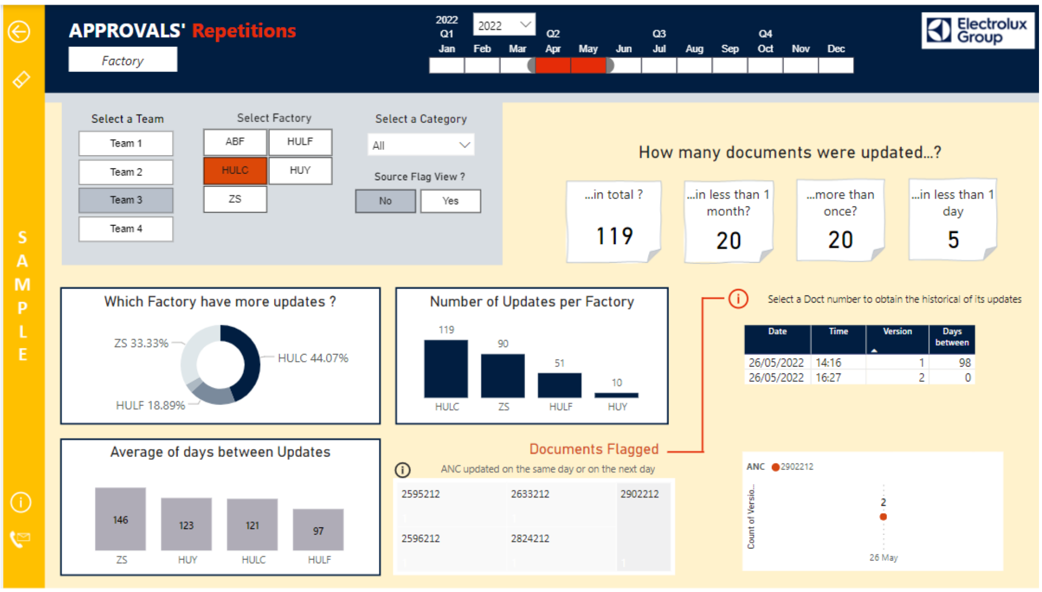


Figure 23 - Approval Repetitions sub-report focusing on Factory

PaperSaving

Consider this final report as a supplementary addition. Its purpose is to offer insights into a Papersaving project the company is actively pursuing, aiming to minimize the number of pages in their user manuals guide for both cost savings and environmental objectives.

The visuals and underlying logic are similar to the two previously mentioned reports. The key distinction lies in the variables employed for this report, which center on the {DIM Document} table, specifically focusing on the number of pages and the various types of User Manuals (UMs).

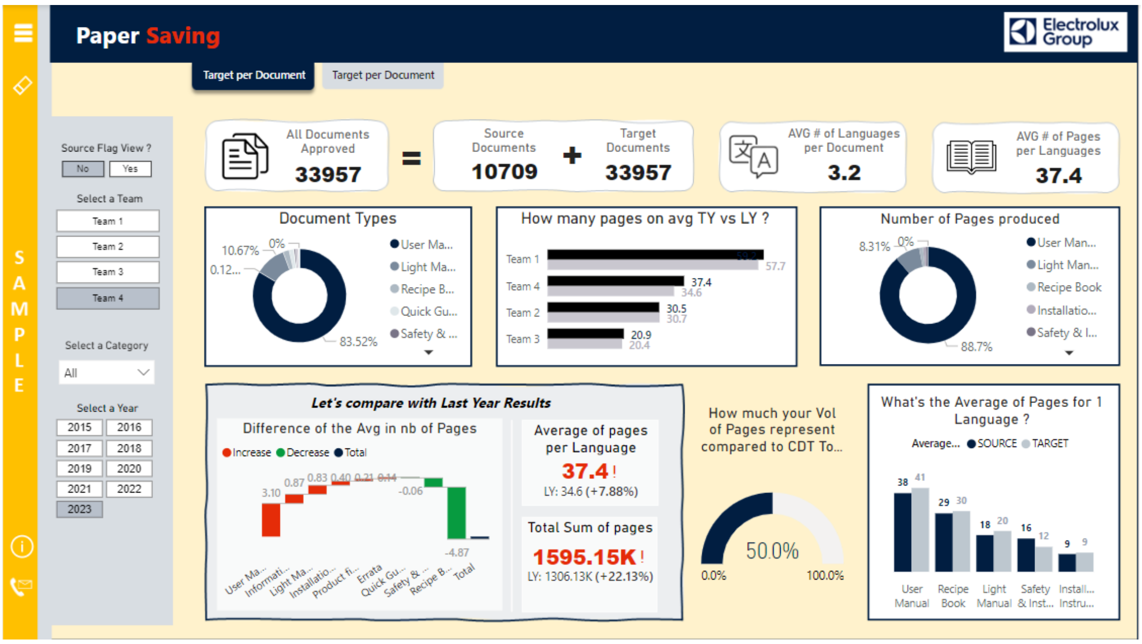


Figure 24 - PaperSaving Perspective, pbi project

3.4.2. Technical Aspects

These dashboards have been carefully designed to meet the specific needs of the team. Users can easily customize their views using various dimensions and criteria to extract relevant insights. To enhance user navigation, especially for those less familiar with Power BI features, two additional technical pages have been incorporated. The first is a Help Page (Figure 25), accessible through a question mark icon present in all reports. Clicking on it redirects users to a summarized page containing tips and tricks to effectively use and understand the insights in the reports. Additionally, there's a Navigation Menu, powered by bookmarks, enabling users to switch seamlessly between reports with a single click. This feature proves particularly useful in configurations where the Power BI file is published on Power App.

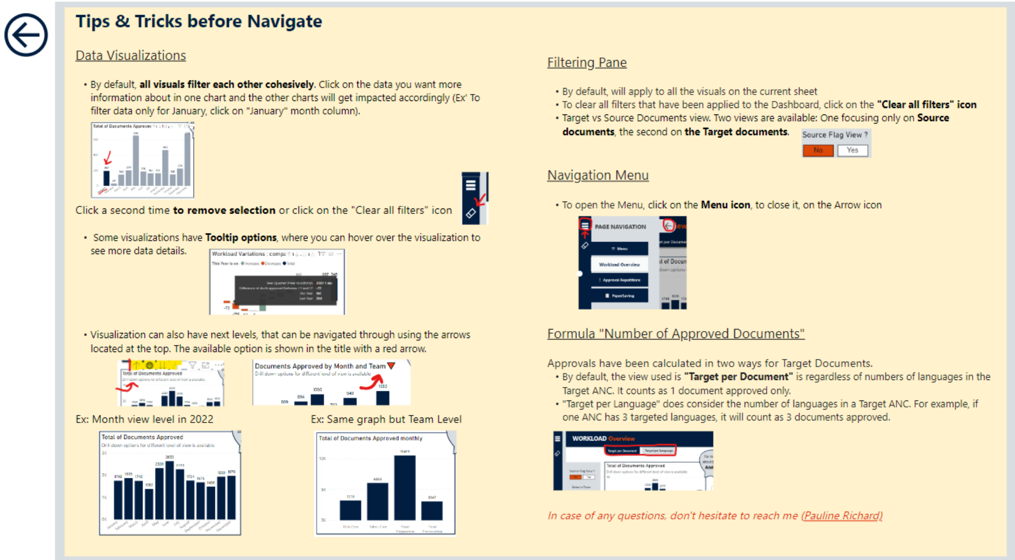


Figure 25 – Help page (hidden), pbi project

Finally, To minimize text clutter and enhance visual clarity, tooltips were added throughout the reports. In Power BI, tooltips are small pop-up windows that display additional information about a data point when the user hovers over it. Tooltips can be used to provide additional context or details about a data point, making it easier for users to understand the data. In our context, tooltips are accessible by hovering over the question mark next to each page title, providing further context and terminology explanations to enhance data comprehension (Figure 26).

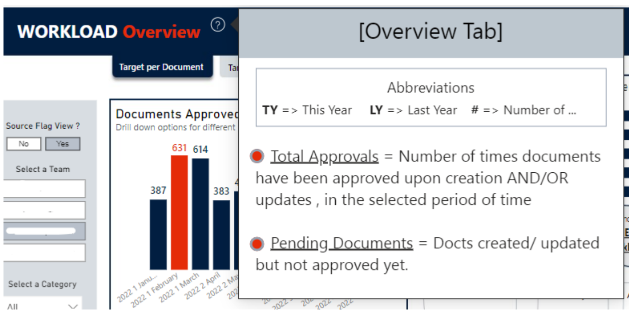


Figure 26 - Example of a tooltip on Workload Report, pbi project

4. RESULT AND DISCUSSIONS

The successful implementation of the dashboard involved a close collaboration with the managers of each team. They provided valuable input on the desired content, presentation, and understanding of the terminology and relationships between the variables. Additionally, the CDT team previously relied on manual data entry in Excel. Ensuring that the data, automatically preprocessed and measured in Power BI, matched the manually ones calculated Excel, was crucial to gaining everyone's trust in the accuracy of the information in Power BI and its suitability for decision-making.

Another key success factor was adhering to change management procedures and actively involving a significant portion of the team in the discussions, enabling them to gradually adopt the new tool. Power BI's power comes with a level of complexity that requires an adjustment period for users unfamiliar with it. For instance, utilizing drill-down features on visuals or understanding that all visuals are interactive and can be used to filter each other’s (outside of the slicers functions), is still relatively new and unnatural to the team. This is partly why the Help page (Figure 25) was added to the report. The goal is to prevent users from feeling lost while navigating the reports.

Since the latest version was published, and the kickoff meeting occurred in September 2023, we've observed consistent monthly engagement with the dashboard by managers (Figure 27). Currently, the dashboard is only accessible to the top five managers in the CDT Team. The report's usage metrics, automatically generated by Power BI, reveal that more than 20 views happen per day, and different people are using it too. This data serves as an indicator of the successful adoption and regular utilization of the dashboard by the management team. However, it's important to note that since the number of users having access to it is relatively low, it does impact the metrics measured.

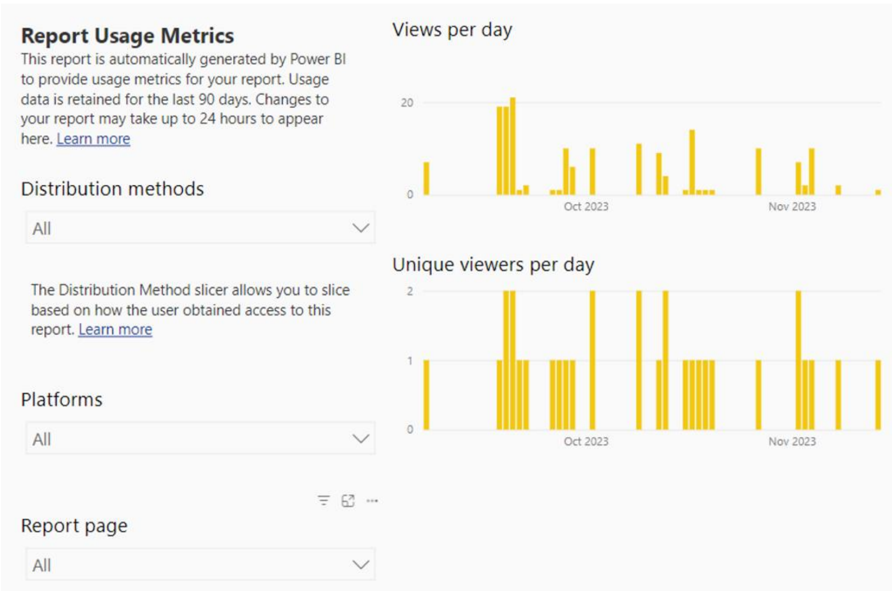


Figure 27 - Report usage metrics, pbi project, extracted on the 15.11.2023.

5. CONCLUSION

In conclusion, my internship project at Electrolux focused on improving insights through the implementation of a business intelligence solution, specifically in the form of a Power BI report. This experience has significantly deepened my understanding of the complexity of data modelling, as well as the critical importance of data quality and accessibility, and highlighted the significance of structuring work to effectively define and achieve intended goals. This multi-faced project offered a diverse range of knowledge and experience across various domains. The biggest difficulties were in identifying all the variables relevant for analysis, compartment the reports into meaningful and structured information, and selecting the appropriate visuals to effectively convey the findings. The Master's courses played a crucial role in my professional development, particularly the Business Intelligence class, which provided the foundation for embarking on this project. Additionally, the Data Mining, Data Visualization courses and the others contributed to a holistic understanding of Data Analytics and fostered out-of-the-box thinking.

Addressing the initial question of how BI solutions can assist a multinational company in leveraging its data and developing analytical insights, we can affirm that the implementation of this Power BI report has resulted in various positive outcomes and benefits. Firstly, there has been an increased emphasis on data-driven approach rather than intuition-based decision-making, by providing users with access to a diverse array of data. Secondly, the improved ability to identify trends and patterns throughout the year particularly stands out. The team was successfully able to recognize seasonality trends in the approval workflow, enabling managers to anticipate future workload more efficiently. Additionally, as a data analyst, I've experienced firsthand the positive impact of collaboration between analysts and business users. The implementation of a BI solution has broken down barriers between analysts and business users, allowing us to team up and pull insights from the data more effectively. Finally, the BI solutions have made it easier for everyone in the organization to make informed decisions. At every level, people now have access to the data they need whenever they need it, leading to better decision-making throughout the entire company.

Therefore, this internship provided me with valuable real-world experience, highlighting the challenges and opportunities within the digital landscape, from building reliable and functional dashboards to the substantial time and effort required for a successful BI project. Despite facing various challenges, including understanding business requirements, managing interpersonal relationships, and tackling unforeseen issues, this professional experience was crucial for the start of my career in Data Analytics. It significantly contributed to my professional development, allowing me to apply theories learned at NOVA IMS. Above all, it confirmed my passion for working in this field, motivating me to pursue further studies and specialization. I discovered that I thrive in addressing analytical challenges, proposing practical solutions, and exploring new business opportunities. As a conclusion, the internship has been a pivotal moment in my career, generating new possibilities, insights, ideas, and fueling my personal and professional growth as a Data Analyst.

6. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORS

Recognizing the constraints of the implemented solution is essential for refining future projects. Throughout the development process, the primary limitations encountered were addressed through the utilization of DAX, a powerful data analysis expression language. Learning how to write efficient DAX code is critical for enhancing report performance. DAX serves as a gateway to unlocking the full potential of Power BI. Therefore, it is recommended to further develop DAX skills to deliver high-performing Power BI reports to end-users.

Another recommendation would be to utilize the Power BI performance analyzer feature. It is a built-in tool that provides a detailed breakdown of the time taken to execute each query and visualization in a report, as well as the time taken to render and display the data in the visuals. By examining report elements with the performance analyzer, we can identify which visuals are taking the longest to load and which queries are causing the most performance issues. We can then identify opportunities to improve the report performance, whether by optimizing DAX expressions or selecting a more appropriate visual for data visualization.

REFERENCES

- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). John Wiley & Sons.

- Olaniyi, Oluwaseun and Abalaka, Anthony and Olabanji, Samuel Oladiipo, Utilizing Big Data Analytics and Business Intelligence for Improved Decision-Making at Leading Fortune Company (September 14, 2023). Journal of Scientific Research and Reports, Volume 29, Issue 9, Page 64-72, 2023, Available at SSRN: https://ssrn.com/abstract=4571876

- Eisenberg, M., & Dirks, L. (2008). Taylor's Value-Added Model: Still Relevant After All These Years. *iConference UCLA, Los Angeles, CA*. Retrieved from https://faculty.washington.edu/mbe/Eisenberg_Dirks_Taylor_Value-Added_Modified_2008.pdf

- Naveen Kunnathuvalappil Hariharan. (2018). DATA SOURCES FOR BUSINESS INTELLIGENCE. International Journal of Innovations in Engineering Research and Technology, 5(11), 75–80. Available at SSRN: https://ssrn.com/abstract=3914444

- Taylor, R. S. (1982). Value-Added Models for Decision Support Systems. In *Decision Support Systems* (pp. 273-297). Springer, Boston, MA.

- Langit, L. (2007). *Data Mining for Business Intelligence: Concepts, Techniques, and Applications*. Upper Saddle River, NJ: Prentice Hall.

- Sukhobokov, A. A., Gapanyuk, Y. E. (November 2022). The concept of an intelligent data lake management system: Machine consciousness and a universal data model. Procedia Computer Science, 213, 407-414. Available at https://doi.org/10.1016/j.procs.2022.11.085

- AltexSoft. (18 Mar, 2021). ETL vs ELT: Compared and Explained. Retrieved from https://www.altexsoft.com/blog/etl-vs-elt/

- Galaktikasoft. (Dec 12, 2017). OLAP Basics and Multidimensional Model. Overview of Online Analytical Processing Technology. Retrieved from OLAP Basics and Multidimensional Model. Overview of Online Analytical Processing Technology (galaktika-soft.com)

- Feng, G., & Fan, M. (2023). Research on Learning Behavior Patterns from the Perspective of Educational Data Mining: Evaluation, Prediction, and Visualization. *Computers & Education*, 223, 106966.

- Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Hua, X. (2015). Data Mining for Healthcare: A Comprehensive Overview. *IEEE Access*, 3, 1853-1876.

- Hakim, D. K., Gernowo, R., & Nirwansyah, A. W. (2023). Flood prediction with time series data mining: Systematic review. Natural Hazard Research, 139. Retrieved from: Systematic review (sciencedirectassets.com)

- Sadiku, M. N. O., Shadare, A. E., Musa, S. M., & Akujuobi, C. M. (2016, December). Data Visualization. *International Journal of Engineering Research and Advanced Technology (IJERAT)*, 2(12), 11-20.

- Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making Data Visualization More Efficient and Effective: A Survey. *The VLDB Journal*, 29(1), 93-117.

- Parenteau, C., Tatu, A., & Johnson, J. (2016). *Data Visualization: A Practical Introduction*. Princeton University Press.

- Elena Jaramillo, Manuel Munier, Philippe Aniorte. Information Security in Business Intelligence Based on Cloud: A Survey of Key Issues and the Premises of a Proposal. Jul 2013, Angers, France. hal-01082063

- Forbes Business Council. (2023, January 9). Data Privacy vs. Data Security: Four Implications for Business Leaders. *Forbes*. Retrieved from https://www.forbes.com/sites/forbesbusinesscouncil/2023/01/09/data-privacy-vs-data-security-four-implications-for-business-leaders/?sh=69dd6ce86afa

- Rahnama, H. (2022, February 1). The New Rules of Data Privacy. *Harvard Business Review*. Retrieved from https://hbr.org/2022/02/the-new-rules-of-data-privacy

- International Data Corporation (IDC). (2023). *Worldwide Big Data and Analytics Forecast, 2022-2026: A Vendor Assessment*. Retrieved from https://www.marketresearch.com/IDC-v2477/

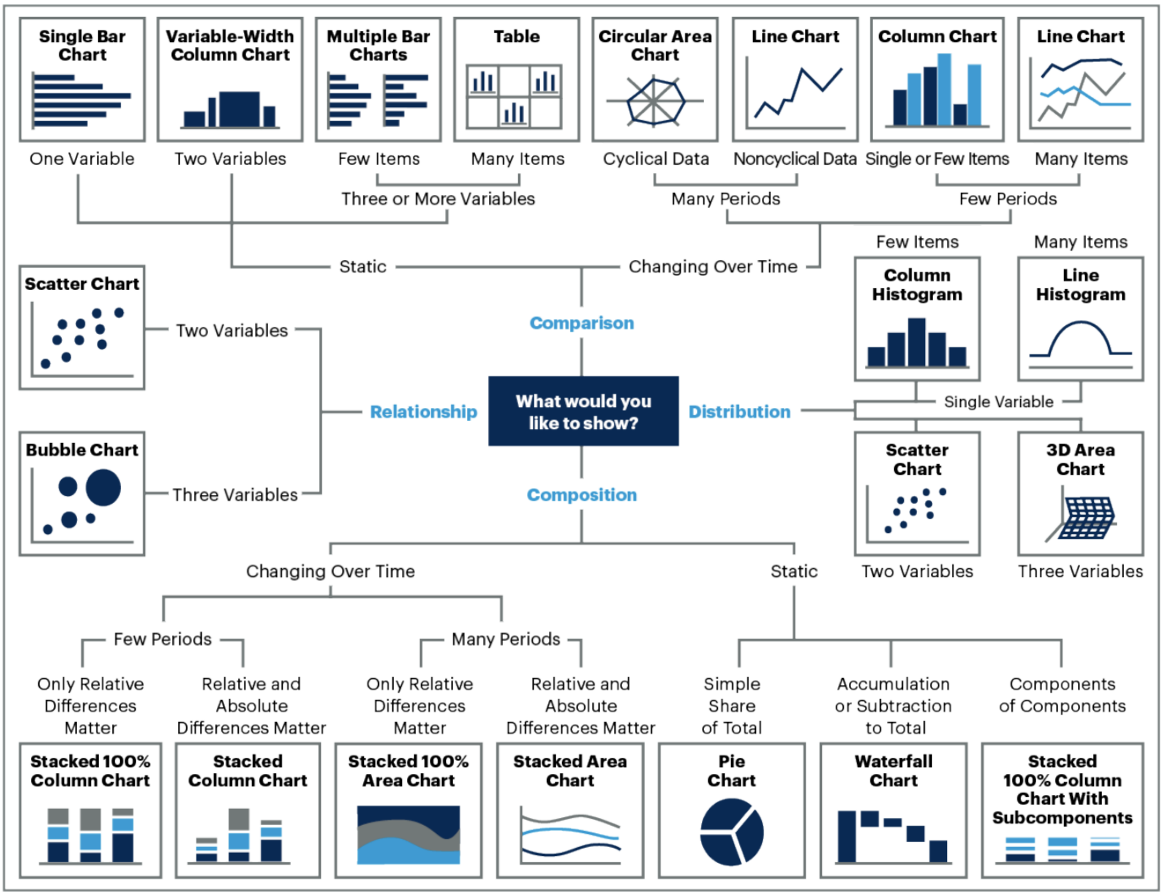
- Synoptek. (2022, February 4). *How business intelligence solutions benefit the retail industry*.

- Moody, D. L., & Kortink, M. (2002*). From ER to DM in 3 easy steps. In Proceedings of the 18th International Conference on Data Engineering* (pp. 750-761). IEEE.

- Astera. (2023, August 8). *Data Warehouse Concepts*. Retrieved from https://www.astera.com/type/blog/data-warehouse-concepts/

- Chen Y, Li C, Wang H. Big Data and Predictive Analytics for Business Intelligence: A Bibliographic Study (2000–2021). MDPI *Forecasting*. 2022; 4(4):767-786. https://doi.org/10.3390/forecast4040042

APPENDIX A - DECISION TREE FOR SELECTING APPROPRIATE GRAPHICS



Source: Gartner

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

Universidade Nova de Lisboa