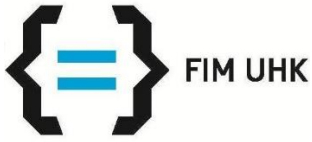
University of Hradec Kralove

Faculty of informatics and management



Dashboards Automation

Master thesis

Author: Bc. Viktorie, Láchová
Major: Information management

Thesis supervisor: prof. RNDr. Petra Poulová, Ph.D.

Hradec Králové April 2024

Declaration:

I declare that the presented thesis is my work, which I created independently using the methods, sources and properly cited literature that is listed in the Bibliography chapter.

In Hradec Králové on April 10th 2024 Viktorie Láchová

Acknowledgment:

I would like to thank Prof. RNDr. Petra Poulová, Ph.D. for the methodical guidance of the work and her valuable advice that contributed to the fulfillment of the key objectives.

Annotation

This master thesis is dedicated to the field of automation of dashboards in a corporate and business environment, thereby offering a detailed view of the field of business intelligence (BI) from a theoretical and practical point of view.

It addresses the critical role of BI in light of current business operations and needs and emphasizes the importance of properly connecting theoretical principles with practical implementation. The practical part allows insight into the development processes of a dashboard displaying data from the Showpad platform. Emphasizes the alignment of features with business goals and a detailed development process based on efficient gathering of dashboard requirements, through data management, design, deployment, and dashboard adoption.

It not only improves business processes and contributes to growth, but also increases efficiency by properly developing and applying BI strategies and optimization methods in the enterprise environment. This leads to a more informed decision-making process in the company.

Keywords:

Dashboard Automation, Business Intelligence, Data Management, Data Modeling, MS power BI, ETL, Optimization, REST API, Showpad Platform

Anotace

Tato diplomová práce se věnuje oblasti automatizace dashboardů v podnikovém a obchodním prostředí, čímž nabízí detailní pohled na oblast business intelligence (BI) z hlediska teoretického i praktického.

Zabývá se kritickou rolí BI ve světle současných obchodních operací a potřeb a zdůrazňuje význam správného propojení teoretických principů s praktickou realizací. Praktická část umožňuje nahlédnout do procesů vývoje dashboardu zobrazujícím data ze Showpad platformy. Zdůrazňuje synergii funkčnosti, obchodních cílů a podrobného procesu vývoje založeného na efektivním shromažďování požadavků dashboardu prostřednictvím správy dat, návrhu, nasazení a přijetí dashboardu.

Nejenže zlepšuje obchodní procesy a přispívá k růstu, ale také zvyšuje efektivitu správným vývojem a aplikací BI strategií a optimalizačních metod v podnikovém prostředí. To vede k informovanějšímu rozhodovacímu procesu ve společnosti.

Klíčová slova:

Automatizace dashboardu, Business Intelligence, Správa dat, Modelování dat, MS power BI, ETL, Optimalizace, REST API, Platforma Showpad

Abstract

This master thesis presents a comprehensive exploration of dashboard automation within the enterprise workspace. It articulates a dual approach, starting with a robust theoretical foundation that encompasses the multifaceted aspects of Business Intelligence (BI), revealing its importance in the modern business environment. The information weaves together performance metrics for BI investments and key stakeholder roles while demystifying complex BI techniques and methodologies. Considerable emphasis is placed on optimization strategies and dashboard design, which can significantly increase the usefulness and effectiveness of BI tools in an organizational environment.

Moving from theory to application, the work dives into the practical aspect, where the main task involves the development of a dashboard based on data from the Showpad Platform. This segment sheds light on the painstaking process of gathering requirements directly from the business project owner and ensuring that dashboard features are aligned with strategic business goals. The thesis details the data extraction process using Python, followed by sophisticated data transformation methods that solidify the backbone of a dynamic BI system. The culmination of this work is the strategic deployment of these techniques, which will not only make data visualization more efficient but also enable end users to quickly make informed decisions.

The result of this thesis is a dashboard, which is evidence of the convergence of theoretical knowledge and practical application and offers substantial added value for the field of dashboard automation and potential end-users. The work asserts the potential of automated dashboards to revolutionize data interaction in enterprises, catalyzing better dissemination of business intelligence and strategic decision-making.

Contents

List of Abbreviations

AI	Artificial Intelligence
APIs	Application programming interfaces
BI	Business Intelligence
BRD	Business Requirements Document
CEO, CIO, CFO	Chief Executive, Chief Financial, Chief Information Officer
DAX	Data Analysis Expressions
ETL	Extract, Transform, Load
HR	Human resources
IT	Information Technology
KPIs	Key Performance Indicators
ML	Machine Learning
NLP	Natural Language Processing
REST	REpresentational State Transfer
RLS	Row-Level Security
ROI	Return on Investment
JSON	JavaScript Object Notation
UAT	User Acceptance Testing
XAI	Explainable Artificial Intelligence

1 Introduction

In a rapidly evolving world, our era is defined by technological advancement. The so-called "information explosion" is flooding lives with data and technology and changing the way the world works. Among the technology's many advantages is its inexhaustible memory – a stark contrast to the limitations of the human brain, which makes technologies the assets of our time. These technological assets propel businesses to new heights, with digitization no longer a luxury but a survival imperative (STAMFORD, 2021). Digital transformation seamlessly integrates IT into business processes, enabling companies to scale and compete effectively. In contrast to historical reliance on direct competition for market dominance, contemporary success hinges on active engagement with innovative technologies (Ge, 2023).

The abundance of data presents both a challenge and an opportunity for organizations to make informed decisions without squandering resources or increasing risk. Enter data management tools and pivotal assets that present information intuitively and visually. Among such solutions is the dynamic dashboard, which is a way of displaying several types of visuals in one place. While dashboards offer a high-level view, reports dive deeper into specific datasets and provide in-depth static analysis (Patel, 2023). They complement each other because dashboards give the "now" while reports tell the story of "how" and "why" (Knezovic, 2024). However, BI tools may understand these terms differently. Microsoft claims that a dashboard is a single page with a certain snippet of data, where essential information is highlighted, and a report further expands this information with details (Microsoft, 2023).

This thesis will explore the crucial steps beyond coding - strategic planning, user engagement, and iterative refinement that ensure a solution’s true impact. Delving into the intricacies of a project built on the Showpad platform, a real-world sales enablement tool, this work will highlight the importance of aligning technology with business strategy and pave the way for more efficient and effective digital transformations in the future.

2 Objectives

The topic of the diploma thesis "Dashboards Automation" covers a wide range of aspects. Therefore, it is crucial to specify the direction and goal that this thesis will follow throughout and to try to fulfill the goal with specific steps.

The key objective of this master thesis is to develop, design, and deploy a comprehensive dashboard, which will be published in the corporate workspace. The commercial operations team aims to create a prototype dashboard that highlights key aspects of their operations standing on the Showpad platform.

This dynamic hub will empower decision-makers by providing actionable insights. It will look for answers to topics and questions such as platform login frequency, and what marketing materials are popular with internal people and clients, and will focus on active users in the sales team, where further attention will be focused on training them in the coach module. Other topics and business questions related to the dashboard are discussed in more detail in the practical part of the work.

In addition, the work will also explore and map the complex life cycle processes and business intelligence methodologies and processes used, which will be further implemented in the second part of the work, which will also reveal potential challenges and limitations in the Showpad dashboard development process.

3 Methodology

This thesis employs a systematic methodology, which is a key part of this thesis, as it provides a structured approach and system contributing to the fulfillment of the objectives. Within the theoretical part, the methodology contributes to the collection and evaluation of literary sources to perfectly map the main topics for the practical development of the dashboard.

1. Collection, analysis, and synthesis of literature, where a search was carried out within online databases such as Google Scholar using selected keywords. Filters such as language restrictions, and publication date of resources were applied. In this way, relevant literature such as articles from various sources, books, and websites of technology giants were obtained.
2. This work was partially supported by AI-based tools such as language model chatgpt-4, Microsoft Copilot, or Grammarly for certain tasks such as providing information on specific topics and proofreading for grammatical errors. The accuracy of the AI's outputs was validated from multiple sources.
3. A method was mapped and created in the theoretical part in the form of a complex workflow of the BI lifecycle, marked as Figure 3, compiled from various sources, which will serve as a model for fulfilling the goal of the work in practical application.

The second part outlines the methodology employed in the hands- on development of the dashboard. To protect the confidentiality of information about the company and employees in which the project was conducted, the following methods are described in general and do not mention specific details. The entire project, i.e., the thesis, is managed by the agile method, resulting in faster, more adaptable solutions that meet dynamic user demands (PMI Team, 2022).

1. The first method is requirements collection, which was done by gathering data by filling out the business requirements document and subsequently creating the user stories. The BDR should include the project summary, project objectives and scopes, business & technology requirements, constraints, and key stakeholders. This guide governs the compiled document used to collect requirements (Schwartz, Business Requirements document, 2022).
2. Next, data extraction and storage will be covered, where Python scripts were developed to extract data from the Showpad RESTful (REpresentational State Transfer) API (Application programming interfaces). There is no need for multiple people to have access to the resources, so the file-based storage in JSON (JavaScript Object Notation) format method was used to save the extracted data (IBM Team, nedatováno).
3. Subsequent data transformation and manipulation that used ETL (Extract, Transform, Load) processes within Python, Microsoft Power Query Editor in MS power BI Desktop, and Tabular Editor tools (Y, 2024). A spreadsheet application for advanced data modeling and DAX (Data Analytics Expressions) without accessing any data (Tabular Editor, 2024).
4. An iterative design process to improve the dashboard elements. Using design principles like typography and color theory to improve the user experience.
5. Performed Unit, Integration, Data Accuracy, and optimization of the dashboard for the performance with the use of coding and best practices design.
6. Detailed dashboard deployment process, including server setup and security aspects. It outlines strategies for ongoing maintenance and updates of the dashboard system about newly released features in the market.

All methods and findings were carefully documented following the academic standards. Each method was selected for its relevance to the research questions and objectives, ensuring a consistent and systematic approach to the automated development and analysis of BI dashboards.

4 Theoretical Foundations

The business landscape has undergone a profound transformation over the past three decades, with rapid advancements in technology reshaping every angle of operations. Technology is at the heart of every successful business, whether it is through products, marketing, sales, design, or employee management (Lupenko, 2022). Let's consider a company in the era predating digital transformation. The business landscape was vastly different, devoid of the internet's expansive reach and the transformative power of data. As technology evolved, data metamorphosed into a reliable decision-making tool, and businesses began to flourish. Today, the IT landscape is rapidly expanding, propelled by artificial intelligence and machine learning, driving innovation across all sectors. Every enterprise, from the fast-food chains to the industrial gear manufacturers, is now deemed a tech company. (Dol, 2021) What was once considered impossible is now commonplace and integral to a company's operations.

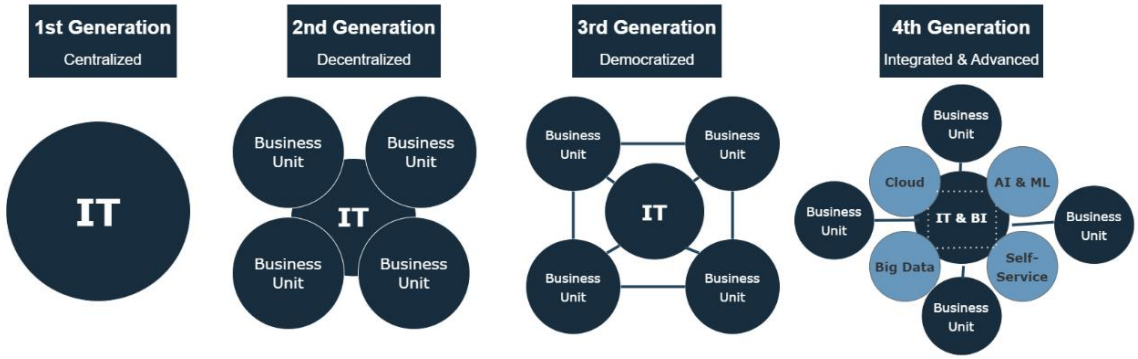
4.1 Business Intelligence

The term "business intelligence", hereinafter referred to as BI, is frequently used to describe a variety of technologies that offer quick, simple-to-understand access to insights about the company. BI combines various IT methodologies, techniques, tasks, and tools to achieve desired outcomes. The latest BI tools combine data literacy, AI (Artificial Intelligence), and ML (Machinery Learning) to bridge the gap between data, insights, and action, allowing the environment to generate itself given enough input from the user.

The beginnings of BI can be traced back to the decision support systems in the 1960s. With each generation, there is a step closer to realizing this promise. The diagram for each generation, which is inspired by the Qlik website, can be found below in Figure 1. The first generation of BI can be dated between the 50's to the 80's, when BI took shape with the development of computer technology. The second generation of BI was in the 80's to 90's (Qlik, 2022). The third generation was introduced in 2000 and continues to this day,

however, the fourth diagram has not been described yet. To get an overview of the fourth generation, ChatGPT AI technology was used. After the prompt *"What could 4th generation BI diagram look like if the first generation was just centralized IT in a circle, the second generation remains a circle with IT and decentralization adds more circles with business units that touch each other. The third generation is like the second, only the circles move away from each other and form a bond between them."* An idea generated led to the creation of the fourth diagram called "Integrated & Advanced".

Figure 1 – Evolution of Business Intelligence



Source: The Thesis author's processing according to (ChatGPT-4, 2023)

The current era focuses on cloud computing, advanced analytics, and AI technologies. BI is becoming a predictive and prescriptive tool for enterprises. Over the years, each generation has brought innovation and improvement to tools, enabling them to become a pillar of business decision- making.

The greatest value of BI lies in its ability to support data-driven decision- making. Present data informs about the current situation in the company, it is often live data or data that may be updated at night for the latest information the next day. It enables optimization, sometimes as well as simulation, and decision modeling to provide the best possible analysis of business decisions and actions. There are distinct kinds of BI that aim to bring different values to the user.

- BI report & dashboard – It is a conventional form of BI, which incorporates equipment for growing very bendy reports, in addition to the strong ones, and will be created as a part of the practical section of this work.
- Data visualization - Interactive BI dashboards with great data visualizations make the data more usable by revealing its shape, and trends and placing data in an ideal context, which will be specifically used and demonstrated in the dashboard created.
- Advanced analytics - It complements human intelligence, increases data literacy, and enables more users to derive value from their data. This is about incorporating AI and ML into BI tools.

BI aims to provide marketers with a clear view of the situation. BI tools access and analyze data sets and display the results in reports, summaries, dashboards, tables, graphs, and maps to provide users with detailed information about the state of their business (Kusnetzky, 2018).

4.1.1 Performance Indicators for BI Investments

To ensure the investment is sound and yields the desired returns, it is crucial to understand and monitor specific metrics and indicators. This chapter delves into the topics that are essential for evaluating the success of a BI project. This includes revenue, costs, productivity, customer retention, and more. The baselines and benchmarks for KPIs and track progress over time should be established. A few interesting metrics and topics were defined in a web article, which are included in this chapter such as ROI (Return on Investment) KPIs (Mukhtiar, 2023).

One of the first KPIs (Key Performance Indicators) is ROI. The most direct way to measure the success of a BI investment is to calculate ROI, which evaluates the usefulness and profitability of a project.

Data Quality Improvement: Various tools and techniques can be used to measure data quality, data cleaning, and data validation.

When data are inaccurate, incomplete, or inconsistent, reporting and analysis become less reliable.

User Adoption Rate: This metric tracks how effectively the employees and users are adopting and using her BI tools and reports. It is possible to check the logins or the kind of information they search most often. An effective BI tool can lead to increased productivity and employee job satisfaction by simplifying data analysis and reducing manual workload.

Operational Efficiency Gains: This involves an assessment of how the BI system has improved various operational aspects, such as reducing process times, improving customer service, or optimizing the supply chain.

Scalability and future readiness: The ability of a BI system to scale with the organization and adapt to future technological changes is a critical factor for long-term success.

By closely monitoring these metrics and indicators, businesses can ensure that their BI funding is not only justified for project initiation but also profitable over the long term, adapting to changing business desires and technological advancements.

4.1.2 Stakeholders in BI

This chapter lists the numerous positions and subject-matter specialists that are essential to business intelligence procedures. Each stakeholder in BI brings unique skills to the table, and their collaborative efforts lead to an effective utilization of BI in driving organizational success.

Executive leadership (CEO, CIO, and CFO) is a key in shaping the strategic path of an agency using BI. They want robust access to strategic thinking, market trends, and the ability to combine BI insights with high-level decision- making.

Department Heads and Managers throughout various departments like sales, marketing, and HR utilize the BI to make department-precise decisions. Their number one purpose is to enhance and support departmental overall

performance and align their techniques and processes with the organization`s ordinary objectives.

IT Professionals and Support Teams represent the spine of BI infrastructure. They want technical information in BI tools, facts management, and problem-fixing competencies to make certain clean operations and cope with technical challenges. Their major contribution is retaining a sturdy and green BI gadget that is handy and purposeful for all users.

BI consultants play a vital role in BI and need to have deep knowledge in this area. The BI experts are to prepare powerful implementation and enhancement of BI systems, tailored to the exact requirements of the organization. BI Developers are tasked with designing and constructing BI solutions. They must possess strong technical talents in programming, database management, and excellent expertise in commercial enterprise processes. Their purpose is to broaden user-friendly, green BI equipment and dashboards that meet the precise necessities of the organization.

Data engineers focus on building and managing the statistical infrastructure necessary for powerful BI. They want skills in database design, statistical modeling, and ETL processes. Data Scientists and Analysts examine and interpret complicated datasets to offer actionable insights. They need to have information in statistical analysis, statistics visualization, and vital thinking. Their number one position is to find tendencies and styles in statistics that could reveal commercial enterprise choices and strategies.

End users are the everyday customers of BI tools. They require knowledge of how to use BI tools to generate actionable insights. Expected results are insights into day-to-day operational decisions that contribute to organizational performance and efficiency.

Data governance and compliance officers ensure that BI methods adhere to crime and moral standards. They want radical expertise in information management, compliance, and change management.

BI involves a diverse range of stakeholders, each of whom brings essential skills that contribute significantly to the successful use of BI. Collaboration between these stakeholders is critical to the BI process, ensuring that data is not only efficiently collected and analyzed but also translated into actionable insights that drive strategic decisions.

4.1.3 Trends and Importance of BI in Today's World

This chapter highlights the unique benefits of BI, along with how it is important to businesses today and where is expected to go in the future. Most businesses looking for success and competitiveness in today's dynamic business environment are discovering that BI is a key part of their overall strategy. The distinctive profits gained by implementing BI in a company are connectedness, collaboration, and awareness for everyone. Furthermore, it is the speed that causes competitive advantages in market response. It brings trusted data, accuracy, and scalability. A big bonus and also a result of BI are multi-platform and multi-user workspaces available to everyone. The concept should be that data arrives before you even know you need it, so you can ask questions you have never thought of and get information you did not know you needed. Complex IT solutions and outdated reporting tools are replaced by amazing tools for automated data storytelling. Every year the direction of BI changes and the platforms and companies follow the latest trends.

In an article on the big data framework website, the following trends were selected for last year (2023): data-driven culture, cyber security, data storytelling, sustainability, optimization, advanced analytics, NLP (Natural Language Processing) and XAI (Explainable Artificial Intelligence) (Big Data Framework, 2024). Some of those trends are also predicted for 2024 and will thus continue to develop are: data security, embedded Analytics, NLP, and XAI. The newly added trends predicted for 2024 are data literacy, Self- service analytics, and Data Lakehouse (Calzon, 2023). The future of business intelligence promises intelligent, unbiased, fair, transparent, and responsible AI and ML that can be trusted to uncover the value of data without causing harm.

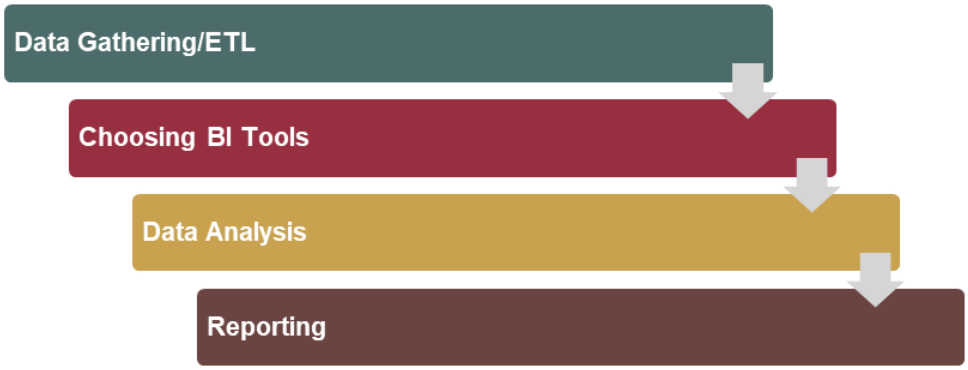
4.2 Essentials of BI: Techniques, Processes and Methods

This chapter covers a comprehensive adventure through the multifaceted world of BI. This survey covers a wide spectrum, from the basic aspects of requirement gathering for development to data management and storing to mapping dashboards of development, deployment, and also the adoption of fresh solutions. The goal is to provide a holistic view of BI, integrating both the theoretical frameworks and practical applications that form the core of BI. The techniques are carefully broken down, with each chapter methodically following the sequence of basic BI tasks.

4.2.1 BI Lifecycle Overview

This chapter delves into the intricate process of BI a technology-driven process that encompasses a wide range of activities and procedures from collecting, storing, analyzing, and providing access to data. The simple Business Intelligence Lifecycle, as depicted in Figure 2, is inspired by the "sctiger" website, and the lifecycle is split into 4 phases:

Figure 2 - Business Intelligence Lifecycle

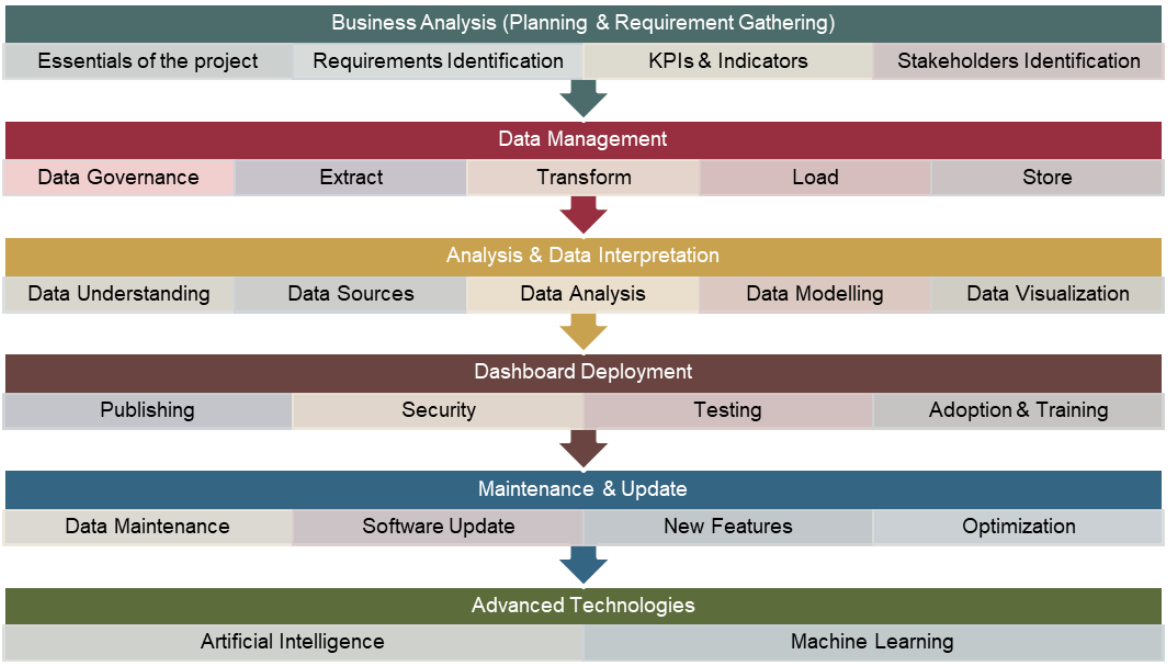


Source: The Thesis author's processing according to (Baktawar, 2022)

Data Gathering/ETL is the initial phase, where the necessary data for the specific use case is identified. Choosing BI tools is the next step, where the specific tools or programming languages to perform operations on large data sets are chosen and utilized. Next is the data analysis phase which involves a detailed examination and interpretation of the gathered data to extract valuable

insights. Reporting is the final phase of the Business Intelligence Lifecycle (Baktawar, 2022). Considering the current progress in BI, which emerges from the 'Trends and Importance of BI in Today's World' chapter and other resources listed in Figure 3, a holistic approach was created that covers the entire end-to-end BI lifecycle workflow, including post-deployment activities. Larger organizations or those that need more complex data ecosystems will benefit from such a detailed cycle. A risk management perspective can also be an advantage, where security and data governance phases are added, big topics of the last few years. The basic difference between the BI Lifecycle Workflow image and the Comprehensive one is the greater detail in phases, which can also be perceived as a checklist of whether the phases are fulfilled or at least considered during the development phases.

Figure 3 - Comprehensive BI Lifecycle Workflow



Source: The Thesis author's processing according to (EffectiveSoft Team, 2023) (Taylor S. , nedatováno) (Kapoor, 2023)

The sequence of phases in the "Comprehensive BI Lifecycle Workflow" figure reflects a logical sequence of steps. While this order provides a structured approach, it is important to note that the process is not rigid. Depending on the specific needs and circumstances of a project, these phases can be rearranged, repeated,

or even skipped. In the following sections, there will be a deeper dive into each of these disciplines, providing a more detailed explanation of their roles and importance The dashboard creation step pattern will be used in this practical part of the thesis.

4.2.2 Business Analysis (Planning and Requirement Gathering)

Central to achieving effective BI solutions is the role of business analysis, which is understanding business needs, capturing, documenting, and managing requirements, a critical but often overlooked aspect of the IT world (REHAN AKBAR, 2008). The requirements are the pillars that shape the entire development, product, stakeholders, and project direction from the beginning. The information basis that must be obtained and mapped is the project overview, project owner, project goals, project scope, project timeline, Key Performance Indicators (KPIs), stakeholder identification, resource allocation.

Here are some widely recognized methods for collecting the requirements. In the practical part, the Business Requirement Document (BRD) will be tested and used as a primary tool for data gathering. The BRD is a comprehensive document that outlines the business solutions necessary for the project to deliver value (Schwartz, Business Requirements Document, 2022).

Furthermore, there are methods like the user stories. User Stories are simple, short descriptions from the customer's perspective, describing a customer's actions as they use or interact with the product (Blake, 2021).

The MoSCoW method prioritizes tasks or features into four categories: Must have, should have, could have, and Will not have (Croft, 2023). This method helps with reaching an ordinary understanding with stakeholders on the importance they place on the delivery of each requirement.

Interviews and questionnaires seem like the basic and traditional methods; however, they need to be used properly. They provide fact-based information and analytical insights from respondents (Keydifferences, 2023).

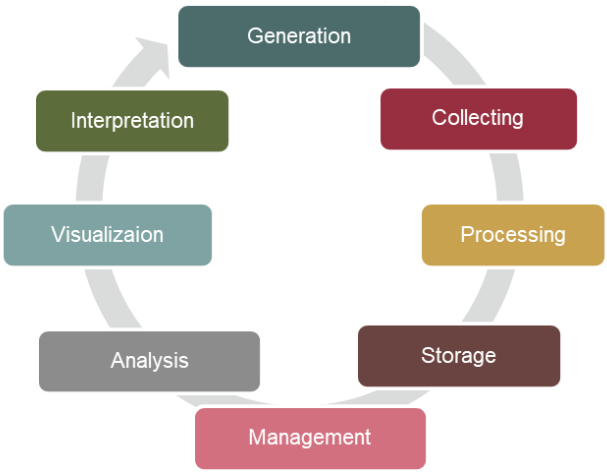
Questionnaires can be more beneficial if you need to ask stakeholders the same question across the board.

4.2.3 Data Management Cycle

BI systems are the cornerstone of decision-making in modern organizations. They provide the tools and methodologies to transform vast amounts of raw data into meaningful, actionable insights. At the heart of every BI system, is the Data Management Cycle. This cycle serves as the backbone of BI, underpinning the overarching goal of transforming raw data into strategic assets.

The data management cycle is a continuous process that ensures data quality, availability, and security. This cycle is critical because it ensures that the data used in decision-making processes is accurate, reliable, and timely. It works with other key components of the BI system, such as data governance policies and data analysis tools. Together, these elements create a robust and efficient system that can handle the complex data needs of today's organizations as seen in the Figure below.

Figure 4 - Data Management Lifecycle



Source: The Thesis author's processing according to (McGill, 2023)

The ETL process plays a pivotal role in data integration (Chaudhari, 2023). It involves three crucial steps:

- Extract: Data is gathered from multiple sources.

- Transform: Data is normalized, cleansed, deduplicated, and transformed into a consistent format.
- Load: Data is loaded into a target repository like a data warehouse.

For example, to illustrate the application of data collection and integration in a marketing context, consider a fictional case study involving an analysis of the effectiveness of a retail company's marketing materials. The company looks for an understanding of how varied materials contribute to sales and brand engagement across different channels. The ETL process in this scenario might include:

- Data extraction from diverse advertising channels – social media analytics, e- mail marketing campaign statistics, internet site visitors, and responses to in-keep promotional substances.
- Transforming these facts to reconcile diverse metrics – together with engagement prices, click-thru prices, and conversion metrics – right into This step includes resolving inconsistencies inside the fact's shape and resolving inconsistencies.
- Loading harmonized facts into the crucial facts warehouse. Here, the incorporated facts set offers a complete view of the overall performance of the diverse advertising substances throughout all channels.

The Data Management Cycle and the ETL process are vital to Business Intelligence systems. They transform raw data into strategic assets for informed decision-making. The ETL process, crucial for data integration, improves data quality and usability. As demonstrated in the case study, these processes can provide comprehensive insights into operations, helping identify trends and patterns. Furthermore, this chapter will deal with some particular processes from the data life cycle.

4.2.3.1 Data Extraction & Storage

This chapter will explore the foundational processes of data extraction, which is the initial stage in the BI lifecycle. Data extraction in BI is a process

of gathering and measuring data from a variety of sources to get a complete and accurate picture of the area of interest. RESTful API utilization for accessing data will be discussed there, detailing the methodologies for accessing this data. The connection between the databases and data warehouses and their importance in the context of API data extraction will be explained. Additionally, the role of Python as a versatile tool for preliminary data manipulation will be highlighted. This process will be practically applied in the following practical part, highlighting the extraction and subsequent transformation of data.

Database and Data Warehouses as Foundations for Data Extraction

In the architecture of BI database management and data warehousing are essential components. They serve distinct roles and operate under different paradigms, yet they complement each other to manage and analyze data effectively. A robust database or data warehouse serves as the foundation for effective data extraction because it is the source of the data people work with.

Data extraction via APIs fundamentally relies on the underlying data storage systems, i.e., databases and data warehouses. The integrity, structure, and quality of these storage systems directly affect the ease and effectiveness of the data extraction. Databases provide real-time, operational data, while data warehouses offer historical insights. The choice of source depends on the API's end goal— whether it is for immediate, transactional purposes or analytical insights. When a developer uses an API to extract data, they rely on the structure and organization of databases and data warehouses. These systems need to be designed for optimal accessibility and querying, which includes considerations for API endpoints.

Python comes into play as a powerful API interface tool thanks to its rich ecosystem of libraries for data manipulation and extraction, such as API call requests and Pandas for data manipulation. Python scripts can automate the process of extracting data through APIs, handle data transformation, and ensure that data is ready for use by various BI tools.

Web Services (APIs) as a Gateway for Data Acquisition

Application programming interfaces (APIs) are the key to connecting to databases. APIs allow you to connect apps to cloud platforms via Backend- as- a- Service, hide database complexity, organize code, design reusable components, and more. APIs access data from various sources, such as databases (Showpad), and data warehouses are used for a variety of purposes. A well-designed API provides a robust and flexible interface, which is crucial for real-time data retrieval and analysis. It ensures that data is stored in a structured and consistent manner, making it easier to retrieve, update, and analyze the data. These can range from inefficient data retrieval and analysis, and increased risk of data breaches, to difficulties in scaling up the operations as the volume of data grows (Postman Team, 2022). In the practical part, data will be extracted from the Showpad database, which as a RESTful API. This will be theoretically analyzed for a better idea of the data, and thus preparation for the subsequent implementation will take place.

API Documentation and Selection: Understanding the importance of comprehensive documentation and selecting APIs that align with the data needs and security requirements of the BI process. Showpad has detailed documentation on how to connect to the database on their website, where it is mentioned that the RESTful client-server API a software architecture should help access server-side data by representing it in a simple format (JSON or XML) (Showpad Team, 2023). RESTful API has become one of the most widely used approaches for building web APIs, making it an extremely capable tool on the Internet.

Endpoint Configuration: Configuring API endpoints meticulously ensures accurate data retrieval, covering aspects like the base URL, paths, and parameters. Underlying everything is the base endpoint, which for Showpad is: 'https://{{subdomain}}.api.showpad.com'. It is important to test API endpoints before integrating them with BI tools. Tools like Postman provide easy-to-use interfaces for making requests to API endpoints and checking responses. This process helps to validate the data structures, response formats, and error handling. Issues like «404 Not Found» and «401 Unauthorized» often occur.

Such errors often indicate incorrect endpoint configuration, such as a typo in the URL or a missing authentication header.

Security Measures: Implementing robust security measures such as OAuth for authentication, managing API keys carefully, and ensuring data privacy during the transfer. In addition to ensuring access to data, it is also important to protect it from unauthorized use or disclosure. The Showpad API uses OAuth2 as an authentication and authorization method and authorizes a user to use a provided access token. Personal API tokens for working with data can be generated in the API menu in the admin panel of the online platform. An API Token named 'ShowpadDataToken' was created for this purpose. Once you get a token, it should be then used as a header for any future API call. In code 1 below, one can see what the specific header for the Showpad API with the token looks like in the code below.

Code 1 - API Request Headers Configuration

```
url = "https://{{subdomain}}.api.showpad.com"
headers = {
    "Accept": "application/json",
    "Authorization": "Bearer ‘ShowpadDataToken’"
}
```

Source: The Thesis author's processing

Testing Phase: Before starting to extract data blindly, it is necessary to verify the functionality of the token and access, to further look at the structure of the data in the database and what calls will be created to obtain it. API connection testing is critical for guaranteeing the interactions with the Showpad database. This part describes the process of testing Showpad API authentication and authorization using the Showpad client server for testing. The first testing will be for managing events, where the GET call will be tried "https://subdomain.showpad.biz/api/v3/exports/events.json". Below in Figure 5, the token worked with a successful response can be found.

Figure 5 - Showpad API Environment: API Testing Response Example

Source: The Thesis author's processing

Writing Effective API Queries: What remains consistent across all systems is that the API is a means of how clients call providers. Operations are performed using various HTTP methods such as GET, POST, PUT, DELETE, OPTIONS, and PATCH. The downside to REST APIs is that their extensive metadata creates large payloads that can sometimes cause more trouble than they are worth. For optimized call and data extraction, it is best practice to always consider what data is available, what data is required, and whether all or only part of the data needs to be extracted.

Data Extraction with Python

The hands-on part will apply these principles specifically and use a RESTful API to extract data that is then supported with Python scripting. Examining the documentation for extracting data from Showpad also gathered some interesting information that must be considered. For a better illustration of the practical part of the work, below is the code used for the extraction of one part of the data called assets. The Python script below is designed to fetch and process data from a specific URL. It begins by initializing an empty list of all_assetsItems to store the fetched items. The script then enters a loop where it sends a GET request to the URL with the specified headers and parameters. If the response status code is 200, indicating a successful request, the script proceeds to extract the 'items' from the 'response' in the returned data. These items are then appended to the all assetsItems list. If no items are returned in the response, the loop breaks, indicating

that there are no more items to fetch. The script also updates the 'offset' parameter by the number of items fetched in the current batch. This allows the next request to skip the items that have already been fetched. If the response status code is not 200, indicating an unsuccessful request, the script prints 'Failed to fetch data' and breaks the loop.

Code 2 – Data Extraction

```
all_assetsItems = []
while True:
    response = requests.get(url, headers=headers, params=params)
    if response.status_code == 200:
        data = response.json()

        items_data = data['response']['items']
        if not items_data:
            break
        all_assetsItems.extend(items_data)
        params['offset'] += len(items_data)
    else:
        print ('Failed to fetch data')
        break
```

Source: The Thesis author's processing

In the same way, other codes will be prepared for extracting data from the Showpad database.

4.2.3.2 Data Transformation & Load

In the realm of data processing, the extraction phase signifies the initial step of transformation. After the extraction phase, further data adjustments can be made in some cases referred to as the post-extraction phase, which can include, for example: the selection of extracted data only necessary for the needs of the projects, which already changes the raw dataset. This is the transition to the transformation phase, where the goal is to convert raw, unprocessed datasets into a structured format that is ready for BI applications.

The basic transformations that occur during this phase must include:

- Data Cleaning: This involves identifying, correcting, and removing errors and inconsistencies in the data, and looking for duplicates or missing values

to improve data quality and accuracy. This includes reducing the volume of data by removing redundant or irrelevant data, making it more manageable and easier to understand.

- Data Formatting: This means converting data from one format or structure into another standardized format. For example, this could involve changing data types such as converting categorical data into numerical data or splitting columns based on specific delimiters.
- Data Normalization: This entails adjusting the values in the dataset to a common scale, without distorting the differences in the ranges of values or losing information (Ninja, 2023). Normalization is particularly important when dealing with parameters of different units and scales, as it allows for fair comparison between atypical features. It is a critical step in many machine learning algorithms as it can significantly improve their performance.
- Data Compression: This includes reducing the size of the data without losing its ability to convey information. Data compression techniques can be lossless (where the original data can be perfectly reconstructed) or lossy (where some information from the original data is lost after compression). Data compression is crucial in managing large datasets as it helps save storage space and allows faster data processing and transmission.
- Data Integration: This involves combining data from diverse sources into a unified view to provide a more comprehensive understanding of the data.
- Derived Metrics and Measures: New data columns or measures may be derived or calculated from existing data to provide more depth to the analysis.
- Data Aggregation/Grouping: This means summarizing or grouping data to create a condensed view. For example, rolling up sales data by region or direct manager.

Transformation with Python

Python's strength lies in its simplicity and readability, which makes it an excellent choice for data manipulation tasks. Its extensive range of libraries, such as Pandas and NumPy, provide robust tools that simplify the process of working with data. An explanation of how Pandas is used for its robust data manipulation capabilities to clean and restructure the extracted data will be demonstrated. In general, the basic requirements include installing Pandas and importing into the code after you can start with magic., then anyone can do magic tricks with Python.

The data structure in Showpad is divided into two main parts: Content reporting and Coach reporting. First, there will be a focus on the content part, from the Showpad documentation is obvious there is only one fact table called Events where all user's activities are. The remaining tables are dimension tables, i.e., describing facts. For more information on dimensions, see the "Data Modeling" chapter in this thesis. It will not be necessary to download all the tables for the goals of a project developed in the practical part (Showpad Team, undated). Important to mention is a big bonus of the Showpad content part because there is a relationship diagram for the content, which makes it easier to manipulate the data, on the other hand, what is missing is a parameter that would call only the data that is needed. Another big part is the Coach data module, which also needs to be extracted from Showpad. The very first encounter with the data and the proposal of how Python codes for extraction and transformation could look is difficult because, unlike the content part, the data here is not so well prepared for further manipulation and loading to BI tools. On the other hand, there is a setup of a parameter called 'field', which allows extracting only the data needed.

A few basic functions that will be used in the practical part are loading a JSON file and saving the JSON file as a new modified document. Further, splitting data into metadata and list items, working with columns, or creating loops to load additional data (scroll paging) will be used. Below is a continuation of the code from the last data extraction chapter where the code moves to simple modifications and manipulation of the data. In the next part of the script,

a list of keys that are to be removed from each item in all_assetsItems is defined. A nested loop then goes through each item in all_assetsItems and removes the keys specified in columns_to_remove. The pop method removes the key from the dictionary and avoids a KeyError by returning None if the key is not found. The transformed all_assetsItems is then written to a file in JSON format. The indent=4 argument formats the JSON data with an indentation of 4 spaces for readability. These lines print a message indicating that the items have been fetched and transformed, followed by the total number of items, which is stored in assetsItemsFetched. In summary, this script performs extraction and transformation operations by removing unnecessary keys from the items in all_assetsItems, thereby refining the data for further use.

Code 3 – Data Transformation

```
columns_to_remove =
['source','description','isAnnotatable','expiresAt','releasedAt','archive
dAt,''commentsCount']

for an item in all_assetsItems:
    for key in columns_to_remove:
        item.pop(key, None)

with open('assets_tr.json', 'w') as file:
    json.dump(all_assetsItems, file, indent=4)

print('Assets Items fetched and transformed. Total item:')
print(assetsItemsFetched)
```

Source: The Thesis author's processing

File-Based Data Storage

In the diverse landscape of BI environments, file-based data import remains a critical component, even in the era of data. The ability to integrate static datasets into workflow analysis is a powerful tool, providing flexibility and depth to data- driven insights. Data is often stored in file formats such as CSV (Comma- Separated Values), Excel, and JSON (JavaScript Object Notation). These formats are widely used due to their simplicity, versatility, and compatibility with various BI tools. Each format

has its unique features and use cases, making them suitable for several types of data and analysis requirements.

CSV documents represent tabular data as plain text divided by commas into columns. This format is widespread due to its easy usage and excellent support across many platforms. Furthermore, CSV files offer faster data transfer compared to other file formats, such as Excel, making them an optimal solution for large datasets. Excel files, on the other hand, are known for their advanced features like formulas, charts, graphs, pivot tables, and macros. They provide support for complex calculations, data analysis tools, and conditional formatting. However, the size of Excel files is usually larger than CSV documents, and they require specific software like Microsoft Excel to open. Therefore, they may not be ideal for large volumes of data.

In the practical part, there will be a focus on JSON files. JSON is a lightweight data format that is easy for humans to read and write and simple for machines to parse and generate (Medium - pandaquests, 2022). JSON represents data as a collection of name-value pairs, similar to objects in JavaScript, or it can represent data in arrays. It is an ideal solution for transmitting data between a server and a web application, making it popular in web development and for APIs. In the example below, the metadata object contains information about the assets table from Showpad. The items array contains a list of items from Showpad, each represented as an object with properties for assetId and other details.

Code 4 - JSON Format Example

```
{
    "meta": {
        "code": 200,
        "message": "OK",
        "serverTime": "2024-03-06T14:00:09+00:00",
        "userTimezone": {
            "offset": "+00:00",
            "name": "Europe/Dublin"
        }
    },
    "response": {
```

```
        "count": 6292,
        "items": [
            {
                "assetId": "19b28c3089e1d4e477bc38393cdd073d",
                "divisionId": "26092c300714ba38ecc245b0da65daef",
                "status": "active",
                "type": "document",
                "uploadedAt": "2020-07-22T08:40:47.000Z",
            }
      ]
}
```


Source: The Thesis author's processing

This is a common structure for JSON files, as it allows for complex, hierarchical data to be represented in a readable format. In conclusion, understanding file-based data storage and the unique features of formats like CSV, Excel, and JSON is crucial in the realm of data analysis and BI.

Transformation with Power Query in MS power BI

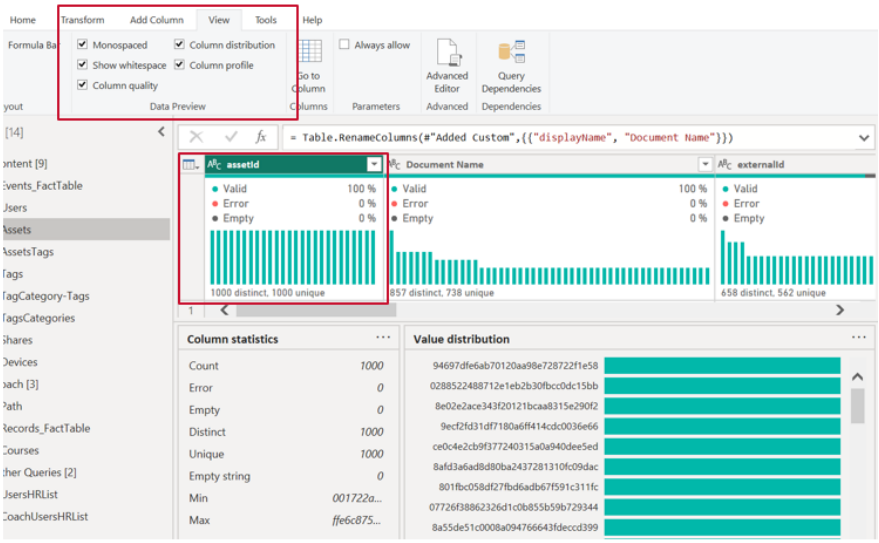
Python's extensive libraries and capabilities make it a great tool for data manipulation and analysis, and its integration with power BI and Power Query allows for seamless data processing, data modeling, final transformation, and visualization. Power Query is a data technology that enables you to explore, connect, combine, and modify data across a wide variety of sources (EDITORIAL TEAM, 2024). It provides an intuitive and interactive interface for data transformation, making it accessible even to non-technical users. Power Query can be used for the key features and operations:

- Data Import & Extraction: Power Query can connect to various data sources, from databases, Excel files, and JSON files to online services and APIs.
- Data Manipulation: Expanding nested data fields such as converting JSON hierarchies into a flat table structure.
- Data Cleaning: Power Query offers tools to handle missing or erroneous data. You can replace values, fill gaps, and even use AI-powered features to detect and correct errors.

- Data Transformation: This includes changing data types, merging columns, splitting text, and creating calculated columns.
- Data Combination: You can append queries to stack data from multiple sources, or merge queries to join data based on common attributes.

A function that is suitable for use in a Data Preview mode, is in the power query, where one can see the quality of the columns at a glance. Here are the key details anyone can get from this preview. For assetId section shows 100% valid data with no errors or empty fields. It indicates 1000 distinct and 1000 unique entries. The detailed column statistics include count and error values along with others.

Figure 6 - Power Query: Data Preview assetID column



Source: The Thesis author's processing

Furthermore, in this phase, other transformations are performed including removing columns, changing data types, filtering rows, etc. In Power Query: The transformations you apply in Power Query are executed each time the data is refreshed. Power Query generates a series of steps (in M language) that are applied in sequence to clean or transform data.

Data Modeling

After loading and transforming data, it can proceed to the creation of a data model in MS power BI, which will provide another view of the data and reveal a lot of secrets that will require further steps. This involves setting up relationships

between tables which allows the creation of calculated columns, and measures, and potentially modifying the data further using DAX.

Stedman defined data modeling as a process of creating a simplified diagram, which contains tools and data elements (Stedman, 2021). It can be said that it is also the preparation of a visual representation using diagrammatic techniques to represent either the entire information system or its parts for communicating connections between data points and structures. The goal of data modeling in BI is to have an efficient organization of data.

By understanding different schema types, data analysts can choose the most suitable schema design for their specific use cases. Normalization refers to the process of organizing and structuring data in a way that eliminates redundant data and improves data integrity. A flat schema is a simple database design where all data is stored in a single table. A Star schema is a type of schema used in data warehousing and dimensional modeling (Geeks For Geeks, 2023). In this schema, a central fact table is connected to one or more-dimension tables based on the common field or column in both fact and dimension tables. The Star schema offers several advantages over the flat schema. A Snowflake schema is an extension of the Star schema (Taylor D. , 2024). In the Adventure Works schema, each dimension table is connected to one or more related tables, forming a hierarchical structure that resembles a Snowflake.

Cardinality refers to the relationships between entities in data models, specifically how one entity relates to the number of items in another. Understanding the cardinalities is essential for designing efficient and exact models. The main types are:

- One-to-One (1:1): Each entity in a relationship can be associated with only one other entity, such as a person who must have his birth certificate.
- One-to-Many (1:N): One entity can be associated with multiple entities in another set, such as an order with more items there.
- Many-to-One (N:1): Many entities can be associated with one entity in another set, such as more people can be born in the Czech Republic.

- Many-to-Many (N: N): Entities in both sets can have multiple associations with each other, such as students following various subjects (Wikipedia, 2023).

A correct definition of these relationships ensures that data models accurately reflect the actual interactions between business entities, which is essential for effective querying and analysis. The most common mistakes when developing a data model can be ignoring business requirements or not aligning the data model with specific business needs. This can lead to irrelevant or missing data in analyses. Also, too much normalization, because while normalization reduces redundancy, an overuse of this can complicate queries and reduce performance. The opposite i.e., insufficient normalization, can lead to redundancy and data inconsistency, which can be problematic for large BI systems.

Thoughtfully incorporating cardinalities and avoiding common data modeling mistakes are key steps in building a robust BI platform. These considerations ensure that the data model serves as a reliable basis for extracting meaningful insights, thereby supporting informed decision-making. As BI systems evolve to handle more complex data and analytical needs, a well- designed data model becomes even more important to effectively navigate the data environment. In the practical part, the snowflake schema will be introduced more and used, because it fulfills the expectations and speed of the model for these purposes. At the same time, it goes hand in hand with the structure of the data extracted from the Showpad database.

Transformation with Tabular Editor

The Tabular Editor is an external tool that can be connected to the MS power BI model. It offers advanced options for data model management and optimization that are not available directly in power BI Desktop. For example, you can create, edit, or delete measures, calculated columns, and tables. The Tabular Editor provides a more robust environment for writing and managing DAX expressions and can help to organize models more efficiently.

The Tabular Editor has a feature called "Best Practice Analyzer" that can audit models for common performance issues and recommend optimizations. This can significantly improve the performance of power BI reports. Using the Tabular Editor, you can perform bulk operations that are cumbersome in power BI Desktop, such as renaming multiple objects at once, changing properties on multiple objects, etc. The Tabular Editor can be used to version control (using Git or other versioning systems) and to deploy changes to power BI models hosted in power BI or Analysis Services. Once changes are in the Table Editor, one can save the changes back to the power BI file. It is important to understand that some changes made in the Table Editor may not be reflected in the power BI Desktop interface but will be present in the model and affect the functionality and performance of the report. The Tabular Editor is recommended for more complex models or when one needs to perform tasks that go beyond the capabilities of the power BI Desktop interface. Note that while the Tabular Editor is a powerful tool, it also requires a good understanding of data modeling and DAX to be used effectively.

4.2.4 Data Intelligence

At the heart of informed decision-making lies the pivotal process of analysis and interpretation—a dual force that unlocks the stories data tells us about the world. Analysis is the act of meticulously dissecting data to extract patterns, while interpretation involves piecing together these patterns into a coherent narrative. Together, they form the bedrock upon which businesses, scientists, and policymakers build strategies and draw conclusions.

4.2.4.1 Data Interpretation

The principle of a well-designed dashboard is to first understand the data. What data types are involved to know how to work with them further? Are there connections between the data, and relationships, are they dependent on each other The entire chapter is inspired by an article by Sarah El Shatby, who explained the data according to a few groups, which are further described below (Shatby, 2023).

Data, categorized by value, can be either qualitative or quantitative, each type having distinct characteristics and applications.

In Python programming, understanding data types, which include primitive types like integers, floats, characters, and Booleans, non-primitive types like arrays, strings, and objects, and specialized types like dates, times, and Enums, is crucial for effective data manipulation and analysis (Severance, 2009) (Sturtz, nedatováno).

Data analysis involves understanding trends over time, distributions, and dependencies between variables, using tools like line charts, histograms (frequency distribution), and scatter plots (correlation analysis). Statistical models can be used for more detailed analysis. One can build models to understand relationships such as regression models.

Data sensitivity involves understanding whether data is public or private, who has access to it, and its potential harm to individuals or institutions if shared; it ranges from low-sensitivity data that can be shared publicly, to medium-sensitive data for internal use like emails, to protected data like health records, all of which are crucial in data-driven decision-making and dashboard optimization.

In the realm of BI, the adage „A picture is worth a thousand words" resonates profoundly and it is necessary to select effective visualization. They reveal trends, patterns, and correlations that might be missed in text-based or numerical data. The process of selecting the right visualization involves identifying the type of data (quantitative or qualitative, discrete, or continuous) and defining the goal (comparing data, showing trends, highlighting distributions).

- bar graphs are ideal for categorical comparisons,
- line graphs for time-based trends,
- histograms for frequency distributions,
- scatter plots for relationship trends,
- box plots are useful for visualizing dispersion and outliers,

- while dot graphs and bubble graphs are used for correlation and multivariate analysis, respectively.

However, it is always necessary to consider the audience and the context and adapt the complexity of the charts to the understanding of the audience. It will be different for people in management, people in analysis, and, for example, the salespeople.

4.2.4.2 Data Mining & Advanced Technologies

This comprehensive chapter explores the synergistic relationship between Data Mining, AI, and ML in the context of BI.

Data mining represents the confluence of statistics and computer science, designed to uncover patterns and extract actionable intelligence from vast data collections. From the digital footprints in cyberspace to the structured records of corporate databases, data mining transcends industries and scales. Within a company, this means searching company data to find the necessary information, and extracting meaningful patterns, and relationships. Insights into customer demographics, recurring patterns in data, or various anomalies or operational errors are often looked for. Within areas such as retail, finance, or even healthcare, data mining is a particularly suitable technique because they are dependent on transactional data (Misiuro, 2022). Retail applications, such as data mining, transform purchase data into a road map of consumer behavior. Healthcare uses it to link symptoms and diagnostics into patterns, improving patient care through predictive analytics. An example of a data mining system is the IBM SPSS Modeler, which is primarily designed for data mining, text analysis, and predictive modeling.

Techniques such as classification, which categorizes data into predefined classes, and clustering, which groups data based on similarity without predefined classes. Then there are techniques like association rule learning to identify relationships between variables in large databases, such as market basket analysis to detect products that are often bought together. Furthermore, it is good to mention

regression analysis, which is used to identify relationships between variables and predicted numerical values. Regression analysis is key to financial forecasting and trend analysis or operational efficiencies. The last technique used is a decision tree, which is a model that uses a tree graph of decisions and their consequences and is also one of the machine learning techniques. It is widely used in risk analysis and operational strategies.

Advancements in AI and ML have significantly bolstered the capabilities of data mining, evolving BI from a historically reactive discipline to a proactive, predictive, and prescriptive analytics powerhouse. AI-driven automation and ML pattern recognition fortify the data mining process, enhancing the discovery of correlations and trends with unparalleled speed and accuracy.

Integrating AI and ML into BI tools empowers automated data processing and the optimization of complex queries, such as those performed in Microsoft Power BI through Quick Measures. AI and ML enhance the analytical processes, transforming BI into a dynamic field that not only interprets data but also anticipates future outcomes and informs business strategies.

AI, despite its potential, presents challenges in BI that include statistical privacy and ethical use. The size of AI-generated insights depends on the integrity of the underlying statistics and highlights the importance of independent and great datasets. Artificial Intelligence and ML have transformed BI into tools for taking initiatives for strategic decision-making. As these technologies continue to evolve, they will further personalize and enhance business strategies, offering a competitive advantage to those who adopt them.

By bringing together the topics of data mining, AI, and ML, this chapter aims to provide a holistic understanding of the modern BI environment. This convergence of these technologies is instrumental in harnessing the full potential of data, offering a competitive advantage to organizations that deftly navigate this integrated terrain.

4.2.4.3 Data Analysis

Data analysis is sometimes done in conjunction with data mining, allowing one to dig deeper into data to discover new patterns and insights.

However, data analytics refers more to the process of analyzing collected data to make informed decision making. There is another system from IBM called SPSS Statistics, which excels at making sense of complex patterns and associations, allowing one to draw conclusions and make predictions in-house or with open- source integration.

It is based on complex mathematical and statistical calculations and operations. Analytics can be divided into a broad spectrum:

- Descriptive analytics focuses on summarizing historical data to show trends and patterns. (Kanade, 2023) It forms the foundation of most dashboard visualizations, and it will form the foundation of the Showpad dashboard as well.
- Predictive analytics uses statistical models and machine learning techniques to predict future outcomes based on historical data. It is crucial for risk assessment and planning.
- Prescriptive Analytics goes beyond prediction to recommend actions. By analyzing past performance and current conditions, it suggests decision options to achieve desired outcomes.
- Diagnostic analytics focuses on understanding the reasons behind past performance (Intellipaat, 2023). It involves drill-down, data discovery, and correlations to show causative factors.

4.2.4.4 Statistical Inference

Beneath the data analysis surface there is a statistical inference, the statistical process that allows us to infer properties of an underlying distribution based on a sample. It is a powerful method that stretches the potential of data and enables to make predictions, test hypotheses, and estimate probabilities. Whether it is determining the efficacy of a new drug or forecasting market trends, statistical inference provides the toolkit for turning data into insights.

The journey of statistical inference begins with the formulation of a hypothesis and follows through to the rigorous application of tests

such as T- tests, ANOVA, and regression analysis. These tests serve as the lens through which it is possible to view the significance and reliability of findings, transforming raw numbers into meaningful conclusions. In the forthcoming sections, there will be deep dive into the realm of analysis and interpretation, exploring how statistical methods guide us from conjecture to certainty. The intricacies of these techniques and the principles that ensure their correct application, providing a comprehensive understanding that transcends mere numbers, towards actionable knowledge will be uncovered.

Data mining and analytics methods are indispensable in extracting value from data, offering a compass for strategic decision-making. By understanding and applying these techniques, organizations can navigate the complexities of their data landscapes, uncover hidden opportunities, and mitigate risks. The integration of these methods into dashboards democratizes data insights, empowering users across all levels of an organization to make informed decisions.

4.2.5 Dashboard Design

In today's data-driven world, the development of data outputs in BI is a critical point. It addresses the challenge of distilling vast oceans of data into meaningful narratives and embedding these insights seamlessly into the decision-making fabric of modern companies. The dashboard is considered by several sources as a graphical user interface and the main goal of a BI dashboard is to provide a clear and immediate overview of relevant business data in an easy- to-digest format. Reports are often created as pages or full informative documents to give precise overviews for some time. According to research created by a team from Databox, that had twenty-four respondents to their questions from the field. Current research suggests six commonly used types of cross-departmental business support dashboards:

- *„Web analytics dashboards,*
- *Marketing dashboards,*
- *Sales dashboards,*
- *Business dashboards,*

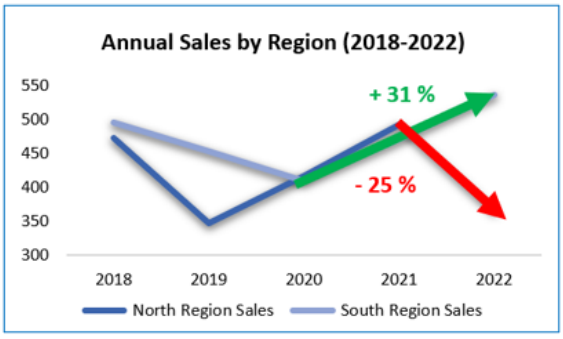
- *HR dashboards,*
- *Customer service dashboards*“ (Analytics, Databox, 2023).

Each dashboard targets different metrics according to the needs of the department and business. As a part of the practical section, a dashboard will be created, which is an interesting example of complexity, and this is not always the right choice. There will be data from marketing, and data about system users and their activity, which goes all the way to the phase of passing on information through the training they have to complete.

Finding the right balance between a visually appealing interface and a feature-rich user experience is a challenge for many developers. The perfect dashboard shows everything with just a click, everything important is at one‘s fingertips. The design provides a consistent overview with sparse, clear underlying data and additional opportunities for detailed information. Elements are displayed in a collapsed view, and you can view the details in a modal window or go to the details page. Below a view of the table of annual sales of two regions, from which not much can be read at once, can be found. After choosing the right chart based on the data, a visualization is created that tells people much more information at a glance, as presented in Figure 7 below on the right side. This choice effectively shows annual sales trends for each region, making it easy to track patterns over the years.

Figure 7 – Table & Visualization Example

Year	North Region Sales	South Region Sales
2018	472	495
2019	347	453
2020	417	409
2021	492	471
2022	367	536



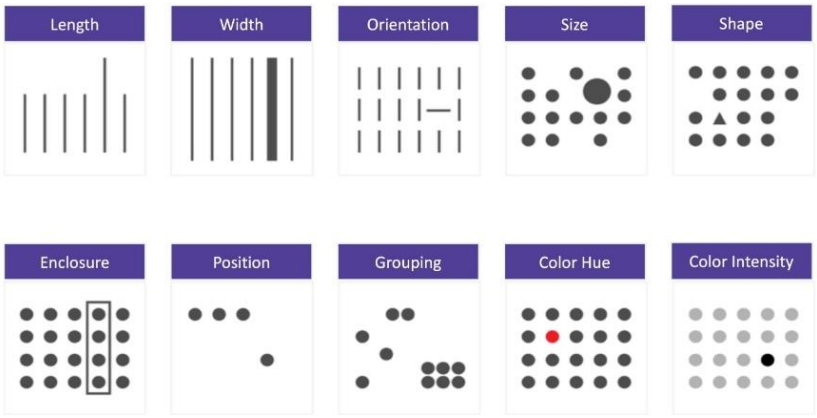
Source: The Thesis author's processing

Another bonus of visualization is the help of memory, which is short-term and is not able to retain more values after reading, for example: the first three lines with values. This example illustrates how data visualization can highlight important trends and insights that might be missed by just looking at numerical tables.

4.2.5.1 Mastering Preattentive Principles

Preattentive attributes are visual elements that instantly stand out and capture attention without conscious effort. It should help the user to grasp the most valuable information in the initial few milliseconds of viewing an object. Their purpose is to highlight key data points, divide two data sets, or guide user focus. There are several types of elements that can influence the user's attention. These elements are summarized in Figure 8 below.

Figure 8 - Preattentive Attributes



Source: (Duncan, 2021)

By making an inappropriate choice, one will attract attention somewhere else. Additionally, there is a need to maintain some consistency in the use of attributes. There may be a situation where the green color in the first graph means positive sales results, where the top performance of the sales department should be highlighted. But in the next graph, the green color will represent the space in the warehouse, and the goal of the graph is to point out the overcrowded spaces in the warehouse. An improper application of the colors will confuse. It is always important to consider what and how to highlight so that it has the desired effect.

4.2.5.2 The Power of Color Palette

Color in dashboard design transcends its role as a mere cautious attribute; it is a multi-functional tool that communicates, differentiates, and enhances the user experience. It includes pre-processing that focuses attention on critical data points, the emotional and cognitive effects of color that evoke emotion and influence mood, and brand considerations where color schemes often match brand guidelines for a cohesive visual identity. Understanding these aspects is key to creating informative and engaging dashboards. There are certain patterns of behavior and explanations of colors that are used in the world, and which are also applied to the field of dashboard design. The fundamental meanings associated with colors are:

- Red: Attention, warning, negative trends (e.g., a decrease in sales).
- Green: Positive trends, growth, success (e.g., an increase in revenue).
- Blue: Trust, reliability, calmness (often used for background elements).
- Yellow: Caution, optimism, or to highlight elements needing attention.
- Orange: A mix of urgency and optimism, often used for calls to action.
- Purple: Creativity, luxury, and sometimes used to indicate advanced metrics.
- Black/Gray: Neutrality, used for background, grids, and secondary elements.
- White: Clarity and space, used for background and to create a clean look.

For the color palette, 4-6 colors are recommended, but it depends on the robustness of the dashboard. To easily create an effective palette, one can consider this guide, which is used in the practical part: One positive accent color, such as a shade of green, indicates growth or success. One neutral background color such as soft gray or light blue for a clean, understated layout. One alert color, such as distinct red for alerts or critical issues. One medium highlight color such as orange or lighter blue for average or stable metrics. One highlight color can be bright blue or yellow to highlight certain data points. One additional color, an easily distinguishable color to match the above for different data representations.

Colors, when used effectively, can improve the usability and readability of a dashboard. The key is to use color purposefully, consistently, and in a way that matches the psychological impact on the viewer. By following these principles, designers can create dashboards that are not only visually pleasing but also effectively communicate the underlying data story.

4.2.5.3 Typography in Dashboards

In a dashboard design, the combination of typography and color is essential to create a clear and cohesive user experience. The key to matching typography with color is to set the color of the text against a different colored background where sharp contrast should be maintained for ensuring legibility. A mismatch between the tone suggested by the color palette and the typeface can cause a misunderstanding or a bad user experience. Basic types of typography:

- Serif fonts: Such as Times New Roman or Garamond are traditional and often used in formal or corporate dashboards to convey confidence and professionalism.
- Sans serif fonts: Like Helvetica or Arial, they are modern and clean, making them a common choice for technical or modern business panels for their legibility and elegant appearance.
- Monospaced Fonts: Like Courier, they resemble typewriters and are often used in technical dashboards that display code or data items for equal spacing and alignment.

A dashboard's effectiveness is enhanced by its ability to communicate clearly and concisely, from the title to descriptions and explanations of KPIs. The name of the dashboard is the first impression, so it should capture the essence of the data insights it offers. The perfect name should be:

1. Concise: Keep it short to ensure it is easily understood immediately.
2. Descriptive: It should reflect the dashboard's purpose and the nature of the data being analyzed.
3. Engaging: A title that sparks interest will encourage deeper exploration.

Instead of a title like "Sales Dashboard," a more descriptive and engaging title could be "2024 Regional Sales Overview: Trends and Opportunities." Integrating typography and color requires a thoughtful approach where functionality meets visual appeal. The goal is to create a dashboard that is not only informative but also intuitive, ensuring users interact with data seamlessly and efficiently.

4.2.5.4 Intuitive Dashboard Layout

Creating an effective dashboard layout requires a synthesis of all design elements. Certain positions on a dashboard are more naturally attention-grabbing. Placing critical information in these spots ensures that they are noticed quickly. In more detail, the upper left corner is where the viewer's gaze usually begins, because in some cultures it is read from left to right. This place is intended for the most important or summary information. Next, there is the middle, which is meant for large central elements that can quickly grab attention, so it is a good place for key visuals or metrics that anchor the dashboard's story. Additionally, the top edge of the dashboard in the upper left corner is also where viewers expect to find essential information, making it ideal for high- level metrics or trends.

The design of the arrangement of the cerebral hemisphere is another theory that could help to create an intuitive and easily digested dashboard. It is an interesting approach, an unproven approach to dashboard layout. This concept makes use of the hypothesized specializations of the cerebral hemispheres:

- Left Side: Place analytical visualizations and data-rich metrics on the left side of the dashboard. This is consistent with the left hemisphere's strengths in logic and analysis.
- Right Side: Use the right side for eye-catching, creative visualizations like infographics or interactive elements. This can make the dashboard not only more visually appealing but also more memorable, utilizing the right hemisphere's capacity for creativity and spatial awareness.

It is also important to consider how the layout will adapt to different screen sizes. A responsive design ensures that the most critical information remains prominent on any device, maintaining the dashboard's effectiveness.

Dashboard layout is not just about aesthetic appeal; it is about creating the path of least resistance to understanding the data. A well-thought-out dashboard layout is a silent guide that takes the users through the data story with intent and purpose.

4.2.6 Dashboard Deployment

Deploying a dashboard effectively is pivotal to leveraging the full power of business intelligence within an organization. This chapter provides a roadmap for the deployment process, covering security, testing, maintaining, optimization, and the adoption and training necessary to maximize the dashboard's value.

4.2.6.1 Dashboard Safety & Security

The safety and security of dashboards is a crucial topic. An unauthorized access or alteration of data can lead to significant losses, both financially and in terms of an organization’s reputation. Therefore, it is essential to understand and implement various security measures, including authentication, authorization, encryption, and regular security audits with a special focus on MS power BI.

Authentication verifies the identity of a user or process. It is usually achieved through methods such as passwords, biometric scans, and token-based and hardware devices. Once authenticated, the user is then authorized. It specifies what data a user is allowed to access and what they can do with that data. For instance, to access the Showpad database, API tokens are used.

Encryption is the process of converting information or data into a code to prevent unauthorized access (Kime, 2023). It is a critical feature of a safe and trustworthy Internet. On the other hand, security audits are systematic evaluations of an organization’s security measures and controls. They help to identify potential vulnerabilities and ensure that all security measures

are functioning in line with the main purpose. During the data extraction, different encoding methods might be used such as UTF-9 or another method Base64 encoding (Chiarelli, 2022). It is not an encryption method but an encoding one that converts binary data into text format.

MS power BI Desktop & Service Safety & Security

MS power BI Desktop, used in the practical part, provides additional security measures, one of which is Row-Level Security (RLS). RLS allows you to control access to data at the row level based on user roles (Microsoft, 2024). This means that different users can have different views of the data based on their assigned roles, adding an extra layer of security. Regular monitoring of access logs and prompt application of security patches are crucial for maintaining the security of a dashboard. Additionally, regular security audits can help to identify potential vulnerabilities and ensure that all security measures are functioning as intended.

In the context of the practical part, the role of a viewer is particularly relevant. Viewers are individuals who are assigned access to view the dashboard but do not have permission to make changes. This ensures that specific individuals can view the dashboard and that a necessary level of security is maintained. It is important to note that no further security measures beyond those mentioned will be implemented in the project of Showpad as the viewer role is sufficient to ensure that only authorized individuals have access to the dashboard.

In conclusion, securing a dashboard involves a combination of various strategies. Whether it is implementing robust security protocols, leveraging features provided by tools like MS power BI, or integrating APIs using tokens for secure data extraction, the goal is to protect sensitive data and prevent unauthorized access.

4.2.6.2 Dashboard Testing

Testing dashboards and BI solutions is a multi-faceted process crucial for ensuring the accuracy, performance, and usability of BI systems. Effective testing strategies guarantee that BI tools deliver insights quickly, respond to user queries efficiently, and oversee real-world usage scenarios without faltering. Testing

can be divided into several topics according to the required measurement and this theoretical framework provides a foundational approach for comprehensive testing:

Performance Testing: A general objective may be to assess the response time and efficiency of the dashboard and the solution behind it, query execution, and report generation. A practical example will be given on the MS power BI performance analyzer, which evaluates the load times of visuals in the practical part.

Accuracy Testing: At the heart of BI is data accuracy. Testing must verify that data extraction, transformation, and loading processes are correctly performed. To verify the fidelity of the ETL, there can be more methods used, such as comparing data with original data sources, in the Showpad case comparing power BI data with Showpad online platform analytics.

User Acceptance Testing (UAT): It involves real users evaluating the solution to ensure it meets requirements and is user-friendly (Fastercapital Team, 2024). It includes testing dashboards, reports, and other user interfaces for intuitiveness and ease of use. Feedback from the project owner will be used to verify the Showpad dashboard, which will be collected in the form of an Excel form.

Security Testing: With BI handling sensitive business data, security testing is indispensable. This testing ensures that data is protected against unauthorized access and breaches. It includes verifying user permissions.

From performance and accuracy to security and usability, each test phase is designed to identify and rectify any issues that could impede the BI system's operation. A well-tested BI solution not only instills confidence in the data-driven decisions it supports but also ensures that the organization can rely on its BI system to withstand the demands of the ever-changing business landscape.

4.2.6.3 Maintaining & Updating

Maintaining and updating the dashboard is an ongoing process that ensures its continuous relevance, accuracy, and effectiveness. This chapter will summarize the best practices for keeping a dashboard operational and reflecting on the latest business data and user needs.

First is routine maintenance, which must have regular checks and updates of data sources, software versions, and user access permissions scheduled.

Next is data monitoring and validation, which involves a search for the anomalies detected in real-time data, which should lead to an immediate investigation and rectification to maintain data accuracy.

It is also possible to set up a feedback loop for the end users, who can give feedback at any time and the comments will then be considered and the dashboard updated.

Feature updates and new updates released by BI software providers must be considered. Evaluate and implement it to improve dashboard functionality and user experience.

Performance Optimization

Optimizing BI dashboards from a technical standpoint requires a focus on data management, system performance, and efficient resource utilization. Here are the key technical areas to consider:

Data management, to maintain data integrity, the implementation of a data verification policy and regular audits are appropriate. Furthermore, data compression must be mentioned, which reduces the size of data for further work, it can also be data models that can go from 10GB to 1GB, and it can increase the refresh rate and query performance. Star and snowflake diagrams are often used for their effectiveness. The last point in data management is data indexing, as it can improve dashboard performance by reducing the time it takes to load data.

Query optimization is the second topic where writing effective queries with optimized joins and aggregations will speed up data retrieval.

Complex subqueries and related subqueries that can slow down performance should be avoided. Next is a focus on strategy caching, which is for storing the results of expensive operations. This can reduce database load and speed up data loading for the commonly accessed data.

Furthermore, the optimization will focus on the performance of the system, where it is primarily about loading times when it is necessary to optimize the loading times of dashboards by minimizing the amount of processed and loaded data at runtime. Dashboard design is also important, as reusing components and queries wherever possible can reduce server load and development time. Also related to this topic is asynchronous loading, where one uses asynchronous operations for loading various parts of a dashboard to ensure that the users can interact with the dashboard without waiting for all components to load.

Other best practices such as version control are when the development dashboards to systematically manage changes and deployments are used.

Technical optimization of BI dashboards requires a thorough approach to data management, system performance, and resource utilization. By focusing on effective data modeling, query optimization, server and network resources, and rigorous monitoring, organizations can ensure that their BI dashboards are not only functional but also powerful and scalable.

4.2.6.4 Change Management

In the era of data-driven decision-making, dashboards have become an essential tool for organizations. However, the successful implementation of a dashboard goes beyond its creation and involves a notable change management process. The first step in the change management process is understanding the need for a dashboard (Dana Miranda, 2022). User acceptance is a critical aspect of the change management process. It is not enough to create a dashboard; users must adopt it and integrate it into their daily work processes. User adoption can be facilitated by involving users in the design process, soliciting their feedback, and making necessary adjustments based on their input.

Once the dashboard is designed and implemented, users must be trained on how to use it effectively. This includes understanding the various dashboard functions, how to interpret the data presented, and how to use that information to make informed decisions. Training can be delivered through workshops, online tutorials, or one-on-one sessions depending on the size and needs of the organization. Even after a dashboard is adopted and users are trained, ongoing support is essential to ensure its continued effectiveness.

MS power BI provides several features that can facilitate the change management process. For example, it allows for easy collaboration and sharing of dashboards, which can help acquire users. It also provides extensive documentation and training resources to help users learn how to use the tool effectively.

In conclusion, the dashboard adoption and training involve a meaningful change management process. It is not just about creating a dashboard, it is about ensuring that users adopt it, that they are trained to use it effectively, and that there is ongoing support to ensure its continued effectiveness.

5 Practical Application

The second main part of the thesis demonstrates the practical application of business intelligence (BI) concepts within the company's leading healthcare industry. Specifically, it focuses on the comprehensive development of dashboards to improve business and marketing decision-making. While maintaining confidentiality, the business solutions team of the company will be presented, which includes a set of BI tools such as MS power BI, Tabular Editor, and the Snowflake data warehouse for optimal data visualization, ETL tasks, and storage. While these systems facilitate decentralized BI solutions across different capacities, there is a key strategy to consolidate data into a centralized Snowflake data warehouse. This consolidation is ready to improve data processing capabilities and provide a holistic view that captures business data from all divisions and not just from the business unit from which the client and owner of this project are. Currently, dashboard development is performed in a decentralized manner by the commercial operations team, ensuring that critical data queries are managed with maximum accuracy and relevance to the specific needs of the sector.

5.1 Showpad Business Analysis

The process of gathering project specifications is a methodical endeavor that sets the foundation for the successful development and deployment of the BI dashboard. First, this chapter includes a summary of a well-crafted Business Requirements Document (BRD), which encapsulates the project's aspirations and criteria from a high-level business perspective. Secondly, it transitions into granulation of these requirements into a few user stories, providing actionable insights and clearly defined objectives for the project development team.

5.1.1 Gathering Showpad Project Specifications

The Business Requirements Document (BDR) serves as a cornerstone of the project planning and strategic phase, capturing the essential elements necessary for the development of a user-centric BI tool. Outlined within the BDR are the following detailed requirements and expectations:

Business Needs and Objectives: The current project, guided by the astute insights of the project owner, revolves around creating a Showpad dashboard that seamlessly blends innovation with user-centric design. The mission is to construct a comprehensive dashboard that captures the intricate nuances of sales and marketing data. This endeavor empowers decision-makers with actionable insights. It aims to optimize engagement strategies across commercial operations, sales management, and training & development spheres. Fueled by a secured funding of $10,000 aspired to transform the data shared on the Showpad platform into compelling narratives using the MS Power BI visualizations tool.

Purpose Clarity: The thematic focus of this dashboard centers around the key metrics that unveil the story behind:

- Showpad Platform Usage,
- Marketing Content,
- Sales Team Engagement,
- Coaching Training Overviews.

Stakeholders & Users Engagement: The tool will cater to the needs of commercial operations, sales, and training & development teams. By analyzing key metrics such as marketing content engagement and sales field team activity, the tool will provide a nuanced view of the effectiveness of materials and initiatives. Their feedback and active engagement are crucial in shaping a dashboard that aligns with their functional requirements and aesthetic sensibilities.

Business Questions: In response to the critical business questions, the dashboard will not only reflect performance metrics but also delve

into the coach completion rates, regional activity, and material interaction. This granularity ensures that every data slice contributes to a comprehensive narrative. Dashboard thematic focus revolves around the key metrics that unveil the story behind:

Marketing Content Engagement:

- Which materials were most popular in the last 14 days/ last month?
- Do the materials change according to the launch of new campaigns and new products on the market? What views do the materials for the new product have?
- Who looked at these materials from the field?

Sales - Field Team Engagement:

- How often is a team active on Showpad?
- Who is the most active, and what materials do they interact with?

Coach Training Overview:

- What is the average completion rate of training for the year 2023?
- Which region or country has the greatest engagement in coach training?

Commercial operations – Showpad Usage:

- What is the activity on the platform?
- What devices are used?
- How many paid licenses are there currently?

Availability, safety, and security: Implementation of Row Level Security (RLS) ensures the protection of sensitive data. (Employees will not see their management activities or the data of their colleagues, which is not the current situation of the dashboard and their end users, however, it is necessary to consider this option in the future)

- Non-sensitive data: People's names, managers, and roles.

- Potentially Sensitive Data: Interaction with materials and activities and dates of hiring sales representatives.

The dashboard will be accessible to the authorized users selected by the stakeholders with data distribution through a workspace, shared once or twice a month. The data will be refreshed as per the desired frequency, twice a month.

After a careful collection of the project owner's requirements, the foundations for implementation were laid. Moving forward, the project remains committed to refining these pillars. Should the need arise, it might be necessary to engage directly with the owner for additional insights or finer details.

5.1.1.1 User Stories

Following the BDR, user stories translate overreaching business needs into specific, user-centered objectives. They highlight the functionality and features of the dashboard from the end-user's perspective, ensuring the final product resonates with its intended audience.

1. As a marketing person, I want to track the viewership of marketing materials uploaded to Showpad. This will help me understand which materials resonate most with the field team, allowing us to optimize content creation and distribution effectively.
2. As a marketing person, I need insights into audience demographics. Distinguishing between regions, countries, and employee roles will guide a targeted approach to content creation. By tailoring materials to specific audiences, the field team with engaging resources for their sales and client interactions can be properly equipped.
3. As a sales manager, I want to monitor how the sales team utilizes the Showpad platform. Measuring engagement with materials and ensuring compliance with mandatory training modules within Showpad Coach is critical. By identifying trends in Showpad engagement rates, one can assess whether engagement is improving or decreasing over time.

4. As a Training & Development specialist, I seek insights into how our training materials resonate with the sales and field teams. By analyzing engagement metrics, we can tailor our training modules effectively. The dashboard will provide real-time visibility into training completion rates, material interactions, and areas where additional support may be needed.
5. As a commercial operations analyst, I need to aggregate and synthesize engagement across all user interactions with the Showpad platform, such as login frequency or paid licenses. Gaining a holistic view of sales and field teams' use of materials and training modules ensures an alignment with the strategic goals. In addition, it is essential to identify the areas for improvement or additional support. The dashboard will unify engagement metrics across marketing content, sales materials, and training modules. Leveraging API integration will automate data collection and real- time analysis for actionable insights in the field.

In conclusion, BDR together with user stories constructs a detailed and pragmatic blueprint that ensures the BI dashboard is robustly aligned with a business strategy while remaining agile and responsive to the user feedback throughout the development process.

5.1.2 Data Sources & Showpad Understanding

The efficacy of BI hinges on the quality and comprehension of the underlying data sources. This chapter elucidates the provenance and cogency of data, outlining the methods by which disparate data streams are transformed into cogent insights. Mastering the data landscape is an endeavor of both discovery and interpretation. The expedition commences with an in-depth look at two primary sources foundational to BI projects:

- Showpad Data: This platform is not merely a repository but a dynamic ecosystem teeming with marketing intelligence, sales interactions, and advanced training modules. It encapsulates a confluence of data points that are critical for devising and refining sales strategies.

- Excel File (HR Details): Within this Excel lies user profiles, roles, and organizational nuances. This file bridges the gap between data and people, providing context to analytics endeavors.

The integration of Showpad's analytical capabilities with the contextual richness of human resource data provides a holistic view. This dual perspective enables a comprehensive approach to business analysis, balancing quantitative assessments with qualitative insights. Showpad's innovative platform exemplifies the confluence of sales and marketing functions, fostering a unified approach to content delivery and training. (Showpad Team, 2023) Showpad is designed to bridge the distance between sales effectiveness and effective advertising and marketing, this platform has transformed the way sales teams interact with customers and consume content. The insights provided herein are intended to equip the reader with the knowledge to fully leverage the potential of diverse data sources, converting raw data into a strategic asset for business intelligence.

5.2 Data Management

This section will set the stage for understanding the critical importance of data management in practice. It will touch on Python data retrieval, the critical role of APIs, and pagination for managing large data volumes.

5.2.1 Data Extraction & Storage

This chapter covers the extraction phase, which is the critical first step in the data management lifecycle. It involves understanding project requirements from the data perspective and retrieving those data, mostly from the Showpad database, setting the stage for further cleaning, analysis, and visualization. If challenges or limitations are encountered during the process, if it is in the project resources, a solution will be applied or a solution will be proposed, and the given situation will be embodied.

5.2.1.1 Understanding Showpad API Access

Showpad has API documentation on its website "https://docs.api.showpad.com/docs/apis", describing that Showpad has a RESTful client-server API. The accessible data will be in simple JSON format.

To access the data, one needs to go through authentication and authorization, in this case, it is the OAuth2 token-based method, where the "ShowpadDataToken" token is available for this project.

It is also necessary to know the configuration of the endpoint. Underlying everything is the base endpoint, namely for the Showpad: "https://{{subdomain}}.api.showpad.com."

After testing the access to the database through the web API interface in the Showpad environment for testing using the appropriate GET call, a header with a token and the necessary parameters can look like this:

Code 5 - Showpad API Request Headers Configuration

```
import requests
import json

url = "https://{{subdomain}}.showpad.biz/api/v3/exports/events.json"

headers = {
    "Accept": "application/json",
    "Authorization": "Bearer "ShowpadDataToken""
    }

query string = {"startedAt": "2024-03-01", "endedAt":"2024-03-
17","pageBased":"true", "limit": "1000"}
```

Source: The Thesis author's processing

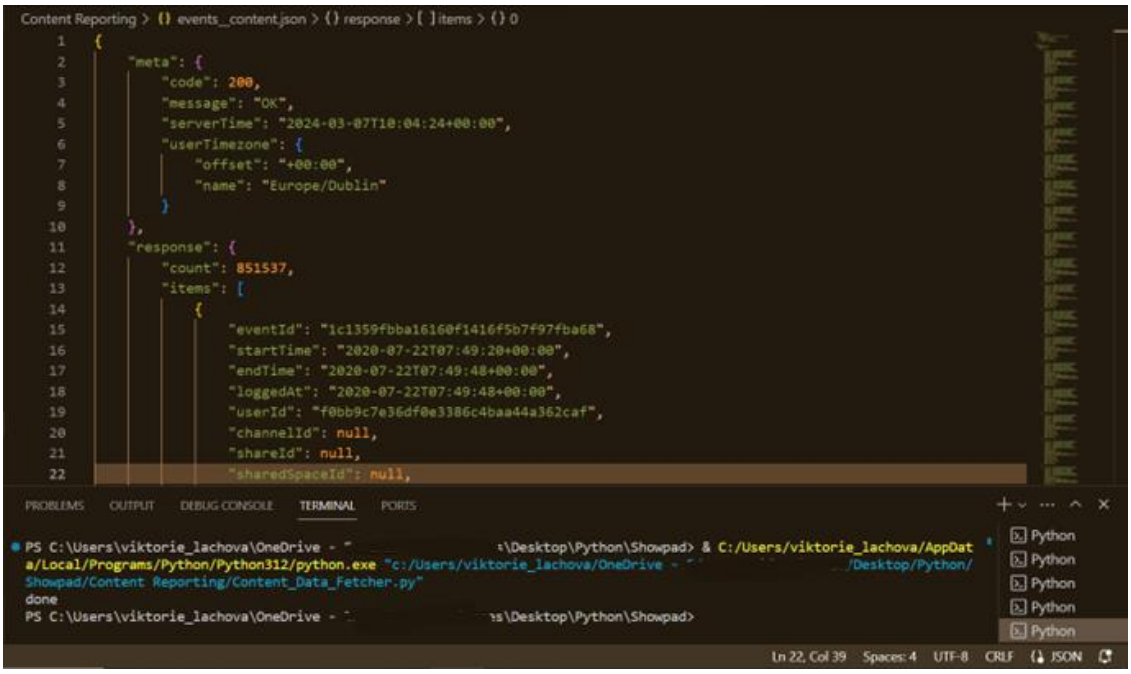
5.2.1.2 Extracting Content Data with Python

In the grand theater of data management, the spotlight falls on a pivotal extraction: pulling the first batch of data based on parameters. The goal of the code is to download the first data based on the set parameters. At the beginning, the Python libraries are defined, then the set header for identity verification continues. Next, the parameters section, where e.g., the limit, which is otherwise

set by default, is defined. Furthermore, the code defines the response from the API, which is then saved as such.

Below an example of the response, can be seen, showing a screenshot of a code editor Visual Studio, highlighting a JSON response from a simple GET call. that the Events “response” section contains data with 851,537 events-items. The response includes metadata under the “meta” key, providing information such as the HTTP status code (200 for success), a success message (“OK”), server timestamp, and user timezone (Europe/Dublin). The JSON structure hints at further data transformations and adjustments to make it useful for specific purposes.

Figure 9 - Showpad API Response



Source: The Thesis author's processing

When analyzing responses, especially concerning findings or insights, it is essential to ask critical questions. Let us break down the considerations:

Some columns are not necessary for the dashboard and therefore no longer need to be pulled from the database, this is the first optimization that can be done. Properly setting the parameters, called "arrays" in Python, will allow one to select only the necessary columns. However, the Showpad

API does not have such a feature. Therefore, all data will be downloaded, and unnecessary data will have to be deleted.

Data comes from an API, considering how to get it efficiently. If the data set is large, it can be paginated. A loop is made to extract all the data. Most of the loops here at Showpad use paging with the use of the limit and offset parameters to load data. When working with large data sets, it is advisable to set a maximum limit, in this case a thousand, to minimize the number of API calls and thus optimize the calls. In this case, Showpad recommends using cursor paging, called 'X- Showpad-Scroll-Id' to fetch more data. For this particular case of the events table, the server returns a cursor leading to the next section in each request, giving all the data linked together. When measuring the time from the start of the API call to the final print, the code ran exactly in 3:04.08 for one month of data for October 2023, i.e., the parameter e.g.: 1.10.2023- 31.10.2023. Data extraction is expected to run around 47:06 minutes for the full 14 months remaining. Below is code 6 as an example for this particular event name table as well and a summarized explanation can be seen. The code fetches event data from the Showpad API, handles pagination, and saves it to a JSON file. It uses the requests library to make HTTP requests and the module to handle JSON data. The retrieved events are stored in the all events list. The script prints a message indicating successful data retrieval or failure.

Code 6 - Fetching Showpad Content Events Data

```
import requests
import json

url = "https://{{subdomain}}.showpad.biz/api/v3/exports/events.json"
headers = {
    "Accept": "application/json",
    "Authorization": "Bearer ShowpadDataToken"
    }
querystring = {"startedAt":"2020-07-22","endedAt":"2024-02-29","pageBased":"true",
"limit": "1000"}
all_events = []

while True:
    response = requests.get(url, headers=headers, params=querystring)
```

```
    if response.status_code == 200:
        data = response.json()
        items = data.get('response',{}).get('items',[])

        if not items:
            print('no more data')
            break

        all_events.extend(items)
        scroll_id = response. headers.get('X-Showpad-Scroll-Id')
        if not scroll_id:
            break
        headers['X-Showpad-Scroll-Id'] = scroll_id
    else:
        print('failed to fetch data')
        break

    with open ('events_Feb2024.json', 'w') as file:
        json.dump(all_events, file, indent=4)
print('Events data fetched', response.status_code)
```

Source: The Thesis author's processing

In the grand theater of data management, the spotlight falls on a pivotal task: extracting the initial batch of data based on the specified parameters. The Python code exemplified here demonstrates this process, emphasizing identity verification headers, parameter definitions, API responses, and simple optimization methods. In this way, the rest of the content data will be extracted and stored in JSON files ready for further development.

5.2.1.3 Extracting Coach Data with Python

Another part of data extraction from Showpad is the data from the coach module, which has a slightly different structure and therefore deserves its own chapter. Those data are nested JSONs that have 'fields' defined as parameters for more efficient and optimized data extraction. One of the new features is a new Python library that will need to be used. This is base64, which is a method for encoding binary data into ASCII characters.

Code 7 - Fetching Showpad User Data

```
import requests
import base64
import json

url =
"https://{{subdomain}}.showpad.biz/api/learn/reporting/v3/json/objects/Checkl
istUser"
headers = {
    "Accept": "application/json",
    "Authorization": "Bearer ‘ShowpadDataToken’"
}

params = {
    "page": 1,
    "resultsPerPage": 100,
    "fields": ["userId", "firstName", "lastName"]
}
```

Source: The Thesis author's processing

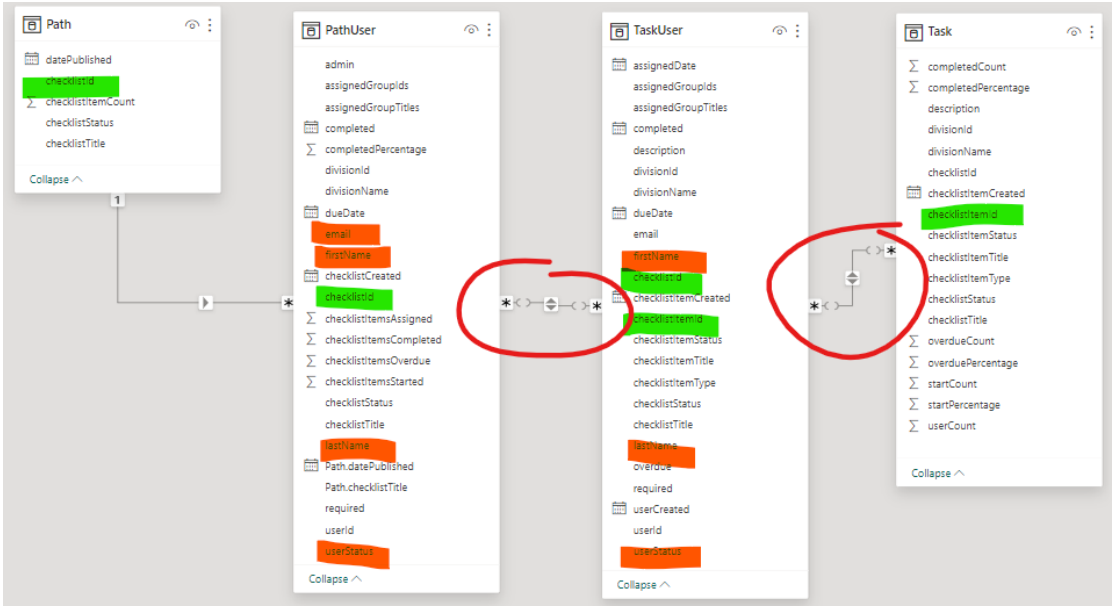
During the extraction table ChecklistUser, there was an issue with the code for ChecklistUser(PathUser). When evaluating the code, it was found that there is a problem with the field "datePublished", which either does not exist, there is no data, or it has a different name. However, the data can be obtained from another table since there are data duplicities across all coach tables. With this method, the rest of the coach tables will also be extracted into JSON files for further work and further transformations.

The data from the coach at the highest granularity are paths that consist of tasks that are assigned to users. However, it can be seen from the documentation that the data called on path (Checklist) and on courses are not sufficiently prepared to create an optimized data model right away. After extracting the first four tables, the critical issues encountered were:

- inconsistent granularity,
- redundant or duplicate data,
- no one central fact table,
- lacking dimensions,
- normalized table (not always good in BI use cases),

- many to many relationships as you can see in the Figure below.

Figure 10 - Coach Data Extraction



Source: The Thesis author's processing

Compared to the content part of coach data, there is a complete challenge in the form of restructuring the model and creating tables, and it may also contribute to the creation of a more suitable data model.

As a part of the solution, an agile step was used, namely the use of a table that is accessible for extraction in the Showpad platform, which is described in the chapter entitled 'Coach Data Model'. The step of temporarily changing the data source for the coach and deviating from the planned path, while the development is closer to the goal, is an opportunity to get the data source in a position where it is needed and more effective. The reason is that modifying a large number of tables and data models directly in power BI Desktop can be inefficient for several reasons such as performance, maintenance, collaboration, data refreshing, reusability, complex transformations, versioning, and others. It is great to use power BI desktop for smaller volumes of data, but for larger volumes, it is more appropriate to either make changes directly in the database, or data warehouse, reconfigure the API, or the second option is to set data flows within power BI service. These dataflows can provide

an efficient way to transform and manipulate data and tables. The benefits would be ETL effectiveness, seamless integration with Power BI desktop reports, and the possibility of scheduling refreshes.

5.2.1.4 Field Team Data

The next part is a list of the active Showpad users, where the data must also be supplemented with other necessary information. These are, for example: employee roles or hire dates. This data comes in an Excel file from the human resources department every month, and it is necessary to manually update the file with active people in the company.

5.2.1.5 Data Storage

This chapter explains the data storage strategy employed in this project, where data has been extracted with Python and preserved in JSON format on a local computer system. This method is an initial step before uploading the data into the MS power BI desktop for further processing.

This storage was chosen due to easy readability and flexibility, and also due to optimizations, it is lightweight and quick data transfer, especially with large volumes of data. Another bonus is compatibility with the power BI desktop, where the JSON data source is already defined, ensuring seamless integration. The disadvantage of this process is the manual management of files and other limitations.

5.2.2 Data Transformation, Load & Data Modelling

During the data extraction phase, a few basic transformations of the data source have already taken place, such as changes to the extracted data or subsequently, some data were deleted because they were not needed for this project. Furthermore, it will be a matter of transformation, some of which will take place in Python, some of which will take place after uploading data to the MS power BI desktop in power query, and some of which will use the Tabular Editor.

5.2.2.1 Data Transformation in Phyton

As a part of the content, the main changes in Python took place only in the Events table. Data from Events was extracted per month due to the fetching limit and amount of data. Therefore, several JSON files are now stored, which must be combined before further transformations. This code below consolidates event data from multiple JSON files into a single file named "all_events_2023.json." It iterates over files in a specified directory, reads the data from each file (assuming they contain lists of events), and appends it to an accumulating list. Finally, the script writes the accumulated data to the output file. It is a useful process for combining data from various sources into a unified format for further analysis or reporting.

Code 8 - Consolidating Showpad Events Data from Multiple Files

```
import os
import json
import glob

directory_path = r'C:\Users\viktorie_lachova\OneDrive -
…\Desktop\Python\Showpad\Content Reporting\events\eventsAllData'

file_pattern = os.path.join(directory_path, 'events_*.json')

all_events_data = []

for file_path in glob.glob(file_pattern):
    with open(file_path, 'r') as file:

        events_data = json.load(file)
        all_events_data.extend(events_data)

output_file_path = os.path.join(directory_path, 'all_events_2023.json')

with open(output_file_path, 'w') as output_file:
    json.dump(all_events_data, output_file, indent=4)

print(f"All events data has been consolidated into {output_file_path}")
```

Source: The Thesis author's processing

After joining the data, the data is further converted and to illustrate the next transformations, another code performs transformations modifying those

Events data. The script starts by loading JSON data from a specified file path into a Panda DataFrame. Next, the script defines a list of column names that are to be removed from the DataFrame. The script then checks if the 'startTime' and 'endTime' columns exist in the DataFrame. If they do, each column is split into two new columns at the 'T' character, separating the date and time. The time is further split at the '+' character to remove the timezone offset. This is another transformation operation as it changes the structure of the DataFrame. Finally, the transformed DataFrame is written back to a JSON file, and the path to the output file is printed. In summary, this script performs transformation operations by removing unnecessary columns and splitting columns in a DataFrame, thereby refining the data for further use. The script demonstrates the power and versatility of the panda's library in handling and transforming data.

Code 9 - Transforming Showpad Events Data for 2023-24

```
import pandas as pd

file_path = r'C:\Users\viktorie_lachova\OneDrive -
…\Desktop\Python\Showpad\Content Reporting\events\all_events_2023.json'
df = pd.read_json(file_path, orient='records')

columns_remove = ['sharedSpaceId', 'sharedSpaceParticipantId',
'metadata']
df.drop(columns=columns_remove, errors='ignore', inplace=True)

if 'startTime' in df.columns:
    df[['startDate','startTime']] =
df['startTime'].str.split('T',expand=True)
    df['startTime'] = df['startTime'].str.split('+', expand=True)[0]
else:
    print('startTime column not found‘)

if 'endTime' in df.columns:
    df[['endDate','endTime']] = df['endTime'].str.split('T',expand=True)
    df['endTime'] = df['endTime'].str.split('+', expand=True)[0]
else:
    print('endTime column not found')

output_file_path = r'C:\Users\viktorie_lachova\OneDrive -
…\Desktop\Python\Showpad\Content Reporting\events\events_tr_23.json'

df.to_json(output_file_path, orient='records', indent=4)
```

```
print(f'Data transformed {output_file_path}')
```

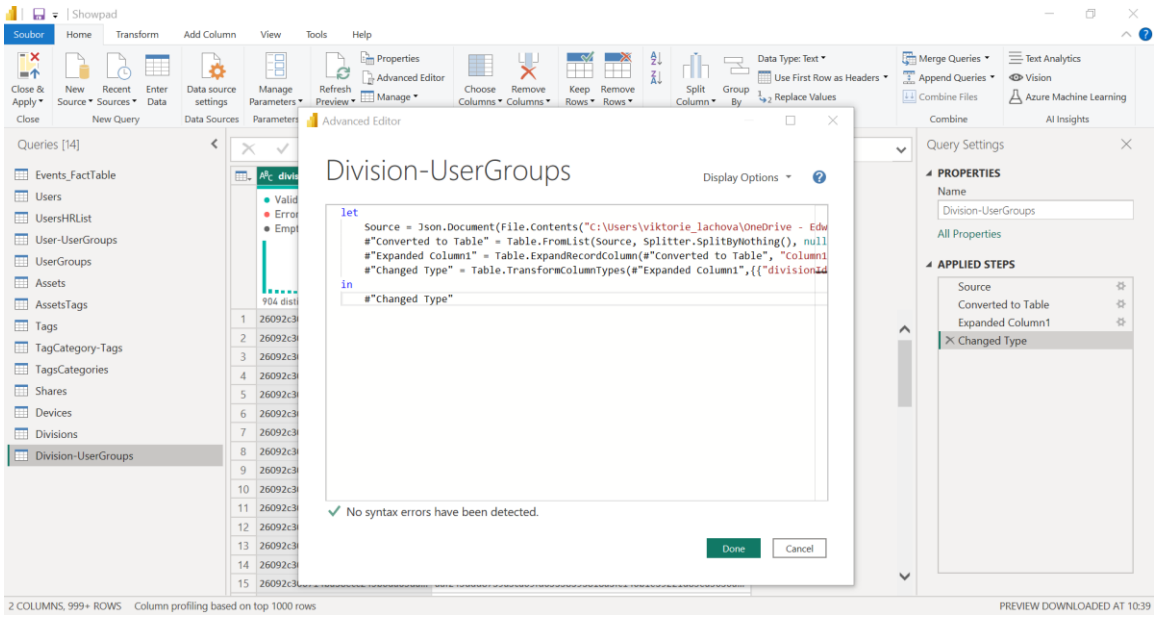
Source: The Thesis author's processing

Data Transformation in Power Query

After a few simple transformations using Python, the data must be loaded into the MS power BI desktop for further manipulation and transformation. As a first step will be uploaded the initial part of the content JSON text files.

Once the data is loaded and opened in power query as above, one can use for example the advanced analysis feature for this case study, where the transformation using M language is written in the screenshot below. This transformation turns the JSON file into a table and then splits the columns to change the data types that are more appropriate for the data.

Figure 11 - Power Query: Advanced Analytics



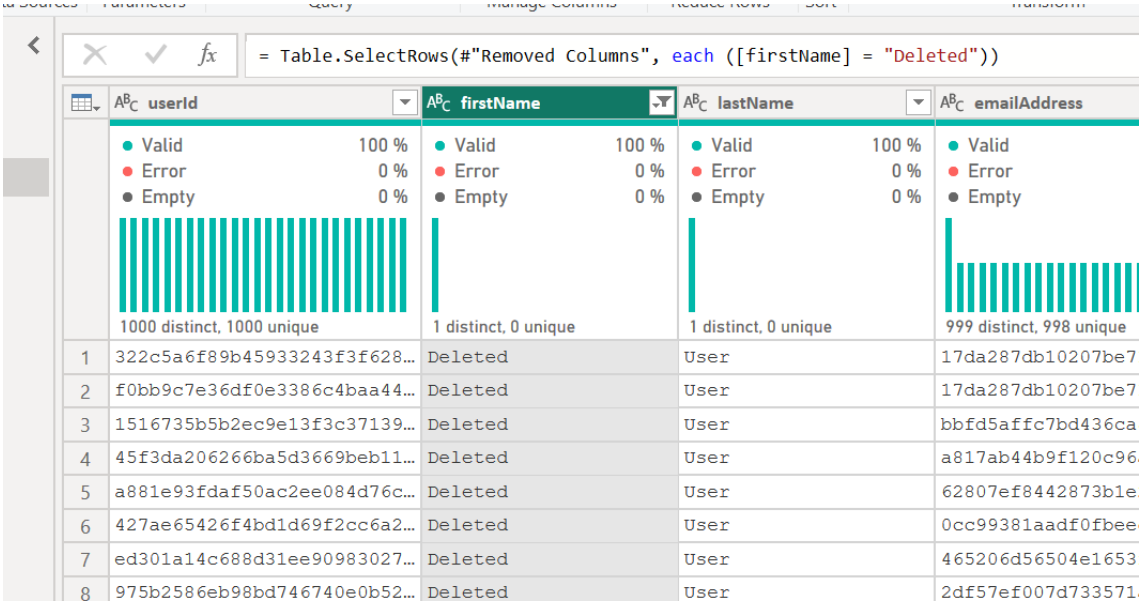
Source: The Thesis author's processing

Simple data transformations are pivotal for ensuring the quality and consistency of data before an in-depth analysis. Here are examples of such transformations used with power query on Showpad data:

Removing empty cells, duplicates, and replacing values: To maintain data integrity, rows or cells that contain no data can be removed and used

in the Assets table. This often involves filtering out records with null values in key fields or this case duplicated in assetId. Another case in the User table was instead of deleting records kept records of people who do not have the license to the Showpad platform anymore and an example can be seen below in the screenshot. FirstName is marked as Deleted, so it was filtered out. sometimes, datasets contain placeholder values that need to be replaced with meaningful data or null values.

Figure 12 - Power Query Transformation: Filtering Records

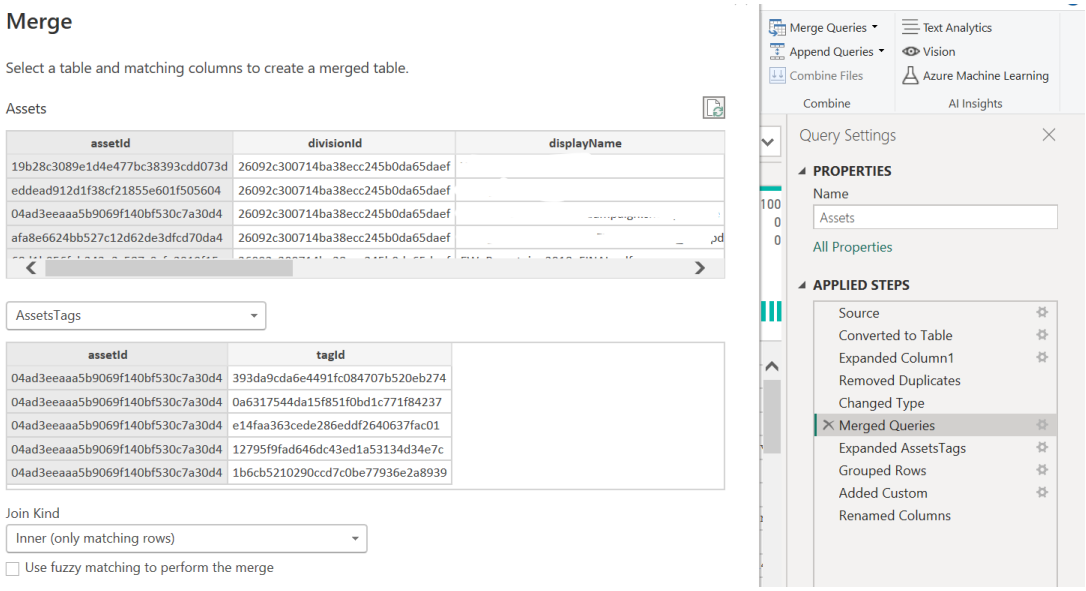


Source: The Thesis author's processing

Error, column renaming, data types: Identifying and correcting errors, such as outliers or incorrect entries, improves data accuracy. To ensure clarity, column names might be standardized or renamed to be more descriptive, facilitating easier identification of data fields. For correct processing, data types may be altered; for example, converting string representations of dates into datetime objects. This type of transformation was used for every table uploaded; it is the most used modification.

Merging Data: Combining data from multiple sources, such as joining tables on a key field, can enrich the dataset. This might involve using SQL JOIN operation as shown in Figure 13, where assetId and Inner join for only matching rows merge is chosen.

Figure 13 - Power Query Transformation: Tables Merge

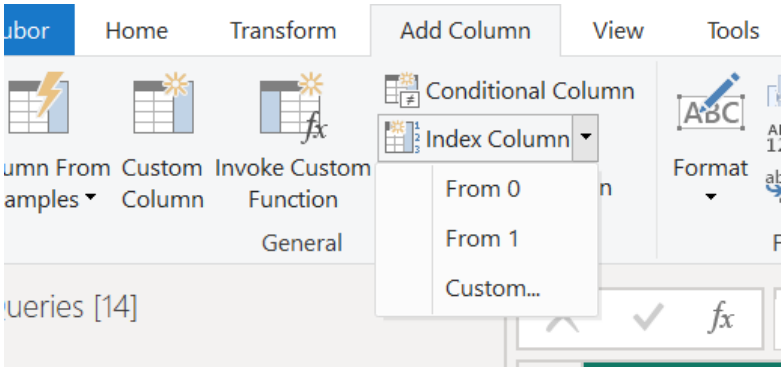


Source: The Thesis author's processing

Grouping Rows: Aggregating data based on a common field can be crucial and it was used in the Asset table where s grouping by assetId was applied.

Adding Index Column: In many instances, particularly for tracking changes and maintaining order, an index column is added. The case in Showpad data is with the Coach fact table, where it is used as an identifier for each row that can aid data manipulation tasks and provide a reference point within the dataset, and one can find it in the Add column menu.

Figure 14 - Power Query Transformation: Adding Index



Source: The Thesis author's processing

Creating calculated or custom columns: Generating new columns based on the calculations from the existing data, such as creating a profit column from revenue and cost columns, can provide additional insights. As illustrated below, this method was used for combining text into a new custom column.

Figure 15 - Power Query Transformation: Custom Column

Source: The Thesis author's processing

Upon completion of the main transformations in Power Query – which included removing empty cells, merging tables, grouping rows, and adding index, columns, among others – the dataset is now primed for the next phase. The next step is to develop a data model.

5.2.2.2 Data Modelling

At the beginning of the development of the data model, it was found that there are 2 fact tables, i.e., content - -events and Coach - -records. When creating one data model with multiple fact tables, queries, and the model may slow down, and therefore it is not a completely optimal solution. It is recommended to create separate data models so that the relationships that are not needed do not arise. Therefore, there will be two models.

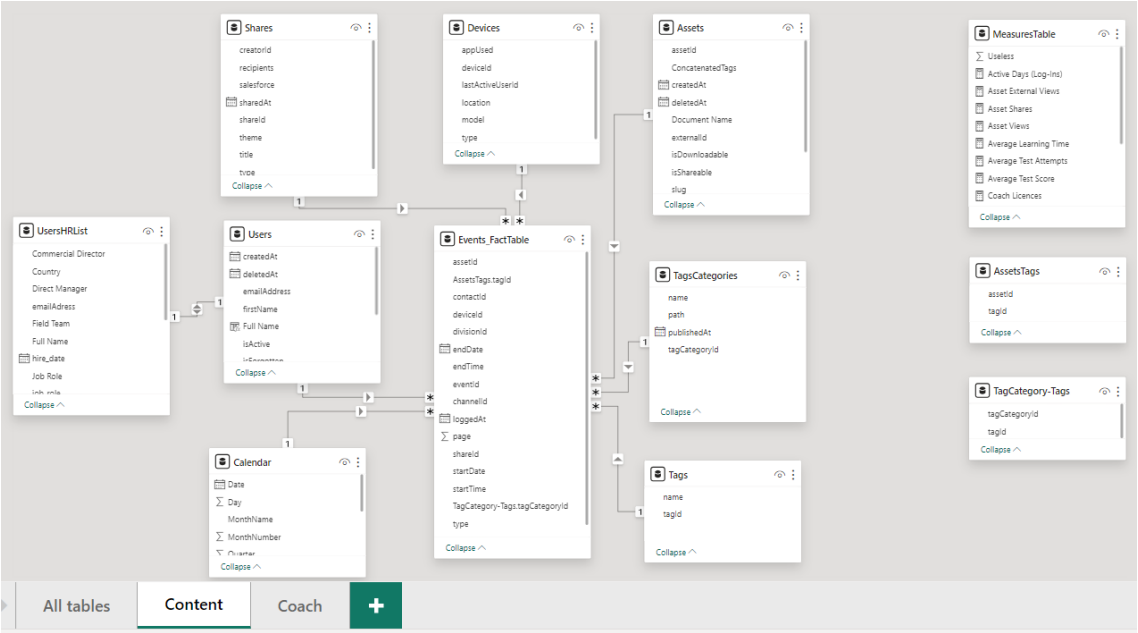
Content Data Model

This model serves as the analytical backbone for interpreting user interactions and content performance. The model includes various interconnected tables, each representing a distinct aspect of the business data:

Events_FactTable: A fact table that records the transactional data, such as user interactions with content. The presence of fields like startDate and endDate provides the ability to analyze the time-based patterns.

Assets Table: Contains details about the content assets themselves, which when combined with event data, can elucidate content usage and popularity.

Figure 16 - Data Model: Content



Source: The Thesis author's processing

Users Table & UsersHRList: Stores user-specific information, potentially outlining different user roles and their activity status. This is central to understanding the user engagement especially connected with UserHRList for more details.

Shares and Devices Tables: These track information about how the content is shared and accessed across different devices, offering insights into content reach and device preferences. In the future, if it is to be combined with the customer's data, a holistic view will be obtained.

Tags and TagsCategories: Used for categorizing the content, which is crucial for segmenting data and understanding the content taxonomy.

MeasuresTable: A calculated table that houses key performance indicators (KPIs), aggregations, and other measures for reporting and analysis.

Calendar Table: A date dimension table, a vital component for a time series analysis and trend identification, created with DAX as a new table for only this MS power BI usage.

This content model supports several functions:

- Relationships between tables enable comprehensive cross-analysis, such as correlating user engagement with content performance.
- The user hierarchy allows for organizational analytics, breaking down the metrics by job roles or regions.
- The Calendar table integration facilitates trend analysis over time, which is key for tracking performance changes and seasonal impacts.
- The association of Assets with Tags and Categories enables detailed content performance reviews, understanding which content types perform best.
- Tables like Devices and Shares provide a basis for analyzing the platforms on which the content is viewed and how it is distributed.
- The schema provides a 360-degree view of the data, from user information to content interaction.
- The normalized structure allows for scalability, accommodating growing amounts of data without significant rework.
- The model's design facilitates the easy addition of new data sources or tables, allowing for flexible adaptations to new business requirements.
- Efficiently designed relationships and indexing can lead to faster query performance in Power BI, making the data model highly performant for large datasets.
- With calculated measures and comprehensive table structures, the model supports robust analytical capabilities, from KPI tracking to deep-dive analytics.

This data model is designed to provide a detailed, query-efficient framework for business intelligence tasks, and allow for complex analyses and strategic insights generation.

Coach Data Model

The second part of the data modeling chapter, the coach module is focused on a view that pertains to a learning management component in Showpad. This data model is designed to analyze and track the educational aspects of a coaching system.

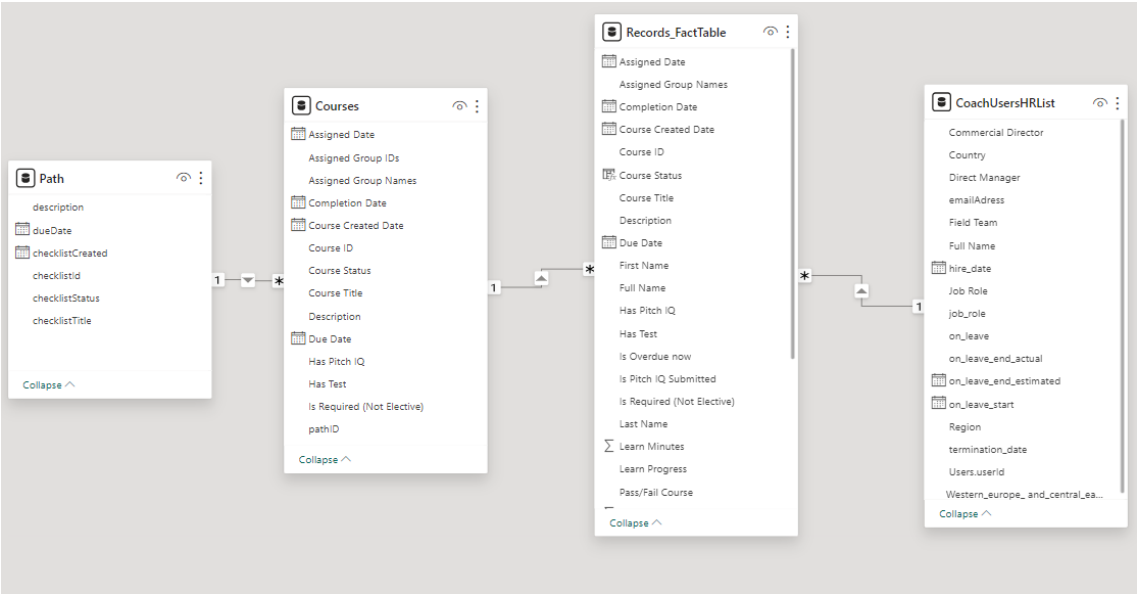
Description of the Data Model:

Records_FactTable: This fact table captures detailed records of user interactions with the courses, like completion status, course performance, and learning progress.

Courses Table: Holds course-related data such as course IDs, titles, and descriptions, as well as dates related to course assignment and completion. This can be pivotal for tracking course engagement and progress.

CoachUserHRList: It is a dimension table containing coach-specific user information, including their roles and hierarchy within the commercial division.

Figure 17 - Data Model: Coach



Source: The Thesis author's processing

The model can be used to monitor the progress of learners through various courses, based on completion rates and other performance metrics. Course Engagement Analysis: With data on assigned and due dates, the analysts can measure engagement levels and timeliness of course completion.

Coach Performance Evaluation: By linking user data to course records, the model can assess the effectiveness of coaches in managing and facilitating courses.

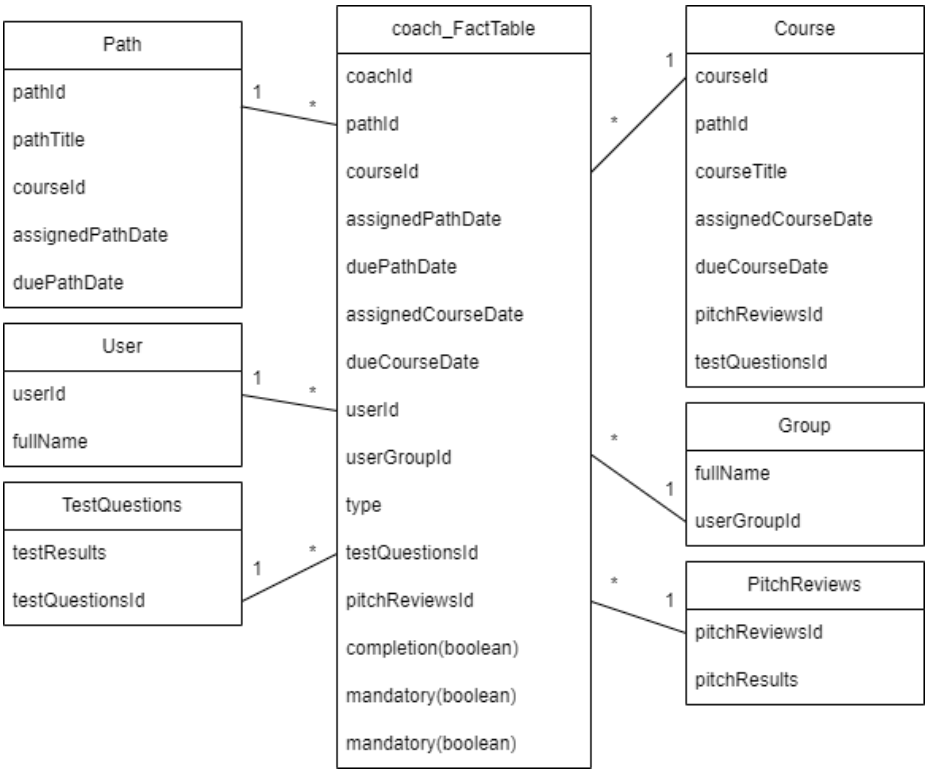
Advantages of this schema:

- Tailored to capture coach interactions, the schema enables deep insights into learning patterns.
- The inclusion of time-related fields like due dates and completion dates facilitates time-to-completion analyses and deadline adherence reviews.
- The schema supports role-specific analysis, which can be critical in understanding the performance and impact of coaches on the learning process.

To solve the data structure of the coach, it was chosen to create a star schema schema. According to the basic principles of creating data models in databases and also in business intelligence, a star schema is usually chosen for advantageous behavior. The modifications listed below should be considered. Additional tables must be created between the M: N links. Next, it would want to evaluate what fact table to create, where to identify the key columns to merge (ChecklistId, CheckItemId), and then add indexes as record IDs. Furthermore, it is necessary to focus on duplications and their modifications, advantages, and disadvantages. For the next part of the data, join queries, joins, etc. will be needed to complete the table. Upon initial inspection, the restructured data model exhibits a commendable level of organization. Its design is a significant stride toward adopting a star schema. Here are several key observations related to this newly proposed diagram.

The Coach Fact Table is in the middle, indicating that it serves as the central table commonly used in the stars and snowflakes scheme. each trainer interaction would be treated as a record that could be separated by a kind or type like course open, path open, test open, or test completed, and this would always be matched by time and date + course results, which would be factual and numeric data.

Figure 18 - Data Modelling: A Refresh of Star Schema for Coach Data



Source: The Thesis author's processing

User, Path, Course, Group, PitchReviews, and TestQuestions are dimensions.

Normalization must be mentioned because the foreign keys in the model are relations pointing from the fact tables to dimensions. in star schema mostly the dimensions are denormalized, but not in snowflake. It depends, that is, which way the database developers of Showpad take here.

The last and essential point here is relationships that are clean and clear, avoiding many-to-many relationships, which is a best practice and optimal solution.

The data models in this chapter - one focused on content interaction and the other two on the educational Showpad coach module - provide comprehensive frameworks for analysis within a coaching and learning environment. The first model offers a multifaceted view including user interaction, content categorization, and device usage. The second model describes the data, from course assignments to learning outcomes. Both models leverage the strengths of Power BI's data modeling to extract meaningful insights from complex datasets,

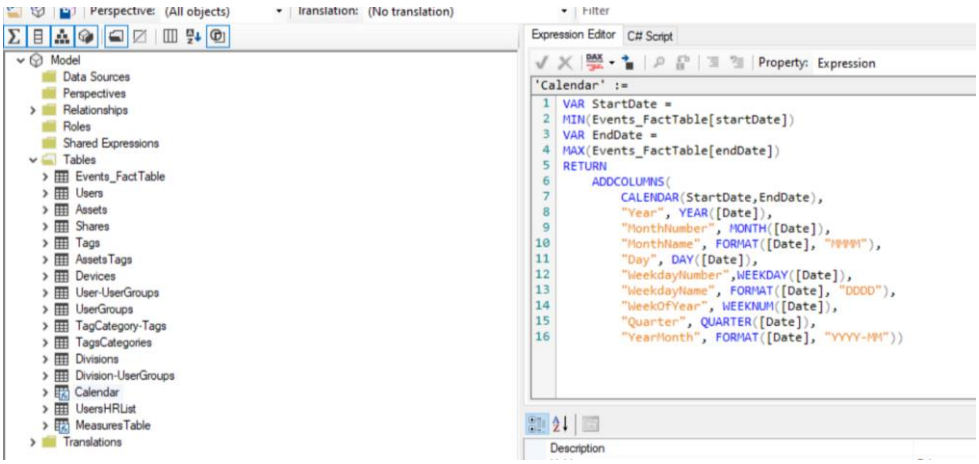
highlighting the versatility and depth of analysis possible within this BI tool. Together, they represent a cohesive analytical strategy for harnessing Showpad data to drive decision-making and business strategy in content-driven platforms. Within this chapter, a restructuring of the data model was also created, which would allow for more efficient querying and data analysis than a highly normalized transactional database. It supports typical data warehouse use cases, where the fact table contains measurable, numeric data and the dimension tables contain descriptive attributes to filter and aggregate the fact table.

5.2.2.3 Transformation with Tabular Editor

While Power BI Desktop is robust, it lacks some advanced features. Tabular Editor 2 fills these gaps and allows users to perform tasks that are not directly achievable in Power BI. After data models are created, a PBI Showpad file can be opened in the editor to allow data editing, DAX scripting, model comparison and optimization, benchmarking, and more.

Below is an example of creating a new calendar table to use a uniform time. In the image, you can see that on the right is the Expression Editor for DAX scripts. Create a new table called Calendar; the DAX script defines various date-related variables and returns them in columns, allowing us to use a consistent time format throughout the dashboard and maintain consistency and granularity.

Figure 19 - Tabular Editor Transformation: Calendar Table



Source: The Thesis author's processing

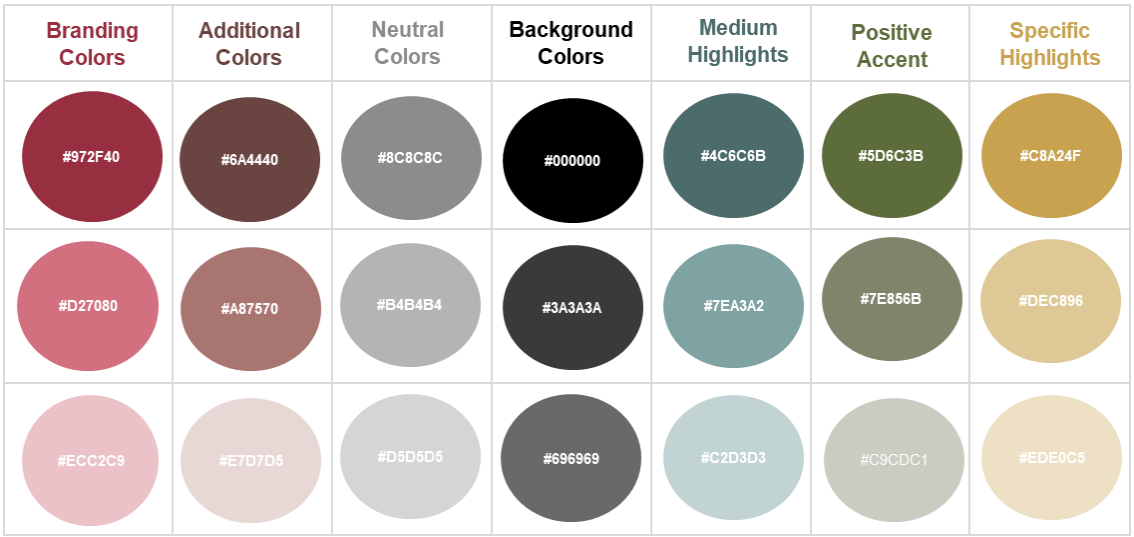
In summary, Tabular Editor is a valuable investment for anyone working with DAX in Power BI. Its productivity features justify the cost, and once you experience its benefits, there is no turning back. For creating measures, new columns, and new tables it is a great user-friendly tool.

5.2.3 Crafting Showpad Dashboard

The development of the dashboard has reached the stage where one needs to start thinking about the design. It is essential to keep in mind the company's existing brand identity. Although the brand's official colors are a key part of the company's image, for this project their palette will be used more as an inspiration. This approach ensures that the branding remains confidential while respecting the original structure of the company's color scheme. The main color of the company is a shade of red and grey.

The color palette created for this Showpad project presents a diverse array of hues that reflect a modern and versatile design aesthetic. With a base of earthy tones, accented with pops of more vibrant colors, as well as an underground burst of both cool and warm tones, this palette is suitable for a variety of applications. Another part of the branding is the logo, to which the palette is adjusted to complement this element.

Figure 20 - Dashboard Color Palette



Source: The Thesis author's processing

The decision to integrate the brand red into the dashboard as an alerting mechanism, rather than solely for branding purposes, is a strategic choice. By doing so, it leverages the strong association users have with the brand color to ensure maximum impact and immediate recognition when attention is required. This palette is a versatile toolkit, with balanced neutrals and eye-catching accents. It highlights a nuanced approach to design that prioritizes both user engagement and operational efficiency. In addition, a font selection should be considered as part of the dashboard design. For our diverse color palette, Segoe UI is the best fit due to its modern and clean design. It would be ideal to use it as shown below:

- **Headings (Titles, cards, KPIs) that will be bold for emphasis.**
- Also, the font for the body (text boxes, labels, labels) will be regular UI Segoe.
- **Segoe UI or Segoe UI Semi-bold is suitable for Slicers and Filters.**

Keeping one typography across the dashboard will create a uniform and professional look. Strategically use size, weight, and color to create a visual hierarchy and guide the user's eye across the dashboard.

The strategic use of these elements enhances the user experience and dashboard effectiveness. The goal is to create a visually appealing and efficient dashboard. Thanks to these insights, it is now possible to design efficient and aesthetically pleasing dashboards.

5.2.4 Dashboard Final Layout

This chapter delves into the complexity of the dashboard layouts used for monitoring and evaluating the engagement and effectiveness of Showpad's features across the various modules. This chapter will demonstrate the detailed visualization strategies employed to represent usage statistics, content interaction, user participation, and completion rates within the Coach module.

5.2.4.1 Analysis Techniques

This chapter will detail the transition from data management to actionable insights, featuring data mining, analytics, and their application in decision-making processes. In the previous chapters whenever the data were extracted from the Showpad database using RESTful API, it falls under the realm of data analytics rather than data mining. This is because it is an analysis of known relationships and works with structured data to extract insights, make predictions, or inform decisions.

An application of descriptive analysis is obvious at first sight, where the focus is on analyzing historical data from Showpad, understanding patterns and trends, and summarizing data points for the dashboard. As a part of history mapping, expectations of data behavior in the future are also formed, which has not yet been defined as one of the requirements for the project and will therefore not be dealt with further in the thesis and this chapter.

5.2.4.2 Dashboard & Showpad Data Interpretation

Each section will articulate how these dashboards serve as critical tools for understanding user behavior, optimizing content, and enhancing training programs. Exploration will also emphasize the design considerations and data representation choices that make these dashboards an invaluable asset for decision-makers.

Showpad Usage – Commercial Operations Team

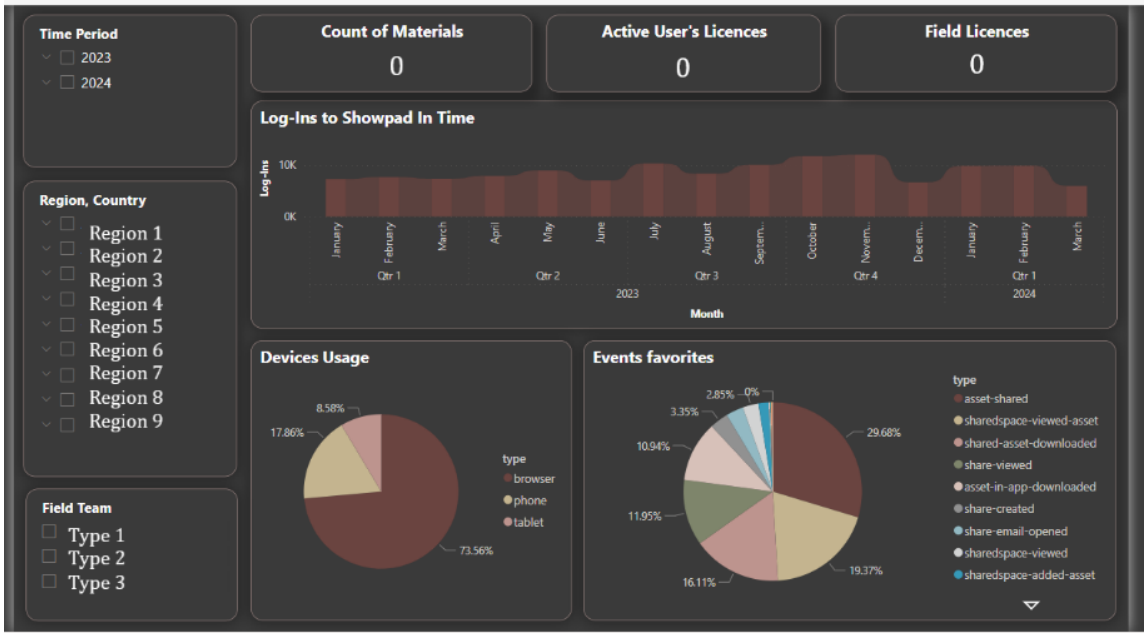
The "Showpad Usage" dashboard page offers a comprehensive view of Showpad utilization at the request of the commercial team that manages the Showpad platform. It shows key metrics such as the number of materials, active user licenses, and login frequency, providing immediate insights into engagement levels and resource distribution.

Showpad logins over time are effectively visualized using bar graphs. The chart displays login counts for specific months in 2023 and 2024,

allowing easy comparison. Notably, activity increased since the launch of the sales training coach module in July 2023.

The first chart at the bottom shows device usage, shedding light on user device preferences, which can be the basis for content format optimization and event planning. Here, the result is the browser, where most of the material is directed, but the phone also has a smaller representation, which must be considered, and the compatibility of the materials evaluated.

Figure 21 - Showpad Dashboard: Showpad Usage Page



Source: The Thesis author's processing

Next, there are popular events where the pie chart has again been chosen for its simplicity in displaying proportions. It represents the popularity of several types of events. The slices indicate the relative importance of individual types of actions, and it can be seen that the most common action is asset sharing, which corresponds to the company's strategy of using Showpad as a source of communication with clients.

Selectors and filter elements are further included in the dashboard, where these interactive elements allow users to customize the view. They are necessary for examining data based on different criteria (eg.: time, region, type of field team). Users can dynamically filter the data

and improve their understanding. Selectors improve user engagement and interactivity and are incorporated into all Showpad dashboard pages.

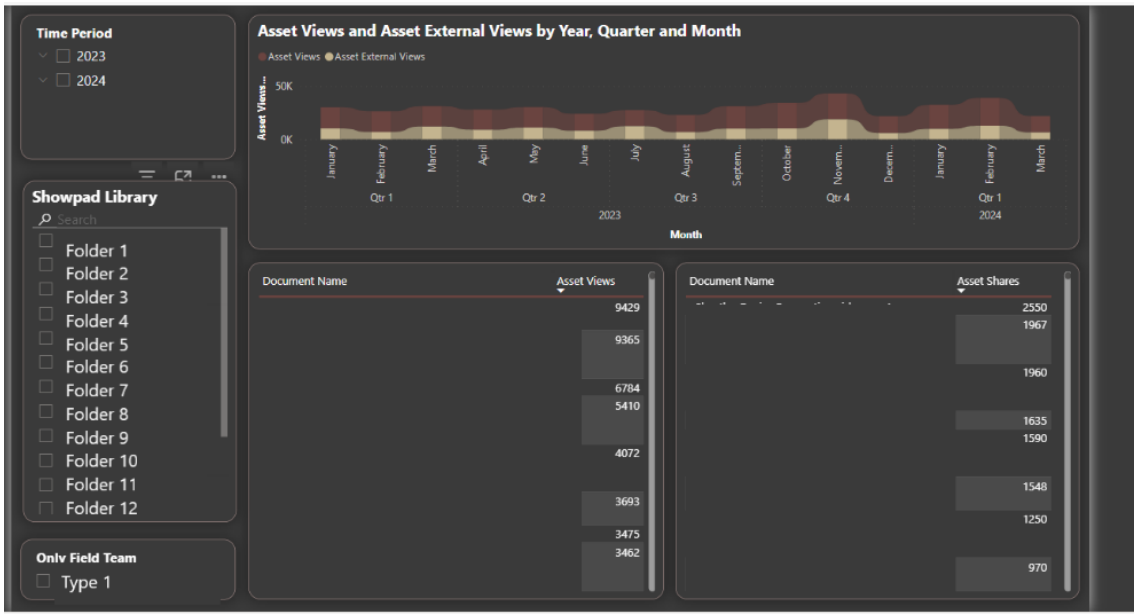
Showpad Content – Marketing Team

The "Showpad Content" dashboard page presents a rich interface for analyzing interaction with assets across the Showpad library and responds to a request from the marketing team to monitor interactions with materials.

Stacked area charts are effective for showing cumulative data over time. They allow viewers to see the total and the contribution of each category's asset views and asset external views. The chart effectively conveys the overall trend in views. It highlights the relative importance of each type of view. The upward trajectory since the launch of the sales training coach module in July 2023 is evident and verified again. It is an essential tool for content managers to understand which materials are resonating with users and external viewers.

On the left, the folder navigation pane provides an intuitive way for users to explore various categories or datasets. It is a simulated structured view of the Showpad Library. It is suitable for organizing and accessing specific subsets of data and it aids users in quickly locating and analyzing the performance of specific content topics. It enhances user experience and ease of use.

Figure 22 - Showpad Dashboard: Content Page



Source: The Thesis author's processing

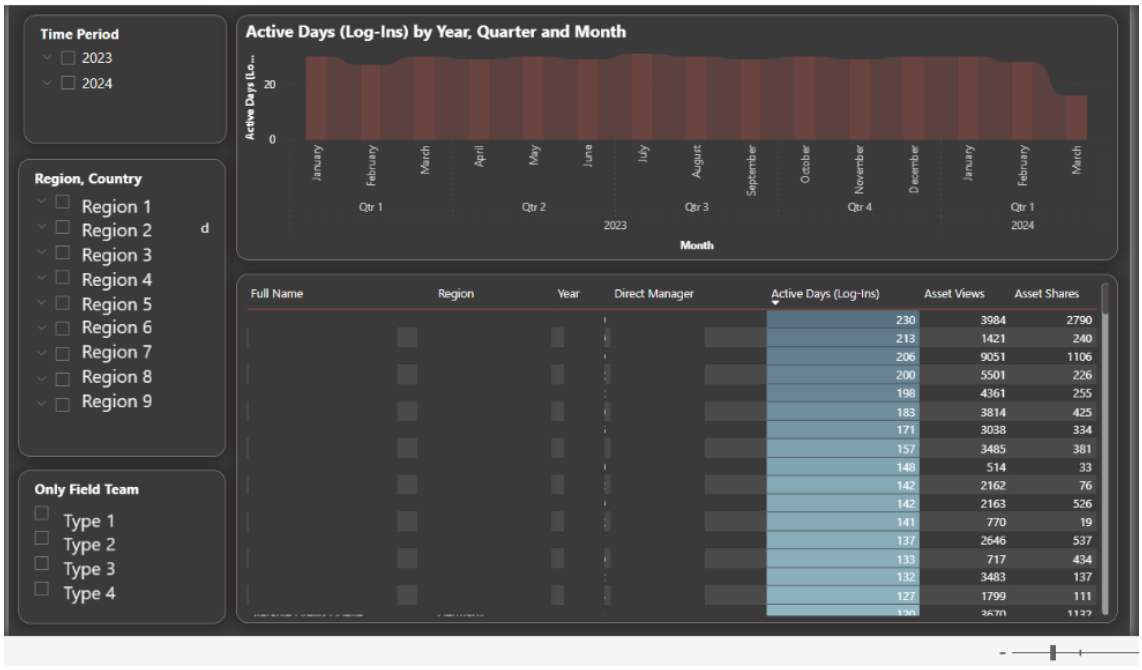
A detailed table offers a more in-depth look, usually, tabular representations are appropriate for displaying specific numerical data associated with individual items in this case the table provides detailed information about document names, views, and shares. It allows viewers to compare values directly. The marketing team can identify top-performing assets and assess their impact.

Showpad User Engagement

This third page of the "Showpad User Engagement" dashboard is focused on user requests from sales management. This page is focused on the activities of the sales team.

A bar graph dominates the visual space and displays the frequency of active days (logins) in different periods (years, quarters, and months), while the corresponding table on the right provides a detailed breakdown by region, manager, and total active days.

Figure 23 - Showpad Dashboard: Users Engagement Page



Source: The Thesis author's processing

The corresponding table at the bottom of the page provides a detailed breakdown by region, manager, and total active days. The spreadsheet also includes metrics such as asset views and shares that translate engagement into useful statistics.

Filters for the time, region, country, and field team type are prominently featured, allowing for tailored data exploration. This versatility ensures that the dashboard can adapt to the varied analytical needs of different users within the organization.

Showpad Coach

The "Showpad Coach" dashboard page is a centralized hub for monitoring and evaluating the coaching and learning process within an organization. It leverages data visualization to provide insights into learning engagement and progress across different paths and courses.

At the top center, there are three key metrics:

- Average Learning Time: Indicates the average time spent on courses (e.g., 13.94 minutes),
- Average Test Attempts: Reflects the average number of attempts made by learners (e.g., 1.90 attempts),
- Average Test Score: Shows the average score achieved in tests (e.g., 86%).

Those KPIs offer immediate quantifiable insights into the effectiveness of the coaching materials and strategies.

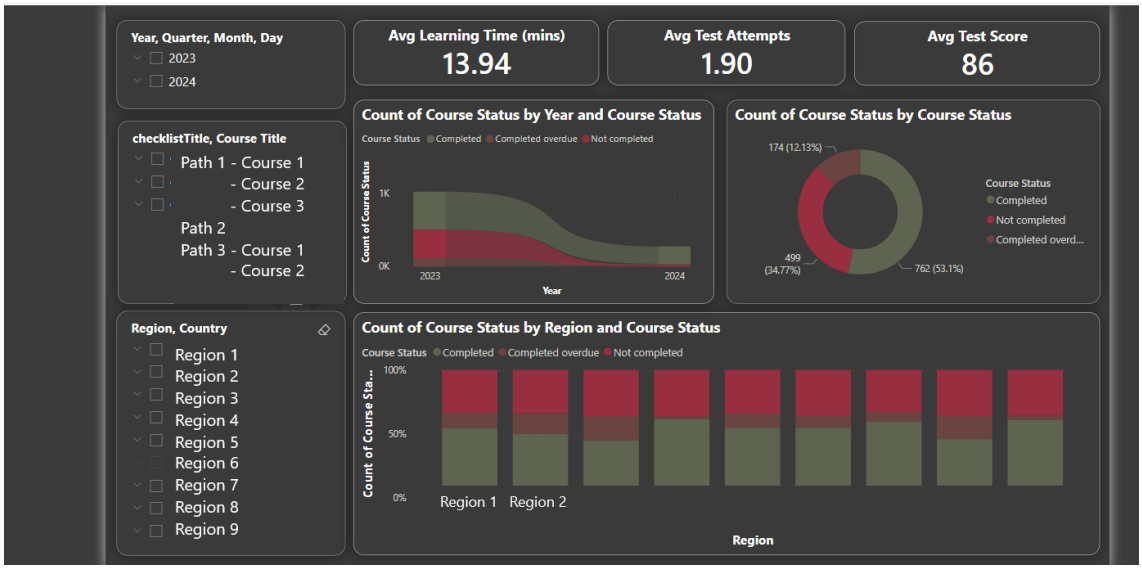
Below these, a Donut chart was chosen to visualize the count of course statuses, which represents the distribution of course statuses:

- Completed Overdue: Courses completed after the deadline,
- Not Completed: Courses not finished,
- Completed: Courses completed on time.

It is vital for identifying areas that may require additional focus or intervention because if the course is not met, a reminder can be sent to specific people who are not listed here, however, the data are prepared and included there for actionable steps.

Complementing the donut chart, a series of bar charts give a granular view of course status by region, allowing for regional performance comparison and targeted analysis. This can inform region-specific coaching interventions.

Figure 24 - Showpad Dashboard: Coach Page



Source: The Thesis author's processing

The dashboards described, such as the Showpad Usage, Content, and Coach dashboards, serve as critical tools for various teams within an organization to monitor and evaluate key metrics, user engagement, and the overall effectiveness of materials and strategies. Visual representations, including bar charts, pie charts, and stacked area charts, along with interactive filtering elements not only facilitated a detailed view of performance but also offered the versatility needed to tailor the analytical process to specific needs.

This chapter highlights the value of strategically designed dashboards in making informed decisions and implementing targeted interventions. As data continues to influence strategic decisions, the insights gained from these dashboards will undoubtedly prove critical to increasing operational efficiency, driving user engagement, and driving the organization toward its goals.

5.3 Dashboard Deployment

This chapter elaborates on one of the final stages of the BI lifecycle, the introduction of the Showpad Power BI dashboard, which is published in the company's workspace. This workspace serves as a centralized repository for the management and analytics teams to access up-to-date business intelligence reports. In this secure environment, the dashboard is carefully customized to support management decision-making processes. Currently, access to these panels is restricted to ensure that statistics and benchmarking remain confidential and only available to authorized personnel.

Since the dashboard was designed primarily for senior management, RLS is not a prerequisite at this stage of the application. The goal is to equip executives with deep knowledge and comprehensive analytical tools to support informed decision-making. However, the architecture of MS power BI Desktop gives the flexibility to adapt if the need arises to expand the approach to include a wider range of employee roles. Such flexibility ensures that RLS measures can be implemented to tailor data visibility to specific roles, protecting data integrity while facilitating a controlled but transparent flow of information.

5.3.1 Showpad Dashboard Testing

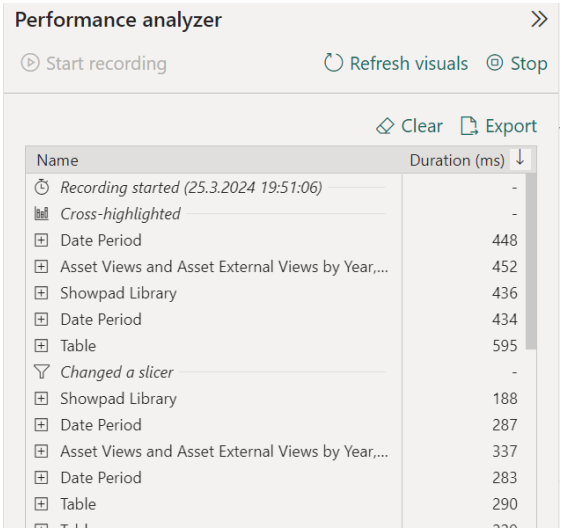
This chapter dives into the critical process of validating and testing the data and functionality presented on the Showpad dashboard. The chapter begins with a discussion of the data validation process where inconsistencies in activity data due to miscalculations were identified. The chapter also discusses dashboard performance evaluated using the MS power BI Performance Analyzer.

Data Accuracy: During the data validation process, the sales numbers presented on the dashboard were compared to the original data in the Showpad platform. It has been found that there are irregularities in the activity. An investigation revealed that the problem was caused

by a miscalculation where Showpad was considering the entire year and only the working days were being considered within the project. After fixing this issue, the activity data in the dashboard matched the source data. When checking the completeness of the data, it was found that for some countries the names of the sellers do not match, or their data is missing. After investigation, it was found that some of them were the new starters in March and not yet updated in an HR list. The adjustments have been made and a date has been set that every week on Monday the data of the entire dashboard will be updated, and the new employees on the 1st of the month and the 15th of the month.

Performance analysis: Below is a screenshot of the power BI performance analyzer that evaluates the speed of visualizations. The durations vary from 198 ms to 595 ms for the content page. It is a satisfactory performance, typical for most visuals. Up to 1000 ms, is still acceptable. Between 1000-2000 ms will already load longer and it would need attention. More than that means a deficient performance. The performance of this page is therefore satisfying.

Figure 25 - Performance Analyzer



Source: The Thesis author's processing

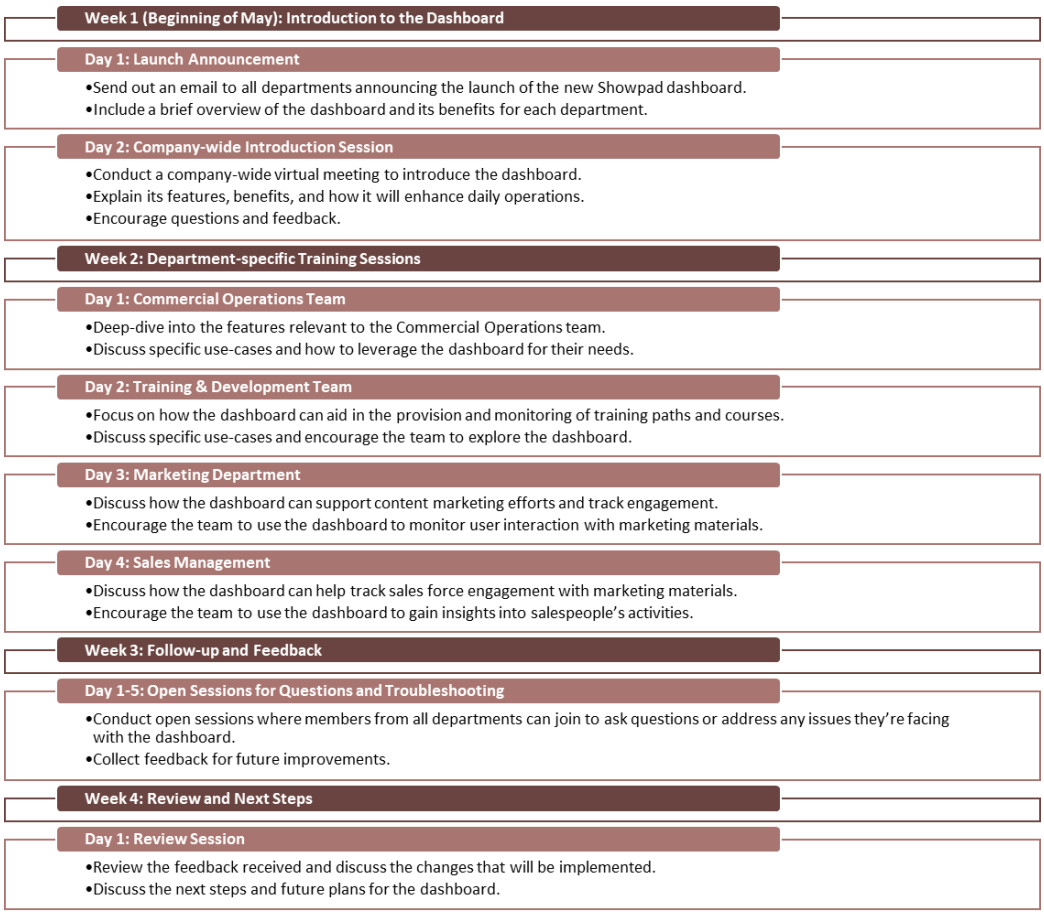
In this chapter, she dives into the critical process of validating and testing the data and features presented in the Showpad dashboard. The journey began with a thorough examination of the data validation process where discrepancies

in activity data due to miscalculations were carefully identified. In addition, we examined the performance of the dashboard and evaluated it using the power BI performance analyzer.

5.3.2 Showpad Dashboard Adoption & Training

In this chapter, a plan was created for the strategic introduction of the new Showpad dashboard within the company. It will dive into the specific activities planned for each department to facilitate dashboard adoption over the three weeks. The adoption strategy is carefully structured to ensure a smooth transition for all teams. Starting with an enterprise-wide announcement, followed by an introductory meeting, and continuing with departmental training, this rollout is designed to promote a thorough understanding and effective use of the dashboard.

Figure 26 - Dashboard Adoption Plan



Source: The Thesis author's processing

The outcome of this adoption plan aims to ensure that each segment of the audience fully understands and utilizes the Showpad dashboard to enhance their daily operations. By tailoring training to audience needs and providing open sessions for questions, the commercial operations team and company expect to see improved efficiency, better data-driven decisions, and increased transparency across all teams. There are, however, inherent risks in any technology adoption process such as resistance to change or misunderstandings, which is why the adoption is prepared and planned in such a comprehensive way so that the risks are the lowest and the chances of using the dashboard daily are the highest.

5.3.3 Showpad Dashboard Maintenance & Updates

To upkeep of the BI dashboard is a critical aspect that ensures its continuous accuracy, and relevance. This chapter outlines an approach to the regular maintenance and updates of the Showpad dashboard, encompassing data sources, software updates, and integration of new features.

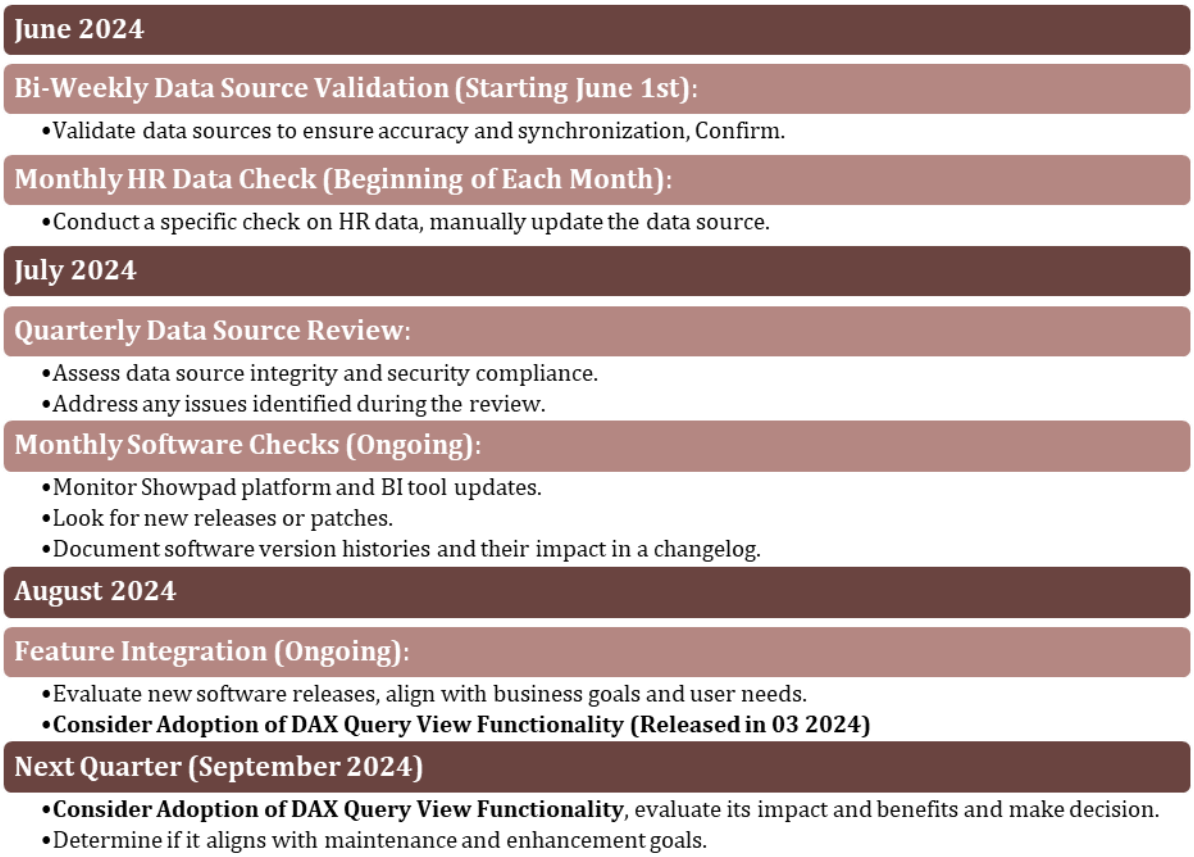
A data source check is scheduled for bi-weekly validation to ensure data validity and correct sync. Once a month, at the beginning, there will be a particular check of HR data and manual updates to the source. Quarterly reviews are done to assess data source integrity and security compliance. Monthly software checks are run for new releases or patches for the Showpad platform and used BI tools. A changelog to document software version histories is maintained and the impact of each update is watched.

Feature integration evaluates new software releases, and quarterly strategic meetings align news to business goals and user needs. One such example is the new DAX Query View feature in Power BI Desktop, released in March 2024 and designed to improve data modeling and analysis. In the practical part, the Tabular Editor is used, which offers a more efficient way of handling and managing calculations. It simplifies the work of data analysts by offering improved data exploration, and optimized performance, and serves as a learning

and debugging tool. In the next maintenance strategy session, next quarter, the adoption of this function will therefore be considered.

For a summary of the above-mentioned dates and better navigation, a Showpad dashboard maintenance and updates timeline is created as shown below.

Figure 27 - Showpad Dashboard Maintenance & Updates Timeline



Source: The Thesis author's processing

6 Results, Findings, Challenges, and Limitations

This chapter serves as a culmination of the thesis's practical part, which stands on a solid foundation in the theoretical part. It is a synthesis of results, findings, challenges, limitations, and at the same time solutions, encountered during the endeavor of fulfilling the thesis goals. The main result is a functional Showpad dashboard. Through iterative design, development, and deployment, the dashboard now provides answers to questions that were collected in BRD. Other various minor results and solutions were used and applied in other parts of the work, which are presented and promptly described further in this chapter.

The first key finding was, that existing literature lacked a comprehensive Business Intelligence (BI) lifecycle workflow. To address this gap, a new, meticulously researched BI lifecycle was crafted, incorporating insights from diverse sources, and aligning with current market needs, including security solutions. This comprehensive framework now serves as a practical guide for BI implementation.

The first key challenge and limitation occurred in the testing phase of the Showpad RESTful API, where certain data in their test environment, particularly from the Coach module, was missing. This absence hindered access to example responses with company data, leading to an unnecessary extension in data extraction time.

In the search for the most efficient solution, it was also discovered that there was no Showpad OpenAPI environment to provide a standardized interface to help developers understand and integrate with the API.

Other key findings emerged in the data extraction and understanding phase, the content part of the data is well structured and user-friendly for further development. However, it lacks the 'field' parameter, which is crucial for extracting only the necessary data. This parameter is set for coach data only.

The subsequent challenge was that, unlike the content data in Showpad, the coaching data was not ready for further development of BI and tools. The critical problems encountered in the work are the absence of a defined fact table with dimensions, inconsistent data granularity, redundant or duplicate data, and many-to-many relationships. This discovery flows with the possibility of comparison with the data structure of the content. Here, a data model in a star scheme was proposed as a solution, which would be more effective for this case according to theoretical knowledge, however, due to the complexity, it would be best to make changes in the data source and not in the MS power BI desktop project.

Another key finding related to the design of the dashboard, as the company's main color is red, which is often evoked by the human brain as something negative or alerting. So, the decision was made to integrate the color red as a warning mechanism instead of a brand identity. It emphasizes a nuanced approach to design that prioritizes user understanding and operational efficiency.

This section summarized the key results. He always emphasizes the situation in which the discovery occurred, what its consequences are, and how it influenced the work.

7 Conclusion

The thesis "Dashboards Automation" described a complex business intelligence (BI) and automation environment within the enterprise workspace. At its core, the master thesis goes through a series of processes, steps, and methods to achieve the key objective. This objective was set to develop, design, and deploy a comprehensive dashboard centered around the Showpad platform for an enterprise. This goal was achieved, and its fulfillment was supported by the corresponding theoretical knowledge.

In the theoretical section, the challenges of business intelligence were penetrated and its necessity as the backbone of a modern competitive strategy was revealed. Exploring the multifaceted nature of BI has illuminated the fundamental methods and frameworks that are changing the evolution of BI solutions. A methodical approach to BI lifecycle management was developed, data extraction using Python and transformations using Power Query and Tabular Editor provided a rich theoretical background for the subsequent practical part and visualization in MS power BI. A critical analysis of the dashboard development process highlighted the importance of synergy between user-oriented strategy, technology, analysis, design, and steps used to ensure optimized dashboard performance.

The practical basis of this work was not only the application of the theory but the discovery of its impact. By carefully detailing the development of the Showpad dashboard, this work illustrated the benefits of user-centric strategies. Realization of the key objective began with requirements gathering using a business requirements document, judicious technology mapping, data extraction, and transformation into actionable knowledge. This iterative refinement is the real essence of the work– a testament to the dynamic and ever-evolving domain of business intelligence.

Key issues were identified in the initial testing phase due to a lack of standardization in the Showpad RESTful API and a lack of data availability from the Coach module. It was also noted that the key "field" parameter needed

for targeted data extraction was missing from the data. Despite the well-structured and user-friendly content data, coaching data was not immediately suitable for BI development, lacked a defined fact table with dimensions, and presented data granularity and redundancy issues. The thesis proposed a star schema data model as a solution to these problems. In addition, the importance of color in the dashboard design was discussed and it was decided to use red as a warning signal rather than part of the brand identity to emphasize user understanding and operational efficiency.

In conclusion, this work produced a deep and detailed exploration of business intelligence dashboard automation, culminating in the successful development and deployment of the Showpad automated dashboard. It demonstrates the power of bridging theory with practice and highlights the transformative potential and adoption of the Showpad dashboard. While the journey ran into several problems, such as data structure limitations and the need for a more standardized API, the work also provided insight into improving user interface design by integrating visual cues for better user interaction. The proposed star schema model marks a leap towards solving data granularity issues, reflecting a deep understanding of the BI domain. The ambition of this work is not only to contribute a meaningful product in the form of an automated dashboard but also to pave the way for continued evolution in the way businesses approach business intelligence. It represents a scientific bridge between the spheres of academic research and practice. Organizations in today's environment are forced to adopt the digital paradigm with agility and foresight. Their task is to deftly navigate the complex terrain of today's data-driven world and demonstrate precision in their actions.

8 Bibliography

Analytics, Databox. (2023, 9 8). *18 Best Dashboard Examples for Any Business*. Retrieved from Databox: https://databox.com/best-dashboards-examples

Baktawar, R. (2022, 4 14). *Business Intelligence Lifecycle.* Retrieved from Sctiger: https://www.sctiger.com/en/blog-news/business-intelligence-lifecycle/

Big Data Framework. (2024, 1 8). *Top Business Intelligence Trends 2023 | Sneak Peek*. Retrieved from Big Data Framework: https://www.bigdataframework.org/knowledge/top-business-intelligence-trends-2023-sneak-peek/

Blake, S. (2021, 6 8). *Use cases vs. user stories*. Retrieved from EasyAgile: https://www.easyagile.com/blog/use-cases-vs-user-stories/

Calzon, B. (2023, 12 20). *Business Intelligence Trends*. Retrieved from Datapine: https://www.datapine.com/blog/business-intelligence-trends/

Croft, D. (2023, 9 24). *Moscow Method*. Retrieved from LearnLeanSigma: https://www.learnleansigma.com/guides/moscow-method/

Dana Miranda, C. B. (2022, 8 7). *Principles of change management*. Retrieved from Forbes: https://www.forbes.com/advisor/business/principles-of-change-management/

Dol, Q. (2021, 1 5). *22 Companies Spearheading Digital Innovation in Their Industries in 2021*. Retrieved from Built In: 22 Companies Spearheading Digital Innovation in Their Industries in 2021 | Built In

Duncan, D. (2021, 4 8). *Webinar Replay: Designing for Data*. Retrieved from Interworks: https://interworks.com/blog/2021/04/08/webinar-replay-designing-for-data/

EDITORIAL TEAM. (2024, 3 22). *Power Query Tutorial*. Retrieved from Under30CEO: https://www.under30ceo.com/terms/power-query-tutorial/

EffectiveSoft Team. (2023, 12 14). *Business intelligence life cycle: from definition to full process description*. Retrieved from EffectiveSoft:

https://www.effectivesoft.com/blog/business-intelligence-life-cycle.html#business-intelligence-life-cycle

Fastercapital Team. (2024, 3 22). *Blockchain prototype: The Benefits of Developing a Blockchain Prototype for Supply Chain Management*. Retrieved from Fastercapital: https://fastercapital.com/content/Blockchain-prototype--The-Benefits-of-Developing-a-Blockchain-Prototype-for-Supply-Chain-Management.html

Ge, C. L. (2023). The Impact of Digital Technology Innovation Network Embedding on Firms' Innovation Performance: The Role of Knowledge Acquisition and Digital Transformation. *SUSTAINABILITY*, 15.

Geeks For Geeks. (2023, 5 8). *Star schema in data warehouses modeling*. Retrieved from Geeksforgeekds: https://www.geeksforgeeks.org/star-schema-in-data-warehouse-modeling/

Hietpas, S. (2021, 1 19). *Tabular Editor Model Development Advantages*. Retrieved from Corebts: https://corebts.com/blog/tabular-editor-model-development-advantages/

ChatGPT-4. (2023, 1 11). *OpenAI*. Retrieved from CHATGPT: https://chat.openai.com/chat

Chaudhari, P. (2023, 11 28). *What is Extract/Transform/Load (ETL)?* Retrieved from Stack overflow: https://stackoverflow.com/questions/3392808/what-is-extract-transform-load-etl

Chiarelli, A. (2022, 2 2). *Encoding Enrcyption Hashing*. Retrieved from auth0: https://auth0.com/blog/encoding-encryption-hashing/

IBM Team. (n.d.). *What is file storage?* Retrieved from IBM: https://www.ibm.com/topics/file-storage#:~:text=File%20storage%2C%20also%20called%20file-level%20or%20file-based%20storage%2C,In%20file%20storage%2C%20data%20is%20stored%20in%20files.

Intellipaat. (2023, 12 5). *Supply chain analytics*. Retrieved from Intellipaat: https://intellipaat.com/blog/supply-chain-analytics/

Kanade, V. (2023, 9 5). *What is decriptive analytics*. Retrieved from Spiceworks: https://www.spiceworks.com/tech/data-management/articles/what-is-descriptive-analytics/

Kapoor, A. (2023, 9 19). *The Lifecycle of BI: From Development to Ongoing Support*. Retrieved from Medium: https://medium.com/nerd-for-tech/the-lifecycle-of-bi-from-development-to-ongoing-support-6b4f6cf3664a

Keydifferences. (2023). *Difference between questionnaire and interview*. Retrieved from Keydifferences: https://keydifferences.com/difference-between-questionnaire-and-interview.html

Kime, C. (2023, 12 7). *Encryption*. Retrieved from Esecruityplanet: https://www.esecurityplanet.com/networks/encryption/

Knezovic, A. (2024, 3 1). *What is a Dashboard? Definition, Types, and Examples*. Retrieved from Blog Udonis: https://www.blog.udonis.co/analytics/dashboard

Kusnetzky, D. (2018, 3 2). *Big Data and Business Intelligence: What's the Difference?* Retrieved from Dataversity: https://www.dataversity.net/big-data-business-intelligence-whats-difference/

Lupenko, V. (2022, 8 23). *Why A Tech-First Strategy Is Key To Success And How To Implement It In Your Business*. Retrieved from Forbes: https://www.forbes.com/sites/forbesbusinesscouncil/2022/08/23/why-a-tech-first-strategy-is-key-to-success-and-how-to-implement-it-in-your-business/?sh=522d4863780d

McGill, J. (2023, 10 24). *Data Cycle.* Retrieved from Leadfuze: https://www.leadfuze.com/data-cycle/

Medium - pandaquests. (2022, 12 28). *JSON for beginners*. Retrieved from Medium: https://medium.com/codex/json-for-beginners-a7308ebdb34d

Microsoft. (2023, 10 11). *Introduction to dashboards for Power BI designers*. Retrieved from Learn Microsoft: https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards

Microsoft. (2024, 3 4). *Service roles new workspaces*. Retrieved from Learn Microsoft: https://learn.microsoft.com/en-us/power-bi/collaborate-share/service-roles-new-workspaces

Misiuro, A. (2022, 11 29). *What are the Basic Business Intelligence Techniques?* Retrieved from Synder: https://synder.com/blog/what-are-the-basic-business-intelligence-techniques/#:~:text=Business%20intelligence%20techniques%20fall%20into%20three%20categories%3A%20data,data%20and%20storing%20it%20in%20a%20centralized%20database.

Mukhtiar, M. (2023, 4 14). *5 Key Metrics for Measuring the Success of Your Business Intelligence Strategy*. Retrieved from Medium: https://medium.com/@rao.mohsin.54/5-key-metrics-for-measuring-the-success-of-your-business-intelligence-strategy-24df35afa97a#:~:text=By%20tracking%20these%20five%20key%20metrics%20%E2%80%94%20data,your%20BI%20solution%20and%20identify%20opportunities%20f

Ninja, N. (2023, 6 6). *https://letsdatascience.com/scaling-and-normalization/*. Retrieved from Let's Data Science: https://letsdatascience.com/scaling-and-normalization/

Patel, K. (2023, 10 16). *What is a dashboard? Definition, uses, types, and examples*. Retrieved from Thoughtspot: https://www.thoughtspot.com/data-trends/dashboard/what-is-a-dashboard

PMI Team. (2022, 2). *What is Agile?* Retrieved from Project Management Institute: https://www.pmi.org/disciplined-agile/agile/whatisagile#:~:text=As%20a%20way%20of%20working%2C%20agile%20is%20an,agile%20teams%20deliver%20work%20in%20small%2C%20consumable%20increments.

Postman Team. (2022, 6 6). *Different types of APIs*. Retrieved from Blog Postman: https://blog.postman.com/different-types-of-apis/

Qlik. (2022). *What is Business Intelligence?* Retrieved from Qlik: What is Business Intelligence? Strategies & Best Practices | Qlik

REHAN AKBAR, M. H. (2008, 8 20-22). *Requirement Gathering and Tracking Process for Distributed Agile.* Retrieved from Researchgate: https://www.researchgate.net/profile/Rehan-Akbar-2/publication/234816539_Requirement_gathering_and_tracking_process_for_distributed_agile_based_development/links/5be16b03299bf1124fbe479c/Requirement-gathering-and-tracking-process-for-distributed-agile-bas

Severance, C. (2009). *Phyton for everybody: Exploring data in phyton 3.* Creative Commons.

Shatby, S. E. (2023, 5 26). *Data Types Complete Guide*. Retrieved from 365 Data Science: https://365datascience.com/trending/data-types-complete-guide/#1

Showpad Team. (2023, 1 29). Retrieved from Showpad: https://www.showpad.com/solutions/

Showpad Team. (2023). *API Fundamentals*. Retrieved from Showpad Developer: https://docs.api.showpad.com/docs/apis/concepts/fundamentals

Schwartz, B. (2022, 9 21). *Business Requirements document*. Retrieved from Project Manager: https://www.projectmanager.com/blog/business-requirements-document

Schwartz, B. (2022, 9 21). *Business Requirements Document*. Retrieved from ProjectManager: https://www.projectmanager.com/blog/business-requirements-document

STAMFORD, C. (2021, 4 28). *Gartner Forecasts Worldwide Hyperautomation-Enabling Software Market to Reach Nearly $600 Billion by 2022*. Retrieved from Gartner: https://www.gartner.com/en/newsroom/press-

releases/2021-04-28-gartner-forecasts-worldwide-hyperautomation-enabling-software-market-to-reach-nearly-600-billion-by-2022

Stedman, C. (2021, 12). *Data Modeling*. Retrieved from TechTarget: https://www.techtarget.com/searchdatamanagement/definition/data-modeling

Sturtz, J. (n.d.). *Basic Data Types in Python*. Retrieved from Real Phyton: https://realpython.com/python-data-types/#boolean-type-boolean-context-and-truthiness

Tabular Editor. (2024, 1 19). Retrieved from SQLBI: https://www.sqlbi.com/tools/tabular-editor/

Taylor, D. (2024, 2 24). *Star Schema vs Snowflake Schema – Difference Between Them*. Retrieved from Guru99: https://www.guru99.com/star-snowflake-data-warehousing.html

Taylor, S. (n.d.). *The Business Intelligence Cycle*. Retrieved from CFI: https://corporatefinanceinstitute.com/resources/business-intelligence/bi-cycle/

Wikipedia. (2023, 10 30). *Cardinality - Data Modeling*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Cardinality_(data_modeling)

work, T. a. (n.d.).

Y, J. A. (2024, 1 2). *Power BI Query*. Retrieved from Wallstreetmojo: https://www.wallstreetmojo.com/power-bi-query/

9 List of Figures

10 List of Codes

11 Attachments

Attachment 1 - Business Requirement Document

Category	Subtopic	Questions	Respondents
Project	Overview	What is the general scope of this project?	Showpad dashboard development project
	Goals	What are the specific objectives we aim to achieve with this project?	
	Owner	Who is responsible for overseeing and managing this project?	Viktorie Láchová
	Timeline	When is the expected started and completion date for this project?	1.2.2024 - 31.3.2024
	Budget	Funding Sources - Are there specific funding sources or budgets allocated for this project?	Secured funding of $10000
	Stakeholders	Audience - Who is the primary audience for this dashboard?	Commercial operations, Sales management, training & development team, marketing team
Purpose	Dashboard Topic	What main topics or data areas should the dashboard cover?	Commercial Operations - Showpad platform Usage, Marketing content engagement, Sales - Field team engagement, Coach trainings overview
	Business Questions	Business Questions - What are the key business questions the dashboard needs to answer?	Commercial operations - Showpad Usage - (Logins), What is device usage? Marketing Content – What materials were most popular in the last 14 days? What views and other interactions do the materials for the new product have? Who looked at these materials from our field? Field Team Engagement - How often is a team active on the Showpad? Who among them is the most active and what materials do they interact with? Coach Training - What is the average completion of training for the year 2023? Which region or country has the greatest engagement in coach training?
	KPIs & Metrics	KPIs/ Metrics - What are the key performance indicators (KPIs) or metrics that need to be tracked?	Field Engagement Rate % - if a Showpad user is active at least once a week, he is considered an active user. When logging into the system once a week, the user will have 100% activity for the week. Average views per user - How many views individual materials have and how many people have seen these materials. Coach Metrics - measurement of average involvement in coach trainings. Field Updates Participation % - What percentage attended the meeting, always calculated from the core time on the call, with full participation of 100%.
Data Discovery	Single vs Multiple Sources	Is the data coming from a single source or multiple sources?	Multiple
	Data Sources	What is/are the source(s) of data for the dashboard?	Showpad Database, Excel files
	Data ownership	Who owns the data sources, and are there any restrictions on their use? Contact?	Showpad platform are owners - database - API - https://docs.api.showpad.com/
	Data Location	Where are the data sources located (on-premises, cloud, third-party)?	
Data Interpretation	Data Types	What types of data (quantitative/qualitative) need to be represented?	Both
	Design	Are there specific branding guidelines (colors, fonts, logos) to follow?	Company Branding & Colors
	Tabs/Sections	How many tabs or sections are envisioned for the dashboard?	Marketin Content, Field Showpad Engagement, Coach Engagement, Field Updates, 5 - Overview Page - Per user all details
	Interactivity Features	Are there specific interactivity features needed (e.g., filters, drill-downs)?	filters -time & regions
	Visualization Preferences	Are there preferences or requirements for specific types of visualizations (charts, graphs, tables)?	They can all be, just stick to easily understandable concepts to the data.
Safety & Security	Row-Level Security (RLS)	Is row-level security required for any of the data/views?	Can be, but its not neseccary- Employee Role, they can't see their management activities or their colleague.
	Data Sensitivity	Are there any sensitive data elements that require special handling or compliance measures?	People's names, their manager and role are not sensitive data. Interaction with materials and activity of sales representatives could.
Publishing & Sharing	Access	Who will have access to the dashboard?	Only people selected by stakeholders will get access. I'll provide s list.
	Distribution	How will the dashboard be distributed or published to the audience?	Workspace, shared once a month.
	Refresh Frequency	How often does the data need to be refreshed?	Twice a month
	Version Control	What mechanisms will be in place for managing versions of the dashboard?	Tabular Editor

Source: The Thesis author's processing

12 Assignment of Master Thesis (eVŠKP)

Zadání diplomové práce

Autor:	**Bc. Viktorie Láchová**
Studium:	I2100099
Studijní program:	N0688A140001 Informační management
Studijní obor:	Informační management
Název diplomové práce:	**Dashboards automation**
Název diplomové práce AJ:	Dashboards automation

Cíl, metody, literatura, předpoklady:

The entire thesis will be dedicated to the automation in the company focused on dashboards. The goal is to automate and to link selected data in to create monthly distributed reports.

Zadávající pracoviště:	Katedra informatiky a kvantitativních metod, Fakulta informatiky a managementu
Vedoucí práce:	prof. RNDr. Petra Poulová, Ph.D.
Datum zadání závěrečné práce:	26.1.2021