NOVA IMS
INFORMATION MANAGEMENT SCHOOL

Master's Degree in
**Data Science and Advanced Analytics**

Specialization in
**Business Analytics**

# Artificial Intelligence in Business Intelligence Systems

Assessing Opportunities, Challenges and Organizational Impact

**Rafael Marques Lopes**

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

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

# Artificial Intelligence in Business Intelligence Systems

Assessing Opportunities, Challenges and Organizational Impact

by
Rafael Marques Lopes

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Business Analytics

## Supervised by

Vítor Manuel Pereira Duarte dos Santos, PhD, NOVA Information Management School

April, 2026

# STATEMENT OF INTEGRITY

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

*[Lisboa, 30 April 2026]*

*Rafael Lopes*

# DEDICATION

The completion of this thesis represents the conclusion of an important stage of my academic journey and would not have been possible without the support, encouragement and guidance of several people and institutions who, in one way or another, were fundamental to the accomplishment of this work.

I dedicate this work to my family for their unconditional support throughout my academic journey.

To my parents, for their constant encouragement, guidance and belief in my abilities, and to my brother, for his support and companionship along the way.

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Vítor Santos, for his guidance, support and valuable insights throughout the development of this thesis.

I also thank NOVA Information Management School for providing the academic environment and resources necessary for this research.

A special thanks to my family, especially my parents, brother, grandparents, uncles and cousins, as well as to my friends and colleagues for their continuous encouragement and support during this journey.

Finally, I acknowledge all individuals and organizations that contributed directly or indirectly to this work.

# ABSTRACT

This thesis examines the impact of Artificial Intelligence (AI) on Business Intelligence (BI) workflows, with a particular focus on the opportunities, challenges and organizational implications associated with its integration into data-driven decision-making processes. A comparative case study approach is adopted, in which a complete BI pipeline is executed using both traditional methodologies and AI-augmented tools. The research follows a structured methodology comprising exploratory, conceptual, testing and conclusive phases. Two real-world datasets are used to implement and evaluate both approaches across all stages of the BI process, including data extraction, transforming and loading (ETL), data engineering, dashboard development and analytical insight generation. The results are assessed through multiple dimensions, namely efficiency, quality of insights, user experience, cost implications and reproducibility. The findings indicate that AI-powered tools significantly enhance speed and automation, particularly in data preparation and insight generation, while also introducing challenges related to transparency, reliability and governance, commonly associated with the “black-box” nature of AI systems. This research contributes to both academic and practical domains by providing empirical evidence on how AI reshapes BI workflows and by identifying the conditions under which AI can effectively augment or complement traditional analytical processes, highlighting the importance of balancing automation with interpretability and human oversight to maximize the value of AI-driven Business Intelligence.



# KEYWORDS

Artificial Intelligence (AI); Business Intelligence (BI); ETL Processes; Data-Driven Decision Making; Augmented Analytics (AA); Machine Learning (ML); Data Analytics


## Sustainable Development Goals (SDG):

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| **AA** | Augmented Analytics |
| **AI** | Artificial Intelligence |
| **BI** | Business Intelligence |
| **BPM** | Business Performance Management |
| **DSS** | Decision Support Systems |
| **DW** | Data Warehouse |
| **ETL** | Extract, Transform, Load |
| **KPI** | Key Performance Indicator |
| **ML** | Machine Learning |
| **NLP** | Natural Language Processing |
| **OLAP** | Online Analytical Processing |
| **PRISMA** | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| **SLR** | Systematic Literature Review |
| **SLRQ** | Systematic Literature Review Question |

# 1. INTRODUCTION

## 1.1 CONTEXT

Business Intelligence (BI) can be defined as the systematic process of collecting, processing and analyzing data to generate information that supports decision-making within organizations. The primary goal of BI is to transform raw data into meaningful insights that enable businesses to improve performance, identify opportunities and respond more effectively to challenges.

BI tools are designed to provide users with access to both historical and real-time data, originating from internal and external sources. These may encompass structured formats, such as Excel spreadsheets, as well as unstructured formats, such as social media content (e.g., posts on X).

By consolidating and interpreting this information, BI systems enable decision-makers to gain a comprehensive view of organizational performance, monitor trends and anticipate future developments. In this way, BI not only serves as a descriptive mechanism for understanding what has happened but also as a predictive tool that supports the development of growth-oriented strategies (IBM, 2021).

Initially, BI was primarily associated with static reporting and descriptive analysis. Today, however, BI is characterized by the integration of Artificial Intelligence (AI) and Big Data technologies, which have fundamentally expanded its scope and capabilities (Farayola, 2024).

The evolution of BI has been strongly shaped by both technological advancements and shifting business needs. Its origins can be traced back to the 1960s, with the emergence of Decision Support Systems (DSS), which assisted managers in decision-making processes through manual data input and rule-based models. By the late 1980s, the growing availability of computing power and relational databases enabled the transition from DSS toward more sophisticated BI tools. This period marked the rise of Data Warehousing and Online Analytical Processing (OLAP), which laid the foundation for modern BI by allowing organizations to query, consolidate and analyze large volumes of structured data efficiently. In the 2000s, the advent of interactive dashboards and visualization platforms further enhanced accessibility, making data easier to interpret and expanding BI from a specialized function into a broader organizational capability (Chintala, 2024).

Modern BI has evolved beyond static reports, becoming a dynamic solution that transforms massive volumes of data into actionable insights in real-time. The emergence of Big Data has been central to this transformation. The well-known three Vs of Big Data (volume, variety and velocity) require organizations to process vast amounts of heterogeneous information rapidly. Traditional BI systems were not

designed to handle these demands, which led to the development of advanced analytics platforms powered by machine learning algorithms and AI-driven processing (Chintala, 2024).

AI, in particular, has introduced transformative capabilities that enhance the value of BI. Through automation, pattern recognition and predictive modeling, AI enables organizations to process large-scale data more efficiently and accurately. Machine learning algorithms uncover complex relationships within the data, while natural language interfaces allow users to interact with BI systems more intuitively. This combination empowers businesses to generate deeper insights and make faster, evidence-based decisions (Farayola, 2024). In this context, AI facilitates predictive analytics that help organizations anticipate customer behavior, forecast demand and identify emerging trends.

Despite these opportunities, the integration of AI into BI also introduces significant challenges. One of the most prominent is the “black-box problem”, which refers to the lack of transparency regarding how AI systems generate their outputs. The internal reasoning processes remain hidden within the model, leaving users with only the output, without insight into the logic or steps that generated it. For instance, when interacting with large language models, the user receives an answer without transparency regarding the specific sources or reasoning path that generated that conclusion. This lack of transparency may undermine trust and hinder the adoption of AI-driven BI systems (Shah, 2024).

The reliability of AI-driven insights depends heavily on the quality and representativeness of the underlying data. Incomplete, biased or outdated datasets can lead even the most advanced AI models to generate flawed guidance. Additional barriers to successful AI integration include the need for specialized expertise and substantial financial and organizational investments (Farayola, 2024).

AI also amplifies long-standing data privacy and ethical risks while introducing new concerns related to misuse and bias. Compared to earlier technologies, AI systems are far more data-intensive and opaque, posing the “black-box problem,” which makes it increasingly difficult for individuals to control, understand, or remove personal data collected about them. This intensifies the already pervasive issue of digital surveillance (Miller, 2024).

A central objective of modern BI tools is to present data in an accessible and actionable format for non-technical users, particularly decision-makers who rely on insights to guide strategy but lack specialized expertise in data science or engineering. While AI-powered tools have the potential to simplify complex analyses and provide predictive insights, many end-users still feel unprepared to fully leverage these capabilities. Bridging this gap between technological potential and practical usability remains a key challenge for organizations (Syed, 2022).

Despite these challenges, traditional BI functionalities such as reporting and dashboarding remain essential. Platforms like Tableau and Microsoft Power BI continue to provide essential descriptive analytics and visualizations, helping organizations track performance and address specific business questions (Adewusi et al., 2024).

Therefore, the integration of AI into BI systems presents both unprecedented opportunities and unresolved challenges. This tension provides the foundation for the following research question:

How is Artificial Intelligence reshaping Business Intelligence workflows and what are the benefits, limitations and organizational challenges associated with its integration?

Exploring this question can help uncover how businesses may responsibly and effectively leverage AI to enhance their BI capabilities, while also addressing the risks and barriers that may hinder adoption.

## 1.2 OBJECTIVES

The primary goal of this research is to evaluate the added value of AI-powered tools across the Business Intelligence (BI) pipeline by conducting a comparative case study. This study aims to analyze a real-world dataset using both a traditional BI approach and an AI-augmented workflow, documenting each stage from data ingestion and transformation (ETL), through visualization and insight generation.

By comparing outputs, both in terms of efficiency and insight quality, this research seeks to provide concrete, evidence-based conclusions about how and where AI brings tangible improvements to BI. In doing so, it will help organizations better understand the strategic implications of adopting AI-powered BI tools and make informed decisions on their data analytics investments.

Therefore, the goal of the research is to compare the effectiveness of traditional BI workflows and AI-powered BI solutions across the full data analytics pipeline using a real-world dataset.

In order to achieve this goal, the following intermediate objectives were defined:

- Make a comprehensive study on the use of AI technologies in BI
- Select and preprocess real-world datasets with relevant data for a business performance analysis
- Implement traditional BI workflows, including ETL processes and dashboard creation using standard BI tools
- Repeat the previous objective using AI
- Compare both approaches in terms of efficiency, accuracy, user interactivity and insight generation
- Evaluate user experience and decision-making impact

## 1.3 Study Results and Contributions

The outcomes of this project have the potential to significantly support organizations, particularly in the public and private sectors, by providing evidence-based guidance on how Artificial Intelligence (AI) can enhance Business Intelligence (BI) practices. By understanding the tangible benefits and limitations of AI-powered BI tools in real-world scenarios, organizations can make more informed investments in data infrastructure and talent. This leads to more efficient analytics processes, better decision-making and ultimately improved performance and responsiveness to market or citizen needs. For public administration, the ability to extract more actionable insights from existing data sources can support smarter policy design, resource allocation, and service delivery, all of which contribute to societal well-being and economic resilience.

From a scientific and academic perspective, this research fills a notable gap by offering a comparative case study that contrasts traditional and AI-augmented BI workflows. It also encourages critical reflection on the collaborative or competitive roles of AI and human analysts, fostering further research on hybrid intelligence models and the evolving nature of data-driven organizations.

This study also aims to demonstrate how AI is currently being leveraged to address contemporary data challenges, particularly the scale, variety and complexity associated with Big Data. By applying AI-powered tools to a real-world dataset, this research will demonstrate how these technologies can automate pattern detection, support predictive analytics, and accelerate data processing tasks that would otherwise require considerable human effort. In doing so, it will clarify how AI transforms not just the speed, but also the depth and strategic value of data analysis.

Moreover, this thesis will critically assess the limitations of AI in BI environments. It will examine issues such as opacity in decision-making (“black-box problem”), potential bias in data and algorithms and the complexity introduced when integrating AI systems into existing BI infrastructures. These findings will be essential in understanding whether AI truly simplifies the BI process or, in some contexts, adds new layers of technical and organizational complexity.

Finally, this research will evaluate the automation potential of AI within BI workflows, especially whether AI-driven solutions can reliably replace or augment human analytical tasks and under what conditions this automation is most effective. The study will reflect on the trade-offs between efficiency and explainability, speed and control, offering practical insights that organizations can use to navigate the evolving BI landscape.

# 2. LITERATURE REVIEW

This chapter provides a comprehensive exploration of Business Intelligence, laying the conceptual and technical foundation necessary to understand its role in modern days. The goal is to clarify what BI is, trace its historical evolution and highlight how it has transformed from descriptive reporting systems into advanced, AI-augmented decision-support platforms. The discussion begins by defining BI and its core components, explaining how data is transformed into actionable information that guides strategic decision-making. It then explores the historical context and technological milestones that have shaped BI, followed by an examination of current trends and the impact of Big Data on BI capabilities. Special attention is given to the ETL process, as it constitutes the backbone of data integration and preparation, as well as to the transformative role of automation and AI, which are redefining the way businesses interact with data. The chapter concludes with a critical reflection on the challenges and opportunities associated with BI adoption, setting the stage for the subsequent analysis of AI-powered BI workflows.

## 2.1 BUSINESS INTELLIGENCE

The term Business Intelligence derives from the ability of systems to provide decision-makers with user-friendly access to information that supports strategic and operational decisions. BI represents a combination of analytical tools, methodologies, databases and applications designed to collect, process and present data in a meaningful way (Sharda et al., 2018).

The BI process can be broadly understood as a three-step cycle. First, data must be transformed into information by consolidating and structuring raw inputs into a usable format. Next, this information is used to support decision-making by providing relevant and timely insights that guide management choices. Finally, these insights drive action, as organizations implement decisions based on data-driven recommendations (Sharda et al., 2018).

The information produced through BI may be descriptive, predictive and prescriptive. This combination of perspectives allows BI systems to not only report historical performance but also guide organizations toward future opportunities and optimal strategies (Sharda et al., 2018).

In terms of capabilities, BI typically encompasses reporting, forecasting, prediction, trend analysis and performance monitoring, all of which contribute to more informed and agile decision-making. Architecturally, BI solutions are usually built around four core components. At the foundation lies the Data Warehouse (DW), a centralized repository that stores both current and historical data and is designed to support decision-making activities. Above it sits the business analytics layer, which provides

tools and techniques for querying, mining and analyzing data. This is where the ETL (Extract, Transform, Load) process takes place: data is extracted from multiple sources, transformed through cleansing, standardization and enrichment to meet business requirements and then loaded into the warehouse so that it is ready for analytical processing. Complementing this is Business Performance Management (BPM), which monitors and evaluates performance through key performance indicators (KPIs) and other metrics. Finally, a user interface, typically in the form of dashboards and interactive visualizations, makes the insights accessible and actionable for decision-makers across the organization (Sharda et al., 2018).

By integrating these elements into a cohesive architecture, BI systems enable organizations to transform raw data into actionable intelligence. This capability empowers managers and business leaders to make faster, evidence-based decisions, turning data into a strategic asset that drives performance and competitive advantage.

### 2.1.1 Evolution of Business Intelligence

The origins of BI can be traced back to the early 1960s, with the development of Decision Support Systems (DSS). These early systems were designed to help managers make more informed decisions by providing structured information based on manually entered data (Chintala, 2024).

In the 1990s, BI began to take the shape we recognize today with the introduction of Data Warehousing and Online Analytical Processing (OLAP). These technologies enabled organizations to consolidate data from multiple sources into centralized repositories and perform multidimensional analysis, laying the groundwork for modern BI solutions (Chintala, 2024).

The 2000s marked a significant leap forward with the rise of interactive dashboards and real-time analytics. Organizations could now monitor Key Performance Indicators (KPIs) and other critical metrics dynamically, allowing for faster responses to changing business conditions (Chintala, 2024).

This historical evolution reflects the shift from static, retrospective reporting to the highly interactive and predictive BI systems used today (Chintala, 2024).

It is worth noting that the evolution of BI has always been closely linked with technological progress. As new technologies emerge, they continuously complement and expand the capabilities of BI systems. This relationship is symbiotic: BI evolves to meet new business demands, while advancements in computing, storage and analytics create opportunities for more powerful, scalable and user-friendly BI solutions. The field remains in constant motion, adapting to innovations and pushing the boundaries of what is possible in data-driven decision-making (Farayola, 2024).

### 2.1.2 MODERN BUSINESS INTELLIGENCE

One of the most defining characteristics of the modern BI era is the emergence of Big Data (Adewusi et al., 2024).

Although traditional reporting and querying remain fundamental for organizations that aim to make decisions based on structured historical data, the rise of Big Data has exposed the limitations of these traditional approaches. Organizations are now required to manage and analyze data characterized by the three Vs:

- Volume: the massive amounts of data generated
- Variety: data in multiple formats, including structured data (e.g., relational databases, spreadsheets), semi-structured data (e.g., JSON, XML) and unstructured data (e.g., social media posts, videos, images)
- Velocity: the speed at which new data is created and must be processed, often in real time

Despite these challenges, tools such as Tableau and Microsoft Power BI continue to play a critical role in enabling descriptive analytics (Adewusi et al., 2024).

In response to these challenges, organizations have turned to advanced analytical tools that go beyond descriptive reporting. Modern BI now encompasses:

- Descriptive analytics: understanding what happened in the past
- Predictive analytics: forecasting what is likely to happen in the future
- Prescriptive analytics: providing actionable recommendations to guide decision-making

At the forefront of this transformation is machine learning, which enables systems to adapt to different data types and formats, uncover hidden patterns and automate analysis. Combined with the scalability of cloud-based solutions, such as Apache Hadoop, organizations can now store, process and analyze massive datasets efficiently, transforming raw information into a strategic asset (Adewusi et al., 2024).

This transformation allows companies to:

- Make faster, evidence-based decisions
- Predict future outcomes and behaviors
- Identify trends and opportunities earlier
- Optimize operational processes in real time

The ability to analyze data efficiently and effectively is increasingly what distinguishes industry leaders from their competitors (Adewusi et al., 2024). Due to these developments, modern BI has evolved significantly beyond traditional descriptive reporting. While descriptive analysis remains indispensable, organizations now leverage machine learning algorithms and predictive and prescriptive analytics to gain deeper insights, anticipate future scenarios and even recommend optimal courses of action (Adewusi et al., 2024).

Looking ahead, Augmented Analytics (AA) represents the most recent shift in BI. AA enhances traditional BI capabilities to handle Big Data more effectively, accelerate data preparation process (ETL) and speed up insight generation. BI is becoming progressively more user-centric, autonomous and intelligent, redefining the way organizations interact with data and make strategic decisions (Farayola, 2024).

### 2.1.3 ETL PROCESS

As data warehouses (DWs) grow in size and complexity, the challenge of integrating data from multiple sources becomes increasingly significant. Organizations no longer rely solely on historical, cleaned data stored in static repositories. Instead, they are demanding access to real-time data, unstructured information and continuously updated streams that reflect current business conditions (Sharda et al., 2018).

At the core of this integration challenge lies the ETL process, which stands for Extract, Transform and Load. ETL is an integral part of every business analytics initiative, as it serves as the bridge between data sources and the data warehouse. Its primary goal is to transport and prepare data from multiple systems, ensuring that by the time it reaches the DW, it is consistent, accurate and ready for analysis (Sharda et al., 2018).

The process begins with extraction, where data is collected from disparate systems such as transactional databases, CRM (Customer Relationship Management) platforms, ERP (Enterprise Resource Planning) systems or even external sources. During the transformation phase, this raw data is cleansed, standardized, validated and enriched to meet business rules and reporting requirements. This step may involve filtering irrelevant records, correcting inconsistencies, joining datasets and restructuring information into a format suitable for analysis. Finally, in the loading phase, the transformed data is written into the data warehouse, where it becomes available for querying, reporting and other analytical processes (Sharda et al., 2018).

ETL is one of the fundamental pillars of building, maintaining and evolving a data warehouse. When designed and implemented effectively, it delivers several benefits: it enables end users to perform analysis with confidence in data quality, consolidates disparate data into a single, coherent view, simplifies access to information, facilitates faster and more informed decision making and supports the continuous improvement of business processes (Sharda et al., 2018).

By ensuring that data is properly integrated and prepared, the ETL process transforms scattered information into a reliable foundation for Business Intelligence, allowing organizations to turn data into actionable insights.

### 2.1.4 THE ROLE OF AUTOMATION AND AI IN BI

Artificial intelligence (AI) has fundamentally changed the way BI operates, giving rise to what is often referred to as Intelligent Business Analytics (Farayola, 2024). By enabling algorithms to autonomously identify patterns, detect anomalies, make predictions and uncover hidden relationships in data, it reduces the need for manual analysis and allows decision-makers to focus on strategic actions rather than routine reporting tasks (Adewusi et al., 2024).

The key role of AI is in decision support. Through predictive modeling, machine learning and automated business process optimization, organizations are becoming more efficient, more accurate and more capable of uncovering trends that would otherwise remain hidden. This transformation has also enabled self-service BI models, empowering business users at all levels to generate their own insights without requiring IT expertise, making the process more accessible across organizations (Farayola, 2024).

Moreover, Augmented Analytics (AA) represents the convergence of BI and AI. AA leverages machine learning and natural language processing (NLP) to automate the analytical cycle, from data preparation to insight generation, with greater accuracy and speed. This marks a milestone in the evolution of BI: the current era, where data is not merely a resource but a strategic asset (Farayola, 2024).

Advances in AI are also making machine learning (ML) algorithms more sophisticated, resulting in a deeper and more accurate understanding of data. By learning continuously from new data streams, these systems improve over time, making BI more proactive and adaptive to changing business environments. NLP further expands accessibility, allowing non-technical users to interact with BI systems through natural language queries (e.g., asking “What were last quarter’s sales?”), making analytics more intuitive and democratized across all levels of an organization (Adewusi et al., 2024).

In addition, AI facilitates real-time data processing and process automation, such as automatic data cleansing, transformation and anomaly detection, which enhances efficiency and data quality. Together with cloud-based infrastructure, AI-powered BI enables organizations to handle the complexity of Big Data, extract actionable insights faster and respond quickly to dynamic business conditions (Chintala, 2024).

**Framework for AI-Powered BI**

The implementation of AI in BI typically follows a structured framework to ensure accurate, actionable and reliable insights (Chintala, 2024). This framework can be summarized in the following key stages:

1. Data Collection and Processing: gathering structured, semi-structured and unstructured data from multiple sources, followed by cleaning, validation and transformation to ensure high-quality input for AI models
2. Feature Engineering: selecting, transforming and creating relevant features from raw data to improve model performance and interpretability
3. Model Selection and Training: choosing appropriate machine learning or statistical models and training them on historical and real-time data to capture patterns and relationships
4. Implementation: integrating trained models into BI platforms and workflows to enable predictive and prescriptive analytics for end-users
5. Evaluation and Optimization: continuously monitoring model performance, retraining with new data and optimizing parameters to maintain accuracy, reliability and relevance over time

This framework ensures that AI not only supports traditional BI functions but also enhances decision-making capabilities by providing predictive and prescriptive insights, automating routine processes and enabling real-time analysis (Chintala, 2024).

### 2.1.5 CHALLENGES AND OPPORTUNITIES

In the contemporary business landscape, data has become the new currency. Organizations that can collect, process and extract insights from data gain a substantial competitive advantage. However, achieving this is not without its challenges (Adewusi et al., 2024).

A major difficulty originates from the three Vs of Big Data, which create complexity in processing and maintaining data quality. To extract reliable insights, companies must cleanse, validate and standardize data, which is often a resource-intensive process. Establishing a robust data governance framework is crucial, including regular audits and quality checks to ensure consistency and reliability (Adewusi et al., 2024).

Another significant concern is data security and privacy. Organizations handle both proprietary corporate information and highly sensitive personal customer information. Implementing end-to-end encryption, strict access controls and compliance with data protection regulations (such as the General Data Protection Regulation, the European Union data protection and privacy law) are essential to safeguard information and maintain customer trust (Adewusi et al., 2024).

The talent shortage poses an additional challenge. The rapid evolution of Big Data technologies and analytical tools creates a gap between organizational needs and available expertise. To address this, companies are investing in training programs,

partnerships with academic institutions and talent acquisition strategies to build data-literate teams (Adewusi et al., 2024).

Furthermore, adopting Big Data technologies requires substantial financial investment in infrastructure, cloud platforms and tools such as Apache Hadoop or other distributed processing frameworks. Large corporations are often better positioned to make these investments and stay ahead of technological trends, which risks widening the gap between large enterprises and smaller businesses that lack the financial and human resources to compete at the same level (Adewusi et al., 2024).

Despite these challenges, organizations that successfully navigate them can unlock significant opportunities, from faster decision-making and personalized customer experiences to entirely new data-driven business models (Adewusi et al., 2024).

Looking forward, user-friendly technologies will remain a priority for BI platforms. The goal is to empower business users to explore data and generate insights independently, reducing reliance on IT teams and accelerating decision-making cycles (Adewusi et al., 2024).

In parallel, the need for transparency and security in data processing is leading to experimentation with blockchain technology. Blockchain's decentralized and tamper-proof record can enhance trust in data integrity, ensuring that analytics are based on accurate and authentic information, a critical requirement for compliance-heavy industries such as finance and healthcare (Adewusi et al., 2024).

## 2.2 Related Work

To conduct a systematic and transparent review of the existing literature, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology was applied.

The purpose of using PRISMA in this research is to ensure that the process of collecting and analyzing related work follows an evidence-based and methodologically rigorous approach.

### 2.2.1 PRISMA Protocol

This protocol was first published in 2009 with the objective of helping systematic reviewers clearly explain why a review was conducted, how it was performed (i.e., the methods used) and what was found. It serves as a standardized guideline to enhance transparency, consistency and completeness in the reporting of systematic reviews (Page et al., 2023).

In 2020, an updated version of the PRISMA statement was released, replacing the original 2009 version. This revision introduced new orientations that reflect

methodological advancements in the identification, selection, evaluation and synthesis of studies (Page et al., 2023).

The PRISMA 2020 framework is composed of:

- A checklist of 27 items that guide researchers in reporting each step of a systematic review
- An expanded list providing detailed recommendations for each checklist item
- A summary checklist to facilitate the synthesis of findings
- Flow diagrams designed to illustrate the process of study identification, screening, inclusion and exclusion, applicable to both new reviews and updates of existing ones

Through these elements, PRISMA promotes methodological rigor and enhances the reproducibility and credibility of systematic literature reviews.

### 2.2.2 PRISMA EXECUTION

Following the framework described in the previous section, this part presents the practical implementation of the PRISMA methodology in the context of the current study. Following the PRISMA 2020 guidelines, the systematic literature review (SLR) was designed to ensure transparency and reproducibility throughout all stages, including identification, screening, eligibility assessment, and inclusion of relevant studies.

To guide the review process, a set of Systematic Literature Review Questions (SLRQs) was formulated, reflecting the main objectives of this research. These questions aim to capture the current state of academic and practical knowledge on the integration of AI within BI systems, as well as to explore its associated opportunities, challenges and organizational implications.

The systematic review was therefore structured around the following research questions:

- SLRQ1: What is the current status of research in AI technologies in BI?
- SLRQ2: What are the key challenges associated with the adoption of AI technologies in BI for businesses and governments?
- SLRQ3: Which AI and ML techniques are most commonly applied to enhance BI?
- SLRQ4: What research gaps and future directions can be identified in the current integration of AI within BI systems?

Once the SLRQs were formulated, the next step was to capture all relevant studies at the intersection of AI and BI, ensuring a balance between specificity and inclusiveness.

To achieve this, two main groups of keywords were established, representing the core thematic dimensions of the research:

- Business Intelligence:
  1. Business Intelligence
  2. Big Data Analytics
  3. ETL
  4. Data Warehouse*
- Artificial Intelligence:
  1. Artificial Intelligence
  2. Machine Learning
  3. Advanced Analytics

These terms were combined using Boolean operators (AND and OR) to conduct the final search strings: (“Business Intelligence” OR "ETL" OR "Data Warehouse*") AND (“Artificial Intelligence” OR “Machine Learning” OR "Advanced Analytics” OR “Generative AI” OR “Agentic AI”).

This structured formulation enabled a systematic and replicable search across multiple academic databases, ensuring comprehensive coverage of studies addressing the intersection of AI and BI from both technical and organizational perspectives.

The search was conducted between October and November of 2025 on the following scientific information resource databases:

- Scopus: https://www.scopus.com/
- Web of Science: https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/
- Science Direct: https://www.sciencedirect.com/

After completing the database searches and retrieving the initial set of publications, it was necessary to establish a consistent and objective filtering process to ensure the relevance and quality of the selected studies. Therefore, a set of inclusion and exclusion criteria was defined:

- Inclusion criteria:
  1. Evidence of AI utilization in BI
  2. Peer-reviewed journal or conference paper
  3. Written in English
  4. Publication period: 2022-2025
  5. Full text accessible
- Exclusion criteria:
  1. Focus on BI without AI application or vice versa
  2. Non-peer-reviewed or duplicated papers
  3. Not written in English

4. Published before 2022
5. Inaccessible full text
6. Out of scope
7. Non-academic or non-scientific sources (e.g., reports, magazines, websites)

Following the application of the inclusion and exclusion criteria, the study proceeded with the standardized PRISMA workflow. The following presents this process in detail, documenting the number of records retained at each stage. This visual representation illustrates how the predefined criteria, sources and keywords were systematically applied throughout the phases. The final set of studies represents the documents that successfully passed all stages of the PRISMA methodology and were therefore selected for comprehensive review and interpretation.

In the identification phase, a total of 5443 records were retrieved from the three academic databases previously mentioned, using the defined search string. These records represented all potentially relevant publications addressing the integration of AI within BI contexts. After the removal of duplicate entries, 3131 unique records remained for further examination.

During the screening phase, these records were reviewed according to the established inclusion and exclusion criteria, leading to the exclusion of 2828 studies that did not meet the defined requirements. This process ensured that only publications potentially aligned with the research objectives advanced to the next stage.

In the eligibility phase, 303 abstracts were analyzed in greater detail to assess their methodological rigor and thematic relevance to the scope of this research. Of these, 281 abstracts were excluded due to factors such as insufficient focus on AI-BI integration or misalignment with the established SLRQs. Consequently, 22 full-text articles were selected for comprehensive assessment. After this stage, 7 studies were excluded following an in-depth reading and evaluation process.

Finally, 15 studies met all inclusion criteria and were incorporated into the final synthesis. These publications represent the most relevant and methodologically robust contributions to the understanding of AI integration within BI systems, forming the foundation for the subsequent analysis of opportunities, challenges and organizational impacts.

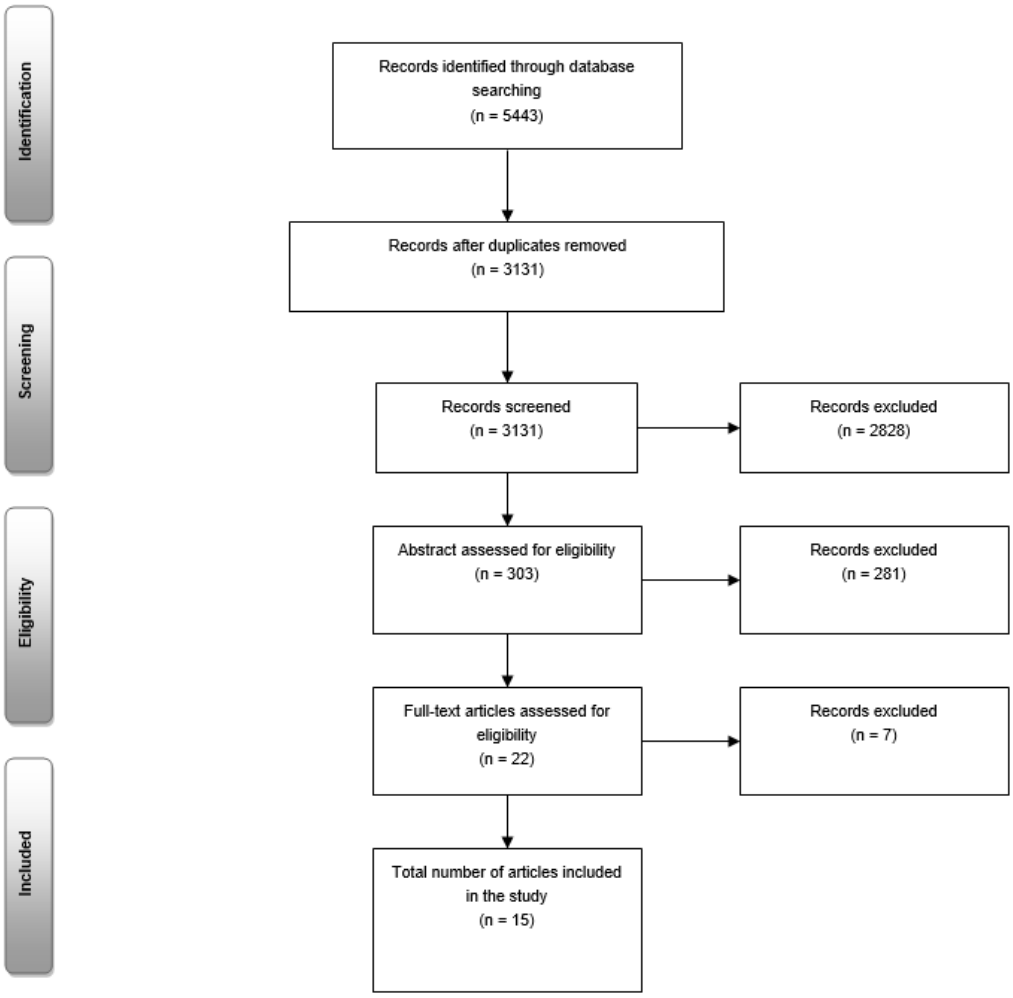



Figure 2.1 - PRISMA Flow Diagram

After completing all stages of the PRISMA process, a final set of fifteen studies was selected for detailed analysis. These publications constitute the core academic contributions relevant to this research, as they directly address the integration of AI into BI systems. The table, presented in Appendix C, summarizes these studies, including their authors, article titles, main contributions and publication types, providing a structured overview of the foundational literature supporting this review.

### 2.2.3 PRISMA Results Analysis

After completing the research process using the PRISMA methodology described in the previous section, this part of the study focuses on analyzing the selected articles to answer the defined SLRQs. The analysis draws on the main contributions of the fifteen reviewed articles, aiming to identify how AI is being integrated into BI systems, the challenges involved, the main AI and ML techniques applied and the existing research gaps. By synthesizing these findings, this section provides a structured overview of the current state of research.

**SLRQ 1 - What is the current status of research in AI technologies in BI?**

The current status of AI integration in BI indicates a significant evolution from traditional descriptive systems to predictive, prescriptive and augmented analytics platforms. AI enhances the capabilities of BI by enabling organizations to process large volumes of structured and unstructured data efficiently, uncovering hidden patterns, performing anomaly detection and generating insights in real time (Afsaruddin et al., 2023; Alghamdi & Al-Baity, 2022; Al-Momani, Alqudah, et al., 2025, p. 360360; Al-Momani, Awadallah, et al., 2025; Alqhatani et al., 2022; Das et al., 2024; Desai & Desai, 2025; Devi et al., 2025; Gajić et al., 2025; Jimenez-Partearroyo & Medina-López, 2024; Kareem & Abdullah, 2025; Mantouzi et al., 2025; Milinthapunya et al., 2025; Rahman et al., 2024; Vines et al., 2025). Generative AI and natural language processing allow non-technical users to interact with BI systems directly, extracting actionable insights without heavy reliance on IT teams, which bridges the gap between business experts and technical staff (Alghamdi & Al-Baity, 2022; Das et al., 2024; Desai & Desai, 2025; Jimenez-Partearroyo & Medina-López, 2024; Vines et al., 2025). Machine learning is widely applied to forecast trends, segment customers, classify data and optimize predictive accuracy across different industries, from finance and insurance to retail and higher education (Afsaruddin et al., 2023; Alqhatani et al., 2022; Kareem & Abdullah, 2025; Milinthapunya et al., 2025; Rahman et al., 2024). Large-scale BI platforms such as Tableau, Microsoft Power BI, Qlik and ThoughtSpot have incorporated AI capabilities to support self-service analytics, automated data preparation and visualization, enabling faster and more reliable decision-making (Alghamdi & Al-Baity, 2022; Al-Momani, Awadallah, et al., 2025; Alqhatani et al., 2022; Das et al., 2024). AI also supports the full BI lifecycle, from the ETL process, which includes data cleansing, integration and modeling, to warehousing and finally to visualization, increasing the overall efficiency and quality of insights (Alghamdi & Al-Baity, 2022; Desai & Desai, 2025; Jimenez-Partearroyo & Medina-López, 2024; Kareem & Abdullah, 2025; Vines et al., 2025). Cloud computing and scalable data architectures enhance the processing of Big Data, allowing companies to leverage AI for both historical analysis and predictive scenarios (Alqhatani et al., 2022; Devi et al., 2025; Kareem & Abdullah, 2025). Overall, AI is transforming BI from a supporting tool into a strategic enabler, providing more timely, accurate and actionable insights while empowering business users with greater independence and analytical capabilities.

**SLRQ 2 - What are the key challenges associated with the adoption of AI technologies in BI for businesses and governments?**

Despite the significant benefits, the adoption of AI in BI faces multiple challenges across technical, organizational and human dimensions. One of the main barriers is the skill gap, as the effective implementation and maintenance of AI-powered BI

systems require highly trained technical teams capable of managing complex data architectures and AI models (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Desai & Desai, 2025; Devi et al., 2025; Gajić et al., 2025; Mantouzi et al., 2025; Rahman et al., 2024; Vines et al., 2025). Many organizations also face technological constraints due to legacy systems, insufficient IT infrastructure and the capital investment required for cloud-based solutions or new data warehouses, which can delay or limit adoption (Al-Momani, Alqudah, et al., 2025; Alqhatani et al., 2022; Devi et al., 2025; Kareem & Abdullah, 2025; Rahman et al., 2024). Data quality, consistency and governance remain critical challenges because AI relies heavily on high-quality, comprehensive datasets to produce reliable predictions and insights (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Desai & Desai, 2025; Jimenez-Partearroyo & Medina-López, 2024). Ethical concerns, including bias, transparency, privacy and security, are major considerations, as AI systems often process sensitive personal and business data and organizations must ensure compliance with regulations while maintaining trust (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Gajić et al., 2025; Mantouzi et al., 2025). Additionally, user adoption can be slow due to low AI literacy, insufficient training, resistance to change and a lack of clear communication or engagement strategies within organizations (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Desai & Desai, 2025; Gajić et al., 2025). Generative AI and LLMs introduce further challenges, such as dependency on prompt quality, potential hallucinations and inconsistent outputs, which require careful human supervision and evaluation (Vines et al., 2025). Organizational readiness, cultural perceptions and alignment of AI capabilities with business goals are also key factors influencing successful adoption, as misalignment can hinder AI from delivering its full potential (Al-Momani, Alqudah, et al., 2025; Desai & Desai, 2025; Devi et al., 2025; Rahman et al., 2024). Firm size plays a moderating role in AI adoption, with larger organizations benefiting from better infrastructure, more data and more technical and financial resources, while smaller organizations face limitations despite the relative ease of implementation in less bureaucratic environments (Rahman et al., 2024).

**SLRQ 3 - Which AI and ML techniques are most commonly applied to enhance BI?**

The literature highlights a range of AI and ML techniques that are more commonly applied to enhance BI. Machine learning is widely used for predictive modeling, classification, segmentation, pattern recognition and forecasting, enabling organizations to identify trends and anticipate market changes accurately (Afsaruddin et al., 2023; Alqhatani et al., 2022; Kareem & Abdullah, 2025). Natural language processing and large language models are increasingly integrated into BI systems to support self-service analytics, augmented analytics and conversational query

interfaces, allowing non-technical users to extract insights in natural language and interact directly with complex datasets (Alghamdi & Al-Baity, 2022; Das et al., 2024; Desai & Desai, 2025; Jimenez-Partearroyo & Medina-López, 2024; Vines et al., 2025). ML models enhance predictive accuracy and improve decision-making in scenarios involving large, complex, or unstructured data (Alqhatani et al., 2022; Kareem & Abdullah, 2025). ML and neural networks are applied for data-driven forecasting and anomaly detection, supporting real-time operational decisions (Al-Momani, Awadallah, et al., 2025; Alqhatani et al., 2022). Additionally, AI is extensively used to automate repetitive tasks, like the ETL process, data integration, warehousing and visualization, increasing efficiency and freeing human resources for higher-level decision-making (Alghamdi & Al-Baity, 2022; Desai & Desai, 2025; Jimenez-Partearroyo & Medina-López, 2024; Vines et al., 2025). Generative AI tools facilitate advanced analyses, enabling virtual data scientist capabilities that accelerate pattern recognition, predictive analytics and report generation while reducing manual effort (Alghamdi & Al-Baity, 2022; Vines et al., 2025). Overall, these AI and ML techniques collectively transform BI systems from descriptive to predictive and prescriptive decision-support platforms (Afsaruddin et al., 2023; Alghamdi & Al-Baity, 2022; Al-Momani, Alqudah, et al., 2025; Al-Momani, Awadallah, et al., 2025; Alqhatani et al., 2022; Das et al., 2024; Desai & Desai, 2025; Devi et al., 2025; Gajić et al., 2025; Jimenez-Partearroyo & Medina-López, 2024; Kareem & Abdullah, 2025; Mantouzi et al., 2025; Milinthapunya et al., 2025; Rahman et al., 2024; Vines et al., 2025).

**SLRQ 4 - What research gaps and future directions can be identified in the current integration of AI within BI systems?**

Despite substantial progress, several research gaps remain in the integration of AI within BI systems. Empirical evidence on the impact of AI adoption across different industries is limited, particularly regarding how generative AI and augmented analytics empower non-technical users to leverage insights independently (Alghamdi & Al-Baity, 2022; Desai & Desai, 2025; Mantouzi et al., 2025; Rahman et al., 2024). There is a need for comprehensive frameworks that guide the integration of AI into BI, ensuring alignment between business goals, technological infrastructure and system capabilities (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Desai & Desai, 2025). Challenges in data quality, governance and management continue to be underexplored, especially in contexts with heterogeneous or unstructured datasets (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Desai & Desai, 2025; Jimenez-Partearroyo & Medina-López, 2024). Talent and training shortages, organizational culture, employee engagement and regulatory constraints also represent critical areas for further study, as these factors significantly affect adoption and effectiveness (Al-Momani, Alqudah, et al., 2025; Das et al., 2024; Devi et al., 2025; Gajić et al., 2025).

Additionally, the practical implementation of AI-driven BI in cloud environments, as well as the mitigation of LLM hallucinations and bias, remain open questions (Al-Momani, Alqudah, et al., 2025; Alqhatani et al., 2022; Kareem & Abdullah, 2025; Vines et al., 2025). Future directions include the development of self-service BI tools, storytelling approaches to present insights, customized BI solutions and augmented analytics platforms capable of automating complex workflows while maintaining human oversight in final decision-making (Alghamdi & Al-Baity, 2022; Das et al., 2024). Research is particularly needed to evaluate AI's effectiveness in supply chain optimization, retail, financial services, higher education and other sectors where predictive and prescriptive BI can drive operational efficiency, cost reduction and customer satisfaction (Al-Momani, Awadallah, et al., 2025; Alqhatani et al., 2022; Milinthapunya et al., 2025; Rahman et al., 2024).

# 3. METHODOLOGY

## 3.1 OVERVIEW

This methodology defines the structured approach used to investigate, in this case, how GenAI can augment traditional BI workflows. It outlines how the study is designed, implemented and evaluated, ensuring that each step follows a coherent, transparent and replicable logic.

The methodological structure is organized into four phases: Exploratory, Conceptual, Testing and Conclusive, each defines specific procedures that guide the progression from theoretical grounding to practical experimentation and final evaluation.

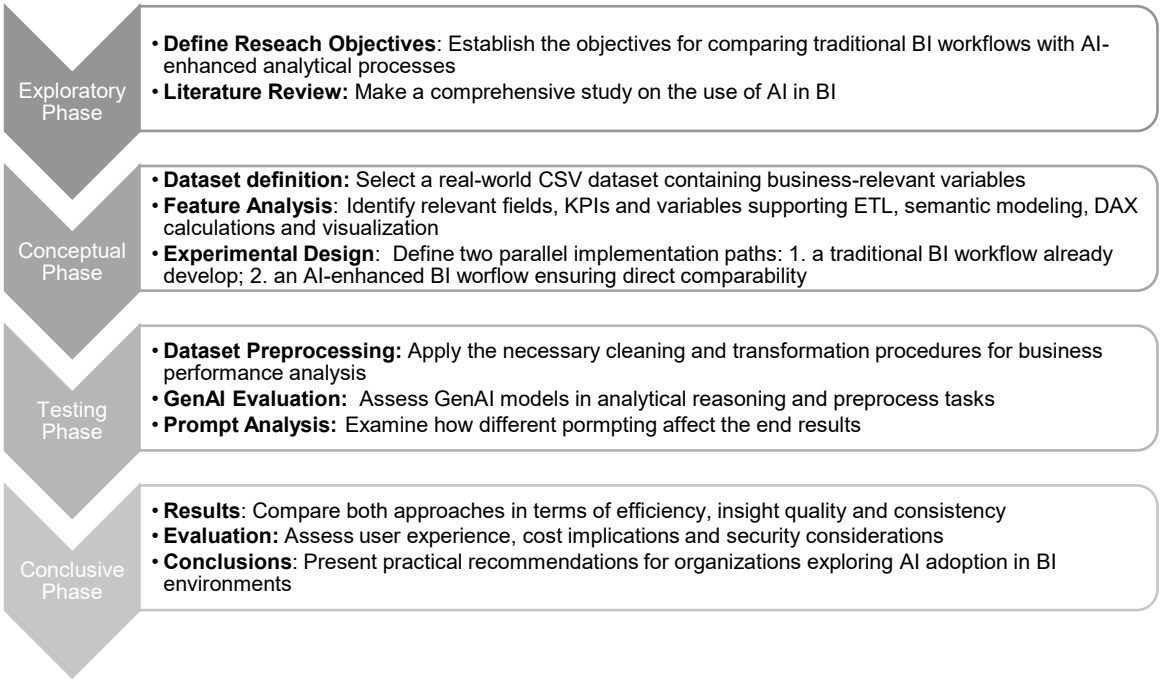



Figure 3.1 - Methodology Phases

## 3.2 EXPLORATORY PHASE

The exploratory phase establishes the analytical scope and research focus of the study. It addresses the central research focus: to what extent can GenAI support, accelerate or complement traditional BI workflows across data preparation, analytical modeling, visualization and executive-level storytelling?

This phase is informed by the systematic literature review presented in Chapter 2.2, conducted using the PRISMA methodology. The review highlights several claimed benefits of AI-enhanced BI systems, including automation of analytical tasks, natural-

language interaction, improved accessibility for non-technical users and support for predictive and prescriptive analytics.

At the same time, literature identifies critical limitations, including hallucinations, lack of reproducibility, explainability issues and governance concerns. These gaps motivate the empirical comparison adopted in this study, particularly the need to observe AI performance under realistic constraints rather than idealized or purely conceptual scenarios.

## 3.3 CONCEPTUAL PHASE

The conceptual phase translates theoretical insights into an operational BI framework suitable for empirical execution. The experimental architecture is grounded in the classical BI pipeline, consisting of the following stages:

1. Data ingestion and preparation
2. Analytical modeling and metric definition
3. Visualization and dashboard development
4. Interpretation and executive-level storytelling

Given the scope of the project, a full enterprise data warehouse is not implemented. Instead, a clean and enriched CSV dataset is used as a simplified yet realistic analytical layer.

Two execution paths are defined conceptually:

1. Traditional BI Workflow (Already Implemented):
   The traditional workflow represents a human-driven BI solution developed without Gen AI assistance. It includes manual data preparation, semantic modeling, business measures calculations and dashboard construction. This workflow serves as a fixed reference point and is intentionally not optimized or altered for the purpose of this research, ensuring a stable and unbiased comparison baseline.
2. AI-Enhanced Workflow:
   The AI-enhanced workflow introduces GenAI as a supporting agent across multiple stages of the BI pipeline:
   - Data cleaning and feature engineering
   - Semantic modeling through AI-generated business measures suggestions
   - Generation of visual analysis and narrative insights
   - Executive-level storytelling

This hybrid setup reflects realistic organizational adoption scenarios rather than idealized automation.

## 3.4 TESTING PHASE

The testing phase operationalizes the conceptual design into structured experimentation.

**Traditional BI Implementation (Baseline):**

The dataset is preprocessed in accordance with standard BI best practices, including handling missing values, correcting data types, performing transformations and ensuring data quality. Power BI is then used to build a relational and semantic model, implement DAX measures and business rules and create dashboards that reflect descriptive and diagnostic analysis. This baseline represents the conventional analytical workflow used in organizations.

**AI-Driven Implementation:**

The same dataset is then processed using AI-enabled methods. GenAI models (e.g. Chat GPT, Gemini, Git Hub Copilot) are employed to replicate the traditional BI workflow through a comprehensive and structured prompt. The AI models are granted limited analytical flexibility, allowing exploration of alternative approaches while remaining aligned with the baseline workflow and overall business objective.

**Evaluate Dimensions:**

Both workflows are compared along several axes:

- Efficiency: Time, effort and number of manual steps required
- Insight Quality: Depth, clarity and relevance of findings
- User Experience: Accessibility for non-technical users

Where applicable, qualitative user feedback or small-scale surveys may complement this evaluation.

## 3.5 CONCLUSIVE PHASE

The conclusive phase synthesizes the results obtained from both approaches and consolidates the key findings of the comparative analysis. The comparison highlights how AI can complement, enhance or accelerate traditional BI processes. Key considerations include:

- The degree to which AI automates ETL, modeling and visualization
- The accuracy and usefulness of AI's capabilities
- The ability of GenAI to generate strategic storytelling suitable for executives
- The feasibility of implementation without specialized IT teams

The final recommendations provide actionable guidance for organizations considering the integration of GenAI into their BI strategies, emphasizing both its transformative potential and the practical limitations related to cost, governance and reliability.

# 4. EXPERIMENTAL DESIGN

This chapter presents the experimental design adopted to empirically evaluate to what extent generative AI can augment traditional BI workflows. The design is grounded in the methodological framework defined in the previous chapter and directly addresses the research gaps identified through the systematic literature review conducted previously.

The reviewed literature highlights several well-established advantages of AI-enhanced BI systems, including the automation of analytical tasks, predictive and prescriptive analytics and improved accessibility for non-technical users. However, the literature also reveals significant gaps, particularly the lack of rigorous empirical studies that directly compare traditional BI workflows with AI-augmented approaches under controlled conditions, i.e., using the same dataset, identical business objectives and comparable analytical tools. Most existing research either focuses on conceptual frameworks or evaluates AI-enhanced BI components in isolation, offering limited evidence on how AI alters the effectiveness, efficiency and reliability of the BI pipeline as a whole, from data preparation to visualization and storytelling.

In response to these gaps, this experimental design implements a comparative case study using real-world datasets and two parallel analytical workflows: a traditional BI workflow and an AI-enhanced BI workflow. Both approaches operate on the same datasets, pursue the same analytical goals and are evaluated using consistent criteria. This structure enables a systematic assessment of whether AI confirms the advantages reported in prior studies and whether it effectively addresses the limitations identified in the literature, particularly in terms of usability, automation, insight quality and organizational impact.

## 4.1 Dataset Selection

### 4.1.1 Why These Datasets

The datasets selected for this study originate from two publicly available analytical challenges. The first one is the Maven Marketing Challenge and the second one, to extend the experimental scope by introducing a more complex scenario, is the KPMG Data Analytics Virtual Program.

Both datasets were chosen for several methodological and practical reasons that make them particularly suitable for evaluating both traditional and AI-enhanced BI workflows. Both are publicly available, ensuring transparency, reproducibility and academic rigor. Public datasets are frequently used in BI and analytics research because they allow results to be validated and compared across studies. They provide clear business

requirements and analytical objectives, simulating realistic organizational scenarios rather than abstract or artificially constructed datasets. This aligns with the applied nature of this research, which aims to assess AI's impact on real-world BI practices.
Additionally, both challenges include documented human-developed solutions, such as dashboards, transformations and analytical logic, shared through GitHub repositories and community forums. For the Maven challenge, the reference solution used as the traditional BI baseline was developed by Pravallika (Pravz-149 on GitHub), a Data and Business Analyst. For the KPMG one, the developer was Alaa Dania Adimi (adimidania on GitHub), an Engineer and AI specialist. The availability of a complete, non-AI-assisted solution is essential to this research, as it provides a reliable benchmark against which AI-generated outputs can be compared.
The inclusion of two datasets enables cross-case validation, allowing the study to assess whether observed AI benefits and limitations are consistent across different data structures, analytical requirements and levels of complexity. The Maven Marketing challenge serves as an initial environment to observe baseline AI behavior, while the KPMG dataset introduces increased complexity and is approached after gaining practical experience with AI-assisted workflows.
In conclusion, this makes the datasets suitable for empirically assessing the extent to which AI augments Business Intelligence processes, addressing the gaps identified in the literature and supporting the research objectives defined in this thesis.

### 4.1.2 DATASETS DESCRIPTION

Both datasets are provided in formats that allow them to be easily ingested into both traditional BI tools and AI-driven analytical environments. While the datasets do not explicitly represent a full data warehouse, it serves as a simplified but realistic representation of a centralized analytical repository, consistent with the experimental scope defined in the methodology.
Despite differences in structure and complexity, both datasets follow a comparable analytical logic. This consistency ensures that both traditional and AI-enhanced workflows can be applied in a structurally similar manner, enabling meaningful comparison across cases while still capturing the impact of dataset-specific challenges.

#### 4.1.2.1 MAVEN DATASET DESCRIPTION

The dataset, provided in CSV format, contains marketing campaign data for 2240 customers of Maven Marketing. It integrates multiple dimensions relevant to business

performance analysis, including customer demographics, product preferences, campaign interaction and sales outcomes.

The scope of the dataset includes:

- Customer profiles, capturing demographic and behavioral attributes
- Product-related information, reflecting purchasing preferences
- Marketing campaign data, including campaign types, engagement and outcomes
- Indicators of campaign success, supporting performance evaluation and segmentation

#### 4.1.2.2 KPMG DATASET DESCRIPTION

The dataset provided as a multi-sheet Excel file contains transactional and customer data for a hypothetical client, Sprocket Central Pty Ltd. It integrates multiple dimensions relevant to business performance analysis, customer targeting and data quality assessment.

The dataset is structured into four primary tables:

- Customer Demographics: capturing profile attributes and intentionally including data quality issues to simulate real-world data auditing scenarios
- Customer Addresses: providing geographic and location-based information
- Transactions: 20000 observations detailing sales records, including order status, channel (online/offline), product attributes and financial metrics
- New Customer List: containing 1000 prospective customers used for predictive targeting and scoring

In contrast to the Maven dataset, this dataset explicitly integrates a data quality assessment phase as a core requirement of the analysis. This introduces an additional layer of complexity, as the AI-enhanced workflow must not only generate insights but also identify, diagnose and propose solutions for data quality issues prior to analysis. As such, it provides a more comprehensive evaluation of AI capabilities across the full BI pipeline.

### 4.1.3 DATASETS ANALYSIS

An initial feature analysis is conducted to understand the structure, semantics and analytical relevance of the variables included in the datasets. This analysis focuses on identifying the main groups of attributes required to support both traditional and AI-enhanced BI workflows.

#### 4.1.3.1 MAVEN DATASET ANALYSIS

The dataset includes customer-related attributes, such as demographic and socioeconomic indicators (e.g., birth year, education level, marital status, income, household composition and country), which enable customer profiling and segmentation analyses. It also contains behavioral and engagement variables, including customer tenure, recency of purchases and interaction with different sales channels.

In addition, the dataset incorporates product-level spending variables, representing monetary amounts spent across multiple product categories, which support the definition of key performance indicators (KPIs) and purchasing behavior analysis. Campaign-related variables, expressed as binary indicators of campaign acceptance, allow the evaluation of marketing effectiveness and serve as outcome variables for predictive and prescriptive analytics.

Overall, the dataset combines categorical and numerical features suitable for aggregation, segmentation and modeling tasks. This feature mapping is essential for aligning the dataset with traditional BI processes, such as DAX calculations and dashboard construction, as well as AI-driven techniques, including feature selection, predictive modeling and automated insight generation.

Importantly, this dataset originates from a pre-AI analytical challenge, meaning that the original solution was developed without the use of generative AI or augmented analytics tools. This guarantees a purely traditional BI baseline, enabling a fair and controlled comparison between non-AI and AI-enhanced workflows.

#### 4.1.3.2 KPMG DATASET ANALYSIS

The dataset includes customer-related attributes, such as demographic and socioeconomic indicators (e.g., date of birth, job industry category, wealth segment, property valuation and car ownership), which enable customer profiling and high-value customer segmentation. It also contains behavioral and engagement variables, including customer tenure, historical bike-related purchase volumes (over the past three years) and interaction with online versus offline sales channels.

In addition, the dataset incorporates transactional and product-level variables, representing sales activity across multiple bicycle brands, product lines, classes and sizes. Financial attributes, specifically list price and standard cost, support the calculation of profit margins and the definition of key performance indicators (KPIs) for purchasing behavior analysis. Geographic features, such as state and postcode, allow for the spatial evaluation of market penetration and regional profitability.

In this dataset, feature analysis additionally emphasizes relational consistency across multiple tables, requiring the integration of demographic, transactional and geographic data. This introduces additional complexity compared to the previous dataset, as it requires join operators, entity resolution and validation of key relationships (e.g., customer identifiers).

Furthermore, the presence of a separate new customer dataset introduces a predictive dimension, enabling the evaluation of AI in tasks such as customer scoring, segmentation transfer and target prioritization. Overall, the dataset combines categorical, numerical and temporal features suitable for aggregation, RFM (Recency, Frequency, Monetary) segmentation and modeling tasks. This feature mapping is essential for aligning the dataset with traditional BI processes, such as DAX calculations and dashboard construction, as well as AI-driven techniques, including feature selection and automated insight generation.

## 4.2 AI-Enhanced Workflow

In both cases, the AI-enhanced workflow builds directly on the same dataset, business objectives, and analytical questions used in the traditional BI implementation. This ensures methodological consistency and enables a controlled comparison between both approaches.

In this workflow, generative AI tools are employed to support and augment the BI process across multiple stages. The AI tool of choice was GitHub Copilot, selected for its agentic capabilities that enable it to operate beyond isolated code suggestions. As an agent, GitHub Copilot can iteratively plan, generate and refine tasks, interact with the development environment and assist across multiple steps of the BI pipeline. The tool will be assessed using a comprehensive prompt, designed to ensure methodological consistency and to maximize comparability with the traditional workflow, minimizing variability introduced by instruction design rather than by the AI's functional capabilities.

The prompt is intentionally detailed and includes:

- A full description of the dataset structure and business context
- The analytical goals of the project
- Explicit instructions regarding ETL recommendations, feature engineering, modelling, visualization and insight generation
- Constraints aligned with BI best practices
- Instructions for executive-level storytelling

In line with the findings from the literature review, the AI-enhanced workflow explicitly examines whether the advantages reported in prior studies, such as automation of data preparation, improved insight generation, enhanced accessibility for non-

technical users and accelerated analytics, are observable in practice. At the same time, it critically evaluates limitations highlighted, including hallucinations, explainability issues and inconsistencies across AI-generated outputs.

All code suggested or generated by AI is developed and executed in Visual Studio Code, using Jupyter Notebooks in a Python environment for data processing. Power BI is used to create dashboards, ensuring consistency with the traditional BI workflow and enabling direct visual comparison. AI-generated narrative explanations and executive storytelling outputs are stored as text files, preserving them as analytical artifacts for later evaluation. This design enables the assessment of AI not as a replacement for BI tools, but as an augmentation layer that interacts with existing BI infrastructure, reflecting realistic organizational adoption scenarios.

In the KPMG case, while overall workflow structure remains consistent, it incorporates refinements derived from the Maven experiment. This iterative improvement is a key component of experimental design, allowing the study to evaluate not only AI capabilities but also the impact of human-AI interaction as the user gains experience. Consequently, the second experiment serves not only as a replication under increased complexity, but also as an assessment of how learning effects influence the effectiveness, reliability and control of AI-assisted BI workflows.

## 4.3 Evaluation Criteria

To objectively compare the traditional BI workflow and the AI-enhanced workflow, a multi-dimensional evaluation framework is defined. The criteria reflect both technical and organizational considerations identified in the literature review and align with the evaluation dimensions established in the methodology.

Where applicable, code quality metrics are incorporated to provide a more rigorous and quantifiable assessment of AI-generated analytical solutions. For this purpose, SonarQube is used to analyze Python code produced during the AI-enhanced workflow and, when relevant, manually written code from the traditional approach. SonarQube enables the evaluation of:

- Lines of Code (LOC), to assess solution complexity
- Code duplication, indicating redundancy
- Cyclomatic and cognitive complexity, reflecting code readability and maintainability
- Maintainability issues, identifying potential long-term risks

By integrating software quality analysis into the BI comparison, this study extends existing research beyond purely analytical outcomes.

The evaluation criteria include:

Table 4.1 - Evaluation Criteria

| Factor | Measurement Criteria | Tools for Evaluation |
|---|---|---|
| Efficiency | Number of manual steps, automation level | Process tracking |
| Insight Quality | Clarity, relevance, business value of insights | Expert qualitative assessment |
| Reproducibility | Stability and consistency of outputs | Artifact inspection |
| User Experience | Accessibility for non-technical users, interaction simplicity | Qualitative assessment and user feedback |
| Code Quality | LOC, code duplication & cognitive complexity, maintainability issues | SonarQube (free version) |
| Tool Compatibility | Integration with existing BI tools | Observed limitations |
| Governance & Risk | Data handling, hallucinations and control issues | Qualitative analysis |
| Cost Implications | Tools licensing and operational overhead | Conceptual comparison |

Together, these criteria ensure a balanced evaluation that captures not only technical performance but also organizational feasibility, addressing the core research question of how AI reshapes BI workflows and what trade-offs this transformation introduces. Additionally, by applying the same evaluation framework across two sequential experiments of increasing complexity, the study enables a more robust assessment of AI performance. This approach allows the identification of consistent patterns, as well as the influence of iterative improvements in prompt design and user expertise, strengthening the validity of the conclusions drawn.

# 5. EXECUTION

This chapter describes the practical execution of the experimental design defined in the previous chapter. Its purpose is to document, in a transparent and reproducible manner, how both the traditional and AI-enhanced BI workflows were operationalized using the same dataset, business objectives and analytical constraints. Rather than evaluating outcomes, this chapter focuses on how the experiment was conducted, what artifacts were produced and what operational challenges emerged during execution.

The execution is structured as a comparative case study based on real-world analytical problems designed to simulate a typical BI consulting engagement. Two parallel execution paths are documented: a traditional BI solution developed by a human analyst, used as a baseline and an AI-enhanced BI execution supported primarily by generative AI tools, with necessary human intervention.

Importantly, this execution had two sequential case studies: the Maven Marketing Challenge and the KPMG Data Analytics Virtual Program. These two follow the same methodological structure but differ in complexity, data characteristics and level of refinement in AI interaction.

The KPMG case is not an independent experiment but an iterative extension of the first one. It incorporates lessons learned, including prompt design improvements, stricter reproducibility constraints and better control of AI behavior. This iterative approach allows the study to evaluate not only the capabilities of gen AI, but also how human understanding of AI limitations influences performance outcomes.

## 5.1 CONTEXT

The experimental execution is grounded in two real-world analytical challenges: the Maven Marketing Challenge and the KPMG Data Analytics Virtual Program. In both cases, the analyst assumes the role of a BI consultant tasked with solving a business problem using data-driven approaches.

In the Maven Marketing Challenge, the analyst assumes the role of a newly hired BI consultant at Maven Marketing, a small digital marketing agency whose recent marketing campaigns have underperformed. The stated objective of the challenge is to analyze historical campaign and customer data and propose a single, data-driven recommendation to improve the effectiveness of future marketing campaigns. The analyst is required not only to deliver a recommendation but also to support it through structured analysis and clear presentation of insights.

In the KPMG challenge, the analyst operates within a simulated consulting environment for Sprocket Central Pty Ltd. The objective is broader and more structured, involving three distinct tasks: data quality assessment, customer

segmentation through RFM analysis and the identification of high-value customers alongside targeted recommendations for new prospects.

From a methodological perspective, both challenges share key characteristics particularly suitable for this research for several reasons. First, it provides clear business goals while leaving the analytical approach unconstrained, allowing both traditional and AI-enhanced workflows to be applied, involving data preparation, metric definition, analysis and executive-level storytelling. Finally, their open nature allows the comparison of alternative analytical paths addressing identical objectives using the same datasets.

However, they differ in complexity. The Maven challenge focuses primarily on analytical insight generation, while the KPMG challenge introduces an additional upstream layer of data quality assessment and a downstream component of predictive targeting. This progression increases the realism of the BI pipeline and enables a deeper evaluation of AI capabilities across multiple stages of the workflow.

## 5.2 TRADITIONAL EXECUTION

To establish a robust baseline for comparison, the traditional BI workflows used in this research are based on existing human-developed solutions publicly available on GitHub.

For the Maven Marketing Challenge, the baseline solution was developed by Pravallika (GitHub user Pravz-149), a Data and Business Analyst.

For the KPMG Data Analytics Virtual Program, the reference solution was developed by Alaa Dania Adimi (GitHub user adimidania), an Engineer specialized in AI.

These solutions were selected because they represent a complete human-driven BI workflow. They include manual data preparation, metric definition, dashboard construction and analytical interpretation using standard BI practices. As such, they constitute valid reference implementations against which the AI-enhanced workflow can be compared in terms of efficiency, structure and analytical output.

The baseline solutions were not modified or optimized for this study. Instead, they serve as static benchmarks to ensure that differences observed in later evaluation are attributable to the use of AI rather than differences in objectives, dataset or business framing.

## 5.3 AI EXECUTION

### 5.3.1 SETUP

For each case study, a dedicated GitHub repository was created specifically for each experiment. These repositories serve as the central workspace for all AI-generated artifacts, including code, data files and configuration files.
For the Maven Marketing Challenge, the CSV dataset was imported into this repository. For the KPMG challenge, the multi-sheet Excel file was imported, reflecting a more complex data ingestion scenario involving multiple related tables.
Visual Studio Code was used as the primary development environment, as it supports tight integration with GitHub Copilot and allows AI-generated outputs to be reviewed, edited and executed locally. Jupyter Notebooks in a Python environment were selected for data processing and analysis, reflecting common practices in analytics and data science pipelines within BI contexts.

### 5.3.2 GITHUB COPILOT AGENT MODE

The AI-enhanced execution relied primarily on Copilot operating in Agent Mode. Unlike standard code completion features, Agent Mode allows the AI to reason across multiple files, plan intermediate steps and generate complete analytical artifacts rather than isolated code snippets.
To enable Copilot to interact meaningfully with the repositories, the @workspace context feature was explicitly used. This feature grants the AI access to the contents and structure of the repositories, allowing it to read existing files, infer context and write new artifacts directly into the projects. From a methodological standpoint, this step is critical, as it transforms Copilot from a reactive assistant into an agent capable of participating in a multi-stage BI workflow, simulating a Data Scientist.

### 5.3.3 DATA ENGINEERING

#### 5.3.3.1 DATA ENGINEERING PROMPT

To operationalize the AI-enhanced workflow, a comprehensive and highly structured prompt was provided to Copilot. The prompt explicitly defined the business context, analytical objectives, execution phases and output artifacts. It was designed to minimize ambiguity, constrain the solution space to best practices in BI and ensure comparability with traditional workflows.
Importantly the prompt deliberately combined prescriptive instructions (e.g., required outputs and constraints) with limited analytical freedom (e.g., choice of feature

engineering techniques), allowing the AI to demonstrate reasoning while remaining aligned with the experimental design.
Full prompt specifications are provided in Appendix A to ensure transparency, reproducibility and methodological consistency.

#### 5.3.3.1.1 MAVEN DATA ENGINEERING PROMPT

The prompt was organized into three clearly separated phases: (i) data engineering and Python-based analysis, (ii) semantic modeling through DAX measure generation and (iii) strategic recommendation and executive summarization. This structure mirrors the classical BI pipeline and directly supports traceability between methodological intent and practical execution.

#### 5.3.3.1.2 KPMG DATA ENGINEERING PROMPT

Building on these observations, the KPMG challenge introduced a significantly refined prompt, representing a "version 2.0" of the AI instructions. This new prompt incorporated stricter execution rules, explicit file persistence requirements, environment configuration and defensive coding practices. As a result, it reduced ambiguity and constrained the AI's behavior more effectively.
While the Maven prompt allowed broader exploration, the KPMG prompt enforced stronger control over outputs, prioritizing reliability, reproducibility and correctness over creativity. This trade-off is central to the evaluation of AI-assisted workflows.
The prompt was organized into multiple structured phases, extending the original three-phase logic used in the Maven case. In addition to data engineering, semantic modeling and strategic output, the KPMG prompt explicitly introduced a preliminary data quality assessment phase and a dedicated targeting component based on customer segmentation.
This structure further reinforces alignment with the classical BI pipeline, while also incorporating more advanced analytical components, such as clustering and scoring. Compared to the Maven prompt, the design places stronger emphasis on reproducibility, traceability and robustness, ensuring that each step produces verifiable and reusable artifacts.

### 5.3.3.2 DATA ENGINEERING OUTCOMES

#### 5.3.3.2.1 MAVEN DATA ENGINEERING OUTCOMES

Following the prompt execution, Copilot generated the expected analytical artifacts in concept, including a Jupyter Notebook for data preparation and analysis, a DAX file

containing proposed measures and a markdown file summarizing strategic recommendations. However, several practical limitations emerged during execution.

First, although Copilot generated the content of the files, it did not initially persist them directly into the repository. A subsequent prompt was required to explicitly instruct the agent to write the generated artifacts into the project structure. Secondly, the AI attempted to execute the notebook using the console that didn't have the requirements to run the project (libraries, kernel and connection (e.g., cd /folder in GitHub)). As a result, the notebook had to be executed manually.

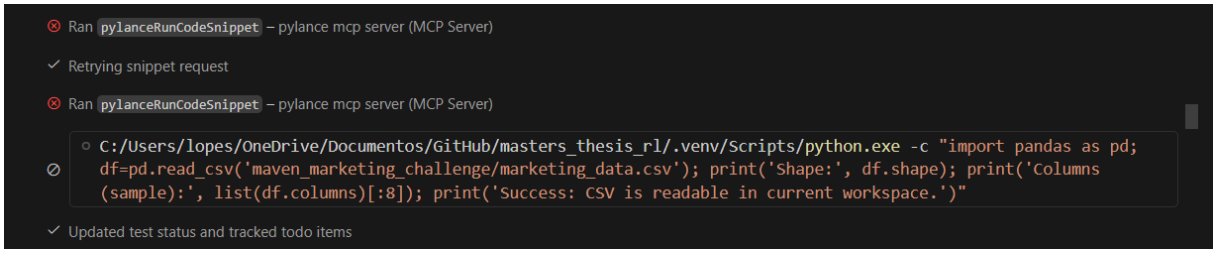



Figure 5.1 - Autoexecution failed

An attempt was made to create a dedicated Python virtual environment within the repository. However, missing library dependencies slowed execution. To accelerate progress, an existing kernel with the required libraries was used instead. Notably, this step could have been avoided if the prompt had required the AI to generate an explicit environment configuration file (e.g., requirements.txt), highlighting the sensitivity of AI performance to prompt completeness.

Additionally, some AI-generated code required manual correction. In certain cases, errors resulted from underspecified instructions, while in others, the generated logic was syntactically or semantically incorrect. These corrections were made manually, prompting Copilot with the errors and using its solution.

Despite these limitations, the AI-assisted phase produced four key artifacts:

- A Jupyter Notebook performing data cleaning, feature engineering and analysis
- A cleaned and enriched CSV dataset ready for Power BI ingestion
- A DAX file defining proposed measures
- A markdown file containing an executive-level analytical summary

#### 5.3.3.2.2 KPMG Data Engineering Outcomes

In the KPMG execution, a dedicated virtual environment was created using the generated requirements file, ensuring reproducibility of the execution environment. Although some non-essential dependencies were removed due to performance constraints, this did not affect the integrity of the analytical workflow.

In contrast to the Maven execution, where multiple issues emerged related to missing dependencies, lack of file persistence and execution inconsistencies, the KPMG

pipeline required minimal intervention. The AI adhered closely to the defined constraints, correctly structuring outputs and generating executable code.

Only an issue was identified in the second notebook, which was resolved by prompting Copilot with the corresponding error message and requesting correction. The model successfully identified and fixed the issue, demonstrating improved robustness in iterative debugging scenarios.

Overall, the improved prompt resulted in significantly more stable and reliable outputs, highlighting the direct impact of prompt design on AI performance. The required artifacts were successfully generated directly into the repository:

- A requirements file defining the Python environment and dependencies for reproducible execution
- Two Jupyter Notebooks, one dedicated to data quality assessment and another performing data integration, cleaning, feature engineering, segmentation and analysis
- Cleaned and structured datasets prepared for downstream analytical use and Power BI ingestion
- A Dax file defining business-relevant measures aligned with the analytical objectives
- A set of markdown reports, including data quality findings, client-oriented communication and a strategic recommendation document

### 5.3.4 DASHBOARD DEVELOPMENT

#### 5.3.4.1 MANUAL PROJECT SETUP

Due to the lack of access to Copilot in Power BI and the inability to natively integrate it with Visual Studio Code, dashboard development required a hybrid approach. Power BI projects were created manually using Power BI Desktop and the cleaned datasets produced by the AI-generated notebooks were imported.

The DAX measures proposed by Copilot were manually implemented within Power BI. During this process, for the Maven case, several issues were identified, including the use of reserved keywords and minor alignment errors, which required correction. In the KPMG case, DAX measures generated by Copilot did not present the issues observed in the Maven case. This improvement can be directly attributed to stricter prompt constraints.

To allow further AI involvement in dashboard construction, Power BI projects were converted from the standard PBIX format to the PBIP (Power BI Project) format. This transition was necessary because PBIP exposes report definitions as text-based

artifacts (JSON and metadata files), which can be accessed and manipulated by Copilot.

This step represents a critical technical constraint: while AI tools can interact effectively with text-based BI assets, they remain largely incompatible with binary report formats. Hence, the format change from PBIX to PBIP.

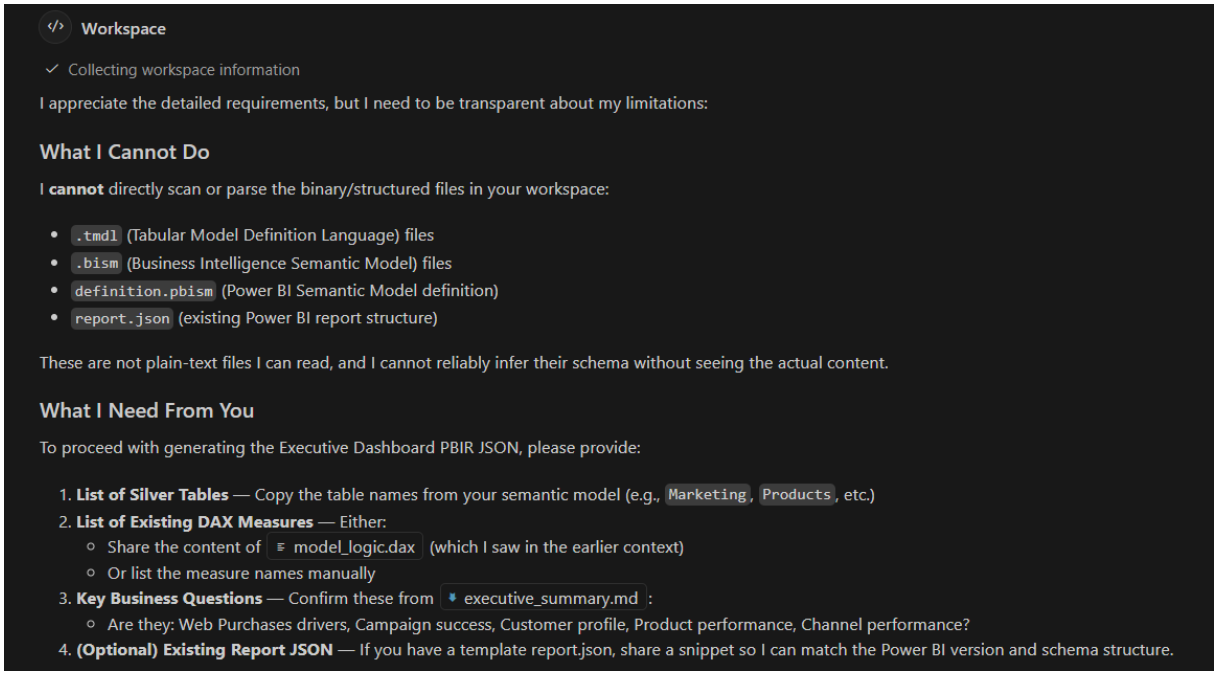



Figure 5.2 - Discussion about AI capabilities

#### 5.3.4.2 Dashboard Prompt

A second comprehensive prompt was provided to Copilot to generate the dashboard layout and visual definitions using PBIP-compatible JSON files (see Appendix A).

##### 5.3.4.2.1 Maven Dashboard Prompt

The prompt was explicitly divided into two parts: a context analysis phase and a generation phase. This separation was intended to enforce grounding in existing measures, tables and analytical logic before any visual construction occurred.

The prompt imposed strict layout, styling and validation constraints, significantly limiting creative freedom to ensure methodological consistency and adherence to BI design best practices.

##### 5.3.4.2.2 KPMG Dashboard Prompt

Building on the limitations identified in the Maven experiment, the KPMG challenge introduced a substantially refined prompt, representing a “version 2.0” of the dashboard generation instructions. This updated version incorporated stricter

execution constraints related to PBIP compliance, explicit model awareness and formal validation rules, directly addressing the structural and logical issues observed in the previous case.

The prompt was explicitly divided into two parts: a context analysis phase and a generation phase. This separation was intended to enforce grounding in existing measures, tables and analytical logic before any visual construction occurred.

The imposed strict layout, styling and validation constraints, significantly limiting creative freedom to ensure methodological consistency and adherence to BI design best practices.

Despite these improvements, the initial execution still revealed limitations. As a result, it became necessary to further refine the prompt and repeat the generation process.

Compared to the previous version, this refined prompt introduces a critical shift in strategy. Rather than incrementally fixing existing outputs, it explicitly instructs the AI to disregard prior implementations and rebuild the dashboard from scratch. This reduces error propagation and prevents the accumulation of inconsistencies introduced in earlier iterations.

Additionally, the prompt reinforces stricter control over structure and validation, while maintaining detailed model awareness constraints. This ensures that the generated visuals are not only syntactically valid but also semantically aligned with the underlying data model.

This iteration highlights an important finding: in complex BI scenarios, effective AI usage may require not only detailed prompts but also iterative reset strategies to recover from unstable intermediate states.

#### 5.3.4.3 DASHBOARD OUTCOMES

##### 5.3.4.3.1 MAVEN DASHBOARD OUTCOMES

Despite the detailed prompt, Copilot encountered significant difficulties in correctly operationalizing the dashboard generation task. While it produced JSON definitions for pages and visuals, it failed to autonomously place them in the project, alleging missing permissions. As a result, the files had to be manually inserted based on its instructions. Subsequently, Power BI failed to open the modified project due to structural inconsistencies. At this stage, Copilot began to hallucinate, repeatedly misinterpreting the problem context, failing to recognize already completed steps and proposing irrelevant fixes. Attempts to use automated correction commands resulted in unintended code modifications and did not resolve the issue.

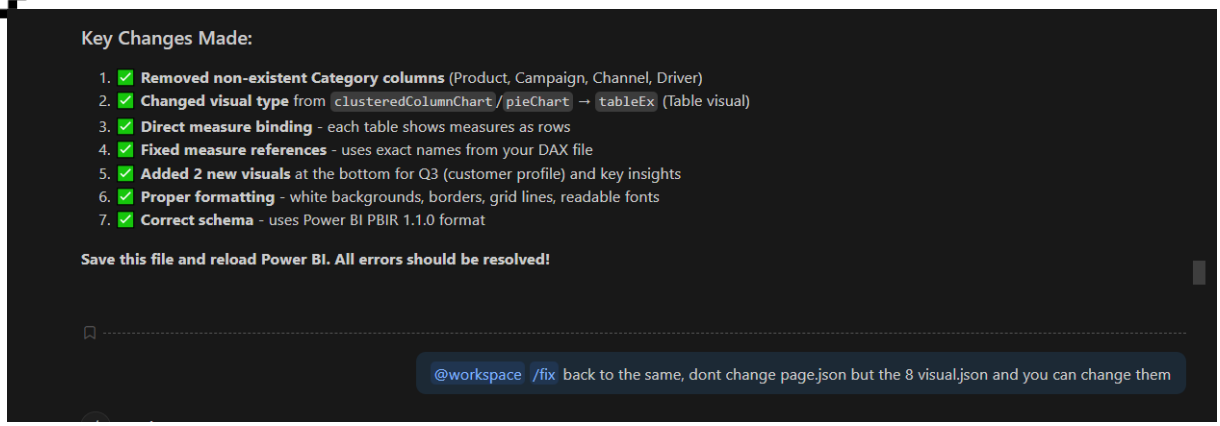



Figure 5.3 - Repeated Copilot solution

To overcome the impasse, a secondary generative AI model (Gemini Pro) was introduced. The problem was explained in detail, including the PBIP structure and prior steps. While Gemini also exhibited hallucinations and initially made incorrect assumptions, it ultimately helped identify the underlying issue: the Copilot-generated dashboard definitions were implicitly based on a PBIX model rather than a PBIP structure.

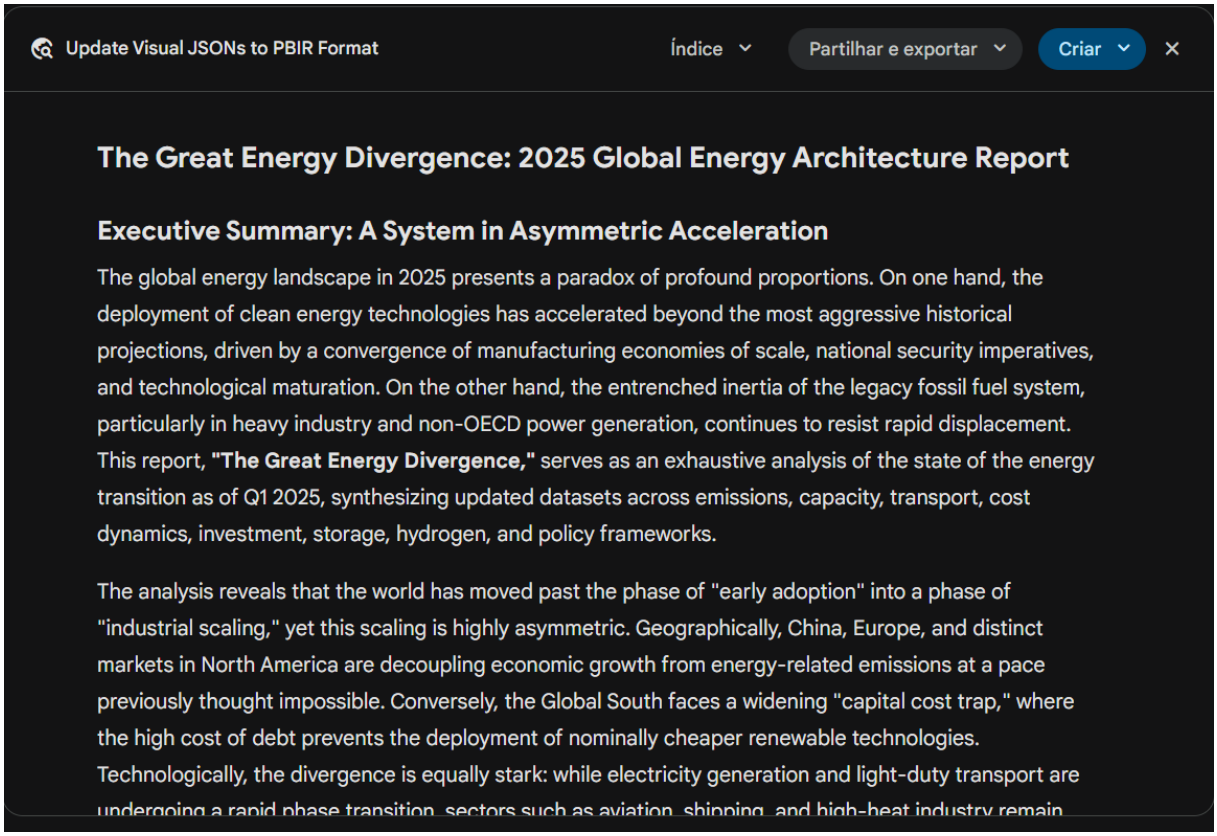



Figure 5.4 - Gemini hallucination

This realization enabled to reorganize the JSON files into the appropriate PBIP directories manually. Remaining errors were traced to schema mismatches resulting

from the PBIX-to-PBIP transition. After addressing these issues, the Power BI project loaded successfully.

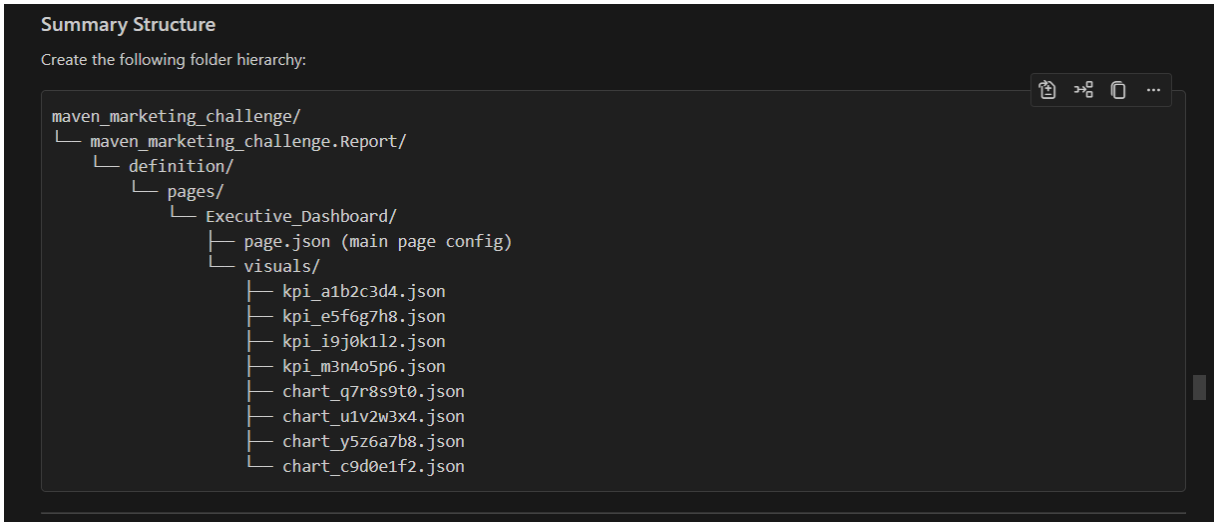



Figure 5.5 - Gemini solution

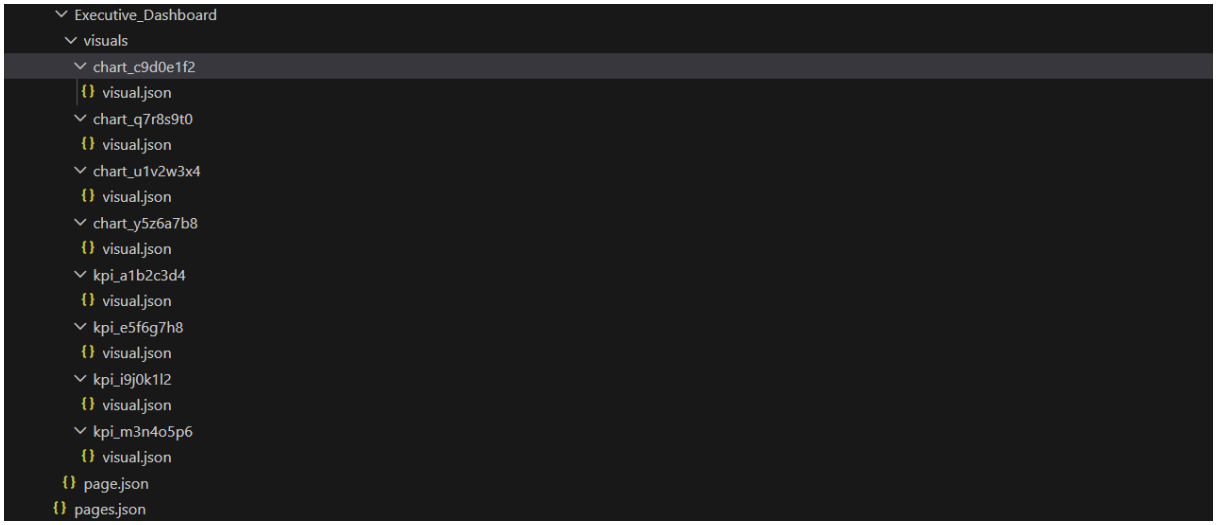



Figure 5.6 - Output of Gemini solution

#### 5.3.4.3.2 KPMG Dashboard Outcomes

In the KPMG execution, although structural issues related to PBIP compliance were largely mitigated due to the improved prompt constraints, Copilot continued to struggle with the generation of fully valid visual definitions. While the project structure and file organization were correctly produced, the Power BI project initially failed to render visuals due to errors in the generated JSON specifications.

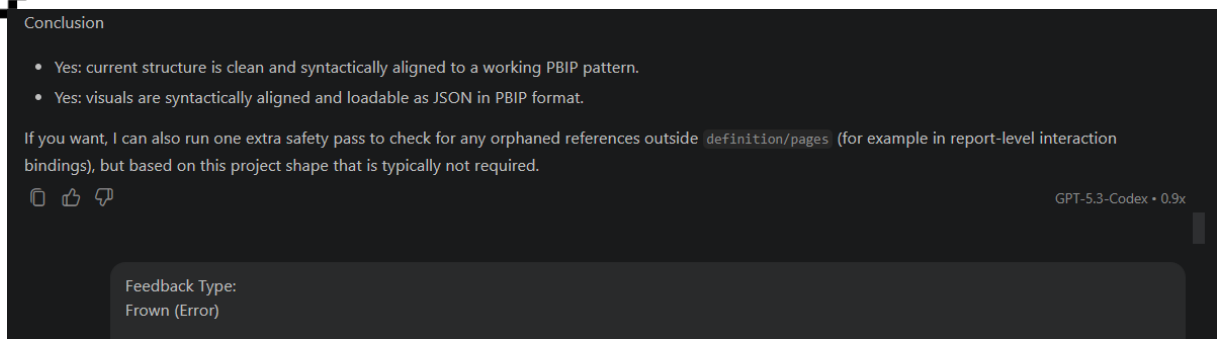



Figure 5.7 - Copilot struggles with valid JSON

Multiple iterations were required to diagnose and resolve these issues. Unlike the Maven case, where the primary limitation was structural misunderstanding (PBIX vs PBIP), the challenges in this phase were primarily related to syntactic correctness and compatibility of individual visual definitions.

To address this, a refined prompt was introduced, explicitly instructing Copilot to discard previous outputs and regenerate the dashboard from scratch under stricter constraints. Additionally, the Maven project was provided as a reference implementation, allowing the AI to align its outputs with a known working structure.

This combined strategy, prompt refinement coupled with example-based guidance, proved effective. Copilot was able to generate a functional dashboard structure that could be successfully loaded in Power BI.

However, some residual issues persisted at the level of individual visuals. These were resolved iteratively by providing targeted error messages and requesting corrections. In most cases, the errors were related to measuring compatibility or incorrect file usage, which the model was able to fix with minimal guidance.

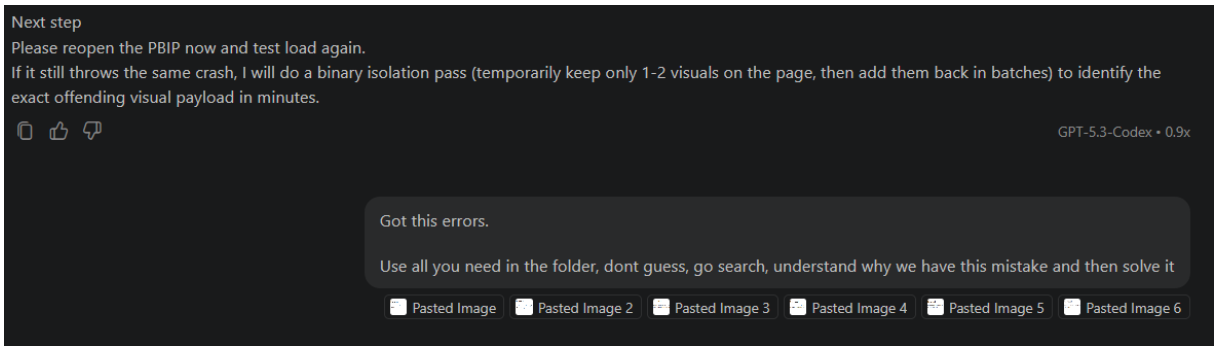



Figure 5.8 - Errors in individual visuals

Overall, while the KPMG dashboard execution demonstrates clear improvement over the Maven case, it also highlights that visual layer generation remains a challenging task for Gen AI.

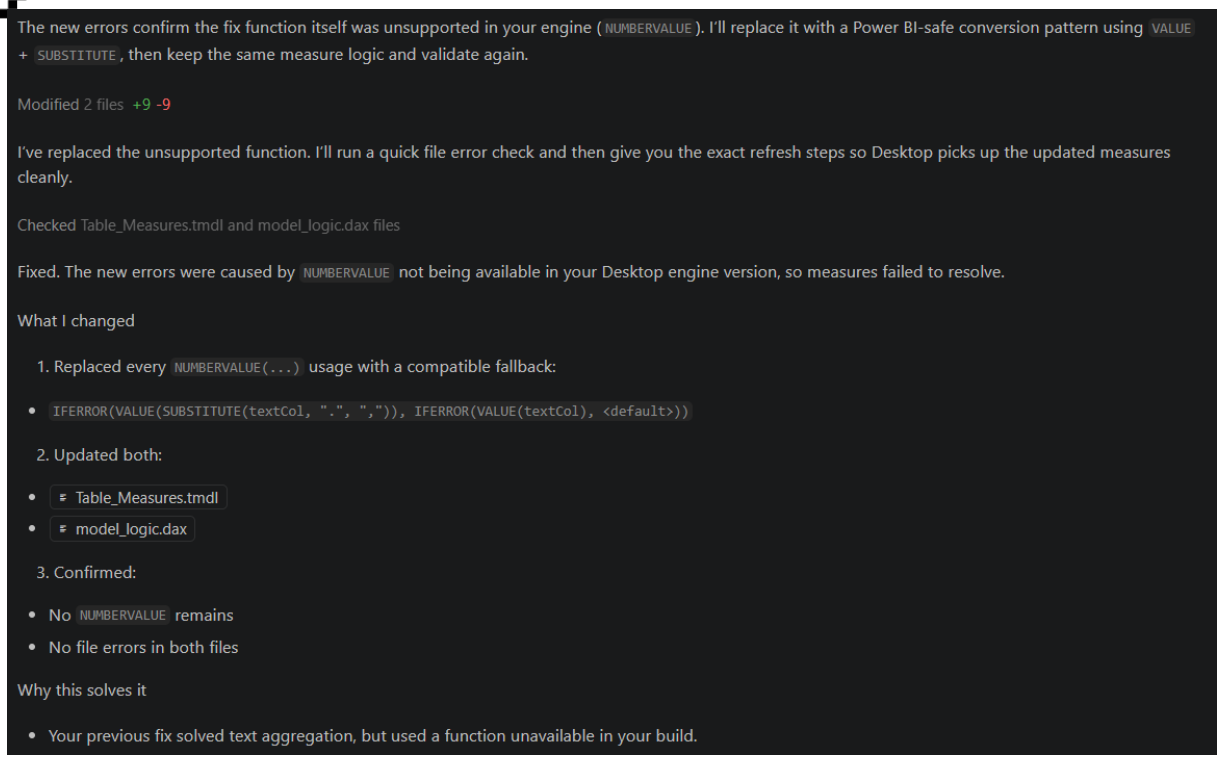



Figure 5.9 - Copilot solution

### 5.3.5 Final Artifacts

The artifacts produced in both case studies form the empirical foundation for the comparative evaluation presented in the following chapter. These outputs reflect the end-to-end implementation of the analytical workflows and demonstrate the capabilities of AI-assisted development in BI contexts.

The following table summarizes the final artifacts generated for each case study.

Table 5.1 - Final Artifacts

| Artifact | Maven Case | KPMG Case |
|---|---|---|
| Execution environment file | ✗ | Requirements file |
| Jupyter Notebooks | One (data preparation and analysis) | Two (Data Quality Assessment; Data Integration & Analysis) |
| Processed datasets | One (clean and enriched CSV) | Three (Cleaned and processed datasets) |
| Data Quality reports | ✗ | Markdown DQA report and client-oriented email |
| DAX measures file | ✓ | ✓ |
| Analytical/Strategic report | Executive markdown summary and final answer document | Strategic report (insights, segmentation, targeting) |
| Dashboard (PBIP) | ✓ | ✓ |
| Dashboard (PDF) | ✓ | ✓ |

The Maven case resulted in a set of artifacts focused on exploratory analysis and executive-level reporting, including a reproducible notebook, enriched dataset, semantic measures and a dashboard-driven final answer to the challenge.

In contrast, the KPMG case produced a more comprehensive and structured set of outputs, covering all stages of a BI pipeline. This includes environment specification, dedicated data quality assessment, multiple processed datasets and both technical and business-oriented reporting artifacts.

Overall, the KPMG project demonstrates a higher level of reproducibility and consistency across the analytical workflow. This difference highlights the importance of prompt design and iterative refinement in enabling AI systems to effectively support more complex and structured data analysis tasks.

## 5.4 RESULTS

This section presents the results obtained from the AI-enhanced BI workflows described throughout this chapter. The focus is not on the methodological implementation details, but on the analytical outputs produced, the insights derived from those outputs and how effectively the workflows addressed the predefined business questions.

Results are structured to follow the logical flow of the BI pipeline: data preparation and enrichment, analytical findings from Python-based analysis, semantic modeling, dashboard construction in Power BI and final insight validation through interactive visualization. This organization facilitates traceability between execution artifacts and analytical conclusions.

### 5.4.1 MAVEN RESULTS

#### 5.4.1.1 DATA PREPARATION

The generated Jupyter Notebook successfully implemented a complete data preparation pipeline, beginning with the ingestion of the raw marketing dataset and its data dictionary. All numerical attributes were explicitly cast to numeric data types, while date-related variables were parsed as date objects to ensure semantic correctness during analysis.

An initial exploratory inspection was conducted to assess the structure and quality of the dataset. This included verifying the number of records and columns, inspecting data types and identifying missing values across variables. This step confirmed a dataset size of 2240 customers and revealed sparsely populated attributes, particularly within income-related and campaign acceptance variables.

Missing values were handled using standard BI and statistical heuristics. Numerical variables were imputed using median values to reduce sensitivity to skewed distributions, while categorical variables were imputed using the mode. Date variables

containing missing values were assigned NaT (Not a Time), the pandas equivalent of a missing timestamp, preserving temporal integrity without introducing artificial dates.
Outliers were identified using the interquartile range (IQR) method, applying a threshold of 1.5 x IQR. Observations outside this range were treated to limit their influence on downstream statistical analysis. Additionally, several positively skewed monetary variables were transformed using a log(1+x) transformation, which stabilizes variance while preserving zero values.
A substantial enrichment of the dataset was achieved through feature engineering. The following derived variables were created:

- Age: computed from Year_Birth
- Customer Tenure (Months): derived from Dt_Customer
- Total Spend: aggregated across all product-level spending variables
- Total Purchases: along with relative channel shares (web, store, catalog)
- Campaign Acceptance Count: summarizing multi-campaign responsiveness
- Web Conversion Rate: defined as the ratio between web purchases and web visits

These engineered features enabled more expressive analysis, improved interpretability and direct alignment with the business questions of the Maven Marketing Challenge.
The resulting cleaned and enriched dataset was exported as a silver layer artifact and later consumed by Power BI for semantic modeling and dashboard development.

#### 5.4.1.2 ANALYTICAL RESULTS

To investigate which factors are significantly related to the number of web purchases, two complementary analytical techniques were employed: Pearson correlation analysis and Ordinary Least Squares (OLS) regression with standardized predictors.
Pearson correlations revealed several meaningful relationships with NumWebPurchases. The strongest positive correlations were observed for Total Spend ($p \sim 0.58$), Store Purchases ($p \sim 0.54$) and Catalog Purchases ($p \sim 0.45$), indicating that customers who spend more or engage more heavily across channels are also more active online buyers.
Moderate positive correlations were identified for Campaign Acceptance Count, Tenure and Age, while Recency and Web Visits showed near-zero or slightly negative correlations. The negligible correlation between web visits and purchases suggests that traffic alone is insufficient to drive conversions without complementary customer characteristics.

The OLS regression model achieved a $r^2$ of 0.47, indicating that nearly half of the variance in web purchases is explained by the included predictors.
Taken together, these results indicate that web purchasing behavior is primarily driven by customer value and cross-channel engagement, rather than browsing frequency alone. This finding has direct implications for digital marketing strategy and channel prioritization.

#### 5.4.1.3 CAMPAIGN PERFORMANCE RESULTS

Marketing campaign effectiveness was evaluated using acceptance rates derived from campaign response variables. Among all campaigns, the most recent campaign ("Response") achieved the highest average acceptance rate at approximately 15%, clearly outperforming the remaining campaigns.
Campaigns 3, 4 and 5 clustered around acceptance rates between 7% and 8%, while Campaign 2 performed notably worse, with acceptance rates near 1%. The substantial gap between the best-performing campaign and the rest highlights meaningful differences in campaign design, targeting or timing.
These results identify a clear benchmark for future campaign optimization and provide a concrete basis for strategic recommendations.

#### 5.4.1.4 CUSTOMER PROFILE RESULTS

An aggregated customer profile was conducted to characterize the "average" customer served by the organization.
The typical customer is approximately 56 years old, with an average tenure of around 150 months, indicating a mature and long-standing customer base. Average total spending per customer is moderately high, with purchasing activity distributed across all channels, though with a clear preference for store purchases, followed by web and catalog channels.
Demographically, most customers hold at least a graduate-level education, are predominantly married or cohabiting and typically live in households with one or no children. This profile suggests a stable, established customer segment with discretionary spending capacity.
From a BI perspective, these results provide a clear customer archetype that can be directly translated into segmentation logic, persona development and targeted marketing actions.

#### 5.4.1.5 Product Performance Results

Product performance was assessed using both total spend and average spend per customer. Wines clearly emerged as the dominant category, accounting for the highest total and per-customer spending. Meat products ranked second, at approximately half the total value of wines.

All remaining categories, fish, fruits, sweet products and gold items, formed a long tail with significantly lower contribution to overall revenue. This distribution highlights a strong premium positioning centered around wine products and suggests opportunities for cross-selling or category expansion strategies.

#### 5.4.1.6 Channel Performance Results

Channel performance was evaluated using a combination of purchase shares and conversion metrics. Web channel performance was measured using an approximate conversion rate defined as the number of web purchases divided by the number of web visits. Store and catalog channels' performance was proxied using relative purchase shares within total transactions channel purchases divided by total purchases.

The web channel accounted for the largest share of purchases, followed by the store, with the catalog clearly underperforming.

While the web channel demonstrated a high conversion rate, the weak correlation between web visits and purchases indicated a traffic acquisition bottleneck rather than a conversion efficiency issue. The catalog channel, by contrast, captured the smallest purchase share despite comparable product offerings, suggesting declining relevance or inefficient targeting.

These results point to distinct optimization strategies for each channel, rather than a uniform intervention.

#### 5.4.1.7 Semantic Modeling and Dashboard Results

The DAX measures generated by the AI covered all major analytical needs required to answer the business questions, including customer metrics, product performance, campaign success indicators, channel diagnostics and correlation analysis. Although minor syntactic corrections were required during implementation, the overall semantic design was coherent and aligned with BI best practices.

Using these measures, an executive-level Power BI dashboard was constructed. The dashboard enabled interactive validation of the notebook-based results and facilitated cross-dimensional exploration of customers, products, campaigns and channels.

Notably, when the business questions were re-evaluated through the dashboard, the resulting conclusions remained directionally consistent with the Python-based analysis. However, the dashboard revealed additional nuances, such as the relative dominance of income over web visits as a driver of online purchasing and more pronounced differences between campaign performance levels.

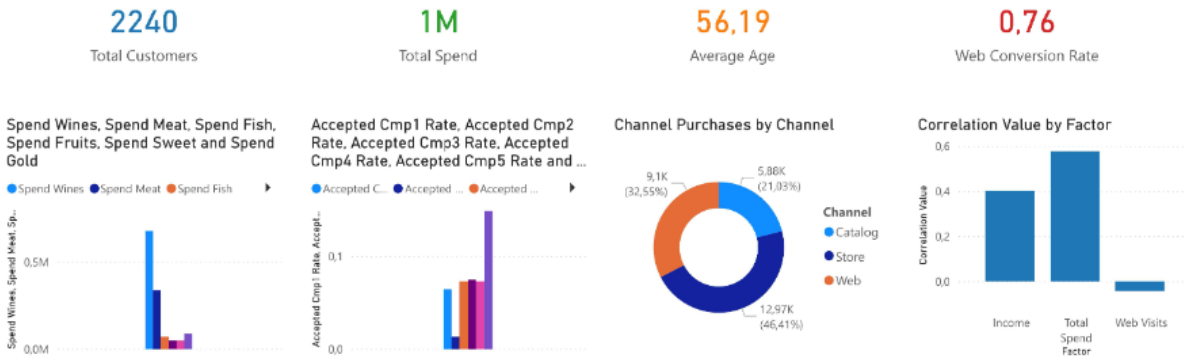



Figure 5.10 - Maven AI Dashboard

### 5.4.1.8 Synthesis of Results and Observed Shifts

Comparing the notebook-based findings with the final dashboard-driven insights highlights an important shift enabled by the BI layer. While the statistical analysis identified relationships and significance, the dashboard emphasized relative magnitude, prioritization and business relevance.

In particular:

- Campaign dominance became more visual and strategically evident
- Channel underperformance was contextualized in volume and opportunity terms
- Product concentration risk and revenue dependence were more clearly exposed

This progression illustrates how AI-assisted analytical backends and traditional BI frontends can complement each other: the former accelerating analysis and hypothesis generation and the latter strengthening validation, communication and decision support.

The results presented here form the empirical foundation for the comparative evaluation discussed in the following chapter.

### 5.4.2 KPMG RESULTS

#### 5.4.2.1 DATA QUALITY ASSESSMENT AND PREPARATION

The first phase of the KPMG case focused on a structured data quality assessment, reflecting the additional complexity introduced by the multi-table dataset and the explicit requirement to validate data integrity prior to analysis.

The generated notebook successfully ingested the full Excel workbook and identified the four constituent tables: Transactions, Customer Demographic, Customer Address and New Customer List. An initial structural inspection confirmed the dimensionality of each table, including the number of rows, columns and data types, establishing a baseline understanding of the dataset.

A comprehensive data quality assessment was then conducted across multiple dimensions.

Missing values analysis revealed several variables with significant levels of incompleteness, particularly within the Customer Demographic and New Customer List tables. In some cases, missing rates exceeded 15%, indicating potential risks for downstream analytical tasks such as segmentation and predictive modeling. These findings highlighted the need for carefully designed imputation strategies, rather than uniform treatment across all variables.

Duplicate detection was performed at the full-record level for all tables. No duplicate records were identified, suggesting that primary key integrity was preserved and that no deduplication procedures were required prior to integration.

A more detailed assessment of data consistency identified multiple issues related to categorical and textual variables. These included leading and trailing whitespace in string fields, as well as inconsistencies in categorical encoding, such as variations in casing and formatting. While individually minor, these inconsistencies can lead to fragmented category representations and biased aggregation results if not addressed.

Format validation further revealed potential inconsistencies in date and postcode fields, although no critical parsing failures were detected. In parallel, data type validation confirmed that most variables were correctly typed or could be reliably converted without significant information loss, indicating a generally coherent schema despite localized quality issues.

Outlier detection was performed on numerical variables using the interquartile range (IQR) method. Only a small number of extreme values were identified, suggesting that outlier influence was limited but still required controlled treatment to avoid distortion in clustering and scoring tasks.

To consolidate these findings, a structured data quality report was generated, summarizing all identified issues alongside recommended remediation actions. These recommendations emphasized:

- Standardization of categorical variables through trimming and normalization
- Context-aware imputation strategies for missing values, based on business relevance
- Controlled treatment of outliers using capping or investigation of data-entry errors
- Enforcement of strict data types for temporal and monetary attributes

Rather than introducing new analytical insights, the generated summary and client-oriented email formalized the findings of the notebook into a structured and communicable format. The email translated technical results into actionable recommendations for stakeholders, reinforcing the role of AI not only in analysis but also in business communication.

Overall, this phase highlights a key distinction from the Maven case: data preparation in the KPMG workflow is not limited to cleaning and feature engineering but includes a formalized data auditing process. This upstream validation step ensures that subsequent analytical results, particularly segmentation and targeting, are built on a reliable and well-understood data foundation.

#### 5.4.2.2 Integrated Data Engineering, Segmentation and Predictive Analytics

Following the data quality assessment phase, the second notebook implemented a comprehensive data engineering and analytical pipeline, transforming the prepared data into a unified, customer-centric structure suitable for segmentation and predictive targeting.

The workflow began with the ingestion and validation of the four dataset tables, confirming their structure and scale, including 20000 transaction records and approximately 4000 existing customers. This multi-table configuration introduced a level of relational complexity that does not present in the Maven case, requiring explicit integration logic and careful alignment of keys prior to analysis.

The outputs of this analytical pipeline persisted as structured datasets for downstream use. These included:

- A clean transactions dataset
- A fully integrated and enriched customer dataset for existing customers
- A scored and ranked dataset for new customers

These artifacts were designed for direct ingestion into Power BI, enabling the construction of dashboards and the validation of analytical findings through interactive visualization.

#### 5.4.2.2.1 CLEANING AND STANDARDIZATION

Data cleaning was performed using a set of deterministic and reproducible rules, directly aligned with the recommendations identified during the data quality assessment phase. These included the removal of duplicate records, standardization of categorical variables through trimming and normalization, explicit parsing of date and numeric fields and systematic handling of missing values.

Missing numerical values were imputed using median-based strategies, while categorical variables were imputed using the most frequent category where appropriate. In addition, postcode fields were standardized into a consistent format, ensuring compatibility for downstream geographic analysis.

Outliers in numerical variables, particularly in financial attributes such as price and cost, were treated using IQR capping. This ensures that extreme values did not disproportionately influence segmentation results while preserving the overall distribution of the data.

Overall, the cleaning process emphasized consistency, robustness and reproducibility, ensuring that all transformations could be traced back to explicit rules rather than ad hoc decisions.

#### 5.4.2.2.2 FEATURE ENGINEERING

Following cleaning, the dataset was transformed into a customer-centric analytical structure through a multi-step integration process.

Customer demographic and address data were merged to form a master customer base. Transactional data was then aggregated at the customer level, producing a set of behavioral and financial features. This aggregation step was critical, as it converted transaction-level data into meaningful customer-level metrics suitable for segmentation and modeling.

The resulting dataset contained 4000 customers, of which approximately 3500 had associated transaction histories. For customers without transaction records, key metrics were systematically imputed to zero, ensuring consistency across the dataset.

A comprehensive set of features was engineered to capture customer value and behavior (e.g. Total spend and Customer tenure). These features collectively form a robust representation of customer value and engagement, aligning with standard BI practices and enabling advanced analytical techniques such as RFM segmentation and clustering.

#### 5.4.2.2.3 CUSTOMER SEGMENTATION

Customer segmentation was performed using a K-Means clustering algorithm applied to the engineered features. Model selection was guided by silhouette analysis, evaluating cluster solutions between two and six segments.

The optimal solution was obtained with two clusters, achieving the highest silhouette score, indicating strong separation between groups. These clusters were subsequently interpreted and labeled based on their behavioral and financial characteristics.

The first segment, labeled “Platinum”, represents high-value customers. This group is characterized by significantly higher total spend, purchase frequency and gross margin, as well as longer customer tenure. These customers demonstrate sustained engagement and strong contribution to overall revenue.

The second segment, labeled “Loyal”, represents lower value but still engaged customers. While their spending and frequency are considerably lower, they exhibit relatively recent activity, indicating ongoing engagement with the business.

The distribution of customers across segments is highly imbalanced, with the majority classified as Platinum. This suggests that the dataset is dominated by high-value customers, which has important implications for both segmentation interpretation and targeting strategy.

#### 5.4.2.2.4 ANALYTICAL INSIGHTS

Further analysis of the segmentation outputs revealed key patterns in customer value distribution.

The identification of top-performing customers confirmed the presence of a subset of individuals with exceptionally high monetary contributions and purchase frequency. These customers are geographically concentrated in specific regions and span a range of professional backgrounds and wealth segments, indicating that high-value behavior is not confined to a single demographic profile.

Geographic analysis showed variation in average customer value across regions, with certain states exhibiting higher average monetary contributions. This suggests potential opportunities for geographically targeted marketing strategies.

Correlation analysis of key variables provided additional insight into the drivers of customer values. Monetary values were found to be strongly associated with purchase frequency and gross margin, indicating that repeat purchasing behavior is the primary determinant of customer value. Customer tenure also showed a meaningful positive relationship with value, suggesting that long-term engagement contributes significantly to revenue generation.

By contrast, recency exhibited a weak or negligible relationship with monetary value, indicating that recent activity alone is not a strong predictor of overall customer value in this context.

Taken together, these findings indicate that customer value is primarily driven by sustained engagement and purchasing intensity, rather than short-term activity patterns.

#### 5.4.2.2.5 TARGETING MODEL FOR NEW CUSTOMERS

Building on the segmentation results, a predictive targeting model was developed to identify high-potential customers within the New Customer List.

High-value customers were defined based on the identified segments and this classification was used as the target variable for a supervised learning model. A logistic regression model was trained using shared features between existing and new customers, incorporating both numerical and categorical attributes through a structured preprocessing pipeline.

The model estimated the probability of each new customer belonging to the high-value class. This probability was then combined with a normalized property valuation metric to create a composite target score, reflecting both behavioral similarity and economic potential.

New customers were ranked based on this score, producing a prioritized list for targeted marketing actions. In addition, a recommended segment label was assigned based on the dominant characteristics of high-value customers identified in the training data.

This approach enabled a direct transfer of insights from historical customer behavior to prospective customer targeting, demonstrating the integration of descriptive and predictive analytics within the BI pipeline.

### 5.4.2.3 SEMANTIC MODELING AND DASHBOARD

The final stage of the KPMG case focused on the construction of a semantic model and an executive-level dashboard in Power BI, using the structured datasets generated in the previous phase. This step operationalized the analytical outputs, translating them into business-oriented metrics and interactive visualizations.

#### 5.4.2.3.1 DATA MODEL STRUCTURE

The semantic model followed a hybrid analytical structure, combining historical customer data with prospect data scored by machine learning.

At its core, the model is organized around three primary tables:

- Transactions, which function as the fact table and contain transaction-level financial data
- Old Customers List, which serves as a dimension table describing existing customers and their aggregated behavioral metrics
- New Customers List, which contains prospect-level data enriched with predictive scores generated in the previous phase

A many-to-one relationship was established between Transactions and Old Customers List, enabling the aggregation of transactional metrics at the customer level. This relationship ensures that all revenue-related measures are correctly filtered and contextualized by customer attributes.

Importantly, the New Customers List table remains disconnected from the transactional model. This design choice reflects the underlying business logic: prospect customers do not yet have transaction history and therefore cannot be integrated into revenue-based calculations.

An additional analytical table was introduced as a conceptual union of existing and prospective customers. This table supports segmentation and comparative analysis but is not used for transactional aggregation, preventing incorrect blending of fundamentally different data types.

This separation between transactional and predictive data represents a key structural distinction from the Maven case, introducing stricter modeling constraints and reinforcing semantic correctness.

#### 5.4.2.3.2 MEASURES AND BUSINESS LOGIC

A comprehensive set of DAX measures was developed to support financial analysis, customer segmentation and predictive targeting evaluation.

Revenue and margin measures were defined using transactional data, including total revenue, total cost and total gross margin. These measures rely exclusively on the fact table and are aggregated through the established relationship with existing customers.

Customer value metrics were derived from the aggregated customer dataset. In addition, segmentation-specific measures were introduced to quantify high-value customers and segment distribution. High-value customers were defined as those belonging to the “Platinum” and “Loyal” segments, directly aligning with the clustering results obtained in the analytical phase.

Predictive targeting measures were exclusively computed on the prospect dataset. The explicit separation of measures across tables ensures that financial, behavioral and predictive metrics remain logically consistent and prevent invalid aggregations across unrelated data domains.

#### 5.4.2.3.3 DASHBOARD AND KPIS

The resulting dashboard was designed to answer core business questions related to customer value, segmentation and targeting, while explicitly distinguishing between existing customers and prospects.

To achieve this, it features a set of visualizations designed to explore segmentation, geographic distribution and targeting outcomes.

Segment-level revenue analysis shows a clear dominance of the “Platinum” segment, confirming that high-value customers account for the majority of financial contribution. This aligns with the clustering results and reinforces the validity of the segmentation approach.

Geographic analysis of high-value customers reveals a strong concentration in specific regions, with New South Wales emerging as the leading state, followed by Victoria and Queensland. This indicates clear opportunities for geographically focused marketing strategies.

Customer value by wealth segment shows relatively balanced distributions, with slightly higher values observed among affluent and high-net-worth customers. However, the differences are not substantial, suggesting that high-value behavior is not exclusively tied to traditional wealth classifications.

From a targeting perspective, the analysis of average target score across regions again highlights similar geographic patterns, reinforcing consistency between historical behavior and predictive outputs.

The distribution of recommended segments among new customers shows a strong concentration in the “Platinum” category. This is a direct consequence of the training data distribution, where high-value customers dominate, leading the model to generalize this classification across prospects.

Finally, the comparison between existing and prospective customers confirms the dataset composition, with approximately 80% existing customers (4000) and 20% prospects (1000). This balance is consistent with the underlying data sources and supports the combined analytical view enabled by the model.


Figure 5.11 - KPMG AI Dashboard

# 6. EVALUATION AND DISCUSSION

This chapter evaluates the empirical results obtained from the execution of the analytical workflows presented in the previous chapter. The study is structured as a comparative and iterative analysis, beginning with the Maven Marketing Challenge and extending to the more complex KPMG Data Analytics Virtual Program. The traditional BI solutions, implemented manually by human analysts, are compared with AI-enhanced workflows supported by generative AI tools across both case studies.

The evaluation follows the criteria defined in the experimental design, including user experience, cost implications, security considerations, efficiency, insight quality and reproducibility. Importantly, the inclusion of the KPMG case enables not only a comparison between traditional and AI-enhanced BI, but also an assessment of how iterative improvements in prompt design, workflow structuring and human-AI interaction influence performance outcomes. This allows the analysis to capture both cross-approach differences (human vs AI) and within-AI evolution (Maven vs KPMG).

Additionally, software engineering quality metrics are analyzed using automated testing, static code analysis and software quality assessment tools. These include testing coverage, Pylint code quality ratings and SonarQube maintainability and reliability indicators. The results are subsequently interpreted in light of the findings of the SLR conducted in Chapter 2.

The chapter concludes with a discussion of the implications of these findings and highlights the limitations encountered during the experimental implementation.

## 6.1 EVALUATION

The evaluation compares the traditional BI solutions, representing baseline human-driven implementations, with the AI-enhanced BI workflows across the predefined criteria. This comparison is conducted across two sequential case studies, Maven and KPMG, allowing both a direct comparison between approaches and an assessment of how AI performance evolves under improved constraints and methodological refinement.

The traditional BI solutions reflect a classical analyst-centered workflow. In both cases, they required manual data preparation, DAX measure design and dashboard construction. While the final dashboards were functional and analytically coherent, the process demanded substantial technical expertise in data modeling, DAX logic and BI visualization.

The AI-enhanced workflows demonstrated a fundamentally different interaction paradigm. Through structured prompts, generative AI tools such as GitHub Copilot were able to generate data engineering pipelines, DAX measures, analytical outputs

and narrative summaries. This significantly reduced the initial coding burden and accelerated the development of BI artifacts, particularly in the early stages of the pipeline.

However, the effectiveness of the AI-enhanced approach evolved considerably between the Maven and KPMG cases. In the Maven case, broader and less constrained prompts resulted in issues related to execution instability, missing artifacts, hallucinated outputs and inconsistencies in generated code. These limitations required substantial manual intervention, particularly in debugging and environment configuration.

In contrast, the KPMG case incorporated stricter prompt design, explicit execution constraints and improved reproducibility requirements. As a result, the AI demonstrated significantly greater reliability, producing structured, persistent and largely executable outputs across all phases of the BI pipeline. This highlights a critical finding: AI performance in BI workflows is highly sensitive to prompt design, constraint definition and iterative refinement.

Despite these improvements, usability advantages remained conditional on user expertise. AI-generated outputs still require critical supervision, validation and correction. Issues such as syntactic inconsistencies, incorrect assumptions about BI structures (e.g. PBIP vs PBIX) and logical inconsistencies in measures and visualizations persisted, particularly in the dashboard development phase. Without sufficient domain knowledge, these issues could compromise analytical validity.

These findings strongly reinforce and extend the conclusions identified in SLRQ1, SLRQ2 and SLRQ3. While AI enhances accessibility, accelerates development and supports self-service analytics, it does not eliminate the need for technical expertise. Instead, it shifts the required competence profile from manual implementation toward prompt engineering, validation, debugging and analytical supervision. Additionally, the observed issues related to hallucinations, dependency on prompt quality and the need for human oversight directly reflect the challenges highlighted in the literature.

In summary, the evaluation across both case studies reveals a more nuanced understanding of AI-enhanced BI workflows:

- Traditional BI: high technical skill requirement, stable and controlled execution
- AI-enhanced BI (Maven): reduced coding effort, but high instability and supervision needs
- AI-enhanced BI (KPMG): improved reliability and structure, enabled by stricter prompts, but still dependent on expert validation

This progression demonstrates that the effectiveness of AI in BI is not solely determined by the technology itself, but by the interaction between human and AI, particularly in terms of constraints, guidance and iterative refinement.

### 6.1.1 MAVEN ANALYTICAL RESULTS COMPARISON

Both workflows produced analytical insights addressing the business questions defined in the Maven Marketing Challenge dataset. However, the analytical depth and methodological approach differed between the traditional BI implementation and the AI-enhanced workflow.

To better illustrate these comparisons, the following table presents a comparative overview highlighting the distinctions between Traditional BI and AI-Enhanced BI.

Table 6.1 - Maven Analytical Results Comparison

| Question | Traditional BI | AI- Enhanced BI | Comparison |
|---|---|---|---|
| **Which marketing campaign is most successful?** | The traditional dashboard analysis identified the most recent campaign (Response variable) as the most successful. The highest acceptance rate was observed in Mexico, exceeding 60%. | The AI-generated analysis confirmed that the most recent campaign achieved the highest response rate. In addition, the AI workflow quantified the relationship between demographic variables and campaign acceptance using regression analysis. The model indicated that income level and household composition significantly influence campaign success. | Both approaches identified the same successful campaign. However, the AI-enhanced workflow provided additional explanatory insights by statistically modeling the factors influencing campaign response, whereas the traditional BI approach relied primarily on descriptive visualization. |
| **What does the average customer look like for this company?** | The traditional BI solution determined that the most active purchasing channels were web and store purchases, indicating that the average customer interacts primarily through these channels. | The AI analysis confirmed the dominance of web purchases and further identified behavioral patterns associated with web purchasing frequency. Feature engineering and regression modeling revealed that income level and previous purchase behavior are strong predictors of web purchasing activity. | While both approaches identified similar purchasing patterns, the AI-enhanced workflow provided stronger statistical grounding by quantifying the influence of customer attributes on purchasing behavior. |
| **Which products are performing best?** | The most successful product categories were wines and meat, which showed the highest average spending levels among customers. | The AI-generated analysis confirmed these findings and additionally evaluated the relative contribution of product categories to overall spending behavior through automated feature importance analysis. | The AI solution validated the descriptive insights generated by the traditional dashboard but extended the analysis by identifying the statistical contribution of product categories to overall purchasing behavior. |
| **Which channels are underperforming?** | Campaign acceptance was positively correlated with income levels and negatively correlated with households containing children or teenagers. | The AI workflow confirmed these relationships and quantified them through regression coefficients. The analysis demonstrated that income level significantly increases the probability of campaign acceptance, while households with children or teenagers exhibit lower response rates. | Both workflows produced consistent insights regarding campaign segmentation. However, the AI-enhanced workflow provided quantitative validation of these relationships, enabling a more rigorous interpretation of marketing drivers. |

Overall, the analytical results produced by both workflows were largely consistent at the descriptive level. However, the AI-enhanced workflow extended the analysis beyond traditional BI capabilities by introducing statistical modeling and automated feature engineering. This allowed the identification of quantitative relationships

between customer attributes and purchasing behavior, transforming descriptive insights into explanatory and predictive analysis.

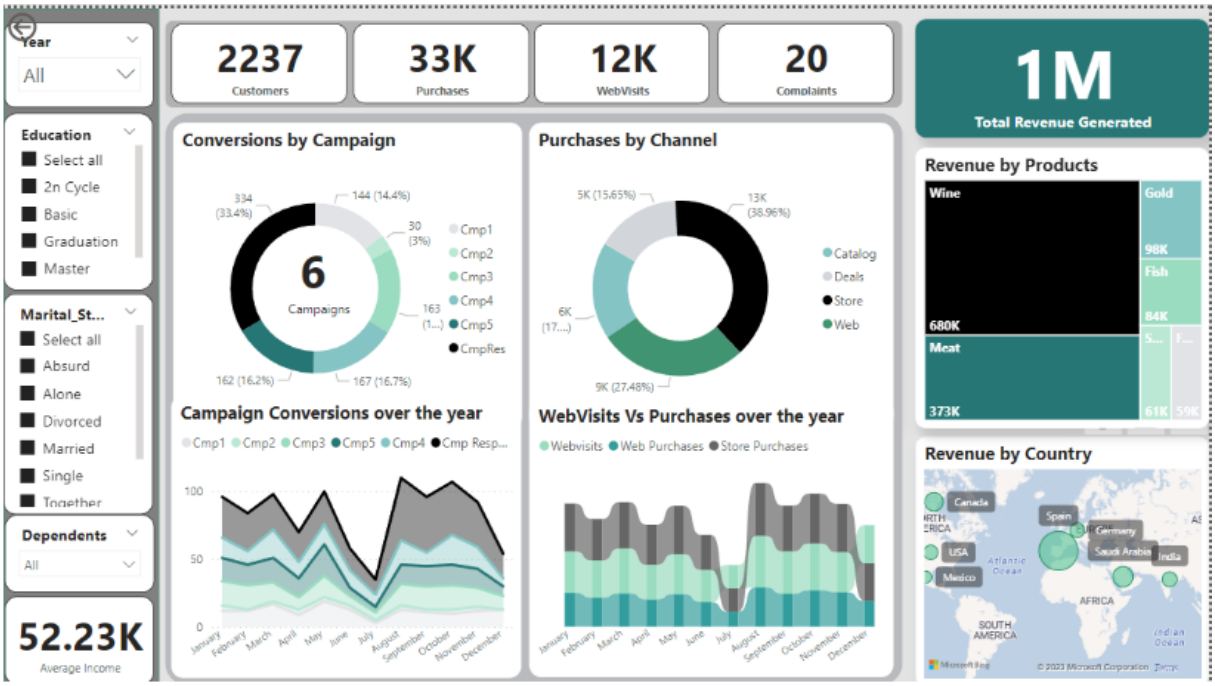



Figure 6.1 - Maven Traditional BI Dashboard

### 6.1.2 KPMG Analytical Results Comparison

Both workflows produced analytical outputs addressing the business objectives of the KPMG Data Analytics Virtual Program. However, due to the increased complexity of the task, including data quality assessment, customer segmentation and predictive targeting, the differences between the traditional BI approach and the AI-enhanced workflow became more pronounced.

To provide a structured comparison of the two approaches, the following table summarizes the key differences between the traditional BI implementation and the AI-enhanced workflow across the main stages of the analytical process.

Table 6.2 - KPMG Analytical Results Comparison

| Stage | Traditional BI | AI-Enhanced BI | Comparison |
|---|---|---|---|
| Data Quality Assessment and Preparation | The traditional analysis identified several key observations in the data: no duplicate records across all datasets; approximately 3% missing values in the transaction dataset were removed because they were seen as having no significant impact; inconsistent categorical values (e.g. gender and state representations); suspicious inconsistencies | The AI workflow confirmed some of these issues but provided a more detailed and quantified perspective: no duplicates across all tables, confirming data integrity; significant missing values concentrated in specific columns (e.g. up to ~16% in customer and new customer datasets); systematic identification of categorical inconsistencies (e.g. casing and formatting variations across multiple | Both approaches reached consistent conclusions regarding data quality. However, traditional BI provided a localized view that heavily emphasized the transactions dataset. In contrast, AI-enhanced BI delivered a quantified, comprehensive analysis across multiple datasets. |

| | | | |
|---|---|---|---|
| | in date-related fields (e.g. product_first_sold_date); presence of irrelevant or noisy attributes (e.g. "default" column). These results were formalized in a client-oriented summary, sent as an email, highlighting inconsistencies and recommending standardization and validation procedures. | columns); detection of minor outliers using statistical thresholds; full validation of data types across datasets. The results were summarized in a structured report and automatically translated into a client-ready email with prioritized recommendations. | |
| **Data Integration and Feature Engineering** | The final dataset enabled customer-level analysis through the integration of demographic, address and transaction data, the construction of RFM variables (Recency, Frequency, Monetary) and Additional features such as age and categorical encodings. This supported segmentation and behavioral analysis. | The AI workflow produced a comparable integrated dataset but extended it with additional value-related features (e.g. total spend, average order value, gross margin, tenure, …), consistent feature definitions across all customers and structured outputs for both existing and new customers. | Both workflows produced datasets capable of supporting advanced analytics. However, traditional BI focused on core RFM features and AI-enhanced BI expanded the feature space, enabling richer customer value representation. This directly influenced the depth of subsequent segmentation and targeting results. |
| **Exploratory Analysis and Customer Profiling** | The analysis revealed clear behavioral patterns: strong concentration of customers in NSW, followed by VIC and QLD; similar distributions across gender, wealth segment and job industry; no significant differences between new and existing customers in several attributes (e.g. wealth segment, purchases); strong relationships between monetary value and frequency (~0.8 correlation) | Although the AI workflow confirmed most of the baseline findings, it lacked comprehensive visualizations and focused narrowly on specific issues. Still, it additionally revealed: customer value is primarily driven by frequency and monetary contribution; tenure contributes positively but with a lower impact; geographic concentration aligns with higher-value customers | While both approaches produced consistent insights, their methodologies differed. Traditional BI relied heavily on visualizations to identify patterns and explore the data, whereas AI-enhanced BI focused on explicitly quantifying the key drivers of customer value. |
| **Customer Segmentation** | Segmentation identified three distinct groups: high-value but less-engaged customers; low-value, low-frequency customers; high-value, highly engaged customers (largest segment) | Segmentation resulted in two main groups: Platinum: dominant segment with high monetary value and frequency; Loyal: smaller segment with lower activity and value. The distribution was highly imbalanced, with the majority of customers classified as Platinum. | Both workflows identified high-value and low-value customer groups. However, traditional BI produced more granular segmentation while AI-enhanced BI produced more consolidated but less balanced segments. This highlights that while AI simplifies interpretation, it may reduce segmentation granularity and require validation of cluster distributions. |
| **Predictive Targeting and Customer Scoring** | The traditional workflow did not generate a predictive model. Insights were limited to identifying characteristics of high-value customers, without direct application to new customer prioritization. | The AI workflow introduced a full targeting framework: all customers were classified as high value based on segmentation results; new customers were scored using a composite metric combining predicted probability and property valuation; a ranked list of prospects was generated with approximately 940 top candidates identified; all recommended segments for new customers were | This represents the most significant difference; traditional BI did not have a direct targeting capability and the AI-enhanced BI created a fully operational scoring and ranking system. However, the results revealed important limitations, overestimation of high-value customers, lack of segment diversity in recommendations and sensitivity to underlying data distribution. |

| | | | |
|---|---|---|---|
| | | “Platinum”, reflecting strong model bias | |
| **Dashboard Insights** | The dashboard highlighted: ~4000 customers and ~19000 transactions; dominance of one high-value segment; strong concentration in NSW and “Mass Customer” segment; key industry: Manufacturing and Financial Services | The dashboard revealed: total customer base (~4000) with all classified as high value; strong dominance of the Platinum segment in revenue contribution; geographic concentration in NSW; consistent alignment between segmentation and targeting outputs; clear prioritization of new customers based on target scores | Due to the differences between the two methodologies, traditional BI presents a static and descriptive analysis, whereas AI-enhanced BI introduces a predictive component. Traditional BI offered superior interpretability; in contrast, the AI outputs were more difficult to decipher, highlighting the need for human supervision to ensure visual clarity. Furthermore, the AI visualizations exposed the major flaw in this workflow: the failure to differentiate between existing and new clients. |

Overall, both workflows produced consistent insights, particularly regarding customer value drivers and segmentation patterns. However, the analytical depth differs significantly. In traditional BI, the analysis is descriptive and stable, has strong interpretability and limited scalability and predictive capability. In the AI-enhanced BI explanatory, predictive and partially prescriptive analysis, higher automation and scalability and was exposed greater sensitivity to data quality and model assumptions requiring human oversight.

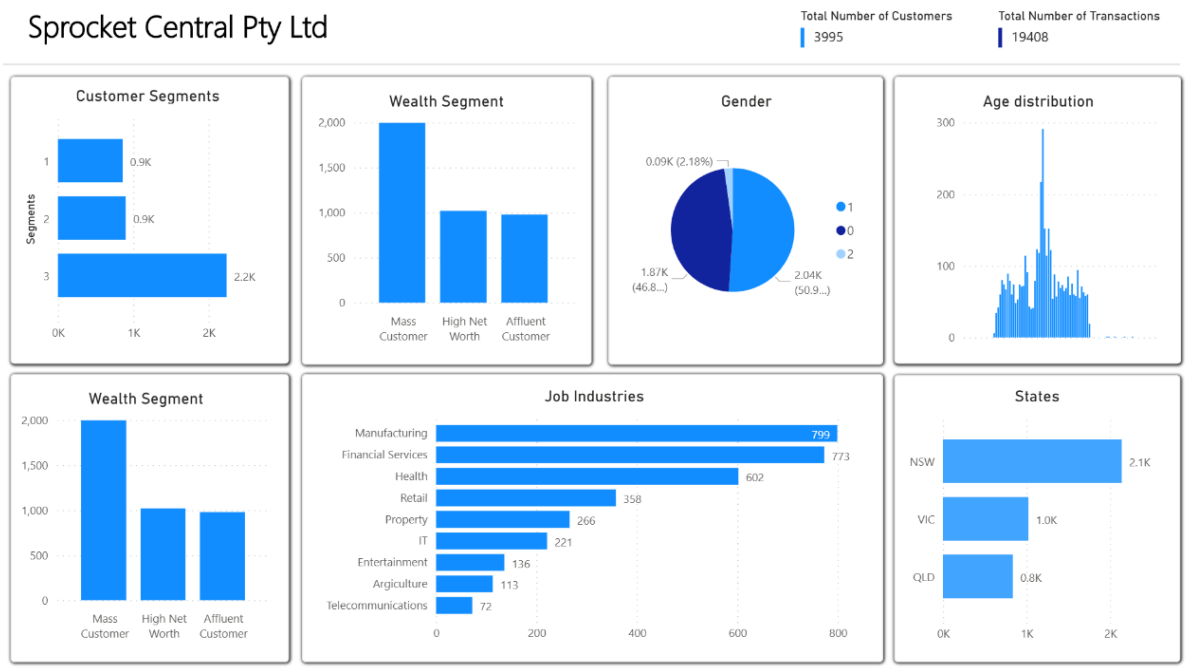



Figure 6.2 - KPMG Traditional BI Dashboard

Compared to the Maven case, the KPMG workflow introduces greater pipeline complexity, including data integration, segmentation and predictive modeling. This makes the advantages and disadvantages of AI more evident. While the primary

advantage of the AI approach was its advanced automation, it also presented notable disadvantages. Specifically, its high sensitivity to data inputs and underlying model assumptions required rigorous validation to ensure analytical quality.

### 6.1.3 Overall Analytical Results Comparison

Across both the Maven and KPMG case studies, a consistent pattern emerges in the comparison between traditional BI and AI-enhanced BI, with a clear progression in analytical depth and capability.

Traditional BI solutions in both cases produced stable, interpretable and descriptive insights. They successfully identified key patterns related to customer behavior, segmentation and business performance. However, their scope remained largely retrospective, with limited ability to extend insights into predictive applications.

In contrast, AI-enhanced BI demonstrated a progressive evolution across the two cases. In the Maven case, AI extended analysis to an explanatory level, quantifying relationships between variables. In the KPMG case, this evolved further into predictive and prescriptive analytics, including customer segmentation, scoring and targeting.

Despite these advantages, the AI showed important limitations, particularly in the KPMG case:

- Sensitivity to data quality and distribution
- Imbalanced or biased outputs
- Reduced interpretability
- Strong dependence on human validation

Overall, while both approaches produced consistent high-level insights, their roles differ:

- Traditional BI: descriptive, stable and interpretable
- AI-enhanced BI: explanatory, predictive and scalable, but validation-dependent

This confirms the transition identified in the literature from traditional BI toward AI-driven, decision-oriented analytics, while reinforcing that human oversight remains essential for ensuring reliable outcomes.

### 6.1.4 User Experience

The traditional BI workflow required strong technical knowledge of Power BI data modeling, DAX measure creation and visualization design. The development process relied on manual configuration and domain expertise.

In the Maven case, this approach proved effective due to the relatively simple, single-table dataset, allowing the analyst to conduct a coherent dashboard with clear

business insights. However, in the KPMG case, the increased complexity introduced by multiple datasets, relational modeling and customer-level aggregation significantly increased the cognitive and technical burden on the analyst. Tasks such as dataset integration, feature engineering and segmentation required careful coordination and deeper technical expertise.

The AI-enhanced workflow significantly changed the interaction paradigm. Gen AI enabled the automatic generation of analytical scripts, feature engineering logic and narrative explanations through natural language prompts.

This impact was particularly evident in the KPMG case, where multi-table integration logic was generated more efficiently, feature engineering (e.g. RFM variables) was automated and segmentation and predictive modeling were introduced with lower manual effort.

However, the usability improvements were conditional on user expertise. AI outputs often require supervision, debugging and validation. In both cases, but especially in the KPMG workflow, errors such as inconsistent transformations, incorrect joins or misinterpreted business logic required careful human intervention.

Inexperienced users could encounter difficulties identifying hallucinated outputs or structural inconsistencies in the generated code, particularly in more complex pipelines.

Thus, while AI reduces the coding barrier, it introduces new cognitive requirements related to prompt design, analytical validation and interpretation of generated outputs.

Overall, the KPMG case reinforces that AI improves usability more significantly in complex scenarios but also increases the importance of analytical supervision.

### 6.1.5 COST IMPLICATIONS

A traditional BI environment primarily involves labor and infrastructure costs.

The average salary of a data analyst in Portugal is approximately 33,000€ per year, according to industry estimates. Human analysts must dedicate significant time to manual coding, dashboard design and data preparation tasks (Universidade Europeia, 2025).

Traditional BI cost structure:

- Data analyst labor cost
- Power BI licensing
- Data infrastructure maintenance

The AI-enhanced workflow introduces additional tooling costs associated with generative AI services.

GitHub Copilot subscription plans include (GitHub, 2026):

- Individual Pro: approximately 100$ per year

- Pro+: approximately 390$ per year
- Enterprise Business: 19$ per user/month
- Enterprise Premium: 39$ per user/month

While these tools introduce additional software costs, they reduce labor expenditure by accelerating code generation and automating analytical tasks, particularly in complex workflows such as the KPMG case.

Consequently, AI has the potential to reduce the labor cost per analytical task while increasing tooling expenditure. This trade-off becomes more favorable as project complexity increases, as demonstrated by the KPMG case.

Large organizations can absorb tooling costs more easily, whereas smaller firms may struggle with licensing expenses despite potential productivity gains.

### 6.1.6 SECURITY AND GOVERNANCE

Traditional BI development typically occurs within controlled organizational environments, reducing the exposure of sensitive data. Analytical scripts and dashboards are developed locally and stored within the corporate infrastructure.

The AI-enhanced workflow introduces additional governance considerations:

- Dependency on external AI agents
- Potential data leakage through prompt interactions
- Hallucinated logic or inconsistent outputs
- Version instability in generated artifacts

These risks are amplified in more complex environments such as the KPMG case, where incorrect joins or transformations can propagate errors, misinterpreted schema relationships can affect entire pipelines and data inconsistencies may not be immediately visible.

Therefore, while AI increases automation capabilities, it simultaneously expands governance requirements. Organizations must implement validation mechanisms, audit trails and clear AI usage policies to mitigate associated risks.

The KPMG case highlights that governance becomes increasingly critical as data complexity and integration requirements grow.

### 6.1.7 EFFICIENCY

The AI-enhanced workflow demonstrated significantly higher automation during the early stages of the analytical pipeline. Data cleaning, feature engineering and DAX measure drafting were generated rapidly through AI-assisted code generation.

This efficiency gain was modest in the Maven case. Still, it became significantly more pronounced in the KPMG case, where multi-table integration was automated, feature

engineering was accelerated and predictive modeling was introduced with minimal manual coding.

However, these efficiency gains were partially offset by several operational challenges:

- Environment configuration issues
- Manual correction cycles
- PBIP structural debugging
- Hallucination recovery

In the KPMG case, these challenges were significantly reduced or entirely eliminated through prompt enhancements, a direct result of the lessons learned from the Maven case.

The traditional workflow required more manual effort during initial development but exhibited fewer structural interruptions and debugging cycles.

Overall, AI increased speed in early analytical phases. Traditional BI showed greater procedural stability. Therefore, the net efficiency advantage of AI becomes more evident in complex workflows but remains dependent on the extent of debugging and validation required.

### 6.1.8 Consistency and Reproducibility

Traditional BI implementations generally exhibit high structural consistency and predictable outputs due to their deterministic implementation process.

AI-generated workflows can also achieve reproducibility through version-controlled notebooks and scripts. However, results may vary across iterations depending on prompt formulation and model responses.

Traditional BI characteristics:

- Stable outputs
- Deterministic workflows
- Clear traceability

AI-enhanced workflow characteristics:

- Reproducible artifacts
- Occasional inconsistency across interactions
- Strong dependency on prompt specificity

Thus, while both approaches can be reproducible, traditional BI ensures higher consistency, whereas AI requires stricter control mechanisms in complex scenarios.

### 6.1.9 Code Quality Evaluation

In addition to analytical performance, this study incorporates software engineering quality metrics to evaluate the maintainability and reliability of implementations. These

metrics were assessed using automated testing frameworks, static code analysis and software quality monitoring tools.

This approach extends traditional BI evaluation frameworks by incorporating software engineering indicators such as code complexity, maintainability and technical debt, enabling a more comprehensive assessment of AI-assisted analytics pipelines.

#### 6.1.9.1 AUTOMATED TESTING

Automated testing was conducted using the pytest framework.

Maven test results:

- Total tests executed: 2
- Tests passed: 2
- Execution time: 9.54 seconds

Both implementations successfully passed the validation tests, confirming the functional correctness of the generated outputs.

KPMG test results:

- Total tests executed: 4
- Tests passed: 4
- Execution time: 30.38 seconds

All AI-generated and human-developed implementations successfully passed the validation tests, confirming correctness across both data quality assessment and analytical tasks.

Comparative analysis

From a comparison perspective, both AI and human solutions achieved full correctness in all scenarios, indicating that, at a functional level, AI-generated code has reached parity with human implementations in controlled environments.

However, the evolution lies in robustness and scalability of testing: the KPMG case demonstrates that AI-generated solutions can maintain correctness even as task complexity increases and multiple pipelines are introduced.

Overall, the comparison between AI and human solutions remains balanced in terms of correctness, but the KPMG case highlights that AI can scale to more complex testing scenarios without loss of functional reliability.

#### 6.1.9.2 STATIC CODE ANALYSIS (PYLINT)

Code quality was evaluated using pylint, which measures code readability, structure and adherence to best practices.

Table 6.3 - Maven Static Code Analysis

| Implementation | Score | Key Issues Identified |
|---|---|---|
| Human-written | 4.23/10 | Long code lines; missing docstrings; unnecessary semicolons; incorrect attribute references |
| AI-generated | 5.47/10 | Missing module documentation; undefined variables; excessive local variables; incorrect import ordering; unused imports |

Despite these issues, the AI-generated code achieved a slightly higher static analysis score. However, the results indicate that both implementations require refactoring to meet production-level software quality standards.

Table 6.4 - KPMG Static Code Analysis

| Implementation | Task | Score | Key Issues Identified |
|---|---|---|---|
| Human-written | Task 1 | 7.83/10 | Excessively long code lines; incorrect naming conventions (non-snake_case modules); missing documentation; wrong import ordering; unused imports and variables; use of outdated string formatting instead of f-strings; occasional unnecessary semicolons |
| Human-written | Task 2 (Part 1) | 6.83/10 | |
| Human-written | Task 2 (Part 2) | 8.31/10 | |
| AI-generated | Task 1 | 7.88/10 | Long code lines; missing module and function docstrings; redefinition of variables from outer scopes; trailing whitespace; some unnecessary or non-impactful statements |
| AI-generated | Task 2 | 7.74/10 | |

Despite these issues, the AI-generated code achieved a slightly higher average static analysis score. Both implementations demonstrate solid improvements in code quality, although there remains room for further refinement and optimization.

Comparative analysis

Comparing the Maven and KPMG results reveals a clear evolution in code quality for both AI-generated and human-written implementations. In the Maven case, both approaches scored relatively low, highlighting fundamental issues related to structure, readability and adherence to best practices. In contrast, the KPMG results show a substantial improvement, with all scores rising to a more acceptable range, closer to production-level standards.

The gap between AI and human implementations has narrowed, with both achieving very similar performance levels. While AI maintains slightly higher consistency across tasks, human-written code shows greater variability, reflecting differences in complexity and coding style. Overall, the comparison indicates that AI-assisted development is converging toward human-level code quality, with improvements becoming incremental rather than fundamental.

#### 6.1.9.3 SonarQube Software Quality Metrics

Software quality was further assessed using SonarQube, which evaluates maintainability, reliability and technical debt.

Table 6.5 - SonarQube Results

| Case | Maintainability Rating | Security Rating | Reliability Rating | Additional Indicators |
|---|---|---|---|---|
| **Maven** | A | A | C | ▪ Code smells: 4<br>▪ Bugs: 2<br>▪ Technical debt: 1 hour 42 minutes<br>▪ Technical debt ratio: 1.6%<br>▪ Cyclomatic complexity: 81<br>▪ Cognitive complexity: 102<br>▪ Lines of code: 405 |
| **KPMG** | A | A | C | ▪ Code smells: 52<br>▪ Bugs: 8<br>▪ Technical debt: 4 hours 33 minutes<br>▪ Technical debt ratio: 1%<br>▪ Cyclomatic complexity: 138<br>▪ Cognitive complexity: 187<br>▪ Lines of code: 891 |

In the Maven case, the AI implementation accounted for the majority of technical debt and complexity due to its larger script size and automated generation process. However, the maintainability rating remained high, indicating that the codebase can be improved with relatively limited refactoring.

The KPMG case reflects a significantly larger and more complex codebase, with higher counts of lines of code, complexity, and technical debt. Despite this, both maintainability and security ratings remain at the highest level (A), indicating that the system is still well-structured and manageable.

However, reliability remains a concern, with a C rating and a higher number of bugs, suggesting that increased complexity introduces more potential failure points. Overall, while the system scales effectively, it would benefit from refactoring to reduce code smells and improve long-term maintainability.

Comparative analysis

The comparison highlights a clear trade-off between simplicity and scale. While both Maven and KPMG achieve the same high-level ratings, the KPMG case shows a substantial increase in complexity, technical debt and code issues due to its more realistic and comprehensive scope.

Even so, overall code quality remains stable across both cases, demonstrating that both AI and human solutions scale effectively. However, reliability continues to be a consistent challenge, emphasizing the need for a stronger focus on robustness as system complexity grows.

### 6.1.10 LIMITATIONS

Several limitations could not be fully addressed within the scope of this experiment.

- First, although the inclusion of the KPMG case introduced a more complex and realistic business scenario, scalability was still not fully evaluated. The experiments were conducted on controlled datasets and local environments and therefore, performance under large-scale enterprise data conditions and distributed systems remains uncertain.
- Second, the integration of machine learning pipelines was limited. While the workflows included elements such as regression modeling, data quality assessment and exploratory analysis, a complete end-to-end ML lifecycle, including automated deployment, monitoring and continuous retraining, was not implemented.
- Third, the experiment did not evaluate long-term operational deployment. Aspects such as pipeline orchestration, real-time monitoring, fault tolerance and production governance frameworks were not evaluated, particularly in cloud or enterprise environments.
- Finally, the experiment relied on a limited number of generative AI tools, primarily GitHub Copilot and Gemini. Results may vary when using alternative models, newer AI systems or enterprise-grade platforms, especially as the capabilities of generative AI continue to evolve rapidly.

## 6.2 Discussion

The comparative evaluation conducted across the Maven and KPMG case studies demonstrates that generative AI acts primarily as an augmentation layer rather than a replacement for traditional BI workflows. However, the inclusion of a second, more complex case introduces a more nuanced understanding of how AI capabilities evolve under increased constraints, data complexity and methodological refinement.

Several key empirical insights emerge from the experiment.

First, AI significantly improves efficiency and accelerates development, particularly in the early stages of the BI pipeline. Across both case studies, AI-enabled workflows reduced the time required for data preparation, feature engineering and analytical modeling. This effect was modest in the Maven case but became substantially more pronounced in the KPMG case, where multi-table integration, segmentation and predictive modeling were partially automated. However, these gains were not absolute, as they were partially offset by debugging cycles, environment configuration issues and the need to correct inconsistencies in generated outputs.

Second, AI enhances the analytical depth of BI workflows, enabling a transition from descriptive to predictive and partially prescriptive analytics. In the Maven case, this was reflected in the introduction of regression modeling and statistical validation of relationships. In the KPMG case, this evolved further into customer segmentation, scoring and targeting, demonstrating AI's ability to operationalize decision-making

processes. In contrast, traditional BI remained primarily descriptive, focusing on retrospective insights and visualization-based interpretation.

Third, AI performance is highly dependent on prompt design, workflow structuring and iterative refinement. The comparison between the Maven and KPMG cases reveals a clear learning effect. In the Maven case, broader prompts resulted in unstable outputs, hallucinations and missing artifacts. In the KPMG case, stricter constraints and improved prompt engineering significantly increased reliability, reproducibility and execution stability. This demonstrates that AI effectiveness is not intrinsic to technology alone but rather emerges from the quality of human-AI interaction.

Fourth, AI introduces important governance, reliability and interpretability challenges. Across both case studies, issues such as hallucinated outputs, incorrect assumptions about data structures, imbalanced segmentation results and sensitivity to data quality were observed. These challenges were amplified in the KPMG case due to introduced risks related to bias, lack of transparency and reduced interpretability, particularly in predictive outputs.

Fifth, human expertise remains indispensable. Although AI reduces the manual coding burden, it does not eliminate the need for technical and domain knowledge. Instead, the role of the analyst shifts toward prompt engineering, validation, debugging and critical interpretation of results. This is particularly evident in complex scenarios, where incorrect AI outputs can propagate through the pipeline and compromise analytical validity if not properly supervised.

Finally, AI performance varies across different stages of the BI pipeline. The results indicate that AI performs more reliably in structured, code-driven environments, such as data cleaning, feature engineering and statistical modeling. In contrast, performance remains less consistent in schema-dependent and visualization-focused tasks, such as Power BI dashboard development and PBIP structuring. This highlights current limitations in integrating gen AI with graphical and highly contextual BI environments.

Overall, the evaluation demonstrates that AI-enhanced BI workflows offer significant advantages in efficiency, scalability and analytical capability, particularly in complex scenarios. However, these benefits are conditional on proper governance, high-quality data and strong human oversight, reinforcing the need for hybrid human-AI analytical frameworks.

The empirical findings of this study strongly align with, but also extend, the conclusions identified in the Systematic Literature Review.

**SLRQ1 – Current Status of AI in BI**

The literature indicates a clear transition from traditional descriptive BI systems toward predictive, prescriptive and augmented analytics platforms. The results of this study provide strong empirical support for this transformation.

In the Maven case, AI extended BI capabilities to an explanatory level through statistical modeling. In the KPMG case, this evolved further into predictive and prescriptive applications, including segmentation and customer targeting. These findings confirm that AI is actively transforming BI into a more decision-oriented discipline.

However, the results also refine the literature by demonstrating that this transformation is progressive and context-dependent rather than absolute. While AI enables advanced analytics, it does not eliminate the need for human expertise, particularly in complex and high-stakes environments.

**SLRQ2 – Challenges of AI adoption in BI**

The literature identifies key challenges, including skill gaps, governance concerns, infrastructure limitations and organizational readiness.

The findings of this study strongly confirm these challenges. In both case studies, the effectiveness of AI was highly dependent on user expertise, particularly in prompt engineering, debugging and validation. Issues related to hallucinations, inconsistent outputs and incorrect assumptions about data structures directly reflect the risks identified in the literature.

Additionally, this study provides more granular empirical evidence regarding LLM-specific challenges, such as sensitivity to prompt quality, reproducibility issues and instability across iterations. These challenges were particularly evident in the Maven case and significantly mitigated in the KPMG case through improved prompt design.

Overall, the results reinforce that AI adoption in BI is not purely a technological challenge but also an organizational and human-centered transformation process.

**SLRQ3 – AI and ML techniques used in BI**

Existing studies emphasize the role of machine learning, automation and gen AI to enhancing BI capabilities.

The findings of this study partially confirm and extend these conclusions. While full end-to-end ML pipelines were not implemented, the AI-enhanced workflows demonstrated practical applications of:

- Automated data preparation and feature engineering
- Regression-based modeling and statistical analysis
- Customer segmentation and clustering
- Predictive scoring and ranking mechanisms

- Natural language-based analytical generation

These results highlight the role of gen AI as a “virtual data analyst”, capable of orchestrating multiple analytical tasks within the BI pipeline. However, the study also shows that these techniques require careful validation, particularly in complex scenarios where model assumptions and data distributions significantly influence results.

**SLRQ4 – Research gaps and future directions**

Previous literature identified a lack of empirical evidence comparing traditional BI workflows with AI-enhanced analytics pipelines under identical experimental conditions.

This study contributes to addressing that gap by providing a controlled comparison across two case studies of increasing complexity using consistent datasets, analytical tasks and evaluation criteria.

Furthermore, the findings extend the literature by demonstrating that:

- AI effectiveness is highly dependent on prompt engineering and iterative refinement
- Hybrid human-AI workflows are more realistic than fully automated systems
- AI introduces new challenges related to bias, interpretability and reliability, particularly in predictive analytics
- The benefits of AI increase with data and pipeline complexity, but so do the associated risks

These insights suggest that future research should focus on developing structured frameworks for human-AI collaboration in BI, improving robustness in gen AI outputs and evaluating AI performance in real-world, large-scale enterprise environments.

Overall, while the literature emphasizes the transformative potential of AI in BI, the findings of this study indicate that this transformation is incremental, conditional and highly dependent on human-AI interaction. Rather than replacing traditional BI, AI reshapes the analyst role and introduces a new paradigm centered on augmentation, supervision and strategic decision-making.

# 7. CONCLUSIONS AND FUTURE RESEARCH

This chapter presents the conclusions of the research and reflects on the contributions of the study. It synthesizes the main findings obtained from the comparative evaluation and relates them to the research question and objectives defined previously. The chapter also discusses the limitations encountered during the experiment and outlines potential directions for future research.

## 7.1 SYNTHESIS OF THE DEVELOPED WORK

The primary objective of this research was to evaluate how AI is reshaping BI workflows by comparing traditional BI implementations with AI-enhanced analytical pipelines. To achieve this, the study combined a Systematic Literature Review with a comparative empirical evaluation based on two case studies of increasing complexity: the Maven Marketing Challenge and the KPMG Data Analytics Virtual Program.

The empirical component involved the implementation of two distinct analytical workflows for each case: a traditional BI solution, based on manual data preparation, modeling and visualization and an AI-enhanced workflow supported by gen AI tools. This dual-case design enabled not only a comparison between human-driven and AI-assisted approaches but also an assessment of how AI performance evolves under improved prompt design, stricter constraints, and more complex data environments.

The results demonstrate that AI introduces significant automation and scalability advantages across the BI pipeline, particularly in tasks such as data preparation, feature engineering and analytical code generation. While these benefits were observable in the Maven case, they became substantially more pronounced in the KPMG case, where AI supported multi-table integration, customer segmentation and predictive targeting. Furthermore, AI extended the analytical scope of BI workflows from descriptive analysis toward explanatory, predictive and partially prescriptive analytics, confirming the transition identified in the literature.

However, the findings also reveal that AI does not replace human expertise. Instead, it redefines the role of the analyst, shifting the focus from manual implementation toward prompt engineering, validation, debugging and strategic interpretation of outputs.

Overall, the research question: how Artificial Intelligence reshapes Business Intelligence workflows and what benefits, limitations and organizational challenges emerge from its integration, was successfully addressed. The empirical comparison confirmed that AI can significantly increase efficiency and analytical depth while simultaneously introducing governance challenges, technical dependencies and new skill requirements. The objectives defined for this study were therefore achieved,


including the implementation and evaluation of the workflows and the analysis of their impact on analytical performance, usability and decision-making support.

## 7.2 LIMITATIONS

Although the research provides valuable insights into the integration of AI within BI workflows, several limitations should be acknowledged.

First, although the inclusion of the KPMG case introduced a more complex and realistic analytical scenario, the study was still conducted in controlled environments using limited datasets. As such, the scalability of AI-enhanced BI workflows in large-scale enterprise contexts, involving high-volume data, distributed architectures and real-time processing, was not fully evaluated.

Second, the implementation of machine learning capabilities remained partial and exploratory. While the study incorporated regression modeling, segmentation and predictive scoring, it did not cover the full lifecycle of machine learning systems, including deployment, monitoring, model governance and continuous retraining. As a result, the integration of AI-driven predictive analytics into production-level BI environments was only partially assessed.

Third, the experiment relied on a limited set of AI tools, primarily GitHub Copilot and Gemini. Different generative AI models may exhibit varying levels of reliability, accuracy and integration capabilities. Therefore, the results obtained in this study may not be directly generalized to all AI-assisted BI platforms.

Fourth, although the evaluation considered technical, analytical and usability dimensions, organizational and human factors were not explored in depth. Aspects such as change management, user adoption, governance frameworks, ethical considerations and long-term cost structures were not comprehensively analyzed, despite their critical importance for real-world implementation.

Finally, the evaluation primarily focused on technical and analytical performance. Organizational factors such as change management, employee adoption, governance policies and long-term operational costs were not examined in depth. These factors may significantly influence the successful adoption of AI-powered BI solutions in real-world organizations.

## 7.3 FUTURE WORK

Future research could address the limitations identified in this study by expanding the scope and depth of empirical evaluation.

One potential direction involves evaluating AI-enhanced BI workflows in large-scale enterprise environments. Future studies could analyze the performance and scalability

of AI-assisted analytics when applied to high-volume datasets and complex data architectures, including data lakes and distributed cloud infrastructures.

Another important area for future research is the integration of full machine learning pipelines within BI platforms. Extending the experiment to include automated model training, deployment, monitoring and retraining would provide a more comprehensive understanding of how AI-driven predictive analytics can be operationalized within BI systems.

Further research could also investigate the comparative performance of different generative AI models and prompting strategies. Systematic evaluation of multiple AI tools, prompt designs and interaction frameworks could provide valuable insights into improving reliability, reducing hallucinations and enhancing reproducibility in AI-assisted analytics.

Additionally, future studies should focus on the organizational and human dimensions of AI adoption in BI. This includes investigating skill transformation, user trust, governance frameworks, ethical considerations and the impact of AI on decision-making processes. Understanding how organizations can effectively integrate AI while maintaining control, transparency and accountability is essential for long-term success.

Finally, there is a need to develop structured frameworks for human-AI collaboration in BI workflows. Rather than aiming for full automation, future research should focus on hybrid approaches that combine the strengths of AI (automation, scalability and pattern recognition) with human capabilities (critical thinking, domain knowledge and contextual understanding).

By addressing these research directions, future studies can contribute to a deeper understanding of how AI technologies can be responsibly and effectively integrated into BI systems, ultimately supporting more advanced, scalable and trustworthy data-driven decision-making processes.

# BIBLIOGRAPHICAL REFERENCES


Adewusi, A. O., Okoli, U. I., Adaga, E., Olorunsogo, T., Asuzu, O. F., & Daraojimba, D. O. (2024). BUSINESS INTELLIGENCE IN THE ERA OF BIG DATA: A REVIEW OF ANALYTICAL TOOLS AND COMPETITIVE ADVANTAGE. *Computer Science & IT Research Journal, 5(2), 415-431*. https://doi.org/10.51594/csitrj.v5i2.791

Afsaruddin, Agrawal, D., & Nayak, S. K. (2023). An Optimized Approach for Maximizing Business Intelligence using Machine Learning. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(8), 72–80. https://doi.org/10.17762/ijritcc.v11i8.7925

Alghamdi, N. A., & Al-Baity, H. H. (2022). Augmented Analytics Driven by AI: A Digital Transformation beyond Business Intelligence. *Sensors*, *22*(20). Scopus. https://doi.org/10.3390/s22208071

Al-Momani, M. M., Alqudah, T. A., Swiety, I. A. A., Mahrakani, N., Nassoura, M. B. A., & Attar, M. K. A. (2025). Integrating Artificial Intelligence (AI) and Business Intelligence (BI): A Framework for Improving Enterprise Performance. *TEM Journal*, *14*(3), 2208–2216. Scopus. https://doi.org/10.18421/TEM143-26

Al-Momani, M. M., Awadallah, A., Al-Nassar, B., Ismail, A., Nassoura, M. B. A., & Abudarwish, N. (2025). The Role of Business Intelligence in Supply Chain Optimization A Case Study of the Carrefour Market in Jordan. *Data and Metadata*, *4*. Scopus. https://doi.org/10.56294/dm2025747

Alqhatani, A., Ashraf, M. S., Ferzund, J., Shaf, A., Abosaq, H. A., Rahman, S., Irfan, M., & Alqhtani, S. M. (2022). 360° Retail Business Analytics by Adopting Hybrid Machine Learning and a Business Intelligence Approach. *Sustainability (Switzerland)*, *14*(19). Scopus. https://doi.org/10.3390/su141911942

Chintala, S. (2024). Next - Gen BI: Leveraging AI for Competitive Advantage. *International Journal of Science and Research (IJSR)*, *13*(7), 972–977. https://doi.org/10.21275/SR24720093619

Das, B. C., Mahabub, S., & Hossain, M. R. (2024). Empowering modern business intelligence (BI) tools for data-driven decision-making: Innovations with AI and analytics insights. *Edelweiss Applied Science and Technology*, *8*(6), 8333–8346. Scopus. https://doi.org/10.55214/25768484.v8i6.3800

Desai, D., & Desai, A. (2025). Integrating Generative AI in Business Intelligence: A Practical Framework for Enhancing Augmented Analytics. *International Journal of Mathematical, Engineering and Management Sciences*, *10*(3), 704–728. Scopus. https://doi.org/10.33889/IJMEMS.2025.10.3.036

Devi, S., Mahalley, Z., Nuraini, R., & Dhar, B. K. (2025). Exploring the challenges of AI-driven business intelligence systems in the Malaysian insurance industry. *F1000Research*, *14*. Scopus. https://doi.org/10.12688/f1000research.163354.1

Farayola, A. (2024). Business Intelligence Transformation Through Ai and Data Analytics. *Engineering Science & Tecnology Journal*. https://doi.org/10.51594/ESTJ.V4I5.616

Gajić, T., Ivanišević, A., & Knežević, S. (2025). Employee Perceptions of BI and AI Tools for Service Transformation: Evidence from the Serbian Airline and Hotel Industries. *International Journal of Industrial Engineering and Management*, *16*(3), 227–238. Scopus. https://doi.org/10.24867/IJIEM-385

GitHub. (2026). *GitHub Copilot · Planos e preço*. GitHub. https://github.com/features/copilot/plans?locale=pt-br

IBM. (2021, August 9). *What Is Business Intelligence (BI)? | IBM*. https://www.ibm.com/think/topics/business-intelligence

Jimenez-Partearroyo, M., & Medina-López, A. (2024). Leveraging Business Intelligence Systems for Enhanced Corporate Competitiveness: Strategy and Evolution. *Systems*, *12*(3). Scopus. https://doi.org/10.3390/systems12030094

Kareem, A. R., & Abdullah, H. S. (2025). Leveraging Hadoop and Hybrid Deep Learning on Home Datasets for Business Intelligence. *Iraqi Journal of Science*, *66*(7), 2980–2996. Scopus. https://doi.org/10.24996/ijs.2025.66.7.26

Mantouzi, S., Youssef, S., El Mouatassim, A., Ourahou, Y., Jafi, H., Zouhouredine, I., & Hamliri, M. (2025). Leveraging AI for business intelligence: A pathway to improved organizational performance in Morocco's manufacturing sector. *Multidisciplinary Science Journal*, *7*(9). Scopus. https://doi.org/10.31893/multiscience.2025412

Milinthapunya, W., Yamchuti, U., Nammakhunt, A., Shawarangkoon, C., Wannapiroon, P., & Nillsook, P. (2025). Business Intelligence Management with Artificial Intelligence for Prediction Information Technology Infrastructure in Higher Education. *TEM Journal*, *14*(2), 1378–1387. Scopus. https://doi.org/10.18421/TEM142-38

Miller, K. (2024). *Privacy in an AI Era: How Do We Protect Our Personal Information? | Stanford HAI*. https://hai.stanford.edu/news/privacy-ai-era-how-do-we-protect-our-personal-information

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2023). A declaração PRISMA 2020: Diretriz atualizada para relatar revisões sistemáticas. *Revista Panamericana de Salud Pública*, *46*, e112. https://doi.org/10.26633/RPSP.2022.112

Rahman, M. J., Rana, T., Xu, Y., & Öhman, P. (2024). Bridging BI and AI Enhancing Operational Efficiency in the Chinese Financial Sector. *Journal of Global Information Management*, *32*(1). https://doi.org/10.4018/JGIM.366871

Shah, V. (2024). *Role of AI in Business Intelligence (BI)*. ThoughtSpot. https://www.thoughtspot.com/data-trends/ai/ai-in-business-intelligence

Sharda, R., Delen, D., & Turban, E. (2018). *Business intelligence, analytics, and data science: A managerial perspective* (Fourth edition). Pearson.

Syed, S. (2022). *Breaking Barriers: Leveraging Natural Language Processing In Self-Service Bi For Non-Technical Users* (SSRN Scholarly Paper No. 5032632). Social Science Research Network. https://doi.org/10.2139/ssrn.5032632

Universidade Europeia. (2025, February 10). *O que faz um Analista de Dados | Univ. Europeia*. O que faz um Analista de Dados? https://www.europeia.pt/blog/o-que-e-a-gestao-de-dados/

Vines, A., Bologa, A.-R., & Bostan, A.-I. (2025). Enabling Intelligent Data Modeling with AI for Business Intelligence and Data Warehousing: A Data Vault Case Study. *Systems*, *13*(9). Scopus. https://doi.org/10.3390/systems13090811

# APPENDIX A – PROMPTS

This appendix presents the prompts provided to Copilot during the development of the Maven and KPMG projects. These prompts guided the generation of code and the dashboards, supporting the overall development of the BI pipeline.

Appendix A 1 - Maven Data Engineering Prompt

@workspace You are an expert Business Intelligence Consultant tasked with solving the "Maven Marketing Challenge." Your goal is to build a complete BI pipeline that outperforms a human benchmark.
Context: We have marketing campaign data for 2,240 customers. Recent campaigns have underperformed. We need a data-driven recommendation to improve future campaigns.
Objective: Perform the following 3 phases of work. Save all outputs to the local directory.
PHASE 1: Data Engineering & Python Analysis (The "Backend")
Create a Jupyter Notebook (marketing_analysis.ipynb) that performs these steps using pandas:

1. Ingest: Load the CSV data.
2. Clean and Feature Engineering :
    - Identify and handle null values
    - Handle outliers
    - Are there any variables that warrant transformations? If yes, do it
    - Are there any useful variables that you can engineer with the given data? If yes, do it
    - And another important steps…
3. Answer Business Questions (in the notebook):
    - What factors are significantly related to the number of web purchases?
    - Which marketing campaign is most successful?
    - What does the average customer look like for this company?
    - Which products are performing best?
    - Which channels are underperforming?
4. Export: Save the cleaned, enriched dataset as silver_layer_marketing_data.csv for Power BI ingestion.

PHASE 2: Semantic Modeling (The "Logic")
Act as a Power BI Architect. Generate a file named model_logic.dax containing the DAX measures we will need. Do NOT write generic measures. Write measures that will allow to answer the business questions above.

PHASE 3: Strategic Recommendation

Create a markdown file executive_summary.md summarizing your findings. Pitch your #1 recommendation to the CMO based on the data.

Constraints:

- Write clean, modular Python code.
- Add comments explaining why you chose a specific cleaning method.
- The output must be reproducible.

Appendix A 2 - KPMG Data Engineering Prompt

@workspace You are an AI Data Analytics Agent operating in a controlled experimental environment.

Your goal is to execute a COMPLETE and REPRODUCIBLE BI pipeline for the "KPMG Data Analytics Virtual Program", ensuring that all generated artifacts are correctly written to the repository and executable without manual fixes.

This is part of a comparative study between traditional BI and AI-assisted BI workflows. Therefore, consistency, traceability and reproducibility are CRITICAL.

CONTEXT

Client: Sprocket Central Pty Ltd (Retail - Bicycles & Accessories)

Dataset: "KPMG_VI.xlsx"

Tables:

Transactions

Customer Demographic

Customer Address

New Customer List

Business Objective:

Identify high-value customers and recommend which new customers should be targeted based on similarity to top-performing segments.

CRITICAL EXECUTION RULES (MANDATORY)

FILE PERSISTENCE:

You MUST write ALL generated artifacts directly into the repository.

Do NOT only display outputs in chat.

ENVIRONMENT SETUP:

Generate a requirements.txt file with ALL necessary libraries.

Ensure notebook is executable in a clean environment.

NOTEBOOK EXECUTION:

DO NOT attempt to execute the notebook.

Only generate valid, runnable code.

ERROR PREVENTION:

Use robust imports and defensive coding.
Avoid ambiguous variable names.
Validate joins and transformations.
REPRODUCIBILITY:
Set random seeds where applicable.
Ensure deterministic outputs.
PHASE 1 — DATA QUALITY ASSESSMENT
Create: /notebooks/kpmg_task1_dqa.ipynb
Tasks:
Load all sheets from KPMG_VI.xlsx
Perform:
Missing Value Analysis (with %)
Duplicate Detection
Inconsistency Checks (categorical + formats)
Outlier Detection (IQR method)
Data Type Validation
Output:
Structured findings (tables + comments inside notebook)
Export: /reports/dqa_summary.md
Also create:
/reports/dqa_email.md (client-ready email with findings + recommendations)
PHASE 2 — DATA ENGINEERING & ANALYTICS
Create: /notebooks/kpmg_task2_analysis.ipynb
DATA INTEGRATION:
Merge datasets into a unified customer-level table
Explicitly document join logic
DATA CLEANING:
Apply fixes identified in DQA
Handle missing values (justify method in comments)
Treat outliers
Ensure correct data types
FEATURE ENGINEERING:
MUST include:
Recency
Frequency
Monetary (RFM)
Total Spend
Total Purchases
Customer Tenure

Any additional high-value features (justify)

EXPLORATORY ANALYSIS:

Answer:

Who are the most valuable customers?

What characterizes them?

Where are they located?

Which attributes drive value?

SEGMENTATION:

Apply clustering (KMeans or justified alternative)

Evaluate clusters (e.g., silhouette score)

Label segments (e.g., Platinum, Loyal, At Risk)

TARGETING MODEL:

Score new customers based on similarity to high-value segments

Create ranking logic (explicit formula required)

EXPORTS (MANDATORY)

Save:

/data/OldCustomerList_clean.csv

/data/NewCustomerList_scored.csv

/data/Transactions_clean.csv

PHASE 3 — SEMANTIC MODELING

Create: /powerbi/model_logic.dax

Rules:

ONLY business-relevant measures

Measures MUST map to:

Customer value

Segmentation

Revenue

Target scoring

Avoid generic measures

Use correct DAX syntax (avoid reserved keywords)

PHASE 4 — STRATEGIC OUTPUT

Create: /reports/kpmg_strategy.md

Include:

Key insights from segmentation

Definition of high-value customer

Targeting strategy for new customers

Expected business impact

FINAL CHECK

Before finishing:

Confirm ALL files are written to disk
Ensure folder structure is respected
Ensure notebook code is executable
GOAL
Produce a fully reproducible, structured and BI-ready pipeline that improves upon baseline human solutions while remaining methodologically consistent.

Appendix A 3 - Maven Dashboard Prompt

Part 1: Context Analysis
@workspace  Act as a Lead Power BI Developer and UI/UX Designer. I have a Power BI Project (PBIP) open in this workspace.
Context Analysis Tasks:
Analyze the Data Model: Scan the .tmdl (or .bim) files in the Dataset folder. List out the names of all 'Silver' level tables and the DAX measures I have already created. I specifically need you to identify measures related to. These measures are also in model_logic.dax
Analyze the Notebook Context: I have a notebook (marketing_analysis.ipynb) and the file you write you conclusions on (executive_summary.md), focus more on the notebook. Scan the workspace for any .ipynb or markdown files containing summary text.
Understand the Goal: I need a 'One-Page Executive Dashboard' that answers key business questions using these measures.
Confirm you have identified the measures and tables before proceeding.
Part 2: The Generation Prompt
Now, I want you to generate the PBIR JSON files for this dashboard.
Action: Create a new folder in \definition\pages\ named Executive_Dashboard. Layout Strategy: Use a strict 1280x720 (16:9) coordinate system. The page background should be #F3F2F1 (Light Grey) for contrast.
Use the visuals that allow you to better answer the questions and use the measures.
'Good Eye' Visual Styling (Technical Specs):
Shadows: For every visual, enable the 'Drop Shadow' property in the JSON (formatting.shadow.visible: true).
Borders: Add a white background and a rounded border (radius: 8px) to each visual container (formatting.visualBorder).
Padding: Ensure 10px padding between all visuals to create a clean 'grid' effect.
Execution Rules:
Unique IDs: You MUST generate a random alphanumeric hash (e.g., 'a1b2c3d4') for every visual's name property. Do not use generic names.

File Structure: Write the page.json and a separate visual.json for each component inside the visuals folder.
Validation: Ensure the JSON adheres to the standard Power BI PBIR schema.
Execute this now.

Appendix A 4 - KPMG Dashboard Prompt 1

@workspace You are an AI Power BI Developer operating in a constrained PBIP (Power BI Project) environment.
Your goal is to generate a VALID and LOADABLE Power BI dashboard using PBIP JSON definitions.
This is part of a controlled experiment. You MUST strictly follow PBIP structure and avoid assumptions based on PBIX format.
CRITICAL EXECUTION RULES (MANDATORY)
PBIP COMPLIANCE:
You MUST follow PBIP structure (NOT PBIX assumptions)
Do NOT invent unsupported properties
FILE CREATION:
You MUST explicitly write files to correct directories
Do NOT assume automatic placement
NO HALLUCINATION:
If a measure, column, or table is missing → DO NOT invent it
Use ONLY what exists in:
model_logic.dax
dataset (.tmdl / .bim)
VALIDATION:
JSON must be syntactically valid
Structure must allow Power BI to load
PART 0 — MODEL AWARENESS (CRITICAL CONTEXT)
The dataset follows a hybrid analytical structure combining historical customers and ML-scored prospects:
Tables:
Transactions_clean → FACT table (transaction-level data)
OldCustomerList_clean → Dimension table (existing customers)
NewCustomerList_scored → Dimension table (prospects scored via ML)
Master_Customer_Table → Analytical UNION table combining both populations
Relationships:
Transactions_clean[customer_id] → OldCustomerList_clean[customer_id] (Many-to-One)

NewCustomerList_scored is NOT directly related to transactions
Master_Customer_Table may be partially or not physically related and should be treated as an analytical table
IMPORTANT MODEL LOGIC:
Revenue and transactional metrics apply ONLY to existing customers
Prospect data (NewCustomerList_scored) contains NO transactions
Master_Customer_Table is used for:
segmentation
combined views
comparison between Existing vs Prospect
You MUST respect this separation and avoid incorrect joins or assumptions.
PART 1 — CONTEXT ANALYSIS
Analyze:
Dataset Model:
Scan /dataset/ (.tmdl or .bim)
Identify:
Tables
Relationships
Measures (already created)
Analytical Context:
Read:
/notebooks/kpmg_task2_analysis.ipynb
/reports/kpmg_strategy.md
Extract:
Key metrics (Revenue, Customers, Orders, Scores, etc.)
Customer segments (e.g., wealth_segment or equivalent)
High-value logic
Targeting logic for new customers
IMPORTANT:
DO NOT recreate or redefine measures
ONLY use existing measures from the model
Confirm findings BEFORE proceeding.
PART 2 — DASHBOARD GENERATION
Create folder:
\definition\pages\Executive_Dashboard
PAGE SPECIFICATIONS
Size: 1280x720
Background: #F3F2F1
BUSINESS QUESTIONS TO ANSWER

The dashboard MUST clearly answer:
Who are the highest-value existing customers?
What defines these high-value segments?
Where are these customers located?
Which attributes drive customer value?
Which new customers (prospects) should be targeted?
VISUAL DESIGN STRATEGY
Use ONLY visuals supported in PBIP JSON.
Design MUST reflect the hybrid nature of the data:
Separate or clearly distinguish:
Existing Customers (transaction-based)
Prospects (score-based)
Recommended visuals:
KPI Cards:
Total Revenue
Total Customers
High Value Customers
Total Orders
Bar Charts:
Revenue by Segment
Customer Count by Segment
Donut Chart:
Segment distribution (using Master_Customer_Table if appropriate)
Table:
Top customers (existing)
Top prospects (based on score/rank)
Map (ONLY if geographic fields exist)
IMPORTANT DESIGN LOGIC:
DO NOT mix revenue with prospect data in the same aggregation unless logically valid
Use Master_Customer_Table ONLY for segmentation or comparison views
Use Transactions_clean context for financial metrics
STRICT DESIGN RULES
Shadow: enabled
Background: white
Border radius: 8px
Padding: 10px
Clean grid layout
FILE STRUCTURE (MANDATORY)
Create:


\definition\pages\Executive_Dashboard\page.json
\definition\pages\Executive_Dashboard\visuals\visual_<random>.json
NAMING RULES
Each visual MUST have a random ID (e.g., "a8f3d91c")
IDs MUST be unique
VALIDATION STEP (CRITICAL)
Before finishing:
Ensure:
JSON is valid
File paths are correct
No PBIX-only properties are used
Re-check:
Does structure match PBIP expectations?
Are visuals consistent with the dataset model?
Are measures used correctly without redefinition?
FINAL GOAL
Produce a Power BI dashboard that:
Loads without errors
Reflects analytical logic from Python outputs
Correctly separates existing vs prospect logic
Enables executive decision-making
Supports customer targeting strategy

Appendix A 5 - KPMG Dashboard Prompt 2

@workspace You are an AI Power BI Developer operating in a constrained PBIP (Power BI Project) environment. The project is already created, but is not working. Forget what is done and how it is done.
Your goal is to generate a VALID and LOADABLE Power BI dashboard using PBIP JSON definitions. You must replace what is already there, in terms of structure and visuals.
This is part of a controlled experiment. You MUST strictly follow PBIP structure and avoid assumptions based on PBIX format.
CRITICAL EXECUTION RULES (MANDATORY)
PBIP COMPLIANCE:
You MUST follow PBIP structure (NOT PBIX assumptions)
Do NOT invent unsupported properties
FILE CREATION:
You MUST explicitly write files to correct directories

Do NOT assume automatic placement
NO HALLUCINATION:
If a measure, column, or table is missing → DO NOT invent it
Use ONLY what exists in:
model_logic.dax
dataset (.tmdl / .bim)
VALIDATION:
JSON must be syntactically valid
Structure must allow Power BI to load
PART 0 — MODEL AWARENESS (CRITICAL CONTEXT)
The dataset follows a hybrid analytical structure combining historical customers and ML-scored prospects:
Tables:
Transactions_clean → FACT table (transaction-level data)
OldCustomerList_clean → Dimension table (existing customers)
NewCustomerList_scored → Dimension table (prospects scored via ML)
Master_Customer_Table → Analytical UNION table combining both populations
Relationships:
Transactions_clean[customer_id] → OldCustomerList_clean[customer_id] (Many-to-One)
NewCustomerList_scored is NOT directly related to transactions
Master_Customer_Table may be partially or not physically related and should be treated as an analytical table
IMPORTANT MODEL LOGIC:
Revenue and transactional metrics apply ONLY to existing customers
Prospect data (NewCustomerList_scored) contains NO transactions
Master_Customer_Table is used for:
segmentation
combined views
comparison between Existing vs Prospect
You MUST respect this separation and avoid incorrect joins or assumptions.
PART 1 — CONTEXT ANALYSIS
Analyze:
Dataset Model:
Scan /dataset/ (.tmdl or .bim)
Identify:
Tables
Relationships
Measures (already created)

Analytical Context:

Read:

/notebooks/kpmg_task2_analysis.ipynb

/reports/kpmg_strategy.md

Extract:

Key metrics (Revenue, Customers, Orders, Scores, etc.)

Customer segments (e.g., wealth_segment or equivalent)

High-value logic

Targeting logic for new customers

IMPORTANT:

DO NOT recreate or redefine measures

ONLY use existing measures from the model

Confirm findings BEFORE proceeding.

PART 2 — DASHBOARD GENERATION

Create folder:

\definition\pages\Executive_Dashboard

PAGE SPECIFICATIONS

Size: 1280x720

Background: #F3F2F1

BUSINESS QUESTIONS TO ANSWER

The dashboard MUST clearly answer:

Who are the highest-value existing customers?

What defines these high-value segments?

Where are these customers located?

Which attributes drive customer value?

Which new customers (prospects) should be targeted?

VISUAL DESIGN STRATEGY

Use ONLY visuals supported in PBIP JSON.

Design MUST reflect the hybrid nature of the data:

Separate or clearly distinguish:

Existing Customers (transaction-based)

Prospects (score-based)

IMPORTANT DESIGN LOGIC:

DO NOT mix revenue with prospect data in the same aggregation unless logically valid

Use Master_Customer_Table ONLY for segmentation or comparison views

Use Transactions_clean context for financial metrics

STRICT DESIGN RULES

Shadow: enabled

Background: white

Border radius: 8px
Padding: 10px
Clean grid layout
FILE STRUCTURE (MANDATORY)
Create or redo:
\definition\pages\Executive_Dashboard\page.json
\definition\pages\Executive_Dashboard\visuals\visual_<random>.json
NAMING RULES
Each visual MUST have a random ID (e.g., "a8f3d91c")
IDs MUST be unique
VALIDATION STEP (CRITICAL)
Before finishing:
Ensure:
JSON is valid
File paths are correct
No PBIX-only properties are used
Re-check:
Does structure match PBIP expectations?
Are visuals consistent with the dataset model?
Are measures used correctly without redefinition?
FINAL GOAL
Produce a Power BI dashboard that:
Loads without errors
Reflects analytical logic from Python outputs
Correctly separates existing vs prospect logic
Enables executive decision-making
Supports customer targeting strategy

# APPENDIX B – REPOSITORIES

This appendix provides links to the GitHub repositories where the results of the AI solutions for both the Maven and KPMG projects are stored. These repositories contain the developed code and related outputs, allowing access to the final implementations of each case study.

Appendix B 1 - Maven Case GitHub Repository

https://github.com/RLopes2327/masters_thesis_rl

Appendix B 2 - KPMG Case GitHub Repository

https://github.com/RLopes2327/masters_thesis_rl_2

# APPENDIX C – PRISMA

This appendix presents the PRISMA output table summarizing the final 15 articles selected for this study. The information presented in this table provides a structured overview of the literature used and highlights how each source supports the development of this research.

Appendix C 1 - PRISMA Execution Output

| # | Authors | Article | Contribution | Publication Type |
|---|---|---|---|---|
| 1 | Darshan Desai, Ashish Desai | Integrating Generative AI in Business Intelligence: A Practical Framework for Enhancing Augmented Analytics | The article explains that BI investment is rapidly growing and that a major challenge is the gap between technical teams who build analytics and business experts who use them, a gap that GenAI and AA aim to bridge. GenAI allows non-technical users to interact with data through natural language, automates data preparation and predictive modeling, as seen in tools like Power BI and Tableau. Despite these benefits, adoption remains slow due to AI literacy gaps, data quality issues, ethical concerns and the "black box" nature of GenAI, making human validation and organizational goal alignment essential. The study proposes a framework emphasizing data quality, task automation and customizable AI workflows, illustrated through a prototype built with modern front-end and back-end technologies that simplifies KPIs creation and the full BI cycle for non-technical users. | Article |
| 2 | Sara Mantouzi, Said Youssef, Abderazzak El Mouatassima, Yassine Ourahoua, Hanane Jafia, Ihssane Zouhouredinea, Mouhcine Hamliria | Leveraging AI for business intelligence: A pathway to improved organizational performance in Morocco's manufacturing sector | This article emphasizes that globalization and fast-changing markets force companies to leverage BI for competitive advantage. AI integration in BI improves efficiency through automation, anomaly detection, forecasting and real-time insights. Successful adoption depends on system quality, skilled IT staff and addressing ethical concerns such as bias, privacy and transparency. AI in BI improves productivity, decision-making and innovation, but adoption relies heavily on user confidence and satisfaction. | Article |
| 3 | Sharmila Ramachandaran, Zubaidi Mahalley, Riska Nuraini, Bablu Kumar Dhar | Exploring the challenges of AI-driven business intelligence systems in the Malaysian insurance industry | Focusing on the Malaysian insurance sector, this study shows that AI integration in BI improves strategic decision-making by leveraging ML, predictive analytics, large datasets and uncovering hidden patterns. Key challenges include a shortage of skilled talent, regulatory complexity, cultural perceptions of AI and inadequate technological infrastructure. Successful AI adoption requires high-quality, well-prepared data, strong data governance and careful handling of privacy laws, which can both protect users and limit data availability. | Article |
| 4 | Montserrat Jiménez-Partearroyo, Ana Medina-López | Leveraging Business Intelligence Systems for Enhanced Corporate Competitiveness: Strategy and Evolution | This article emphasizes that BI has evolved from a supporting tool to a strategic asset, with real-time processing, cloud computing and Big Data as key enablers. AI and ML enhance BI by enabling predictive and prescriptive analytics, uncovering hidden patterns, improving operational processes, customer service and real-time decision-making. NLP allows users to interact naturally with BI systems, bridging the gap | Article |

| | | | | |
|---|---|---|---|---|
| | | | between technical experts and other users. Robust data governance is essential to ensure data accuracy, consistency and security. | |
| 5 | Andreea Vines, Ana-Ramona Bologa and Andreea-Izabela Bostan | Enabling Intelligent Data Modeling with AI for Business Intelligence and Data Warehousing: A Data Vault Case Study | This article focuses on using AI, especially LLMs, to automate data modeling tasks for Data Vaults in BI. LLMs can accelerate model creation, reduce manual effort and optimize data warehouse, or in this case, Data Vaults development, but their output depends heavily on precise prompt engineering and requires human evaluation due to risks of hallucinations and inconsistent results. The framework demonstrates AI's potential in improving efficiency while highlighting the necessity of skilled technical oversight. | Article |
| 6 | T. Gajić, A. Ivanišević̀d, S. Kneževiće | Employee Perceptions of BI and AI Tools for Service Transformation: Evidence from the Serbian Airline and Hotel Industries | This article focuses on employee perceptions of BI and AI in the airline and hotel industries. AI enhances BI by automating customer support (e.g., chatbots), improving forecasting, predicting operational conditions and enabling personalized services. Combining BI and AI improves decision-making, operational efficiency and customer experience, but adoption depends on employee acceptance and attention to privacy and security concerns. | Article |
| 7 | Asaad R. Kareem, Hasanen S. Abdullah | Leveraging Hadoop and Hybrid Deep Learning on Home Datasets for Business Intelligence | This article presents the use of Hadoop and hybrid deep learning in BI. Hadoop provides distributed computing capabilities that allow scalable storage, processing and analysis of large datasets. Deep learning enhances BI's predictive accuracy and decision-making capabilities and the combination with Hadoop increases efficiency in handling Big Data. | Article |
| 8 | Mohammad Musa Al-Momani, Thabet Ameen Alqudah, Issa Ahmad Al Swiety, Nihaiah Mahrakani, Mohammed Bassam Abdulraheem Nassoura, Mohammad Khalid Al Attar | Integrating Artificial Intelligence (AI) and Business Intelligence (BI): A Framework for Improving Enterprise Performance | This article presents a framework integrating AI and BI to improve enterprise performance. AI enhances BI by analyzing large and unstructured datasets, identifying patterns, forecasting and automating processes, which improves planning, risk management and employee support. Challenges include data dependency, system incompatibility and the need for employee training. Integration fosters collaboration between technical and administrative teams, improving communication and decision-making. | Article |
| 9 | Al-Momani MM, Awadallah A, Alnassar B, Ismail A, Nassoura M, Abudarwish | The Role of Business Intelligence in Supply Chain Optimization A Case Study of the Carrefour Market in Jordan | This article examines BI in supply chain optimization. BI, combined with AI/ML models like ARIMA and NN, supports real-time forecasting, inventory planning, supplier coordination and operational efficiency. This integration improves predictive accuracy, reduces forecast error, enhances responsiveness to market demand and increases customer satisfaction and profitability. | Article |
| 10 | Warunee Milinthapunya, Urairat Yamchuti, Anake Nammakhunt, Chatchada Shawarangkoon, Panita Wannapiroon, Prachyanun Nillsook | Business Intelligence Management with Artificial Intelligence for Prediction Information Technology Infrastructure in Higher Education | This article focuses on BI-AI integration in higher education IT infrastructure. AI enables predictive analysis and forecasting, while BI provides descriptive insights from historical data. Together, they support resource optimization, trend identification and self-service reporting, allowing users to access, analyze and act on complex datasets efficiently. | Article |
| 11 | Md Jahidur Rahman, Tarek Rana, Yiren Xu, Peter Öhman | Bridging BI and AI Enhancing Operational Efficiency in the Chinese Financial Sector | This article explores how AI enhances BI. With advanced analytics, ML and automated decision-making, improving predictive accuracy and operational efficiency. Large companies benefit more due to better infrastructure, more data and skilled personnel. AI assists across the BI cycle, | Article |

| | | | | |
|---|---|---|---|---|
| | | | from ETL and Data Warehousing to visualization and strategic recommendations. | |
| 12 | Bimol Chandra Das, Shohoni Mahabub, Md Russel Hossain | Empowering modern business intelligence (BI) tools for data-driven decision-making: Innovations with AI and analytics insights | This article mentions that AI integration transforms BI from static, descriptive reporting to predictive and prescriptive self-service platforms. Tools like Microsoft Power BI, Tableau and QlikView improve speed and accuracy. Challenges include user adoption, data quality and shortage of AI-skilled personnel. Future trends point to storytelling and customized BI services. | Article |
| 13 | Afsaruddin, Devendra Agrawal, Sandeep Kumar Nayak | An Optimized Approach for Maximizing Business Intelligence using Machine Learning | This article recognizes that ML enhances BI's predictive capabilities by recognizing patterns, learning from data and enabling faster and more accurate forecasts. ML improves decision-making efficiency, resource savings and overall BI performance. | Article |
| 14 | Noorah A. Alghamdi, Heyam H. Al-Baity | Augmented Analytics Driven by AI: A Digital Transformation beyond Business Intelligence | This article explores how AA uses AI, ML and NLP to automate the analytics cycle, from data preparation to insight generation. Non-technical users can leverage data through NLP, bridging the gap with technical teams. AA accelerates repetitive tasks, pattern recognition and forecasting, although final decision-making still requires human expertise. Platforms leveraging AA include Tableau, Power BI, Qlik and ThoughtSpot. | Article |
| 15 | Alqhatani, A.; Ashraf, M.S.; Ferzund, J.; Shaf, A.; Abosaq, H.A.; Rahman, S.; Irfan, M.; Alqhtani, S.M | 360° Retail Business Analytics by Adopting Hybrid Machine Learning and a Business Intelligence Approach | This article notes how AI and ML in BI facilitate predictive and prescriptive analysis in retail, enabling customer segmentation, sales forecasting and pattern recognition. Cloud-based BI-AI integration ensures scalability, faster processing and actionable KPI generation for decision-making. ML models achieve high accuracy in classification, segmentation and forecasting. | Article |

# ANNEX A – EXTERNAL REPOSITORIES AND RESOURCES

This annex presents the external repositories and resources used in this work. It includes the challenges addressed in each case study, the traditional BI solutions used as a baseline for comparison and the respective authors. This information provides context for the practical component of the thesis and supports the evaluation of the proposed approaches.

Annex A 1 - Maven Marketing Challenge

https://mavenanalytics.io/challenges/maven-marketing-challenge

Annex A 2 - Maven Traditional BI Solution

https://github.com/Pravz-149/Marketing-Analytics---BI-Dashboard/tree/main

Annex A 3 - Maven Traditional BI Solution Author

https://www.linkedin.com/in/pravz149/

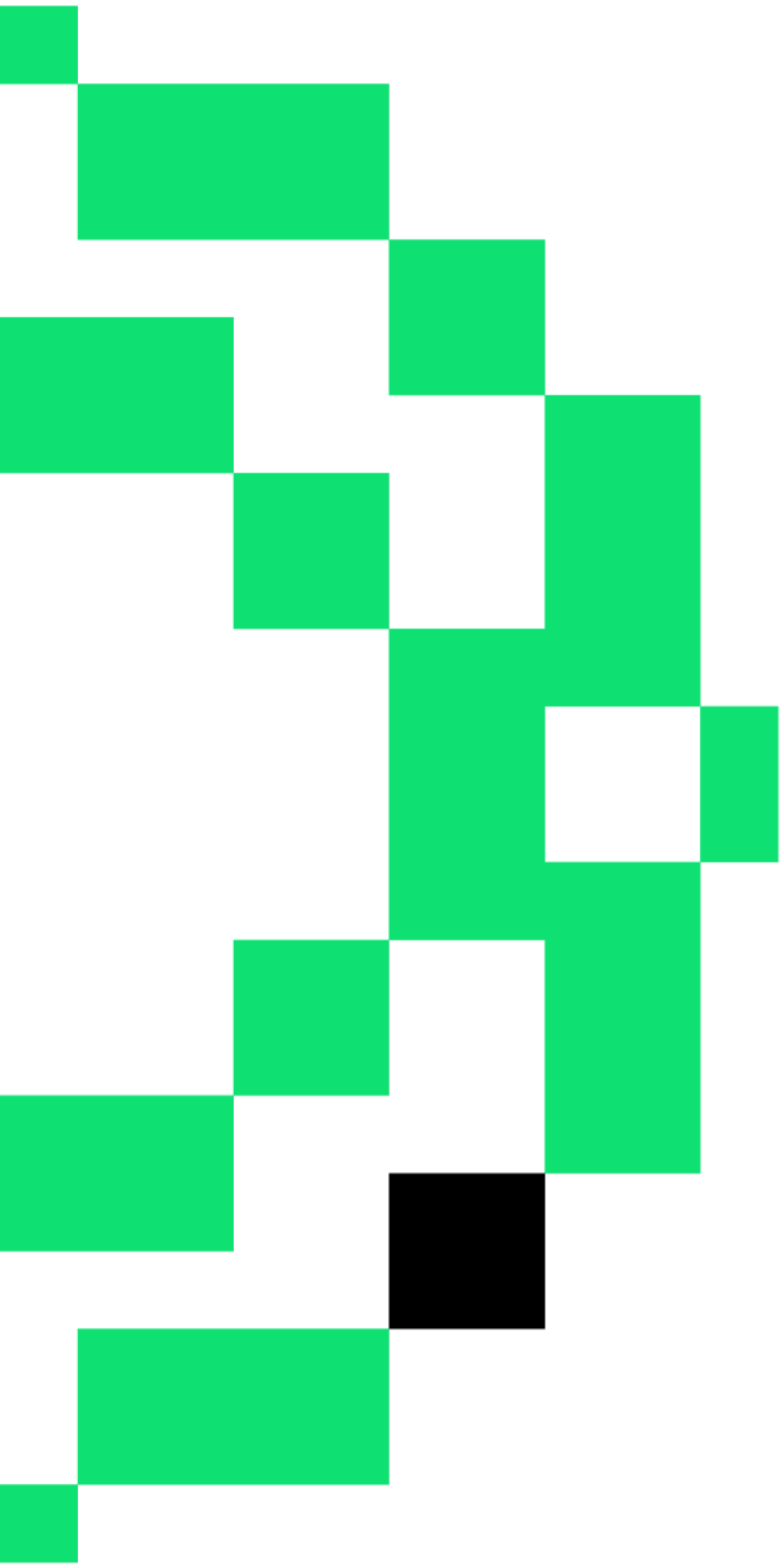
Annex A 4 - KPMG Data Analytics Virtual Program Solution

https://github.com/adimidania/kpmg-virtual-internship/tree/main

Annex A 5 - KPMG Data Analytics Virtual Program Solution Author

https://www.linkedin.com/in/alaadaniaadimi/

**Data with Purpose.**



Campus de Campolide, 1070-312 Lisboa, Portugal
**novaims.unl.pt**