

**School of Computer Science**

**Advanced Database Topics- COMP 8157**

**Summer 2023**

**Professor: Dr Shafaq khan**

**Empowering Business Strategy with Advanced Data Analysis and KPI Visualization**

**Project Milestone Report**

**By**

|  |  |
| --- | --- |
| **Student ID** | **Student Name** |
| 110100089 | Harbhajan Singh |
| 110097177 | Manjinder Singh |
| 110100116 | Navjot Makkar |
| 110088741 | Rohit Kumar |

Table of Contents

Table of Contents 2

**1.** **ABSTRACT** 3

**2.** **INTRODUCTION** 3

**3.** **LITERATURE STUDY** 4

**4.** **PROPOSED MODEL** 6

**5.** **SYSTEM DEFINITION** 6

**6.** **REFERENCES** 7

# **ABSTRACT**

This project aims to conduct an in-depth analysis on "Global Super Store Dataset" by leveraging key performance indicators (KPIs) and employing data visualization techniques. To enhance the reliability of the data, various preprocessing techniques will be applied. While previous research has emphasized the importance of KPIs in gaining insights from data, existing solutions often involve repetitive execution of Python or R scripts manually, resulting in time-consuming processes that are less user-friendly for businesses. To overcome these limitations, our project proposes an end-to-end solution that encompasses **data preprocessing**, **KPI calculation**, and **interactive data visualization using Tableau**. By integrating the powerful features of Tableau, users will have the capability to dynamically perform data analytics, including switching between categories and adjusting date ranges. Furthermore, this approach extends beyond visualizing the data itself to include visual representations of the calculated KPIs, thereby advancing the utilization of KPIs and optimizing the decision-making process through comprehensive data visualization and interactive exploration.

# **INTRODUCTION**

1. **Problem Description**

A survey reveals that 79% of executives acknowledge the insufficiency of relying solely on raw data for driving business decisions, leading to inadequate analytics and subsequent consequences [1]. This indicates **lack of comprehensive data preprocessing techniques, compromising the reliability of the data**. Furthermore, **traditional visualisation methods** such as static charts, graphs, and excel sheets involve manual processing, resulting in delays and limiting the scope of leveraging KPIs. Therefore, there is a need to develop an efficient and user-friendly solution that integrates data preprocessing, KPI calculations, and interactive data visualization.

1. **Motivation**

Recent research has highlighted the critical role of key performance indicators (KPIs) in extracting valuable insights from sales data. **KPIs provide a quantitative measure of business performance**, enabling organizations to assess progress towards their objectives. Moreover, data visualization has become increasingly important for decision-making processes as it allows for intuitive and interactive exploration of complex datasets. However, the existing solutions fail to fully leverage the power of KPIs and do not provide a user-friendly environment for data visualization and analysis. So, we will be proposing a solution that will **utilize the capabilities of KPIs and goes beyond basic visualization of data**.

1. **Solution Statement**

To ensure accurate analysis, we will begin by **employing statistical techniques** such as Z-Score, Interquartile Range (IQR), and Mahalanobis Distance to detect and address outliers in the product data. By treating outliers effectively, we can enhance the overall quality and reliability of the dataset.

Furthermore, we integrate these KPIs into an interactive Tableau dashboard, providing users with **dynamic data exploration** options on quarterly, half-yearly, yearly, or custom basis, switching between different regions, and different categories. The dashboard supports various visualizations tailored to specific analysis needs, allowing users to gain comprehensive insights into sales data trends and patterns. By combining customized KPIs, outlier treatment, and dynamic data visualization capabilities, our solution goal is to empowers businesses to comprehend their data more easily, make better-informed decisions, and maximize the benefits of data analysis.

# **LITERATURE STUDY**

The research paper titled **"Digital transformation: A multidisciplinary reflection and research agenda"[1]** published in Jan 2021 offers a thorough examination of how digital transformation affects businesses and consumers. According to the paper, digital transformation has significantly impacted consumer expectations and behaviours, leading to immense pressure on traditional companies and widespread disruption across various markets. The authors outline **three stages of digital transformation: digitization, digitalization, and digital transformation**. Digitization involves converting analogue information into digital formats, while digitalization focuses on leveraging digital technologies to enhance business processes and create new value propositions. The final stage, digital transformation, entails a complete **rethinking of business models** and the **establishment of new digital ecosystems**.

Within the context of digital transformation, the ability to acquire and analyze big data plays a crucial role in making informed decisions. Even though big data is widely available and collecting it has become easier, many companies struggle to develop the necessary skills to analyse and make use of this data. Accenture's research (2018) reveals that **79% of executives surveyed acknowledged that their most crucial systems and strategies rely on data, but a significant number of them have not made investments to ensure the reliability of this data**. This insight highlights the significance of investing in data analysis capabilities as part of the digital transformation journey.

The paper also presents a research agenda to inspire and guide future investigations into digital transformation. It emphasizes that merely analyzing data is not sufficient; it is **equally important to analyze the data correctly and present the results accurately**. The authors observe that often trends are identified but not accurately communicated, leading to inappropriate business policies. Therefore, this paper underscores the importance of refining the data, conducting analysis according to specific requirements, and effectively visualizing the findings as part of a comprehensive digital transformation strategy.

Overall, the author provides a solid foundation by summarizing key concepts from the research paper and highlighting the importance of data analysis and visualization in the context of digital transformation.

Coming to the data visualisation, another research paper titled **"Research on Python Data Visualization Technology"[2]** published in January 2021 explores the use of Python libraries, particularly **Matplotlib, to create visual representations of data**. The paper emphasizes the importance of visualizing data in diverse fields such as science, medicine, business, and engineering. In the business context, **data visualization is highlighted as a crucial factor in decision-making**. By visualizing data, businesses can identify trends and patterns in significant information like **customer data, sales figures, and other business metrics**. The authors argue that data visualization helps businesses make **better decisions and improve their financial performance**.

The paper recommends that researchers and professionals leverage tools like Matplotlib and Seaborn to create visual representations of data, enabling them to uncover valuable insights and connections within their datasets. However, the paper also highlights that despite the availability of these tools, **the data visualization process remains technical, requiring proficiency in programming languages**. This poses a challenge for many business decision-makers who may **lack the technical knowledge or resources** to fully utilize these tools effectively.

This paper presents the findings emphasizing the importance of data visualization in decision-making across various fields. While recommending the use of tools like Matplotlib, it acknowledges the technical nature of these tools and the potential barriers they pose for non-technical users in the business domain.

Based on our literature review, we have discovered that **key performance indicators (KPIs) are emerging as a popular method for accurately identifying trends and driving effective business decisions**. However, current approaches often involve analyzing raw, untreated data, resulting in limited insights. Additionally, existing visualization methods, such as static charts and numbers, lack flexibility and require technical programming knowledge.

To overcome these limitations, our project aims to enhance the data analysis process. We will begin by performing outlier detection and data refinement to ensure the reliability of the data. This preprocessing step will enable us to create meaningful KPIs that align with both the data and business requirements. Rather than static visualizations, we will leverage the power of Tableau, a dynamic visualization tool, to present the KPIs in an interactive and visually appealing manner. This approach not only expands the possibilities of KPI visualization but also empowers businesses to make more informed decisions.

**The World Economic Forum (WEF) promoted Tableau over other Business Intelligence tools because of its ease of use and seamless connection to other applications** [3]. Tableau offers flexible connectivity, integration, and extensibility with many databases and files, making it easy to integrate with other applications [4][5]. Tableau's combination of user-friendly interface, large dataset handling, customizable dashboards, integration capabilities, and powerful analytics has made it a popular choice for organizations across industries, including the WEF.

To summarize, by combining data preprocessing, tailored KPI calculation, and dynamic visualization using Tableau, we strive to revolutionize the way KPIs are visualized. This innovative approach opens new avenues for businesses to explore and extend the use of KPIs in their decision-making processes, potentially leading to the identification of novel KPIs and, ultimately, better decision-making outcomes.

# **PROPOSED MODEL**

To provide valuable insights into sales performance, we are developing a Tableau dashboard that analyses the sales data of different products sold by a company. In this project, we are referring to sales data available on Kaggle [6] platform.

Furthermore, after downloading data from the available resources that can be in many formats (Comma Separated Values in our case), we will be saving it in the relational database for further processing. In addition to this, some features of the database (MS SQL) will be utilized like SQL query with optimized runtime and views. For ensuring accurate analysis, we will begin by detecting and addressing outliers in the product data using statistical techniques of machine learning such as Z-Score, Interquartile Range (IQR), and Mahalanobis Distance. Once the outlier-treated data is obtained, we will process it to enhance its quality and reliability.

Moreover, we will be posting processing the data to extract meaningful insights, performing calculations for KPIs [7], and facilitating dynamic visualization using Tableau. The dashboard will offer flexible visualization options based on date range, country, and product.

A picture containing text, screenshot, font, diagram

Description automatically generated

**Figure**: Workflow of proposed solution

**By leveraging this dashboard, the company will gain a comprehensive understanding of their sales trends, identify areas of improvement, and make data-driven decisions**.

# **SYSTEM DEFINITION**

The main objective of this solution is to develop robust sales analysis system that delivers valuable insights with respect to the performance of various products sold by the company. It will assist them to devise future business strategies to derive growth of their sales. In this section, we will explain the key components and functionalities of the proposed solution.

1. **Database Management:** For database management, we will be using MS SQL to maximize the capabilities of database tool. To effectively manage and manipulate the sales data, optimized SQL queries for better runtime and views will be utilized.
2. **Data Preprocessing using Statistical Techniques:** To ensure that data used for analysis is accurate, we will be employing outlier detection techniques [1] to handle missing values and detect outliers within the available sales data. Moreover, for these techniques such as IQR (Inter Quartile Range) [8], Z-Score, and Mahalanobis Distance will be used to recognize and address outliers. By preprocessing and refining the data, we are ensuring the quality and reliability of the sales dataset for subsequent analysis.
3. **Data Analysis using KPIs:** After data preprocessing, the refined sales dataset will be utilized to extract meaningful insights and calculate KPIs [7] related to sales performance.
4. **Data Visualization using Tableau:** The solution will leverage Tableau's dynamic data visualization capabilities to create an interactive dashboard [3] [9]. This dashboard will not only visualize the sales data but also showcase the calculated KPIs, enabling business experts to gain deeper insights into sales trends, product performance, and market dynamics within the selected date range.

It will enable businesses to store their data in a robust database management system, refine it through **outlier detection techniques to enhance reliability, extract valuable insights by analyzing the data and calculating KPIs, and finally, visualize the data and KPIs dynamically**. This empowers businesses to make informed decisions based on comprehensive and visually appealing representations of their sales data, ultimately optimizing their decision-making process.

# **REFERENCES**

[1] P. C. Verhoef et al., “Digital transformation: A multidisciplinary reflection and research agenda,” Journal of Business Research, vol. 122, pp. 889–901, 2021, doi: https://doi.org/10.1016/j.jbusres.2019.09.022. Available: <https://www.sciencedirect.com/science/article/pii/S0148296319305478>

[2] S. Cao, Y. Zeng, S. Yang, and S. Cao, “Research on Python Data Visualization Technology,” *Journal of Physics: Conference Series*, vol. 1757, p. 012122, Jan. 2021, doi: <https://doi.org/10.1088/17426596/1757/1/012122>

[3] salesforce, “Tableau Helps the World Economic Forum Shape Global, Regional and Industry Agendas,” *Tableau*, 2022. Available: <https://www.tableau.com/solutions/customer/tableau-helps-world-economic-forum-shape-global-regional-and-industry-agendas>

[4] Tableau, “What is Tableau?” *Tableau*, 2022. Available: https://www.tableau.com/why-tableau/what-is-tableau

[5] Tableau, “Connectivity, integration, and extensibility,” *Tableau*, 2022. Available: <https://www.tableau.com/products/our-integrations>

[6] A. MAHALINGAPPA, “Global Super Store Dataset,” *www.kaggle.com*, 2020. Available: https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset

[7] M. Dahr, A. Khalaf, I. Najm, and M. Ahmed, “IMPLEMENTING SALES DECISION SUPPORT SYSTEM USING DATA MART BASED ON OLAP, KPI, AND DATA MINING APPROACHES,” *Journal of Engineering Science and Technology*, vol. 17, p. 0275, Feb. 2022.

[8] J. Ostrowski and M. József, “Statistical Analysis of Machinery Variance by Python,” *Acta Polytechnica Hungarica*, vol. 17, pp. 151–168, Jan. 2020, doi: https://doi.org/10.12700/APH.17.5.2020.5.8

[9] Syed Muhammad Mubashir Rizvi, “Mastering data visualization: Simplifying complex data for better insights | Data Science Dojo,” *datasciencedojo.com*, May 2023. Available: https://datasciencedojo.com/blog/data-visualization-guide/. [Accessed: Jun. 21, 2023]