

Large-Scale Retail Analytics: Market Basket Analysis & Customer Segmentation for Strategic Decision Making

Prof. Ujwala Khartad¹, Mr. Pranav Ambadkar²

¹Professor, Department of Computer Science & Engineering (Data Science)

²Student, Department of Computer Science & Engineering (Data Science)

Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India

Ujwala.ds@tgpcet.com and pranavambadkar20982@gmail.com

Abstract: *In the modern retail environment, businesses generate large volumes of transactional data that can be leveraged to improve decision-making. This research presents an analytical framework combining customer segmentation and market basket analysis to extract meaningful insights from retail data. The study utilizes a dataset of mobile sales transactions and applies data preprocessing, segmentation techniques, and descriptive analytics using Microsoft Power BI. Customer segmentation is performed based on demographic attributes, while product-level analysis identifies purchasing patterns. The results highlight key customer groups, seasonal sales trends, and product performance. Although advanced association rule mining is not fully implemented, the study demonstrates the potential of data analytics tools in supporting business strategies. The findings provide valuable insights for improving marketing, inventory management, and customer engagement in retail environments.*

Keywords: Retail Analytics, Customer Segmentation, Market Basket Analysis, Business Intelligence, Power BI, Data Visualization

I. INTRODUCTION

Retail organizations today operate in a highly competitive and data-rich environment. Every transaction generates valuable information about customer behaviour, product demand, and sales performance. However, without proper analysis, this data remains underutilized.

This research focuses on transforming raw retail data into actionable insights through customer segmentation and exploratory market basket analysis. By using business intelligence tools, organizations can better understand customer preferences and optimize their strategies.

The primary objective of this study is to develop a structured analytical approach that supports data-driven decision-making in retail businesses.

II. LITERATURE SURVEY

Recent studies emphasize the growing importance of business intelligence in retail analytics. Gupta and Sharma (2023) highlighted how BI tools improve operational efficiency by enabling better visualization and reporting. Verma and Singh (2024) discussed the role of segmentation in enhancing customer targeting strategies.

Traditional market basket analysis techniques, such as the Apriori algorithm, have been widely studied. However, recent trends focus more on visual analytics and dashboard-based exploration, especially in scenarios where datasets do not support multi-item transactions.

Industry reports from Deloitte (2025) and McKinsey (2024) indicate that organizations using data visualization tools experience improved decision-making speed and accuracy.



This study builds upon these insights by focusing on descriptive analytics and interactive dashboards, rather than predictive or algorithmic approaches.

III. METHODOLOGY

The methodology adopted in this study follows a systematic data analytics pipeline aimed at transforming raw retail data into meaningful insights through structured processing and visualization. The analysis is based on a dataset containing mobile sales transactions, which includes information related to products, customers, sales values, and payment methods. Initially, data preprocessing is carried out using Power Query to ensure data quality and consistency. This involves handling missing values, standardizing data formats, and creating derived fields such as revenue and categorized age groups.

Following preprocessing, customer segmentation is performed to group customers based on key attributes such as age, geographic location, and purchasing behavior. This helps in identifying high-value customer segments and understanding variations in consumer patterns. Additionally, an exploratory approach to market basket analysis is applied to examine product-level trends and identify frequently sold items, providing insights into product performance.

To support efficient analysis, a structured data model is developed using a Star Schema approach, consisting of a fact table containing transactional data and a dimension table for date-based analysis. Further, Data Analysis Expressions (DAX) are utilized to compute key performance indicators such as Total Sales, Month-To-Date (MTD), and Year-on-Year (YoY). Finally, the processed data is presented through an interactive dashboard, enabling clear visualization of trends and patterns, and supporting effective data-driven decision-making.

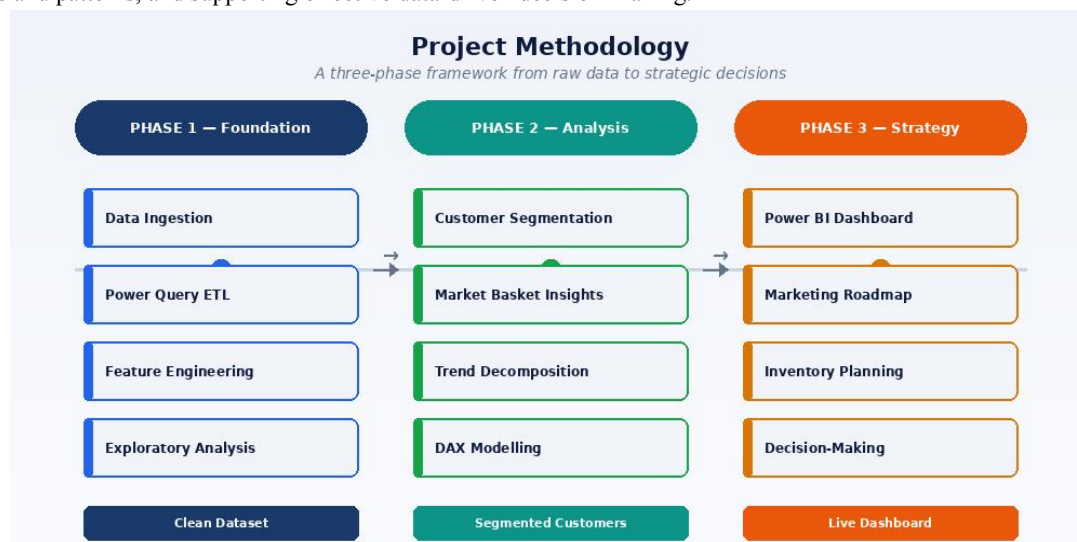


Fig. 1: Methodology Adopted

IV. IMPLEMENTATION

The implementation phase focuses on the practical execution of the proposed methodology using business intelligence tools. The dataset is first imported and prepared using Power Query, where data cleaning, formatting, and transformation are performed. Missing values are handled, and additional features such as revenue and age group classifications are created to enhance the analysis.



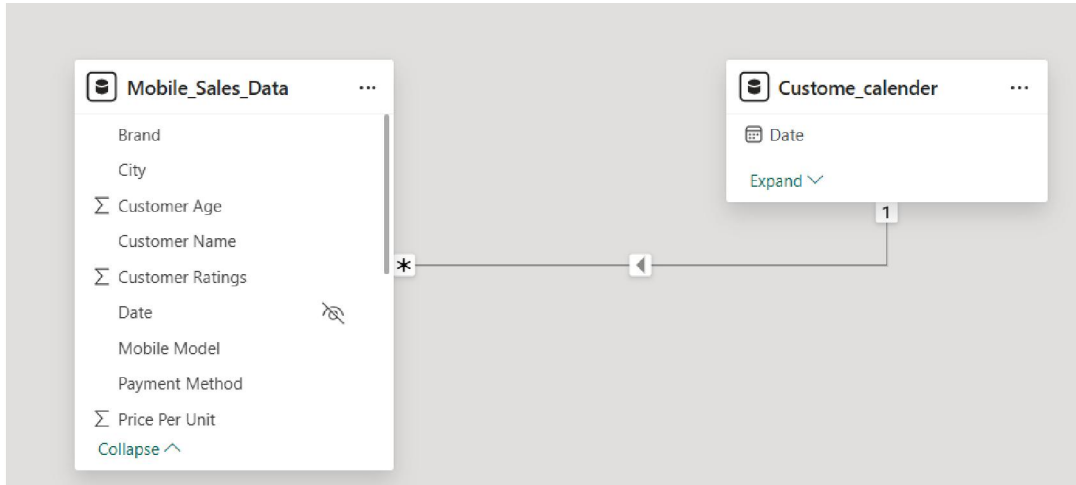


Fig. 2: Data Model

A Star Schema data model is then implemented to structure the dataset efficiently. The model consists of a fact table named *Mobile_Sales_Data* and a dimension table known as the *Calendar* table, which enables time-based analysis and improves query performance. Following this, DAX is used to develop key performance indicators, including Total Sales, Total Quantity, Average Price, Month-To-Date (MTD), and Same Period Last Year (YoY). These measures allow dynamic and context-based analysis of the data.

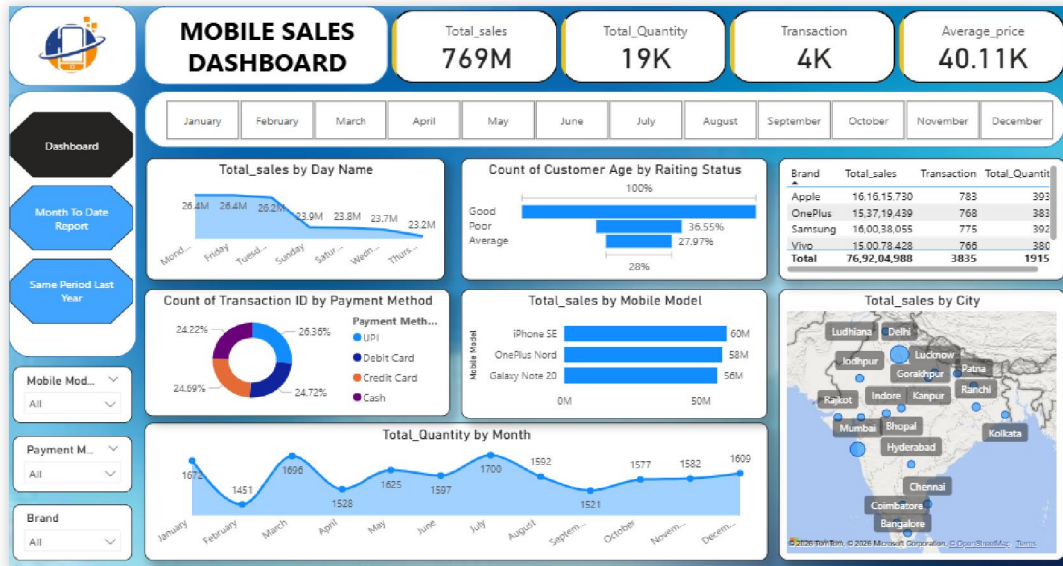


Fig. 3: Dashboard Large-Scale Retail Analytics

An interactive dashboard is developed to present the analytical results in a user-friendly manner. The dashboard includes KPI cards, time-series charts, geographic maps, and interactive filters, allowing users to explore the data from multiple perspectives. The entire system follows a structured workflow, starting from data collection and preprocessing, followed by feature engineering, data modelling, DAX-based analysis, dashboard development, and finally insight generation. This implementation ensures a smooth transition from raw data to actionable business insights.



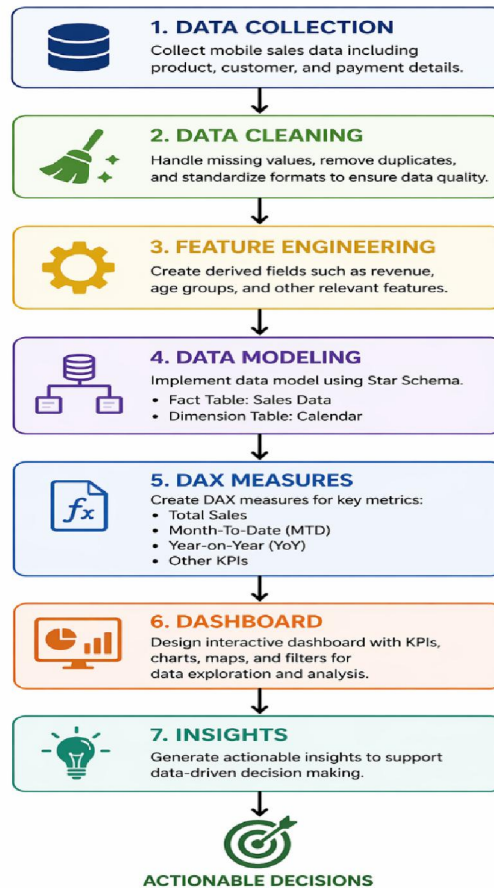


Fig. 4: Dataflow Large-Scale Retail Analytics

V. RESULT & ANALYSIS

The analysis of the mobile sales dataset provides important insights into customer behaviour and business performance. Customer segmentation shows that the age group 26–40 contributes the highest share of revenue, making it a key target segment. Sales trends indicate a noticeable peak during the October to December period, highlighting the impact of seasonal demand. The distribution of payment methods is relatively balanced, suggesting that customers use multiple payment options and businesses must support all channels. Product-level analysis reveals that certain mobile models consistently generate higher revenue, playing a significant role in overall sales. Additionally, year-on-year comparison reflects steady growth, indicating a positive business trend. Overall, the results demonstrate that descriptive analytics and visualization effectively support informed decision-making in retail environments.



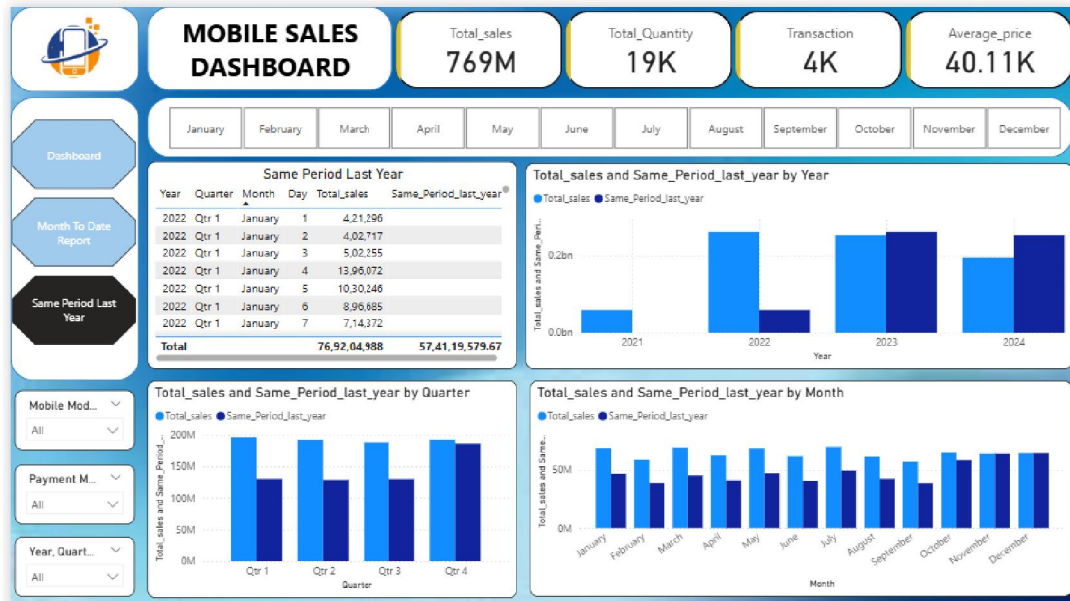


Fig. 5: SPLY Dashboard Large-Scale Retail Analytics

VI. FUTURE SCOPE

The project can be further enhanced by incorporating more advanced analytical capabilities and system improvements. Future work may include the implementation of association rule mining techniques, such as the Apriori algorithm, to identify deeper product relationships. Integration with real-time data sources can enable dynamic and up-to-date analysis. Additionally, predictive analytics models may be developed to forecast sales trends and customer behaviour. Deploying the system on cloud platforms can improve scalability, accessibility, and overall performance, making the solution more suitable for real-world applications.

VII. CONCLUSION

This research demonstrates that effective retail analytics does not necessarily require complex machine learning models. By using structured data modelling, descriptive analysis, and interactive visualization, meaningful insights can be extracted from transactional data.

The use of Power BI enables businesses to convert raw data into actionable knowledge, improving decision-making in areas such as marketing, inventory, and customer engagement.

The study highlights the importance of clarity, simplicity, and usability in data analytics solutions.

VIII. ACKNOWLEDGEMENT

The author would like to thank Mentor for her continuous guidance, support, and valuable feedback throughout the development of this project. Her insights and suggestions played an important role in shaping the overall direction and quality of the work.

The author also expresses gratitude to the Department of Computer Science and Engineering (Data Science) for providing the necessary academic environment, resources, and support required to successfully complete this research. Special thanks are extended to the developers of Microsoft Power BI and related tools, which were essential for data processing, modelling, and visualization. These tools made it possible to transform raw data into meaningful insights and develop an interactive analytical solution.



The author is also thankful to the broader data analytics and business intelligence community for their contributions in the field. Existing research, documentation, and shared knowledge provided a strong foundation for understanding data visualization, customer segmentation, and retail analytics concepts used in this study.

Finally, the author acknowledges all those who directly or indirectly contributed to the successful completion of this project.

REFERENCES

- [1]. S. Gupta and R. Sharma, "Retail Data Analytics Using Business Intelligence Tools," *International Journal of Data Science and Analytics*, vol. 12, no. 2, pp. 45–58, 2023.
- [2]. P. Verma and A. Singh, "Customer Segmentation Techniques in Retail Industry Using BI Tools," *Journal of Business Analytics*, vol. 9, no. 1, pp. 22–35, 2024.
- [3]. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, 1994. (Classic reference for Apriori algorithm — important to keep)
- [4]. Deloitte, "Global Retail Industry Outlook 2025," Deloitte Insights, 2025.
- [5]. McKinsey & Company, "The Future of Retail Analytics and Data-Driven Decision Making," McKinsey Report, 2024.
- [6]. Microsoft Power BI Documentation, "Power BI Guide and Business Intelligence Capabilities," Microsoft Learn, 2025.
- [7]. Statista, "Retail Data Analytics and Digital Transformation Trends," Statista Research Department, 2025.

