



Purchase Pattern Analytics using Market Basket Analysis for Retail Insights

Dnyaneshwari V. Nanekar¹

Student, MBA Department
Dhole Patil College of Engineering, Pune

Prof. Kanif Satav²

Project guide, MBA Department
Dhole Patil College of Engineering, Pune

Author Email: nanekardnyaneshwari@gmail.com | ORCID: <https://orcid.org>

How to Cite this Article:

Nanekar, D. V. (2026). Purchase Pattern Analytics using Market Basket Analysis for Retail Insights. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04). <https://doi.org/10.55041/ijcope.v2i4.705>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i4.705>

Abstract—

Understanding how customers purchase products has become very important in today's retail environment, where businesses aim to improve sales and customer experience through data-driven strategies. With large volumes of transaction data being generated every day, it is possible to identify patterns in customer buying behaviour that were not easily visible earlier. These patterns can help retailers make better decisions related to product placement, cross-selling, and inventory management.

In this project, we analyze customer purchase behaviour using transactional retail data. Each transaction represents a basket of items purchased together, which allows us to study how different products are related to each other. The dataset was first cleaned and transformed to remove inconsistencies, handle missing values, and structure it into a basket format suitable for analysis.

To identify product associations, the Apriori algorithm was applied. This method helps in finding frequently occurring itemsets and generating association rules based on support, confidence, and lift. These metrics help in understanding how strongly products are connected and how likely they are to be purchased together.

The results show that certain products are consistently bought together, indicating strong relationships between them. These insights can be useful for

creating combo offers, improving store layout, and designing targeted marketing strategies. Overall, the study demonstrates how analyzing purchase patterns can help retailers make practical and informed business decisions.

Keywords-

Purchase Pattern Analytics, Market Basket Analysis, Apriori Algorithm, Association Rules, Retail Analytics, Customer Buying Behaviour.



I. INTRODUCTION

The retail industry has become highly competitive, with businesses constantly looking for ways to increase sales and better understand customer behaviour. Today, customers have many choices and their buying decisions are influenced by factors such as product variety, pricing, and convenience. Because of this, retailers need to go beyond traditional methods and use data to understand what customers are actually purchasing and how products are related to each other.

With the growth of digital billing systems and online transactions, a large amount of purchase data is generated every day. Each transaction contains useful information about what items customers buy together, how frequently they shop, and how much they spend. This data creates an opportunity to study purchase patterns and gain insights that can help improve business decisions. However, analyzing such large datasets manually is difficult, and simple analysis methods often fail to capture meaningful relationships between products.

This is where data analytics techniques, especially market basket analysis, become useful. By using algorithms like Apriori, it is possible to identify frequently purchased item combinations and understand the association between products. These insights help retailers in making decisions related to product placement, cross-selling, bundling offers, and inventory planning.

In this project, we focus on analyzing retail transaction data to identify purchase patterns and product associations. The goal is to understand customer buying behaviour and use these insights to support better retail strategies. By applying structured data preprocessing and association rule mining techniques, this study aims to provide meaningful insights that can be directly applied in real-world business scenarios.

II. LITERATURE REVIEW

Purchase pattern analysis has been widely used in the retail sector to understand customer buying behaviour. Earlier studies mainly focused on identifying frequently purchased items using basic statistical methods, but these approaches were limited in capturing relationships between multiple products.

With the advancement of data mining, techniques like the Apriori algorithm [1] became popular for identifying frequent itemsets and generating association rules based on support, confidence, and lift. Later, improved methods such as FP-Growth [2] were introduced to handle large datasets more efficiently.

Research also highlights the importance of data preprocessing, including cleaning, handling missing values, and converting data into basket format, as these steps directly affect the accuracy of results [3].

However, many existing studies focus mainly on generating rules without clearly linking them to business applications. This study addresses that gap by combining pattern mining with visualization and practical interpretation, making the results more useful for retail decision-making.

III. METHODOLOGY

This study follows a structured methodology to analyze customer purchase behaviour using retail transaction data. The dataset consists of transaction-level details such as BillNo, item name, quantity, price, and customer information. The initial step involved understanding the dataset structure, identifying data types, and examining basic distributions.

Data preprocessing was a crucial stage in this study. Missing values in the CustomerID field were handled appropriately, duplicate records were identified, and inconsistencies in item names were cleaned. Outlier analysis was performed on Quantity and Price to detect abnormal values such as bulk purchases or negative entries. The dataset was then transformed into a basket format, where each transaction represents a set of items purchased together.

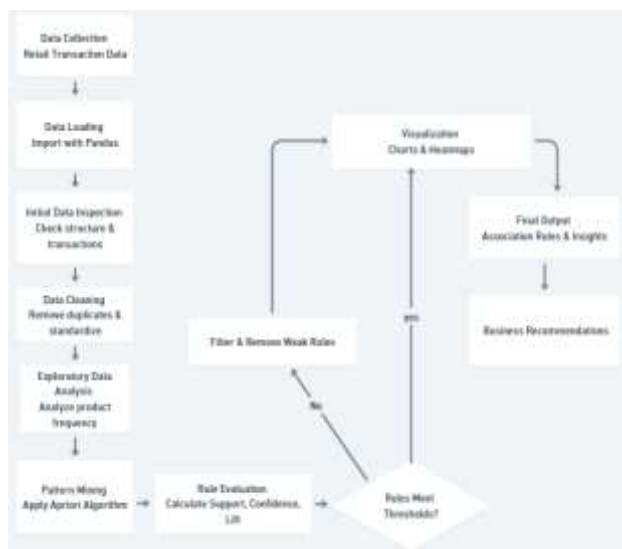
For pattern discovery, the Apriori algorithm was applied to identify frequent itemsets. Minimum thresholds for support and confidence were defined to filter meaningful patterns. Association rules were generated using key metrics such as support, confidence, and lift, which help in understanding the strength and significance of relationships between products.

The results were further analyzed and visualized using charts such as frequent itemsets, product combinations, and purchase trends. These visualizations helped in interpreting customer buying behaviour in a clear and practical way.

Overall, this methodology ensures a systematic flow from data preprocessing to pattern extraction and interpretation, enabling the generation of actionable insights for retail decision-making such as cross-selling, product bundling, and inventory planning.

In addition to the core analysis, feature engineering techniques were applied to enhance the depth of insights. New attributes such as Total Transaction Value, Basket Size, Item Frequency were derived from the original dataset to better understand customer purchasing intensity and product demand. Frequency-based categorization was also performed to group products into different demand levels such as high, medium, and low. Furthermore, a product combination matrix was created to systematically capture co-occurrence patterns between items across transactions. These enhancements not only improved the interpretability of the results but also supported more meaningful visualization and business-oriented analysis.

Fig 1: Market Basket Analysis Pipeline for Purchase Pattern Analytics



IV. RESULTS AND DISCUSSION

In this study, the analysis was carried out using a structured pipeline that included data preprocessing, transformation, and pattern extraction. Unlike prediction-based models, this project focused on identifying relationships between products using association rule mining techniques.

After cleaning and preparing the dataset, the transaction data was converted into a basket format suitable for market basket analysis. The Apriori algorithm was then applied to generate frequent itemsets. To ensure meaningful results, minimum thresholds were defined for key parameters such as support (0.02) and confidence (0.30). Additionally, only rules with lift greater than 1 were considered, as they indicate a positive association between products. The generated association rules revealed several interesting purchase patterns. Frequently bought product combinations were identified, showing that certain items are commonly purchased together within the same transaction. These patterns highlight customer buying behaviour and provide insights into product relationships.

Performance in this study was evaluated based on the strength and usefulness of the generated rules rather than traditional accuracy metrics.

The key evaluation measures included: Support, Confidence, Lift

These metrics helped in filtering strong and reliable rules that can be used for business decision-making.



V. CONCLUSION

This project demonstrates how data analysis techniques can be used to understand customer purchase behaviour in a retail environment. By working with transaction-level data, the study was able to identify meaningful patterns and relationships between products using association rule mining.

The implementation of the Apriori algorithm provided clear insights into which items are frequently purchased together. These insights can help businesses make better decisions related to product placement, cross-selling, and promotional strategies. Instead of relying on assumptions, retailers can use data-driven findings to improve sales and customer experience.

One of the key learnings from this project is the importance of data preparation. Cleaning the dataset, handling missing values, and converting the data into the right format were essential steps that directly influenced the quality of the results. It also highlighted how selecting appropriate thresholds for support, confidence, and lift is necessary to extract useful and actionable rules.

Although the analysis provided valuable insights, there are still opportunities for improvement. Future work can include using larger datasets, experimenting with advanced algorithms like FP-Growth, and implementing real-time recommendation systems to enhance practical applications.

In conclusion, this study shows that purchase pattern analytics can play a significant role in helping businesses understand customer behaviour, optimize strategies, and increase overall efficiency.

VI. REFERENCES

- [1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int. Conf. Very Large Data Bases (VLDB), pp. 487–499, 1994.
- [2] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," 3rd ed., Morgan Kaufmann, 2011.
- [3] T. M. Mitchell, "Machine Learning," McGraw-Hill, 1997.
- [4] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
- [5] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning," Springer, 2013.
- [6] M. Hahsler, B. Grün, and K. Hornik, "Introduction to arules – Mining Association Rules and Frequent Itemsets," R Package Documentation, 2005.
- [7] P. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining," Pearson Education, 2006.
- [8] Python Software Foundation, "Python Documentation," <https://www.python.org/>
- [9] Microsoft Corporation, "Power BI Documentation," <https://learn.microsoft.com/en-us/power-bi/>
- [10] Kaggle, "Online Retail Dataset," <https://www.kaggle.com/>