A PROJECT REPORT

On

A STUDY ON PERSONALIZED RECOMMENDATIONS THROUGH CUSTOMER SEGMENTATION AND MARKET BASKET ANALYSIS IN THE RETAIL SECTOR

DONE FOR

Project Report Submitted in partial fulfilment of the requirement of
PONDICHERRY UNIVERSITY for the award of the degree of
MASTER OF BUSINESS ADMINISTRATION
IN
DATA ANALYTICS

Submitted By
MOHAMMED IRSHAD K
(Reg. No. 23401053)

Under the guidance of
DR. J. RAMA KRISHNA NAIK
Assistant Professor
&
Mr. SHANTHAKUMAR
IT Department, Fresh2Day

DEPARTMENT OF MANAGEMENT STUDIES
SCHOOL OF MANAGEMENT
PONDICHERRY UNIVERSITY
PUDUCHERRY – 605 014

MAY 2025

DEPARTMENT OF MANAGEMENT STUDIES
Pondicherry University, Pondicherry – 605 014, India
Ph: (O) 91-413-123123, Mobile:12345-12345

Dr. J. RAMA KRISHNA NAIK, **M.B.A., Ph.D., UGC-PDF.**
Assistant Professor

C E R T I F I C A T E

This is to certify that the project titled **'A STUDY ON PERSONALIZED RECOMMENDATIONS THROUGH CUSTOMER SEGMENTATION AND MARKET BASKET ANALYSIS IN THE RETAIL SECTOR'** submitted for the award of Degree of Master of Business Administration, in the Department of Management Studies, Pondicherry University, Pondicherry, India, is a record of bonafide project work carried out by **MOHAMMED IRSHAD K** under my supervision.

Place: Pondicherry
Date:

(Dr. J. RAMA KRISHNA NAIK)

Countersigned by
Dr. R. Kasilingam

Professor & Head,
Department of Management Studies,
Pondicherry University

Project Viva Voce Examination

Date of Viva: _____

(Signature & Name of Viva Examiner)

DECLARATION

I hereby declare that the project titled, **"A study on personalized recommendations through customer segmentation and market basket analysis in the retail sector"** is an original work done by me under the guidance of **Dr. J. Rama Krishna Naik, Assistant Professor,** Department of Management Studies, Pondicherry University, and Mr. Shanthakumar, IT Department, Fresh2Day, Gopalapuram, Chennai, Tamil Nadu. This project or any part thereof has not been submitted for any Degree / Diploma / Associateship / Fellowship / any other similar title or recognition to this University or any other University.

I take full responsibility for the originality of this report. I am aware that I may have to forfeit the degree if plagiarism has been detected after the award of the degree. Notwithstanding the supervision provided to me by the Faculty Guide, I warrant that any alleged act(s) of plagiarism in this project report are entirely my responsibility. Pondicherry University and/or its employees shall under no circumstances whatsoever be under any liability of any kind in respect of the aforesaid act(s) of plagiarism.

Mohammed Irshad .K
Reg. No. 23401053
Department of Management Studies
Pondicherry University

Date:
Place: Pondicherry 605 014

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my research guide **Dr. J. Rama Krishna Naik**, Assistant Professor, Department of Management Studies, Pondicherry University for his incessant encouragement and support extended throughout the research period.

Special thanks to Mr. Shanthakumar, IT Department,Fresh2Day, Gopalapuram, Chennai, Tamil Nadu, for his guidance and advice throughout the period.

Many thanks to those faculty members who helped me in sharpening my thinking by cheerfully providing challenging comments and questions.

I thank all my classmates of Pondicherry University who had given me moral support.

MOHAMMED IRSHAD K
MBA DATA ANALYTICS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF GRAPHS

ABSTRACT

This project focuses on building a Personalized Recommendation System for Retail using customer transaction data collected from Fresh2Day software. The dataset includes customer IDs, product categories, and sales information. A Customer Segmentation was performed using the RFM (Recency, Frequency, Monetary) model to identify high-value customers. Customers were scored between 3 to 15 based on their RFM values and segmented into High, Medium, and Low value groups.

To validate if these segments differed in purchasing behaviour, an ANOVA (Analysis of Variance) test was conducted, revealing statistically significant differences in spending among segments. Focusing on the high-value customers, Market Basket Analysis (MBA) was performed using the Apriori algorithm with a minimum support of 1% and confidence of 50%. These metrics help identify product combinations frequently purchased together.

Based on these associations, an RFM-based Recommendation Engine was built. The engine recommends products that a customer has not yet purchased but are likely to be of interest based on the buying patterns of similar high-value customers. This approach supports personalized cross-selling, enhancing the customer experience and potentially increasing revenue.

CHAPTER – 1
INTRODUCTION

1.1 INTRODUCTION:

The retail industry is one of the most dynamic and customer-centric sectors in the global economy. With increasing competition and the growing influence of digital platforms, retailers are focusing more on understanding consumer behaviour to enhance customer experience and drive sales. Data-driven insights have become essential for making informed decisions in areas such as marketing, product placement, and customer engagement. One of the most powerful tools in modern retail is the recommendation system. These systems suggest relevant products to customers based on their purchase history, preferences, or the behaviour of similar customers.

To segment customers based on their value to the business, RFM (Recency, Frequency, Monetary) analysis is widely used. RFM is a marketing technique that evaluates customer behaviour by analysing how recently a customer made a purchase (Recency), how often they purchase (Frequency), and how much they spend (Monetary). Customers are scored on these metrics to identify high-value and loyal customers, helping businesses target them more effectively. Market Basket Analysis (MBA) is another important technique used in retail analytics. It identifies combinations of products frequently bought together, using association rules. With algorithms like Apriori, MBA reveals valuable insights such as “customers who bought product A also tend to buy product B.” These insights are measured using metrics like support and confidence, which help assess the strength and frequency of these associations.

Integrating RFM segmentation with Market Basket Analysis enables businesses to generate smarter, more personalized product recommendations. This combined approach enhances the accuracy of suggestions and ensures that valuable customers receive product recommendations that align with their purchasing behaviour, leading to improved customer engagement and increased revenue.

1.2 COMPANY PROFILE:

Fresh2Day is Chennai's leading food and grocery store, offering a wide range of over 200 handpicked products across categories like fresh fruits and vegetables, rice, dals, spices, bakery items, packaged foods, and beverages. With a strong focus on quality, affordability, and convenience, Fresh2Day aims to deliver a premium and hassle-free shopping experience to its customers. The store enhances in-store shopping with features like cut fruit boxes and juice bars, while also providing an online ordering option for doorstep delivery. Combining traditional grocery essentials with modern retail technology, Fresh2Day continues to redefine food shopping in Chennai.

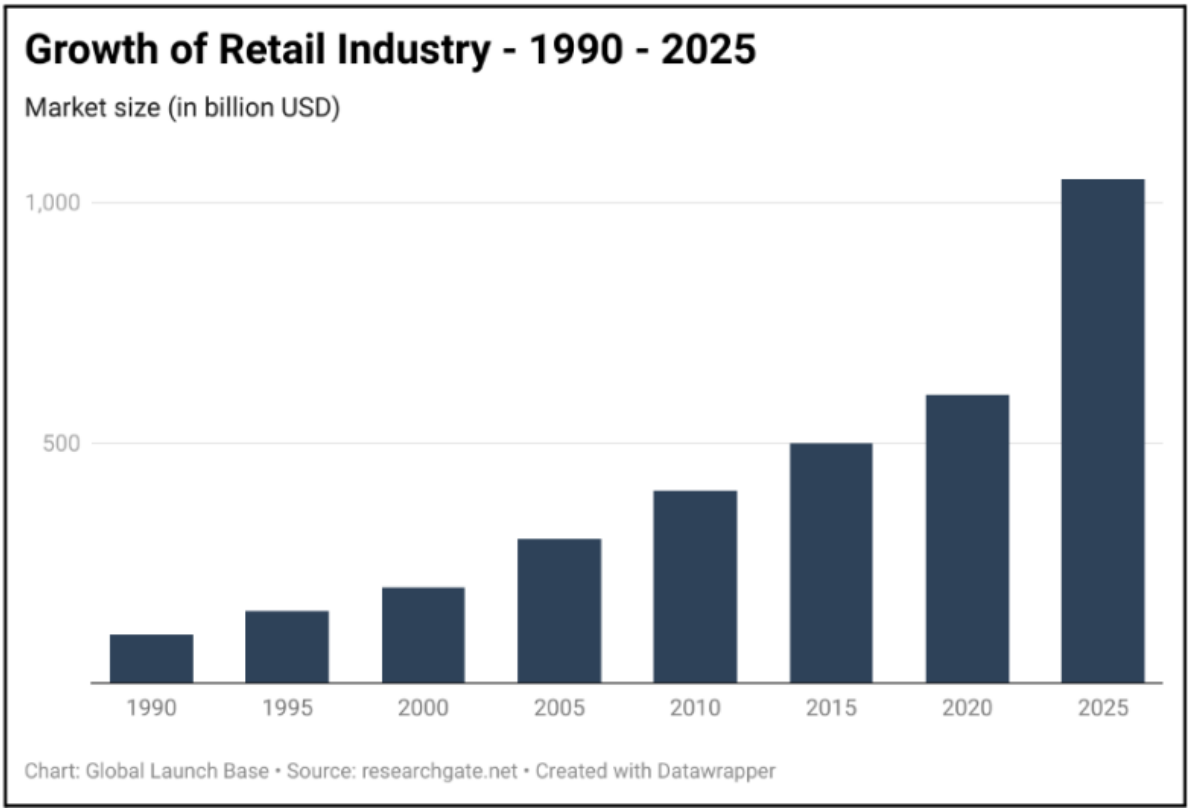
Industry: Retail
Founded year: 2021
Founder: Karthik Annamalai
Headquarters: Chennai
Products:

- Fresh Fruits
- Vegetables,
- Rice and Dals,
- Spices and Seasonings
- Bakery Products, Beverages,

1.3 INDUSTRY PROFILE

The retail industry plays a vital role in the economic development of a country by generating employment, driving consumption, and contributing to GDP. It refers to the sale of goods and services to the final consumer for personal use. With the evolution of consumer needs and preferences, the retail industry has transformed from traditional Kirana (small neighbourhood) stores to organized supermarkets, hypermarkets, and now, digital platforms.

India's retail sector is one of the largest in the world and is projected to reach over USD 1.5 trillion by 2030. The growth is driven by factors such as a rising middle class, increased disposable income, rapid urbanization, and technological advancement. India's demographic advantage, with a young population and growing digital literacy, makes it a lucrative market for retailers. Additionally, government initiatives like "Digital India" and FDI (Foreign Direct Investment) liberalization in multi-brand retail have further accelerated its expansion.



Among the various formats, supermarket retail chains are witnessing robust growth in India. These organized retail outlets provide a wide assortment of food and non-food items, offered in a self-service format. The convenience, availability of branded products, competitive pricing, and hygienic shopping environment make supermarkets a preferred choice for urban

consumers. Supermarket chains also enable effective supply chain management and allow retailers to use consumer data for targeted marketing.

Omni-Channel Retailing: Retailers started adopting omni-channel strategies, integrating physical stores with online platforms. This approach allowed consumers to research products online and make purchases either online or in-store. Retailers like Tata Group's Croma and fashion brands implemented these strategies to provide a seamless shopping experience.

Omni-channel retailing has gained significant traction in India as the country's retail landscape continues to evolve and adapt to changing consumer preferences and technological advancements. Indian retailers have embraced omni-channel strategies to provide customers with seamless and personalized shopping experiences across various channels. Here's how omni-channel retailing is shaping up in India:

- E-commerce Pioneers: India's booming e-commerce sector has been at the forefront of omni-channel adoption. Leading e-commerce platforms like Flipkart, Amazon India, and Snapdeal have expanded their offerings to include not only online marketplaces but also BOPIS (Buy Online, Pick Up In-Store) options and partnerships with brick-and-mortar stores for faster deliveries.

- Traditional Retailers Going Digital: Traditional retailers have recognized the importance of digital transformation. Many established brick-and-mortar brands have launched their own e-commerce websites and mobile apps to cater to the growing online shopper base. For instance, retail chains like Shoppers Stop, Reliance Retail, and Tata-owned Croma have developed omni-channel strategies to bridge the gap between offline and online shopping.
- Click-and-Mortar Stores: Some Indian retailers have adopted the "click-and-mortar" approach, wherein they maintain both physical stores and robust online platforms. This approach allows customers to seamlessly move between online and offline channels. Brands like Nykaa, a cosmetics and beauty retailer, have successfully combined their e-commerce presence with offline stores to offer customers a comprehensive shopping experience.

- Fashion and Lifestyle Industry: The fashion and lifestyle sector in India has been a key driver of omni-channel strategies. Leading fashion brands like Myntra, Lifestyle, and

FabIndia have integrated their online and offline operations, offering customers options like trying on clothes in-store before ordering online or making online purchases and returning items to physical stores.

- Mobile-First Approach: With the increasing penetration of smartphones in India, mobile apps have become a crucial aspect of omni-channel strategies. Retailers focus on creating user-friendly and responsive apps that allow customers to browse, shop, and track orders on their mobile devices.
- Personalization and Loyalty Programs: Indian retailers are leveraging customer data to offer personalized recommendations, discounts, and promotions. Loyalty programs that span across online and offline channels encourage repeat business and brand loyalty.
- Challenges: While omni-channel retailing is on the rise in India, challenges such as inadequate technology infrastructure, supply chain complexities, and varying levels of digital literacy in different regions still exist. Retailers need to tailor their omni-channel strategies to suit the diverse Indian market.
- Government Regulations: Indian regulations around e-commerce and foreign direct investment (FDI) have influenced how some companies implement omni-channel strategies, particularly in the multi-brand retail segment. Regulations may impact how international brands and retailers establish their presence in India.

- Future Outlook: As technology continues to advance and consumer expectations evolve, omni-channel retailing is expected to play an increasingly vital role in the Indian retail sector. Retailers will likely continue to invest in technology, data analytics, and personalized experiences to create a seamless and convenient shopping journey for customers across various channels.

Grocery and Fresh Retail: The grocery segment saw significant transformation with players like BigBasket and Grofers offering online grocery delivery services. Additionally, numerous hypermarkets and supermarkets entered the scene, catering to consumers' demand for fresh produce and convenience.

Key Characteristics:

- Essential Goods: Grocery and fresh retail primarily deals with essential food items and household consumables that people need on a regular basis.

- Perishable Items: A significant portion of the products sold in this segment are perishable, which requires careful management of inventory, supply chain, and storage to maintain product quality and reduce wastage.
- Wide Product Range: Grocery and fresh retail encompasses a wide range of products, from staples like rice, flour, and lentils to fresh produce, dairy products, bakery items, frozen foods, and packaged goods.
- Frequent Purchases: Consumers tend to visit grocery stores regularly for their daily or weekly needs, making it an integral part of their routine.
- Customer Loyalty: Successful grocery retailers often build strong customer loyalty due to their consistent and reliable supply of essentials.

Challenges:

- Quality Maintenance: Maintaining the quality and freshness of perishable items is a significant challenge, requiring proper storage, transportation, and inventory management.
- Wastage Management: Reducing food wastage due to spoilage or expiration is a top concern in this sector.
- Price Sensitivity: Grocery shoppers are often price-sensitive and seek the best value for their money. Retailers need to strike a balance between quality and affordability.
- Competition: The grocery sector is highly competitive, with multiple retailers vying for consumers' attention. This can lead to slim profit margins.
- Supply Chain Complexity: Managing the supply chain for perishable products requires a robust distribution network and efficient transportation to ensure timely deliveries.

Modern Trends and Innovations:

- Online Grocery Shopping: The rise of e-commerce has led to the growth of online grocery platforms that allow customers to order essentials and have them delivered to their doorstep.
- Contactless Shopping: The COVID-19 pandemic has accelerated the adoption of contactless shopping methods, such as curbside pickup and home delivery.
- Smart Inventory Management: Retailers are using technology and data analytics to optimize inventory, reduce wastage, and streamline the supply chain.
- Private Labels: Many retailers are introducing their own private-label brands for grocery items, offering quality products at competitive prices.
- Fresh Formats: Some retailers are focusing on "fresh" formats, offering a wide selection of high-quality produce and perishables.

Key Players:

In India, the grocery and fresh retail sector is served by various players:

- Kirana Stores: Small neighborhood stores that play a crucial role in serving local communities with everyday essentials.
- Supermarkets: Larger stores that offer a broader range of products and are known for organized displays and air-conditioned shopping environments.
- Hypermarkets: Larger than supermarkets, hypermarkets provide an extensive range of products, including groceries, electronics, clothing, and more.
- Online Grocers: E-commerce platforms like BigBasket, Grofers, and Amazon Pantry offer online grocery shopping and home delivery.
- Specialty Stores: These focus on specific categories, such as organic foods, gourmet products, or health-conscious items.

The grocery and fresh retail segment is an essential and dynamic part of the retail industry, continually adapting to consumer preferences, technological advancements, and changing market dynamics.

CHAPTER – 2
REVIEW OF LITERATURE

2.1 LITERATURE SURVEY:

1. Agrawal, R., Imieliński, T., & Swami, A. (1993)

Topic: *Mining Association Rules Between Sets of Items in Large Databases*
This seminal paper introduced the **Apriori algorithm**, a key technique for uncovering association rules from transaction datasets. It explains how frequent itemsets can be extracted from large databases using iterative approaches. This algorithm serves as the backbone of **Market Basket Analysis (MBA)**, making it highly relevant for retail businesses to understand purchasing patterns and recommend items frequently bought together.

2. Hughes, A. M. (1996)

Topic: *The Complete Database Marketer*

Hughes laid the foundation for **RFM segmentation**, a marketing technique to rank customers based on Recency, Frequency, and Monetary value. He demonstrated how businesses could increase retention and engagement by focusing on valuable customers. His model supports personalization, especially in retail, by allowing tailored strategies based on customer behavior.

3. Berry, M. J. A., & Linoff, G. S. (2000)

Topic: *Mastering Data Mining: The Art and Science of Customer Relationship Management*
The authors explore data mining for **Customer Relationship Management (CRM)**, including RFM and association rules. They show practical applications of predictive modeling and segmentation in retail settings, offering insights into how data-driven decisions can improve targeting and cross-selling opportunities.

4. Chen, M.-S., Han, J., & Yu, P. S. (2002)

Topic: *Data Mining: An Overview from a Database Perspective*
This paper presents an overview of major data mining techniques, including **clustering**, **classification**, and **association rule mining**. The authors discuss their effectiveness in discovering hidden patterns in retail databases, supporting applications like segmentation, recommendation, and market analysis.

5. Kumar, V., & Reinartz, W. (2006)

Topic: *Customer Relationship Management: A Databased Approach* Focusing on data-driven CRM strategies, this book highlights how **RFM segmentation** can optimize marketing efforts and enhance customer loyalty. It emphasizes the importance of understanding customer value and predicting long-term behavior, which supports your approach to identifying high-value customers.

6. Bose, I., & Mahapatra, R. K. (2007)

Topic: *Business data mining — A machine learning perspective* This paper examines data mining tools used in business intelligence, specifically in retail. It compares traditional statistical models with modern **machine learning approaches**, including association rule mining and recommendation engines, stressing their use in improving decision-making and customer targeting.

7. Kim, Y. A., & Ahn, H. (2008)

Topic: *A recommender system using GA K-means clustering in an online shopping market* The authors proposed a hybrid recommendation system using **K-means clustering** and **Genetic Algorithms (GA)**. They segment customers based on shopping behavior to make personalized recommendations. This reinforces the idea that combining clustering with recommendation systems enhances personalization.

8. Chiu, C.-M., Hsu, M.-H., Lai, H., & Chang, C.-M. (2012)

Topic: *Re-examining the influence of trust on online repeat purchase intention* Although focused on consumer trust, this study indirectly supports the importance of personalized and reliable recommendation systems. It suggests that personalized marketing—based on behavioral data like RFM or MBA—can build trust and influence repeat purchases.

9. Kang, H. J., & Park, S. C. (2013)

Topic: *Integrated Modeling of Customer Segmentation and Lifetime Value for Improved Personalization* This study combines **RFM segmentation** with **Customer Lifetime Value (CLV)** modeling to deliver personalized marketing strategies. The authors show how

combining segmentation with value estimation leads to more effective targeting, directly supporting your strategy of applying MBA to high-value customers only.

10. Sharma, R., & Kaur, A. (2020)

Topic: *Market Basket Analysis for Online Retail Using Apriori Algorithm*
In this more recent work, the authors apply the **Apriori algorithm** to a real-world e-commerce dataset to discover item associations. Their findings confirm that MBA helps retailers understand co-purchase behavior and improve product placement and bundling, aligning with your recommendation approach.

CHAPTER – 3
RESEARCH METHODOLOGY

3.1 NEED AND SIGNIFICANCE OF THE STUDY

In the competitive retail landscape, understanding customer behaviour is crucial for enhancing sales and customer satisfaction. This study uses RFM segmentation to identify high-value customers and Market Basket Analysis (MBA) to discover product combinations frequently purchased together. By focusing on these insights, Fresh2Day can deliver personalized product recommendations, improve cross-selling, optimize inventory, and strengthen customer loyalty. The study helps transform raw sales data into actionable strategies for smarter retail decision-making.

3.2 STATEMENT OF THE PROBLEM

Retailers often struggle to identify their most valuable customers and understand their purchasing behaviour, resulting in missed opportunities for personalized marketing and increased sales. Fresh2Day, despite having rich customer transaction data, lacks a systematic approach to segment customers and recommend relevant products. This study addresses the problem by applying RFM segmentation and Market Basket Analysis (MBA) to uncover key customer segments and product associations, enabling more targeted and effective product recommendations.

3.3 OBJECTIVES OF THE STUDY

1. To segment customers based on RFM (Recency, Frequency, Monetary) values to identify high-value customers.
2. To apply Market Basket Analysis (MBA) using the Apriori algorithm to identify product associations and frequent item combinations.
3. To develop a personalized product recommendation system for high-value customers based on their purchasing patterns.

3.4 RESEARCH DESIGN:

Quantitative Research Design: This research aims to quantitatively analyze customer behavior and sales patterns through the application of the RFM model and Market Basket Analysis (MBA). By collecting transactional data, including customer ID, product categories, and sales amounts, the study seeks to identify purchasing trends and preferences. The focus is on high-value customers, identified through RFM segmentation, to explore frequent itemsets and product associations using the Apriori algorithm.

3.5 DATA COLLECTION:

Secondary data: The data for this study was extracted from Fresh2Day's retail management software, covering transactions from 13 different outlets. It is a customer-based sales dataset, capturing detailed records of purchases across various product categories.

3.6 DATA ANALYSIS:

Descriptive Statistics:

Basic statistical measures such as mean, median, total, and count were calculated for customer spending, purchase frequency, and recency. These metrics provided a general overview of sales trends and customer purchasing patterns.

RFM Segmentation:

The Recency, Frequency, Monetary (RFM) model was applied to categorize customers based on how recently, how often, and how much they purchased. Based on their total RFM scores, customers were segmented into high, medium, and low-value groups to identify top contributors to revenue.

Statistical analysis:

A one-way ANOVA was conducted to determine whether there are statistically significant differences in average spending between the different RFM-based customer segments.

Market Basket Analysis (MBA):

Using the Apriori algorithm, frequent product combinations were identified within the high-value customer group. This helped uncover product pairing trends and revealed strong association rules.

Product Recommendation Modelling:

An RFM-based recommendation system was developed, using the insights from MBA to suggest products that customers haven't purchased yet but are frequently bought together with their previously purchased items.

Visualization:

Visual tools such as bar charts, heatmaps, and network graphs were used to depict customer segmentation, item associations, and sales distribution across categories.

3.6.1 TOOLS USED:

- **Python :**

Used for data cleaning, preprocessing, RFM segmentation, Market Basket Analysis (MBA), and statistical testing such as ANOVA.

- **Pandas & NumPy:**

For data manipulation, summarization, and numerical operations.

- **Microsoft Excel:**

Utilized for initial data exploration, basic filtering, pivot tables, and validating calculated metrics.

- **Tableau:**

Used to build interactive dashboards and visualize customer segmentation and product performance insights.

3.7 SOURCE OF DATA:

The analysis relies on Secondary Data Source.

3.8 DATA COLLECTION PROCEDURE

The data used for this study was extracted directly from the Fresh2Day software, which serves as the Point-of-Sale (POS) and inventory management system for the retail chain. The dataset comprises customer-based sales transactions collected from 13 different retail outlets operated by Fresh2Day.

3.9 DATA DESCRIPTION:

Fresh2day_202531_31_Sales:
Shape (478443, 12)

Columns:

Bill Date: The date on which the transaction occurred.

Cust cid: Unique identifier for each customer.

Outlet Name: Name of the Fresh2Day store where the purchase was made.

MAINCATEGORY: Broad classification of products (e.g., Grocery, Dairy). breakdown under main category (e.g., Rice, Milk).

BRANDS: The brand associated with the purchased product.

CATEGORY: Functional or business grouping of the item (e.g., Fresh Fruits, Bakery & Cakes).

Item Code: Unique code assigned to each item for product-level identification.

Item Name: Name of the product purchased.

Qty: Quantity of the product purchased in the transaction.

Item Rate: Unit price of the item.

Total Amt: Total transaction value (Qty × Item Rate).

3.10 LIMITATIONS OF THE STUDY:

Limited Scope of Data: The analysis is based on historical transaction data from 13 Fresh2Day outlets of 1month and may not reflect customer behaviour across all regions or during different time periods.

Category-Level Analysis: Market Basket Analysis was performed at the category level rather than the individual product level, which may overlook item-specific associations.

Assumption-Based Recommendations: Recommendations are based on item associations (Apriori) and may not consider personal preferences or external influences such as promotions or seasonality.

Exclusion of Low-Value Customers: The study focuses only on high-value customers, which might exclude potential insights from medium or low-value segment.

CHAPTER – 4

DATA ANALYSIS AND INTERPRETATION

4.1 Packages and library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from efficient_apriori import apriori
from mlxtend.frequent_patterns import apriori, association_rules
from mlxtend.preprocessing import TransactionEncoder
from scipy.stats import f_oneway
import datetime
```

Pandas : Used for data manipulation and analysis, especially with structured data like tables (DataFrames).

Numpy : Provides support for large, multi-dimensional arrays and mathematical functions.

Matplotlib : A plotting library used to create static, animated, and interactive visualizations.

Seaborn: A statistical data visualization library built on top of Matplotlib; it makes attractive and informative graphics easier to create.

Eficient_apriori : A fast implementation of the Apriori algorithm for finding frequent itemsets and association rules in transaction data.

Mlxtend.frequent_patterns : Offers tools for mining frequent patterns and generating association rules using Apriori in a more customizable way.

Mlxtend.preprocessing import TransactionEncoder: Converts a list of transactions into a format suitable for association rule mining (i.e., a boolean matrix).

Scipy.stats import f_oneway: Used to perform a one-way ANOVA test, which checks if there are statistically significant differences between the means of three or more groups.

Datetime: Provides functions to manipulate dates and times.

4.2 Data loading

```
[2]: df = pd.read_csv('Fresh2day_202531_31_Sales',skiprows=3)
```

```
/var/folders/c_/8r42hlvx2w36c1j39xf03hkw0000gp/T/ipykernel_20874/3052439813.py:1: DtypeWarning: Columns (19,21,22,23,24,25,26,27,31,35,37,39,
43,44,45,51,52,55,59,68,77,84,85,89,92,94,96) have mixed types. Specify dtype option on import or set low_memory=False.
  df = pd.read_csv('Project data_Raw.csv',skiprows=3)
```

```
[3]: df.head()
```

[3]:

	MAINCATEGORY	SUBCATEGORY	BRANDS	CATEGORY	FAMILY	Bill Date	Outlet Name	Location Name	Order Type	Counter	...	Item Weight in GMs	Price Level Name	Rate Edit Reason	Max Conversion Type	Conv
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
1	Processed Food	Jam / Spreads	NaN	Ready Foods	NaN	2025-03-02	FRESH2DAY-PALAVAKKAM	Main Location	NaN	BAKOF	...	1.0	NaN	NaN	NaN	
2	Processed Food	Jam / Spreads	NaN	Ready Foods	NaN	2025-03-02	FRESH2DAY-PALAVAKKAM	Main Location	NaN	BAKOF	...	1.0	NaN	NaN	NaN	
3	FRUITS & VEGETABLES	Regular Fruits	F2D	FRESH FRUITS	NaN	2025-03-14	FRESH2DAY - KILPAUK	Main Location	NaN	CL4	...	0.0	NaN	NaN	NaN	
4	FRUITS & VEGETABLES	Regular Fruits	F2D	FRESH FRUITS	NaN	2025-03-14	FRESH2DAY - KILPAUK	Main Location	NaN	CL4	...	0.0	NaN	NaN	NaN	

The dataset named Fresh2day_202531_31_Sales was loaded using the pandas library.

4.3 Data cleaning

- Columns in raw data

```
df.columns
```

```
Index(['MAINCATEGORY', 'SUBCATEGORY', 'BRANDS', 'CATEGORY', 'FAMILY',
       'Bill Date', 'Outlet Name', 'Location Name', 'Order Type', 'Counter',
       'PBill No(T)', 'Cust Name', 'Bill Time', 'Outlet ERP Code', 'User Name',
       'Item Code', 'Item Name', 'HSN_SAC Code', 'AliasName', 'Qty',
       'Free Qty', 'Bill Amt(WOT Charges)', 'Item Amt', 'Item Rate',
       'Item Disc Amt', 'Bill Disc Amt', 'SMS_ALERT_NO', 'Total Disc Amt',
       'Tax%', 'Cust cid', 'Cust_Code', 'Tax Amt', 'Item_Master_Base_UOM',
       'Service Tax%', 'Service Tax Amt', 'Amt With Tax', 'CESS %', 'CESS Amt',
       'CGST %', 'CGST Amt', 'IGST %', 'IGST Amt', 'SGST %', 'SGST Amt',
       'Total Amt', 'Gross Margin', 'Void Qty', 'Void Amt', 'I_U',
       'Item Remarks', 'Sales Type', 'I_U Qty', 'GoodsValue', 'Combo Qty',
       'Supplier', 'Total Qty', 'Item Conversion',
       'RateDiff(Ori Grathan Billed)', 'RateDiff(Ori Lesthan Billed)',
       'Piece_Rate_WOTtax', 'Batch No', 'Expiry Dt', 'Tender Type',
       'Item Free Qty in(Weight)KG', 'EAN Code', 'Group Name', 'Type Name',
       'Extra Cess Amount', 'Item Conversion Qty', 'Online_order_number',
       'SO_Channel', 'GST BILL NO', 'Item GST UOM', 'Calamity Cess Amount',
       'Customer GST No', 'Item MRP', 'Gross_Profit%', 'Purchase amount WOT',
       'GST Tax Amt WOT Cess', 'GST Cess Amt', 'Location Type',
       'Customer Address', 'Customer City', 'Freight Charge', 'Landing Cost',
       'Purchase Rate WOT Tax', 'Item Qty in(Weight)KG', 'Item Weight in GMs',
       'Price Level Name', 'Rate Edit Reason', 'Max Conversion Type',
       'Max Conversion Qty', 'Bill Qty with Max Conv', 'Min Conversion Qty',
       'Bill Qty with Min Conv', 'Unique Trans Refno', 'Conversion Rate'],
      dtype='object')
```

- Taking necessary columns for analysis

```
[6]: df_new = df[['Bill Date','Cust cid','Outlet Name','MAINCATEGORY', 'SUBCATEGORY', 'BRANDS', 'CATEGORY','Item Code', 'Item Name','Qty','Item Rat
[7]: df_new.head()
```

[7]:

	Bill Date	Cust cid	Outlet Name	MAINCATEGORY	SUBCATEGORY	BRANDS	CATEGORY	Item Code	Item Name	Qty	Item Rate	Total Amt
1	2025-03-02	1N60006307	FRESH2DAY-PALAVAKKAM	Processed Food	Jam / Spreads	NaN	Ready Foods	38784.0	F2D LEMON PICKLE 300 GM	1.00	120.0	99.00
2	2025-03-02	1N60006307	FRESH2DAY-PALAVAKKAM	Processed Food	Jam / Spreads	NaN	Ready Foods	38793.0	F2D TOMATO PICKLE 300 GM	1.00	120.0	99.00
3	2025-03-14	1N60006307	FRESH2DAY - KILPAUK	FRUITS & VEGETABLES	Regular Fruits	F2D	FRESH FRUITS	156.0	BANANA KARPOORAVALI	0.60	80.0	48.24
4	2025-03-14	1N60006307	FRESH2DAY - KILPAUK	FRUITS & VEGETABLES	Regular Fruits	F2D	FRESH FRUITS	156.0	BANANA KARPOORAVALI	0.09	80.0	6.96
5	2025-03-14	1N60006307	FRESH2DAY - KILPAUK	FRUITS & VEGETABLES	Herbs & Seasoning	F2D	FRESH VEGETABLES	88.0	GARLICK	0.66	169.0	112.22

- Data information

```
[8]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 478443 entries, 1 to 478443
Data columns (total 12 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   Bill Date     478443 non-null  object
 1   Cust cid      460387 non-null  object
 2   Outlet Name   478443 non-null  object
 3   MAINCATEGORY  478443 non-null  object
 4   SUBCATEGORY   478443 non-null  object
 5   BRANDS        449367 non-null  object
 6   CATEGORY      478438 non-null  object
 7   Item Code     478443 non-null  float64
 8   Item Name     478443 non-null  object
 9   Qty           478443 non-null  object
 10  Item Rate     478443 non-null  object
 11  Total Amt     478443 non-null  object
dtypes: float64(1), object(11)
memory usage: 43.8+ MB
```

- Removing null values

```
[11]: df.isna().sum()
```

```
[11]: Bill Date           0
      Cust cid        18056
      Outlet Name         0
      MAINCATEGORY        0
      SUBCATEGORY         0
      BRANDS          29076
      CATEGORY            5
      Item Code           0
      Item Name           0
      Qty                 0
      Item Rate           0
      Total Amt       32768
      dtype: int64
```

```
[12]: df = df.dropna()
```

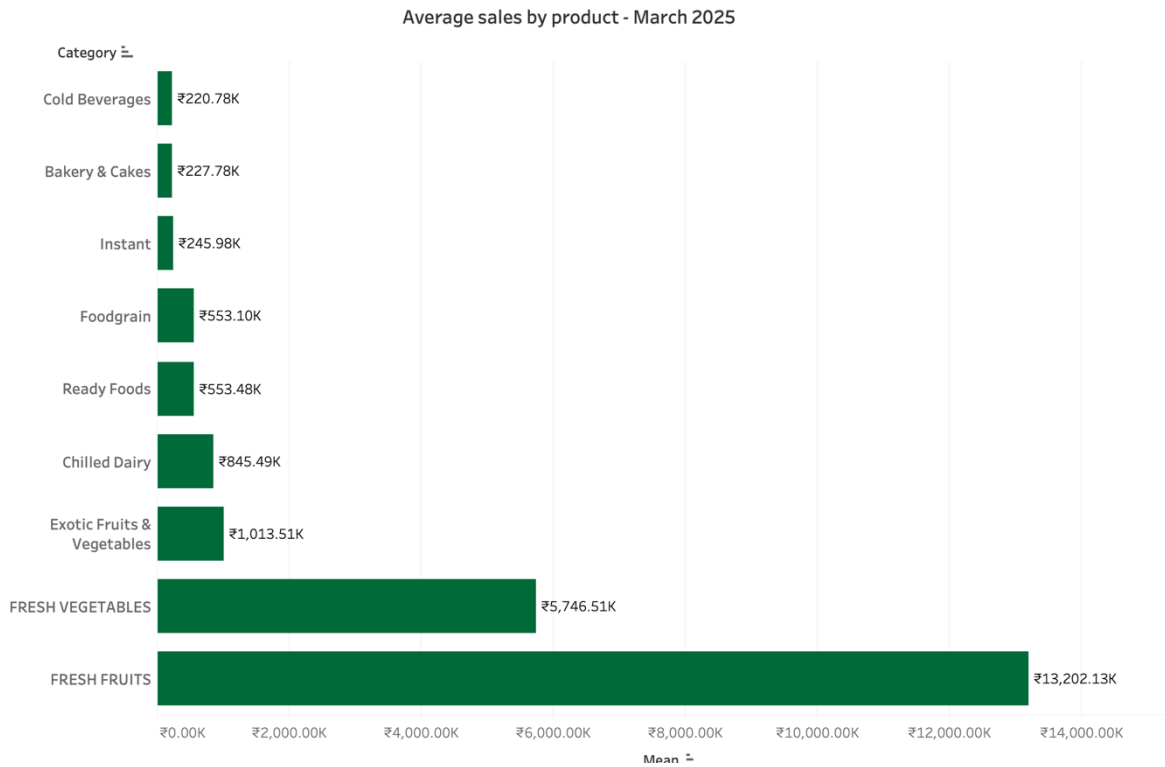
```
[13]: df.isna().sum()
```

```
[13]: Bill Date       0
      Cust cid        0
      Outlet Name     0
      MAINCATEGORY    0
      SUBCATEGORY     0
      BRANDS          0
      CATEGORY        0
      Item Code       0
      Item Name       0
      Qty             0
      Item Rate       0
      Total Amt       0
      dtype: int64
```

4.4 Average sales by Product category

Category	Average Sales
Bakery & Cakes	₹227.78K
Chilled Dairy	₹845.49K
Cold Beverages	₹220.78K
Exotic Fruits & Vegetables	₹1,013.51K
Foodgrain	₹553.10K
FRESH FRUITS	₹13,202.13K
FRESH VEGETABLES	₹5,746.51K
Instant	₹245.98K
Ready Foods	₹553.48K

Table – 4.1



Graph – 4.1

Fresh2Day's average sales were highest in its core categories—Fresh Fruits (₹13.2K) and Fresh Vegetables (₹5.7K)—highlighting strong alignment with its brand focus. Exotic Fruits & Vegetables also performed moderately well, while other categories like Cold Beverages, Bakery, and Instant Foods contributed minimally, indicating they serve more as complementary products.

4.5 RFM Segmentation

[16]:
```
snapshot_date = df['Bill Date'].max() + pd.Timedelta(days=1)
```

[17]:
```
rfm = df.groupby('Cust cid').agg({
    'Bill Date': lambda x: (snapshot_date - x.max()).days,
    'Item Code': 'count',
    'Total Amt': 'sum'
})
```

[18]:
```
rfm.columns = ['Recency', 'Frequency', 'Monetary']

# RFM Scoring (1–5 scale)
rfm['R_Score'] = pd.qcut(rfm['Recency'], 5, labels=[5, 4, 3, 2, 1])
rfm['F_Score'] = pd.qcut(rfm['Frequency'].rank(method='first'), 5, labels=[1, 2, 3, 4, 5])
rfm['M_Score'] = pd.qcut(rfm['Monetary'], 5, labels=[1, 2, 3, 4, 5])

rfm['RFM_Segment'] = rfm['R_Score'].astype(str) + rfm['F_Score'].astype(str) + rfm['M_Score'].astype(str)
rfm['RFM_Score'] = rfm[['R_Score', 'F_Score', 'M_Score']].astype(int).sum(axis=1)
```

[19]:
```
rfm['RFM_Score'] = rfm[['R_Score', 'F_Score', 'M_Score']].astype(int).sum(axis=1)
```

[20]:
```
def segment_customer(score):
    if score >= 10:
        return 'High Value'
    elif score >= 5:
        return 'Medium Value'
    else:
        return 'Low Value'

rfm['Segment'] = rfm['RFM_Score'].apply(segment_customer)
```

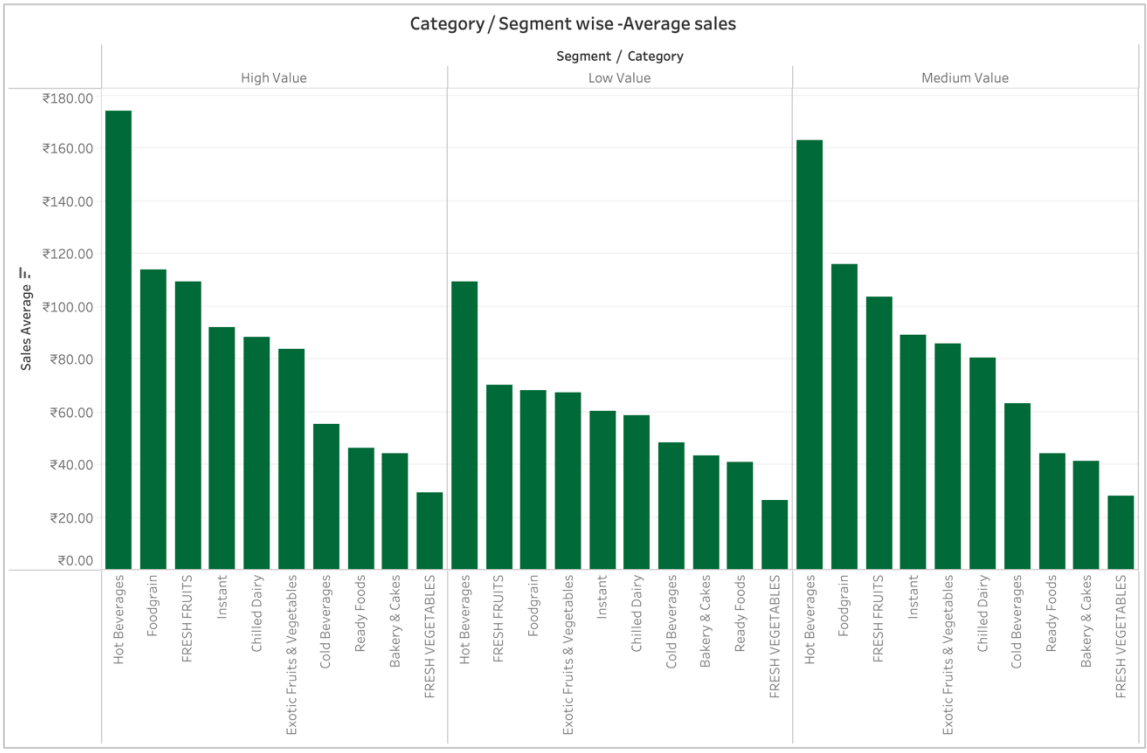
Cust cid	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_Segment	RFM_Score	Segment
11N10000	16	21	505.63	2	5	4	254	11	High Value
11N1001	24	1	60.00	1	1	1	111	3	Low Value
11N10012	31	2	215.33	1	1	2	112	4	Low Value
11N1002	5	5	121.80	4	3	2	432	9	Medium Value
11N10030	10	3	602.07	3	2	4	324	9	Medium Value

In this RFM segmentation process, customers were evaluated based on their Recency (days since last purchase), Frequency (number of transactions), and Monetary value (total amount spent). Each metric was scored on a 1–5 scale, and the combined RFM Score (ranging from 3 to 15) was used to classify customers into segments: Low Value (score ≤ 5), Medium Value (score > 5 and ≤ 10), and High Value (score > 10). This approach helps identify and target customers based on their engagement and spending behavior

4.6 Customer segment and Category wise Average sales

	Segment		
Category	High Value	Low Value	Medium Value
Hot Beverages	₹174.06	₹109.42	₹162.81
Foodgrain	₹113.81	₹68.05	₹115.90
FRESH FRUITS	₹109.28	₹70.11	₹103.52
Instant	₹91.88	₹60.43	₹89.02
Chilled Dairy	₹88.39	₹58.64	₹80.31
Exotic Fruits & Vegetables	₹83.79	₹67.12	₹85.88
Cold Beverages	₹55.11	₹48.17	₹63.25
Ready Foods	₹46.17	₹40.82	₹44.18
Bakery & Cakes	₹44.13	₹43.26	₹41.20
FRESH VEGETABLES	₹29.31	₹26.29	₹28.05

Table – 4.2



Graph – 4.2

High, Medium, and Low—Hot Beverages**,** Foodgrain, and Fresh Fruits have the highest average sales, likely due to higher price points. In contrast, Fresh Vegetables, despite being a core product for Fresh2Day, show lower average sales, possibly because of their lower unit prices. This suggests that while some categories generate high average bills, volume-driven categories like vegetables may be underrepresented in mean-based comparisons.

4.7 One-way ANOVA – Monetary and segments

[30]:
```
# Grouping the data
high = rfm[rfm['Segment'] == 'High Value']['Monetary']
medium = rfm[rfm['Segment'] == 'Medium Value']['Monetary']
low = rfm[rfm['Segment'] == 'Low Value']['Monetary']

# One-way ANOVA
f_stat, p_value = f_oneway(high, medium, low)

print("F-statistic:", f_stat)
print("p-value:", p_value)
```

```
F-statistic: 35.530649271366194
p-value: 3.86761285877884e-16
```

The one-way ANOVA test was conducted to determine if there is a significant difference in monetary values among the High, Medium, and Low Value customer segments. The results show an F-statistic of 35.53 and a p-value of approximately 3.87e-16, which is far below the typical significance level of 0.05.

This indicates a statistically significant difference in average monetary values between at least two of the customer segments, confirming that the RFM-based segmentation effectively distinguishes customers based on their spending behaviour.

4.8 Market basket analysis

```
[22]: # Prepare data: list of transactions
      transactions = df.groupby('Cust cid')['CATEGORY'].apply(list).tolist()
```

```
[23]: te = TransactionEncoder()
      te_data = te.fit_transform(transactions)
      df_te = pd.DataFrame(te_data, columns=te.columns_)
```

```
[24]: # Apriori and Rules
      itemsets = apriori(df_te, min_support=0.01 , use_colnames=True)
      rules = association_rules(itemsets, metric='confidence', min_threshold=0.5)
```

```
top_rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']]
```

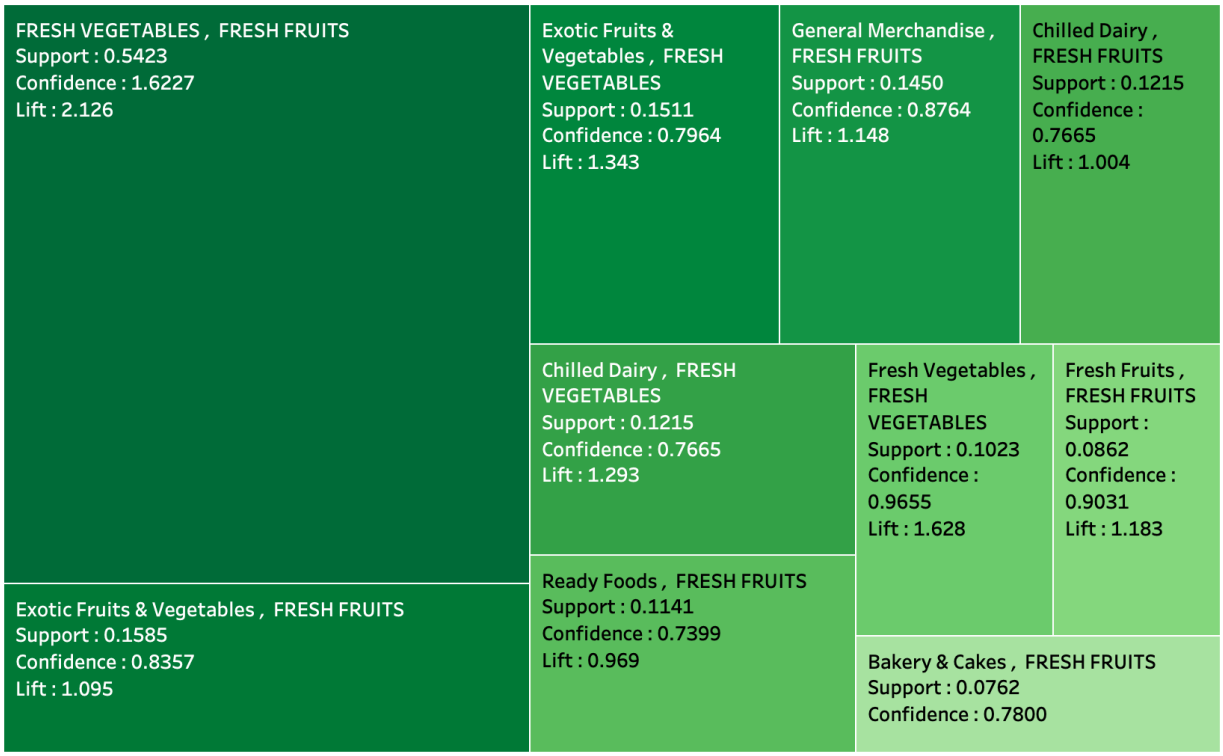
	antecedents	consequents	support	confidence	lift
220	(Fresh Vegetables, Masala)	(FRESH VEGETABLES)	0.014401	1.000000	1.686694
417	(Fresh Vegetables, Masala, FRESH FRUITS)	(FRESH VEGETABLES)	0.013570	1.000000	1.686694
381	(Exotic Fruits & Vegetables, Fresh Vegetables,...	(FRESH VEGETABLES)	0.011175	0.997033	1.681689
499	(Exotic Fruits & Vegetables, Fresh Vegetables,...	(FRESH VEGETABLES)	0.010676	0.996894	1.681455
372	(Foodgrain, Exotic Fruits & Vegetables, Fresh ...	(FRESH VEGETABLES)	0.010244	0.996764	1.681235
322	(General Merchandise, Chilled Dairy, Fresh Veg...	(FRESH VEGETABLES)	0.010211	0.996753	1.681217
387	(Foodgrain, Fresh Vegetables, FRESH FRUITS)	(FRESH VEGETABLES)	0.018326	0.996383	1.680593
286	(Exotic Fruits & Vegetables, Fresh Vegetables,...	(FRESH VEGETABLES)	0.016796	0.996055	1.680040
250	(FRESH FRUITS, Bakery & Cakes, Fresh Vegetables)	(FRESH VEGETABLES)	0.016131	0.995893	1.679767
467	(Exotic Fruits & Vegetables, Fresh Vegetables,...	(FRESH VEGETABLES)	0.015898	0.995833	1.679666

In the Market Basket Analysis, transaction data was grouped by customers and transformed into a format suitable for mining frequent item sets using the Apriori algorithm. Association rules were then generated based on these item sets to identify strong product combinations that customers tend to purchase together, using metrics like support, confidence, and lift. This helps in understanding purchasing patterns and can be used to enhance product recommendations.

4.9 Item sets by Association rule

Itemset	Confidence	Lift	Support
Bakery & Cakes , FRESH FRUITS	0.780	1.022	0.076
Fresh Fruits , FRESH FRUITS	0.903	1.183	0.086
Fresh Vegetables , FRESH VEGETABLES	0.965	1.628	0.102
Ready Foods , FRESH FRUITS	0.740	0.969	0.114
Chilled Dairy , FRESH FRUITS	0.767	1.004	0.122
Chilled Dairy , FRESH VEGETABLES	0.767	1.293	0.122
General Merchandise , FRESH FRUITS	0.876	1.148	0.145
Exotic Fruits & Vegetables , FRESH VEGETABLES	0.796	1.343	0.151
Exotic Fruits & Vegetables , FRESH FRUITS	0.836	1.095	0.159
FRESH VEGETABLES , FRESH FRUITS	1.623	2.126	0.542

Table – 4.3



Graph – 4.3

This treemap visualizes frequently co-purchased item sets based on Market Basket Analysis. Each block represents a product combination, with its size indicating support (how often the combination occurs) From the chart, "Fresh Vegetables & Fresh Fruits" stands out as the most frequent and strongly associated pair, with the highest support and a strong lift of 2.126, indicating that customers who buy one are highly likely to buy the other.

4.10 Rule-based recommender system

[40]:
```
def get_recommendations(user_products, rules_df):
    recommended = set()
    for _, row in rules_1.iterrows():
        antecedent = set(row['antecedents'])
        consequent = set(row['consequents'])
        if antecedent.issubset(user_products):
            recommended.update(consequent)
    # Remove any product the user already bought
    recommended = recommended - user_products
    return list(recommended)
```

[42]:
```
user_products = set(df[df['Cust cid'] == '11N10000']['CATEGORY'])
recommendations = get_recommendations(user_products, rules)
print("Recommended products:", recommendations)
```

```
Recommended products: ['FRESH FRUITS']
```

Rule-based recommendation system uses Market Basket Analysis results (from the Apriori algorithm) to suggest relevant product categories to customers. It compares a customer's past purchases (in this case, product categories) with the association rules generated from customer data. If the items a customer bought match the antecedent (left-hand side) of a rule, the system recommends the corresponding consequent (right-hand side) items—excluding anything the customer has already purchased. For example, if a customer has bought certain categories and those match a known rule, it recommends associated categories like 'FRESH FRUITS' that are commonly bought together.

CHAPTER – 5

FINDINGS, SUGGESTIONS AND CONCLUSION

5.1 FINDINGS

- The RFM model was used to classify customers into High, Medium, and Low value groups based on purchase behaviour. This helped identify loyal, high-spending customers. Targeting these segments allows for more effective marketing strategies.
- A one-way ANOVA test showed statistically significant differences in spending across the segments. This confirms the RFM method accurately separates customers by monetary value. It supports data-driven decision-making for customer targeting.
- Market Basket Analysis identified frequent item combinations, with "Fresh Fruits" and "Fresh Vegetables" being the strongest pair. These insights help in planning product bundles and improving in-store or online layout. They also support effective cross-selling.
- A recommendation engine was developed using association rules from high-value customers' purchases. It suggests products customers haven't bought but are likely to need. This boosts personalization and encourages repeat purchases.

5.2 SUGGESTIONS

- Since high-value customers contribute significantly to revenue, Fresh2Day should implement loyalty programs, exclusive offers, or early access promotions for this segment. Personalized engagement will help retain these valuable customers and increase their lifetime value.

- Market Basket Analysis revealed strong associations, especially between Fresh Fruits and Vegetables. Fresh2Day can introduce combo deals or curated baskets featuring these pairs, which can improve cross-selling and average bill value.

- Insights from frequent item pairings should guide store layout (placing related items close) and stock prioritization. This can streamline customer experience and reduce missed sales due to out-of-stock issues on popular pairs.

- The current rule-based recommendation system should be integrated into Fresh2Day's mobile app or website. Real-time, personalized product suggestions can enhance user engagement and drive online conversions.

5.3 CONCLUSION

This project demonstrates a practical application of data analytics in the retail industry by leveraging customer transaction data to derive meaningful insights. Through RFM segmentation, it categorizes customers based on their purchasing behaviour, helping retailers like Fresh2Day identify and prioritize high-value segments. The use of Market Basket Analysis further supports strategic decisions around product bundling and cross-selling by identifying commonly purchased product combinations.

In the context of retail, such data-driven methods are essential for personalizing the shopping experience, optimizing inventory, and improving marketing effectiveness. As the industry increasingly shifts towards digital platforms and personalized services, the models used in this project provide a strong foundation for future innovations. Going forward, integrating real-time data, expanding the analysis to include all customer segments, and moving towards product-level recommendations can make these systems even more effective. This approach not only enhances customer engagement but also supports long-term growth and competitiveness in the evolving retail landscape.

Reference

Websites:

https://www.fresh2day.com/

https://www.assosia.com/sectors-channels/retail

https://www.geeksforgeeks.org/apriori-algorithm/

https://www.uniphore.com/glossary/rfm-analysis/