**Apu Datta**

Data Mining and Visualization

**Project: 2, Question 3: Customer Segmentation and Market Basket Analysis - Business Report**

**Executive Summary:**

This project focuses on understanding customer behavior and purchasing patterns using the Sample Superstore dataset. By combining customer segmentation with market basket analysis, the goal is to help decision makers identify distinct customer groups and uncover product combinations that drive sales. The findings enable more targeted marketing, optimized inventory management, and improved sales strategies.

**Business Problem and Objectives:**

The Superstore dataset contains transactional data with details such as customer demographics, product categories, sales amounts, and order dates. However, without analytical processing, these records do not directly translate into strategic insights. The objective of this project is to segment customers into meaningful groups based on their purchase behavior and to discover associations between products that can inform cross selling and upselling strategies.

**Data Summary:**

The dataset includes attributes such as Customer ID, Order Date, Ship Date, Product Category, Sub-Category, Sales, Quantity, and Profit. Additional features were engineered such as Recency, Frequency, and Monetary value to support RFM analysis. The dataset also underwent preprocessing to handle missing values, remove duplicates, and format dates for time-based analysis.

**Methodology:**

The analysis began with customer segmentation using K Means clustering applied to RFM features. Optimal cluster numbers were determined using the Elbow method, Silhouette score, and Davies Bouldin index. Each cluster was profiled to create customer personas. For market basket analysis, the Apriori algorithm was applied to identify frequent item sets and generate association rules based on support, confidence, and lift. These results reveal which products are commonly purchased together.

**Key Findings**

The segmentation revealed distinct customer groups including high value repeat buyers, moderate spenders, and occasional shoppers. High value customers displayed a strong preference for technology and office supplies. The market basket analysis showed strong product pairings such as printers with ink cartridges and office chairs with desks. These associations can be used to design targeted promotions.

**Visual Insights:**

Several visualizations support the findings, including customer cluster plots, comparisons of different clustering methods, summaries of association rules, and evaluation metrics for cluster quality. These visuals make it easier to communicate insights to business stakeholders.

**Business Recommendations:**

Based on the analysis, it is recommended to focus marketing campaigns on high value customer segments with tailored offers. Bundling products that frequently appear together in association rules can encourage larger purchases. Moderate and occasional shoppers could be targeted with loyalty programs or seasonal discounts to increase engagement.

**Limitations and Next Steps:**

This analysis is limited to historical data from a single dataset and does not account for external market factors. Future work could integrate demographic data, web browsing behavior, and social media interactions to enrich customer profiles. Predictive modeling could also be applied to forecast customer lifetime value and churn probability.

**Data set and deployment**

**Dataset\_kaggle:**<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final/suggestions>

**Cloud\_Streamlit:**

Github\_repository:

**Appendix:**