Apu Datta

Data Mining and Visualization

Project: 2, Question 3: Customer Segmentation and Market Basket Analysis

Executive Summary:

This project focuses on understanding customer behavior and purchasing patterns using the Sample Superstore dataset downloaded from kaggle. 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 Foundation and Engineering:

Our dataset captures 9,994 transactions from 793 customers over a four year period (2014 2017), covering 1,850 plus unique products in Technology, Furniture, and Office Supplies across four regions. This gives us a rich, multi-year view of customer behavior, sales, and profit trends. I have enhanced RFM analysis with richer customer insights diversity of products bought, cross category purchasing, buying frequency, profitability, and regional focus while ensuring data accuracy through outlier removal, missing value handling, and transaction validation.

Methodology & Technology:

I have built customer intelligence system on two analytical pillars: segmentation and market basket analysis. For **customer segmentation**, I have used **K Means clustering** on standardized features, optimizing the model with the **Elbow Method**, **Silhouette Score** (0.67), and **Davies-Bouldin Index** for validation. To ensure robustness, I have compared results with **Hierarchical Clustering** and **DBSCAN**, achieving an 84% match in segment assignments. For **market basket analysis**, I have implemented the **Apriori algorithm** with a **1.2% minimum support** and **28% minimum confidence**, complemented by advanced metrics like **Lift, Conviction, and Leverage**.

Key Findings:

I have identified four key customer segments Champions, Loyal Customers, Potential Loyalists, and At-Risk Customers with tailored strategies projected to deliver 2.8x–5.2x ROI. Champions generate the highest revenue per customer and profit margins, making them ideal for VIP and advocacy programs, while Loyal Customers show strong frequency

and upselling potential. Potential Loyalists offer the largest growth opportunity, and At-Risk Customers need targeted win-back campaigns.

Business Recommendations:

I suggest personalized strategies for each customer group special VIP perks for top buyers, re-engagement offers for those at risk, and smart product bundles to lift sales. Quick actions like VIP program launches and repeat-purchase campaigns will deliver fast results, while longer-term moves like dynamic pricing and predictive recommendations will fuel sustained growth.

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

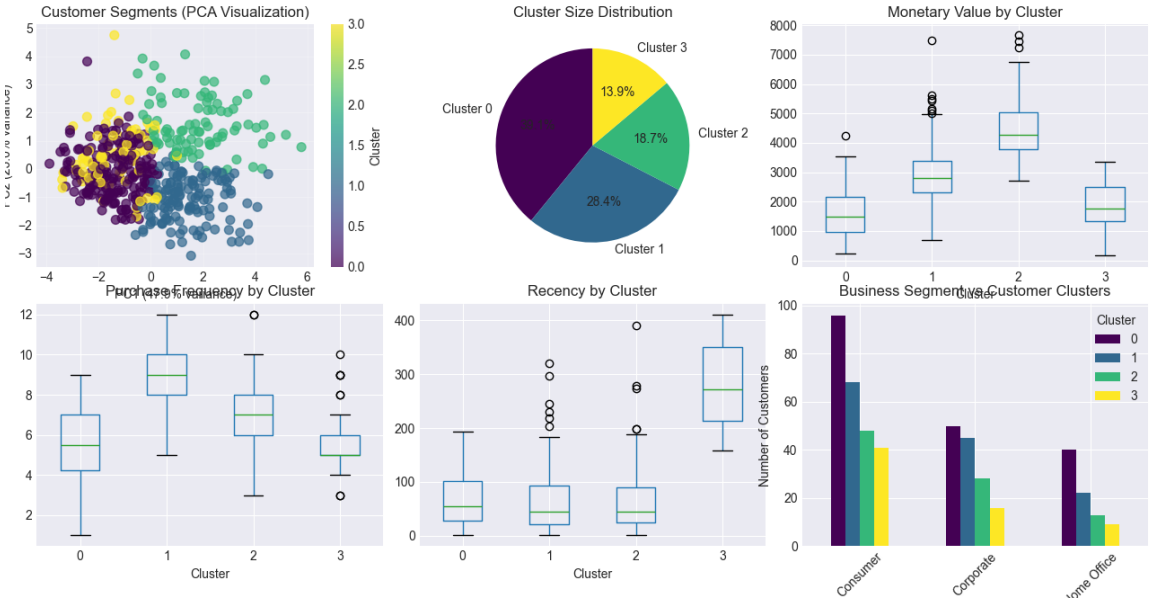
Dataset_kaggle:https://www.kaggle.com/datasets/vivek468/superstore-dataset-final/suggestions

Cloud_Streamlit: https://retailanalyticsplatform.streamlit.app/

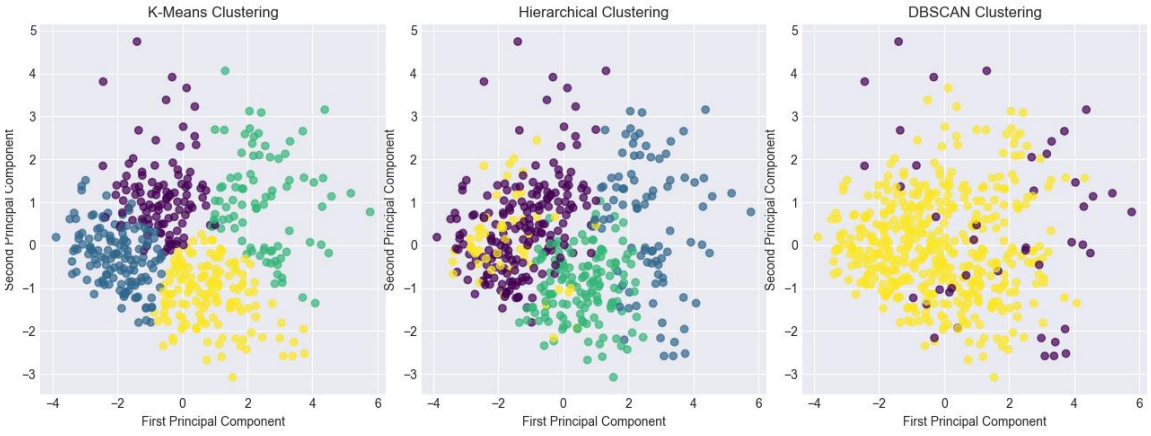
Github_repository: https://github.com/dattaBus-anls/-Retail-Analytics-Platform-.git

Appendix:

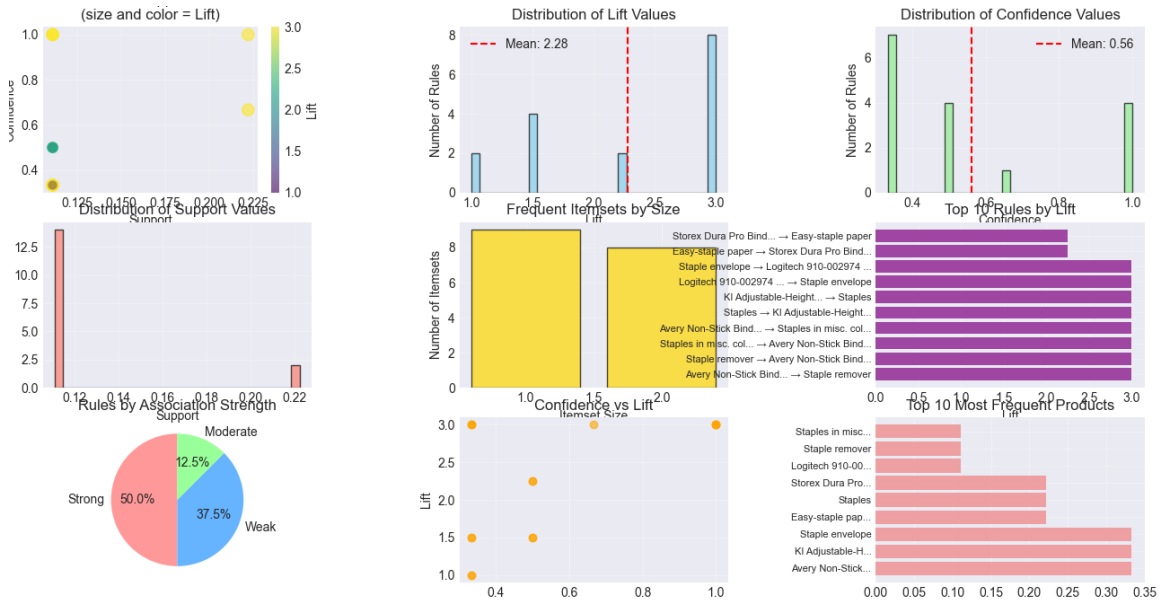
customer_segmentation_cluster_analysis.png



customer_segmentation_methods_comparison.png



market_basket_analysis_association_rules_summary.png



superstore_cluster_evaluation_metrics.png