Customer Segmentation Analysis Report

Objective: The goal was to segment customers based on transaction behaviors and product preferences.

Data Used: `Customers.csv`, `Products.csv`, and `Transactions.csv`.

Data Preprocessing

- **Data Cleaning**: Converted datetime columns, dropped redundant columns, separated product/brand names.
- **Feature Engineering**: Focused on RFM (Recency, Frequency, Monetary) values, AOV (Average Order Value), CLV (Customer Lifetime Value), product/brand preferences, and regional information.

Datasets Tested

- **RFM Dataset**: Included RFM values, AOV, CLV, product/brand preferences.
    - **Results**: Poor clustering performance, silhouette score: 0 to 0.15, DBI: 2.4.
- **RFM2 Dataset**: Key features like R, F, M, AOV, and CLV with additional features like AvgQuantity, Tenure.
    - **Results**: DBI improved to 1.016, silhouette score: 0.23 to 0.27.
    - **Best Features**: Removal of AvgQuantity improved results.
- **RFM3 Dataset**: Added "brand_weights" and "region_rank" for regional/product-specific information.
    - **Results**: Better than RFM but worse than RFM2, suggesting noise from additional categorical features.

Clustering Algorithm

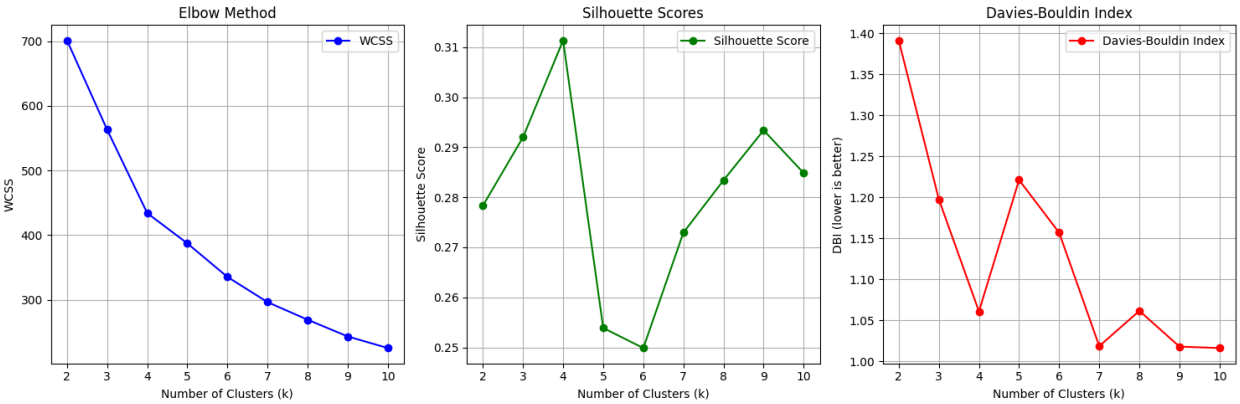
- **K-Means**: Selected RFM2 dataset, used **Elbow Method** and **Silhouette Score** to find 7 clusters.
    - **Results**: DBI = 1.016, silhouette score = 0.278, WCSS = 300.
- **DBSCAN**: Poor performance due to noise, many outliers.

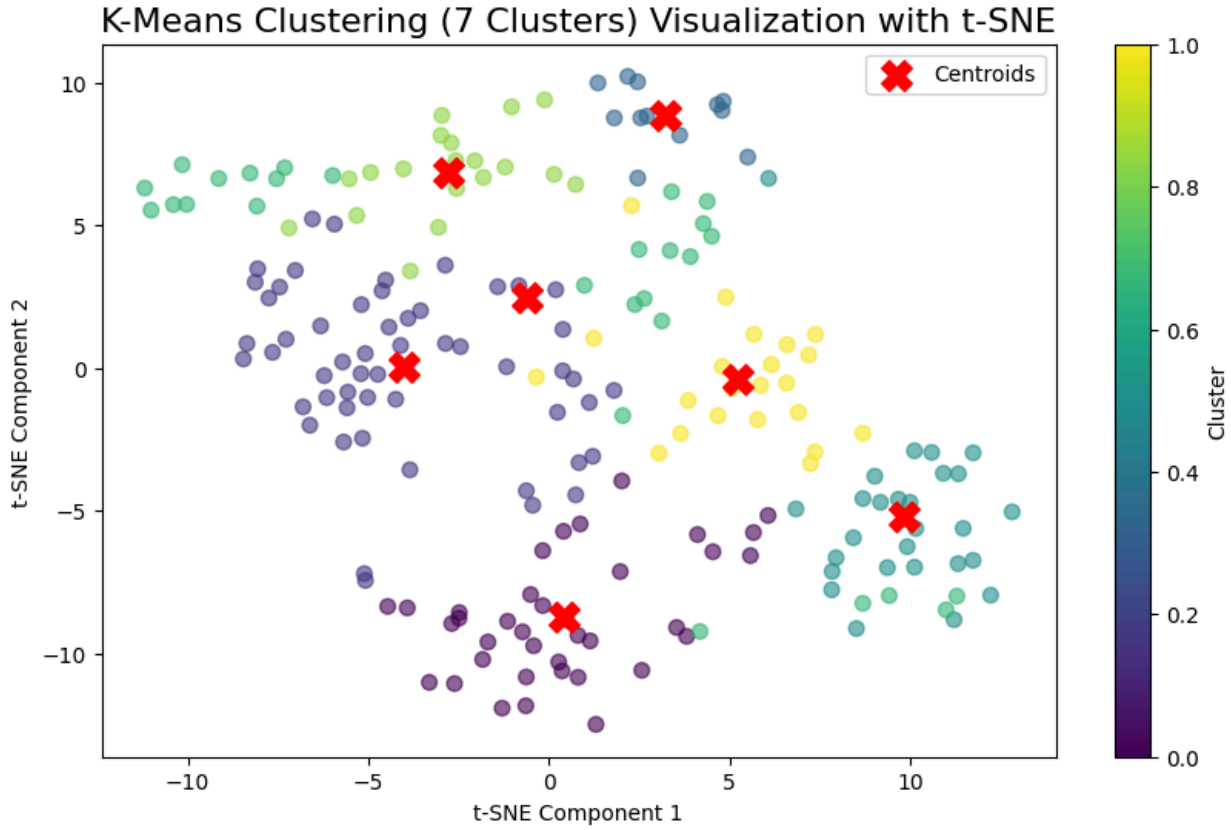
Visualization

- Used **PCA** and **t-SNE** for dimensionality reduction.
    - **Best Visualization**: t-SNE showed better cluster separation.

Conclusion

- **Best Model**: K-Means with 7 clusters on **RFM2 dataset**.
- **Performance**: DBI = 1.016, silhouette score = 0.278.
- **Best model performance metrics**-


Elbow Method
WCSS
Number of Clusters (k)
Silhouette Scores
Silhouette Score
Number of Clusters (k)
Davies-Bouldin Index
DBI (lower is better)
Number of Clusters (k)


K-Means Clustering (7 Clusters) Visualization with t-SNE
Centroids
t-SNE Component 1
t-SNE Component 2
Cluster