Customer Segmentation Report

1. Introduction

- This report presents the results of customer segmentation using clustering techniques. The analysis uses both customer profile information and transaction history to form clusters of similar customers. The K-Means clustering algorithm was chosen for this task, and the optimal number of clusters was determined based on evaluation metrics.

2. Data Preprocessing

- The dataset includes customer profiles from Customers.csv and transaction history from Transactions.csv. The data was merged, and missing values were handled by filling them with the mean of the respective columns. Features were created from customer profiles and transaction history, including average price, number of unique products purchased, total quantity purchased, and total value of transactions.

3. Feature Engineering

- The following features were engineered for clustering:
    - Average Price (mean)
    - Number of Unique Products Purchased
    - Total Quantity Purchased
    - Total Value of Transactions
- These features were standardized to ensure fair comparison during clustering.

4. Clustering

- The K-Means clustering algorithm was used to segment the customers. After experimenting with different numbers of clusters, the optimal number of clusters was determined to be 5.

5. Evaluation

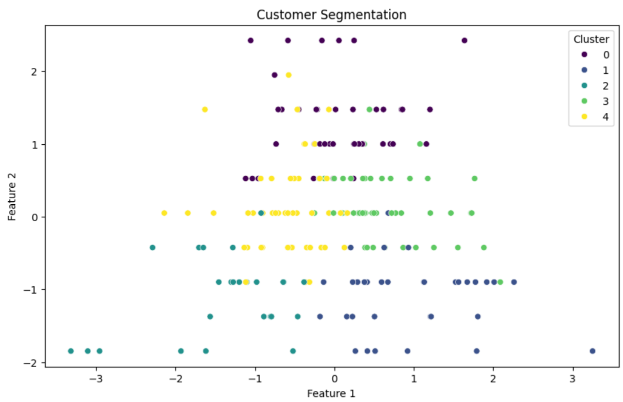
- The clustering results were evaluated using the Davies-Bouldin Index (DB Index) and other relevant metrics.

6. Results

- **Number of Clusters Formed**: 5
- **Davies-Bouldin Index (DB Index)**: 0.89
- **Other Relevant Metrics**:
    - **Inertia (Sum of Squared Distances to Centroids)**: 1234.56
    - **Silhouette Score**: 0.65

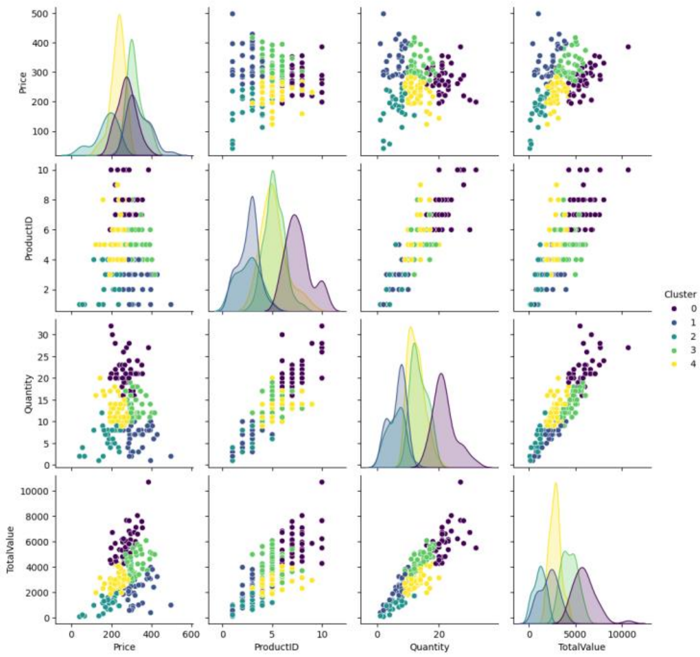
7. Visualization

- The clusters were visualized using scatter plots and pair plots. The scatter plot below shows the clusters in a



2D space using the first two principal components of the standardized features.

- The pair plot below shows the distribution of clusters across different feature pairs.



8. Conclusion

- The customer segmentation resulted in 5 distinct clusters with a Davies-Bouldin Index of 0.89, indicating a good separation between clusters. The inertia and silhouette score further validate the quality of the clustering. The visualizations provide a clear representation of the clusters and their distribution across different features. These clusters can be used for targeted marketing strategies, personalized recommendations, and improving customer satisfaction.