Customer Segmentation Report

1. Introduction

Customer segmentation is an essential task in data analysis that allows businesses to group customers based on shared characteristics. This segmentation helps companies better understand their customers and tailor marketing strategies.

This report presents the results of customer segmentation using the K-means clustering algorithm, based on customer profile information and transaction data.

2. Data Overview

The data was sourced from two CSV files:

- **Customers.csv**: Contains customer profile information, including region.
- **Transactions.csv**: Contains transaction data, including the total value and quantity of each purchase.

These two datasets were merged on the CustomerID column.

3. Data Preprocessing

- The transaction data was aggregated by customer, summing up the TotalValue and Quantity for each customer.
- Categorical features like Region were encoded using one-hot encoding.
- Numerical features (TotalValue, Quantity) were standardized using StandardScaler for clustering purposes.

4. Clustering Results

4.1 Clustering Algorithm

The **K-means clustering** algorithm was used to segment customers into clusters. The algorithm was run for cluster values between 2 and 10, and the **Davies-Bouldin Index (DB Index)** was calculated for each clustering result. The optimal number of clusters was determined by selecting the clustering with the lowest DB Index.

4.2 Optimal Number of Clusters

The optimal number of clusters based on the Davies-Bouldin Index was found to be **X clusters**.

4.3 DB Index

The DB Index for the optimal clustering was **Y**.

5. Cluster Visualization

The clusters were visualized using a scatter plot, with the TotalValue and Quantity (both scaled) as the axes. Each point represents a customer, and the color indicates the cluster they belong to.

(Include your generated cluster visualization image here)

6. Conclusion

This segmentation provides valuable insights into customer behavior, helping businesses understand purchasing patterns. Further analysis could include profiling each cluster to derive actionable insights (e.g., high-value customers, frequent buyers, etc.).

7. Clustering Metrics

- **Number of Clusters**: X
- **Davies-Bouldin Index**: Y
- Other metrics such as Silhouette Score (if computed) could further enrich the analysis.