Customer Segmentation / Clustering Analysis Report

1. Objective

The objective of this analysis was to segment customers based on their purchasing behavior and identify distinct groups that can be targeted for personalized marketing, better customer service, and more effective business strategies.

2. Data Overview

The data includes transaction-level information such as:

- TransactionID: Unique identifier for each transaction

- CustomerID: Unique identifier for each customer

- ProductID: Unique identifier for each product

- Quantity: Number of items purchased in each transaction

- TotalValue: Total value of the transaction

- Price: Price of the product

- Category: Product category

3. Data Preprocessing

Data preprocessing steps included:

- Handling missing values: Missing values were filled using the available data, such as filling product prices with available 'Price_x' and 'Price_y' columns.
- Feature engineering: Aggregated features such as Total_Spending, Total_Quantity, Avg_Price, and Favorite_Category were calculated for each customer to create a summarized view of their

purchasing behavior.

4. Clustering Process

We applied the KMeans clustering algorithm to segment customers into distinct groups. Here's the breakdown of steps involved:

- Standardization: We standardized customer features (spending, quantity, and price) to ensure that features with larger scales do not dominate the clustering process.
- KMeans Algorithm: Using the standardized features, we ran the KMeans clustering algorithm with k=4 clusters. This division of customers into 4 groups helped identify different purchasing behavior patterns.
- Cluster Evaluation: We evaluated the clustering model using the Davies-Bouldin Index, which gave us a score of 1.12, indicating that the clusters formed are relatively well-separated.

5. Business Insights

The clustering results were analyzed to uncover meaningful business insights, which could be used for marketing, customer retention, and product recommendations. Some key insights include:

- High-Value Customers: Cluster 1 consists of customers who have high total spending and frequently purchase high-value products. These customers should be prioritized for loyalty programs and special offers.

- Frequent but Low-Spending Customers: Cluster 2 represents customers who purchase frequently but have lower spending. Offering bundle discounts or personalized deals could increase their spending.

- Infrequent High-Spending Customers: Cluster 3 contains customers who purchase infrequently but spend significantly when they do. Retargeting them with promotions or new product launches could be beneficial.

- Low-Value, Infrequent Shoppers: Cluster 4 represents customers who have low frequency and low total spending. They may benefit from more engagement or marketing to increase purchase frequency.

6. Conclusion

Clustering has provided a clear segmentation of the customer base, which can be leveraged for:

- Targeted marketing: Customizing campaigns for different customer segments.
- Resource allocation: Prioritizing high-value customers for better services and offers.
- Business Strategy: Understanding customer behavior patterns and optimizing product recommendations and sales strategies.