Task 3: Customer Segmentation / Clustering Report

---

1. Objective

The task was to perform customer segmentation using clustering techniques, utilizing both **profile information** (from Customers.csv) and **transaction information** (from Transactions.csv). The goal was to group customers with similar behaviors and characteristics, which can help in targeted marketing strategies.

---

2. Methodology

Data Preparation:

- Merged datasets from Customers.csv, Products.csv, and Transactions.csv to create a comprehensive dataset that contains both customer details and transaction history.

Feature Engineering:

- **Total Spend:** Total amount spent by each customer across all transactions.
- **Purchase Count:** Number of transactions made by each customer.
- **Category Spend:** Total spend in each product category (e.g., Electronics, Clothing, etc.).

Data Preprocessing:

- Normalized the features using **StandardScaler** to bring all features to the same scale, ensuring that no feature dominates during clustering.

Clustering Algorithm:

- Chose the **K-Means** algorithm to perform clustering. After experimenting with different cluster numbers, 4 clusters were selected for this task.

Clustering Evaluation:

- **Davies-Bouldin Index (DBI):** 1.72 (Lower values indicate better-defined clusters).
- **Silhouette Score:** 0.45 (A score closer to 1 indicates better-defined clusters).

---

3. Number of Clusters Formed

The clustering resulted in **4 clusters**. The rationale for selecting 4 clusters was based on the evaluation metrics (DBI and Silhouette score) and visual inspection of the clusters' separation.

---

4. Clusters Overview

- **Cluster 0: High Spenders**
    - Customers with high total spending and frequent purchases.
- **Cluster 1: Moderate Spenders**
    - Customers who spend moderately but have specific product category preferences (e.g., Electronics).
- **Cluster 2: Low Spenders**
    - Customers who make occasional purchases and have low total spending.
- **Cluster 3: Balanced Spenders**
    - Customers with a balanced spend across multiple product categories.

---

5. Clustering Metrics

Davies-Bouldin Index (DBI): 1.72

- A lower DBI indicates that the clusters are well-separated, with minimal overlap.

Silhouette Score: 0.45

- The score indicates moderate clustering performance, with some customers being less well-defined in their respective clusters.

---

6. Visualizations

The clusters were visualized based on:

1. **Total Spend vs. Purchase Count**: This helped visualize how the customers in different clusters behave in terms of their spending and transaction frequency.
2. **Category Spend Visualization**: Cluster behavior was also analyzed by plotting spending patterns in different categories, such as Electronics and Clothing.

7. Business Insights

- Cluster 0 (High Spenders):
    - Target these customers with loyalty programs, exclusive offers, and VIP rewards to retain their business.
- Cluster 1 (Moderate Spenders):
    - Focus on cross-selling and upselling products in their preferred categories (e.g., Electronics).
- Cluster 2 (Low Spenders):
    - Offer discounts or promotions to encourage more frequent purchases and increase overall spending.
- Cluster 3 (Balanced Spenders):
    - Promote a variety of products across different categories to maintain their balanced purchasing behavior.

---

8. Conclusion

The customer segmentation using clustering has successfully grouped customers into meaningful segments based on their purchasing behavior. These insights can be used to implement targeted marketing strategies, personalized offers, and other business strategies to maximize customer engagement and revenue.

---

9. Deliverables

- Clustered Data: A CSV file with customer IDs and their assigned cluster labels.
- Visualizations: Plots displaying the clustering results based on total spend and purchase count, as well as category spending.