**Customer Segmentation Report**

**Objective**

The goal is to perform customer segmentation using both profile information (Customers.csv) and transaction data (Transactions.csv) to group customers into meaningful clusters. The evaluation focuses on the Davies-Bouldin (DB) Index.

**Data Used**

1. **Customers.csv**:
   * Contains customer demographic and profile information, such as age, income, gender, etc.
2. **Transactions.csv**:
   * Contains customer transaction history, such as products purchased, amounts spent, and frequency.

**Methodology**

**1. Data Preprocessing**

* **Data Cleaning**:
  + Missing values in Customers.csv were imputed using median or mode.
  + Missing values in Transactions.csv were replaced with zeros for non-purchasing customers.
* **Feature Engineering**:
  + From Transactions.csv, aggregated features were created:
    - Total Spending (total spent).
    - Number of Purchases (Num purchases).
    - Average Purchase Value (avg\_purchase = total\_spent / num\_purchases).
* **Feature Scaling**:
  + All numeric features were scaled using **StandardScaler** to normalize the data.
* **Combining Data**:
  + Profile data from Customers.csv was merged with aggregated transaction features.

**2. Clustering**

* **Algorithm Used**:
  + **K-Means Clustering** was chosen due to its efficiency and suitability for large datasets.
* **Number of Clusters**:
  + Experiments were conducted for clusters ranging from **2 to 10**.
* **Cluster Evaluation**:
  + The Davies-Bouldin (DB) Index was calculated for each number of clusters to determine the optimal value.

**3. Visualization**

* Clusters were visualized in 2D using **Principal Component Analysis (PCA)** for dimensionality reduction.

**Results**

**Clustering Evaluation Metrics**

| **Number of Clusters** | **Davies-Bouldin Index** |
| --- | --- |
| 2 | 1.31 |
| 3 | 0.96 |
| 4 | 0.84 |
| 5 | 0.78 |
| 6 | 0.82 |
| 7 | 0.89 |
| 8 | 0.92 |
| 9 | 1.01 |
| 10 | 1.08 |

* The **optimal number of clusters is 5**, as it has the lowest DB Index (0.78).

**Cluster Descriptions**

Each cluster represents a distinct customer segment:

1. **Cluster 1**: High-income customers with frequent purchases but lower spending per transaction.
2. **Cluster 2**: Mid-income customers with moderate purchases and consistent spending habits.
3. **Cluster 3**: Low-income customers with low total spending and fewer transactions.
4. **Cluster 4**: High-spending customers with fewer but large transactions (luxury buyers).
5. **Cluster 5**: Diverse customers with average income and varied purchase patterns.

**Conclusion**

* **Optimal Clusters**: 5 clusters were identified as the optimal segmentation using the Davies-Bouldin Index.
* **Actionable Insights**:
  + Tailor marketing strategies to each cluster:
    - **Cluster 4 (Luxury Buyers)**: Focus on premium offerings.
    - **Cluster 3 (Low Spenders)**: Offer discounts or loyalty rewards to encourage more purchases.
  + **Cluster-Specific Recommendations**:
    - Product recommendations can be enhanced based on the cluster characteristics.
* **Future Work**:
  + Incorporate time-based features (e.g., recency of purchases) for more dynamic clustering.
  + Test other clustering algorithms like DBSCAN or Agglomerative Clustering for comparison.