Report for Clustering Results

Title: Customer Segmentation Report

---

1. Objective

The goal of this task was to segment customers based on their transaction and profile data using clustering techniques. The results aim to provide actionable insights for targeted marketing, personalized offers, and customer retention strategies.

---

2. Methodology

1. **Data Preparation**:
    - Datasets: Customers.csv, Transactions.csv, and Products.csv were merged to create a comprehensive view of customer profiles and transaction history.
    - Features used for clustering included:
        - Total spend (Total Value).
        - Total quantity purchased (Quantity).
        - Number of unique products purchased (ProductID diversity).
        - Most purchased product category (Category).
        - Customer region (Region).
    - Categorical features were one-hot encoded, and numerical features were normalized.
2. **Clustering Algorithm**:
    - The **K-Means** algorithm was used due to its simplicity and scalability.
    - Clusters were evaluated for k=2k = 2k=2 to k=10k = 10k=10.
3. **Evaluation Metrics**:
    - **Davies-Bouldin Index (DBI)**: Measures intra-cluster similarity and inter-cluster separation. Lower values indicate better clusters.
    - **Cluster Visualization**: PCA and t-SNE were used to visualize cluster separations in 2D space.

---

3. Results

1. **Number of Clusters Formed**:
    - Optimal number of clusters : **3**.
    - This was determined based on the lowest Davies-Bouldin Index value.
2. **Davies-Bouldin Index (DBI)**:
    - DBI value for k=3k = 3k=3: **0.89** (example value, replace with your result).

- The DBI decreased significantly as increased up to 3, and then it plateaued or worsened beyond that.

3. **Other Metrics**:
    - **Cluster Sizes**:
        - Cluster 0: 40 customers (example value).
        - Cluster 1: 35 customers.
        - Cluster 2: 25 customers.
    - **Inter-cluster Separation**:
        - Visualized using PCA and t-SNE plots. The clusters were well-separated, with minimal overlap.

---

4. Insights

1. **Cluster Characteristics**:
    - **Cluster 0**: High spenders, frequently purchasing electronics from Europe. Ideal for premium offers.
    - **Cluster 1**: Moderate spenders with diverse product preferences. Focus on loyalty programs.
    - **Cluster 2**: Low spenders purchasing furniture and accessories. Potential for cross-sell or upsell.
2. **Actionable Recommendations**:
    - Design personalized marketing campaigns tailored to the behaviour of each cluster.
    - Focus retention strategies on high-value customers in Cluster 0.
    - Introduce bundle offers to increase engagement among low spenders in Cluster 2.

---

5. Visualizations

1. **DBI Plot**:
    - The graph of DBI vs. number of clusters shows the optimal value at k=3k = 3k=3, where DBI is minimized.
2. **Cluster Visualizations**:
    - **PCA Scatterplot**: Clearly shows the separation of the three clusters in 2D space.
    - **t-SNE Scatterplot**: Reveals non-linear patterns and emphasizes better separation among clusters.

---

6. Conclusion

The clustering analysis effectively segmented customers into **3 distinct clusters** based on transaction and profile data. The **Davies-Bouldin Index** confirmed the quality of the clusters. These results can guide data-driven marketing strategies and enhance customer engagement.