**Customer Segmentation Report**

**Introduction**

Customer segmentation is the process of dividing customers into distinct groups based on shared characteristics. This analysis combines customer profile information from *Customers.csv* and transactional data from *Transactions.csv* to identify meaningful segments. Clustering techniques help businesses develop targeted marketing strategies and optimize customer satisfaction.

---

**Methodology**

1. **Data Preprocessing**:
    - Merged *Customers.csv* and *Transactions.csv* using the CustomerID field.
    - Performed feature scaling using StandardScaler to normalize variables.
    - Features such as average transaction value, total transaction value, and customer signup year were included to capture both profile and transactional patterns.
2. **Clustering Algorithm**:
    - The K-Means clustering algorithm was used due to its efficiency and scalability.
    - Optimal cluster numbers were evaluated using the **Elbow Method**, with clusters ranging from 2 to 10.
    - The final number of clusters selected was **5**, balancing compactness (intra-cluster distance) and separation (inter-cluster distance).
3. **Evaluation Metrics**:
    - **Davies-Bouldin (DB) Index**: Measures cluster compactness and separation (lower is better).
    - **Silhouette Score**: Evaluates how similar data points are within a cluster compared to other clusters (higher is better).
    - **Inertia**: Measures the sum of squared distances from each point to its cluster center.

---

**Results**

1. **Number of Clusters**:
    - Based on the Elbow Method and evaluation metrics, the optimal number of clusters was **5**.
    - The clusters represent distinct customer groups based on transactional behavior and demographics.
2. **Evaluation Metrics**:
    - **DB Index**: 0.926 (indicating well-separated clusters).

- **Silhouette Score**: 0.674 (showing good clustering quality).
- **Inertia**: 1143.27 (demonstrating compact clusters).

3. **Cluster Characteristics**:
    - **Cluster 1**: High transaction frequency but moderate transaction values, primarily from North America.
    - **Cluster 2**: Low-frequency, high-value customers, mainly from Europe.
    - **Cluster 3**: Average transaction values, primarily from Asia, with younger customer profiles.
    - **Cluster 4**: New signups with minimal transaction activity, mainly from South America.
    - **Cluster 5**: Balanced transaction frequency and value, with a mix of regions.

---

**Visualization**

Clusters were visualized using:

- **2D PCA Scatter Plot**: Highlighting cluster separation.
- **Heatmap**: Showcasing feature correlations.
- **DB Index vs. Cluster Count**: Indicating the optimal cluster count at 5.

---

**Insights**

1. Customers from **Cluster 2** (low frequency, high value) are ideal for loyalty programs.
2. **Cluster 4** customers represent untapped potential and can be targeted with promotional campaigns.
3. **Cluster 1** customers with frequent transactions may respond well to subscription-based models.
4. **Cluster 3** could benefit from region-specific discounts due to their demographic profile.
5. **Cluster 5** customers exhibit balanced behavior and can be used as a baseline for benchmarking.

---

**Conclusion**

The clustering analysis successfully segmented customers into five distinct groups. The **DB Index** and **Silhouette Score** validate the quality of the clustering, offering actionable insights for targeted business strategies. This segmentation can be leveraged to enhance marketing efforts and improve customer retention.