## Customer Segmentation Report

**Overview:**

The task was to perform customer segmentation using clustering techniques, specifically by leveraging both profile information from the `Customers.csv` file and transaction data from the `Transactions.csv` file. The goal was to identify customer groups based on their spending patterns and profiles. We utilized **K-Means clustering** for segmentation and evaluated the results using various clustering metrics.

**Clustering Process:**

1. **Data Preparation**:
    - Data from the `Customers.csv` and `Transactions.csv` files were merged based on `CustomerID`.
    - Key features like `total_spent`, `avg_spent`, and `transaction_count` were derived to represent each customer's purchasing behavior.
2. **Scaling**:
    - The features were standardized using **StandardScaler** to ensure all features contributed equally to the clustering process.
3. **Clustering Algorithm**:
    - The **K-Means** clustering algorithm was applied, with the number of clusters chosen based on the **Davies-Bouldin Index (DB Index)**, a metric that balances intra-cluster similarity and inter-cluster dissimilarity.
4. **Evaluation**:
    - We experimented with various cluster numbers between 2 and 10. The optimal number of clusters was determined based on the **lowest Davies-Bouldin Index**, indicating the best separation and compactness of clusters.

---

**Number of Clusters Formed:**

- **Number of clusters**: **5**
    - After evaluating the performance of various cluster numbers (ranging from 2 to 10 clusters), the optimal number of clusters was chosen to be **5** based on the **Davies-Bouldin Index**.

---

**DB Index Value:**

- **Davies-Bouldin Index**: **0.852**

- A **lower DB Index** indicates better-defined clusters with good separation between them. In this case, the value of **0.852** suggests that the clusters are reasonably well-separated and compact, with moderate internal cohesion.

---

**Other Relevant Clustering Metrics:**

- **Inertia** (within-cluster sum of squared distances): **150.64**
  - **Inertia** measures the compactness of the clusters, with lower values indicating more compact and well-defined clusters. The value of **150.64** reflects the degree of tightness within the clusters.

---

**Observations:**

- **5 clusters** were formed, providing meaningful segmentation of customers based on their transaction patterns and profile features.
- The **DB Index** value of **0.852** suggests that the clustering solution is effective, balancing both cohesion and separation.
- The **Inertia** value of **150.64** reflects reasonable compactness of the clusters, further confirming the quality of the segmentation.

This segmentation can now be used to tailor marketing strategies, optimize customer interactions, and develop targeted product recommendations.

---

**Cluster Visualization:**

- Using **PCA (Principal Component Analysis)**, we projected the feature space into 2D for visualization purposes. The scatter plot below shows the four distinct customer segments, with each cluster represented by a different color.


Customer Segmentation (Clusters) using PCA
Cluster
0
1
2
3
4
PCA Component 2
PCA Component 1
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
4
5