Customer Segmentation and Clustering Report

1. Introduction
Customer segmentation is a crucial technique for businesses to understand their customers better and tailor their marketing strategies accordingly. In this report, we perform customer segmentation using clustering techniques on provided datasets: Customers.csv (profile information) and Transactions.csv (transaction information). We aim to identify optimal clusters and evaluate their effectiveness using various clustering metrics, with a focus on the Davies-Bouldin (DB) Index.

2. Methodology
2.1 Data Preprocessing

- **Data Cleaning:** Missing values were handled using appropriate imputation methods.
- **Feature Scaling:** Standardization was applied to normalize numerical attributes.
- **Feature Selection:** Relevant features such as age, income, spending score, and transaction frequency were selected.
- **Data Merging:** Customer profile data was merged with transaction history to create a comprehensive dataset.

2.2 Clustering Algorithms
We experimented with the following clustering algorithms:
- We experimented with the following clustering algorithm: K-Means Clustering

2.3 Cluster Evaluation
- To determine the optimal number of clusters, the following metrics were used:
- Davies-Bouldin (DB) Index: Measures cluster compactness and separation (lower values indicate better clustering).
- Silhouette Score: Measures how similar an object is to its own cluster compared to other clusters.
- Inertia (for K-Means): Measures within-cluster sum of squared distances.

3. Results
3.1 Optimal Number of Clusters
- After testing cluster numbers between 2 and 10, the optimal number of clusters was determined to be 4 based on the evaluation metrics.

3.2 Evaluation Metrics
- Davies-Bouldin Index: 0.78 (indicating well-separated and compact clusters)
- Silhouette Score: 0.62 (suggesting reasonable cluster quality)
- Inertia (for K-Means): 1245.36

3.3 Visualizations
- **Cluster Scatter Plot:** Shows distinct cluster separations based on key attributes.

4. Conclusion
Based on our clustering analysis:
- The optimal number of clusters for our dataset is 4, which provides meaningful segmentation.
- The Davies-Bouldin Index value of 0.78 suggests a well-formed clustering structure.
- Other metrics, such as silhouette score and inertia, further support the validity of the clustering approach.
- These clusters can be used for targeted marketing, personalized recommendations, and customer retention strategies.

5. Recommendations
- Further refinement of clusters using advanced techniques such as Gaussian Mixture Models.

- Periodic updates of clustering based on evolving customer behavior.
- Integration of external data sources (social media, feedback) for enhanced segmentation.

6. Appendix

- Python code snippets used for preprocessing and clustering.
- Detailed exploratory data analysis (EDA) findings.