Customer Segmentation / Clustering

1. Objective

The objective of this task is to perform customer segmentation using clustering techniques, utilizing customer profile information from Customers.csv and transactional details from Transactions.csv.

2. Methodology

a. Feature Engineering:
    - Aggregated transaction data to create customer-level profiles:
        - TotalValue: Total revenue generated by each customer.
        - Quantity: Total quantity of items purchased.
        - TransactionID: Number of transactions made by the customer.
        - ProductID: Number of unique products purchased.
    - Normalized the aggregated features using StandardScaler for consistent scaling.

b. Clustering Algorithm:
    - Used KMeans Clustering with 5 clusters (chosen empirically).
    - Evaluated the clustering performance using the Davies-Bouldin Index (DB Index).

c. Dimensionality Reduction:
    - Applied Principal Component Analysis (PCA) to reduce high-dimensional data to 2 dimensions for visualization.

d. Visualization:
    - Generated a scatter plot to visualize the clusters in a 2D space, where each cluster is represented by a unique colour.

---

3. Results

3.1 Key Metrics

- Davies-Bouldin Index: 0.9174 (Lower values indicate better clustering quality).
- Number of Clusters: 5.

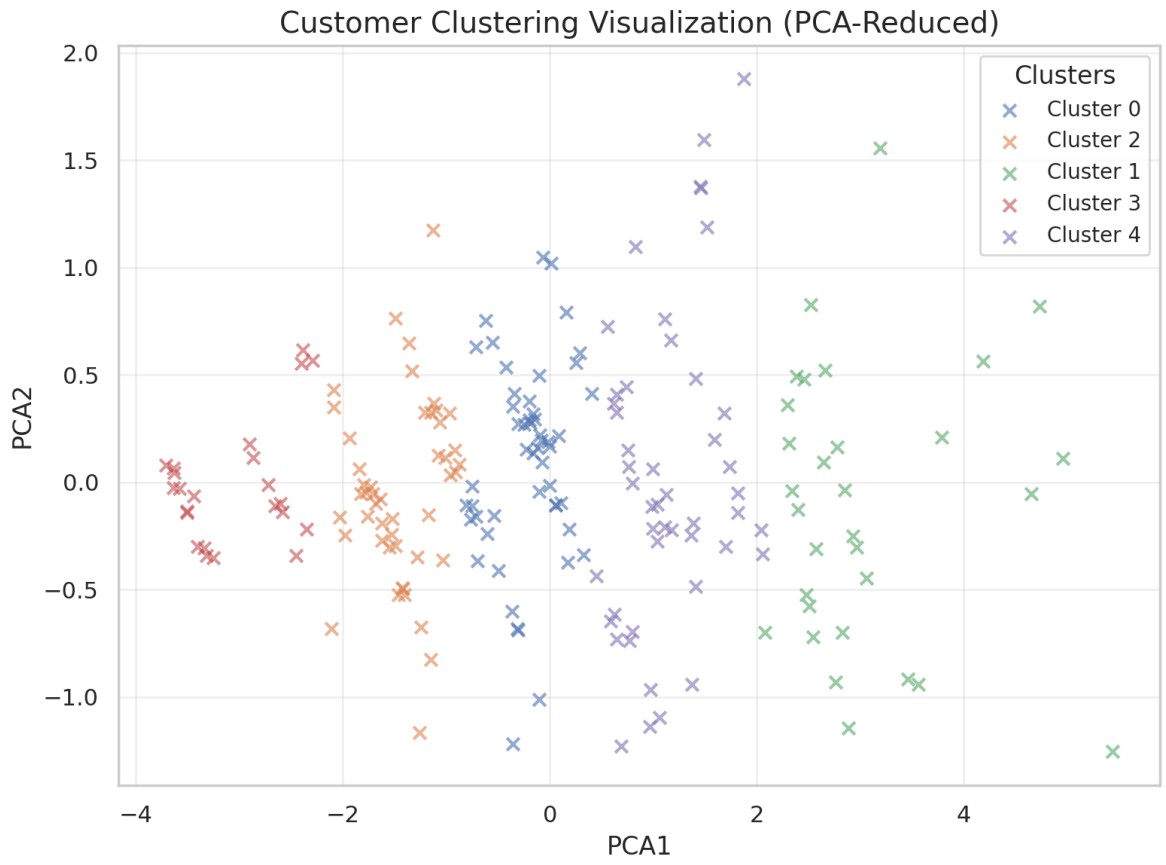
3.2 Clustered Data (Preview):

CustomerID	TotalValue	Quantity	TransactionID	ProductID	Cluster
C0001	3354.52	12	5	5	0
C0002	1862.74	10	4	4	2
C0003	2725.38	14	4	4	0
C0004	5354.88	23	8	8	1
C0005	2034.24	7	3	3	2

---

4. Visualization

The following scatter plot shows the customer clusters after applying PCA for dimensionality reduction:



---

5. Conclusion

- Customers were successfully segmented into 5 distinct clusters based on their purchasing behavior and profile attributes.
- The clusters can be used to derive targeted marketing strategies and personalized recommendations for each group.
- The Davies-Bouldin Index of 0.9174 indicates reasonable cluster quality, with distinct boundaries between groups.
