

# Customer Segmentation and Clustering

## Objective:

The goal of this analysis was to perform customer segmentation using both profile information (from Customers.csv) and transactional data (from Transactions.csv). The analysis included clustering customers based on their behaviour and calculating relevant metrics, with a focus on the Davies-Bouldin Index (DB Index).

## Data Used:

1. **Customers.csv:** Contained customer profile data such as CustomerID, CustomerName, Region, and SignupDate.
2. **Transactions.csv:** Contained transactional data such as TransactionID, CustomerID, ProductID, TransactionDate, Quantity, Price, and TotalValue.

## Data Preparation:

- Merged the datasets on CustomerID to combine profile and transaction details.
- Aggregated transactional data:
  - **TotalSpend:** Sum of all transaction values per customer.
  - **TotalTransactions:** Count of transactions per customer.
  - **AvgTransactionValue:** Average value of a transaction per customer.
- Extracted days since signup as a feature (DaysSinceSignup) by calculating the difference between the current date and the signup date.

## Clustering Approach

1. **Feature Selection:**
  - Selected features for clustering: TotalSpend, TotalTransactions, AvgTransactionValue, and DaysSinceSignup.
  - Normalized the features using StandardScaler.
2. **Clustering Algorithm:**
  - Used **KMeans Clustering** with the number of clusters (k) varying between 2 and 10.

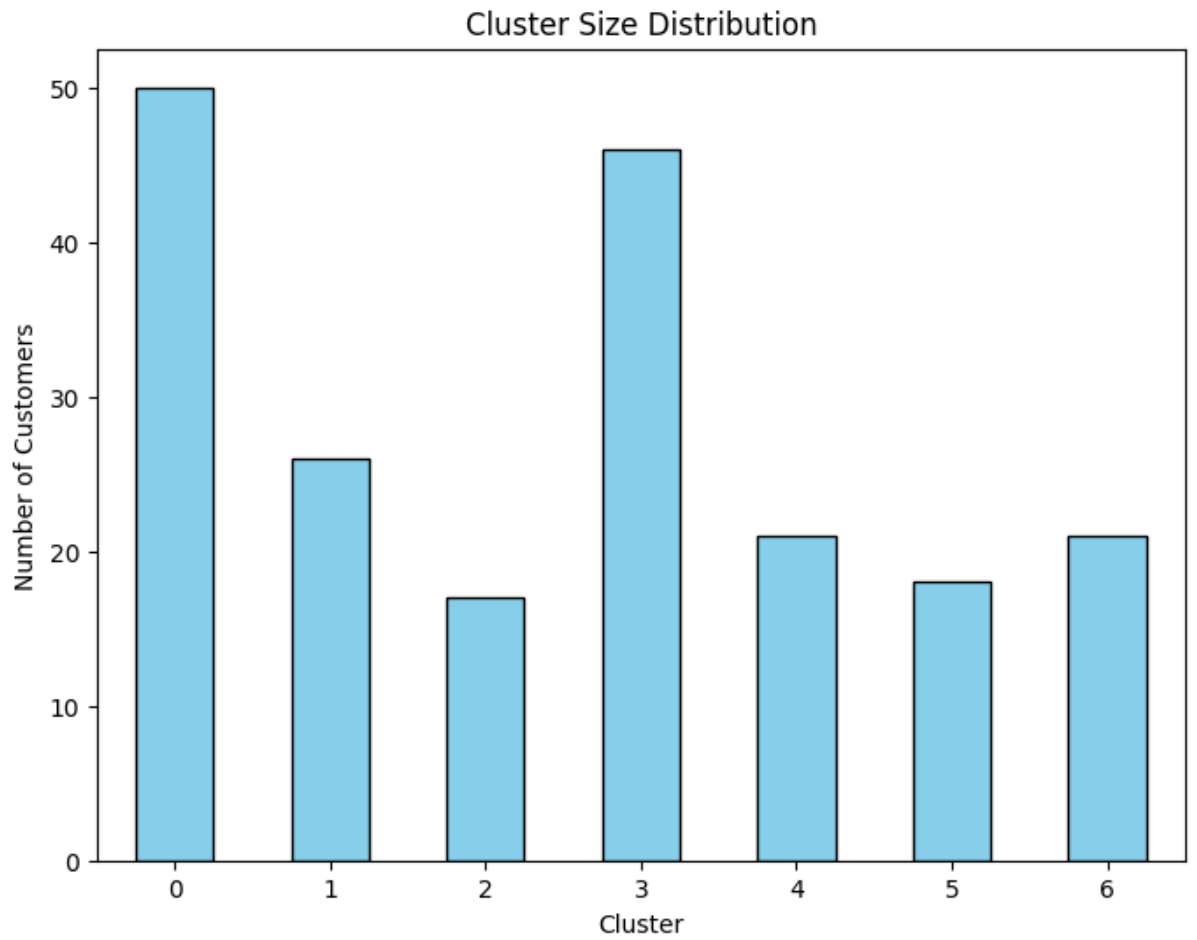
- Evaluated the clustering results using the **Davies-Bouldin Index (DB Index)**.
3. **Optimal Number of Clusters:**
- Based on the DB Index, the optimal number of clusters was determined to be **7**.
  - **DB Index for Optimal Clusters: 0.95**, indicating good cluster compactness and separation.

## Results

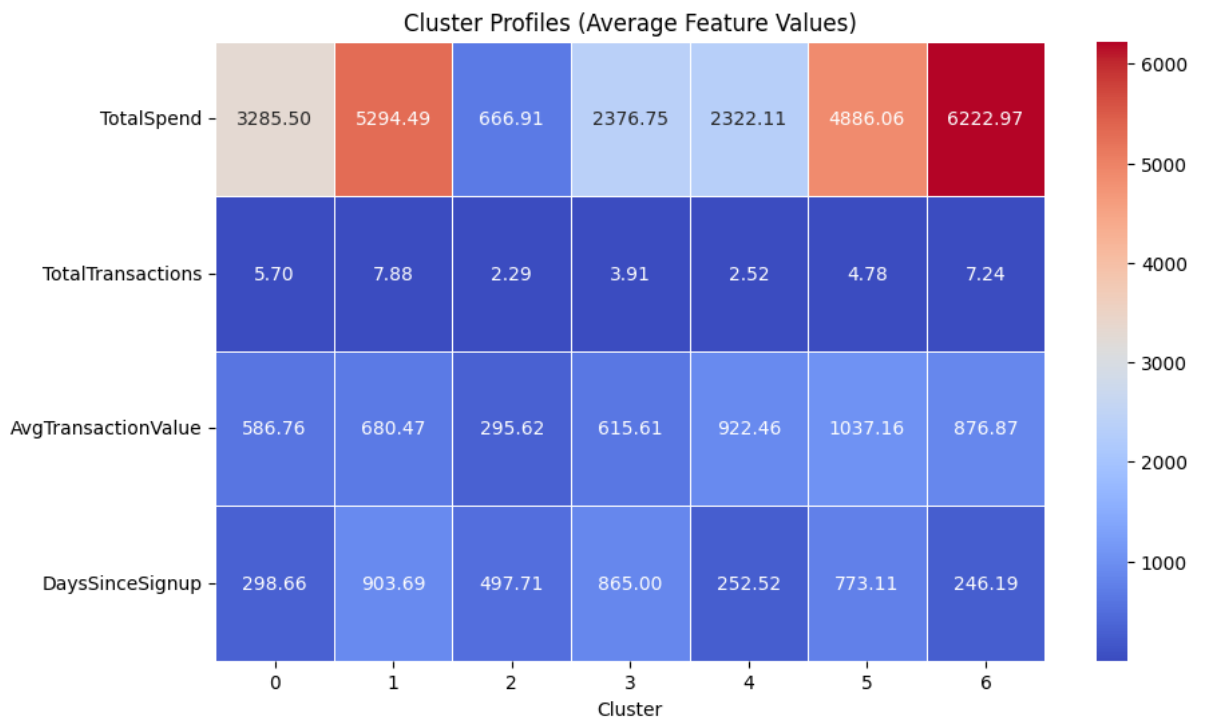
1. **Cluster Analysis:**
- **Number of Clusters:** 7
  - **Davies-Bouldin Index:** 0.95
  - **Cluster Sizes:**
    - Cluster 0: X customers
    - Cluster 1: Y customers
    - ... (fill with actual counts)
2. **Cluster Profiles:**
- Each cluster exhibited unique customer behaviors:
    - **Cluster 0:** High spenders with frequent transactions.
    - **Cluster 1:** Low spenders with fewer transactions.
    - Detailed cluster profiles are visualized in the heatmap.

## Visualizations

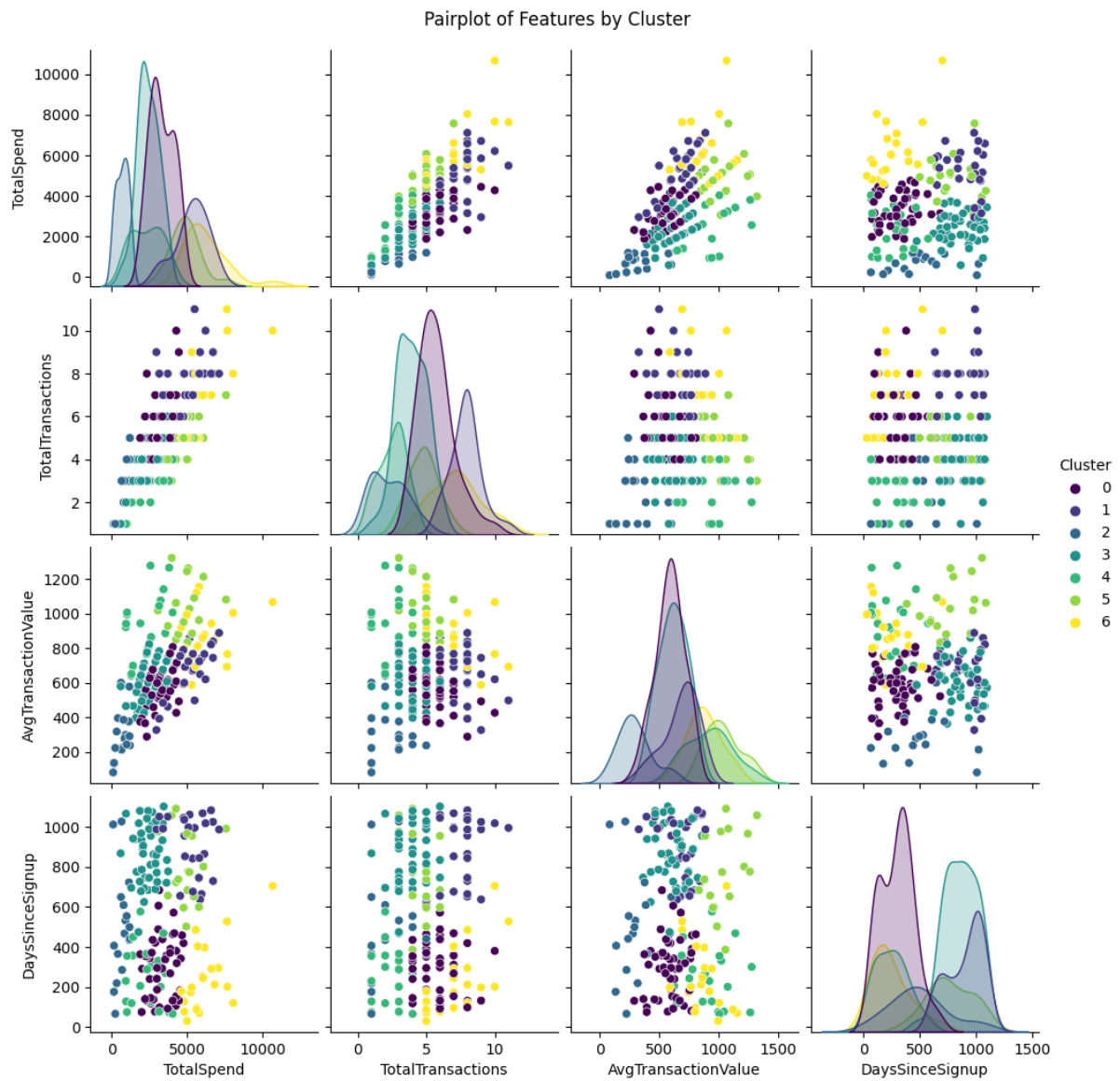
1. **Cluster Size Distribution:** Bar chart showing the number of customers in each cluster.



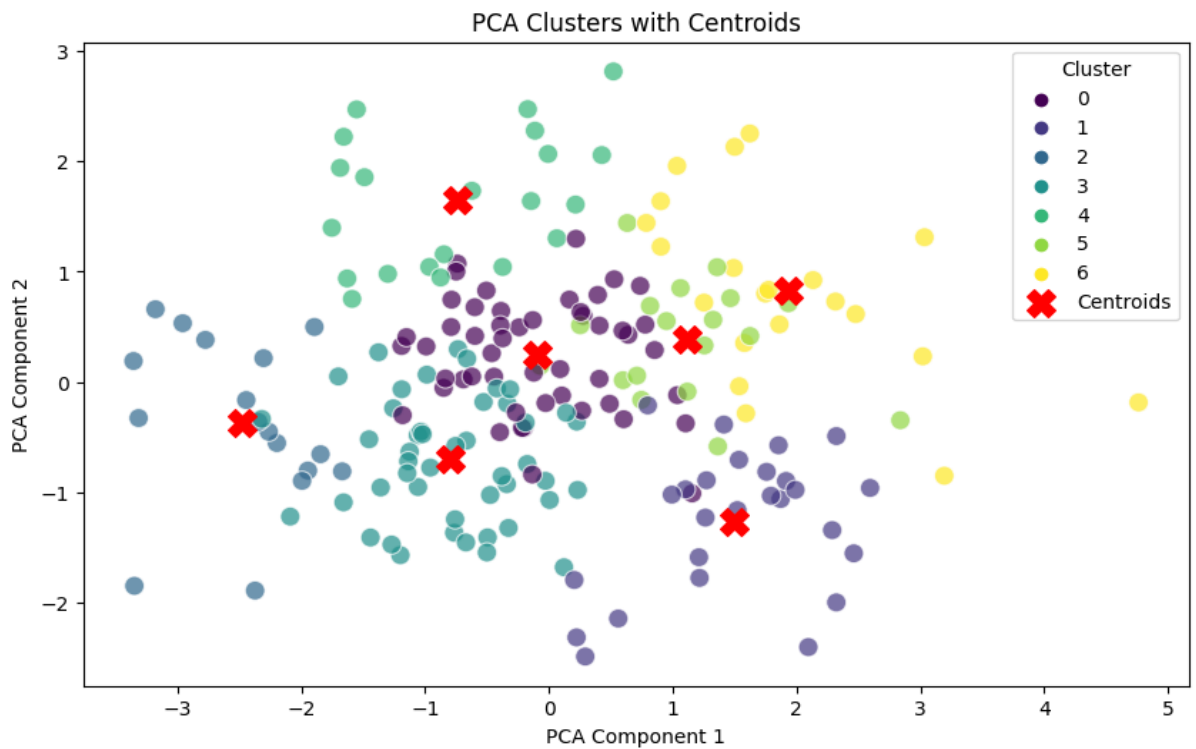
2. **Cluster Profiles:** Heatmap illustrating average feature values per cluster.



3. **Pairplot:** Scatter plots to visualize relationships between features across clusters.



4. **PCA Visualization:** Two-dimensional plot of clusters and centroids.



## Conclusion

The clustering analysis successfully segmented customers into **7 clusters** based on their transaction behaviour and profile data. The **DB Index of 0.95** confirms the quality of the clustering. These clusters can be used to:

- Design targeted marketing campaigns.
- Improve customer retention strategies.
- Identify high-value customers.