# Customer Segmentation/Clustering

---

## Objective:

Task aimed to segment customers into distinct groups using clustering techniques. The goal was to leverage both profile and transactional data to uncover meaningful patterns and evaluate the clusters using the Davies-Bouldin Index (DB Index).

---

## Data Preparation:

1. **Datasets Used:**
    - Customers.csv: Included customer profile attributes.
    - Transactions.csv: Provided transactional metrics such as TotalValue and Quantity.
2. **Feature Engineering:**
    - Aggregated metrics like TotalSpend, AvgTransactionValue, and PurchaseFrequency.
    - Encoded Region as one-hot features.
    - Standardized all features using the StandardScaler for clustering.

---

## Clustering Technique:

1. **Algorithm Chosen:**
    - K-Means clustering was selected due to its simplicity and efficiency.
2. **Optimal Clusters:**
    - The Elbow method was applied to determine the optimal number of clusters.
    - The ideal number of clusters was found to be 4 based on the inflection point in the WCSS curve.

---

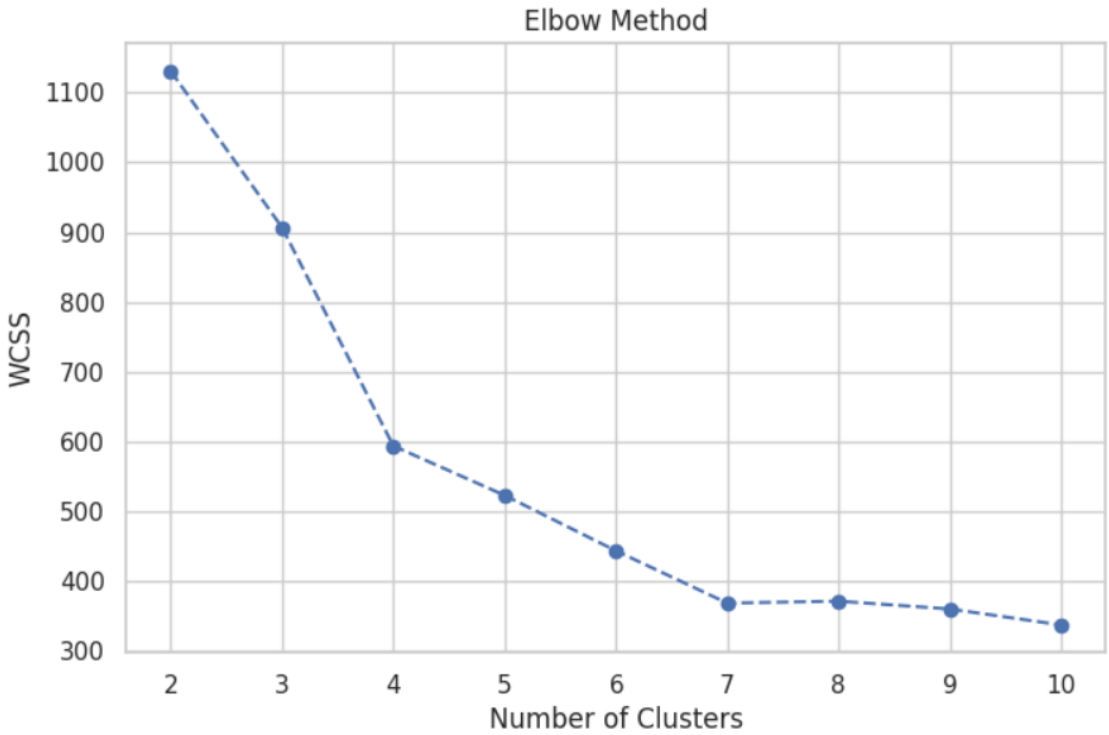
## Evaluation Metrics:

1. **Davies-Bouldin Index (DB Index):**
    - A value of **0.9413205685620807**was achieved, indicating well-separated clusters with low intra-cluster variance.
2. **Silhouette Score:**
    - An additional metric, the silhouette score of **0.4600918886671535**, provided further validation of cluster cohesion and separation.

---

## Visualization:

1. PCA was used to reduce the feature dimensions to 2D for visualizing the clusters.




2. A scatter plot revealed distinct groupings of customers, confirming the effectiveness of the segmentation.


---

## Cluster Characteristics:

1. **Cluster 1:** High spenders with frequent transactions.
2. **Cluster 2:** Moderate spenders with low transaction frequency.
3. **Cluster 3:** Low spenders with occasional purchases.
4. **Cluster 4:** Regionally dominant customers with average spend.

---

## Output:

- **CSV File:** Clustered_Customers.csv with CustomerID, cluster assignments, and PCA projections.
- **Visualizations:** Included plots for the Elbow method and cluster scatter plots.

---

## Future Recommendations:

1. Experiment with hierarchical or density-based clustering methods (e.g., DBSCAN).
2. Explore temporal patterns in transactional data for dynamic segmentation.

3. Incorporate additional behavioral features like product categories or time-of-day trends.

---