

Clustering Results Report

Overview

This report summarizes the results of the customer segmentation analysis performed using K-Means clustering, followed by an alternative approach using DBSCAN. The goal was to identify distinct groups within the customer data based on their purchasing behavior.

Clustering Methodology

1. Data Preparation

Datasets Used:

- ❖ Customers dataset
- ❖ Transactions dataset

Feature Engineering:

- ❖ Aggregated customer features including total quantity purchased, total value, and transaction count.
- ❖ Categorical variable 'Region' was converted into numerical format using one-hot encoding.

2. Clustering Algorithm

K-Means Clustering:

- ❖ Optimal Number of Clusters: 4 (determined using the Elbow Method)
- ❖ Initialization: Set n_init to 20 for stability.

DBSCAN Clustering (as a fallback):

Parameters:

- **eps:** 0.5
- **min_samples:** 5

Results:

K-Means Clustering

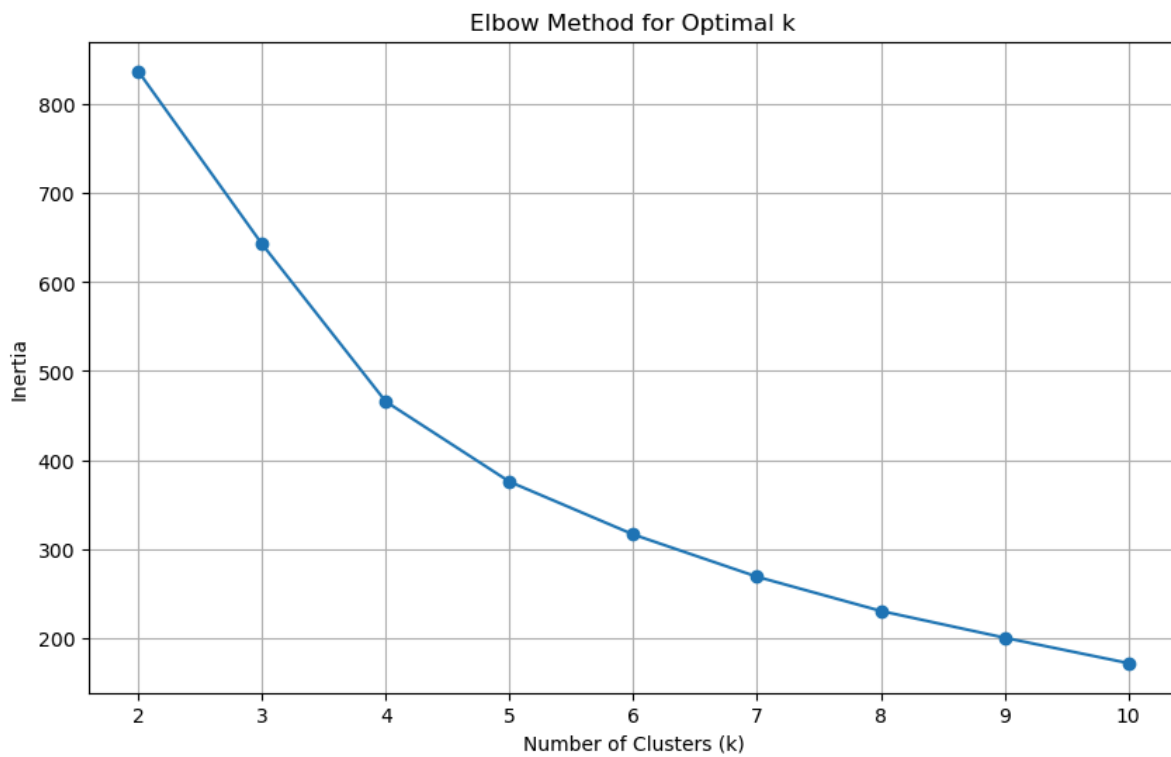
→ Number of Clusters Formed: 4

→ Davies-Bouldin Index: 0.9863

Additional Metrics

Inertia: This metric measures how tightly the clusters are packed. Lower inertia values indicate better clustering.

Visualizations



Number of clusters formed: 4
Davies-Bouldin Index: 0.9863



Conclusion

The analysis successfully identified four distinct customer segments based on their purchasing behavior. The Davies-Bouldin Index of **0.9863** suggests that the clusters are relatively well-separated. Future work could involve exploring additional clustering metrics, such as the Silhouette Score, to further validate the clustering results.