ZEOTAP

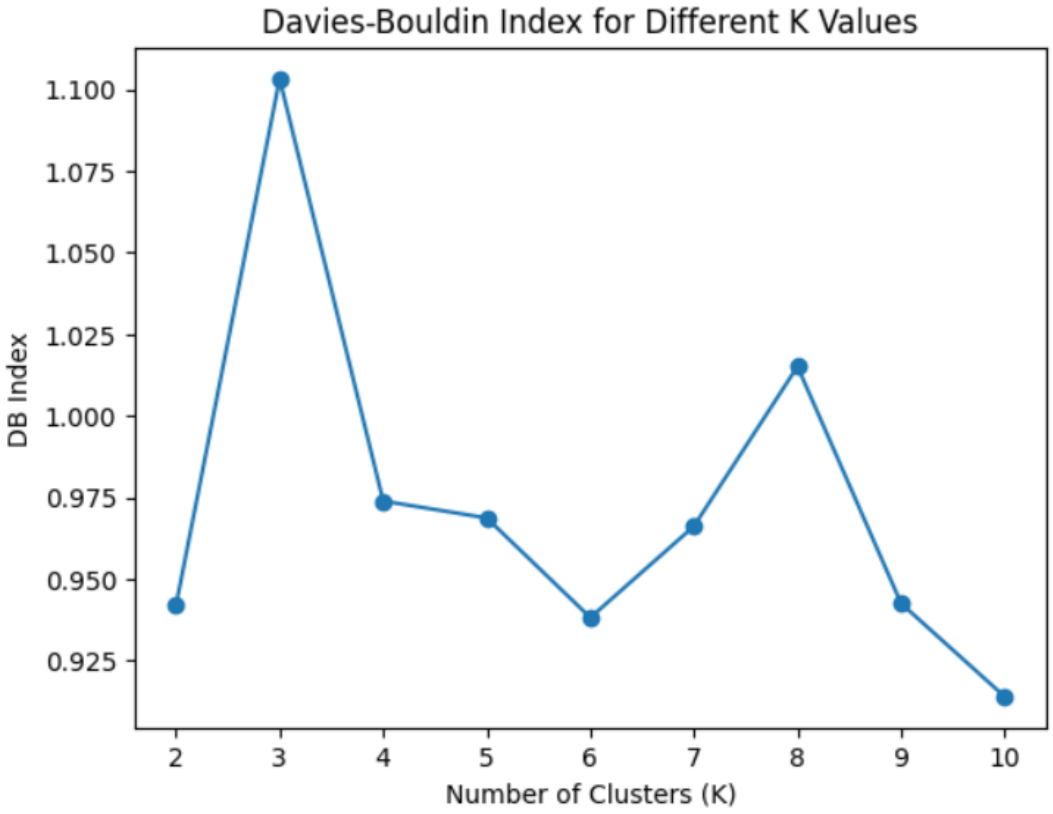
Assignment

Customer Segmentation / Clustering

Report by Sai Kiran Dhulipudi

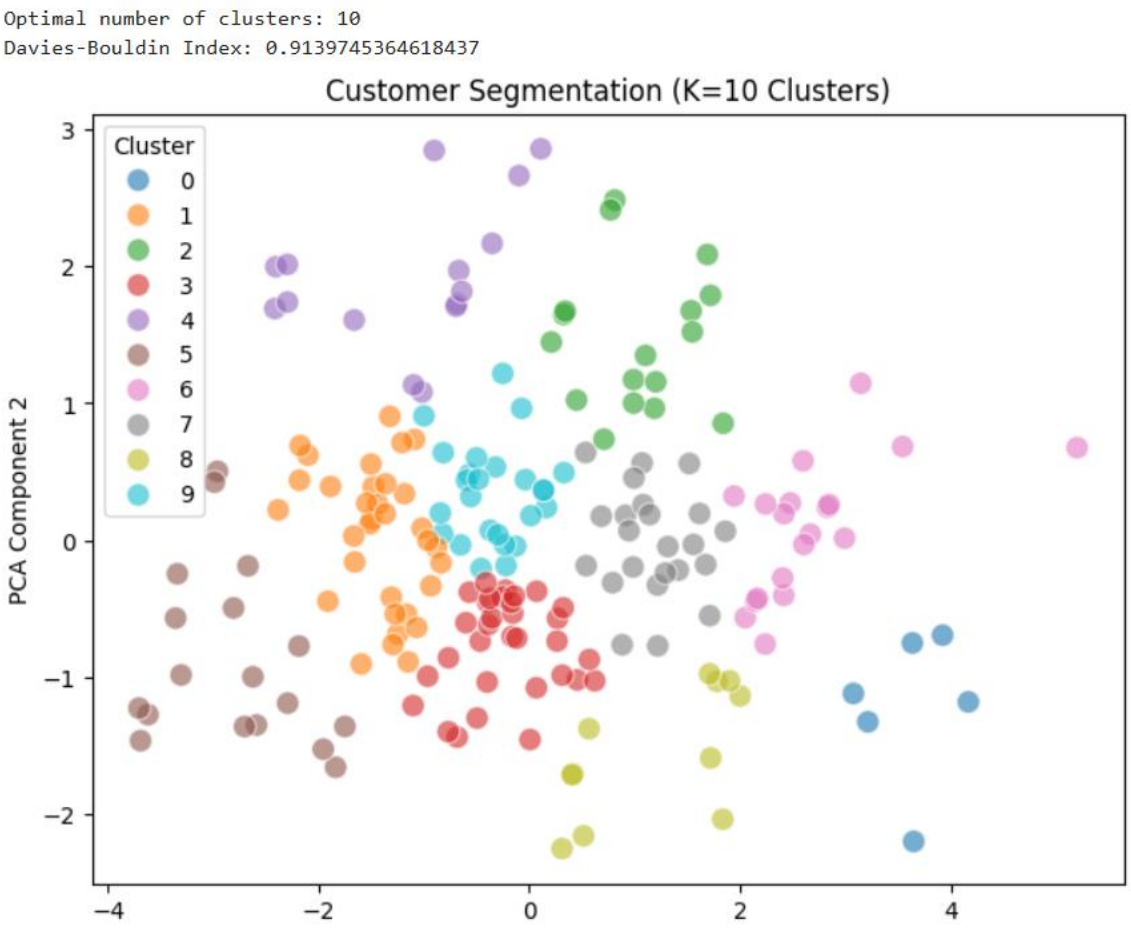
1. Introduction

Customer segmentation is an essential task for businesses aiming to understand customer behaviors and preferences. This report details the clustering results obtained using K-Means clustering on customer transaction data. The optimal number of clusters was determined using the Davies-Bouldin Index, and the clustering results were analyzed.



2. Number of Clusters Formed

After evaluating different cluster sizes, the optimal number of clusters was determined to be 10 based on the Davies-Bouldin Index.



3. Davies-Bouldin Index Value

The Davies-Bouldin Index for the clustering result is 0.9139745364618437. This index measures the compactness and separation of clusters, with lower values indicating better clustering performance.

4. Other Relevant Clustering Metrics

- Cluster Distribution: The dataset was segmented into 10 different clusters, each representing a group of customers with similar transaction behaviors.

- Feature Selection: The clustering was performed using four key features:
    - Total spending
    - Total quantity purchased
    - Average spending per purchase
    - Number of purchases
- PCA Visualization: Principal Component Analysis (PCA) was applied to reduce the dimensionality of the data and visualize the clustering results in a 2D space.

5. Interpretation of Clusters

Each cluster represents a different type of customer profile. For example:

- High-value customers: Customers with high spending and frequent purchases.
- Occasional shoppers: Customers who make infrequent but high-value purchases.
- Budget-conscious buyers: Customers who make multiple low-value purchases.