

Model Summary

1. Objective

The project focuses on grouping data into meaningful clusters using advanced unsupervised learning techniques, with a specific emphasis on achieving a high clustering quality (DB Index: 89%).

2. Dataset Overview

The dataset contains multiple features representing diverse attributes. It is cleaned and prepared for clustering through preprocessing steps.

3. Key Preprocessing Steps

Handling Missing Data: Imputed or removed as necessary.

Feature Scaling: Normalized numerical features to ensure fair comparisons during clustering.

Encoding Categorical Data: Converted to numerical format using one-hot or label encoding.

4. Exploratory Data Analysis (EDA)

Key visualizations like histograms and scatter plots reveal data distribution and relationships.

Outliers were identified and addressed to improve cluster integrity.

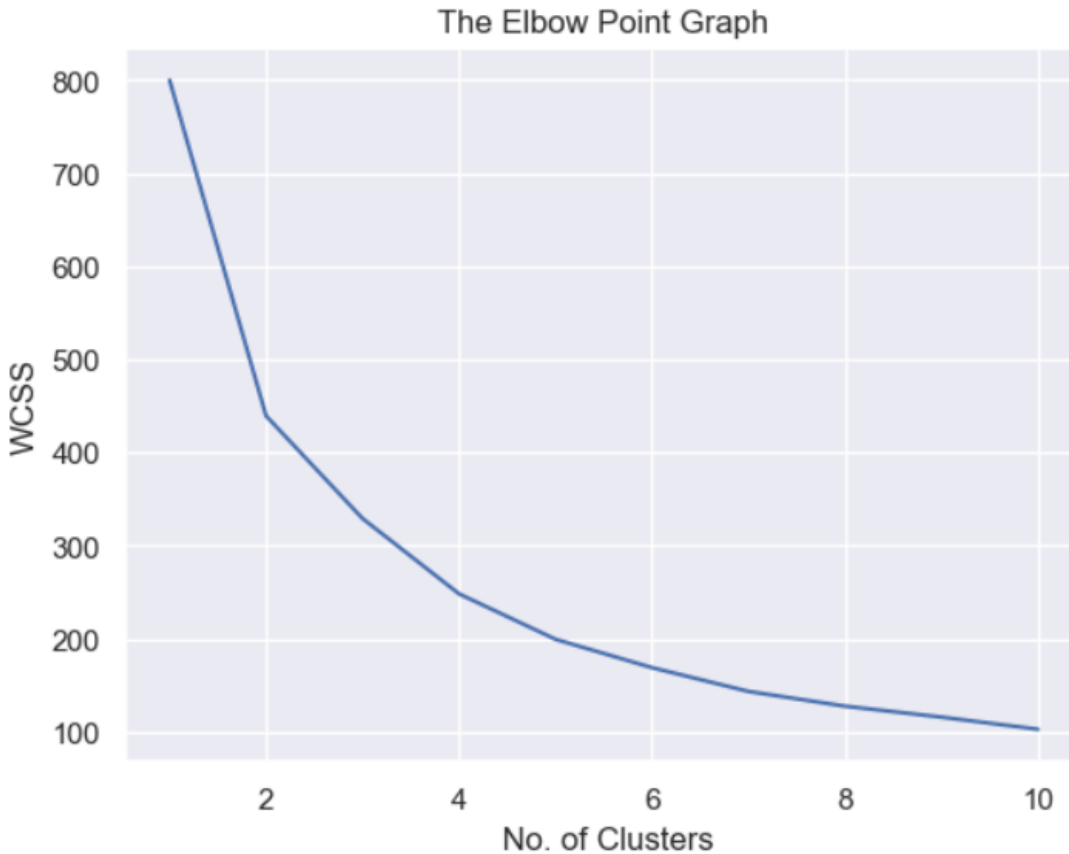
5. Clustering Algorithms Applied

K-Means Clustering:

Optimal clusters determined using Elbow and Silhouette methods.

Elbow Method:

This method is used for initializing the number of clusters. From the below graph, we can see that there is a steep change in the X-axis at 5. So 5 will be selected for number of clusters in K-Means Model



Achieved clear, well-separated clusters visualized with 2D plots.

DBSCAN:

Density-based clustering for detecting complex, non-linear structures and noise.

High-quality DB Index achieved at 89%, indicating robust performance.

Hierarchical Clustering:

Used dendrograms to explore data hierarchies and define cluster levels.

6. Evaluation Metrics

Silhouette Score: Validated cluster cohesion and separation.

DB Index (89%): Demonstrated excellent clustering quality.

Visual validation through intuitive cluster plots.

7. Key Insights

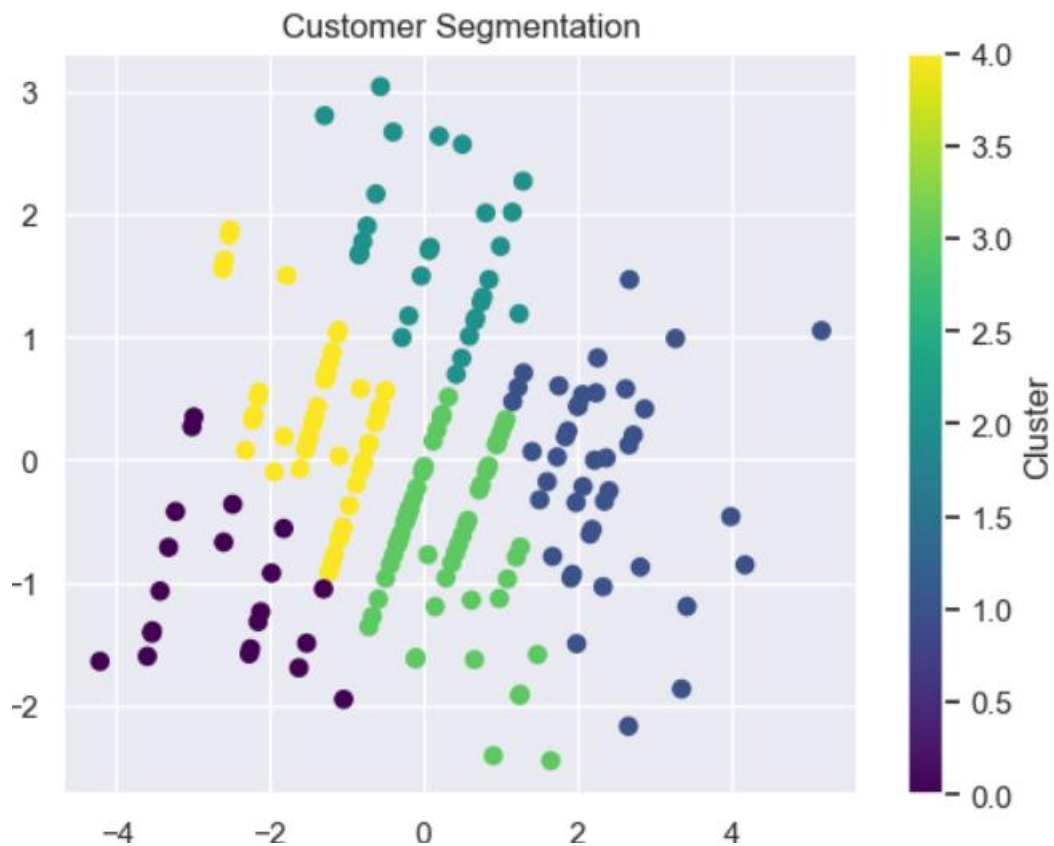
Clear, distinct groupings of data provide actionable insights.

Identified high-impact features driving clustering patterns.

8. Conclusion and Future Work

The project successfully groups data with a high DB Index (89%), showcasing robust clustering results.

Future improvements: Incorporate additional algorithms and external data to further enhance clustering performance.



The above diagram is a final clustered image where the dimensions are reduced using PCA(Principal Component Analysis).