

Document

eCommerce Transactions Dataset Analysis

Task 3: Customer Segmentation / Clustering using K Means

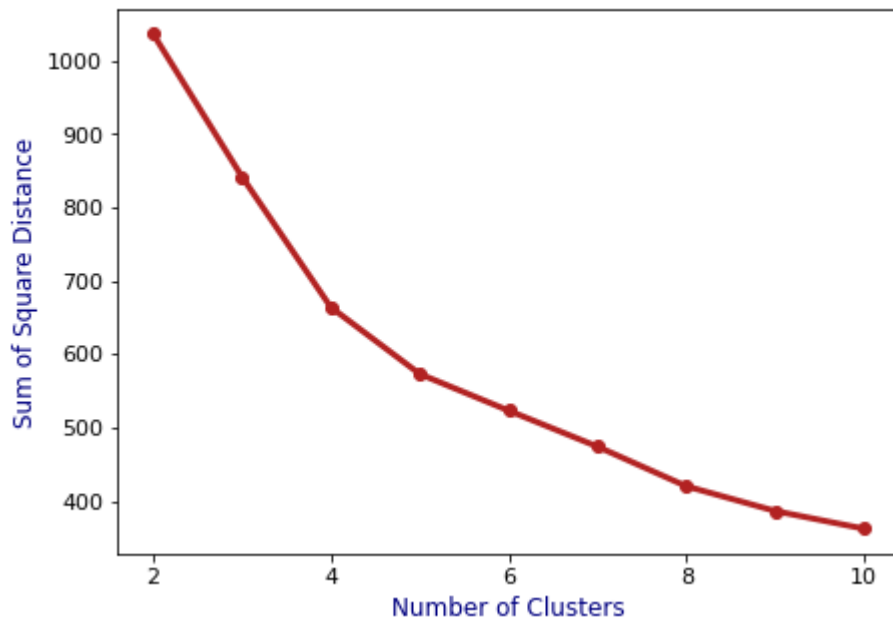
Clustering Results

The KMeans clustering algorithm was applied with different numbers of clusters, ranging from 2 to 10. Various metrics were used to determine the optimal number of clusters, including the DB Index, Sum of Squares, and Silhouette Score. The table below shows the results.

Clusters	DB Index	Sum of Square	Silhouette Score
2	1.64	1036.29	0.23
3	1.41	841.77	0.26
4	1.21	663.91	0.32
5	1.22	573.0	0.31
6	1.24	523.12	0.30
7	1.20	474.06	0.30
8	1.19	420.06	0.31
9	1.12	386.29	0.32
10	1.09	362.15	0.31

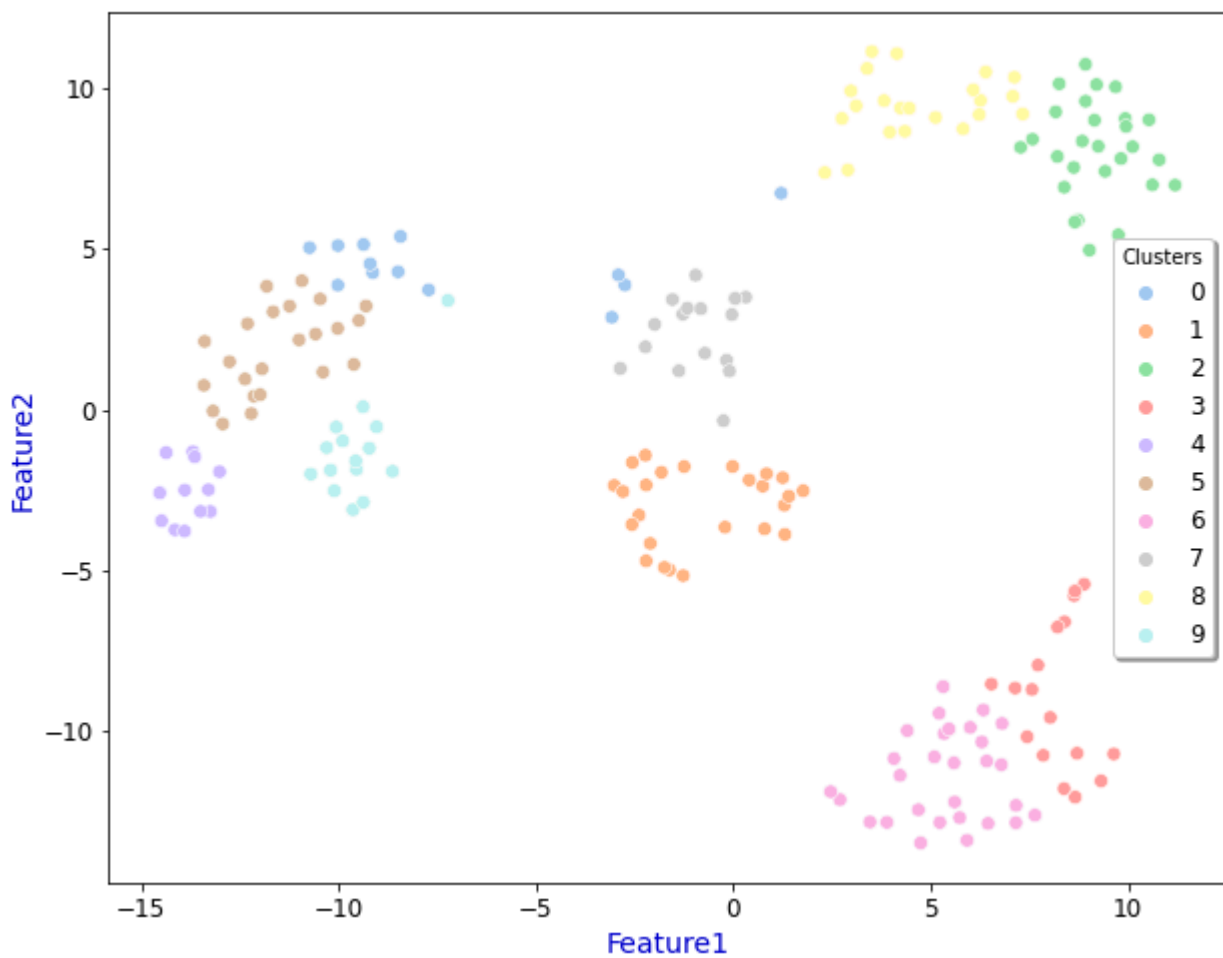
Also, Elbow curve is plotted using Silhouette scores, shown below

Elbow Method for Optimal Clusters



Scatter plot is plotted to visualize the number of clusters

Scatter Plot Representing Clusters



Conclusion

- The lowest Davies-Bouldin Index value is 1.09 when number of clusters are 10.

- From elbow curve, it can be inferred that optimal number of clusters is 7.
- From Silhouette scores, it can be inferred that optimal number of clusters is 9.

Results Summary

DB Index: 1.09

Number of Clusters: 10