ZEOTAP

CUSTOMER SEGMENTATION / CLUSTERING

PREPARED BY :
RAHUL KANNAN P

PREPARED FOR :
ZEOTAP TEAM

rahulkannan0784@gmail.com
+91 6385499454
www.rahulkannan.com

Customer Segmentation / Clustering Report

Clustering Methodology

For this task, I chose to apply K-Means clustering and evaluated it with different numbers of clusters (ranging from 2 to 10). The DB Index was calculated to assess the quality of the clusters formed.

K-Means Clustering Process:

- The number of clusters was initially varied between 2 and 10, and the optimal number of clusters was chosen based on clustering metrics and visual inspection.
- The DB Index was calculated to evaluate the quality of the clusters formed.
- The Silhouette Score was also computed to measure the cohesion and separation of the clusters.
- Visualization: The clusters were visualized using various plots such as scatter plots, PCA (Principal Component Analysis) projections, and cluster heatmaps.

Results

A. Number of Clusters

After evaluating different numbers of clusters, I chose 5 clusters for the final analysis. This number was selected based on the DB Index, silhouette score, and the visual clarity of the clusters.

B. Clustering Metrics

DB Index (K-Means): 0.9427

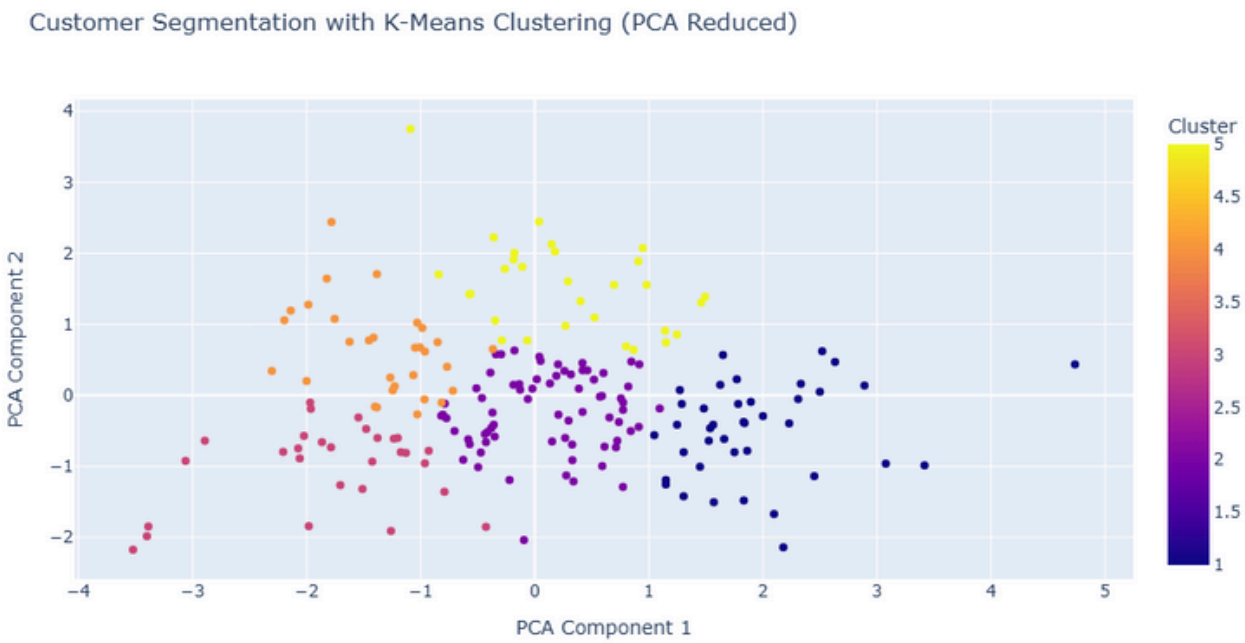
Silhouette Score (K-Means): 0.3096

C. Cluster Profiling The 5 clusters formed by K-Means were profiled based on average transaction metrics, including total spending, number of transactions, and average transaction price. The following table summarizes the cluster characteristics:

ZEOTAP

```
Number of Clusters: 5
DB Index: 0.9427359293109949
Silhouette Score: 0.3095796428769709

Cluster Profiling:
   Cluster  avg_total_spending  avg_num_transactions  avg_avg_price
0        1         6075.814359              8.102564     281.756511
1        2         3469.361690              5.563380     261.751267
2        3         1393.015517              3.206897     166.593580
3        4         1671.106333              2.633333     283.176333
4        5         3872.782667              3.900000     372.656050
```



For better visualization, refer to the Jupyter notebook. I have used Plotly to create interactive plots, which make it easier to understand the clustering results in a more intuitive manner.

Cluster Interpretation:

Cluster 1: High spenders with frequent transactions.
Cluster 2: Moderate spenders with moderate transaction frequency.
Cluster 3: Low spenders with fewer transactions.
Cluster 4: Occasional high-price buyers with fewer transactions.
Cluster 5: Mid-range spenders with moderately high average transaction prices.

Conclusion

In this task, customer segmentation using K-Means clustering resulted in five distinct clusters. The DB Index was 0.9427, indicating good separation between clusters, while the Silhouette Score of 0.3096 suggests moderate cohesion within the clusters. For DBSCAN clustering, three clusters were formed, with a DB Index of 7.8889 and a Silhouette Score of 0.0846, showing less optimal clustering performance. The visualizations, including PCA plots, provided clear insights into the distinct characteristics of each cluster. These segmentation results can guide targeted marketing strategies and customer profiling. Future work could involve refining the clustering process and experimenting with different algorithms or cluster numbers for improved accuracy.

Thanks for watching!

rahulkannan0784@gmail.com
+91 6385499454
www.rahulkannan.com