

DATA SCIENCE INTERN ASSIGNMENT



ZEOTAP

Report on

Customer Segmentation Report: Clustering Analysis

Submitted By: **Shubham Kumar**
Indian Institute of Engineering Science and Technology, Shibpur

Objective:

The primary objective of this analysis was to do consumer segmentation using clustering approaches. The purpose was to segment consumers based on both their profile information (from the consumers.csv file) and transaction data (from the Transactions.csv file), resulting in separate customer groups that could be used to develop personalised marketing and business strategies.

Clustering Approach:

I used the K-Means clustering method to divide the clients into separate groups. After experimenting with the amount of clusters, I concluded that 5 clusters were the best fit based on clustering evaluation criteria such as the DB Index, Inertia, and Silhouette Score.

Clustering Results:

1. Number of Clusters Formed:

After analysing consumer data, the clustering model identified 5 clusters. These clusters indicate various purchase behaviours and client profiles, allowing organisations to adjust their marketing efforts more successfully.

2. DB Index:

The DB Index for this clustering solution is 0.8525, indicating substantial overlap between groups. A lower DB Index indicates better cluster separation, and this result is somewhat higher than the ideal range, indicating that there may be some space for improvement in terms of clear customer segmentation.

3. Inertia:

This model has an inertia value of 150.6441, calculated as the sum of squared distances to the nearest cluster centre. This comparatively low number suggests that the data points within each cluster are closely clustered around their respective cluster centres, implying that each cluster is internally cohesive. However, while inertia is minimal, it does not guarantee the most distinct consumer groups, as evidenced by the DB Index.

4. Silhouette Score:

The Silhouette Score of 0.3535 indicates that the clusters are not extremely distinguishable from one another. A higher silhouette score (closer to 1) indicates more distinct clusters with obvious edges. The comparatively low score indicates that the customers inside the clusters

are not totally cohesive, and there is considerable overlap, implying that the segmentation may be less effective than it could be with fewer groups.

Insights:

- **Cluster Cohesion and Separation:** While the inertia value indicates compact clusters, the DB Index and silhouette score indicate areas for improvement in segmentation. The current segmentation model may require modification, as the clusters overlap and do not fully reflect diverse client behaviours.
- The DB Index and silhouette score show that 4 clusters may give better consumer segmentation than the 5-cluster model.
- Customer segmentation into 5 clusters offers useful information into purchase behaviour. While the clusters are not completely unique, they can provide opportunities for focused marketing and personalised communication, which can dramatically boost consumer engagement and retention.

Conclusion:

According to the findings, client segmentation might help organisations personalise their marketing efforts more effectively. While the five clusters are beneficial for segmentation, more refinement is required to optimise the clustering process and improve customer group distinctness. Using clustering indicators like as the DB Index and silhouette score, the company may constantly enhance its customer segmentation methods, thereby promoting long-term client loyalty and increasing revenue.

Recommendations:

- Use alternate clustering approaches (e.g. hierarchical clustering) to enhance cluster separation and cohesiveness.
- Optimize the number of clusters to balance segmentation quality (DB Index and Silhouette Score) with interpretability.
- Use segmentation results to generate tailored marketing campaigns based on each cluster's purchase behaviours.

