

Comprehensive Report on Customer Segmentation

1. Introduction

Customer segmentation is a crucial technique in marketing that divides a customer base into distinct groups, or clusters, based on various characteristics and behaviors. This report explores three different clustering methods—K-Means, Hierarchical Clustering (HC), and DBSCAN (Density-Based Spatial Clustering of Applications with Noise)—applied to a dataset containing purchasing behaviors of FLO customers.

2. Methods and Analysis

2.1. K-Means Clustering

K-Means is a widely-used clustering algorithm that partitions data into a pre-specified number of clusters. In this analysis:

- The algorithm was applied to the dataset, specifying 8 clusters.
- After clustering, the data was analyzed to determine the characteristics of each cluster.

Results:

- **Number of Clusters:** 8
- Each cluster represents a group of customers with similar purchasing behaviors.
- The clusters were evaluated using metrics such as mean, median, count, and standard deviation for each cluster.

2.2. Hierarchical Clustering (HC)

Hierarchical Clustering, specifically the agglomerative approach, builds clusters based on a connectivity matrix. The linkage matrix was created using different methods (e.g., 'ward', 'average') to assess their impact on clustering:

- A dendrogram was generated to visualize the hierarchical relationships between data points.
- The dataset was ultimately segmented into 3 clusters using the Ward method.

Results:

- **Number of Clusters:** 3
- Clusters were formed based on the degree of similarity among data points.
- Cluster characteristics were evaluated similarly to K-Means, focusing on customer segmentation.

2.3. DBSCAN Clustering

DBSCAN is a density-based clustering method that can identify clusters of arbitrary shapes and handle noise in the data:

- Parameters such as `eps` and `min_samples` were set to define the density threshold for forming clusters.

- This method detected 7 clusters within the dataset.

Results:

- **Number of Clusters: 7**
- DBSCAN identified clusters based on the density of data points, with some points classified as noise.
- The resulting clusters were analyzed using statistical metrics to understand the customer distribution across clusters.

3. Comparative Analysis

The results from the three clustering methods revealed different insights:

- **K-Means (8 Clusters):** Provided the most granular segmentation, dividing customers into 8 distinct groups. This method is effective for larger datasets where clear partitions exist.
- **Hierarchical Clustering (3 Clusters):** Offered a broader view by grouping customers into 3 larger clusters, suitable for understanding higher-level relationships.
- **DBSCAN (7 Clusters):** Identified clusters based on density, with the ability to handle noise. This method is ideal for discovering clusters in datasets with varying densities.

4. Conclusion

Each clustering method provided unique insights into the customer base:

- **K-Means** was useful for fine-grained segmentation, identifying detailed customer segments.
- **Hierarchical Clustering** highlighted broader customer groups, offering a macro-level view of the customer base.
- **DBSCAN** effectively handled noise and revealed clusters based on density, providing a flexible approach to segmentation.

These results demonstrate the importance of selecting an appropriate clustering method based on the specific characteristics of the data and the goals of the analysis. Each method has its strengths, and their combined insights can lead to a more comprehensive understanding of customer behaviors, enabling more targeted marketing strategies and product development efforts.