

MALL CUSTOMER SEGMENTATION REPORT

AMAT – 592 (Jahnavi Kalyanam [001612364])

1. INTRODUCTION

This project aims to address the challenge of customer segmentation, which is crucial for businesses seeking to tailor their marketing strategies and improve customer engagement. By analyzing customer data, including demographic information and spending behavior, businesses can identify distinct groups with similar characteristics. The goal is to solve the problem of identifying actionable customer segments and providing insights that can drive more targeted, effective marketing efforts.

2. DATA DESCRIPTION

The dataset contains information about 200 customers with the following features:

- **CustomerID:** Unique identifier for customers.
- **Gender:** Male (encoded as 1) and Female (encoded as 0).
- **Age:** Customer age in years.
- **Annual Income (k\$):** Customer's annual income in thousands of dollars.
- **Spending Score (1-100):** A score assigned to customers based on their spending behavior and loyalty.

3. PREPROCESSING

The dataset underwent thorough preprocessing to ensure the quality and reliability of the clustering analysis. First, missing values were checked across all features, and none were detected, confirming the dataset's completeness. Next, the categorical variable Gender was encoded using label encoding, where "Male" was represented as 1 and "Female" as 0, making the data suitable for numerical computations. Finally, all features were normalized to a uniform scale, ensuring fair clustering by eliminating the influence of differing measurement units across variables.

4. METHODOLOGY

Cluster analysis was chosen for this project to identify and segment customers into distinct groups based on their similarities in demographic and behavioral characteristics. The clustering analysis employed two distinct approaches: K-Means Clustering [from scratch] and Hierarchical Clustering [using sklearn], each offering unique methods for grouping customers based on their attributes.

K-Means Clustering began with the random initialization of centroids, followed by iterative updates to minimize the distance between each data point and its nearest centroid. This process continued until the centroids stabilized, ensuring optimal grouping of customers. The performance of this clustering method was evaluated using the Silhouette Score, achieving a value of **0.458**, indicating moderately well-defined clusters.

Hierarchical Clustering took an alternative approach by employing agglomerative clustering. This method grouped customers prioritizing minimum variance within clusters. The resulting clusters were visualized through a dendrogram, which provided a hierarchical perspective on customer relationships and merging processes. The Silhouette Score for this method was **0.446**, slightly lower than that of K-Means, yet still indicative of meaningful clustering.

MALL CUSTOMER SEGMENTATION REPORT

AMAT – 592 (Jahnavi Kalyanam [001612364])

Both methods provided valuable insights into customer segmentation, with K-Means offering slightly better-defined clusters and Hierarchical Clustering adding a hierarchical understanding of customer groupings.

5. CLUSTER ANALYSIS

Three clusters were identified, characterized as follows:

5.1 Cluster Centroids:

K-means

Cluster	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1.0	4.87	4.63	6.16
1	11.0	7.05	4.89	3.88
2	11.0	2.97	4.85	8.21

Hierarchical Clustering

Cluster	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
Cluster 0	1.0	4.87	4.63	6.16
Cluster 1	11.0	6.78	4.71	4.07
Cluster 2	11.0	2.91	5.10	8.41

5.2 Overall Means for Comparison

Feature	Mean
Gender	5.40
Age	5.01
Annual Income (k\$)	4.73
Spending Score (1-100)	6.02

5.3 Cluster Characteristics:

5.3.1 K-Means Clustering

Cluster 0 consists of middle-aged individuals with slightly below-average income and a moderately high spending score. These individuals are budget-conscious but still willing to spend on value-driven products, making them a good target for promotions and discounts.

MALL CUSTOMER SEGMENTATION REPORT

AMAT – 592 (Jahnvi Kalyanam [001612364])

Cluster 1 consists of older individuals with average income but a low spending score. The group mainly consists of customers who are cautious with their spending and prioritize essential and practical purchases.

Cluster 2 is made up of younger individuals with above-average income and the highest spending score. The cluster includes customers who are affluent and enthusiastic spenders, likely to engage in discretionary purchases, premium products, and lifestyle-oriented offerings.

5.3.2 Hierarchical Clustering

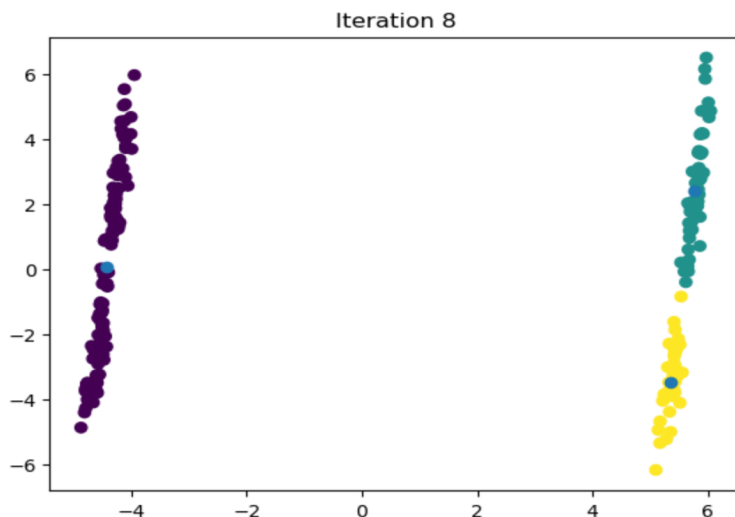
Cluster 0 consists of middle-aged customers with slightly below-average income and a slightly above-average spending score. These individuals are budget-conscious but still willing to spend strategically.

Cluster 1 is characterized by predominantly, older customers with below-average income and a significantly lower spending score. This group represents cautious spenders who focus on essentials rather than discretionary purchases.

Cluster 2 includes predominantly younger customers with above-average income and the highest spending score. These customers are affluent, spend-eager, and likely to engage in discretionary and lifestyle purchases.

6. VIZUALIZATIONS

6.1 K-Means Clustering:

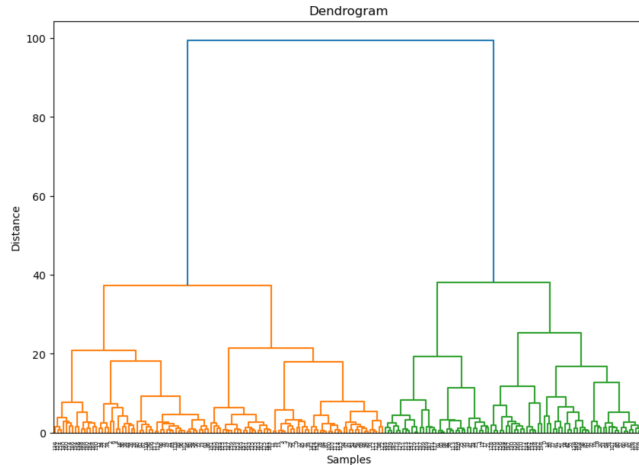


The data points are clearly divided into three distinct vertical clusters, each represented by a different color (purple, teal, and yellow). The clustering has converged effectively, with minimal overlap between the clusters.

MALL CUSTOMER SEGMENTATION REPORT

AMAT – 592 (Jahnavi Kalyanam [001612364])

6.2 Hierarchical Clustering Dendrogram:



The dendrogram shows two distinct clusters, separated by a significant distance, indicating strong dissimilarity. Each cluster further divides into smaller, closely related subgroups. The clear separation suggests the dataset contains two main groups.

7. EVALUATION

The K-Means clustering algorithm achieved a Silhouette Score of 0.458, while Hierarchical Clustering produced a slightly lower score of 0.446. The Silhouette Score measures how well-defined the clusters are, with higher values indicating that data points are more tightly grouped within their respective clusters and better separated from other clusters. In this case, K-Means clustering performed marginally better, compared to those generated by Hierarchical Clustering. While both methods provided reasonably good clustering, the higher score for K-Means indicates a slight edge in cluster quality.

8.TARGET CUSTOMER RECOMMENDATION

The ideal target customers are those in Cluster 2, consisting primarily of younger individuals with above-average incomes and the highest spending scores. This group demonstrates a strong inclination toward spending. Marketing efforts should focus on offers, products, and campaigns to effectively engage and attract this segment, as they possess both the financial capacity and enthusiasm for such purchases.

9. CONCLUSION

This project successfully segmented customers into meaningful clusters. The **K-Means clustering** technique produced slightly better-defined clusters compared to Hierarchical clustering, as reflected in the higher Silhouette Score. This indicates that K-Means was more effective in partitioning the customers. This segmentation offers businesses insights into customer behavior, allowing for tailored marketing strategies that can enhance engagement and boost sales.