

Customer Segmentation Analysis Using 3D Clustering

Student: Daniel Sorban

Exam Session A: 23.01.2026

Abstract

The goal of this project is to identify distinct groups of mall customers to improve marketing strategies. Using a dataset of customer demographics and spending habits, we applied machine learning algorithms (K-Means and Hierarchical Clustering). Unlike simple analyses that only look at income, we included **Age** as a third dimension. Our analysis revealed six distinct customer profiles, allowing for more personalized and effective advertising campaigns.

1 Introduction

In retail, treating every customer the same is inefficient. "Customer Segmentation" is the process of dividing customers into groups based on common characteristics.

This project analyzes the **Mall Customers** dataset. Our objective is to find hidden patterns in the data that are not obvious to the naked eye. We focus on three key variables:

- **Age:** How old the customer is.
- **Annual Income:** How much they earn (in thousands of \$).
- **Spending Score:** A score from 1-100 assigned by the mall based on customer behavior.

2 Methodology

We used the R programming language to process the data and build the clusters. The process consisted of three main steps.

2.1 Data Preparation

The variables in our dataset have very different ranges. For example, Age goes up to 70, but Income goes up to 137. If we ignore this, the computer will think Income is more important just because the numbers are bigger.

To fix this, we used **Scaling** (Standardization). This adjusts all values so they are on the same "playing field," allowing the algorithm to compare Age and Income fairly.

2.2 Finding the Number of Groups

One of the hardest parts of clustering is deciding how many groups (k) actually exist. We used two methods to be sure:

1. **The Dendrogram (Tree Diagram)** We used Hierarchical Clustering to build a "family tree" of customers. It groups similar people together step-by-step.

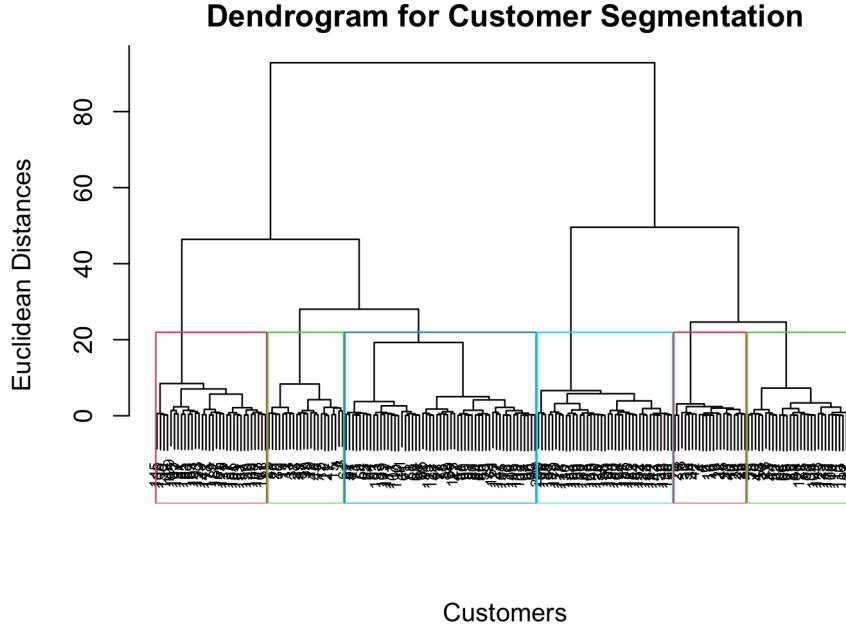


Figure 1: The Dendrogram suggests 6 distinct groups.

As seen in Figure 1, the tree splits naturally into 6 main branches.

2. **Silhouette Analysis** We mathematically checked the quality of the clusters using the Silhouette Score. This measures how "tight" each group is. The analysis confirmed that **6 clusters** gave the best mathematical result.

3 Results and Analysis

We applied the K-Means algorithm with $k = 6$. By visualizing the results in a 3D space (Age vs. Income vs. Score), we identified the following customer types:

1. **The Savers (High Income, Low Spend):** These people earn a lot but rarely spend. Marketing should focus on "value" propositions to unlock their wallet.
2. **The Big Spenders (High Income, High Spend):** The VIP group. They are the most valuable customers and should receive exclusive luxury offers.
3. **The Impulse Buyers (Low Income, High Spend):** Younger people with less money who love to shop. They respond well to sales and discounts.

4. **The Sensible Ones (Low Income, Low Spend):** Careful shoppers with tight budgets.
5. **The Middle Class (Older):** Average income and spending, typically aged 40+.
6. **The Middle Class (Younger):** Average income and spending, but under 30.

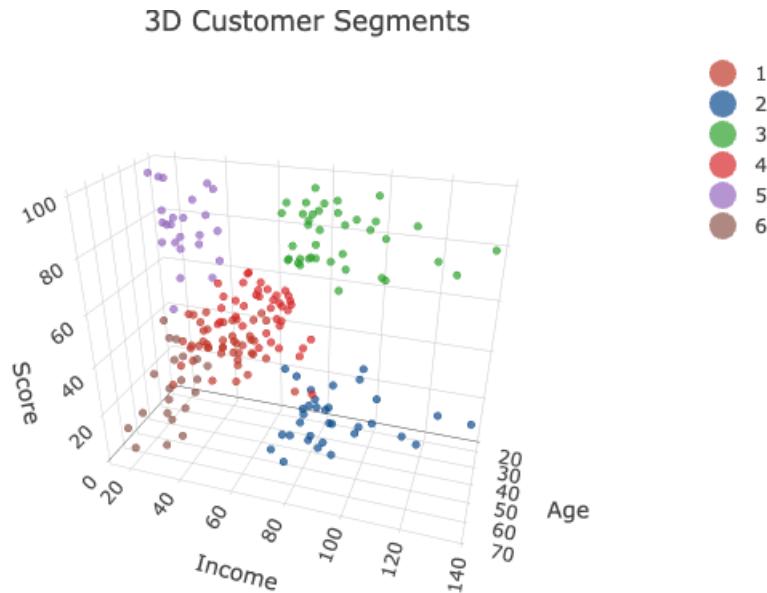


Figure 2: 3D Visualization of Customer Segments.

The 3D visualization (Figure 2) was crucial because it showed that "Average" customers actually split into two groups based on **Age**. A simple 2D analysis would have missed this important detail.

4 Conclusion

This project successfully organized the mall customers into six clear segments. The application allows the marketing team to stop guessing and start targeting. For example, they can send TikTok ads to the "Young Impulse Buyers" and email catalogs to the "Older Middle Class." This data-driven approach saves money and increases sales.

References

- [1] Elsheikh, J. (2024). *Clustering Analysis on Mall Customers*. Kaggle. Available at: <https://www.kaggle.com/code/jeeelsheikh/clustering-analysis-on-mall-customers> [Accessed 22 Jan. 2026].
- [2] Kaggle Datasets. (2016). *Mall Customer Segmentation Data*. Available at: <https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>

- [3] MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*.