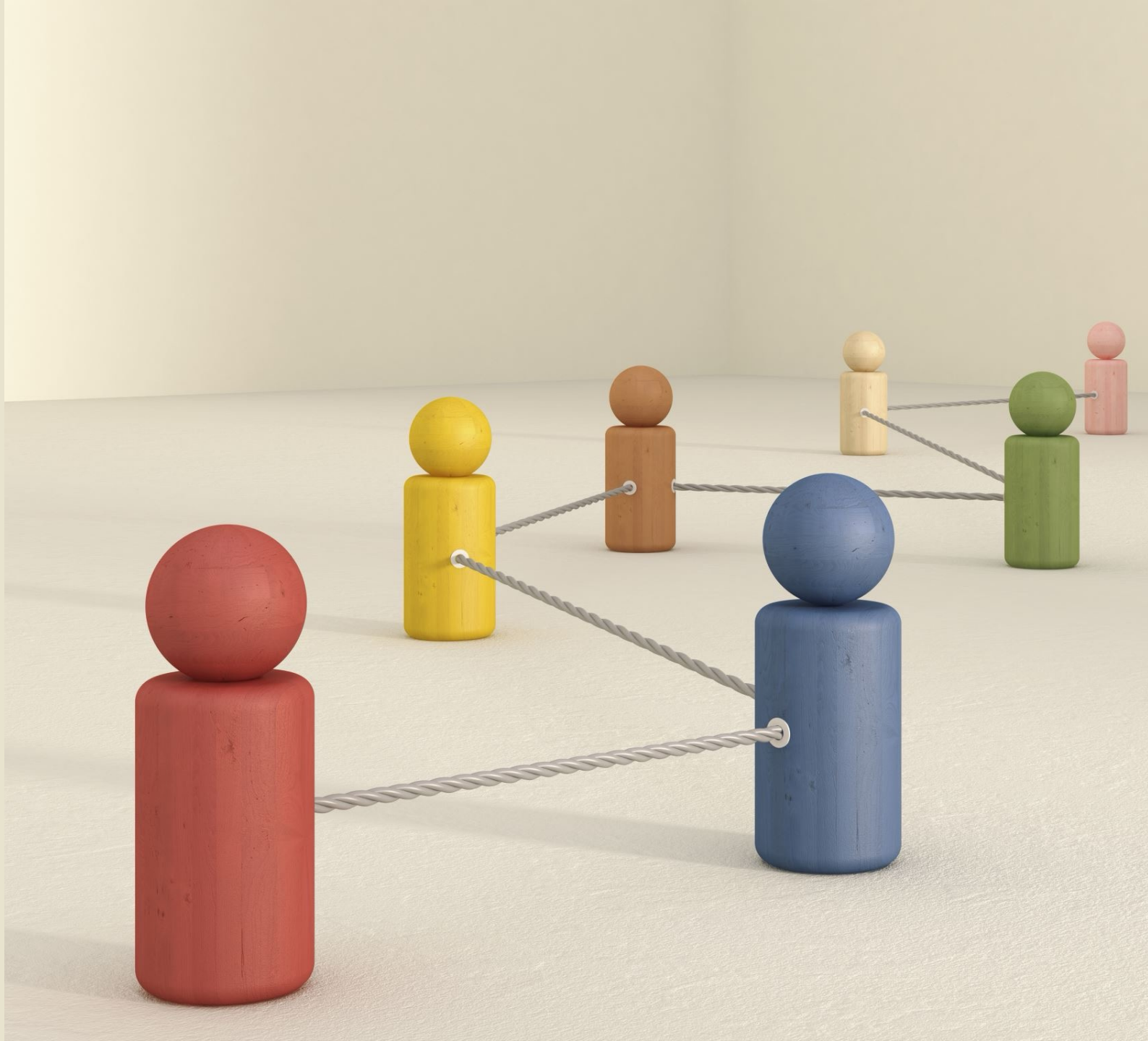
CUSTOMER SEGMENTATION

- **Team Members**
- Harris Hamid
- Bhavya Harshitha Ch
- Bhavana Sriharika K
- Vyshnavi



Introduction

- Customer segmentation is a critical process in marketing and retail analytics. The goal of this project is to analyze customer behavior at a mall and group similar customers together based on their age, annual income, and spending patterns. By using unsupervised learning techniques, we aim to identify patterns and clusters that can guide targeted marketing strategies, product placement, and customer engagement.
- **Problem Statement:**
  How can we segment customers effectively to understand spending behavior and income patterns without prior labels, and which clustering algorithm produces the most meaningful segmentation?
-

BACKGROUND

- Unsupervised learning is a type of machine learning used to identify patterns in datasets without pre-labeled outcomes. Clustering is a common technique in unsupervised learning, grouping similar data points together.
- **Algorithms Used:**
- **K-Means:** Centroid-based, requires specifying number of clusters. Efficient for well-separated, spherical clusters.
- **Hierarchical Clustering:** Builds a dendrogram to merge clusters iteratively. Useful for visualizing cluster relationships.
- **DBSCAN:** Density-based, identifies clusters and outliers, does not require a predefined number of clusters but sensitive to density parameters.
- **Affinity Propagation:** Finds examples automatically, selects clusters without specifying k, based on message passing between points.

METHODOLOGY

Data Preprocessing:

- Loaded Mall_Customers.csv.
- Checked for null values.
- Exploratory analysis using histograms, countplots, and Andrews curves.

Exploratory Data Analysis (EDA):

Studied distribution of Age, Annual Income, and Spending Score by gender.

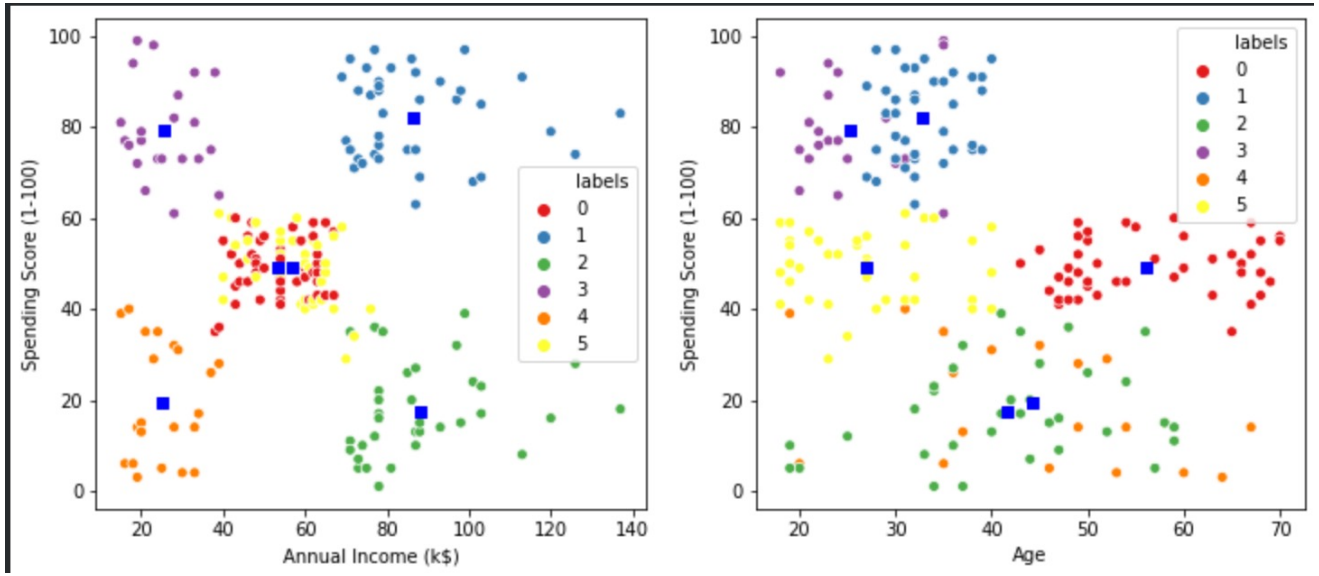
- Analyzed median income and spending by age groups.
- Calculated correlation between Age and Spending Score.

Clustering Analysis:

- **K-Means:**
  - Determined optimal k using Elbow method and Silhouette score.
  - Applied clustering for Age vs Spending Score, then all numeric variables.
- **Hierarchical Clustering:**
  - Built dendrogram using Ward linkage.
  - Compared cluster formations with K-Means.
- **DBSCAN:**
  - Tested multiple combinations of eps and min_samples.
  - Identified outliers and clusters.
- **Affinity Propagation:**
  - Tuned preference parameter.
  - Selected clusters based on silhouette score.

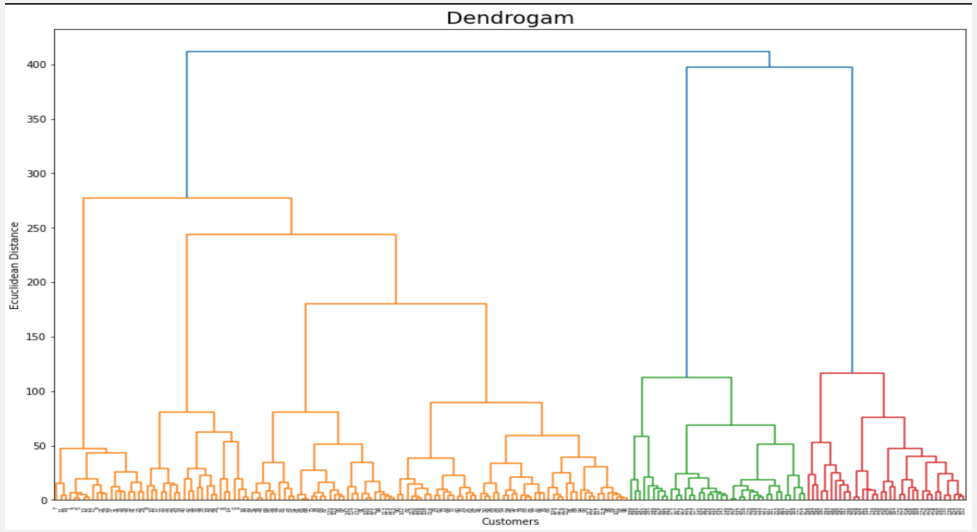
RESULTS

- **K-Means (6 clusters):**
  - Younger clients with medium income and spending.
  - High-income clients with low or high spending.
  - Low-income clients with low or high spending.

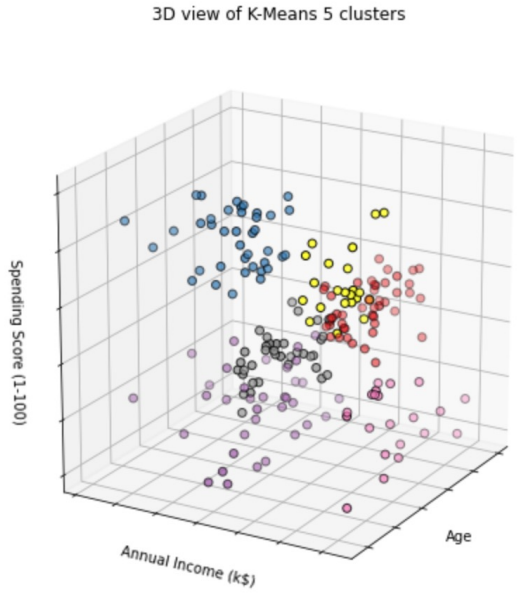


RESULTS

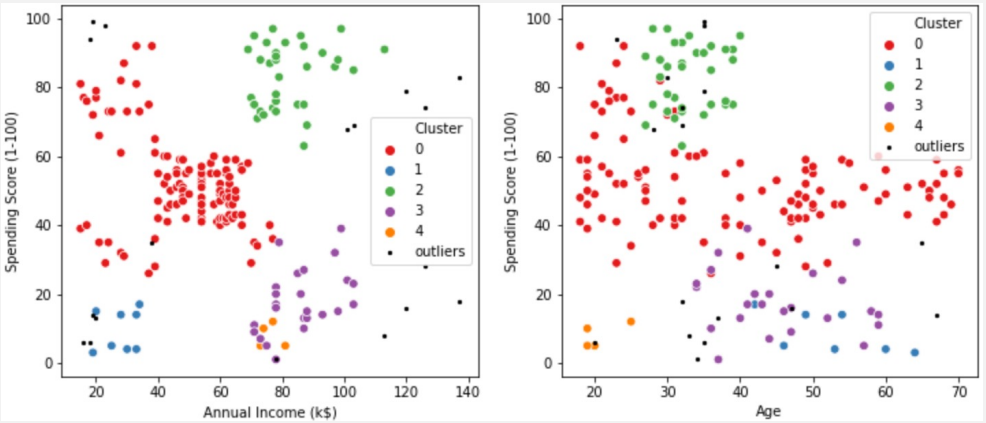
- **Hierarchical Clustering:**
  - Clusters similar to K-Means, Age less significant.



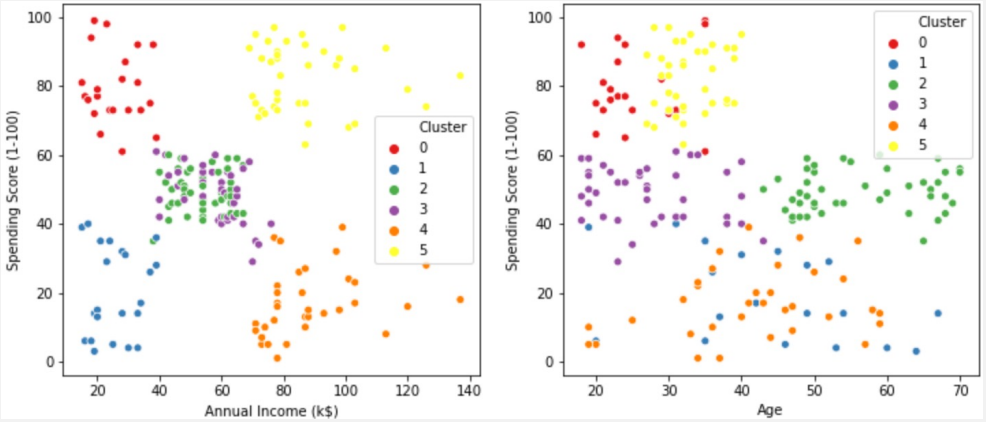
K-Means (5 clusters):
Similar segmentation, slight reduction in cluster granularity.



DBSCAN



Affinity Propagation:



RESULTS

DBSCAN:

Identified 5 clusters with 18 outliers.

Struggled with clusters of varying densities.

Affinity Propagation:

Generated 6 clusters automatically, comparable to K-Means.

INSIGHTS:

Income and Spending Score are stronger drivers for segmentation than Age.

K-Means and Affinity Propagation provide the most meaningful clusters.

DBSCAN has limitations for datasets with mixed-density clusters.

CONCLUSION

- Unsupervised clustering allows effective customer segmentation without prior labels. In this study:
- K-Means and Affinity Propagation produced interpretable and actionable clusters.
- Hierarchical clustering confirms similar patterns but offers dendrogram visualization.
- DBSCAN detected outliers but struggled with mixed densities.
- **Key Takeaways:**
- Marketers can use identified clusters for targeted promotions.
- Age is less predictive of spending behavior than Income and Spending Score.
- Combining multiple clustering methods helps validate results and improves segmentation accuracy.

REFERENCES

- Mall Customer Dataset
- Scikit-learn Documentation (Clustering & GMM)
- Seaborn and Matplotlib Visualization Libraries