



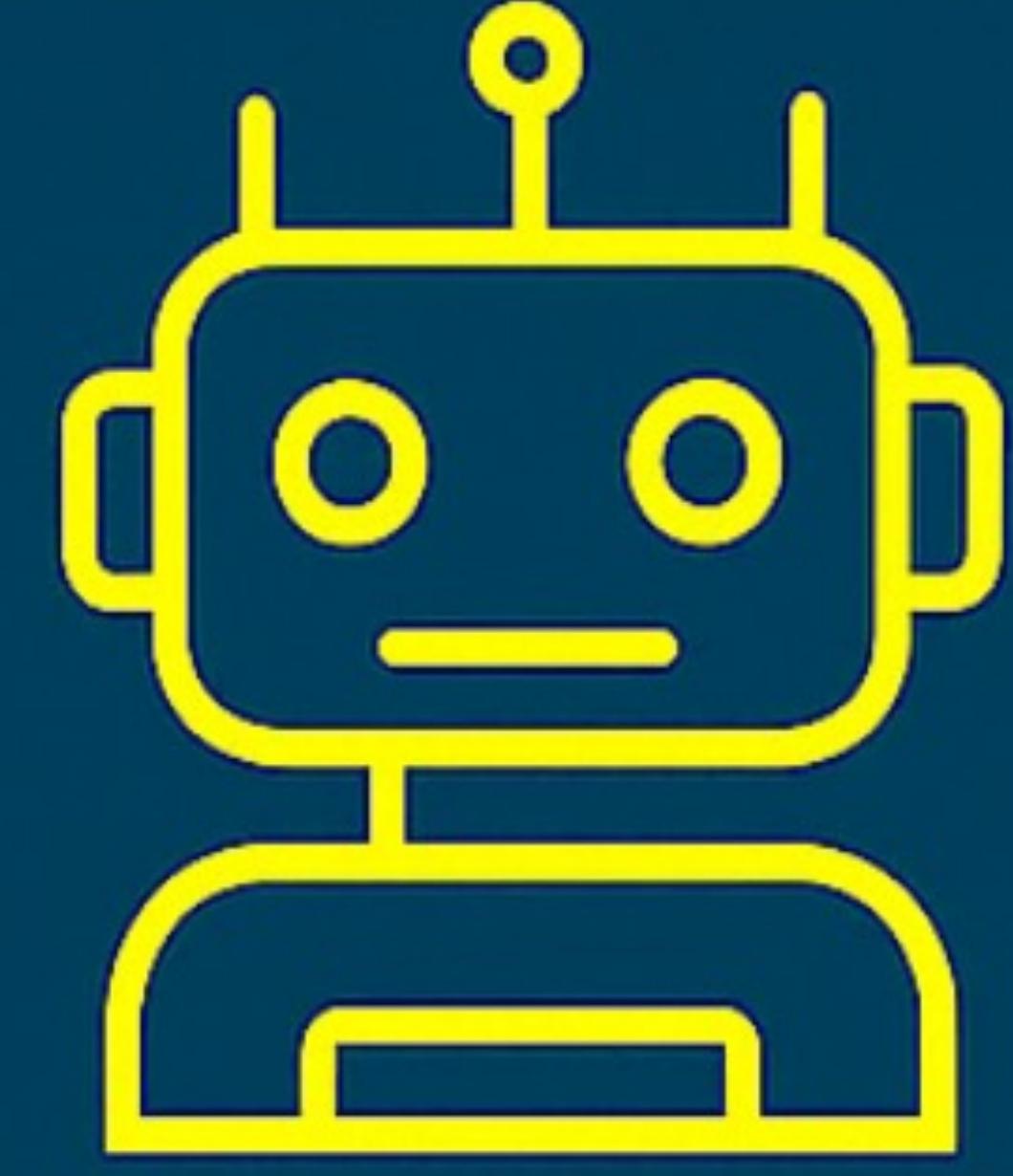
Customer Segmentation Project

Team Name: The Segmennts

Track: AI/Data Science

Team Members: Harris Hamid, Bhavya Harshitha Chennu, Bhavana Sriharika Kondapalli, Muni Vyshnavi Sirimavilla

University of Missouri-Kansas City Hack-A-Roo



Abstract

Customer segmentation helps businesses understand and target customers effectively.

In this study, we applied unsupervised machine learning techniques to group mall customers based on Age, Annual Income, and Spending Score.

Using K-Means, Hierarchical Clustering, DBSCAN, and Affinity Propagation, we identified meaningful clusters that reflect different spending behaviors.

Results show that Income and Spending Score are the most important factors influencing customer segmentation, enabling more personalized marketing strategies.

Introduction

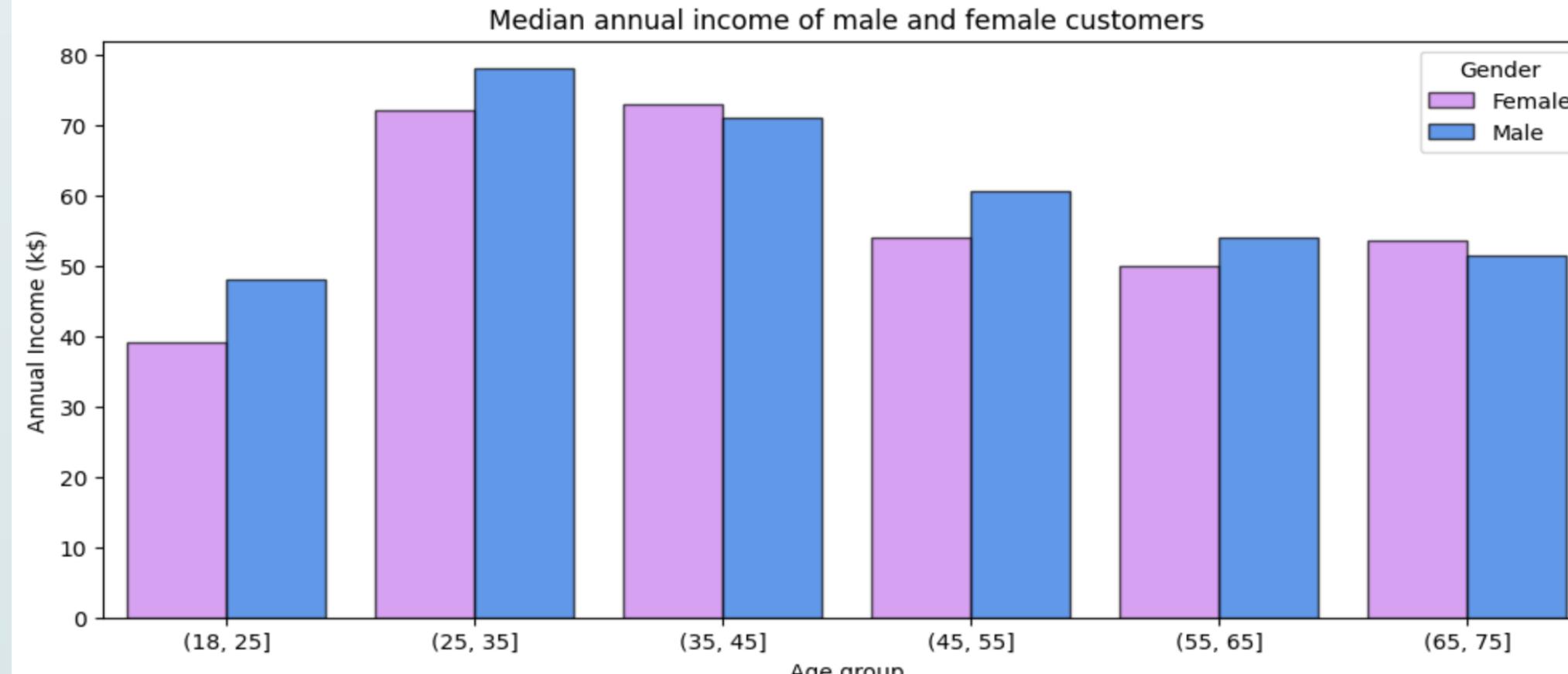
Understanding customer behavior is vital for modern retail and marketing analytics.

Segmenting customers allows businesses to tailor promotions and product placement based on distinct behavioral patterns.

This project aims to find the best clustering method for segmenting mall customers without predefined labels.

Problem Statement:

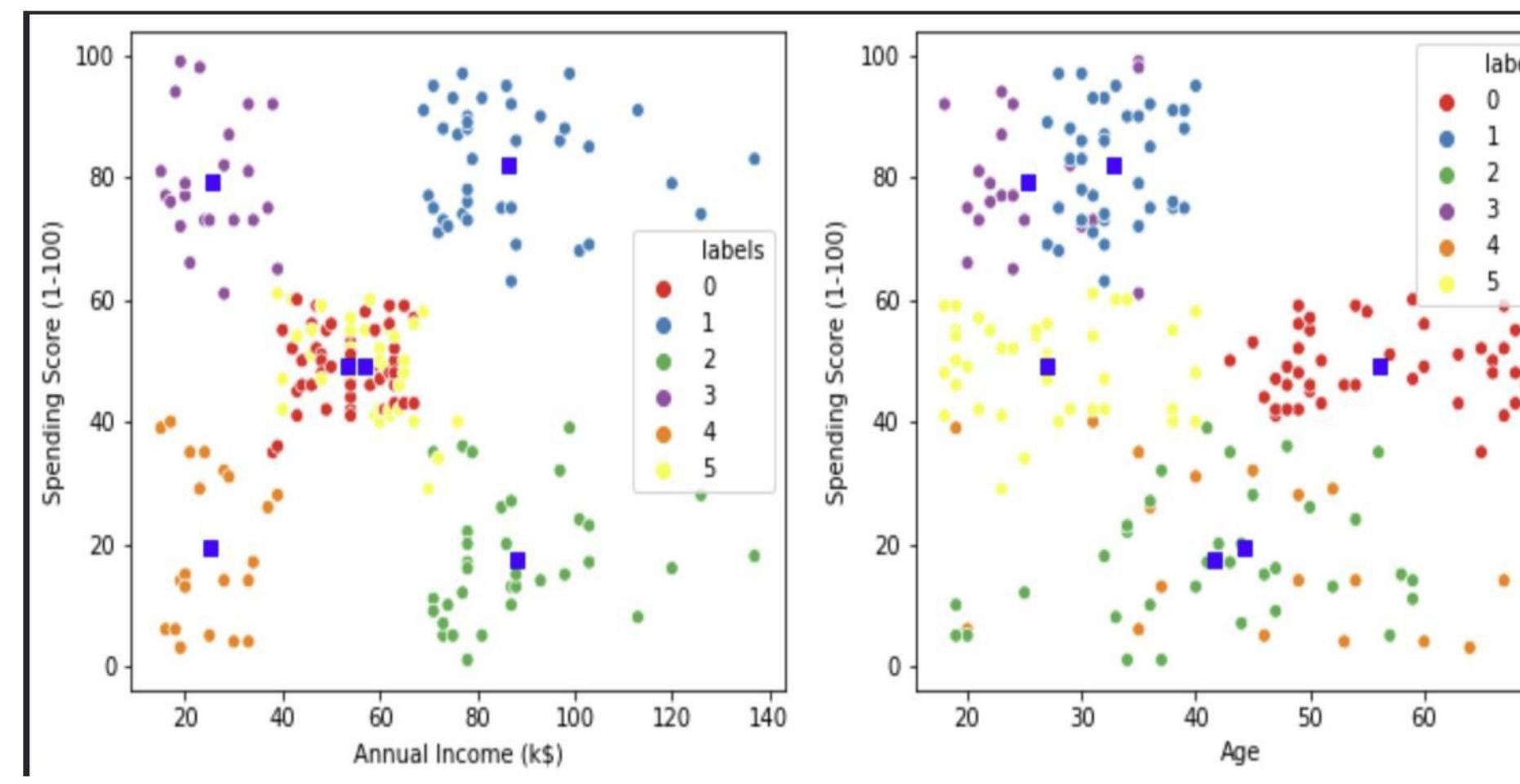
How can customers be effectively segmented to understand spending patterns and income levels using unsupervised learning?



Methodology

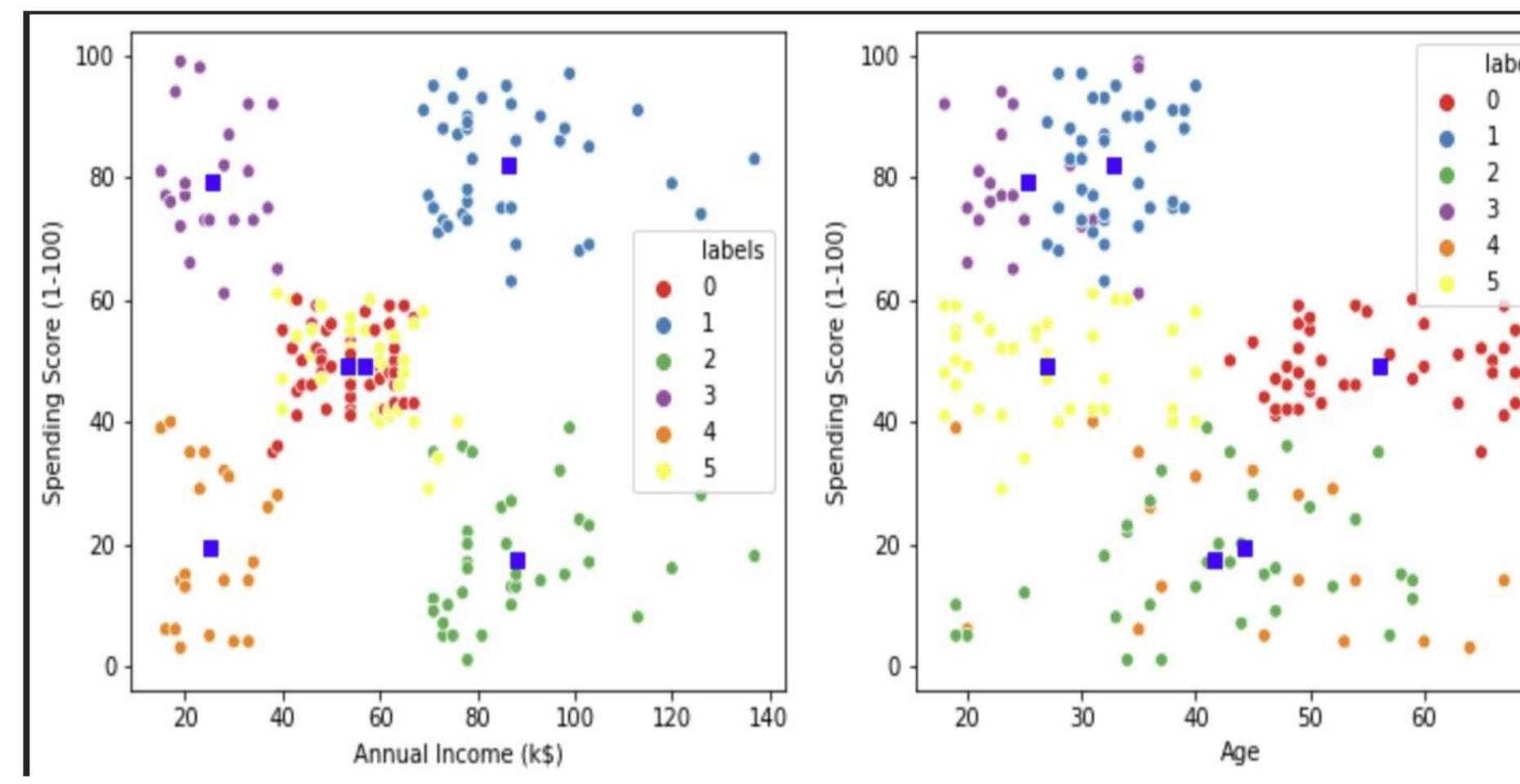
1. Data Preprocessing

- Loaded and cleaned dataset, checked for null values
- Analyzed distribution of Age, Income, and Spending Score
- Created visualizations using histograms and countplots.

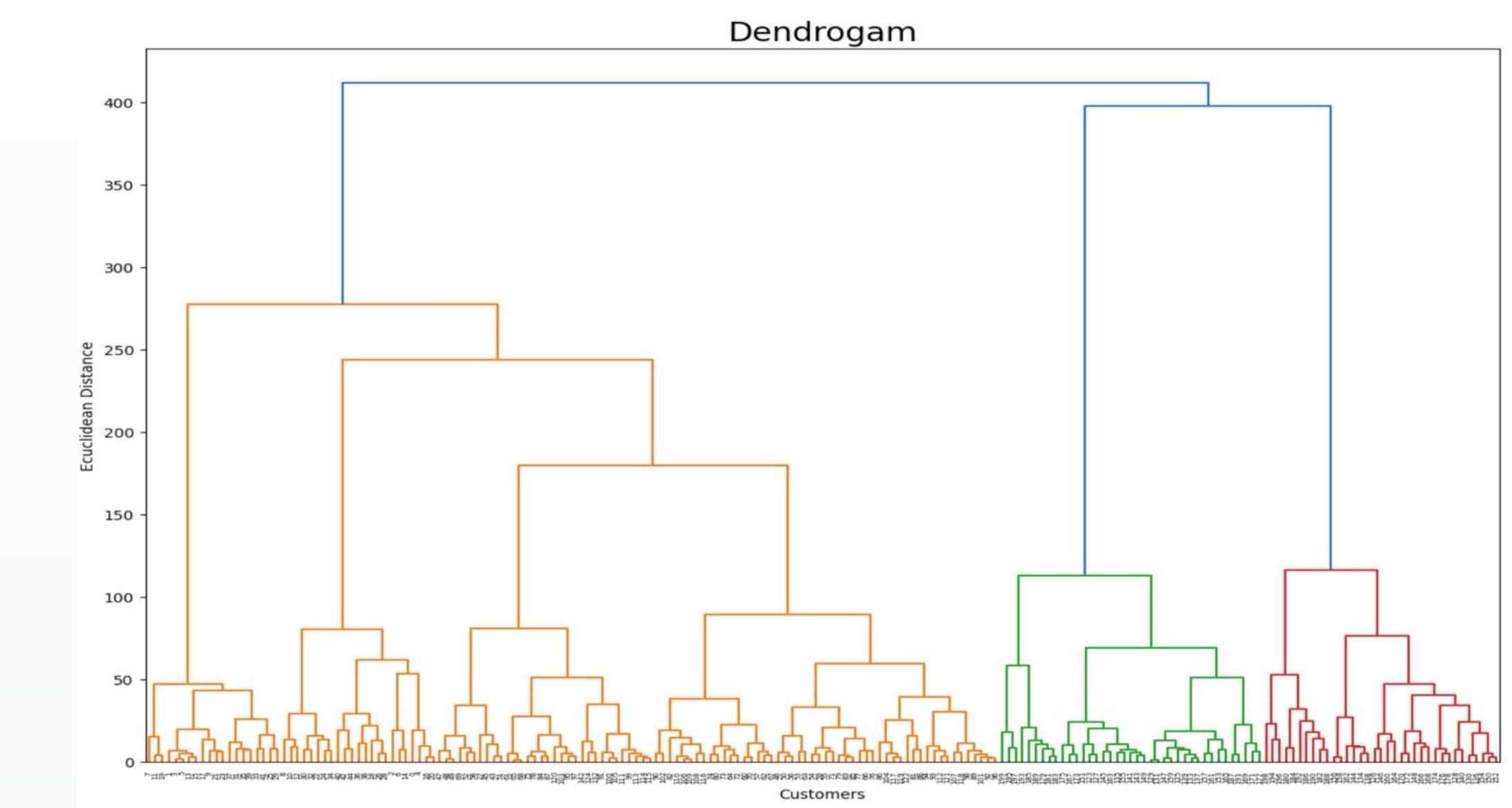


2. Clustering Techniques

- K-Means: Determined optimal k using Elbow and Silhouette methods.
- Hierarchical Clustering: Used Ward linkage and dendrograms for visual analysis.
- DBSCAN: Tuned eps and min_samples to identify density-based clusters and outliers.
- Affinity Propagation: Automatically identified number of clusters by tuning preference.



- K-Means Clustering:** Formed five clear customer groups ranging from low-income, low-spenders to affluent, high-spending consumers.

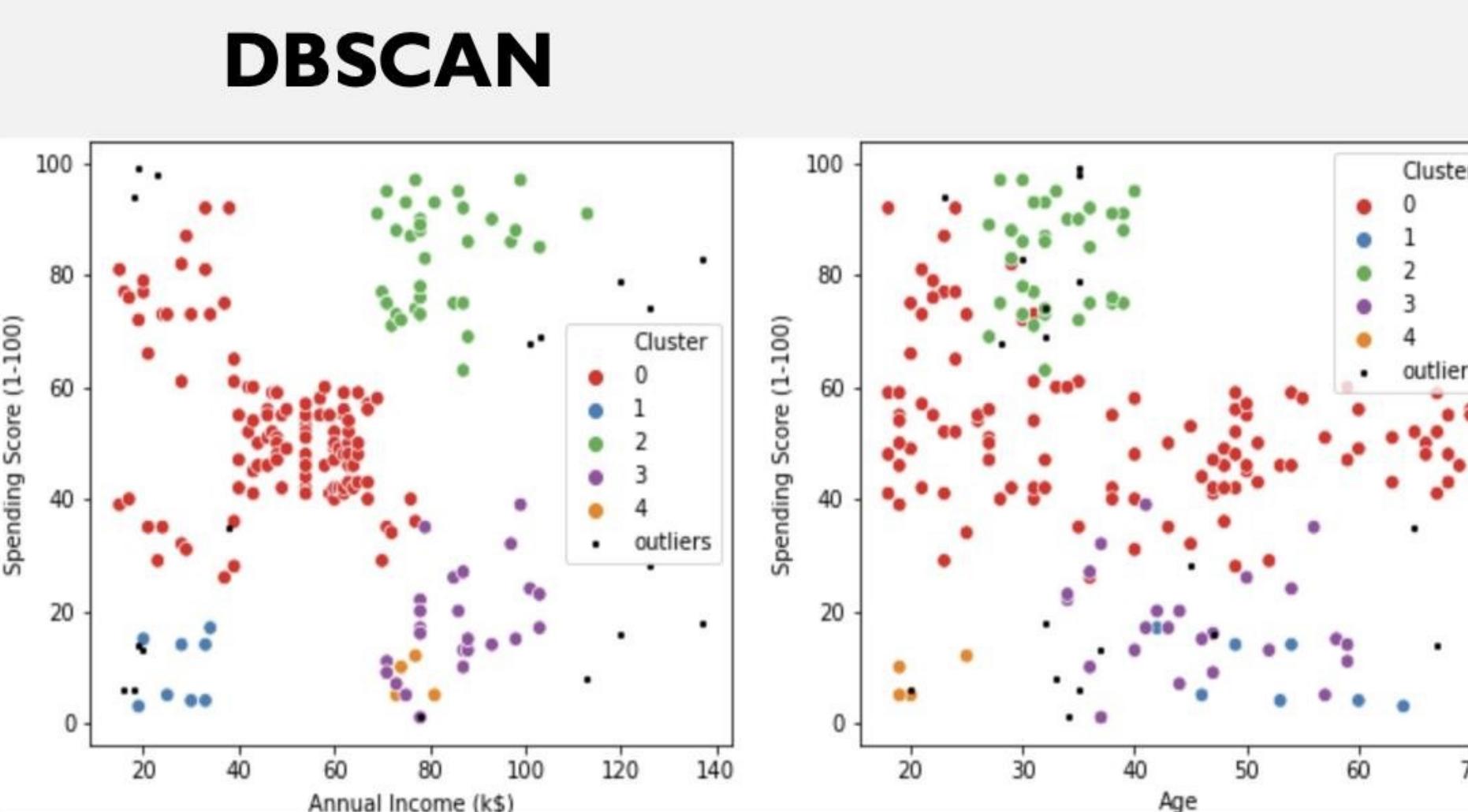


- Hierarchical Clustering:** Confirmed similar patterns, illustrating cluster relationships through a dendrogram.

The analysis revealed that customer segmentation is primarily driven by income and spending patterns, while age plays a lesser role. Among the clustering methods, K-Means and Affinity Propagation provided the most interpretable and well-separated clusters, capturing distinct consumer groups based on spending habits. Hierarchical clustering supported similar patterns with visual clarity, whereas DBSCAN was effective for detecting outliers but less suited for varied-density data. Overall, the results highlight the effectiveness of unsupervised learning in uncovering actionable customer insights.

3. Evaluation

- Compared cluster interpretability and silhouette scores across all models.

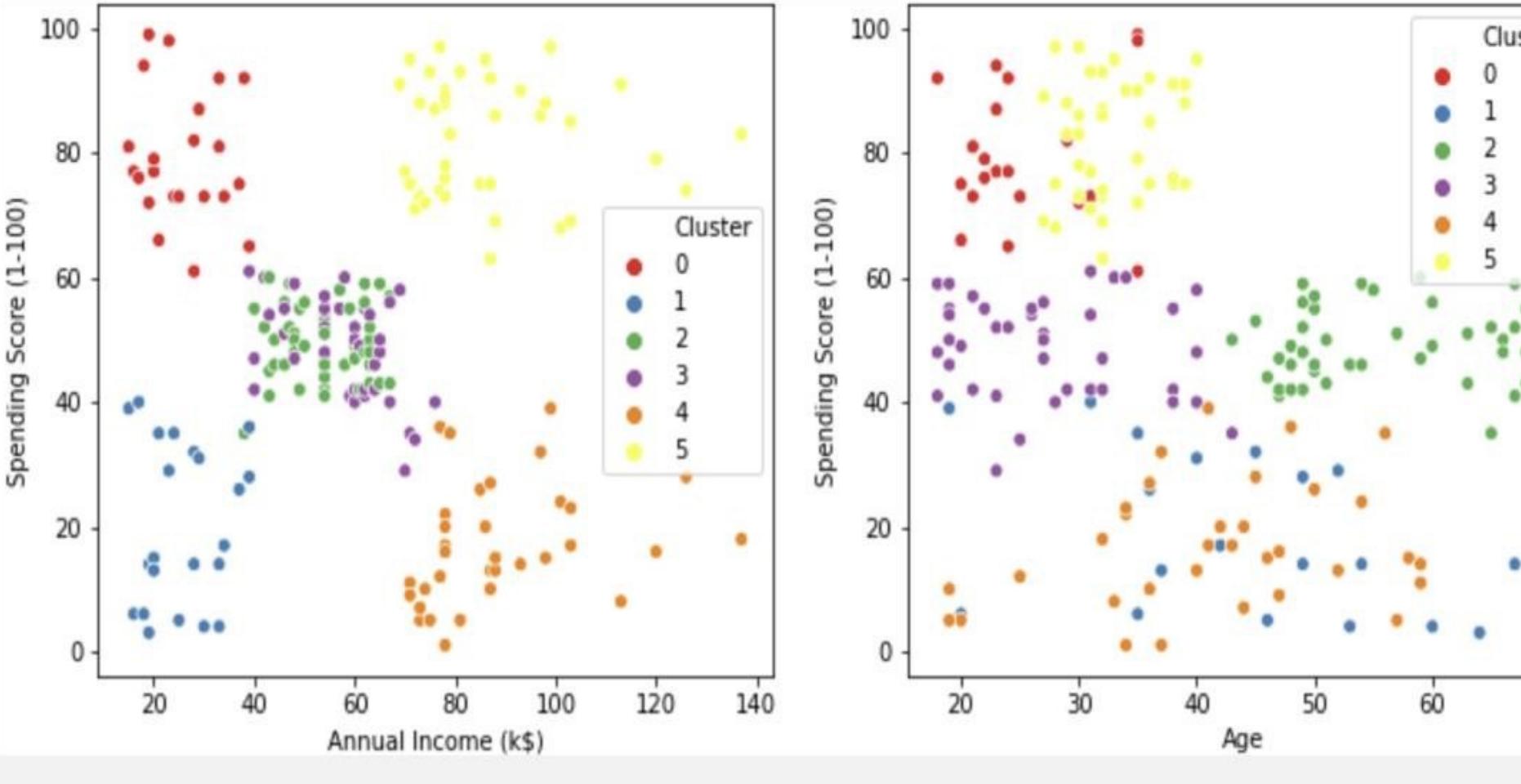


Results

- Overall, income and spending score showed the strongest influence on segmentation outcomes, highlighting young adults (20–40 years) with higher income and spending scores as the most profitable group.
- Income and Spending Score are stronger indicators of customer behavior than Age.
- K-Means and Affinity Propagation gave the most meaningful and consistent results.
- DBSCAN effectively detected anomalies but less stable on mixed-density data.
- Hierarchical Clustering validated other models and added interpretability through visualization.

DBSCAN: Identified outliers effectively but struggled with clusters of mixed density.

Affinity Propagation



Affinity Propagation: Produced interpretable clusters that reinforced the validity of K-Means results

Conclusion

- Unsupervised clustering successfully segmented customers without prior labels.
- K-Means and Affinity Propagation produced the most interpretable and stable clusters, while Hierarchical Clustering confirmed their structure visually.
- DBSCAN was useful for identifying outliers but less effective for mixed-density data.

Takeaways:

- Marketers can target promotions more accurately using income-spending-based clusters.
- Age is less predictive of spending behavior.
- Combining multiple clustering methods enhances segmentation reliability.

Acknowledgements

We would like to express our sincere gratitude to the University of Missouri-Kansas City for providing the platform and resources to participate in Hack-A-Roo.