



CUSTOMER SEGMENTATION

ABSTRACT

This project analyses customer segmentation using visualizations and clustering algorithms to uncover distinct customer segments based on gender, age, annual income, and spending score.

Spandan Ghosh

Introduction

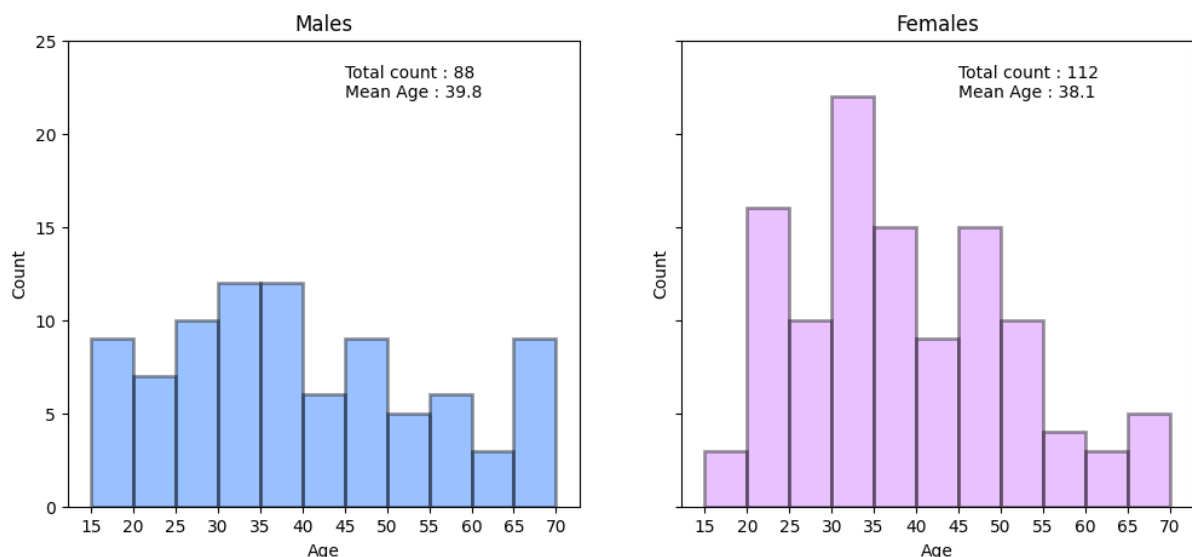
Customer segmentation is a crucial task for businesses as it helps identify distinct groups of customers with similar characteristics and behaviours. This report presents an analysis of customer segmentation. The goal is to explore various visualization techniques and apply clustering algorithms to uncover meaningful customer segments.

Data Description

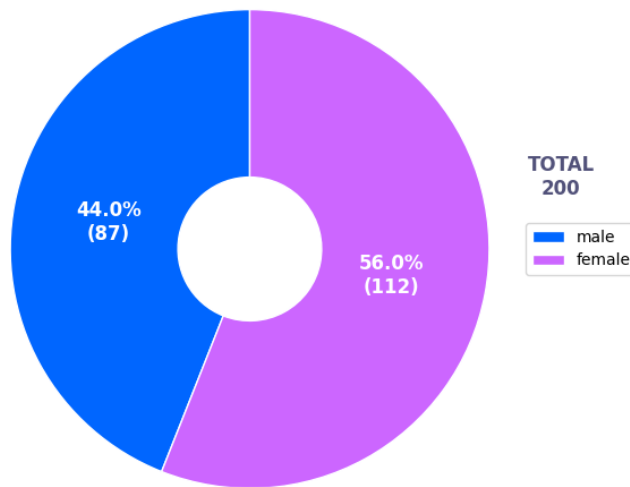
The dataset used for this analysis contains information on customers, including their CustomerID, Gender, Age, Annual Income, and Spending Score. The CustomerID serves as a unique identifier, while Gender indicates the customer's gender. Age represents the customer's age, Annual Income represents their yearly income, and Spending Score denotes their spending behaviour.

Exploratory Data Analysis

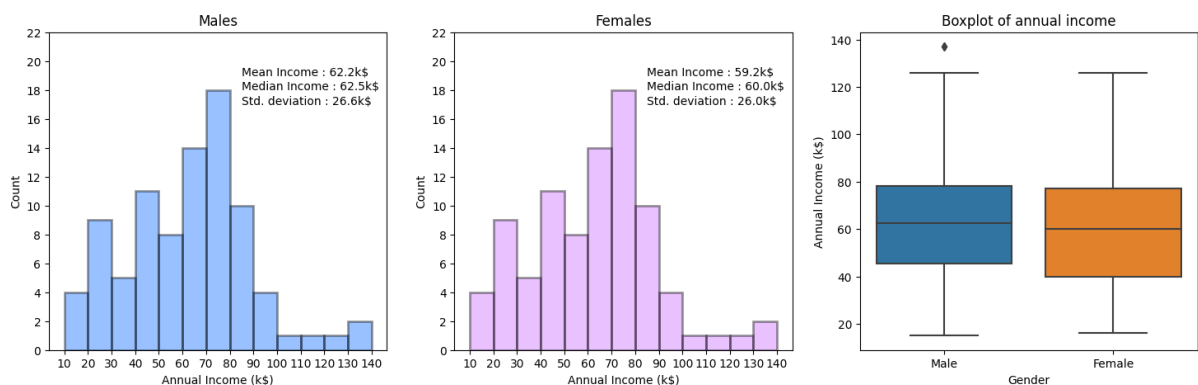
- i. Bar Chart Comparing Mean Age between Male and Female Customers: This visualization provides a comparison of the average age for male and female customers.



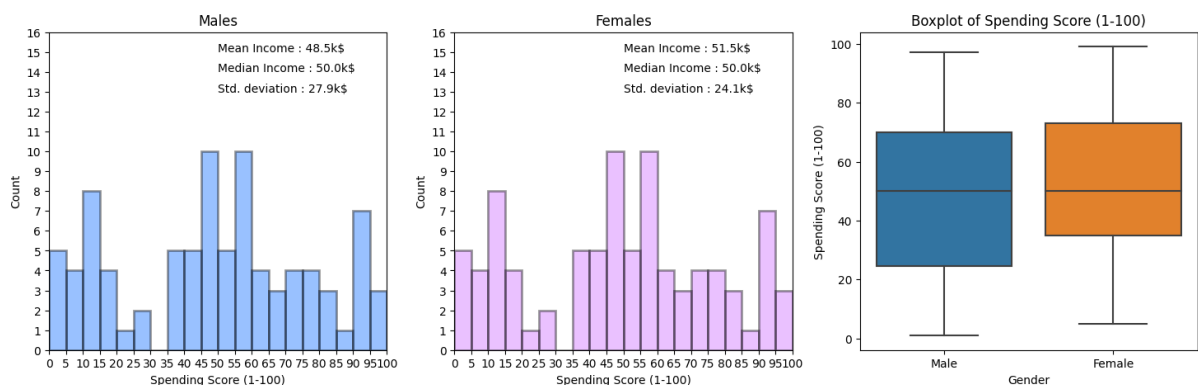
- ii. Pie Chart Stating the Percentages of Male and Female Customers: This pie chart displays the distribution of male and female customers as a percentage of the total.



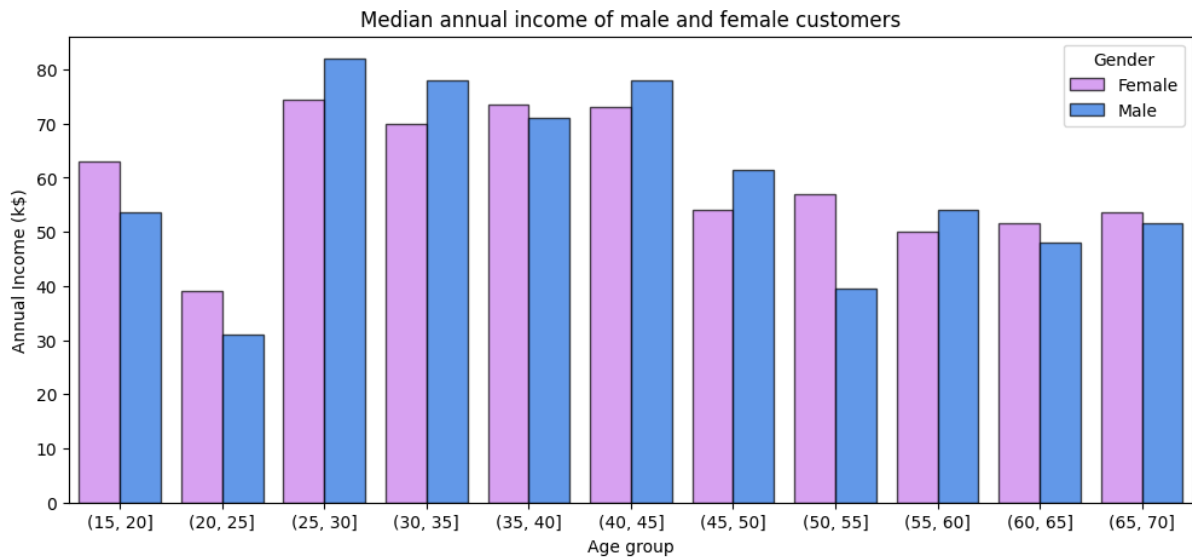
- iii. Bar Chart and Box Plot Comparing Male and Female Based on Income: This visualization compares the mean income, median income, and standard deviation of male and female customers.



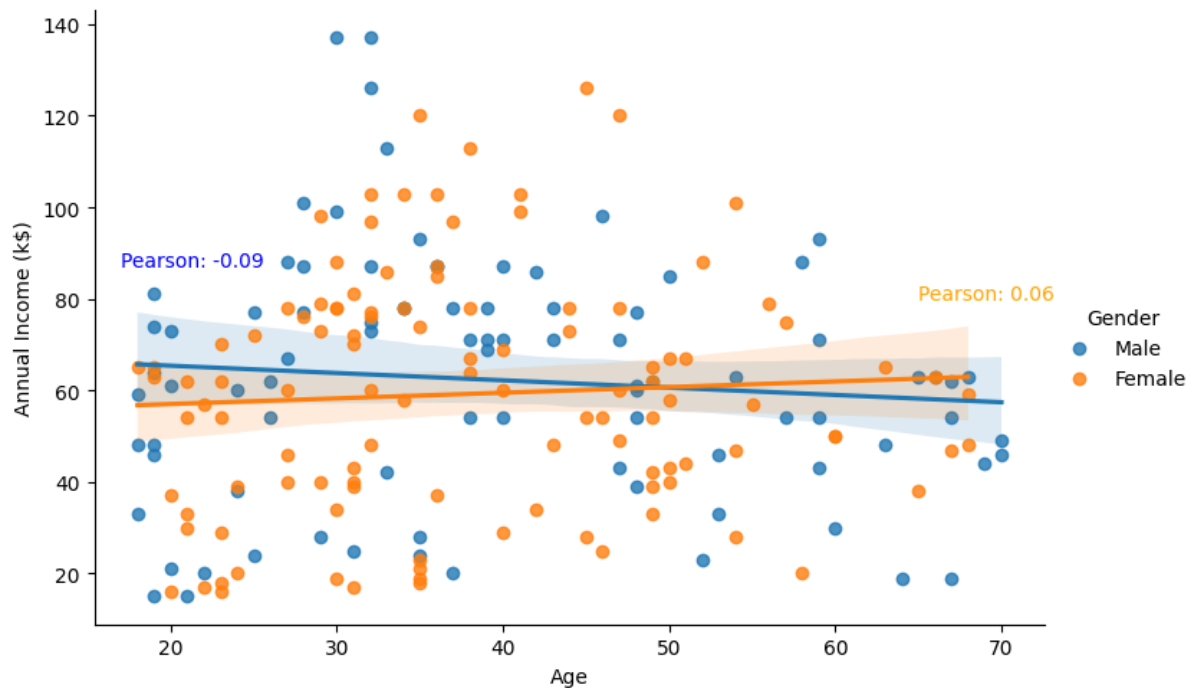
- iv. Bar Chart and Box Plot Comparing Male and Female Based on Spending Score: This visualization compares the mean spending score, median spending score, and standard deviation of male and female customers.



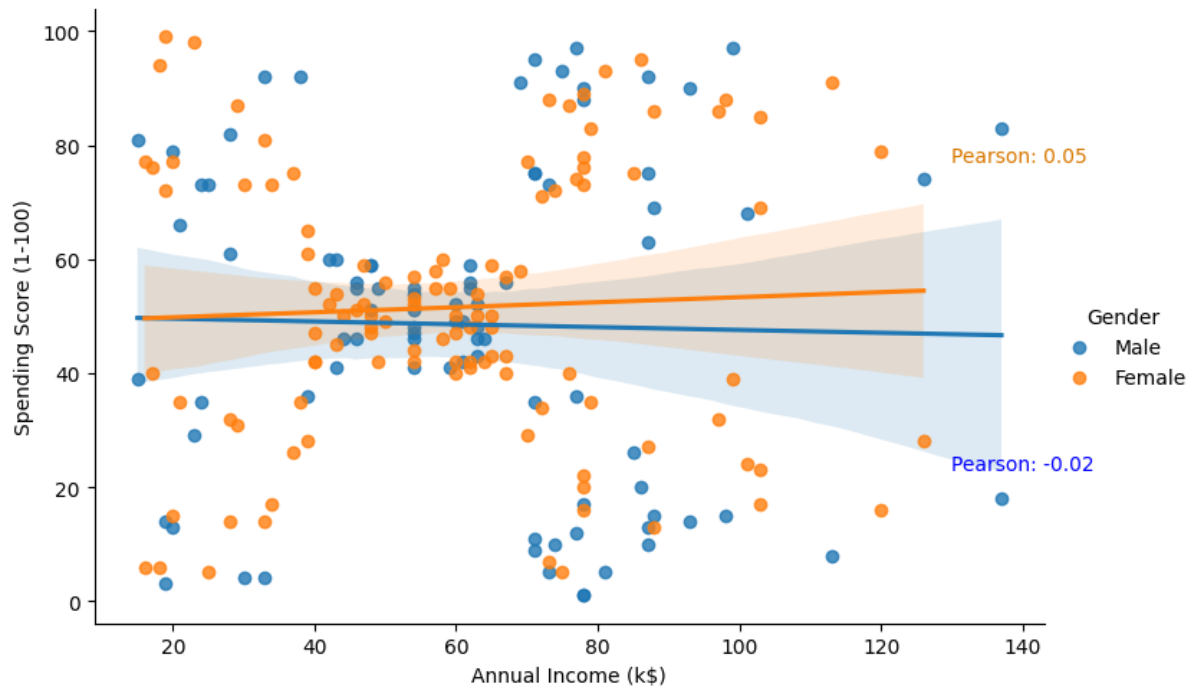
- v. Multiple Bar Chart Comparing Median Annual Income of Male and Female Customers: This chart showcases a comparison of the median annual income for male and female customers.



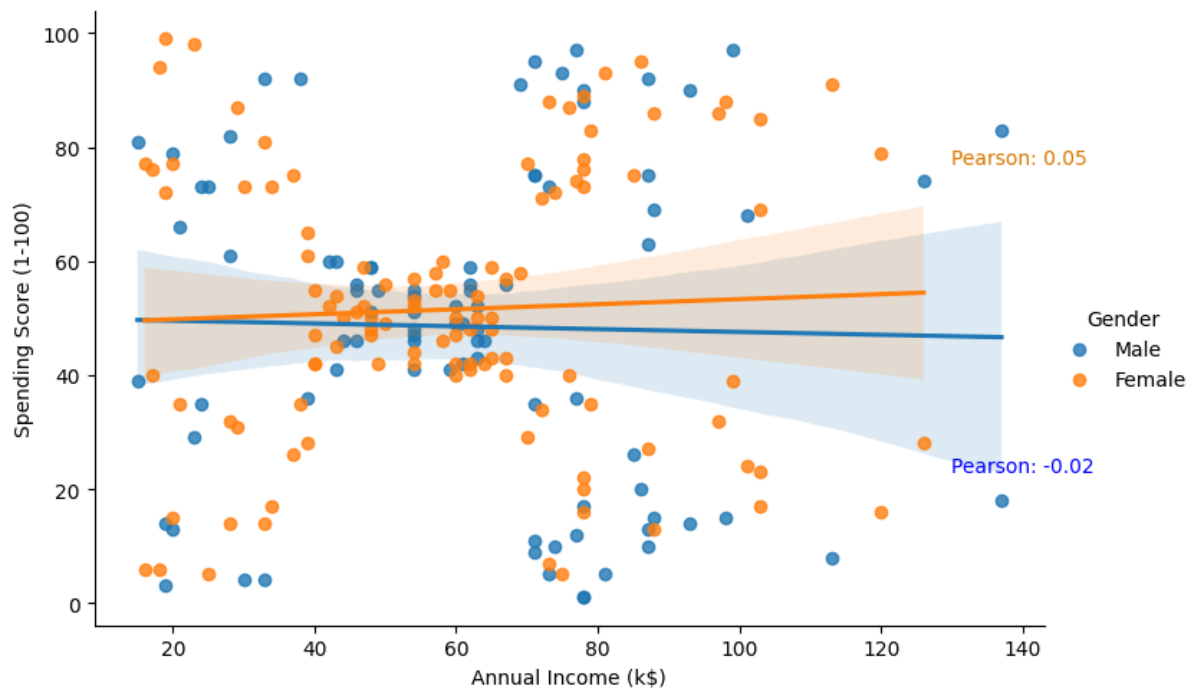
- vi. Pearson Correlation of Annual Income and Age between Male and Female: This analysis calculates the Pearson correlation coefficient between annual income and age for male and female customers.



- vii. Pearson Correlation of Spending Score and Age between Male and Female: This analysis calculates the Pearson correlation coefficient between spending score and age for male and female customers.



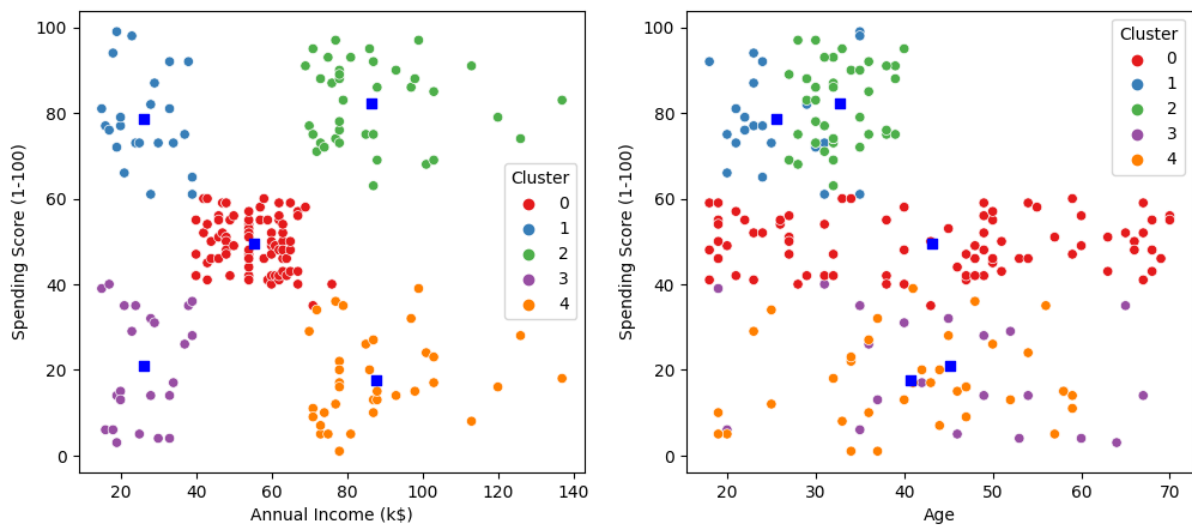
- viii. Pearson Correlation of Annual Income and Spending Score between Male and Female: This analysis calculates the Pearson correlation coefficient between annual income and spending score for male and female customers.



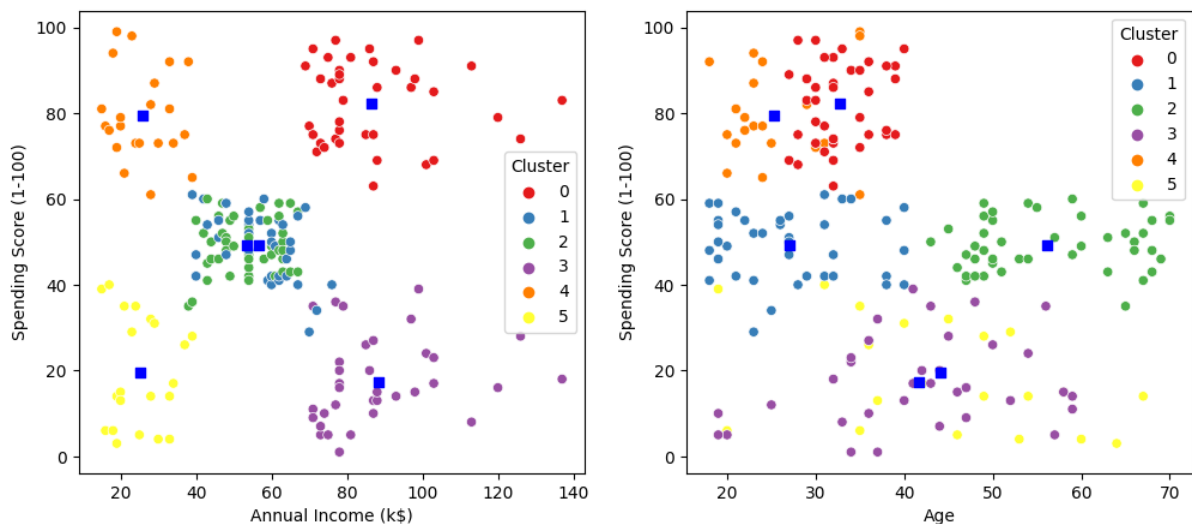
Customer Segmentation Algorithms

1. K-means: K-means is a clustering algorithm that aims to partition data points into K clusters. By using the elbow method and silhouette score, we determined that the optimal number of clusters for our dataset is 5-6. We plotted a 2D plot with annual income and spending score as axes and another 2D plot with spending score and age. Additionally, we created a 3D plot with spending score, annual income, and age.

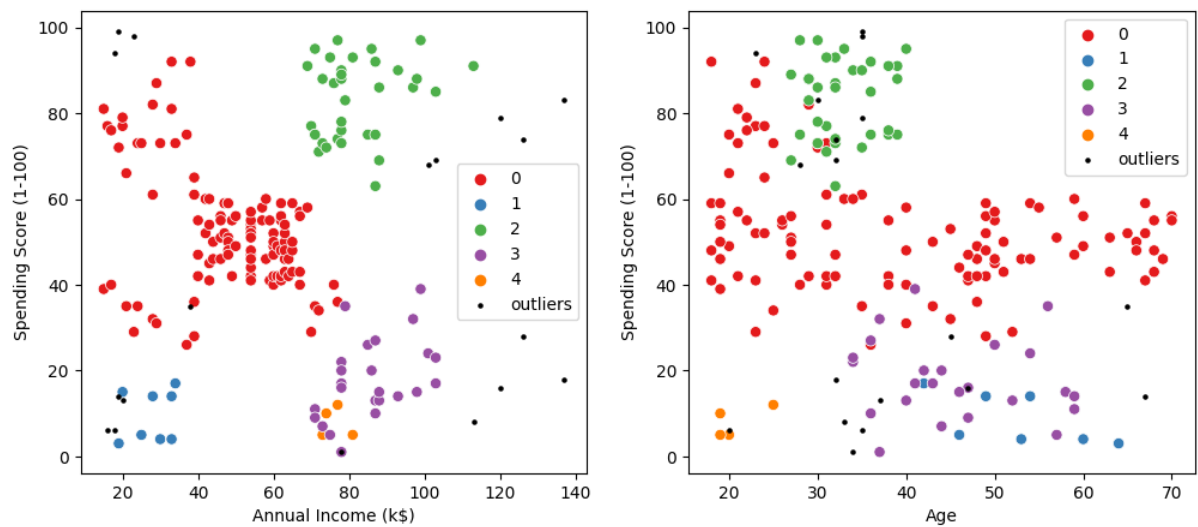
5 Clusters:



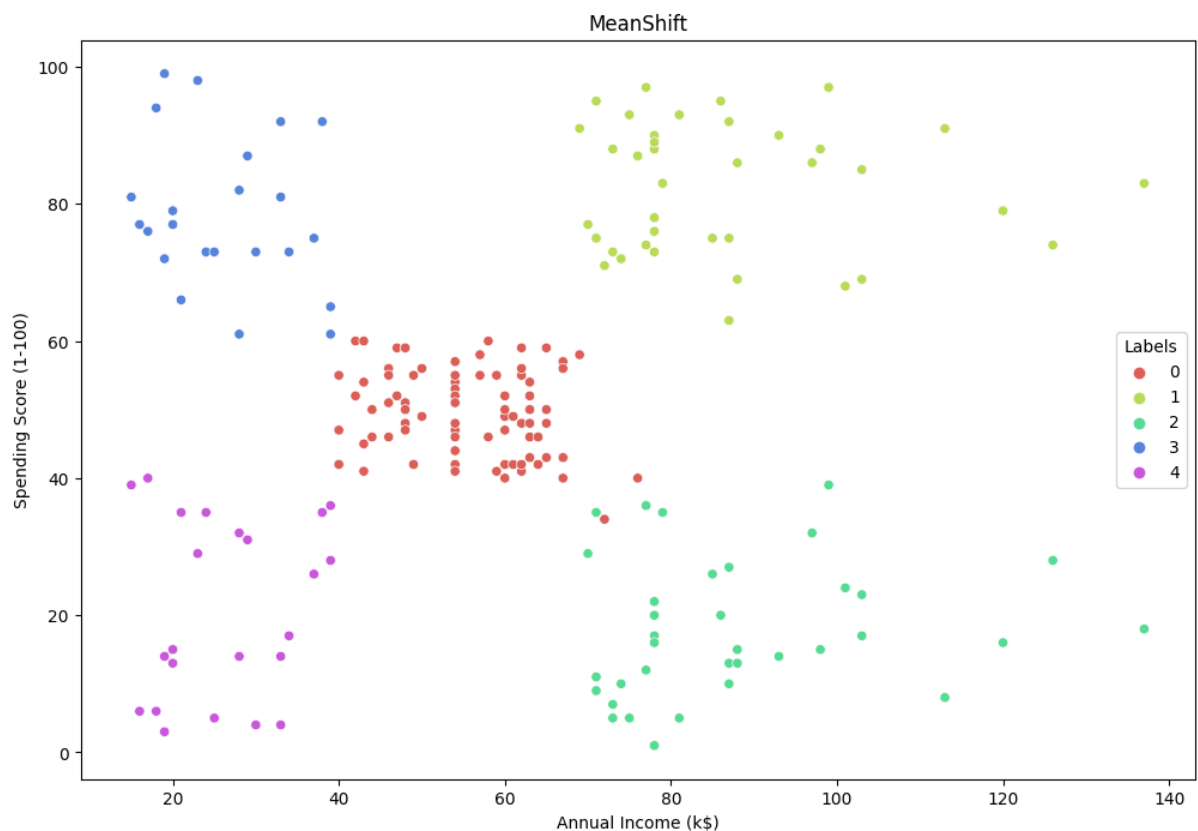
6 Clusters:



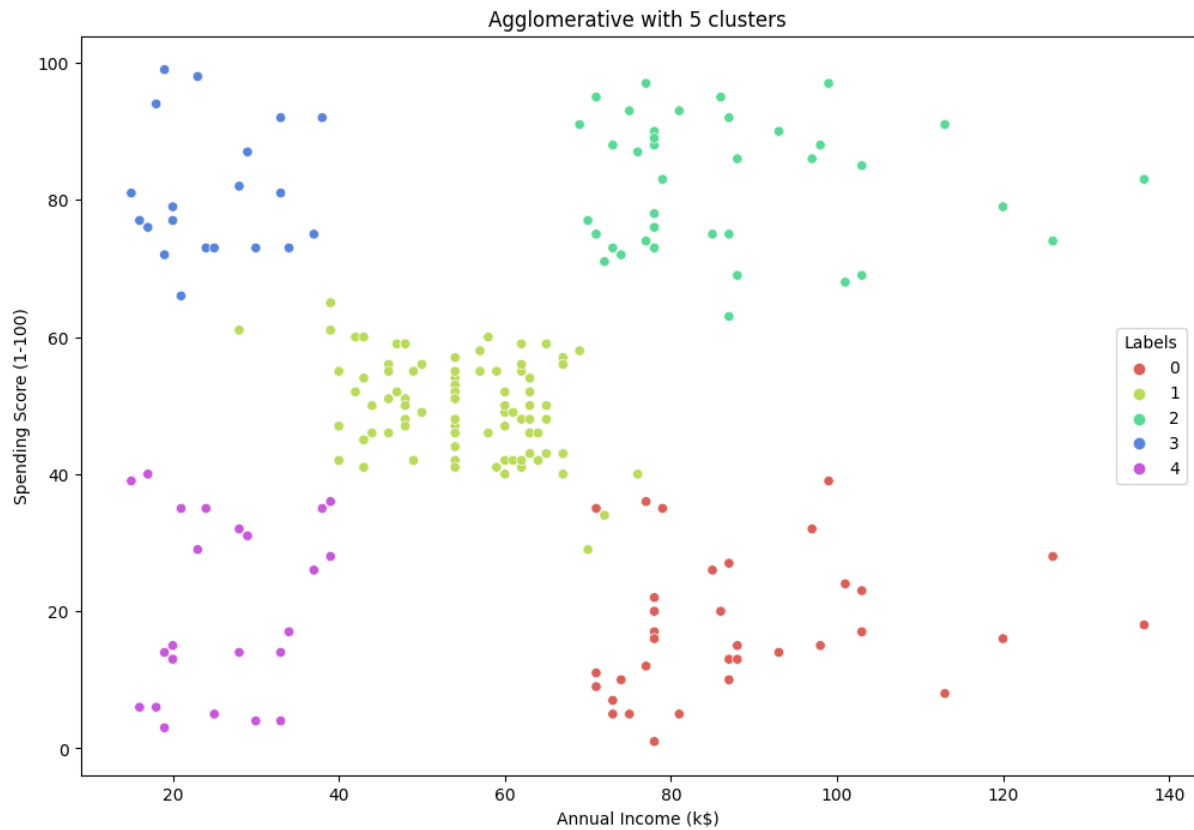
2. DBSCAN: DBSCAN is a density-based clustering algorithm that groups together data points based on their density. By applying this algorithm, we found that the required number of clusters for our dataset is 5. We created a plot representing the resulting clusters.



3. MeanShift: MeanShift is an iterative algorithm that discovers clusters by estimating the density and finding modes in the data. We estimated the bandwidth and plotted the resulting 5 clusters using this algorithm.



4. Agglomerative Clustering: Agglomerative clustering is a hierarchical clustering algorithm that merges data points into clusters based on proximity. We plotted the five resulting clusters on the same graph and created a dendrogram using the complete linkage method.



Conclusion

Customer segmentation is a powerful technique that enables businesses to gain insights into their customer base. In this project, we explored various visualizations to understand the distribution and relationships between different customer attributes. We applied four clustering algorithms - K-means, DBSCAN, MeanShift, and Agglomerative Clustering - to identify distinct customer segments based on their characteristics. These segments can help businesses personalize their marketing strategies, enhance customer satisfaction, and improve overall business performance.