Student ID: 23036717
Student Name: Jafar Urchintala
GitHub Link: https://github.com/Jafarurchintala/ADS-Assignment-Clustering-and-Fitting.git

Customer Segmentation and Analysis

Introduction:

In this report, we conduct an analysis of customer data from a mall to uncover insights and patterns that can inform segmentation strategies and enhance marketing efforts.

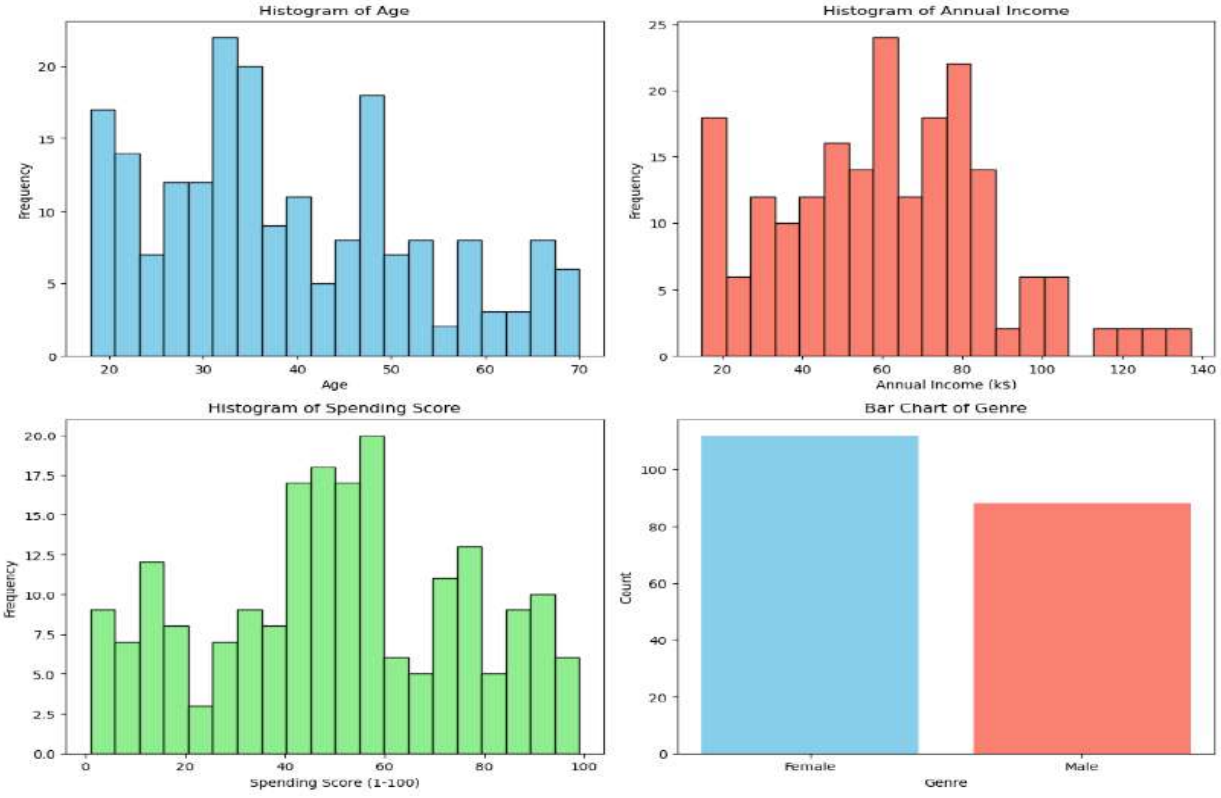
Data Overview:

The dataset contains information about customers, including their CustomerID, Genre, Age, Annual Income and Spending Score. Before proceeding with the analysis, we conducted initial data exploration to understand its structure and characteristics. The dataset consists of 200 entries and does not contain any missing values.

Descriptive Analysis:

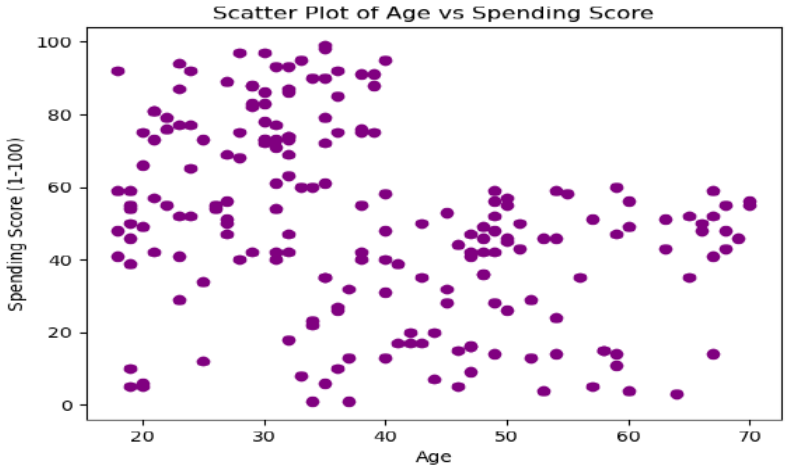
We began by calculating statistical moments, including mean, median, standard deviation, skewness, and kurtosis, for key variables. The distributional features, central tendency, and variability of the data are all revealed by these statistics.

	Mean	Median	Standard Deviation	Skewness	Kurtosis
Annual Income (k$)	60.56	61.5	26.264721	0.321843	-0.098487
Spending Score (1-100)	50.20	50.0	25.823522	-0.047220	-0.826629
Age	38.85	36.0	13.969007	0.485569	-0.671573



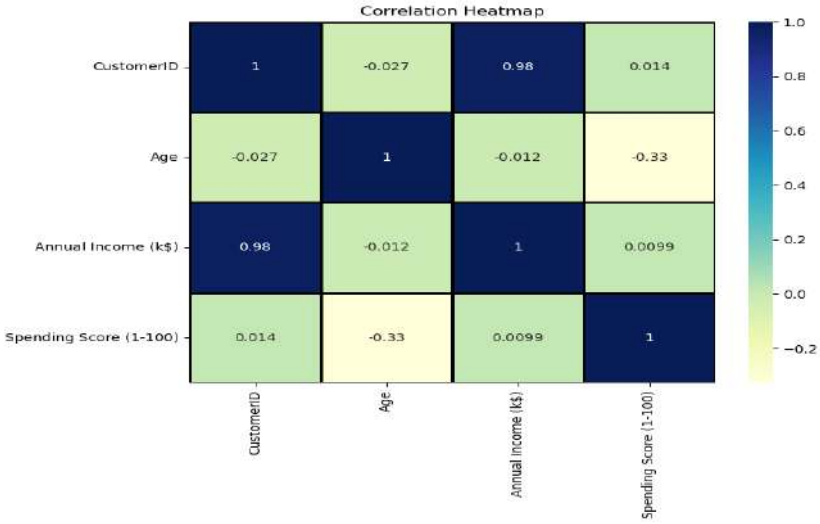
We visualized distributions of Age, Annual Income, and Spending Score using histograms. These visualizations allow us to understand the frequency distribution of each variable and identify any patterns or outliers. Furthermore, we created a bar chart to visualize the distribution of Genre (gender) among customers.

The histograms reveal that Age is relatively evenly distributed, with a slight skew towards younger customers. Annual Income shows a broader distribution, with a peak around $50,000. Spending Score appears to have a bimodal distribution, with peaks at low and high spending scores, indicating the presence of distinct customer segments. And the bar chart depicting the distribution of Genre (gender) among customers shows a relatively equal distribution between male and female customers.



A scatter plot of Age vs. Spending Score was generated to explore the relationship between these two variables. This visualization helps us understand if there is any discernible pattern or correlation between age and spending behaviour.

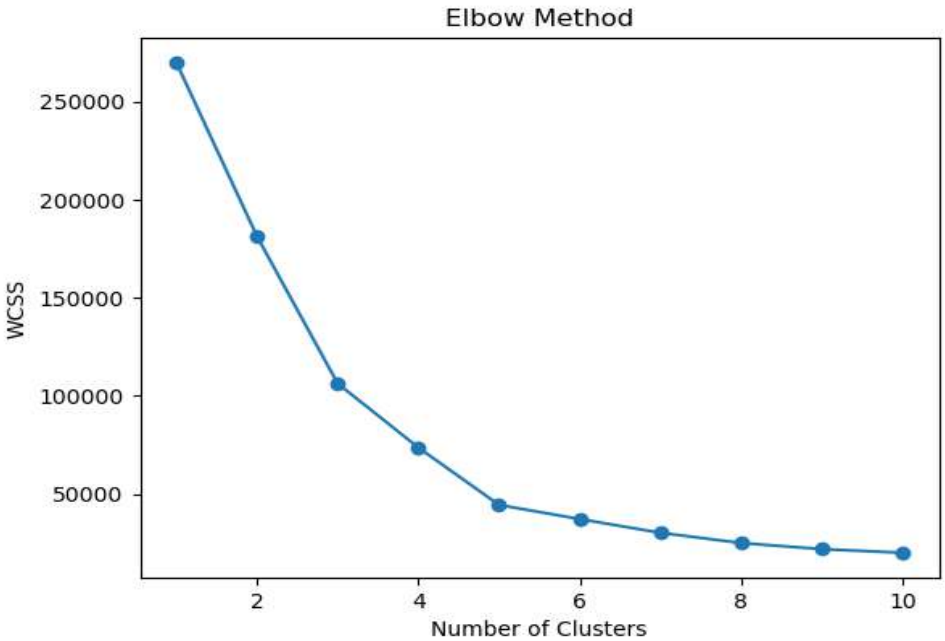
It appears that age may not be a reliable indicator of spending behaviour on its own because the scatter plot does not clearly indicate a linear link between spending score and age.



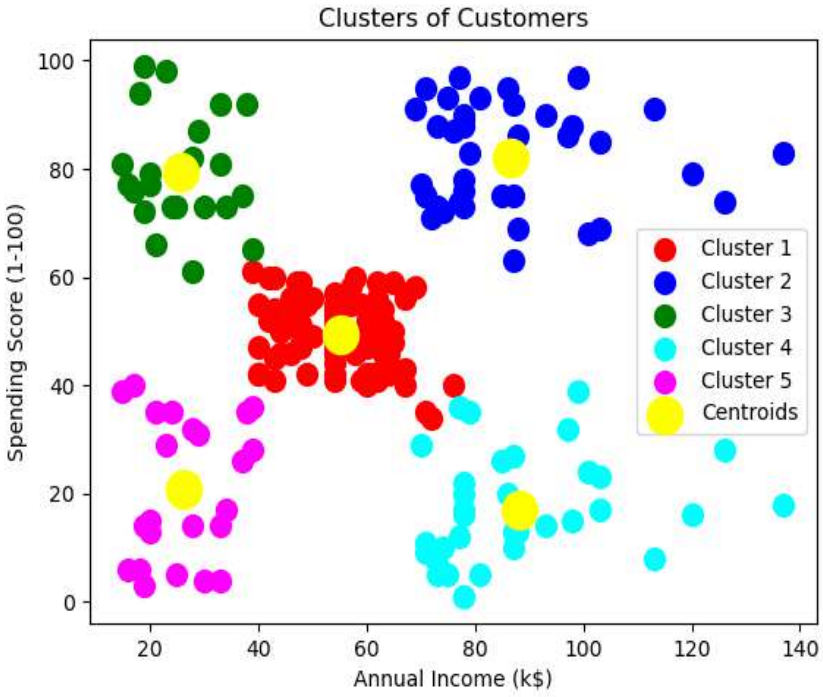
Generated a correlation heatmap to understand the correlation between numerical features. Found no strong correlation among the variables.

Customers with greater yearly earnings typically have higher spending scores, as seen by the moderately positive connection we find between annual income and expenditure score. There is no discernible relationship between Age and Spending Score or Annual Income.

Customer Segmentation:



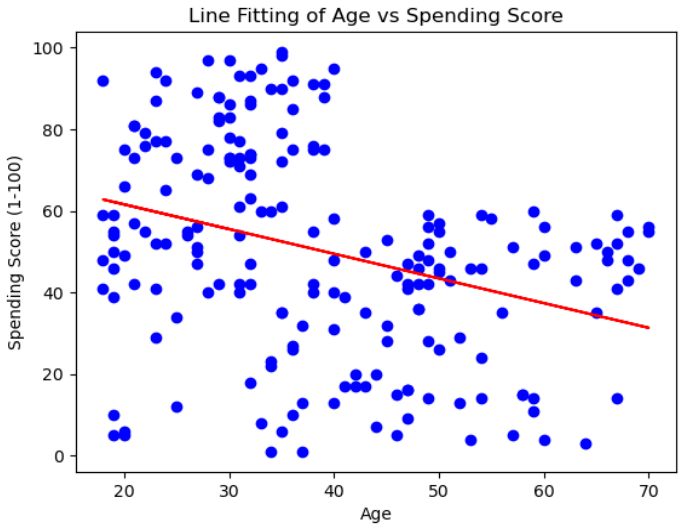
K-means clustering technique was used to divide up the client base according to expenditure score and annual income. The elbow approach was employed to ascertain the ideal number of clusters.



With centroids standing in for the cluster centres, the clustered segments were displayed on an Annual Income vs. Spending Score scatter plot.

According to their spending patterns and income levels, the five different consumer clusters are depicted in the scatter plot. Low-income and low-spending customers are represented by Cluster 1, whereas high-income and high-spending customers make up Cluster 3. Intermediate segments with a range of income and spending levels are represented by clusters 2, 4, and 5.

Predictive Analysis:



A linear regression model was fitted to predict Spending Score based on Age. The regression coefficients and model performance metrics were interpreted to understand the relationship between age and spending behaviour. This predictive analysis provides insights into how age influences customers' propensity to spend.

Indicating that younger consumers typically have higher spending scores, the regression line plot reveals a minor negative association between age and spending. Notwithstanding, the correlation is somewhat feeble, implying that additional variables could potentially impact consumer expenditure patterns.

Conclusion:

Conclusively, the examination of patron data obtained from the shopping centre yields significant understanding of patron choices and behaviour. Businesses may more successfully target particular client segments by segmenting their customer base according to their income levels and purchasing patterns. Based on the segmentation data, businesses were able to create customised experiences and targeted marketing efforts for each of the five separate consumer groupings, which varied in terms of income and spending patterns.

Furthermore, statistical analysis uncovered important trends and relationships within the data. Descriptive statistics highlighted the central tendencies and distributions of key variables, offering a deeper understanding of customer characteristics. The scatter plot of Age vs. Spending Score and the linear regression analysis provided insights into the relationship between age and spending behaviour, allowing businesses to better anticipate and respond to customer needs.

Overall, the insights gathered from this study can assist companies in increasing revenue, enhancing consumer engagement, and optimising their marketing tactics. To stay competitive in the ever-changing retail market, segmentation techniques will need to be explored and improved upon moving forward, along with constant data analysis and monitoring.