Student ID: 23036717
Student Name: Jafar Urchintala
GitHub Link: https://github.com/Jafarurchintala/ADS-Assignment-Clustering-and-Fitting.git
Dataset Link: https://www.kaggle.com/datasets/shwetabh123/mall-customers?rvi=1

Customer Segmentation and Analysis Report

Introduction:

In the modern business landscape, understanding customer behaviour and preferences is crucial for creating targeted marketing strategies and delivering personalized experiences. Customer segmentation, the process of categorizing customers into distinct groups based on shared characteristics, allows businesses to tailor their products and services to meet the unique needs of each segment.

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 (gender), Age, Annual Income (in thousands of dollars), and Spending Score (ranging from 1 to 100). 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 such as Age, Annual Income, and Spending Score. These statistics provide insights into the central tendency, variability, and distributional characteristics of the data.

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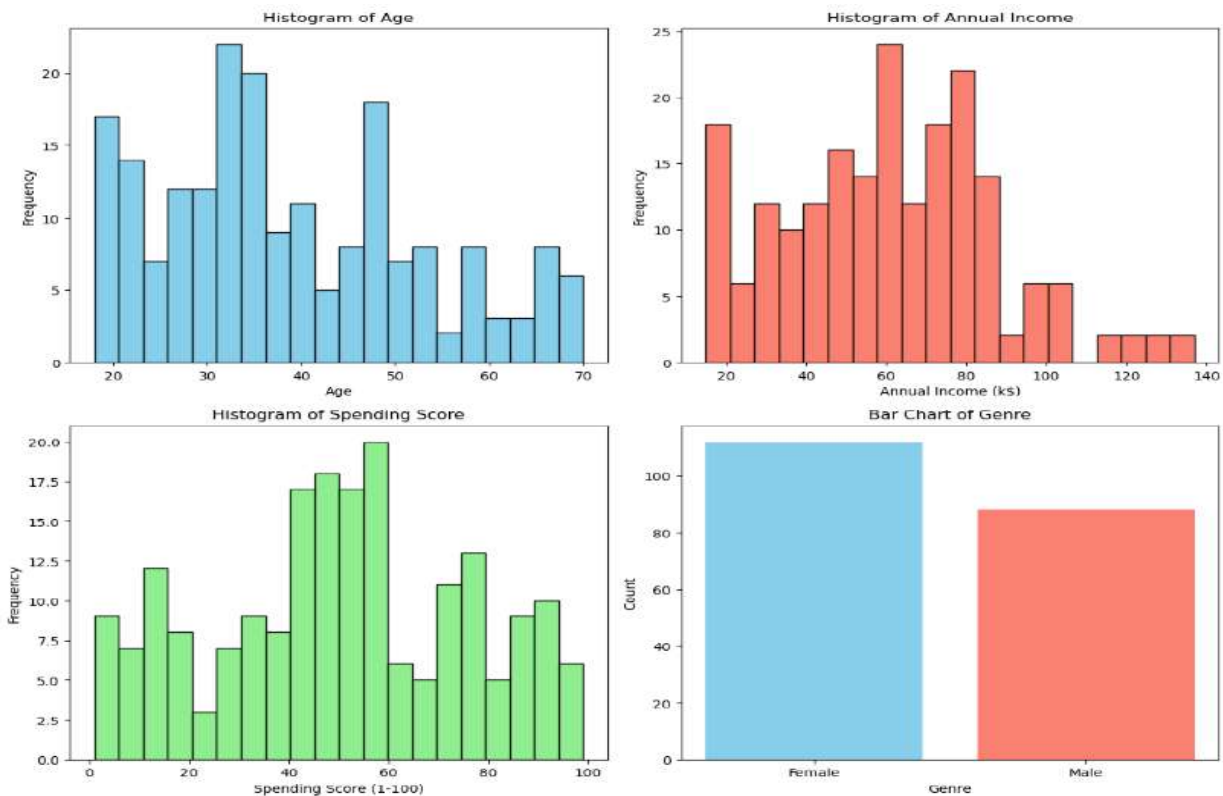
We visualized the distributions of Age, Annual Income, and Spending Score using histograms. These visualizations allow us to understand the frequency distribution of each variable and identify any patterns or outliers. Furthermore, we created a bar chart to visualize the distribution of Genre (gender)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

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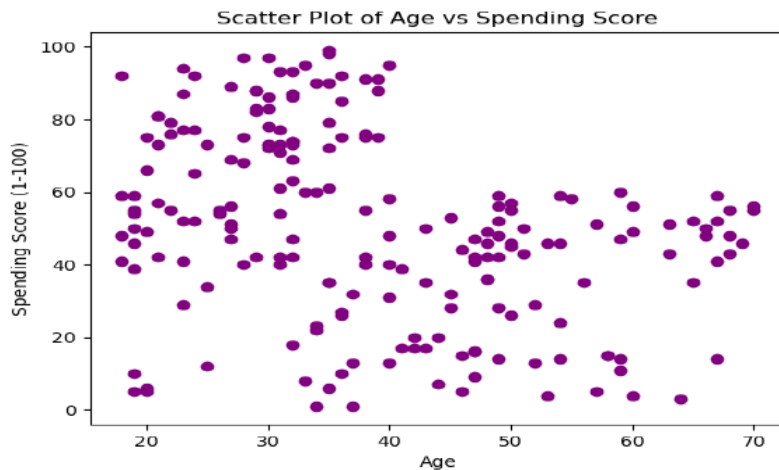
A scatter plot of Age vs. Spending Score was generated to explore the relationship between these two variables. This visualization helps us understand if there is any discernible pattern or correlation between age and spending behaviour.

The scatter plot shows no clear linear relationship between age and spending score, suggesting that age alone may not be a strong predictor of spending behaviour.


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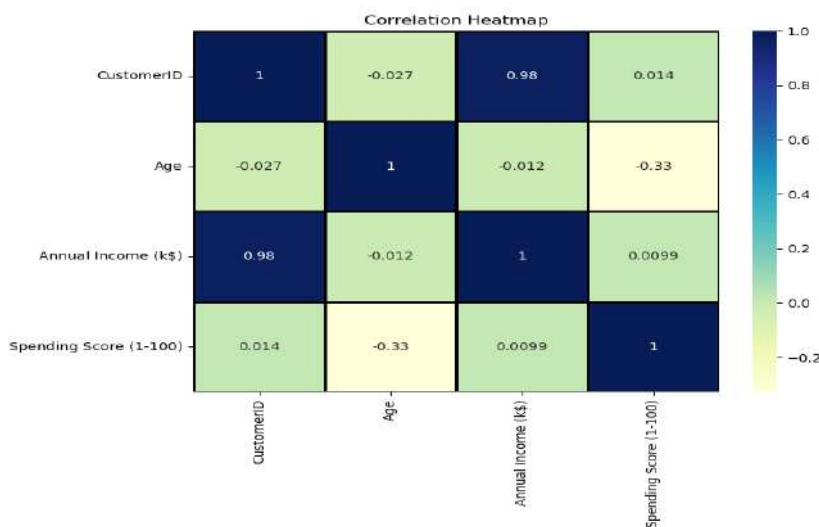
Generated a correlation heatmap to understand the correlation between numerical features. Found no strong correlation among the variables.

We see there is a moderate positive correlation between Annual Income and Spending Score, indicating that customers with higher annual incomes tend to have higher spending scores. No significant correlation is observed between Age and either Annual Income or Spending Score.

## Customer Segmentation:
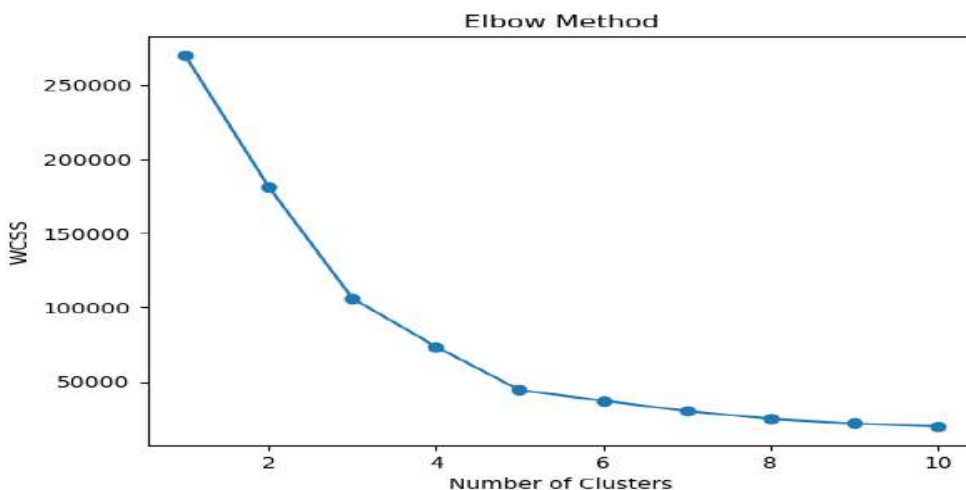
Elbow Method

To segment customers based on their Annual Income and Spending Score, we employed the K-means clustering algorithm. We utilized the elbow method to determine the optimal number of clusters, which was found to be 5.

Clusters of Customers
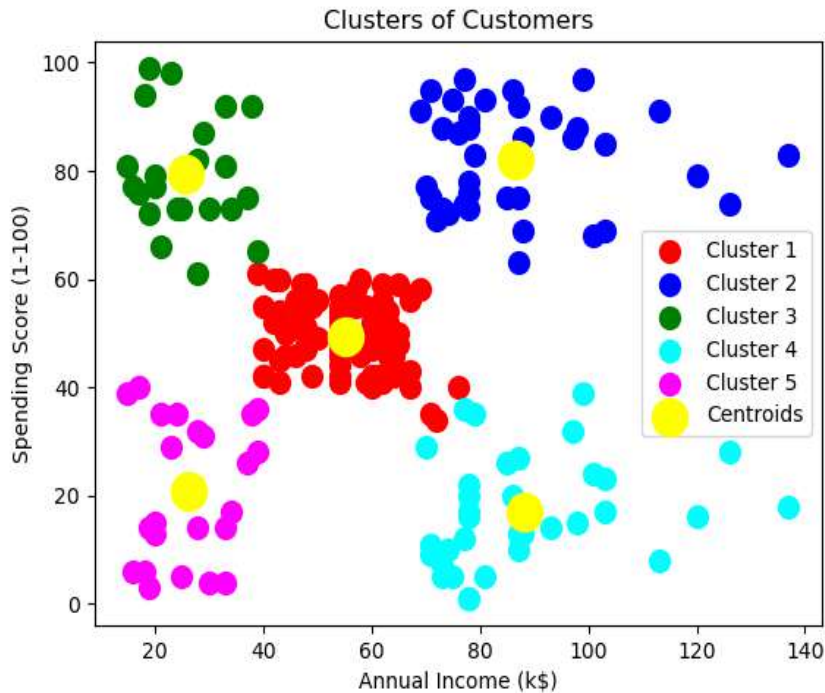
The clustered segments were visualized on a scatter plot of Annual Income vs. Spending Score, with centroids representing the cluster centres.

The scatter plot illustrates five distinct clusters of customers based on their spending behaviour and income levels. Cluster 1 represents customers with low income and low spending scores, while Cluster 3 consists of high-income customers with high spending scores. Clusters 2, 4, and 5 represent intermediate segments with varying income and spending levels.

Predictive Analysis:

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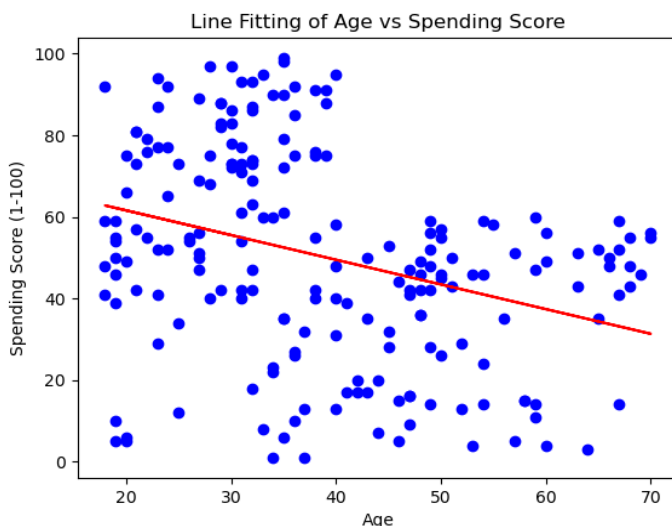
A linear regression model was fitted to predict Spending Score based on Age. The regression coefficients and model performance metrics were interpreted to understand the relationship between age and spending behaviour. This predictive analysis provides insights into how age influences customers' propensity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

The regression line plot shows a slight negative relationship between age and spending score, indicating that younger customers tend to have higher spending scores. However, the relationship is relatively weak, suggesting that other factors may also influence spending behaviour.

Conclusion:

In conclusion, the analysis of customer data from the mall provides valuable insights into customer behaviour and preferences. By segmenting customers based on their spending behaviour and income levels, businesses can tailor their marketing strategies to target specific customer segments more effectively. The segmentation analysis revealed five distinct customer segments with varying spending patterns and income levels, enabling businesses to develop targeted marketing campaigns and personalized experiences for each segment.

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 gained from this analysis can help businesses optimize their marketing strategies, improve customer engagement, and drive revenue growth. Moving forward, continued exploration and refinement of segmentation strategies, coupled with ongoing data analysis and monitoring, will be essential to maintain a competitive edge in the dynamic retail landscape.