REPORT FOR PROJECT

Title: Customer Segmentation

Problem Statement

Customer Segmentation is a popular application of unsupervised learning. Using clustering, identify segments of customers to target the potential user base. They divide customers into groups according to common characteristics like gender, age, interests, and spending habits so they can market to each group effectively.

Use K-means clustering and visualize the gender and age distributions. Then analyse their annual incomes and spending scores.

Dataset Used

The dataset name is Mall_Customers.csv consists of 5 columns which are CustomerID, Gender, Age, Annual Income (k$), Spending Score (1-100) where Gender is a categorical value and rest all features are numeric.

The size of the dataset is (200, 5) which is 200 rows and 5 columns.

	A	B	C	D	E
1	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
2	1	Male	19	15	39
3	2	Male	21	15	81
4	3	Female	20	16	6
5	4	Female	23	16	77
6	5	Female	31	17	40
7	6	Female	22	17	76
8	7	Female	35	18	6
9	8	Female	23	18	94
10	9	Male	64	19	3
11	10	Female	30	19	72
12	11	Male	67	19	14
13	12	Female	35	19	99
14	13	Female	58	20	15
15	14	Female	24	20	77
16	15	Male	37	20	13
17	16	Male	22	20	79
18	17	Female	35	21	35
19	18	Male	20	21	66
20	19	Male	52	23	29
21	20	Female	35	23	98

Algorithms

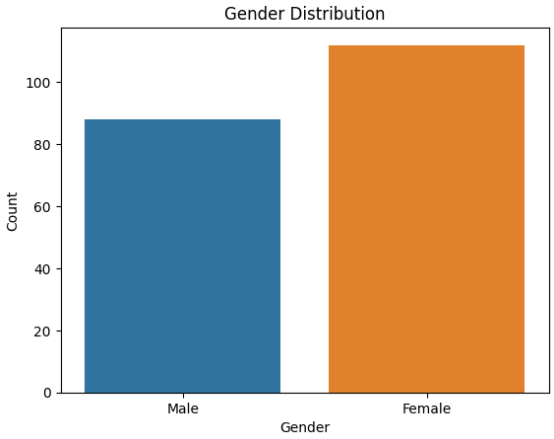
K-means algorithm is used in this project to analyze and form clusters of customers based on their income and spending score features.

Environment (Libraries and Technologies):

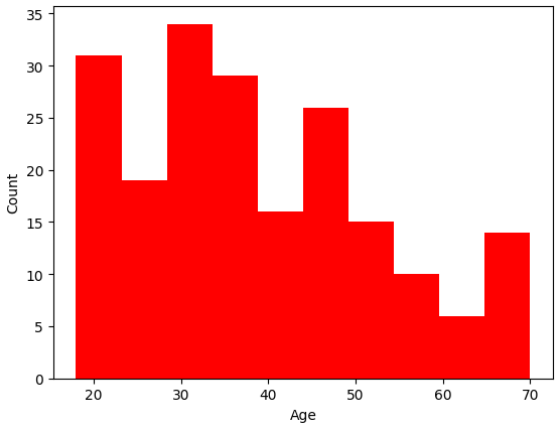
Numpy, Pandas, Matplotlib, Seaborn, Jupyter Notebook, Google Colab.

Analysis

1. From the *Count plot*, it is observed that the number of Female customers is more than the total number of Male customers.



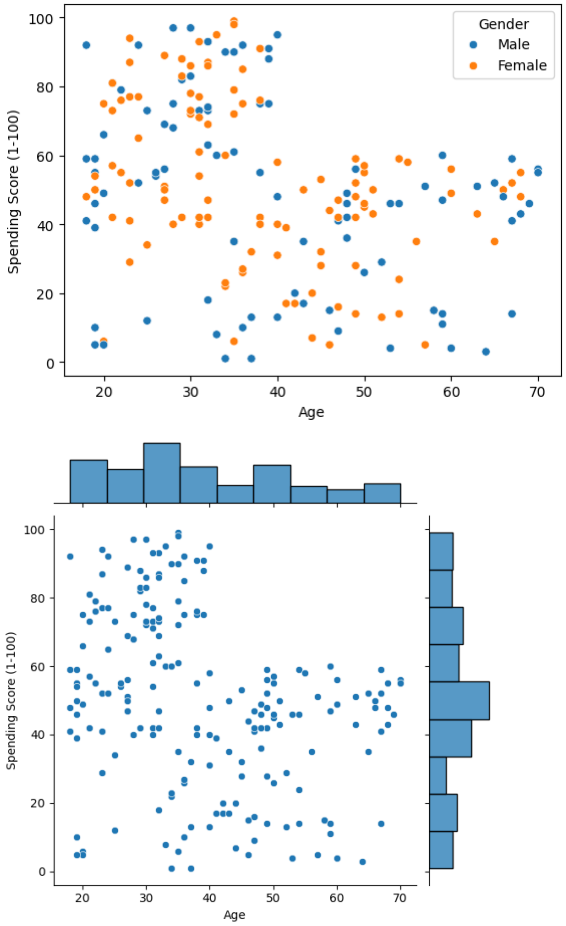
2. From the *Histogram* it is evident that there are 3 age groups that are more frequently shop at the mall, they are : 15-22 years, 30-40 years, and 45-50 years.



3. *Age vs Spending score*

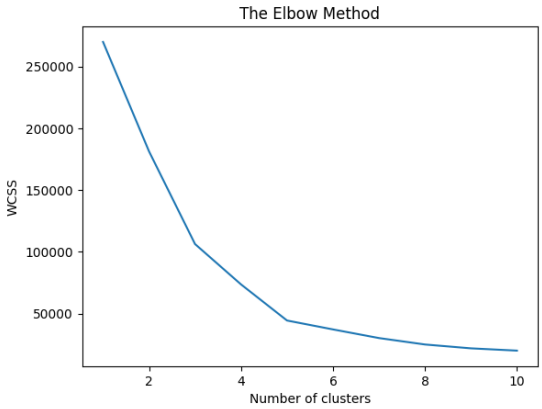
3.1. From the Age vs. Spending Score plot, we see that customers with a spending score above 65 are aged between 15 and 42 years. The scatter plot also shows that among these customers, there are more females than males.

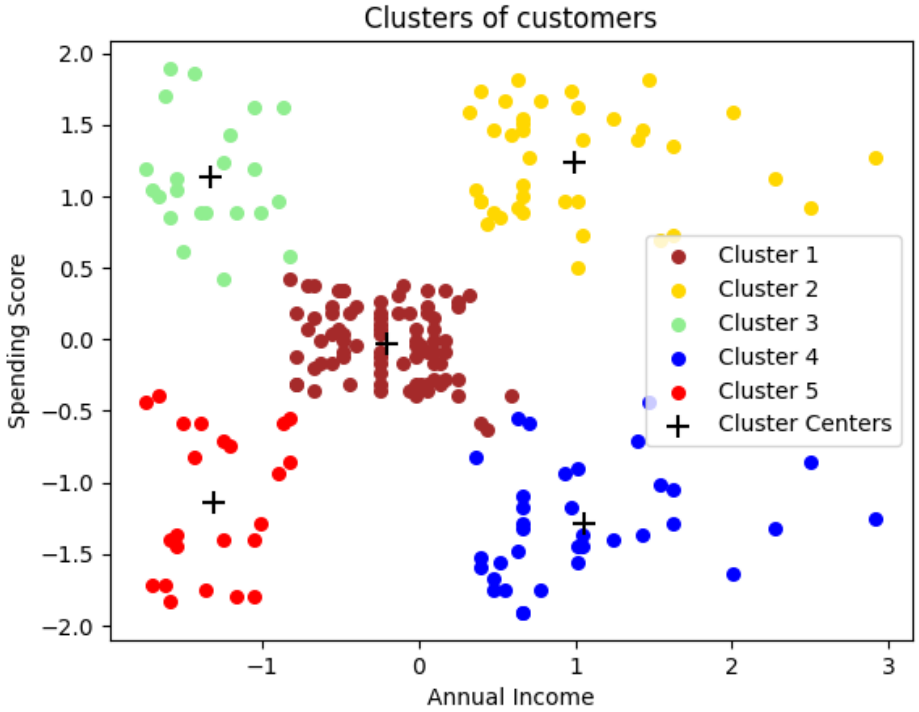
3.2. Customers with an average spending score (40-60) are aged between 15 and 75 years, with a roughly equal number of males and females in this age group.



4. *Cluster Analysis*

Using Elbow method, we can obtain the number of clusters possible that is 5.





Cluster 1: Average Income, Average Spending Score: Target these customers based on the mall's policy.

Cluster 2: High Income, Low Spending Score: Gather feedback and improve advertising to convert these customers into Cluster 5.

Cluster 3: Low Income, Low Spending Score: Do not target these customers as they need to save money.

Cluster 4: Low Income, High Spending Score: Offer low-cost EMI options to attract these customers.

Cluster 5: High Income, High Spending Score: Send new product alerts to these loyal customers to boost revenue.