BA 890 Individual Research
Mall Customer Segmentation
Ziqi Shan

Introduction
People visit shopping malls to enjoy variety while shopping, for socialization and for entertainment. As the mall itself, it's important to understand the customer behavior and what type of customers have more purchase power. This project aims to cluster customers in order to find a target customer group so that the marketing department can plan the strategy accordingly.

Dataset
The dataset totally has 200 data points and 5 features:
Customer ID:
Age: the age of the customer, from 18 to 70
Gender: the sex of the customer, female and male
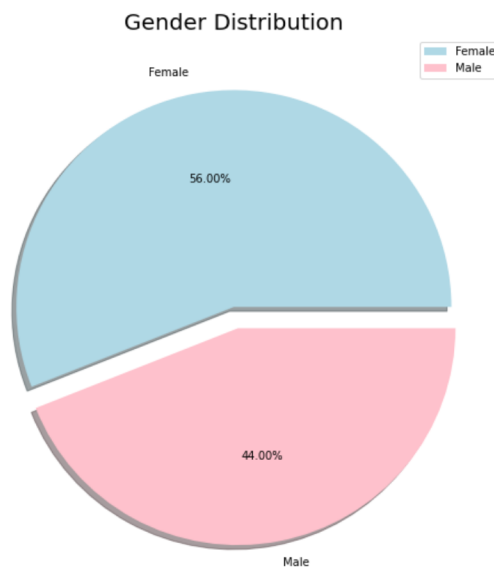Annual Income: annual income of the customer in k$, from 15k to 137k
Spending score: score assigned by the mall based on customer behavior and spending nature

The dataset has no missing value, and is good to explore.
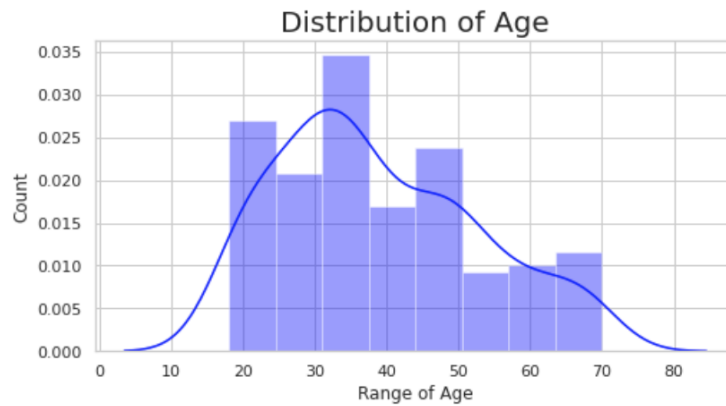
Exploratory Data Analysis
1. **Gender**
   We have 56% female customers and 44% male customers, female are slightly more than male.
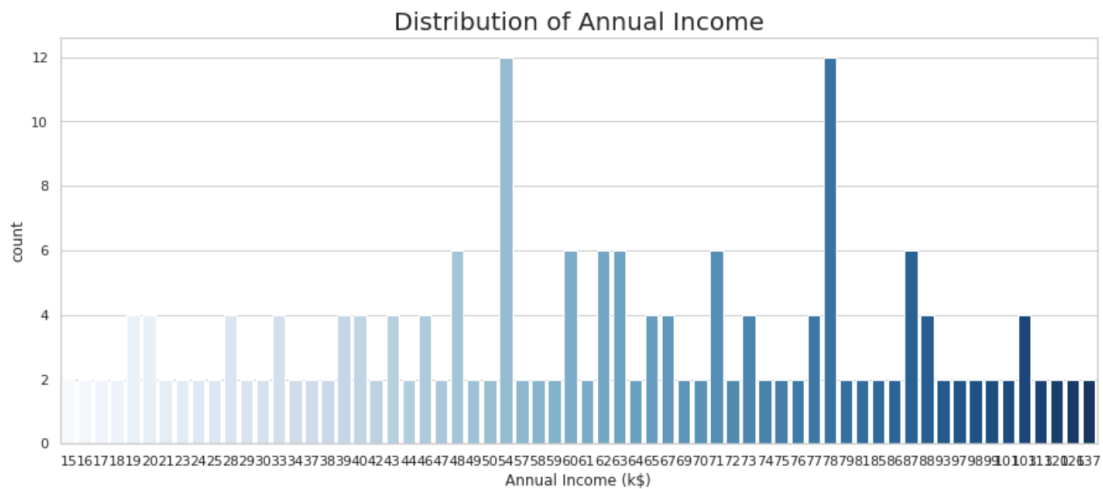

Gender Distribution

2. **Age**
   From the graph, we can see that most regular customers for the mall are around 30-39 years old. Senior group(50-70) is the least frequent customer in the mall.

Distribution of Age

3. **Annual Income**

The annual income ranges widely from 15k to 137k. There are few customers who earn more than 100k one year. Most of them have yearly earnings around 50-80k. It's worth noting that there are a large number of customers who have their annual income of 54k or 78k.


Distribution of Annual Income

4. **Spending Score**

Generally, most customers have a spending score in the range of 30-60. It's interesting to notice that the distributions of 0-10 score and 90-100 are similar, which indicates that the mall satisfies the variety of customers with varying needs and requirements available in the mall.

Distribution of Spending Score (1-100)
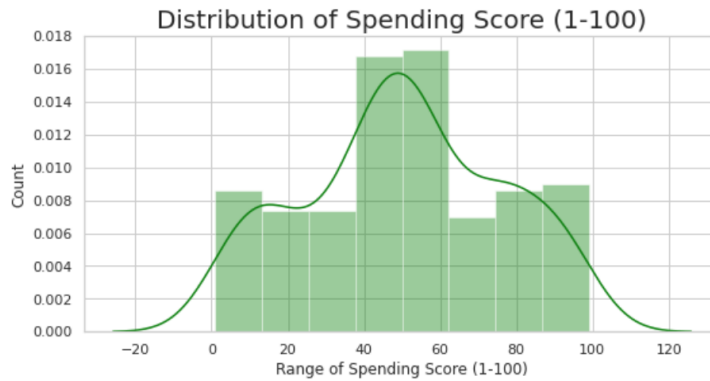
5. Correlation Check

It's good to see that there is no obvious correlation among variables. Therefore, all features could be further processed.

Heatmap for the Data
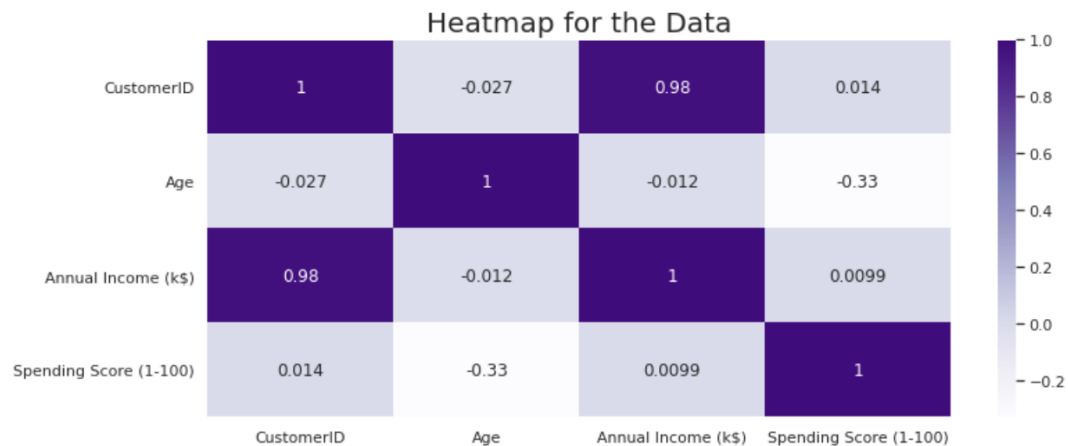
6. Gender - Spending Score

Since there is no significant difference in the average score between male and female, I will remove this variable from further modeling. However, it's worth knowing that most male customers have a spending score range from 25 to 70, meanwhile, female customers have the score range from 35 to 75. The score range for females is slightly higher than male, indicating that women are the main consumers in the mall.


Gender vs Spending Score

7. **Age - Spending Score**

Customers between the ages of 20-40 have higher spending scores, meaning they have more purchase power than other age groups.


Age - Spending Score

8. **Annual Income - Spending Score**

People with annual earnings of 40k-65k tend to have a middle purchase power(with score 40-60). However, people with annual income of 0-40k and 65k-140k have different performance, either have high scores(60-100) or low scores(0-40).

Annual Income (k$) - Spending Score
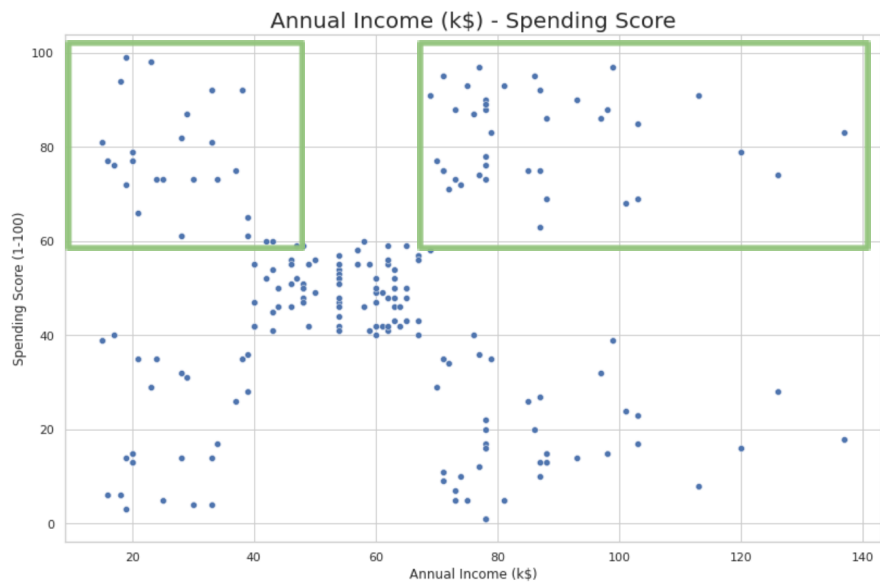
Model

K-means clustering is the main method that is used. In order to choose the optimal number of clusters, the Elbow method is conducted. According to the graph, I decide to choose k=2.

The Elbow Method
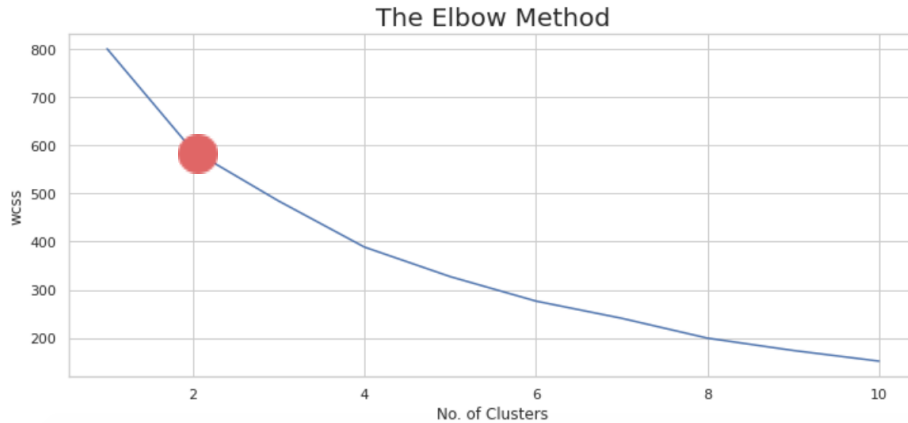
After clustering all the customers into two groups, let's see how features differ from each other.

1. **Age**

It's clear that group 0 contains mainly customers aged from 18-40, on the other hand, group 1 has people ranging from 40-70 years old.


Distribution of Age
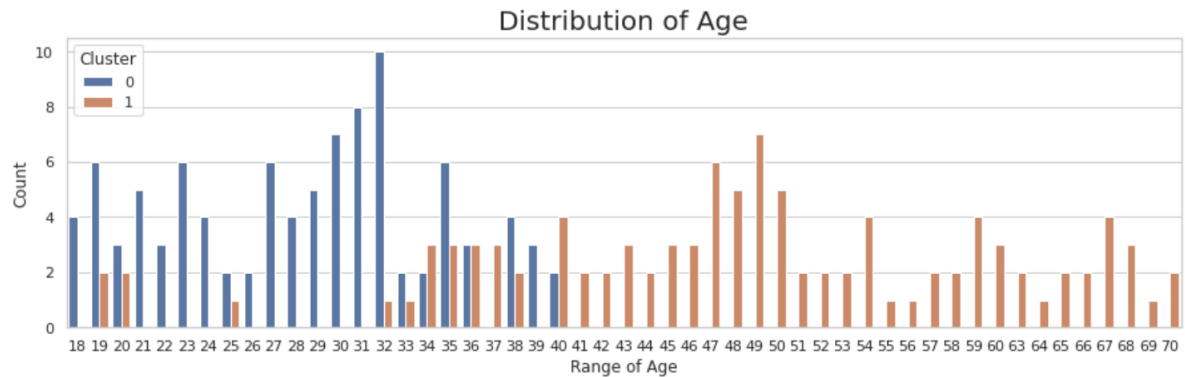
2. **Annual Income**

There is no significant difference between the two groups. On average, group 0 has mean annual income of 60.50k and group 1 has 60.63k. However, group 1 contains more people with 43-54k annual income. Meanwhile, group 0 has more customers with 69-76k annual earnings.

Distribution of Annual Income (k$)
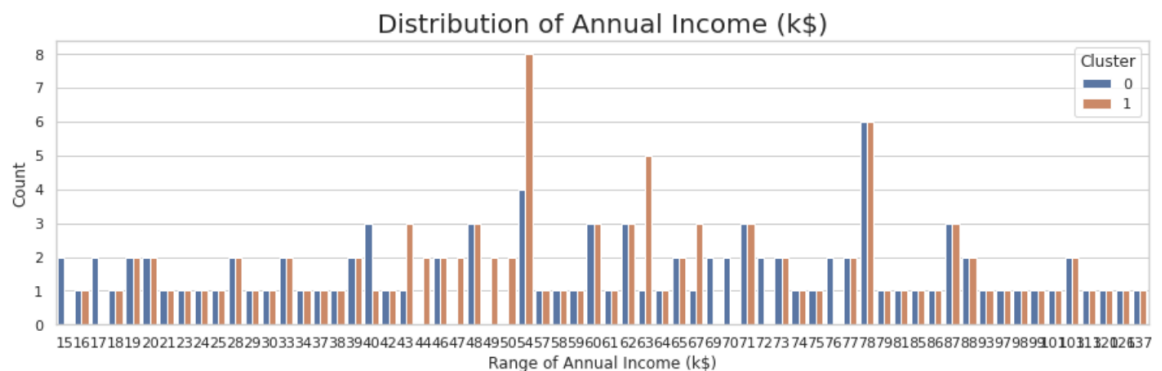
3. **Spending Score**

According to the graph, it seems that group 0 is the target group that we are interested in, with higher spending score.

Distribution of Spending Score (1-100)
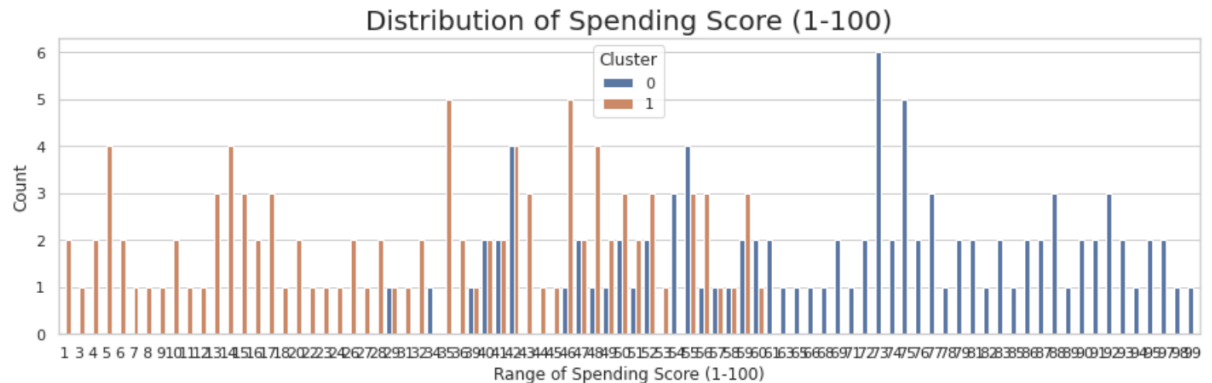
Takeaways and Recommendation
1. It seems that there is no significant difference between male and female customers according to the spending score, so the gender-based audience could be ignored.
2. According to the analysis above, the consumers between 18-40 age would have more purchase power compared to other age groups. Therefore, making special campaigns to this age group of people could have a positive impact on the mall's profit.
3. Although we could not have a clear clue about the annual income effect based on the clustering method, from the analysis compared the annual income and spending score, we could see that people with middle-income(40-60k dollars annual earnings) are also at a middle level of spending score. One possible marketing strategy is to run campaigns to target these middle-income groups to increase the number of these types of customers.
4. It's worth noting that customers with high annual income perform differently. Some of them have a high spending power, however, still a portion of customers did not spend so much in the mall. This might be because low-spenders are not satisfied with the shopping. Possible strategy is to improve the consumer service and product quality to stick with these high-income customers.
5. The data provide limited information for analysis.If we have other features like customer shopping habits or the reviews/suggestions about the mall, we could even formulate some more specific marketing strategies to increase the profit.