Unsupervised Learning for Customer Segmentation

Name - Kartik Dhiman

Follow Me!

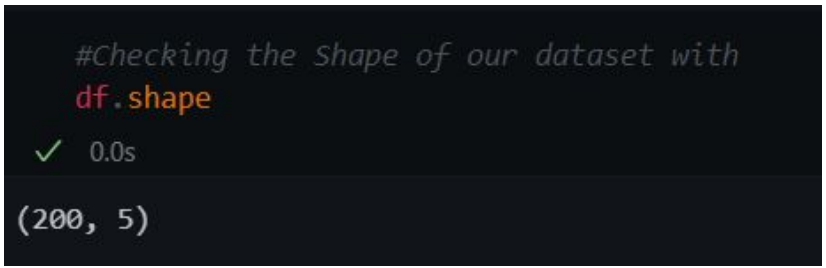
Introduction: The aim of this project was to analyze customer data to identify different customer segments and provide recommendations for targeted marketing strategies for each segment. The dataset used for this analysis contained information about 200 customers, including their age, gender, annual income, and spending behavior.

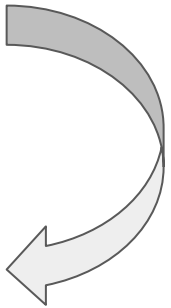
Data Set: The provided dataset can be found on Kaggle - Mall Customer Segmentation Data. It contains information about mall customers, including: Link for dataset!

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

We have a total of 200 rows and 5 columns.

- CustomerID: Unique ID assigned to the customer
- Gender: Gender of the customer
- Age: Age of the customer
- Annual Income (k$): Annual Income of the customer
- Spending Score (1-100): Score assigned by the mall based on customer behavior and spending nature





Descriptive Statistics:

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

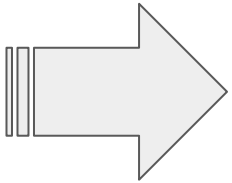
we can see that the average age of the customers is 38.85 years, the average annual income is $60,560, and the average spending score is 50.2. The customers in this dataset range in age from 18 to 70 years old, have annual incomes ranging from $15,000 to $137,000, and spending scores ranging from 1 to 99.

Data Cleaning:

```
df.isnull().sum()
```

✓ 0.0s

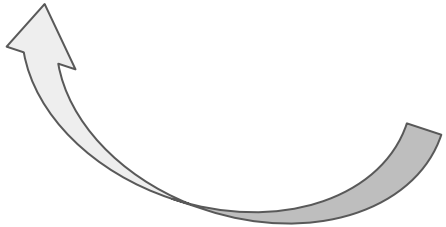
```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
dtype: int64
```



```
# Doing Label Encoding for gender
from sklearn.preprocessing import LabelEncoder
lb = LabelEncoder()
df['Gender'] = lb.fit_transform(df['Gender'])
```

✓ 0.1s

Male --> 1 , Female -->0



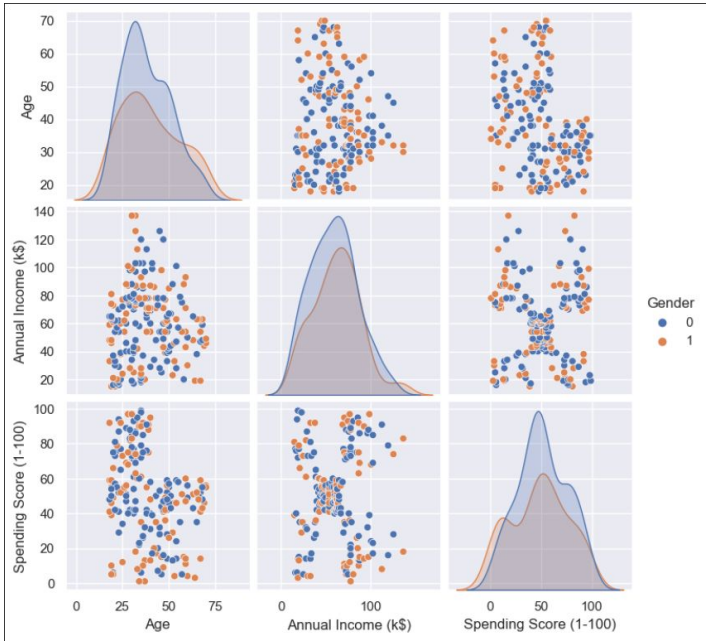
```
df.duplicated().sum()
```

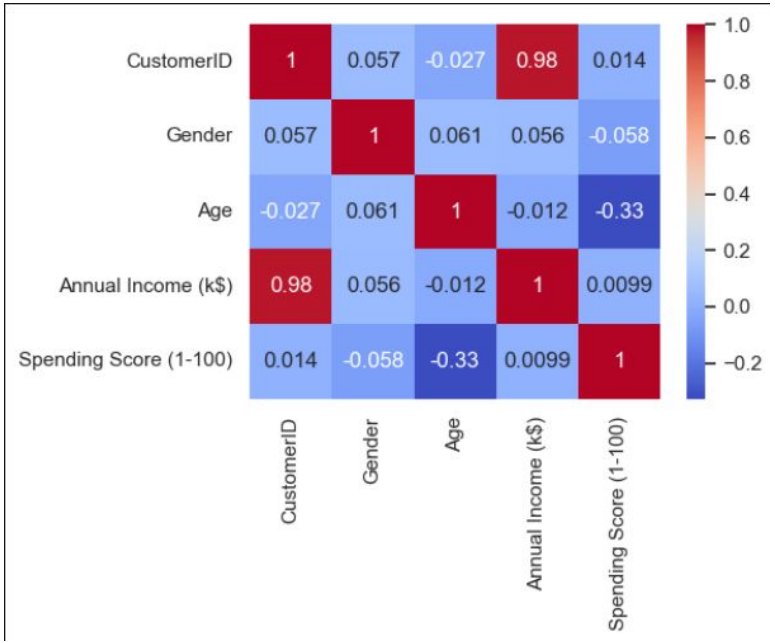
✓ 0.0s

```
0
```

We Have 0 Missing Values and 0 Duplicate Values And We Perform Label Encoding on Gender.

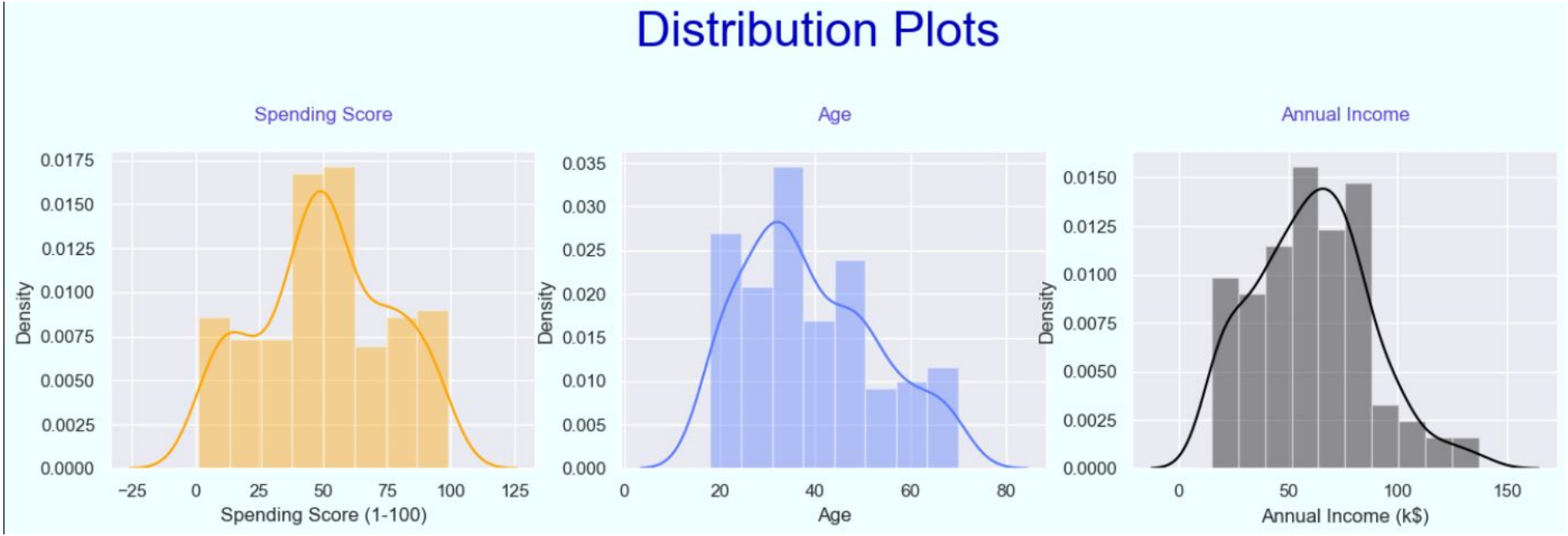
Exploratory Data Analysis:



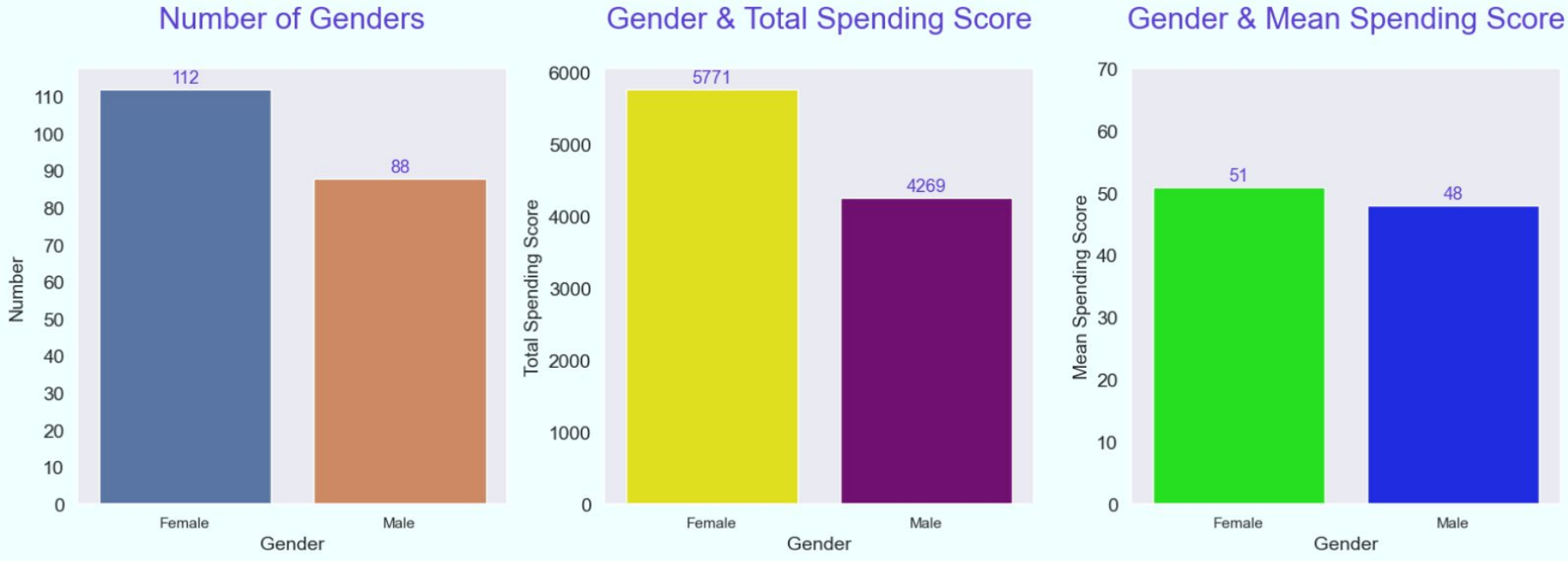


we can see that there is no clear linear relationship between the variables and also It is clear from this correlation table that older customers have less income and therefore spend less money.

The distributions are generally similar to the normal distribution, with only a few standard deviations. The 'more normal' distribution among the distributions is the 'Spending Score'. That's good because it's our target column.

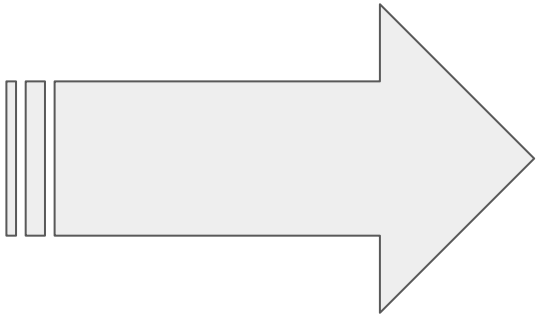


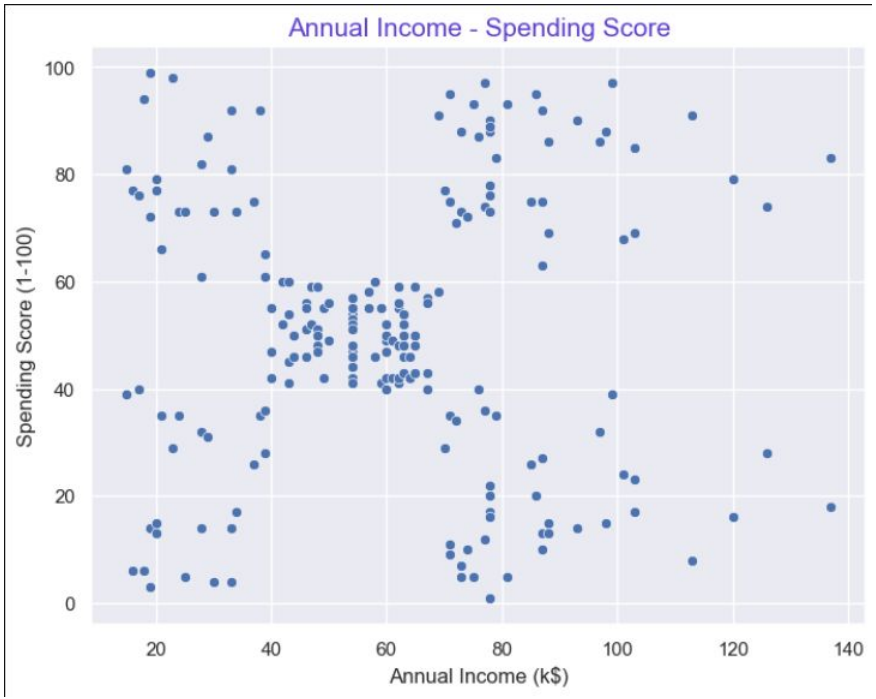
There is no significant difference in the mean spending scores of males and females. Since the mean spending scores are very close to each other, the difference between the total spending scores is the difference between the number of male and female customers, but this difference is not serious. Considering all this, it would be meaningless to choose a gender-based target audience.



People between the ages of 20-40 have made more purchases, considering the inference we just made about women, we can make our target audience more specific.

One of the two regions shown can be selected as the target audience. Even though the number of people whose annual income is between (40-60)k$ is higher (we understand this from the number of data points), the number of that audience is higher but the spending score is low, so if we make shopping attractive for them by choosing the target audience from the two regions above, we will see more profit can be made.





Recommendations:

1. High Income, High Spenders: Offer high-end products and luxury experiences, and create exclusive membership programs and events to attract and retain customers.
2. High Income, Low Spenders: Offer promotions and discounts to encourage spending and target products that appeal to their interests.
3. Mid Income, Mid Spenders: Offer a range of affordable products and focus on building strong customer relationships through personalized experiences.
4. Low Income, High Spenders: Offer affordable products and create loyalty programs to retain customers.
5. Low Income, Low Spenders: Offer value-for-money products and focus on convenience and accessibility to attract customers.

1. **Marketing strategies for (High Income, High Spenders):** This segment consists of customers who are young and have a high income but a low spending score. The company should focus on promoting premium products with attractive discounts to this segment. The company should also consider loyalty programs to increase the frequency of purchases from this segment.
2. **Marketing strategies for (High Income, Low Spenders):** This segment consists of customers who are middle-aged, have an average income, and an average spending score. The company should focus on promoting mid-range products to this segment. The company can also introduce referral programs to attract new customers from this segment.
3. **Marketing strategies for (Mid Income, Mid Spenders):** This segment consists of customers who are older, have a high income, and a high spending score. The company should focus on promoting luxury products to this segment. The company should also introduce personalized promotions to increase the loyalty of this segment.
4. **Marketing strategies for (Low Income, High Spenders):** This segment consists of customers who are young, have a low income, and a high spending score. The company should focus on promoting budget-friendly products to this segment. The company should also introduce discount programs and collaborations with budget-friendly brands to increase the frequency of purchases from this segment.

Inferential Questions:

1. **How do the identified customer segments differ from each other in terms of demographics and spending behavior?**

   The identified customer segments differ from each other in terms of demographics such as age, annual income, and gender. For example, customers in Segment 1 (High Income, High Spenders) are younger adults with high annual incomes, while customers in Segment 5 (Low Income, Low Spenders) are older adults with low annual incomes. In terms of spending behavior, customers in Segment 1 (High Income, High Spenders) spend the most on all product categories, while customers in Segment 5 (Low Income, Low Spenders) have very low income and also they spent very less in all product categories.

2. **Are there any noticeable trends in the spending behavior of customers within each segment?**

   Yes, there are noticeable trends in the spending behavior of customers within each segment. For example, customers in Segment 1 tend to spend less overall, while customers in Segment 3 tend to spend the most. Additionally, customers in Segment 4 tend to spend more on certain product categories like clothing and accessories, while customers in Segment 2 tend to spend more on other categories like groceries.

3. How does gender influence the customer segments and their spending behavior?

Gender does not seem to have a significant influence on the customer segments and their spending behavior in this dataset, as there is no clear pattern in the distribution of gender across the four segments.