

Customer Segmentation using Machine Learning

Customer Segmentation is one the most important applications of unsupervised learning. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. In this machine learning project, we will make use of [K-means clustering](#) which is the essential algorithm for clustering unlabeled dataset.

Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.

Datasets: Load dataset Mall Customer dataset.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df=pd.read_csv("Mall_Customers.csv")
df.head()
```

	CustomerID	Genre	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

Shape of the dataset is 200 rows and 5 column.

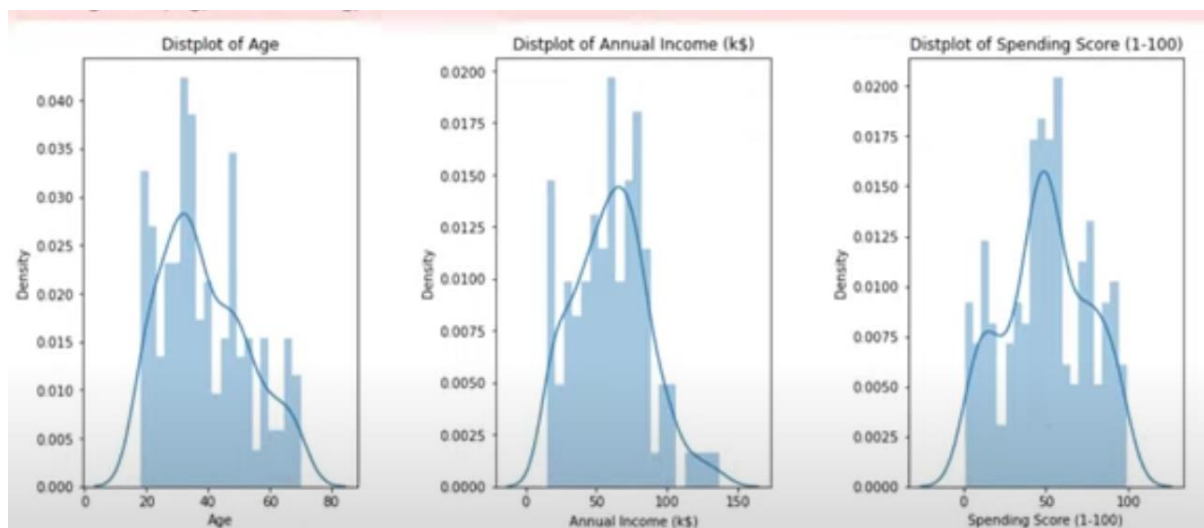
Missing Values: No missing value.

Statistics of Data:

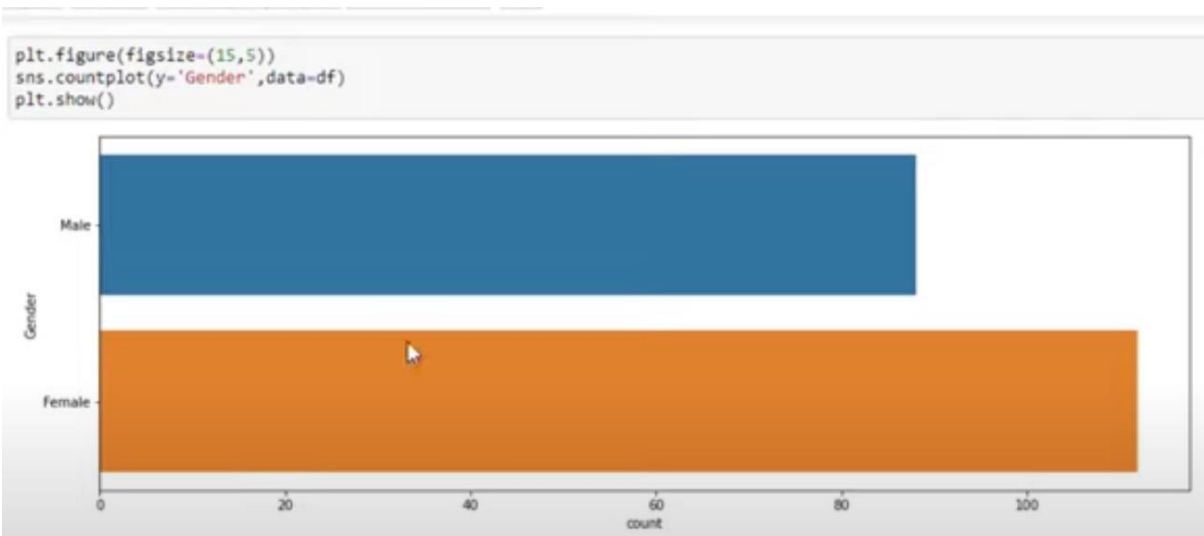
```
df.describe()
```

	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

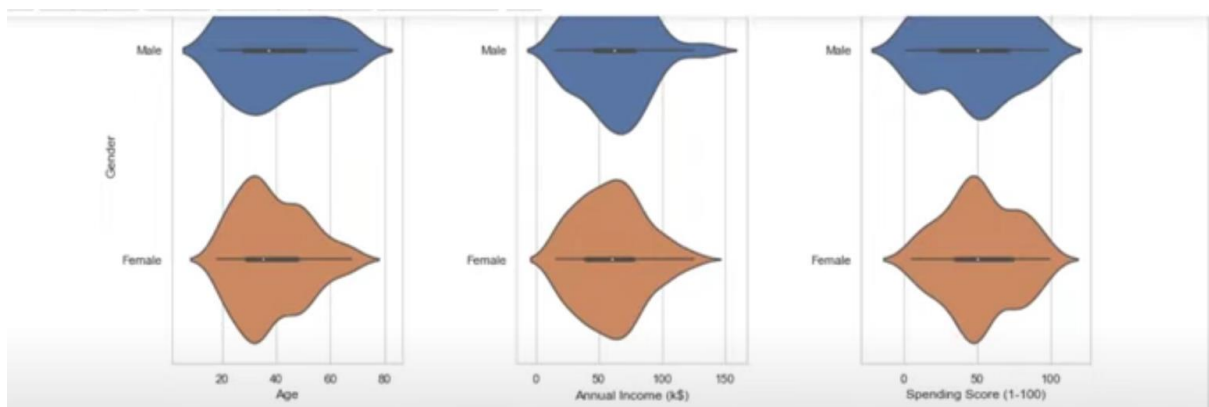
Exploratory Data Analysis: Create an distribution plot graph of Age, Annual Income and Spending Score.



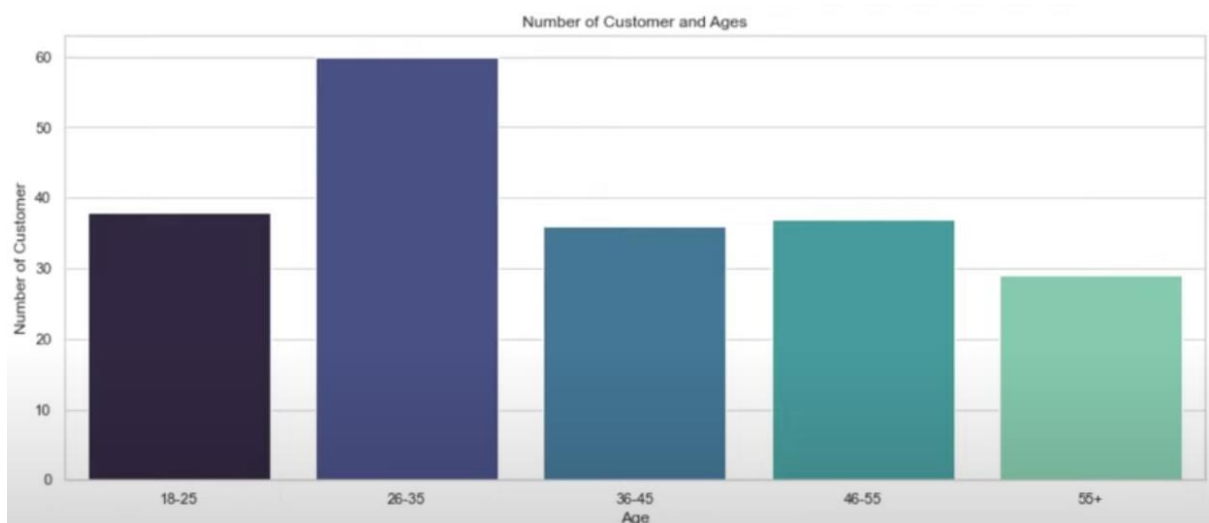
Create Countplot graph between Male and Female



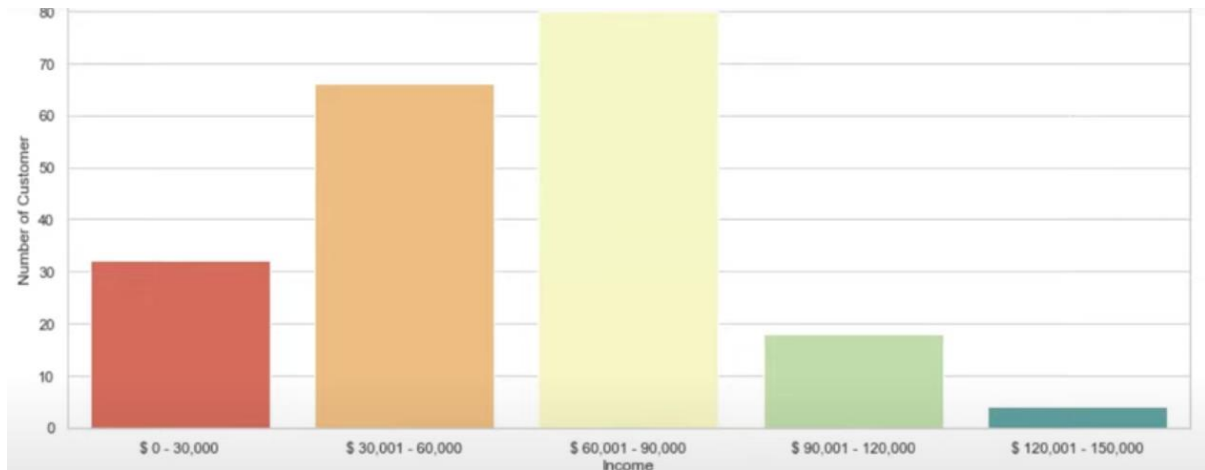
Create Violin plot between Age, Annual Income, Spending score based on Gender.



Create an barplot between various Age groups and Number of Customers.



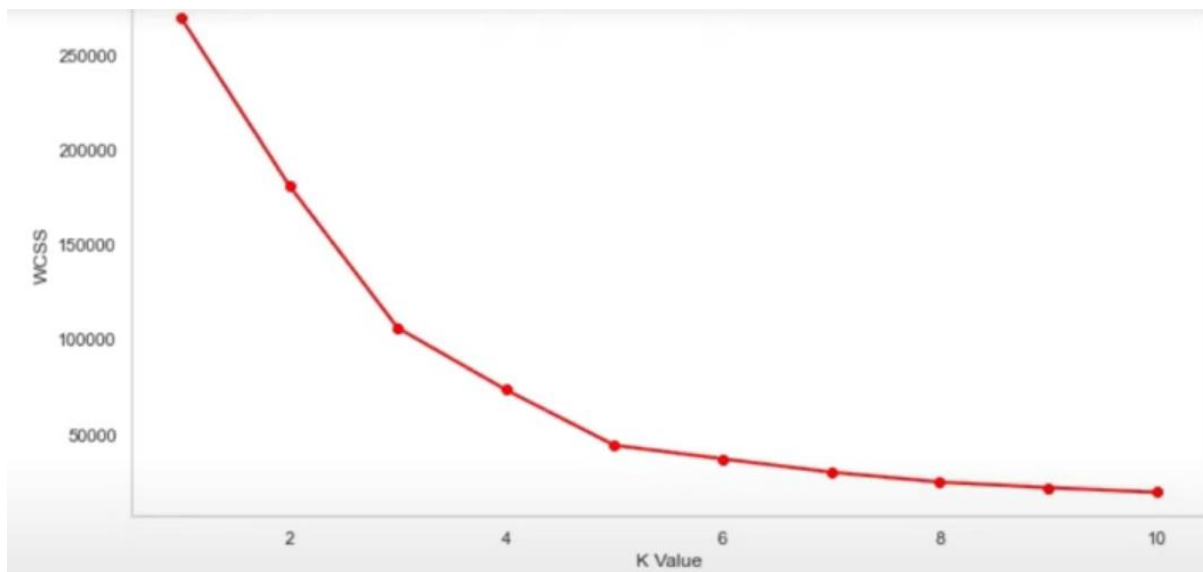
Create an barplot between Aunnal Income and Numbers of Customers.



Create the cluster data between Aunnal Income and Spending Score.

```
X2=df.loc[:, ["Annual Income (k$)", "Spending Score (1-100)"]].values

from sklearn.cluster import KMeans
wcss = []
for k in range(1,11):
    kmeans = KMeans(n_clusters=k, init="k-means++")
    kmeans.fit(X2)
    wcss.append(kmeans.inertia_)
plt.figure(figsize=(12,6))
plt.grid()
plt.plot(range(1,11),wcss, linewidth=2, color="red", marker="8")
plt.xlabel("K Value")
plt.ylabel("WCSS")
plt.show()
```

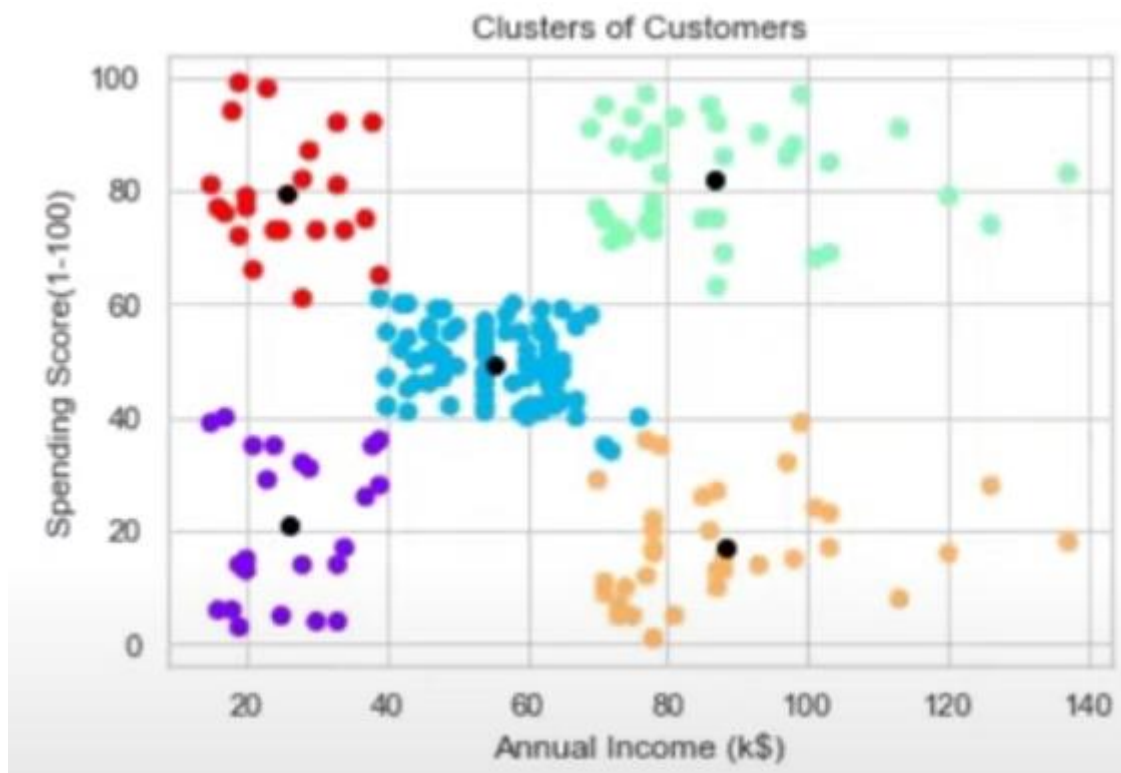


```
kmeans = KMeans(n_clusters=5)
label = kmeans.fit_predict(X2)
print(label)
```

```
[0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0
 4 0 4 0 4 0 1 0 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3
 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3 2 3]
```

```
print(kmeans.cluster_centers_)
```

```
[[26.30434783 20.91304348]
 [55.2962963  49.51851852]
 [86.53846154 82.12820513]
 [88.2         17.11428571]
 [25.72727273 79.36363636]]
```



Cluster 1(Blue) – This cluster represents the customer_data having a high annual income as well as a high annual spend.

Cluster 2(Red) –This cluster denotes a high annual income and low yearly spend.

Cluster 3(Violet) –This cluster denotes the customer_data with low annual income as well as low yearly spend of income.

Cluster4(Green) –These clusters represent the customer_data with the medium income salary as well as the medium annual spend of salary.

Cluster 5(Orange) –This cluster represents a low annual income but its high yearly expenditure.

CONCLUSION:

We developed this using a class of machine learning known as unsupervised learning. Specifically, we made use of a clustering algorithm called K-means clustering. We analyzed and visualized the data and then proceeded to implement our algorithm.

