```
#Importing the Libraries to iterate the functions
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
#Importing the dataset
df = pd.read_csv("Mall_Customers.csv")
```

```
#To Access all data from the csv file
df.info(verbose = True, null_counts = False)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Dtype
---  ------                  -----
 0   CustomerID              int64
 1   Gender                  object
 2   Age                     int64
 3   Annual Income (k$)      int64
 4   Spending Score (1-100)  int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```
#To Check the Index values of our Dataset
df.columns
```

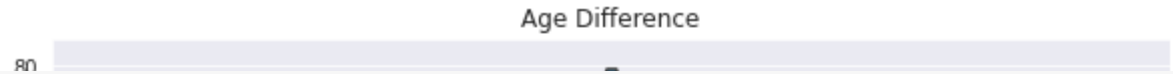
```
Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k$)',
       'Spending Score (1-100)'],
      dtype='object')
```

```
#To check these values with the split up values
df.head()
```

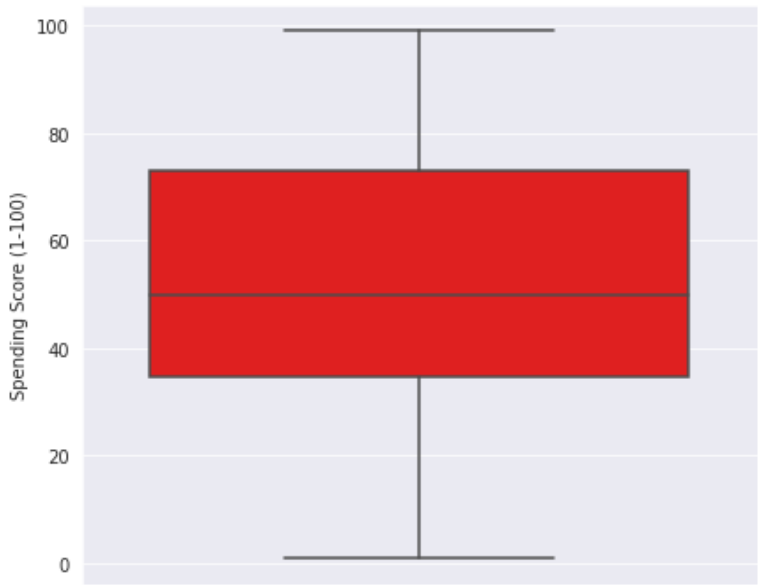
	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77

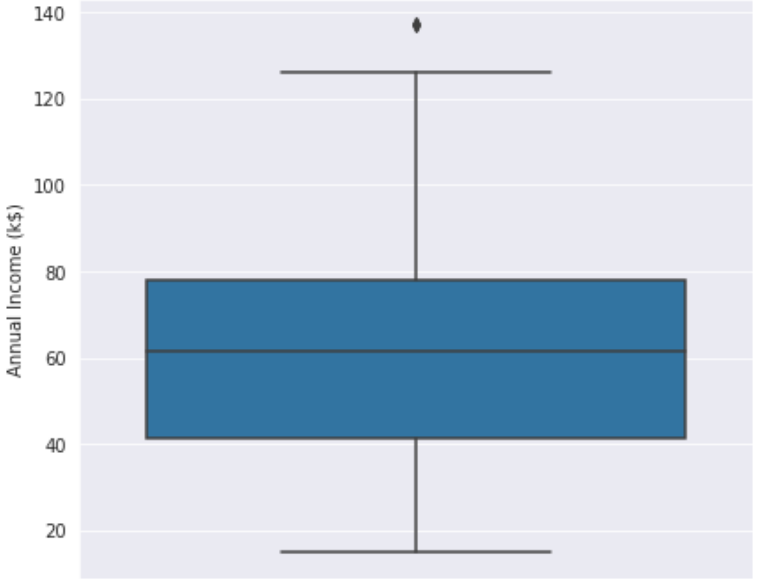
```
#To Line up these values from dataset
df.drop(["CustomerID"], axis = 1, inplace = True)
```

```
#To plot the age difference from the total customer dataset
plt.figure(figsize = (10,6))
plt.title("Age Difference")
sns.axes_style("dark")
sns.violinplot(y=df["Age"])
plt.show()
```

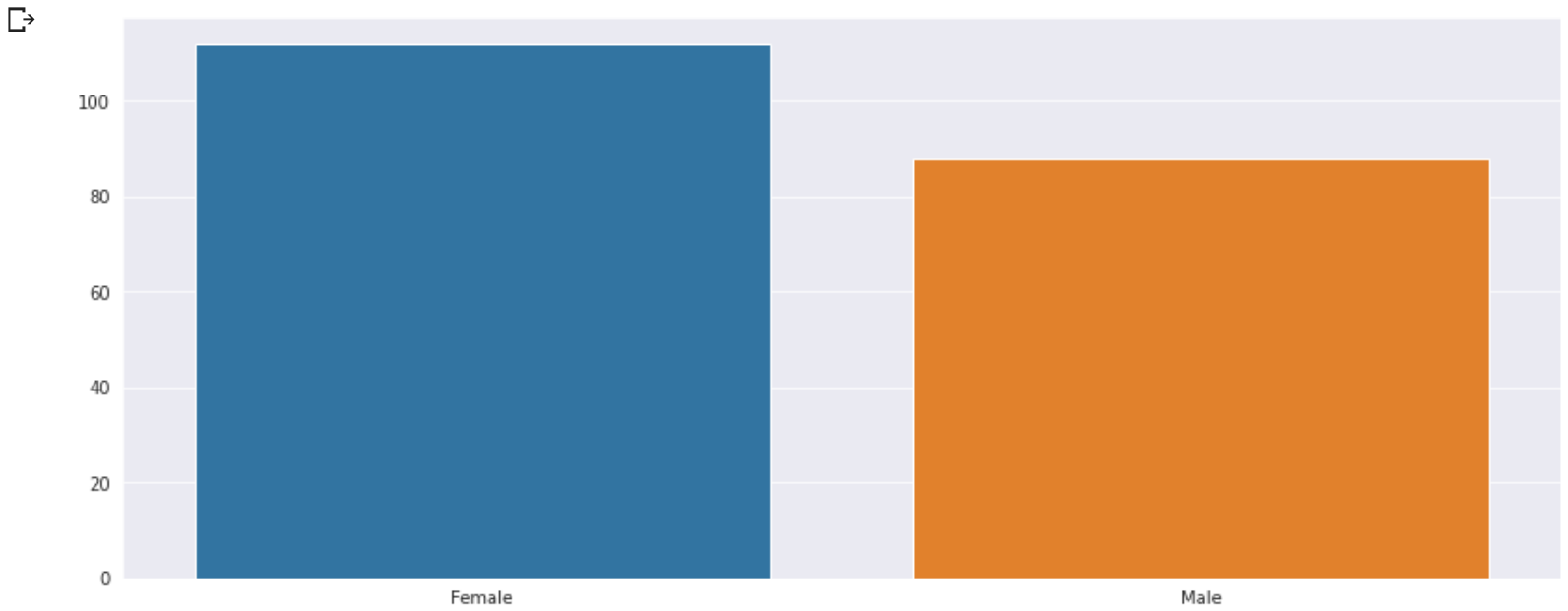



```
#To get the Amount Spent by customers
plt.figure(figsize=(15,6))
plt.subplot(1,2,1)
sns.boxplot(y=df["Spending Score (1-100)"], color = "red")
plt.subplot(1,2,2)
sns.boxplot(y=df["Annual Income (k$)"])
plt.show()
```

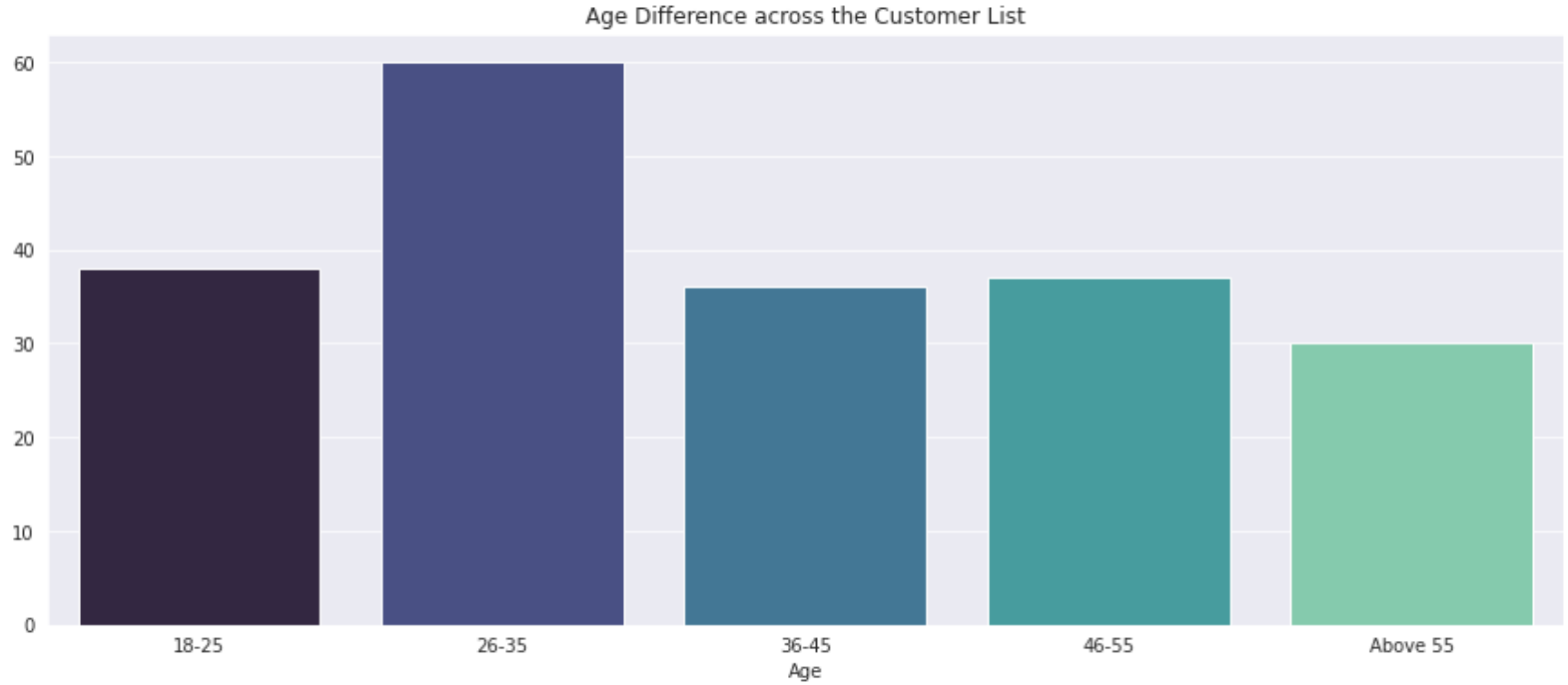




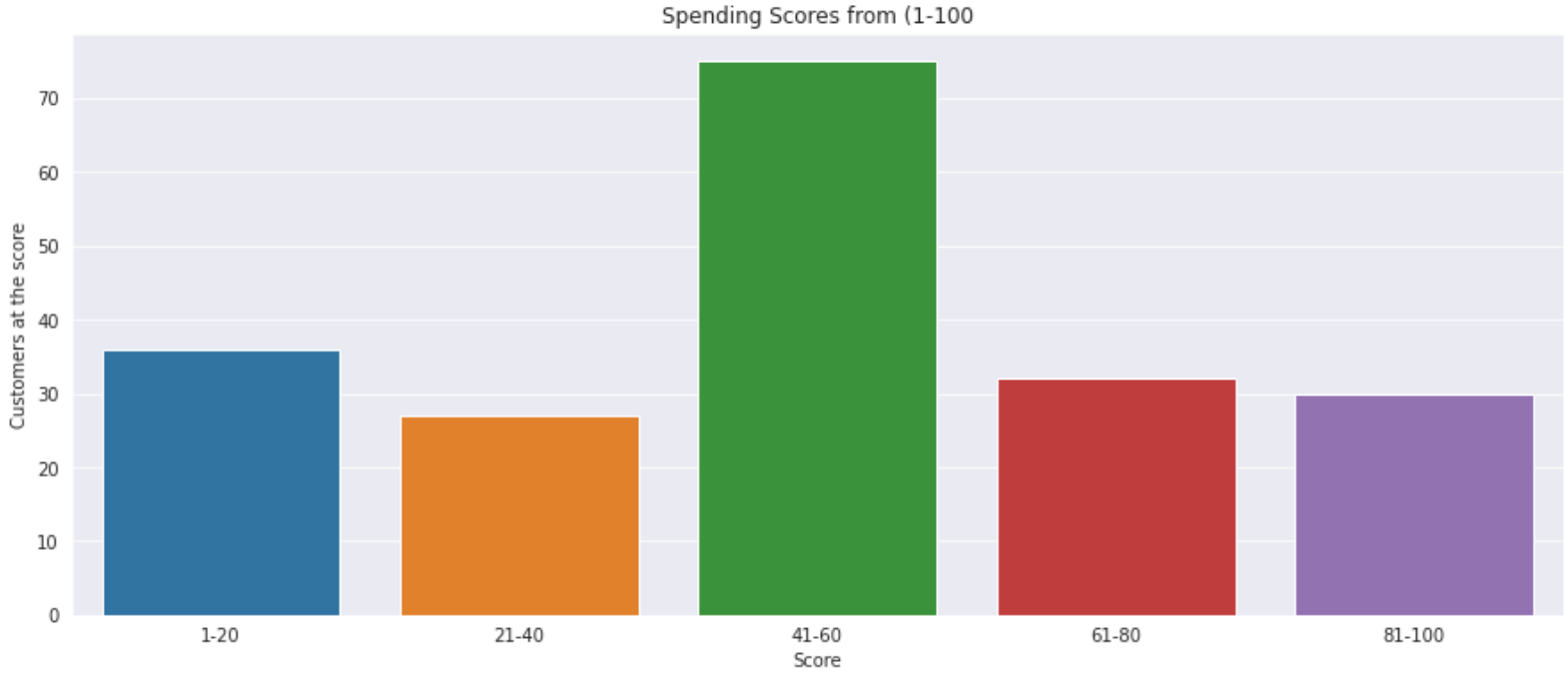

```
#To Identify the Gender Ratio
Gender = df.Gender.value_counts()
sns.set_style("darkgrid")
plt.figure(figsize=(15,6))
sns.barplot(x=Gender.index, y=Gender.values)
plt.show()
```



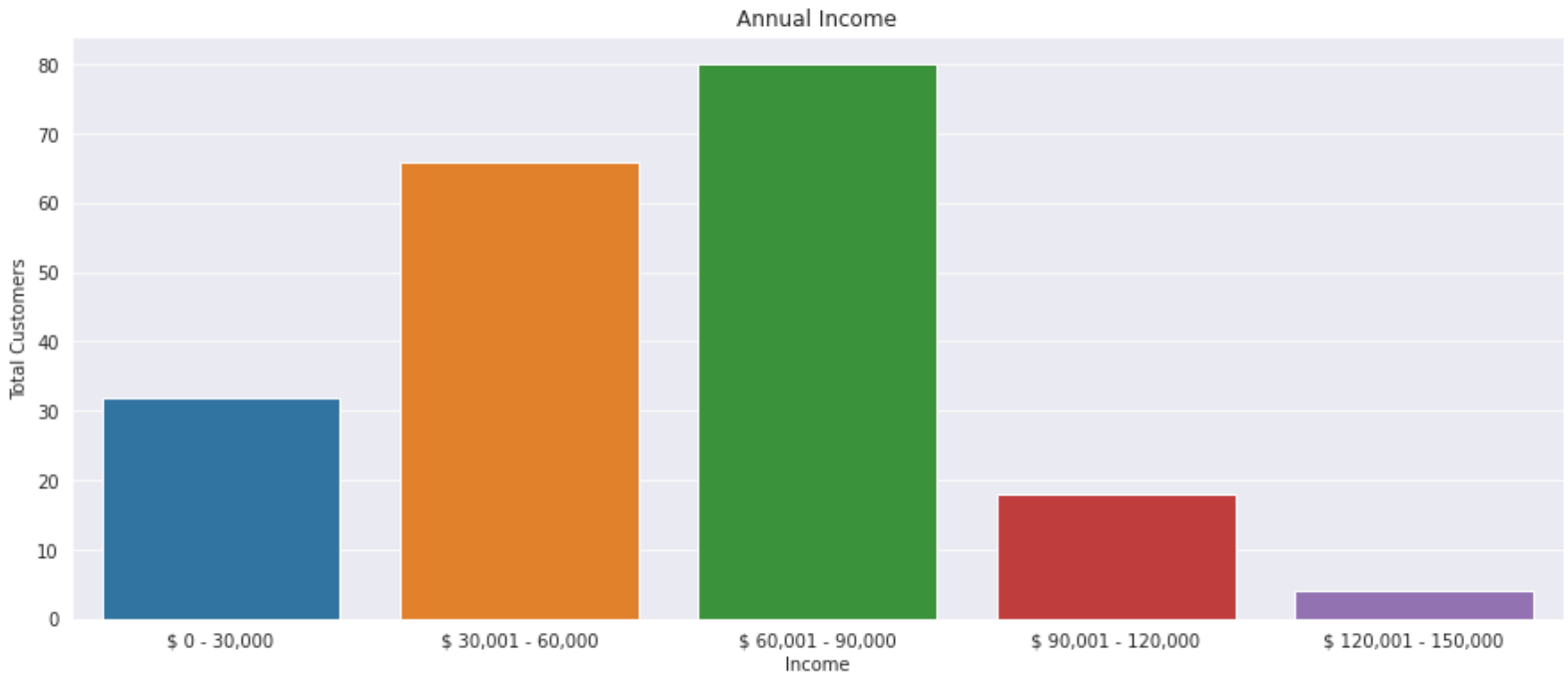

```
#To check the distribution via agegroups
age18_25 = df.Age[(df.Age <= 25 ) & (df.Age >= 18)]
age26_35 = df.Age[(df.Age <= 35 ) & (df.Age >= 26)]
age36_45 = df.Age[(df.Age <= 45 ) & (df.Age >= 36)]
age46_55 = df.Age[(df.Age <= 55 ) & (df.Age >= 46)]
age55_above = df.Age[(df.Age >= 55 )]

x = ["18-25","26-35","36-45","46-55","Above 55"]
y = [len(age18_25.values),len(age26_35.values),len(age36_45.values),len(age46_55.values),len(age55_above.values)]

plt.figure(figsize=(15,6))
sns.barplot(x=x,y=y,palette = "mako")
plt.title("Age Difference across the Customer List")
plt.xlabel("Age")
plt.show()
```




```
#To check the spending score distribution
ss1_20 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 1) & (df["Spending Score (1-100)"]<=20)]
ss21_40 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 21) & (df["Spending Score (1-100)"]<=40)]
ss41_60 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 41) & (df["Spending Score (1-100)"]<=60)]
ss61_80 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 61) & (df["Spending Score (1-100)"]<=80)]
ss81_100 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 81) & (df["Spending Score (1-100)"]<=100)]

ssx = ["1-20","21-40","41-60","61-80","81-100"]
ssy = [len(ss1_20.values),len(ss21_40.values),len(ss41_60.values),len(ss61_80.values),len(ss81_100.values)]

plt.figure(figsize = (15,6))
sns.barplot(x=ssx, y=ssy, palette = "tab10")
plt.title("Spending Scores from (1-100")
plt.xlabel("Score")
plt.ylabel("Customers at the score")
plt.show()
```



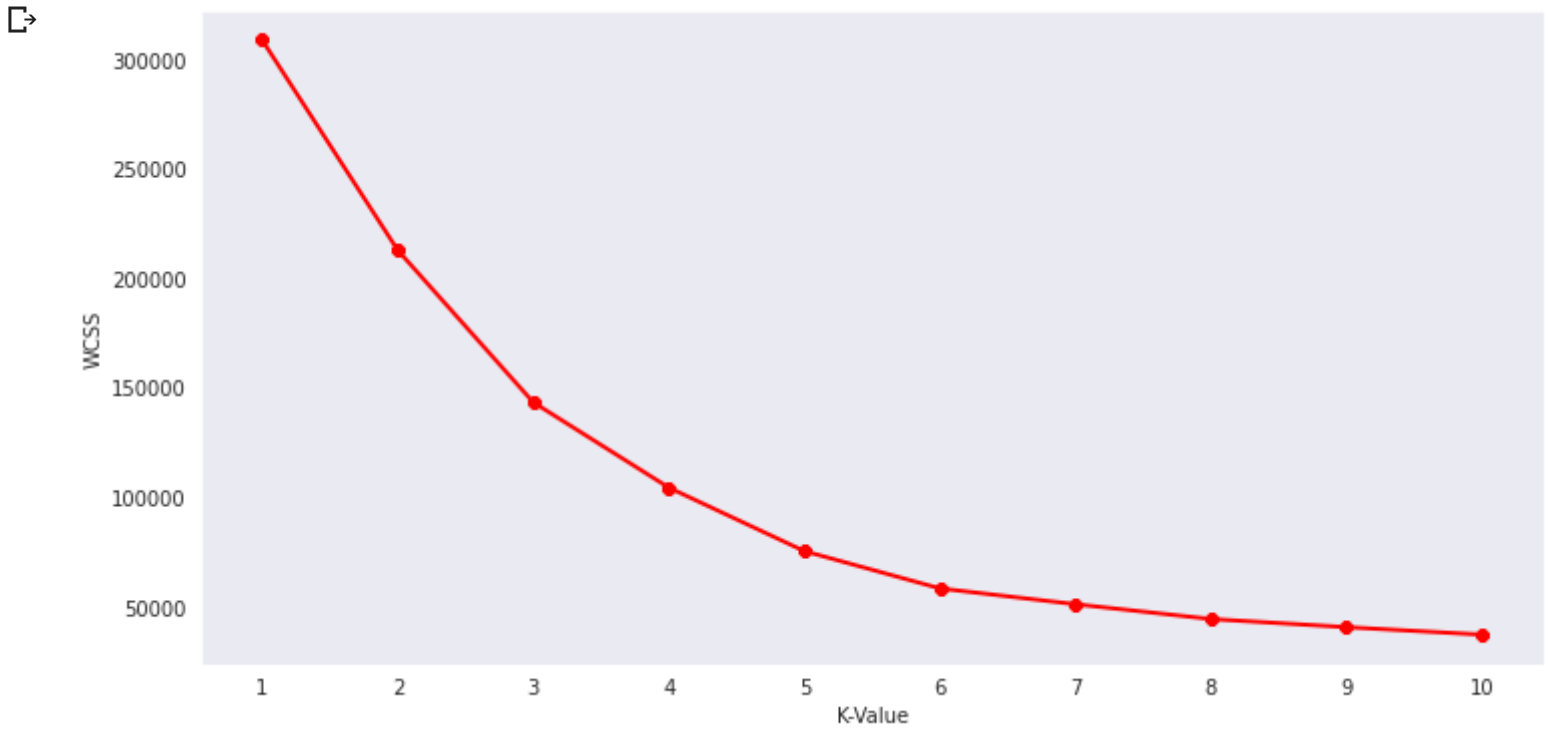

```
#To check the Annual Income Distribution
ai0_30 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 0) & (df["Annual Income (k$)"]<=30)]
ai31_60 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 31) & (df["Annual Income (k$)"]<=60)]
ai61_90 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 61) & (df["Annual Income (k$)"]<=90)]
ai91_120 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 91) & (df["Annual Income (k$)"]<=120)]
ai121_150 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 121) & (df["Annual Income (k$)"]<=150)]

aix = ["$ 0 - 30,000","$ 30,001 - 60,000","$ 60,001 - 90,000","$ 90,001 - 120,000","$ 120,001 - 150,000"]
aiy = [len(ai0_30.values),len(ai31_60.values),len(ai61_90.values),len(ai91_120.values),len(ai121_150.values),]

plt.figure(figsize = (15,6))
sns.barplot(x=aix, y=aiy, palette = "tab10")
plt.title("Annual Income")
plt.xlabel("Income")
plt.ylabel("Total Customers")
plt.show()
```




```
#To apply K-Means in our dataset
from sklearn.cluster import KMeans
```

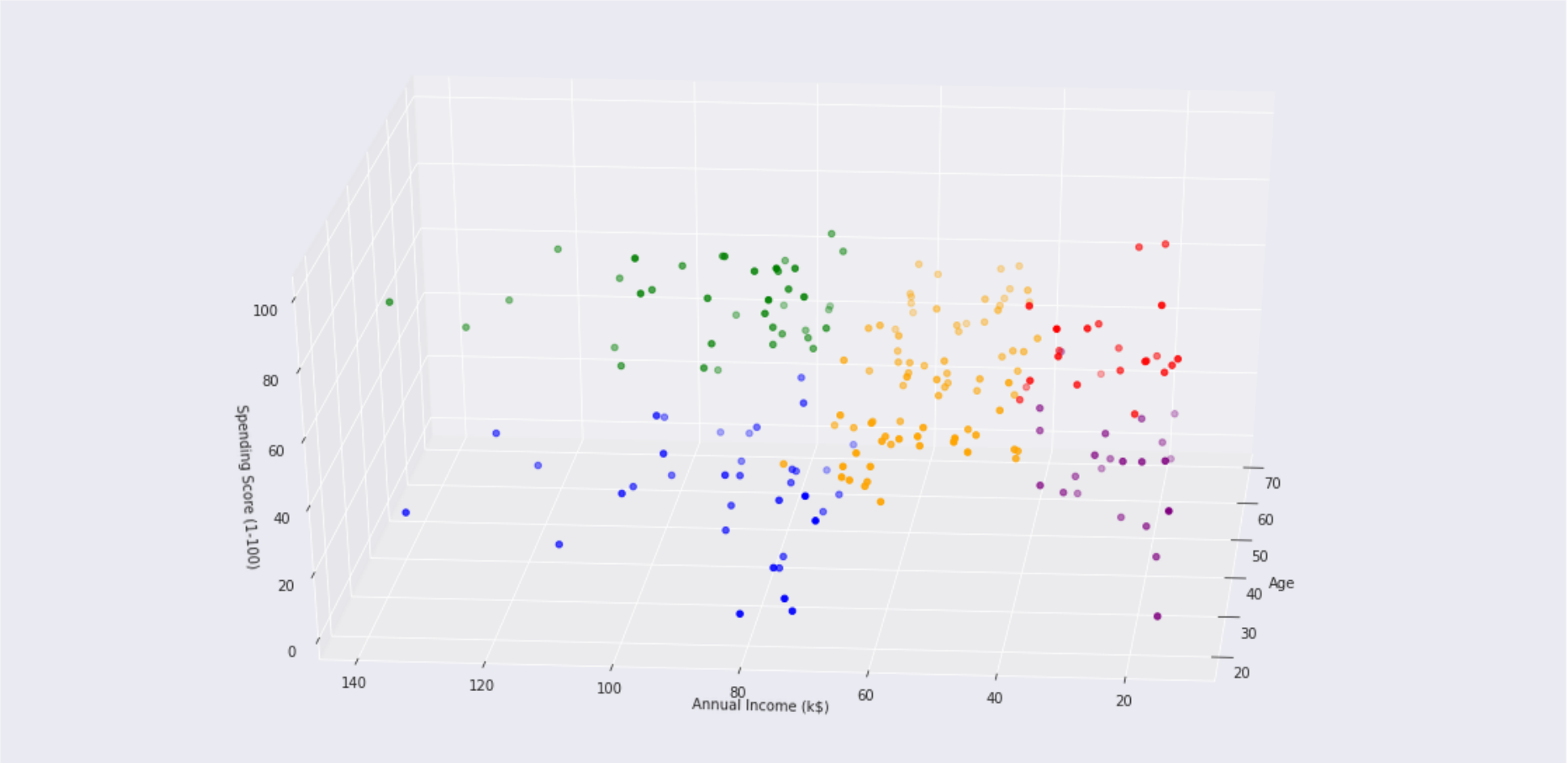
```
#Creating Variables and Assigning Values
import numpy as np
wcss = []
for k in range(1,11):
  Kmeans = KMeans(n_clusters = k, init = "k-means++")
  Kmeans.fit(df.iloc[:,1:])
  wcss.append(Kmeans.inertia_)
plt.figure(figsize = (12,6))
plt.grid()
plt.plot(range(1,11),wcss,linewidth = 2, color = "red", marker = "8")
plt.xlabel("K-Value")
```

```
plt.xticks(np.arange(1,11,1))
plt.ylabel("WCSS")
plt.show()
```




```
#To Create an algo for applying datasets
km = KMeans(n_clusters = 5)
clusters = km.fit_predict(df.iloc[:,1:])
df["label"] = clusters

from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

fig = plt.figure(figsize=(20,10))
axis = fig.add_subplot(111,projection = '3d')
```

```
axis.scatter(df.Age[df.label == 0], df["Annual Income (k$)"][df.label == 0],df["Spending Score (1-100)"][df.label ==0], c="blue")
axis.scatter(df.Age[df.label == 1], df["Annual Income (k$)"][df.label == 1],df["Spending Score (1-100)"][df.label ==1], c="red")
axis.scatter(df.Age[df.label == 2], df["Annual Income (k$)"][df.label == 2],df["Spending Score (1-100)"][df.label ==2], c="green")
axis.scatter(df.Age[df.label == 3], df["Annual Income (k$)"][df.label == 3],df["Spending Score (1-100)"][df.label ==3], c="orange")
axis.scatter(df.Age[df.label == 4], df["Annual Income (k$)"][df.label == 4],df["Spending Score (1-100)"][df.label ==4], c="purple")
axis.view_init(30,185)
plt.xlabel("Age")
plt.ylabel("Annual Income (k$)")
axis.set_zlabel("Spending Score (1-100)")
plt.show()
```


Spending Score (1-100)
100
80
60
40
20
0
140
120
100
80
60
40
20
Annual Income (k$)
70
60
50
40
30
20
Age