**CODE FOR THE PROJECT**

import pandas as pd

import numpy as np // Importing Libraries

import matplotlib.pyplot as plt

import seaborn as sns

df=pd.read\_csv("E:\Exposy\Mall\_Customers.csv") // inputting the dataset

df.head()

df.describe

df.isnull().sum()

df.drop(("CustomerID") , axis=1 , inplace=True)

plt.figure(1,figsize=(13,5))

n=0

for x in ['Age','Annual Income (k$)','Spending Score (1-100)']: // Displaying the distplot for

n+=1 age , Income , Spending

plt.subplot(1 , 3, n) score

plt.subplots\_adjust(hspace=0.5 , wspace=0.5)

sns.distplot(df[x] , bins=20)

plt.show()

plt.figure(figsize=(10,6))

plt.title("Ages Frequency")

sns.axes\_style("dark") // Displaying the axis plot for Age

sns.violinplot(y=df["Age"])

plt.show()

plt.figure(figsize=(15,6))

plt.subplot(1,2,1)sns.boxplot(y=df["Spending Score (1-100)"], color="orange")

plt.subplot(1,2,2) // Displaying

sns.boxplot(y=df["Annual Income (k$)"]) the boxplot for

plt.title("") male and

plt.show() female

genders = df.Gender.value\_counts()

sns.set\_style("darkgrid") //Displaying barplot of gender vs

plt.figure(figsize=(10,4)) spending Score

sns.barplot(x=genders.index, y=genders.values)

plt.show()

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)] //Displaying barplot of age vs

age46\_55 = df.Age[(df.Age <= 55) & (df.Age >= 46)] no of customers

age56\_65 = df.Age[(df.Age <= 65) & (df.Age >= 56)]

age65above = df.Age[df.Age >= 65]

x = ["18-25","26-35","36-45","46-55","56-65","65+"]

y = [len(age18\_25.values),len(age26\_35.values),len(age36\_45.values),len(age46\_55.values),len(age56\_65.values),len(age65above.values)]

ss1\_25 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 1) & (df["Spending Score (1-100)"] <= 25)]

ss26\_50 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 26) & (df["Spending Score (1-100)"] <= 50)]

ss51\_75 = df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 51) & (df["Spending Score (1-100)"] <= 75)]

ss76\_100= df["Spending Score (1-100)"][(df["Spending Score (1-100)"] >= 76) & (df["Spending Score (1-100)"] <= 100)]

ssx = ["1-25", "26-50", "51-75", "76-100"]

ssy = [len(ss1\_25.values), len(ss26\_50.values), len(ss51\_75.values), len(ss76\_100.values)]

plt.figure(figsize=(15,6))

sns.barplot(x=ssx, y=ssy, palette="nipy\_spectral\_r")

plt.title("Spending Scores")

plt.xlabel("Score")

plt.ylabel("Number of Customer Having the Score")

plt.show()

plt.figure(figsize=(15,7)) // Displaying barplot b/w Spending

sns.barplot(x=x, y=y, palette="rocket") score vs No of customers

plt.title("Number of Customer and Ages")

plt.xlabel("Age")

plt.ylabel("Number of Customer")

plt.show()

ai0\_25 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 0) & (df["Annual Income (k$)"] <= 25)]

ai26\_50 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 26) & (df["Annual Income (k$)"] <= 50)]

ai51\_75 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 51) & (df["Annual Income (k$)"] <= 75)]

ai76\_100 = df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 76) & (df["Annual Income (k$)"] <= 100)]

ai101\_125= df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 101) & (df["Annual Income (k$)"] <= 125)]

ai126\_150= df["Annual Income (k$)"][(df["Annual Income (k$)"] >= 126) & (df["Annual Income (k$)"] <= 150)]

aix = ["$ 0 - 25K$", "$ 25,001 - 50K$", "$ 50,001 - 75K$", "$ 75,001 - 100K", "$ 100,001 - 125K$","125,001 - 150K$"]

aiy = [len(ai0\_25.values), len(ai26\_50.values), len(ai51\_75.values), len(ai76\_100.values), len(ai101\_125.values),len(ai126\_150.values)]

plt.figure(figsize=(15,6))

sns.barplot(x=aix, y=aiy, palette="Set2")

plt.title("Annual Incomes") // Displaying the barplot b/w

plt.xlabel("Income") Annual Income vs No of Customers

plt.ylabel("Number of Customer")

plt.show()

from sklearn.cluster import KMeans

wcss = []

for k in range(1,11):

kmeans = KMeans(n\_clusters=k, init="k-means++")

kmeans.fit(df.iloc[:,1:])

wcss.append(kmeans.inertia\_) // Calculating the Value of K using the WSS

plt.figure(figsize=(12,6)) and plotting K against WSS ( Elbow

plt.grid() Method )

plt.plot(range(1,11),wcss, linewidth=2, color="black", marker ="8")

plt.xlabel("K Value")

plt.xticks(np.arange(1,11,1))

plt.ylabel("WCSS")

plt.show()

km = KMeans(n\_clusters=5)

clusters = km.fit\_predict(df.iloc[:,1:])

df["label"] = clusters

from mpl\_toolkits.mplot3d import Axes3D // Implementing K means and displaying the

import matplotlib.pyplot as plt 3D axis that shows the segmentation

import numpy as np of the customers

import pandas as pd

fig = plt.figure(figsize=(20,10))

ax = fig.add\_subplot(111, projection='3d')

ax.scatter(df.Age[df.label == 0], df["Annual Income (k$)"][df.label == 0], df["Spending Score (1-100)"][df.label == 0], c='blue', s=60)

ax.scatter(df.Age[df.label == 1], df["Annual Income (k$)"][df.label == 1], df["Spending Score (1-100)"][df.label == 1], c='red', s=60)

ax.scatter(df.Age[df.label == 2], df["Annual Income (k$)"][df.label == 2], df["Spending Score (1-100)"][df.label == 2], c='green', s=60)

ax.scatter(df.Age[df.label == 3], df["Annual Income (k$)"][df.label == 3], df["Spending Score (1-100)"][df.label == 3], c='orange', s=60)

ax.scatter(df.Age[df.label == 4], df["Annual Income (k$)"][df.label == 4], df["Spending Score (1-100)"][df.label == 4], c='purple', s=60)

ax.view\_init(30, 185)

plt.xlabel("Age")

plt.ylabel("Annual Income (k$)")

ax.set\_zlabel('Spending Score (1-100)')

plt.show()