import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.cluster import KMeans

# loading the data from csv file to a Pandas DataFrame

customer\_data = pd.read\_csv('/content/Mall\_Customers.csv')

# first 5 rows in the dataframe

customer\_data.head()

# finding the number of rows and columns

customer\_data.shape

# getting some informations about the dataset

customer\_data.info()

# checking for missing values

customer\_data.isnull().sum()

X = customer\_data.iloc[:,[3,4]].values

print(X)

# finding wcss value for different number of clusters

wcss = []

for i in range(1,11):

kmeans = KMeans(n\_clusters=i, init='k-means++', random\_state=42)

kmeans.fit(X)

wcss.append(kmeans.inertia\_)

kmeans = KMeans(n\_clusters=5, init='k-means++', random\_state=0)

# return a label for each data point based on their cluster

Y = kmeans.fit\_predict(X)

print(Y)

# plotting all the clusters and their Centroids

plt.figure(figsize=(8,8))

plt.scatter(X[Y==0,0], X[Y==0,1], s=50, c='green', label='Cluster 1')

plt.scatter(X[Y==1,0], X[Y==1,1], s=50, c='red', label='Cluster 2')

plt.scatter(X[Y==2,0], X[Y==2,1], s=50, c='yellow', label='Cluster 3')

plt.scatter(X[Y==3,0], X[Y==3,1], s=50, c='violet', label='Cluster 4')

plt.scatter(X[Y==4,0], X[Y==4,1], s=50, c='blue', label='Cluster 5')

# plot the centroids

plt.scatter(kmeans.cluster\_centers\_[:,0], kmeans.cluster\_centers\_[:,1], s=100, c='cyan', label='Centroids')

plt.title('Customer Groups')

plt.xlabel('Annual Income')

plt.ylabel('Spending Score')

plt.show()