# Data Preprocessing Template

# Importing the libraries

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

# Importing the dataset

dataset = pd.read\_csv('Mall\_Customers.csv')

x = dataset.iloc[:, [3,4]].values

#using the elbow method to find optimal no of clusters

from sklearn.cluster import KMeans

wcss=[]

for i in range(1,11):

kmeans=KMeans(n\_clusters=i,init='k-means++',n\_init=10,max\_iter=300,random\_state=0)

kmeans.fit(x)

wcss.append(kmeans.inertia\_)

plt.plot(range(1,11),wcss)

plt.title('The elbow method')

plt.xlabel('No of clusters')

plt.ylabel('wcss')

plt.show()

#fitting the kmeans

kmeans=KMeans(n\_clusters=5,init='k-means++',n\_init=10,max\_iter=300,random\_state=0)

y\_kmeans=kmeans.fit\_predict(x)

#visualising the cluster

plt.scatter(x[y\_kmeans==0,0], x[y\_kmeans==0,1],s=100,c='red',label='Careful')

plt.scatter(x[y\_kmeans==1,0], x[y\_kmeans==1,1],s=100,c='blue',label='Standard')

plt.scatter(x[y\_kmeans==2,0], x[y\_kmeans==2,1],s=100,c='green',label='Target')

plt.scatter(x[y\_kmeans==3,0], x[y\_kmeans==3,1],s=100,c='black',label='Careless')

plt.scatter(x[y\_kmeans==4,0], x[y\_kmeans==4,1],s=100,c='pink',label='Sensible')

plt.scatter(kmeans.cluster\_centers\_[:,0],kmeans.cluster\_centers\_[:,1],s=300,c='yellow',label='centroids')

plt.title('clusters of clients')

plt.xlabel('annual income')

plt.ylabel('spending score')

plt.legend()

plt.show()