import pandas as pd

import matplotlib.pyplot as plt

df=pd.read\_csv("mall.csv")

df

x=df.iloc[:,3:]

x

df.columns

plt.title('unclustered data')

plt.xlabel('Annual')

plt.ylabel('spending')

plt.scatter(x['Annual Income (k$)'],x['Spending Score (1-100)'])

from sklearn.cluster import KMeans

km = KMeans(n\_clusters=3)

x.shape

km.fit\_predict(x)

#ssm sum sq

km.inertia\_

sse=[]

for k in range (1,16):

km=KMeans(n\_clusters=k)

km.fit\_predict(x)

sse.append(km.inertia\_)

sse

plt.title('elbow method')

plt.xlabel('value of k')

plt.ylabel('sse')

plt.grid()

plt.xticks(range(1,16))

plt.plot(range(1,16),sse,color='r')

from sklearn.metrics import silhouette\_score

silh=[]

for k in range (2,16):

km=KMeans(n\_clusters=k)

labels =km.fit\_predict(x)

score=silhouette\_score(x,labels)

silh.append(score)

silh

plt.title('silhouette method')

plt.xlabel('value of k')

plt.ylabel('silh')

plt.grid()

plt.xticks(range(2,16))

plt.bar(range(2,16),silh,color='r')

km = KMeans(n\_clusters=5)

labels=km.fit\_predict(x)

labels

plt.title('clustered data')

plt.xlabel('Annual')

plt.ylabel('spending')

plt.scatter(x['Annual Income (k$)'],x['Spending Score (1-100)'],c=labels)

km.inertia\_

km.labels\_

df[labels==4]

km.predict([[70,29]])