* **MODUL 11**

1. Lakukan proses analisis data dengan menggunakan metode Klasterisasi K-Means untuk data berikut:
2. Segmentasi Pelanggan adalah salah satu aplikasi paling penting dalam dunia bisnis, secara khusus untuk analisis Marketing. Dengan menggunakan teknik pengelompokan (clustering), perusahaan dapat mengidentifikasi beberapa segmen pelanggan yang memungkinkan mereka untuk menargetkan basis pengguna (konsumen) potensial. Dalam penlitian ini, akan digunakan pengelompokan K-means yang merupakan algoritma penting untuk pengelompokan dataset yang tidak berlabel (unsuperpised). Sebelum terlalu jauh dalam projek ini, sebenarnya apa itu segmentasi pelanggan. Gunakan dataset : Pengunjung\_Mall.csv yang dapat anda download dari classroom.

* **Kode**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import datetime as dt

import sklearn

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

from sklearn.metrics import silhouette\_score

df = pd.read\_csv("Documents/Tugas Kuliah/5P42/Data Science/Praktek/Pengunjung\_Mall.csv")

df

df.info()

df.describe()

dataset = pd.read\_csv("Documents/Tugas Kuliah/5P42/Data Science/Praktek/Pengunjung\_Mall.csv")

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

from sklearn.cluster import KMeans

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\_)

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

plt.title('Metode Elbow')

plt.xlabel('Jumlah Clusters')

plt.ylabel('WCSS')

plt.show()

kmeans = KMeans(n\_clusters = 5, init = 'k-means++', random\_state = 42)

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

plt.scatter(X[y\_kmeans == 0,0], X[y\_kmeans == 0,1], s = 100, c = 'yellow', label = 'Cluster 1')

plt.scatter(X[y\_kmeans == 1,0], X[y\_kmeans == 1,1], s = 100, c = 'cyan', label = 'Cluster 2')

plt.scatter(X[y\_kmeans == 2,0], X[y\_kmeans == 2,1], s = 100, c = 'black', label = 'Cluster 3')

plt.scatter(X[y\_kmeans == 3,0], X[y\_kmeans == 3,1], s = 100, c = 'green', label = 'Cluster 4')

plt.scatter(X[y\_kmeans == 4,0], X[y\_kmeans == 4,1], s = 100, c = 'blue', label = 'Cluster 5')

plt.scatter(kmeans.cluster\_centers\_[:, 0], kmeans.cluster\_centers\_[:, 1], s = 250, c = 'brown', label = 'Centroids')

plt.title('Cluster Pelanggan')

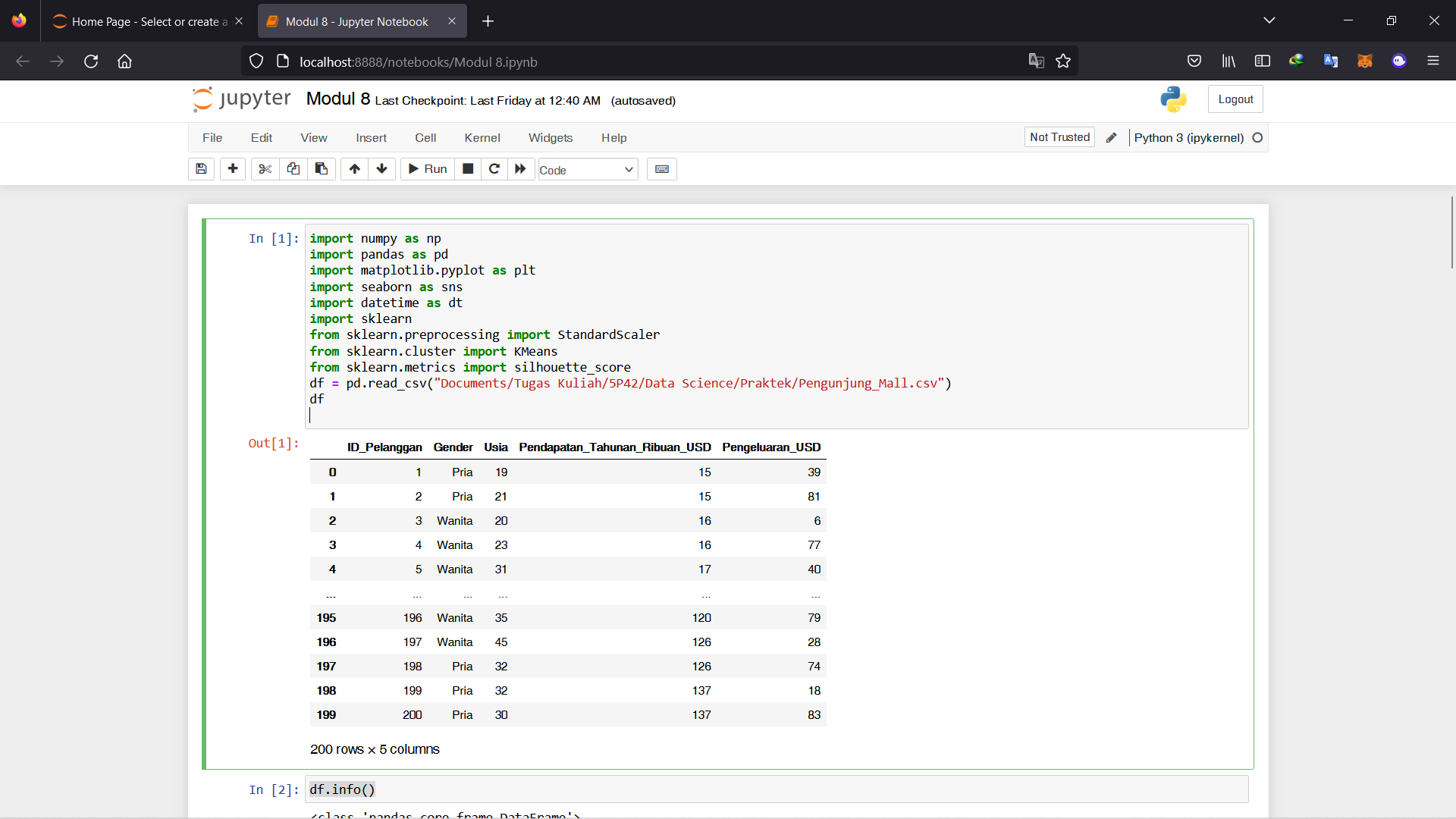
plt.xlabel('Pendapatan Tahunan (perjuta)')

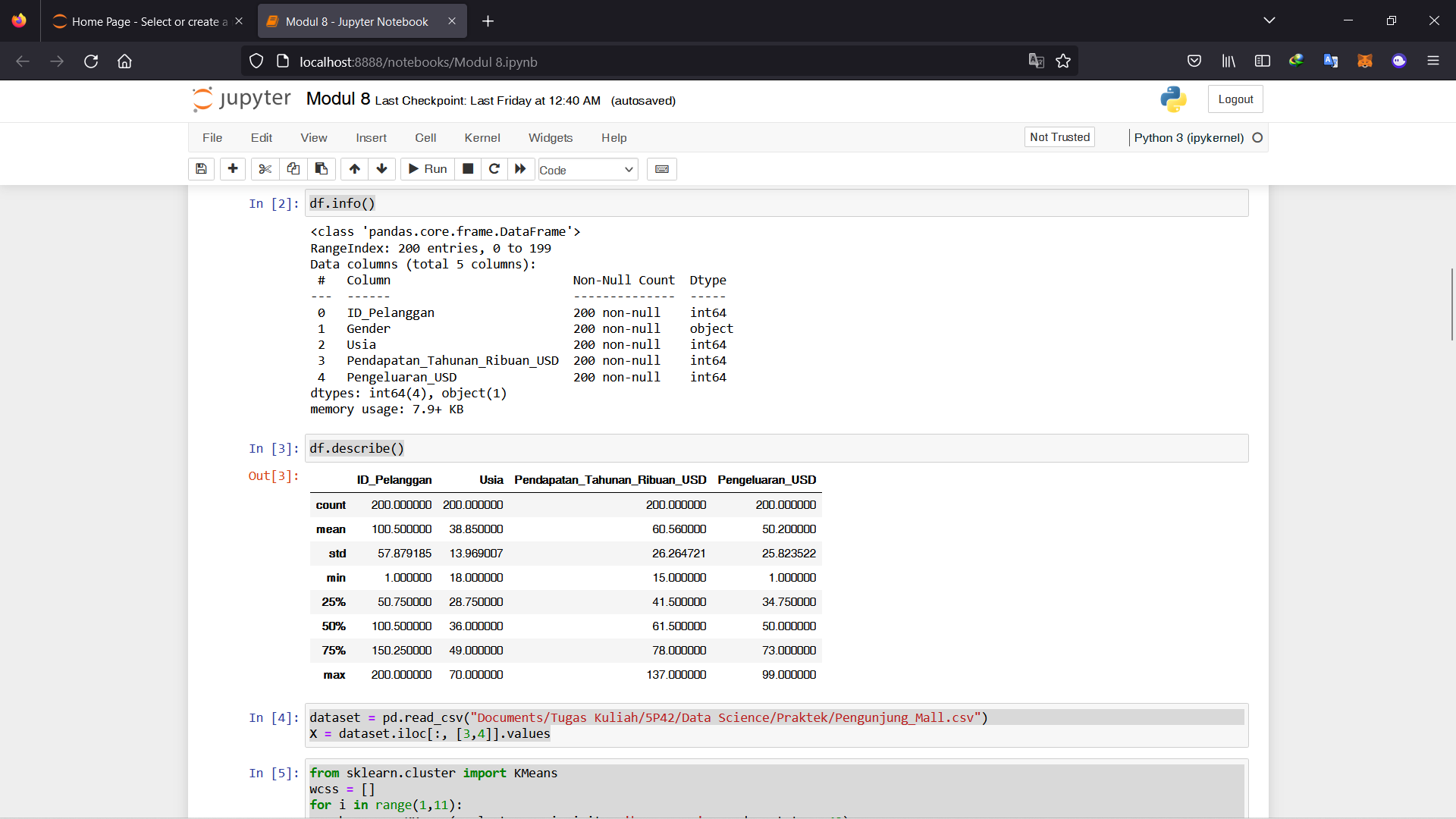
plt.ylabel('Rating Pengeluaran')

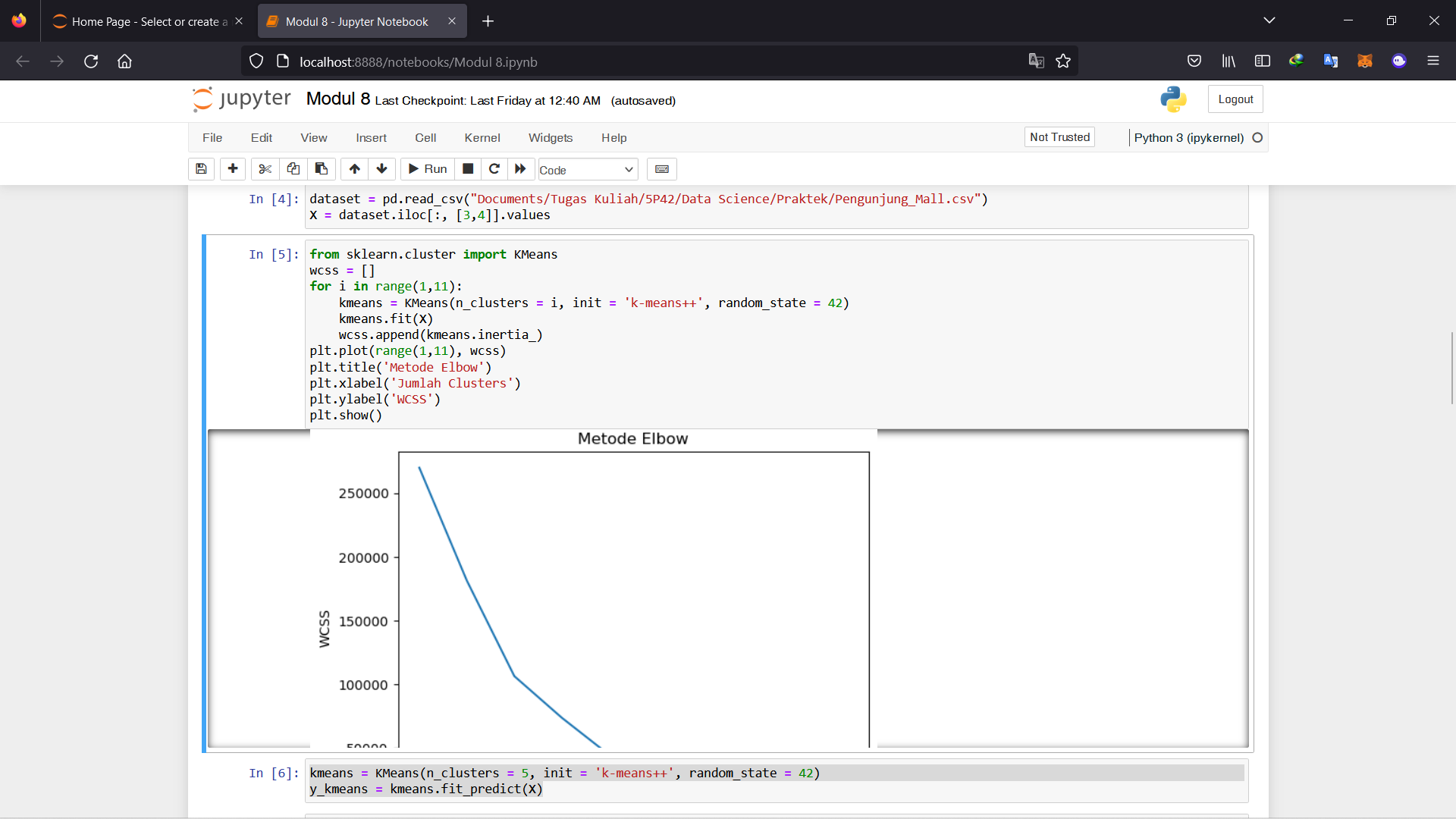
plt.legend()

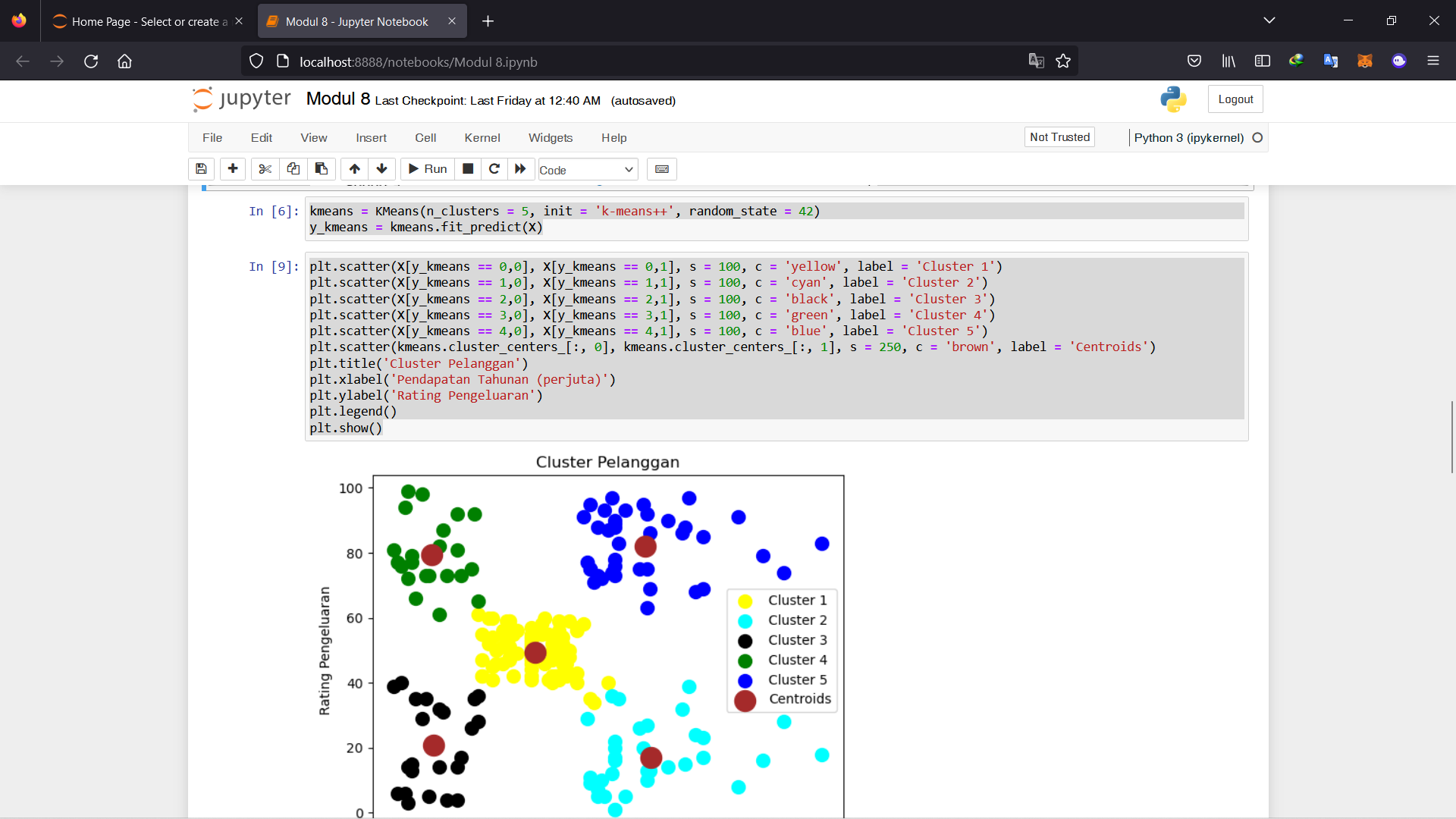
plt.show()

* **Hasil**









1. Dari dataset yang berisi penghasilan satu keluarga dan jumlah anggota keluarga. Lakukan analiasis clustering tingkat penghasilan per jumlah anggota keluarga. Berikut isi dari dataset penghasilan-per-keluarga.csv:

* **Kode**

dp = pd.read\_csv("Documents/Tugas Kuliah/5P42/Data Science/Praktek/penghasilan-per-keluarga.csv")

dp

dp.info()

dp.describe()

dp.isna().sum()

dp = np.array(dp)

print(dp)

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

x\_scaled = scaler.fit\_transform(dp)

x\_scaled

kmeans = KMeans(n\_clusters = 2, random\_state=7)

kmeans.fit(x\_scaled)

print(kmeans.cluster\_centers\_)

print(kmeans.labels\_)

new\_dp = pd.DataFrame(data=dp, columns=['JumlahAnggotaKeluarga',

'Penghasilan (juta rupiah)'])

new\_dp['kluster'] = kmeans.labels\_

new\_dp

* **Hasil**

