**PRAKTIKUM**

**BIG DATA ANALYTICS**

**RESPONSI**

****

**Disusun Oleh :**

**NAMA** **: Raden Isnawan Argi Aryasatya**

**NIM** **: 195410257**

**JURUSAN** **: Informatika**

**JENJANG** **: S1**

**KELAS : 5**

**Laboratorium Terpadu**

**UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA**

**(UTDI)**

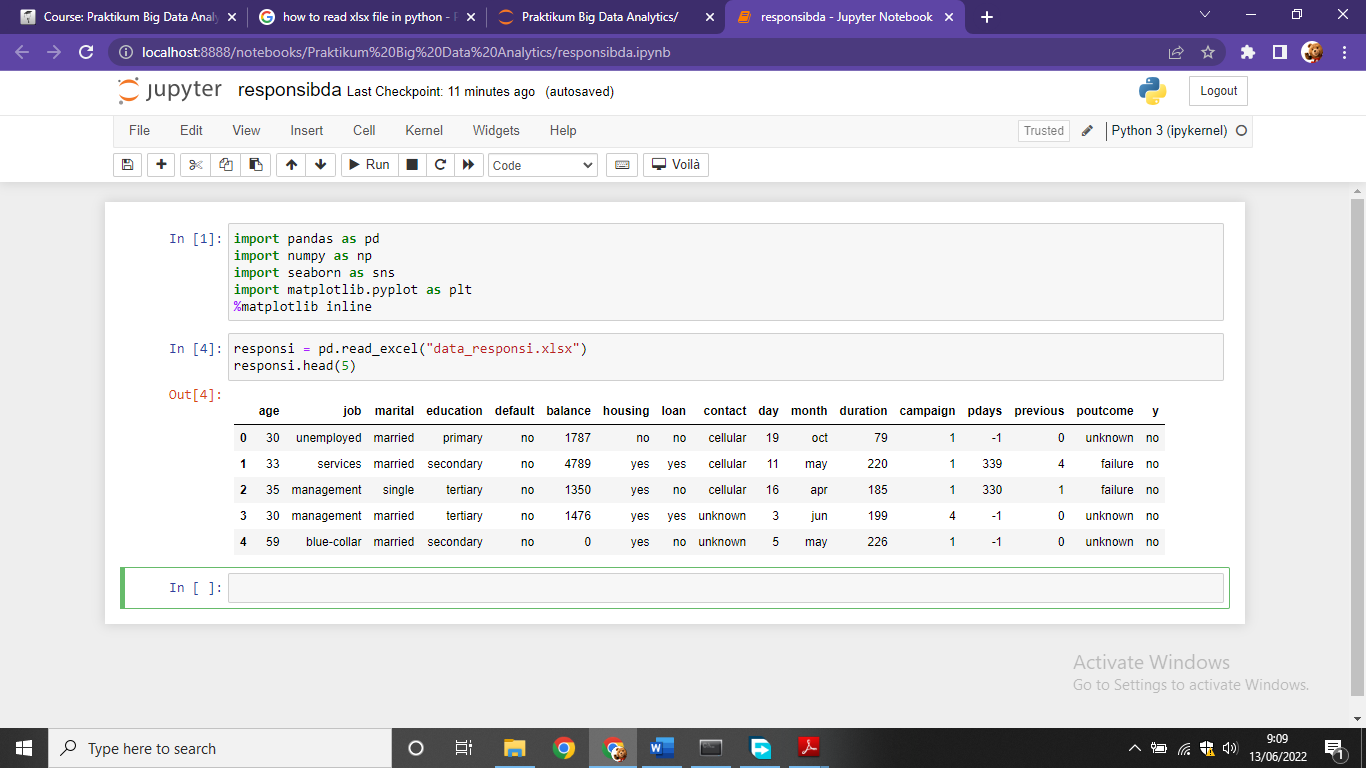
**2021/2022**

RESPONSI

1. **Lakukan ploting data untuk variabel Age dan balance dgn ketentuan sumbu age usia dan sumbu y balance**

Jawab:

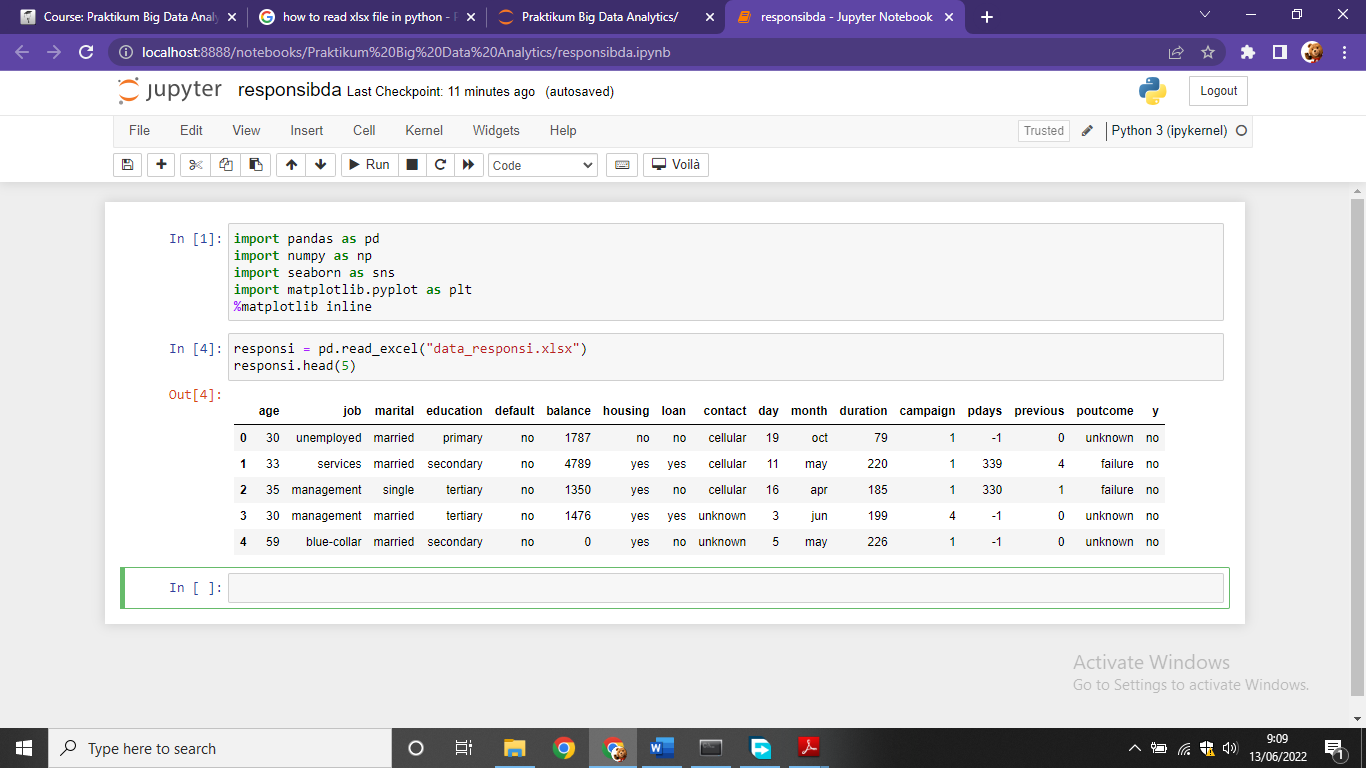
Pertama, import libraries yang diperlukan



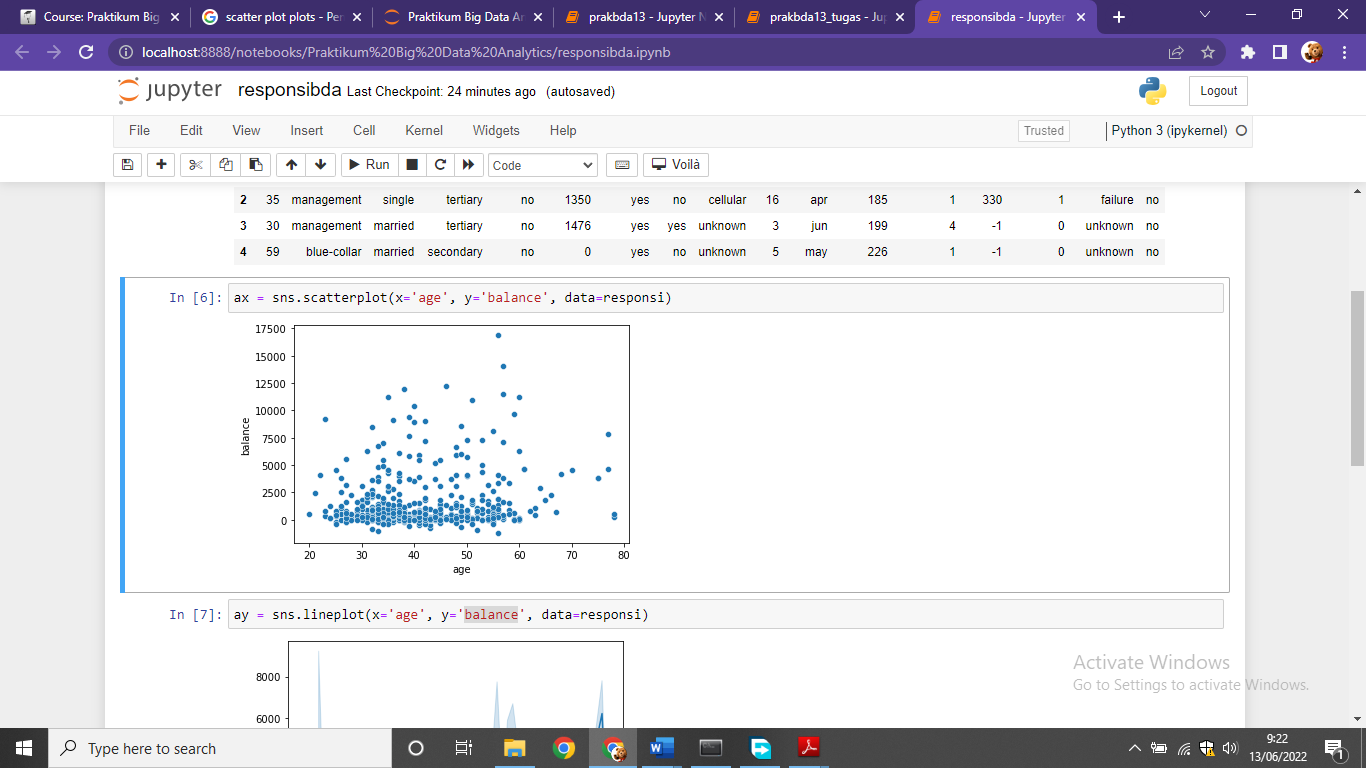
Penjelasan:

* import numpy yang diwakili dengan variabel bernama np yang digunakan untuk bekerja dengan array
* import pandas dengan diwakili variabel pd sebagai tool untuk memanipulasi dan menganalisis data
* import seaborn dengan diwakili variabel sns
* deklarasi sns.set tanpa memasukkan value pada argumen
* kita juga akan menggunakan plotly yang merupakan library plot open source interaktif yang mendukung lebih dari 40 jenis chart unik yang mencakup berbagai grafik penggunaan statistik, keuangan, geografis, ilmiah, dan 3 dimensi.

Baca dataset dan tampilkan 5 data teratas



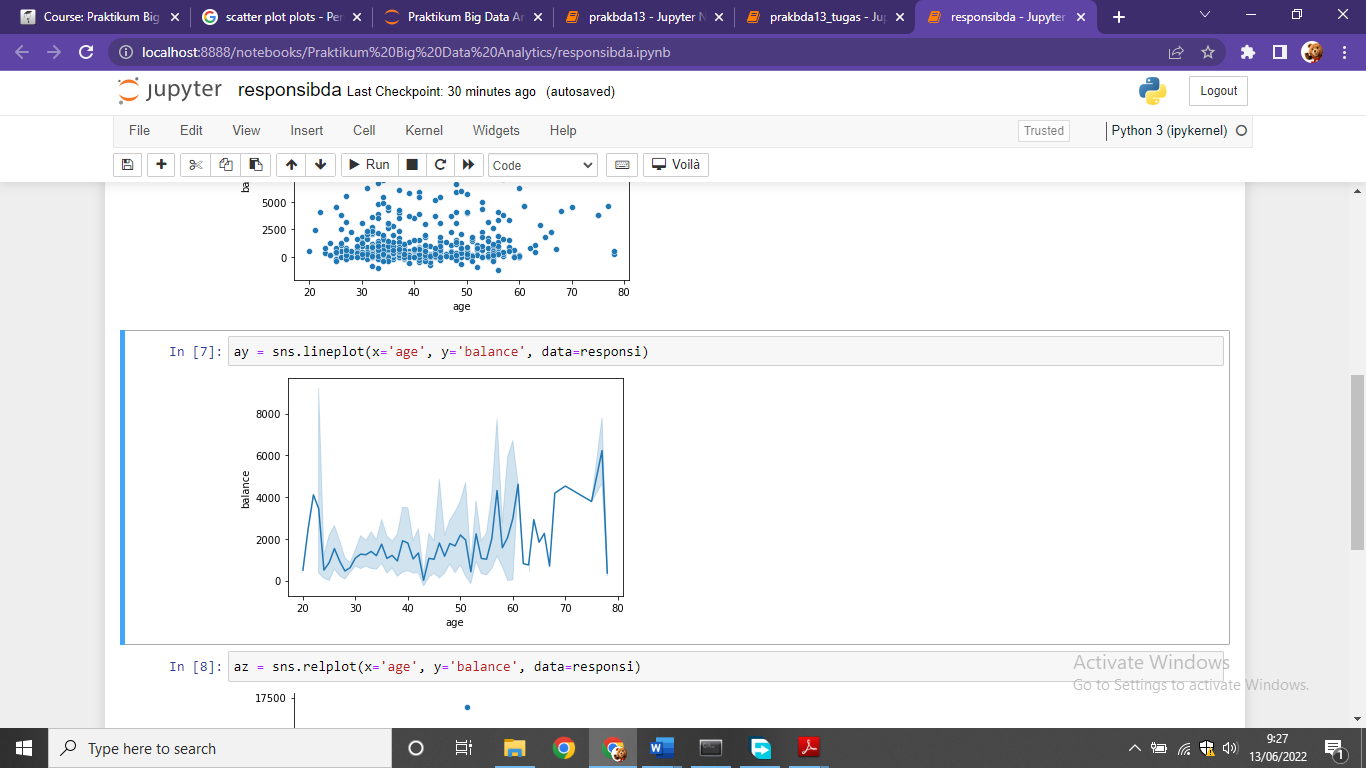
Tampilkan variabel ‘age’ dan ‘balance’ dengan scatter plot



Penjelasan:

Dengan mengambil dua variabel yaitu age (x) dan balance(y), kita tampilkan sebaran data dengan fungsi seaborn yaitu scatterplot. Scatter plot menggunakan dot (titik) untuk merepresentasikan nilai dari dua variabel numerik yang diplot pada sisi horizontal (x) dan vertikal (y). Dari grafik di atas, bisa disimpulkan bahwa mayoritas orang memiliki balance sebesar 25000 kebawah. Orang yang balance nya mencapai 150000 keatas hanyalah di kategori usia 50-60.

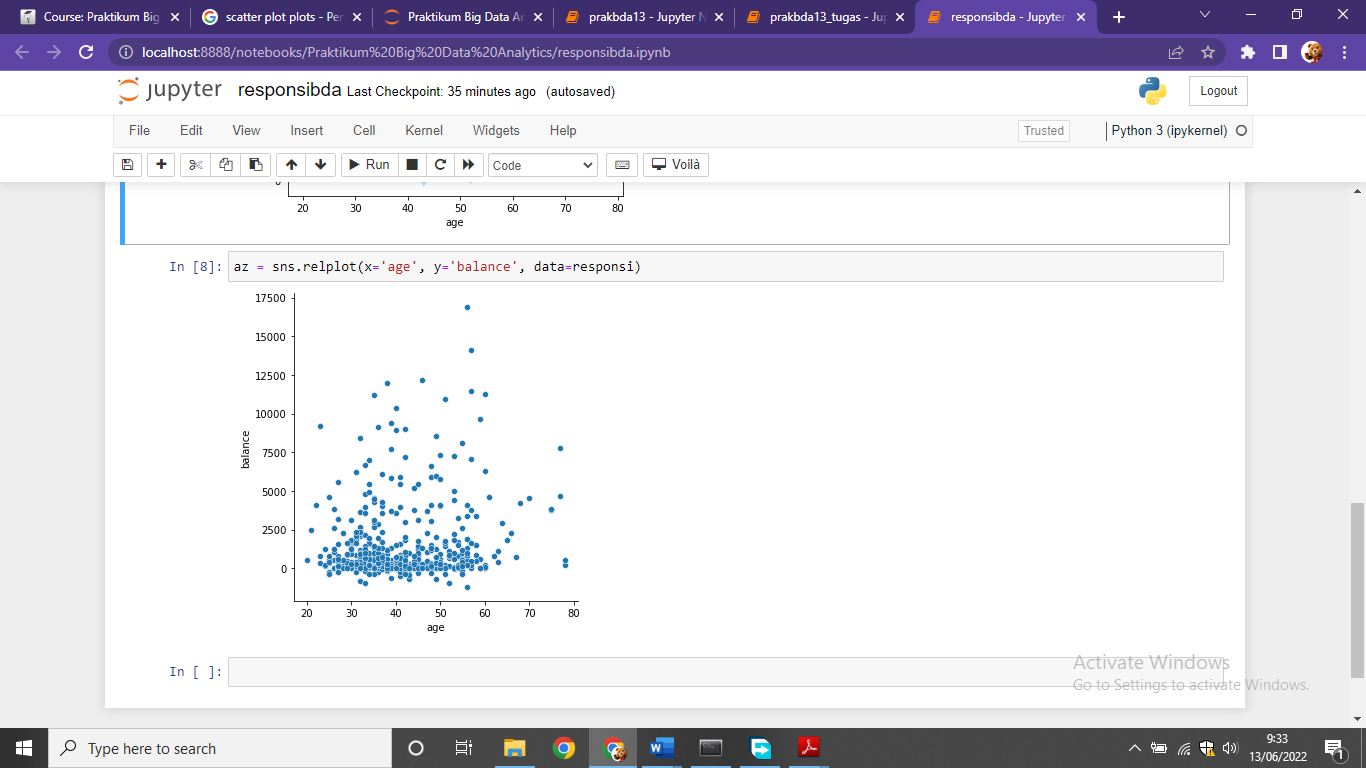
Tampilkan variabel ‘age’ dan ‘balance’ dengan line plot



Penjelasan:

Dengan menggunakan dua variabel yang sama dengan tadi, yaitu age dan balance, kita tampilkan grafik datanya menggunakan lineplot. Lineplot() merupakan salahsatu fungsi visualisasi data seaborn yang merupakan jenis plot dasar yang berguna untuk menampilkan grafik informasi berupa rangkaian titik data yang terhubung dengan segmen garis lurus. Dari grafik tersebut bisa kita simpulkan bahwa balance tertinggi dimiliki oleh usia 80, balance terendah dimiliki oleh usia 40an, dan mayoritas usia memiliki balance sebesar 20000an.

Tampilkan variabel ‘age’ dan ‘balance’ dengan rel plot



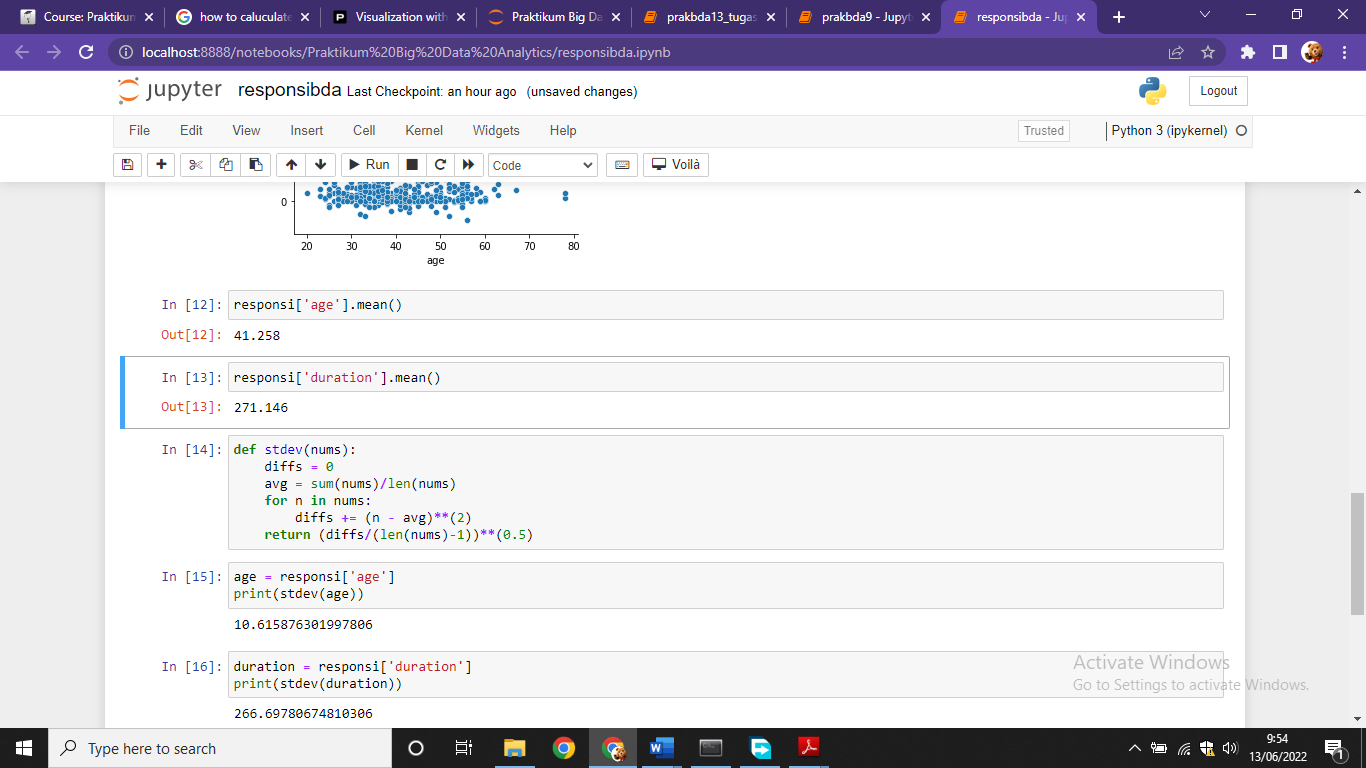
Penjelasan:

Pada grafik tersebut, kita menganalisis relasi antara 2 quantitive variable dengan relplot. Dengan kata lain, Pada kode di atas kita menggunakan relplot yang memiliki fungsi untuk menyediakan akses ke beberapa fungsi sumbu yang berbeda yang menunjukkan hubungan antara dua variabel dengan pemetaan semantik dari subset. Pada kode di atas, kita menunjukkan relationship di antara dua kolom yaitu ‘age’ dan ‘balance’.

1. **Tampilkan statistic deskriptif yang terdiri nilai rata2, dev standar, nilai max, min untuk variabel age dan duration**

Jawab:

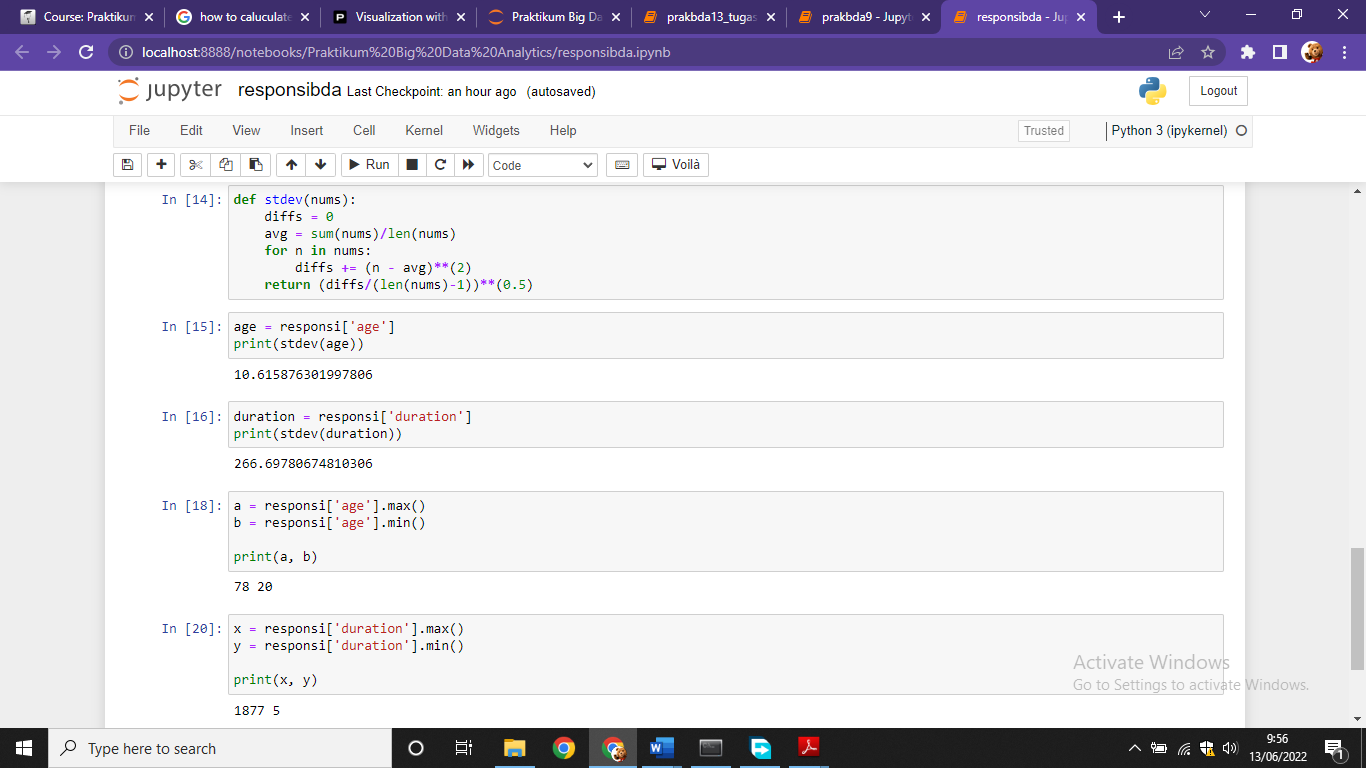
Rata-rata



Penjelasan:

Pada kode tersebut kita mencari rata-rata pada dua variabel dengan fungsi mean(). Variabel age memiliki rata-rata 41,258 dan variabel duration memiliki rata-rata 271,146.

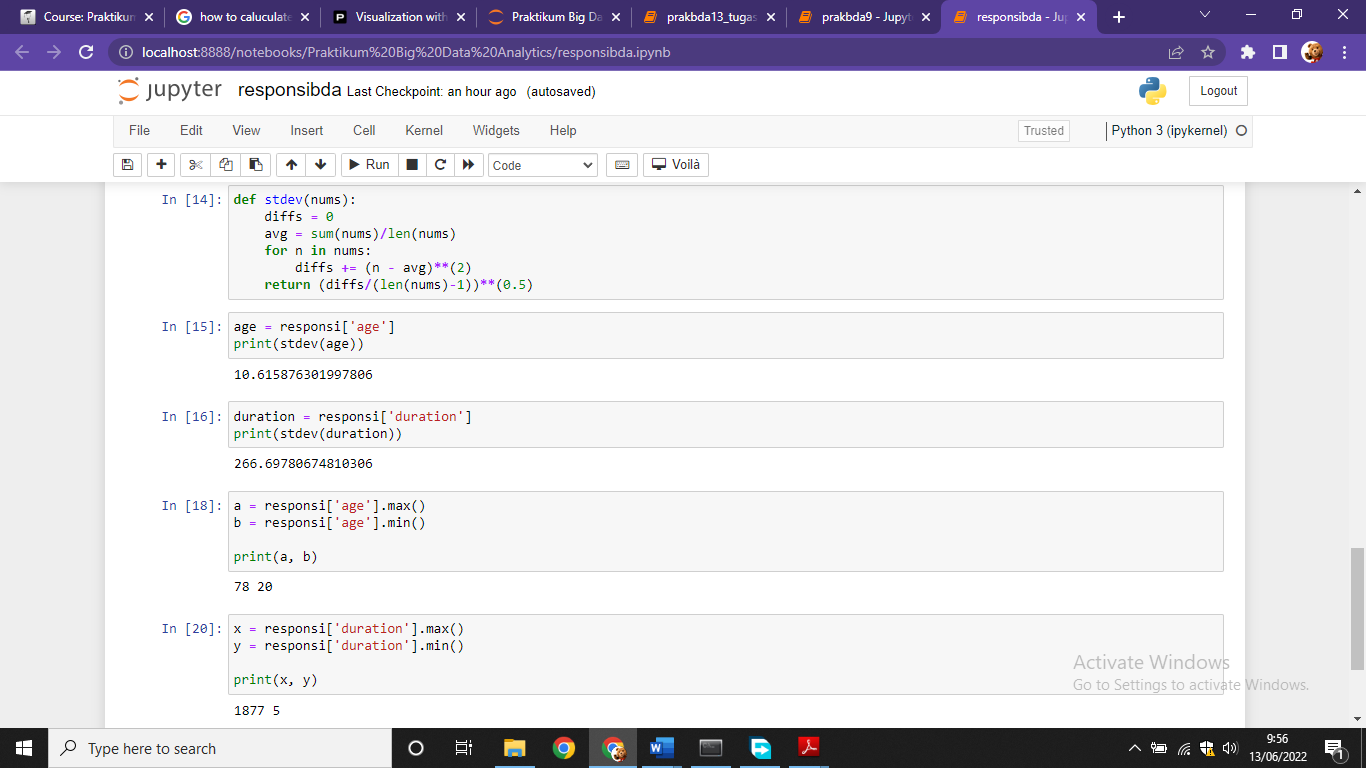
Deviasi standar



Penjelasan:

standar deviasi adalah nilai statistik yang dipakai guna menentukan seberapa dekat data dari suatu sampel statistik dengan data mean atau rata-rata data tersebut. Pada kode di atas kita membuat standar deviasi terlebih dahulu, kemudian kita aplikasikan rumus tersebut pada variabel age yang menghasilkan 10,615876301997806 dan pada variabel duration yang menghasilkan angka sejumlah 266.69780674810306.

Max dan min



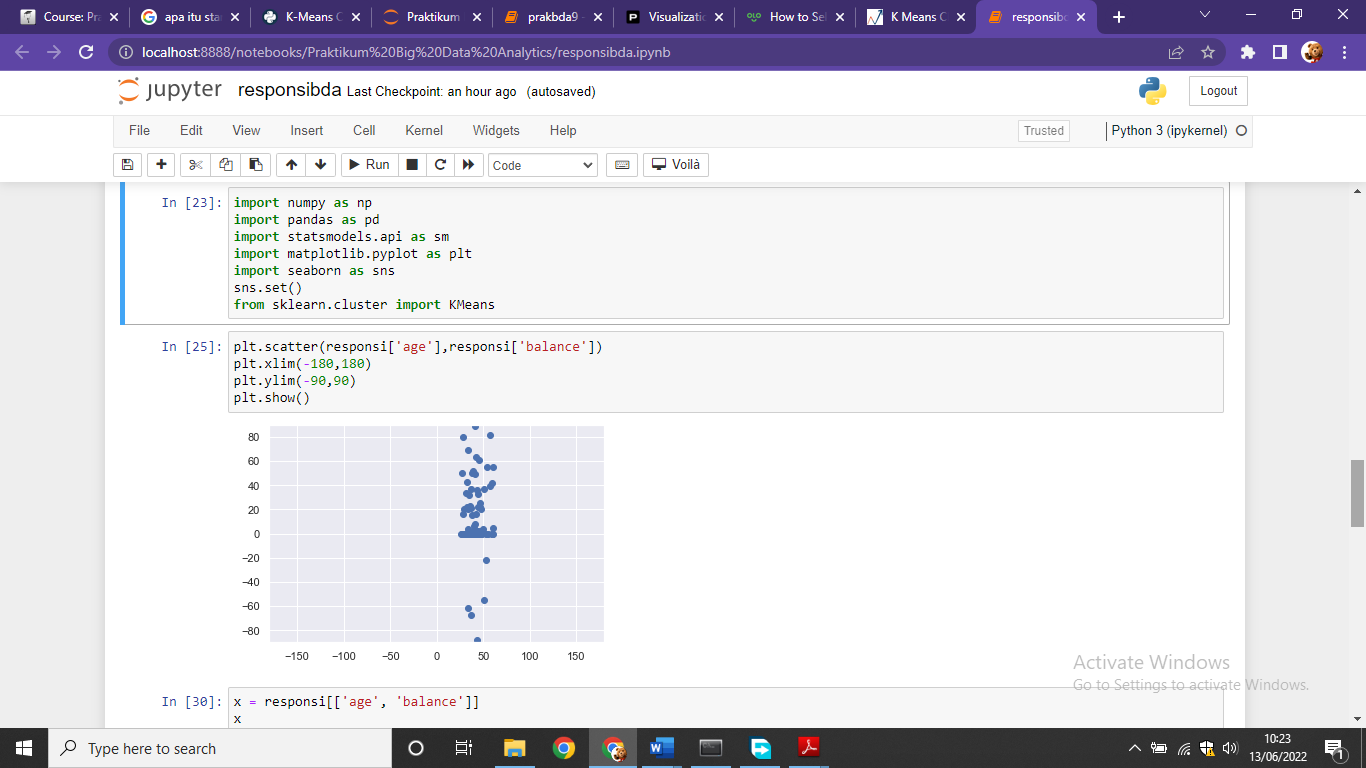
Penjelasan:

Pada kode di atas kita mencari nilai paling tinggi dan paling rendah pada variabel age dan duration. Untuk mencari nilai tertinggi, gunakan max(). Untuk mencari nilai terendah, gunakan min().

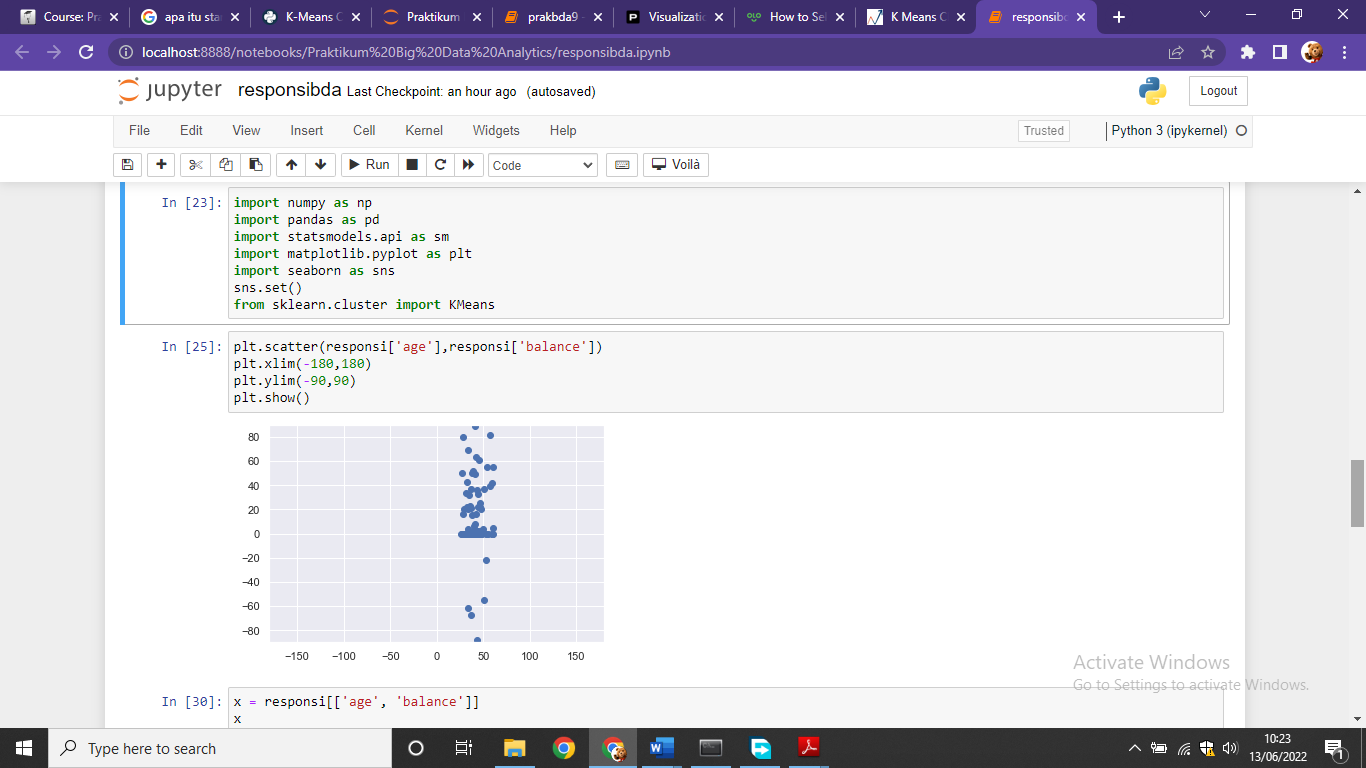
1. **Lakukan clustering menggunakan K-Mean untuk variabel USIA dan balance menjadi 3 kelompok**

Jawab:

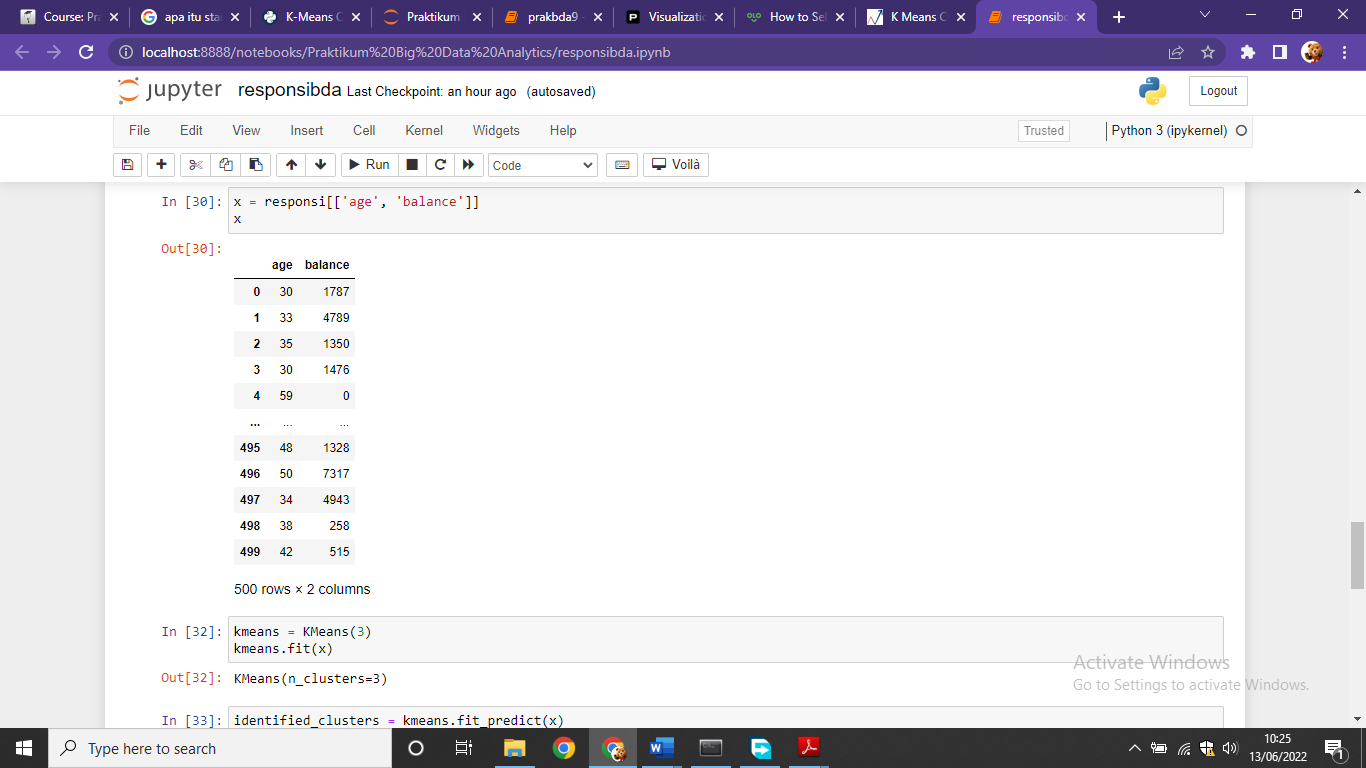
Import seluruh libraries yang diperlukan, terutama library sklearn.cluster untuk mengambil metode KMeans



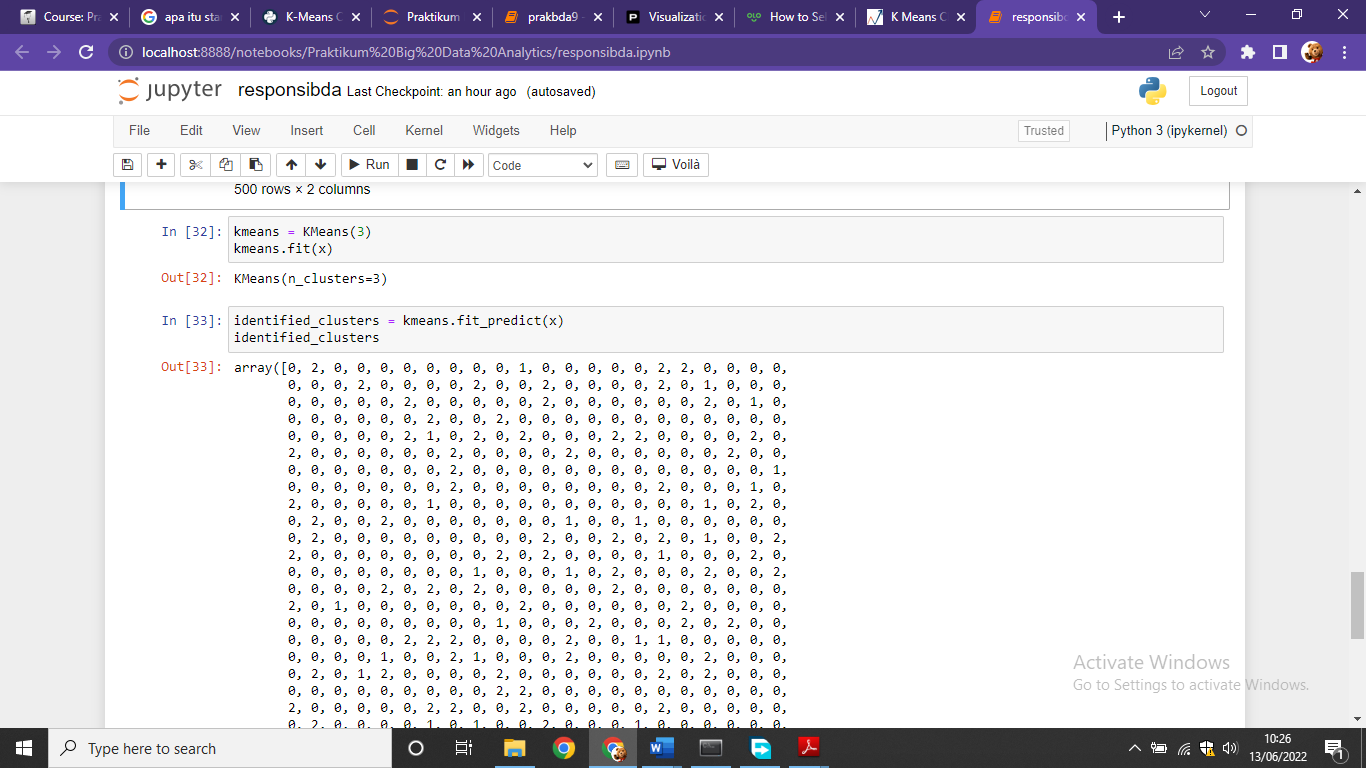
Load data dan plot data untuk melihat sebaran



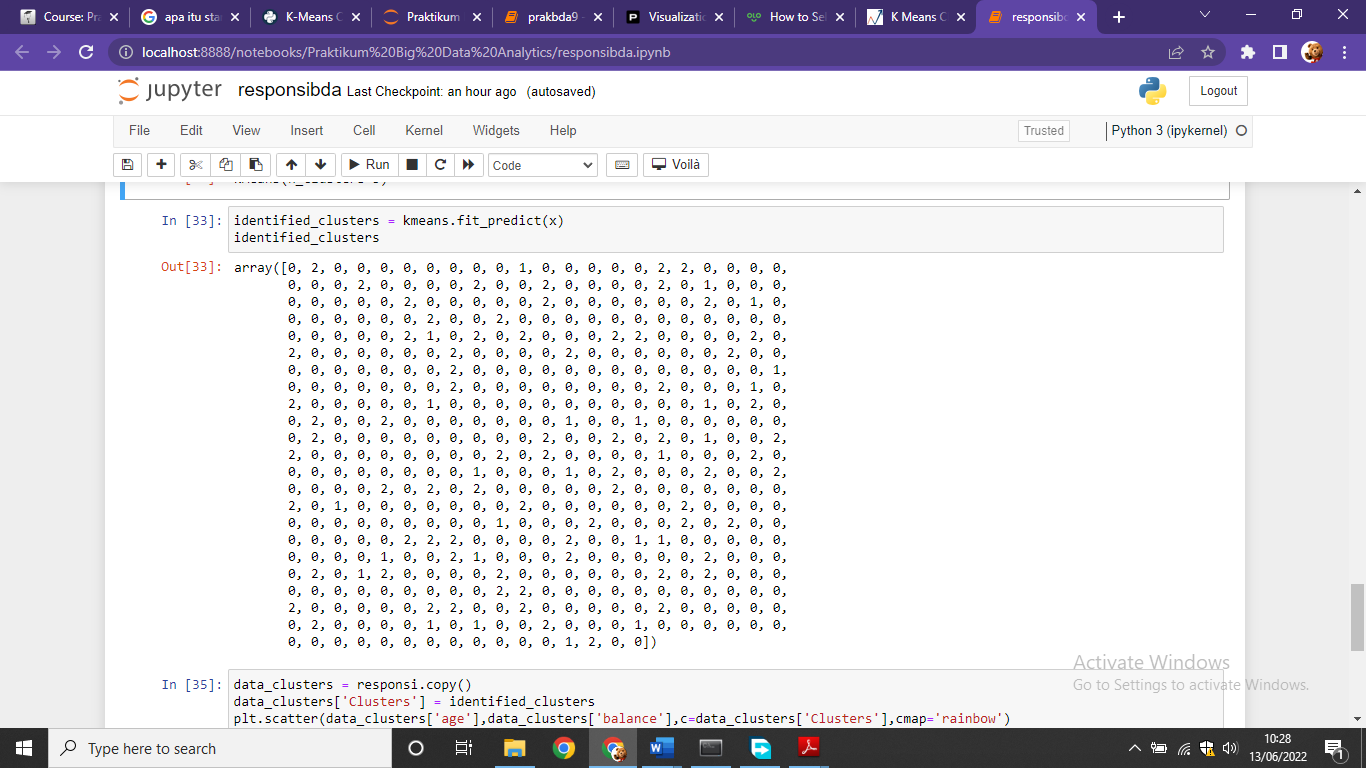
Pilih dua variabel yang akan di-cluster



Buat cluster sebanyak 3 kelompok



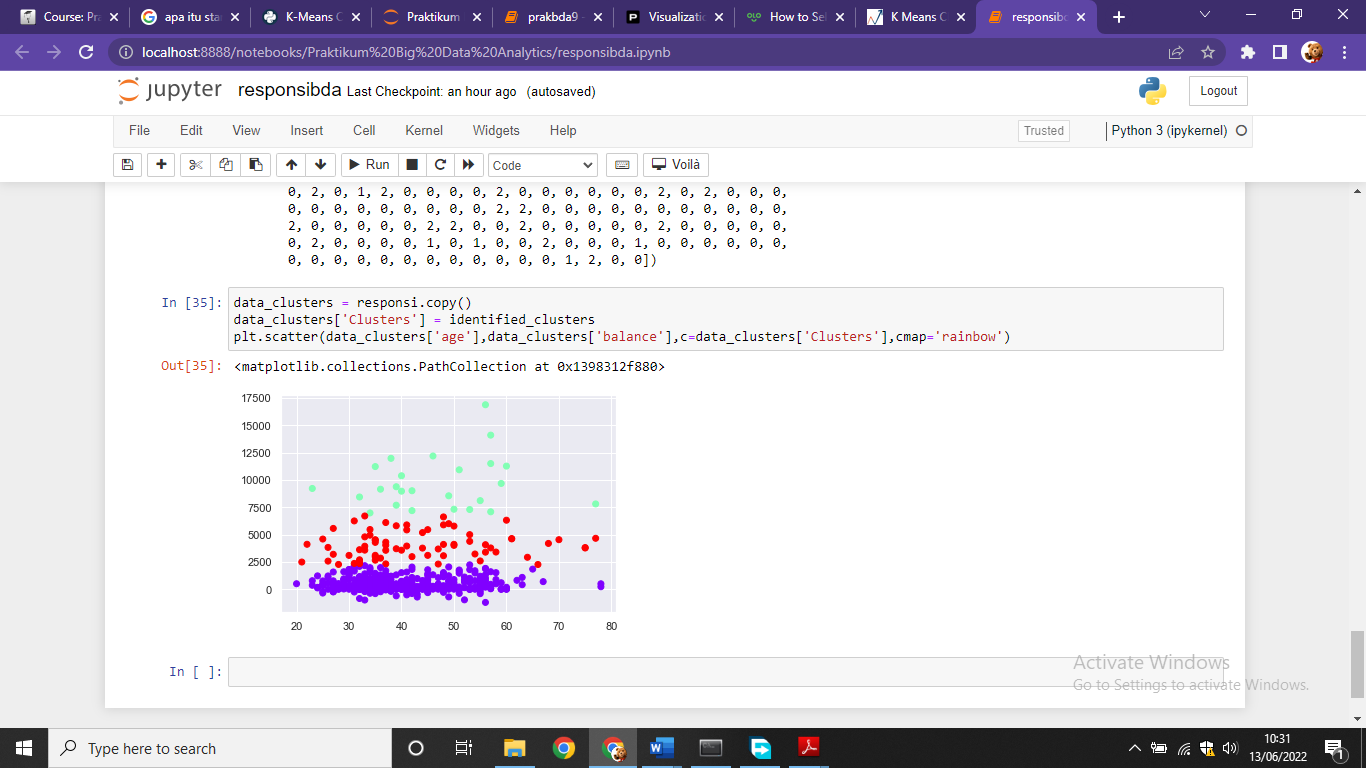
Prediksi pengelompokkan cluster dengan kmeans.fit\_predict(x)



Penjelasan:

Bisa kita lihat pada array tersebut, kmeans.fit\_predict(x) mengelompokkan data menjadi 3 cluster yaitu 0,1,2.

Menampilkan visualisasi sebaran cluster 0,1, dan 2.



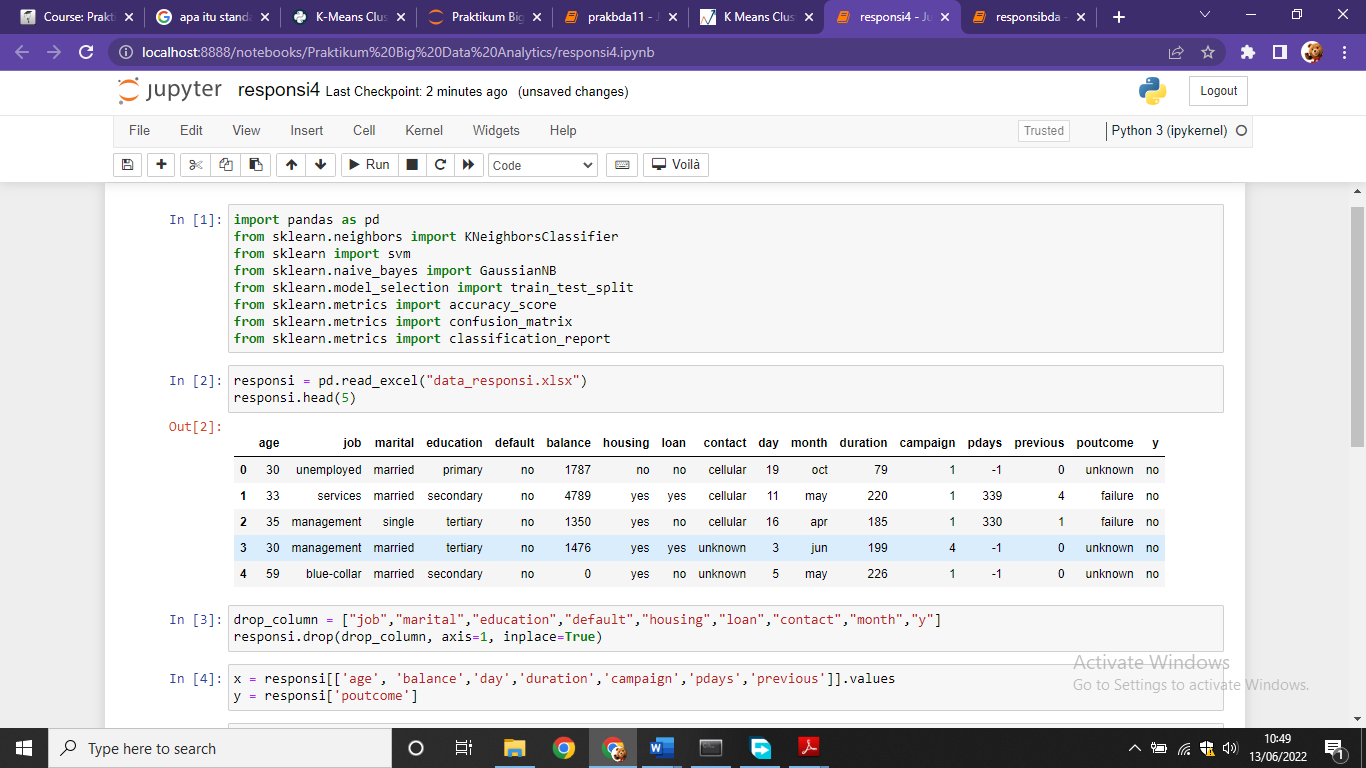
Penjelasan:

Pada sebaran cluster tersebut, cluster dibeda-bedakan dengan 3 warna. Visualisas cluster tersebut kita hasilkan dari array cluster yang telah kita bentuk tadi yang berisi cluster 0,1,2 sesuai dengan hasil cluster yang telah diprediksi oleh kmeans.fit\_predict(x).

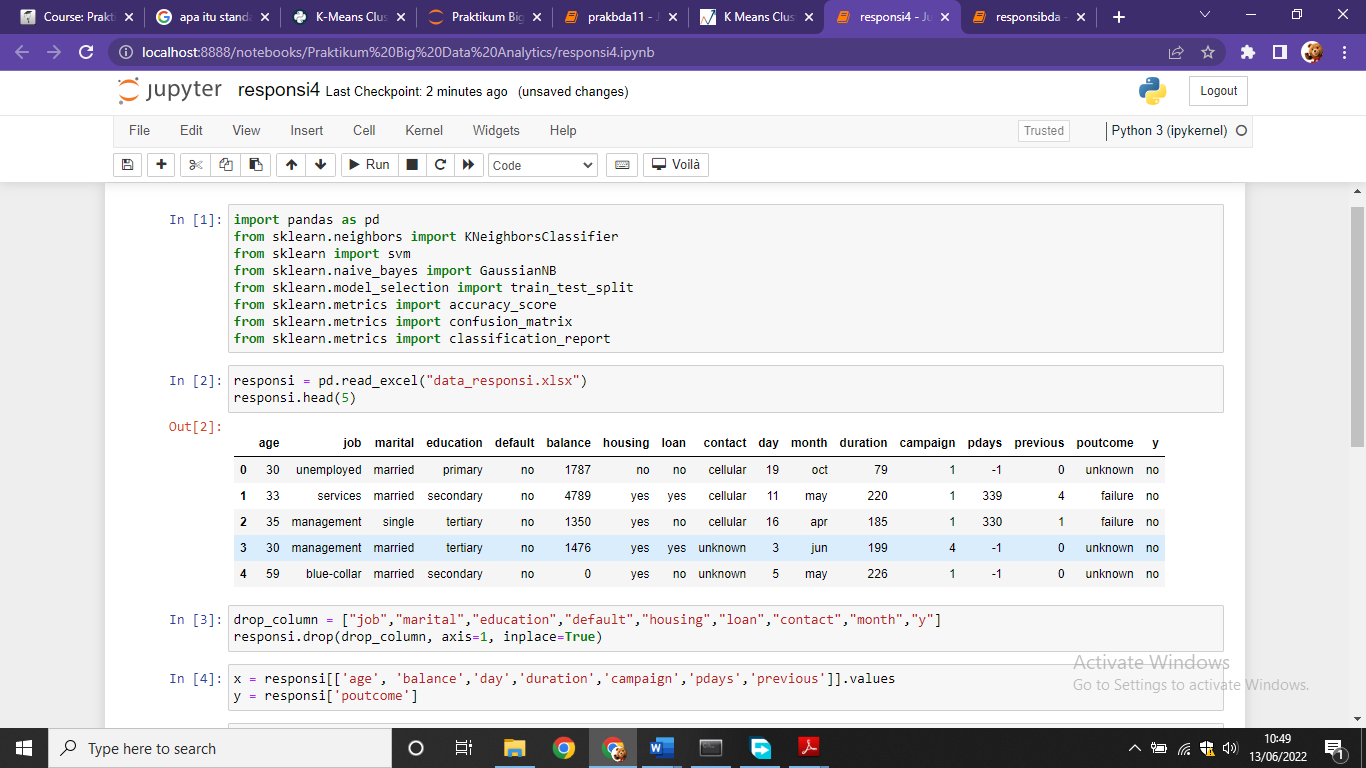
1. **Lakukan klasifikasi untuk variabel X (age, job, marital, balance dan education) sedangkan untuk target Y adalah loan**

Jawab:

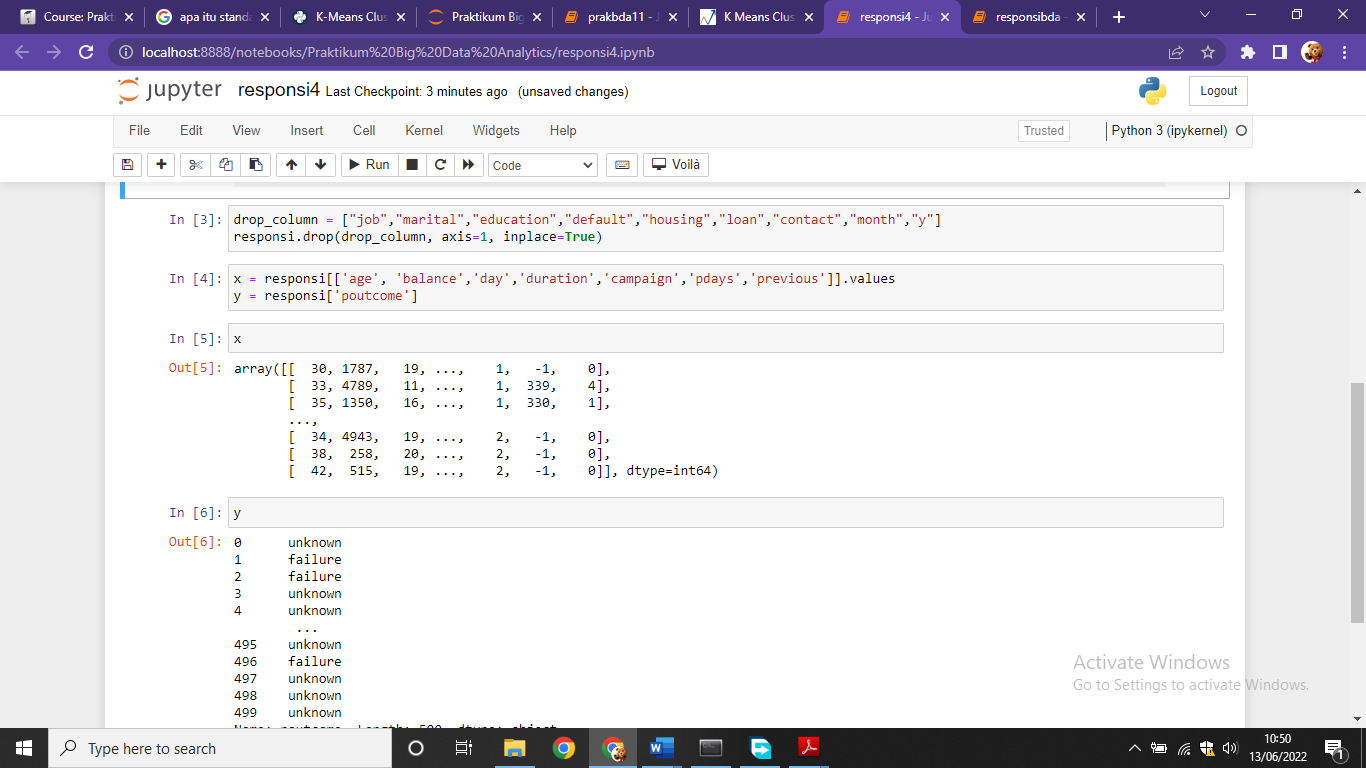
Import libraries yang diperlukan



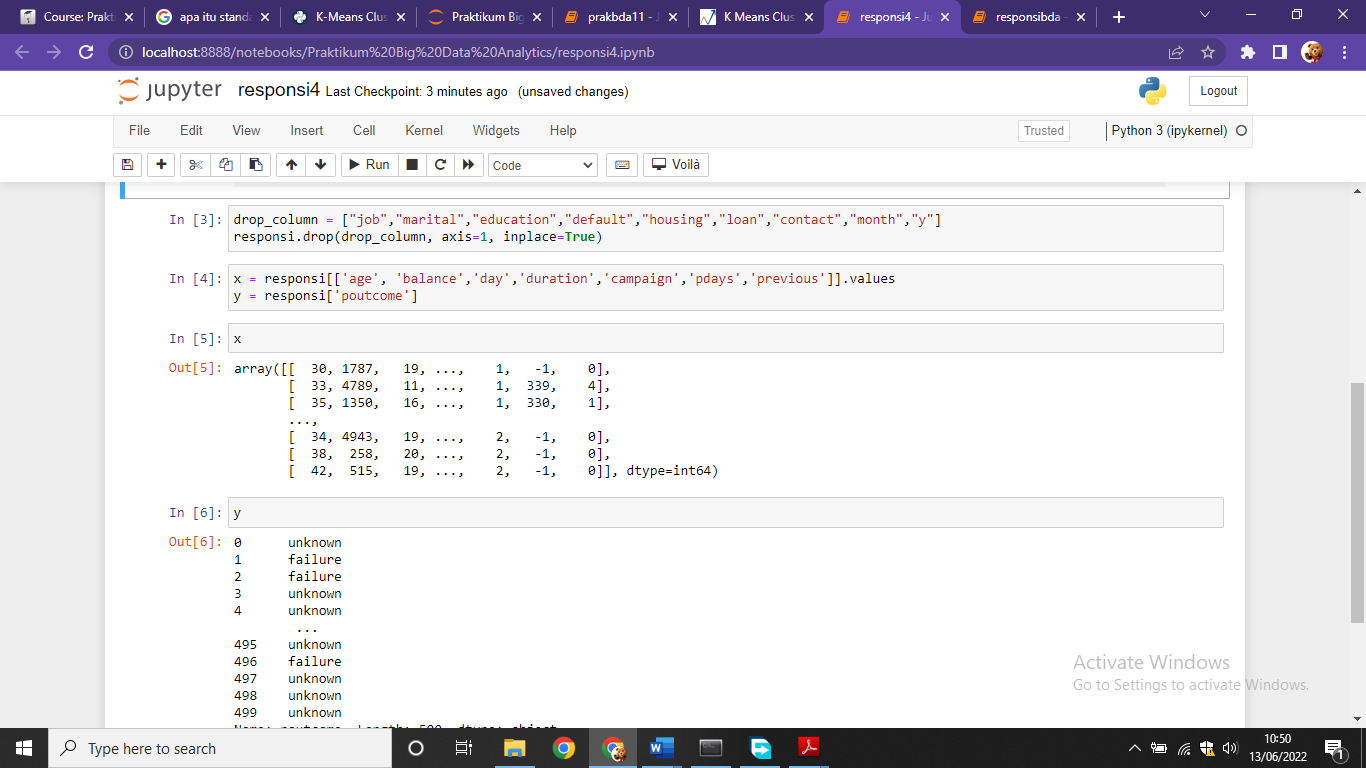
Tampilkan data



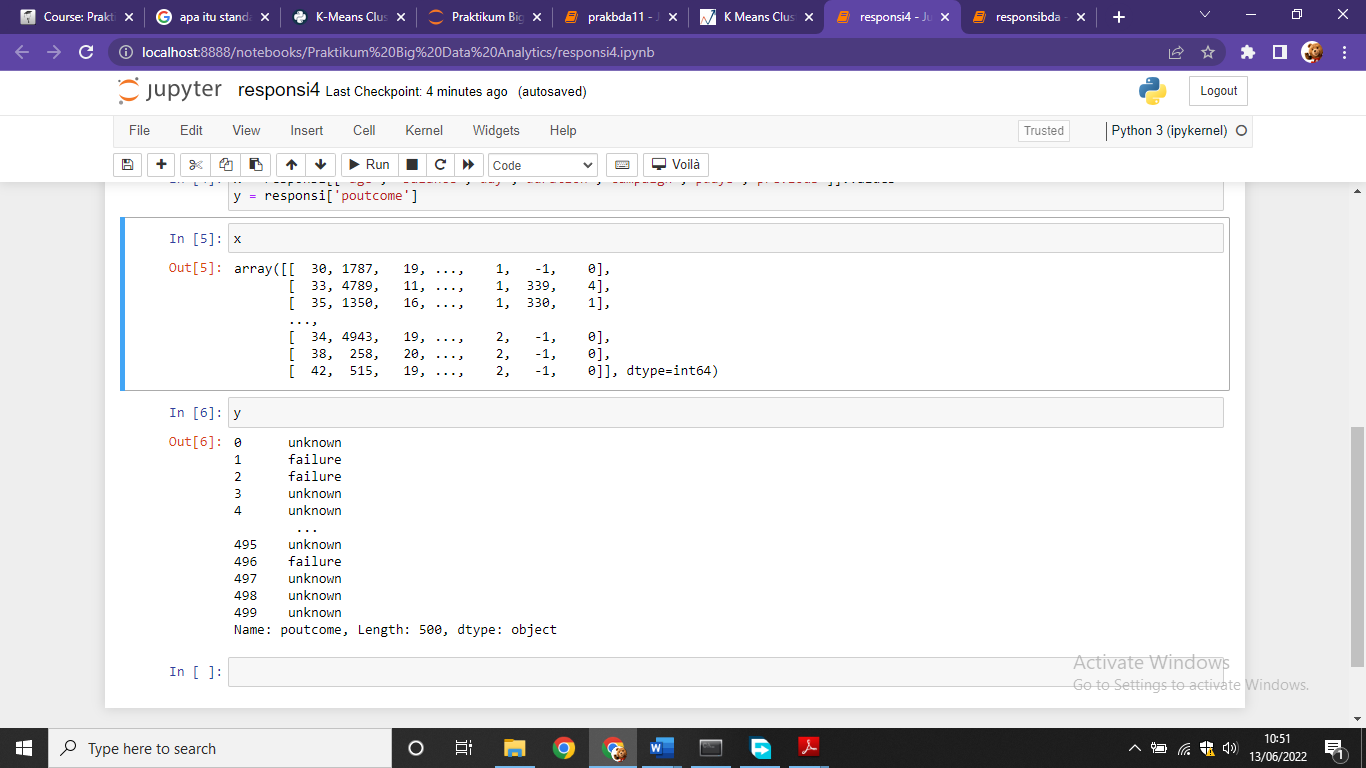
Drop kolom yang tidak digunakan



Masukkan data yang akan diproses ke x, dan data target ke y



Tampilkan x dan y

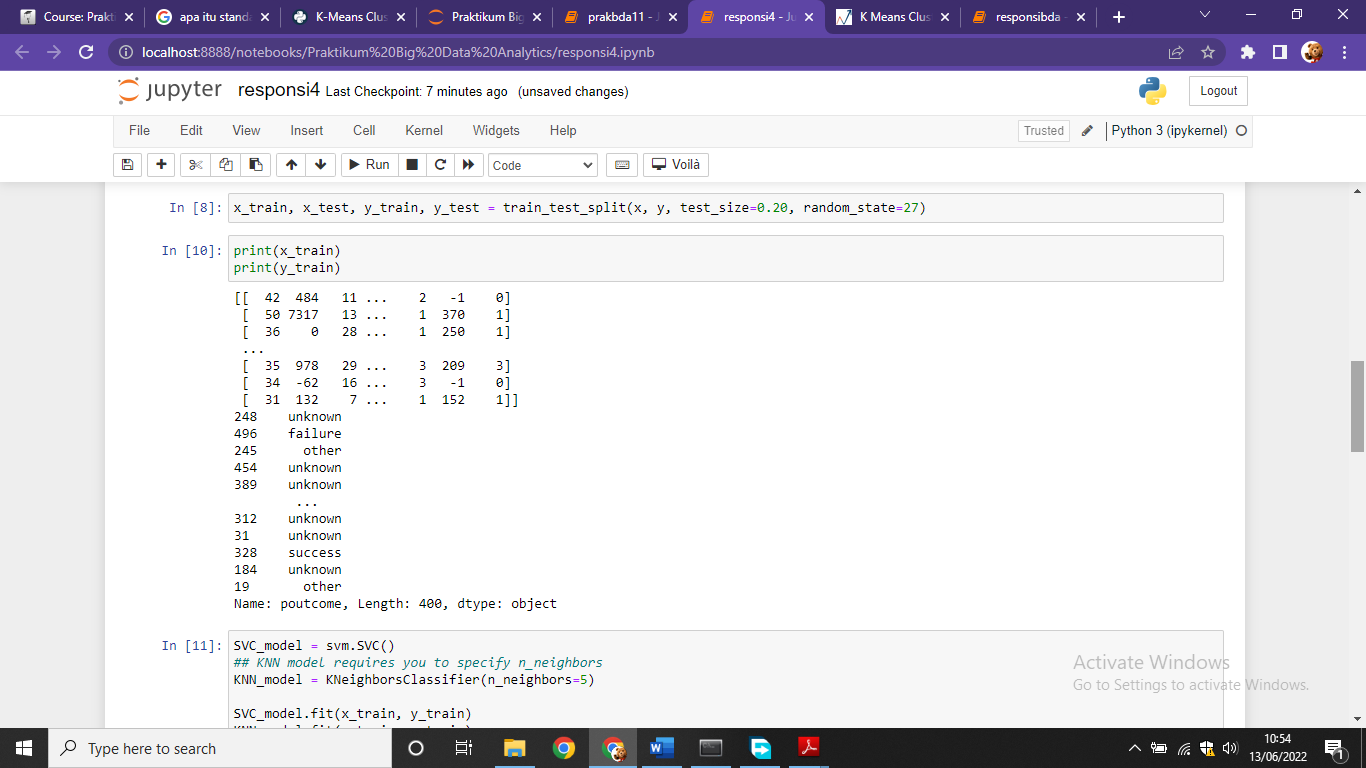


Split data

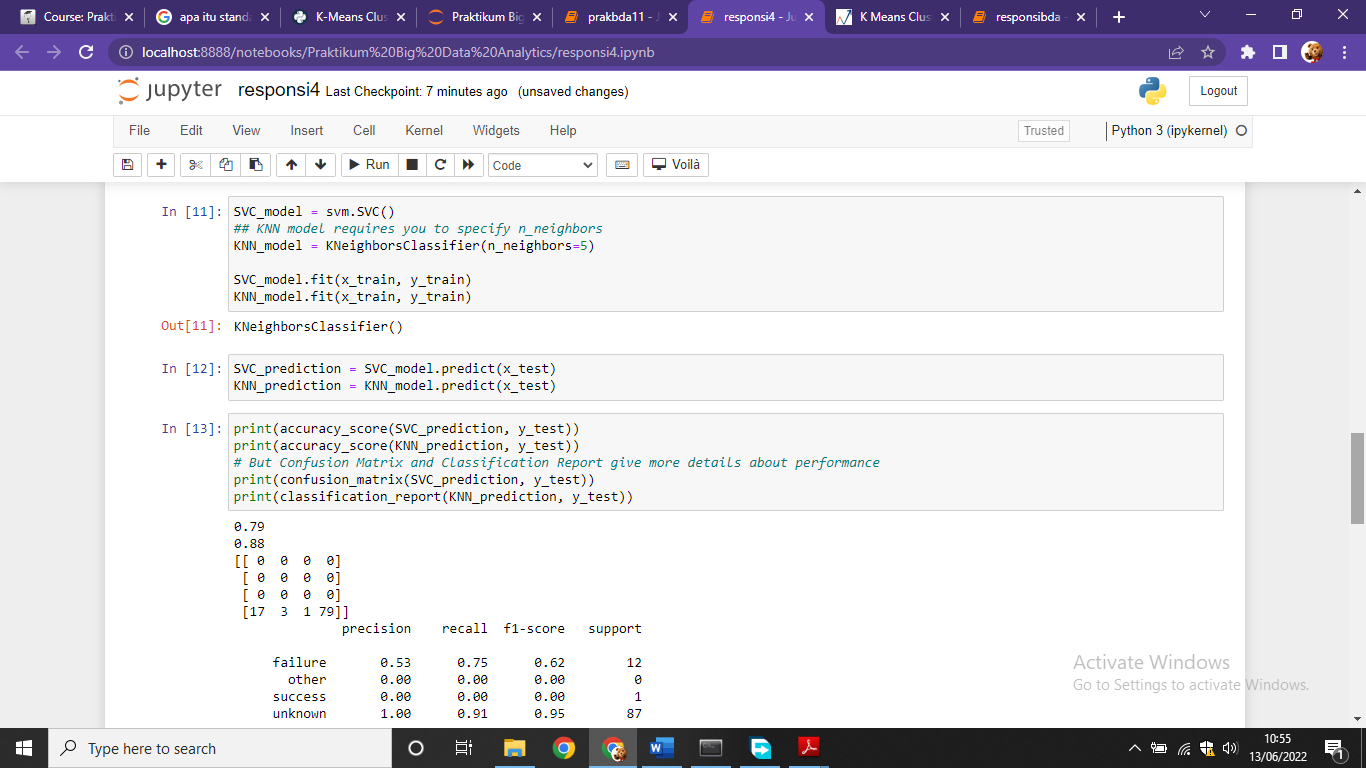
(di halaman selanjutnya)

(di halaman selanjutnya)

(di halaman selanjutnya)



Proses klasifikasi dengan 2 metode

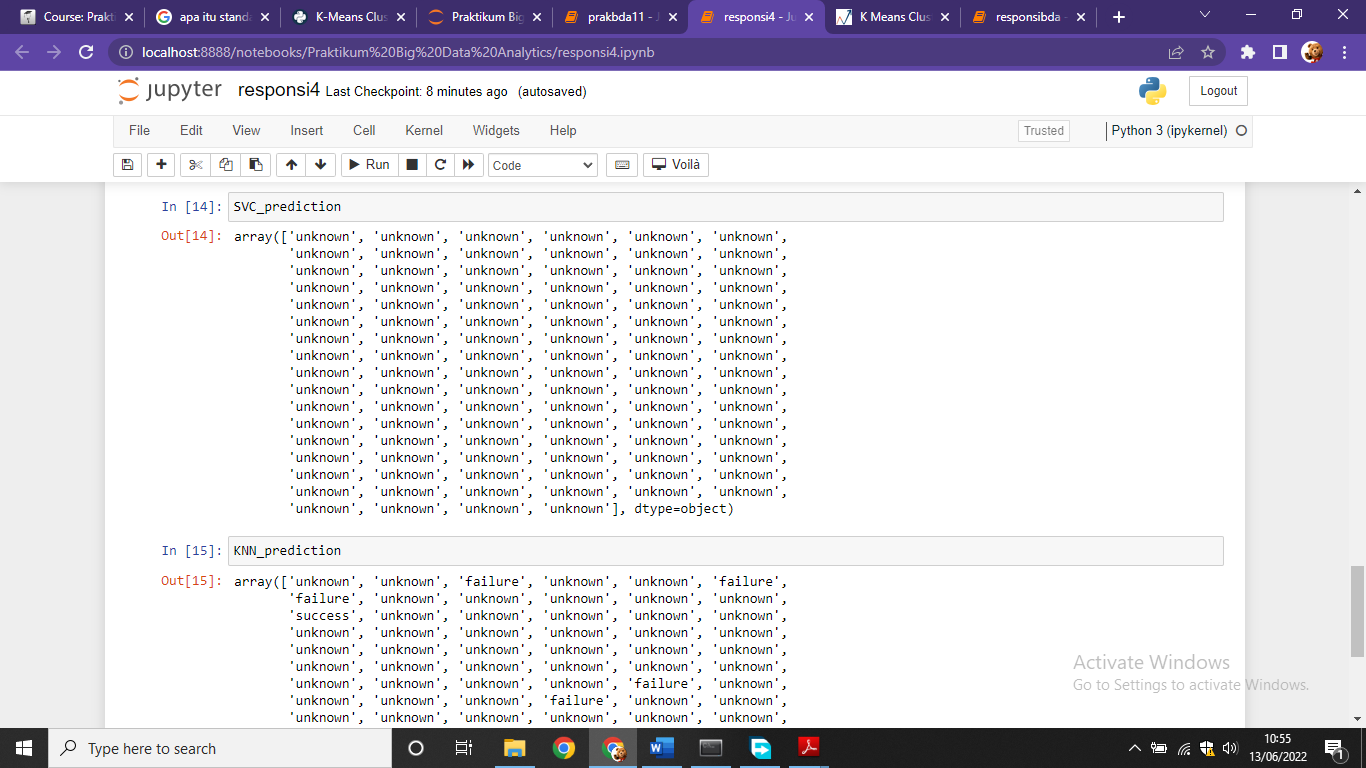


Tampilkan prediksi akhir dari svc dan knn

(di halaman selanjutnya)

(di halaman selanjutnya)

(di halaman selanjutnya)



1. **Dari hasil jawaban diatas lakukan penjelasan dari hasil yang anda dapatkan**

(sudah saya beri penjelasan di sebelah bawah setiap screenshot codingan dan grafik)

**Terima Kasih**