**Analisis Big Data**

**Nama : Alfianri Manihuruk**

**NIM : 120450088**

**Prodi : Sains Data**

Dataset:

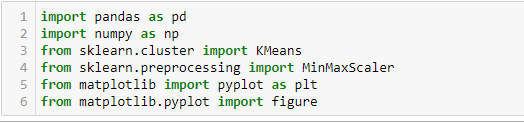
https://drive.google.com/file/d/1sJ\_dVlPCGY1jX4dcMZoXEYZKQBJXe\_3v/view?usp=sharing

Lakukan clustering K-means dengan data diatas. Berikan pendapatan anda mengenai hasil yang Anda dapatkan.

**Clustering K-means**

1. **Mengimpor Library**

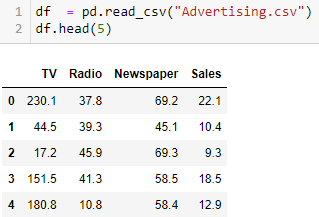
untuk metode clustering kali ini, saya menggunakan library seperti panda, numpy, matplotlib dan lain-lain, namun untuk Library utama adalah Sklearn dengan alat bantu Pandas untuk pengelolaan ekspor dan impor file serta matplotlib untuk pembuatan grafik, untuk cara pemanggilannya dapat di tampilkan sebagai berikut.



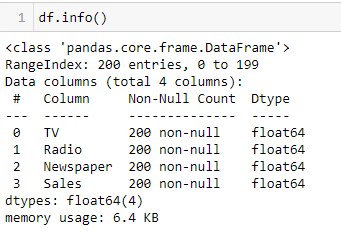
1. **Memanggil dan Memahami data**

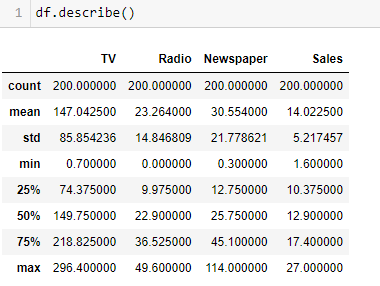
Langkah selanjutnya, yaitu meng-input dataset. Dataset yang akan digunakan pada tulisan ini adalah dataset Advertising. data ini merupakan sekumpulan data Advertising Berikut adalah langkah untuk meng-input dataset dan melihat detail informasi tentang dataset.

Untuk kali ini kita menggunkan data dengan ekstension csv, untuk pemanggilannya dapat di lakukan sebagai berikut.



Selanjut kita melihat memahami data dengan melihat ringkasan datanya, dengan menggunakan perintah sebagai berikut.

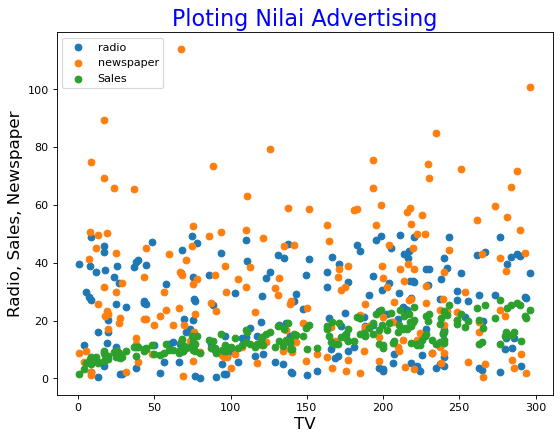




Selanjutnya kita dapat menvisualisakan data yang kita miliki sebelum dilakukan clustering, untuk perintahnya dapat dilakukan sebagai berikut.

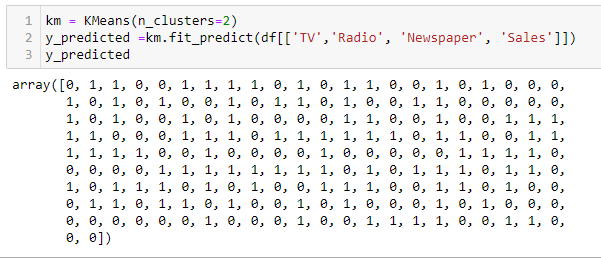


Dimana pada visualisasi untuk sumbu x yang digunakan adalah atribut (‘TV’) dan untuk sumbu y-nya adalah atribut (‘Radio’, ‘Newspaper’, ‘Radio’), untuk hasilnya dapat di lihat sebagai berikut.



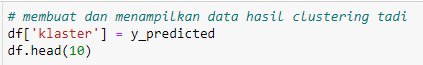
1. **Prediksi**

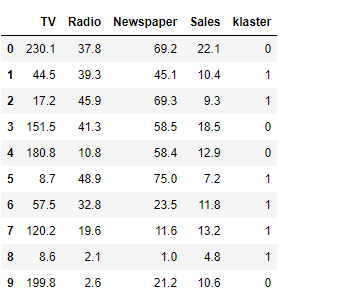
pada tahap ini kita akan melakukan prediksi untuk untuk melakukan clustering dari data yang kita miliki. Nah, hal terpenting adalah tidak hanya menghitung y\_predicted saja melainkan memberi label kembali datanya. Percuma saja jika kita tidak mampu memetakan kembali siapa saja yang masuk kategori klaster ‘0’ dan ‘1’.



Untuk penentukan klasternya, disini saya menggunakan 2 label yaitu ‘0’ dan satu dengan sentroid yang dibuat secara random oleh library yang telah panggil di atas tadi.

Untuk tahap selanjutnya kita menambahkan kolom baru untuk memberikan label hasil klaster tadi kedalam tabel data sebelumnya, untuk perintah dan hasilnya sebagai berikut



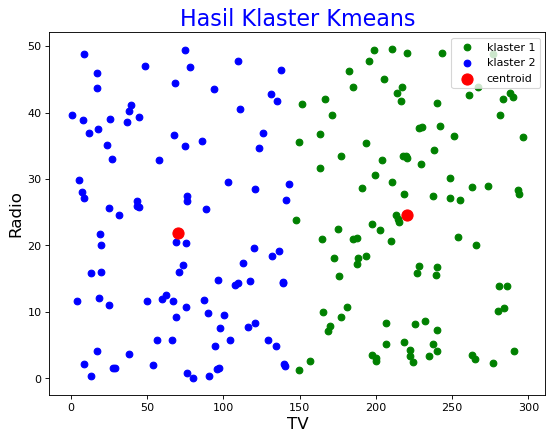


1. **Finishing**

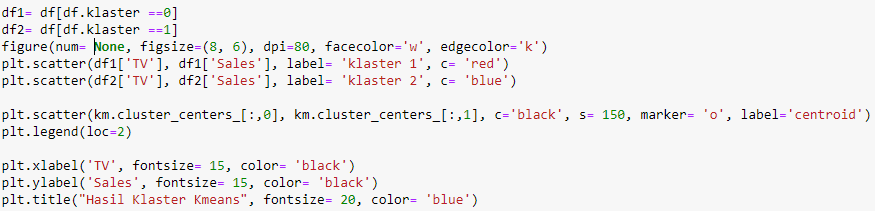
pada tahap ini kita kembali memvisualisasikan data hasil clustering tadi dalam bentuk grafik

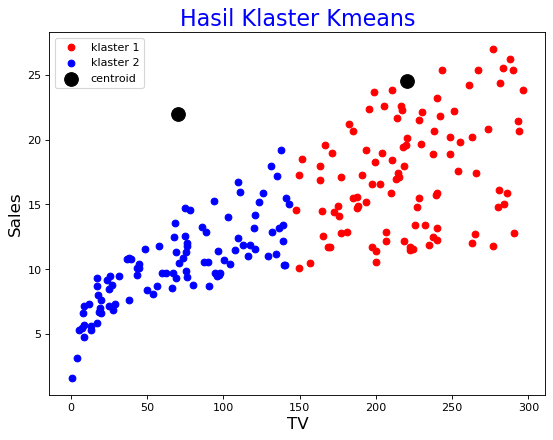
1. **Visualisasi atribut TV dan Radio**





1. **Visualisasi atribut TV dan Sales**

****



1. **Visualisasi atribut TV dan Newspaper**

****

