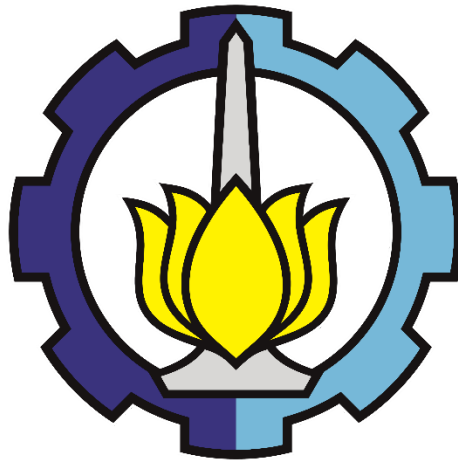


# CLUSTERING

## KECERDASAN KOMPUTASIONAL - B



Dosen : Ibu Dr. Diana Purwitasari, S.Kom., M.Sc.

Kelompok 3

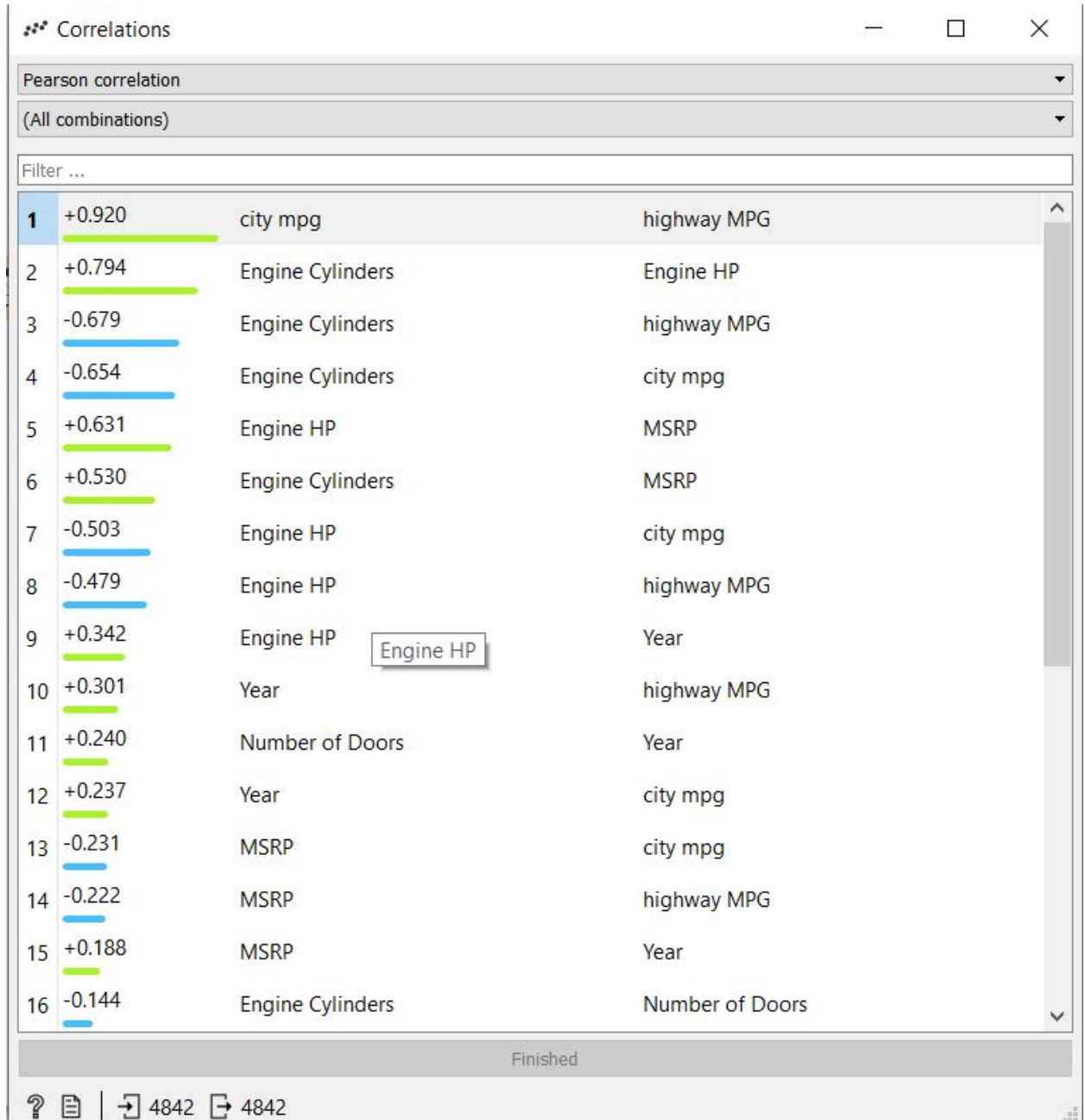
Kevin Angga Wijaya	05111840000024
Angelita Titiandes Br. Silalahi	05111840000088
Aflakah Nur Farhana	05111840000120

DEPARTEMEN TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA

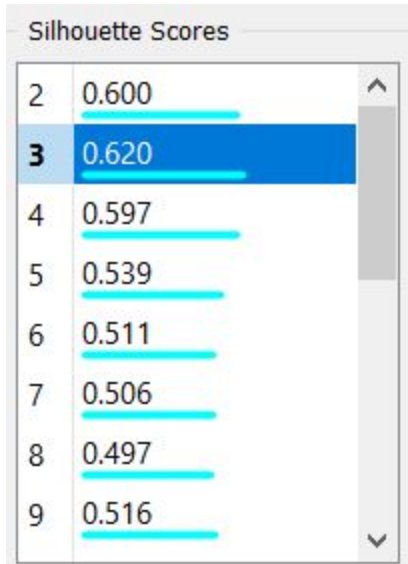
2020

**1. Apa kombinasi fitur yang memberikan cluster terbaik menurut indikator Silhouette score?**

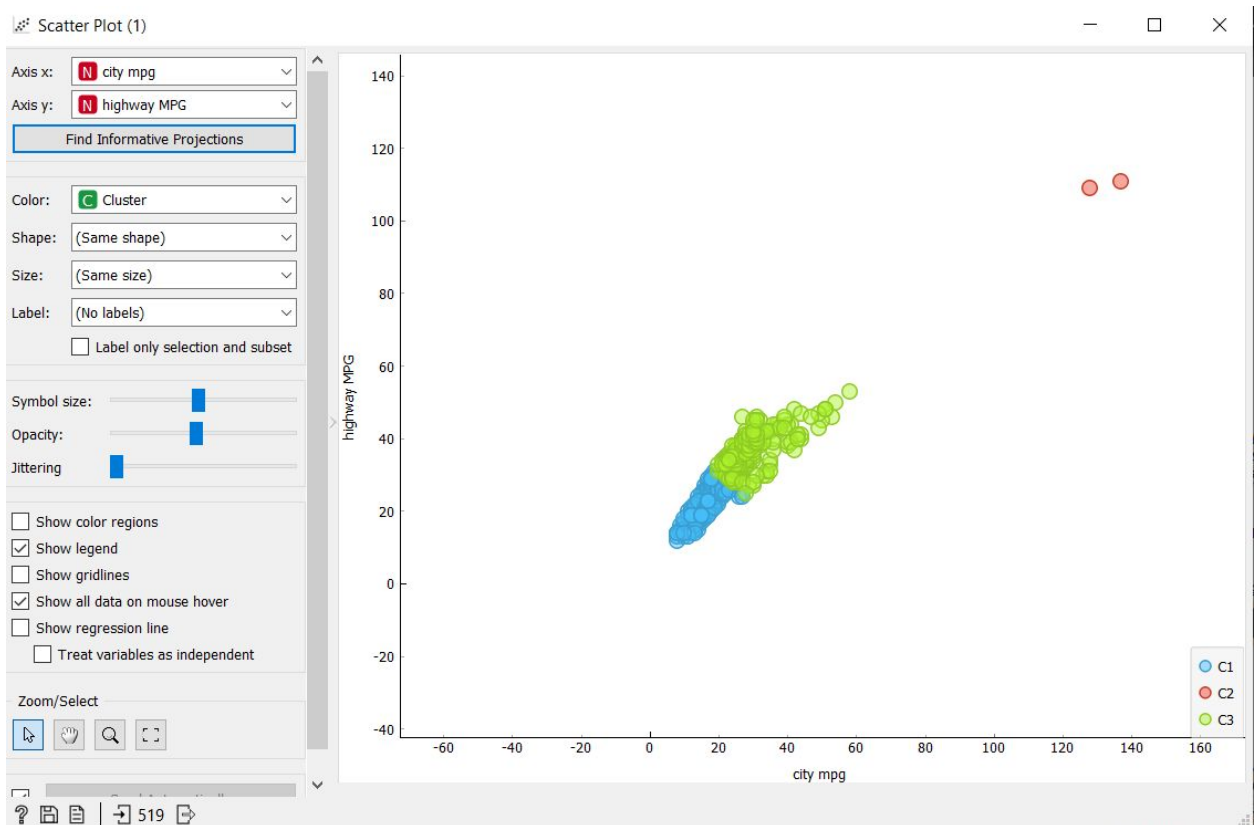
Berdasarkan widget correlation di orange, dapat kita lihat kolom kolom yang memiliki korelasi tertinggi sebagai berikut



Dapat kita lihat bahwa city MPG dan highway MPG memiliki korelasi yang paling dekat. Setelah dilakukan pembersihan data duplikat, silhouette scorenya dilihat dan hasilnya adalah sebagai berikut



Dapat dilihat dengan pembagian menjadi 3 cluster silhouette score nya cukup tinggi, yakni 0,620 dengan hasil pembagian sebagai berikut



Untuk pasangan Engine Cylinders dan Engine HP serta pasangan Engine Cylinders dan Highway MPG memang memiliki korelasi yang cukup tinggi. Hanya saja, ketika di cluster menggunakan k-means, pembagian clusternya menjadi cukup banyak sehingga dianggap bukan merupakan yang terbaik.

**2. Bagaimana hubungan korelasi antar fitur yang terpilih?**

Highway MPG dan City MPG merupakan pemakaian bahan bakar dari kendaraan itu sendiri di medan yang berbeda yakni 'highway' dan 'city'. Kedua hal ini berbanding lurus jika dilihat dari data yang ada dan hasil clustering yang ada di soal sebelumnya, sehingga korelasinya pun cukup tinggi hampir mencapai angka 1.

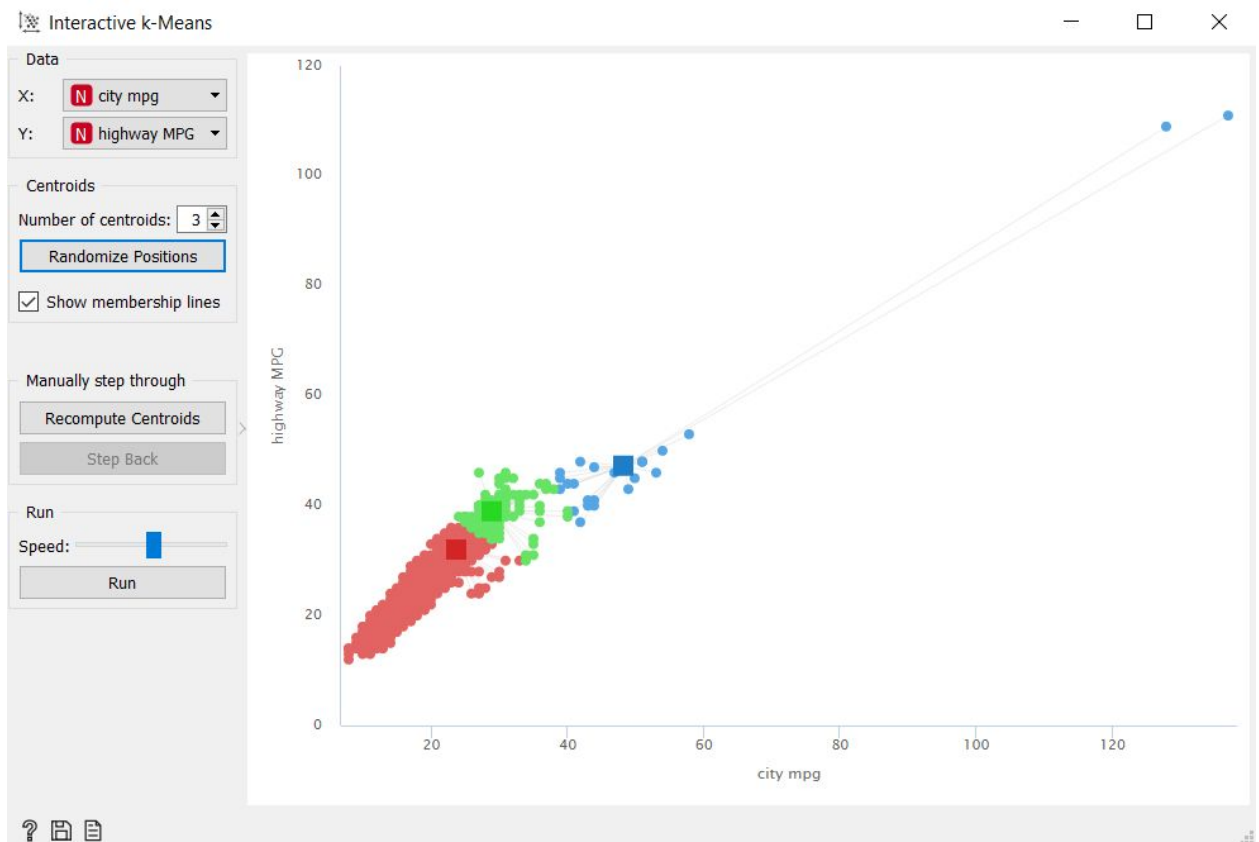
**3. Apakah ada proses cleaning data (impute, dll)**

Dalam percobaan code dengan python dilakukan pembersihan data yang duplikat dan hanya menggunakan kolom yang bersangkutan.

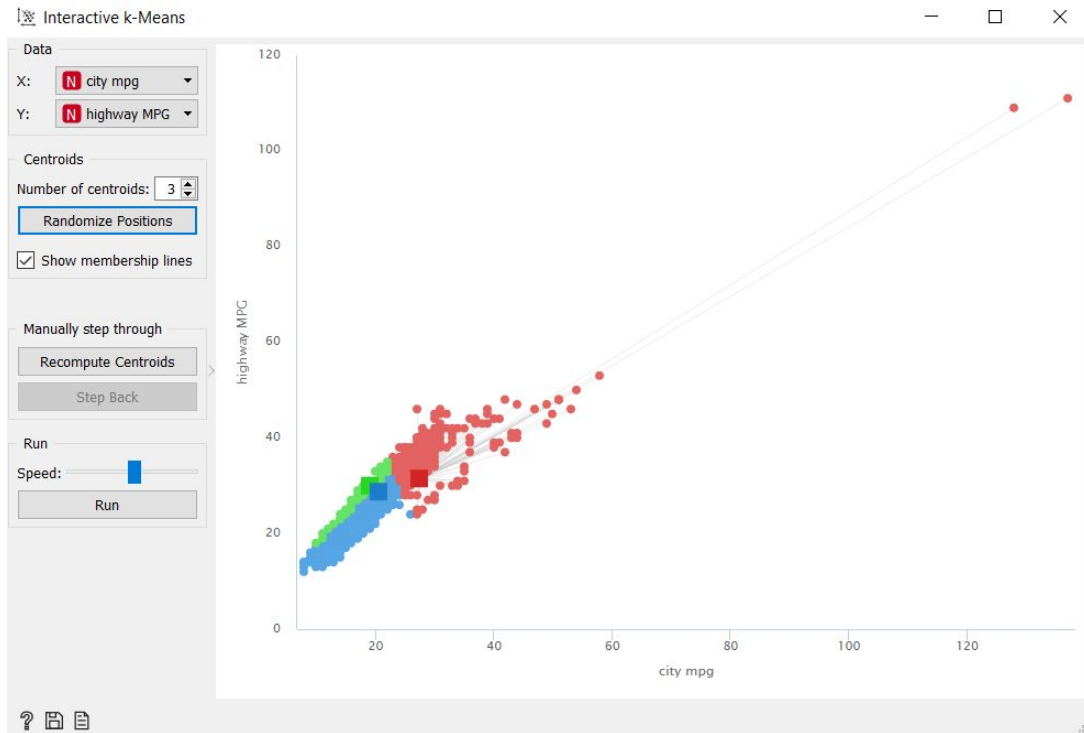
Untuk percobaan menggunakan orange, dilakukan impute untuk membersihkan data dengan null value,

**4. Bagaimana hasil cluster jika centroid terbentuk secara random atau dipilih yang terbaik K Means ++?**

Dipilihkan K-Means

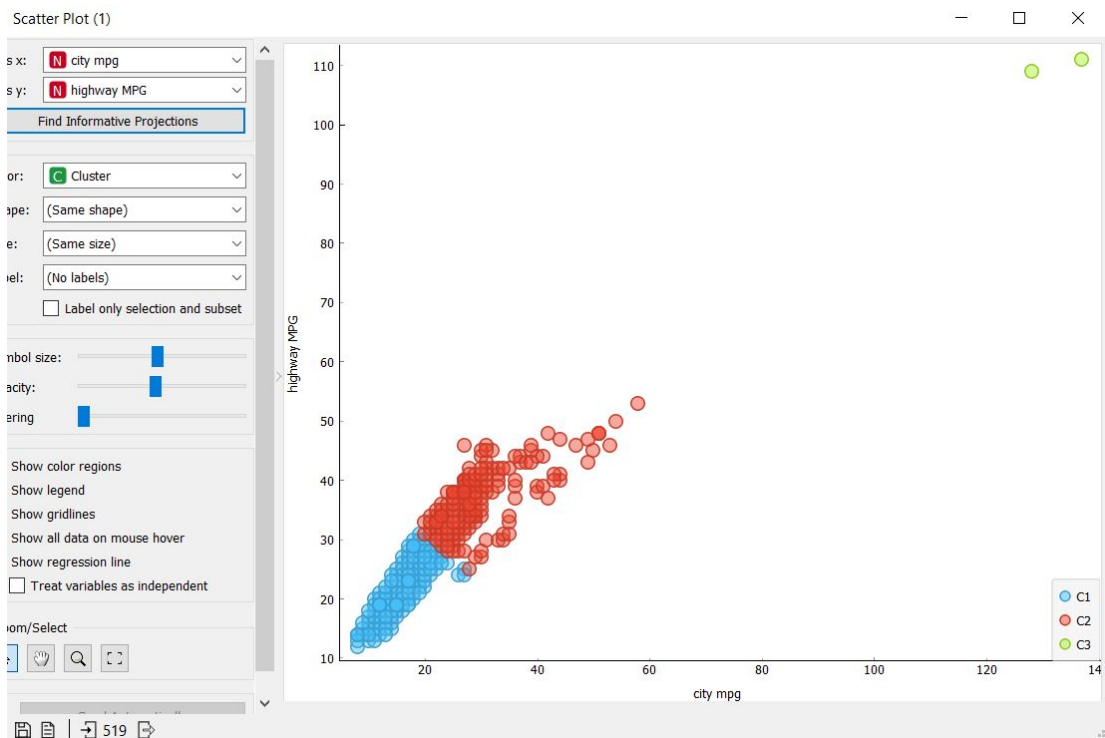


Salah satu hasil randomize

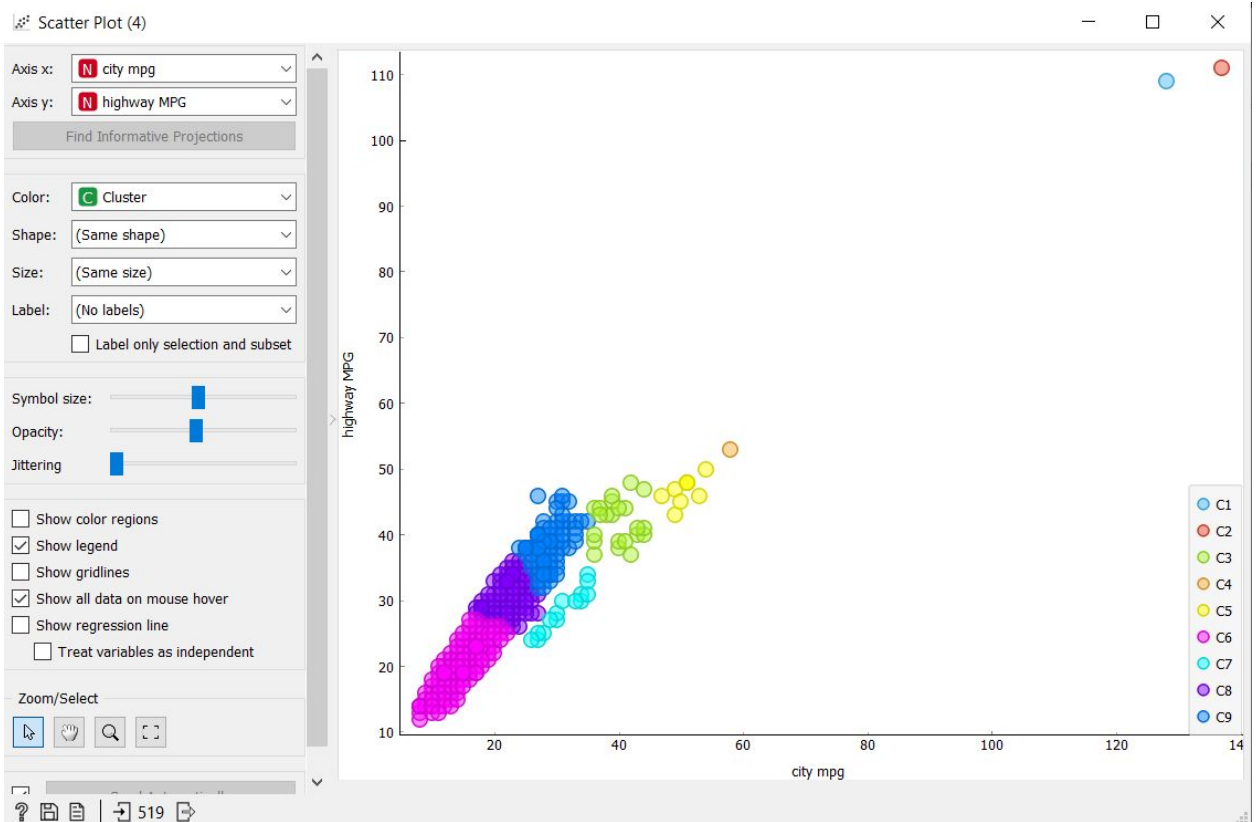


## 5. Bandingkan K Means dgn Hierarchical (divisive, agglomerative)?

Hasil clustering K-means untuk data city-MPG dan highway-MPG



## Hasil clustering Hierarchical untuk data city-M

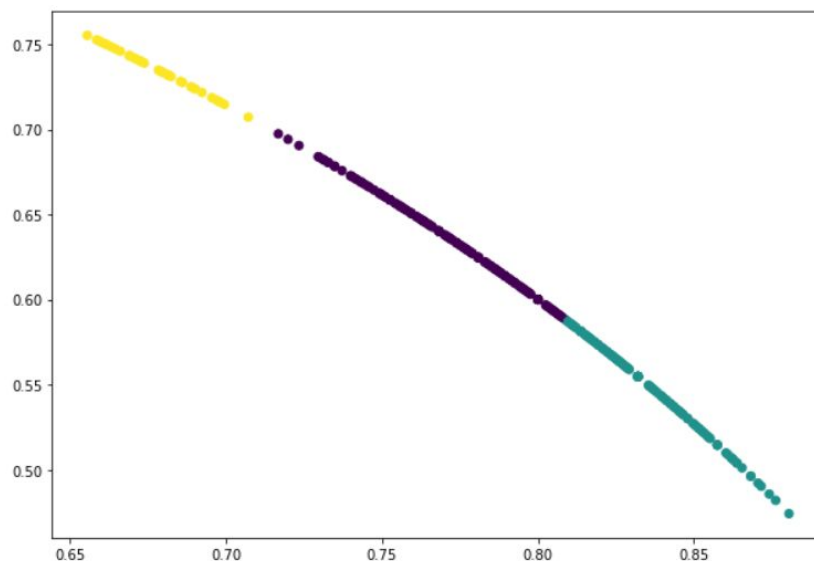


PG dan highway-MPG

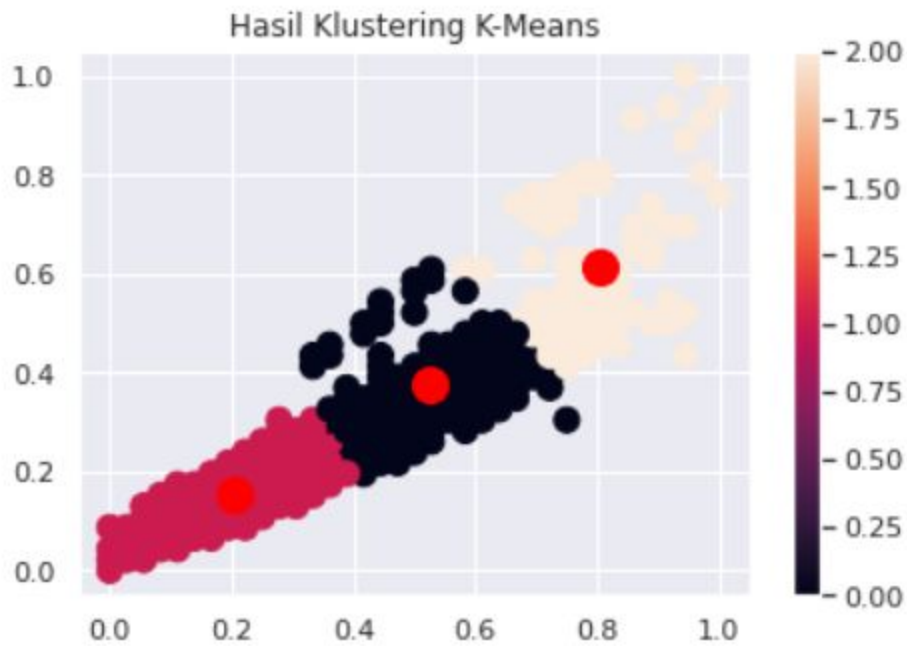
## 6. Bandingkan dengan algoritma jika menggunakan SkLearn

Tujuan: utk Data Cars mana yang paling sesuai dalam melakukan pengelompokan? partitional atau hierarchical

Hasil Hierarchical menggunakan SK-Learn



Hasil K-means menggunakan SKLearn



Untuk data cars, **metode yang paling sesuai adalah dengan menggunakan K-means.** Karena metode Kmean lebih efektif untuk mengolah data dalam jumlah banyak. Penyebaran scatter plot dengan metode k-means dalam kasus cars ini lebih tersebar. Apabila di modelkan dalam bentuk 3D scatter plot yang dihasil dengan dengan metode hierarchical menumpuk. Hal tersebut memperlihatkan bahwa metode K-mean lebih bembagi cluster sari data cars secara lebih efektif