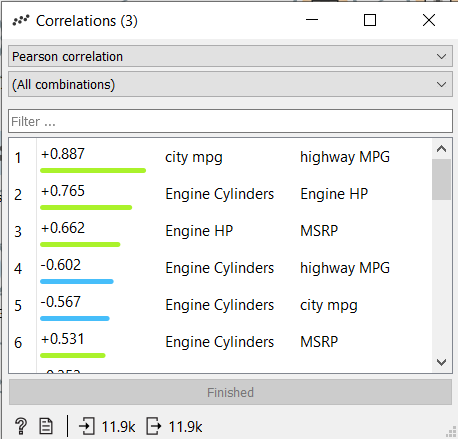
Muhamad Haris Wicaksono (5111840000029)

M Ridho Daffa Ardista (5111840000065)

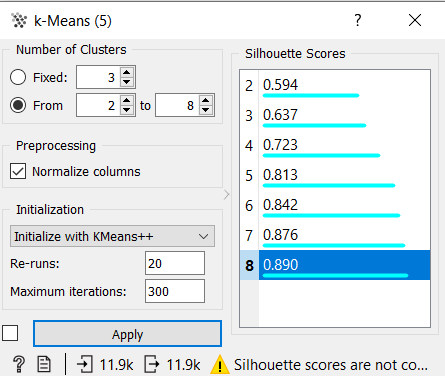
Kenji Hikmatullah (5111840000074)

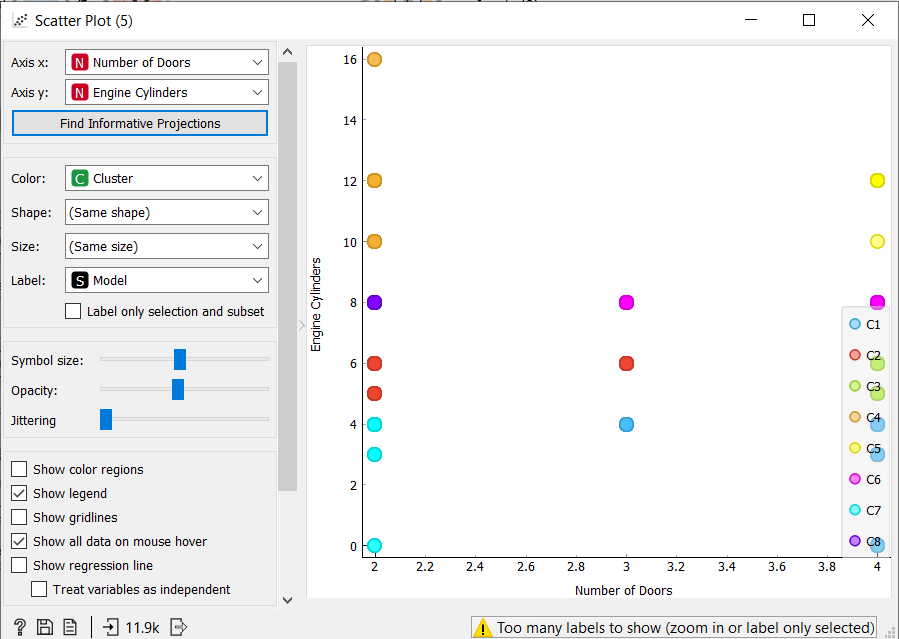
**1. Apa kombinasi fitur yang memberikan cluster terbaik menurut indikator *Sillhoutte score*?**

Pertama kami melihat dapat melihat korelasi terbesar

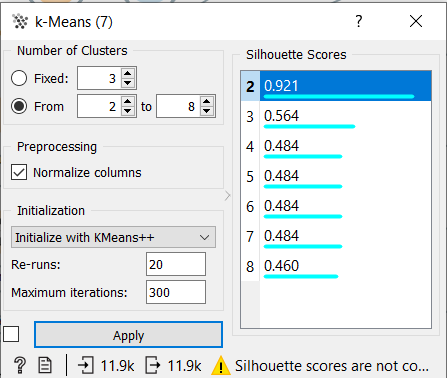
**

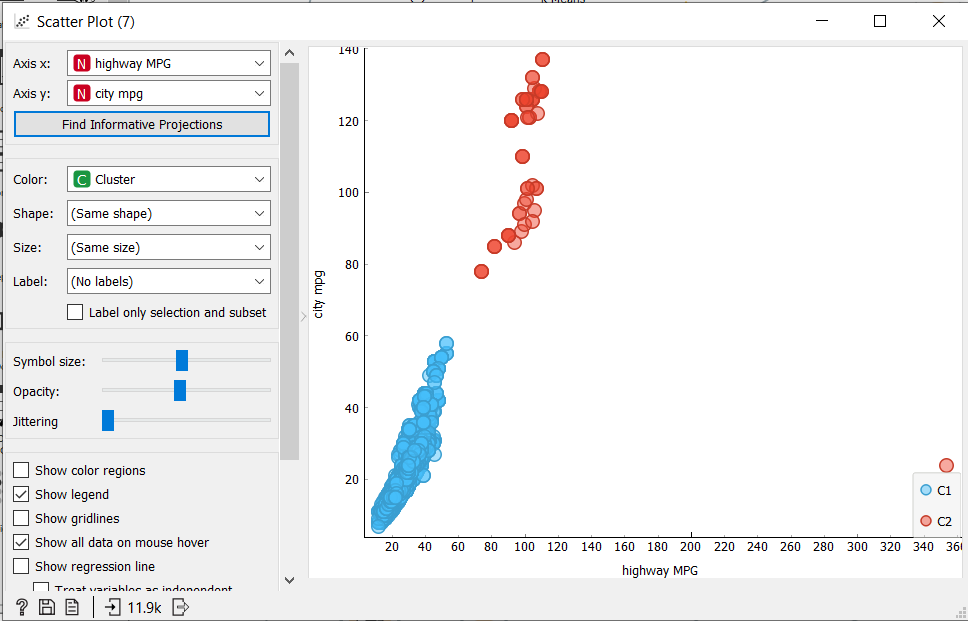
Pada fitur *Engine Cylinder* dan *Number of Doors*



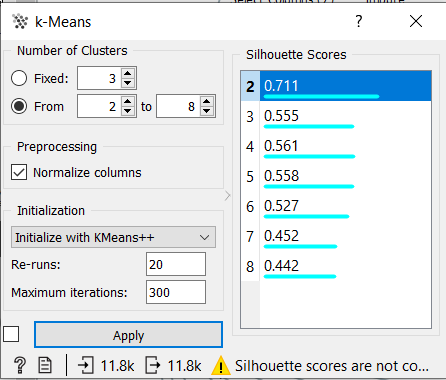


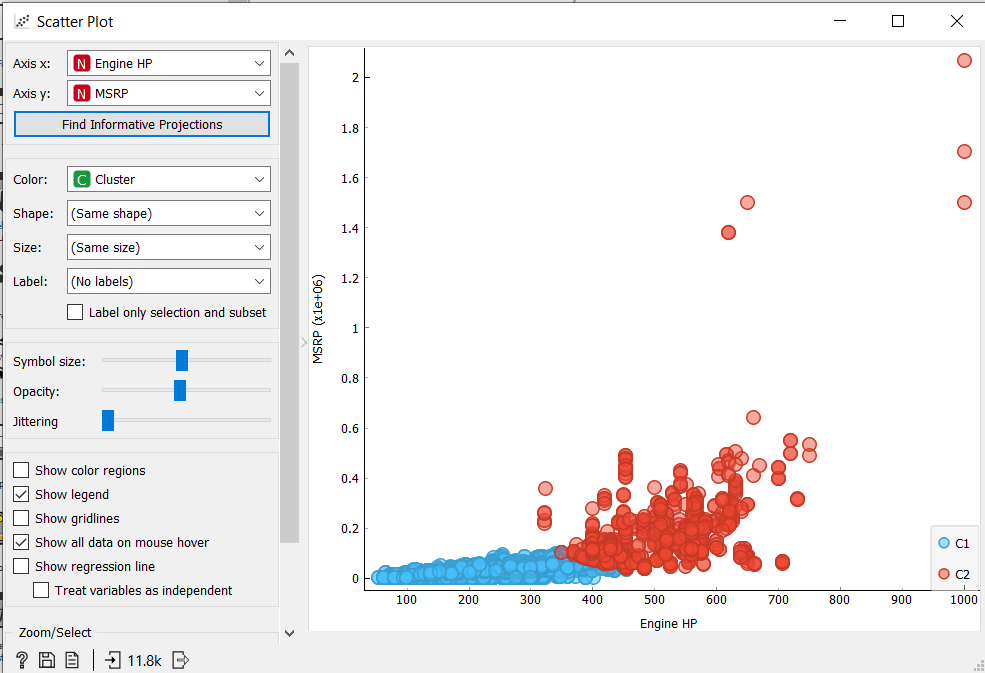
Pada fitur *city mpg* dan *highway MPG*





Pada fitur *Engine HP* dan *MSRP*

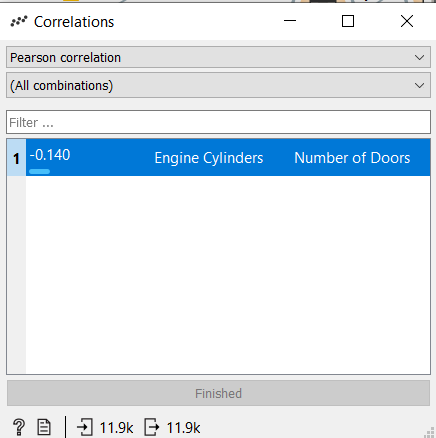


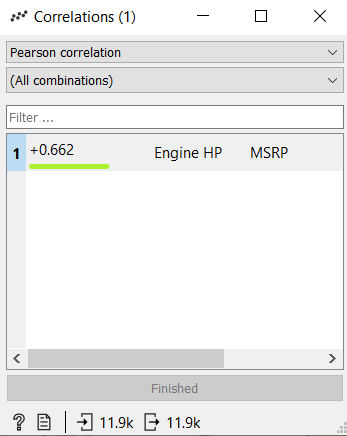
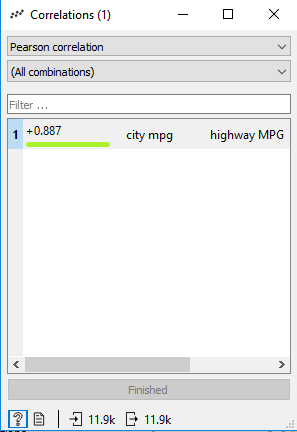


Setelah membandingkan dari berbagai fitur nilai dari *silhoute score* yang paling baik antara *city mpg* dan *highway MPG*

**2. Bagaimana korelasi antar fitur-fitur yang dipilih?**

Dibawah ini adalah nilai-nilai correlations antara fitur-fitur yang kami pilih pada no 1. Kalo dilihat dari nilai-nilai di bawah hubungan korelasi antara fitur-fitur



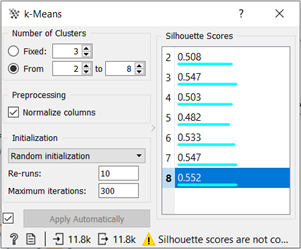
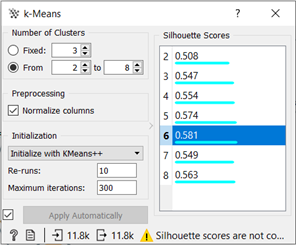


**3. Apakah ada proses *cleaning data* (impute, dll)?**

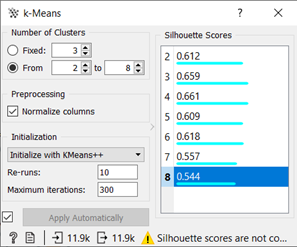
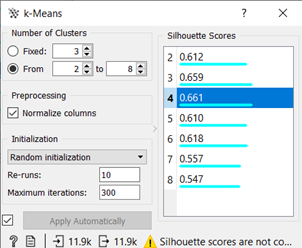
* ada

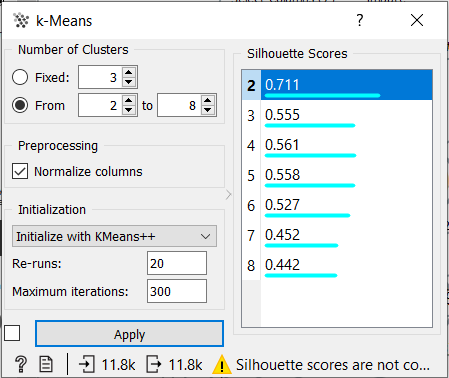
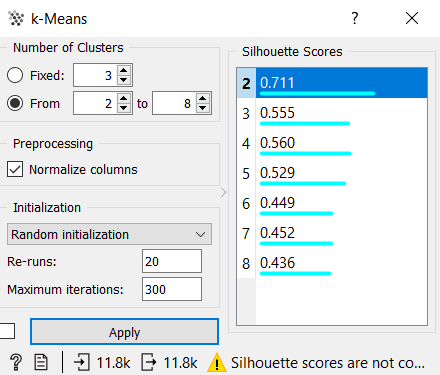
**4. Bagaimana hasil cluster jika centroid terbentuk secara random atau dipilihkan yang terbaik kmeans ++?**

Pada fitur *Engine Cylinder* dan *Engine HP*



Pada fitur *MSRP* dan *Year*





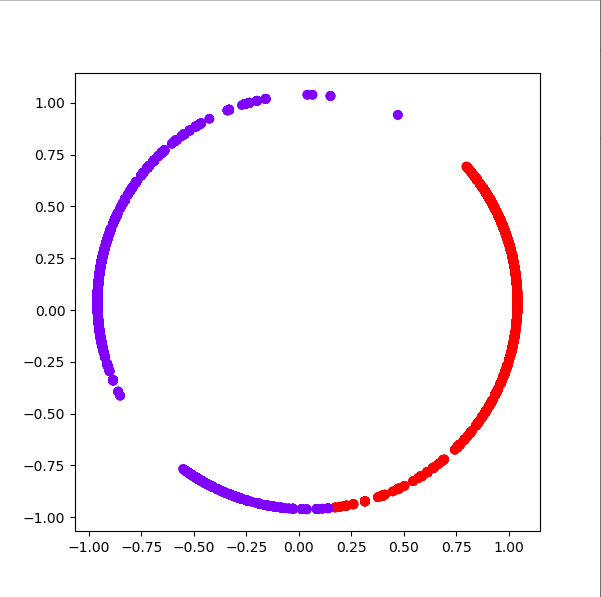
Rata-rata nilai silhouette lebih tinggi

**5. Bandingkan K Means dengan Hierarchical (divisive, agglomerative)!**

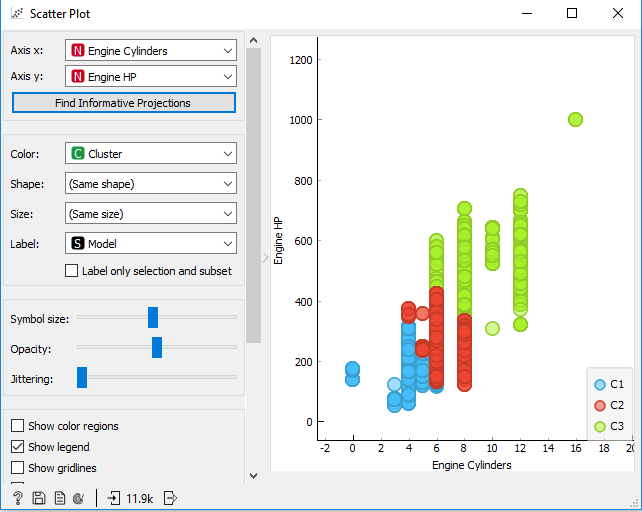
Waktu kompleksitas K means linier O(n) sedangkan Hierarchical O(n).

**6. Bandingkan dengan algoritma jika menggunakan SkLearn**

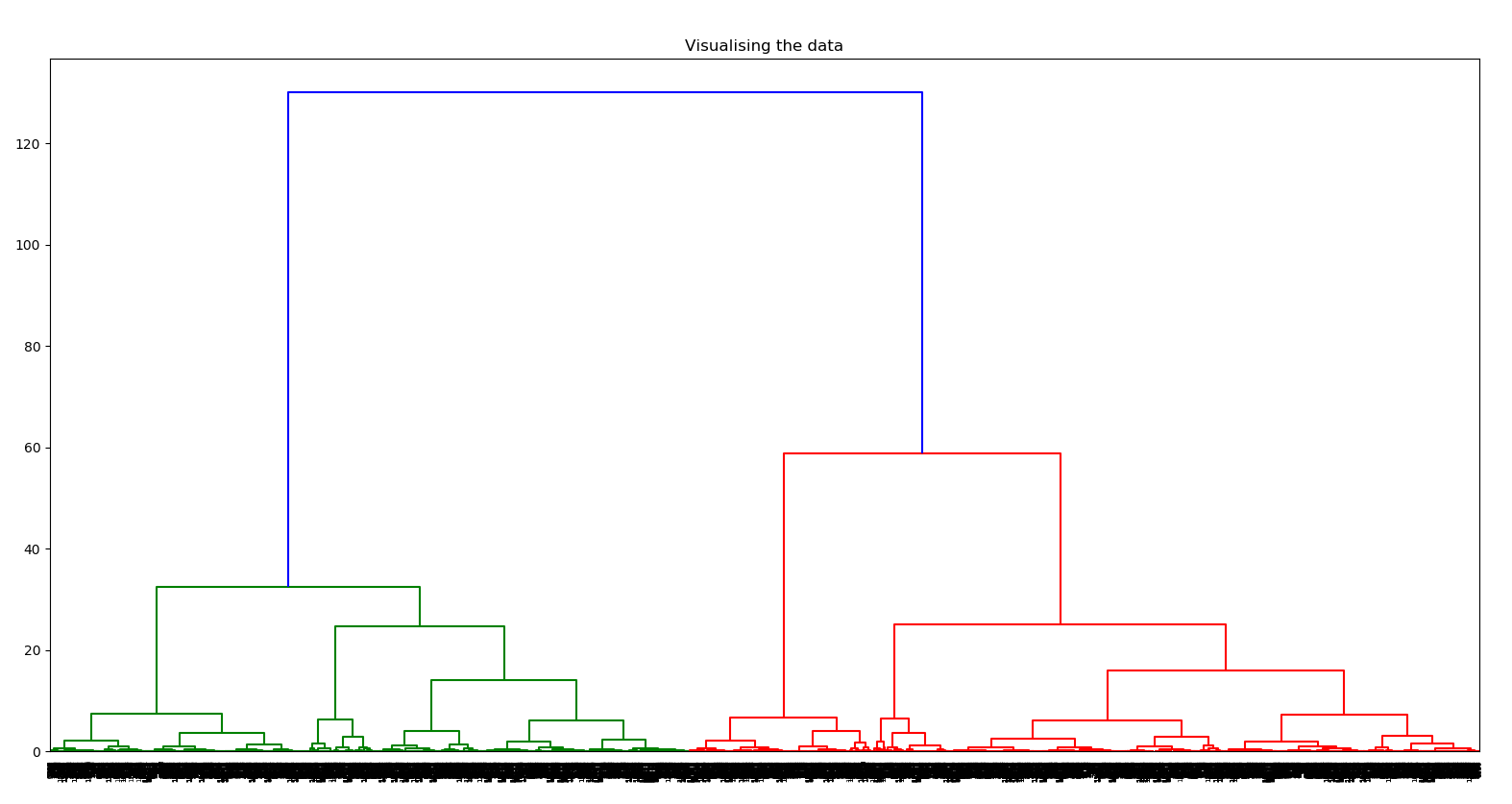
**Kelompok kami melakukan pengcekan pada fitur Engine HP dan Engine Cylinder**



Agglomerative Hierarchical Clustering



Scater plot Kmeans



Dendogram untuk hierarchical clustering Agglomerative

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

Saat melakukan percobaan fitur fitur diatas