

## **KELOMPOK 4**

1. Ni Putu Ayu Triana	1908561031	Aachen
2. Desak Putu Sri Wulandari	1908561019	Aachen
3. Khaerul Anwar	1908561020	Aachen
4. Ni Komang Santi Cahyani	1908561017	Aachen
5. Winda Kuncorowati	T201810001	Apollo
6. Muhammad Rafly Ramadhani	02511940000050	Apollo

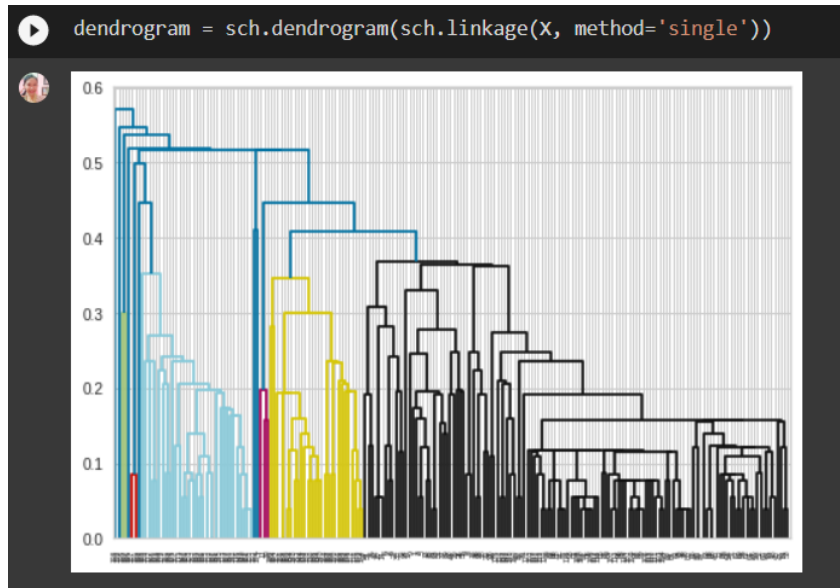
**Orbit Future Academy  
2022**

## Sesi Pagi

- Membuat dendrogram, model clustering, visualisasi hasil clustering, dan rata-rata cluster, tapi dengan ketentuan:

method/linkage = 'single', affinity = 'manhattan' (jarak manhattan)

### Dendrogram

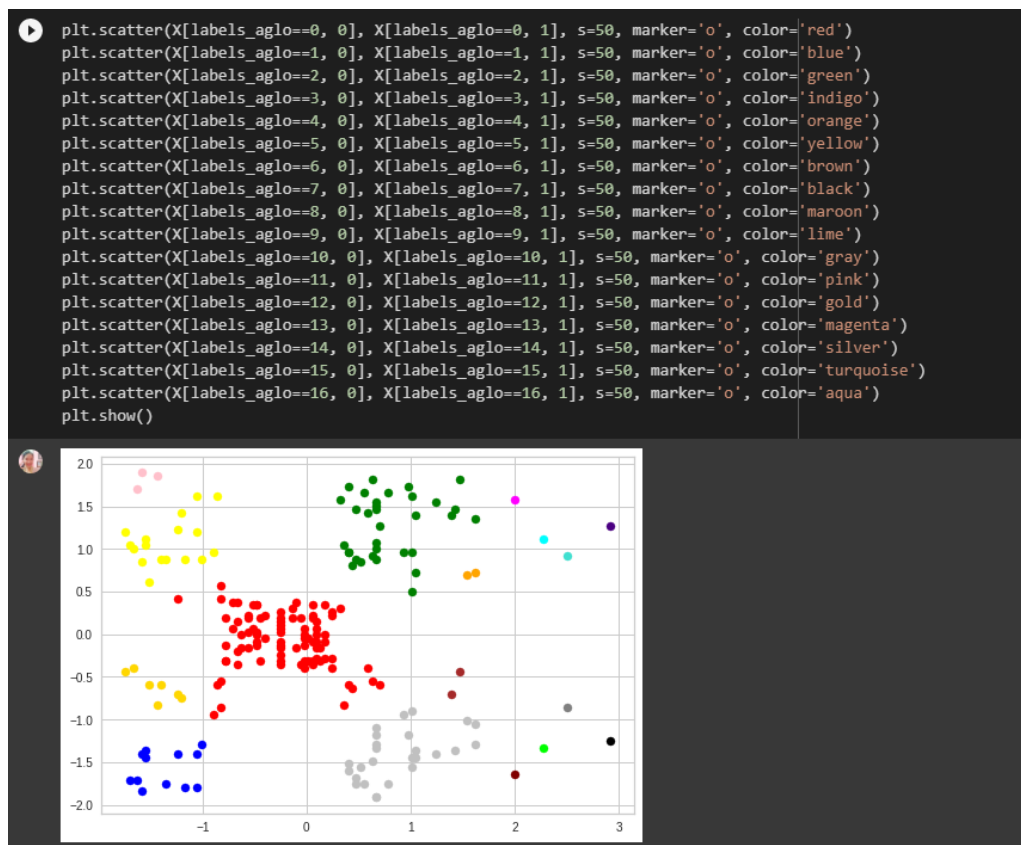


### Model

```
aglo = AgglomerativeClustering(n_clusters=17, affinity='manhattan', linkage='single')
aglo.fit(X)
labels_aglo = aglo.labels_
labels_aglo
```

```
array([[12,  5,  1,  5, 12,  5,  1, 11,  1,  5,  1, 11,  1,  5,  1,  5, 12,
        5, 12, 11, 12,  5,  1,  5,  1,  5, 12,  0, 12,  5,  1,  5,  1,  5,
        1,  5,  1,  5,  0,  5,  0,  5,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,
        0,  0,  0,  0,  2,  0,  2,  0,  2, 14,  2, 14,  2,  0,  2, 14,  2,
        14,  2, 14,  2, 14,  2,  0,  2, 14,  2,  0,  2, 14,  2, 14,  2, 14,
        2, 14,  2, 14,  2, 14,  2,  0,  2, 14,  2, 14,  2, 14,  2, 14,  2,
        14,  2, 14,  2, 14,  2, 14,  2, 14,  2,  6,  2, 14,  2,  6,  2, 14,
        4, 14,  2, 14,  4,  8, 13,  9, 16, 10, 15,  7,  3])
```

### Visualisasi



## Tabel hasil Cluster Agglomerative Clustering

```
##Menambahkan hasil clustering ke dataframe
data['cluster_aglo'] = labels_aglo
data
```

	Genre	Age	Annual Income (k\$)	Spending Score (1-100)	cluster_aglo
CustomerID					
1	Male	19	15	39	12
2	Male	21	15	81	5
3	Female	20	16	6	1
4	Female	23	16	77	5
5	Female	31	17	40	12
...	...	...	...	...	...
196	Female	35	120	79	16
197	Female	45	126	28	10
198	Male	32	126	74	15
199	Male	32	137	18	7
200	Male	30	137	83	3

200 rows × 5 columns

## Rata-rata per cluster

```
#Mencari rata-rata per-cluster
for i in range(max(labels_aglo)+1):
    print("cluster ke-", i)
    print("rata2 pemasukan = ", data[data["cluster_aglo"]==i]["Annual Income (k$)"].mean(),
          ", rata2 pengeluaran = ", data[data["cluster_aglo"]==i]["Spending Score (1-100)"].mean())

cluster ke- 0
rata2 pemasukan = 54.72222222222222 , rata2 pengeluaran = 48.46666666666667
cluster ke- 1
rata2 pemasukan = 24.583333333333332 , rata2 pengeluaran = 9.583333333333334
cluster ke- 2
rata2 pemasukan = 81.06060606060606 , rata2 pengeluaran = 83.0
cluster ke- 3
rata2 pemasukan = 137.0 , rata2 pengeluaran = 83.0
cluster ke- 4
rata2 pemasukan = 102.0 , rata2 pengeluaran = 68.5
cluster ke- 5
rata2 pemasukan = 25.823529411764707 , rata2 pengeluaran = 78.17647058823529
cluster ke- 6
rata2 pemasukan = 98.0 , rata2 pengeluaran = 35.5
cluster ke- 7
rata2 pemasukan = 137.0 , rata2 pengeluaran = 18.0
cluster ke- 8
rata2 pemasukan = 113.0 , rata2 pengeluaran = 8.0
cluster ke- 9
rata2 pemasukan = 120.0 , rata2 pengeluaran = 16.0
cluster ke- 10
rata2 pemasukan = 126.0 , rata2 pengeluaran = 28.0
cluster ke- 11
rata2 pemasukan = 20.0 , rata2 pengeluaran = 97.0
cluster ke- 12
rata2 pemasukan = 22.428571428571427 , rata2 pengeluaran = 34.42857142857143
cluster ke- 13
rata2 pemasukan = 113.0 , rata2 pengeluaran = 91.0
cluster ke- 14
rata2 pemasukan = 83.42307692307692 , rata2 pengeluaran = 13.76923076923077
cluster ke- 15
rata2 pemasukan = 126.0 , rata2 pengeluaran = 74.0
cluster ke- 16
rata2 pemasukan = 120.0 , rata2 pengeluaran = 79.0
```

## Kesimpulan:

Hasil clustering menggunakan method single dan affinity mahattan tidak bagus karena beberapa titik yang berdekatan memiliki warna yang berbeda. Jika jumlah cluster dikurangi, maka titik yang berjauhan memiliki warna yang sama (1-12), jika jumlah cluster ditambah, maka akan ada cluster yang tidak memiliki anggota ( $>17$ ). Dilihat dari dendrogram, jarak terpendek masing - masing data tidak dapat didefinisikan karena membandingkan antar data sehingga cluster yang dibentuk banyak dan tidak dapat didefinisikan jumlahnya. Jika data ini dicari menggunakan jarak terdekat masing - masing data, maka hasilnya tidak bagus (dibuktikan pada pengimplementasian code), berbanding terbalik dengan method "average" (dicoba oleh Winda) dan "complete" (tugas individu) yang menghasilkan cluster lebih baik meskipun memiliki afiinity yang sama.