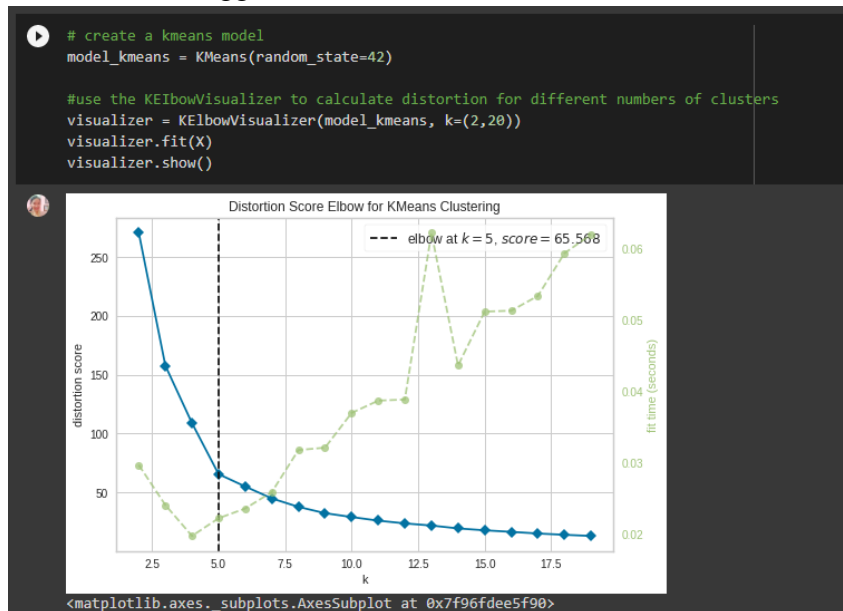# KELOMPOK 4

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 Siang

- Menghitung silhouete score untuk model agglomerative clustering dan membandingkannya dengan K-Means, mencari rata-rata income dan pengeluaran per cluster K-Means dan menginterpretasikannya.

Menghitung silhouete score model agglomerative dan K-Means

Visualisasi menggunakan KElbowVisualizer

```
# create a kmeans model
model_kmeans = KMeans(random_state=42)

#use the KEIbowVisualizer to calculate distortion for different numbers of clusters
visualizer = KElbowVisualizer(model_kmeans, k=(2,20))
visualizer.fit(X)
visualizer.show()
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x7f96fdee5f90>
```
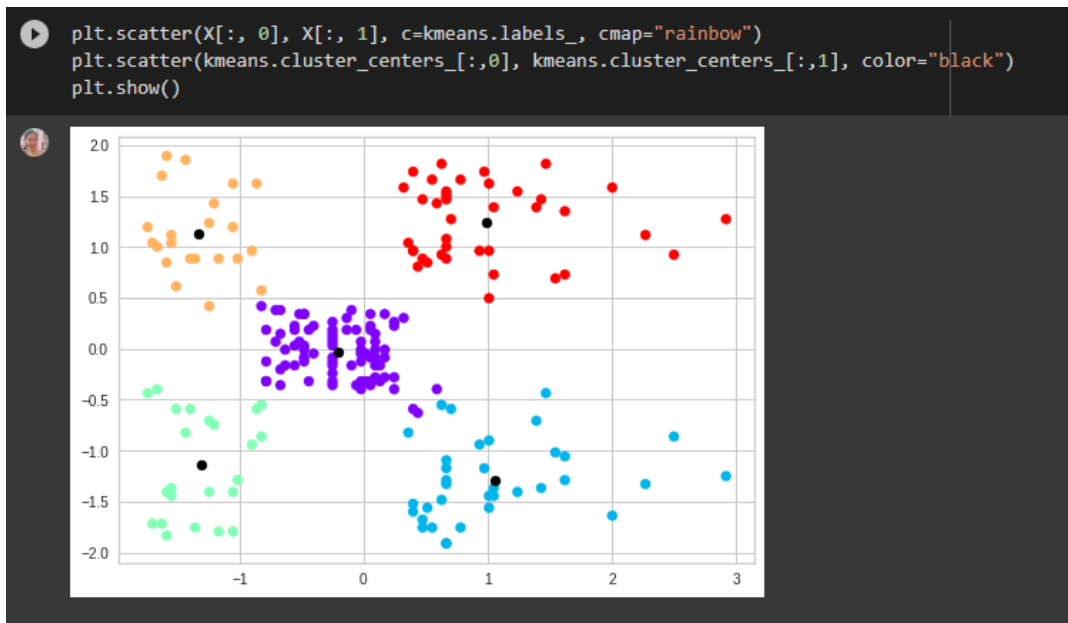
Menggunakan 5 cluster

```
#Gunakan k=5
kmeans = KMeans(n_clusters=5, random_state=42)
kmeans.fit(X)
print(kmeans.cluster_centers_)
```

```
[[-0.20091257 -0.02645617]
 [ 1.05500302 -1.28443907]
 [-1.30751869 -1.13696536]
 [-1.32954532  1.13217788]
 [ 0.99158305  1.23950275]]
```

Menampilkan sebaran data dengan k-means

```
plt.scatter(X[:, 0], X[:, 1], c=kmeans.labels_, cmap="rainbow")
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], color="black")
plt.show()
```


Perhitungan silhouette score Agglomerative Clustering dan K-Means

```
[ ]  label_kmeans = kmeans.labels_
     label_kmeans

     array([2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3,
            2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 0,
            2, 3, 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,
            0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 4, 1, 4, 0, 4, 1, 4, 1, 4,
            0, 4, 1, 4, 1, 4, 1, 4, 1, 4, 0, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4,
            1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4,
            1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4, 1, 4,
            1, 4], dtype=int32)
```

Evaluasi silhouette score

```
[ ]  ss_agglo = silhouette_score(X, labels_agglo, metric='euclidean')
     ss_kmeans = silhouette_score(X, label_kmeans, metric='euclidean')

     print('nilai silhouette score untuk Agglomerative Clustering = ', ss_agglo)
     print('nilai silhouette score untuk Kmeans Clustering = ', ss_kmeans)

     nilai silhouette score untuk Agglomerative Clustering =  0.5314268013240602
     nilai silhouette score untuk Kmeans Clustering =  0.5546571631111091
```

Hasil Silhouette K-Means lebih baik dibandingkan Agglomerative clustering, karena hasil K-Means lebih mendekati 1

Data hasil clustering dengan K-Means

```
[ ] data['cluster_kmeans'] = label_kmeans
    data
```

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

200 rows × 6 columns

Rata-rata per Cluster

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

cluster ke- 0
rata2 pemasukan =  55.2962962962963 , rata2 pengeluaran =  49.51851851851852
cluster ke- 1
rata2 pemasukan =  88.2 , rata2 pengeluaran =  17.114285714285714
cluster ke- 2
rata2 pemasukan =  26.304347826086957 , rata2 pengeluaran =  20.91304347826087
cluster ke- 3
rata2 pemasukan =  25.727272727272727 , rata2 pengeluaran =  79.36363636363636
cluster ke- 4
rata2 pemasukan =  86.53846153846153 , rata2 pengeluaran =  82.12820512820512
```

Hasil :

Cluster 0 adalah kelompok keuangan menengah ke-atas, baik secara pemasukan maupun pengeluaran

Cluster 1 adalah kelompok yang kaya dan hemat, pengeluarannya sangat kecil

Cluster 2 adalah kelompok keuangan rendah bawah, baik secara pemasukan maupun pengeluaran

Cluster 3 adalah kelompok yang boros, pemasukannya rendah namun pengeluarannya cukup besar

Cluster 4 adalah kelompok yang kaya, namun pengeluarannya juga lumayan