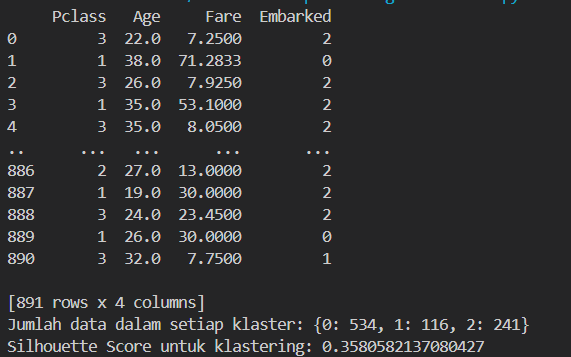
K-MEAN CLUSTER

Deskripsi data

|  |  |
| --- | --- |
| Ticket Class | 1 = 1st, 2 = 2nd, 3 = 3rd |
| Age | Usia dalam tahun |
| fare | Tarif penumpang |
| embarked | Pelabuhan Embarkasi |

Disini saya akan mencoba menbuat cluster pada data titanic passanger dengan 3 cluster, dengan data:

* Pclass : tiket class (kelas 1,kelas 2 dan kelas 3)
* Age : usia penumpang
* Fare : tarif penumpang
* Embarked : Pelabuhan naiknya peunmpang (0 = Cherbourg, 1 = Queenstown, 2 = Southampton)

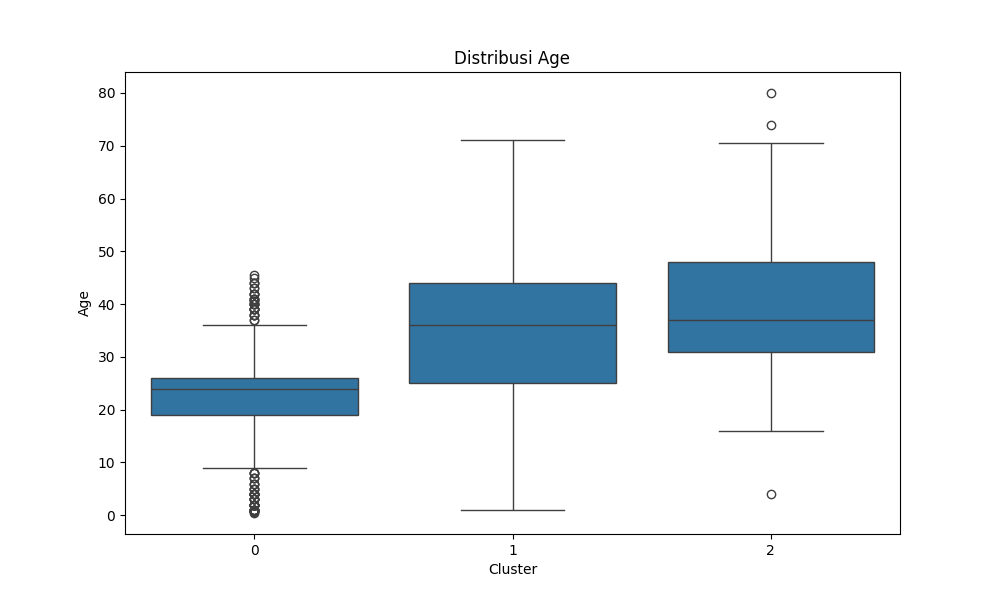


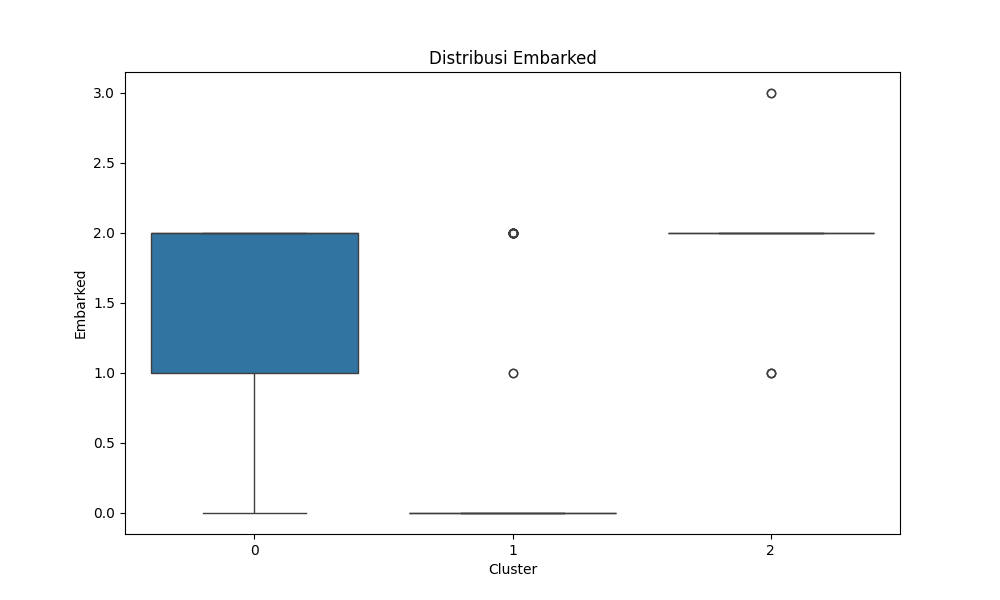
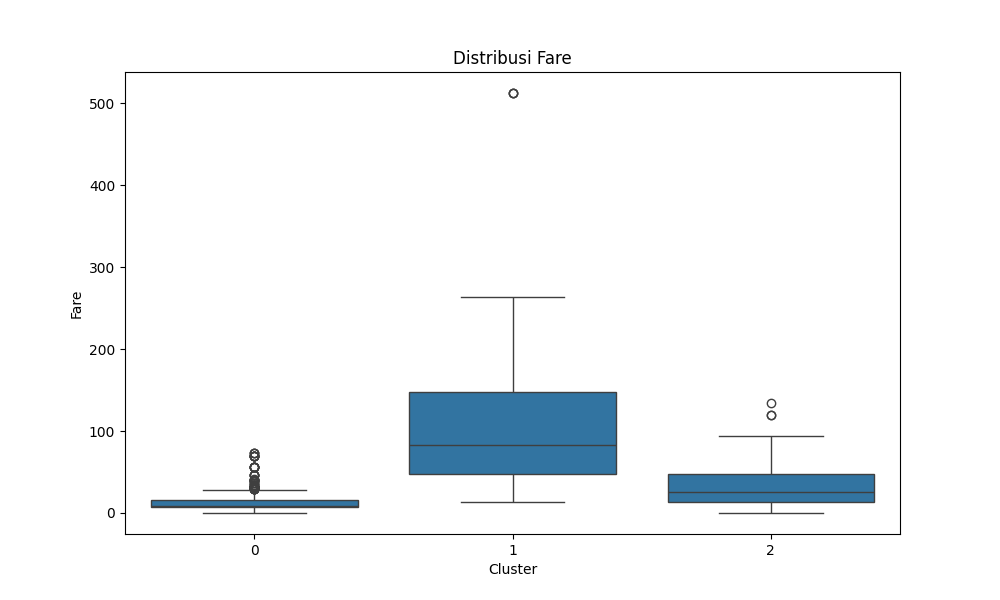
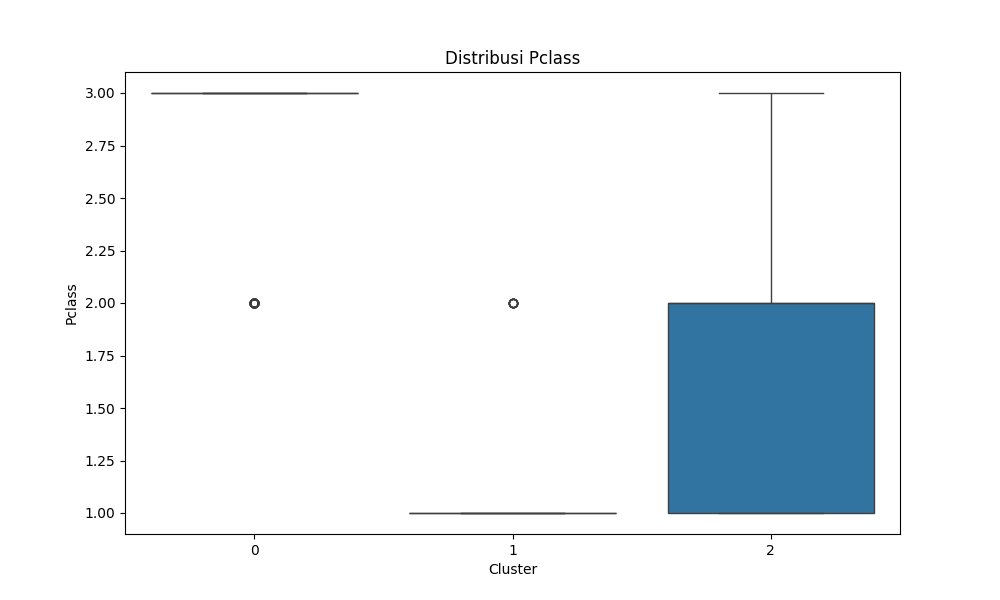
Dari output diatas bisa di dapat, Dari data cluster 1-3 itu ada penyeberan berbeda-beda, diantaranya cluster 1 sebanyak 534 data, cluster 2 sebanyak 116 data dan cluster 3 sebanyak 241 dengan jumlah data 891.

Nilai silhouette Score 0.35 artinya pengclusteran data kurang baik, bisa dilakukan pengclusteran dengan iterasi beberap kali juga bisa dengan metode pendekatan berbeda seperti Euclid

Dari hasil distribusi penyebaran data yang paling baik pada 3 cluster adalah usia penumpang

Jadi bisa disimpulkan secara sederhana untuk hasil pengclusteran data pada data titanic passanger masih kurang baik.



dari output diatas berikut penjelasan sederhana dari saya :

* Didapat Nilai silhouette Score **0.35** artinya *pengclusteran data kurang baik*, dan hal ini dari referensi yang saya dapat bisa dilakukan pengclusteran dengan iterasi beberap kali agar penclusteran lebih optimah, juga bisa dengan metode pendekatan berbeda seperti selain Euclid distance ada manhattan distance dan minkawski distance
* Dari hasil distribusi penyebaran data yang paling baik pada 3 cluster adalah usia penumpang,dengan tiap cluster data bisa berada

kesimpulan

* Jadi bisa disimpulkan secara sederhana untuk hasil pengclusteran data pada data titanic passanger masih kurang baik.

demikian intepretasi dari saya kurang nya mohon maaf dan mohon bimbingannya