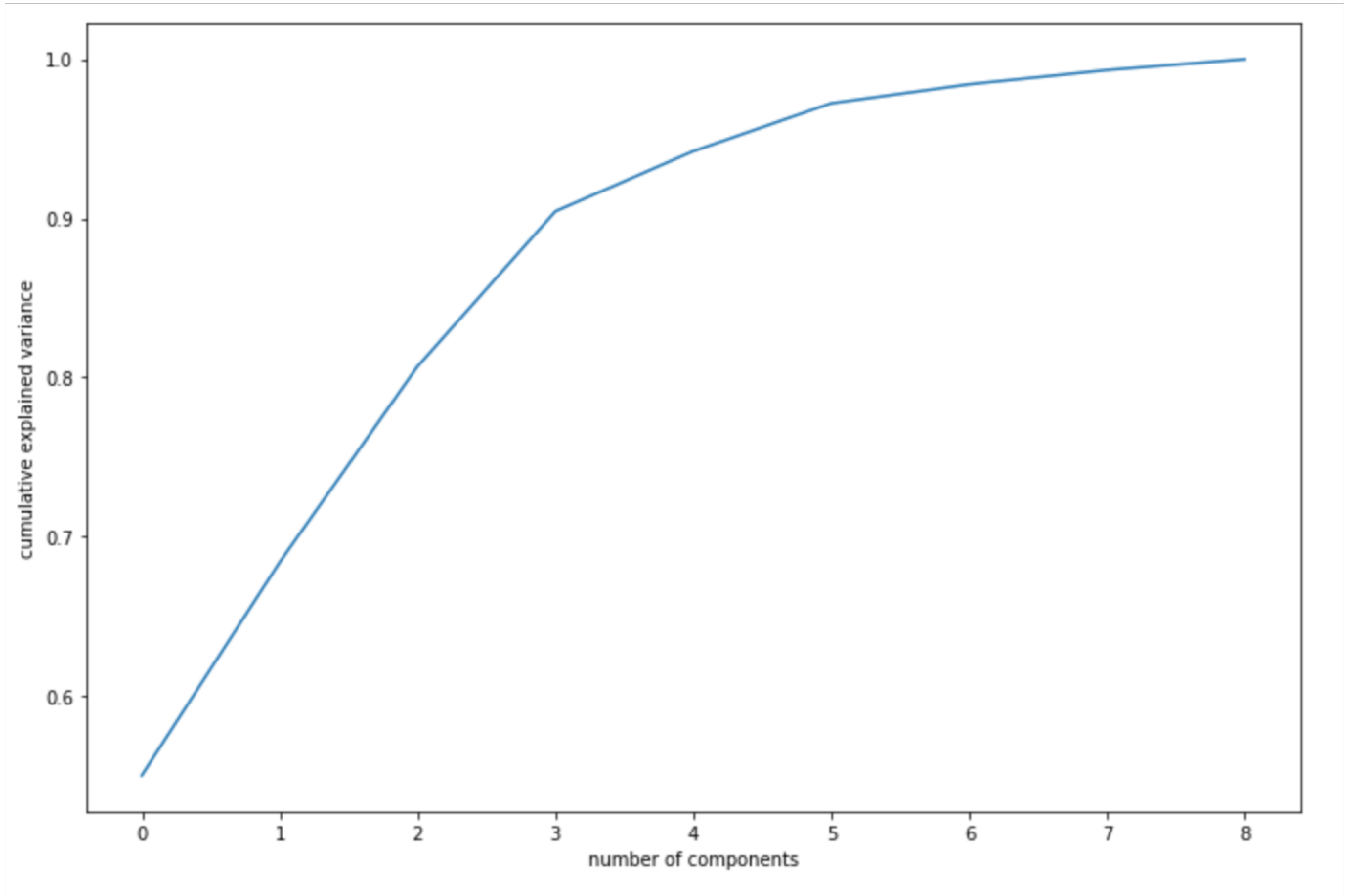
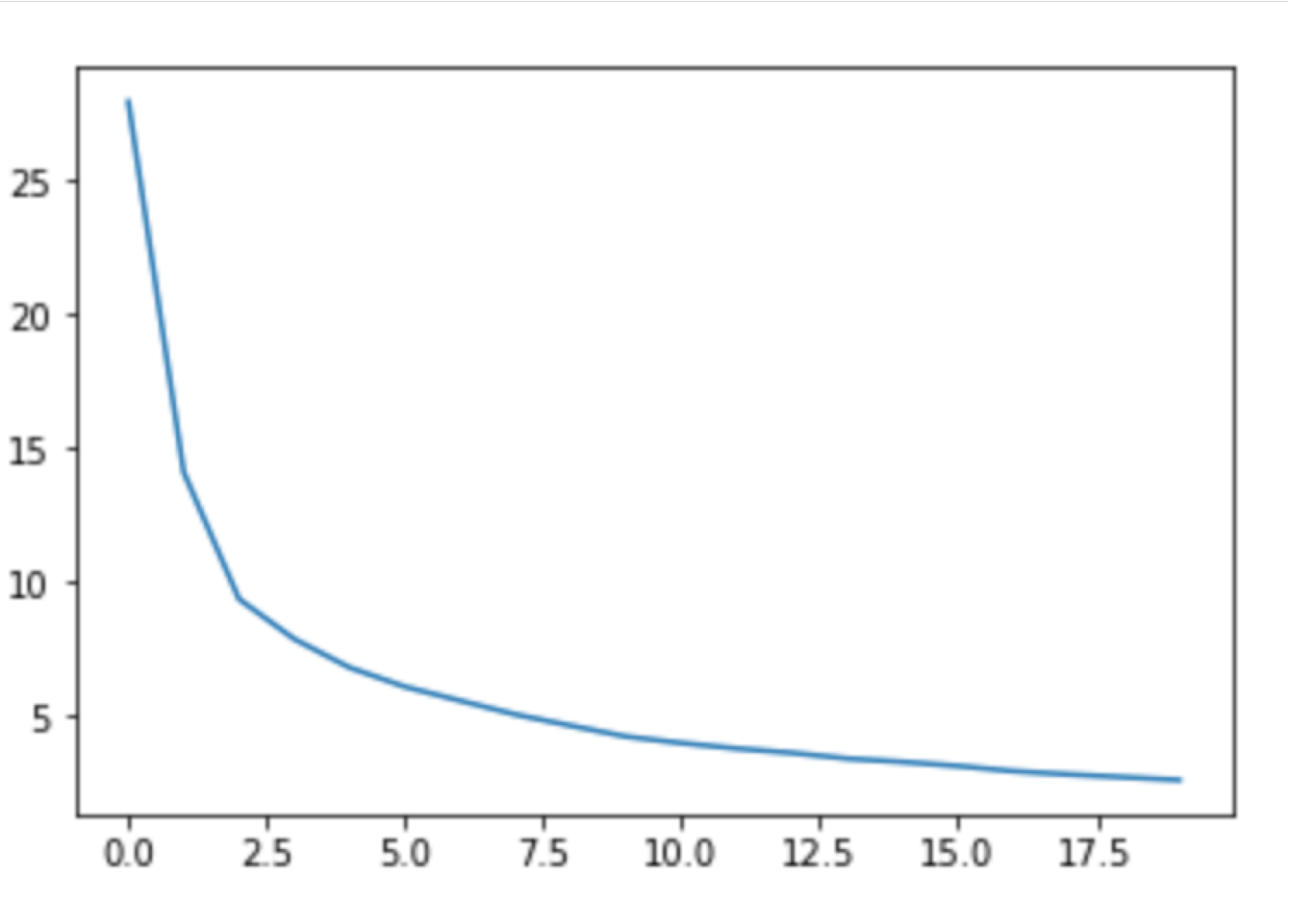
PCA AND CLUSTERING

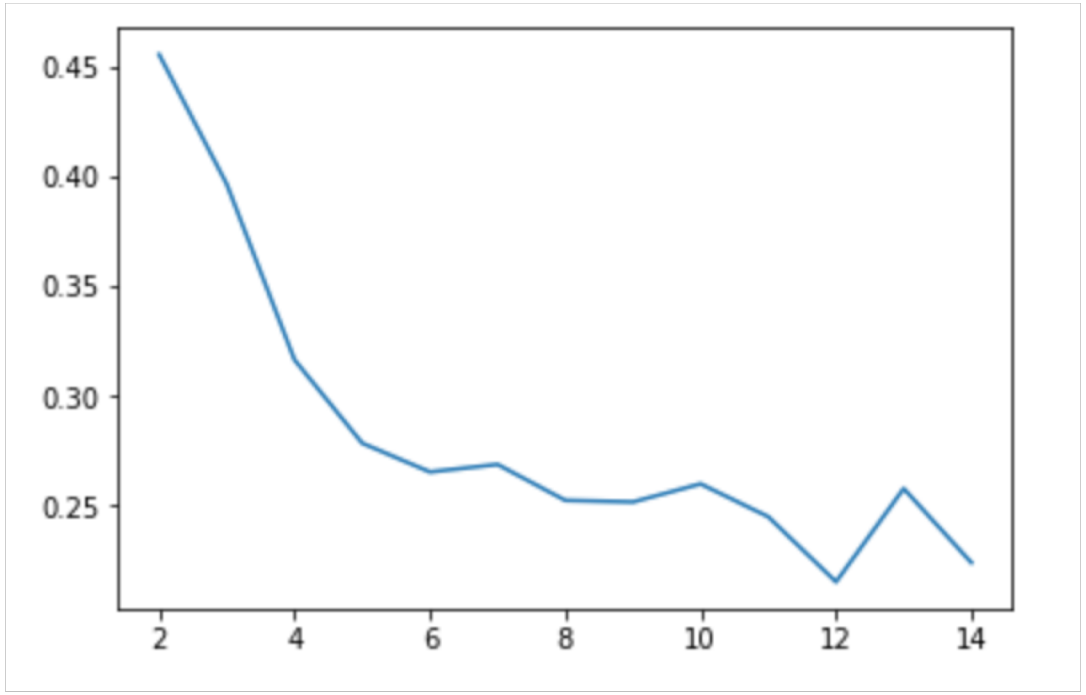


1. IDEALLY WE WOULD WANT TO CATCH +90 VARIENCE.

   SO WE SHOULD USE 4 OR 5 COMPONENTS.BUT IDEALLY 5 IS BETTER. AS MORE IS BETTER


25
20
15
10
5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

Silhouette score

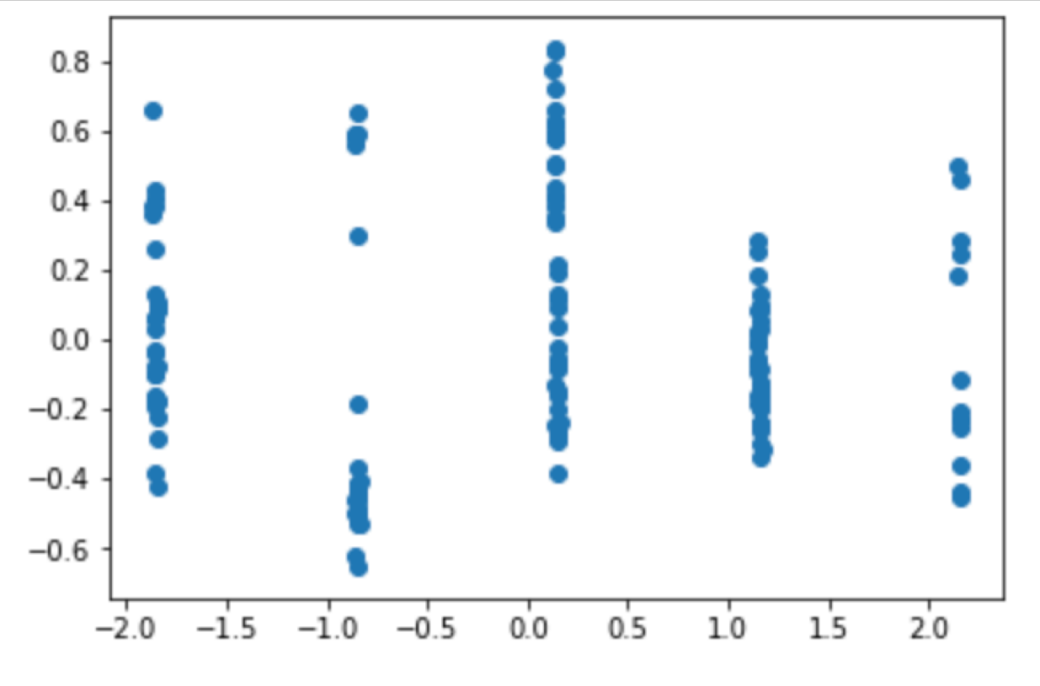


The **silhouette** value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation)

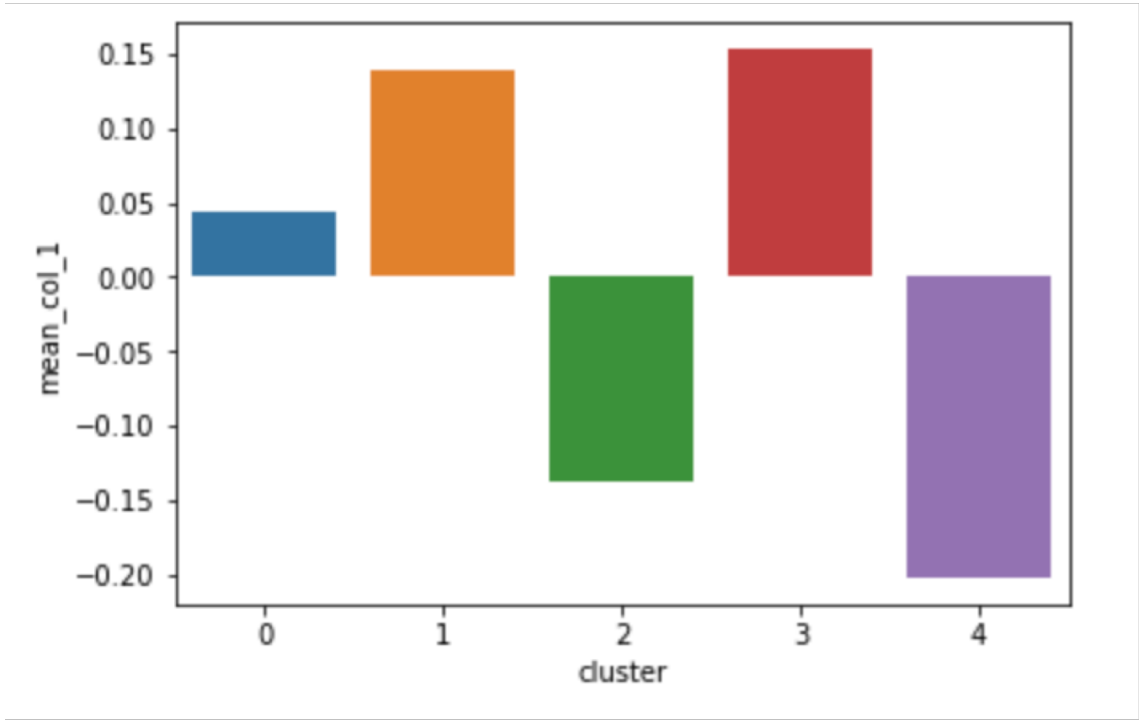
here we could choose 3 or 4 or 5. 5 clusters are better.

	Feature	PC1	PC2	PC3	PC4
0	child_mort	-0.476463	0.281733	0.100012	0.029081
1	exports	0.170333	0.466644	-0.390524	-0.188488
2	health	0.133886	0.044510	0.781911	-0.512983
3	imports	0.081620	0.394257	-0.302775	-0.606656
4	income	0.306981	0.404405	0.068367	0.372823
5	inflation	-0.088009	-0.001602	-0.057029	0.212621
6	life_expec	0.420535	-0.137520	0.007857	0.148021
7	total_fer	-0.571087	0.369246	0.174836	0.173062
8	gdpp	0.338243	0.476078	0.309637	0.310448

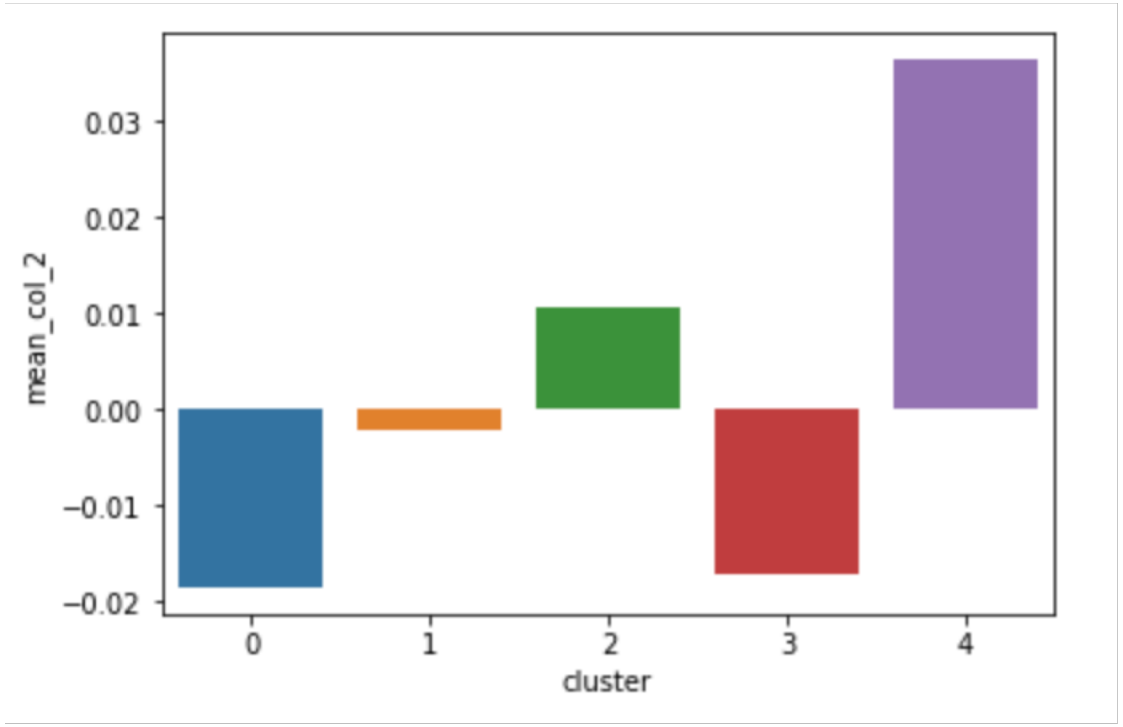
We would like 4 PC components



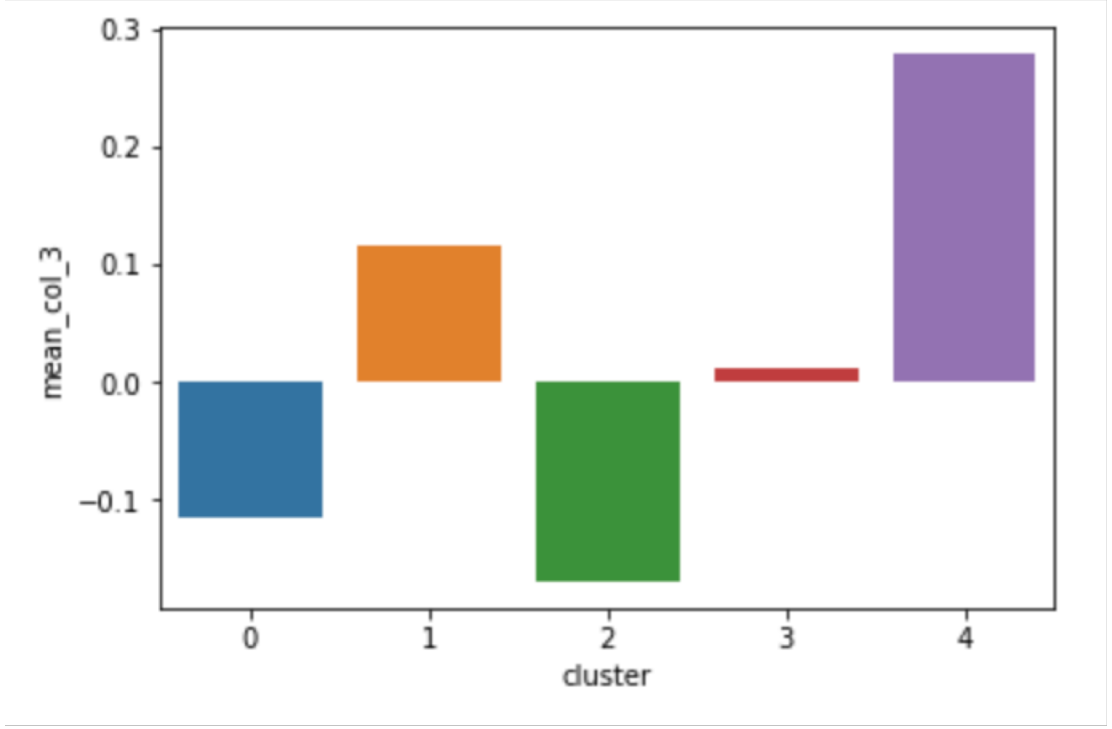
Using k-means and find the 5 clusters



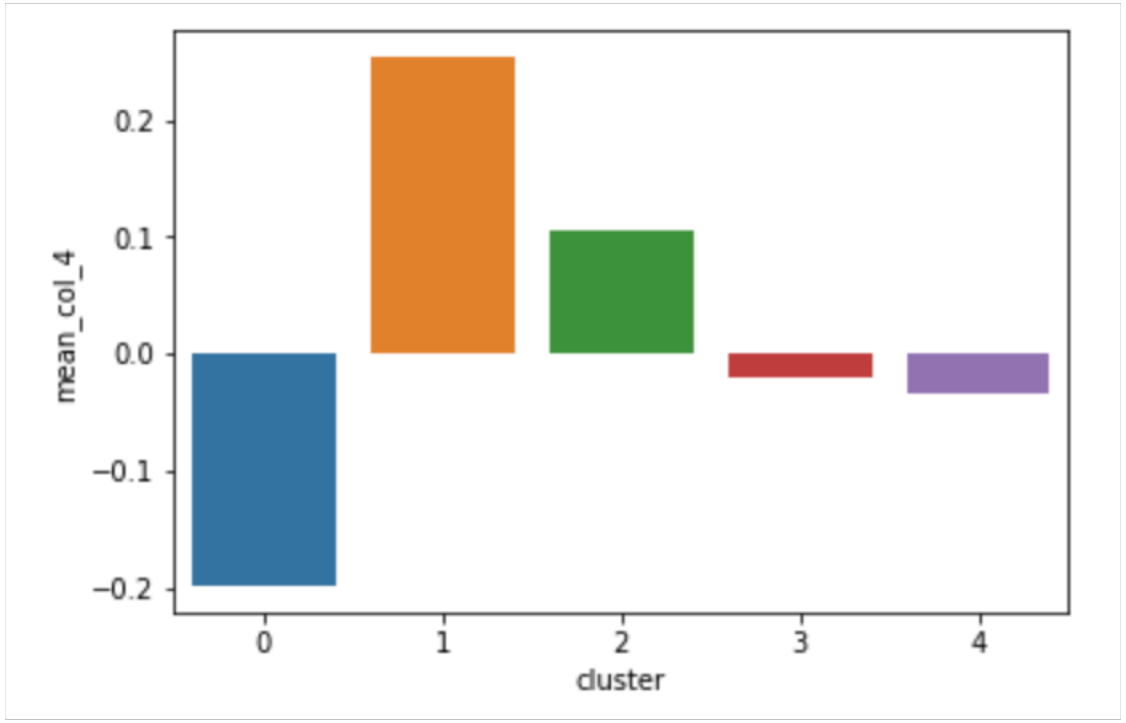
From col_1 mean we could see that cluster 5 and cluster 3 require financial assistance. In financial sector.



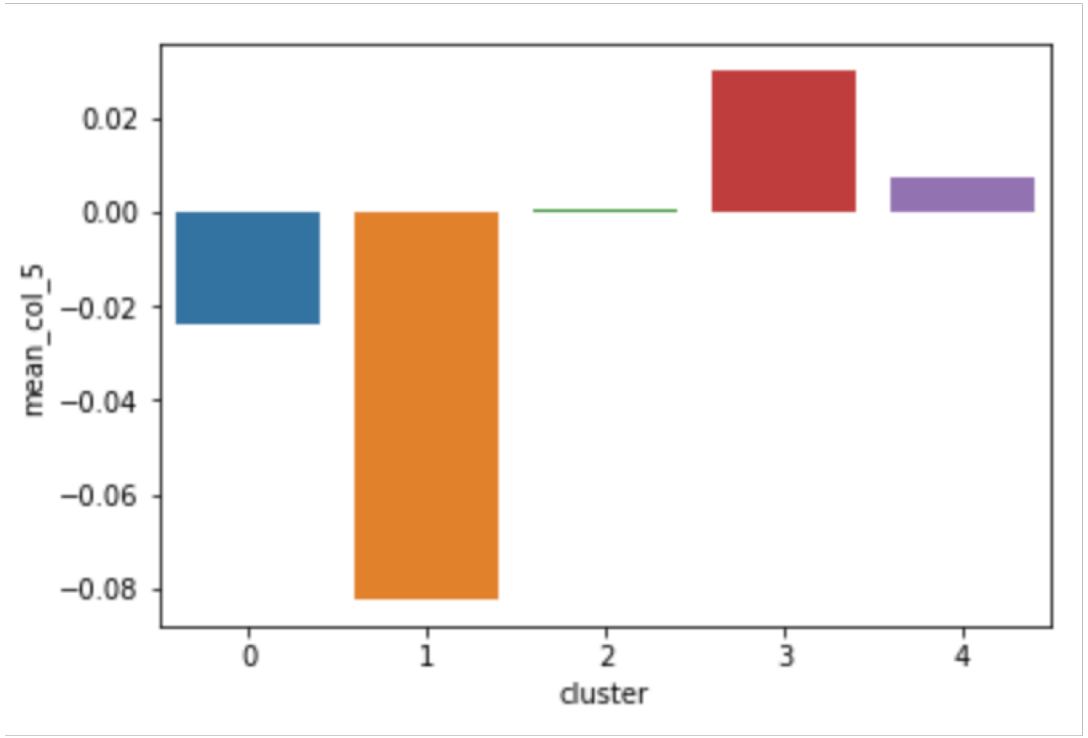
From col_2 mean we could see that cluster 1 and cluster 4 require financial assistance. In health sector



From col_3 mean we could see that cluster 1 and cluster 3 require financial assistance.



From col_4 mean we could see that cluster 1 and cluster 5 require financial assistance.



From col_5 mean we could see that cluster 1 and cluster 2 require financial assistance

conclusion

- We have specified the step used in clustering.
- We feel that 5,3 required financial assistance.
- We feel that 1,4 required medical assistance.
- And k-means clustering is efficient than Hierarchical clustering