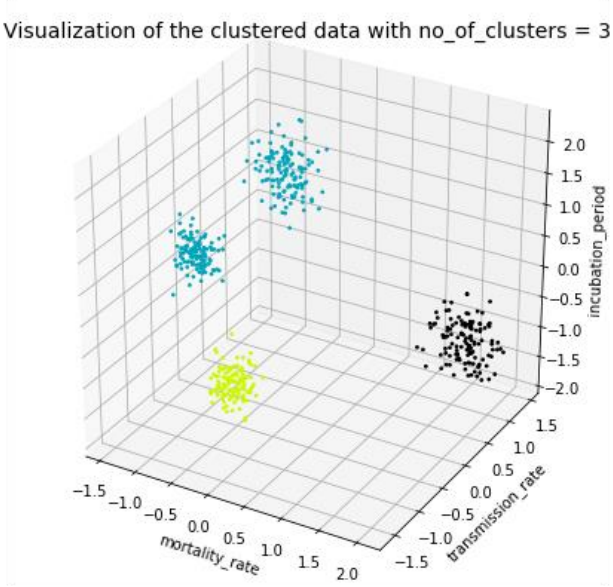
[Mini Project 3] **Project Code : DS2**

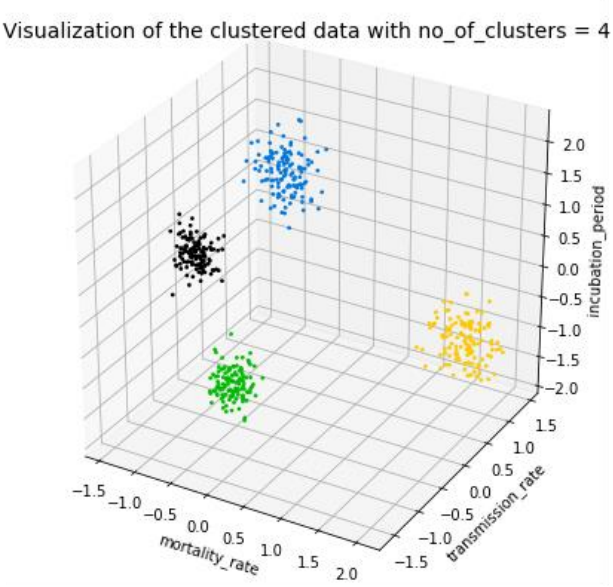
Coronavirus Data Clustering using Single Linkage Hierarchical Clustering Technique
REPORT

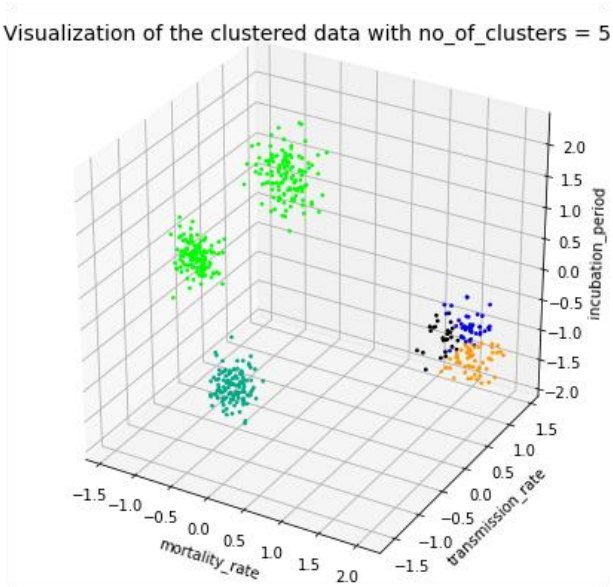
- Pulkit Singhvi (17EE10035)

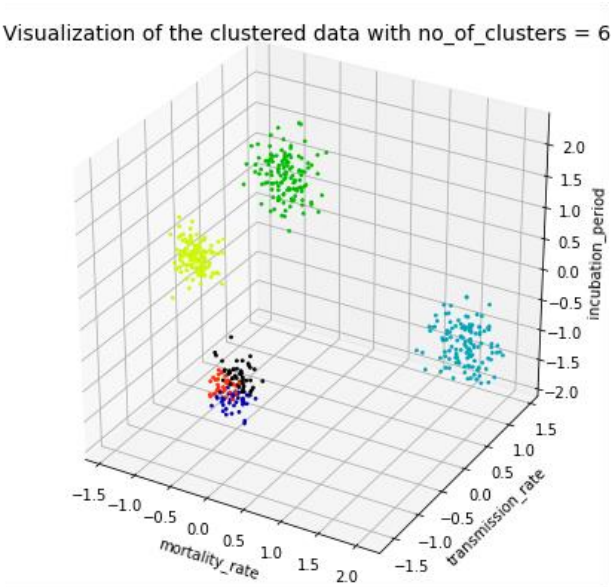
Optimal number of clusters **= 4**
Max. Silhouette Score (achieved at **Optimal_k** = **4)** = **0.769**

No. Of Clusters	Silhouette coefficient
3	0.71069
4	0.76906
5	0.58485
6	0.62957









Analysis of Silhouette Scores:

- For k = 3, the blue chunks are put in the same cluster but they are well apart from each other. But there is no overlapping among different clusters.
  Hence, silhouette score is less compared to that for k=4 but more than those for k=5 and k=6.

- For k = 4, we have fairly dense clustering with clusters well apart from each other which is reflected in the silhouette score too with it being the highest among the 4 cases considered.

- For k = 5, we have overlapping clusters as well as two far apart chunks colored in green are categorized in the same cluster. Hence silhouette score is less compared to that for k = 4.

- For k = 6, we have overlapping clusters but the two far apart chunks which were put in the same cluster for k=5 are now categorized differently. Hence, silhouette score is less than that for k=4 but more compared to that for k=5.

Analysis of Jaccard Similarity Scores:
With the mapping being bijective and the Jaccard score of 1 observed for each mapping, we can conclude both K-Means clustering and Single linkage hierarchal clustering gave the same results for optimal_k .