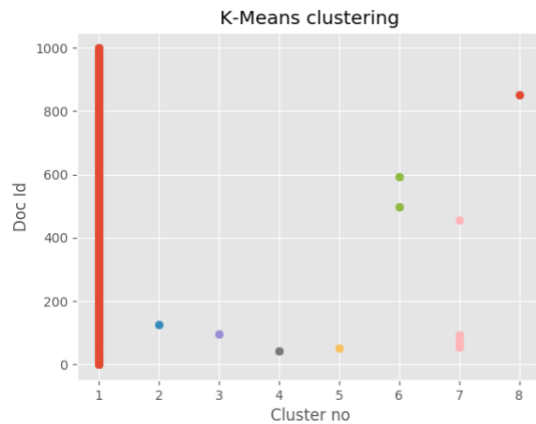


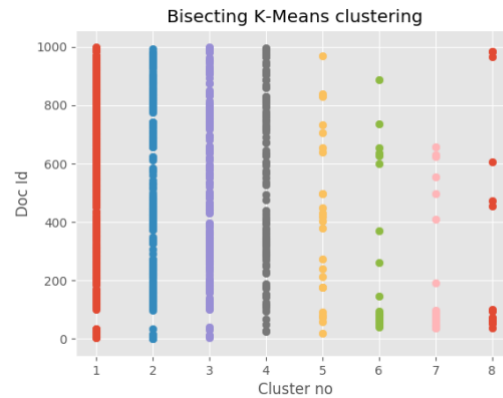
Bonus Question:

Used matplotlib library in python to generate the plots. Code used to generate plots can be found on next page.

K means



Bisecting K means



Description

1. Used matplotlib library in python to generate the plot.
2. X axis is the document number and Y axis is the cluster number the document is assigned to.
3. For K-means, as you can see from the plot, most of the documents fall into cluster and many clusters have only one document in them.
4. But in the bisecting K-means, the documents are well spread across all the clusters which is clearly visible in the plot.
5. Each document is represented as a circle in the plot.
6. Points in each cluster are represented with different color.