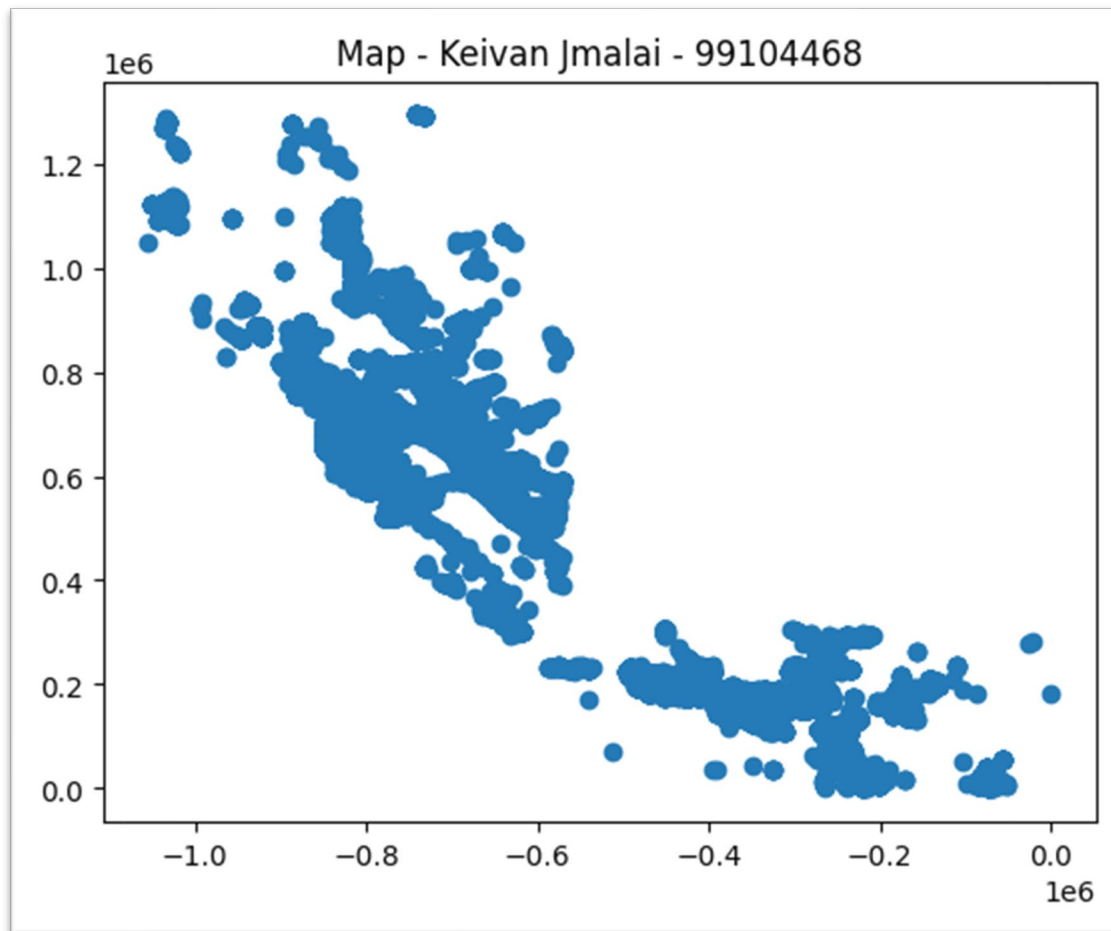


Part 1:

Taking all x and y component of the nodes out of Lng and Lat. This process took about 5 min. However, after that the plotting process was fast. Less than 1 min.



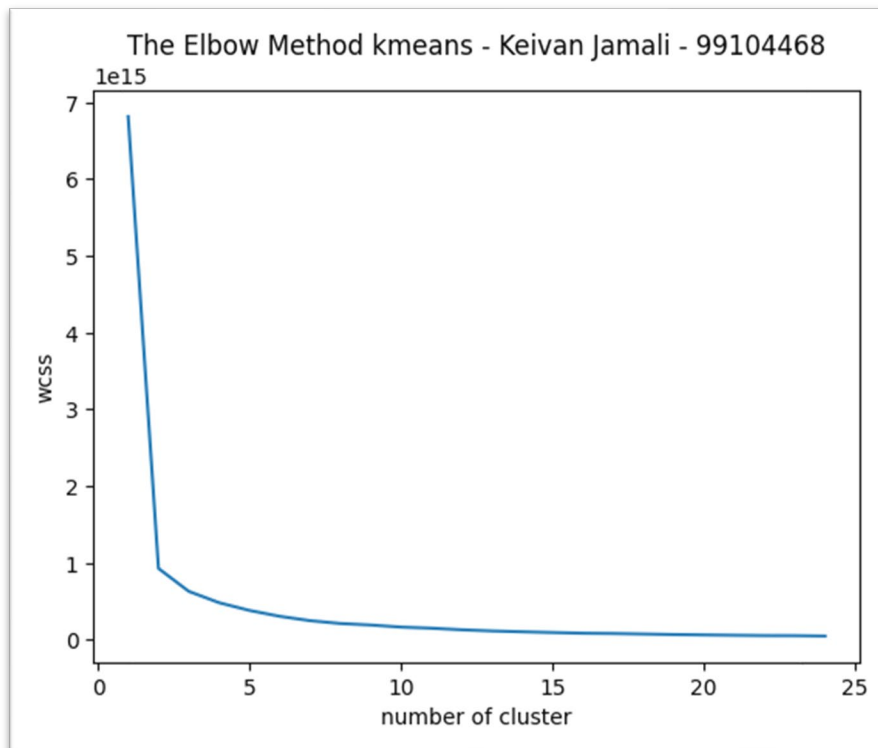
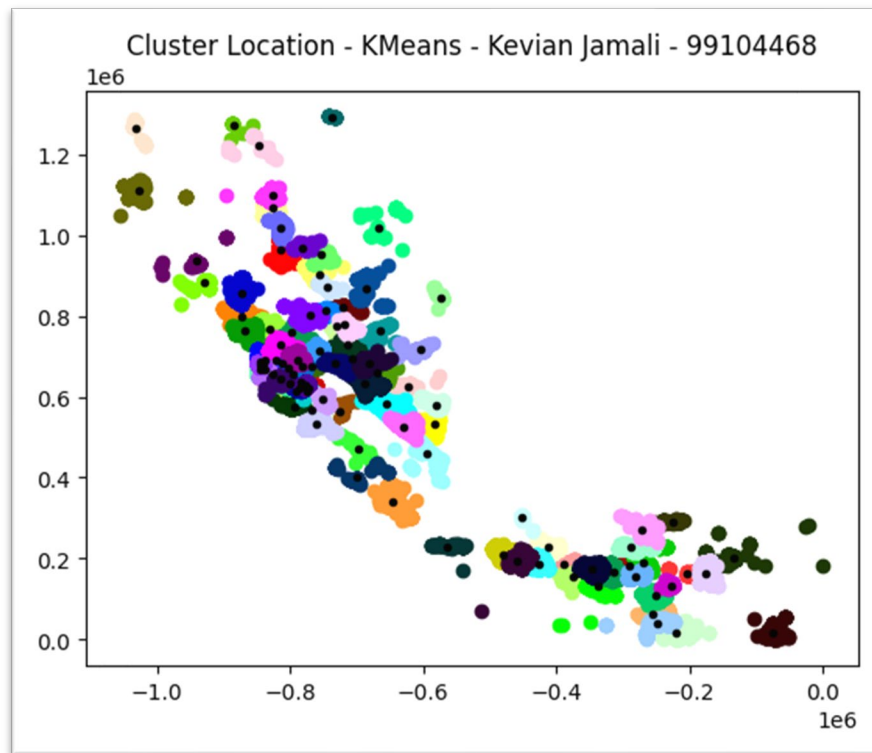
Part 2:

In this part I calculate distances in meter using harvest function, resulting in 936 nodes near than 50 km to zone 8.

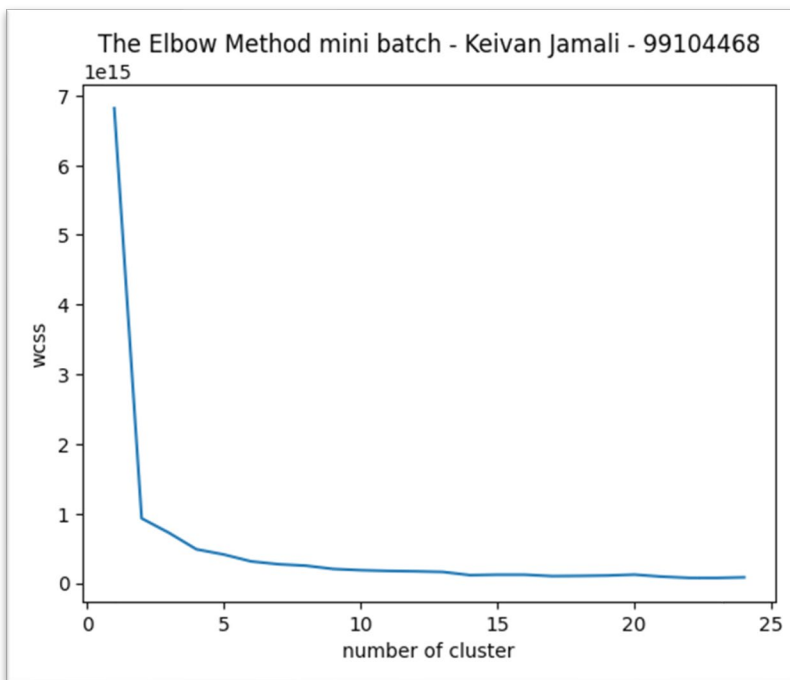
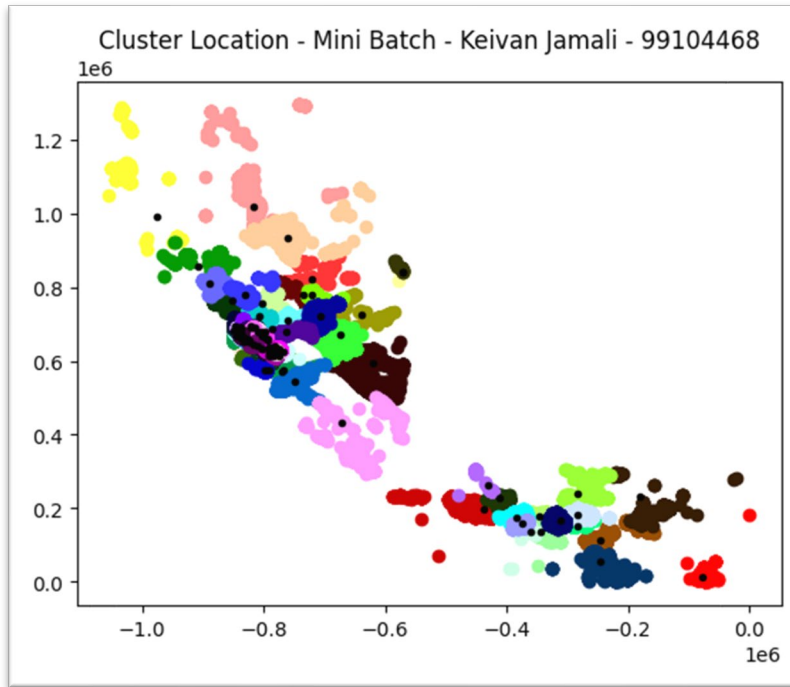
Part 3:

First making the data for training, containing all x and y components. Training all models including KMeans, Minibatch kMeans, DBSCAN and plot them.

I used Elbow method to find the optimum cluster numbers for kmeans which calculate to be 3 or 4.



The same for minibatch kmeans.



For choosing parameters, I try different items and by experiment mostly choose parameters. For example, choosing batch size for minibatch model, for the fast and also good result I choose 128. I also test some other cases, leading to the 128-batch size. For finding eps and min\_samples, I did the same thing and by testing different scale of it and measuring the score of the model, I found with eps=150 and min\_samples=300 will result in the best shape.

For the last part again by using some for loop I found the eps which results in 100 clusters.

All time are tracked in the code, in each modeling we have time tracker. As obviously, by largening the cluster number it will took longer to execute. One other conclusion is that minibatch is super faster compared to kmeans and DBSCAN is faster too. Bu I didn't that much difference between DBSCAN and minibatch kmeans.

#### Part 4:

We make function out of these codes and use them to find time-plot. With 1000 in each step.

