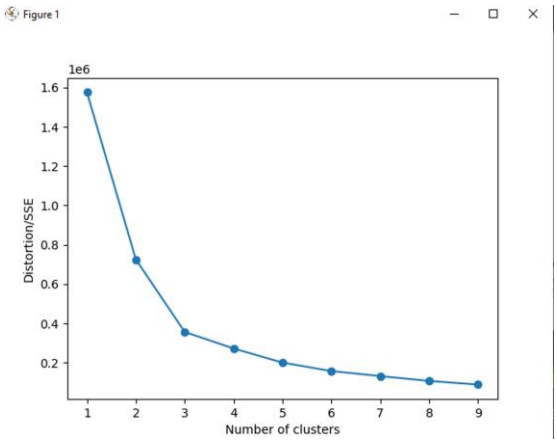
Assignment 2

Part 1



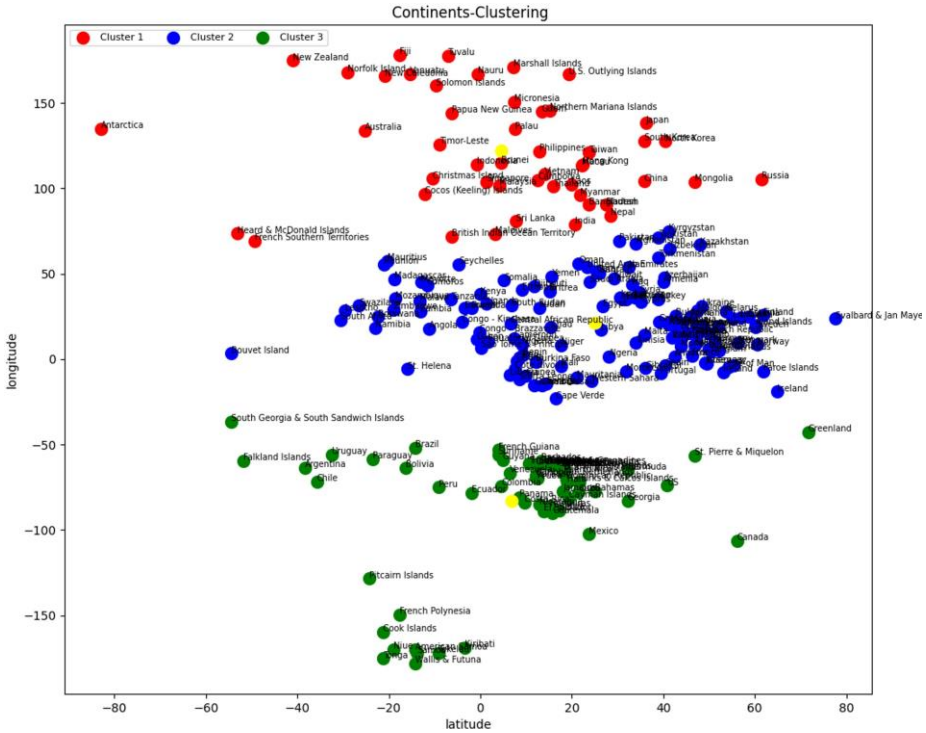
a,b)Using Elbow method,

To determine the optimal number of clusters, we have to select the value of k at the "elbow" ie the point after which the distortion/inertia start decreasing in a linear fashion. Thus for the given data, we conclude that the optimal number of clusters for the data is **3**.

```
MACHINE LEARNING\Assignment2\A2.py"
Coordinates of cluster centers:  [[  4.51565086 121.86887525]
 [ 25.11346528  21.57271323]
 [  6.74638591 -82.962124  ]]
Sum of squared distances to centroids 356365.8708194744
Number of iterations run:  5
```

From part a), we knew that 3 is the optimal number of cluster via the Elbow method and the iterations runs at 5 instead of 300{maximum limit) .This is the best clustering parameter.

Part C)



Cluster	Centroid	Continent
0	4.51565086 121.86887525	Asia/Oceania
1	25.11346528 21.57271323	Africa/Europe
2	6.74638591 -82.962124	South America/North America

Part 2

a)

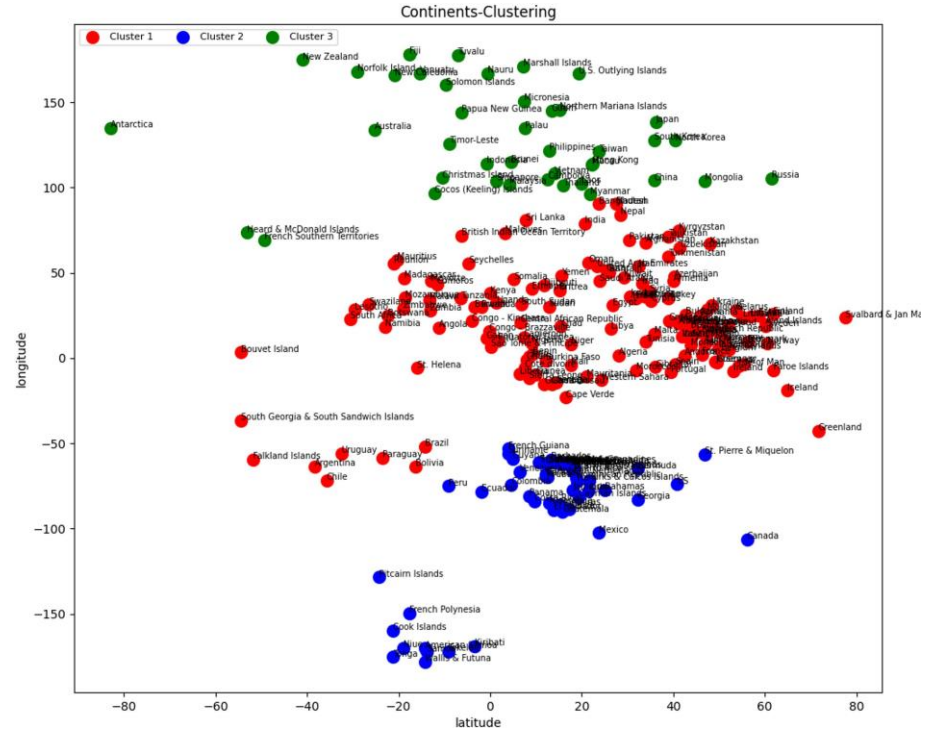
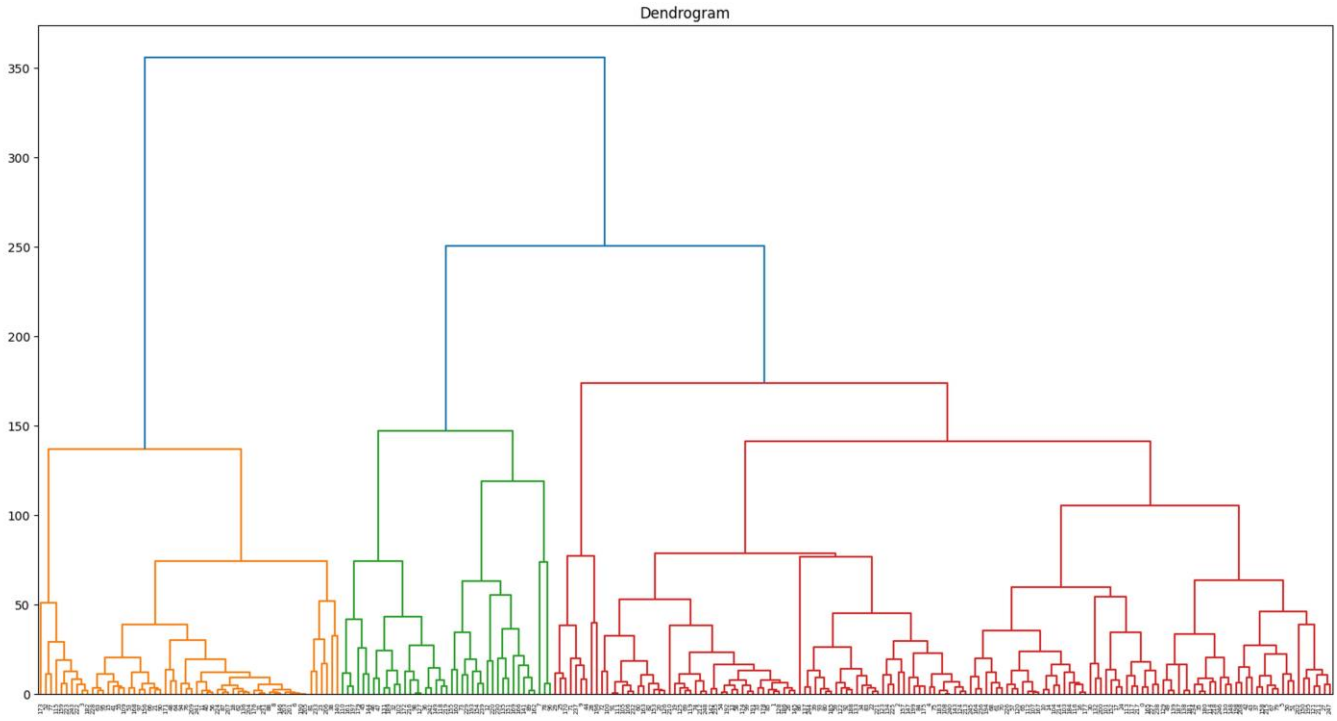


Figure 1



At cluster distance around 250,we have a perfect cut which shows 3 dissimilar clusters.the parameters used are {n_clusters = 3, affinity = 'euclidean', linkage = 'complete').

b)

Cluster	Continent
0	Asia/Oceania
1	Africa/Europe
2	South America/Antartica

Part 3

Kmeans provides the most accurate grouping since for eg, countries like lndia,Nepal and Sri Lanka are indeed belongs to Asia side.

Cluster	Continent
0	Asia/Oceania
1	Africa/Europe
2	South America/North America