

Analysis of Covid Dataset using K-Means and K-MEdoids

Dataset

Dataset was taken from kaggle.com. Dataset represents the number of covid cases and deaths in different states in a day. Data of 566 different states was recorded.

	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths
0	8/28/2020	Abruzzo	Italy	4/2/2021 15:13	3725	472
1	8/28/2020	Acre	Brazil	4/2/2021 15:13	24462	608
2	8/28/2020	Adygea Republic	Russia	4/2/2021 15:13	3327	27
3	8/28/2020	Aguascalientes	Mexico	4/2/2021 15:13	5471	398
4	8/28/2020	Aichi	Japan	4/2/2021 15:13	4417	61
...
562	8/28/2020	Zaporizhia Oblast	Ukraine	4/2/2021 15:13	1487	25
563	8/28/2020	Zeeland	Netherlands	4/2/2021 15:13	1036	72
564	8/28/2020	Zhejiang	Mainland China	4/2/2021 15:13	1277	1
565	8/28/2020	Zhytomyr Oblast	Ukraine	4/2/2021 15:13	3063	61
566	8/28/2020	Zuid-Holland	Netherlands	4/2/2021 15:13	18636	1343

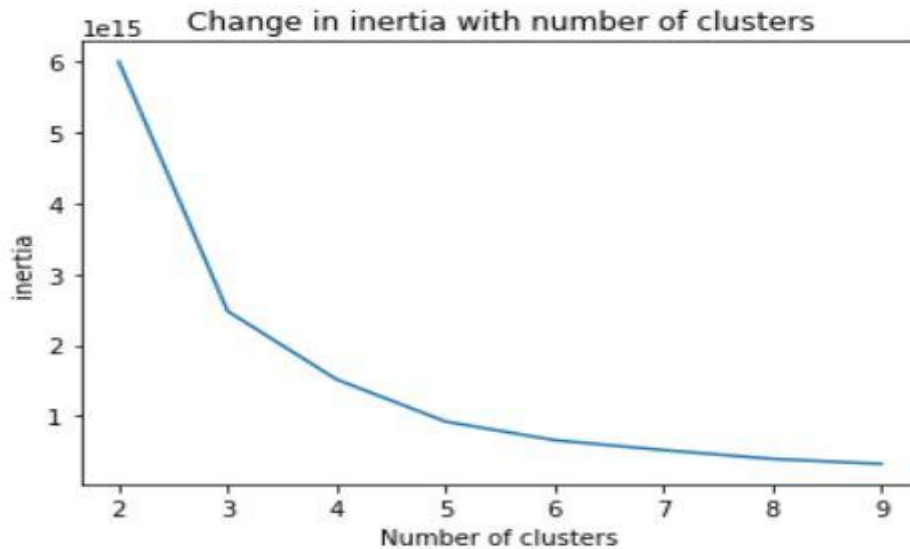
Dataset

	Confirmed	Deaths
count	567.000000	567.000000
mean	32719.188713	1153.680776
std	84699.229026	3303.856596
min	0.000000	0.000000
25%	1131.000000	20.000000
50%	5845.000000	132.000000
75%	19597.500000	776.500000
max	796209.000000	36837.000000

Description of Dataset

K-Means

K means Clustering was implemented and results are summarized as follows. Clustering was performed with 4 Clusters. Elbow Graph shows variation of inertia with Number of Clusters.



	Confirmed	Deaths
Cluster label		
0	267365.526316	10293.157895
1	7867.527083	347.722917
2	696995.000000	18028.400000
3	98578.523810	3198.698413

Mean of data per Cluster

Analysis of Clusters

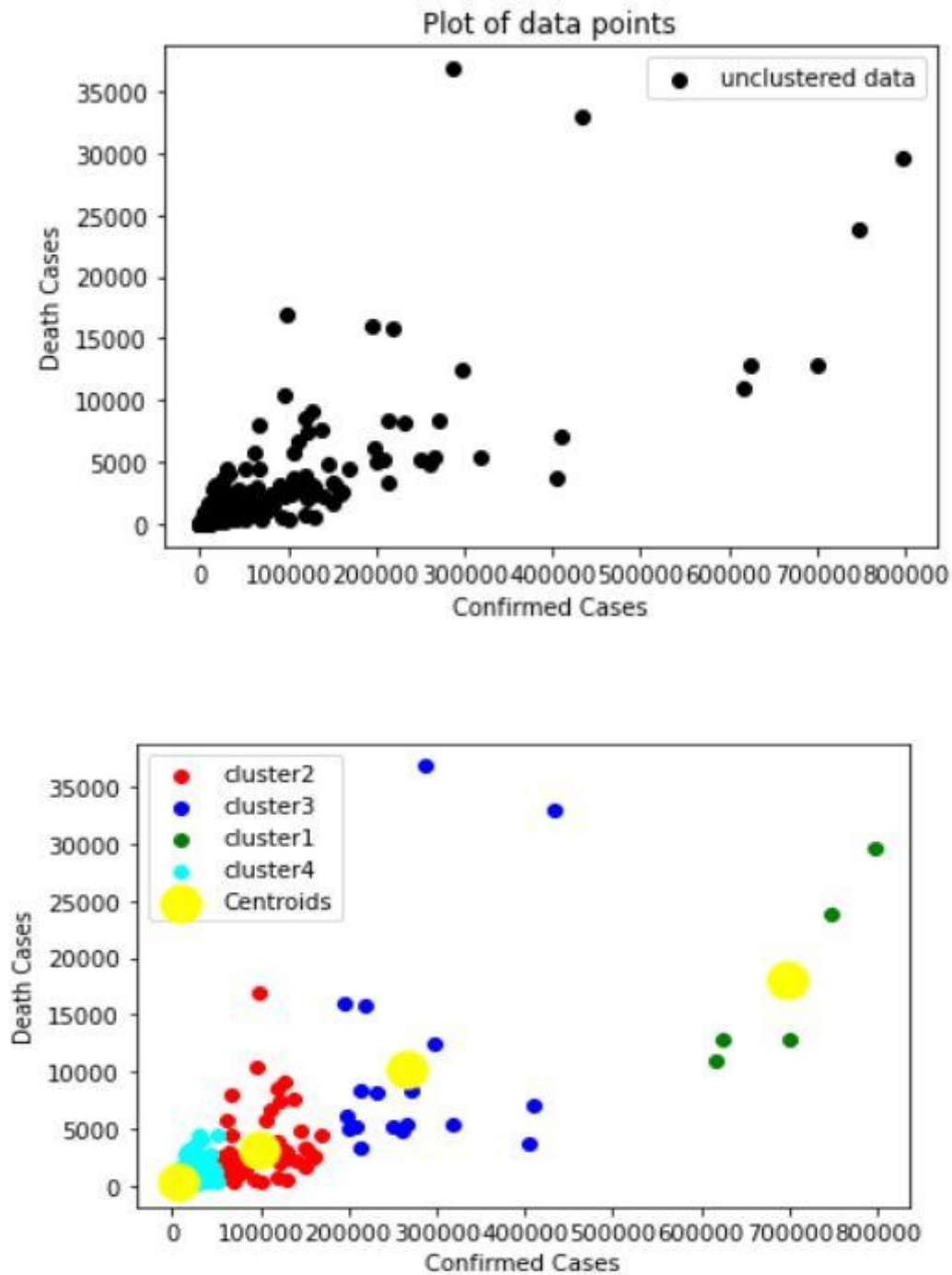
Cluster 1 includes the states with the second highest rate of Confirmed Cases and Death cases.

Cluster 2 includes the states with the highest rate of Confirmed Cases and Death cases.

Cluster 3 includes the states with the medium rate of Confirmed Cases and Death cases.

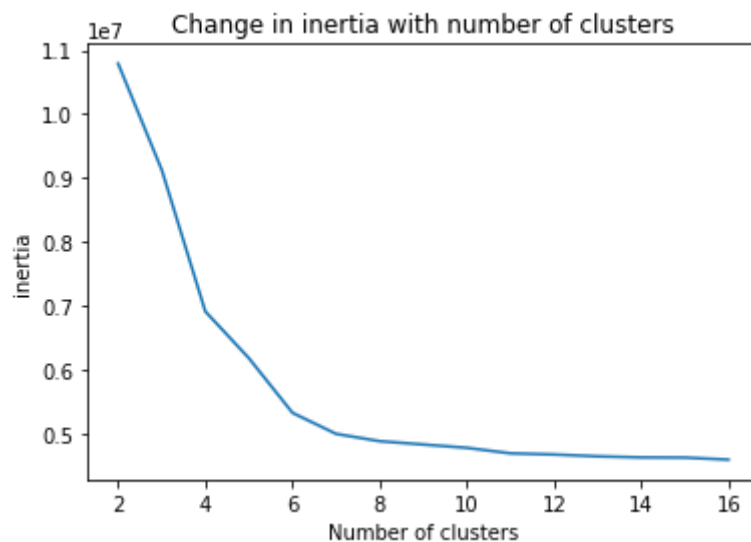
Cluster 4 includes the states with the lowest rate of Confirmed Cases and Death cases.

The Results before and after implementing K Mean clustering is Shown below in the scattered plots. Data points and Centroids are shown in the plots.



K-Medoids

K medoids Clustering was implemented and results are summarized as follows. Clustering was performed with 4 Clusters. Elbow Graph shows variation of inertia with Number of Clusters.



	Confirmed	Deaths
Cluster label		
0	267365.526316	10293.157895
1	7867.527083	347.722917
2	696995.000000	18028.400000
3	98578.523810	3198.698413

Analysis of Clusters

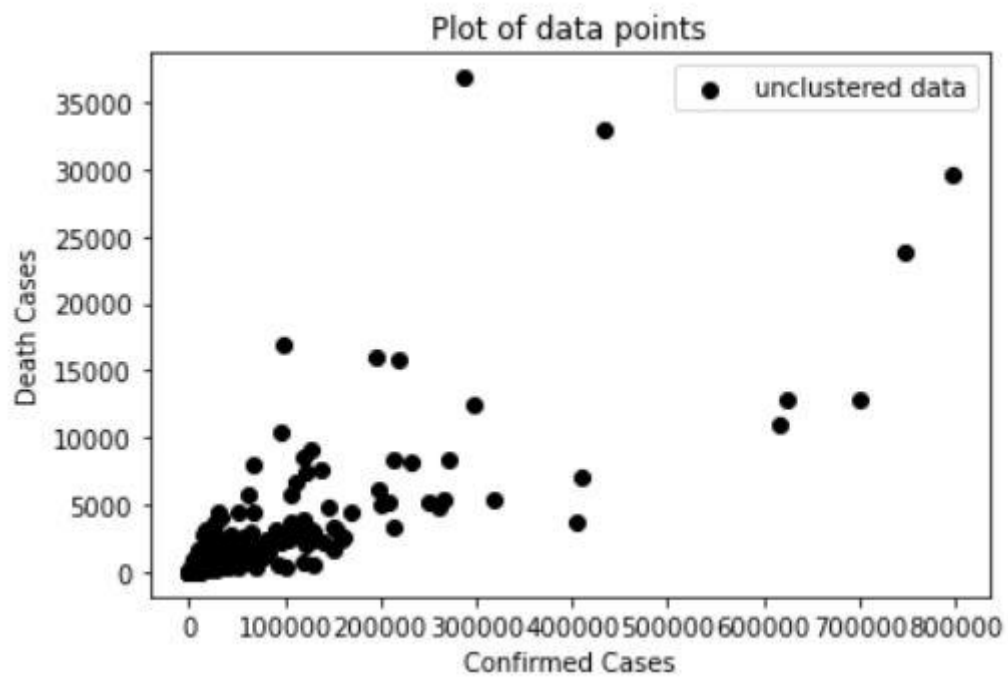
Cluster 0 includes the states with the second highest rate of Confirmed Cases and Death cases.

Cluster 1 includes the states with the lowest rate of Confirmed Cases and Death cases.

Cluster 2 includes the states with the highest rate of Confirmed Cases and Death cases.

Cluster 3 includes the states with the medium rate of Confirmed Cases and Death cases.

The Results before and after implementing K Medoids clustering is Shown below in the scattered plots. Data points and Centroids are shown in the plots.



Centers found by your model:

```
[[ 2229 65]  
 [ 18688 851]  
 [272039 8462]  
 [ 96636 2195]]
```

