

District wise clustering of covid-19 patients in Maharashtra

Miss.Komal Sanjay Phadtare¹, Miss. Amruta kisan Nimbalkar²

Abstract:

According to the official website of Public Health Department, Maharashtra, 34 out of the total 36 districts in the state have reported cases of COVID-19. In this paper we try make the groups of different districts according to similarity of the number of covid-19 patients, death and recovery patients by using cluster Analysis. The objective of this study is to stop the chain of transmission thus reducing the morbidity and mortality due to COVID-19. This also help to the cluster containment strategy would be to contain the disease within a defined geographic area by early detection of cases, breaking the chain of transmission and thus preventing its spread to new areas.

Keywords: Covid-19 patients, Cluster analysis, Distance matrix, Hierarchical clustering, Dendogram.

Introduction:

The number of coronavirus cases increasing at an alarming rate in Maharashtra. As per the latest data available at the Ministry of Health and Family Welfare, The state is nearing the 14,541 mark as the number of confirmed COVID-19 cases in the state has risen on May 5. Among the total number of infected, 2,411 have recovered while 583 patients succumbed to the disease. According to the official website of Public Health Department, Maharashtra, 34 out of the total 36 districts in the state have reported cases of COVID-19. In this paper we try make the

groups of different districts according to similarity of the number of covid-19 patients, death and recovery patients by using cluster Analysis. This study helps to stop the chain of transmission thus reducing the morbidity and mortality due to COVID-19. The data for analysis is taken from the press note of Public Health Department dated -05/05/2020.

Methodology:

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). There are different types of clustering method but in this paper we use Hierarchical clustering method to classify the districts/corporation according to the covid-19 patients, total death and recovery patients in Maharashtra. The Hierarchical clustering method is an alternative approach to partitional clustering for grouping objects based on their similarity. Hierarchical clustering can be performed with either a distance matrix or raw data.

Hierarchical Clustering Method:

Hierarchical clustering starts by treating each observation as a separate cluster. Then, it repeatedly executes the following two steps:

(1) identify the two clusters that are closest together, and

(2) merge the two most similar clusters.

This iterative process continues until all the clusters are merged together.

The main output of Hierarchical Clustering is a dendrogram, which shows the hierarchical

relationship between the clusters. The distance between two clusters has been computed based on the length of the straight line drawn from one cluster to another. This is commonly referred to as the Euclidean distance. For the cluster analysis we use R software. The following data we have used for cluster analysis:

Observation Table 1 : District/Corporation wise Distribution of COVID-19 cases in Maharashtra

District /corporation	Total cases	Total deaths	Recovered
Mumbai MNC	9310	361	1424
Thane	64	2	27
Thane MNCs	514	8	54
Navi Mumbai MNC	254	4	34
Kalyan-Dombivali MNC	228	3	30
Ulhasnagar MNC	4	0	0
Bhivandi-Nizampur MNC	22	2	0
Mira-Bhayandar MNC	152	2	80
Palghar	46	1	15
Vasai-Virar MNC	158	4	57
Raigad	41	1	23
Panvel MNC	64	2	0
Nashik	21	0	6
Nashik MNC	31	0	2
Malegaon MNC	330	12	3
Ahmednagar	35	2	24
Ahmednagar MNC	7	0	0
Dhule	8	2	0
Dhule MNC	24	1	0
Jalgaon	46	11	1
Jalgaon MNC	11	1	0
Nandurbar	18	1	0
Pune	102	4	0
Pune MNC	1796	106	378
Pimpri-Chinchwad MNC	120	3	48
Solapur	3	1	0
Solapur MNC	126	6	22
Satara	79	2	12
Kolhapur	8	0	4
Kolhapur MNC	6	0	0

Sangli	32	0	26
Sangli-Miraj-Kupwad MNC	2	1	0
Sindhudurg	2	1	1
Ratnagiri	10	1	5
Aurangabad	3	0	0
Aurangabad MNC	310	10	23
Jalna	8	0	1
Hingoli	52	0	1
Parbani	1	1	0
Parbani MNC	1	0	1
Latur	19	1	8
Osmanabad	3	0	3
Beed	1	0	0
Nanded	3	0	0
Nanded MNC	28	2	0
Akola	7	1	0
Akola MNC	48	5	13
Amaravati	1	1	1
Amaravati MNC	57	9	0
Yavatmal	91	0	10
Buldhana	24	1	20
Washim	1	0	1
Nagpur	2	0	0
Nagpur MNC	172	2	52
Bhandara	1	0	0
Gondia	1	0	1
Chandrapur	1	0	0
Chandrapur MNC	3	0	0
Other States	29	5	0
Total	14541	583	2411

Statistical Analysis :

We have used R-software to obtain this cluster analysis. The R-program which we have used is given below:

```
#Program
```

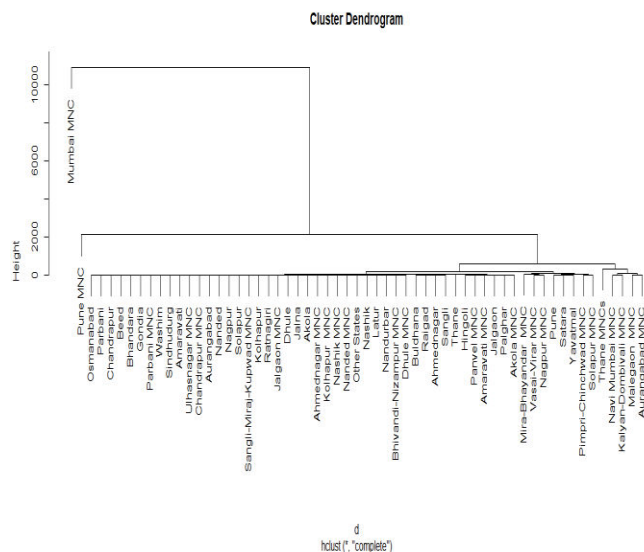
```
data=read.delim(file.choose(),header=T)
```

```
data;
```

```
# Give row names and column names
cnames <- c("Districts","Totalcase","Totaldeaths",
,"Recovered")
colnames(data)<- cnames
```

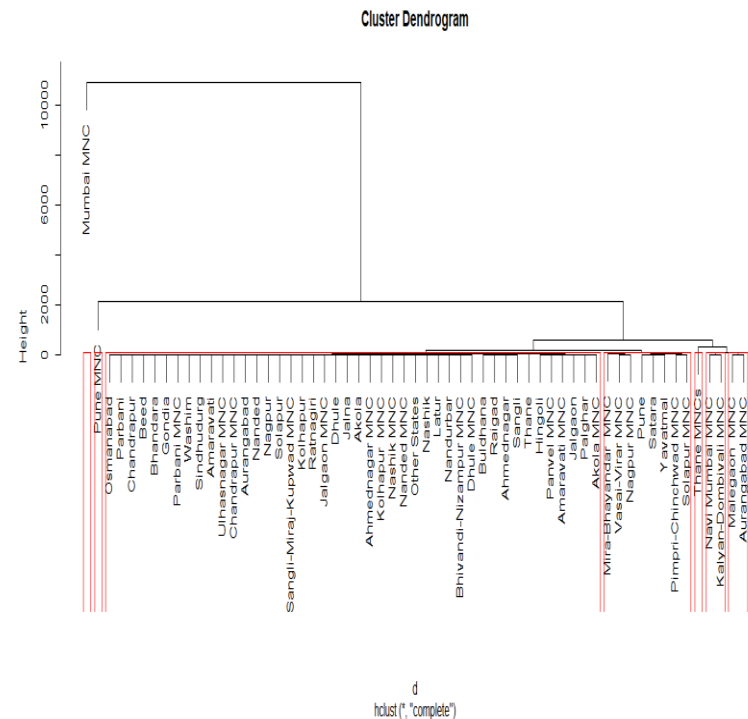
```
rownames(data)<-data$Districts
data;
# Hierarchical Clustering
# Create distance matrix
d <- dist(data, method = "euclidean")
fit <- hclust(d)
# display dendrogram
plot(fit,label=data$Districts)
groups <- cutree(fit, k=7) # cut tree into 7
clusters
# draw dendrogram with red borders around the 7
clusters
rect.hclust(fit, k=7, border="red")
```

Output:



Graph 1: It shows the different clusters of districts / corporations according to similarities.

Firstly, we got this dendrogram in output which gives number of clusters by default. So, we have merged these clusters into 7 clusters. Then we get following output:



Graph 2: It represents merged 7 clusters of District / Corporation

From the above graph 2, we observed that the seven cluster by using Hierarchical clustering method are as follows :

Cluster-1: Mumbai MNC

Cluster-2: Pune MNC

Cluster-3:Thane MNC

Cluster-4:Navi Mumbai MNC and Kalyan Dombivali MNC

Cluster-5: Malegaon MNC and Aurangabad MNC

Cluster-6: Mira-Bhayandar MNC, Vasai-Virar MNC, Nagpur MNC,Pune,Satara,Yavatmal,

Pimpri-chinchwad MNC, Solapur MNC.

Cluster-7: Osmanabad, Parbani, Chandrapur, Beed, Bhandara, Gondia, Parbani MNC, Washim, Sindhudurg, Amaravati, Ulhasnagar MNC, Chandrapur MNC, Aurangabad, Nanded, Nagpur, Solapur, Sangli-Miraj-Kupwad MNC, Kolhapur, Ratnagiri, Jalgaon MNC, Dhule, Jalna, Akola, Ahmednagar MNC, Kolhapur MNC, Nashik MNC, Nanded MNC, Other States, Latur, Nandurbar, Bhivandi-Nizampur MNC, Dhule MNC, Buldhana, Raigad, Ahmednagar, Sangli, Thane, Hingoli, Panvel MNC, Amaravati MNC, Jalgaon, Palghar, Akola MNC.

From graph 2, In the cluster -1 there is only one location that is Mumbai Municipal Corporation (MNC), with the highest distance from other corporation. We can conclude that the Mumbai MNC has highest covid-19 patients, Total number of death as well as Total number of recovery patients. Further that in cluster -2 also only one location Pune Municipal Corporation (MNC), which show high covid-19 patients, Total death and recovery patients. Locations in cluster-3, cluster-4, cluster-5 and cluster-6 are also having more covid-19 patients, deaths and recovery patients. But locations in cluster -7 are having low covid-19 patients, death and recovery patients as compare to other clusters corporations.

Conclusion:

From the above graph 2, we observed that district/corporation Mumbai Municipal Corporation, Pune Municipal Corporation, Thane Municipal Corporation, Navi Mumbai MNC, Kalyan Dombivali MNC, Malegaon MNC, Aurangabad MNC, Mira-Bhayandar MNC, Vasai-Virar MNC, Nagpur MNC, Pune, Satara, Yavatmal, Pimpri-chinchwad MNC, Solapur MNC .having most Cases of covid-19 patients, death and recovery patients .The districts/ corporations in cluster-1 to cluster-6 has become more vulnerable as far as further transmission of the COVID-19 virus is concerned. It is necessary to break further transmission of the disease in the area and it becomes imperative to take stringent precautionary measures like declaring some villages as containment/red zone and surrounding areas as a buffer zone.

References:

1. <https://www.indiatoday.in/india/story/coronavirus-cases-in-india-rise-to-275-here-s-a-state-wise-breakdown-1658080-2020-03-21>
2. <https://statisticsbyjim.com/regression/interpret-r-squared-regression/>
3. <https://m.economictimes.com/news/politics-and-nation/number-of-covid-19-patients-in-maharashtra-rises-to-135/articleshow/74841619.cms>
4. https://en.wikipedia.org/wiki/Nonlinear_regression
5. <https://www.worldometers.info/coronavirus/country/india/>
6. <https://www.arogya.maharashtra.gov.in>