CS685: Data Mining

Assignment 1

Shilpa Chatterjee
Roll No. :- 20111057

1 Analysis of Covid19 Data Districtwise

For the analysis I have used raw data files (1-14) obtained from `https://api.covid19india.org/documentation/csv/` and the time frame that has been considered is 15th March,2020 - 5th September,2020.

1.1 Top 5 Covid affected districts in India

From the case count obtained from the Covid19 website the top five most affected districts of India are **Pune, Delhi, Mumbai, Thane, Bengaluru Urban**.

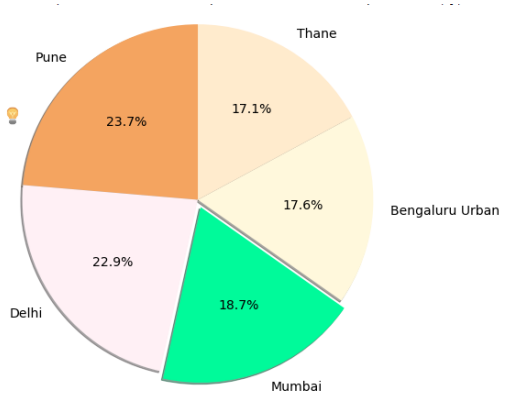


Figure 1: Top 5 most covid affected districts in India

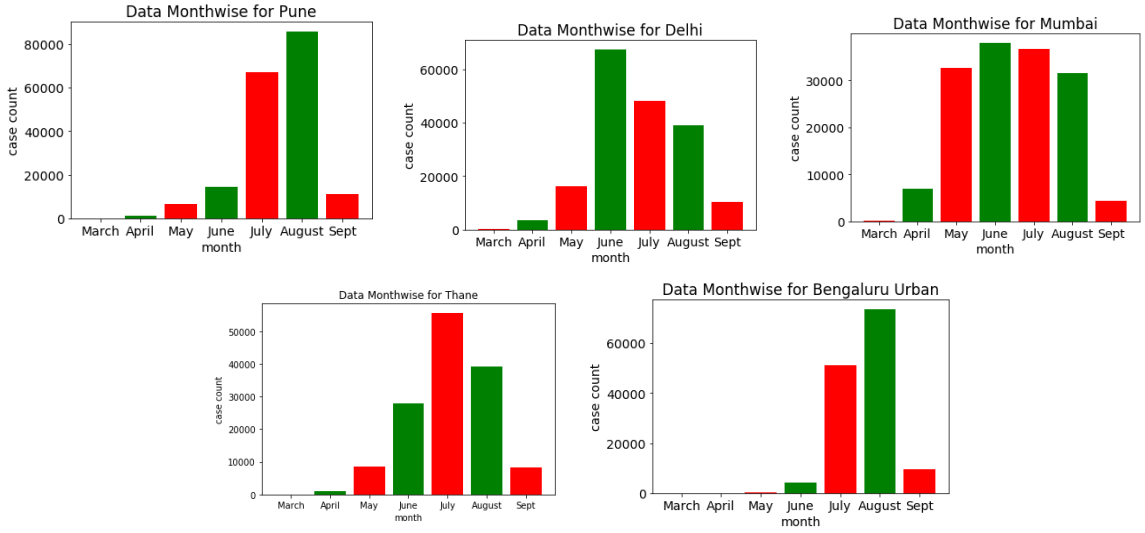


Figure 2: Monthwise Covid19 confirmed cases of top 5 affected districts in India

1.2 Top 5 districts of India which have high standard deviation from its neighbors

Standard Deviation tells us how varied the distribution of data is from the mean. Large Standard Deviation signifies high variation in data.
The top five districts of India with high devation of cases from its neighbors are **Raigad,Faridabad, Ghaziabad, Satara, Sonipat**.

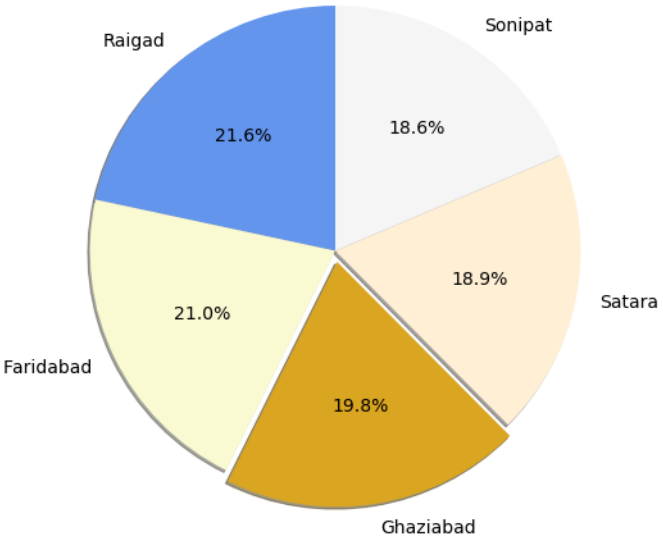


Figure 3: Top 5 districts in India with high Standard Deviation from its neighbors

1.3 Top 5 districts of India which have high standard deviation from the other districts in the state

The top five districts of India with high devation of cases from the other districts of the same state are **Ahmednagar, Kolhapur, Aurangabad(Maharastra), Solapur, Palghar**.

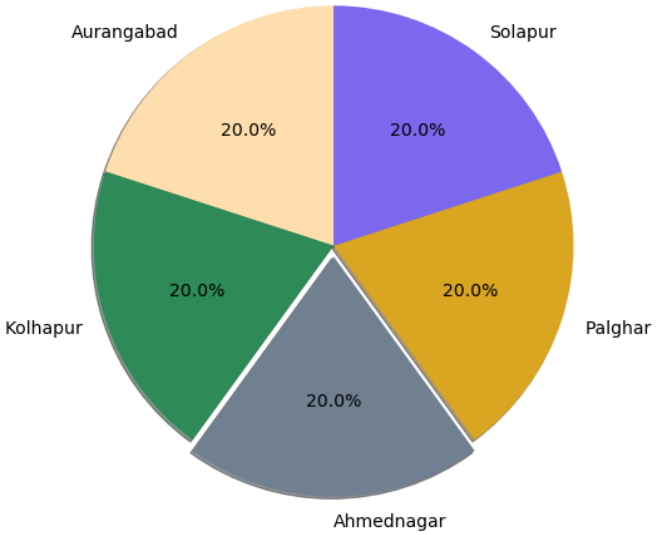


Figure 4: Top 5 districts in India with high Standard Deviation from districts in the state

1.4 Top 5 districts of India which have more cases than mean of other neighbouring districts (z-score)

Zscore signifies how far the raw data is away from the mean(average). Here higher zscore means that a district has more number of cases than the average number of cases of its neighboring districts.
The top five districts of India with higher covid19 case count from the mean covid confirmed case count of other neighbouring districts are **Assam, Bengaluru Urban, Bhopal, Delhi, Lucknow**.

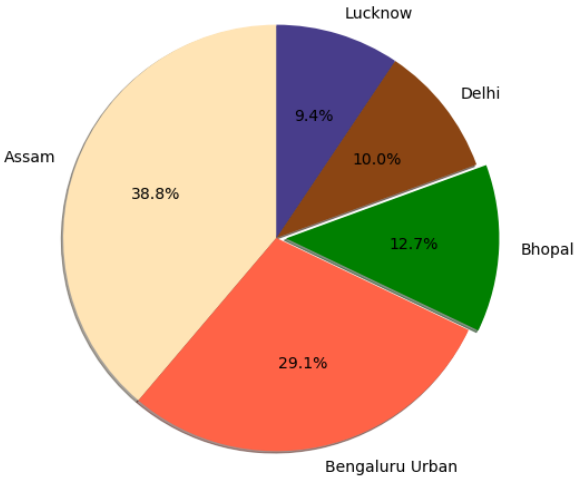


Figure 5: Top 5 districts in India with higher covid19 cases than mean of other neighboring districts (hotspots)

1.5 Top 5 districts of India which have more cases than mean of other districts of the state (z-score)

Higher zscore here signifies that a district has more number of cases than the average number of cases of the other districts in same state.
The top five districts of India with higher covid19 case count from the mean covid confirmed case count of other districts in the state are **Puducherry, Bengaluru Urban, Chennai, Raipur, Patna**.

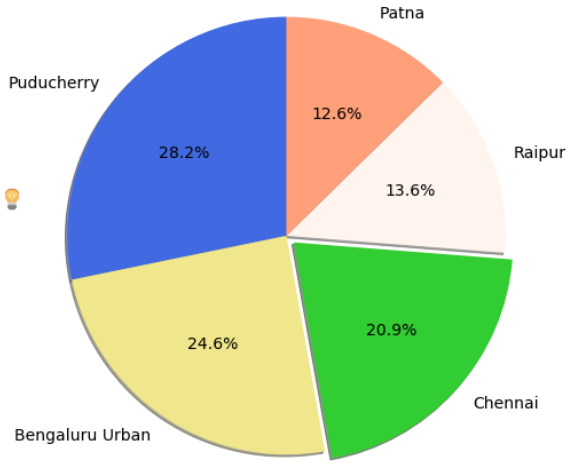


Figure 6: Top 5 districts in India with higher covid19 cases than mean of other districts in the same state (hotspots)

2 Analysis of Covid19 Data of India

2.1 Confirmed Case Analysis of India

For this analysis I have used confirmed covid19 case count of the time period 15th March,2020 - 5th September,2020.

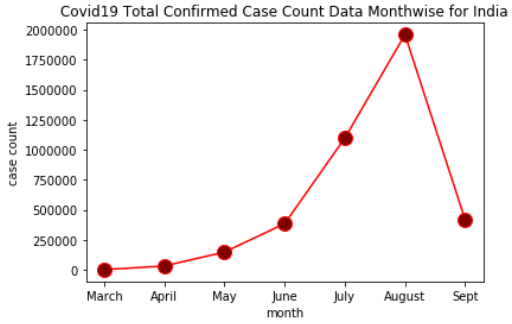


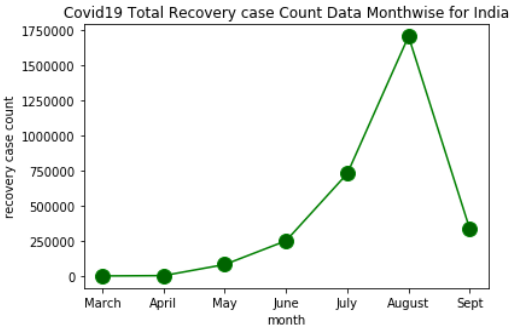
Figure 7: Confirmed Covid19 cases monthwise of India

The steep downfall of the number of covid19 cases in the month of September,2020 is because I have considered only 5 days in the month of September,2020.

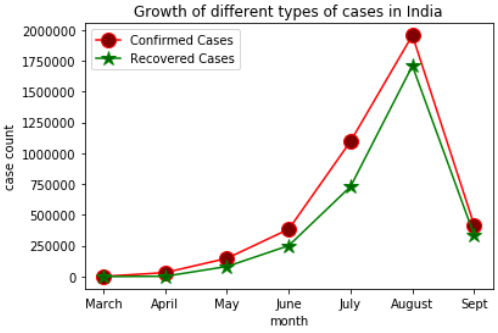
From the graph it can be observed that the number of covid19 cases has been constantly increasing since the month of March,2020 and has reached its peak in the month of August,2020 (in the time frame that I have considered for the analysis) .

2.2 Recovered Case Analysis of India

For this analysis I have used recovered covid19 case count of the time period 15th March,2020 - 5th September,2020.



(a) Recovered Covid19 cases monthwise of India



(b) Recovery Rate vs Confirmed Cases Rate Monthwise

Figure 8: Growth of different type of cases monthwise in India

The steep downfall of the number of recovered covid19 cases in the month of September,2020 is because I have considered only 5 days in the month of September,2020.

From the graph(Figure 8(a)) it can be observed that the recovery rate graph has a similar curve as that of the confirmed case graph which can also be viewed in Figure 8(b).