



CS685: DATA MINING REPORT ON ASSIGNMENT 1

Abhas Kumar-20111001

M.Tech CSE

23 September 2020

Abstract

This document reports the significant results obtained while working with COVID-19 data for the duration between 15 March 2020 and 5 September 2020, presents analysis of these results and finally draws few conclusions based on the analysis.

RESULTS AND ANALYSIS

1. During 15 March to 5 September 2020, 10 most affected districts of our country just on the basis of total confirmed cases are **Pune, Delhi, Bangaluru Urban, Chennai, Godavari, Kolkata, Kamrup, Ahmedabad, Lucknow, Patna** in decreasing order.

And 10 least affected districts are **Aizawal, Dadra and Nagar Haveli, East Sikkim, Imphal, Solan, Papum Pare, Leh, East Khasi Hills, Dimapur, South Goa** in increasing order.

State/union territory	District	Cases	Tests(L)	Positive Ratio
West Bengal	Kolkata	42543	1.41	0.30
Assam	Kamrup Metropolitan	32472	1.24	0.26
Maharashtra	Pune	194559	8.50	0.23
Andhra Pradesh	East Godavari	65704	3.90	0.17
Karnataka	Bengaluru Urban	144757	10.10	0.14
Tamil Nadu	Chennai	140685	10.30	0.14
Bihar	Patna	22318	1.70	0.13
Delhi	Delhi	188193	17.50	0.11
Uttar Pradesh	Lucknow	31502	3.70	0.09
Gujarat	Ahmedabad	32351	8.60	0.04

Figure 1: Showing Districts with test positive ratio in decreasing order

- Whereas top 10 districts on the basis of test positive ratio ratio(confirmed cases/total tested) are **Kolkata, Kamrup, Pune, East Godavari, Bengaluru Urban, Chennai, Patna, Lucknow, Ahmedabad** in decreasing order (Figure 1).

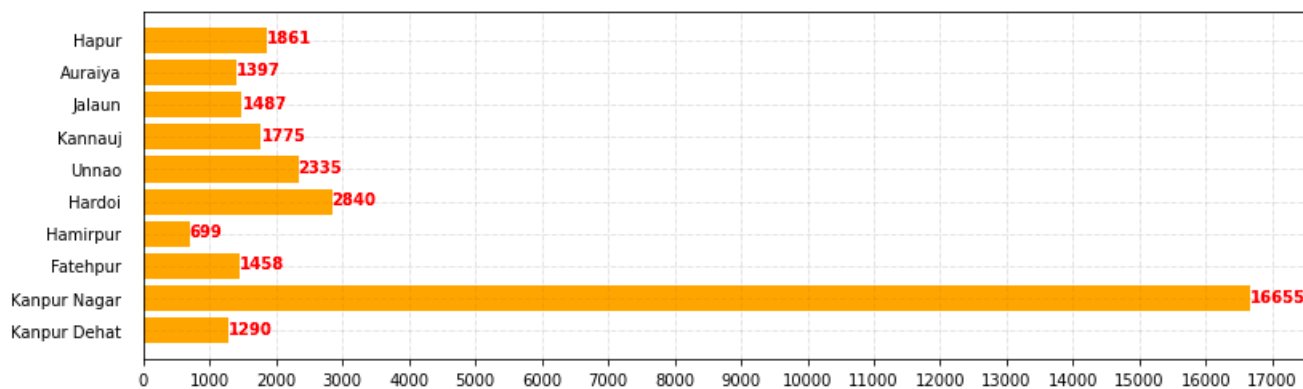


Figure 2: Kanpur with its Neighbor districts

- Kanpur nagar alone had **16655** covid cases which was more than than all its neighbor district's cases combined, that was **15142**(Figure2).

Region	Number of States/UT	Total Cases	Average Cases
South India	5	1332750	266550
North India	9	660664	73407
West India	5	1083601	216720
East India	5	484832	96966
Central India	2	113549	56774
Northeast India	7	119562	17080

Table 1: Cases in Major 6 regions of our Country

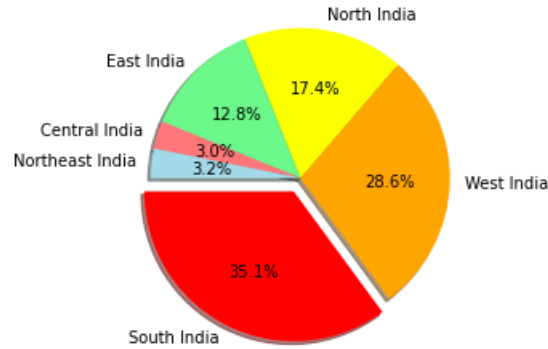


Figure 3: Portion of total cases each of 6 regions had

CONCLUSIONS

1. Ratio of Positives to tested is a better parameter than absolute number of confirmed cases for any region. Although **Pune** had most confirmed cases but it had only **23 positive cases out of 100** tested, whereas **Kolkata** had **30 positives per 100**.
2. Southern and Western part of India are most affected regions. Southern and Western part with total of 10 states had **63.7%** of total cases in India while rest 4 regions with 23 states had only **26.3%** cases.
3. Districts having metropolitan cities like **Delhi, Mumbai, Puna, Chennai, Bangaluru, Kolkata** are most affected regions which is in accordance with the fact that they have huge population and also large number of testing have been done in these major cities.