EY Techathon 2021:
#iSolve4aBillionChallenge

Problem statement: Prioritize vaccine delivery using AI/ML

We are Team **Huskarls** from Birla Institute of Technology and Science, Pilani

Team Members:

1. Ayushi Gaur (Email: p20190023@pilani.bits-pilani.ac.in)
2. Ayush Kumar (Email: f20170761@pilani.bits-pilani.ac.in)

Understanding the Problem Statement

Problems faced in Rollout of Vaccine

With limited supply of COVID-19 vaccines available at first and large demographic to cater to, a proper rollout plan is needed to break the infection chain in an efficient manner. This includes:

- Analysis of number of cold storage facility at state and district level.
- Analysis of district wise population to identify people with age more than 50 years and/or with pre-existing condition such as Asthma, Diabetes etc. and density of population in various region.
- Focusing more on the area with high transmission rate at present to contain it.
- Identifying the frontline workers and the type of employments unable to physically distance.
- The country was in lockdown for a long period hence we need to identify the population associated with education and economically critical sector.

Since there are multiple criteria to decide the priority, a proper metric should be devised to do the ranking.

With proper analysis of data and using appropriate algorithms we can dynamically classify the rollout plan of vaccine for different districts depending on the availability of the vaccine and the dosage required in each area.

Solution

Designing a Model to provide insights on vaccine delivery

The two main challenges faced from the solution point of view are:

Creation of an Absolute Dataset

Accurate Algorithm for Analysis

Dataset

Cold storage facilities available in every state

Population aged > 50 years

Students missing school due to lockdown

Health workers, frontline rescue team

People with Comorbidities

Current COVID Cases

Our compiled Dataset extensively covers all major aspects of the problem.

District-wise population demographic

The algorithm That would help us do the task:

- Compilation and Cleaning of Data
- Data Analysis: Finding out the relation between different attributes which would help us in finding a suitable metric
- Extracting the main features using PCA
- Clustering using DBSCAN and KNN: Using the elbow curve we found out the number of levels districts can be divided for proper analysis.

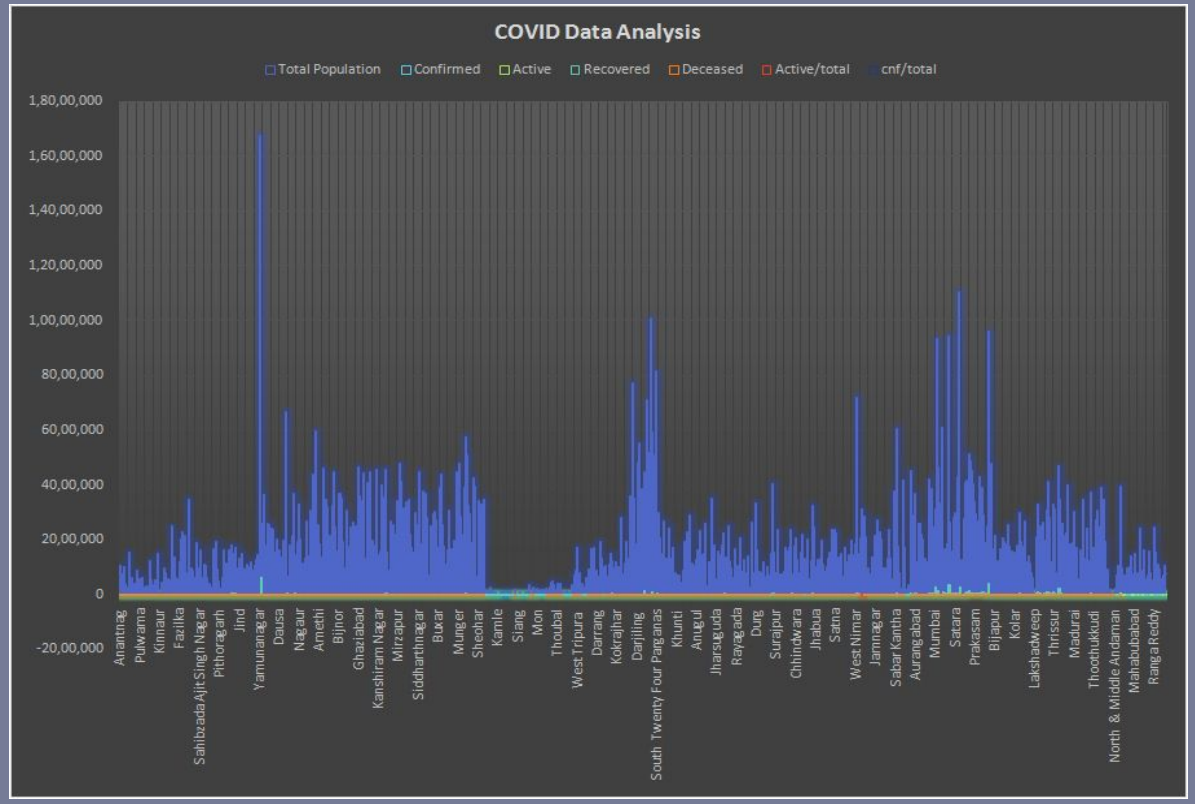
Ranking Algorithm

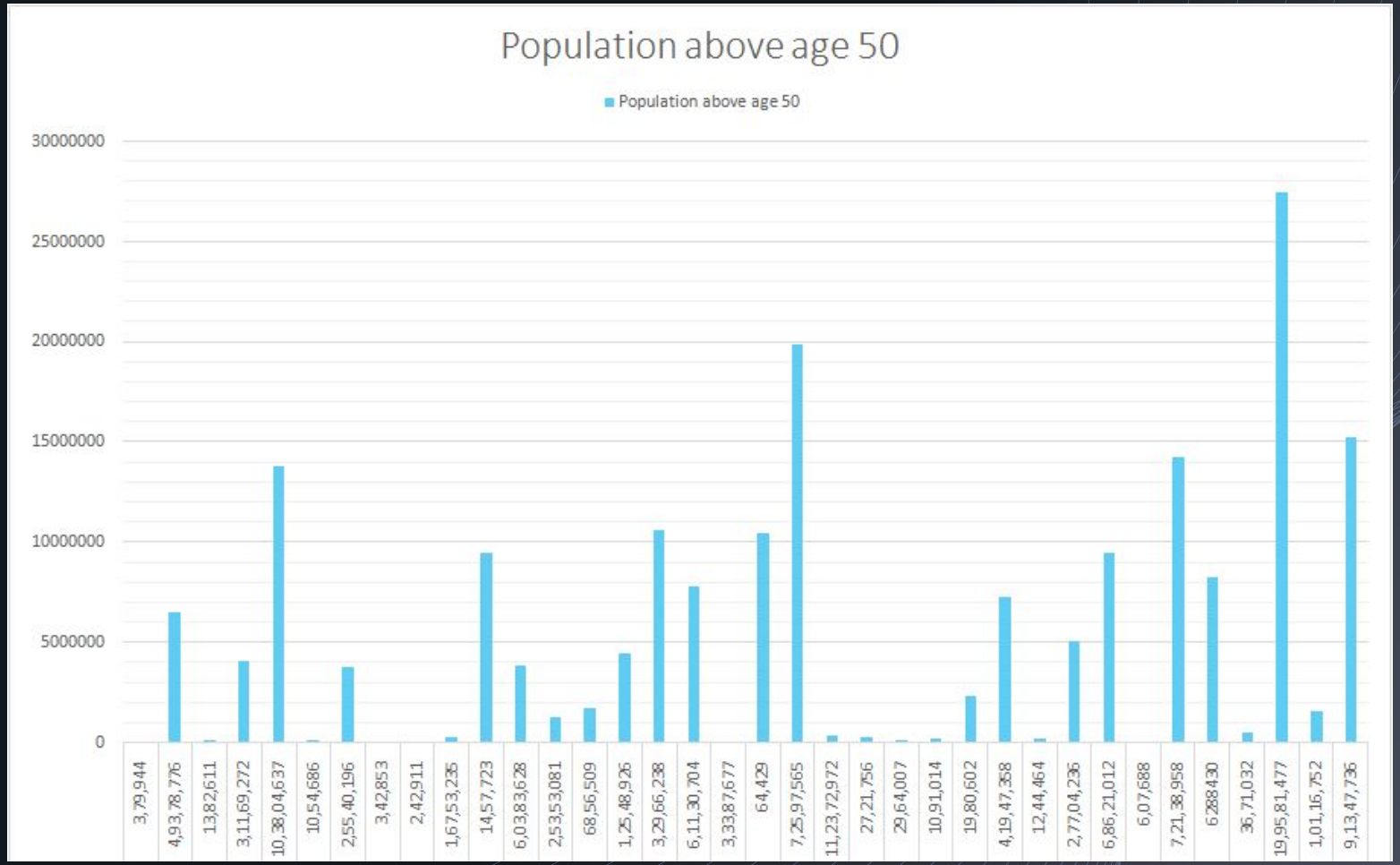
- Using **SEIR** (**S**usceptible → **E**xposed → **I**nfected → **R**emoved) analysis, we figure out the population susceptible for rapid spread which can be prioritised. (We consider the population which hasn't contracted the disease yet as susceptible and combining it with rate of active cases form a priority)
- We also factor comorbidities, population above 50 and ease of access to cold storage for the ranking.

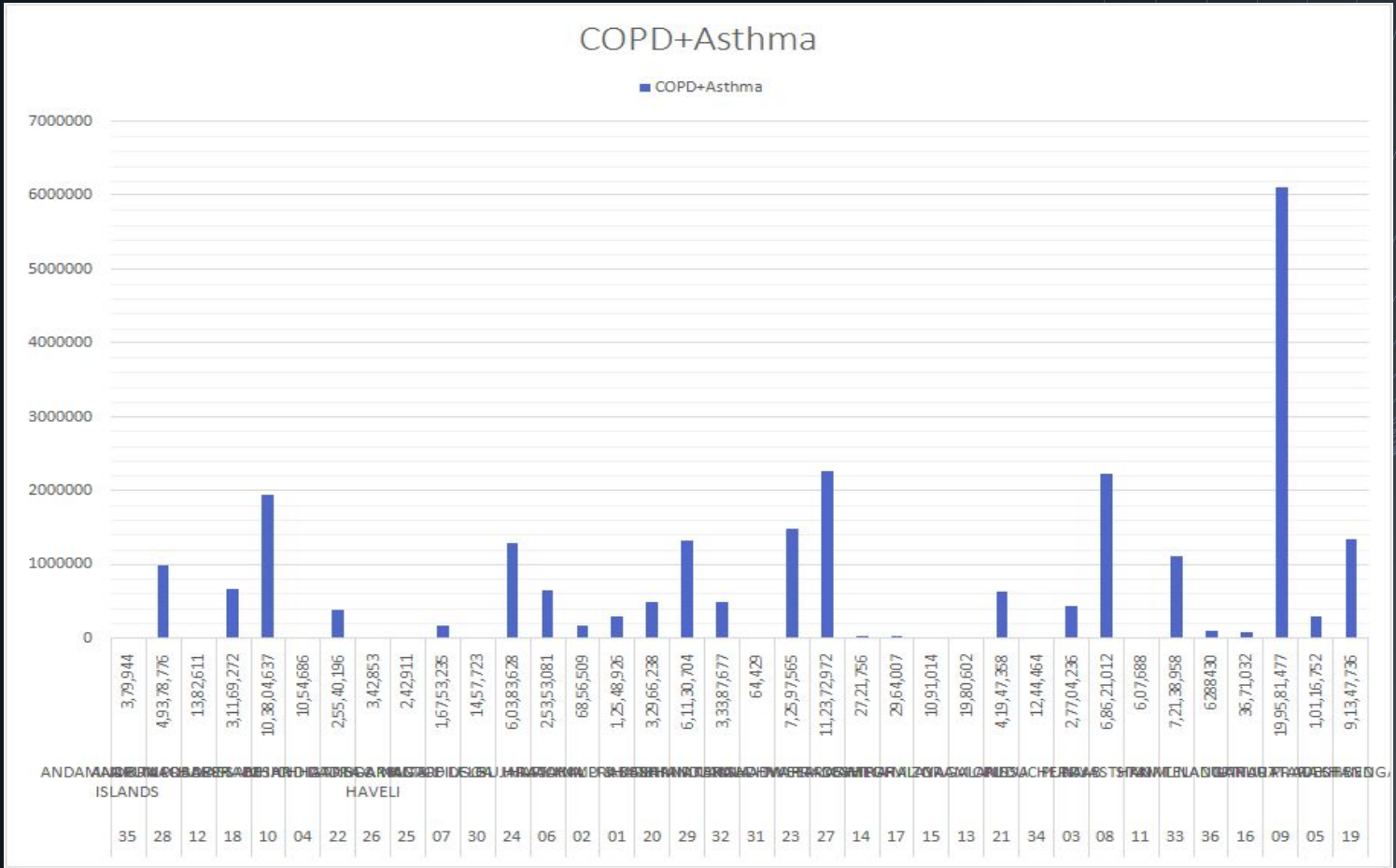
Methodology

- The whole analysis has been done in Python and the clustering algorithm has been implemented using scikit-learn library.
- Used Seaborn and matplotlib for plotting the data and for further analysis.
- We created Dashboard in Excel(using pivot table) to analyse data.

Output


COVID Data Analysis
Total Population
Confirmed
Active
Recovered
Deceased
Active/total
cnf/total
1,80,00,000
1,60,00,000
1,40,00,000
1,20,00,000
1,00,00,000
80,00,000
60,00,000
40,00,000
20,00,000
0
-20,00,000
Anantnag
Pulwama
Kinnaur
Fazilka
Sahibzada Ajit Singh Nagar
Pithoragarh
Jind
Yamunanagar
Dausa
Nagaur
Amethi
Bijnor
Ghaziabad
Kanshiram Nagar
Mirzapur
Siddharthnagar
Buxar
Munger
Sheohar
Kamle
Siang
Mon
Thoubal
West Tripura
Darrang
Kokrajhar
Darjiling
South Twenty Four Parganas
Khunti
Anugul
Jharsuguda
Rayagada
Durg
Surajpur
Chhindwara
Jhabua
Satna
West Nimar
Jamnagar
Sabar Kantha
Aurangabad
Mumbai
Satara
Prakasam
Bijapur
Kolar
Lakshadweep
Thrissur
Madurai
Thoothukkudi
North & Middle Andaman
Mahabubabad
Ranga Reddy


Population above age 50
Population above age 50
30000000
25000000
20000000
15000000
10000000
5000000
0
3,79,944
4,93,78,776
13,82,611
3,11,69,272
10,38,04,637
10,54,686
2,55,40,196
3,42,853
2,42,911
1,67,53,235
14,57,723
6,03,83,628
2,53,53,081
68,56,509
1,25,48,926
3,29,66,238
6,11,30,704
3,33,87,677
64,429
7,25,97,565
11,23,72,972
27,21,756
29,64,007
10,91,014
19,80,602
4,19,47,358
12,44,464
2,77,04,236
6,86,21,012
6,07,688
7,21,38,958
6288430
36,71,032
19,95,81,477
1,01,16,752
9,13,47,736


COPD+Asthma
COPD+Asthma
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
3,79,944
4,93,78,776
13,82,611
3,11,69,272
10,38,04,637
10,54,686
2,55,40,196
3,42,853
2,42,911
1,67,53,235
14,57,723
6,03,83,628
2,53,53,081
68,56,509
1,25,48,926
3,29,66,238
6,11,30,704
3,33,87,677
64,429
7,25,97,565
11,23,72,972
27,21,756
29,64,007
10,91,014
19,80,602
4,19,47,368
12,44,464
2,77,04,236
6,86,21,012
6,07,688
7,21,38,958
6,28,430
36,71,032
19,95,81,477
1,01,16,752
9,13,47,736
ISLANDS
HAVELI
35
28
12
18
10
04
22
26
25
07
30
24
06
02
01
20
29
32
31
23
27
14
17
15
13
21
34
03
08
11
33
36
16
09
05
19

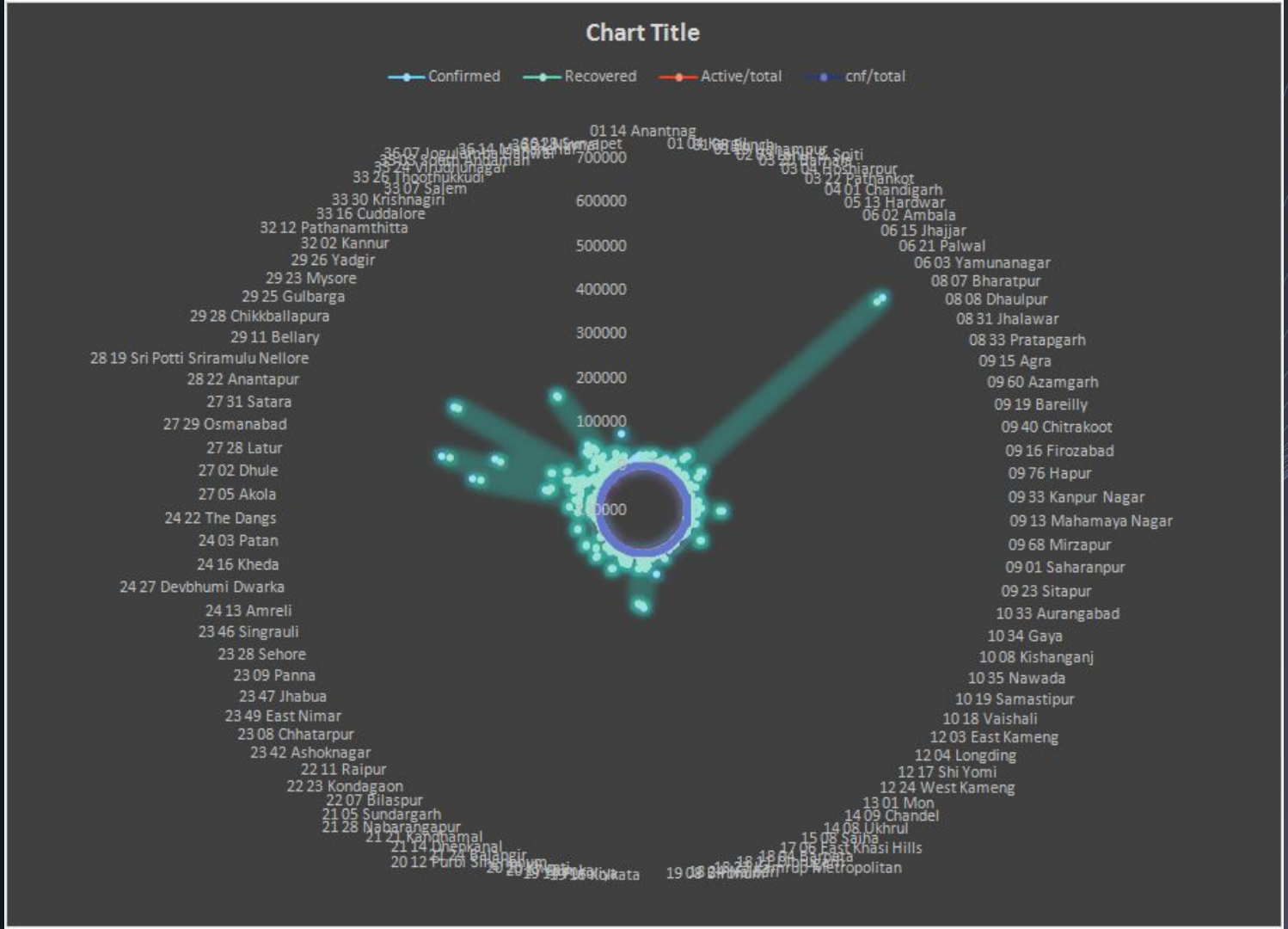


Chart Title
Confirmed
Recovered
Active/total
cnf/total
700000
600000
500000
400000
300000
200000
100000
01 14 Anantnag
03 22 Pathankot
04 01 Chandigarh
05 13 Hardwar
06 02 Ambala
06 15 Jhajjar
06 21 Palwal
06 03 Yamunanagar
08 07 Bharatpur
08 08 Dhaulpur
08 31 Jhalawar
08 33 Pratapgarh
09 15 Agra
09 60 Azamgarh
09 19 Bareilly
09 40 Chitrakoot
09 16 Firozabad
09 76 Hapur
09 33 Kanpur Nagar
09 13 Mahamaya Nagar
09 68 Mirzapur
09 01 Saharanpur
09 23 Sitapur
10 33 Aurangabad
10 34 Gaya
10 08 Kishanganj
10 35 Nawada
10 19 Samastipur
10 18 Vaishali
12 03 East Kameng
12 04 Longding
12 17 Shi Yomi
12 24 West Kameng
13 01 Mon
14 09 Chandel
14 08 Ukhrul
15 08 Saiha
17 06 East Khasi Hills
Metropolitan
Kolkata
20 12 Puri
21 14 Dhenkanal
21 21 Kandhamal
21 28 Nabarangapur
22 07 Bilaspur
22 23 Kondagaon
22 11 Raipur
23 42 Ashoknagar
23 08 Chhatarpur
23 49 East Nimar
23 47 Jhabua
23 09 Panna
23 28 Sehore
23 46 Singrauli
24 13 Amreli
24 27 Devbhumi Dwarka
24 16 Kheda
24 03 Patan
24 22 The Dangs
27 05 Akola
27 02 Dhule
27 28 Latur
27 29 Osmanabad
27 31 Satara
28 22 Anantapur
28 19 Sri Potti Sriramulu Nellore
29 11 Bellary
29 28 Chikkballapura
29 25 Gulbarga
29 23 Mysore
29 26 Yadgir
32 02 Kannur
32 12 Pathanamthitta
33 16 Cuddalore
33 30 Krishnagiri
33 07 Salem
33 26 Thoothukkudi

Appendix

GITHUB Repo: https://github.com/quasayush/EY-Techathon.git

- Click Here to go to our GITHUB Repository
- Interactive dashboard is available in GITHUB Repo.
- Population demographic data taken from https://censusindia.gov.in/2011census/
- Cold Storage data taken from https://www.indiastat.com/table/agriculture-data/2/cold-storages/32430/1230615/data.aspx
- School Students Data taken from https://data.world/inderz/india-district-level-school-report-card
- Pulmonary diseases data taken from here
- SEIR related paper:

Mwalili, S., Kimathi, M., Ojiambo, V. *et al.* SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Res Notes* 13, 352 (2020).