

Deliverable 1

1) Team

a) Members

Aleena Jimmy George (801254680)
Anupama Ramesh (801261344)
Gayathri Jayagopal (801259246)
Santhan Pal Bandari (801259376)
Sneha Chillumula (801161898)

b) Communication plan

Git Repo : <https://github.com/j-gayathri/BigDataProject>

Meeting notes :  GroupMeetingNotes

Meeting platform : Zoom

2) Selection of domain and data

Link : [COVID-19 Data Lake](#)

The above is a dataset with up-to-date curated dataset on the spread of COVID-19

3) Business Problem or Opportunity, Domain Knowledge:

COVID-19 has been a concern for over a year now. The data collected regarding the COVID helps in a lot of research. The dataset includes a lot of information such as the number of covid positive cases, the count of positive cases identified every day, the death toll, and a lot more. Firstly, the information is required to learn about the growth of the cases. Lockdowns were imposed on several parts of the world and a lot of restrictions were put on social meetings or day-to-day activities. The effectiveness of the restrictions imposed can be seen from the past data.

Furthermore, the rate of increase in positive cases, especially in the past, alerts us about the seriousness of the situation and the measures to be taken to curb the further spread, if any.

Finally, the data helps us in projecting the future positive count and be prepared in terms of medical services to be provided.

4) Research Objectives and Question(s):

Objectives that the research caters to:

- Generate better forecasts of hotspots and trends
- Dashboards could be built to track infections and collaborate to efficiently deploy vital resources like hospital beds and ventilators

Questions that answered:

- What are the county-level predictions for number of deaths
- What county-level predictions are allocated to hospitals within counties proportional the their total number of employees
- What is the threshold number of cumulative predicted deaths for a hospital