

2. DATASETS

2.1 DATA ASSEST

I. CONTEXT & LINKS

The first COVID-19 case was reported on 30 January in a student who arrived in Kerala state from Wuhan. Then 2 more cases were reported in the next 2 days in Kerala again. For almost a month, no new cases were reported in India, however, on 8th March, five new cases of corona virus in Kerala were again reported and since then the cases have been rising affecting 14 states. Coronavirus is a family of viruses that can cause illness, which can vary from common cold and cough to sometimes more severe disease. SARS-CoV-2 (n-corona virus) is the new virus of the corona virus family, which first discovered in 2019. It is a contiguous virus which started from Wuhan in December 2019 which later declared as Pandemic by WHO due to high rate spreads throughout the world. Pandemic is spreading all over the world; it becomes more important to understand about this spread.

From [World Health Organization](#) - On 31 December 2019, WHO was alerted to several cases of pneumonia in Wuhan City, Hubei Province of China.

So daily level information on the affected people can give some interesting insights when it is made available to the broader data science community.

[Johns Hopkins University has made an excellent dashboard](#) using the affected cases data. Data is extracted from the google sheets associated and made available here.

- COVID-19 cases at daily level is present in covid_19_india.csv file
- Individual level details are present in IndividualDetails.csv file and is obtained from [this link](#)
- Population at state level is present in population_india_census2011.csv file
- Number of COVID-19 tests at daily level in ICMRTestingDetails.csv file
- Number of hospital beds in each state in present in HospitalBedsIndia.csv file and is extracted from this [link](#)

- Travel history dataset by <https://www.kaggle.com/dheerajmpai/covidindiatravelhistory>

The number of new cases are increasing day by day around the world. This dataset has information from the states and union territories of India at daily level.

State Wise data fetched from [Ministry of Health & Family Welfare](#)

ICMR Testing Data comes from [Indian Council of Medical Research](#)

- <https://www.covid19india.org/> provides a crowd sourced data.
- Their API provides district level number of cases.
- The API also provides data on tests, facilities, no. of hospitals.
- More information about them can be found here : <https://www.covid19india.org/about>
- <https://api.covid19india.org/> has the list of all the API's that they have made available.
- Here is the github repo of the project : <https://github.com/covid19india/covid19india-react>.
- <https://github.com/imdevskp/covid-19-india-data>
- Population at state level is present in `population_india_census2011.csv` file.
- Number of COVID-19 tests at daily level in `ICMRTestingDetails.csv` file.
- Number of hospital beds in each state is present in `HospitalBedsIndia.csv` file and is extracted from this [link](#).

II. ACKNOWLEDGMENT

- Original source of data was https://developers.google.com/public-data/docs/canonical/countries_csv and https://developers.google.com/public-data/docs/canonical/states_csv. Data was originally released under a Creative Commons 4.0 license.
- Thanks to [covid19india.org](https://www.covid19india.org/) for making the individual level details and testing details available to general public.
- Thanks to Indian [Ministry of Health & Family Welfare](#) for making the data available to general public.
- Thanks to [Wikipedia](#) for population information.

III. DATA VISUALIZATION WITH PYTHON PROGRAMMING LANGUAGE:

1. Data Analysis with Python Language **(Python Notebook)**.
2. Comparison of COVID-19 Cases with other countries **(Table Wise)**.
3. Spike in India for COVID Cases **(Line Chart)**.
4. Age Group Analysis **(Pie Chart)**.
5. Indian COVID-19 Patients Outcome Age Wise **(Box-Plot)**.

TOOLS & OTHER GRAPHS FEASIBLY USED IN CAPSTONE DATA SCIENCE FINAL PROJECT

- MATPLOTLIB
- PANDAS
- NUMPY
- SCIPY
- SEABORN
- FOLIUM (WORLD LEVEL & INDIA LEVEL INTENSITY OF THE COVID-19 SPREAD)
- GEOSPATIAL DATA & CO-ORDINATES DATA
- AREA PLOT
- BAR CHART
- SCATTER PLOTS WORD CLOUD