*A Project Report*

On

**COVID-19 TRACKER**

*Submitted By*

19BEC002 (ADITI JHA)

19BEC005 (APARNA KHOLIA)

19BEC015 (HENI PRAJAPATI)

19BEC042 (SREYA GADEPALLY)

*Under the guidance of*

DR. UMA



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

**Purpose**

The outbreak of COVID-19, an infectious disease spread through a newly discovered strain of coronavirus has brought the entire world to a standstill in a matter of a few months from the day it was declared a pandemic - 30 January, 2020 - by the WHO (World Health Organization). This pandemic has called for extreme measures that has led the to complete shutdown of most countries. People have been encouraged to stay in quarantine in order to avoid human contact.

This project aims at providing important results and information by crunching up COVID-19 time-series data. It aims to process dynamic data sets and provide specific results that inform the user of the various aspects of effects of the corona virus disease. It is a python implementation of growth rate and other trend analysis.

**COVID-19 Tracker – 1: plots death growth rate of a country w.r.t the number of confirmed cases; lists high-risk countries for travel, for internships or other educational and professional purposes.**

This piece of code processes the time\_series\_covid19\_confirmed\_global.csv – a csv file that lists all the countries and the number of confirmed cases each day starting from 22 January, 2020 – and time\_series\_covid19\_deaths\_global.csv – a csv file that lists all the countries and the number of confirmed deaths each day starting from 22 January, 2020 – dynamic data sets. The links to the above two datasets is given below.

<https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv>

<https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv>

Countries with appreciable Human Development Index (HDI) has been parsed from the link below.

<https://en.wikipedia.org/wiki/Developed_country#Rankings>

The dynamic data sets are organized into a 2D data frame (2D array) using pandas library of python3. List data structure is used to store lists of countries.

**Output**:

* A trends graph of death growth rate of any country that is fed by the user. It plots the death rate with respect to the number of confirmed cases everyday starting from 22 January, 2020.
* The time series data of specific countries with high HDI is processed to produce a short, finalized list of high-risk-countries that are to be avoided by students while travelling for internships over a period of next two years. A country whose death growth rate has been more than 2% of the number of confirmed cases at any point of time comes under the high-risk-country category (not a standardized measure).

**Results and Observations**:

Since the results vary pertaining to the constantly changing death rates, a constant observation and analysis of the output suggests that most European countries are constantly under high-risk category.

Very few countries have a decreasing death growth rate over the period of observation. These results suggest that the spread of coronavirus is growing by each passing day resulting in more deaths than cured cases.

**COVID-19 Tracker – 2: lists high-risk states in India and the US for a given age group**

This code tells the user whether a state is a high-risk state, for the given age group. The data is taken from the Bangalore Mirror and the official website of John Hopkins University. Links are as given below:

* <https://coronavirus.jhu.edu/>
* <https://bangaloremirror.indiatimes.com/>

The criteria for labelling the states as High risk, Average risk and Low risk is set as:

* if the number of confirmed cases is greater than or equal to 1000, the state is considered high risk
* if the number of confirmed cases is less than 1000 but greater than or equal to 200, the state is considered to be at an average risk.
* if the number of conformed cases is less than 200, then the state is considered to be at a low risk

The data provided by various states had various age groups defined by them. So, the age groups that contained the given age group i.e. 15<=age<=25 have been considered.

The Data Structure Lists and CSV file for storage have been used.

The various states that were considered are:

New York (USA) Massachusetts (USA)

Pennsylvania (USA) California (USA)

Michigan (USA) Florida (USA)

Texas (USA) Georgia (USA)

Maryland (USA) Maharashtra (India)

Karnataka (India)

**Output**:

From the given data, we can see that the status of various states for the age groups are:

**High Risk**:

* Massachusetts for ages 20-25
* Pennsylvania for ages 20-25
* California for ages 20-25
* Florida for ages 15-24
* Georgia for ages 18-25
* Maryland for ages 20-25
* Maharashtra for ages 21-25

**Average Risk**:

* Massachusetts for ages 15-19
* California for ages 15-17
* Michigan for ages 20-25
* Texas for ages 20-25
* Georgia for ages 15-17
* Maryland for ages 15-19
* Maharashtra for ages 15-20

**Low Risk**:

* New York for ages 15-25
* Pennsylvania for ages 20-25
* Michigan for ages 15-20
* Texas for ages 15-19
* Karnataka for ages 15-25

**Results and Observations**:

From the above output, it can be inferred that people who fall in the age group of 20-25 years are affected the most. So, the age group of 20-25 is considered to be the high-risk age group. Hence, it is not recommended for students in this age group to travel extensively. The states of California and Georgia are most affected and states of New York and Michigan are comparatively less affected.

**COVID-19 Tracker – 3: calculate the average number of worldwide confirmed, recovered and death cases of COVID-19 week wise.**

The piece of code for this problem processes the data file time\_series\_covid\_19\_confirmed.csv which contains country names, dates and number of confirmed cases each day. Similarly, time\_series\_covid\_19\_death.csv contains country names, dates and number of death cases each day and time\_series\_covid\_19\_recovered.csv contains country names, dates and number of recovered cases each day.

The data was taken from the official site of Kaggle. The links are as follows:

1. Confirmed : <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#time_series_covid_19_confirmed.csv>
2. Death : <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#time_series_covid_19_deaths.csv>
3. Recovered : <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#time_series_covid_19_recovered.csv>

The aforementioned data sets contain data from 22nd January, 2020 to 5th April, 2020.

The software library used for this problem is pandas.

The data is organized into two-dimensional data frames with columns of potentially different data types.

To do the sum of all the values of column, the function used is df.sum() and for the sum of values of rows, df[“column name”] is used in the given code.

**Output**:

The output first shows the worldwide average number of confirmed cases starting from week1 to week11, then for death cases and then finally for recovered cases.

**Results and Observations**:

It is observed that as the days passed the effect of the virus is increasing as seen from the drastic increase in number of confirmed cases from 2029 to 1102302, number of death cases from 53 to 58511 in 10 weeks. On the other hand, the number of recovered cases has gone from 50 to 227080 in the same time period.

A very important inference from the output that has to be noted is that the number of recovered cases over this time period are more than the number of deaths. This shows progress towards curing COVID-19.