**COVID -19: The Pandemic**

* **TEAM MEMBERS:**

1. Madugula Yashwanth (19bec025)
2. Mohammed Abdul Sohail (19bec024)
3. Mounika Rayudu (19bec035)
4. Munagapati Anupama (19bec026)
5. Pranay Sai Teja (19bcs102)
6. Segu Pavan Kumar (19bcs097)
7. Sriram Shivaganesh (19bec043)
8. Sujith Joseph (19bec029)

* **PROBLEM STATEMENTS:**

1. To list out the high risk countries which are facing the Covid-19 Virus.
2. Plotting a graph of age group by deaths for selected states in United States (US) and trying to identify which age group is most affected.
3. Finding out the Average number of days it took for a confirmed case to turn

to a death state in any country and also finding the average number of days it took a confirmed case to turn to a recovery state in any country. Also displaying the country which took maximum number of days in either case.

* **The reference links used in our project are as follows:**

1. https://github.com/datasets/covid-19/blob/master/data/countries-aggregated.csv

(This reference is used in both Problem Statement 1 & 3)

1. <https://www.worldometers.info/coronavirus/country/us/>

(This reference is used in creating dataset for Problem Statement 2)

* **DATA FILE NAMES:**

1. Dataset for Problem\_Statement\_2:(An Excel Sheet Created by our team using the reference link mentioned). (Note: The data was extracted on 23rd April 2020)
2. 'countries-aggregated.csv'.(Used for Problem statement 1 & 3).

*PROBLEM STATEMENT 1*

* **PURPOSE:**

1. The main theme of writing this code is that to list out the countries at 'HIGH RISK' due to Covid-19 for students looking for internship for next 2 years.

* **OBJECTIVES:**

1. To Identify the HIGH RISK TRAVEL destination countries.

1. To find confirmed and death cases in a country and display the graph of death rate for that country.( Formula: Date Rate= (Number of deaths)/ (Number of confirmed cases) \* 100)
2. To print all the countries which are at high risk due to covid-19 and meant for internship.

(The countries having a death rate of 4 or above and in which a minimum of 200 deaths have occurred are taken into consideration.)

* **DESCRIPTION & OBSERVATIONS:**

1. Covid-19 has been effecting our world's economy in a drastically way, it was not affecting directly but the consequences due to covid-19 effected the abroad studies as well. As the covid-19 looms in all states of every country, there is a concern for the students to select safer countries for their internships and higher studies.
2. As to know, which state is at high risk and the confirmed cases and death cases in a particular country, the code we have implemented is very much useful.
3. In this code, we tried to sort out the countries which are at High risk (death rate>4.0) and display the countries that come under high risk.
4. This code is also helpful in displaying graph for the death rate of the required country and also confirmed and death cases of a particular Country which the student wants to visit for Internship.
5. From the output,

i) We can conclude the countries at high risk (for Example France, Italy, US, Belgium, etc) cannot be selected for internship by UG students.

ii) We can also summarize the death rate Statistics of the country which the student wishes to visit.

* **DATAFILES USED:**

1. Name of Dataset taken for consideration: 'countries-aggregated.csv'.
2. Reference Site: https://github.com/datasets/covid-19/blob/master/data/countries-aggregated.csv

* **DATA TYPES DESCRIPTION:**

1. To read the dataset we used the concept of pandas and made use of the 'read\_html' attribute. A great Advantage of using Pandas Data Frame is, it has various attributes which helps in accessing the columns of the dataset and modifying it as we wish.
2. After that we sorted the countries based on deaths (Criteria: minimum 200) and storing the implemented data in a list. list.sort(reverse=False) and extracted the percentage of countries having 'maximum Death Rate at any point of time'.('An important thing to note here is the present Death Rate of a country might be low, but it was high at a particular time which makes it a 'HIGH RISK COUNTRY')
3. To categorize the high risk countries, we used 'Death Rate concept'(Formula mentioned) and sorted the countries using the attribute df.'sort\_values'(it’s an attribute of Data Frame).These countries come under high risk countries for internships.
4. For plotting the graph we used matplotlibrary , pyplot module, which has great attributes like 'legend(),xlabel(),ylabel(),etc.'
5. To sum it up the Data Types used were Lists(zip and sort were the attributes used), Strings, Pandas(its many attributes such as pop, max are used), Data Frame(to display table and access contents),Pyplot module from MatplotLib (to display graph) were used in writing this code.

* **FINAL RESULT DESCRIPTION:**

1. The Code confirms whether the country (which student wants to visit) is at HIGH RISK or not, Displays No. of confirmed and Death cases in that country, the sorted high risk countries Table and the Graph showing death rate of the country is displayed.

*PROBLEM STATEMENT 2*

* **PURPOSE:**

1. The main intention of writing this code is to identify the deaths by age group in selected 8 States of World’s Leading Country US.

* **OBJECTIVES:**

1. Selecting some states of a country and collecting the information about deaths occurred in every Age Group. (Criteria for selection: Deaths > 500 for Cases >10,000)
2. States New York, Michigan, Massachusetts, Washington, Chicago (a city in Illinois), Ohio, Maryland and Colorado are considered for Observation. Deaths occurred in age groups 0-19,20- 29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+ are recorded from those States.( Refer Dataset for Problem\_Statement\_2.(data recorded on 23rd April 2020))
3. Plotting the Graph Age Group (on x-axis) vs No. of Deaths (on y-axis).

* **DESCRPITION & OBSERVATIONS:**

1. As to know, which age group has been worst affected by this Covid-19, the code we have implemented is very much useful.
2. This code clearly shows us the graph of Age Group vs. Deaths of the required state we wish to view.
3. From the plot, we could know which Age Groups are worst affected by the Virus.
4. From the Output, we could conclude that 80+ Age Group has been worst affected. The reason might be this age group people possess weak immune system and might be suffering from common chronic diseases.

* **DATAFILES USED:**

1. Name of Dataset taken for consideration: ‘Dataset for Problem\_Statement\_2’ (An excel sheet created by us from the data obtained from ref link).
2. Reference Site: <https://www.worldometers.info/coronavirus/country/us/>

* **DATA TYPES DESCRIPTION:**

1. To read the dataset which is in Excel sheet, we have used the module ‘xlrd’ and made use of concept of Pandas Data Frame that has helped us to access the Index (States) and Columns (Age group).
2. After that we have made the use of empty lists so as to append the data (deaths) corresponding to each list (Countries).
3. For plotting the graph, we have made the use Matploblibrary, pyplot module that has attributes like plot (), xlabel (), ylabel (), title (), legend ().

* **FINAL RESULT DESCRIPTION:**

1. This code shows us the Graph of Age group vs. Deaths of the required State. From this graph, we could conclude that Age Group 80+ has been affected drastically.

*PROBLEM STATEMENT 3*

* **PURPOSE:**

1. The main theme of writing this code is to find the average number of days it took for a confirmed Case to convert into a death state and back into recovered state.

* **OBJECTIVES OF CODE:**

1. Finding the average number of days it took for a confirmed case to turn in to a death

And also which country took the maximum number of days.

1. Finding the average number of days it took for a confirmed case to turn in to a recovered

State and also which country took the maximum number of days.

* **DESCRIPTION &OBSERVATIONS:**

1. In this code we tried finding the average days took to turn the confirmed case to

Death State and back to Recovered state by finding average days.

1. To get the average, no of days taken for first confirmed case to turn to a Death Case is considered. Similar procedure is followed for 7 cases and the average of days is taken.
2. This code is also helpful in getting the status i.e. days took for confirmed case to change

Into death and confirmed case to recovered for a particular country and even for the countries

which were inserted in the code. (Note that not all but few well-know countries are taken into consideration.)

* **DATAFILES USED:**

1. Name of Dataset taken for consideration: 'countries-aggregated.csv'.
2. Reference Site: https://github.com/datasets/covid-19/blob/master/data/countries-aggregated.csv

* **DATA TYPES DESCRIPTION:**

1. To access the contents of webpage considered, pandas module is used. Its attribute pandas.read\_html () helps in accessing contents of the page.
2. The Table is accessed using 'Pandas Data Frame' , its attributes such as iloc[] are used in getting required information
3. Use of functions makes the code easier to process, lists are being used to store and append data as and when required.
4. All in all Strings, Lists, Pandas, Data Frame (its attributes) and Functions were helpful in solving the problem.

* **FINAL RESULT DESCRIPTION:**

1. The code outputs the days taken by a country (given by user as input) to turn into death state as well as recovered state. It also outputs the country (from the countries taken into consideration) which took the maximum no of days to turn into either Death or Recovered State.