

CS102 Data Structures : Virtual Class Hackathon

Date : START Date : 11th April 2020
END Date : 25th April 2020

Team Formation : Maximum upto Eight Students in a Team. Teams with less number of students are also allowed.

Final Delivery : Sharing the deliverables on GitHub account

Deliverables:

- 1. RESEARCH PAPER/ARTICLE – one copy per TEAM covering all the three problems.**
- 2. Python PROGRAM/S – one copy per TEAM**

EVALUATION :

- 20 MARKS for RESEARCH ARTICLE**
- 30 MARKS for PROGRAM/S (10 mark for each problem Statement)**

PROBLEM DESCRIPTION :

Objective : To understand the distribution of data in COVID-19 Cases and perform various data processing operations using DATA STRUCTURES concept

Introduction : World Health Organisation (WHO) has declared COVID-19 as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020. The COVID-19 infectious disease is fast spreading across the countries, impacting the health of large numbers of people, and thus requires immediate actions to prevent the disease at the community level across the GLOBE.

As a Student community, it is a concern for the community as many of our students travel across the states and countries for their project work, and internship work in future. The result of this RESEARCH cum Data TRENDS study of this Class Virtual hackathon is therefore crucial. The final outcome is aimed towards understanding the infectious disease data pattern in the GLOBAL population such that the impact of COVID-19 has minimum effect on our students' Project/Internship selection planning process.

LIMITATION : The result is limited to only Data Exploration, Data observation and TRENDS. Exercise is not focusing on Forecasting or prediction modelling.

Problem Statements

1. Problem Statement 1:

- 1. Identity the countries as HIGH RISK TRAVEL destination countries for Internship or Project work for next two years.**

Criteria for selection :

- Use Dataset as Time series data.**
- Use confirmed cases and death cases data. The data set is growing dynamically.**
- Draw a trend graph of death growing rate for that country. The graph may change as the data set is dynamic. Collect the death growing rate (fatality or mortality).**

- $\text{Death growing rate} = \frac{\text{Number of Deaths}}{\text{Cases Confirmed}} * 100$
- If the death growing rate is more than 2% of the confirmed cases at any point of time, identify the country as a HIGH RISK COUNTRY.
- List out the country names

Problem Statement 2:

2. Identify the TOP FIFTEEN(15) countries as HIGH RISK Age Group Countries for Study

Criteria for selection :

- Age group to be selected between 15 to 25 ie $15 > \text{AGE} < 25$.
- Find out the number of people died in that age group.
- List the TOP Fifteen countries which come under this age group
- List the countries as HIGH RISK AGE Group COUNTRY for Studies.
- Draw the plot of Age/Number of deaths (x & y axis)

Note: These HIGH RISK COUNTRY and HIGH RISK AGE GROUP is defined only for this Academic EXERCISE purpose and do not come under any standard guidelines

Problem Statement 3:

- 3.1 Find out the Average number of days it took for a confirmed case to turn to a death state in any country. Which country took the maximum number of days?

- 3.2 Find out the Average number of days it took a confirmed case to turn to a recovery state in any country . Which country took the maximum number of days?

Recommended Sites to look for Dynamic DATA SETs of COVID-19

- John Hopkins University Site:
- WHO site
- Koggle
- Gitub

Select an authenticate site before finalising your data set. I have also included few sample DATA SETS.

Examples :

1. <https://data.world/covid-19-data-resource-hub/covid-19-case-counts>
2. <https://github.com/CSSEGISandData/COVID-19/issues/1250>
3. <https://github.com/datasets/covid-19/tree/master/data>
4. <https://data.world/covid-19-data-resource-hub/covid-19-case-counts>
5. <https://github.com/CSSEGISandData/COVID-19/issues/1250>
6. <https://github.com/datasets/covid-19/tree/master/data>
7. <https://github.com/CSSEGISandData/COVID-19>
8. https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#time_series_covid_19_confirmed.csv

Research Article Preparation Guidelines:

1. Sections to be included in the WORD document :

1. Enter Team details in the First Page (ROLL Number, Problem Selected and Data File names and REFERENCE names)

2. Second Page onwards :

1. Purpose of the RESEARCH topic Study
2. What are the important observations found in the Data Study done in this exercise
3. Name of Data Files used and its reference to the site used for the study.
4. List out what Data Types (eg:Strings, Arrays, Lists, Stacks, Queue, Dictionaries, Trees) or Data structures are used in Solving the problems
5. Final Results description
6. Name of the Document : Give a Proper Name and Include all your Roll NUMBER in the FIRST PAGE

Note:

Font to be used for documentation

1. Arial 14 for Heading.
2. Arial 12 for Text Body
3. Word document format
4. Number of WORD Pages to be Limited to Maximum Five.
5. Content should be written by the TEAM in their own words. NO plagiarism is allowed. If found, TEAM/individual will be disqualified from the evaluation.