**Project Phase-1: Planning**

Harsh Anand

Navya Akshitha Sangishetty

Nikitha Kakuru

Sampath Kumar Kuram

Department of Information Technology, Arizona State University

IFT 598 Data Visualization & Reporting for IT

Asmaa Elbadrawy

November 13, 2022

**Dataset Description:**

​​The dataset on which we will be working in this project is the "Novel Corona Virus 2019 Dataset," in which the World Health Organization expressed alarm about a new virus that has yet to be discovered and how it affects humans. When daily-level information on the afflicted persons is made available to the larger data science community, it can provide some fascinating insights.

This dataset contains daily statistics on the number of impacted patients, fatalities, and recovery from the 2019 new coronavirus. Thus the number of instances on any day represents the total number of cases.

The dataset contains the following Attributes/Columns:

* Observation Date MM/DD/YYYY - The observation showcases the number of covid cases reported daily.

Datatype: Interval

Domain: Range (01/22/2020 to 05/29/2021)

* Province/State – It showcases the names of provinces or states of different countries which are considered for observation in which few can be empty when there is missing data.

Datatype: Categorical

Domain: List (input is from the Province/State column)

* Country/Region – This column contains the names of the countries considered for observation.

Datatype: Categorical

Domain: List (input is from the Country/Region column)

* Confirmed – This column provides the cumulative number of confirmed cases of COVID - 19 till the date of the last observation conducted.

Datatype: Ratio

Domain: List (input is from the Confirmed column)

* Deaths - This column provides the cumulative number of deaths due to COVID - 19 till the date of the last observation conducted.

Datatype: Ratio

Domain: List (input is from the Deaths column)

* Recovered - This column provides the cumulative number of recovered cases of COVID - 19 till the date of the last observation conducted.

Datatype: Ratio

Domain: List(input is from the Recovered column)

* Longitude - it contains the longitudinal geographical location countries of confirmed, recovered, and death numbers.

Datatype: Interval

Domain: List(input is from the Longitude column)

* Latitude - it contains the latitudinal geographical location of confirmed, recovered, and death numbers.

Datatype: Interval

Domain: List(input is from the Longitude column)

**Prospective Dashboard Users**

* **Government**

The government relies heavily on data for decision-making and action to combat the COVID-19 pandemic. These data enable the government to set priorities and make quick and effective decisions in response to rapidly changing COVID-19 situations.

* **Health Services**

The dashboard will let physicians efficiently assess patient volumes and case severity in order to prioritize clinical care and allocate scarce resources more effectively.

* **General Public**

The dashboard provides everyone, from business and government leaders to the general public, with the information they need to make informed decisions.

**Potential Questions**

1. Top N countries/states with highest covid cases.
2. Countries/states with the highest number of death.
3. Daily cases, recovery, and death counters.
4. Daily trends of the total number of covid cases by countries/state.
5. Daily trends of the total number of covid recoveries by countries/state.
6. Daily trends of the total number of covid death by countries/state
7. Trend showing weekly cases, recoveries, and death by countries/state.
8. Trend showing monthly cases, recoveries, and death by countries/state.
9. Trends showing rate of change in confirmed and death cases.
10. Recovery percentage trends by country/state.

**References**

<https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset>

**Mural Link**

<https://app.mural.co/t/team55769/m/team55769/1668114250029/2c4c7e84aa826b9ffd995492d5ce26b252362fda?sender=ucaba08cce9c57a8852bd7940>