

## **Title:**

- Visualization of Historic Covid Data and its impacts using RStudio.

## **Objective:**

- This project helps people to learn how the covid pandemic affected the world and let them gain knowledge on which states have maximum and minimum rates of different scenarios like death rate, recovered rate, active rate. The main objective of this project is to simplify the data and create a visualization by which the people can understand in a broader perspective.

## **Motivation:**

- During the pandemic we have faced and suffered lot of consequences. Along with us the people who collected every individual data had suffered a lot compared to us. Many of them gathered the information and turned that into a dataset which can be used and be useful in future. This situation makes me to think that the collected data will be understandable only to who created it or who uses but not to the common people and who are newly using the data. And more over people cannot spend that much time to go through the dataset and check for each and every detail of the particular region or state or anything else. It will take some days or weeks to go through the whole data in the dataset. So, I decided to make it easier by converting the data into visualization by considering the some of the scenarios. This would make it easier for the people who cannot understand the dataset or if someone wants particular scenario, he can visualize the data for easy understanding. This can reduce the time and make more comfortable to understand.

**“A picture is worth a thousand words”**- is our inspiration and motivation to work up on the available dataset to get it into pictorial representation.

## **Team Members:**

We are two members in our team.

1.Preethi Uppuluri

2.Mohan Chandra Rangu

## **Proposed methodology:**

The proposed methodology is what we learned in the class:

KDD (Knowledge Discovery in Database):

### **1. Learning the application domain**

- we have learned and gained the knowledge of the dataset and able to understand what it is saying to the us.

### **2. Creating a target dataset**

- we have spent some quality time to go through some datasets that are available on google sources and we decided to choose selected the covid\_19\_clean\_complete Dataset.

### **3. Data cleaning and pre-processing**

### **4. Data reduction and projection**

### **5. Choosing the data mining function (summarizing, clustering, classification and regression)**

### **6. Choosing the data mining algorithms**

### **7. Data Mining (process of analyzing large databases) hyperparameters? Methods?**

### **8. Interpretation (what is the meaning of results?)**

### **9. Using discovered knowledge (using the obtained knowledge in real world ?**

- **We will follow the 3-9 steps to make our project successful.**

### **Tool:**

- we believe the RStudio provides user friendly environment to achieve above 9 steps.

### **Expected Outcome:**

We cannot have single outcome from one dataset. Multiple scenarios will arise in brain after studying the dataset. So our project outcome includes:

- Visualization of data of W.H.O Region and Death rate.
- Visualization of data of Death rate vs Country.
- Visualization of data of Recovered rate vs Country.

We are sure that we will can get some more new ideas while doing the project after knowing what is probability and how can it be used in RStudio.