Final Project Proposal

IST 652 – Scripting for Data Analysis

Team Members :

We will be working in a team of two members i.e., the same group as for the advanced topic presentation. Our team members include:

1. Anvitha Shashidhar (aks100@syr.edu)
2. Chirag Sable (csable@syr.edu)

Topic of Investigation: Covid Data Analysis

For this project, we will be investigating the Covid data of India, USA and comparing it to that of the world.

Data sets with references :

We plan to use these three datasets for our final project:

1. Covid Cases in India :

<https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india>

It is a dataset on the COVID-19 pandemic in India, which includes time-series data tracking the number of people affected by the virus in each state at daily level

1. Covid Cases in USA :

<https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-usa>

The dataset is called COVID-19 in USA1. It has information from 50 US states and the District of Columbia at daily level. The dataset provides information on the number of people affected by the virus in each state at daily level.

1. Covid Cases Worldwide:

<https://www.kaggle.com/datasets/themrityunjaypathak/covid-cases-and-deaths-worldwide>

This dataset contains the covid cases of the entire world. It has fields such as total number of cases, total recovered, deaths and other related fields.

Data Acquisition and Analyses :

We will be working on the following questions :

1. Covid Positive Ratio Analysis - Which country has the most COVID positive ratio in percentage

For this analysis, the attributes will be the total number of positive tests, recovered number of patients and the population

1. How does the situation of covid look like in the two major countries of the world i.e India and the USA – For this we will be doing different kinds of exploratory data analysis with the help of some visualizations
2. Death Analysis – Analyses on how many people lost their lives with respect to the world.
3. How has the COVID-19 pandemic impacted different age groups, genders and ethnicities in India, USA, and the world? Are there any significant differences in the data?

Potential Development Tasks :

1. We intend to use Regressions and Forecasting to get an understanding of the possible situation of the future.
2. Conducting statistical analysis can help in identifying relationships between different variables and detecting significant differences between groups.
3. Since the COVID-19 dataset is time-series data, we can conduct time-series analysis can help in predicting future trends and patterns.