Database Management Systems (COP5725)

Fall 2021

Instructor: Dr. Markus Schneider

TA: Kyuseo Park

Project Deliverable 1

Title: COVID-19 impact trends across the US and comparison with

rest of the world

Bodke, Gauri Pandharinath	<g.bodke@ufl.edu>
Gopinath, Ashwin	<ashwingopinath@ufl.edu>
Saoji, Nikhil Mukesh	<n.saoji@ufl.edu>
Viswanadha, Srinivas Nishant	<viswanadha.s@ufl.edu>

Pledge:

On our honor, we have neither given nor received unauthorized aid in preparing this project report 1.

For scoring use only:

	Maximum	Received
Rubric 1	10	
Rubric 2	10	
Rubric 3	10	
Rubric 4	15	
Rubric 5	25	
Rubric 6	25	
Rubric 7	5	

Table of Contents

Project Deliverable 1

Group 09

1   Introduction

1.1   Overview

Corona Virus (COVID-19) has changed our lives in the last 18 months. The world had to come to a standstill with absolutely no direction on how to face human adversity. Everyone stayed home watching the world events unfold in a way no one had envisioned. Food markets and utility stores were raided, some lost jobs and had no income; companies faced huge financial losses; the travel & tourism sector and the F&B industry are still trying to recuperate their losses. More importantly, many of them fell sick and some of them succumbed to death. Over the last six months, the world has collectively come together and found a way to mitigate the virus by vaccinating people that generate antibodies that fight the virus. Nevertheless, as we speak, people are still being victims of the virus, and fear of falling sick or death always looms around.

1.2   Project Description

The impact of COVID-19 was across the globe and across all kinds of people as well as different Industries. The data keeps changing every second and it gets difficult to handle such huge volumes of data. It also gets difficult for various health organizations, the state organizations, to keep track of data and analyze the impact of COVID-19. The United States of America is one of the countries which were impacted on a large scale. As of today, close to 43 million have been impacted by COVID-19 and there have been approximately 700 thousand deaths all over the US.

The aim of our project is to analyze the impact of COVID-19 on the United States as a country as well as determine how individual states have performed against COVID-19. We also look in-depth regarding the trends of vaccinations, rate of hospitalization, rate of positivity in each state as well as the country. We then go on to compare these results with the rest of the world and see how the United States as a country has performed against COVID-19. This analysis also helps the country to observe the impact and helps them use the existing data to be better prepared for future contingencies.

1.3   Motivation for the project

Our ancestors faced adversities in their times, and they had no technology or equipment to be better prepared. But against all odds, they have survived all adversaries and are one of the reasons why we are where we are today. Currently, COVID-19 is one such testing obstacle for the human race, and with the current technology in hand, it's our responsibility to make the best use of it and ensure the impact is as minimal as possible. Our product is to design a website that keeps track of the data on the impact of COVID-19 on the United States and gives an opportunity for the government, health organizations, and the public to be aware of the situation. It's impossible to manually keep records of the continuous exponential increase of data. Thus, a computerized database system can solve many problems with a very minimum resource requirement. These systems are not only high-speed, but they can store millions of records without a need to manage them. These database

systems provide the user with a high functionality, high-reliability framework, which can be leveraged through real-time access.

In our application, we need to store more than a million records of COVID-19 related data since March 2020. This data will be used in real-time by a web application to display trends and descriptive reports of COVID-19 data to various user groups. Database Systems allow us to store complex data and provide an ability to perform complex computations on this data to provide new and quality insights for better decision making.

2   Users & Stakeholders

For our project, we have identified four different users who will benefit from our analysis of COVID-19 data and be able to prepare and understand the impact of COVID-19 on the Unites States of America. Users using our system will be:

1. World Health Organization (WHO)

2. Hospitals and medical centers in the US

3. General Public

4. The US Government and the State Governments

2.1   World Health Organization (WHO)

WHO is an international organization which belongs to a specialized agency of the United Nations responsible for international public health. The goal of the organization is to give strong advice, provide direction and help trace outbreaks and provide necessary support as and when required. The trend queries on the overall impact of COVID-19 across the United States would allow WHO to:

- Analyze the impact and give them insight on how the rate of cases, rate of hospitalization, and also the rate of vaccination for each state and the country as a whole
- Use the impact on the US as a case study and help the developing countries to fight COVID-19
- Maintain a single repository to track data and monitor the situation regularly and provide direction as and when needed

2.2   Hospitals and Medical Centers

Hospitals and Medical Centers play a vital role in the whole COVID-19 situation all over the world. They are responsible for the diagnosis, prescriptions, and vaccinations for all the patients and apart from this, they are also responsible for the infrastructure and basic necessities and security for all the employees as well as customers. Some of the usefulness of the information is:

- Detailed report and inferences on the number of hospitalizations, beds, positive cases, deaths, admitted patients, and vaccinations that occurred during a specific time interval

- Trends of increased cases and deaths give foresight on what to expect and prepare infrastructure and other essentials accordingly

2.3   General Public

The public was the victim of the devastating COVID-19 virus in the Unites States and the world as a whole. Due to the ease of spreading the virus, it was important for the public to be aware of situations and ensure they play their part for a safer environment and for the benefit of their personal health as well as for the benefit of others. The data allows them to:

- Stay abreast of the latest COVID-19 information and accordingly take self-measures
- Keep track of information all over the country and make travel plans accordingly
- Be a catalyst and spread awareness based on the data to help extended family and friends and mitigate damage and exposure

2.4   The US Government and the State Governments

The local authorities are responsible for taking social measures that ensure the safety of the public. They have the power and resources to take necessary actions and issue standard operating measures to put the situation under control. But to do so, they will need valid appropriate data and insights on the threshold values or the situations in when the instructions need to be issued. Thus, they use the data to:

- Issue guidelines when there is a high volume of cases for the operation of restaurants, public buildings, and so on where the foot traffic might be high
- Prepare beforehand on the necessary logistics for the vaccinations and work with medical centers to raise awareness in the least vaccinated zones.

3   Application Goals in terms of Trend Analysis

The application goals regarding the analysis of trends can be different for each user.

- **World Health Organization (WHO):**

WHO can use the information to make new observations based on the current COVID-19 scenario and ensure they provide right directions and measures to ensure the impact is as minimal as possible. As a global organization they can compare the scenarios in the United States and the rest of the world and design a guidebook or procedures to help others mitigate the impact from the virus. They can also plan for sustainable solutions for the future to deal with COVID-19 like viruses.

- **Hospitals and Medical Centers:**

This trend analysis would help the medical field in the preparation of COVID-19 medical infrastructure and to provide better arrangements for patients. They can also infer the rate of vaccinations and accordingly prepare the stock for vaccinations and raise awareness if required.

- **General Public:**

The information of trends would make people aware of the daily cases and most affected COVID-19 regions. They can also check information for each state in the United States and with respect to other countries and plan things like travel accordingly. This can make the public more responsible and precautionary towards COVID-19.

- **The US Government and the State Governments:**

The trends would help the government to impose precautionary measures in the required states and take necessary actions in certain zones to control COVID-19. Also, the US government can play a vital role in the distribution of vaccines and taking measures on travel by comparing COVID-19 cases in other countries.

4   Real-World Dataset

For our platform, we will be using the COVID-19 data on daily statistics regarding the number of tests, positive cases, deaths, hospitalization, vaccination and so on. It is a freely available, centralized, and up-to-date repository hosted on the AWS cloud. It is curated by the AWS data lake team itself by working with well-recognized partners like John Hopkins University and New York Times. Hence, this dataset is reliable and accurate. It has all the data from the beginning of the pandemic till August 2021. It contains county and state-wise data across the United States and contains a global dataset that has information about various countries in the world on the above-mentioned parameters. All tables together include more than a million records. In addition, we will be using regional census data to help us with our analysis. Some of the tables and corresponding attributes in the dataset are as follows:

1. Global COVID-19 data:

It provides daily statistics on the number of tests, positive cases, deaths, hospitalization, etc. It has attributes related to the following: *Date, Country, Number of tests performed, Number of newly confirmed cases, Number of deaths, Hospitalization count, ICU count*

2. The United States COVID-19 data:

It provides daily statistics for individual states in the United States. It has attributes related to the following: *Date, State, Number of tests performed, Number of positive results, Deaths, Hospitalization count, ICU count, Ventilator count*

3. State-wise vaccination data:

It includes daily vaccination count across various states in the United States. Along with the census data, it will help us calculate the vaccination rate and cumulative vaccination count. It has attributes related to the following: *Date, State, Vaccination count, Number of people vaccinated, Number of vaccines distributed, Percentage of usage of shared vaccine doses*

The above data collected from multiple sources is sufficient to do the trend analysis explained in the next section.

5   Complex Trend Queries

Trend analysis is the focus of this application system. Depending on the user, there will be different types of trends presented on the dashboard. The focus will be on trends based on various time ranges and scales. The general trend queries that will be presented in all the user views are:

- **How has the positivity rate changed over time between regions?**

This query would give us a trend on how the positivity rate has changed across various states in the country (and various countries in the world) over a time period. For example, analyzing this query from Jan 2021 to Feb 2021 would allow the government to identify the containment zones, track the increase in the number of containment zones, and take preventive measures accordingly. The user will be able to see a line graph showing the fluctuations in the positivity rate over a time period. It will have time on the x-axis and positivity rate on the y-axis, and each line graph will represent a single state (country).

- **How has the death rate varied over time?**

This query will showcase the trend on how the death rate has varied across various states in the country (and various countries in the world) over a time period. It will help us understand if there is significant stress on the medical system. The user will be able to see a line graph showing the death rate over the years. It will have time on the x-axis and positivity rate on the y-axis, and each line graph will represent a single state.

- **Depicting the rise and fall of COVID-19 cases across all the states in the United States:**

This query allows us to analyze the overall increase or decrease of COVID-19 cases for various states in the country over a given time interval. For example, analyzing this query from Aug. 2020 to Jan. 2021 would help us understand whether cases were rising, stagnant, or decreasing across various states. This would allow the stakeholders mentioned, like the local government and the public to stay aware of the situation and allow them to take necessary precautions. The graph would have time on the x-axis and the y-axis would have positive cases with each line showing the considered state.

- **What is the rate of hospitalization over a given time interval?**

The output of this query shows the rate at which patients who have been infected with COVID-19 are being hospitalized for a given state over a given time interval. Assuming the rate of hospitalization is high allows the hospitals who are one of our stakeholders to be prepared to receive more patients in the future and will also indicate the government to take strict actions to reduce the hospitalization rate. A similar kind of analysis can be done if there is a decline in the rate of hospitalization. The x-axis contains time, and the y-axis contains different values depicting the overall hospitalization rate.

- **How many of the positive cases have been admitted in ICU or in a critical state on the ventilator?**

When a person tests positive, he/she could have severe symptoms due to which they'll have to get hospitalized. This query provides insights on how many of the patients who are tested positive and admitted to the hospital are in ICU or are on the ventilator. It shows a line graph where on the Y-axis, we will represent percentage of patients in ICU or on Ventilator out of total hospitalization and the X-axis will contain the time interval. If the rate of patients in the ICU is increasing, it allows the stakeholders to understand the seriousness of the situation.

- **What is the number of vaccinations per million people within the states in the US and how has the US performed against the rest of the world?**

The aim of this query is to render data of the number of vaccinations that have taken place for every million people across various states in the US. It would allow the stakeholders to understand the willingness of the people to get vaccinated. It also tells them which states require more effort to raise awareness about the necessity of vaccination so that most of the people can get vaccinated. We also check on how the US has performed against the rest of the world on the number of vaccine doses injected. In the graph, the x-axis will have time and the y-axis will have the population in terms of millions.

- **Calculating the Pandemic Growth Factor (PGF) and predicting the number of future cases:**

The number of cases for the day 'N' divided by the no of cases on day 'N-1' will give us the growth factor (PGF). By calculating the average PGF for the last 5 days we can predict the no of cases for the next 10 days using the below formula,

Number of cases after N days = Number of cases today * (PGF ^ n)

6   Public & Proprietary Software

The overall application consists of three major components: Frontend, Backend, and Database. The frontend contains all the dashboards a user can interact with inside a web browser. End users will interact with it to get the trend they want. The backend refers to the server-side of an application that processes user requests and acts as a middleware between the client-side and the database. Here the database system acts as one of the most important components as it will store the dataset and perform highly complex computations according to the user inputs. The technologies we will be using to develop this system are listed below:

- Frontend:
    - HTML: A markup language to define the basic structure of a web page
    - CSS and Bootstrap: A CSS framework to create responsive and visually appealing web pages that render well on different types of devices.
    - JavaScript: A scripting language that helps to build interactive web pages
    - D3.js: A JavaScript library that is widely used to plot dynamic and interactive plots and graphs on web-based dashboards. We will use this library to plot complex trend queries for individual users.
- Backend:
    - Node.js: A runtime environment that allows server-side execution of JavaScript. We will use this to write the application logic and to connect with our database server.
    - Express: A widely used Node.js framework for backend development of web applications. One of the main reasons for using this framework is its lightweight nature. It will allow us to focus more on the database part instead of getting stuck in the intricacies of managing codes and files in a full-fledged framework.
- Database:
    - Oracle Database: One of the most advanced RDBMS systems that is used to store structured data and to perform complex computations. We will use this database to store our dataset and dynamically evaluate complex queries according to user inputs.