Covid-19 Visualization and Analyzation Application

Group Alright

Yun Miao, Jingqi Guan, Zhe Li, Ting Deng

1. Introduction

Purpose of This Document

The Covid-19 Visualization and Analization Application is an up-to-date Covid-19 online interactive dashboard, that analyzes data from accredited resources, to provide secured, unbiased and precise data with proper interpretations to the public on a daily basis.

Our goal is to spread the correct information, raise awareness among the public, and reduce the impact of misinformation on our fight against the pandemic. Since the pandemic started, our society has not only suffered from the disease itself, but also has gotten worse off from misinformation, a lack of data transparency, and poorly interpreted data. These issues prevent us from going back to normal life, but also make our society less united. Considering all of these problems, we want to do something to solve these problems to fight against the Covid-19, and help to make our society a better place.

Background Information and Literature Review:

The Covid-19 pandemic has been with us for almost 2 years, and the World Health Organization (WHO) suggests that the virus could continue for years. Covid-19 has a large reproduction number, according to WHO, the reproduction rate of the original alpha variant is estimated to fall between 1.4 and 2.4, which means in a population, an infectious case can cause 1.4 to 2.4 secondary cases. Reproduction number, also known as R0, is a common term in epidemiology to indicate how contagious a virus is, it evaluates the spread speed of a disease. If

a R0 is greater than 1, the disease will continue, if R0 less than 1, the disease will end eventually. Yet as we know, Covid-19 has a large reproduction number, the pandemic spiked once it landed in North America. It started with a positive case of 286 on January 24th 2020, and has spread over multiple regions and increased dramatically to 7659 on March 11th, the WHO declared Covid-19 is a global pandemic on the same day. Until today, the Covid-19 virus has changed to multiple variants such as beta, delta, gamma and mu variants. Each variant has a different reproduction number that has different impacts on the pandemic. The reproductive rate of the delta variant spread 2 more times faster than the initial alpha variant; the lab report shows its reproduction number ranged from 3.8 to 8.0, with an estimated mean of 5.08. Hence we can easily observe another dramatic spike in Summer 2021. The total count of infected cases shifted to 275 million cases, with a solid daily confirmed case near 550,000, and death cases reached 5.15 million.

Intended Audience

Since the Covid-19 spread across the United States rapidly and causes thousands of cases per day, our intended audience is vast, covering all US residents who are affected by the Covid-19 pandemic and deeply hope the pandemic ends. Precisely, the intended audience to the dashboard includes our governors, policy makers, travellers, and all of the US residents similar to us.

Governors and policy makers need precise and timely information to make quick decisions to encounter the changing pandemic situation. Travellers need detailed destination Covid-19 information to compensate for the travel plan, such as daily cases and government policies, travel requirements. All of the US residents are all in this pandemic together, we need to work tight together to get our lives back to normal.

The dashboard takes the data published by John Hopkins University as a primary data source that covers the United States and is updated everyday. The datasource contains detailed Covid-19 data subject to each US county, and each state's vaccination rate. A US Covid-19 case map is attached to the dashboard that enables users to click for detailed datas and each county's Covid-19 cases. The dashboard users can also find states with most and least infected cases per request. Infected cases count and vaccination rate subject to each state will be plotted by the dashboard for a clear comparison upon the user changing the request.

Compared to other dashboard applications, we built our dashboard portable, interactive and can be easily accessed through a web browser that does not require downloading a phone application. The Covid-19 Visualization and Analyzation Application is designed for web browsers that can present nationwide Covid-19 data with minimized processing time that is also interactive to the users. The dashboard costs 40 hours of work to complete. We employed Amazon Web Services to host and secure our dashboard, and Tableau to process the data.

Plan for Introduction:

1.1: Project goals and objectives:

Goals: build an interactive dashboard that visualizes and analyzes the Covid-19 dataset and vaccination dataset, plot data from desired states for comparison, and provide precise information to our target users.

Objectives: Covid-19 data and vaccination data across the United States that are updated everyday.

1.2 problem and motivation

Problems: Misinformation and misleading Covid-19 data interpretation make the public worse off, and did not help with combating the pandemic, not enough awareness for our current situation. Technically, to consistently provide the Covid-19 dashboard service, we need a secured platform to host our dashboard and save our data. Also, we need a proper model that responds with a decent processing time to analyze the Covid-19 data since our dashboard is interactive and allows users to send requests to our web server to get desired Covid-19 information.

Motivation: providing unbiased information and properly interpreted data to the public, especially people in need of correct information, and raising the awareness to fight against Covid-19 that helps to end the pandemic.

1.3 Project application and impact

Application: Our dashboard can be applied to multiple fields subject to our targeted users, which includes governor and policy makers, travellers and all US residents.The dashboard can be used as a reliable reference when the government need to come up with a new policy to combat the pandemic, and can also be used as an unbiased indicator for all users to get an overall knowledge of our situation. Moreover, travellers can use our dashboard as a collective of Covid-19 information to compensate for destination government information with respect to the pandemic.

Impact: We create a secured dashboard to the public that provides unbiased and accurate Covid-19 information to the society that has been suffering from Covid-19 pandemic and being distorted by misinformations. With proper advertisements, our dashboard will have more users, and they will be better off by the unbiased data and proper interpretations from our dashboard. In the competitive market, misinformations will be offset by our dashboard.

1.4 Project results and expected deliverables

Results: a nicely constructed dashboard that provides consistent, secured, and exhaust information to users, and also suitable for web browsers.

Expected deliverables: a dashboard based on web browser and requires SSO per visit.

1.5 Market research

Multiple companies provides Covid-19 dashboarding services, including medias such as New York Times (https://www.nytimes.com/interactive/2021/us/california-covid-cases.html), World Health Organization (https://covid19.who.int/), and each level of CDC government. However, media are often biased, and thus provide not properly interpreted data to the public. The MicroSoft Powerful BI and the government dashboarding services lack interaction and are too wordy to read. The World Health Organization provides the best dashboarding services in our opinion, however it's not focused on the United States. Considering all these drawbacks of the existing dashboarding services, we created our own dashboard that only analyzes the Covid-19 data in the United States, which is unbiased, and focuses on user interaction.

1.6 Project report structure

We will provide a detailed explanation about background and relative works, system requirement and analysis, system design, implementation, testing and implementation in the following part of the project report.

2. Background and Objective

In this section, we will describe the background and objectives of this project in detail.

2.1 Background

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease spread worldwide, leading to an ongoing pandemic. There are 257,282,155 confirmed cases and 5,148,350 deaths as of today. Many countries have been working together to reduce the spread. It's critical for policy makers to understand the COVID spread statistics more easily to help me make better policies fast.

Data visualization is an interdisciplinary field that deals with graphic representation of data. It's a rather efficient way of communicating when the volume of data is numerous or it's associated with a time series or a geographical pattern. A mapping between the actual data and the graphic representation has to be created carefully to preserve the original meaning and convey the message more effectively. It requires a deep understanding of statistics as well as design skills and computing skills.

Data analytics is the process of systematically analysis of data or statistics using computations. It's used for finding meaningful patterns in data. Then, the meaningful patterns can be referenced towards effective decision making. It requires skills consisting of statistics, computer programming.

2.2 Objective

Obviously, it's crucial to utilize data visualization and data analytics to help understand of COVID-19 patterns. Here, we explain the objectives in several directions.

First, the project can help people track the spread. People need to keep track of COVID stats to calculate their risks of whether to do a long distance travel or whether they can meet their close relatives. There are multiple sources of data since not only you need to track the new cases and deaths each day, you also need to discover a trend by yourselves for your current location and the

location you want to visit. It's a lot of work and the information is scattered in different places. It's important to provide a single place with good visualization for people to keep track of the spread more easily and inform them to make better decisions.

Second, we aim to help people understand the trend of COVID-19. We utilized data visualization techniques to create a plot of COVID-19 trend in a map of the United States. Users can use the map to understand new cases and deaths information more directly. Plus, they can change the date in the date selector which adds another dimension to help them view the trend along with the geographical information.

Third, the project can potentially help future researchers and policy makers. We collected the COVID-19 dataset in different sources and cleaned them in a way that's easy to host on any cloud services. The data we collected can be used by any researcher for their ongoing research as a cleaned data source.

3. Organization

In this section, we will explain the organization of this project.

3.1 Project Group

There are several groups of works in the project. We will describe each one in detail in the following paragraphs. In the end, there will be a table describing each person's responsibilities.

- Data collection
- Data cleaning
- Cloud architecture
- Cloud security
- Data visualization

- Web app

- SSO

| Name	Initials	Responsibility (roles)
Yun Miao	YM	AWS architecture and data visualization
Jingqi Guan	JG	Data collection and data cleaning
Zhe Li	ZL	AWS architecture, web apps, and SSO
Ting Deng	TD	AWS security and data collection

3.2 Expected Users

The target users are listed below:

- Average American who wants to keep track of COVID-19 information to help them make better decisions.

- Policy makers who need to have a deep understanding of the COVID-19 trends to help them calculate the trade-offs to make better policies.

- Researchers who are required to monitor the spread of COVID-19 and predict the future trend and directions of the spread.

4. Development Process

This project follows the development process as shown in the Figure 1 below. There are three main stages: Data storage and Processing, Application Development, Protection & Security. We used Amazon S3 for data storage, Tableau for integration and dashboard creation,

Tableau online for html creation, Google SSO for authentication, AWS cloudfront for deploying

our application and AWS CLi to generate the signed URLs for security.
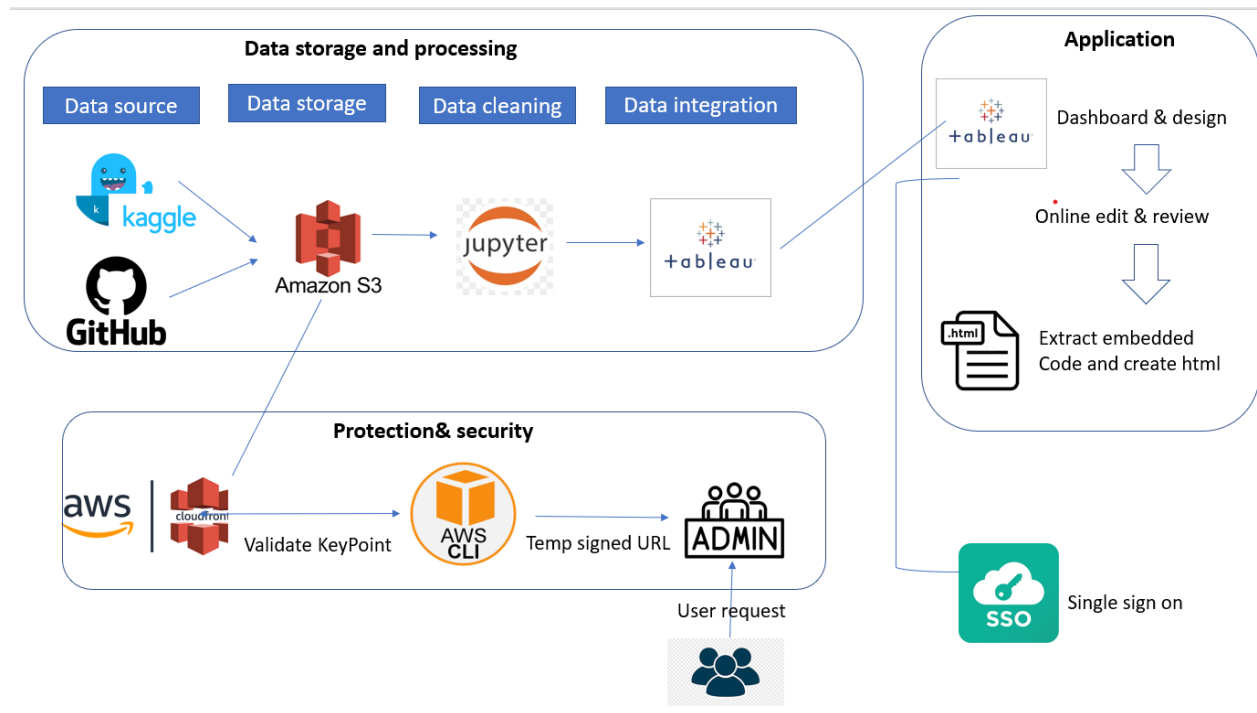
5. Deliverables

To	Output	Planned time	Promised time	Late	Delivered time	Notes
Class Presentation	228 presentation .pptx	11/10/2021	5 days	0 day	11/15/2021	None
Class Presentation	228 tableau workbook	9/31/2021	4 weeks	1 week	11/15/2021	None
Class Presentation	Web application	10/10/2021	4 weeks	0 day	11/15/2021	None
Term Project	Report	11/15/2021	1 week	0 day	11/21/2021	None
Term	Github	11/15/2021	1 week	0 day	11/21/2021	None

Project	Repository					
Term Project	NIST Form	11/15/2021	1 week	0 day	11/21/2021	None

6. Project Risks

In the process of the project, we thought of some possible project risks that could happen and also overcame some risks that happened. There are some risks that were noting down:

A. Purpose and need are not well designed. The possibility is medium and the possible risk that is caused by this problem is very high. In order to decide a topic that everyone is interested in, We have a meeting with teammates to discuss and identify the goals of the project, the project deliverables, and the target.

B. Data structure risk. There are many missing values or null values in datasets that would cause the analysis to be not accurate. While doing data loading, check our datasets carefully to check for missing values for key features. After checking, the average missing value percentage for a column is about 5% out of 20,000 rows. We decided to drop the missing values because they do not impact a lot on our analysis.

C. Communication risk that each role could have been doing duplicated work. Affected by the pandemic, we had meetings online. Communication risk is one of the most common risks. We realized it and worked hard to overcome it. The project leader communicates with teammates on a timely basis and everyone delivers certain results of their own findings. In a shared document, we noted down specifically each person's responsibility.

D. Time conflicts risk that the time of each part will take longer than expected. To overestimate the time that we need for specific parts and arrange priority tasks that will

impact teammates' parts. This way we can have time to schedule later. Also, we help

each other understand the project lifecycle well so that everyone can manage the time

well.

E. Performance risk that the project result is unlikely to achieve the result that we want. We

saved our work for every step and connected to a data warehouse prior to doing any

analysis or data cleaning. If the result is not what we wanted, we can test again by

restarting the steps and find out what goes wrong.

7. Communication

Collaborators:

Group: Alright

Zhe Li   013835783   zhe.li01@sjsu.edu

Jingqi Guan   015786225   jingqi.guan@sjsu.edu

Yun Miao   015970825   ting.deng@sjsu.edu

Ting Deng   014558310   yun.miao@sjsu.edu

All source code and finished documentation will be uploaded to Github repository.

Repository URL: https://github.com/JoeyLi01/Team-Alright-project

8.Project Plan

Our project plan is described as specific topics and timeline for each topic. There are two main parts: time schedule and test plan. Our time schedule is based on five big questions: project idea and topic, project design & deliverables, Data processing which includes data extraction, data cleaning and data integration, Application setup such as dashboard design and embedded code, Security & protection part to secure the application from outside users. Also, we planned to perform a few tests on coding and the system. Firstly, we tested the coding in order to check whether coding is working and make sure bugs and errors are found. Therefore, we displayed the application, tested different numbers to see results, and compared with google data to see confirmed cases & death. The coding ran well and the data matched.

After this, we performed a system test to analyze what improvements are needed for the system. By listing the target users and project purpose and doing a small survey to find the most liked features, we decided to add top category features on the application so that users can see the top and bottom states for the confirmed cases.

Reference

1. Ying Liu, Joacim Rocklöv: The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus, *Journal of Travel Medicine*, Volume 28, Issue 7, October 2021.

   DOI: https://academic.oup.com/jtm/article/28/7/taab124/6346388

2. Nicky Phillips: The coronavirus is here to stay — here's what that means, *Nature*, 16 February 2021.

   DOI: https://www.nature.com/articles/d41586-021-00396-2

3. Michael Greshko: COVID-19 will likely be with us forever. Here's how we'll live with it. *National Geographic*, January 22, 2021

   DOI:https://www.nationalgeographic.com/science/article/covid-19-will-likely-be-with-us-forever-heres-how-well-live-with-it

4. Nithya C Achaiah, Sindhu B Subbarajasetty, and Rajesh M Shetty: R0 and Re of COVID-19: Can We Predict When the Pandemic Outbreak will be Contained? *Indian Journal of Critical Care Medicine,* 2020 Nov; 24(11), page: 1125-1127.

   DOI: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7751056/

5. Hamid Mukhtar, Hafiz Farooq Ahmad, Muhammad Zahid Khan, and Nasim Ullah: Analysis and Evaluation of COVID-19 Web Applications for Health Professionals: Challenges and Opportunities. *Healthcare,* Volume 8, Issue 4, Dec 2020.

   DOI: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7712438/

6. Mahdi Salehi, Mohammad Arashi, Andriette Bekker, Johan Ferreira, Ding-Geng Chen, Foad Esmaeili1 and Motala Frances: A Synergetic R-Shiny Portal for Modeling and Tracking of COVID-19 Data. *Technology and Code, Public Health.* January 27 2021.

   DOI: https://www.frontiersin.org/articles/10.3389/fpubh.2020.623624/full