

# COVID-19 Data Extractor - IS 411

## **Project Summary:**

Our project focuses on delivering our users COVID-related health data. For instance, we may include the vaccination rate, hospitality rate, and the death rate of COVID-19 cases in the United States for the past two and a half years in our database. Our users can extract any of the data that they would like.

In addition, we also plan to offer users the option to generate data visualizations based on the data that they choose to extract. It could be a scatter plot, a histogram and maybe even a map.

## **Description:**

Our project aims to provide users with accessible ways to find COVID-related data. In the website that we design, users will be able to get data just by typing a couple keywords instead of typing an entire SQL query, although that is also an option for using our website. For instance, if the client types "Champaign, IL", our program could be able to automatically generate all the COVID-related data from our database related to Champaign.

## **Usefulness:**

The application will be used to query and update the number of COVID-19 infected people from different countries and regions. Users will be able to select any keywords they like to search for information conveniently. There is a famous COVID-19 world map here: <https://coronavirus.jhu.edu/>. Our website will be more concise than this one, so users can find the information they want more easily instead of having to search all over the site for what they want. Additionally, our website will be able to provide data precise to the county, which this website does not offer.

## **Realness:**

Our data is from the Google COVID-19 Open Data, which is one of the most comprehensive collections of up-to-date COVID-19-related information. Consisting of data from more than 20,000 locations worldwide, it contains a rich variety of data types from hundreds of data sources including authoritative sources (governments), research/educational sources (universities), general sources (news media and publications), and crowdsourcing (volunteers and other contributors).

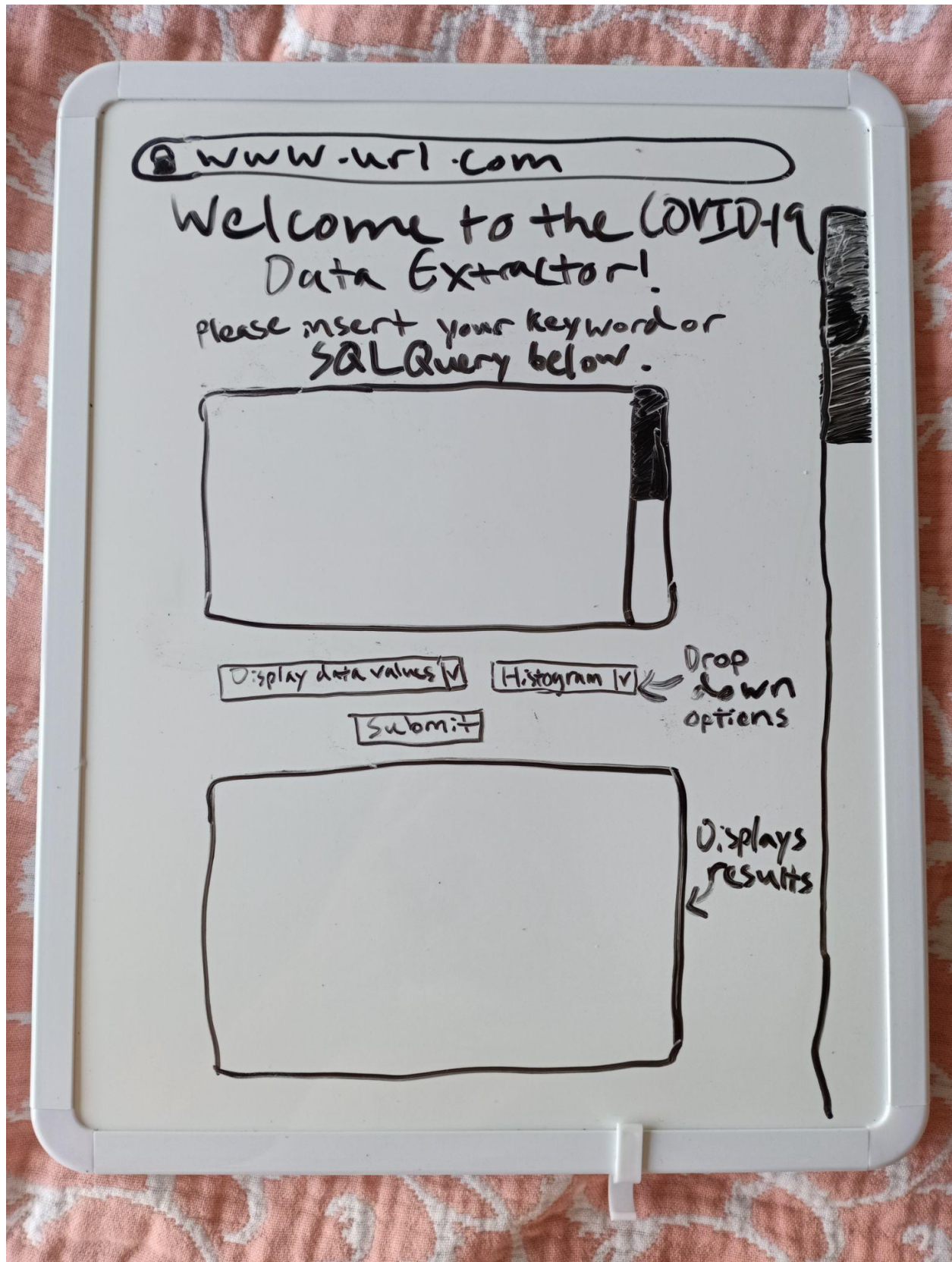
## **Functionality:**

Our database stores essential information about COVID-19, including health information of citizens of various states in the United States, the vaccination states of infected patients, etc. Some attributes that would be stored include economy (country), health indicator of the region, vaccinations, age, and sex.

The primary function of the web application would be to enable users to search by keyword to find matching data stored in our database. Users can also use SQL queries to filter more advanced data. Our website's complex and innovative feature is to provide a visualization tool to display the keyword/query results by map or other graph types.

A good creative component that can improve the functionality of our application is a data visualization option. This would allow users to create graphs like scatter plots, histograms, bar charts, and maps, etc. of the data that they are trying to extract and display from the database. This will improve the functionality of our application by allowing users to have a more comprehensive perspective and understanding of the data they are interested in. We plan to achieve this by incorporating python data visualization libraries like matplotlib and seaborn into our project.

Low Fidelity UI Mockup:



**Project work distribution:**

1. Justin Xiao: Front-end data visualization (which is our creative component)
2. Lucy Jiang: Front-end website development
3. Jian Yin: Back-end Database design and implementation
4. Hanlin Wang: Back-end design and implementation