

Analyzing Intervention Effectiveness During COVID-19 Pandemic

Brittany Roberts, Sebastian Rolett, & Kevin Arnold
DSC 450-T301 Fall 2023

<https://brittlaur.github.io/>
<https://emberanimus.github.io/DSC450/>
<https://kevinqrnold.github.io/>

Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data.

[CDC COVID-19 Home](#) – Central for Disease Control and Prevention's COVID-19 home page.

[CDC Data & Surveillance FAQ](#) – Information on how the dataset being used was obtained

[Trends in United States COVID-19 Hospitalizations, Deaths, Emergency Department \(ED\) Visits, and Test Positivity by Geographic Area](#) – Graphical representation of hospitalizations and deaths due to COVID-19

[World Health Organization COVID-19 Home](#) – Information regarding WHO's COVID-19 guidance

[Timeline: WHO's COVID-19 Response](#) – timeline of WHO's response guidance

Which Data?

What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

<https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4>

We will be examining the Center for Disease Control and Prevention's COVID-19 case surveillance database. According to the website, "COVID-19 case reports are routinely submitted to CDC by public health jurisdictions using national standardized case reporting forms."

Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.

Our analysis will include identifying trends around chronological, geographical, and demographical information contained within the data. The intent is to show the timing and impact of the introduction of public health safety measures and to hypothesize further trends leading to exposure, hospitalization, and even death had those safety measures not been rolled out.

What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

Our approach will be creating a supervised learning regression task to predict the levels of further impact the COVID-19 pandemic would've had on infections, hospitalizations, and deaths had safety measures, including vaccines, not been developed and introduced.

We will first perform exploratory data analysis to summarize and cleanse the data to be able to present it in a useable format. We will then create an initial model to predict our target which is the numerical feature of number of cases.

Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

The dataset, as of 9/20/23, contains over 100 million entries. The size and scope of the raw data may present system capacity issues when attempting to perform wrangling and cleansing operations. The raw data file itself took 15+ minutes to download from the CDC's website which was the only option to obtain it due to restrictions within the provided API.

Concluding Remarks

Tie it all together. Think of this section as your final report's abstract.

As mentioned in the 'What Method?' section above, our intent is to hypothesis the level of impact the COVID-19 pandemic would've had on the United States population had control and prevention measures not been implemented by local, state, and federal authorities. Comparisons of existing positive exposure rates, including hospitalizations and deaths, will be made to statistics resulting from models created to show how much of a greater impact the pandemic would've had.

Variables like location, age, and sex of those infected will be factored into our analysis to show which clusters of individuals based on certain characteristics would've been more adversely affected had no measures been put in place. The impact of vaccination on the risk of mortality related to COVID-19 will play a major factor in our analysis as well. Other measures such as mask mandates, public closures, and hospital availability will also be considered during our modeling in order to understand what multi-variate factors influenced the existing data as well as our model's outputs.