

INFO 330 Project Proposal

Database application we plan to make:

We plan to build a database application that tracks details about COVID-19 by demographic through patient data, vaccination status of patients, and mutations of the virus in each country in the world. Our application will combine globally compiled data on the spread, health impacts, and prognosis of COVID-19 to summarize expected outcomes for a demographic (age, sex, or ethnicity) based on available information (country, vaccines available). We are trying to highlight where and for what demographic the prognosis for COVID-19 is worst, so help can be directed there.

Target outputs from database application:

We plan to market our database to the WHO, an organization formed by the UN in 1948 to promote the wellbeing of people around the world. The WHO has an ongoing initiative to provide universal health coverage globally. In times of emergency, like the pandemic, the WHO was essential in distributing medical resources to those in need.

There is no doubt that some demographics are more susceptible to COVID-19 than others. This is true for the pathology of all diseases, not just COVID-19. The reasons that some demographics are at higher risk comes down to anthropological, cultural, and political factors of the place they live – essentially, it is a community problem, not a biological one. According to the Mayo clinic, significant differences in positive case rates between groups could be attributed to factors like comorbidities (having concurrent diseases worsens prognosis), lack of access to healthcare, national/local poverty or upheaval, treatment stigmas relating to religion or politics, and type of work the person does (in person workers were more susceptible), just to name a few.

With the pandemic now over, vast amounts of data are available on the case rates per country and the performance of the world overall. It is true that each country in the world is at a different place in their development index. And there is more than one reason why that is the case. So it is unrealistic for us to paint a perfect picture of the factors that affect each country in the world. What we can do is determine where significant differences exist between demographics in recovery from COVID-19. These significant differences are indicators of social, cultural, or medical barriers that these demographics face in healthcare, which may still exist post-pandemic.

Our database will compare patient data from various locations around the world and determine which ethnicities, genders, and age groups were most affected in each country for which data is available. In this article by the NIH, it is mentioned that new data science methods are breaking the ice in treating COVID-19, and tools like it are needed in the future. Problems of most relevance are finding successful contact tracing methods, analyzing global response to COVID-19, assessing economic impacts, mining patient data, mining scientific literature, and

mining social media responses. We aim to address the subject of analyzing global response by mining patient data. Our database will deliver statistics on patient health, but not the anthropological context of those statistics, which would require further research outside the scope of our database.

Main entities:

- 1) Variants of COVID-19
- 2) Countries with COVID-19 case data
- 3) Vaccines in use
- 4) Patient cases
 - a) In the US
 - b) In Bangladesh
 - c) In Mexico
 - d) Globally, with comorbidities

INFO330 A

12/26/25

Citations:

Business Needs:

Demographic Susceptibility:

<https://www.mayoclinic.org/diseases-conditions/coronavirus/expert-answers/coronavirus-infection-by-race/faq-20488802>

NIH Business Needs:

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8607150/#s4>

About the WHO:

<https://www.who.int/about>

Datasets:

Global Case count:

https://www.kaggle.com/datasets/imdevskp/corona-virus-report?select=full_grouped.csv

Variants of COVID-19:

<https://www.kaggle.com/datasets/lumierebatalong/covid-19-variants-survival-data>

Vaccinations by manufacturer

https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress?select=country_vaccinations_by_manufacturer.csv

Vaccinations by country

https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress?select=country_vaccinations.csv

Income support by country

<https://www.kaggle.com/datasets/pranjalverma08/coronavirus-covid19-indepth-dataset?select=income-support-covid.csv>

Mexican patient details

<https://www.kaggle.com/datasets/meirizri/covid19-dataset>

Patients with comorbidities global

<https://www.kaggle.com/datasets/shirmani/characteristics-corona-patients>

Bangladesh Patient symptoms

<https://www.kaggle.com/datasets/takbiralam/covid19-symptoms-dataset>

US Patient details

<https://www.kaggle.com/datasets/arashnic/covid19-case-surveillance-public-use-dataset>