

# FINAL PROJECT – DSC540-T302

## SUMMARY

Data is critical to support countries in managing the global coronavirus (COVID-19) pandemic, I've learned a lot from this project. I started off with my csv that I pulled from Kaggle and that was fairly easy because up until now I had worked consistently with csv files using pandas. The third milestone had us web scraping which took a lot of time figuring out what the table name was then getting it to pull in correctly. I had never web scrap using python, and this was a bit challenging because I tried looking through the books and google and was having problems finding a good answer. Turns out one of my lists needed to be nested inside of another list and once that was done it became much easier to work with. All I had to do was switch some N/A values into 0 and then remove rows that had pulled from other tabs in the table from the webpage. The fourth milestone also took some work to pull the data as I tried several different ways to call the api. I ended up asking for help on stack overflow and someone provided an example that worked for me. I also found out that I could make it a lot easier if I tried something from your suggestions to one of the students (saved me a lot of time). To fulfill milestone 5, I went back to each dataset and saved it off as a csv to start with. I pulled those csv's into this file but found some of the cleaning I did on them originally didn't stick so I had to re-clean a few things before combining. I tried merge but ended up with a ton of different columns and since my first two csvs were practically the same column-wise I did combine first using my first one as my source of truth. I went ahead and select to use SQLite for this project and figuring out how to create a table. Carla and Riley posted a link to a tutorial that helped you do it via command line, which was great because I am comfortable CLI. Thankfully this was easy to follow so I was able to create my database then import my csv as my table. I pulled that all back into this file then worked on my visualizations and saved plt images locally. I had a few issues and hiccup with these and ended up reverting back and forth on saving and deleting version I've created because they were getting extremely distorted. This is an area where I need some work as I have a lot of python basics down, but I feel like I could improve in this area. Overall, this class and this project have challenged my python skills and how I think about my data. It has made me realize why so much time is spent in this area and given me a lot of ideas on where I could improve. I also think that I can improve more with api(s) and web scrapping with python.

## 5 VISUALIZATIONS





