

Karmic Tuner - PM6
 Mark Meuse, Nic Shepard, Ture Carlson
 CS5200 - Database Management Systems

At the start of this course we set out to aggregate data on Covid-19 cases with other county based datasets to make a tool that could take discrete sources and combine them into something greater than the sum of its parts. Behind the backdrop of statewide lockdowns, we imagined a tool that could be used by policymakers and public health officials to map the ways people were getting sick, how they were moving, and how policy decisions like stay at home orders tangibly affected the same. Specifically, we defined our product with this value proposition and these goals:

Value Proposition

“MOVD is a public tool for those responding to the pandemic who are overwhelmed with information and who are looking for a simple tool that provides actionable insights into the relationships between mobility and Covid-19 deaths, provides a localized view of the impact and response to the virus by county, and the effectiveness of legislative measures to decrease the rate of infection by restricting mobility and enforcing social distancing.”

Deliverables:

- *Given the date of a policy action provided by the user, display the impact it had on a population's mobility and infection rates for the next two weeks.*
- *Given a location such as a county, return the history of infection.*
- *Analyze localized impact based on demographics such as urban-rural ranking, population age, race, and socio-economic factors.*
- *Perform CRUD operations on the database*

Overall, the product that we have built has stayed true to what we originally intended it to do. *MOVD* provides meaningful insights for the stated goals and has even evolved to incorporate more political and economic queries that have become relevant as the pandemic becomes increasingly politicized across the American landscape.

While there are many things that went well for us throughout this project, such as the preponderance of reputable, well maintained, datasets, the ways that we helped each other grow as computer scientists through bringing complementary skill sets and approaches, and the excitement of trying to understand one of the most defining moments of our generation, we did make many changes to our project due to structural challenges. For instance, we had these stated goals for expansion in PM1:

We hope to expand our model to include new ways to relate Covid-19 infection rate with:

- *Data outlining actions taken by federal, state, and municipal governments*
- *Data outlining prior underlying medical conditions in conjunction with Covid-19 cases*
- *More precise demographic data*
- *Expansion of data to include sources beyond the United States*

To begin, we were not able to find the datasets we had hoped for actions taken by local or state governments. While we came across one, federally based, dataset, it was neither rigorous nor comprehensive enough for our purposes. Because of this, we eventually got rid of both our User and Event tables in PM3 once we understood the difficulty of finding this data and the scope of our front-end framework we were to build. Similarly, both structural challenges of finding data and the increasingly political direction of our project led us not pursue details about the medical records of patients and to take a US-centric approach as we delved into the data county by county, state by state. For this reason, we also added tables for S&P 500 data, Unemployment Data , and President Trump’s approval ratings.

While there were a lot of things that were very positive about this project, if we were to do it again we would like to create mechanisms to automate information retrieval, cleaning, and loading into the database. While we made it a point to do as much cleaning as we could in SQL, challenges of loading our data often forced us to have to make changes to the csv files themselves by using command line tools and Python. If we had the resources, these tasks would have lent themselves very well to a few simple scripts that would ensure the relevance of the data as cases across the country spike but our data remains too old to analyze it. In addition, currently we only have two scripts that create and insert data. If we were to do it again, we would add a third staging table script to better organize our data and reduce lengthy loading times for our many rows of data.

While this project will likely be retired after the semester, we are grateful for the opportunity to have earned the hands-on experience that only countless hours struggling with MySQL can provide. *MOVD* remains fresh in our minds as Mark tests the software for a medical company that builds ventilators, Nic settles into his new job as an Application Engineer for a software company that supports clinical trials, including those for Covid related research, and we all search for answers on the best way to get through this.

Final EER Diagram:

