

One of the biggest challenges we faced during the project was in the data selection phase in making sure that after data cleaning there was still enough data to be able to draw conclusions; the most prevalent example of this was filtering for time. We wanted to make comparisons across specific years in our datasets, but finding datasets that included the same range of years was difficult. We ran into issues early on, datasets that initially contained a sufficient number of observations became too small once we filtered them by year. Overall, we chose to reevaluate our topic and instead base it on the data available to us, and introduced datasets that did not align with our initial topic but also did not present the same timeframe challenge.

Once we landed on the final datasets, during the data analysis stage, it was difficult to narrow down which specific variables we wanted to focus on. We knew that we wanted to include the mental health dataset that had statistics for multiple different disorders, but we had a myriad of demographic factors to compare them to. We initially left the dataset broad for data analysis, thinking that if we analyzed more factors, we would be able to narrow them down through our exploration to those that appeared to be correlated. However, this did not work as planned as the number of variables made it difficult to coordinate with other team members to ensure that our project was streamlined and cohesive. Since we were all looking at different factors for analysis, we had to narrow it down after all of the analysis was done and could have saved a ton of time in this stage by cutting down our variables beforehand.

The biggest lesson we learned was just how integral comments are when working in a large group with only one shared github repository. Using descriptive and concise comments in the code made it much easier to follow and we were easily able to understand what was being done in each chunk. This was a really useful tool since we would often be working on sections of the project individually, but with comments it was easy to follow the code and understand what stage of the project had been completed. Comments were also beneficial because we all have different coding experience. With only one member being a computer science major, being able to follow each other's code allowed us to learn new things and have a deeper understanding of how best to code.

One technical challenge we experienced was writing code that could handle updates to the original datasets. Another challenge was connecting and saving to MySQL. A few members of our group had some problems connecting to MySQL which made it difficult to run the ETL pipeline and perform the analysis. We addressed this challenge by pushing a change that saves

and downloads the data from the cloud. This allowed all of our group members to download the data on their machines and continue working on the project.

For future projects, narrowing down the topic and choosing datasets before beginning analysis would be a huge improvement and save a lot of time. We started out with a general idea that we wanted to investigate the correlations between mental health disorders and demographic factors, but our topic was more broad and open to evaluating almost any relationship. After initial data analysis we chose to evaluate the relationships that we believed had the most significant implications. It would have been more ideal to have a more narrow topic and choose datasets that would be useful. However, considering that we were limited by number of observations, year range, and available datasets, we believe that we did the best we could given the challenges we encountered.

Together we learned how to effectively communicate as a group, including setting deadlines for each project stage and ensuring that everyone was in agreement. Most of the project was completed outside of class, so learning to communicate effectively allowed us to be able to work on the project individually while keeping each other informed about what was completed and what still needed to be done. We were also able to discuss any challenges we had in working with the datasets and developing the ETL pipeline. Since we were working on the project for a few weeks, establishing expectations was helpful to make sure we stayed on track. Aside from just learning how to communicate as a team, we also learned how to work as a team. We were able to communicate to understand each other's strengths and weaknesses. We all have different coding backgrounds, and it was helpful to delegate tasks and discuss who wanted to take on each role during the project. This was also helpful in ensuring that each member was doing their fair share of work, since we knew who was assigned to each task. Keeping an open line of communication was vital for the completion of our project.

As previously mentioned, there were multiple variables in our datasets that we did not dedicate much time and analysis to. Expanding the depth of the project by exploring the correlation between mental health disorder rates, such as schizophrenia, and global demographic factors, such as fertility rates would be a useful expansion of this project. We could also take a modeling approach to the project and examine the correlation between multiple factors at once, using one variable as the given dependent variable. Since there are so many categories in our datasets, there is a lot of potential for detailed expansion.

Communication is a skill that can always be improved upon, especially when working with a large group of people, as is being open-minded to others ideas and opinions. These are skills that we can bring into our future careers because we will always be interacting with other employees and will most likely experience working in a group at some point in our careers. Being able to respectfully state your opinion or idea and listening to others is a valuable trait in a team member, and one that we developed as a group while working on this project. We also learned how to receive and give constructive feedback which ultimately will help us personally, professionally, and help us create a better project.