I shared the circle roll with Mel which involves preprocessing our data and creating the database. Once this roll was complete, I assisted in constructing the Google Slides final presentation. On the ReadME you will find a comprehensive breakdown of the preprocessing and database building. The greatest challenges of the rolls were: initially finding enough cohesive data that has enough similar variable to merge, in cleaning up the data we ran into numerous issues we needed to further research to resolve i.e., learning the TRIM function on postgress, and some members of the team were out a few days for various reasons.

Our team overcame the initial hurdle of finding datasets by focusing on what “issue” or “problem” we wanted to solve in our project. Once narrowed down, we were able to focus our search on a specific topic which opened resource possibilities such as going straight to California government websites. The greatest strength of our team and the key to cohesive teamwork is communication. Our team was active on slack in chats to assist teammates that needed help, Mel and I would zoom call to work through our roll together, and we used our time in class wisely to focus our goals for the week to ensure our project was progressing for each deadline.

Our team proposed for portions of California’s Public Health and General annual spending to be used to make clean energy and/or environmentally conscious infrastructure changes to reduces the risk of asthma in California residents. We have used a machine learning model to investigate why certain California counties may be above (True) or below (False) California's collective asthma rate of 8.8% and what environmental factors may contribute, with the assumption that black communities were above the asthma rate and counties with more electrical vehicles would be below the asthma rate. The first model used was BalancedRandomForestClassifier was chosen because its use of boolean values provides an accurate score consistently. The second model used was Neural learning which was chosen because it can learn over time to be near or completely accurate after X amount of epochs. We found that black communities were more likely to be in above average counties but we also found that clean energy does not effect the asthma rathe. We suggested expanding the parameters to includes other clean energy sources, which will also improve the machine learning model as one of the “issues” is we did not have enough data for the machine learning model to train properly.