Report Stage 3

Stage 3 was about machine learning and regression models. As stated by the question, I calculated the linear regression, polynomial regression and rmse error. I created a function named linear regression that helps me print all the needed data and graphs. I am writing images on a folder and showing them so that the viewer can see it on Github. While looking through the graphs, we can see different best-fit lines and different prediction for cases and deaths of different geography. For the counties, two function named prediction_cases and prediction_deaths are there to help them create all the necessary dataframe and calculate the linear and non-linear models.

For the point of no return, the state of Washington has a lot of beds and with the ongoing cases right now and the prediction, Washington is not going to reach the point of no return any time soon. Even in its worst time, Washington had 2500 extra beds. I also had a 'Difference' column in the dataframe 'merged_df' which shows the difference between the capacity of the hospital in that state and twenty percent of total cases in two weeks. No state had a negative data so none of them are going for point of no return any time soon. (For the total capacity of the beds, the number of licensed beds and potential capacity were added.

For my hypothesis, I got p-value less than 0.05 and I rejected the null hypothesis. My hypothesis is stated in the Jupyter notebook.

The prediction line's part was fun for this project. Predicting what could happen was something nice. Applying machine learning in this stage was very knowledgeable.