

Progress Report 1

Tyler Fusco, Mari Leonard, Gabrielle Berasi

Current Status:

At this stage of the project, we have completed a lot of background research regarding the project proposal and familiarized ourselves with the dataset we will be working with. We have identified what each of the variables refers to and have been hypothesizing about which variables will be important to our machine learning model. We have also identified possible roadblocks in the data set, such as missing data entries, and have been working on identifying the best way to deal with these missing entries to maintain the accuracy and effectiveness of our models. We have also individually given presentations on research papers that conducted similar projects to ours. We have also started to use the data to perform some preprocessing and exploratory data analysis.

Future Plans:

The next steps for the project are going to be developing and tuning the machine learning models as well as creating visualizations for our outputs. This will be a team effort, and we are probably going to divide the machine learning models amongst ourselves, so we each develop one or two models. We also will need to generate visualizations for our models for our presentation. We will also need to prepare for our final presentation and for our upcoming client meetings. We plan to meet either in person or over zoom to make sure we stay on the same page, and update the GitHub as well.

Potential Issues:

Some potential issues could be the previously mentioned missing entries in the dataset. We were told that the client may have some specific ways he would like us to deal with these missing entries, so we will have to tweak our models to align with his specifications. Until then, we will have to pick our own way of dealing with them, whether that be removing rows with missing data, plugging in median or mode values, or something else. Another issue we have faced is the possibility of overfitting the data since the dataset is so small.

