**Self-Assessment**

I was responsible for setting up the repository which includes naming the repository and adding other team members. I worked with every team member to ensure that everyone has his or her own branch to work from. I always ensure that the main branch is free from code in progress and that all ongoing work is kept on the branches separate from the main branch. I monitored the repository on github to keep only tested, working code, READMEs, and other documents in the main branch. I was also responsible for writing the READMEs for each segment of the project and the final README for presentation. I participated in other roles of the project such as working on the machine learning model and the presentation.

**Team Assessment**

We work well together on the project as a team. Our communication protocol includes discussion about the project in breakout room during class and weekly zoom meetings outside class to work on deliverables and to discuss different aspects of the project. We keep ourselves updated on the progress of the project on Slack as well. We had a challenge during preprocessing of our model which we sought help from the LA from and we finally employed Random Forest feature importance to process the data for feature of importance to drop redundant and unimportant columns. The strength of our team is good communication. We communicated well which helped in coordination of the whole project and beating the weekly submission deadlines.

**Data Scientist Salary Analysis:** Python and Pandas in Jupyter notebook were used to create a data frame to implement a machine learning to predict how much salary a data scientist would get paid based on certain skillsets. Linear Regression model was used to create training and testing groups from the dataset and data exploration was conducted which includes visualizing the data, getting the shape of the data, checking for missing values, and using the describe function to get the overall statistics of the data. 70/30 method was used to split the data for training and testing and the test set result was scored. Random Forest Regressor was used to select all the features of columns that have highest correlation with the target and Feature importance was used to remove all unimportant columns. The Linear Regression machine learning model provided some insight that certain skillset might impact the salary a data analyst would get paid.