Relax Data Science Challenge:

The goal of this project was to determine an ‘adopted’ user based on other features. An adopted user is a user that has logged into the product on 3 separate days in at least one seven day period. Once I found the adopted users, my next step was to do some exploratory data analysis that helped me figure out some relationships between the target variable ‘adopted’ column, and the other features. When looking at the variances and covariances between the features and the target variable, there weren’t any strong relationships between them. What I did then was I did some feature engineering. My first new feature was creating dummy variables from the creation source column as well as creating a new column that was based on the difference between last_session_creation_time and creation time. I called this column, days_from_creation. This new variable had a .81 correlation with the target variable. I took all my numerical data and used it as the input to run my learning algorithms. My best model was when I was using the RandomForestClassifier, which gave me an accuracy score of .968. The best feature importance was in fact the new feature I created, days_from_creation which showed a strong correlation when doing exploratory data analysis. For more information, my code is in the github repository.