Springboard Project 20.1A (Ultimate Take-home Challenge)

This project is a sample take-home assignment designed as practice for the interview process for data-science jobs. The aim is to identify factors that predict user adoption, where user adoption is defined as logging into the product on three separate days in at least one seven-day period.

I start with data wrangling, which entailed the removal of an unneeded column and checking for missing data. From there, I reformatted the time stamps in the data frame so that I could perform the necessary wrangling to count the number of adopted users. I then added a binary column to indicate whether a user was adopted or not. This column ultimately became the dependent variable in analysis. The final wrangling step was the merging of two datasets to unite independent variables with the dependent one.

Next was exploratory analysis. I explored time-series data in monthly intervals and found that there was an overall increase as time went by. Users were also clustered into a relatively small number of organizations. (Organization ID was another independent variable.) However, I did not have contextual information about organization IDs available, so further conclusions were hard to make. Finally, I checked the balance of the several variables and conducted one pre-processing step – one-hot encoding – was also performed.

For analysis, I used a logistic regression. This type of model generally works well in classification problems like this one, and it also provides coefficients for each variable. These coefficients help identify which predictors have the most predictive value. After building an initial model and checking global accuracy, I tuned model parameters and checked precision and recall to evaluate performance. With those steps satisfied, I produced the variables (and their coefficients) that had the most predictive value with respect to customer adoption.