Bank Marketing Analysis

By Steven Jasper

Good morning. My name is Steven Jasper and I will be presenting the results of my analysis on the bank marketing data concerning our Term-Deposit Subscription campaign.

Problem Statement

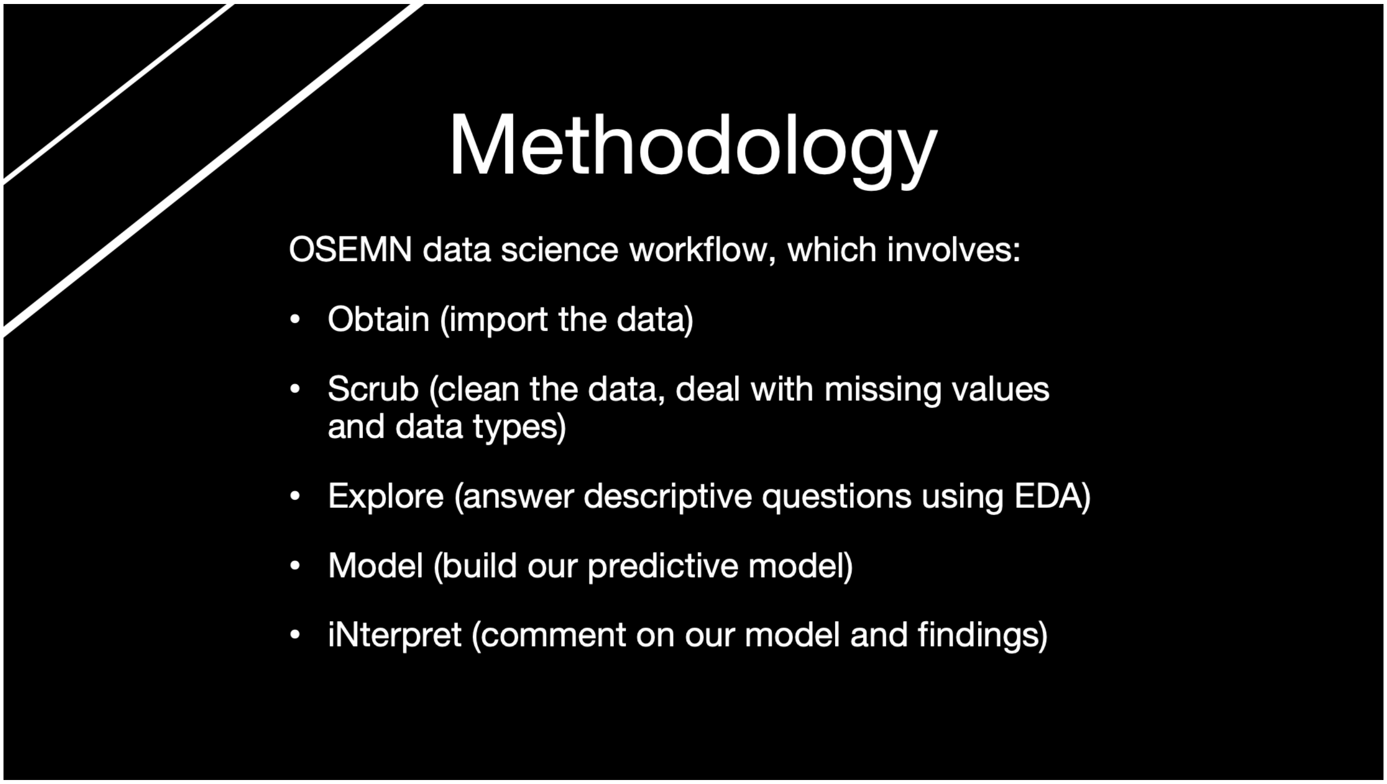
What type of customer is more likely to subscribe to a term-deposit?

Our problem statement is simply; are there any distinct characteristics that make a customer more likely to subscribe to a term deposit? I will explore this first by taking a look at our cleaned data, then we will get into how we determined significant features using machine learning.

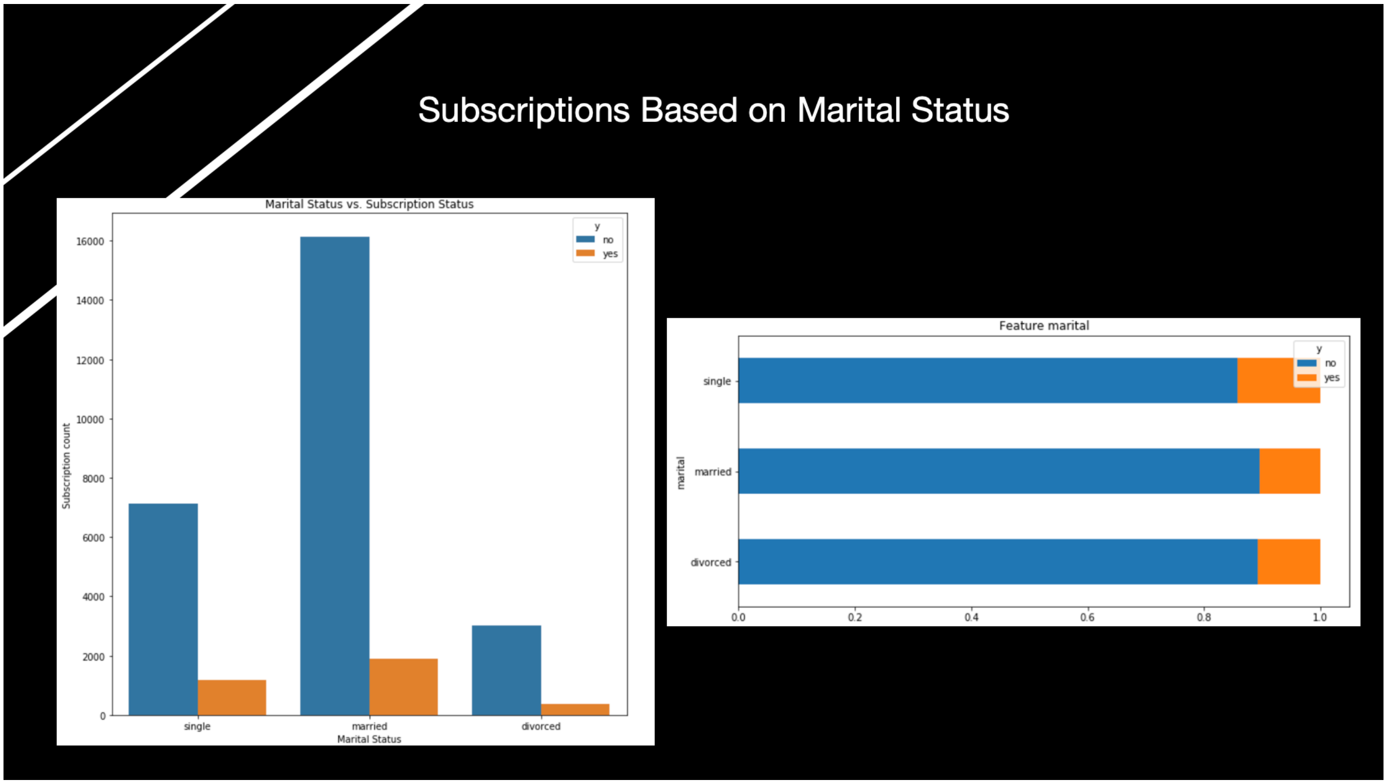
Business Value

- Increase revenue from term-deposits
- Better determine what customer to spend time on sales
- Create reusable pipeline to increase confidence in sales predictions

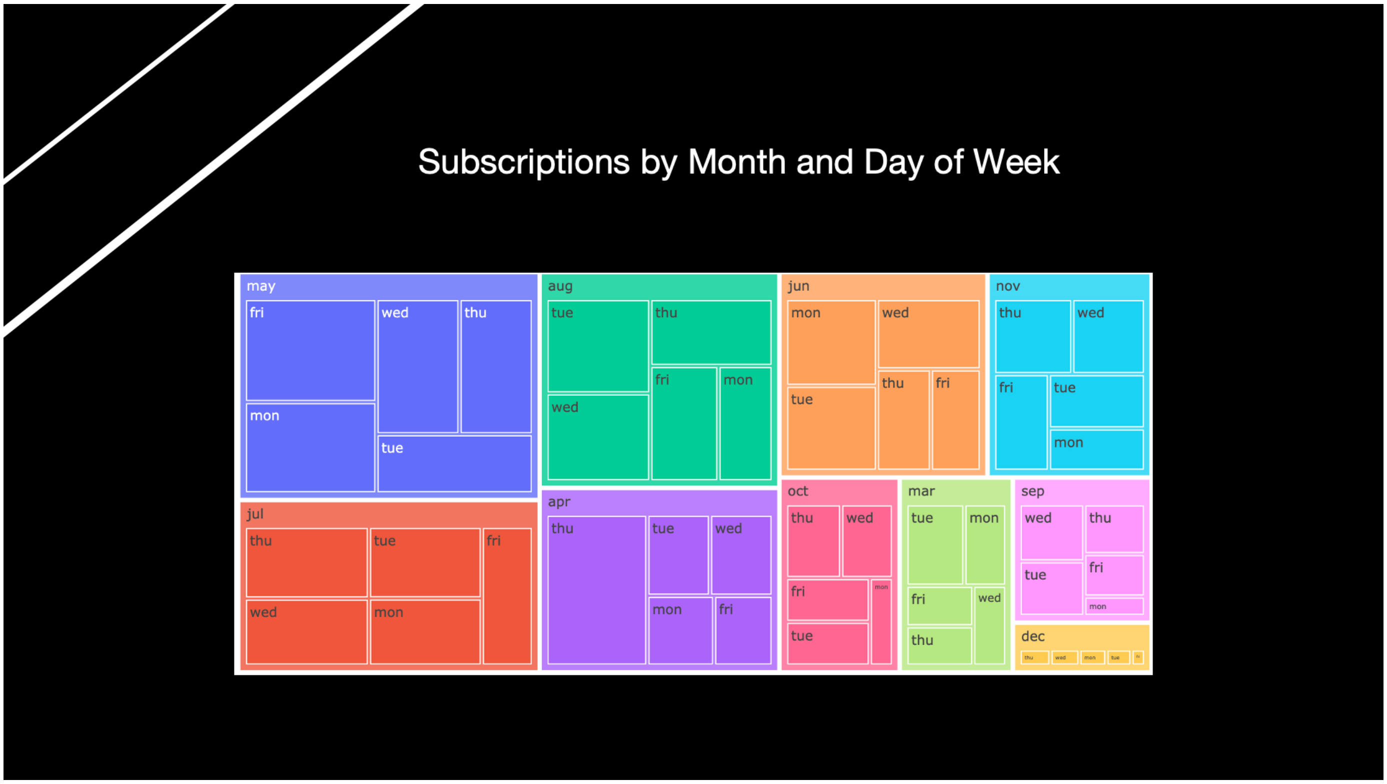
The business value here is simple, we will increase revenue from term-deposits, determine where and when to focus our campaign efforts, and create a reusable system that can be used to more accurately predict outcomes with additional data.



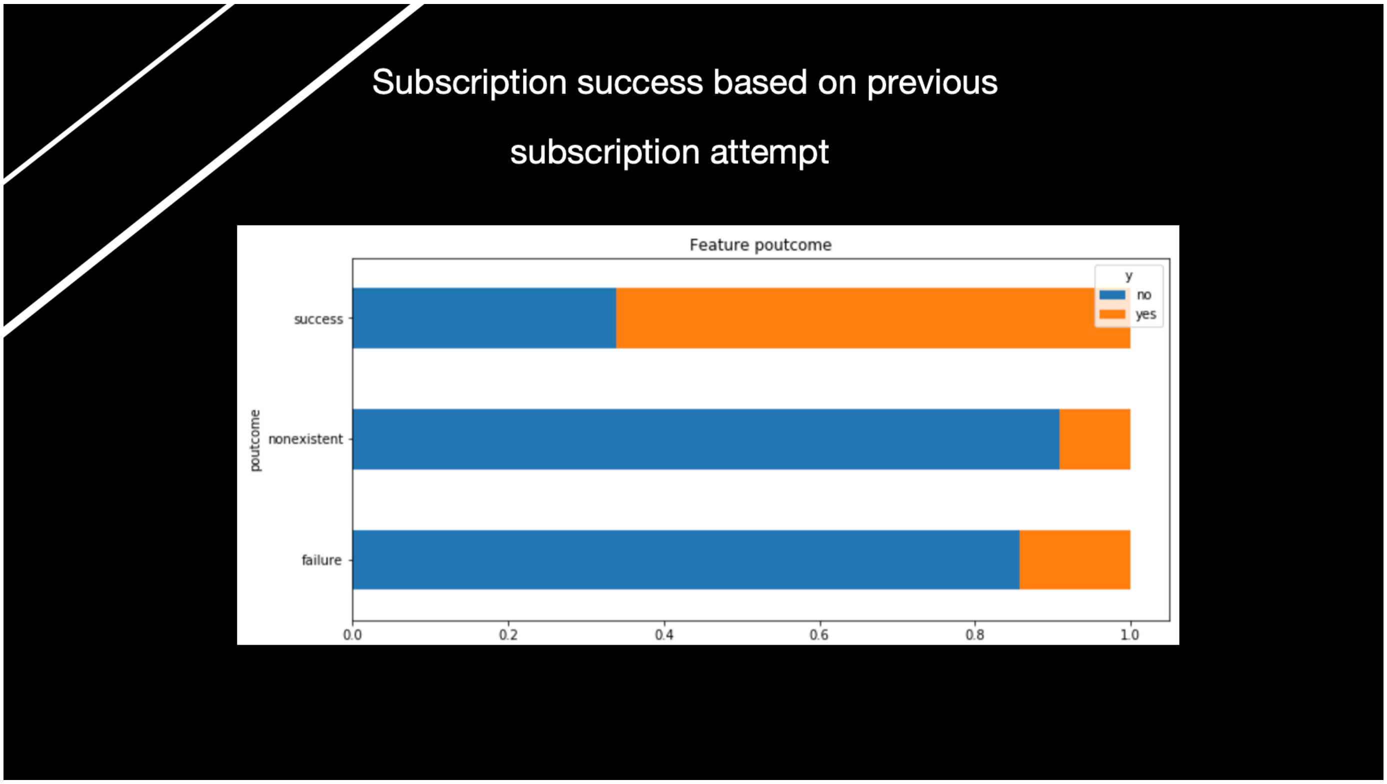
I utilized the OSEMN data science work flow for this analysis. I began by obtaining the data, once the data was obtained I scrubbed, explored and began modeling our data. Finally we took outcomes from our model and interpreted these to produce our findings.



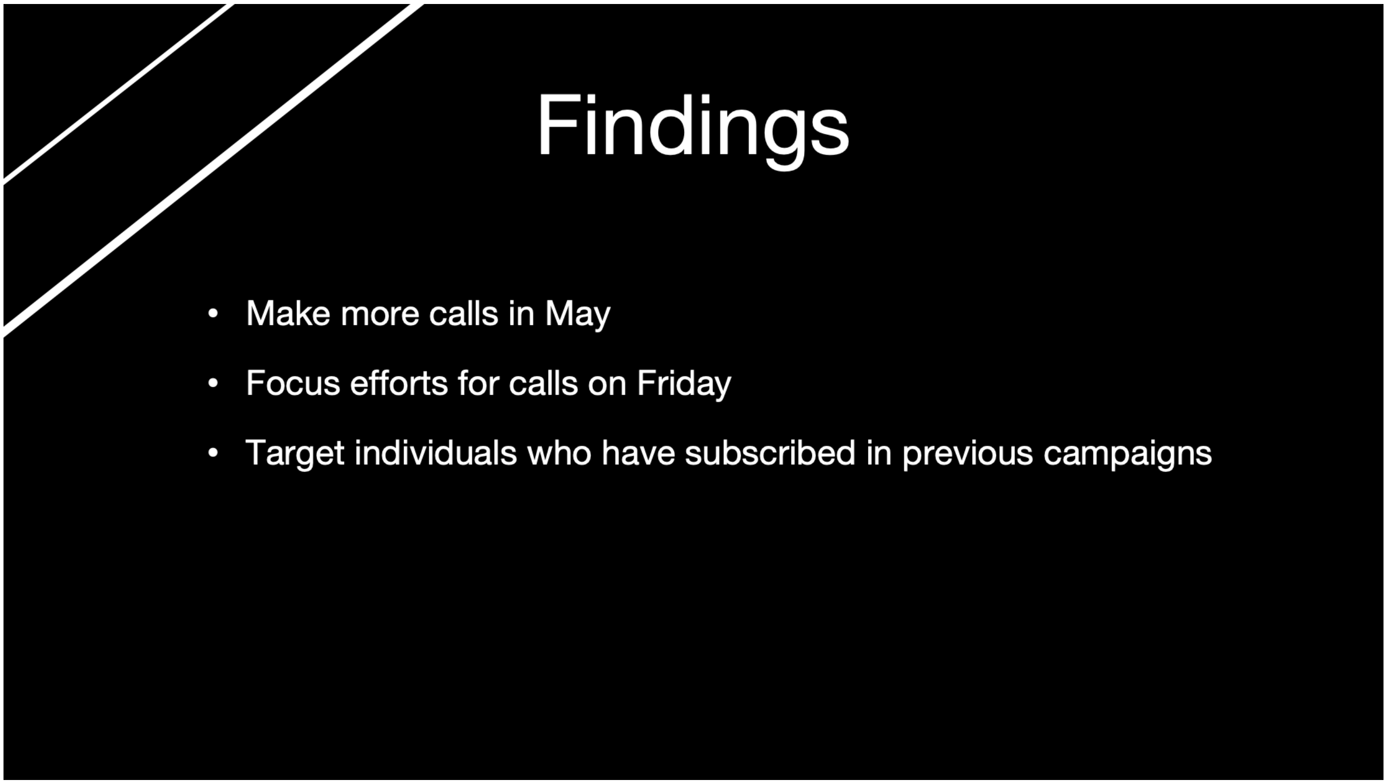
These graphs illustrate our subscriptions based on our marital statuses. The left graph shows the data, but also shows a clear case of class imbalance, therefore we include the right graph to put this more into perspective to show that out single customers were more likely to subscribe.



This Tree Graph shows the subscriptions by month and day of week, the largest box is May which represents the fact that May had the most subscriptions taken out. We also see that Friday is the largest inside of our May box. This graph displays each month/day combination.

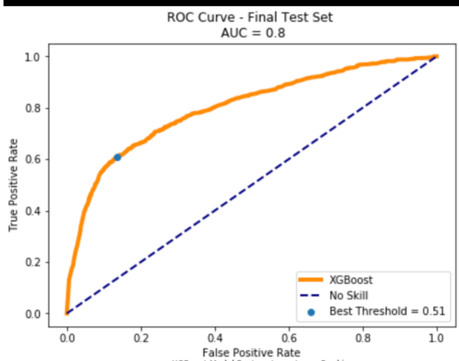


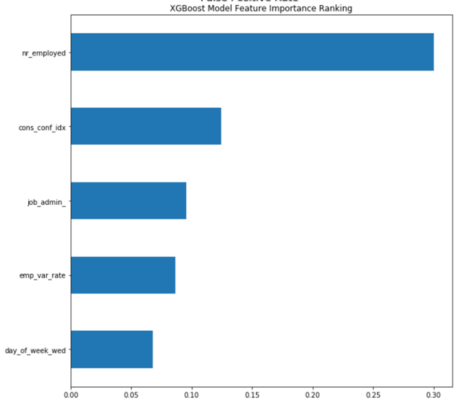
This graph signifies our subscription rate based on previous subscription attempts. We can see that it is pretty clear that individuals who have taken out a term-deposit in a past campaign are much more likely to take out another.



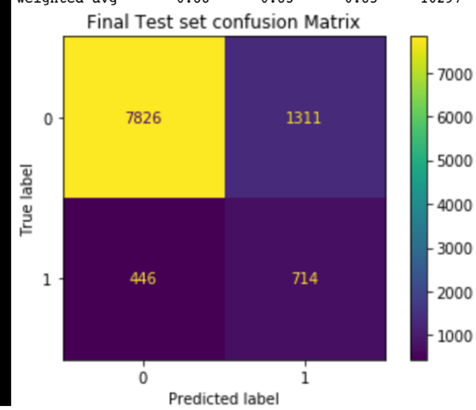
In order to have the highest chance for an individual to agree to open a term-deposit subscription we would want to make more calls in May, call on a Friday, and ensure that they have subscriped in previous campaigns

Model Performance

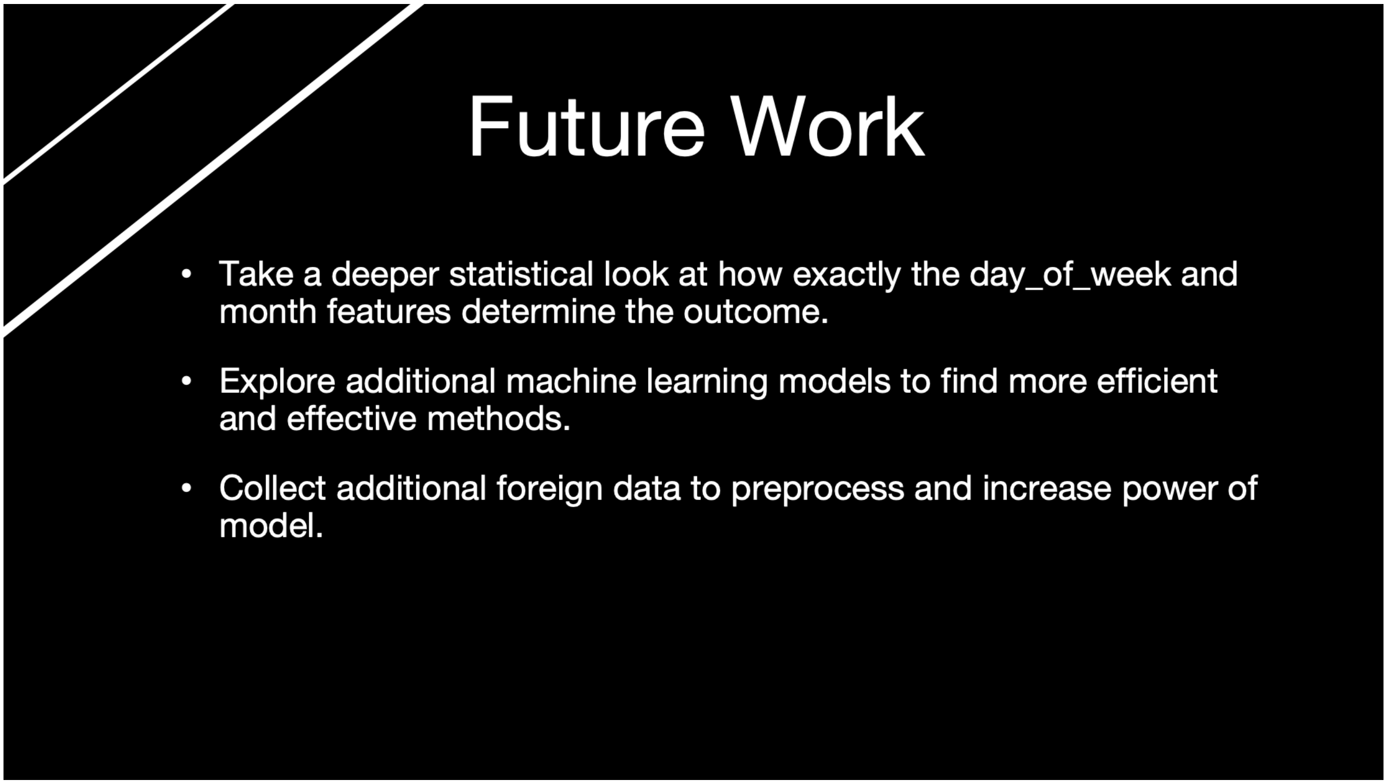




	precision	recall	f1-score	support
0	0.95	0.86	0.90	9137
1	0.35	0.62	0.45	1160
accuracy			0.83	10297
macro avg	0.65	0.74	0.67	10297
weighted avg	0.88	0.83	0.85	10297



These graphs show different metrics of success for our XGBoost model, out of the 6 models tested the XGBoost had the best statistics and efficiency among them. The top left image you can see our ROC curve which the dotted line shows how well the model would be at predicting completely randomly, whereas the orange line is where our model performed, while not over fitting our data we were able to produce this ROC curve on a ‘second test set’ which preformed quite well as you can see on the top right. The aim for this analysis was to minimize the number of False Negative which means we would like to have a higher recall value. Our recall value for our secondary test run was .62, quite good. Below you can see the breakdown of our True Positives, True Negatives, False Positive and false Negatives. In regards to significant features the graph in the bottom right shows which features XGBoost determined were significant in the model.



Work I would like to complete in the future is to conduct a deeper statistical look at how exactly our month/day of week feature affects the outcomes, I would also like to explore other options for classification machines for this data set. Collecting additional data to preprocess and increase the power of our model is an obvious take away.

Questions?