1. To clean the data, we had to fill the NaN values for a few columns. The columns with NaN values present are: `avg_rating_by_driver`, `avg_rating_of_driver`, and `phone`. By looking at the summary statistics for the average rating columns, we see that the mean is not too far away from the median and that the standard deviation is low. We thus fill NaNs in those columns with the mean of that column. For phone, we fill NaN values with the mode, which happened to be IPhone. We drop the date columns and define the `retained` column for predictions. We are now ready for modeling
2. For this project I used a Random Forest Classifier. Random forests algorithm is among the most popular machine learning methods thanks to its relatively good accuracy, robustness and ease of use. It is also often used for feature selection. I considered other ensemble algorithms like gradient boosted trees as well and further testing could be done to pin down the best algorithm.

   First I split the data into 80% training and 20% testing. I used 5-fold cross validation with the base parameters of scikit-learn's random forest classifier and our training set to estimate the accuracy, which was 0.76. I then trained the model with the training set. I tested the model with the training set and I then used scikit-learn's classification report to get the classification statistics. Overall we achieve 0.75 average precision, recall and f1-score for both classes, with the positive class, i.e., the user is retained, having the lowest scores in all categories, indicating it's more difficult to predict retention than no retention. Our model achieved an AOC of 0.8. Finally, we look at the feature importances of our model. The top features in descending order are `avg_dist`, `weekday_pct`, `surge_pct`, `avg_rating_by_driver`, and `avg_rating_of_driver`. One concern with the model is that we performed no hyperparameter tuning, which would certainly improve the performance of the model and is a valuable next step in model tuning.
3. By looking at the feature importances, I suggest Ultimate:

   1. Attract riders to take longer trips during the first 30 days after signup.
   2. Provide better service to get higher ratings from the riders.
   3. Persuade riders to take trips during weekdays.