Part 3

1. 69.22 % users were retained.
2. The problem is to predict if a user will be active. It is a classification problem. I used to logistic regression and random forest model to make such prediction. The accuracy score of logistic regression is 69.224% and the accuracy score of random forest model is 82.232%.
3. The first two important features are average distance and the most of last trip date. It means the shorter average distance, the more likely that the user will be retained and the month of the user’s last active date is closer to current month, the more likely the user will retain.

Suggestion: Ultimate can provide more nearby customers to rider in order to simulate rider to be active.