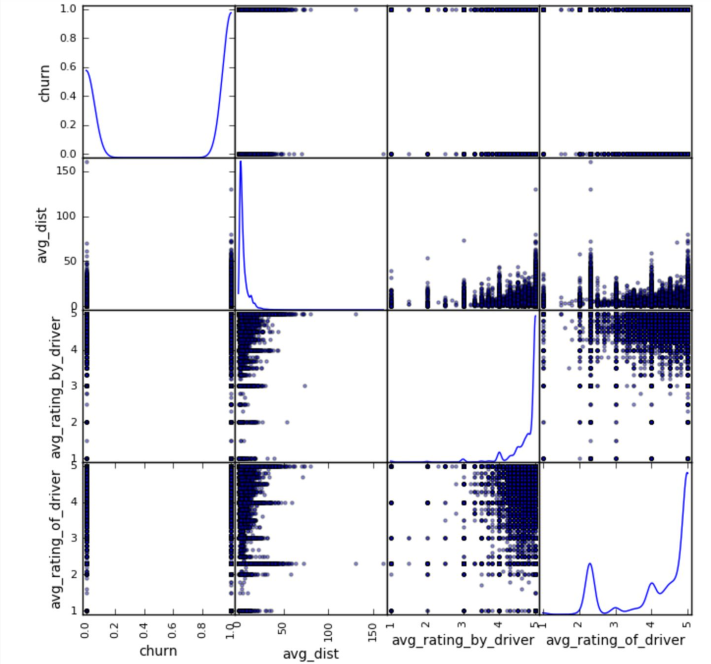
Churn Prediction Model

Chris Seymour, Chris Van Dusen, Hao Wei

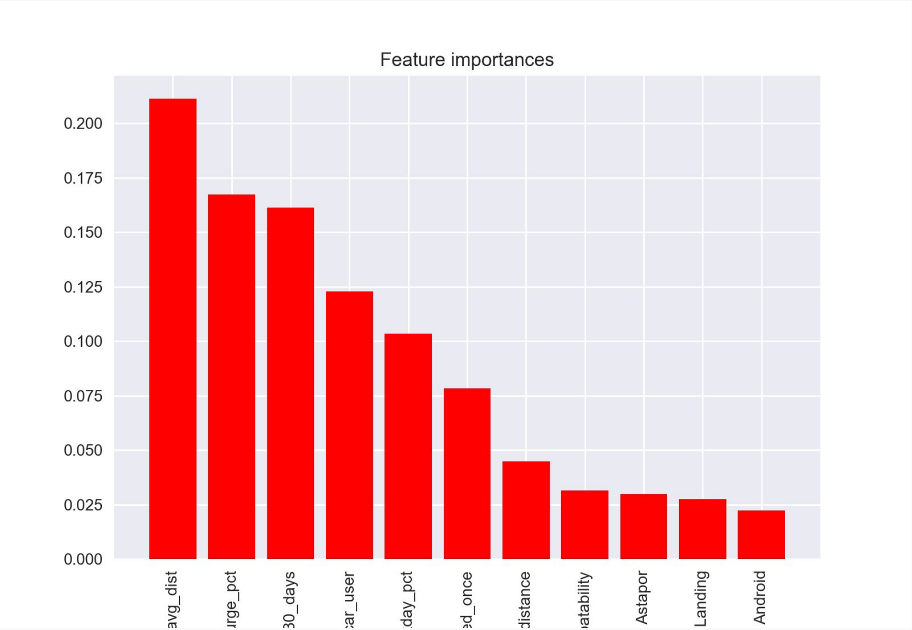
Exploratory Data Analysis

- The data was all from one month of signups: January 2014
- Relatively small amount of data
- Assumed that there is no missing data (presence of missing data turned out to be predictive)
- No variable for Churn, so we made a binary variable for Churn: True if last trip date before June 1, 2014, False if after June 1, 2014



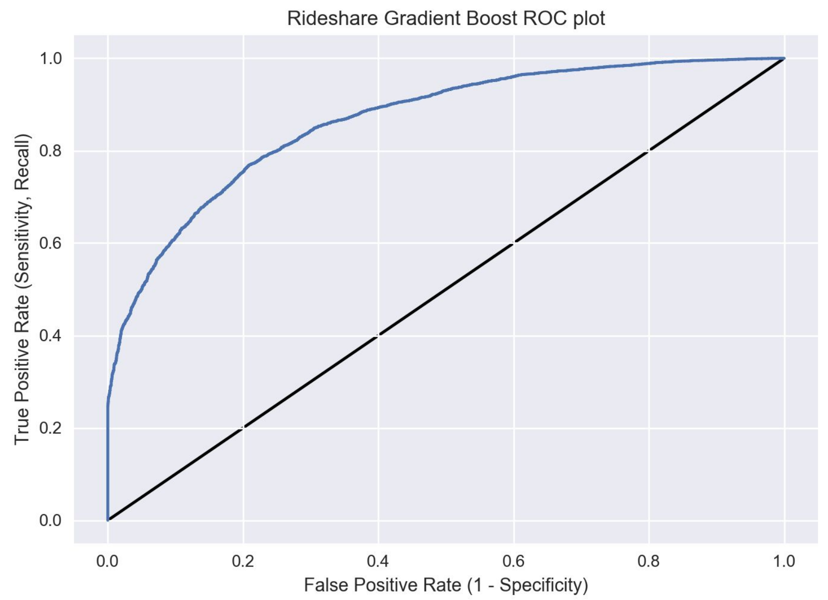
Most important features:

- Weekday percentage
- Surge aggregates
- Average distance
- It is relatively uncommon for users or drivers to fail to give a rating. However, when there is no rating, it is a very strong predictor for churning.



Results

- We chose accuracy for scoring, though one could make the case for either precision or recall
- Best model was Gradient Boosting
- Performed at .7894 accuracy



Possible Strategies:

- Because weekday traffic is so important, target employers for rideshare deals
- Stop using so much surge pricing! Reduces demand not just during the surge pricing hours, but in general
- Options for reducing surge pricing:
  - No surge pricing for a user’s first ten rides
  - No surge pricing during catastrophic events