Data Science Challenge - Churn Prediction

A San Francisco-based ride sharing company is interested in predicting rider churn. To help explore this question, we have provided a sample dataset of a cohort of users who signed up for an account in January 2014. The data was pulled several months later; we consider a user retained if they were “active” (i.e. took a trip) in the preceding 30 days (from the day the data was pulled). Assume the latest day of `last_trip_date` to be when the data was pulled. The data is `churn.csv`.

Here is a detailed description of the data:

```
city: city this user signed up in
phone: primary device for this user
signup_date: date of account registration; in the form `YYYYMMDD`
last_trip_date: the last time this user completed a trip; in the form `YYYYMMDD`
avg_dist: the average distance (in miles) per trip taken in the first 30 days after signup
avg_rating_by_driver: the rider’s average rating over all of their trips
avg_rating_of_driver: the rider’s average rating of their drivers over all of their trips
surge_pct: the percent of trips taken with surge multiplier > 1
avg_surge: The average surge multiplier over all of this user’s trips
trips_in_first_30_days: the number of trips this user took in the first 30 days after signing up
luxury_car_user: TRUE if the user took a luxury car in their first 30 days; FALSE otherwise
weekday_pct: the percent of the user’s trips occurring during a weekday
```

We would like you to use this data set to help understand **what factors are the best predictors for churn**, and offer suggestions to operationalize those insights to help this ride sharing company. Therefore, your task is not only to build a model that minimizes error, but also a model that allows you to interpret the factors that contributed to your predictions.

Work Flow

1. Perform any cleaning, exploratory analysis, and/or visualizations to use the provided. data for this analysis.
2. Build a predictive model to help determine whether or not a user will churn.
3. Evaluate the model.
4. Identify / interpret features that are the most influential in affecting your predictions.
5. Discuss the validity of your model.

Deliverables

- Code you used to clean data, explore data, build model, validate the model.
- Documentations, including the following points:
    - How did you computed the features and target?
    - What model did you use in the end? Why?
    - Alternative models you have considered. Why are they not good enough?
    - What performance metric did you use? Why?
    - Based on insights from the model, actionable plans to reduce churn.