



Case Study Churn Predictor

By Camilla Nawaz, Jeremy Gozlan,
Matthew Wong, and Urmi Mukherjee



Our Objective

We were given data from a ride sharing company from January to June 2014 and we wanted to find the best predictors for retention of those customers.



At first Glance

Defining Churn:

We considered any customer that did not use the ride sharing application within the last 30 days as churning.

Some other factors we initially thought might have the most signal:

- **Surge Percentage and Average Surge**
 - Paying higher surge prices on average will deter customers from using our app.
- **Weekday Percent Usage**
 - Users who mostly use the app on weekdays might be using it for commuting and might be less likely to churn.
- **Possibly Average Distance Travelled**
 - Travelling further might deter customers from using our app as often.

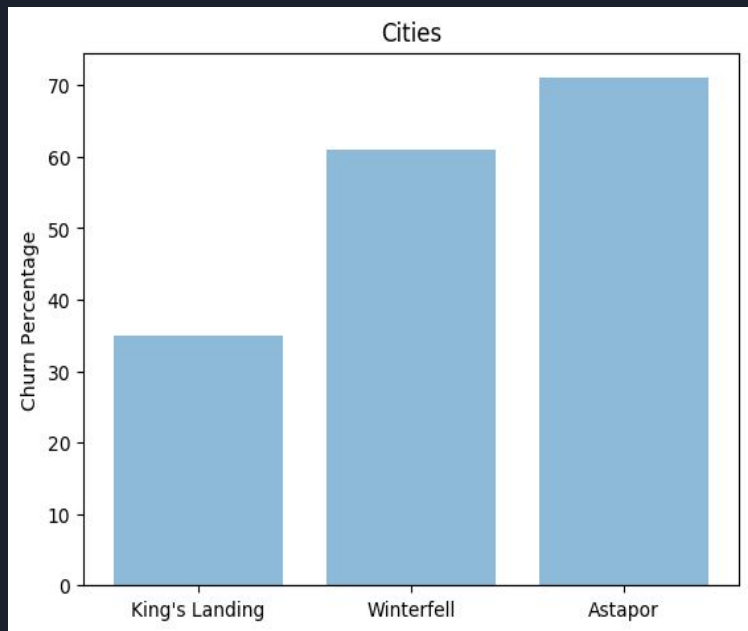


Our data:

- Our data had very few missing values
- We had to add whether or not the customers had used the app in the last 30 days
- There were a lot of missing values for the average rating of the rider in our data, so we dropped the data
- Most of the other features had all the data completed or very few missing values (~1%), so we dropped the rows where we had missing values

Exploring Our data

- The churn rate for King's Landing is significantly lower than that of the other two cities. (~35% in King's Landing, ~61% in Winterfell, ~71% in Astapor)
- The average driver rating is also slightly higher in King's Landing for those recorded. (4.84/5 King's in King's Landing, 4.53/5 in Winterfell, and 4.63/5 in Astapor)



Our First Model: Logistic Regression

Features Used:

- Average distance
- Average surge multiplier
- Weekday use percentage
- Number of trips in the first 30 days.

Accuracy:
0.659

Adding a few features

Features Used:

- Luxury car user, Iphone user, Surge * distance, average rating by driver, average distance, average surge, King's Landing, dummified bins for weekday percentage

Accuracy:
0.693

Our Second Model: Random Forest

Features Used:

- Average rating by driver
- Average surge multiplier
- City: King's Landing
- Phone
- Luxury car use

Accuracy:
0.75



Our Third Model: Gradient Boost

Features Used:

- Average rating by driver
- Average surge multiplier
- City: King's Landing
- Phone
- Luxury car use

Accuracy:
0.76

A decorative graphic on the right side of the slide, featuring a series of dark gray rectangular blocks arranged in a descending staircase pattern. Two blocks are highlighted: a light green one and a blue one, both positioned on the right side of the staircase.



Insights and Recommendations:

- The ratings and churn rate of King's Landing are comparably different than those of the other cities even though the averages of many of the other features are similar to those of other cities.
- We should survey those users in King's Landing and see why their experience using our app is better and try to replicate those environments in the other cities.
- iPhone users are also the majority of our customers, and they are more likely to use luxury vehicles so possibly targeting those users for surveys and product improvements could provide us with more insightful data.
- Possibly look to expand to other cities like King's Landing
- King's Landing represents a smaller proportion of our customers, so maybe that city is smaller or has better public transportation. Possibly gathering data on other details might provide more insight into our data.

Other considered features:

- We could not distinguish between the surge rates on weekdays and weekends, so that would be useful to have in the future.
- We would also like more information on the number of trips that a user takes because we only have information on the first 30 days and the last trip they took
- Also looked into whether or not there were a difference in where users did not rate their driver and did not find

