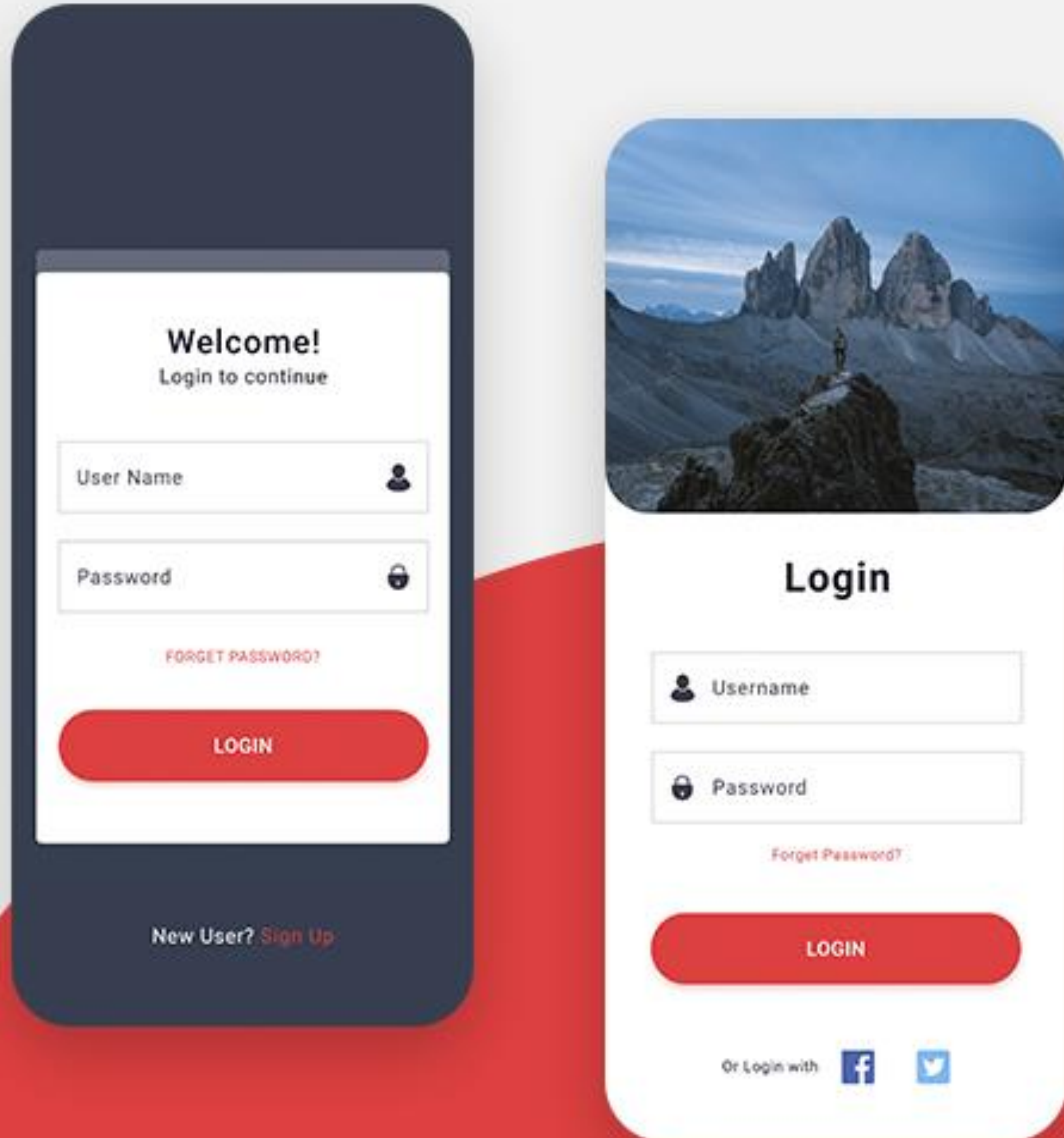


ULTIMATE INC. TAKE HOME CHALLENGE

REPORT

NOVEMBER 07, 2022

Authored By: Amit Kukreja

Table of Contents

1.0 Login Pattern

Based on the login data available from January 1 to April 14, 1970, the user logins were found to have a weekly pattern as shown in Chart 1.

- The *solid blue* line indicates the mean (expected) logins per hr at a given time on a given day of the week, for e.g. around 12 noon on Monday, the expected logins are about 60 per hr.
- The *light blue bands* around the solid blue line show 1 standard deviation around the expected logins i.e. at 12 noon on Monday, 65% of the time the expected logins are between 35/hr to 75/hr.

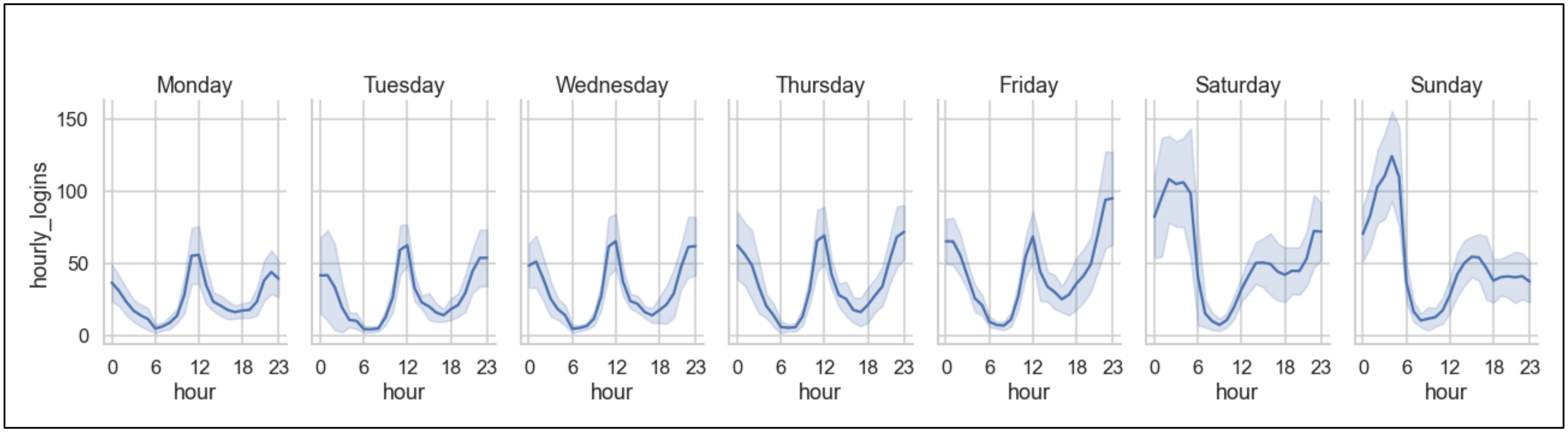


Chart 1: User Login pattern through the week

As one can see from the chart 1 above, Monday through Thursday the logins follow a similar pattern with a daily peak around noon. On weekends, the logins surge past midnight and peak at 4-5am the next day. More details on daily cycles in charts 2 and 3.

1.1 Daily Cycles

Weekdays (Monday - Thursday)

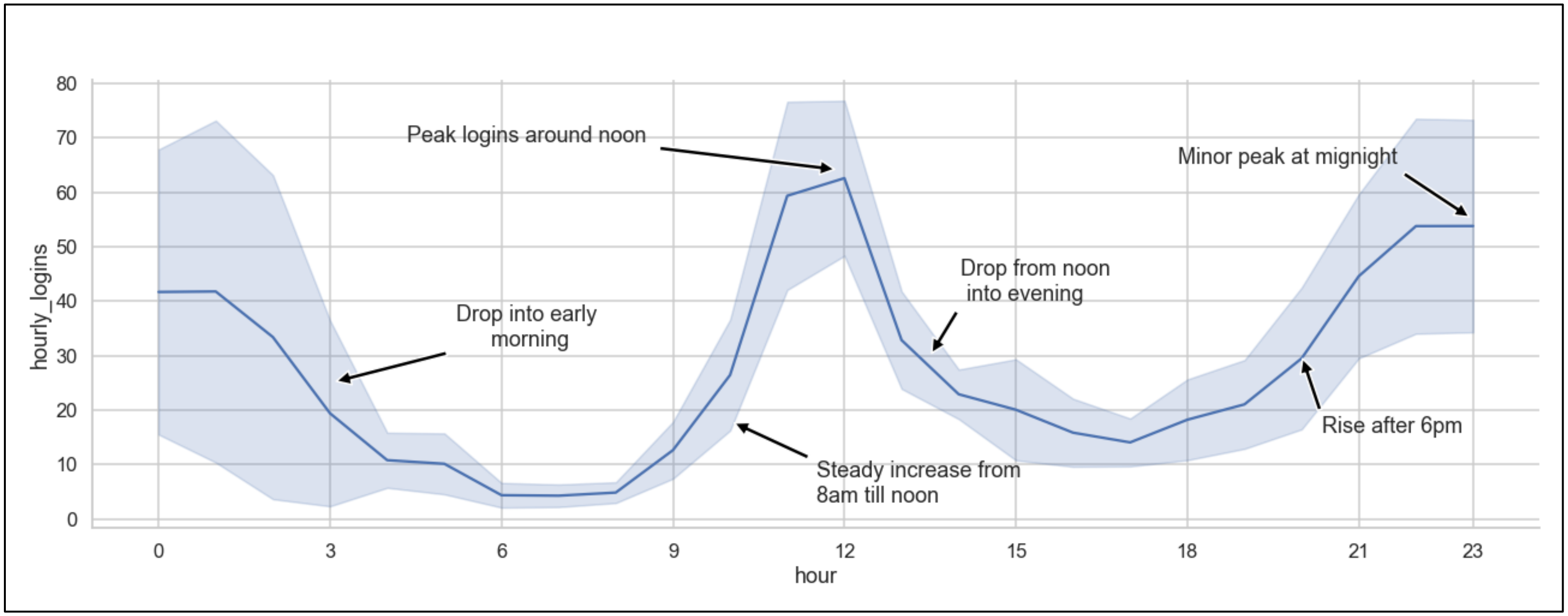


Chart 2: Daily login cycle on weekdays (Mon-Thur)

The daily pattern for weekdays is repetitive as shown on Chart 2.

1. There is a peak around midnight (40-60 logins per hour) followed by a drop into early morning (6am).
2. The logins start rising after 8am and peak again at 60-80 logins/hr around noon.
3. Post noon, the logins gradually drop to below 20/hr by 5pm.
4. The logins start climbing again after 6pm and peak around midnight.

Login cycle on Weekends (Friday – Sunday)

The pattern for weekends (Friday to Sunday) is somewhat different than weekdays as shown in chart 3.

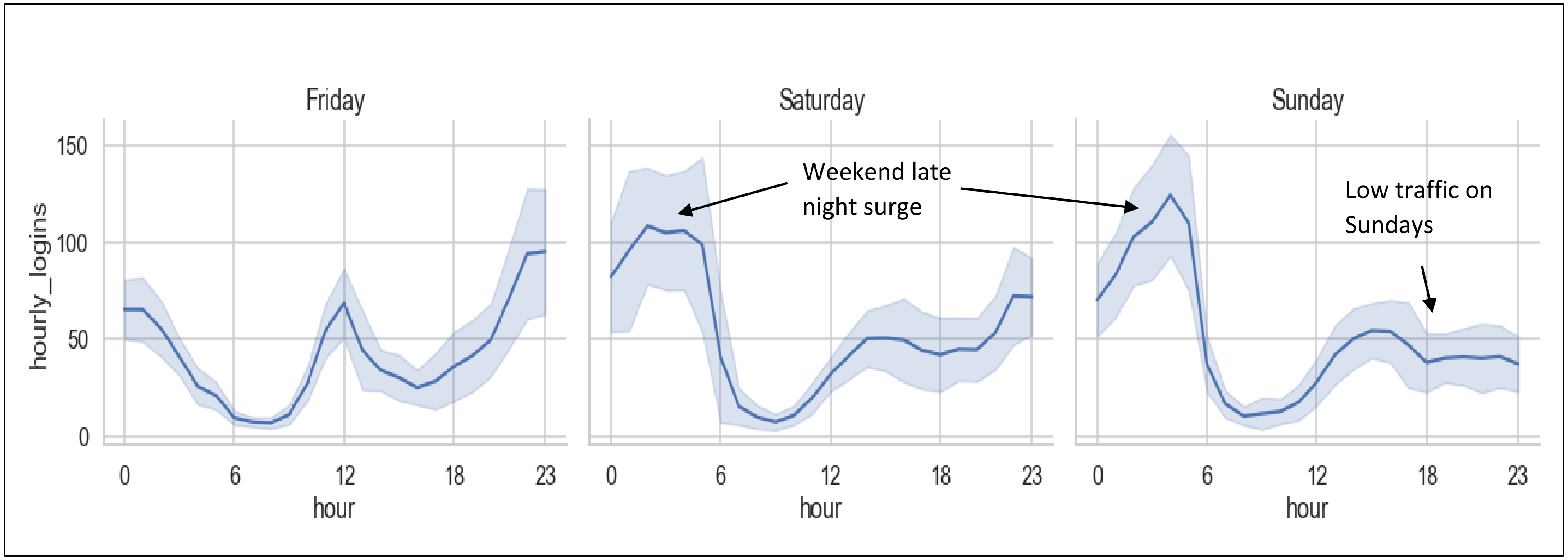


Chart 3: Login pattern on weekends (Fri-Sun)

1. Friday and Saturday nights see a surge in logins after 9pm that continues into early morning the next day. The hourly logins surge past 100/hr after midnight and peak between 100-150/hr around 4-5am the next morning.
2. Sundays see lower traffic during the day and the evenings are much less busy.

1.2 Data Quality

The data quality is quite clean, there was no missing data. Data was available from January 1 to April 16, 1970.

- There was 14 weeks (week 2 to 15) of complete data for all days of the week.
- For week 1 and 16, partial data was available. This meant that for some days e.g. Monday, 15 observations were available and for others e.g. Tuesday, 14 observations were available. However, given that a consistent pattern was found across weeks 2 to 15, this minor difference in no. of observations available does not impact the outcome of the analysis.

Python code for the analysis is attached in a separate file named *Ultimate_Inc_Data_Analysis_Challenge_ver2.0.ipny*

2.0 Experiment and metrics design

Experiment: Encouraging driver partners to be available in both cities, by reimbursing toll costs

Criteria for Success: I would compare the ratio of drivers serving both cities between test set (set of drivers offered the toll reimbursement) and base set (set of drivers not offered any incentive) as the success metric.

I would choose this metric as it indicates whether the incentive is encouraging more drivers to serve both cities.

<u>Design of the experiment</u>

Step 1: Randomly choose drivers for the test set.

- I would ensure that the no. of drivers in test set is at least 50, but not more than 10% of the total population of drivers. As it is an experiment, I would not want drivers not chosen for the experiment to learn that some of their fellow drivers are receiving an incentive which they are not. So it would make sense to keep the ratio less than 10%, but still be enough to test for statistical significance.
- If the base of drivers is less than 300, then we could randomly choose 50 drivers every week and give them the incentive, so there would be a chance that many

drivers get the opportunity to receive the incentive over the course of the experiment. However, this option would be slightly tedious to execute, as we would have to effectively communicate to drivers that this incentive is only for the week.

Step 2: Offer the chosen drivers the toll reimbursement incentive for inter city trips.

Step 3: Conduct the test over 30 days, so we have min. 30 observations. Record the daily ratio of drivers serving both cities in both the test set and base set.

Step 4: Calculate the means for both sets and use the t-test for difference in means to test for statistical significance.

<u>Interpreting the results</u>

The **null hypothesis** in this case would be that 'Providing an incentive doesn't change ratio of drivers who serve both cities" i.e. there is no difference in the means of the two distributions.

I would use a p-value threshold of 0.05 to check for significance. A p-value less than 0.05 would indicate statistical significance in the difference of the means, so we reject the null hypothesis. If p-value is greater than 0.05, we do not reject the null hypothesis.

3.0 Predictive Modeling

Key Objectives

1. Visualize & analyse data to understand what factors impact user retention.
2. Create a predictive model for user retention.
3. Understand key features that affect user retention and advise Ultimate on how it could leverage these insights.

3.1 Understand User retention through Data Visualization

After 6 months of signup, 37.6% (18,804 out of 50,000) of the users stay active. The tables & charts below help understand the profile of users who remain active.

Table 1: Active users by city

status	active	inactive	Total	percent_active
city				
Astapor	4228	12306	16534	25.57
King's Landing	6363	3767	10130	62.81
Winterfell	8213	15123	23336	35.19

In Table 1, we can see that King's Landing has the highest percentage of active users at 62.81%, but it has the lowest user signup at 10,130 users. Winterfell has the highest users signup, but has a less than average retention rate of 35.19%

Table 2: Active users by phone

status phone	active	inactive	Total	percent_active
Android	3146	11876	15022	20.94
iPhone	15658	19320	34978	44.77

In Table 2, we see that not only do more iPhone users signup, but they are twice as likely to stay (45%) as compared to Android phone users (21%).

Table 3: Active users by usage of Ultimate-Black

status ultimate_black_user	active	inactive	Total	percent_active
Not used ultimate-black	9307	21839	31146	29.88
Used ultimate-black	9497	9357	18854	50.37

Table 3 shows that users who used Ultimate black have a 50% chance of staying active, much higher than those who did not experience Ultimate black.

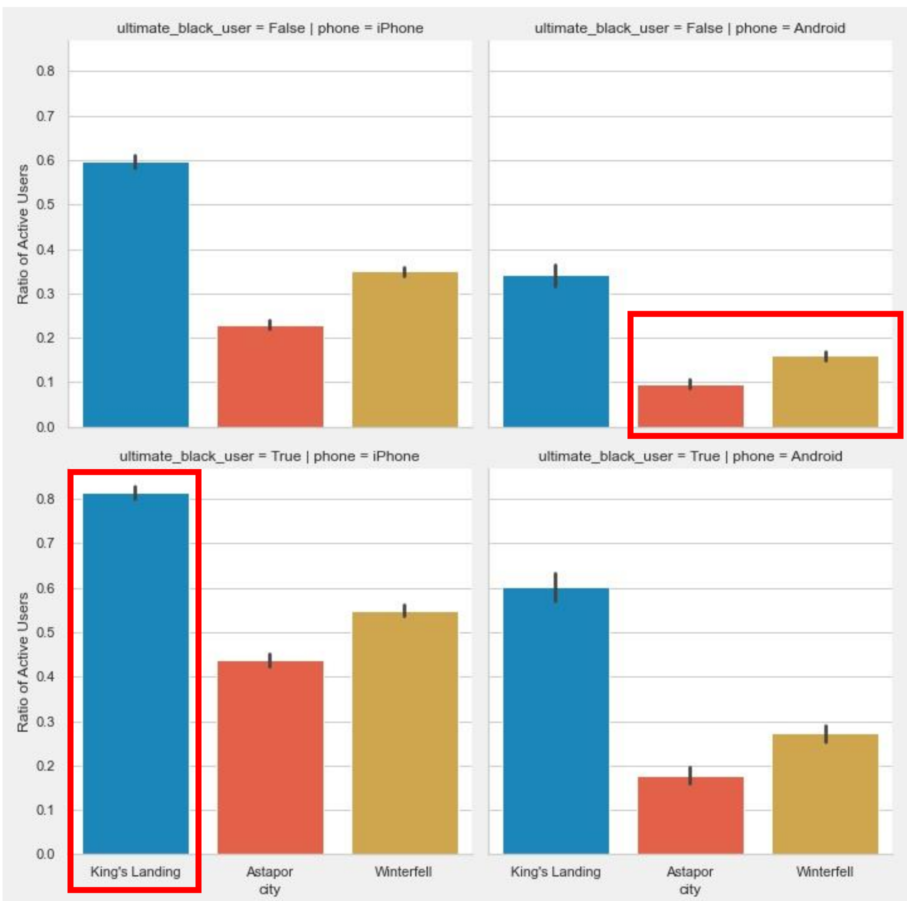


Chart 4: Ratio of active users by city | phone | ultimate black

Chart 4 above separates users on 3 axis – city, phone and ultimate black. It offers some interesting insights, for e.g.

- An iPhone user from King's Landing who has used Ultimate black has an 81% chance of staying
- An Android user from Astapor who hasn't used Ultimate black has only a 10% chance of continuing
- An Android user from Winterfell who hasn't used Ultimate black has only a 15% chance of staying

Table 4: Average values of some features by active status

status	trips_in_first_30_days	avg_dist	avg_surge	surge_pct	weekday_pct	avg_rating_of_driver	avg_rating_by_driver
active	3.306318	5.114788	1.073809	9.152797	61.389034	4.593679	4.762801
inactive	1.658482	6.207939	1.075339	8.666739	60.647032	4.606309	4.787415

In Table 4, we see that the trips in first 30 days is a good indicator of retention. Users who drop off from the service, usually take less than 2 trips on average in the first 30 days, while those who continue take more than 3 trips on average in the first 30 days. Also active users seems to take somewhat shorter distance trips (5.1 miles) compared to inactive users (6.2 miles).

3.2 Predictive Model for user retention

The task is to predict whether a user will be active or not by their 6th month. So, this is a classification task with a binary target variable. The intuition is to build models based on classifiers such as Logistic Regression, Decision Trees, and Gradient Boosted decision trees.

While a key metric is recall for the users that remain active, recall alone shouldn't be the only criteria. We do want to understand users that will not remain active and therefore we should not try to maximise recall at the expense of lowering precision or overall accuracy. As such, balanced metrics like **'overall accuracy'** or '**AUC'** would be the right metrics to evaluate the performance of the models.

3 models were built and tested based on the following classifiers:

1) Random Forest

2) Logistic Regression

3) XG Boost

Evaluation of Models

The results of different evaluation metrics are summarized in Table 5.

Metric	XG Boost Model	Random Forest Model	Regression Model
Overall Accuracy	79.42 %	78.4%	72.3%
AUC score	0.8602	0.8503	0.7653
Recall – Active Users	0.67	0.65	0.50
Recall – Inactive Users	0.87	0.86	0.65

Table 5: Evaluation metric of 3 different predictive models

- Predictive model based on XG Boost classifier makes 79.4% of the predictions accurately.
- The accuracy is higher for inactive users at 87%, with 67% of active users being predicted correctly.
- This model also has the highest AUC score of 0.8602. The ROC_AUC curves can be seen in Chart 5.
- The performance of the Random Forest Model is a close second to the XG Boost Model.
- There is a trade-off between increasing recall of active users and precision, as evident in Chart6. As recall increase, precision falls consistently.

Chart 5: ROC AUC curves of different models

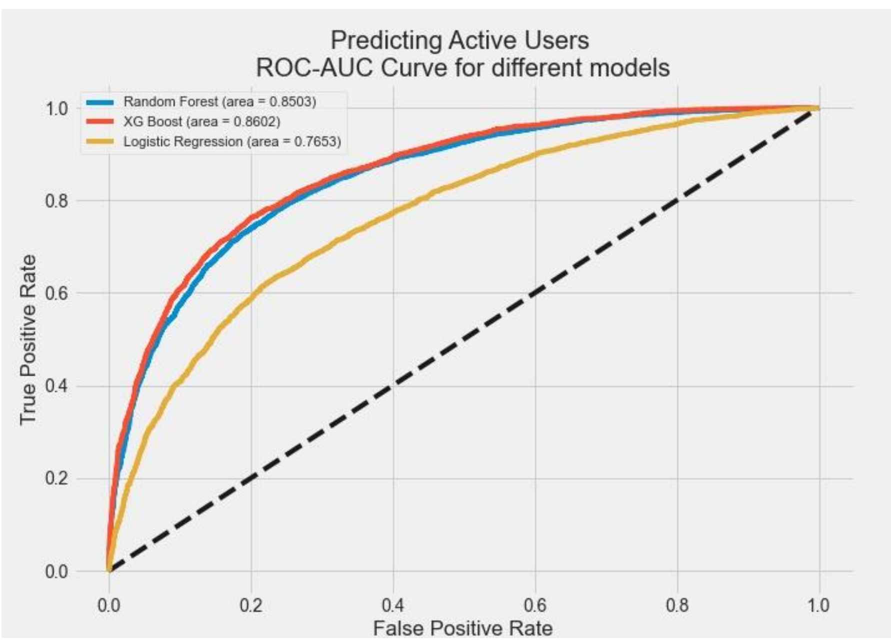
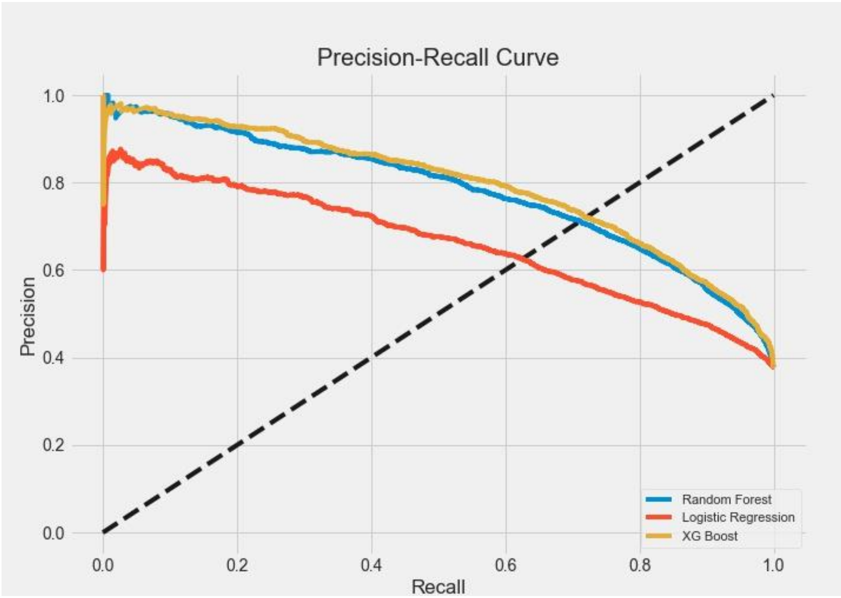


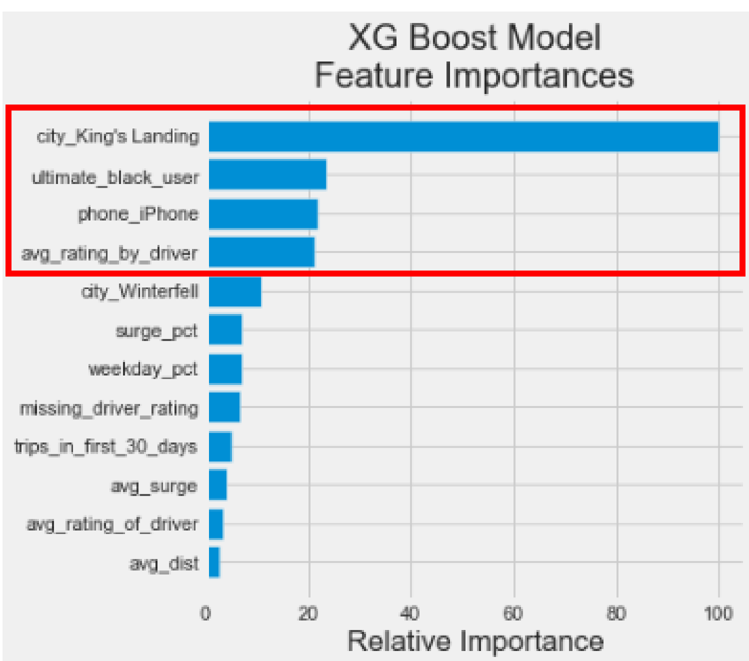
Chart 6: Precision Recall curves of different models



3.3 Insights from Predictive Model

There are some extremely useful insights that can be gathered from the most important features that XG Boost model used for predictions, as shown in Chart 7.

Chart 7: Feature Importances for Predictive Model



1) **City**: City of King's Landing was the most important feature. We saw during EDA that 63% of users from City of King's Landing remained active after 6 months. However, this city had the lowest user signup. Ultimate should focus and invest on growing their customer base in this city, as it would give them a loyal set of customers.

2) **Ultimate black usage**: This was the second most important feature. Again, we saw during EDA that 50% of Ultimate black users stay active. Ultimate Inc. could provide some incentive to users for experiencing this service as it is likely to pay dividends by making customers more sticky.

3) **Phone**: This was the third most important feature, with iPhone users twice more likely to remain active (45%) as compared to Android users (21%). While Ultimate cannot influence the phone used by a particular user, it might help to understand whether iPhone brand represents a certain demographic that is more predisposed to ride-hailing service e.g., iPhone users might be younger as compared to Android users. It would also help to understand if the Ultimate app for these two platforms offers any differential experience that might be causing this behavior.

4) **Avg. rating by driver**: The ratings given by drivers to users was the 4th most important parameter. We found that a larger % of users with a 4.6 to 4.8 rating given by drivers are likely to stay active. So Ultimate would do well to 'listen' to its driver partners!