

# Ultimate Take Home Challenge

## 1. Exploratory Data Analysis

- a. I was able to summarize the logins to 15 minute aggregates. I then created three visualizations. The first showing the most popular days of the week. The next was the average logins by time of day. The final graph shows the average logins by time of day and by day.
- b. A note on the data quality is that the timestamp given is all from the year 1970 which doesn't seem right.

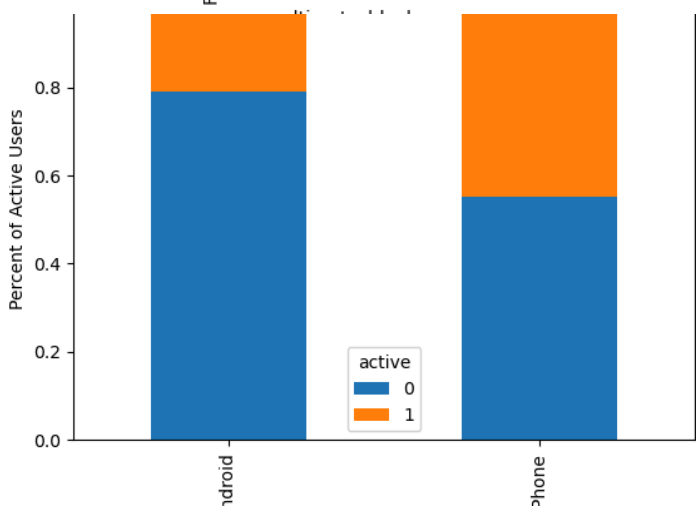
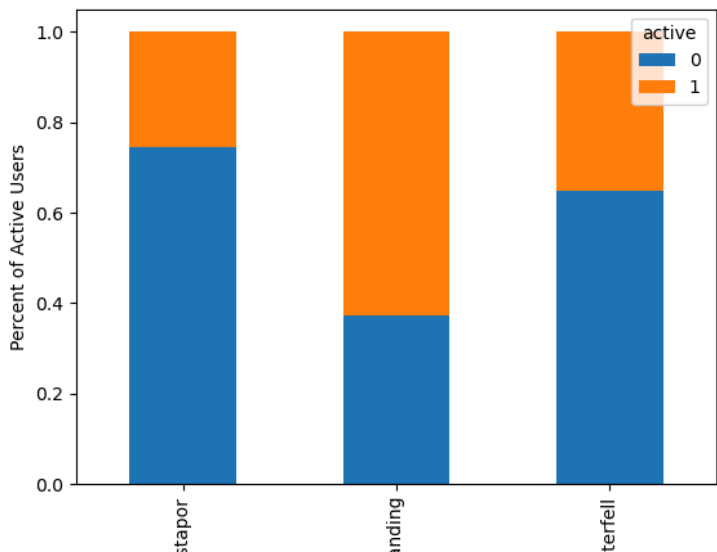
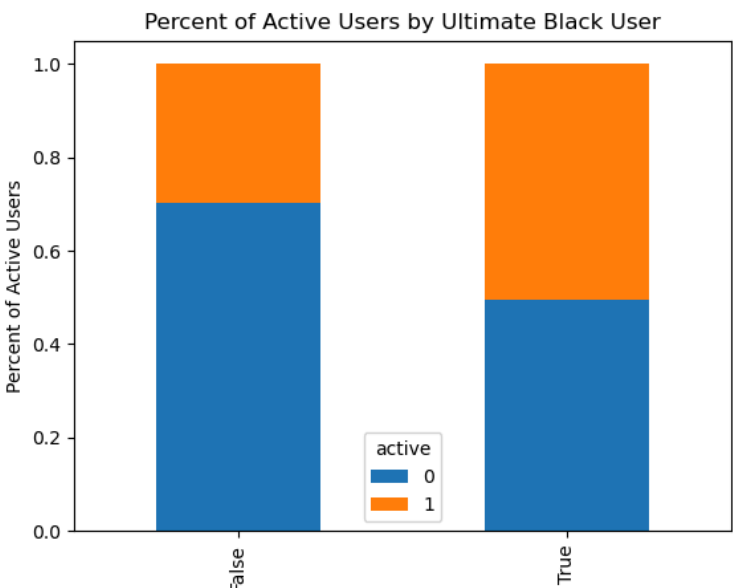
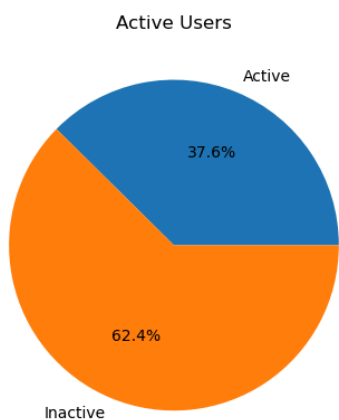
## 2. Experiment and metrics design

- a. The key measurement of success I would choose for this experiment would be the percent of drivers who serve in both cities. I would choose this measurement because it closely reflects the goal of drivers to work in both areas and we can accurately find differences before and after the reimbursement goes into effect.
- b. To implement the experiment I would select a subset of the drivers to be a part of a trial group. For this group only I would tell them in the app that they would be reimbursed for trips taken across the bridge. I would then conduct a T-test on the observed differences in the percentage of drivers who serve both cities between the two groups.
- c. If the P value is less than 0.05 then I would conclude that the experiment was statistically significant and the proposed changes would work at large. When rolling out

the changes to all of the drivers I would tell the city operations team to make sure the metrics of percentage of drivers who serve both cities continues to align with their long term business goals. For example, overall profits could decrease from drivers going back and forth across the bridge too much, racking up large toll fees for the company.

3. Predictive Modeling

- a. During the exploratory data analysis stage I found that 37% of the customers were still active. I also found some correlated variables that could be insightful to decision makers.



- b. To build a predictive model I first standardized and one hot encoded the dataframe to ready it for my model. I then create a train test split so my evaluations would not be overfit on data I train on. I then Evaluated a couple different models with grid search hyperparameter tuning. The best model I found was a random forest classifier. It had an accuracy score of 0.784 which means it is able to identify whether a user will be inactive or not greater than  $\frac{3}{4}$  of the time.
- c. To be able to gain insights from the model Ultimate could use targeted marketing campaigns to encourage likely inactive customers to ride again. It could also take a look at correlated and important features to improve areas that lead customers to stop using the service. For example finding out why users in Astapor and Winterfell are more likely to cancel that Kings Landing.