03/23/2025

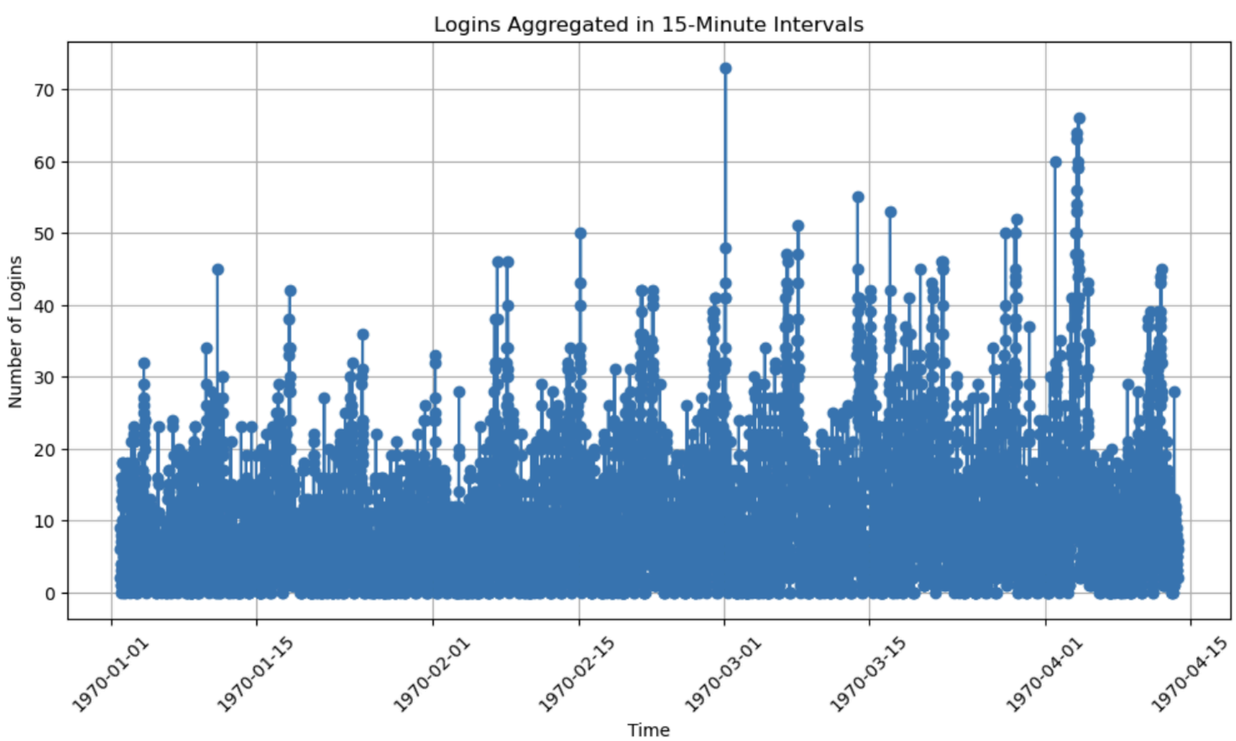
Sarah Berkin

Project Summary: Rider Retention Analysis for Ultimate

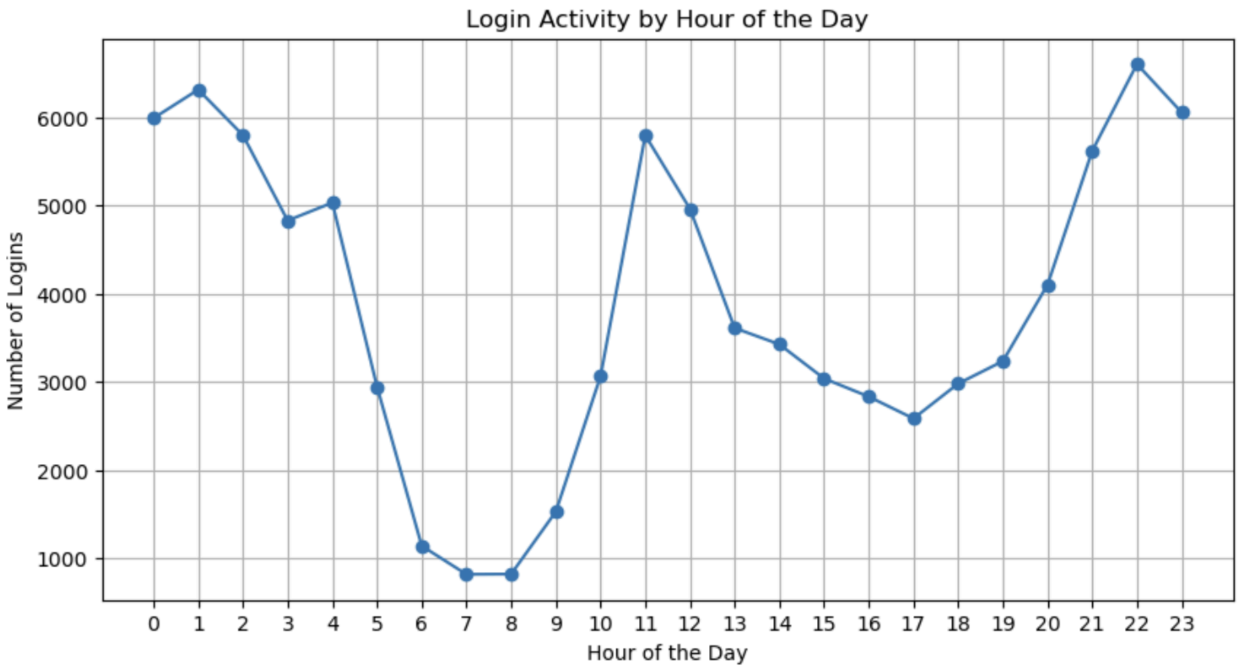
Part 1: Login Time Series Analysis

Using a dataset of login timestamps, I aggregated user activity into 15-minute intervals and visualized trends to uncover usage patterns. The analysis revealed clear daily and weekly cycles in user demand: 1) Daily peaks are overnight and around lunchtime, while the least activity is during morning rush hour commute and 2) Weekly lowest activity is on Monday, with each day gradually increasing as the weekend peak approaches. These cycles provide foundational insights into where to focus efforts in service optimization.

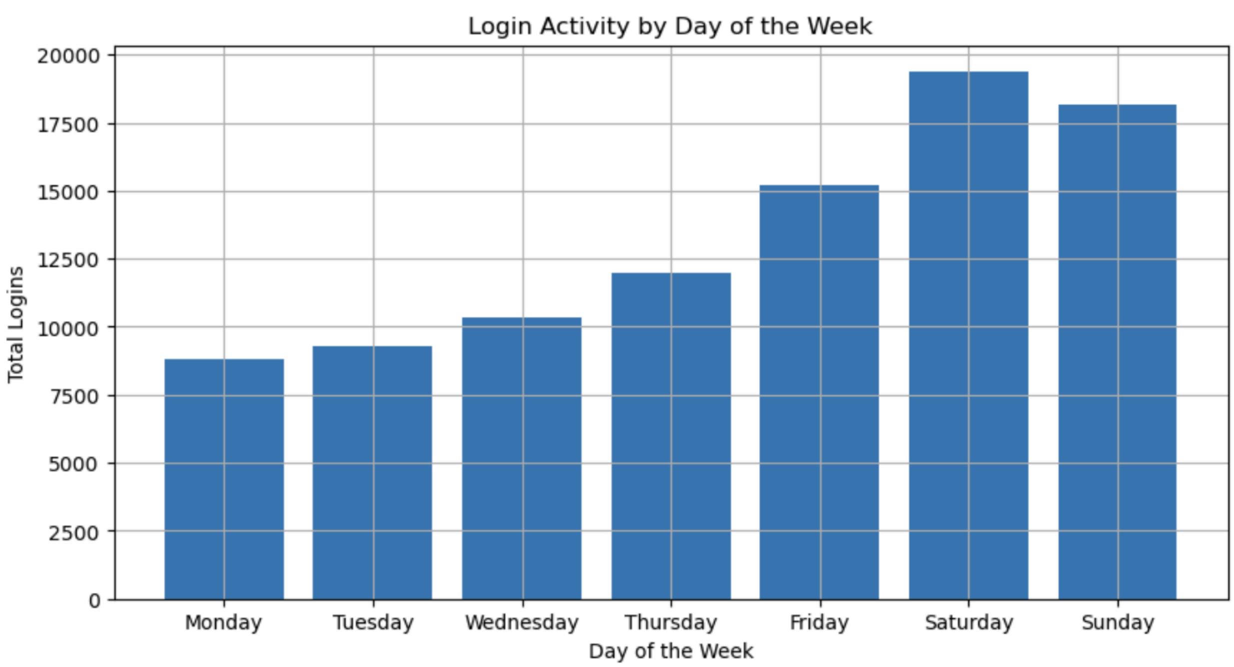
Visual of full dataset



Daily cycles



Weekly Cycles



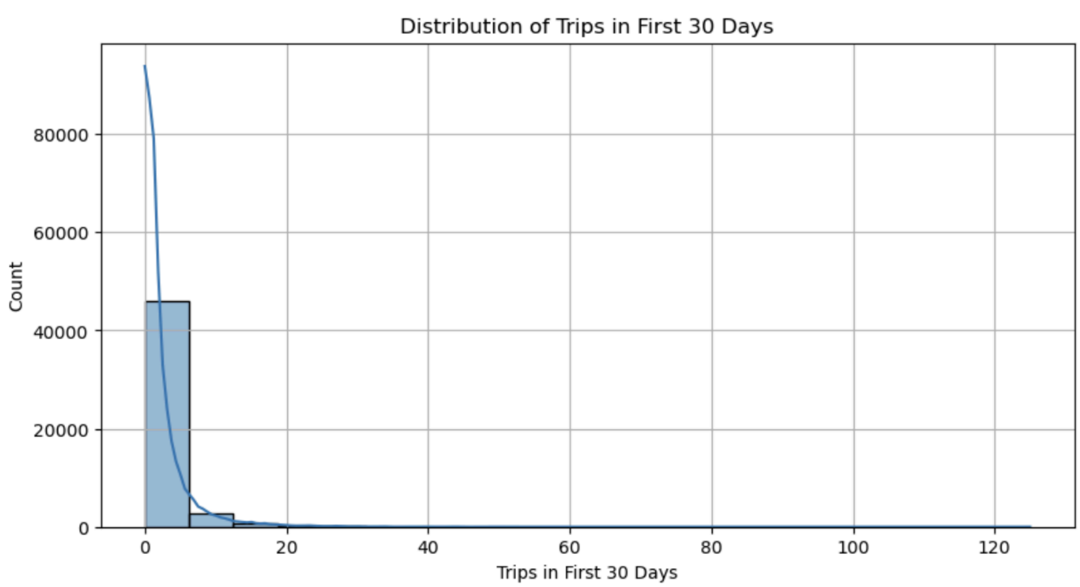
Part 2: Experiment & Metrics Design

1) The key success metric selected was **average revenue per driver from cross-city trips**, as it aligns directly with Ultimate's business objectives.
2) I propose an A/B experiment to assess the impact of toll reimbursement on driver behavior between two cities.
    a) This can be implemented with a randomized control trial. Drivers would be randomly split into two groups: a control group with no reimbursement and a treatment group receiving full toll reimbursement for cross-city trips. Over a 4–6 week period, we would track revenue per driver from these trips to assess behavioral change.
    b) To determine statistical significance, a two-sample t-test would be used to compare the average cross-city revenue between the two groups.
    c) If the treatment group shows a meaningful and statistically significant increase in revenue, we can conclude that toll reimbursement has a positive impact on cross-city driving. In that case, we would recommend expanding the program. It's important to monitor for any confounding factors such as seasonal demand or geographic distribution. Long-term driver satisfaction and cost-effectiveness of the incentive should also be evaluated before scaling city-wide.

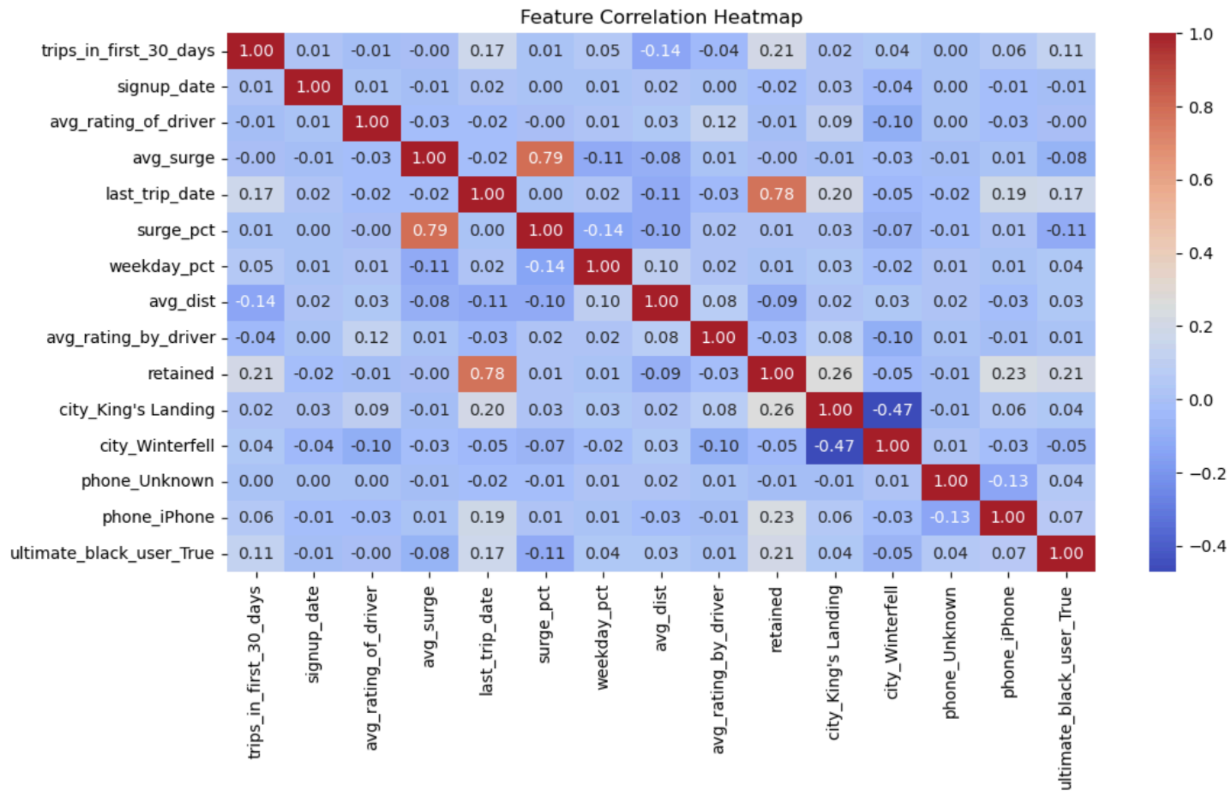
Part 3: Predictive Modeling of Rider Retention

1) After performing data cleaning and defining retention as users who had a trip in the last 30 days before the dataset end date, we found that 37.61% of users were retained. This indicates a significant chunk of the user base that potentially could be room for improvement.

Trips in 1st 30 days



Feature Correlation Heatmap

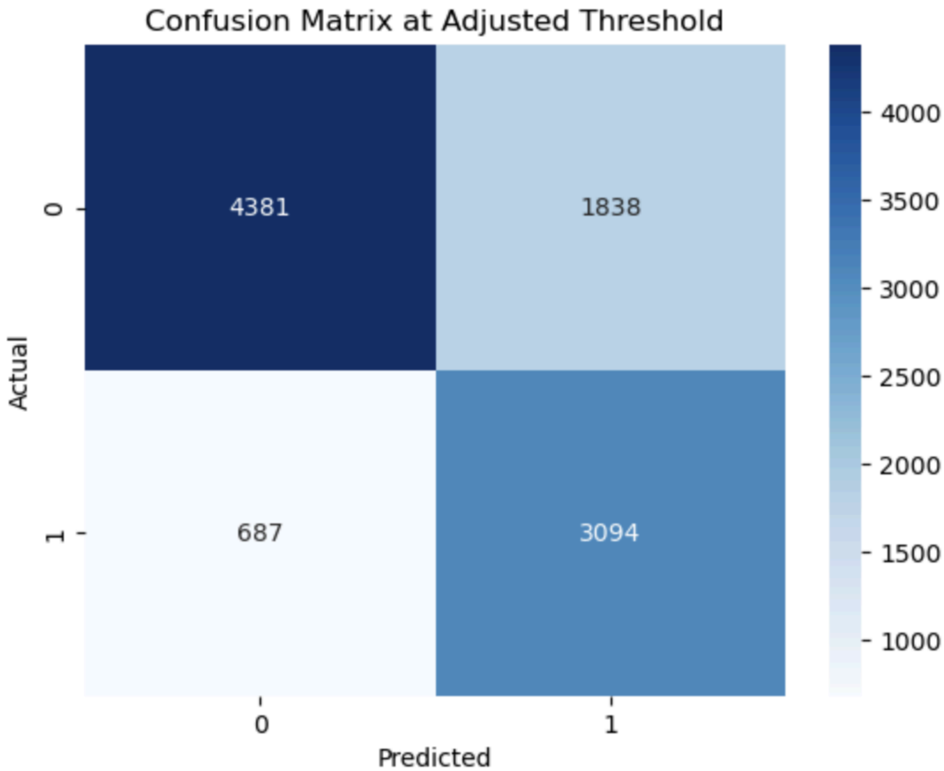


2) I built a Gradient Boosting Classifier to predict whether a user would remain active six months after signup. Feature engineering was guided by exploratory analysis, focusing on early user activity, device type, service tier, and geography. To better capture retained users, I adjusted the classification threshold to favor **recall over precision**, ensuring high-risk users were not overlooked. The model achieved an accuracy of **78%**, with strong performance in identifying churn risks and high-retention profiles.

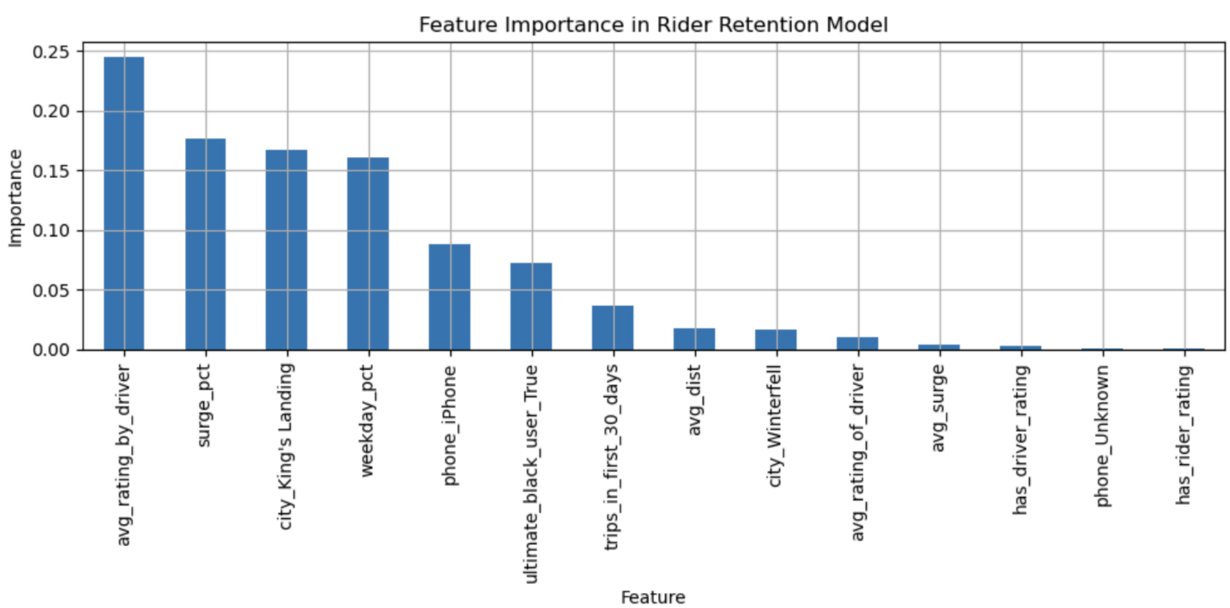
Results of the final model

```
Adjusted Threshold: 0.3
Accuracy: 0.75

Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.70      0.78      6219
           1       0.63      0.82      0.71      3781

    accuracy                           0.75     10000
   macro avg       0.75      0.76      0.74     10000
weighted avg       0.77      0.75      0.75     10000
```

Confusion Matrix



Feature Importance



Business Impact

The model uncovered actionable insights that Ultimate Technologies could take:

- Early trip activity is a strong signal of long-term retention.
- Users of premium services (Ultimate Black) are more likely to stay engaged.
- City-specific behavior patterns suggest opportunities for localized marketing strategies.

These findings support targeted onboarding, segmentation-based campaigns, and proactive engagement efforts to drive long-term rider retention.