

Summary

- Solution Overview
- Data Exploratory Analysis
- Data Processing
- Feature Selection
- Visualization and Insights
- Policy Recommendation
- Business Impact
- Next Steps

Business Problem

Rider Cancellations create bad experiences for both Drivers and Riders on the Lyft platform and impact Lyft financially. How to leverage Advanced Analytics to guide rider behaviors and develop cancellation policy?

Solution

First

- Increase factors via Feature Engineering
- Split ride data into shared ride and standard ride
- Identify key factors via machine learning

Second

- 4 Insights and visualization

Last

- Cancellation policy recommendation and business impact

Data Exploratory Analysis

Time Span

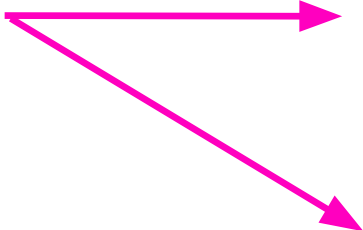
42 days

(19/04/14 - 19/05/26)

Location



1,367,790



463,203 (shared)



904,587 (standard)

Drivers

36,385



Riders

530,756

MAX

Cancellations

87



MAX

Requests

107

Data Processing

Feature Engineering – Add new features:

- Total number of requested rides for each rider
- Total number of cancelled rides for each rider
- Cancellation rate of each rider:
 $\text{total number of cancelled rides} / \text{total number of requested rides}$
- Weekend/Weekday flag
- Rush hour flag:
7am to 10am & 4pm to 7pm during weekday
- ETA difference:
 $\text{post-match ETA} - \text{pre-match ETA}$
- Matching time span:
 $\text{accepted time} - \text{request time}$

Feature Selection

Goal: To **identify significant factors** that affect cancellation behaviors of riders

Methodology: Logistic regression & Tree based Xgboost model

Factors included in both models:

- o Upfront fare
- o ETA to rider pre match
- o ETA to rider post match
- o Rider request number
- o Cancel penalty(\$1,\$3,\$5)
- o Ride eta diff'ride wait time before accept
- o Weekday flag
- o Rush hour flag
- o Total count of ride requests for each rider
- o Total count of cancellation for each rider
- o Cancellation rate of each rider
- o Cancellation flag

Model Performance

Logistic Regression

- shared ride: 0.9367612854690354
- Standard ride: 0.927485938479283

XgBoost Tree

- Standard ride:
 - F1 score: 0.82944774841525093
- shared ride:
 - F1 score: 0.83234842238473947

Decision Tree and Logistic Regression

Standard Ride

TOP 3 important factors and correlation direction:

- 1.**Cancellation rate** (positively correlated) – riders tend to cancel more if they have high cancel rate from past ride history.
- 2.**ETA difference** (positive correlated) - the larger the difference between ETA of pre match and ETA of post match, the more likely the rider is going to cancel a ride
- 3.**Upfront fare** (negative correlated) – Riders tend to cancel more if they have low upfront fee

Decision Tree and Logistic Regression

Shared Ride

TOP 3 important factors and correlation direction:

- 1.**Cancellation rate** (positively correlated) - the higher the cancellation rate in a rider's ride history, the more likely the rider is going to cancel a ride
- 2.**ETA difference** (positively correlated) - the larger the difference between ETA of pre match and ETA of post match, the more likely the rider is going to cancel a ride
- 3.**Matching period** (positively correlated) - shared riders are time sensitive, because matching period require longer matching time frame

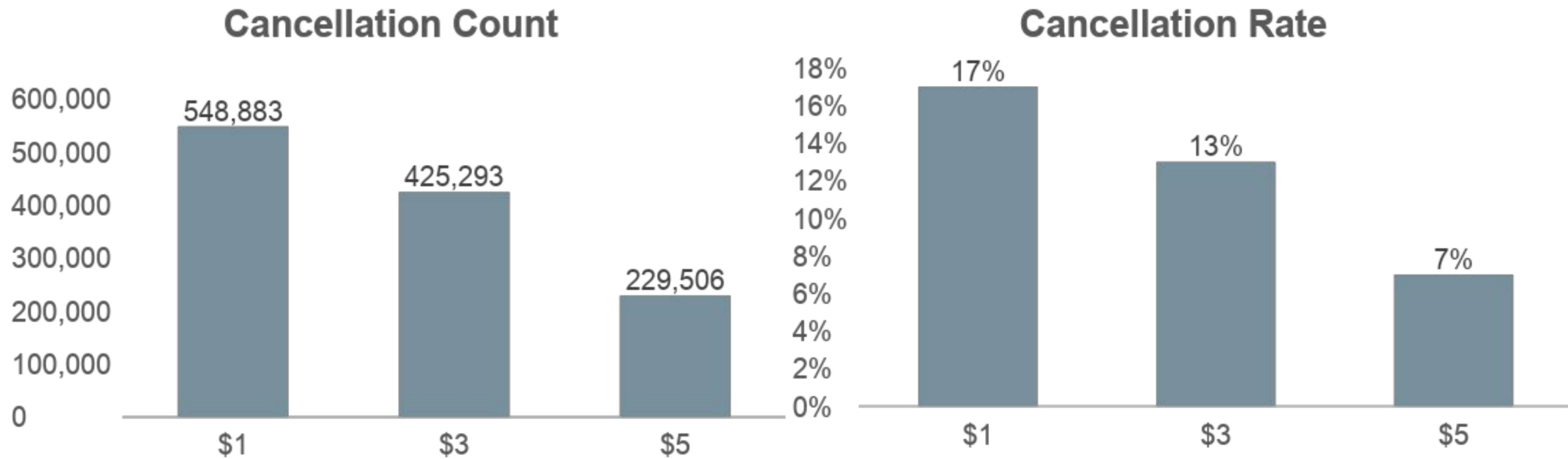
What did we find?



Insight 1

Behavioral Impact

Experiment penalty affects cancellation rate and count with statistical significance as illustrated below. **So Lyft should set up cancellation policy.**



Insight 2

Financial Impact

Estimated Net Loss due to cancellation is \$1,985,474

Estimated Net loss = Total upfront fee - Total penalty charged

Penalty	Total Upfront Fee	Total Number of Ride	Total Penalty Received	Estimated Net Loss
1\$	\$1,073,695	77,528	\$77,528	\$996,167
3\$	\$857,312	61,252	\$183,756	\$673,556
5\$	\$493,341	35,518	\$177,590	\$315,751
Sum	\$2,424,348	174,298	\$438,874	\$1,985,474

Insight 3

Time Sensitivity

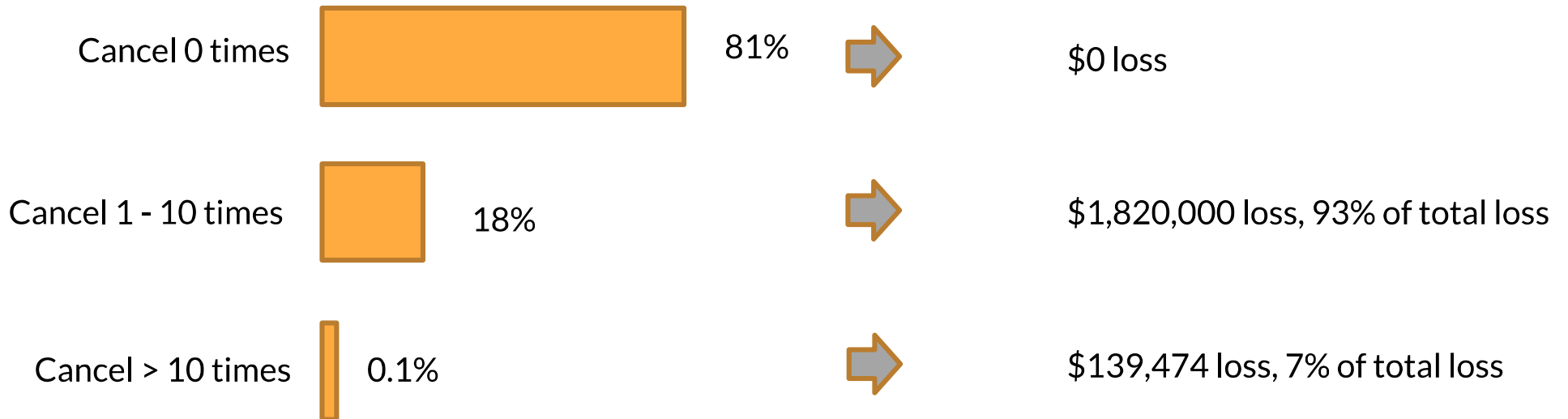
Mitigate cancellation behaviors by reducing differences between pre-match eta and post-match eta, which is the second most important feature in decision tree model for both standard ride and shared ride

One important reason behind cancellation is that riders are time sensitive. Lyft values rider's time. We can enhance rider experience by providing better prediction of ETA.

Insight 4

Cancellation Distribution

81% of riders cancel 0 times. **18%** of riders cancel < 10 times and make up of **93%** of total loss. **0.1%** of riders cancel 10+ times, cause **7%** of total loss



*All rides are in 40 days period

Cancellation Policy



Cancellation Policy

Rider Segmentation

Low value riders: cancel more than 10 times in 42 days experiment

Some of these cancellation behaviors can be considered as malicious. For example, a rider cancel 87 times in 42 days!

These low value riders consist of 0.1% of total riders, but they account for 7% of total financial loss.

These low value riders with high cancellation rate are more likely to cancel in the future. Lyft should apply higher penalty since higher penalty correlates with lower cancellation rate.

Since **low value riders** disrupt the marketplace by frustrating drivers and causing financial loss to Lyft, we should apply higher penalty for them.

Cancellation Policy

Rider Segmentation

High Value Riders: cancel 1 – 10 times in 42 days experiment

These riders have long lifetime value, who make cancellation occasionally.

As we know riders are time-sensitive, Lyft should create friendly policy for those with a reasonable reason to cancel.

Since higher penalty correlates with lower cancellation rate, Lyft should intervene the cancellation behaviors by using appropriate penalty.

More flexible cancellation penalty for **High Value Riders** who cancel their rides for reasonable situations, such as ETA longer than expected

Cancellation Policy Recommendation

Standard Rides

Cancel 1 mins after request: \$2
Cancel 3 mins after request: \$3
Cancel 5 mins after request: \$5

Rational
Riders

Cancellation Penalty of \$8

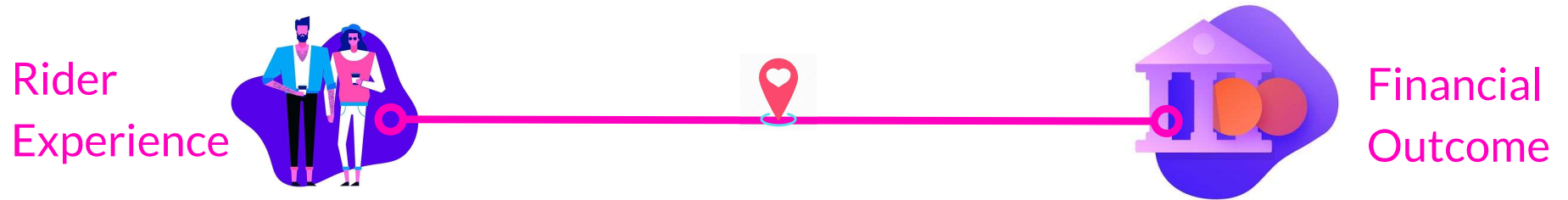
Low-value
Riders

Shared Rides

Cancel 1 mins after request: \$1
Cancel 3 mins after request: \$3
Cancel 5 mins after request: \$5

Cancellation Penalty of \$6

Business Impact



- 2 cancellation policy covering both Rational Rider and Low Value Rider
- Impact cancellation behavior of 97,794 riders, 18.1% of total rider, who account for estimated \$1,985,474 financial loss
- Maintain good relationship with rational riders, 18% of total rider, with flexible cancellation policy, creating better experience for riders and future revenue for Lyft.
- Penalize 0.1% low value rider, who cancel 10 + times in 40 days, which make up of 7% of financial loss in order to prevent further damage to the company.

Next Steps

- Combine with driver data to study the cancellation penalty on driver's behaviors
- Experiment to test the effect of new cancellation policy

What question do you have?

