Lyft Data Challenge 2019

Ideal Speculation: Jen Sheng Wong, Zhihao Guo

Table of Contents

- Problem Statement
- Feature Engineering
- Methodology
- Conclusion

Problem Statement

Problem Statement

- What is a driver churn?
- What factors affect driver churn?
- How to reduce churn?
- Test hypothesis!

Definition of driver churn

- Assumption:

  Drivers who have more than 10 days of inactivity have churned

- Last day is 2016/06/27
- Look for driver whose last day of activity was before 2016/06/17

Why 10?

- We call 10 the "window"
- Find a window such that the gap between not giving any ride is smallest
- Assign `True` or `False` to `driver_id_rest_too_long`

driver_id	0	1	2	3	4	5	6	7	8	has_churned
002be0ffdc997bd5c50703158b7c2491	35.0	45.0	32.0	10.0	10.0	29.0	46.0	47.0	23.0	0
007f0389f9c7b03ef97098422f902e62	1.0	1.0	9.0	5.0	0.0	6.0	5.0	2.0	2.0	0
011e5c5dfc5c2c92501b8b24d47509bc	7.0	4.0	1.0	5.0	7.0	5.0	5.0	0.0	0.0	1

Why 10?

- Did not drive on week 2, 3, 4, 5
- Drove again on week 6, 7
- Did not drive on week 8

Has he churned?

No! Is he taking a break?

driver_id	0	1	2	3	4	5	6	7	8	has_churned
8d5a99f2032a336a57222948915edf09	31.0	26.0	0.0	0.0	0.0	0.0	2.0	1.0	0.0	1

Why 10?

- For a window of 7, 320 drivers “took a break” for 7 days
- For a window of 10, 197 drivers “took a break” for 10 days
- For a window of 30, 14 drives “took a break” for 30 days

driver_id	0	1	2	3	4	5	6	7	8	has_churned
002be0ffdc997bd5c50703158b7c2491	35.0	45.0	32.0	10.0	10.0	29.0	46.0	47.0	23.0	0
007f0389f9c7b03ef97098422f902e62	1.0	1.0	9.0	5.0	0.0	6.0	5.0	2.0	2.0	0
011e5c5dfc5c2c92501b8b24d47509bc	7.0	4.0	1.0	5.0	7.0	5.0	5.0	0.0	0.0	1

Feature Engineering

Feature Engineering

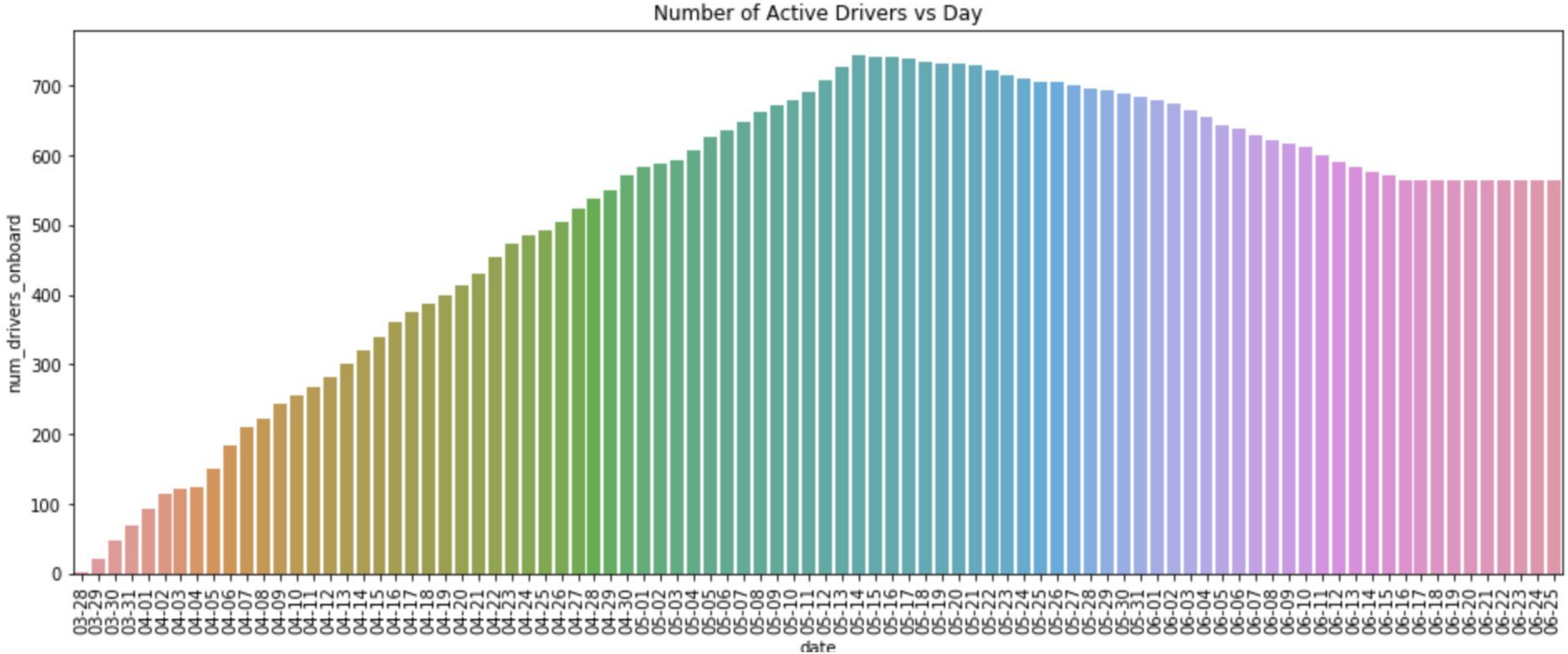
Features	Explanation
fare	The revenue generated by the ride. Calculated from the formula: $\text{Fare} = (\text{base fare} + \text{cost per mile} \times \text{miles traveled} + \text{cost per min} \times \text{mins traveled})(1 + \frac{\text{prime time}}{100}) + \text{service fee}$
drop_requested_lag	lag between a driver dropping off a passenger and picking up next passenger
ride_prime_time	Prime Time applied on the ride
driving_period	The time elapse between the first a driver drove till the last day he drove. The driver may rest in the middle.

Feature Engineering

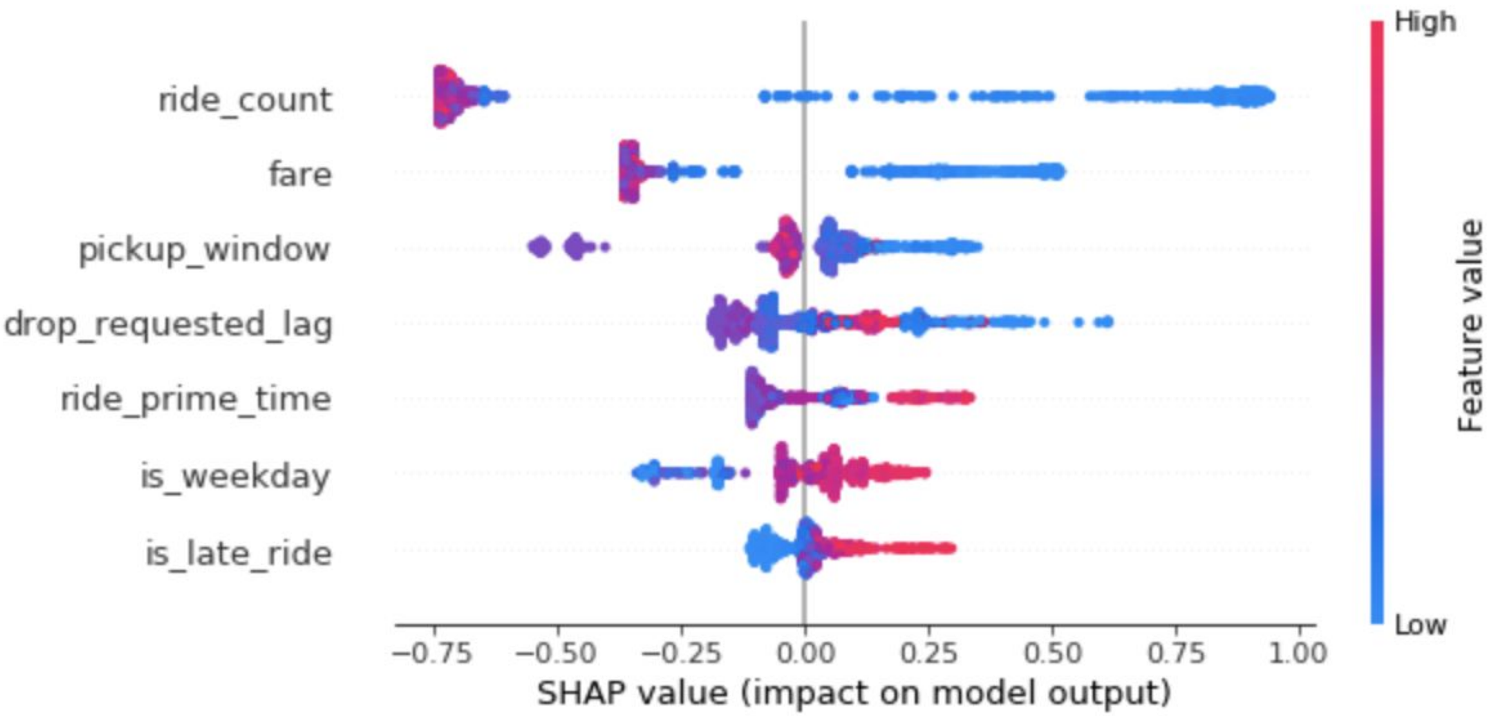
Features	Explanation
unique_days	The number of days a driver actually drove
is_weekday	The percentage of the drives that happened on weekdays for a driver
is_late_ride	The percentage of rides that took place after 11pm for a driver
has_churned	According to our assumption, whether this driver has been churned. Binary.

Methodology

Definition of driver churn



Factors affecting driver churn



Factors affecting driver churn

From 1st Challenge,

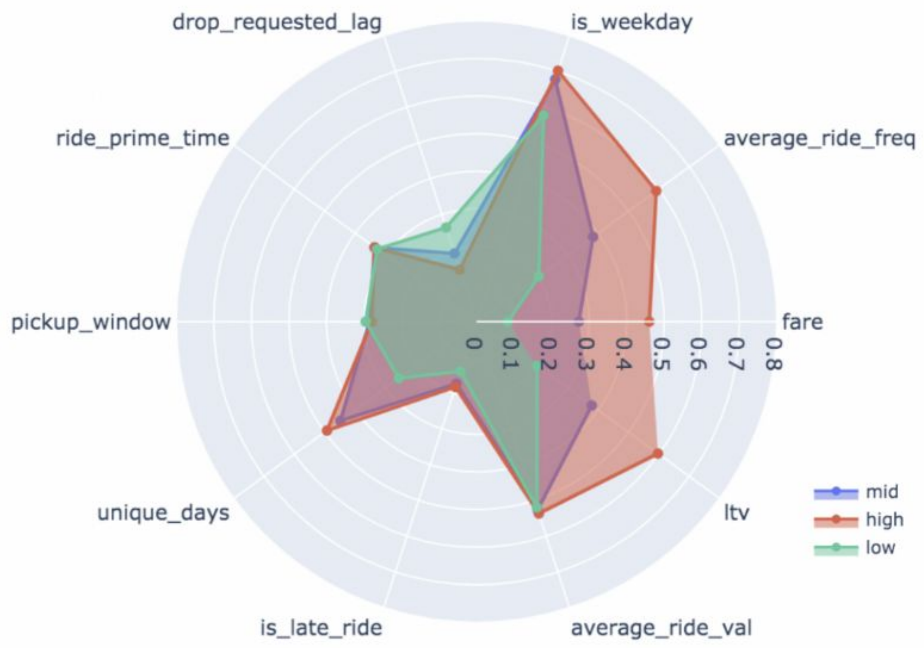


Figure 4: Radar Chart of Features

Steps to reduce churn

- Increase driver's take-home pay!

- Offer a bonus based on "streaks", such as rewarding them if they drive more consistently. This can help minimize their gap of `average_ride_freq` with high value drivers (as seen in the radar plot)

Hypothesis

- If we increase their take-home pay, churn rate will reduce
- Assume Lyft now takes a 20% cut from how much a driver makes per trip
- What if Lyft takes only 15% cut?

Hypothesis: Why impactful?

- Drivers will be lured to drive for a longer period
- Lower churn rate = $$$$$

Experiment: Original

- The original churn rate is around 30%
- Fit an XGBoost model to predict `has_churned (label)`
- Average AUC of 0.86 on 5 fold validation set

Experiment: Setup

- Increase fare for drivers who are predicted to churn by 6%
- Quick and easy to increase in take-home pay
- Why use 6%? 0.8 * 1.06 = 0.85
- Drivers take home 85% of a fare

Experiment: Setup

- Holding all features unchanged, predict `has_churned (predicted)` using the original trained XGBoost

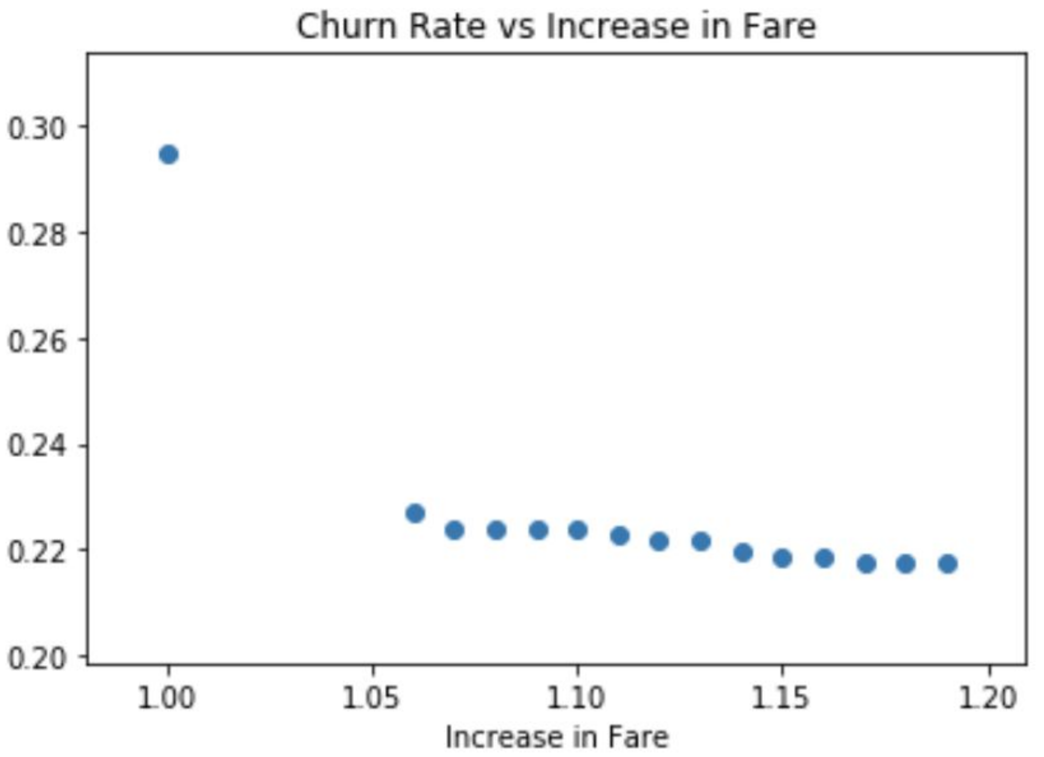
- Compare `has_churned (label)` with `has_churned (predicted)`

- Calculate overall `has_churned (predicted)`

Experiment: Result

- Repeat experiment from increase by 6% to 20%
- Translates to take-home pay of 85% to 96%

Experiment: Result



Experiment: Metrics

- AUC to predict has_churned, or better, (will_churn)
- Reduce in a driver's "break"
- Change in behavior: from `has_churned = 1` to `has_churned = 0`

Experiment: How long?

- Observe activities of drivers who have high probability to churn
- Run experiment for 2 or 3 months to check if increasing their take-home pay convinced them to stay

Experiment: Make decision?

- We are able to run the models and calculated the change in the churn rate.
- Do the statistical test to see if the change is significant or not.
- If it is significant, then we conclude that the change does affect the churn.

Conclusion

Conclusion

- Increase in a driver’s take-home pay reduces churn for drivers!

Acknowledgement

Thank you Lyft (and Deitrick especially)!

The End

Questions?