

## OVERVIEW

### Value Criterion

To maximize value to Lyft, a driver will consistently drive high value rides which best fit demand from customers. This can be achieved by maximizing:

1. *Time*: Maximizing the number of rides. This means increasing driver retention, especially for frequent drivers.
2. *Supply*: Increasing driver presence during peak times to meet currently unmatched demand (represented by high prime time value in the data), improving customer experience and capturing potentially lost rides.

### Thesis

Through a statistical and visual analysis of driver data and supply-side information, we have identified certain cohorts of drivers (stratified by number of rides, active days, weekend versus weekday, primetime-heavy) and structure in the data that lead us to a lifetime value function  $LTV(driver) = time * supply$  as defined in “Implications for Lifetime Value”. We defer categorizing drivers to each component of the function, which we believe is more informative and less ad hoc.

## DATA ANALYSIS METHODS AND RESULTS

### Definitions and Data Assumptions

#### Definitions

- o Employed days: Number of days between last accepted ride and onboarding date.
- o Active days: Number of days over the employment period in which a driver accepted at least one ride.
- o Lifetime: Aggregate metric over all of a driver’s employed days.
- o Cleaned: Analysis without drivers who on-boarded to Lyft but didn’t accept any rides.
- o Lines: green line is the mean, red line is the median.

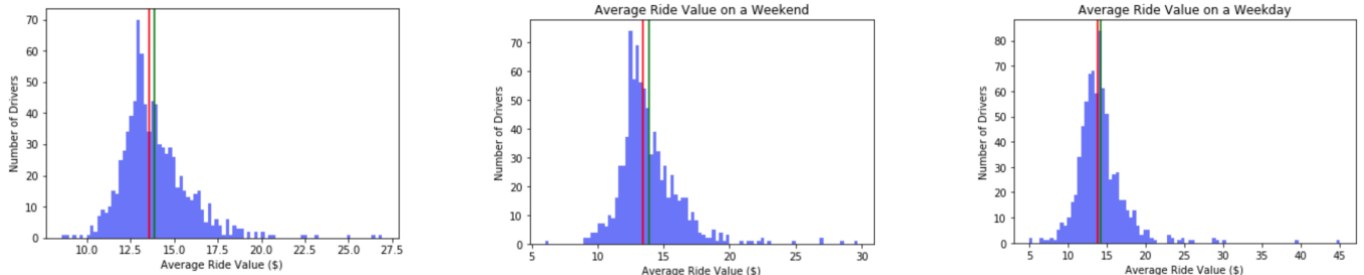
#### Assumptions

- o Missing rides: Certain drivers do not have any ride data but do have onboard dates.
- o Sample window: The data spans a 3 month period. We assume drivers who had a ride within the last week of this period are still active and have not quit driving for Lyft.

### General Analysis to Understand Basic Driver Behavior

#### Preliminary Monetary Analysis

Figure 1. Aggregate average ride value, average weekend ride value, and average weekday ride value for each driver.

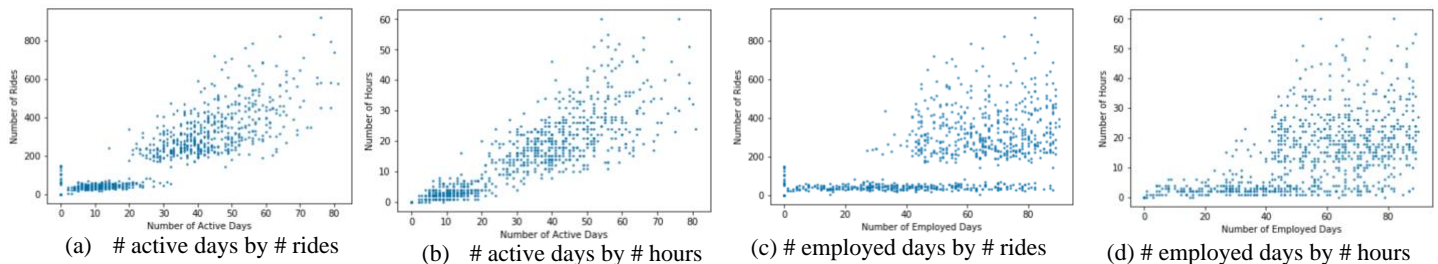


(a) Histogram of average ride value (\$) (b) Histogram of weekend average ride value (\$) (c) Histogram of weekday average ride value (\$)

The aggregate average cost of a ride is \$13.89, while on the weekend it is \$13.86 and \$14.15. These histograms are unimodal due to the fact that a driver cannot see the duration and destination of a rider before they accept, so there are no demonstrated preferences.

### Total Active versus Total Employment Days

Figure 2. Relationship between a driver’s active and employed days and number of rides accepted and hours spent driving.



(a) # active days by # rides

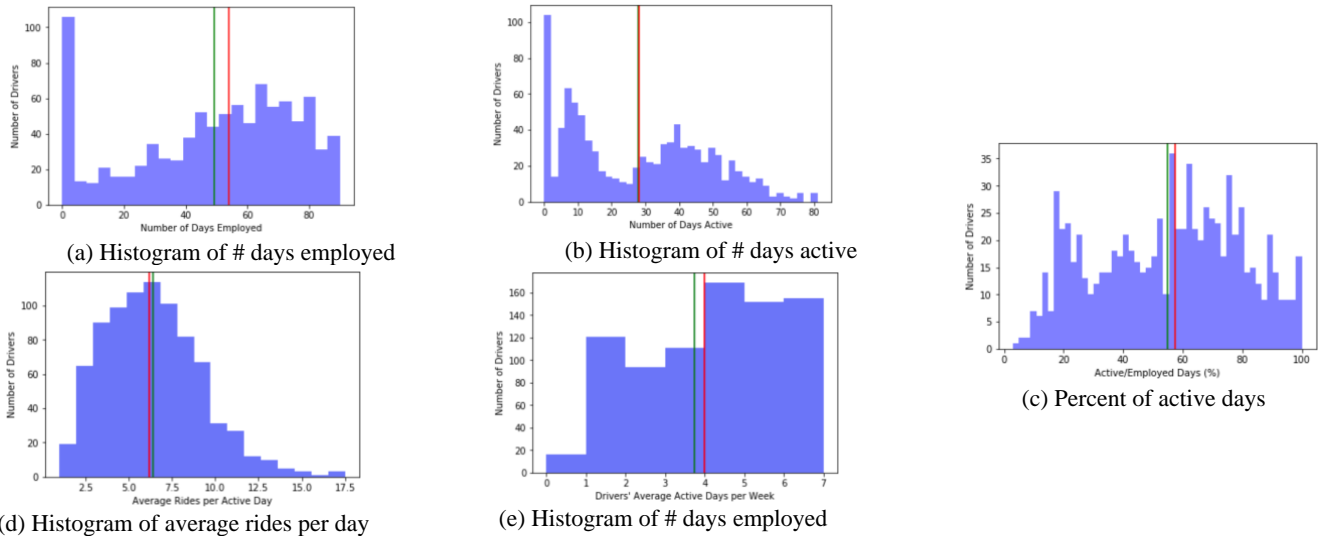
(b) # active days by # hours

(c) # employed days by # rides

(d) # employed days by # hours

Figure 2 illustrates the clear difference between using active days versus employed days as a component metric of lifetime. Active days are better correlated with the number of rides and the number of hours than employed days, as drivers have sporadic driving patterns and do not drive at the same frequency. Based on this, active days will be the ongoing measure of time with Lyft. In this dataset, the average employment lifetime of a driver is 49.2 days and the average active lifetime of a driver is 27.9 days (55.1 and 31.2 days, respectively, with cleaned data). However, many drivers are active at the end of the 3 month period.

Figure 3. Histograms of all driver's active (a) and employment days (b), percent of active days (c), average rides on an active day (d), and average active days per week (e).

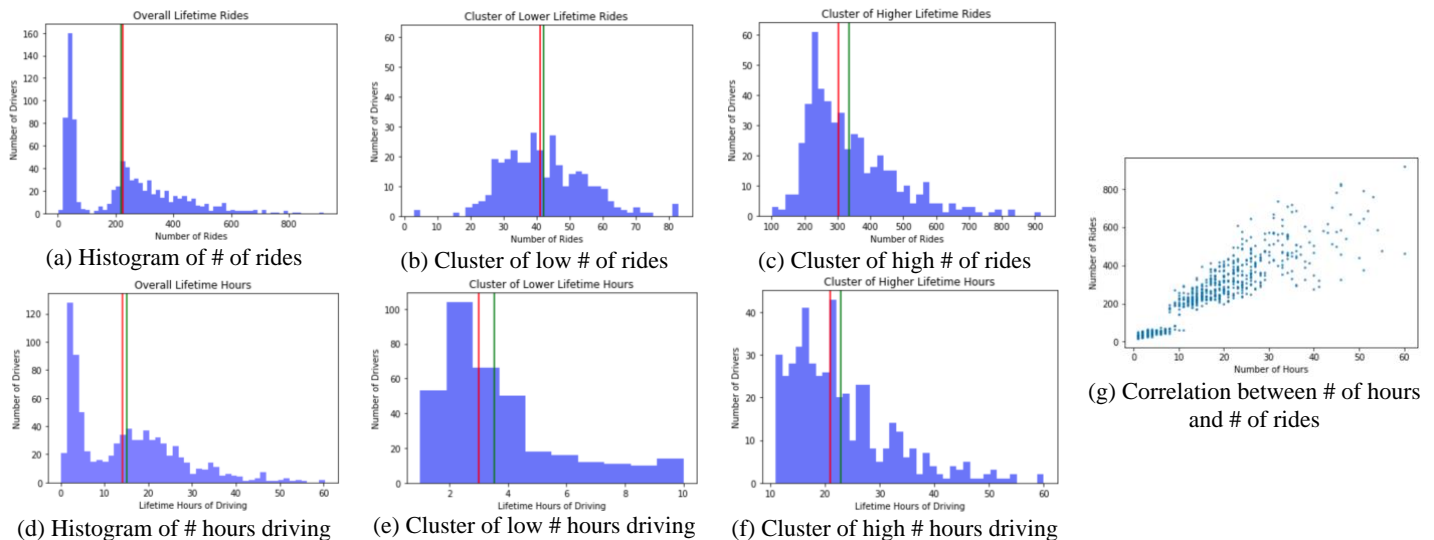


We first notice several drivers who have 0 rides in 3(a) and 3(b). We assume this is because of incomplete data, but give business suggestions if this is not. Henceforth, we will ignore this set of drivers. From 3(b), there are clearly two cohorts of drivers, corresponding to the first peak and the rest of the graph: those who quit after approximately 10 active days and those who stay longer. The first peak might be due to those who joined solely for a signing bonus or other promos and subsequently left, or those who tried driving for Lyft and did not like it.

3(c) shows that there isn't a clear separation between consistently active drivers and drivers who are less active, although relative peaks occur approximately every 20% increase in active days. 3(d) shows a mean number of active rides per day of 6.39 and 3(e) shows the average active days per week to be 3.74.

#### Lifetime Number of Rides Accepted and Lifetime Number of Hours Driving

Figure 4. Histograms of lifetime number of rides and number of hours driving and a scatterplot of the relationship between hours and rides.

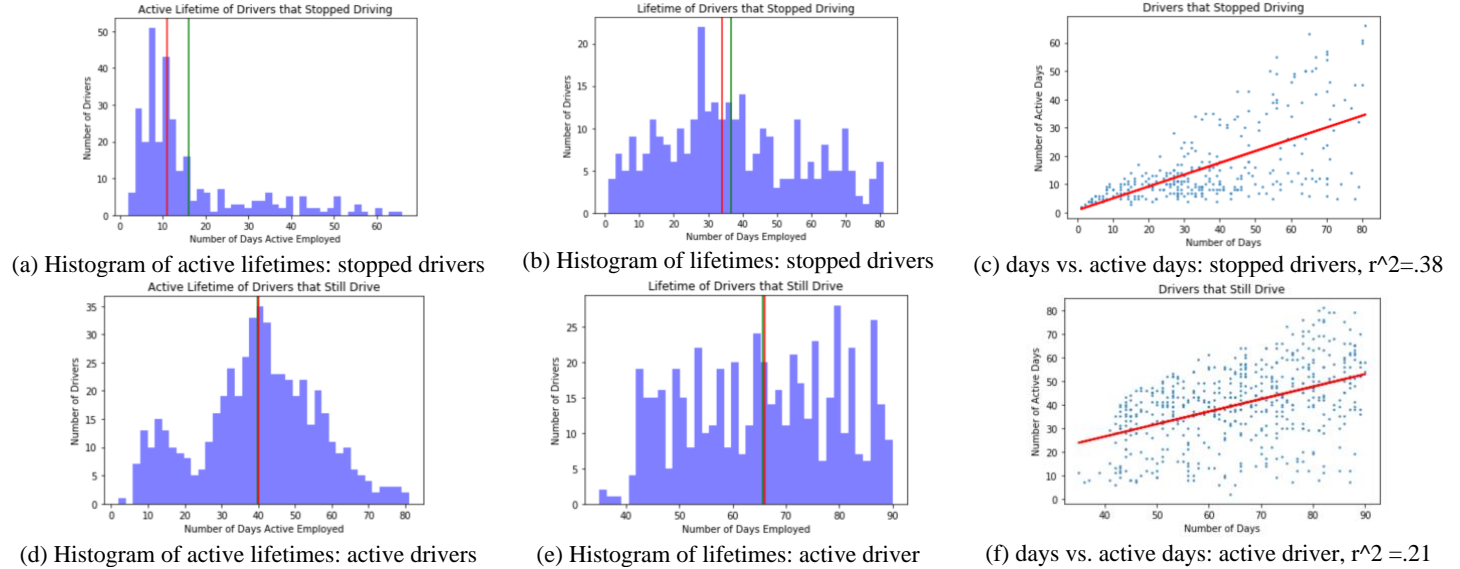


Similar to the number of active days, there are clear cohorts of drivers who have accepted a greater number of rides and driven a greater number of hours. There is a strong positive correlation between hours driven and rides accepted and 4(g) has two cohorts: one with hours 0-10 with very few rides and the rest of the data with many rides and > 10 hours.

### Projected Lifetime of a Driver

In order to get a good estimate for driver lifetime value, we should be able to get a good estimate of a driver's lifetime. Since the data terminates after three months, taking the mean of the data does not yield expected lifetime. Instead, we must extrapolate with the definition that a driver is active if they had a ride within 7 days of the last ride datapoint.

Figure 5. Histograms of all driver's lifetime active and employment days stratified on active status.



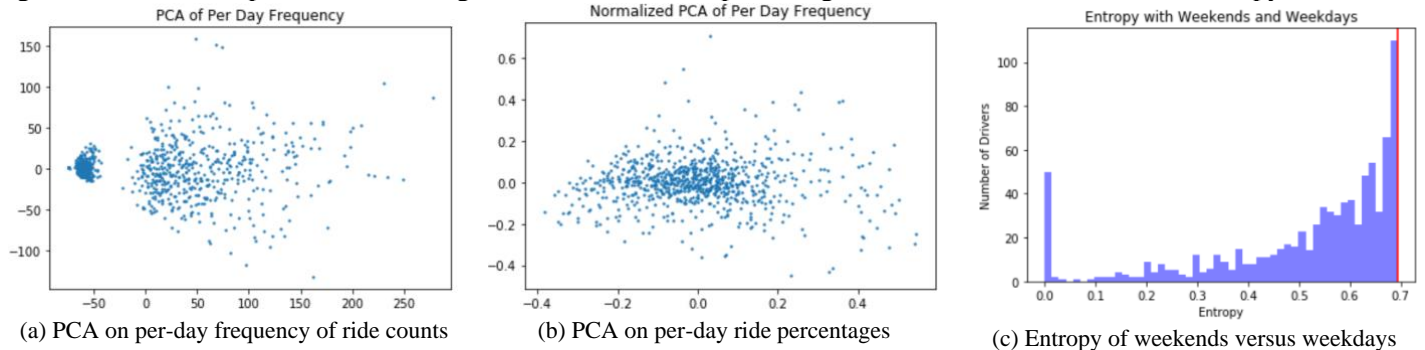
Similar to the claim made by Figure 2, active days have a lot of structure, but employment days do not. In this light, we see active days and employment days are not well correlated.

In looking at active drivers which left 5(a), we see an initial peak which matches the previous graphs which contains roughly approximately 28% of drivers, which we contend are those who weren't interested in driving or were incentivized only by initial promos. After that, we see a steady drop rate of approximately .5% of drivers. As for drivers who are still driving 5(d), we see sizable variance. This is likely because drivers have varying start dates in the dataset but all cut off at the same time, so the amount of data for their trajectory varies per driver.

Thus, based on the 28% initial dropoff within the first 20 active days and a .5% decay rate after, we approximate the expected driver lifetime as  $.28 * 10 + .78 * 1/.005 = 158$  days, which is based on the two partitions and a geometric tail end distribution. Unfortunately, as the dataset only has 3 months data, it is difficult to ascertain long term driver behavior and the tail end of the distribution is very hard to predict, so we can only really make claims about the initial dropoff.

### Weekend versus Weekday Drivers

Figure 6. PCA decomposition and histogram of weekend ride percentages. The red line indicates max entropy.



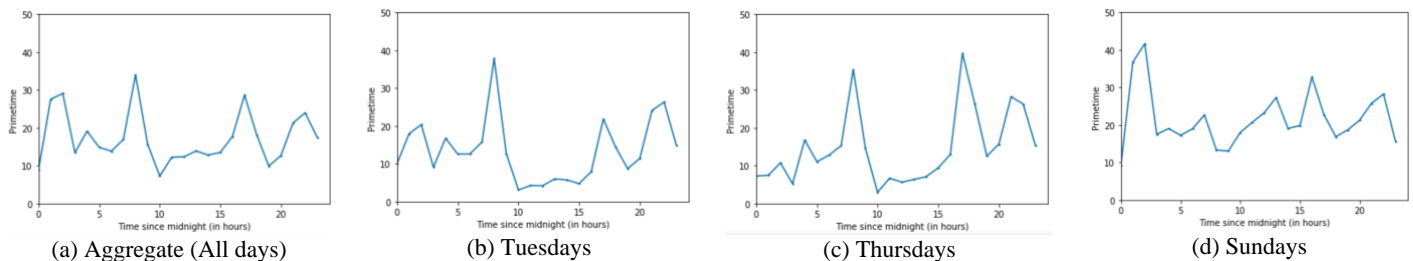
We used PCA, an algorithm that determines important underlying factors of a dataset, on ride counts of each day to better understand driving patterns. The results turned out to be fairly interpretable: the original results in 6(a) show the most significant factor is a nearly-uniform vector (indicating count) and the second most significant is a vector positive on weekends and negative on weekdays, indicating weekend structure. After normalizing and switching from counts on each day to percentage of rides on each day in 6(b), weekends is now the most significant structure and the second a vaguely interpretable result relating to Fridays. This means there is clear structure and different driving patterns because of weekends but no clear categorization or separation.

Additionally, looking at the entropy in probabilities of weekend vs. weekday shows a massive spike at 0. This low-entropy value indicates very structured data, so many drivers only drive on weekdays or only on weekends.

### Prime Time Drivers

Prime Time, activated when the demand for rides is higher than the drivers available<sup>1</sup>, can be thought of as representing the difference between actual and fulfilled demand, or  $\text{Prime Time} = c * (\text{true demand for rides} - \text{rides fulfilled})$ .

Figure 7. The average primetime-based unfulfilled demand curve over time of rides taken in specific time periods.

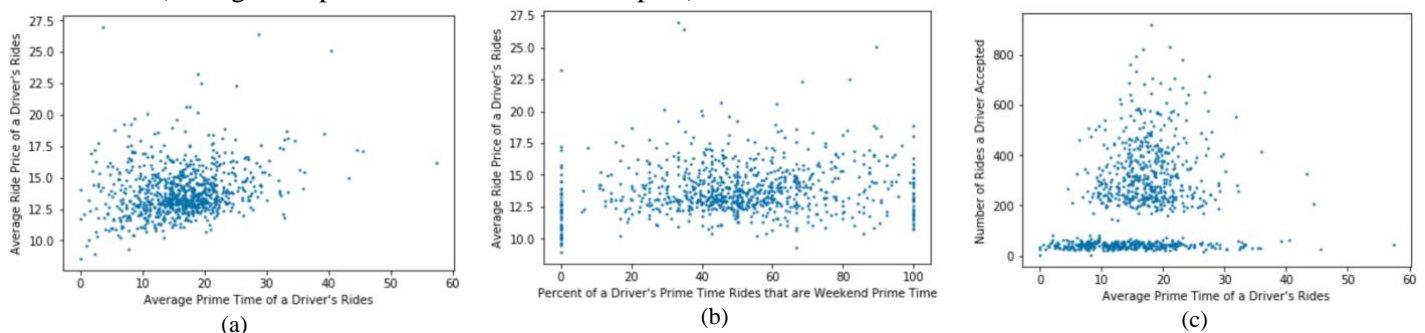


As shown in 7(a), there are generally 4 significant peaks: late-night peaks for weekends at the beginning and end of the day, and morning/ evening rush peaks for weekdays. For the first half of the weekday, the average morning rush prime time is higher, and in the second half, the average evening rush prime time is higher.

### Prime Time Driver Analysis

In this analysis, average prime time is calculated as the mean of prime time bonuses on all a driver's rides. We can see in 8(a) that there is no significant correlation between drivers with high average prime time and average price. 8(b) shows the tendency of a driver's primetime rides (defined as prime time bonus > 0) to be on a weekend versus weekday. It shows a clear separation between the drivers that drive primetime only on weekdays, anytime, or only on weekends. Interestingly enough, the average ride price is higher for drivers who only drive primetime on weekends, pointing to an unmet demand for weekend drivers. Finally, 8(c) shows that for drivers with many rides, the average prime time is lower (especially at the extreme). This shows a cohort of drivers who don't "chase prime time bonuses" and accept a greater array of rides. This is as compared to the band of low-rides at the bottom, which have a greater variance in average prime time value.

Figure 8. Different comparisons between prime time metrics (% prime time rides, average prime time bonus) and lifetime value metrics (average ride price, number of rides accepted).



<sup>1</sup> <https://help.lyft.com/hc/en-us/articles/115012926467-Prime-Time-for-drivers>

## IMPLICATIONS FOR LIFETIME VALUE

### Function Definition

$LTV = time * supply$ , where  $supply = 1 + \frac{average\ prime\ time}{50} + \frac{weekend\ percentage}{2}$  and

$Time: (\# of\ rides\ per\ active\ day) * (expected\ number\ of\ active\ days) * \left( \left( 1 + \frac{active\ days}{employed\ days} \right) * 0.5 \right)$ .

*Time (Measure of frequency of driving):* The base component is the number of rides per day times the expected number of days. We then boosting those who drive more frequently with a multiplier, as a long sporadic lifetime is not as valuable.

*Supply (Optimizing supply curve):* We give a multiplier based on the average prime time for a driver and a percentage boost to weekend drivers, as we see there is a need for more drivers then. Notably, we do not factor in prices as based on Figure 8(a), there is not much correlation between price and prime time. We give preference to prime time as while more expensive rides are more lucrative, it captures length of ride more than the demand curve.

One attractive property to this approach is if the expected number of active days can be accurately estimated, then a sample window can be used to estimate the lifetime function, allowing for more sophisticated business decisions. One way to calculate expected number of active days is to use the geometric distribution approach above, but more long term data is needed to refine this.

Figure 8. Distribution of driver lifetime values in aggregate and averaged per day.



## MAJOR BUSINESS TAKEAWAYS

### Increasing Driver Satisfaction to Increase Retention

Mitigating the underlying causes would increase the lifetime value of each driver and make the high on-boarding costs and bonuses sustainable.

#### *Increasing Retention for New Drivers*

If the datapoints where drivers register but don't ever drive are genuine and not because of a lack of data, running a survey or sending push notifications might help keep them engaged.

#### *Increasing Incentives for Steady Drivers*

From Figure 5(a), it appears that after the initial decline of drivers, the rate at which drivers drop out is nearly constant. There could be a few reasons for this decline, including a bad experience with a passenger, another job opportunity, or general burnout. The first step would be to exit survey these drivers and identify why they are leaving. Using incentives and bonuses staggered over every 10 active days for a period of 50 days might be one way to continue to incentivize drivers.

#### *Matching Demand*

From the prime time analysis, it seemed that there was unmet demand for drivers, specifically on weekends. One solution may be to ask drivers to pre-schedule the time they will drive in pre-designated high-demand areas to secure a supply of drivers for a small bonus. This reduces the need for high primetime fares which discourage users and high overhead for Lyft to find drivers at high-demand times.