Lyft Data Challenge Findings

Prithvi Kannan and Harsh Chobisa

Team databruins

# Summary

The amount of total revenue a driver brings to Lyft is controlled by two factors: how frequently they drive and the average revenue collected per ride. We theorized that a “smarter” driver could bring in more money by driving at optimal times or by knowing the best routes from experience. This is, however, not the case. The average driver brings in $12.66 per ride, and, according to our data, it is clear that this value is not heavily influenced by what time riders drive, how long they’ve been driving, or how frequently they seek out prime rides. Thus, since the majority of drivers bring in approximately the same revenue per ride, “driving smart” seems to be a fictitious concept: drivers can only increase their total revenue brought in by driving more. Thus, a driver is more valuable to Lyft only if they drive more frequently or more extended periods of time.

# Driver Lifetime Value

The average driver lifetime value (LTV) is $4573.723881. The methodology for calculating LTV is the product of average ride value, average number of rides per week, and average projected lifetime. These three quantities represent.

12.69565217 \* 25.92646319 \* 13.89541914 = **4573.723881**

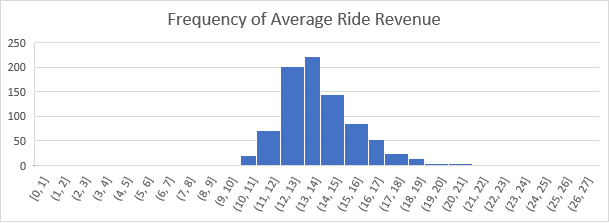
## Average Projected Lifetime

To calculate the driver’s lifetime, we want to know the average duration between a driver’s first and last ride. We chose this measure of lifetime because it was best modeled by our data, [TODO: add more]. For computational simplicity, we are measuring lifetime in weeks (as opposed to days or another more granular measure). To determine the start of a driver’s life with Lyft, we consider the date of the driver’s first ride, rather than the onboard date, since an onboarded driver who has not yet driven has zero value to Lyft. We observed that no drivers are onboarded after week 7 (except for 2 drivers for week 8). Filtering by “cohorts” of drivers from each onboarding week (1-7), we noticed that the many drivers are still driving by the 13th week, which is the end of our data window. The chart below shows how many drivers from each cohort (c0, c1, etc.) remain after each week.

To account for our limited duration of data, our approach considers the rate at which drivers drop from each cohort over time, determined to be approximately 5.75 drivers per cohort per week. Although the data is quite noisy, the linear trendline, y = 5.75, captures the behavior of most cohorts.

## Average Ride Value

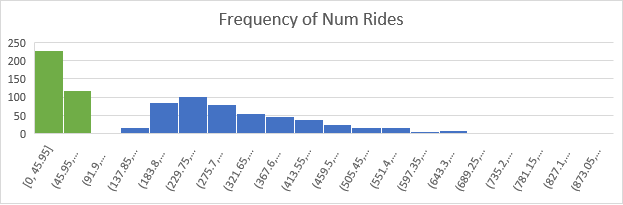
We were able to empirically determine that Lyft drivers are not able to find more valuable rides. The factors that affect ride price are length, duration, and primetime. As a Lyft driver, you are unable to predict the length or duration of a ride prior to accepting. This is supported by a distribution on the histogram of drivers’ average ride revenue. The unimodal nature of the plot indicates that all drivers’ average ride revenue converges to the population’s average ride revenue. If certain drivers were better than others because they were able to find more valuable rides, then we would see a bimodal distribution: one for the naïve driver and one for the value-optimized driver. From this, we observe a mean ride value of $13.89541914/ride.



## Average Number of Rides per Week

To reach the quantity of rides per week, we calculated each driver’s total quantity of rides and divided it by the driver’s number of weeks as an active driver. We assume that for each driver, their number of rides will not vary from week to week, so we chose to use the driver’s *average* number of rides to week.

Plotting a histogram of drivers’ average rides per week yields a bimodal distribution, indicating that there are two groups of drivers: those who “start and drop” (green) and those who “start and stay” (blue).



This segmentation is supported by a scatterplot between the number of days driven versus the number of rides given. We assume that number of days driven is proportional to weeks as an active driver. This shows a group (“tried and dropped” from above) who have drove less than 100 rides.

As a sanity check, it makes sense that the amount lost per week per cohort would be linear, as that would mean the loss as a percentage of the current cohort is increasing. For example, if a cohort has 100 drivers and 10 drop in week 1, that is a 10% loss, whereas if by week 5 the cohort has 50 and another 10 drop, that is a 20% loss. This makes sense as drivers become more likely to drop as time goes on.

If our average cohort size is 146 drivers, and we expect to lose 5.75 drivers per week, that means we expect the whole cohort to leave after 25.4 weeks. Using this data, we estimate that the average number of weeks in a driver’s lifetime is 12.7. Obviously, we know that there are certain drivers who have been with Lyft for years, but the data tell us that those are outliers and that the average lifetime of a driver is around 3 months. Considering many drivers sign up with Lyft when they are between jobs or on some other break, the 12.7 week estimate for driver lifetime seems accurate.

# Business Insights

## Primetime

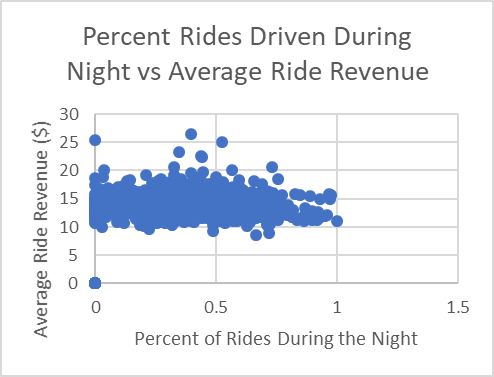
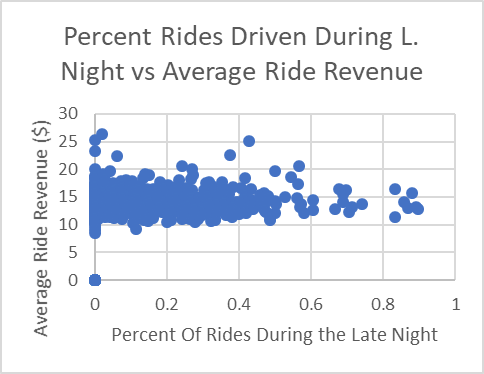
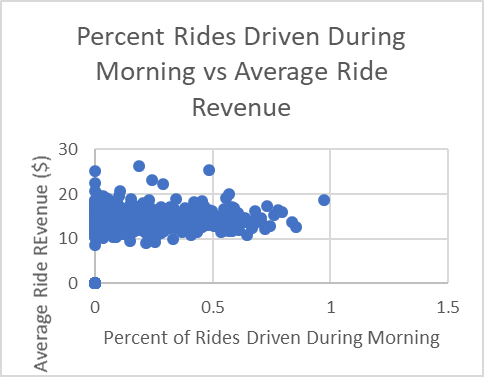
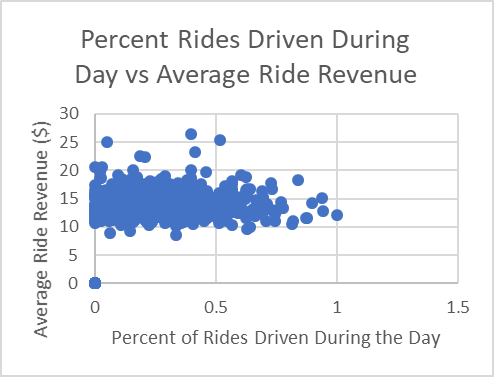
We hypothesized that drivers that drove more frequently during prime time would be more valuable for Lyft, as these drivers would theoretically be making more money per ride because of the surge pricing. However, it is clear from this graph that there is no correlation between the percentage of prime rides given and the average ride revenue of a driver.

## Time of Day Driven

To understand the effect of driving at different times of day on driver value, we classified time of day into 4 categories: Morning (5am-11am), Afternoon (11am-5pm), Evening (5pm-11pm), Late Night (11pm-5am). We hypothesized that rides during the middle of night would be more expensive than rides during the day due to limited supply of drivers an off-hours. However, our data shows at regardless of which time window we look at, the average ride revenue stays the same.

A major caveat in this logic is our pricing model. For this scope of this project, we are using a simplified model that considers duration, distance, and primetime, but does not explicitly consider time of day. We know that Lyft uses a more complicated algorithm to price rides based on supply and demand, but we believe our finding that time of day is not correlated with ride price is still valid since primetime is related to the time of day (more primetimes during morning and evening, less during afternoon and late night).

On the next page are charts of average ride revenue for each of the 4 groups aforementioned.



## Drive Frequency

One factor that we thought may influence a driver’s average ride revenue is the number of days driven. We hypothesize that a driver who casually drives a few days a week would have a lower average ride revenue than a driver who drives everyday because a more regular driver would be able to optimize their rides. However, the data suggests otherwise, as there is no correlation between average ride revenue and the number of days driven. We attribute this to the driver’s inability to find a more valuable ride as discussed earlier in the report.

# What’s Next For Lyft

We know that an average driver’s lifetime value is $4573. Now what? For all parties, it is beneficial to have a higher LTV - the driver makes more money and Lyft can amortize onboarding costs. As discussed prior, there are two major components driving LTV, amount driven and value per ride. Driver’s can only change their amount driven, so Lyft needs to do more to retain their drivers longer.

We recommend implementing a driver loyalty rewards program, where drivers can get bonuses or prizes for being with the company longer. This would incentivize drivers to continue driving Lyft even after they get another job or whatever other reason they drop off, and boost revenues for Lyft.