# Analysis of Predicted Driver Value

Jack Morris and Elaine Shyu

Through our analysis of the data provided we were able to conclude that the average driver has a lifetime of approximately 42 days and will have an average lifetime value of $4173. We arrived at these figures from adapting a predictive model that determines if a driver is still active based on their history of break-times. We were also able to conclude that prime-time pricing played little role in identifying the top performing drivers. Finally, our methods also allowed us to distinguish between full-time and part-time drivers and determine which attributes of an individual's driving history can be used to classify the category they fall into.

## 1. Driver's Lifetime Value

### 1.1 Methodology

First, the cost of each ride was calculated. The distance in meters was converted to miles and the time in seconds was converted to minutes. The distance and times rates were multiplied by these values respectively and added to the service fee. Based on Lyft's prime time policy, that was multiplied on top of the base price. The maximum between this value and 5 was taken, and then the minimum of the resulting value and 400 was returned as the cost of the ride.

Next, the amount of money a driver had earned in total was calculated. The rides that each driver had given were extracted from the data, after which the number of rides a driver had given, the average cost of the rides, and the total amount of money they had earned was calculated. This data was used to analyze the lifetime value of drivers explained further in the next.

## 2. Projected Lifetime of Driver

Using the ride data, a list of accepted times of all rides that a given driver had given was created. The difference between every two timestamps within this list was taken to calculate the time between any two rides given by the driver. This is used to help determine whether a driver is taking a break or whether they have quit relative to the last drop off time data point of the entire set, which is June 26, 2016. An assumption was made that the data set ends on this day.

### 2.1 Driving Consistency

To determine if a particular driver will continue driving, the most important components to consider were the differences between rides of a particular driver. Establishing whether a driver has quit driving for Lyft altogether or is taking a break can be evaluated by comparing the average break-time between each of the rides with the difference between their last drop off and the end of the data set.

#### 2.1.1 Driver Lifetime Heuristics

When predicting the lifetime of a driver (i.e. the time that they are still completing rides at any frequency) we first consider some simple approaches to deciding if a driver was active at the time that data collection stopped. Figure 1 is a visualization of one particular driver's total number of rides over time. We can be fairly certain that

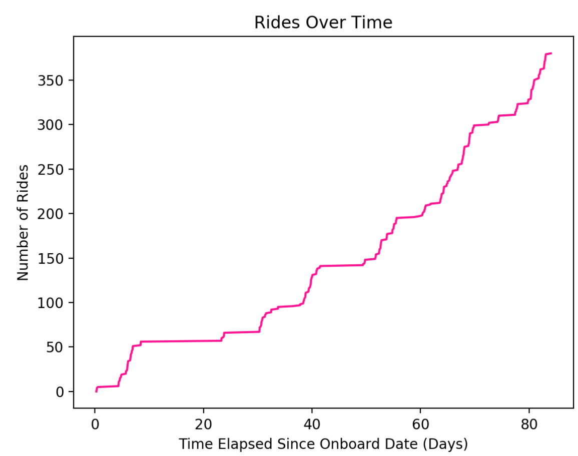


**Figure 1.** Ride Volume Over Time

the driver in question here was still driving with Lyft at the time data collection ended since they appear to have driven one or two days before. However, if data collection ended 20 days after this driver's onboarding date, we would not be as sure because the driver had not driven at all for the prior 10 days. This brings us to the first challenge in deciding whether or not a driver is still active: interpreting the duration of ”break-times”. In particular, analyzing the time between rides or ”break-times”. To do this, we aggregate all break-times among all drivers, while distinguishing the last break-time for each which is defined as the time between the accepting of their last recorded ride and the last time any data was recorded (6/26/2016). In order to identify which drivers have stopped driving with Lyft, we examine the time between their last ride and the last time that data in this data set was recorded. We will call this quantity the last break-time for a driver.

### 2.1.2 Break-time Distribution Estimation

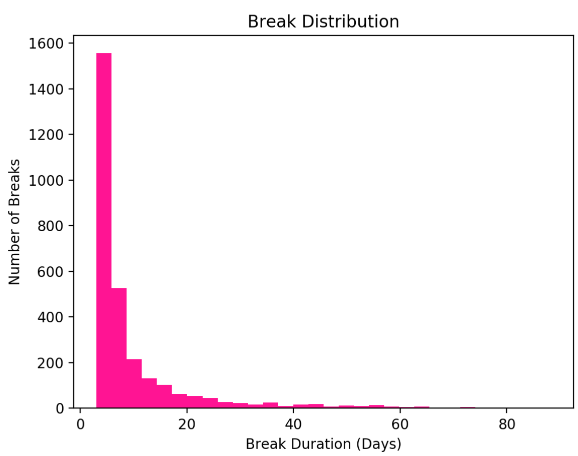


**Figure 2.** Break Distribution

Depicted in Figure 2 is the distribution of all break-times that lasted more than three days. Excluding break-times of drivers of less than three days allows for more weight to be placed on abnormally long breaks that were not indicative of an individual stopping altogether. Our goal is to best identify how population break-times and individual driver break-times are distributed in order to assess whether or not any given final break indicates that the corresponding driver has stopped driving for the foreseeable future.

Now, Figure 2 shows that the distribution of all break durations appears to be approximately exponential. In addition, using the exponential distribution is a natural choice as the memory-less property suits our given observations and our assumptions about the nature of a break-time. Thus, in order to infer the distribution of break-times we let

$$\hat{\lambda}_{total} = \frac{1}{\overline{\mathbf{B}}}$$

where $\mathbf{B}$ is the set of all break-times greater than three days. This is the Maximum Likelihood Estimate (MLE) for $\lambda$. Now, using the distribution $\exp(\hat{\lambda}_{total})$ we can test the final break-time of an individual driver, i.e. find

$$\mathbb{P}(X \geq f_b^D | X \sim \exp(\hat{\lambda}_{total}))$$

where $f_b^D$ is the final break-time of a driver, $D$. With our given distributional assumptions this is the same as finding the probability that any given break that driver $D$ takes is longer than or equal to the break they are currently on, assuming they ever return from this break. However, we also need to account for the fact that $D$ has their own individual break-time habits. Thus, we will also assume that this distribution is exponential with parameter $\hat{\lambda}_D = \overline{\mathbf{B}_D}$, again where $\mathbf{B}_D$ is the set of all break-times for $D$, excluding break-times of less than three days and the final break-time.

Using this information we can construct a method for deciding whether or not a given driver has stopped driving completely or now. Without getting too technical, this function depends on a cutoff probability (referred to as $\varepsilon$ in our code), the probability explained earlier, and the same probability but computed just accounting for the break patterns of that given driver.

Now, we can designate the group of drivers that our function decides to be inactive in order to gain information about the expected lifetime of drivers.

## 2.2 Driver Lifetime Calculation

After applying the decision function to the entire list of drivers we can compute the expected lifetime of a driver simply as the average time between the onboard date and the last ride completed, since we have inferred that the final break of those drivers indicates that they are no longer actively driving. This computation results in a mean driver lifetime of 31.66 days and standard deviation of 17.005 days. These results are visualized in Figure 4. However, these lifetime projections are only for those drivers with a substantial gap between their last ride and the 26th of June.

Another relationship of interest is that between rides per day and projected lifetime. While we do not have a projected lifetime for those driver’s that we predict to still be driving after the 26th of June, we can potentially extrapolate a relationship from the driver’s that we predict to have quit. This relationship has potential for predictability since a high number of rides per day may show that a driver is ambitious and will therefore continue driving for longer. Conversely, low rides per day may indicate little interest in continuing with Lyft.

In order to predict lifetimes for all drivers and not just that stratum. In order to incorporate all drivers from this case study, we will consider the amount of time that each of these individuals were driving and extrapolate from there. As we can see there is a clear divide

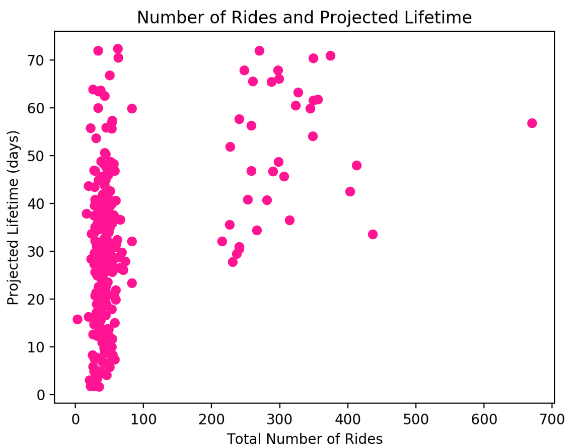


**Figure 3.** Ride Volume and Projected Lifetime

present here, those with less and more than 175 rides (we choose this cutoff since it describes the rest of the data as well). Those in the cluster having 175 or more rides averaged a 50.9 and those under 175 averaged 28.5.

Using these figures and a naive assumption that this approximation will scale linearly, we compute the projected lifetime ($PL$) of a driver as:

$$\mathbb{E}(PL(d)) = \frac{W_{>175} \cdot D_{>175}| + W_{<175} \cdot D_{<175}}{D_{>175} + D_{<175}} = 41.69 \, days$$

Where $W_{>175}$ is the expected number of days a driver that has completed more than 175 rides and $W_{<175}$ is the same, but for any drivers having completed 175 or less rides. These numbers are taken from the drivers that our decision function classified as no longer active drivers. $D_{<175}$ is the number of drivers in this set who drove less than 175 rides and $D_{175<}$ is defined similarly.

## 2.3 Lifetime Value Computations

In order to compute projected lifetime value of a driver we could just multiply our expected lifetime by the mean ride cost across all drivers. However, this will

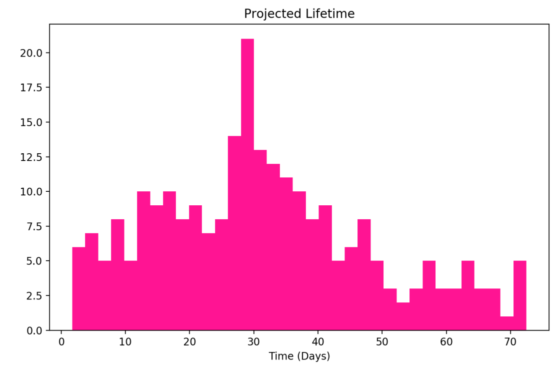


**Figure 4.** Lifetime Projections

yield a projected lifetime value that does not account for several factors including average ride cost changing over the driver's lifetime, the breaks patterns specific to that individual driver, and more. With these in mind, we can still use average cost per ride to compute projected lifetime value. Thus, we have projected lifetime value of \$4173.4. Which is computed with a similar weighted sum as before.

## 2.4 Main Factors in Lifetime Value

Of course, number of rides is the most important factor in an individual driver's lifetime value (as seen in Figure 7 these two correlate highly), but there are several other factors whose relationship with lifetime value deviate slightly from expected. The first of these being time elapsed since on-board date.

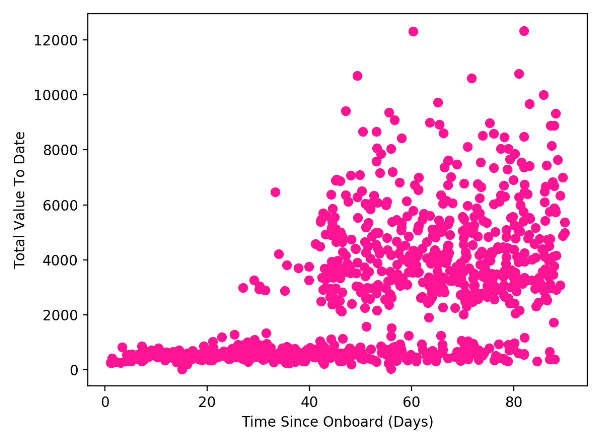


**Figure 5.** Total Value vs Time Driving

As depicted in Figure 5 there is certainly a distinction between those drivers not yet completing a large number of rides and those who have brought in \$3000 or more. Restricting our attention to the upper cluster of higher earning drivers, something interesting to note here is that the lifetime value of these drivers has little relation to the time that has elapsed other than all of these drivers had been driving for 30 or more days by the end of the data collection time period. A possible explanation for this is that top earning drivers happen to be the individuals that are confident of their decision to start driving full-time once they have on boarded.

## 3. Segments of Drivers

In terms of segments of drivers, two main ones are apparent: part-time versus full-time drivers. More broadly, these two could even be classified as those who continue driving with Lyft and those who quit after a few weeks.

### 3.1 Part-time versus Full-time

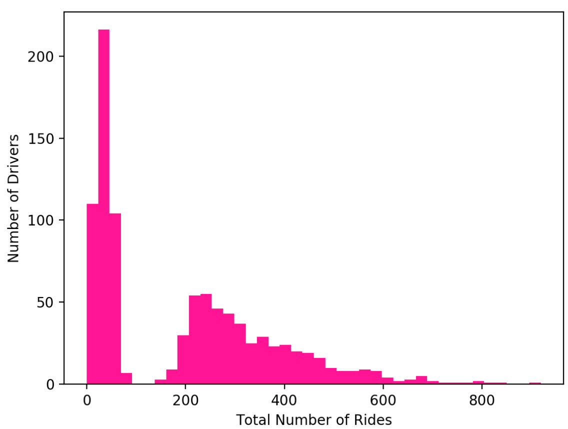


**Figure 6.** Ride Volume

By visualizing the ride volume, or how many Lyft drivers achieved a certain amount of rides in the data set time period, two main groups with peaks at 100 and 250 rides respectively can be seen. Reference to Figure 6. This suggests one segment of drivers is full-time versus part-time Lyft drivers; those that fall near the first peak are part-time, and those on the higher end near the second peak are full-time. The split number is around 175 rides, suggesting that perhaps drivers who are providing more rides than that can be considered as full-time.

The divide between full-time and part-time drivers is made even more apparent in Figure 6 demonstrating the expected causal relationship between number of rides and total value for any particular driver. In addition, this scatter plot shows the two clusters of drivers more clearly separated, those having less than earned less than $2000 those having earned more.

Another interesting relationship is that between total value and number of rides. While there are some outlier

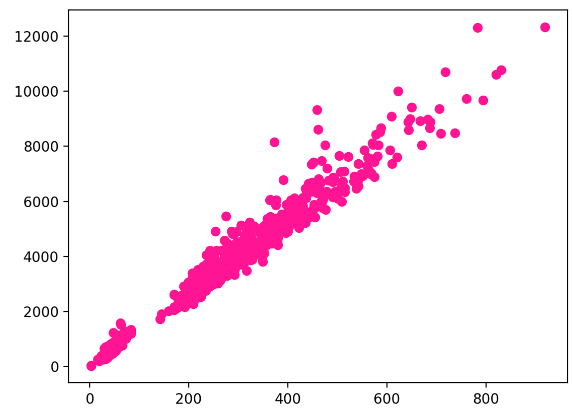

**Figure 7.** Ride Volume (#rides) vs. Total Value (dollars)

top performing drivers bringing in more than $8,000 during this time period, the majority of full-time drivers brought in between $3000 and $5000.

### 3.2 Prime Time Consideration

Since prime time benefits can increase the price of a ride so substantially, it stands to reason that a driver would want to maximize the portion of their rides that have prime time apply to them and therefore, the top earning drivers would have the highest portion of prim time rides. However, upon visualizing the data, we found prime time to play very little role in a driver's value. As

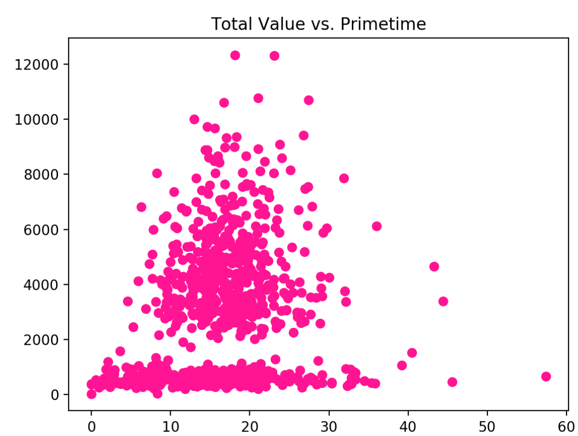


**Figure 8.** Total Value to Date vs. Prime Time

seen in Figure 8, the highest earning cluster of drivers only has about 15% of their rides being affected by any prime time benefits, while several of the drivers with the most prime time rides are those below the $2000 threshold. This was likely the result of those drivers driving very few rides that all happened to be getting them prime time. It is possible that prime time is rare enough or in such confined areas that waiting for or going to a specific location just for prime time rides would not be worth it for full time individuals.

## 4. Actionable Recommendations

As seen in just about every figure and chart in this report, there is a clear divide between the two segments of drivers: full-time and part-time. Clearly, the full-time drivers are worth more to Lyft and this isn't just for shear volume. They will also deliver a consistent volume of rides with fairly standard break patters. The type of consistency and volume present in these drivers will and likely already does bring much value to Lyft.

The biggest challenge moving forward in this domain would be to continue adding drivers that would be considered as full-time. However, we suspect that there is a large segment of current Lyft drivers (as in this data set) who drive minimally part time and constitute the lower cluster of riders that can be seen in this analysis. Encouraging drivers to cross the threshold into full time would likely bring the most value to Lyft. This could be done through bonuses, extra in app privileges (such as preferred access to scheduled rides), and other benefits to be gained after passing and maintaining a certain figure (either dollars or rides) per week.

Additionally, there is a chance that some of the drivers in the lower cluster are also driving with ride sharing competitors, so convincing them to conduct more of their business with Lyft (via the previously mentioned benefits) would help add too the full time driver group and eliminate competition.