Lyft Data Challenge Report

Team Name: Turkey Sausages
Authors: Nilai Vemula and Terry Luo
Date: 15 September 2019

Abstract

Lyft's goal is to optimize their operation by identifying the value of a driver and determining which types of drivers are more valuable to the company. The provided data sets describe 937 San Francisco drivers who started driving with Lyft in Q2 2016 and the rides they gave in that 90-day period. We analyzed this data to determine which drivers quit in that 90-day period and used that information to cluster the Lyft driver population. Four distinct groups of drivers emerge, distinguished by their commitment to Lyft and how much they drive. The 90-day value of a typical driver in each of these clusters varied from $642 to $4,606. By segmenting the driver population, we recommend strategies to target these clusters and increase their respective value to Lyft. To improve retention, we conclude that Lyft should offer incentivization programs starting at 50 rides and employee benefits to drivers that tend to quit within their first three months of driving. In addition, we suggest methods to encourage drivers to work for Lyft over competitors.

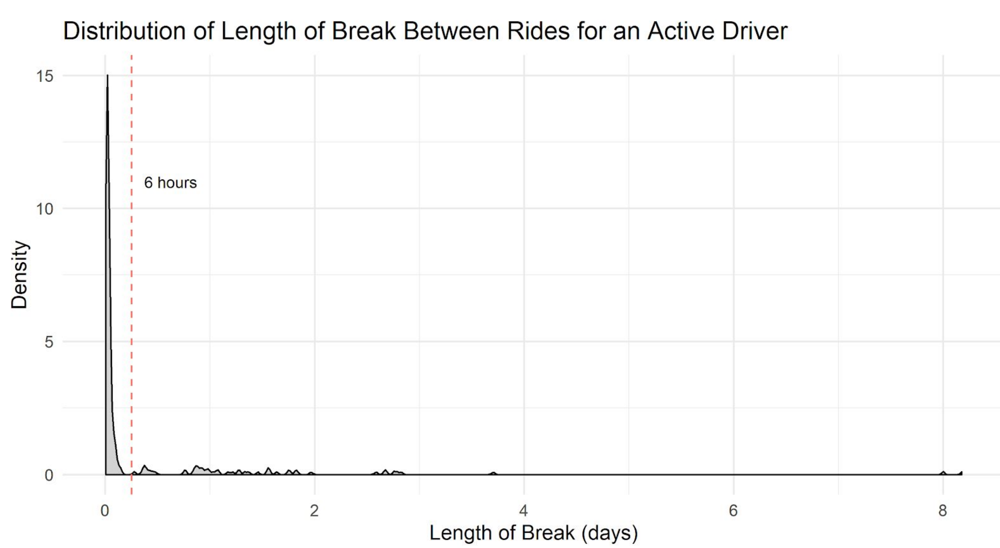
Driver Attributes

Our first step was to join the provided data. After joining, we discovered only 837 of the original 937 drivers had given rides for which we have complete timestamp information for, so we narrowed our analysis to these 837 drivers.

Due to the fact that the provided data spanned only three months, calculating a driver's lifetime value proved to be a challenge. Some drivers clearly worked longer than the given period, making it difficult to calculate anything relating to a "lifetime". What we can do, however, is first determine which drivers quit within these three months and which drivers remained active.

Quitters

We first noted the last timestamp given in the datasets: `2016-06-27 00:50:50 PDT`. Then, we found the time difference between each driver's last recorded ride and the end of the data logging period, which we call the last break. If this last break is some duration such as 50 days, it is easy to conclude that the driver quit, as he or she has no recorded rides in the last 50 days of the dataset. What if the last break was 4 days, or 4 hours?



To answer this question for every driver, we must know more about their driving habits. Do they consistently drive every day or do they tend to take longer breaks? After calculating the difference in time for every ride a driver made, it is clear that there are two types of breaks. On the left is a density plot displaying the distribution of breaks for a sample driver[1].

This data suggests that drivers often work in "sessions" — periods of

[1] Driver ID: 002be0ffdc997bd5c50703158b7c2491

hours at a time when they are continuously looking for rides. The large spike on the left corresponds to all the breaks a driver makes when searching for a new ride in a single session, which is usually on the order of minutes. The smaller peaks to the right represent the breaks between sessions. If a driver actually quit, then their last break should be considerably longer than other breaks between their driving sessions.

After looking at the break distributions for all the drivers, we determined the cutoff between the two types of breaks to be five to six hours. This is corroborated by Lyft's time limit policy, which prevents drivers from driving for longer than 14 hours at a time without taking an uninterrupted six hour break[2].

We can now categorize each break longer than 6 hours as a "break between sessions." We considered the last break a driver took to be significant if it was:

1. greater than two standard deviations above the mean length of breaks between sessions, and
2. greater than or equal to the longest recorded break a driver previously took.

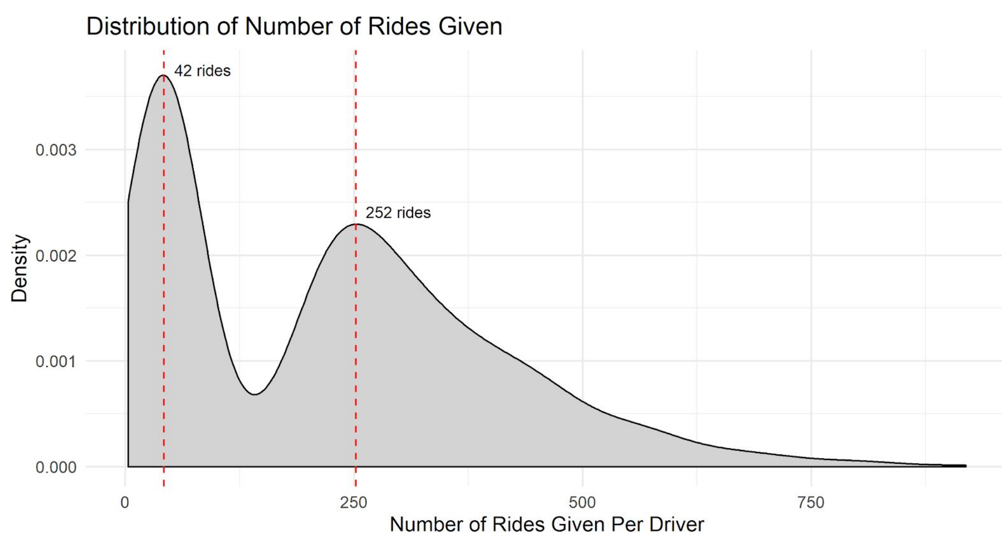
After filtering out all the drivers whose last break was significant, we determined that out of these 837 drivers, 259 (30.9%) of them quit within the 90-day period. The average career length of these quitters was 34 days.

Other Attributes

We continued to extract attributes from the data to describe each driver. An explanation follows the attributes that are not intuitive based on name.

- **Career length (days):** how long it has been since a driver onboarded. This is only relevant for drivers who quit within the 90-day period.
- **Average Ride Duration (minutes)**
- **Average Ride Distance (miles)**
- **Number of Rides Given**
- **Average Eagerness (scaled 0 to 1):** a measure of how "eager" a driver is to take a ride. It is inversely proportional to the time difference between when a ride is requested by the user and when the driver accepts that ride. The driver with an eagerness of 1 is the quickest driver on average to accept a ride.
- **Average Response Time (scaled 0 to 1):** a measure of how far drivers are willing to travel for a ride. It is inversely proportional to the time difference between when a driver accepts a ride and when they arrive at the pickup location. The driver with a response time of 1 is quickest driver on average to arrive at the pickup location.
- **Average Prime Time (weighted by ride duration, 0 to 100)**

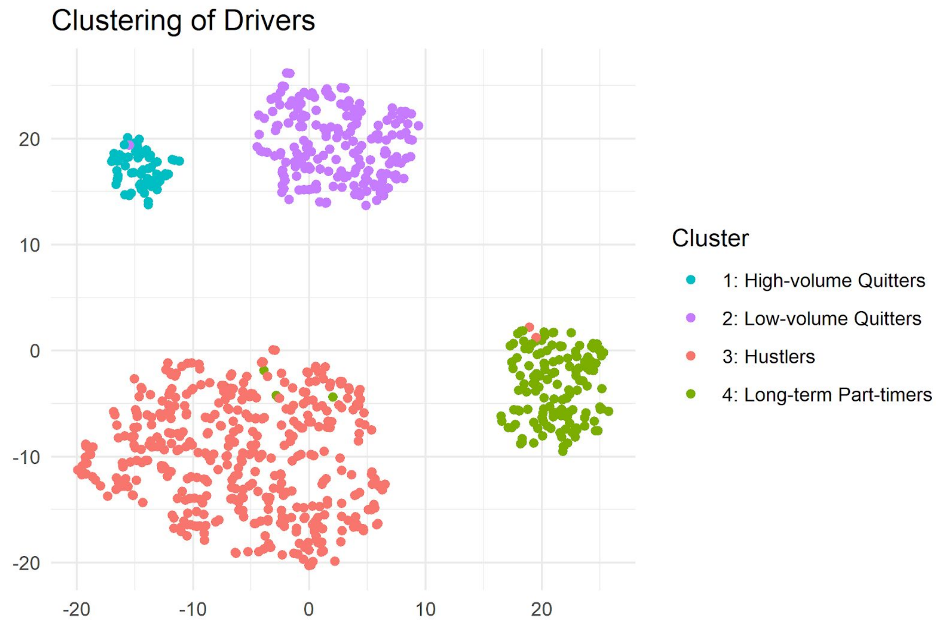
Of these attributes, the distribution of Number of Rides Given was very interesting as it was bimodal. We noted that this distinction between giving many or few rides could be a result of treating Lyft driving as either a full-time or part-time job. We hypothesized that drivers could be classified as either quitting or non-quitting and by high-volume of rides or low-volume of rides given.



At this point, it became clear that our population of drivers could be separated into distinct groups with different properties. We calculated these attributes for each driver in an attempt to group them by PCA (Principal Component Analysis) clustering.

[2] help.lyft.com/hc/en-us/articles/115012926787-Taking-breaks-and-time-limits-in-driver-mode

Clusters



It is important to note our clustering methodology was unsupervised, meaning we did not influence the model by indicating which attributes were more important or how to segregate drivers. Thus, the PCA analysis helps to confirm that these four types of drivers do exist. The following figure and table define our clusters and provide descriptive statistics for each cluster. The results show two clusters of quitters and two clusters of drivers who did not quit in the 90-day period. Groups of quitters and non-quitters can additionally be separated by number of rides given per day, as mentioned earlier.

Note: The cluster naming was done after all analysis was completed and the quitting attribute was merged into the cluster name. All clusters were homogenous in quitters or non-quitters.

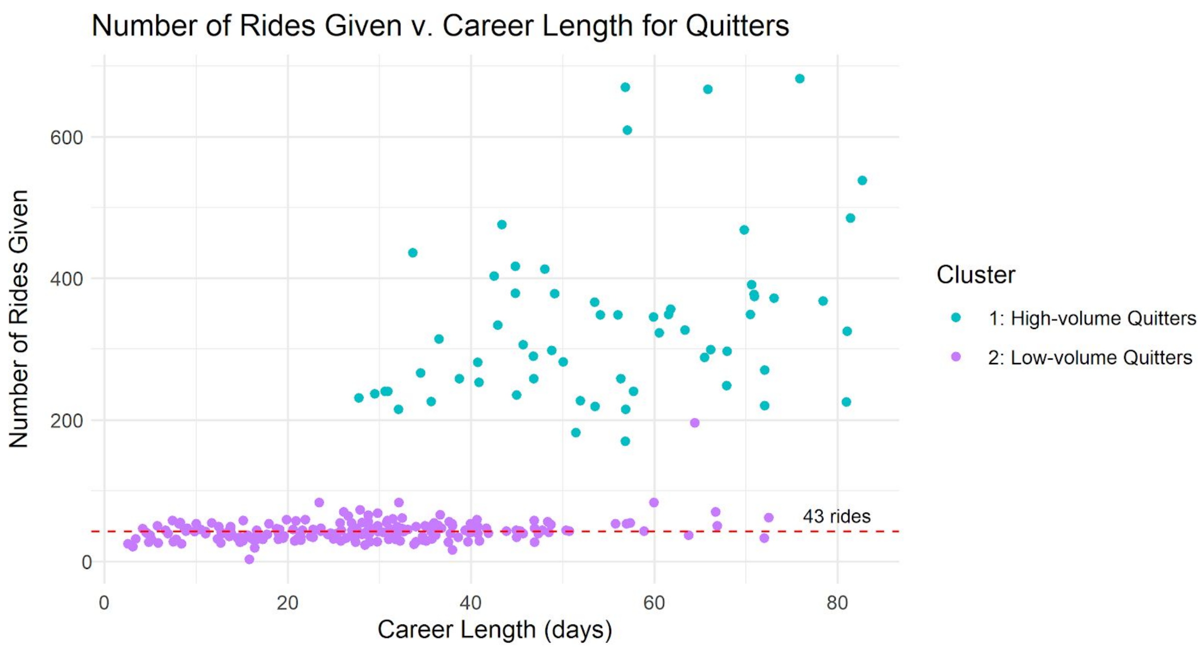
Cluster Name	Average Earnings per Ride (USD)	Average Ride Distance (miles)	Average Ride Duration (mins)	Average Eagerness	Average Responsiveness	Average Number of Rides Given	Average Prime Time Weighted by Ride Duration	Size	Number of Days Worked	Average Number of Rides Per Day Worked
High-volume Quitters	13.60	4.22	14.49	0.69	0.66	335.43	19.42	61	40.95	8.51
Low-volume Quitters	13.56	4.57	14.04	0.67	0.67	43.07	14.97	198	10.75	4.65
Hustlers	13.49	4.28	14.30	0.68	0.69	341.57	17.71	436	45.29	7.62
Long-term Part-timers	14.23	5.09	14.73	0.63	0.65	44.33	14.07	142	13.18	3.88
Average	**13.64**	**4.48**	**14.33**	**0.67**	**0.67**	**220.08**	**16.57**	**837**	**31.36**	**6.35**

We have shown that not all drivers act alike, but it is necessary to determine if this results in differential value for Lyft. The Hustlers clearly make the most total revenue for Lyft, but that is simply due to the high volume of rides they give and their commitment to driving in the long-term, making this point rather uninteresting. So, let us consider earnings when the number of rides is controlled for. In this case, Long-term Part-timers stand out. They make $0.59 more than the average driver per ride. One explanation for this is their speed. By using average ride duration and distance, we found that Long-term Part-timers drove at an average speed of 20.73 mph, compared to 18.72 mph of the average driver. Essentially, these drivers are making more money per ride by choosing slightly longer rides and driving a little faster.

Another group to note is the High-volume Quitters, who average the highest weighted Prime Time per ride. We believe this group is also very important. Although these drivers make close to the average earnings per ride, their willingness to drive more during Prime Time is beneficial for Lyft as a company. In general, if rides in an area are subject to Prime Time pricing, then there are a lot of riders compared to Lyft drivers and other competitors such as Uber. This is a common situation where a rider will compare prices between

Uber and Lyft and choose the lower price. Having more drivers available during busy hours is crucial in controlling more of the market by increasing supply.

That leaves just the Low-volume Quitters. Exclusively analyzing the two clusters of quitters exposes an important distinction between them. Here, the number of rides given by the quitters is plotted against career length:



The High-volume Quitters give a varying amount of rides before quitting. Since the high-volume drivers seem to be driving full time, we suspect that some of these drivers are looking for more traditional jobs and chose to drive for Lyft during their job search. Otherwise, it would not make sense for a full-time Lyft driver to quit — these drivers tend to work more days than average and give more rides than average.

The low-volume group, however, consistently quits after giving around 43 rides. It does not seem to matter whether the 43 rides were completed in 10 or 50 days. This trend suggests another factor is causing these low-volume drivers to quit. For example, these drivers may become uninterested in Lyft after working for a while or after realizing that Lyft is no longer economically advantageous for them. Ultimately, the decision to quit is likely a personal decision for many drivers and more data about the driver population is necessary to better cater to these individuals' needs.

Additionally, the average number of rides per day worked is an interesting statistic. It was much lower than we expected, especially for drivers that we considered to be driving full-time. Even among the Hustlers, the average number of rides per day worked was less than 8. In hindsight, this makes more sense as a large proportion of drivers work for more than one rideshare service. It is fairly common for drivers to choose the service that offers the greatest pay at any given point in time.

Finally, although eagerness and average response time were useful attributes for characterizing each driver, there was not a significant difference in these values between clusters.

Lifetime Values

Lifetime value is defined as the sum of all the calculated ride fares for every ride that a driver gives across their lifetime. This number is simple to calculate for the drivers who quit within the 90-day period as we know exactly which rides they did in their lifetime, equivalent to the product of their average ride fare and number of rides given. This is represented as:

$$\text{Lifetime Value} = \Sigma\ \text{Ride Fare} = \text{Average Ride Fare} \times \text{Number of Rides Given}$$

This was calculated to be **$1523.09** on average.

For non-quitters, lifetime value is a function of career length, a value that we do not know as drivers continue to drive beyond 90-days.

$$\text{Value Per Day} = \text{Average Number of Rides per Day} \times \text{Average Ride Fare}$$

On average, drivers that stick with Lyft long-term generate $40.36 per day. Therefore, lifetime value would be $40.36 multiplied by their career length in days. The 90-day value of non-quitters is **$3632.63** on average.

If we analyze lifetime value in the context of our clusters, similar patterns from before emerge. The High-volume Quitters earned $86.36 per day of their career on average. The Low-volume Quitters earned only a fraction of that, about $29.85 per day of their career. The two clusters of drivers who did not quit also show similar disparities. The Hustlers made $51.18 per day on average and $4606.51 over the 90-day period. The Long-term Part-timers made $7.14 per day on average and $642.42 over the 90-day period.

Business Recommendations

One of the best ways to increase a driver's value for Lyft is stopping a driver from quitting. If we are able to turn drivers who quit early into regular drivers — not only does Lyft make more money, but it also minimizes further onboarding costs. We recommend adding some type of incentive to drivers around the 50 ride mark to encourage low-volume quitters to keep driving. Since quitting after around 43 rides is extremely common, we believe this is a great way to specifically target a group that has the potential to generate a lot more value. For this solution, we also recommend continuous rewards for reaching frequent milestones. If only a 50 ride bonus was announced, it is very possible that drivers will give 50 rides and stop again. A tiered system should be set up that incentivizes low-volume drivers to keep driving.

In addition, it was mentioned that we suspect High-volume Quitters were able to find another job, thus causing them to quit driving. Their extremely high average earnings of $80 per day represent a large loss to Lyft when these drivers eventually quit. To mitigate this, one possibility is to make Lyft driving more similar to a traditional job by treating them like valuable employees rather than independent contractors. One example of a benefit that would greatly improve driver retention is to offer a 401k matching program with a long vesting period. This type of program would encourage drivers to stay with Lyft for a long period of time to ensure they maximize their retirement savings. However, we recognize that more data is needed in order to perform a cost-benefit analysis of such decisions. It’s crucial to note that our goal is to convert these High-volume Quitters to Hustlers while still preserving the high average earnings they originally had.

Our last recommendation has to do with drivers that work for more than one rideshare company. We suggest a feature that incentivizes drivers to take continuous rides with Lyft. For example, Lyft could offer a small increase in pay if a driver accepts another ride through the Lyft app shortly after dropping off a rider. If the driver feels pressured to make a decision between Lyft and other rideshare services, a small bonus could go a long way. This can also help Lyft gain market share in competitive markets.