

Lyft Data Challenge Report

Team Name: Turkey Sausages

Authors: Nilai Vemula and Terry Luo

Date: 15 September 2019

Introduction

Lyft's goal is optimize their operation by identifying the value of a driver and by determining which types of drivers are more valuable to the company. This report describes a method of using a set of data describing 937 San Francisco drivers who started driving with Lyft in the second quarter of 2016 and the rides they gave in that 90-day period to determine which drivers are quitting in the 90-day period as well as clustering the Lyft driver population into groups driven by their commitment to Lyft and their driving frequency. By segmenting the driver population, we can recommend strategies to target these clusters and increase their value to Lyft.

Driver Attributes

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 three months for which we have data for, making it be very difficult to calculate anything relating to a "lifetime". What we can do, however, is first try to determine which drivers quit within these three months and which drivers were still 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. Fig.1 is a density plot displaying the distribution of breaks for a sample driver:

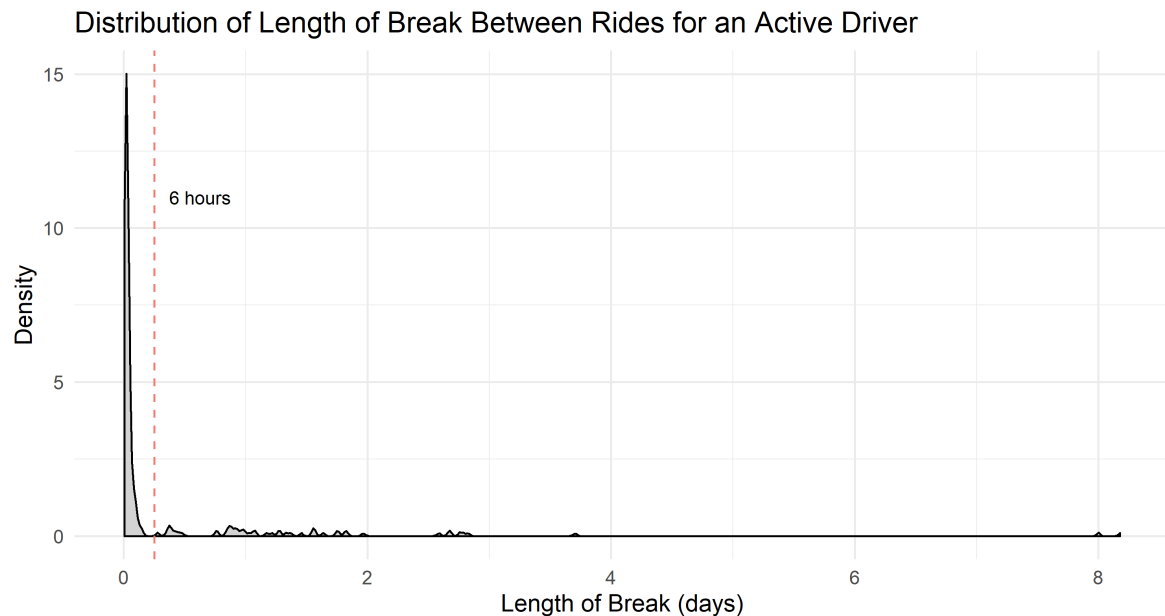


Fig. 1: The distribution of the length of breaks between rides for driver "002be0ffdc997bd5c50703158b7c2491." The red line delineates breaks inside a driving session from breaks between driving sessions.

This data suggests that drivers often work in "sessions" — periods of 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 5-6 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 6 hour break. [1]

We can now categorize each break longer than 6 hours as a "break between session." 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 session breaks, and
2. at least as long as the longest recorded break between sessions a driver took.

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

Other Attributes

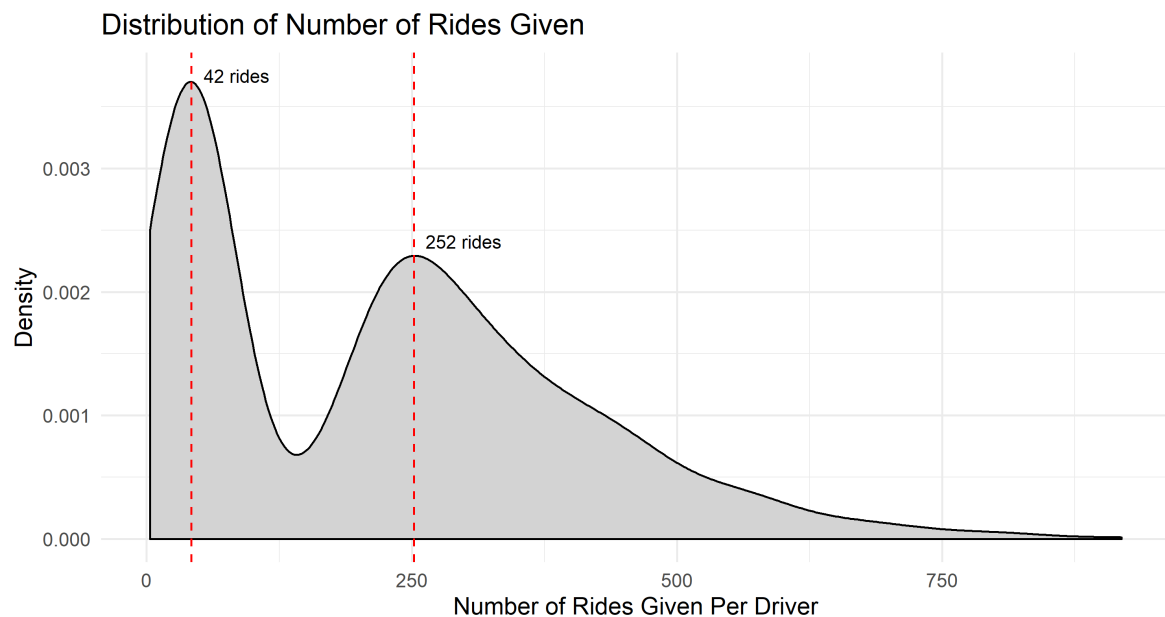
We continued to extract attributes from the data to describe each driver. An explanation follows the attributes that are not intuitive based on name.

- **Career length (days):** how long it has been since a driver onboarded. This is only relevant for drivers that quit. It is not applicable for drivers that are still working.
- **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 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.



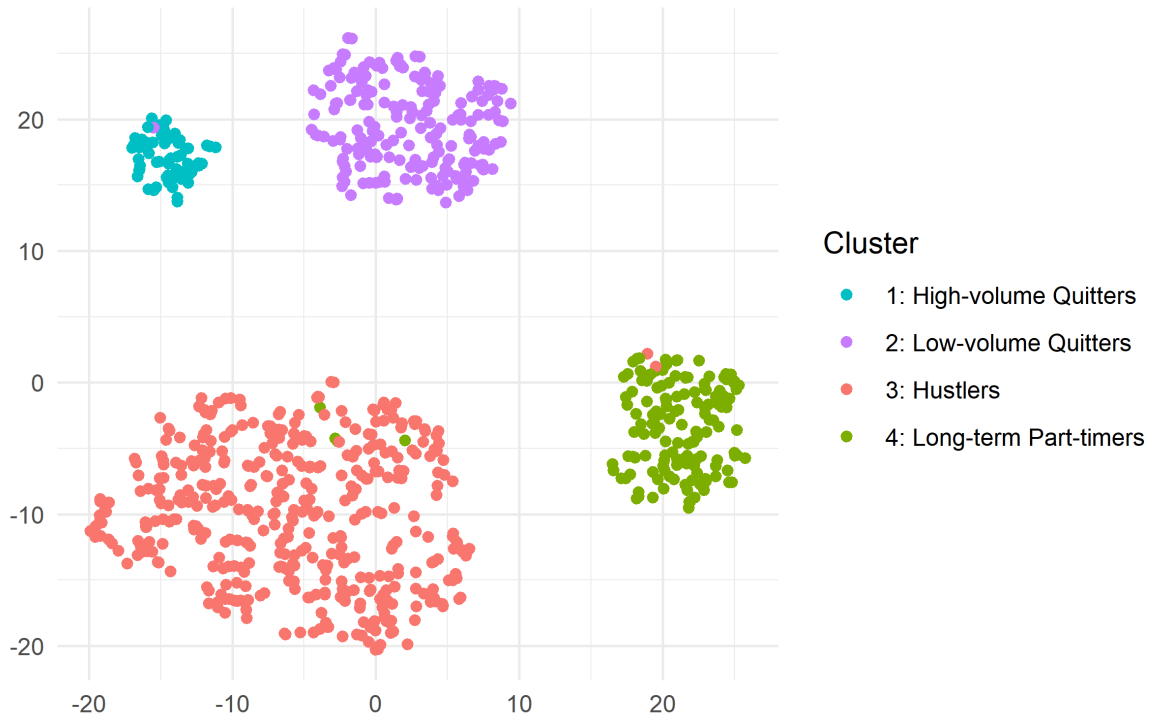
We noted that this distinction between giving many or few rides could be a result of treating Lyft driving as either a full-time or part-time job. We hypothesized that that drivers could be classified by either quitting or non-quitting and by high-volume of rides or low-volume of rides given. Additionally, we believe these two factors are independent.

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

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 that did not quit in the 90-day period. Additionally, in each of these groups, there is a group that gives a lot of rides and a group that does not give as many rides.

Clustering of Drivers



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	9
Low-volume Quitters	13.56	4.57	14.04	0.67	0.67	43.07	14.97	198	10.75	5
Hustlers	13.49	4.28	14.3	0.68	0.69	341.57	17.71	436	45.29	8
Long-term Part-timers	14.23	5.09	14.73	0.63	0.65	44.33	14.07	142	13.18	4
Average	13.64	4.48	14.33	0.67	0.67	220.08	16.57	837	31.36	6

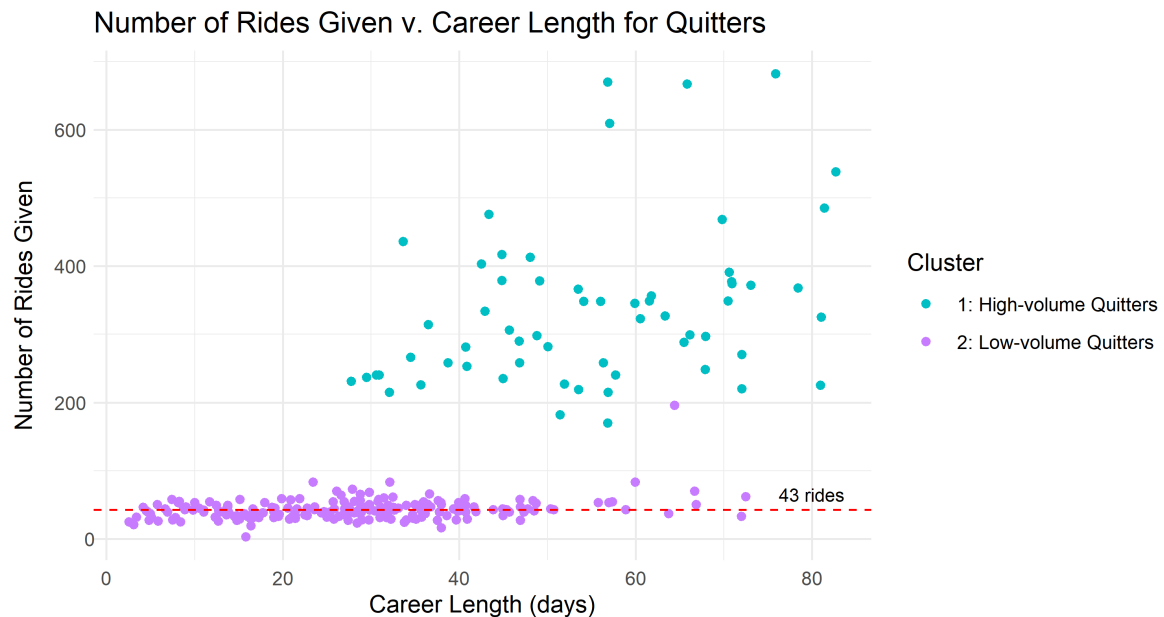
Note: The cluster naming was done after all analysis was completed and the quitting attribute was merged into the cluster name - all clusters were homogenous in quitters or non-quitters.

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 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 towards 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 drivers in Lyft 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. Looking at the two clusters of quitters by themselves yields more surprising results. 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 an example to describe these drivers are those looking for more traditional jobs who chose to drive for Lyft between jobs. It would not make sense for a full-time Lyft driver to quit otherwise given that this segment of the population tends to work more days than average and give more rides than average. The low-volume group, however, consistently quits after around 43 rides, regardless of career length. They drive less often and complete only half as many rides as their high-volume quitter counterparts. 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.

Lifetime Values

The lifetime value of a driver that quit within three months is calculable since the data for the entire lifetime is available — it is the sum of each ride's earnings. We are going to assume that Lyft's revenue is proportional to the amount of money that a driver makes based on the ride fare formula. This was calculated to be **\$1523.09** on average.

For non-quitters, lifetime value is a function of career length. 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.

Business Recommendations

One of the best ways to increase a driver's value towards Lyft is stopping a driver from quitting. If we are able to turn drivers that quit early into regular drivers - not only does Lyft make more money, but 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 40 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.

In addition, it was mentioned that we suspect High-Volume Quitters were able to find another job, thus causing them to quit driving. To mitigate this, one possibility is to make Lyft driving more similar to a traditional job and treating them like valuable employees rather than contractors. One example of a benefit that would greatly improve driver retention is to offer a 401k matching program with a long vesting schedule. However, we recognize that more data is needed in order to perform a cost-benefit analysis of such decisions.

Citations

1. <https://help.lyft.com/hc/en-us/articles/115012926787-Taking-breaks-and-time-limits-in-drive-r-mode>