Lyft Data Challenge 2019

Steel-Cut Oats: Kelly Chen, Michelle Lee

We would like to thank Lyft, Inc. for giving us this opportunity to compete in the Lyft Data Challenge. We had a lot of fun working with the data and hope that you will find our work insightful.

Introduction:

We were tasked with analyzing data on Lyft drivers to calculate their Lifetimes and Lifetime Values and identify segments of drivers with higher performance in order to make actionable recommendations to Lyft. We extracted and analyzed data from the three separate data tables: driver_ids, ride_ids, and ride_timestamps. There is data present for 1020 Lyft drivers, 937 of whom are identified by a unique Driver ID. Each driver's onboarding date and ride history data was provided.

We found that the first month of driving was most critical to employee retainment. Drivers that failed to complete approximately 90 rides within the first month are not likely to become frequent, long-term Lyft drivers. Moreover, the amount of Lyfts driven per day is strongly positively correlated with Lifetime Value, indicating that consistency is a strong factor in Lifetime Value. However, Prime Time Status was not strongly correlated with Lifetime Value. From these observations, we have several actionable recommendations.

Explanation of Our Approach:

1. **Data Cleansing**

Before we attempted to analyze the data, we started by cleansing it to find holes and determine which variables of the given data set we wanted to use. The missing values we found were excluded from their respective calculations. For example, 9262 rides were missing data for distance, duration, and prime time status and were thus excluded from calculations for Lifetime Value, Rides per Day, Driver Onboarded/Day, Onboard Date, Lifetime Revenue, and Percent of Rides with Prime. 8683 rides were missing timestamp data for all five events and were omitted from calculations of Driver Lifetime.

The other 579 rides lacked data for distance, duration, and prime time status. 1 ride of the 193502 sets of ride data had an invalid, negative duration measurement and was thus omitted from (specific measurement).

In addition to omitting certain pieces of data for calculations, we manipulated and combined the given data tables to help in our calculations. In our code, the primary dataframe we have is lyft_data, which includes all the columns from the original driver_ids and ride_ids data sets. Lyft_data also includes timestamps for each ride sorted by the factors from the *event* column of the ride_timestamps dataframe.

2. **Finding Summary Statistics**

To get a bigger picture of the data, we found some basic summary statistics of noteworthy variables. We were provided data about 1020 unique Lyft drivers and 202748 unique rides. "n" refers to the amount of data was provided and usable.

Rides Data

Variable	n	% Omitted Data	Minimum	Maximum	Median	Mean
Duration (s)	202748	0%	2	28204	726	858.9
Distance (m)	193315	4.652%	1	724679	4020	6962
Prime Time (%)	193315	4.652%	0	500	0	17.31

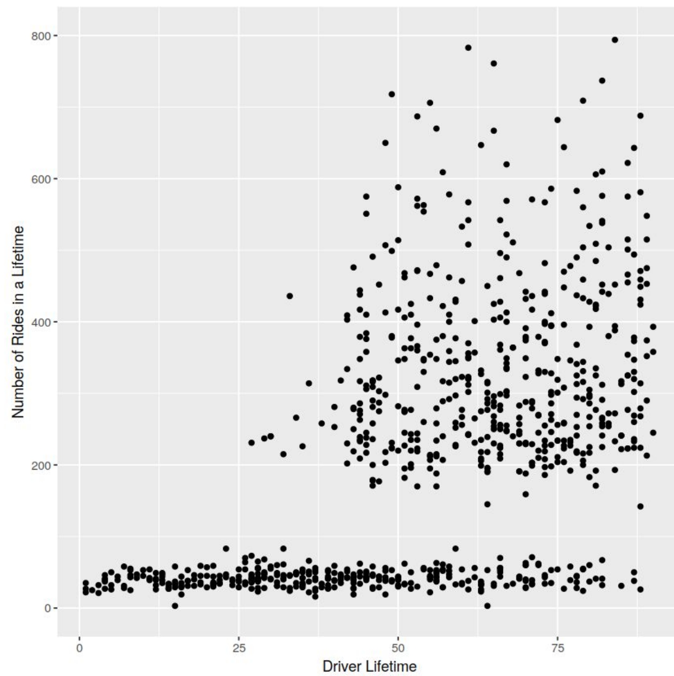
Driver Statistics

Variable	n	% Omitted Data	Minimum	Maximum	Median	Mean
Onboard Date	937	8.137%	3/28/2016	5/15/2016	4/7/2016	4/18/2016
# Drivers Onboarded/Day	837	17.941%	2	36	20	19.12
Driver Lifetime	837	17.941%	0	91	55	50.59
Lifetime Revenue	837	17.941%	0	12641.3	2701	2771
Lifetime Value	837	17.941%	0	221.07	48.7694	53.84
Total Rides	920	9.804%	0	919	205	200
Rides Per Day	837	17.941%	0.047	17.5	3.57	3.899
% Rides w/Prime	920	9.804%	0	75.36	32.19	30.05

3. Driver Lifetime and Employee Retainment

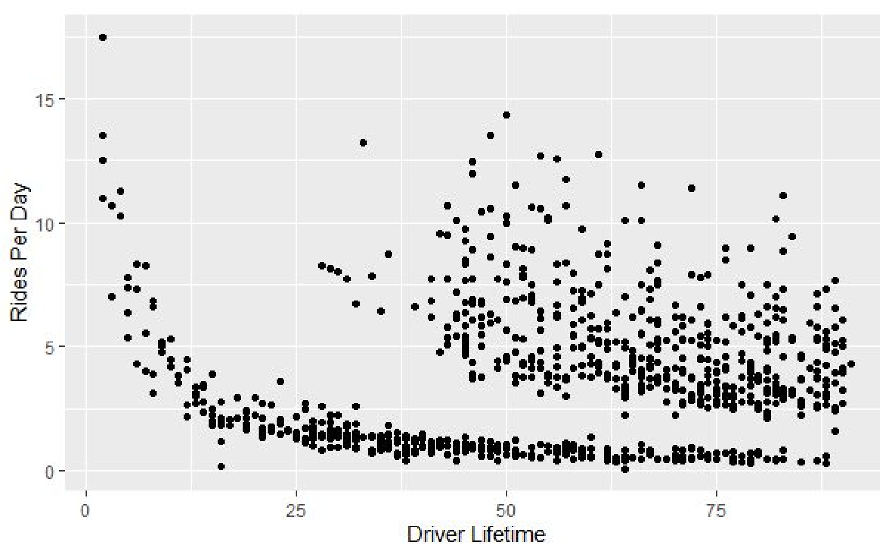
Lyft's business model depends on the company's management of its employees: the drivers. As a company that offers flexible hours and minimal training prior to onboarding, we believe it is important to investigate what elements help increase the retainment and lifetime value of Lyft's drivers, particularly since only approximately 20 drivers are onboarded daily.

The first component of determining a driver's lifetime value is calculating how long they have driven for (driver lifetime). In the code, we grouped the data by *driver_id* to determine the most recent ride given by each driver. This is interpreted in the data as the latest timestamp in the *dropped_off_at* column of our "lyft_data" dataframe. Then, we found the lifetime of each driver by subtracting the *driver_onboard_date* from the *dropped_off_at* column. **The calculated average driver lifetime was about 46 days**. Of course, drivers who signed up for Lyft but never gave rides would affect the data. To get a better idea of Lyft's driver behavior overtime, we created a scatter plot depicting driver lifetime and the total number of rides a driver gives in their lifetime.



From the graph above, we see that the data is split into two distinct groups. One group ranges from about 90-800 rides total, while the second group stays consistently below the 90 rides total mark. Also, drivers tend to try out driving for Lyft for about 27 days, before they decide if they want to be a frequent Lyft driver or not. If they haven't completed about 90 rides in 27 days, they most likely will not become a frequent, long-term Lyft driver. Approximately 68% of drivers who made it past a 27-day lifetime became regular Lyft drivers. This information is significant for Lyft because by looking at how many rides each driver has completed by the 27 day mark, Lyft can predict which drivers will be reliable employees in the long run.

In addition to comparing total rides in a lifetime to driver lifetime, we also graphed and observed patterns between rides per day and driver lifetime, shown in the scatter plot below. Our objective was to further analyze how often Lyft drivers needed to drive to be a long-term, reliable employee.



The general path of the data points is in exponential decay, suggesting that drivers drive less overtime. Similar to the previous graph, there are two distinct groups of data. Beginning at approximately the 27 day mark for driver lifetime, drivers either drive an average of less than two rides per day or an average of between 2.5 to 17.5 rides per day. So, only a portion of drivers who make it past a driver lifetime of about 27 days decide to drive full time. For both these groups of data, there's a gradual decline in average rides per day given as driver lifetime gets longer.

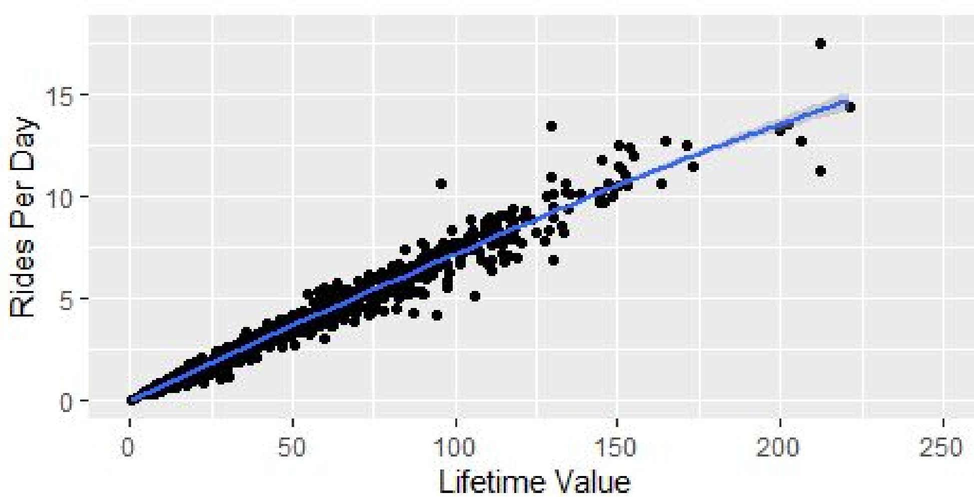
With the observations we made from the given data, we have a few actionable recommendations for Lyft to monitor and analyze their drivers' activity. First, we suggest that Lyft observe data from roughly the first month of each Lyft driver. If they average more than 2.5 rides per day, there is a higher probability that the driver will be long-term and regular Lyft driver. Also, we recommend that Lyft regularly survey their drivers after the one month period to see what their driver's plans for the future are (to become a full time driver and/or to continue driving for Lyft or not) as individual responses may be more accurate than predictions from the data. By targeting drivers who plan to drive on a frequent, long term basis, Lyft can refine their pool of drivers and enhance the experience for both riders and drivers.

4. Recommended Lifetime Value and Improving Employee Performance

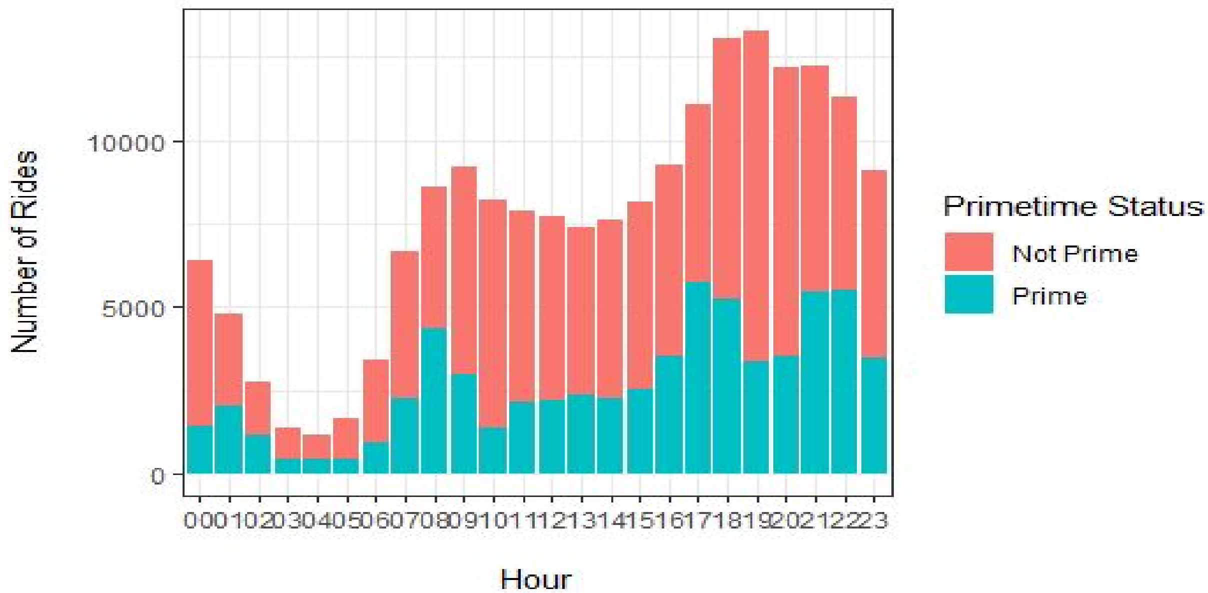
From the data, we were able to estimate that the average Lifetime Value for a Lyft driver is **$53.84/day.** The Lifetime Value for each driver was calculated by dividing the total revenue earned by each driver and by the driver's lifetime. The mean of every driver's Lifetime Value was then calculated to obtain the average Lifetime Value of a Lyft driver.

The significance of the Lifetime Value is that it is strongly correlated with higher revenues for Lyft. Obtaining this value allows for us to examine drivers to determine what types of drivers and driving habits are correlated with higher Lifetime Values.

From there, we began examining driving habits and found that consistency is key to achieving higher Lifetime Values. The below graph shows a strong positive relationship between driving more consistently (Rides Per Day) and Lifetime Value with a correlation coefficient of 0.982. As we have identified that the first month of driving is most critical to encouraging long-term driving with Lyft, we recommend that the company find ways to incentivize driving in the first month and help drivers incorporate it as a job, not just an occasional source of income. Potential ideas include bonuses for every nth ride driven or driving a certain number of times daily or vouchers to help pay for gas.



We also looked into how Prime Time status of drives affected Lifetime Value and found that the relationship was incredibly weak, with a correlation coefficient of 0.272, indicating that Lifetime Value of drivers was unlikely to be related to Prime Status of the rides. This may be attributed to how most rides do not have Primetime Status, even during peak hours of Lyft activity when users would expect Primetime prices. On average, only 30.05% of rides have Primetime status in a driver's lifetime. We recommend that Lyft utilize Primetime status more effectively, perhaps increasing Primetime status when drivers work most often (like 6 to 8 PM on the below graph), to increase overall Lifetime Value of all drivers. Lyft could also release information regarding Primetime rides to drivers during peak hours as an incentive to increase driver consistency.



5. **Future Work**

In the future, we would like to expand our analysis to further delve into the driving habits of Lyft drivers, examining other factors such as age, occupation, hours spent driving daily, geographical location, and how these may impact projected Lifetime and Lifetime Value of Lyft drivers.