Data Challenge 2019

A Driver Lifetime Value Report

Team This is Fine

Zhangyi Fan | Kai Lu

***"Calculating CLV is a matter of what story to tell and to have the right story to relate past behavior to future value"* - Peter Fader**

Treating Drivers as Customers

We begin by stating a key decision in our driver lifetime value model selection process. We decided to treat drivers as customers as they share similar characteristics and behavioral patterns.

To help justify the reasoning behind this decision, we state that it is a fact that drivers, like traditional customers, can choose or choose not to drive during any given time period (as opposed to customers choosing to buy or not buy). Additionally, drivers and customers both have some lifetime where the two become inactive. This would be equivalent to drivers ending their relationship with Lyft or customers completely stop making transactions at a store.

Non Contractual Setting

Now that we have defined our drivers as customers, we must realize the type of transactional setting we are dealing with here which is crucial to calculating the value of our drivers.

The perks of driving with Lyft is that drivers can go as long as they want without driving and still stay an active driver. According Lyft's policies, as long as the driver's insurance documents and everything like that are up-to-date, a driver is allowed to drive on the platform. However, this introduces a particular challenge:

How do we differentiate between those drivers who have ended their relationship with Lyft versus those who are simply in the midst of a long hiatus between driving for Lyft?

As opposed to a contractual setting (e.g. Spotify subscription service) where we observe the time at which the customers "die", Lyft drivers operate in a non contractual setting where the time at which a driver becomes inactive is unobserved. This makes calculating driver lifetime value very tricky.

Solution - the BG/NBD Model

The Beta geometric/Negative Binomial Distribution model was introduced by Peter Fader's Paper "Counting Your Customers" the Easy Way and is one of the best models for dealing with calculating customer lifetime value in non contractional settings. To predict number of future transactions, the model treats the customer purchasing behavior as a coin tossing game.

To Put Simply: Each customer has a buy coin that controls the probability to purchase, and a die coin that controls the probability of a customer to quit and never purchase again.

To Translate this Model in Lyft Terms: Each driver has a ride coin that probability that he or she drives at least one ride for some day x, and a die coin that controls the probability of the driver to quit and never drive for Lyft again.

Model Assumptions in Lyft Terms

Since we've tweaked the model to make it work for calculating driver lifetime value for Lyft drivers, we need to explain the assumptions the model is making under these new terms and justify why this is the right model for our problem.

Assumption 1

When a driver is alive, the number of active days made by a driver in a time period of time t is described by a Poisson distribution with active days rate λ.

This is equal to saying that for each day each driver tosses a ride coin and depending on the result, he drives or not. The number of active days we observe in the period depends on each driver's probability distribution around λ which is different from driver to driver. The expected value of active days for each driver is thus their parameter λ.

Assumption 2:
Heterogeneity in active day rates among drivers follows a Gamma distribution.

This means that drivers differ in driving behavior, which makes sense. It is a safe assumption to make that Lyft drivers differ in ride-taking behavior.

Assumption 3:
After any active day of rides, a driver becomes inactive with probability p and their dropout point (when they become inactive) is distributed with a geometric distribution for a given time period.

For example, if a driver's p=0.2, the probability of being inactive after 3 active days is 0.13. This is intuitive as the more drivers drive, the less likely they are to become inactive.

Assumption 4:
Heterogeneity across driver dropout probabilities has a Beta distribution with two parameters α and β.

The Beta distribution is the best for representing a probabilistic distribution of probabilities in the case where we don't know what the probability is in advance. In this case, since we don't know what the dropout probabilities of Lyft drivers are at any given point, we use the Beta distribution is estimate it.

Assumption 5:
Active day rate and dropout probability independently vary across drivers.

It is safe to assume that individually, each Lyft driver has their own live/die probabilities and drive/not drive probabilities. Everyone has two unique coins.

Main Factors that Affect DLV

For our BG-NBD Model, only 4 factors matter in calculating driver lifetime value (DLV).

- **Frequency** represents the number of active days the driver has had - 1.
- **Age** represents the age (days) of the driver since they made their first ride
- **Recency** represents the age (days) of the driver when they made their most recent ride
- **Monetary Value** represents the average amount of revenue drivers generate for Lyft in an active day

Driver Example:
Say we have data on a driver A from 100 days ago. Let's say driver A onboarded 100 days ago; drove 68 active days and was last active 3 days ago.

Her Frequency would thus be 67, Recency would be 97, and Age would be 100. The BG/NBD model would predict high future active ride days for this driver as she has a high frequency and a small difference between recency and age. In driving terms, this means that she is a frequent driver and has recently completed a ride.

On the other hand, suppose some driver B had Frequency 50, Recency 40, and Age 100. The model would now predict low future active ride days as driver B hasn't driven recently (inactive for 100-40=60 days) and has most likely "died".

Calculating Revenue per Active Day

The final piece to the model is the Monetary Value column for each driver. To calculate this column, we need to first explain how we calculated revenue per trip. The formula we used is as follows:

$$Trip\ Revenue = (Base\ Fare + Cost\ per\ Mile \times Ride\ Distance + Cost\ per\ Minute \times Ride\ Duration) \times (1 + Prime\ Time\ Multiplier) + Service\ Fee$$

For total revenue per day, we aggregated each driver's trip revenues from all their trips and then divided each by the number of active days each driver had. This yielded the average revenue per active day for each driver, which becomes our Monetary Value column input to the model.

The Story We Want To Tell

From before, we alluded to Peter Fader's quote on telling the right story that relates past behavior to future value. The story we want to tell is not as complex, but essentially, a two part novel:

$$\begin{aligned} DLV_i &= expected\ number\ of\ future\ active\ days_i \times expected\ active\ day\ revenue_i \\ &= f(Frequency_i, Recency_i) \times g(Monetary\ Value_i) \end{aligned}$$

Part 1: From Frequency, Age, and Recency, we forecast the expected number of active days for each driver in some time period of length t.

Part 2: Then, we estimate the most likely monetary value of each active day for each driver and putting everything together we can calculate the lifetime value for each driver.

Our Driver Clusters

Before we implement our BG-NBD model, we need to realize that Lyft Drivers are undoubtedly different, and therefore, we believe that in order to properly project the driver-lifetime-value, one generalizable approach alone is not enough. But instead it is more suitable for us tweak the model parameters to better represent the different kinds of driver segments that are present.

Specifically we envision Lyft drivers to be classified into three different types defined 2 parameters ***(A for activity and S for stickiness).*** We define **activity** by the frequency of rides happening within a time interval, serving as a proxy of the commitment of drivers, while we define **stickiness** by how long does the driver remain active in our system, a proxy of loyalty.

From the two parameters above, we envision three main segments by which Lyft Drivers falls into:

- **High Yield Engaged Drivers (high A and S)**
- **Low Yield Recurring Drivers (low A & high S)**
- **Passerby Drivers (low S)**

The Age Parameter

An important parameter underlying the BG/NBD model is the "Age" of a driver. The idea of this Age essentially refers to how long do we expect a driver to stay with us. We view the lifetime of Lyft drivers as the main generalizable variant within different segments of drivers. Furthermore while the BG/NBD model accounts of elements such as frequency, the supposed lifetime of the driver is something we will have to deduce.

Our Life Time Approach: We believe that drivers within each of these segments could be generalized by a decaying function that symbolizes the funnelling out/fading of drivers in a group. From that decay function, we define the average lifetime of drivers by a median lifetime, where in we are essentially finding the time period (week 1..week 50..week 100) where in only 50% of the drivers are active.

Defining Segments: K-Means & PCA

Framing it as an unsupervised learning problem, we decided to generate our driver segments through a combination of feature engineering, variable reduction and clustering.

In line with previous assumptions, we looked into 4 broad level of features:

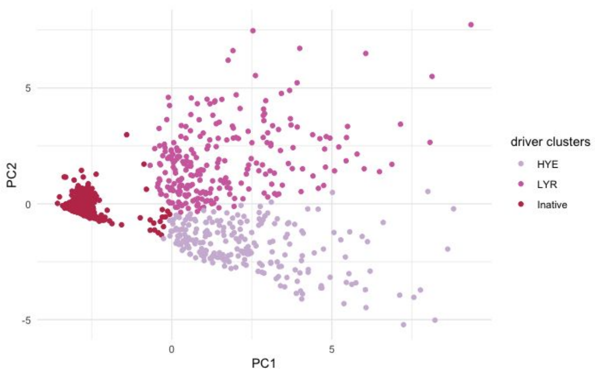
- **Weekly Count:** This echoes our activity parameter. This feature refers to the number of rides a driver initiated at a weekly interval. We created 13 variable under weekly count (week1_count..week13_count), which covers the span of this dataset.
- **Weekly Activity:** This echoes our stickiness parameter. This feature is a boolean, where in the value is true if a driver initiated a ride in that week. This feature is frequency agnostic and only focuses on active or not. We similar also created 13 variables under this field.
- **Distance:** We added in the distance parameter to account for possible differences in driver habits with regards to how far they drive. Distance also plays a role in the revenue incurred by the drivers.
- **Weekday/Weekend:** We also added a field on weekday activity and weekend activity. We presume that this field can also help determine the commitment level of the driver (full time drivers are likely going to drive on both weekday and weekends, while part time may only do so in weekends)

Dimension Reduction: This leads to a combined feature space of 30 variables. To better generalize and avoid the prospect of running dimensionality issues, we decided to perform a PCA on top of these features.

Clustering: We then performed a K-means clustering (K = 3) on the first 4 Principal Components, which accounts for **70% of the total variation** within the dataset.

Driver Lifetime Value | Clusters & Time Parameter

Cluster Result



Interpretation: Simplifying the clusters for the visualization, our algorithm seems to be able to capture 3 main segments across the two main principal components. To summarize the custer pivots, we classify those on the left hand side as inactive drivers as they typically have lower ride counts along with activity. Those classified as HYE (High Yield Engaged) and LYR(Low Yield Recurring) typically differ in that drivers in HYE tend to drive more and are also typically acquired later. In terms of activity, the two remain rather similar. Weekend/Weekday along with distance are not as informative and rather correlated with other variables We will expound upon these further in the recommendation.

The Lifetime Model

As aforementioned, we assumed that the average lifetime of a driver segment can be approximated by some form a decaying function. To model that in practice, we have a couple of other assumptions to clarify

Assumption 1:

We assumed that the decay function exists with a functional form of:

$$log(P(Driver\ is\ Active|\ Time)) = B_1 * Week\ Since\ Onboarded$$

This function could be interpreted as percentage decay of active drivers as every week passes. We believe that the Beta value which is the rate of decay would differ among different driver segments outlined above.

Assumption 2:

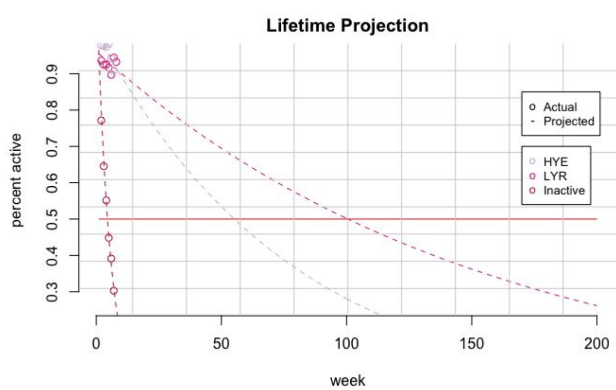
Given that we only have approximately three months of data, we are essentially assuming a rate of constant decay will persist later on. This is a strong assumption but one that will simplify the model drastically

Model Specification:

Instead of utilizing the full stretch of the three months of data, we looked at the date at which the segments of drivers are acquired, from there we modelled a span of time that accommodates most drivers. This a trade off between time span and number of observations on which we prioritized the latter since each segment only has about 200-300 drivers **(Inactive: 350; HYE : 227; LYR : 252)** and we wouldn't want to lose generalizability further by funneling out some drivers to get a bigger span.

Our final model is then based upon 7 weeks of data for HYE and Inactive segment and 8 weeks of data for LYR segment (they are generally acquired earlier)

Lifetime results

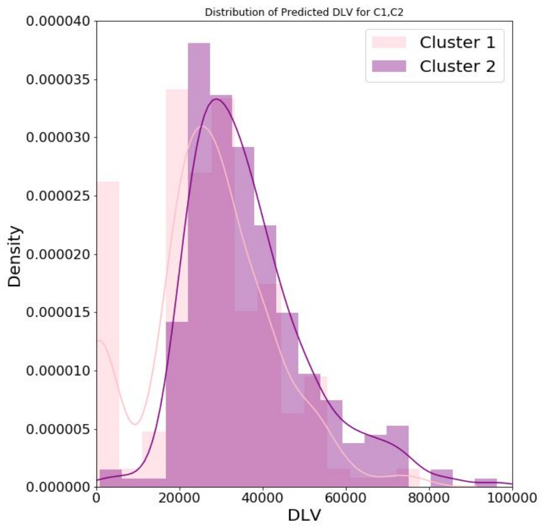


After running the decay function, the result are as follows:

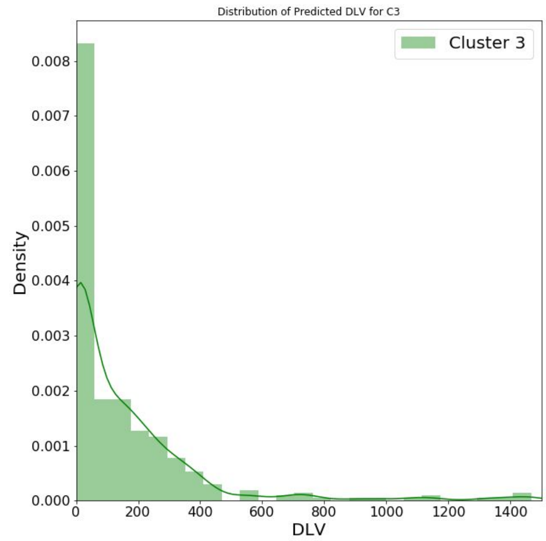
- **High Yield Engaged Drivers: 55 weeks**
- **Low Yield Recurring Drivers: 100 weeks**
- **Passerby Drivers: 4.5 weeks**

From there, we will feed these value in as a parameter for our BD/NGD model

Predicted DLV for Clusters 1 and 2



Predicted DLV for Cluster 3



Our Recommendations:

1. Behavioral Driver Segmentation

When we predict CLV for drivers, one of the biggest things we are trying to understand is who our drivers are and how we can devise behavior specific strategies to talk to each type of driver. From our segmentation we defined three segments: High Yield Engaged Drivers (Cluster 1), Low Yield Recurring Drivers (Cluster 2), and Passerby Drivers (Cluster 3).

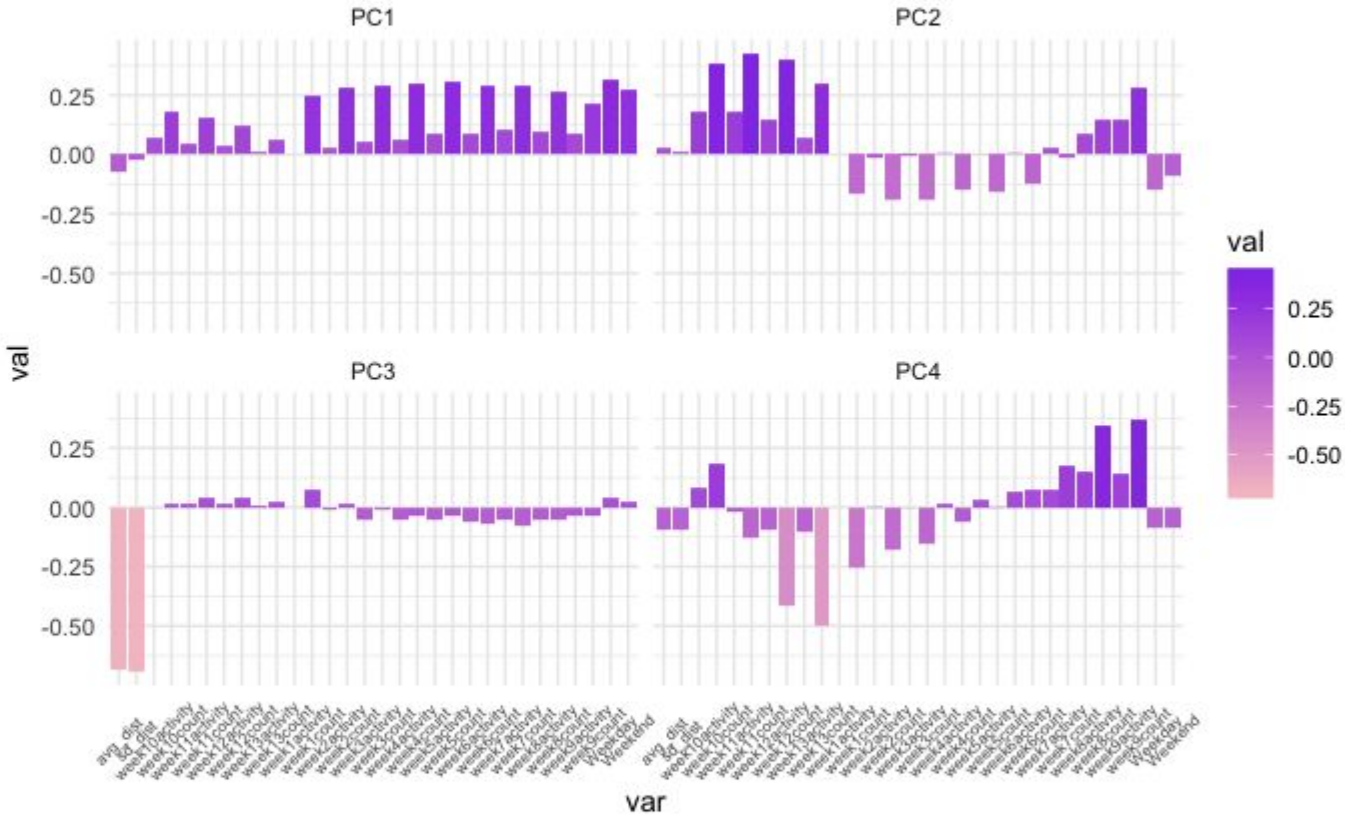
From the predicted DLVs, we see differences in DLV distribution for each: Compared to Cluster 2 drivers, Cluster 1 drivers have much more drivers that fell into the leftmost bin and overall had less predicted DLV than Cluster 2 drivers who were segmented to be stickier and predicted to last longer. On the other hand, the BG/NBD model predicted extremely small DLV for all Cluster 3 drivers. Since it is clear that we have three different driver segments, we need to devise segment-specific strategies for each. For example, for Cluster 3 drivers, these are our low Recency drivers who likely would benefit from reinforcement and nudges from Lyft to encourage driving again. For our Cluster 1 and 2 drivers, we would want to give them recognition for their consistency to reward good driving behavior. These are merely simple examples and only scratch the surface of possibilities. Our main point is that by devising segment-specific strategies, we will be able to apply the suitable solutions to each segment's frustrations.

2. Marketing Strategy Review

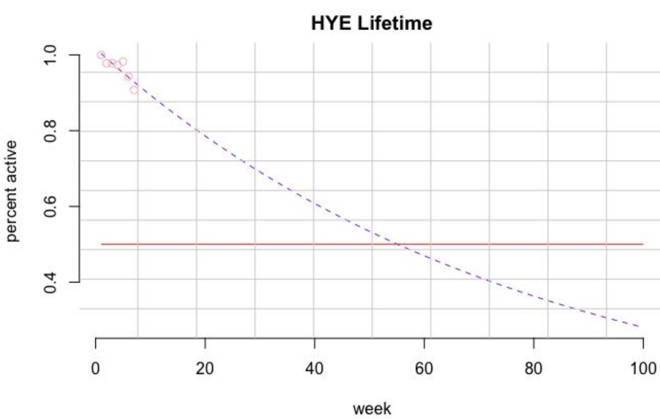
Though above we stated that it is important to devise segment-specific strategies when talking to each cluster of drivers, we have to realize that resources are often limited. That said, we want to focus on who our best drivers are first. While we would presume that acquiring the most active driver (Cluster 1) might be most beneficial for the company as a whole, these might not be our best drivers actually. Based on our DLV calculation across the 3 segments of drivers, our efforts may be better spent on relatively lower yield but more loyal/sticky drivers. Furthermore we also deduce that these set of drivers in Cluster 2 (LYR) represent a bigger population and market size than drivers in Cluster 1(high commitment full time drivers). This directly influences the market acquisition cost associated with these segments (lower for LYR than HYE), making those that are in the LYR segment even more appealing. In summary, we have found that the drivers in Cluster 2 are Lyft's best customers.

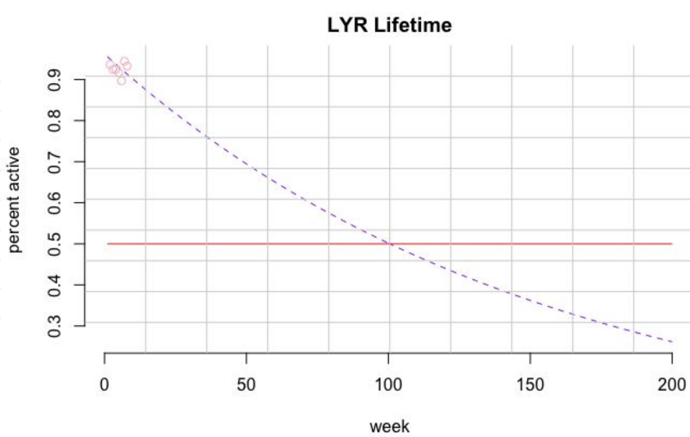
Appendix

PCA Result



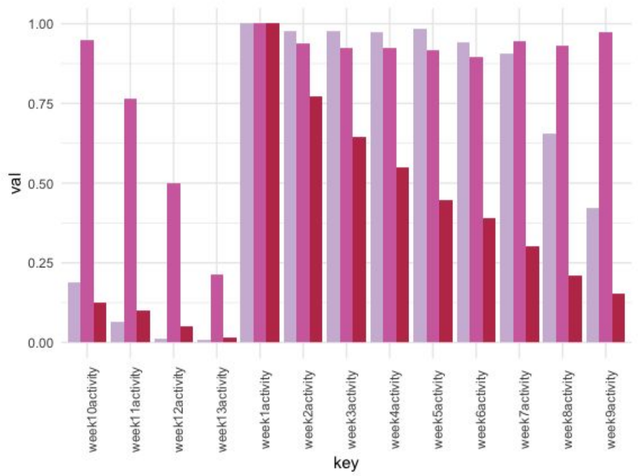
Lifetime Projection



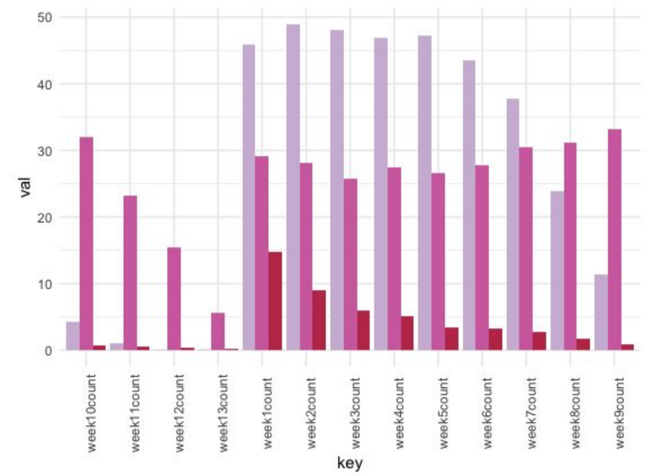


Appendix

By Cluster Visuals: Weekly Activity



By Cluster Visuals: Weekly Count



By Cluster Visuals: Onboard Date