Analysis and Computation of Driver's Lifetime Value and Its Contributing Factors

Presented by:
Khanh Thuy Mai Nguyen,
Svitlana Midianko
(Team "Girlnervans")

September 16th, 2019

Summary of the Conclusions Made

In the following report, we analysed and computed Retrospective Monthly Average Lifetime Value (ALTV) of all the drivers. Based on this, we conducted **statistical significance tests** to study the contribution of specific factors that yield higher ALTV for the driver. For this, we subset the data of all drivers (ranked according to their ALTV) into the bottom and top quantiles, which let us derive the following conclusions regarding significant characteristics of drivers who bring more value to the company. Specifically, our analysis suggests that:

1. Drivers who have a **higher frequency of 'prime' time rides** bring more added value to the company;
2. Drivers who have **shorter trips tend to yield higher ALTV**, thus also contribute more

This has been confirmed by statistically significant p-values resulted from the difference in mean tests between two quantiles.

Thus, we suggest that Lyft analyses the marketing channels used to target top 25% of the drivers and utilize them more frequently.

Average Lifetime Value (ALTV)

Definition of terms

We decided to focus on the analysis of the retrospective ALTV since the projected LTV is hard to derive from the available data. The indicator ALTV takes into account the revenue generated by the driver subtracting corresponding acquisition costs. Because the information on acquisition costs is not available publicly, we will assume that such cost is constant for each driver, thus is not relevant for the relative measure in the scope of the research question as it brings consistent change for all the drivers. Thus, in the scope of this paper, we will refer to ALTV as the value which excludes the total acquisition cost.

Driver's Monthly Average Lifetime Value

Although we computed the Total Retrospective Revenue generated by all the drivers, we suggest that this measurement is not indicative of the true driver's value as each driver has worked different time period, as onboard_date suggests.

Instead, we suggest relying on ALTV, which can be computed as a sum of all ride fares generated over the month period. This approach let us to the ranked list of all the drivers (see Appendix A) based on their average lifetime value.

Driver's Projected Lifetime Value

Various techniques could estimate the projected time that the driver will stay with Lyft. Due to limited data, we assumed that one would be interested in knowing the value that will be generated in the short-term business period - 1 year. This can be computed by multiplying ALTV and relatively will yield the same ranking of drivers and ALTV list shown in Appendix A.

Contributing to the ALTV Factors

Given the available data, we hypothesised that there are two main contributing factors:

1. Frequency of 'prime' rides, which are the rides given during the prime time;
2. Frequency of shorted trips compared to the longer rides.

To prove that the factors indeed contribute, we conducted statistical significance tests on the mean difference between the average prime time of Quantile 1 and Quantile 4 drivers as according to ALTV ranking.

Working with top and bottom quartiles should help one look for patterns, and since the sample sizes of both quantiles are relatively large (210 observations), we hope to reduce the impact of bias and noise in the model.

Significance of Prime Time

Prime time contributes significantly to the final total ride's fare; thus, we hypothesized that drivers who have a higher frequency of high 'prime' trips, should have higher ALTV generated. We conducted a test on the difference in means, which are two average prime times of both top 25% and bottom 25% drivers. The choice of the one-sided test should always be carefully thought-through, and here, we firmly believe that one should examine only whether 'prime' riders would contribute to higher driver's value. Thus, we suggest that the difference between the top 25% and bottom 25% drivers average prime times is positive.

We computed the average prime time for both quantiles and confirmed the above assumption, as the average prime time for top 25% drivers (Quantile 1) and the bottom 25% (Quantile 4) are 18.36 and 13.76 accordingly.

Therefore, the null and alternative hypotheses are as follows:

$$H0 = \mu(Q1) - \mu(Q4) = 0$$

$$Ha = \mu(Q1) - \mu(Q4) > 0$$

The observations are independent - because there sample consists of less than 10% of the whole population data, we can assume this independence, we can assume the normality of distribution.

With a significance level of 0.05 and t-score of 7.43 we obtained p-value of 0.00001. Since it is smaller than indicated above significance level, one can reject the null hypothesis in favor of the alternative hypothesis, meaning that there is indeed a statistically significant difference between the averages.

Looking at figure 1, one can see the tendency of data points to rise with the increase in ALTV.

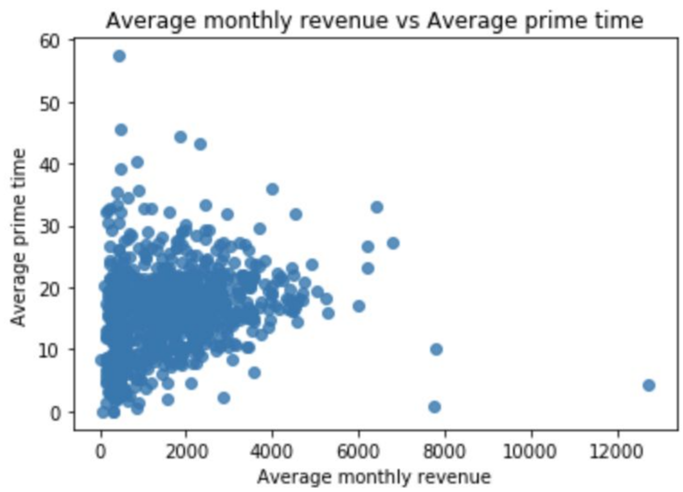

Figure 1. Relationship between the ALTV and Average prime time of the driver. One may see that as the ALTV increased, the average prime time tends to increase too, meaning that the drivers who have a higher frequency of prime rides yield more value.

Significance of average ride distance

Same concept is being applied to the ride distances. Here we perform statistical significance test on differences in average meters driven by button 25% and top 25% drivers. Thus the null hypothesis takes a sceptical view that there is no difference in average distances of button and top drivers. Whereas the alternative hypothesis assumes that there is a statistical significant difference between there in either side, making it a two-sided test. The average distance for bottom 25 and top 25 are: 7468.29 (meters) and 7048.5 (meters) accordingly.

$$H0 = \mu(Q1) - \mu(Q4) = 0$$

$$Ha = \mu(Q1) - \mu(Q4) != 0$$

With t-score of -2.46, a significance level of 0.05 we obtained a p-value equal to 0.007353. This concludes that the average distance of higher-valued drivers is usually smaller as compared to those who yield less value.

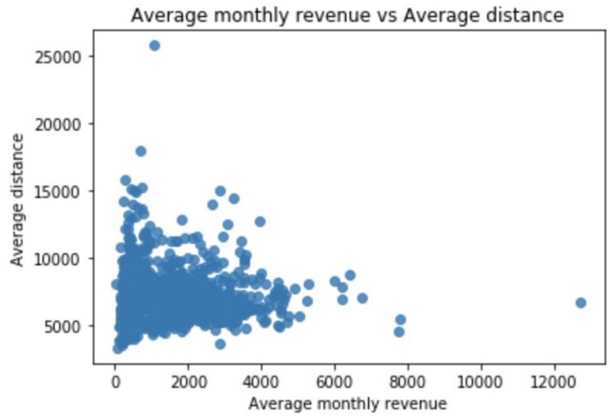

Figure 2. Relationship between ALTV and average distance. Here, one can see that riders who bring less monthly revenue (have smaller ALTV) tend to have longer rides on average.

The rationale for this can be that fare includes high service and base fee. Thus, the marginal utility of the additional meters driven tend to diminish because of the high starting base price.

Recommendations

Important aspect mentioned in the report is that the data on Driver Acquisition Cost was not given. We recommend collecting information on such data as well as the marketing channels used for finding top and button valued drivers. One should use the marketing strategies to target drivers who are situated in the top quintile and abandon those that are used for tagring the button ones.

Appendix

[GitHub Repository with the notebook]

Appendix A: Drivers Ranked In Descending Order According to Monthly Average Lifetime Value

In [21]:
```
# Look at the drivers making least revenue
df2.sort_values("revenue").head()
```

Out[21]:

	driver_id	driver_onboard_date	revenue	short_ride	long_ride
430	7b625f643d0775f0ac4898e33235377b	2016-04-04 00:00:00	24.60	1	0
447	7ff85c5c0e9324e28d1e0d0589c364bd	2016-04-19 00:00:00	36.50	1	1
841	fd831ca1d79ae5c6fc3a679a22e5b8cf	2016-05-07 00:00:00	205.88	3	1
713	d31eded9263eab43f614eecc6a52a0f5	2016-05-07 00:00:00	214.38	4	2
390	706466935b9e1d04e4e116be7ce90ea9	2016-04-01 00:00:00	223.62	3	3

In [22]:
```
# Look at the drivers making most revenue
df2.sort_values("revenue").tail()
```

Out[22]:

	driver_id	driver_onboard_date	revenue	short_ride	long_ride
461	844e9be5a30d8d9c1f8e9ddb086ff717	2016-04-15 00:00:00	10852.25	200	195
367	6b65c06851e944351dd285a1eb729499	2016-05-08 00:00:00	11052.32	198	154
274	4eb382d1f7d50fae1294964263d1ce82	2016-04-06 00:00:00	11085.73	245	167
189	3788dc9e91f1548816ce8b5af07ddadc	2016-04-26 00:00:00	12604.96	220	213
316	5ccc0e6dc9c7475caf785cdce7b8eb7a	2016-04-05 00:00:00	12640.07	231	249

Contact details:
Khanh Thuy Mai Nguyen - khanh@minerva.kgi.edu
Svitlana Midianko - svitlana@minerva.kgi.edu