**Value Generating Merchants:**

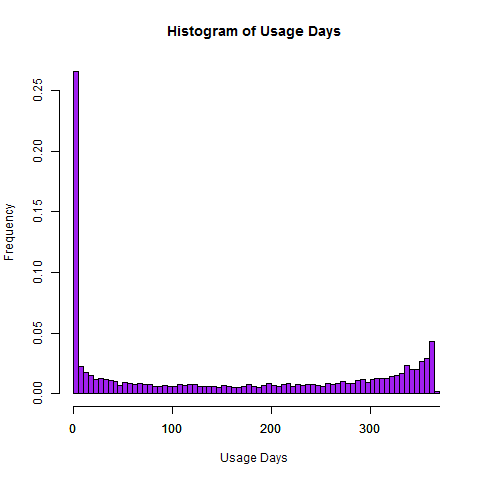
**Some Evidence from User Transaction Data**

The success of an internet business (COMPANY) that produces software to process payment transactions is tightly linked to the revenue generating success of new businesses that use the payment software product. Examining the company transactions of new businesses in their first year over the years of 2015-2017 can provide a signal on COMPANY’s future prospects. While the analysis finds that many new merchants sign up and experiment with the company’s payment platform they often opt out before one year. If the COMPANY needs to focus its efforts, there is evidence that it could be generated through marketing and product development for merchants selling more costly products and that have a high frequency of transactions.

**Many experiment with COMPANY’s platform, but decide to discontinue usage**

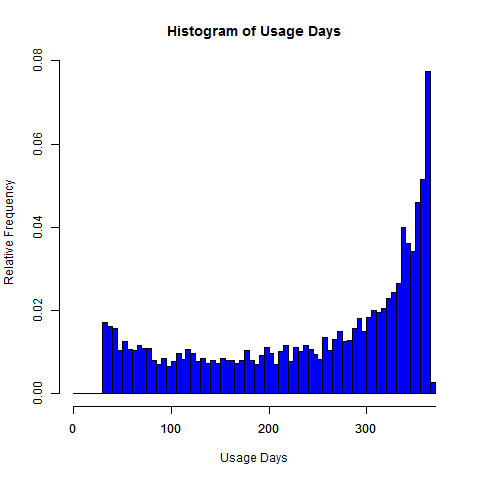
Many new merchants appear to experiment with using the COMPANY’s platform, but discontinue using after a single transaction (**Figure 1**). However, even longer-term users of the platform exit from using the platform before a full-year is completed. Removing users that were considering using the platform on a trial basis for a month or less (or where the observation is truncated due to the sample period drawn) it is still found that only around 12 percent of users are on the COMPANY platform for a full year (**Figure 2**). Moreover, transaction usage is not very high even among those using the Stripe platform for an extended period with few merchants having positive transactions on any given day. In fact more than 60% of all merchants that are on the platform for more than 30 days have less than 10 total transactions.

**Figure 1**



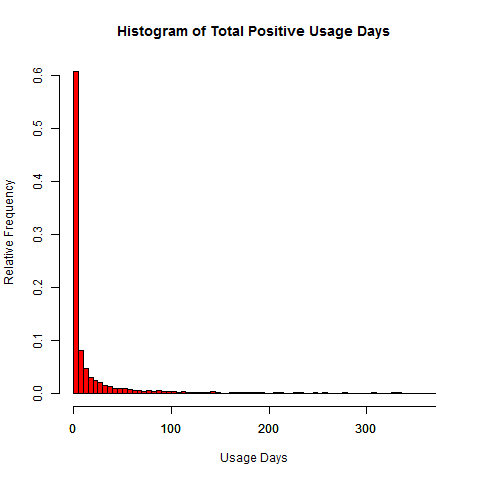
Note: Distribution of number of days each merchant user is observed in the sample up to 1-year.

**Figure 2**



Note: Distribution of number of days each merchant user is observed in the sample up to 1-year. Sample eliminates all users that used the platform for less than 30 days and users where the maximum date in sample – observed start date of user is less than 365 meaning the data was cut-off.

**Figure 3:**

****

Note: Distribution of number of days each merchant user is observed in the sample up to 1-year. Sample eliminates all users that used the platform for less than 30 days and users where the maximum date in sample – observed start date of user is less than 365 meaning the data was cut-off.

**Revenue Generation is in High-Value and High-Frequency Transactions**

Merchants that generate the most revenue are those with high value-high frequency transactions. These merchants are also ones with the greatest average growth per day. However, nearly 60% of merchants have very few transactions per day (<0.15 meaning that they have positive sales values around once week) and the average amount per transaction is worth less than $10 (**Table 1**). Merchants with high frequency of transactions (at least one transaction per day) and high value transactions (greater than $10 per transaction) account only for 10% of the sample.

The daily transaction day may be too volatile and also could have numbers that are distorted due to many days with 0 transactions. Collapsing the data to weekly (or monthly) could reduce problems related to the volatility of sales on a daily basis. In **Table 2** we observe that nearly 50% of merchants have less than 0.4 transactions in a week meaning that they have transactions less than once every other week. However, 25% of the sample has between 0.4 to 2.3 transactions per week and transaction values of over $1. Moreover about 22% of the sample has at least 2.3 transactions per week and average values of over $10. Moreover average weekly revenue growth is about 8%. These merchants also have average revenue per week that is about 6 times that of any other group accounting for about $3025 per week in revenues per merchant. In the analysis going forward we will focus on this set of merchants and also examine predicting revenue growth.

**Table 1: Average Daily Revenue and Revenue Growth**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Average # of Transactions Per Day | Average Amount per Transaction ($) | Average Revenue Per Day ($) | Revenue Growth per Day (%) | Percent of Sample |
| <0.15 | 1 | 0.20 | -0.04 | 49.4 |
| <0.15 | 1 to 10 | 4.88 | -0.03 | 18.9 |
| <0.15 | 10+ | 39.92 | 0.17 | 5.2 |
| 0.15 to 1 | 1 | 4.30 | 0.04 | 0.2 |
| 0.15 to 1 | 1 to 10 | 15.60 | 0.04 | 5.8 |
| 0.15 to 1 | 10+ | 95.42 | 0.20 | 9.4 |
| 1+ | 1 | 7.47 | 0.45 | 0.1 |
| 1+ | 1 to 10 | 125.41 | 0.26 | 0.7 |
| 1+ | 10+ | 807.82 | 2.63 | 10.2 |

Note: Sample excludes merchants who were only observed for less than 30 days. Average daily revenue = sum of transaction amounts divided by number of days firm is observed in the sample.

**Table 2: Average Weekly Revenue and Revenue Growth**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Average # of Transactions Per Week | Average Amount per Transaction ($) | Average Revenue Per Week ($) | Growth per Week (%) | Percent of Sample |
| <0.4 | 1 | 0.78 | -0.16 | 34.2 |
| <0.4 | 1 to 10 | 22.83 | 0.00 | 11.7 |
| <0.4 | 10+ | 248.03 | 1.96 | 3.5 |
| 0.4 to 2.3 | 1 | 6.47 | -0.05 | 0.7 |
| 0.4 to 2.3 | 1 to 10 | 46.06 | 0.15 | 10.5 |
| 0.4 to 2.3 | 10+ | 464.37 | 24.35 | 14.6 |
| 2.3+ | 1 | 6.91 | 0.66 | 0.2 |
| 2.3+ | 1 to 10 | 167.36 | 1.68 | 2.2 |
| 2.3+ | 10+ | 3025.48 | 8.96 | 22.4 |

Note: Sample excludes merchants who were only observed for less than 30 days. Average weekly revenue = sum of transaction amounts divided by number of days firm is observed in the sample.

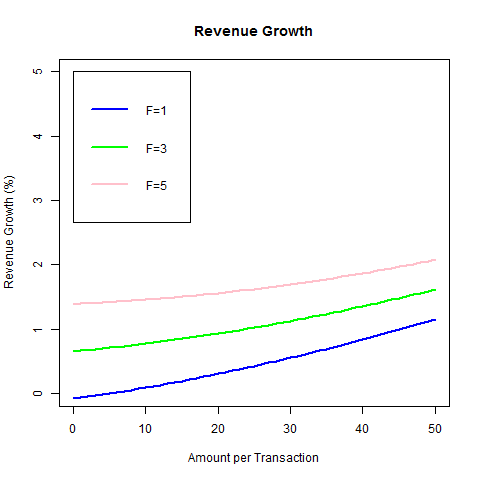
**High Value-High Frequency Transactions are Only Part of the Story of Revenue Growth**

High growth merchants that use COMPANY’s platform have the potential to drive long-term revenues for the COMPANY. From the above tables there is a clear relationship between frequency of transactions, average transaction value and revenue growth.

Using the average value per transaction and the frequency of transactions to predict growth shows that these two variables have a positive and significant relationship with growth (**Figure 3, Appendix Table 1**). Moreover, this effect is non-linear.

Firms with low transaction values (along the x-axis) are predicted to have negative growth. The turning point for growth comes when the average amount per transaction is $30. At this point, greater frequency of transactions (represented by the different colored lines) and average value per transaction results in increasingly higher growth rates. For example, an average transaction value of $75 and an average frequency of 10 transactions per day is related to growth rates that are 5% and are indicative of “high growth” merchants.

**Figure 4**



**Extensions to the Analysis**

***Power of Predictions (Cross-sectional Data)***

We examined using fit and train models for daily averages and weekly averages which collapsed the data over the period of observation. These models generally performed poorly with negative R^2 estimated for test models and high mean-squared errors. Of linear regression, random tree regressor, and random forest regressor the random tree regressor performed the best.

***Power of Predictions (Panel Data)***

The poor accuracy of the models when applied to the cross-sectional data suggested that we could try to exploit the panel data. We constructed

**Other analysis**

The analysis could be simplified by constructing a 0/1 indicator for whether a merchant is high value and using a logistic model to predict high value merchants. Some suggestions for constructing the 0/1 indicator include:

* High value merchant = 1 if the merchant has at least X1 days or X2 weeks of positive growth.
* High value merchant = 1 if the merchant has at least X1 days or X2 weeks of positive growth that is > 5%.

Suppose that we have a sales team that are going to target high growth merchants and ensure that these merchants are happy and are retained as part of the business. However, the sales team is quite constrained in time compared to the number of merchants that we predict as high growth. If this is the case then we would to try and minimize Type I errors (FALSE POSITIVES). On the other hand if the marketing team is not constrained we would apply a less conservative estimate and instead try to minimize Type II errors (FALSE NEGATIVES) hoping to make sure that we catch the number of clients that are high value.

**Value per Transaction and Frequency are Only a Small Part of the Puzzle of Merchant Revenue Growth**

The analysis presented here greatly simplifies the relationship between the frequency of transactions and the monetary value per transaction. The fit of the regression was not high as the r-squared was 0.128. Models that are better at predicting growth could include lagged values of transaction frequencies and controls for seasonal variations in the data. However, from a strategic and policy standpoint these types of regressions would not convey much about the types of firms that have high growth. Examining whether high values per transaction are generated from multiple product purchases versus single purchases, looking at industry and country of origin of firms could assist in providing more information on high growth merchants to target through product development and marketing.

Delving deeper into why merchants decide to stop using the COMPANY platform also is important as nearly 88 percent of the merchants in the sample decide to stop using the platform before completing one year even when they use the platform for over a month. Recent statistics by the U.S. Bureau of Labor indicates that new 75 percent of new businesses survive the first year suggesting that the large share of exits is probably not from merchants going of business. Initial examination of the determinants of exit showed that those that nearly stayed with the platform for a year were just as likely to have high values per transaction as those that decided to exit. On the upside, those with a high frequency of transactions had a higher probability of staying with the COMPANY platform over the course of a year.

**Appendix Table 1: Regression Estimates to Predict Average Growth**

Results

====================================================

Dependent variable:

---------------------------

Growth in Weekly Revenues

----------------------------------------------------

Txn per week 0.370\*\*

(0.146)

Txn per week Sq. -0.0003

(0.0003)

Txn per week Cu. 0.00000

(0.00000)

Amt Per Txn 0.019

(0.020)

Amt Per Txn Sq. 0.0002\*\*\*

(0.00001)

Amt Per Txn Cu. -0.000\*\*\*

(0.000)

Amt Per Txn\*Txn per Week -0.003\*\*\*

(0.001)

Constant -0.452

(1.981)

----------------------------------------------------

Observations 6,306

R2 0.347

Adjusted R2 0.346

Residual Std. Error 143.816 (df = 6298)

F Statistic 478.485\*\*\* (df = 7; 6298)

====================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Improvements to Project**

1. Coding:
   1. Would have executed it in Python/R as requested.
2. Variable construction:
   1. Smoothed the data over time (probably by week rather than using daily data to come up with the averages). The reason is that I was looking at the data and there was a lot of companies that did not have transactions in any given day.
   2. Integration of present discounting into the transaction amounts over time to get the values more in real terms and adjust for inflationary aspects.
   3. Creating an index of product variation that could be derived from variation in transaction amounts.
   4. Separating users based on distribution of purchase of frequency: (total purchases over a month for firm i/total purchases for all firms over a month)
   5. Time between purchases = Time/frequency of purchases
3. Modeling:
   1. Trying out different types of modeling exercises to better analyze the fit of the data and predictions and test out sensitivity of the results.
      1. In particular, how does analysis change on who is a high value client depending on whether I use daily, weekly, monthly averages of the data?
      2. Is it possible to use lagged data to predict future data. Concern over serial correlation so use some sort of panel time series process. For example first differencing with GMM might work to examine how changes in transaction frequency and transaction amounts as well as initial (mean values) correlated over time with the revenue of users. Include time fixed effects to account for trends in data.
      3. Is it possible to correct for sample selection on the time series data that comes from drop-out as the data that is unbalanced. Here then we could have predicted high value clients accurately. Probably the easy way would be just to have an indicator of probably the user didn’t survive until the end, but this doesn’t exactly correct for the entire problem and we would be a losing a lot of information by not including the exact drop out time. Might be able to use some weights for time in sample.
   2. I probably would have tried to analyze survival times as I was concerned with the number of customers that dropped out and didn’t make it to one year using the COMPANY payment platform. In particular when did they drop out and who is dropping out? (i.e. is it those with high frequency of transactions and/or those with high amounts per transaction?) The data is technically truncated.

Truncation depends on the probability that a person is still a COMPANY user in period t.

1. Writing/Presentation
   1. Visualizing more of the data (i.e. bar graphs of segmentation)
   2. Supplementary background information. Perhaps seeing how compares to other data that is similar.
   3. Thinking more about what we could gain from additional information. More data on user/firm identifiers (do they sell a diverse set of products), do they sell products across markets, what types of products, what is their marketing strategy that determines their frequency of transactions (more endogenous than the perhaps the good itself)?
2. What other things if you had more data?
   1. Comparison of COMPANY firms with some indicators of how it compares to distribution of all firms