on a Telecom Subscription Data

Derek Ren Follow
Oct 23, 2018 · 5 min read

Transforming from a product company to a service company.

One thing and now sort of a cliche I learned during my master study are that every company nowadays, especially technology companies, is ***transforming from products to services*** in the business market. Few examples could be:

Amazon started Prime Pantry plan that you could subscribe for varieties of homecare products; ***HP initiated printer as a service (PaaS)*** a long time ago, providing personalized enterprise printing service instead of just selling printers; and of course, music, video, movies subscriptions with the booming of ***streaming services.***

Therefore from the analytical perspective, companies start to ***look at customer lifetime value (CLV) for better segmentation and targeting*** besides traditional revenue metric.

· · ·

The data used in this article is from ***Kaggle: Telco Customer Churn***. Looking through the kernel, I found that lots of the notebooks are focusing on building up machining learning model to predict the churn. However, ***there are few articles analyzing from business context or customer value's angle.*** Therefore, three questions were analyzed through:

> *What is general churn/ survival pattern of the telco service?*

> *How to estimate and how much is the CLV?*

> *Are there any different patterns across different contract? (month-to-month, 1-year, 2-year)*

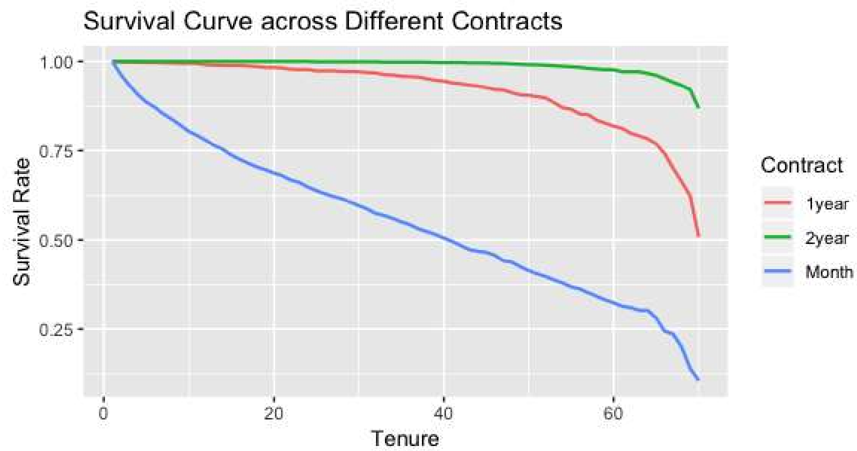
. . .

Customer Survival Analysis

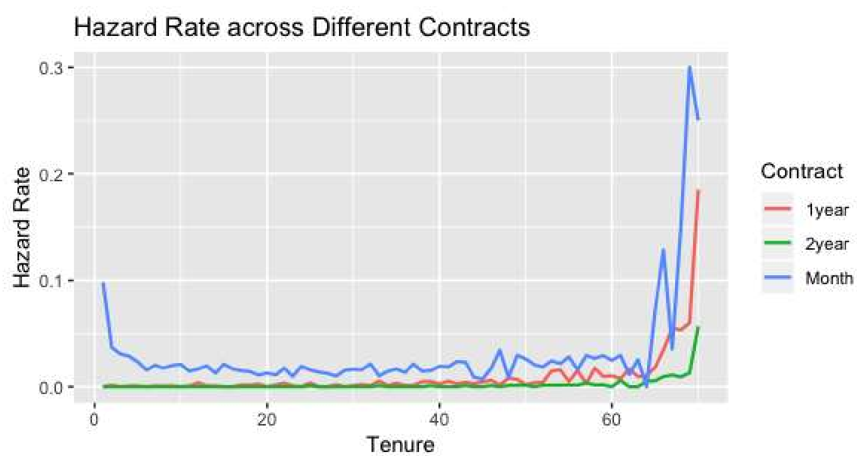
As we talk about subscription services, ***customer retention*** is the key we look at. And it could be interpreted from different ways:

- ***Survival (St):*** How many/ percent of customers survived by t periods?
- ***Churn:*** How many/ percent of customers churn by t periods?
- ***Hazard (ht):*** Given the customer went through t-1 period, what is the probability and he/she will churn in t period?

I won't cover the restrictive math formulas here, which could refer to ***Wikipedia*** ***(Survival Analysis)*** or my Github notebook. Here I used one of the major methodologies ***Kaplan-Meier*** to calculate and plot out the survival curve and hazard curve.



As expected, the survival rate decreases as the tenure increases. By the beginning of the period, the curve of Month-to-month contract decreases much more rapidly than the rest of the other two, which indicates that ***without the lock-down contract of long period, people are more likely to churn by the beginning of the product use.*** Overall, the longer the contract is, the less likely customers are going to churn.



As we look through the hazard rate, ***the month-to-month contract demonstrates a basin pattern***, with upper tails on two sides. One-year and two-year contract flatten out the tail at the begining, with little spike at each of the renewal period. (every 12 months, 24-months, etc) However, ***all of the contract experienced a high churn rate around 70 weeks.*** (6 years) This could because of the incomplete sample of the data, or made by some specific customer policy.

. . .

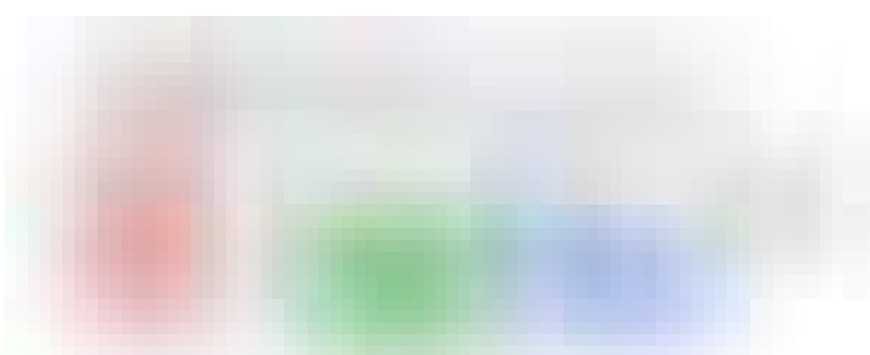
Customer Lifetime Value

For those who are not familiar with customer lifetime value, CLV is present value of the projected future cash flow from the customer relationship. The simplest formula is listed below

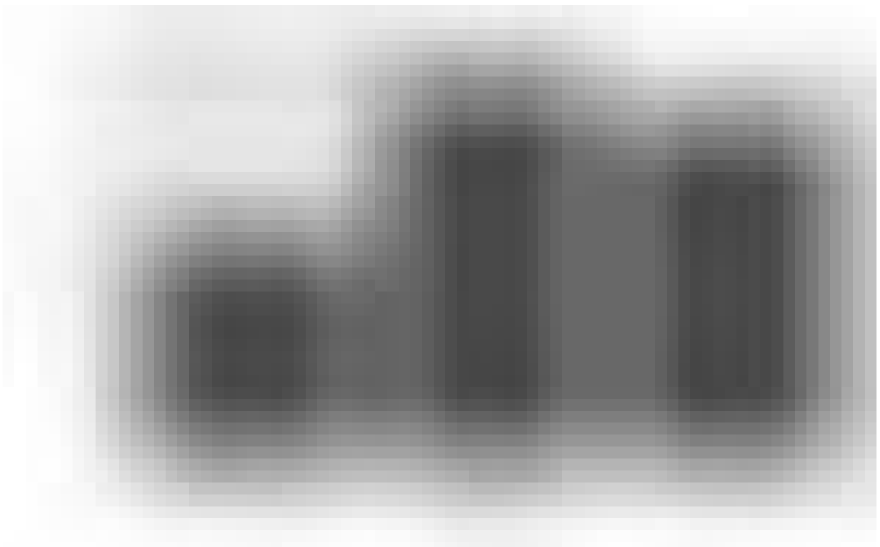
CLV with constant retention rate

- T : Total T period customers went through
- m: Monetary value company get from customers by each period. (Gross margin more precisely)
- r: retention rate ***(In real world, it is not a constant number)***
- d: discount rate

In our case, I first use the result of KP survival analysis to calculate the retention rate of each period (1–70 weeks). Beyond 70 weeks, I make a simple assumption that it keeps a constant retention rate 80%. Based on the monthly charges of each customers, CLV was calculated and the density plots for each type were plotted.
(Detailed R function can refer to my Github)



By looking at the distribution of CLV, we find that those contracts that maintain long customer relationship in long-term would generate more revenue. For month-to-month, most of all CLV are concentrated below 4k, while for one-year, two-year contract, there are fair proportion of CLV that are between the range of 4k to 6k.



What's more, if we look at the average CLV of these three different types. **We would find that one-year contract on average generates the highest CLV (3132), which is 68.2% more than that of month-to-month subscription.** However, it's worth noticing that two-year doesn't generated higher CLV than the one-year. And I made a few hypothesis:

- ***Monthly customer spending:*** Since there's no a big difference in the churn rate between 1-year and 2-year, probably in current database, 1-year customers spend much more monthly than 2-year customers, which let ***m*** dominates the CLV formula
- ***Retention rate beyond data:*** In our data, we only have 72 period and we made a naive assumption about the future. Probably in future, the churn rate of 2-year is much less than that of 1-year

. . .

Conclusion:
1. By ***turning in to services company and build long-term***

relationship with customers, companies could potentially generate more value.
2. ***Survival analysis*** could be helpful method to understand the ***subscription retention pattern and product performance***.
3. ***CLV*** can be used to quantify the ***long-term value of customers segment, justify pricing strategy and more beyond.***

More problems can be worked out from this dataset. For the full analysis of this work, please check out my ***Github*** for the Jupyter Notebook.

Feel free to leave any comment or feedback. And it is hoped that it is starting point of my Medium Blog and Data Sciences Life.