09/02/2020

IEOR 4650: BUSINESS ANALYTICS

VIBHOR MALIK – vm2599

ONKAR KULKARNI – ok2294

In partial fulfillment of the requirements of the course Business Analytics at Columbia IEOR.

Columbia University in the City of New York.

Table of Contents

Table of Figures

Introduction

Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of clients or customers. Telephone service companies, Internet service providers, pay-TV companies, insurance firms, and alarm monitoring services, often use customer attrition analysis and customer attrition rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Companies from these sectors often have customer service branches which attempt to win back defecting clients, because recovered long-term customers can be worth much more to a company than newly recruited clients.

Companies usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another company or service provider, involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care facility, death, or the relocation to a distant location. In most applications, involuntary reasons for churn are excluded from the analytical models. Analysts tend to concentrate on voluntary churn, because it typically occurs due to factors of the company-customer relationship which companies control, such as how billing interactions are handled or how after-sales help is provided.

Predictive analytics uses churn prediction models that predict customer churn by assessing their propensity of risk to churn. Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn. In this project, we use exploratory data analysis to draw insights from data on users from a Telecom company. Then we develop a Weibull model to validate our initial insights and find new ones about the customers. We then attempt to divide the customers into to segments based on the churn timing characteristics. We finally give insights into the characteristics of both subsets of customers.

Data Description

The data we use is from a hypothetical Telecom company. The data contains details of 7000 of the company's customers. Each row represents a customer, each column contains a customer's attributes. The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Exploratory Data Analysis (EDA)

Customer Demographic Analysis

We analyze the churn characteristics for customers based on demographic factors such as Gender, Senior Citizenship, and their family environment.

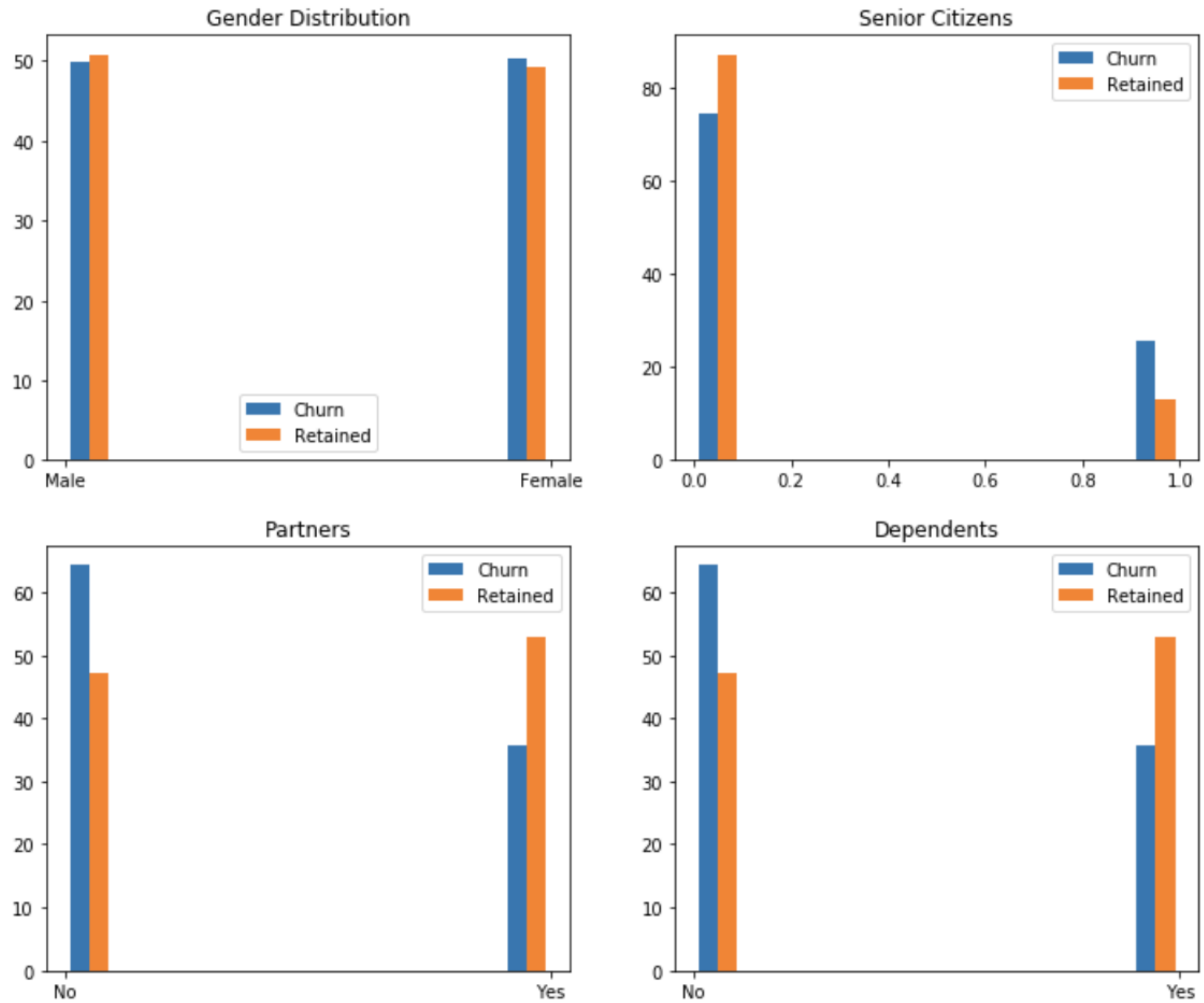

Figure 1 Customer Churn Characteristics based on different demographics

Inferences:

i. Customers with a family(a partner or dependents) have a lower chance of churn.
ii. Senior citizen customers are more likely to churn.

Customer Tenure Analysis

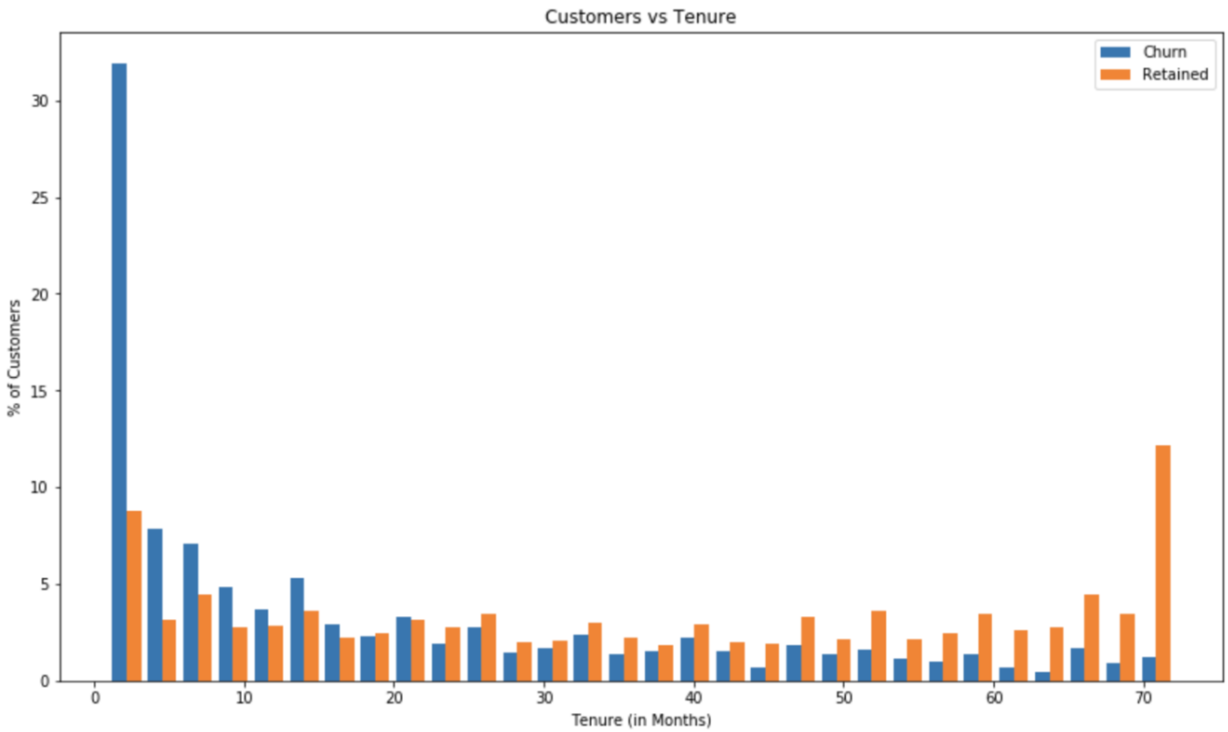

Figure 2 Customer Churn Characteristics based on tenure

Inferences:

i. Analyzing the customers tenures revealed that a very large proportion of the customers that churn leaves within the first month of service. This could be related to dissatisfaction with the service or could be related to fraudulent accounts that leave the service after a month of use without paying the bill.

ii. Another interesting observation is that customers have a much higher chance of churning within the first 20 months.

Customer Analysis based on Services used

Inferences:

i. Customers who use fiber optic internet service are more likely to churn.
ii. Customers who do not use internet services are less likely to churn.
iii. Customers who do not use online security are more likely to churn.
iv. Customers who do not use online backup are more likely to churn.

v. Customers who do not use device protection backup are more likely to churn.

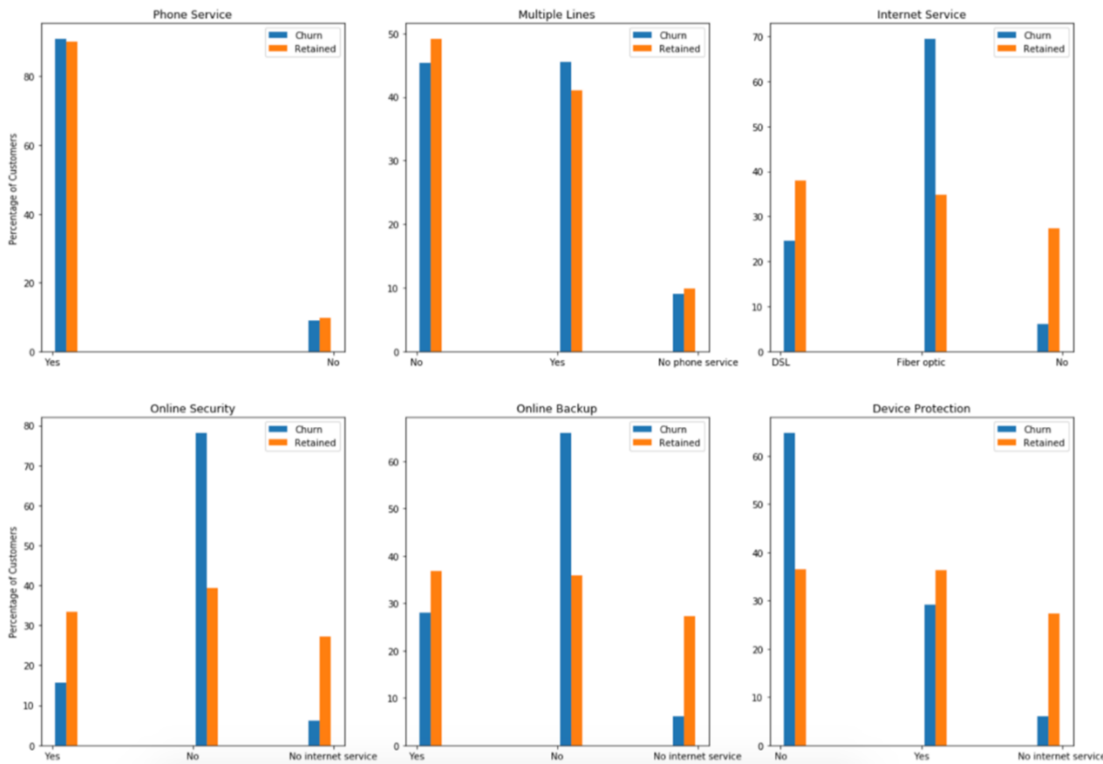

Figure 3 Customer Churn Characteristics based on Services used

Customer Analysis based on Payment Methods and Charges

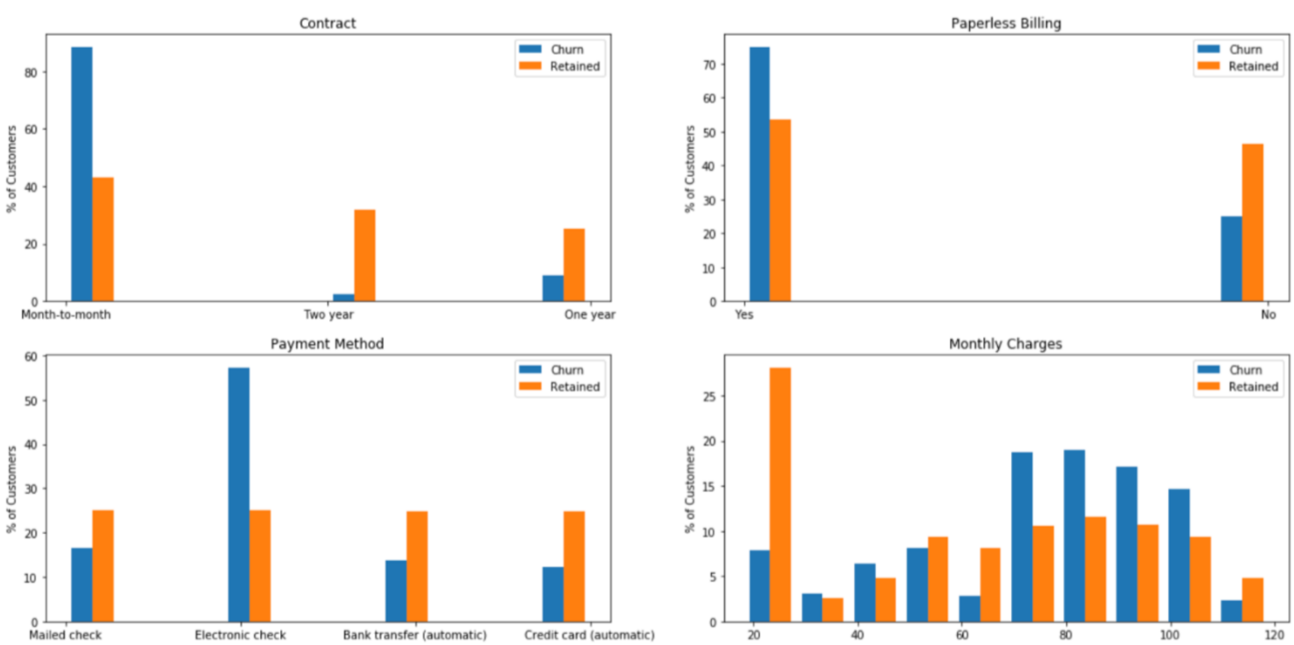

Figure 4 Customer Churn Characteristics based on payment methods and monthly charges incurred

i. Customers with month-by-month contracts have much higher probability of churn.

ii. Customers paying by electronic checks have higher churn probability
iii. Customers with paperless billing have a higher probability of churn.
iv. Customers with higher monthly charges, in the $70 - $120 range, have a higher probability of churn.
v. Total Charges are 0 for a very high percentage of Churn customers, reinforcing the earlier hypothesis that a lot of customers simply register for a month and leave the service without paying the bill

Churn Timing Analysis

One-Segment Weibull Model

To understand the customer churn characteristics, we developed a Weibull model to predict the customer churn timing. We started with a simple Weibull model with homogeneous lambda and c for all customers. However, in order to model the customer churn timing using this model, we need to make two assumptions:

a. We assume that all customers will eventually churn.
b. All customers that have not churned in the observed period are assumed to have been censored!

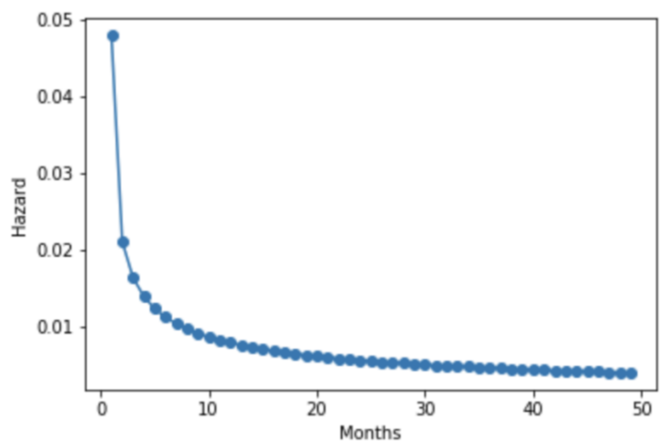

Figure 5 Variation of Hazard with Customer Tenure (in months)

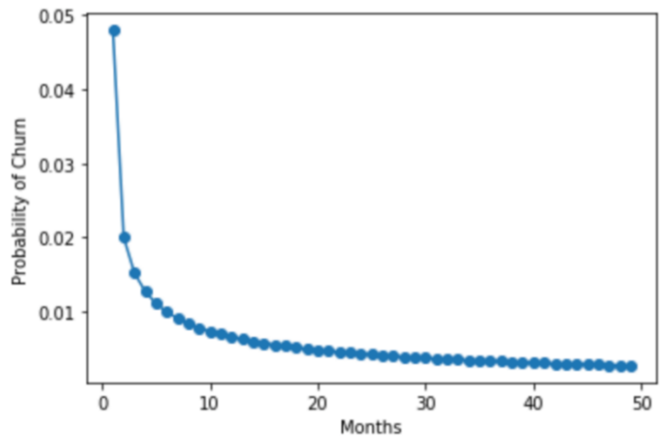

Figure 6 Variation of Probability of Churn with Customer Tenure (in months)

The value of C for the above estimated model is lesser than 1, indicating that hazard decreases with time. This did not make sense intuitively as we expect hazard of customer to leave the service should increase with increasing time. We therefore tried out a two-segment model next.

Two-Segment Weibull Model

The CDF of a Weibull model is given by:

$$F(t) = 1 - e^{-(t\lambda)^c}$$

Here, c affects how the rate of churn changes with time while lambda affects the overall rate of churn. For the two-segment model we assume that "c" parameter is same for both groups while lambda is different for each group. Therefore, both groups have different rates of churn but time similarly affects these rates in both groups.

We also incorporate covariates in the model to analyze how the different customer characteristics impact the churn timing. These will be covered in more detail in the next section.

Results:

C for Weibull model = 1.32
Lambda for Segment 1= 0.23,
Lambda for Segment 2 = 0.0065
Probability for a new customer to belong to segment 1 = 0.124

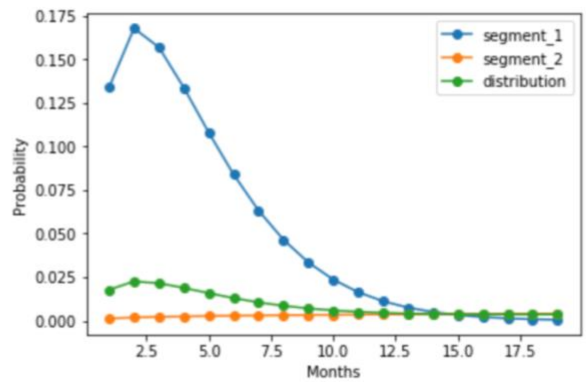

Figure 7 Customer Churn Probability variation with Tenure (in months)

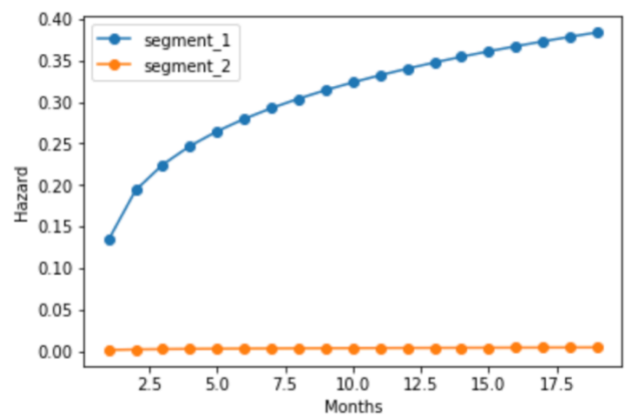

Figure 8 Customer Churn Hazard variation with Tenure (in months)

From the results we can infer:

i. The 'C' >1 indicating that churn probability increases with time, which is in line with expectations.
ii. It is clear from the graphs and the lambda values that segment 1 customer have a higher probability of churning earlier than segment 2 customers.

The above results make it clear that the customers can be divided into two segments, segment one with higher churn rate and segment two with lower churn rate and longer tenures.

Posterior Analysis

One of the hypotheses we had proposed in section 3b was that quite a few customers leave after being with the company for only 1 or 2 months. We had also predicted that this may be due to fraudulent customers who only avail the service for 1-2 months and leave without paying their bills. In order to verify this, we conduct a posterior analysis of the Weibull model and try to answer the following question:

How does the probability of customer of churn change with if the customer is with the company for 1 month or with the company for 2 months?

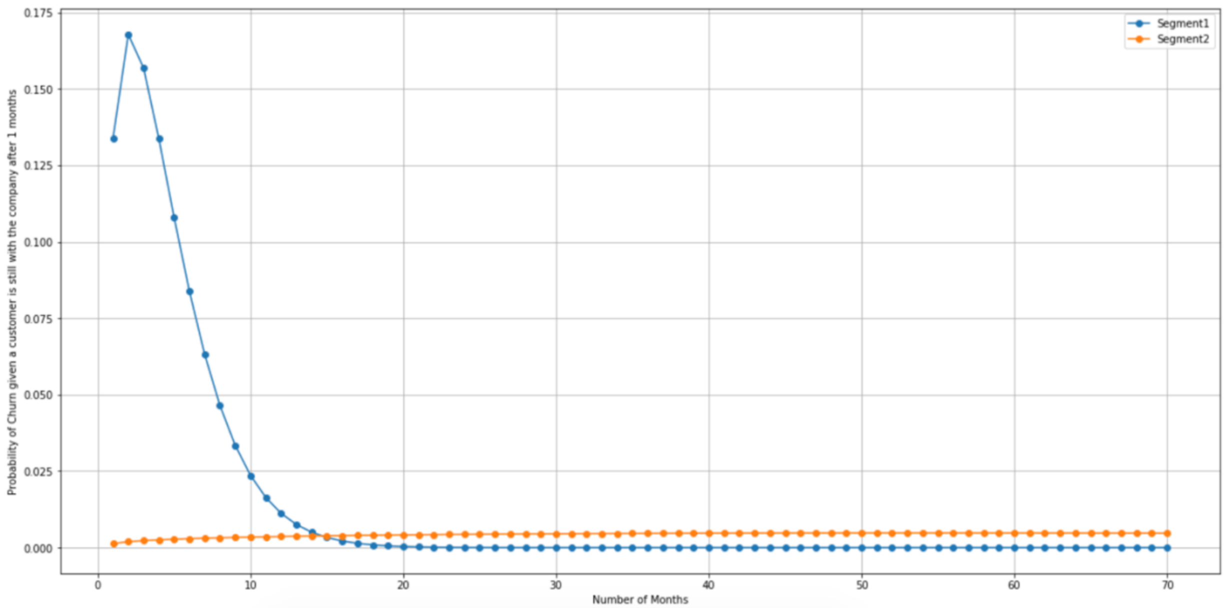

Figure 9 Probability of Churn given a customer is still with the company after 1 month

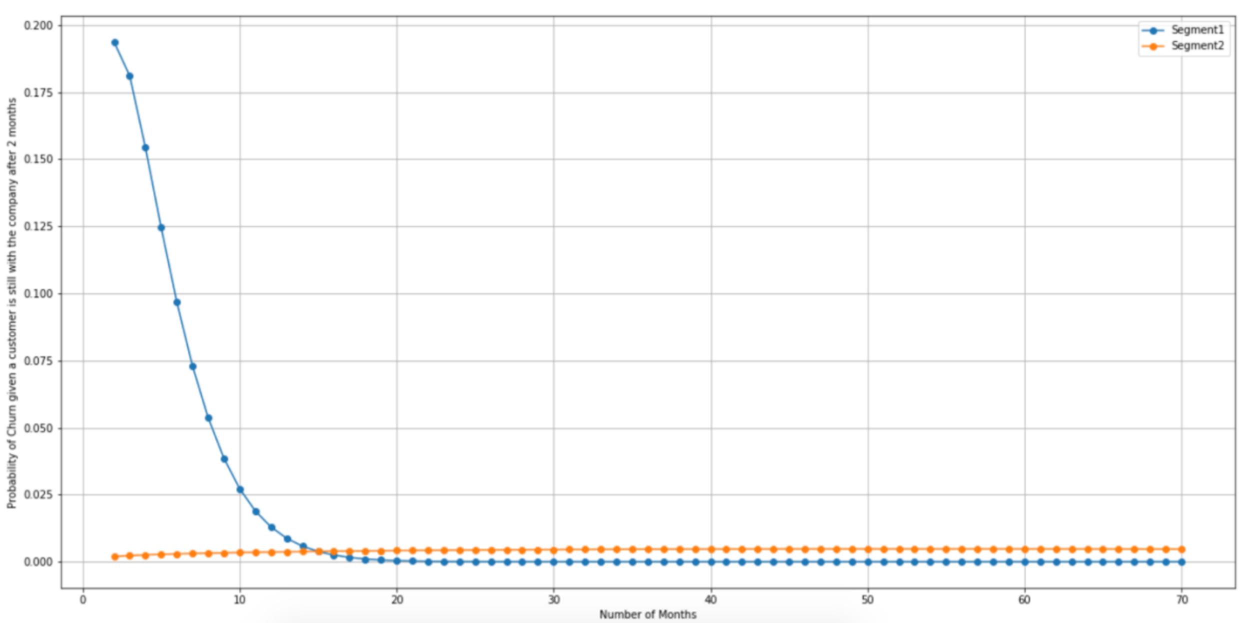

Figure 10 Probability of Churn given a customer is still with the company after 2 months

F

higher chance of churn in the next month, after which the probability of churn drops dramatically.

However, in Fig 10 we see that the probability of churn for a customer if he is with the company for 2 months drops dramatically thereafter, though the probability of churn in the 2nd month itself is higher in absolute terms.

This analysis gives further credence to our hypothesis of fraudulent customers leaving the service within 1 or 2 months. It could also be that the service is lacking certain features to tie-in more customers beyond this early phase.

Customer Characteristics

Next, in order to further understand the customer behavior, we introduce covariates in the Weibull model presented earlier.

```
Optimization terminated successfully.
         Current function value: 3.182433
         Iterations: 69
         Function evaluations: 81
         Gradient evaluations: 81
                              Weibull_2C Results
==============================================================================
Dep. Variable:                 tenure   Log-Likelihood:                -22379.
Model:                     Weibull_2C   AIC:                         4.478e+04
Method:            Maximum Likelihood   BIC:                         4.485e+04
Date:                Tue, 01 Sep 2020
Time:                        13:01:48
No. Observations:                7032
Df Residuals:                    7022
Df Model:                           9
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
seniorcitizen        0.4811      0.122      3.945      0.000       0.242       0.720
partner              1.4876      0.098     15.236      0.000       1.296       1.679
dependents          -0.2234      0.103     -2.172      0.030      -0.425      -0.022
OnlineSecurity       0.5136      0.093      5.538      0.000       0.332       0.695
OnlineBackup         0.9049      0.090     10.068      0.000       0.729       1.081
DeviceProtection     0.7482      0.095      7.847      0.000       0.561       0.935
TechSupport          0.2052      0.096      2.136      0.033       0.017       0.393
StreamingTV          0.6203      0.099      6.244      0.000       0.426       0.815
StreamingMovies      0.6995      0.099      7.035      0.000       0.505       0.894
```

Figure 11 Results of Two Segment Weibull Model with Covariates

Based on the coefficients of the covariates in the model we can conclude that the following factors significantly affect the churn rate:

iii. Senior citizen customers have an increased churn rate.
iv. Customers with online security, online backup, device protection, tech support, streaming TV and streaming movie services have a higher churn rate

v. Customers with partners have a lower churn rate.

NOTE: Other customer features such as payment method, contract type, etc. were included in the model, however, they did not meet the p-value threshold of 0.05.

Conclusions

i. From the exploratory analysis and Weibull model we can conclude that senior citizen customers have a higher rate and probability of churn. The company should either focus on better retaining such customers or pivot their strategy to move away from senior citizens due to their higher churn rate.
ii. The company should focus more on customers with partners or families as they have low churn rates and probability of churn.
iii. The company should analyze its strategy for online security, online backup, device protection services. Customers with these services have a lower probability of churn, as seen in the EDA, however customers with these services have accelerated churn rate which is alarming.
iv. The company should analyze its strategy for streaming movie and streaming TV services as these should be the latest cutting-edge offerings, but the model shows that they also accelerate the churn rate.
v. The EDA in section 3 revealed that customers had very high churn probability within the first couple of months of tenure. This hypothesis was further strengthened by the results of the posterior analysis of the model.

Future Work

The scope of this project was restricted to understanding the customer churn timing characteristics and understanding the customer traits which affected the rate of churn. Based on the conclusions drawn an interesting area to investigate could be to further examine the causes of early stage churn, which may require more data. Potentially increasing the retention rate of such early stage customers by introducing new services/schemes could be instrumental in boosting revenues and overall customer retention. Another aspect of this investigation could be to identify potentially fraudulent activity by customers who leave within one or two months of joining the service without paying.

Another possible area of investigation could be to further investigate and understand how factors such as payment methods, contract type, etc., which we have shown do not affect the churn rate significantly, impact the consumer behavior.