

CUSTOMER CHURN FORECASTING

Liu, Amy

Abstract

Customer churn prediction is a common analytical task in many industries, particularly when customer acquisition costs are high and customer lifetime value is paramount. This report analysis the dataset 'Telecom Churn', which covers app. 3000 user information and 20 features. by manipulating this data, applying Data Visualization and Machine Learning methods, 3 business questions are answered, and analysis is presented in the report and the Tableau dashboard will provide business insight to the management :

Q1. Formulate a clear business question that can be addressed by mining the data. Estimate the size of the problem in dollars – possible losing due to reduced phone call minutes.

Q2 Forecast customer churn. Show model performance metrics, explain which the key metric for model performance evaluation is. Comment on the quality of the models.

Q3 Visualize the effects of the top 5 predictors of customer churn and find the consumer portfolio

The report is powered by MS Azure and Tableau, in the payoff metric, the excel what if analysis is also applied. the published material can be found here:

Tableau (Data Analysis Dashboard):

https://public.tableau.com/views/TelecomChurnDataAnalysis/TelecomChurnDashboard?:embed=y&:display_count=yes&publish=yes

Azure (Modeling):

<https://gallery.cortanaintelligence.com/Experiment/Customer-churn-3>

- **Q1. Formulate a clear business question that can be addressed by mining the data. Estimate the size of the problem in dollars.**

The business case is established on the article drowned by Erik Heinrich (<http://fortune.com/2014/06/23/telecom-companies-count-386-billion-in-lost-revenue-to-skype-whatsapp-others/>), the case we address here is: under the consumer behavior changing and the prosperity of free WIFI, trying to use the data set to construct the revenue that will transferred into internet usage.

Certainly, there are many ways to conduct the revenue of internet usage, but here, we will only discuss the possible telecom service, such as call minutes being replaced by internet service such as skype, LINE or What's app.

“According to Ovum, consumer use of OTT VoIP will grow at a compounded annual rate of 20 percent between 2012 and 2018 to reach 1.7 trillion minutes. That translates to \$63 billion in lost revenue in the final year of its forecast.”

Build on this data set, we might come out the total lost for such current issue:

			\$0.25 per minute			
	Monthly	Annually	plan by AT&T	Revenue	Reduced 20%	
Total Day minutes	599190	7190280	0.25	\$ 1,797,570	\$	1,438,056
Total Even minutes	669868	8038416	0.25	\$ 2,009,604	\$	1,607,683
Total Night minutes	669507	8034084	0.25	\$ 2,008,521	\$	1,606,817
			Sum	\$ 5,815,695	\$	4,652,556
			Loss of losing 20% of revenue from call service	\$ 1,163,139		

By the calculation, a telecom company may loss over million for losing revenue from decreasing of phone call minutes. This problem should be able to be solved by implementing attractive LTE data pass plans to retain their subscribers.

- **Q2 Forecast customer churn. Show model performance metrics, explain which the key metric for model performance evaluation is. Comment on the quality of the models.**

Before started, the feature “phone-number” was eliminated due to it is basically directly link to the result. Few columns are domified including: International Plan, voice mail plan and area code, because they are presented by not over 5 levels of factors.

Five models are applied to the prediction, which are:

Logistic regression, Decision Tree, Random Forest, Neural Network and SVM

The performance outcome shows:

Model	Accuracy	Precision	Recall	F-Score	AUC
Logistic Regression	0.86	0.50	0.14	0.21	0.79
Decision Tree	0.954	0.92	0.74	0.82	0.92
Random Forest	0.956	0.91	0.76	0.83	0.91
Nueral Network	0.861	0.50	0.67	0.57	0.87
SVM	0.856	0.47	0.20	0.28	0.78

Among 19 features, three are considered

NO Contribution to the prediction, which are:

Total Night Min,

Total Night Charge,

Total Int Charge

Feature	Score
NumberServiceCalls	0.04
TotalDayMin	0.035
TotalDayCharge	0.035
VoiceMailPlan	0.028
NumberVMsg	0.023
InternationalPlan	0.012
AccountLength	0.005
TotalEveMin	0.001
TotalEveCharge	0.001
TotalIntMin	0.001
TotalIntCalls	0.001
TotalNightMin	0
TotalNightCharge	0
totalIntCharge	0
TotalDayCalls	-0.005
TotalNightCalls	-0.005
State	-0.006
TotalEveCalls	-0.007
AreaCode	-0.008

The payoff calculation will base on the research done by Aditya Kapoor (Kapoor, 2017) and market information from glass door, we assume the new contract will cover 12 months service and provide 15% discount for retaining customer. The assumption of the calculation is that for retaining customer who may terminate the service, the company will provide them extra discount, and the true positive customer will eventually leave the service.

Item	Cost
Monthly revenue per customer	\$ 62
Monthly gross Profit per customer	\$ 34
Acquistion cost for a new customer	\$ 315
Customer lifetime value	\$ 1,782
Marketing (marketing specialist hr rate \$32 *2 + advertisement \$30)	\$ 94

Save			Loss		Profit
TP	15% discount for 12 months. $0.15 * 34 * 12$	\$ 61.20	Customer life time value	\$ 1,782.00	\$(1,720.80)
TN	NA	\$ -	NA	\$ -	\$ -
FP	NA	\$ -	15% discount for 12 months + Marketing	\$ 155.20	\$ (155.20)
FN	15% discount for 12 months + Marketing"	\$ 155.20	Customer life time value	\$ 1,782.00	\$(1,626.80)

The calculated pay off for the model is:

Random Forest			
	Units	Unit Pay Off	Caculated Pay off
TP	106	\$ (1,720.80)	\$ (182,404.80)
TN	849	\$ -	\$ -
FP	11	\$ (155.20)	\$ (1,707.20)
FN	34	\$ (1,626.80)	\$ (55,311.20)
Sum		\$	(239,423.20)
Best Cut off			0.98
Best pay off		\$	(228,410.00)

\$ (239,423.20)		(239,423.20)	
0.1	\$ (262,438.80)	0.90	\$(231,324.00)
0.2	\$ (247,972.40)	0.91	\$(230,854.00)
0.3	\$ (243,034.40)	0.92	\$(230,384.00)
0.4	\$ (240,575.20)	0.93	\$(230,196.00)
0.5	\$ (239,423.20)	0.94	\$(229,914.00)
0.6	\$ (237,425.20)	0.95	\$(229,726.00)
0.7	\$ (235,582.40)	0.96	\$(229,256.00)
0.8	\$ (234,050.00)	0.97	\$(229,068.00)
0.9	\$ (231,324.00)	0.98	\$(228,410.00)

Conclusion:

With evaluation of pay off matrix, we can find that the best cut off for the model is 0.98%, and since the church will however leading to the loss, if the company may provide better fitting package to retain customers, the life time value will be able to fill in the deficit. Base on my model, the precision will be the key metric. The accuracy of the model reaches 95%, which is significantly larger the naive guessing is approx. 86%. The church rate in 2017 is 1.9% across four carriers, the 95% accuracy might yet able to catch certain method applied models across the industry, but consider features being covered by this data are not comprehensive, this model might be considered acceptable but there is a large room for improvement.

► Q3. Visualize the effects of the top 5 predictors of customer churn.

According to the model output, we can find the Top 5 feature importance are:

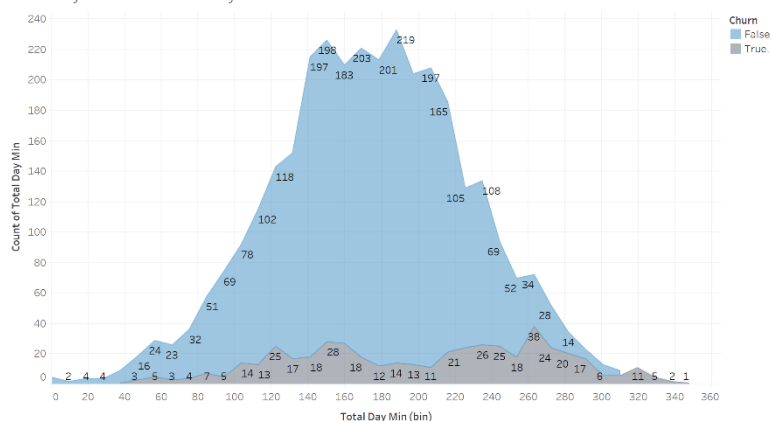
Total Day Min, Total Day Charge, Number Service Calls, International Plan, Total Int Calls

Random Forest

Feature	Score
TotalDayMin	0.059
TotalDayCharge	0.059
NumberServiceCalls	0.059
InternationalPlan	0.047
TotalIntCalls	0.029

1. Total Day Min:

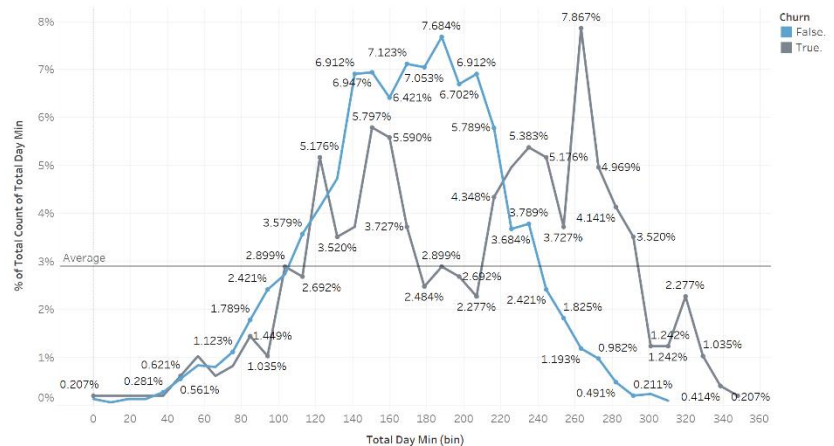
Total Day Min distribution by Churn



First, we investigate the total day min distribution. The data reveals that most of users are central to around 13 mins to 230 mins of usage. Churn customers show lesser concentration of Day min.

And with comparison of percentage distribution of churn in each total day min. We can also find churn record dominates total day min over 300 mins. It might be red flags that the company is not wellly satisfy user with high total day min.

Churn rate by Total Day Min



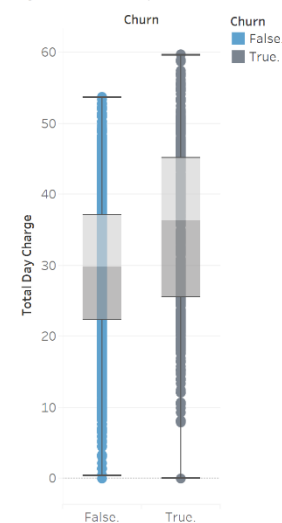
2. Total Day Charge:

Total day charge records show similar pattern with total day min. The churn cluster concentrate in a higher number of total day charge. Figure▶▶▶

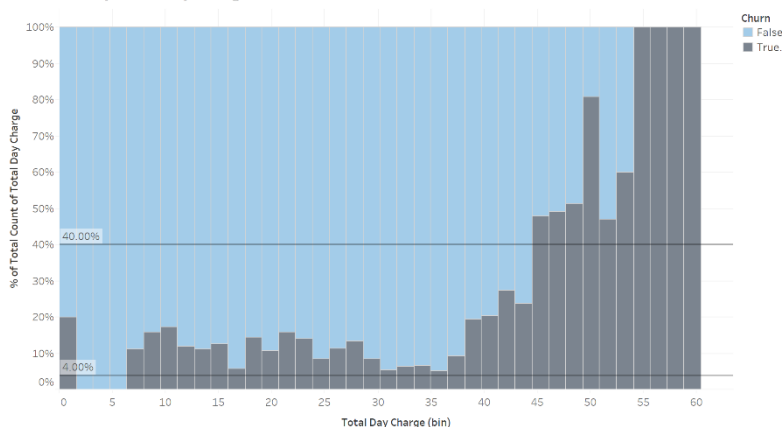
Also, if we look into the churn rate across each total day charge, we can also find that the churn rate surpasses subscription rate once the total day charge over 45, which could also being set as a warning.

Figure ▼▼▼

Totally Day Charge by Churn Box plot



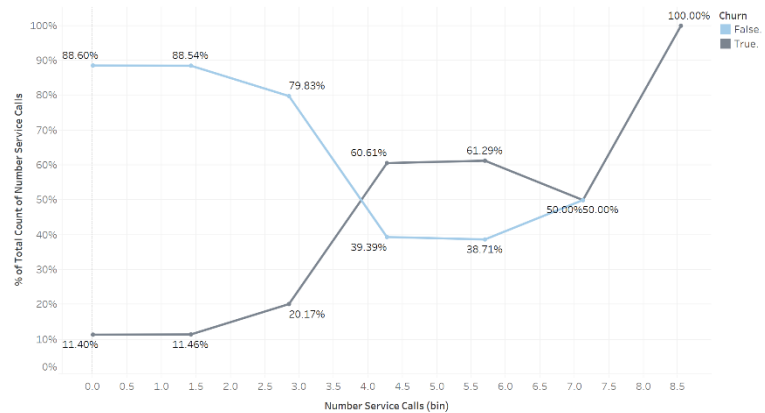
Churn rate by Total day charge



3. Number service calls

The cross of subscription rate and churn rate here locates in 4.0 service call. Even though we are not give elaboration of meaning of this feature, still we should set both 4.0 and 7.0 as alert that people may terminate service once their service calls reach these two number, the later one is almost a hundred percent guarantee of churn. Certainly, a larger data set might change the story.

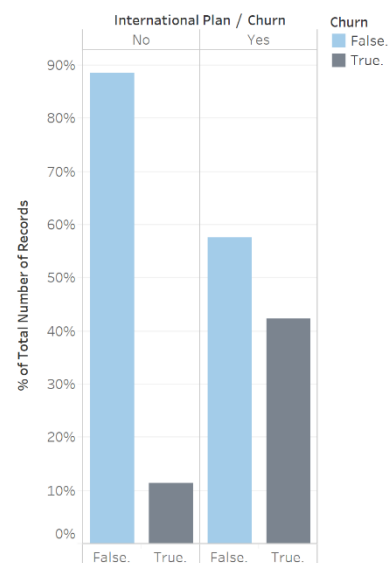
Churn rate by number service call



4. International Plan

Compared to users who joined international plan, users not covered by international plan reveals higher probability of terminating the service. We might interpretive this as the sign that the user is not planning a long-term subscription. Still, certainly, it might because user purely doesn't need it.

Churn rate by international plan



5. Total Int call

The distribution of total int call presents no significant clue of where the point should the company pay attention to. However, we can find most of number of total int call is under 10. The company might be able to provide better package to users with international call demand under 10 calls to raise subscription rate.

Total international call rate by Churn

