# SDM Telco Churn:

Describe the process by which you cleaned, processed, and partitioned data as necessary?

Below are the variable tables which I have considered to calculate the churn rate of telephone only, Internet only and internet & telephone customers respectively. To analyze and model them separately, I choose to subset the data. First subset excluding the customers who do not have internet service from the initial data set, which will give me only telephone customers. Second subset excluding the customers who do not have telephone service from the initial data set, which will give me only Internet service customers. Finally, the third one will be customers who have telephone service and have one of the either (DSL/Fiber optics) internet service. I have engineered payment method as automatic or not and Churn variable to num 1 and 0.

What predictors do you think contributes to the churn of (i) only telephone customers, (ii) only Internet service customers, and (iii) customers who subscribe to both phone and Internet services? Explain the rationale for your answer.

The first dataset has 1526 data points of which 113 customers have churned. Converted the required categorical columns to factors.

**Churn of subscribers of telephone services only**

|  |  |  |
| --- | --- | --- |
| Variable | Effect | Rationale |
| gender | ? | Specific gender cannot influence churn rate, but I want to see if there is any bias. |
| Senior citizen | - | Senior citizens usually dislike change, thus churn rate should be low. |
| dependent | - | Usually, dependents also will be in the same plan. Thus, churn rate could be low for moving as a group |
| tenure | - | Longer the customer is with the telco company, lesser are the chances to move. |
| MultipleLines | - | Easy for customers to have all lines with one company. Churn rate would be low if he/she has a bunch lo lines to move. |
| Contract | -/+ | Longer contracts yield to lesser churn rates as people usually don’t want to break contacts and move. |
| PaymentMethod\_auto | -/+ | Automatic payment methods, this explains the customer want to have a long tenure with the company, thus low churn rate. |
| MonthlyCharges | + | More are the charges more is the chance of customer moving away to cheaper options. |
| Excluded Variables | | |
| customerID | 0 | This alphanumeric number has no effect on churn. |
| Partner | 0 | No effect on churn, both could be in the same plan or different |
| InternetService, OnlineSecurity, OnlineBackup,DeviceProtection, TechSupport, StreamingTV, StreamingMovies | 0 | Not related to telephone service |
| Paperlessbilling | 0 | Has no effect on churn |
| Total Charges | 0 | This is combination of contract and monthly charges. |

The Second dataset has 682 data points of which 170 customers have churned. Converted the required categorical columns to factors.

**Churn of subscribers of Internet services (DSL) only**

|  |  |  |
| --- | --- | --- |
| Variable | Effect | Rationale |
| gender | ? | Specific gender cannot influence churn rate, but I want to see if there is any bias. |
| dependent | - | Usually, dependents also will be in the same plan. Thus, churn rate could be low for moving as a group |
| tenure | - | Longer the customer is with the telco company, lesser are the chances to move. |
| OnlineSecurity, OnlineBackup,DeviceProtection, TechSupport, StreamingTV, StreamingMovies | - | Each item of this set is an extra service provided by the company; thus, more are the features less is the churn rate of customers. |
| Contract | -/+ | Longer contracts yield to lesser churn rates as people usually don’t want to break contacts and move. |
| PaymentMethod\_auto | -/+ | Automatic payment methods, this explains the customer want to have a long tenure with the company, thus low churn rate. |
| MonthlyCharges | + | More are the charges more is the chance of customer moving away to cheaper options. |
| Excluded Variables | | |
| customerID | 0 | This alphanumeric number has no effect on churn. |
| Senior citizen | 0 | Senior citizens are much into internet services. |
| Partner | 0 | No effect on churn, both could be in the same plan or different |
| Multiple lines | 0 | Not related to Internet service |
| Paperlessbilling | 0 | Has no effect on churn |
| Total Charges | 0 | This is combination of contract and monthly charges. |

This final dataset has 4835 data points of which 1586 customers have churned. Converted the required categorical columns to factors.

**Churn of subscribers of both telephone and internet services**

|  |  |  |
| --- | --- | --- |
| Variable | Effect | Rationale |
| gender | ? | Specific gender cannot influence churn rate, but I want to see if there is any bias. |
| Senior citizen | - | Senior citizens usually dislike change, thus churn rate should be low. |
| dependent | - | Usually, dependents also will be in the same plan. Thus, churn rate could be low for moving as a group |
| tenure | - | Longer the customer is with the telco company, lesser are the chances to move. |
| MultipleLines | - | Easy for customers to have all lines with one company. Churn rate would be low if he/she has a bunch lo lines to move. |
| InternetService | -/+ | Fiber optic is better alternative when compared to DSL in this high speed era, thus churn for Fiber could be low |
| OnlineSecurity, OnlineBackup,DeviceProtection, TechSupport, StreamingTV, StreamingMovies | - | Each item of this set is an extra service provided by the company; thus, more are the features less is the churn rate of customers. |
| Contract | -/+ | Longer contracts yield to lesser churn rates as people usually don’t want to break contacts and move. |
| PaymentMethod\_auto | -/+ | Automatic payment methods, this explains the customer want to have a long tenure with the company, thus low churn rate. |
| MonthlyCharges | + | More are the charges more is the chance of customer moving away to cheaper options. |
| Excluded Variables | | |
| customerID | 0 | This alphanumeric number has no effect on churn. |
| Partner | 0 | No effect on churn, both could be in the same plan or different |
| Paperlessbilling | 0 | Has no effect on churn |
| Total Charges | 0 | This is combination of contract and monthly charges. |

Create training and test data sets with a 75:25 split using a random seed of 1024. Use the training data to train three logit models with the variables you identified in Question 2. Combine the outputs of the three modes using stargazer.

Only Telephone customers churn rate model:

**tele\_logit** <- glm(Churn ~ gender + SeniorCitizen + Dependents + tenure + MultipleLines + Contract + PaymentMethod\_auto + MonthlyCharges, family=binomial (link="logit"), data=train\_tele)

Only Internet customers churn rate model:

**inter\_logit** <- glm(Churn ~ gender + Dependents + tenure + OnlineSecurity + OnlineBackup + DeviceProtection + TechSupport + StreamingTV + StreamingMovies + Contract + PaymentMethod\_auto + MonthlyCharges, family=binomial (link="logit"), data=train\_inter)

Both telephone and Internet customers churn rate model:

**both\_logit** <- glm(Churn ~ gender + Dependents + MultipleLines + InternetService + tenure + OnlineSecurity + OnlineBackup + DeviceProtection + TechSupport + StreamingTV + StreamingMovies + Contract + PaymentMethod\_auto + MonthlyCharges, family=binomial (link="logit"), data=train\_both)

A close-up of a document

Description automatically generated with medium confidence

Pi/1-Pi = e^(B0 + BiXi)

What are the top three predictors of churn of

**(i)only telephone customers**

Contract

P(one year contact) = 0.048

P(Two year contract)=0.036

Customers contract period impacts the churn rate, the probability of a customer with 1 year contract to churn is 95% less than that of a customer with monthly contracts. Similarly, the probability of a customer with 2 year contract to churn is 96% less than that of a customer with monthly contracts.

Being a senior citizen

P=0.338

A senior citizen customers probability to churn is 33.8% more than that of a customer who is not a senior citizen.

Having multiple lines with the company

P=0.11

A customer with multiple lines has 89% less probability to churn than customer with single line.

**(ii)only Internet service customers**

Contract

P (one year contact) = 0.39

P (two year contract) = 0.23

Customers contract period impacts the churn rate, the probability of a customer with 1 year contract to churn is 60% less than that of a customer with monthly contracts. Similarly, the probability of a customer with 2 year contract to churn is 76% less than that of a customer with monthly contracts.

Online security

P=0.095

A customer with online security added in plan has 90% less probability to churn than customer without online security.

Tech support

P=0.108

A customer with tech support added in plan has 89% less probability to churn than customer without tech support.

**(iii) customers who subscribe to both phone and Internet services.**

Contract

P (one year contact) = 0.107

P (two year contract) = 0.029

Customers contract period impacts the churn rate, the probability of a customer with 1 year contract to churn is 89% less than that of a customer with monthly contracts. Similarly, the probability of a customer with 2 year contract to churn is 97% less than that of a customer with monthly contracts.

Fiber optic

P=0.75

A customer with Fiber optic Internet connection has 75% more probability to churn than customer with DSL internet connection.

Having multiple lines with the company

P=0.59

A customer with multiple lines has 59% more probability to churn than customer with single line.

Fit your models using test data, and compute recall, precision, F1-score, and AUC values for each of your three models. Create a table with these values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Recall | Precision | F1 - Score | AUC |
| **tele\_logit** | 0.974359 | 0.9318801 | 0.9526462 | 0.5839537 |
| **inter\_logit** | 0.8538462 | 0.8604651 | 0.8571429 | 0.7019231 |
| **both\_logit** | 0.8525641 | 0.786052 | 0.8179582 | 0.7153263 |