**TELCO CHURN - Swathi Banna**

**Data Preprocessing:**

* The following variables, which have no effect on the Dependent Variable (Churn), are dropped.

CustomerID

PhoneService (MulitpleLines is considered)

gender

Partner

TotalCharges (Tenure and Monthly charges are considered)

* There are 11 missing values in TotalCharges. However, we need not consider these as the corresponding column is dropped.
* Three new feature variables are created:

onlyphone (customers availing only phone services) 1526 values

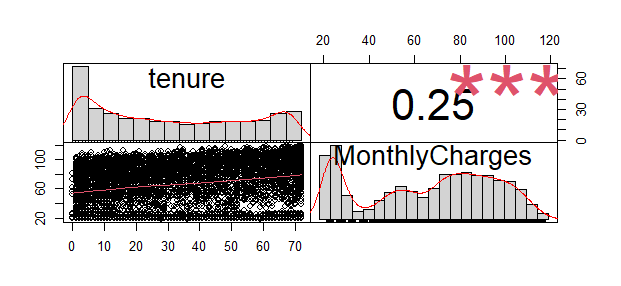
onlyint (customers availing only internet services) 682 values

phint (customers availing both phone and internet services) 4835 values

* All the categorical variables are converted to factors. Levels are changed from “Yes” or “No” to 1 and 0 respectively.

|  |  |  |
| --- | --- | --- |
| **Predictor Table DV: Churn** | | |
| **Predictor** | **Effect** | **Rationale** |
| SeniorCitizen | + | Churn is higher for Senior Citizens as they tend to stop the subscription due to meagre funds or they might be deceased. |
| Dependents | - | Customers having dependents will continue the subscriptions as they have more necessities. |
| tenure | - | The longer the tenure, the higher is the loyalty. So, Churn will be less. |
| MultipleLines | NA | This is included in the newly created features phint, onlyphone and onlyint. Hence dropping this feature. |
| InternetService | NA | This is included in the newly created features phint, onlyphone and onlyint. Hence dropping this feature. |
| OnlineSecurity | - | These features are related to “Only internet” and “Both Phone and Internet” customers.  People availing these services along with the Internet Service will have lower churn. When the product mix is high, customers tend to be loyal, and Churn is less. |
| OnlineBackup | - |
| DeviceProtection | - |
| TechSupport | - |
| StreamingTV | - |
| StreamingMovies | - |
| Contract | +/- | Customers having yearly subscription will have lesser Churn when compared to monthly payments [the chances of default in payment will be less in yearly/two-year subscriptions] |
| PaperlessBilling | + | Customers receiving physical bills will not default much when compared to those with Paperless bills. When bills are received in the mails, generally they forget to pay and then gradually become inactive. |
| PaymentMethod | +/- | Automated payment methods have lesser Churn as the payment defaults will be lesser. |
| MonthlyCharges | + | Customers with higher monthly charges may shift to Competitors’ products. So, Churn will increase. |
| phint | +/- | Used for finding the solutions to the given questions |
| onlyphone | +/- | Used for finding the solutions to the given questions |
| onlyint | +/- | Used for finding the solutions to the given questions |

**Correlation Matrix**: (For numeric variables)



There is no correlation between the numeric variables.

**Analyzing categorical variables:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total | **Services** | | | | Sr Citizen | **Contract** | | | **Payment** **Method** | | | |
| Only Phone | Only Internet | | Both | Monthly | One yr | 2 yrs | Bank | CC | eChk | MailChk |
| **Churn** | 1869 | 113 | 170 | 1586 | | 476 | 1655 | 166 | 48 | 258 | 232 | 1071 | 308 |
| **Active** | 5174 | 1413 | 512 | 3249 | | 666 | 2220 | 1307 | 1647 | 1286 | 1290 | 1294 | 1304 |

* 4835 customers have both phone and internet services and the churn is also highest in this category (1586).
* Out of the total churn, Sr citizen churn is around 25%.
* Customers having monthly subscriptions have the highest churn (88%) when compared to the ones with one or two year subscriptions.
* Customers paying through electronic checks have highest churn (57%)

**Train and Test Datasets:**

Train dataset – 5282 values; Test dataset – 1761 values

From the train and test datasets, new subsets were created for only phone, only internet and both. The models were created and tested with these subsets.

**Models:**

**logit.phone** <- glm(Churn ~ SeniorCitizen + Dependents + tenure + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges, family=binomial (link="logit"), data=train.phone)

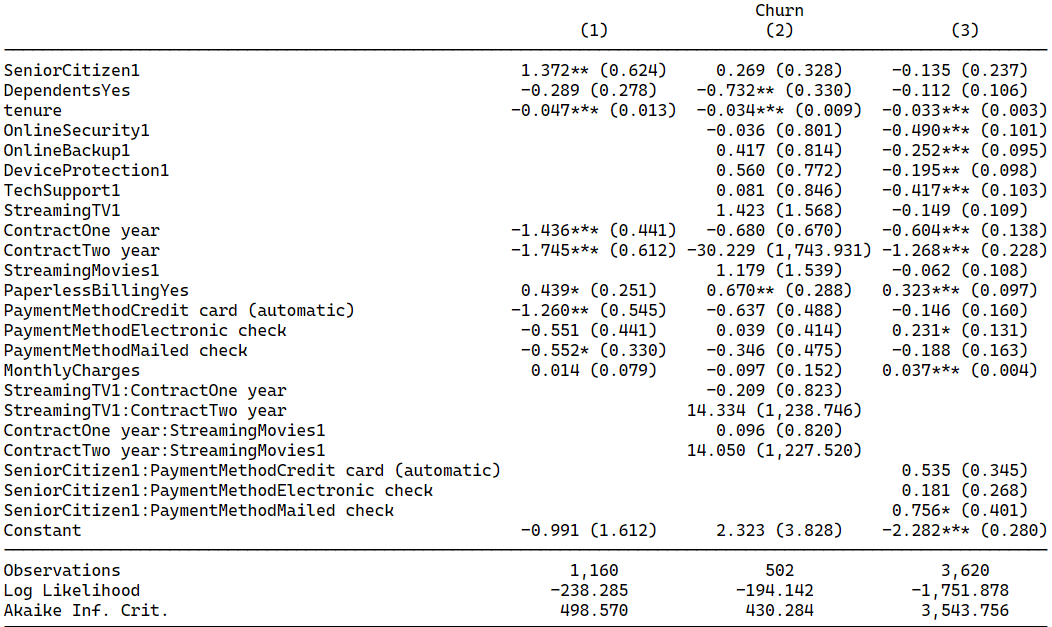
**logit.int** <- glm(Churn ~ SeniorCitizen + Dependents + tenure + OnlineSecurity + OnlineBackup +

DeviceProtection + TechSupport + StreamingTV\* Contract + StreamingMovies\*Contract +

PaperlessBilling + PaymentMethod + MonthlyCharges, family=binomial (link="logit"), data=train.int)

**logit.phint** <- glm(Churn ~ Dependents + tenure + OnlineSecurity + OnlineBackup + DeviceProtection + TechSupport + StreamingTV + StreamingMovies + Contract + SeniorCitizen\*PaymentMethod + PaperlessBilling + MonthlyCharges, family=binomial (link="logit"), data=train.phint)

stargazer::stargazer(logit.phone, logit.int,logit.phint,type="text",single.row = TRUE)

****

**Top 3 Predictors:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Only Phone** | | **Only Internet** | | **Both Phone & Internet** | |
| **Predictor** | **Marginal Effect** | **Predictor** | **Marginal Effect** | **Predictor** | **Marginal Effect** |
| Contract | Two Year 🡪 82.6% decrease in odds  One Year🡪 76.3% decrease in odds | Streaming Movies | 246.21% likely increase in odds | Contract | Two Year 🡪 71.9% decrease in odds  One Year🡪 45.4% decrease in odds |
| Senior Citizen | 294% likely increase in odds | Contract (Two year) | 92.86% likely decrease in odds | Online Security | 60.26% likely decrease in odds |
| Payment Method (Credit Card) | 72% likely decrease in odds | Dependents | 52.1% likely decrease in odds | Tech Support | 34.2% likely decrease in odds |

**Interactions**:

When we consider interaction terms in “Both phone and internet” category, we can observe the following:

* If a Senior citizen is paying through Mailed Check, the odds that Churn will occur increases by 112%. (exp of 0.756). If the payment is with automated credit card payment, it will increase by 70% and with Echeck, it is 19% increase. It might be because of the reduced income and less availability of funds. So, automated bank payments are the best option for this group.
* In case of Paperless billing, the odds that Churn will occur increases by 38%. The bills will reach the inbox of the customer and he/she might not look at it. Hence, there are chances of default in these cases and gradually the customer becomes inactive.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Only Phone** | **Only Internet** | **Both** |
| **Confusion Matrix** | 0 1  0 303 38  1 10 15 | 0 1  0 106 23  1 19 32 | 0 1  0 704 108  1 146 257 |
| **Accuracy** | 86.88% | 76.66% | 79.09% |
| **ROC Plot** | Chart, line chart, scatter chart  Description automatically generated | Chart, scatter chart  Description automatically generated | Chart, line chart  Description automatically generated |
| **AUC** | 0.74 | 0.72 | 0.75 |
| **Recall** | 0.6 | 0.62 | 0.63 |
| **Precision** | 0.28 | 0.58 | 0.70 |
| **F1 ratio** | 0.38 | 0.60 | 0.66 |
| **Multi Collinearity** | No multi collinearity among features. | Multi collinearity observed for monthly charges. | No multi collinearity. |
| **Independence** | data: logit.phone  DW = 1.9417, p-value = 0.1595  No Auto-correlation observed. | data: logit.int  DW = 2.035, p-value = 0.6581  No Auto-correlation observed. | data: logit.phint  DW = 1.9829, p-value = 0.3037  No Auto-correlation observed. |