TELCO CHURN PROJECT_AR

Dataset contains information about 7043 telco subscribers of telephone and/or Internet Services.

Let's look at how to predict reasons for churn for three sets of customers-

1.Telephone Users,

2.Internet Service Users, &

3.Users of both services.

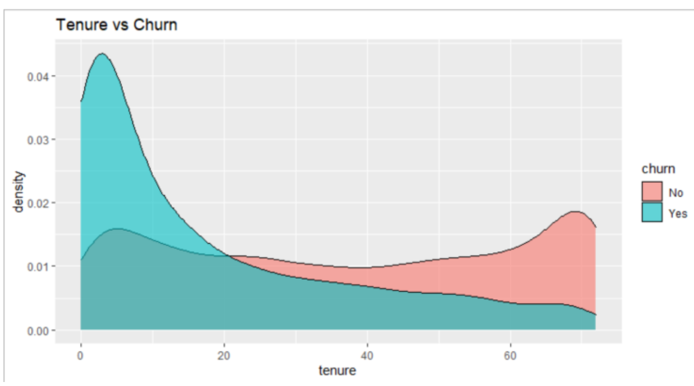
DATA CLEANING & FEATURE ENGINEERING

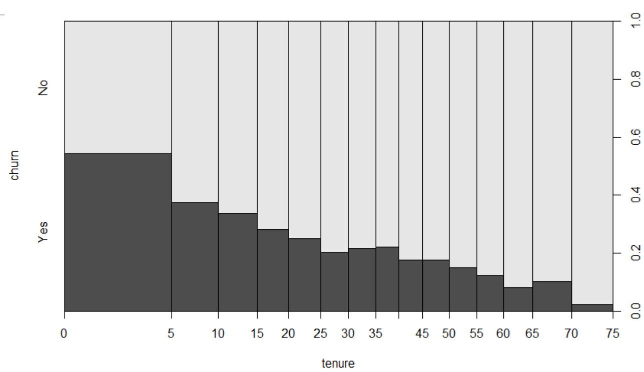
First things first, let's clean the data and make it useful for analysis. The following steps were done to clean the data in the primary step.(Code can be found in the R file).

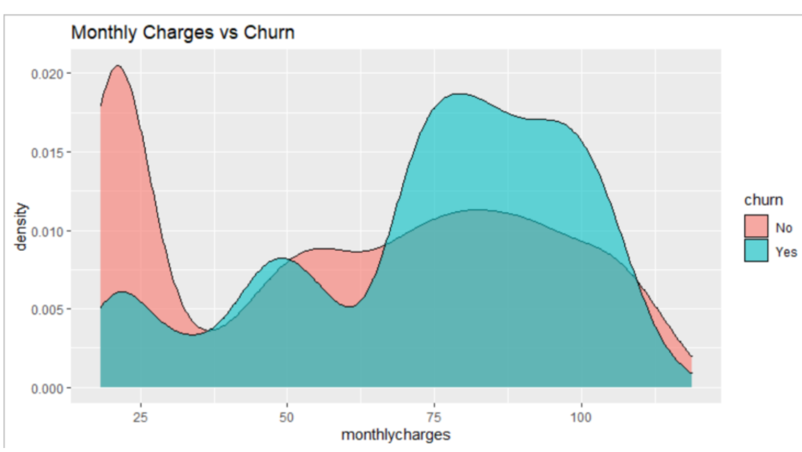
- Removed correlated variables as shown in the predictor table.
- All binary columns with "Yes/No" were changed to "1/0" to make it useful for analysis.
- Created new data frame for useful variables for modelling and further feature engineering and subsetting.
- Factorized the character column features which are required for analysis-Gender, Contract, Payment Method, Senior Citizen ,Internet Service.
- Regarding Data Partition, created three new subsets of the data for phone only customers, internet only customers, customer who use both services using tidyverse and then created 75-25 splits for training and testing the model as shown in last class.

EXPLORATORY DATA ANALYSIS

Let's go ahead and plot some visualizations to understand the data better.







BASIC OBSERVATIONS-

- Customers churn is highest in 1 to 10 months of usage, and after 60 months in the dataset.
- Customer churn is highest when monthly charges are between $1 to 30/month, and between $70 to $100.

CHOOSING VARIABLES FOR MODELS

The logical reasoning behind choosing variables for models is as given below in the predictor table-

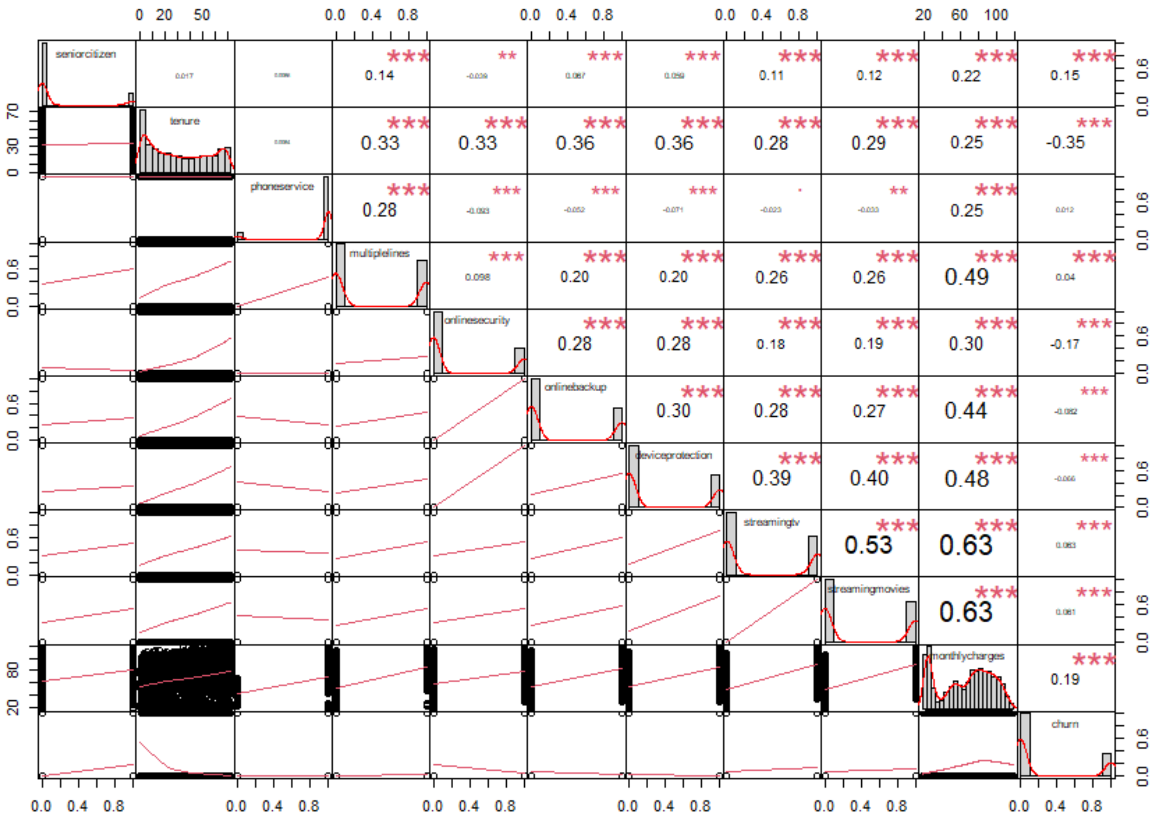
PREDICTOR TABLE

Green => Useful for Analysis for that specific customer type with hypothesized direction of effect on Churn.

Red => Not useful for Analysis for that specific customer type.

Y VARIABLE - Churn -1/0				
Predictor Variables	**RATIONALE**	**telephone**	**internet**	**both**
customerID	None. Unique ID.	None	None	None
gender	Phone usage and internet usage amounts might vary by gender, so we can include to understand bias and preference to check effect on churn.	?	?	?
SeniorCitizen	Can be used to understand bias/preference and Seniors Citizens might use only phone over internet.	?	None	None
Dependents, Partner	Customers who have dependents/partners will have multiple lines which is already available in the data.	None	None	None
tenure	Longer/Loyal customers tend to churn less.	-	-	-
PhoneService, InternetService	DATA FILTER. Used for feature Engineering, not modeling.	None	None	None
MultipleLines	Customers with multiple lines will have more difficulty in churning and might churn less for phone customers.	-	None	None
OnlineSecurity	Security might be important in deciding to stay with current provider/move to another provider.	None	-	-
OnlineBackup	Customer preferences for a provider's service will be an important decision point for churn.	None	-	-
DeviceProtection	Customer preferences for a provider's service will be an important decision point for churn for internet / customers using both services.	None	-	-
TechSupport	The above three variables are more granular and a type of tech support, so this can be excluded.	None	None	None
StreamingTV	Customers who opt for bundled services tend to churn less.	None	-	-
StreamingMovies	Customers who opt for bundled services tend to churn less.	None	-	-
Contract	People under contract will churn less in reality.	-	-	-
PaperlessBilling	Just an option. Shouldn't affect churn significantly.	None	None	None
PaymentMethod	Auto Pay customers might churn less since they are used to the provider's services and have enrolled in auto pay.	?	?	?
MonthlyCharges	Rapid Increase/ Irregular inconsistent might make customers churn.	?	?	?
TotalCharges	Correlated & calculated from Monthly Charges and tenure.	None	None	None

PERFORMANCE ANALYTICS CHART



We can observe that, for the chosen variables, none of them are highly correlated to churn, whereas some features/services such as multiple lines, streaming & technical services highly affect the price, as expected.

WHAT IS CAUSING CHURN & WHICH VARIABLES HAVE THE HIGHEST MARGINAL EFFECT?

I am using logit models for the 3 Y variables i.e., to predict churn for the different types of customers – Phone Users (Model 1), Internet Users (Model 2) and Users of both services (Model 3).

```
#logit Models on churn for the 3 Y Variables

phone <- glm(churn ~ gender+ seniorcitizen+tenure+multiplelines+ contract+paymentmethod+
               monthlycharges, family=binomial (link="logit"), data=trainphone)

net <-glm(churn ~ gender+ tenure+onlinesecurity +onlinebackup+deviceprotection+ streamingtv+
            streamingmovies+contract+paymentmethod+ monthlycharges, family=binomial (link="logit"), data=train_internet)

both <-glm(churn ~ gender+ tenure+onlinesecurity +onlinebackup+deviceprotection+ streamingtv+
             streamingmovies+contract+paymentmethod+ monthlycharges, family=binomial (link="logit"), data=trainboth)

library(stargazer)
stargazer(phone,net,both, type="text", single.row=TRUE)
```

Model Outputs and Beta Coefficients in the next section.

STARGAZER OUTPUT OF MODELS

```
===========================================================================================
                                                     Dependent variable:
                                       ----------------------------------------------------
                                                            churn
                                              (1)               (2)               (3)
-------------------------------------------------------------------------------------------
genderMale                               -0.019 (0.246)    0.515** (0.256)   -0.071 (0.083)
seniorcitizen
tenure                                  -0.048*** (0.013) -0.035*** (0.008) -0.031*** (0.003)
multiplelines                            -0.562 (1.280)
onlinesecurity                                              0.256 (0.428)   -0.407*** (0.122)
onlinebackup                                                0.645 (0.400)    -0.176 (0.115)
deviceprotection                                           1.045** (0.416)   -0.088 (0.117)
streamingtv                                               1.888*** (0.656)    0.221 (0.168)
streamingmovies                                            1.612** (0.636)    0.153 (0.167)
internetserviceFiber optic                                                  0.955*** (0.343)
contractOne year                        -1.338*** (0.436) -1.219*** (0.423) -0.501*** (0.136)
contractTwo year                        -1.640*** (0.593) -1.903** (0.786)  -1.347*** (0.224)
paymentmethodBank transfer (automatic)    0.421 (0.348)     0.328 (0.454)    -0.107 (0.150)
paymentmethodCredit card (automatic)     -1.006* (0.543)   -0.063 (0.423)     0.060 (0.152)
paymentmethodElectronic check             0.398 (0.369)    0.777** (0.341)  0.374*** (0.126)
monthlycharges                            0.045 (0.235)   -0.148** (0.058)    0.005 (0.013)
Constant                                 -2.038 (4.687)    3.189** (1.522)   -0.780 (0.624)
-------------------------------------------------------------------------------------------
Observations                                 1,144              512              3,626
Log Likelihood                              -227.849          -203.086         -1,759.796
Akaike Inf. Crit.                            475.699           434.172          3,549.591
===========================================================================================
Note:                                                        *p<0.1; **p<0.05; ***p<0.01
> |
```

INFERENCES FOR CHURN FROM STARGAZER OUTPUT BETA COEFFICIENTS

Note - Representation in Green implies these factors bring down churn whereas the ones in light red represent factors that are influencing and increasing customer churn.

Phone only customers:

Predictor	**β**	**exp(β)**	**Interpretation**
Two-year contract	-1.640	0.194	Having a 2-year contract reduces the odds of churn by 80.6% over customers with month-to-month service.
One year contract	-1.338	0.262	Having a 1-year contract reduces the odds of churn by 73.7% over customers with month-to-month service.
CC payment	-1.006	0.365	Customers who pay by credit card (e.g., autopay) have 63.4% less odds of churn than people who pay using other means.
Bank Transfer(Auto) payers	0.421	1.523	Customers who pay by auto bank transfer have 52.3% more odds of churn than people who pay using other means.
Electronic Check payers	0.398	1.488	Customers who pay by electronic check have 48.8% more odds of churn than people who pay using other means.

Internet only customers:

Predictor	β	exp(β)	Interpretation
Two-year contract	-1.903	0.163	Having a 2-year contract reduces odds of churn by 83.7% over customers with month-to-month service.
One-year contract	-1.219	0.507	Having a 1-year contract reduces odds of churn by 49.3% over customers with month-to-month service.
CC Payers	-0.063	0.938	Customers who pay by credit card (e.g., autopay) have 6.1% less odds of churn than people who pay using other means.
Streaming TV	1.888	6.606	Customers who are using the streaming TV service have a whopping **560.6% more** odds of churn than people who use regular services!
Streaming Movies	1.612	5.012	Customers who are using the streaming movies service have a whopping **401.2% more** odds of churn than people who use regular services!
Device Protection	1.045	2.843	Customers who are using the device protection service have **184.3% more** odds of churn than people who use regular services
Electronic Check payers	0.777	2.174	Customers who pay by electronic check have 117.4% more odds of churn than people who pay using other means.
Online Backup	0.645	1.906	Customers who are using the online backup service have 90.6% more odds of churn than people who use regular services.
Male	0.515	1.673	Male customers have 67.3% higher odds of churn than female customers. This might be due to most male users and above factors influencing churn, not just the gender.
Bank Transfer(Auto) payers	0.328	1.388	Customers who pay by auto bank transfer have 38.8% more odds of churn than people who pay using other means.
Online Security	0.256	1.292	Customers who are using the online security service have 29.1% more odds of churn than people who use regular services.

Both phone and Internet customers:

Predictor	β	exp(β)	Interpretation
Two-year contract	-1.347	0.260	Having a 2-year contract reduces odds of churn by 73.9% over customers with month-to-month service.
One-year contract	-0.501	0.605	Having a 1-year contract reduces odds of churn by 39.4% over customers with month-to-month service.
Fiber Optic customers	0.955	2.598	Customers who are using the internet fiber optic service have 159.8**% more** odds of churn than people who use DSL internet service.
Electronic Check payers	0.374	1.453	Customers who pay by electronic check have 45.3% more odds of churn than people who pay using other means.
Streaming TV	0.221	1.247	Customers who are using the streaming TV service have 24.7**% more** odds of churn than people who use regular services.
Streaming Movies	0.153	1.165	Customers who are using the streaming movies service have 16.5% more odds of churn than people who use regular services!

METRICS OF ML MODELS

Type of model	**Recall**	**Precision**	**F1 score**	**AUC**	**Accuracy**
Phone	0.99	0.91	0.95	0.5	91.8%
Internet	0.95	0.77	0.85	0.63	77.1%
Both	0.84	0.80	0.82	0.73	76.8%

The estimates below suggest that the phone model has greater predictive accuracy, and internet and both models almost have similar accuracy.

Model 1 –

```
> print(recall_phone)
[1] 0.994302
> print(precision_phone)
[1] 0.9188482
> print(f1_phone)
[1] 0.9561644
> print(auc_phone)
[1] 0.5
> print(Accuracy(predictedz,testphone$churn))
[1] 0.9188482
> table(testphone$churn,predictedz)
   predictedz
      0
  0 351
  1  31
```

Model 2 –

```
> print(recall_net)
[1] 0.9586777
> print(precision_net)
[1] 0.7733333
> print(f1_net)
[1] 0.8560886
> print(auc_net)
[1] 0.6324001
> print(Accuracy(predictedzz,test_internet$churn))
[1] 0.7705882
> table(test_internet$churn,predictedzz)
   predictedzz
      0   1
  0 116   5
  1  34  15
```

Model 3 –

```
> print(recall_both)
[1] 0.8410256
> print(precision_both)
[1] 0.8078818
> print(f1_both)
[1] 0.8241206
> print(auc_both)
[1] 0.7386946
> print(Accuracy(predictedzzz,testboth$churn))
[1] 0.7684036
> table(testboth$churn,predictedzzz)
   predictedzzz
      0   1
  0 656 124
  1 156 273
```

ACTIONABLE RECOMMENDATIONS FOR TELCO TO REDUCE CUSTOMER CHURN

For Phone only customers :

Promote contract offers via ads and discounts. Issue cashbacks/ small discount for users paying through Credit Card autopay to decrease customer churn.

For Internet only customers :

Similar to phone customers, tie down new customers to contracts and CC Autopay through cashbacks and discounts.

Infrastructure & Technology level of Streaming Movies, Streaming TV, Online Backup, Online Security & Device Protection needs to be massively improved. A technical root cause analysis is needed because these factors are influencing the highest churn among internet customers.

For customers of both services :

Similar to phone customers, tie down new customers to contracts and CC Autopay through cashbacks and discounts.

Fiber Optic service customers have high churn rate over DSL customers. Technology and Infrastructure of fiber optic service needs to be improved. At the same time, consider offers and discounts to lure new customers to sign up for this service post improvement.

Streaming TV and Streaming movies infrastructure and content needs to be improved via good customer recommendations and overall quality as these have high churn rate.