Aryan Sharma
TelcoChurn

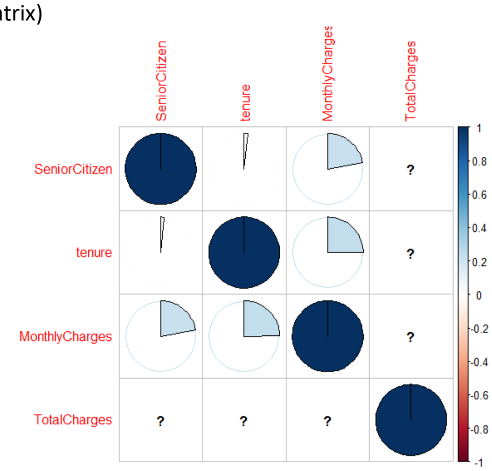
Exploratory Data Analysis & Processing

The data has 21 variables and 7043 observations. Out of the 21 variables a lot of the variables are character variables, we can either change them to factor variables or can change them in numeric sequences. Here I have changed the character variables into numeric variables, for e.g.: Partner, Dependents, etc. From the main dataset I have created subsets defined by whether a customer is only having internet, telephone connection or both. We are predicting the churn rate of the customers of the telco services and there are 2 different services and some customers have opted for both the services.

Correlation:

There are only 4 variables in the data which are numeric. Hence, we are calculating whether they have any correlation between them or not.

```
correlation_matrix <- cor(data[c("SeniorCitizen", "tenure","MonthlyCharges","TotalCharges")])
corrplot(correlation_matrix, method = "pie")
print(correlation_matrix)
```



There is a correlation between total charges and monthly charges, hence we are only taking Monthly charges the variables.

Predictor table:

Predictor	Phone Connection only	Internet Connection only	Both Connections	Rationale
Gender	+/-	+/-	+/-	Gender can be a factor in churning rate and can be classified as a human factor as some gender may take quick decision
Partner & Dependents	+/-	+/-	+/-	The number of family member can lead to more lines and hence changing a family plan can be taxing and may lead to sustaining
Streaming movies	-	-	-	If extra services like streaming is offered by provider, then it can lead to less churn rate
Online Backup	+	-	-	Having a backup of data might help in retaining customers and it won't be helpful for phone connection only customers
Online Security	-	-	-	If the service provider is providing good security over data and other services then churn rate might decrease as security is a major factor in data privacy nowadays
Multiplelines	-	()	-	Generally, multiple lines only effect phone connection and intenet is not there, multiple lines mean family plan & attrition rate is less when customer has family plan

Contract	+/-	+/-	+/-	If the contract is for a longer duration, then the churn rate of customer is less which is good and vice versa is also true
Techsupport	-	-	-	If the tech support is good and solve all the customer related problems then the churn rate would be less
SeniorCitzen	-	-	-	Churn rate should be less in senior citizens as they do not want to change what they have
Monthly charges	+	+	+	If the monthly bill amount is more then the churn rate can be higher
Paperlessbilling and Paymentmethod	+/-	+/-	+/-	This can be an indivilusitic factor depending upon a person whether or not they want a type of service and they might churn if that service is not available

Regression Analysis

We have taken 3 models for this analysis, Model 1 has only the people who have phone connection.

```
phonemodel = glm(Churn ~ gender + SeniorCitizen + Partner*Dependents +
                     tenure*MultipleLines + Contract + PaperlessBilling +
                     PaymentMethod + MonthlyCharges+TechSupport, data=trainset_phone,
                   family=binomial (link="logit"))

internetmodel = glm(Churn ~ gender + SeniorCitizen + Partner*Dependents +
                        tenure*MultipleLines + OnlineSecurity + OnlineBackup + DeviceProtection +
                        TechSupport + StreamingTV + StreamingMovies + Contract +
                        PaperlessBilling + PaymentMethod + MonthlyCharges,
                      data=trainset_internet, family=binomial (link="logit"))

bothmodel = glm(Churn ~ gender + SeniorCitizen + Partner*Dependents +
                    tenure*MultipleLines + OnlineSecurity + OnlineBackup +
                    DeviceProtection + TechSupport + StreamingTV + StreamingMovies +
                    Contract + PaperlessBilling + PaymentMethod + MonthlyCharges,
                  data=trainset_both, family=binomial (link="logit"))
```

	Dependent variable:		
		Churn	
	(1)	(2)	(3)
genderMale	-0.041 (0.251)	0.178 (0.253)	-0.079 (0.083)
SeniorCitizen1	1.025 (0.684)	0.193 (0.329)	0.194* (0.104)
Partner	-0.550 (0.549)	0.580* (0.323)	-0.092 (0.107)
Dependents	-0.482 (0.508)	-0.497 (0.532)	-0.073 (0.210)
tenure	-0.055*** (0.016)	-0.039*** (0.009)	-0.041*** (0.005)
MultipleLines	-1.247 (1.422)		-0.099 (0.141)
OnlineSecurity		-0.106 (0.812)	-0.610*** (0.103)
OnlineBackup		-0.038 (0.831)	-0.289*** (0.096)
DeviceProtection		0.109 (0.830)	-0.147 (0.099)
ContractOne year	-1.252*** (0.447)	-0.671* (0.384)	-0.490*** (0.138)
ContractTwo year	-1.532** (0.635)	-1.809** (0.795)	-1.176*** (0.234)
PaperlessBilling	0.285 (0.265)	0.141 (0.277)	0.355*** (0.098)
PaymentMethodBank transfer (automatic)	0.444 (0.358)	0.346 (0.423)	-0.184 (0.155)
PaymentMethodCredit card (automatic)	-0.999* (0.545)	-0.036 (0.423)	-0.031 (0.154)
PaymentMethodElectronic check	0.355 (0.380)	0.645* (0.332)	0.274** (0.128)
MonthlyCharges	0.035 (0.237)	-0.024 (0.157)	0.035*** (0.004)
TechSupport		-0.498 (0.854)	-0.541*** (0.105)
StreamingTV		0.530 (1.601)	-0.087 (0.110)
StreamingMovies		0.545 (1.566)	-0.141 (0.109)
Partner:Dependents	0.646 (0.789)	-0.366 (0.677)	0.126 (0.247)
tenure:MultipleLines	0.037 (0.025)		0.011** (0.005)
Constant	-1.769 (4.733)	0.256 (3.917)	-2.061*** (0.271)
Observations	1,144	511	3,626
Log Likelihood	-223.812	-206.951	-1,738.440
Akaike Inf. Crit.	479.625	453.902	3,520.881
Note:			*p<0.1; **p<0.05; ***p<0.01

Interpretation

1. Only Phone Line Customers

Predictor	Beta	Exp Beta	Rationale
Multiple Lines	-1.247	0.2873	Churn rate of customer with multiple lines would be reduced to 71.27%
Contract Year One	-1,252	0.2858	Churn rate of customer with multiple lines would be reduced to 71.42%
Contract Year two	-1.532	0.2160	Churn rate of customer with multiple lines would be reduced to 78.4%

2. Only Internet Line Customers

Predictor	Beta	Exp Beta	Rationale
Dependents	-0.497	0.6083	Churn rate of customer with multiple lines would be reduced to 78.4%
Contract Year One	-0.671	0.5112	Churn rate of customer with multiple lines would be reduced to 78.4%
Contract Year two	-1.809	0.1637	Churn rate of customer with multiple lines would be reduced to 78.4%

3. Customers who have both internet as well as phone line

Predictor	Beta	Exp Beta	Rationale
Online Security	-0.610	0.5432	Churn rate of customer with multiple lines would be reduced to 46.68%
Contract Year One	-0.490	0.6125	Churn rate of customer with multiple lines would be reduced to 39.75%
Contract Year two	-1.176	0.3086	Churn rate of customer with multiple lines would be reduced to 69.14%

4. F1 Score and Precision/Recall

Predictor	Precision Score	Recall	F1 Score	AUC Score
Phone Service User Only	0.9186	0.9971	0.9562	0.7094
Internet Service User Only	0.8201	0.9268	0.8702	0.6528
Both Internet and Phone Service user	0.8142	0.8279	0.8279	0.6797

Recommendations:

Explore additional features or transform existing ones to capture more complex relationships in the data. Consider interaction terms, polynomial features, or domain-specific transformations to better represent the underlying patterns. Regularization helps control model complexity and prevents extreme parameter estimates, leading to better performance on unseen data.