Telco Churn

Predictor Table:

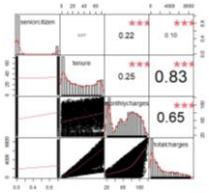
DV: Churn

Predictors	Expected Effect	Rationale
Dependents, Monthly charges	+	Customers with more dependents will eventually have higher bills and would be most likely looking for better options every time.
Tenure, Contract	-	Higher the time the customer is with a telecom, the more he is habitual of the service and hence lesser churn probability.
Multiple lines	-	More lines would mean more initial time, effort and cost for setup which may lead customers to stick to the plan.
Payment Method	+	The more the effort to pay manually, the more likelihood to cancel the service at any time. (From automatic to manual)
Streaming tv	-	If customers use this, they are less likely to churn as they might be using exclusive content which they would want to stay updated on.
Device protection	-	Customers who avail this have tech support and secure options included and are less likely to churn.
Online Backup	-	Data is of utmost importance for anyone, and this service might keep customers retained as they will be able to access it ubiquitously.
Excluded Factors		
Gender, Senior citizen, partner	NA	Demographics and age might not affect churn.
customerID	NA	Not considered as it's a unique identifier.
Paperless Billing	NA	This is covered under payment method.
Total charges	NA	Since monthly charge and tenure is taken, this can be calculated to an approximation.

Cleaning and Processing

Converting gender, partner, contract, payment method to factor.
Partitioned the data to subsets for three categories asked: phone, internet, and both.
For phone, phone service=1 and internet service=0,
For internet, phone service=0 and internet service=1,
For both, phone service=1 and internet service=1.
Data is quite imbalanced. For classification, we will randomly split it to training and test sets.
As there were missing values in total charges, multiplied tenure and monthly charges to fill up those values.
Converted all the yes/no/no service values (multiple lines, online security, online backup, device protection, tech support, streaming tv, streaming movies) to 0 and 1 factors.
Converted internet services DSL, fibre optic to 1 and no service to 0 as factors.
Tenure and total charges are correlated as R>0.7 hence not taking total charges.

```
> table(d$churn)

  No  Yes
5174 1869
```



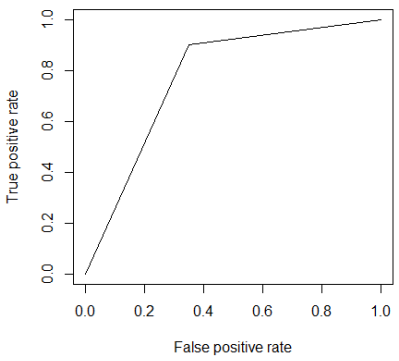
Modelling

```
logit_p  <- glm(churn ~ dependents + tenure + multiplelines + contract +
          paymentmethod + monthlycharges, family=binomial (link="logit"), data=train_p)

logit_i  <- glm(churn ~ dependents + tenure + contract + onlinebackup +
          deviceprotection + paymentmethod + monthlycharges +
          streamingtv, family=binomial (link="logit"), data=train_i)

logit_b  <- glm(churn ~ dependents + tenure + contract + multiplelines +
          onlinesecurity+ onlinebackup+ deviceprotection+ techsupport+
          paymentmethod + monthlycharges+ streamingtv, family=binomial (link="logit"), data=train_b)
```

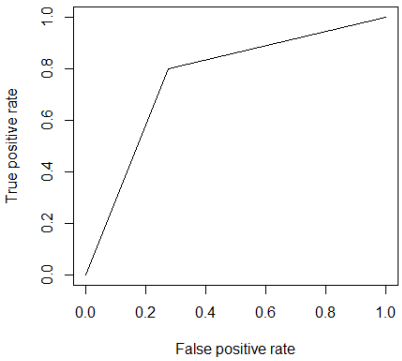
3. Stargazer

	churn		
	(1)	(2)	(3)
dependents	-0.419 (0.294)	-0.871*** (0.314)	-0.138 (0.105)
tenure	-0.047*** (0.013)	-0.027*** (0.007)	-0.037*** (0.003)
multiplelines	-0.558 (1.277)		0.183* (0.097)
onlinesecurity			-0.487*** (0.101)
contractOne year	-1.291*** (0.439)	-1.039*** (0.372)	-0.485*** (0.135)
contractTwo year	-1.539** (0.603)	-2.015*** (0.650)	-1.061*** (0.220)
onlinebackup		-0.095 (0.281)	-0.264*** (0.095)
deviceprotection		0.093 (0.318)	-0.199** (0.099)
techsupport			-0.434*** (0.103)
paymentmethodCredit card (automatic)	-1.400** (0.598)	-0.387 (0.433)	0.047 (0.142)
paymentmethodElectronic check	0.041 (0.453)	0.341 (0.349)	0.302*** (0.116)
paymentmethodMailed check	-0.396 (0.349)	-0.624 (0.402)	0.023 (0.151)
monthlycharges	0.048 (0.235)	-0.020 (0.023)	0.038*** (0.004)
streamingtv		0.437 (0.396)	-0.180* (0.108)
Constant	-1.635 (4.700)	0.955 (0.772)	-2.246*** (0.260)
Observations	1,144	512	3,626
Log Likelihood	-226.791	-223.587	-1,765.814
Akaike Inf. Crit.	473.582	471.174	3,561.628

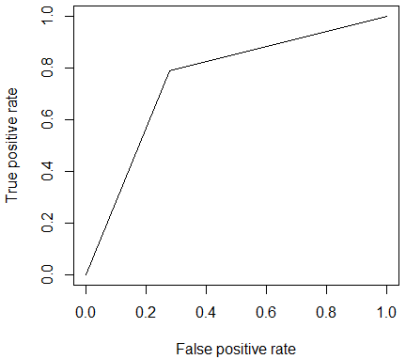
Phone(0.044)



Internet(0.338)



Both(0.333)



Assumptions

We know that glm models are robust to Linearity, normality and heteroscedasticity assumptions, so we will test multicollinearity and independence.

	Multicollinearity	Autocorrelation
Phone	GVIF Df GVIF^(1/(2*Df)) dependents 1.040122 1 1.019864 tenure 1.529060 1 1.236552 multiplelines 6.813726 1 2.610311 contract 1.449810 2 1.097306 paymentmethod 1.104554 3 1.016712 monthlycharges 6.788375 1 2.605451	lag Autocorrelation D-W Statistic p-value 1 0.02033319 1.958907 0.466 Alternative hypothesis: rho != 0
Internet	GVIF Df GVIF^(1/(2*Df)) dependents 1.035539 1 1.017614 tenure 1.585760 1 1.259270 contract 1.333852 2 1.074675 onlinebackup 1.297910 1 1.139259 deviceprotection 1.622178 1 1.273647 paymentmethod 1.279348 3 1.041913 monthlycharges 4.147051 1 2.036431 streamingtv 2.689287 1 1.639904	lag Autocorrelation D-W Statistic p-value 1 0.0252037 1.942276 0.536 Alternative hypothesis: rho != 0
Both	GVIF Df GVIF^(1/(2*Df)) dependents 1.032848 1 1.016291 tenure 2.230639 1 1.493532 contract 1.673992 2 1.137466 multiplelines 1.375193 1 1.172686 onlinesecurity 1.100552 1 1.049072 onlinebackup 1.206224 1 1.098282 deviceprotection 1.303736 1 1.141812 techsupport 1.162589 1 1.078234 paymentmethod 1.220549 3 1.033775 monthlycharges 2.610974 1 1.615851 streamingtv 1.708719 1 1.307180	lag Autocorrelation D-W Statistic p-value 1 0.02159047 1.956225 0.21 Alternative hypothesis: rho != 0
Status	Pass as all values <10	Pass as all values ~2

Interpretations

Phone Service	**Multiple lines**: the odds of churning for a person with multiple lines is 0.57 times that of odds of person with no multiple lines (43% less) **Contract**: the odds of churn for people with 2yr contract are 0.21 times of the odds of people with monthly contract (78 % less). **Payment Method**: the odds of churn for people with automatic credit card payments are 0.24 times of the odds of people with bank transfer (76% less likely to churn)
Internet Service	**Streaming tv**: the odds of churning of a person with this service is 1.54 times of the odds of a person without streaming tv facility. (54% more) **Contract**: the odds of churn for people with 2yr contract are 0.13 times of the odds of people with monthly contract (87% less likely to churn). **Payment Method**: the odds of churn for people with automatic credit card payments are 0.67 times of the odds of people with bank transfer (33% less)
Both	**Online backup**: the odds of churn of a person with this service is 0.76 times of the odds of a person without backup facility (24% less). **Contract**: the odds of churn for people with 2yr contract are 0.34 times of the odds of people with monthly contract (66% less). **Payment Method**: the odds of churn for people with automatic credit card payments are 1.04 times of the odds of people with bank transfer (4% more), *the odds of churn for people with e-check payments are 1.35 times of the odds of people with bank transfer* (35% more)

Based on explainability

Metrics

Phone

```
> cat("Accuracy =
Accuracy =  0.67
> cat("Precision =
Precision =  0.19
> cat("Recall = ",
Recall =  0.9
> cat("F1-score =
F1-score =  0.31
> cat("AUC = ", ro
AUC =  0.76
```

Internet

```
> cat("Accuracy =
Accuracy =  0.74
> cat("Precision
Precision =  0.43
> cat("Recall = "
Recall =  0.8
> cat("F1-score =
F1-score =  0.56
> cat("AUC = ", r
AUC =  0.76
```

Both

```
> cat("Accuracy =
Accuracy =  0.75
> cat("Precision =
Precision =  0.6
> cat("Recall = ",
Recall =  0.79
> cat("F1-score =
F1-score =  0.68
> cat("AUC = ", ro
AUC =  0.76
```