Telco Churn – Classification Analysis

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

Data Pre-processing

- Converted columns names to lowercase to achieve consistency.
- Dropped 11 rows which had blank/missing values in total_charges column.
- Dropped customerid as it is not related to our analysis.
- **Generated a new feature** named streamingtype which combined streamingtv and streamingmovies representing four factor levels named TV, Movies, Both, None.
- Converted seniorcitizen, gender, partner, dependents, phoneservice, multiplelines, internetservices, onlinesecurity, oninebackup, deviceprotection, techsupport, streamingtype, contract, paperlessbilling, paymentmethod and churn variable into factor.

Data Partitioning

- To obtain data for customers subscribed for **phone service**, I filtered master dataframe into with condition **phone service = yes and internet service == no.**
- To obtain data for customers subscribed for **internet service**, I filtered master dataframe into with condition **phone service = no and internet service != no.**
- To obtain data for customers subscribed for **both phone and internet service**, I filtered master dataframe into with condition **phone service = yes and internet service != no.**

2. 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 table below display the predictors used in the analysis along with its impact, explanation and relevance with respect to customers opted for phone services, internet services and both.

Predictors	Impact	Relevance	Explanation/Rationale
Tenure	**+/-**	Phone Service Internet Service Both	Customer generally prefer to stick to one service provider for years. Higher tenure can depict customer loyalty towards the provider. In some cases customer may prefer new providers who offer comparatively better services at low rates hence I thought this variable could be crucial in the analysis.
Multiple Lines	**+/-**	Phone Service Both	Customer subscribed for multiple lines may have external and internal lines based on their purpose and quality of service. Hence people may utilize them for business needs making it a vital feature in churn analysis.
Online Security	**+/-**	Internet Service Both	Online security is crucial for factors such as data privacy and online payments etc. Customer satisfied with online security may continue to opt the service hence included it in the analysis.
Online Backup	**+/-**	Internet Service Both	Online Backup is crucial for customers as it enable them to store data into the cloud. Such service can satisfy customers and are crucial for churn analysis.
Device Protection	**+/-**	Internet Service Both	Customers may prefer device protection plans that offer more coverage in less cost. Such extra services can satisfy customers and are crucial for churn analysis.
Tech Support	**+/-**	Internet Service Both	Tech Support is crucial for new customers, senior citizens for getting accustomed with the plans and billing hence a vital feature for analysis churn behaviour.
Streaming Type	**+/-**	Internet Service Both	Customers opted for internet service can subscribe for different streaming services such as TV, Movies
Contract	**+/-**	Phone Service Internet Service Both	Depending on services subscribed and utilization purpose, customer may prefer different types contracts availability of which could decide churn behaviour of the customer.
Paperless Billing	**+/-**	Phone Service Internet Service Both	Customer may opt for paperless billing service for paying their bills. This service will enable them to track their monthly statements from anywhere.

Payment Methods	**+/-**	Phone Service Internet Service Both	Customer may have different preferences such as credit cards, electronic checks, debit cards for paying their monthly bills and hence can be vital in determining churn behaviour.
Total Charges	**+/-**	Phone Service Internet Service Both	Customers may decide to churn if company charge them huge total charges as compared to the competitors.
Senior Citizen	**+/-**	Phone Service Internet Service Both	Generally senior citizens tend to prefer sticking to the same plan for years. Senior citizens may prefer phone services more than internet services hence I have included this variable to check the marginal effect.
Partner	**+/-**	Phone Service Internet Services Both	Generally customers tend to prefer family plans and services benefitting the subscribers. Hence customers with partners prefer flexible services and contract which could be vital for churn analysis.

Variables customerID, Monthly Charges, Gender, Dependents are irrelevant or correlated with other variables and hence are not included in the analysis and are not included as the predictors.

3. 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. (3 points).

- **Model 1 - Customers subscribed for phone service.**

Logit_ps <- glm(churn ~ tenure + seniorcitizen + partner + multiplelines + contract + paperlessbilling + paymentmethod + totalcharges, family=binomial (link="logit"), data=train_ps)

- **Model 2 - Customers subscribed for internet service.**

Logit_is <- glm(churn ~ tenure + partner + onlinebackup + onlinesecurity + deviceprotection + techsupport + streamingtype + seniorcitizen + contract + paperlessbilling + paymentmethod + totalcharges, family=binomial (link="logit"), data=train_is)

- **Model 3 - Customers subscribed for both services.**

Logit_b <- glm(churn ~ tenure + partner + onlinebackup + onlinesecurity + deviceprotection + techsupport + streamingtype + multiplelines + seniorcitizen + contract + paperlessbilling + paymentmethod + totalcharges, family=binomial (link="logit"), data=train_b)

	churn		
	(1)	(2)	(3)
tenure	-0.062	-0.073**	-0.101***
	(0.076)	(0.029)	(0.010)
seniorcitizenNo	0.518	0.621*	0.211**
	(0.688)	(0.329)	(0.101)
deviceprotectionNo		0.177	0.131
		(0.305)	(0.096)
techsupportYes		-0.742**	-0.524***
		(0.313)	(0.103)
streamingtypeMovies		-0.148	-0.285**
		(0.481)	(0.145)
streamingtypeNone		-0.348	-0.374***
		(0.459)	(0.130)
streamingtypeTV		0.220	-0.270*
		(0.487)	(0.145)
partnerYes	-0.112	0.202	0.067
	(0.281)	(0.266)	(0.090)
multiplelinesYes	0.308		0.311***
	(0.514)		(0.095)
onlinebackupYes		-0.512*	-0.148
		(0.282)	(0.095)
onlinesecurityYes		-0.400	-0.570***
		(0.318)	(0.102)
contractOne year	-1.174***	-1.117***	-0.672***
	(0.403)	(0.429)	(0.139)
contractTwo year	-1.584***	-1.873**	-1.317***
	(0.561)	(0.740)	(0.228)
paperlessbillingYes	0.321	0.287	0.266***
	(0.257)	(0.272)	(0.097)
paymentmethodCredit card (automatic)	-0.977**	-0.406	0.026
	(0.488)	(0.454)	(0.143)
paymentmethodElectronic check	-0.206	0.164	0.420***
	(0.440)	(0.389)	(0.119)
paymentmethodMailed check	-0.490	-0.506	-0.061
	(0.335)	(0.439)	(0.155)
totalcharges	0.001	0.001	0.001***
	(0.003)	(0.001)	(0.0001)
Constant	-0.814**	0.464	0.375**
	(0.345)	(0.612)	(0.190)
Observations	1,140	510	3,624
Log Likelihood	-236.877	-206.031	-1,749.030
Akaike Inf. Crit.	497.754	448.061	3,536.060
Note:			*p<0.1; **p<0.05; ***p<0.01

4. What are the top three predictors of churn of (i) only telephone customers, (ii) only Internet service customers, and (iii) customers who subscribe to both phone and Internet services. Explain using marginal effects how much each predictor contributes to churn occurrence. (3 points)

- **Customers subscribed for phone service.**

Senior Citizen	Odds of a non-senior citizen person getting churned is **1.6781** times of the odds of a senior citizen if all the other variables are invariant
Paperless Billing	Odds of customer opted for paperless billing service getting churned is **1.378** times of the odds of a customer who has not opted for paperless billing
Multiple Lines	Odds of customer opted for multiple lines getting churned is **1.36** times of the odds of a customer who has not opted for multiple lines.

- **Customers subscribed for internet service.**

Senior Citizen	Odds of a non-senior citizen person getting churned is **1.861** times of the odds of a senior citizen if all the other variables are invariant
Paperless Billing	Odds of customer opted for paperless billing service getting churned is **1.332** times of the odds of a customer who has not opted for paperless billing
Streaming Type	Odds of customer subscribed for streaming TV getting churned is **1.246** times of the odds of a customer who has opted for streaming both TV and movies.

- **Customers subscribed for both services.**

Payment Method	Odds of customer opted for payment method as electronic checks getting churned is **1.522** times of the odds of a customer who has opted Bank Transfer
Multiple Lines	Odds of customer opted for multiple lines getting churned is **1.364** times of the odds of a customer who has not opted for multiple lines.
Paperless Billing	Odds of customer opted for paperless billing service getting churned is **1.304** times of the odds of a customer who has not opted for paperless billing.

5. 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. (2 points)

	Model 1 – For customers opted for phone service	Model 2 – For customers opted for internet service	Model 3 – For customers opted for both services
Precision	**0.28**	**0.5122**	**0.59**
Recall	**0.068**	**0.4884**	**0.61**
F1 Score	**0.11**	**0.5**	**0.608**
AUC value	**0.527**	**0.665**	**0.7128**