Churn Prediction Model Building Process

By grouping the data set into some of the categorical variables, I have arrived at the conclusion that there is a significant difference in the average monthly payment between the customers that subscribe to internet service and those that not. With this, it must be the case that customers that pay more should get access to more features that the telecommunications company offers. If customers do not get their money's worth, then it makes sense for them to churn out of their subscription.

```
In [87]: test.groupby('InternetService').mean() #significant difference in monthly charges
```

Out[87]:

	SeniorCitizen	tenure	MonthlyCharges
InternetService			
DSL	0.103825	32.814208	57.702186
Fiber optic	0.269406	32.853881	91.600228
No	0.020408	31.438776	20.826020

Doing the same process also enabled me to identify the significant (internet feature) variables in the data set. As seen in the grouped tables below, the significant variables are OnlineBackup, DeviceProtection, StreamingTV, and StreamingMovies. Not only that these show that customers that access the service's features tend to pay more, but it also shows that tenure is longer for customers subscribing to multiple features. Intuitively, this makes sense as customers with a lot of features but want to churn are faced with high switching costs and may be better off not churning out instead.

	SeniorCitizen	tenure	MonthlyCharges
OnlineBackup			
No	0.180095	22.293839	71.300474
No internet service	0.020408	31.438776	20.826020
Yes	0.209424	44.481675	81.547382

	SeniorCitizen	tenure	MonthlyCharges
StreamingTV			
No	0.174359	24.184615	63.270256
No internet service	0.020408	31.438776	20.826020
Yes	0.212560	40.985507	88.320048

	SeniorCitizen	tenure	MonthlyCharges
DeviceProtection			
No	0.192982	23.824561	69.971711
No internet service	0.020408	31.438776	20.826020
Yes	0.195402	44.643678	84.289655

	SeniorCitizen	tenure	MonthlyCharges
StreamingMovies			
No	0.176190	25.685714	65.825952
No internet service	0.020408	31.438776	20.826020
Yes	0.213542	40.656250	87.481771

Having this, identifying whether internet subscribing customers would churn or not would depend on how many of these features do they have. The model identified the highest paying customer in the MonthlyCharges column and compared the monthly payment of all other customers with respect to the maximum. The identification criteria is as follows:

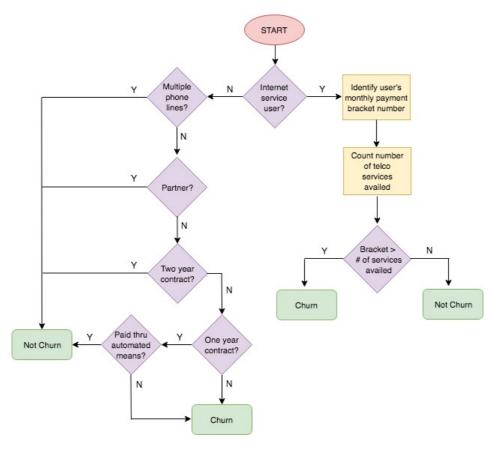| Bracket	Percentile Range	Minimum Monthly Charge	Maximum Monthly Charge	Required Criteria*
1	25%	0.0	29.4	at least one feature
2	50%	29.5	58.8	at least two features
3	75%	58.9	88.1	at least three features
4	100%	88.2	117.5	all features

*For this model, being subscribed to the phone service is considered as a feature and is taken into account in the identification criteria.

For the non-internet service users, tenure served as the 'predictor variable' in identifying whether customers would churn or not. By grouping the data set into its categorical variables, the significant variables which could be used for analysis are MultipleLines, Partner, Contract, and PaymentMethod. Insights among the following variables are as follows:

- Customers subscribing to multiple lines have longer tenure
- Partners of the telecommunications company have longer tenure
- Customers with longer contracts with the company have longer tenure, and generally pay less (on a per month basis) vis-à-vis customers with shorter contracts
- Customers that pay through automated means (i.e. bank transfer, credit card) have longer tenure vis-à-vis customers that do not

MultipleLines	SeniorCitizen	tenure	MonthlyCharges
No	0.090909	26.000000	54.406926
No phone service	0.120690	30.965517	41.510345
Yes	0.246445	40.184834	83.816588

Contract	SeniorCitizen	tenure	MonthlyCharges
Month-to-month	0.207273	19.334545	67.363273
One year	0.125000	43.557692	68.773077
Two year	0.082645	53.173554	57.715702

Partner	SeniorCitizen	tenure	MonthlyCharges
No	0.165385	25.784615	62.551154
Yes	0.154167	39.904167	68.323333

PaymentMethod	SeniorCitizen	tenure	MonthlyCharges
Bank transfer (automatic)	0.141509	42.632075	69.102830
Credit card (automatic)	0.168142	42.725664	65.334071
Electronic check	0.253012	26.090361	75.438855
Mailed check	0.034783	22.634783	47.220870

MODEL FLOWCHART

*For this model, being subscribed to the phone service is considered as a feature and is taken into account in the identification criteria.