ISP CHURN ANALYSIS

Yusuf Ahmed - 10130098

NOVEMBER 30, 2018

Contents

Table of Figures

1. Introduction

This report will inform an Internet Service Provider (ISP) why their customers are churning (leaving the service for another provider). The outline below highlights the step-by-step process used to reach this understanding.

1. **Literature Review.** This will provide context and understanding of churning of ISPs.
2. **Exploratory Analysis.** Statistics and trends will be explored to develop an understanding of the customers' behaviour.
3. **Data cleaning.** The data will then be cleaned of missing values and outliers.
4. **Attribute selection.** Exploratory analysis will indicate which attributes can be naively selected.
5. **Scaling.** Data will be normalized or discretized to simplify further analysis
6. **Clustering.** This will provide understanding of the prediction difficulty.
7. **Prediction.** Select predictors that fit the objective of identifying why customers left the ISP.
8. **Results.** Predictors will be measured, analyzed and iterated. The results should clearly indicate why customers are leaving the ISP.
9. **Recommendations.** The report will conclude with a few recommendations based on the analysis regarding customer retention and targeting.

As the problem requires determination of whether customers churn or not, this will be a classification problem.

2. Literature Review

When studying a dataset, it is helpful to have some background knowledge. This can help form hypotheses and understand relationships between attributes.

My literature review provided two key findings. Khan et al. identified that billing and usage features were more predictive than demographic features. This indicates that attributes like *PaymentMethod*, *PaperlessBilling* and *StreamingTV* may be key to the prediction, in contrast to attributes like *Partner* and *Dependents* [1].

Do et al. shared similar findings – usage data is highly predictive of churn rate. In this example usage was measured by time spent online. The closest attributes to this in the provided dataset are *StreamingTV* and *StreamingMovies* [2].

3. Exploratory Data Analysis

The objective in exploring the dataset is to discover patterns, spot anomalies and build hypotheses on what may be happening. This can be done through univariate or multivariate visualization. Exploratory Data Analysis will be done primarily using the Pandas library in Python. Through this analysis, data types, troublesome data and visualization will be explored.

Right away, we see that there are 7,042 rows and 21 columns in the dataset. Roughly 26% of the records in this dataset churned. The columns correspond to 21 attributes, one of which is the class attribute. Three of these attributes are numerical (tenure, monthly and total charges). The remainder are categorical data points, except for contract type, which is ordinal. Interestingly, the *Senior Citizen* column uses "0" and "1" rather than "yes" and "no".

Troublesome data can come in the form of **missing values, wrong data, sampling issues or bias.** Missing values can be visualized in Pandas. Figure 1 below visualizes missing values in a bar chart. As the bar chart shows, each attribute has most of the data filled in.

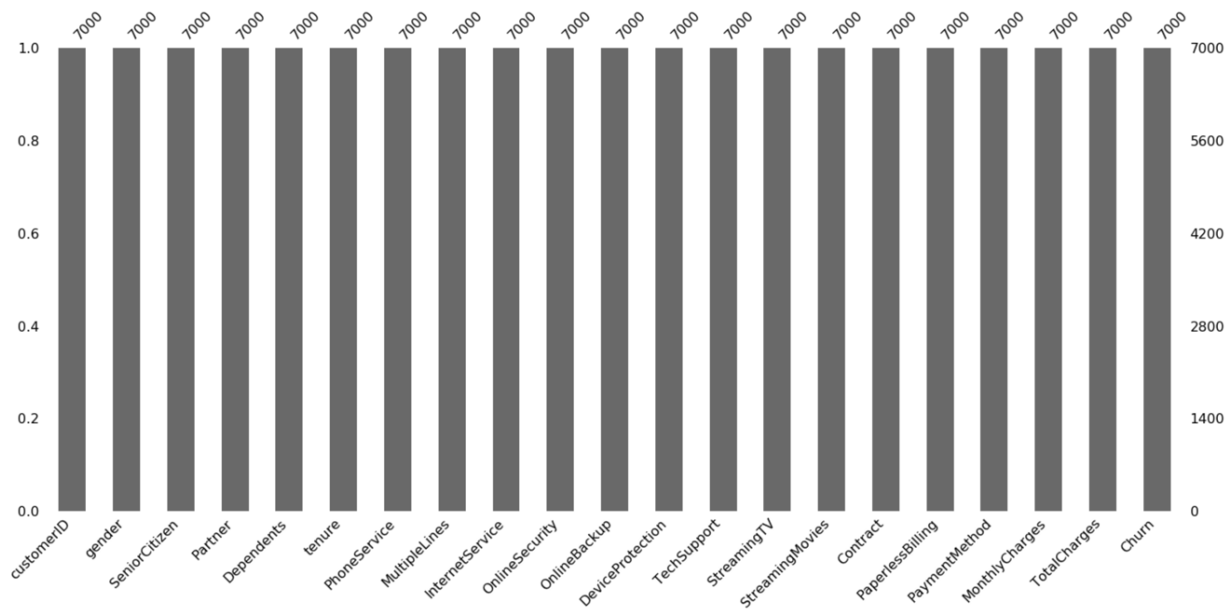


Figure 1: Chart of missin values

A closer look shows that there are missing values in the column *TotalCharges*. *TotalCharges* is an interesting column as it comes directly after *MonthlyCharges*. My initial thought was to divide the two to get total number of months, but it turns out that that is captured in the *Tenure* column. Both columns relating to charges are clearly correlated, meaning only one of the two columns need to be included. As a result of missing data and correlated attributes, *TotalCharges* will be removed from the dataset for all further analysis.

High correlation between parameters can harm future predictions for certain predictors. A correlation matrix below scores the correlation between each attribute, from 0 to 1.

	SeniorCitiz	MonthlyCh	TotalCharg	gender (to	Partner (tc	Dependent	PhoneServ	MultipleLi	InternetSe	OnlineSecu	OnlineBacl	DevicePro	TechSuppo	Streaming	StreamingI	Contract (t	PaperlessE	PaymentM	Churn (to number)
SeniorCitizen		0.220173	0.102411	-0.00187	-0.01648	-0.21119	0.008576	0.113791	-0.03231	-0.2109	-0.14483	-0.1571	-0.22377	-0.13013	-0.1208	-0.14255	-0.15653	-0.0937	0.150889
MonthlyCharges	0.220173		0.651065	-0.01457	-0.09685	-0.11389	0.247398	0.4907	-0.32326	-0.62123	-0.71048	-0.51344	-0.59759	-0.42307	-0.4246	-0.07419	-0.35215	-0.07435	0.193356
TotalCharges	0.102411	0.651065		4.78E-05	-0.31907	0.064653	0.113008	0.412495	-0.17569	-0.15323	-0.53773	-0.07739	-0.14101	-0.0757	-0.07257	0.450306	-0.15783	0.222694	-0.19948
gender (to numbe	-0.00187	-0.01457	4.78E-05		0.001808	0.010517	-0.00649	-0.00945	-8.63E-04	-0.00343	0.01223	0.005092	9.85E-04	0.001156	-1.91E-04	1.26E-04	0.011754	-0.00521	-0.00861
Partner (to numb	-0.01648	-0.09685	-0.31907	0.001808		-0.45268	-0.01771	-0.11731	-8.91E-04	-0.08185	0.090753	-0.09445	-0.06907	-0.08013	-0.07578	-0.29481	-0.01488	-0.13311	0.150448
Dependents (to n	-0.21119	-0.11389	0.064653	0.010517	-0.45268		-0.00176	-0.01966	0.04459	0.190523	0.062775	0.156439	0.180832	0.140395	0.12582	0.243187	0.111377	0.123844	-0.16422
PhoneService (to	0.008576	0.247398	0.113008	-0.00649	-0.01771	-0.00176		0.67507	0.387436	0.125353	0.12977	0.138755	0.12335	0.171538	0.165205	0.002247	-0.0165	-0.00407	0.011942
MultipleLines (to	0.113791	0.4907	0.412495	-0.00945	-0.11731	-0.01966	0.67507		0.186826	-0.06684	-0.13062	-0.01307	-0.06668	0.030195	0.028187	0.083343	-0.13325	0.025676	0.03631
InternetService (t	-0.03231	-0.32326	-0.17569	-8.63E-04	-8.91E-04	0.04459	0.387436	0.186826		0.607788	0.650962	0.662957	0.609795	0.71289	0.70902	0.099721	0.138625	0.008124	-0.04729
OnlineSecurity (tc	-0.2109	-0.62123	-0.15323	-0.00343	-0.08185	0.190523	0.125353	-0.06684	0.607788		0.621739	0.74904	0.791225	0.701976	0.704984	0.389978	0.334003	0.2138	-0.33282
OnlineBackup (to	-0.14483	-0.71048	-0.53773	0.01223	0.090753	0.062775	0.12977	-0.13062	0.650962	0.621739		0.601503	0.617003	0.604117	0.606863	0.035407	0.260715	0.003183	-0.07421
DeviceProtection	-0.1571	-0.51344	-0.07739	0.005092	-0.09445	0.156439	0.138755	-0.01307	0.662957	0.74904	0.601503		0.76797	0.763279	0.766821	0.390216	0.276326	0.191746	-0.28146
TechSupport (to r	-0.22377	-0.59759	-0.14101	9.85E-04	-0.06907	0.180832	0.12335	-0.06668	0.609795	0.791225	0.617003	0.76797		0.737578	0.737123	0.41844	0.310749	0.216878	-0.32985
StreamingTV (to r	-0.13013	-0.42307	-0.0757	0.001156	-0.08013	0.140395	0.171538	0.030195	0.71289	0.701976	0.604117	0.763279	0.737578		0.809608	0.327951	0.203907	0.117618	-0.20574
StreamingMovies	-0.1208	-0.4246	-0.07257	-1.91E-04	-0.07578	0.12582	0.165205	0.028187	0.70902	0.704984	0.606863	0.766821	0.737123	0.809608		0.330993	0.211818	0.123869	-0.20726
Contract (to num	-0.14255	-0.07419	0.450306	1.26E-04	-0.29481	0.243187	0.002247	0.083343	0.099721	0.389978	0.035407	0.390216	0.41844	0.327951	0.330993		0.176733	0.358913	-0.39671
PaperlessBilling (t	-0.15653	-0.35215	-0.15783	0.011754	-0.01488	0.111377	-0.0165	-0.13325	0.138625	0.334003	0.260715	0.276326	0.310749	0.203907	0.211818	0.176733		0.10148	-0.19183
PaymentMethod	-0.0937	-0.07435	0.222694	-0.00521	-0.13311	0.123844	-0.00407	0.025676	0.008124	0.2138	0.003183	0.191746	0.216878	0.117618	0.123869	0.358913	0.10148		-0.26282
Churn (to numbe	0.150889	0.193356	-0.19948	-0.00861	0.150448	-0.16422	0.011942	0.03631	-0.04729	-0.33282	-0.07421	-0.28146	-0.32985	-0.20574	-0.20726	-0.39671	-0.19183	-0.26282	

Figure 2: Correlation matrix - all attribtues

The matrix is colour-coded so highly correlated variables are coloured red (positive) and green (negative). Cells in yellow indicate low correlation. There are a few interesting observations:

- As expected, monthly and total charges are correlated.
- Many add-on internet services (online security, online backup, device protection, tech support, streaming movies and TV) are highly correlated. The correlation likely stems from similarities shared in between customer usage of these additional services. It may be redundant to include each of these attributes, but they will be included for now in case they have collective predictive power. This will be tested through a neural network.
- Add-on internet services are also highly correlated with monthly charges. This makes sense, as additional internet services will lead to higher monthly charges.
- *PhoneService* and *MultipleLines* are highly correlated as well. *PhoneService* indicates whether a customer has phone service or not. *MultipleLines* provides the same information, with an additional class for customers who have multiple lines. Having both would be redundant, so *MultipleLines* will be selected from the two columns as it is more informative.

Troublesome data can also come in the form of inaccurate data, or outliers. This can be quickly observed by looking at some summary statistics for total charges vs. other attributes.

Table 1: Summary statistics for numeric data

	tenure	MonthlyCharges
count	7043.000000	7043.000000
mean	32.371149	64.761692
std	24.559481	30.090047
min	0.000000	18.250000
25%	9.000000	35.500000
50%	29.000000	70.350000
75%	55.000000	89.850000
max	72.000000	118.750000

The table above shows that the average customer spent about 32 months (2.67) years with the ISP, at a rate of $64.76 per month. The longest tenure is 72 months (6 years), and the shortest is zero months. Customers who left in zero months must have quit the service in their first month. Monthly charges range from $18.25 to $118.75.

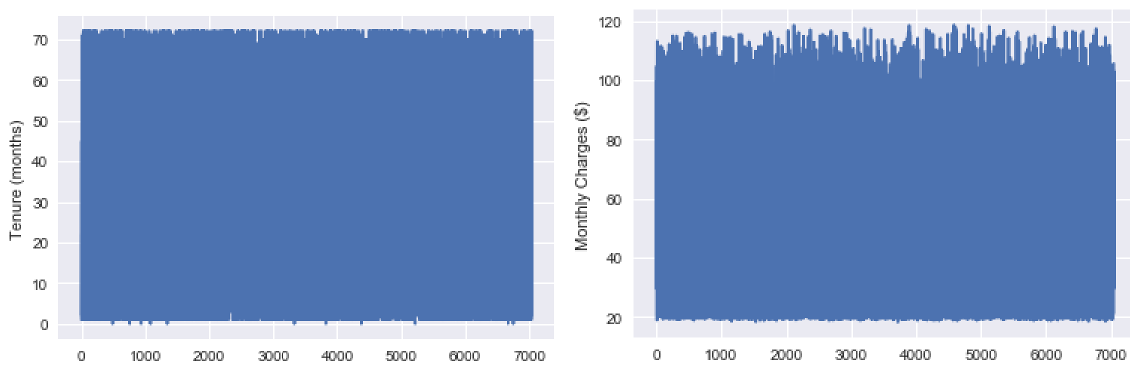


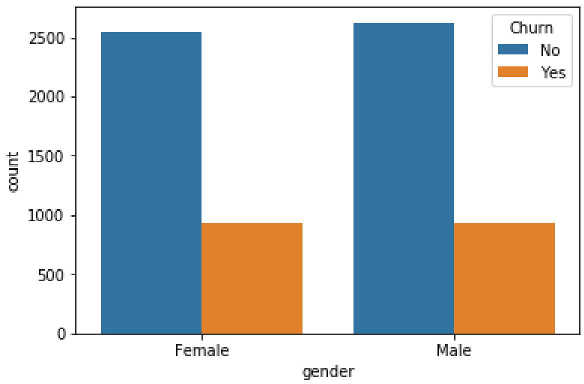
Figure 3: Tenure and monthly charges values

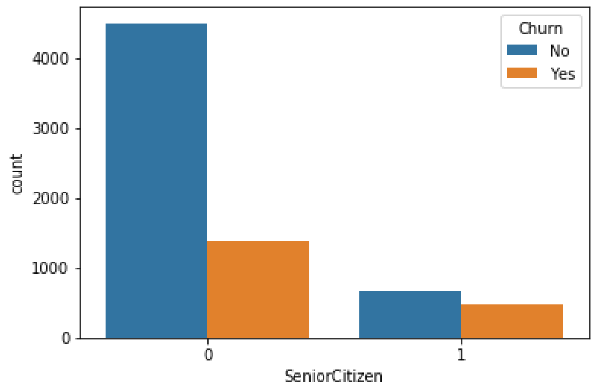
The range of values can be visualized as well, although the minimum and maximum values paint a clearer picture. Based on the reasonable range of tenure and monthly charges, it is assumed that there are no outliers in the data. Outliers can be visualized through box plots as well, as will be seen later.

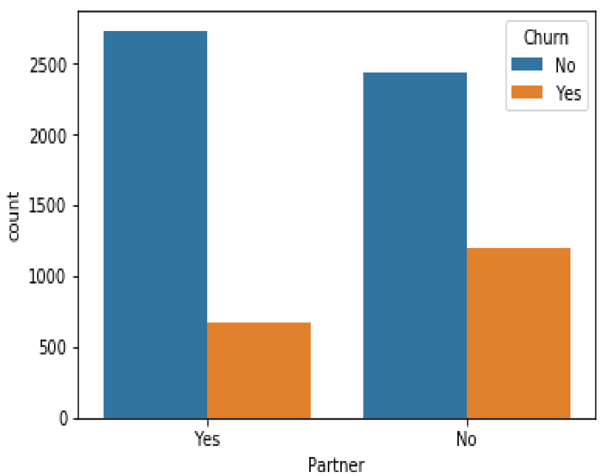
Before proceeding with the next step of the analysis, the attributes in the dataset will be analyzed to complete naïve attribute selection. *Customer ID* may be a useful attribute if the model is being used to target certain customers, but it is not useful in this context. This column will be removed from the dataset.

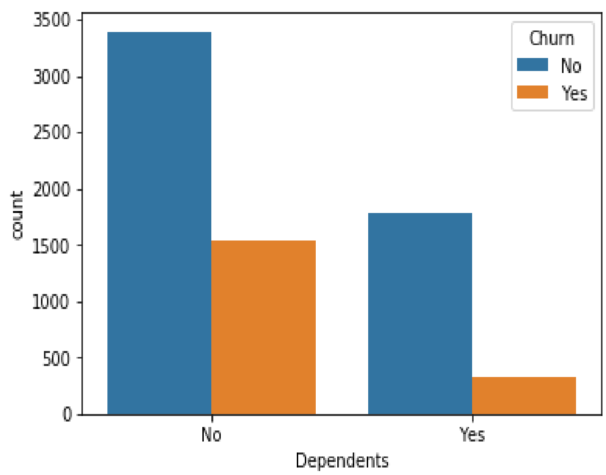
There are two columns which deal with the method of payment: *PaperlessBilling* and *PaymentMethod*. My first thoughts are that they appear to be redundant. ISP users that use paperless billing do so through electric or mailed checks, credit card or automatic bank transfer. To my surprise, customers that don’t use paperless billing pay through the same four methods. This refutes my theory that having both columns is redundant. This is justified by seeing that the two columns have a correlation of 0.101

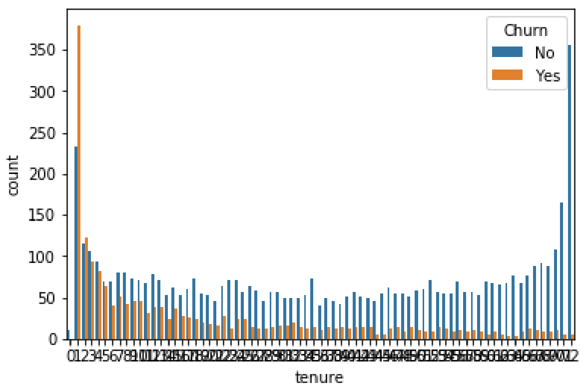
The relationship between each attribute and churn can be visualized through bar charts, as can be seen below.

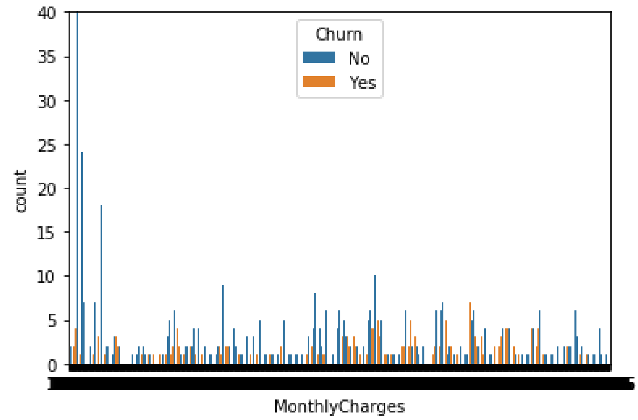


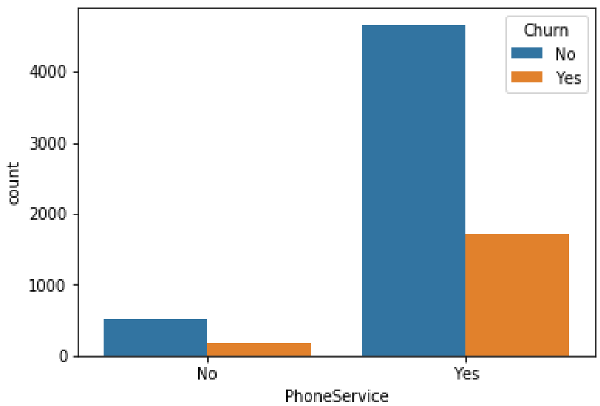


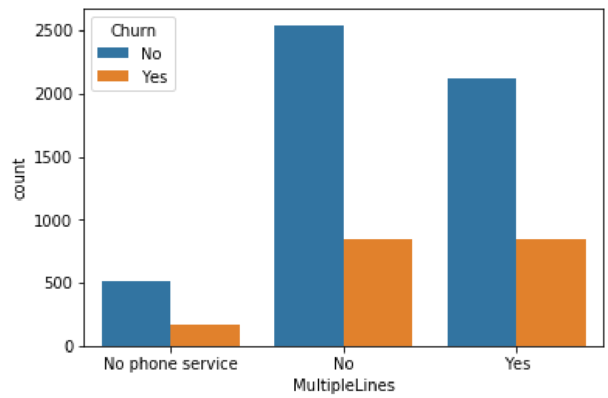


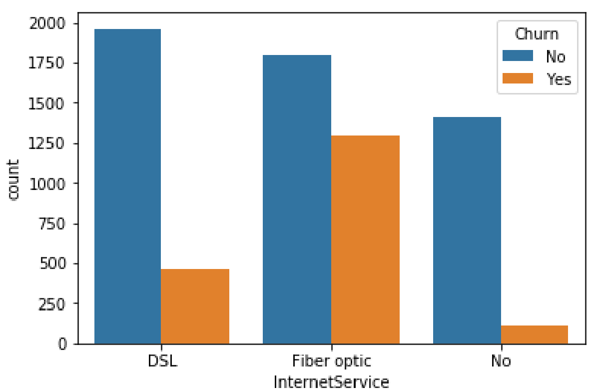


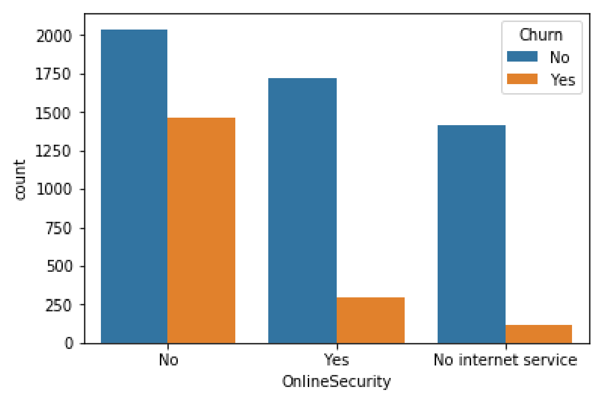

Churn
No
Yes
count
0
50
100
150
200
250
300
350
tenure

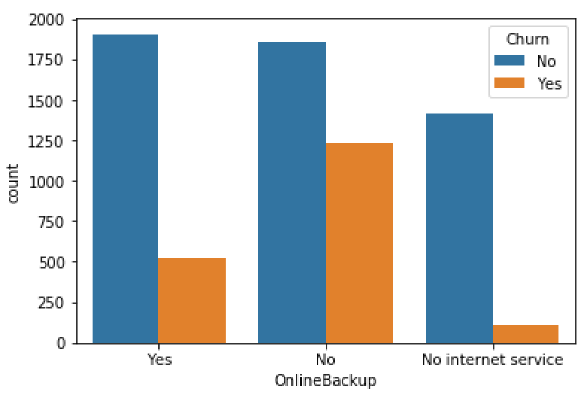

Churn
No
Yes
count
0
5
10
15
20
25
30
35
40
MonthlyCharges

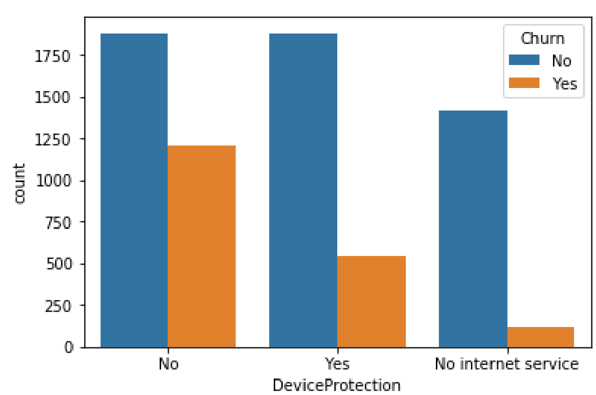

Churn
No
Yes
count
0
1000
2000
3000
4000
No
Yes
PhoneService


Churn
No
Yes
count
0
500
1000
1500
2000
2500
No phone service
No
Yes
MultipleLines


Churn
No
Yes
count
0
250
500
750
1000
1250
1500
1750
2000
DSL
Fiber optic
No
InternetService


Churn
No
Yes
count
0
250
500
750
1000
1250
1500
1750
2000
No
Yes
No internet service
OnlineSecurity


Churn
No
Yes
count
0
250
500
750
1000
1250
1500
1750
2000
Yes
No
No internet service
OnlineBackup


Churn
No
Yes
count
0
250
500
750
1000
1250
1500
1750
No
Yes
No internet service
DeviceProtection

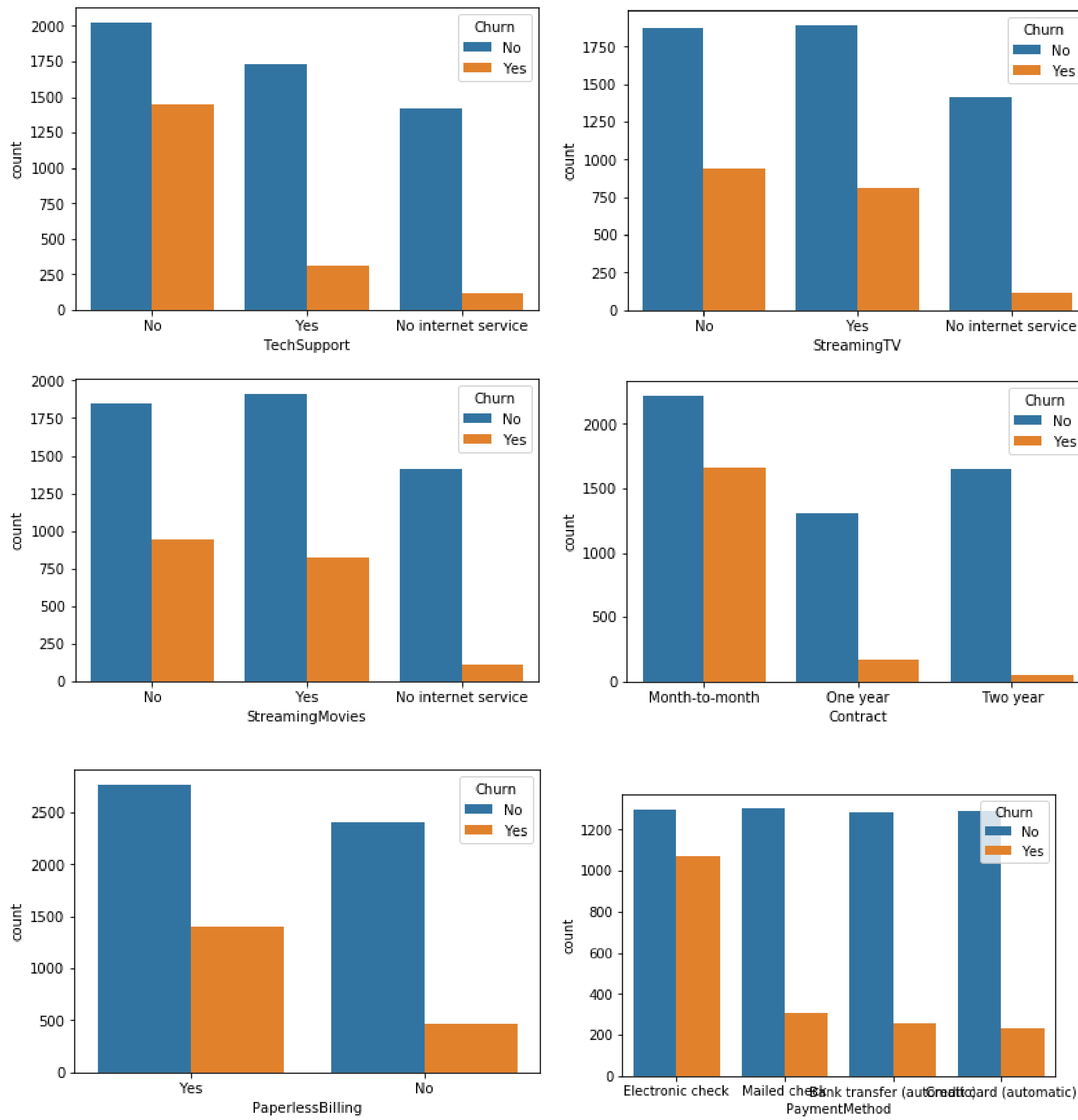


Figure 4: Attribute relationship with churn

Binning each of the attributes by their classes and highlighting the number of customers who churned or stayed provides some interesting insights:

- Customers who use additional services (tech support, device protection, online backup/security) are less likely to churn
- Customers on month-to-month contracts churn much more frequently than customers on two-year contracts
- Customers who pay through electronic checks churn much more frequently than customers who use other payment methods (mailed check, bank transfer, credit card)

- Longer tenured customers are more likely to stay with the ISP
- Customers who use DSL as opposed to fiber optics are more likely to stay with the ISP

4. Clustering

The next step of the analysis is to group all the data points into clusters. This will provide insight into the dataset, especially with regards to how difficult prediction may be later. A k-means cluster is used

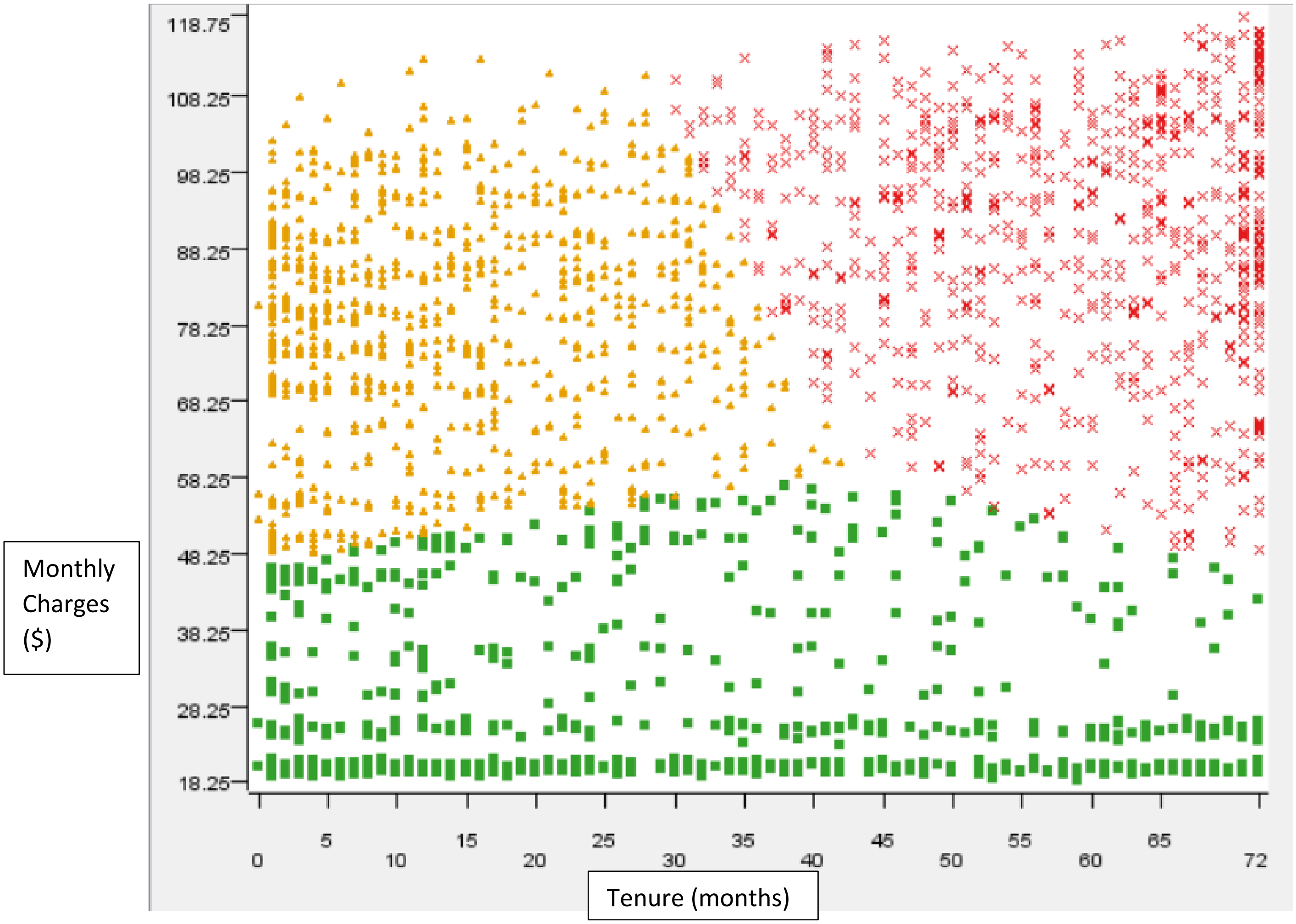


Figure 5: K-means cluster (k=3)

After selecting k clusters, the algorithm is run by iteratively moving the cluster centre to the mean of each cluster until k clusters appear. After clustering for several values of k, it was determined that k-=3 produced the clearest clusters, as shown above. There are three distinct clusters shown:

1. High-paying, short-term customers
2. High-paying long-term customers
3. Low-paying customers

The clean clustering indicates that prediction will be relatively simple for this problem. There is very little overlap amongst clusters, suggesting simple prediction. The three clusters will be useful in future analysis of customer segmentation.

5. Predictors

Two predictors will be used and optimized to identify the most accurate model that can predict customer churn.

1) Random Forests – this will be used first as it will provide guidance as to which attributes should be tossed to begin with. It can also return a list of ranked attribute importance, based on scores derived from the several trees being run. This will be key to answering why customers are leaving the ISP.
2) Neural Networks – this model will act as a check to the random forests and investigate whether there is a complicated relationship between attributes. The neural networks will indicate whether the attributes' predictive power is determined collectively rather than individually. As this is a relatively small dataset (7,034 records) it will not take too long to run.

5.1 Random Forest

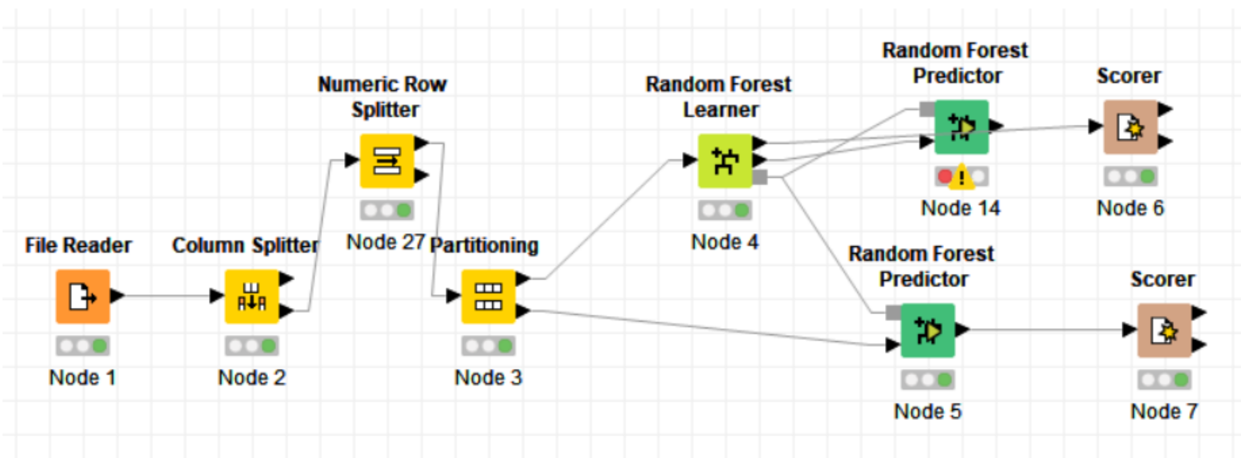


Figure 6: Random forest workflow in KNIME

5.1.1 Total Customer Analysis

A random forest model can be modified for the number of models, tree depth, node size and scoring method. These variables are optimized below, resulting in an accuracy of 79.4%:

Table 2: Random forst parameters

Variable	Value
Number of Models	500
Tree Depth	1,000
Minimum Node Size	1
Scoring Method	Gini Index

These were selected after several iterations for each variable. The value was increased until there was no noticeable improvement to the prediction accuracy. It would be possible to have 100,000 models and a tree depth that contains 20,000 levels but that would be a wasteful if it did not improve prediction accuracy.

Table 3: Random forest attribute statistics

Row ID	#splits (level 0)	#splits (level 1)	#splits (level 2)	#candidates (level 0)	#candidates (level 1)	#candidates (level 2)
gender	0	6	2	25	48	62
SeniorCitizen	0	4	2	23	44	52
Partner	0	1	5	22	34	64
Dependents	1	2	11	19	35	67
tenure	11	25	20	19	56	59
PhoneService	0	5	6	17	47	54
MultipleLines	1	8	10	22	46	83
InternetService	6	6	11	15	36	80
OnlineSecurity	6	8	7	15	42	60
OnlineBackup	6	8	3	25	50	55
DeviceProtect...	7	6	3	24	37	61
TechSupport	5	7	5	22	55	58
StreamingTV	6	8	12	20	37	75
StreamingMo...	4	0	11	16	40	61
Contract	2	8	10	12	35	51
PaperlessBilling	4	8	14	19	47	63
PaymentMethod	12	10	23	31	35	80
MonthlyCharges	17	19	27	28	40	67
TotalCharges	12	13	36	26	36	64

The table above shows attribute statistics – the #splits column shows how many times the attribute was selected as the splitting attribute and the #candidates column shows how many times it was a candidate to be the splitting attribute. To account for varying candidacy, the attributes will be ranked on their selection ratio, taken as $\frac{\#splits\ (level\ 0{+}1{+}2)}{\#candidates\ (level\ 0{+}1{+}2)}$

The ranked importance looks like:

Table 4: Ranked attribute importance

Attribute	Selection Ratio (%)
tenure	61%
Contract	45%
Monthly Charges	33%
Tech Support	31%
Device Protection	31%
Online Security	30%
Payment Method	30%
Online Backup	23%
Internet Service	18%
Multiple Lines	15%
Senior Citizen	15%
Streaming Movies	13%
Streaming TV	10%

This aligns with the findings made in plotting the relationship between each attribute in relation to churn, where it appeared that contract type, tenure, payment method and additional services were indicative attributes. The remaining attributes (gender, senior citizenship, having a partner, dependents or multiple lines) all score 5% or less.

Increasing the number of models from 100 to 1000 incrementally improves prediction accuracy from 79.1% to 79.2%. Using information gain rather than information gain ratio or gini index increases accuracy further to 79.4%.

Tenure, *Contract* and *MonthlyCharges* are the three most predictive attributes according to the table above. Each of the three will be analyzed further.

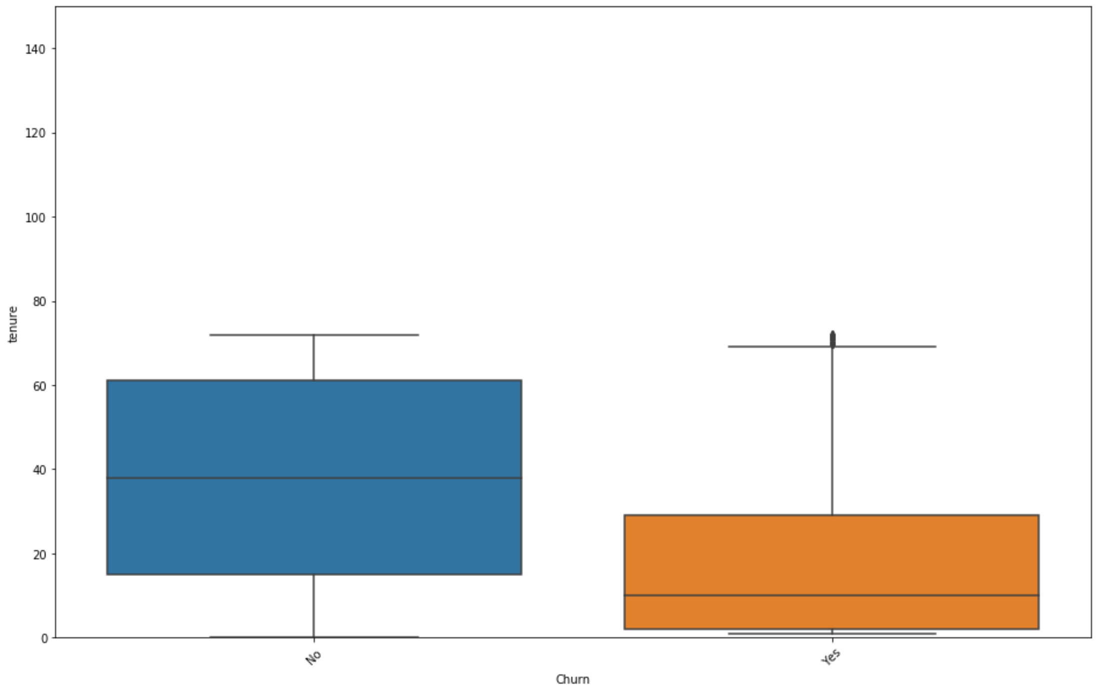


Figure 7: Churn vs. tenure box plot

The box plot above shows the difference in tenure between customers that churn and don't churn. As can be seen, customers that churn spend much less time with their ISP – 18 months on average, compared to 38 months for customers who remained with the ISP.

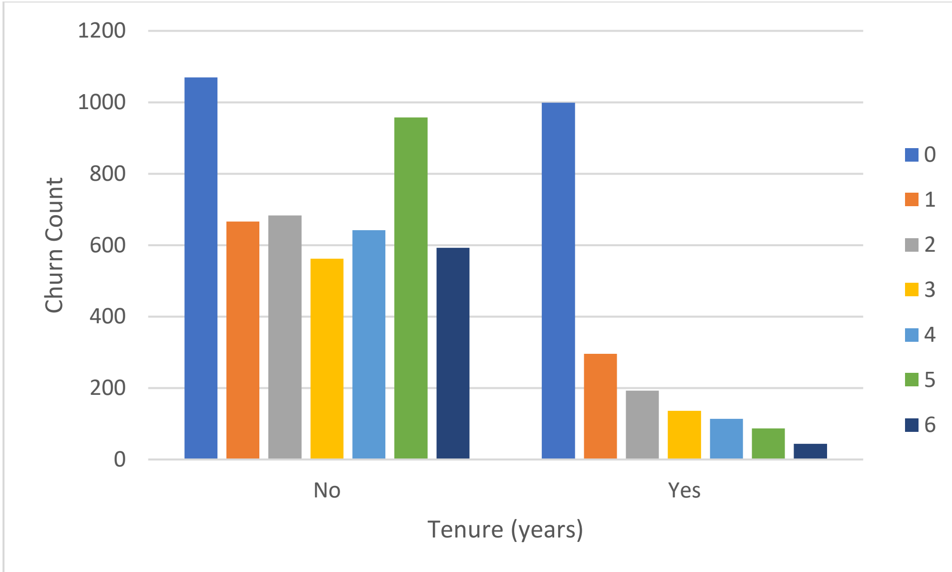


Figure 8: Tenure vs. churn pivot chart

The chart above further highlights the relationship between tenure and churn. Customers are most likely to churn in their first year, with the churn rate dropping with each additional year spent with the ISP.

Table 5: Tenure churn rate

Tenure (years)	Churn Rate
0	48%
1	31%
2	22%
3	19%
4	15%
5	8%
6	7%

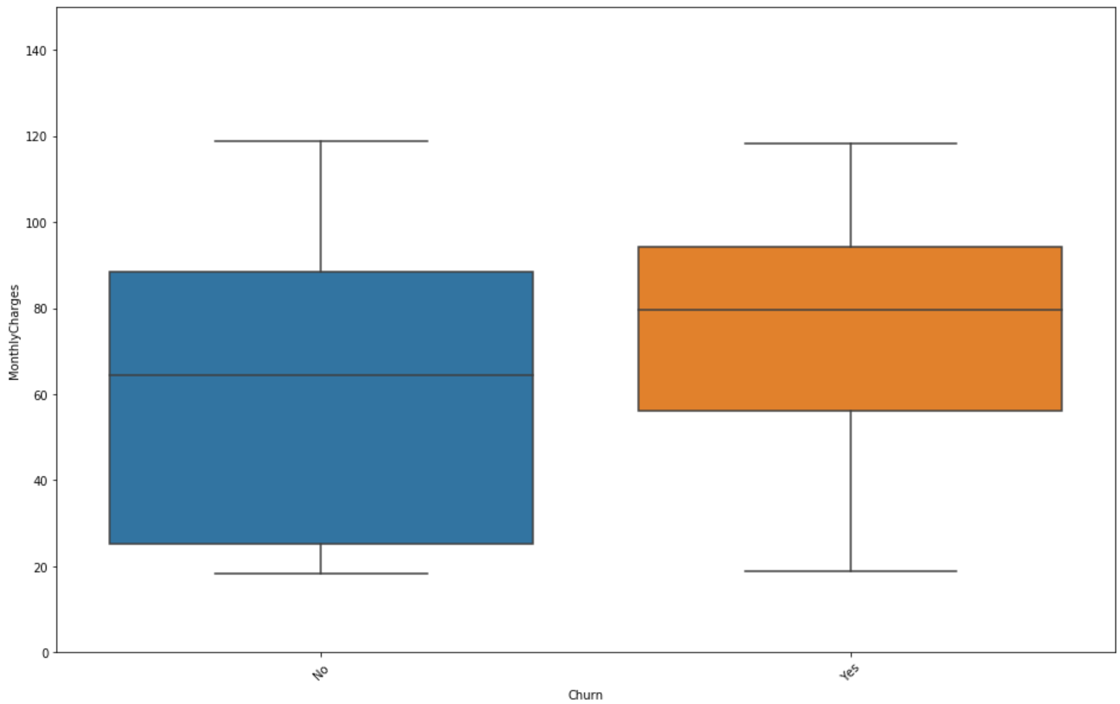


Figure 9: Churn vs. monthly charges box plot

It appears customers who churn pay more for their ISP ($74/mo. Vs. $61/mo.). The trend in churn is less noticeable for monthly charges than it is for tenure.

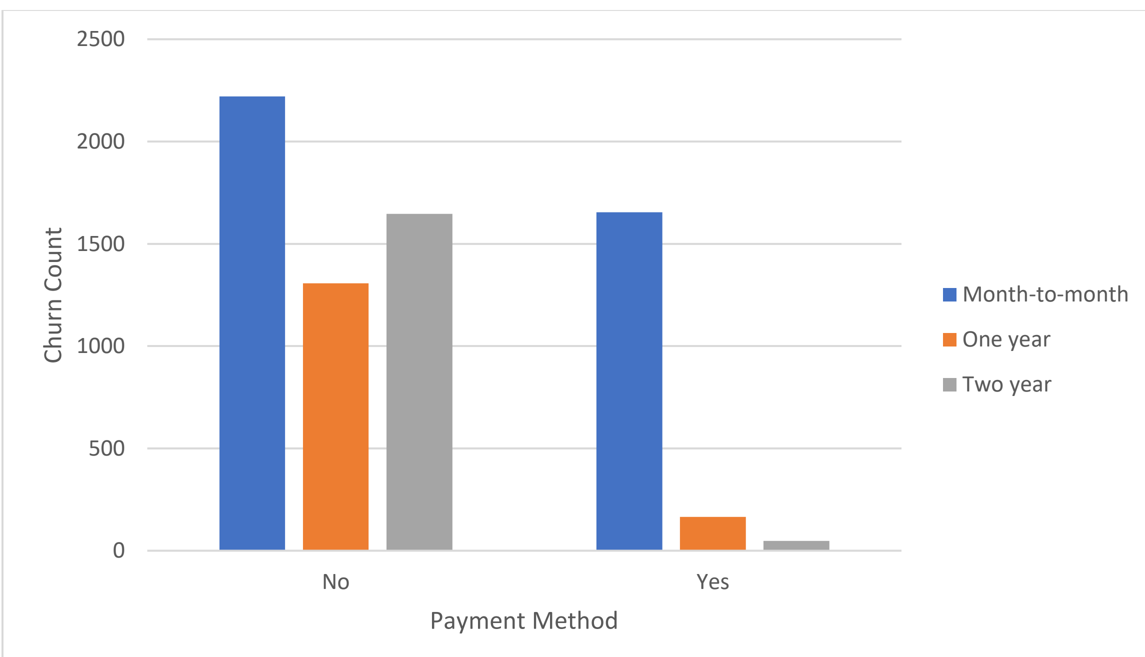


Figure 10: Payment method vs. churn box plot

As was seen earlier, customers are much more likely to churn on a month-to-month contract rather than a one-year or two-year contract.

Table 6: Contract type churn rate

Contract Type	Churn Rate
Month-to-month	42.7
One-year	11.3
Two-year	2.8

Based on the findings from the random forest attribute scores and the analysis that followed, it can be said that the customers most likely to leave have:

- Short tenure
- Pay high monthly charges
- Pay on month-to-month contracts

Amongst the three indicators of churn, tenure with the ISP is the most indicative, as the random forest and churn rates show. The analysis will be repeated for customers who have spent more than a year with the ISP.

5.1.2 2 Month + Customer Analysis

The model improves to 83.9% accuracy when the data is filtered to customers who have spent more than a year with the ISP. This is to be expected as these customers will share more in common, making them easier to predict.

Table 7: Ranked attribute importance for customers with ISP for more than one year

Attribute	Selection Ratio %
Payment Method	56%
Monthly Charges	48%
Contract	36%
tenure	34%
Online Security	31%
Internet Service	31%
Paperless Billing	21%
Tech Support	20%
Online Backup	15%
Multiple Lines	14%
Streaming TV	12%

According to the random forest attribute statistics, the three most important attributes for customers who have been with the ISP for more than a year are:

- Payment method
- Monthly Charges
- Contract

Payment Method

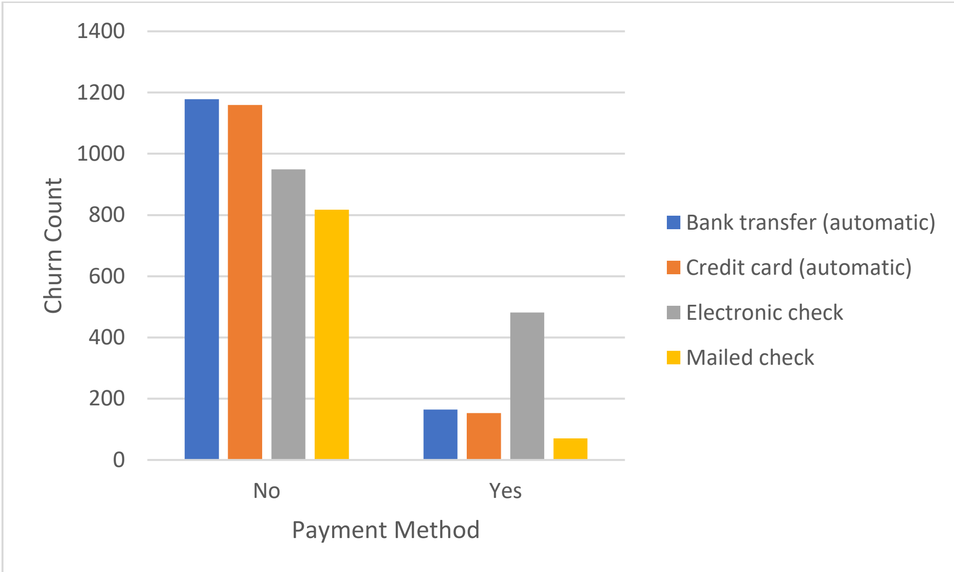
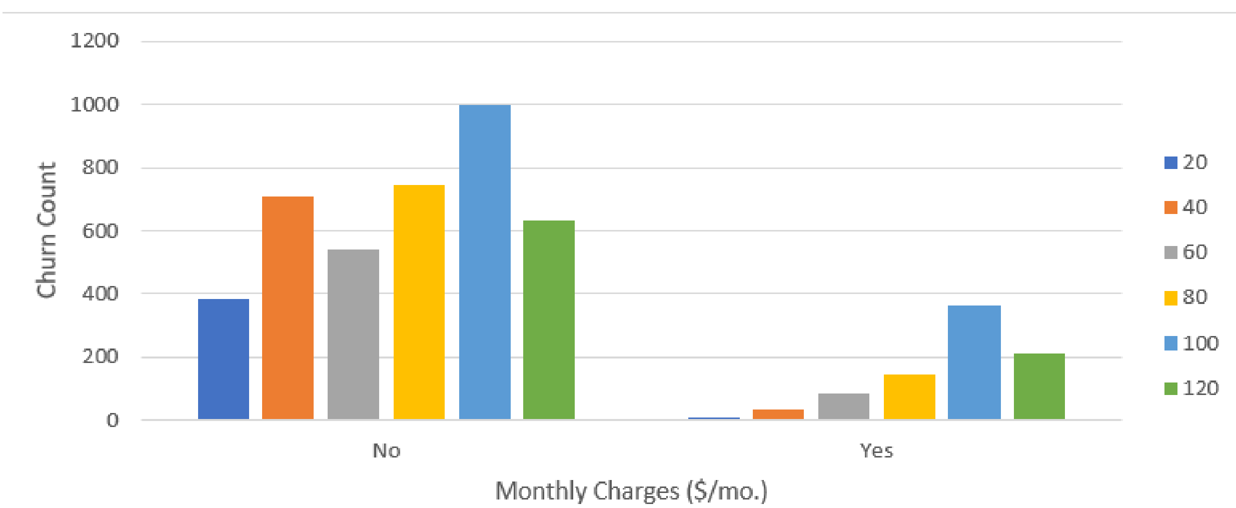


Figure 11: Payment method vs. churn count pivot chart

Most customers who churned after one year with the ISP used electronic checks to pay for the service. As the table below shows, customers who pay through electronic churn at a higher rate than other forms of payment.

Monthly Charges ($/mo.)	Bank transfer (automatic)	Credit card (automatic)	Electronic check	Mailed check
Churn Rate	12%	12%	34%	8%

Monthly Charges



Monthly charge data was binned into segments of $20/month. The group labelled "$20" pay $20 or less per month. The noticeable trend is that customers who pay more monthly charges churn at a higher rate. The table below shows the rate at which each binned group churns at.

Table 8: Monthly Charges churn rate

Monthly Charges	$20	$40	$60	$80	$100	$120
Churn Rate	3%	5%	14%	16%	27%	25%

The data can be split into three clear tiers.

- Tier 1: $40/month and below. 3-5% churn rate
- Tier 2: $40-$80/month. 14-16% churn rate
- Tier 3: $80-$120/month. 25-27% churn rate

Contract

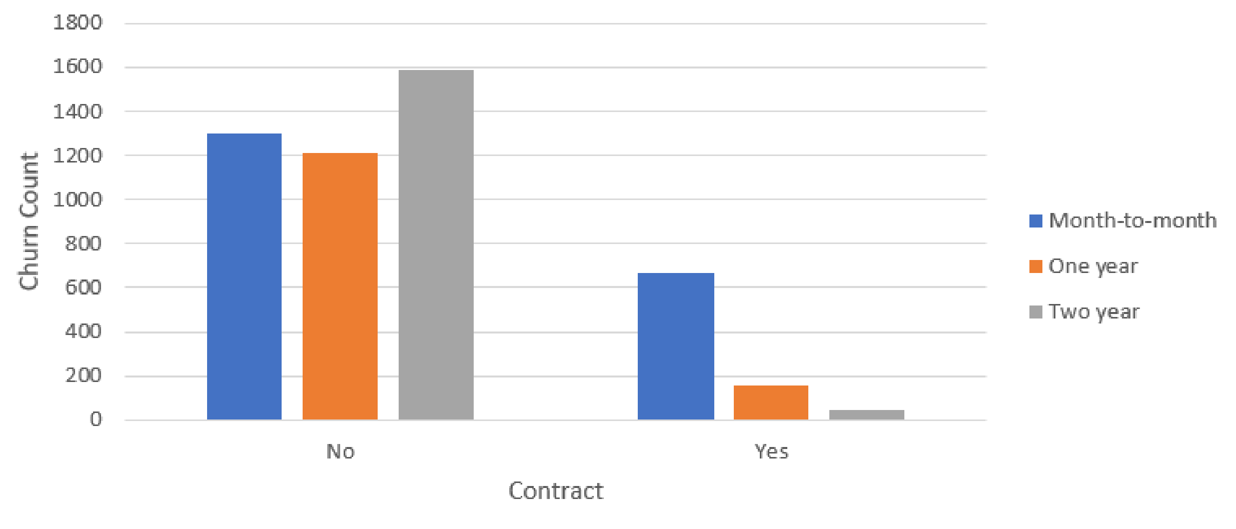


Figure 12: Contract vs. churn rate box plot

Table 9: Contract type churn rate

Contract Type	Month-to-Month	One year	Two year
Churn Rate	34%	12%	3%

Contract data for customers who leave after at least one year with the ISP closely matches contract data for total customers: churn rate is greatest for customers who use a month-to-month contract.

Internet Service

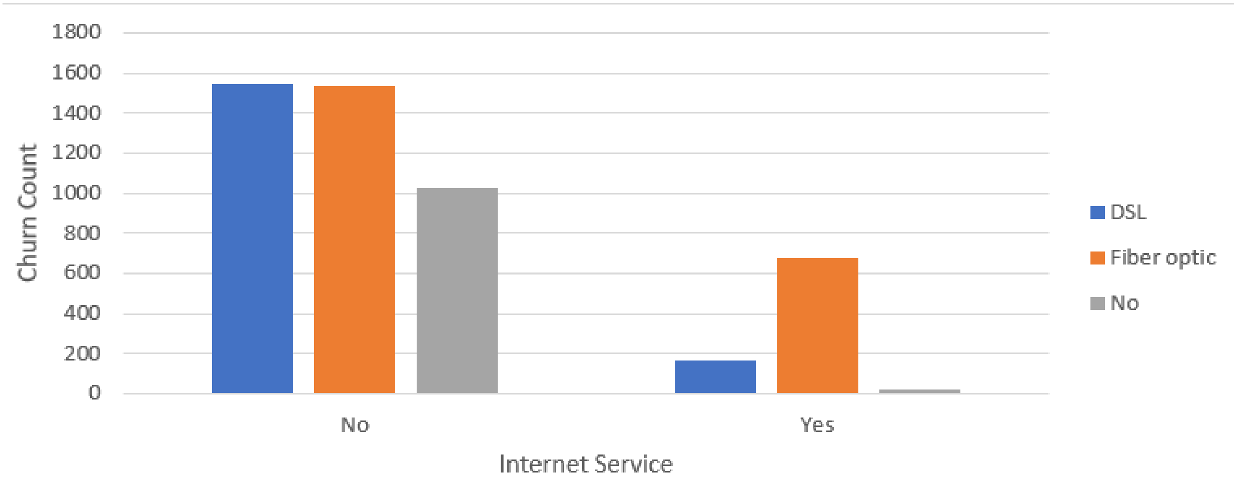


Figure 13 Internet service vs. churn pivot chart

Customers who use fiber optics churn more frequently than customers who use DSL.

Additional Services

Three additional services are some of the most predictive attributes: Online security, online backup and tech support.

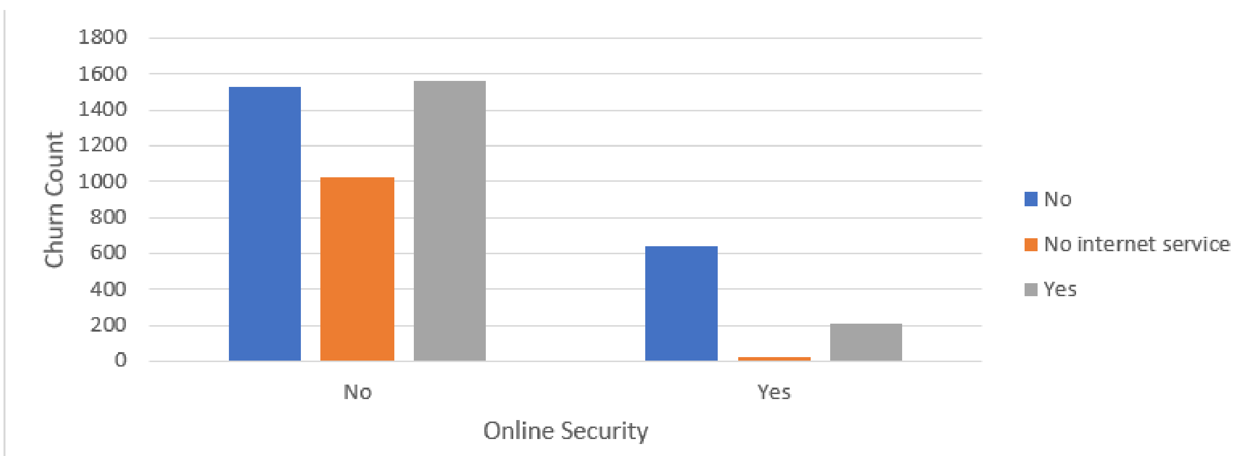


Figure 14: Online security vs. churn - pivot chart

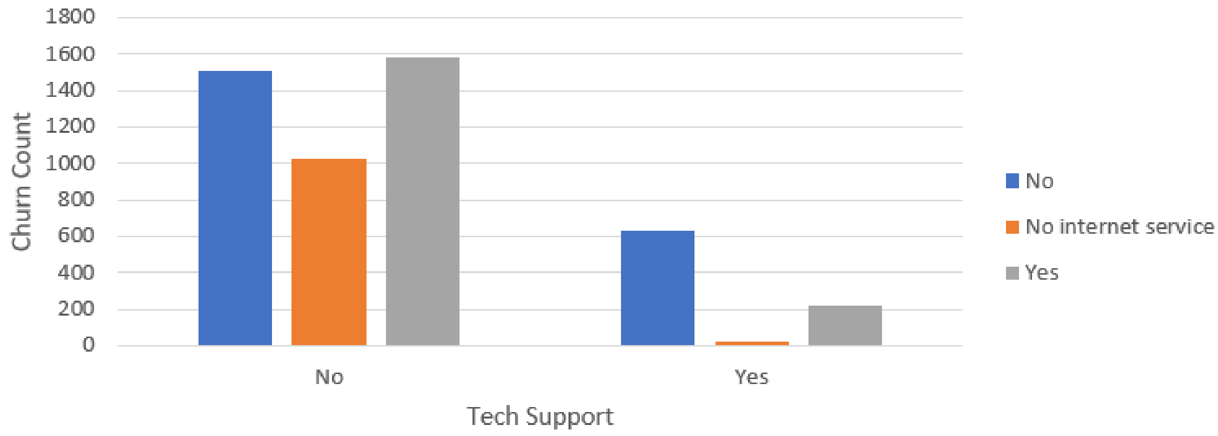


Figure 15: Tech support vs. churn - pivot chart

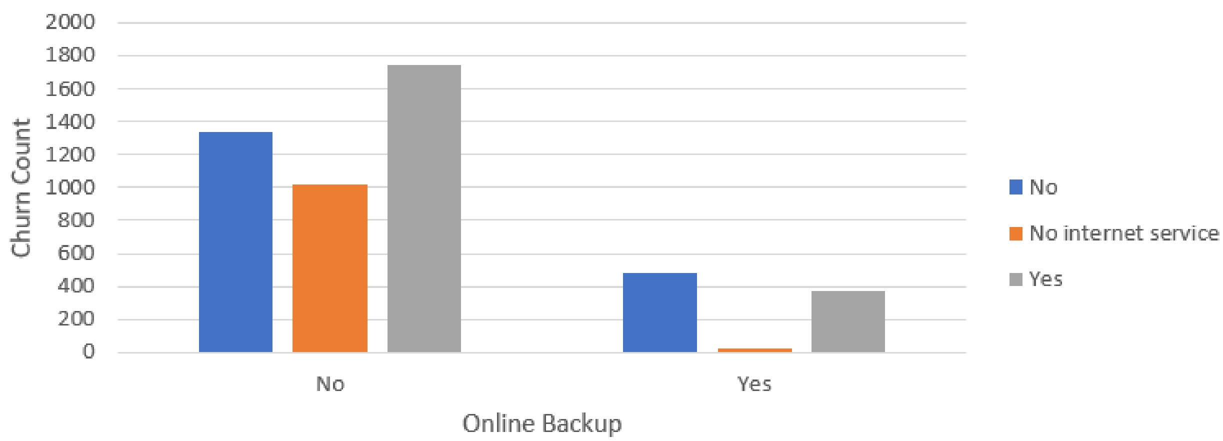


Figure 16: Online backup vs. churn - pivot chart

Two observations can be made:

- Customers who have "no internet service" rarely churn
- Customers who use additional services are less likely to churn

5.2 Neural Networks:

Whereas the random forests provide an indication as to which attributes are of greatest importance, neural networks can indicate whether there is a complicated relationship between attributes. A neural network will be a better predictor if several attributes have collective, but not individual predictive power.

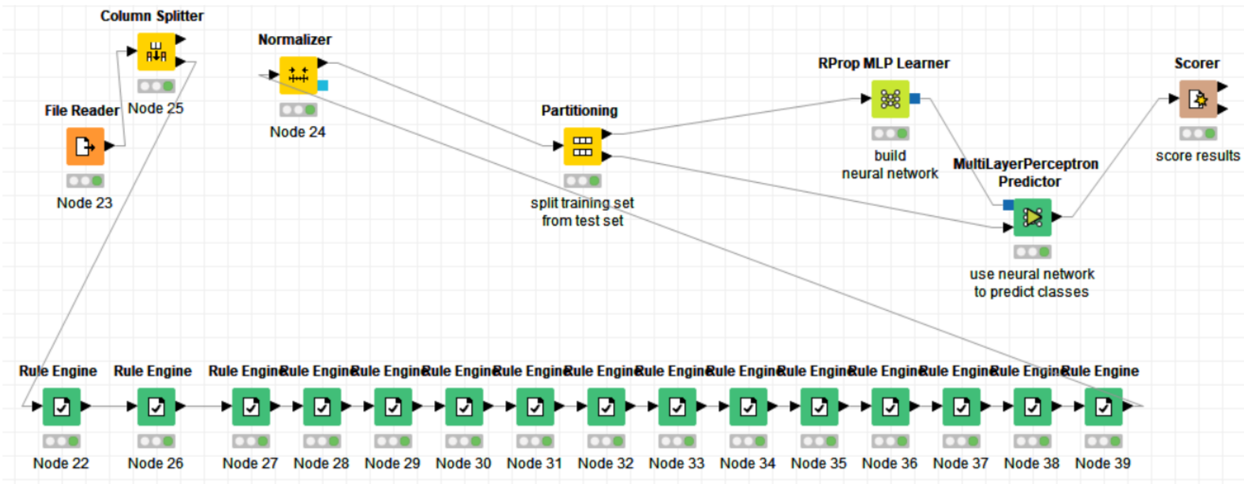


Figure 17: Neural network workflow in KNIME

The neural network requires numeric inputs. The rule edges allow the user to assign a number to each categorical variable. For example:

- "yes" = 1 and "no" = 0.
- "Month-to-month" = 0, "One year" = 1, "Two year" = 2.

The rule edges are followed by a normalizer. The neural network is optimized to result in the highest possible accuracy, as shown in the table below.

Table 10: Neural network parameters

Variable	Value
Maximum number of iterations	50
Number of hidden layers	1
Number of hidden neurons per layer	10
Use seed for random initialization	Yes

This results in a prediction accuracy of 80.34%. This is slightly greater than that of the random forest. The high prediction accuracy suggests that:

- Churn rate is a non-linear function of inputs
- Many attributes in the dataset have small predictive power that are very telling collectively
- A complicated relationship exists between attributes

These findings will be further considered in the following analysis, where slightly predictive attributes that haven't been explored will be analyzed further. Multiple attributes will be analyzed because the neural network suggests that many of the attributes have predictive power.

6. Further Analysis

- Phone users
- Internet service costs

- Customer profiles.
  - Average churn rates for each customer profile (phone user, internet user, additional services). Some kind of table.
- Look at multiple lines and dependents. Do they churn?
  -

Two more trends will be analyzed further. They are:

- Add-on internet service costs
- Phone users

6.1 Internet Service Costs

The correlation matrix at the start of the analysis showed that customers with additional internet services had higher monthly charges. This applied to device protection, online backup, online security, tech support and streaming movies/TV. Monthly charges were also shown to be correlated with churn rate. The chart below visualizes churn count for customers paying at different monthly rates – with and without device protection. Customers with no internet service have been filtered out of the chart.

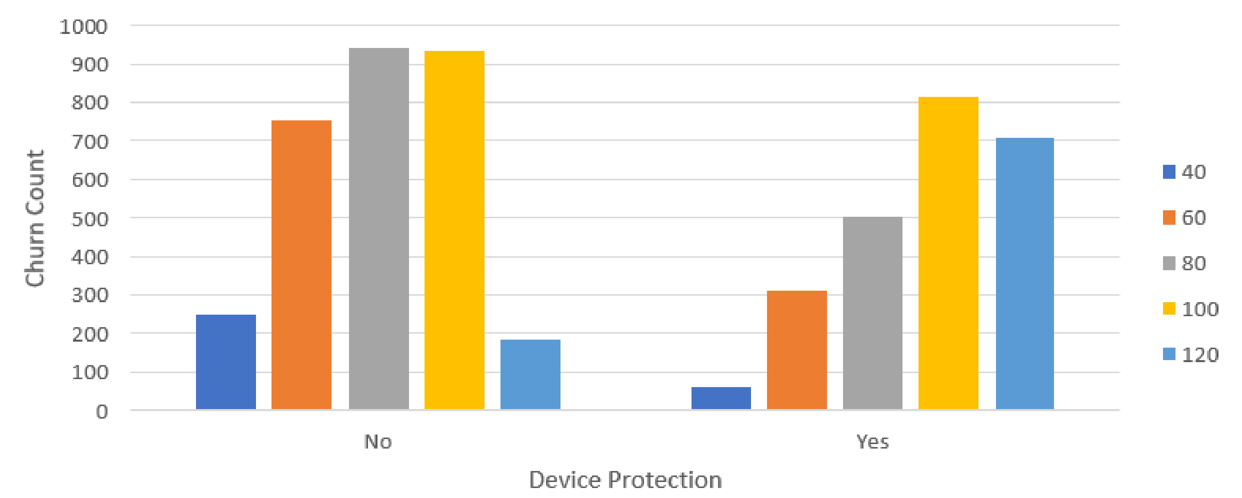


Figure 18: Device protection vs. churn count/monthly charges - pivot chart

The chart shows that churn rate increases greatly with higher costs. The table below shows how significant this trend is.

Table 11: Monthly charegs churn rate

Monthly Charges ($/mo.)	$40	$60	$80	$100	$120
Churn Rate	20%	29%	35%	47%	79%

Nearly 80% of customers who pay between $100 and $120 a month are churning. My hypothesis is that there is stiff price competition for competing ISPs, leading customers to churn at high rates. This holds true for customers who have been with the ISP for less and more than one year.

6.2 Phone Users

In comparing churn rate to additional internet services, an interesting trend appeared. It was seen that customers who had “no internet service” were the least likely to churn. After analyzing the spreadsheet, it appears that customers with “no internet service” have phone service instead. It can be said that customers who have a phone service with the ISP rarely churn.

7. Further Recommendations:

In conclusion, there are four main reasons why customers are leaving the ISP:

1. **Onboarding.** Customers are finding alternatives to the ISP within their first year, and within 18 months on average.
2. **Contract Type.** Month-to-month contracts allow customers the chance to leave at a much more frequent basis.
3. **Prices.** Customers are being priced out and not seeing the value in additional internet services or the ISP at higher prices.
4. **Payment Method.** Customers who spend more than a one year with the ISP churn primarily based on their form of payment – those who pay via electronic check churn at 3 times the rate of any other form of payment. It should be noted the customers who spend more than a year with ISP also churn due to the aforementioned causes of high prices and month-to-month contracts.

The ISP can use this analysis to make some changes aimed at reducing churn rates. Based on the identified churn causes, it is recommended that they **offer their customers a two-year contract with a discounted first-year price**. This tackles three causes of churn at once: first-year churn, month-to-month contracts and high prices. A contract of this nature will help keep customers with the ISP at the time when they are most likely to leave, help transition customers away from month-to-month contracts and offer their customers a price point that will help retain them.

Another recommendation is for the ISP to **offer bundled additional internet services on 2-year contracts**. The data shows that customers are leaving the ISP when they pay high rates for additional services. The ISP can raise the value of these services by grouping them together, like packaging online backup and security with long-term deals. This is critical for the ISP as they must be able to retain their highest paying customers – those who utilize their additional services. The ISP can hit two birds with one stone by retaining their most profitable customers while signing them on to long-term contracts by bundling these additional services together.

Alternatively, the ISP can simply **reduce prices on their additional internet services.** To prevent 79% of customers who pay over $100/month from leaving the ISP, they can either improve their additional internet services or reduce the cost. This will assist in fending off the probable price competition from competitors who take their most profitable customers.

It is in the ISP’s interest to **encourage automatic payments.** Transitioning payments from electronic checks to automatic payments will help reduce the churn rate, as electronic checks are the greatest indicator of churn for customers who have been with the ISP for over a year. This transition can be facilitated by simplifying and promoting automatic payments via bank transfer and credit card.

The final recommendation is to **utilize segmented and targeted marketing to drive changes in customer behaviour-based findings of the analysis**. The clustering analysis showed that there are three primary types of customers. The table below presents how they should be targeted based on the findings of the analysis:

Table 12: Target marketing strategy

Customer Type		Targeted Marketing Strategy	Objective
Monthly Charges	Duration with ISP		
Low	-	Promote new deals on bundled additional internet services	Move this segment to become more profitable customers
High	Low	Promote discounted rate on first year of 2-year contracts	Prevent customers from churning in their 1-2 years
High	High	Inform customers of newer, simplified automatic payments	This segment appears to value convenience, as their biggest cause of churn is slower methods of payment. Simplify their experience.