The Strategies to Increase Revenue for a Telecommunication Firm

Section 1 Group A

Team member:

Shijun Wei

Zhenghao Zhou

Zhe Wang

Zhitao Zheng

Yumeng Zou

Table of Contents

Executive Summary

This report explored three ways to increase revenue for a telecommunication firm to help it grow its competency in the competitive telecommunication industry. We gathered a Kaggle dataset containing information on a Telco firm's customer churn. Regression analysis and random forests were deployed to study the influence elements of tenure, monthly charges and churn rate. Customer lifetime value was also calculated as a guide to differentiate marketing strategies. To increase revenue, the firm would need to increase its customers' tenure, increase their monthly pay, improve the churn rate and target those with higher CLV. We then gave managerial recommendations based on the findings.

To achieve these, the firm can first provide incentives for choosing longer contract terms. Secondly, if the firm wants to increase its short-term revenue, then it can spend more effort in incentivizing customers to sign up for the additional services. However, if the firm aims to increase long-term revenue by increasing tenure or decreasing churn rate, it can only pay attention to those services that mattered. Another way to decrease churn rate is to send out target retain offers to those with high probability of churn in order to utilize resources more efficiently. Lastly, by comparing CLVs of different groups, we could target marketing those who would bring higher CLVs to the firm to leverage resources to maximize profit. We also encourage the firm to collect more information about its customers so that we can run additional analysis to provide more detailed insights.

Introduction

The continuing advance of mobile technology and rising demand for faster data transmission keep driving innovation and competition within the telecommunication industry. The customer churn rate of the telecommunications industry in the United States in 2018 is

21%[1]. In this report, we will help a telecom firm to enhance its revenue so that it can reinvest in itself to improve competence and market share.

This telecom firm provides services primarily through phone, internet, streaming and other add-on services. Same as other telecom companies, this firm makes money through subscription plans. Of all the customer markets, the residential market, which this firm focuses on, is arguably the toughest. Compared to corporate, residential customers are less stable and more price-sensitive. To increase revenue, we broke down the problem based on its business model into increasing tenure, monthly charges, and churn prediction.

Problem Formulation

In order to increase the revenue of the company, we divided our work in three directions. The scheme of our presented work is illustrated in Appendix 1. Firstly, improving tenure is the priority. Traditionally, to retain customers, telecom companies have implemented strategies to improve their customer experience such as creating frictionless experience and sending out surveys to gather feedback. In this report, we aimed to help the firm identify specific areas to prolong tenure using regression analysis. Secondly, we also wanted to analyze what decided monthly charges with regression analysis, since the variable is related to the value of a customer. We hope to delve into the model and locate variables that could affect monthly charges and ultimately increase revenues.

Thirdly, we predicted the churn rates. Customer churn continues to present a significant and costly challenge to the telecom industry. This firm had a higher churn rate of 26.54% than the industry average. It costs on average $315 to acquire a new customer in most Telecom industries.[2] When a customer leaves, the firm loses not only the future revenue from this

[1] https://www.statista.com/statistics/816735/customer-churn-rate-by-industry-us/
[2] https://www.performancemagazine.org/june-smartkpi-telecom-sub-acquisition/

customer, but also the resources it spent to acquire the customer in the first place. Since churn is a lagging indicator, the firm needs to be proactive and implement tactics to prevent customers from churning. Therefore, predicting churn rate is crucial for better allocation of resources. Lastly, Customer Lifetime Value (CLV), a prediction of the net profit attributed to the entire future interaction with a customer[3], for each customer were also calculated in this report. CLV allows businesses to quantify the monetary worth of each user on their platform, thus helping it determine optimal capital outlays per user.

Data Characteristics

The dataset contains information on a Telco company's customer churn, and the original data was collected by IBM[4]. The dataset is accessible from Kaggle[5]. The raw data contains 7043 rows (customers) and 20 columns (features). Each row represents a customer's information, each column contains customer's attributes which are described in Appendix 2.1. The first step of our analysis began with exploratory data analysis (EDA), which provided some basic understanding of the data and established some hypotheses for development of the models.

Numerical Features

For all the 20 variables, besides *Churn*, there are 3 numerical features: *Tenure*, *MonthlyCharges* and *TotalCharges*. The tenure span of all the customers in the data set is quite varied from 0 to 72 month with an average of 32 month. The range of monthly charges of customers is from $19 to $119 with an average of $65. More facts of numerical features are listed in the table in Appendix 2.2. Further, since the total charges of a customer is based on the monthly charges and his/her tenure month, only monthly charges and tenure were analyzed.

[3] https://www.qualtrics.com/experience-management/customer/customer-lifetime-value/
[4] https://developer.ibm.com/technologies/data-science/patterns/predict-customer-churn-using-watson-studio-and-jupyter-notebooks/#
[5] https://www.kaggle.com/blastchar/telco-customer-churn

According to Appendix 2.3, new clients are more likely to churn. Therefore, relationships between the variables and tenure would be further analyzed.

Categorical Features

The rest of the variables are 16 categorical features with 6 binary features (Yes/No) : *Gender, Partner, SeniorCitizen, Dependents, PhoneService, PaperlessBilling*; 9 features with three unique values each (categories): *MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract* and 1 feature with four unique values: *PaymentMethod*.

Analysis of the relationship between 6 binary features and the churn rate are taken into account. From Appendix 2.4, the customers who are non-senior citizens, without dependents, without partners, with phone services, with *PaperlessBilling* are more likely to churn. But there is no difference between gender in terms of churn rate.

For 9 categorical features with three unique values, three features are more important: *MultipleLines*, *InternetService* and *Contract*. As shown in Appendix 2.5, customers with no phone service have a lower churn rate. Customers using fiber optics have a higher churn rate. And customers with month-to-month contracts have a higher churn rate. The rest are 6 additional services if the customers use the internet service: *OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies*. The number of total customers and churn customers are shown in Appendix 2.6. The service of *OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport* successfully lowered the number of customers who churn while the *StreamingTV, StreamingMovies* has little help to lower the churn rate. From Appendix 2.7, for the payment method feature with four unique values, the customers whose payment method is electronic check are more likely to churn as well.

Through the EDA analysis, the relationship between features as well as churn rate are demonstrated in the heatmap in Appendix 2.8. With the initial findings, three models were built to provide more sophisticated insights for managerial recommendations.

Model Development

Model 1: Linear Regression model on *Tenure*

First, we wanted to see which attributes had significant effects on the tenure of the customers, and how the company can leverage the results to keep their customers longer. The first model considered all demographics and payment-related attributes, and ignored the attributes related to phone and internet services. Then, customers are segmented into three groups to examine different behaviors between different service groups. The first group is customers who were only using the phone services. The second group is customers who were only using the internet services, and the last group is customers who used both phone and internet services. For each group, attributes on customer demographics, payment, and relevant services are included in the model. The models can be found in Appendix 3.1.

Model 2: Linear Regression model on *MonthlyCharges*

This model was built to explore the relationship between different attributes and monthly charges, since the more money a customer is willing to pay, the more revenue for the company. By analyzing the coefficients, we were able to define contribution levels for those attributes to monthly charges. The total charge variable was excluded from the model as it has a linear relationship with tenure and monthly charge. Then null values were also eliminated from the monthly charge column and dummy variables were created to run the model. The linear model ran independently for three groups: customers who only used phone service, customers who only used internet service, and customers who used both of them. Accuracy tests of the three models showed a minimum value of 93.3%. Detailed models can be found in Appendix 3.2.

Model 3: Random Forest model for Churn Prediction

To predict customers' churn behavior based on their given information in the dataset, we used the Random Forest Classifier technique. We started with data cleaning that dropped 11 NA values in the *TotalCharges* column. Categorical variables were turned into dummy variables and labeled with numbers. We choose variables based on weak correlations with the churn column (Appendix 2.8) and high feature importance in the model (Appendix 3.3.1). *SeniorCitizen, InternetService, PaymentMethod, Tenure, MonthlyCharges* were selected in this model. The dataset was divided into one third of test data and two third of train data. Then, the results in Appendix 3.3.2 shows that the AUC score is 0.82, which means 82% of our churn prediction is correct. Detailed model performance and selection are in Appendix 3.3.

Extension of Model 3 - CLV

Based on the churn rate obtained from Model 3 and monthly revenue from each customer, Customer Lifetime Value (CLV) was also studied. CLV calculation incorporated the discount rate of future revenue, length of a customer's lifetime, gross margin, and customer's retention rate.

The formula used to calculate CLV is

$$\text{CLV} = \sum_{t=1}^{T} \frac{mr^{t-1}}{(1+i)^{t-1}}$$

Assumptions: Gross margin(including initial customer acquisition cost) = 72%[6]

Discount rate = 8% annually

Average CLV were compared within *SeniorCitizen* and *PaymentMethod* categories.

[6] https://csimarket.com/Industry/industry_Profitability_Ratios.php?ind=905

Results and Limitations

For the linear regression models on tenure, the results showed that gender did not have significant impacts on tenure in any customer segment. There were also some common trends in all three customer groups. Customers who had a partner tend to have longer tenure than those who didn't; customers who did not have dependents tend to have slightly longer tenure than those who had dependents; customers who paid with automatic payment methods tend to have longer tenure than those who paid with checks. The longer the contract period, the longer the tenure, thus customers with two-year contracts have the longest tenure on average. For customers who were only using the phone service, monthly charges did not have a significant impact on their tenure, but on average, those who had multiple phone lines stayed ten month longer than those who didn't. For customers who were only using the internet service, only the online backup service in the additional services and a significant impact on tenure. On average, if a customer did not use the online backup service, the tenure would decrease four months. Lastly, for customers who were using both phone and internet services, additional services of multiple lines, online security service, online backup service, device protection service, and technical support services all had significant impacts on their tenures. Customers who used these additional services stayed with the company longer than those who didn't. In addition, there are some limitations for the models, since the adjusted R-squared were around 0.60. There is still room for improvement in the models to bring more accurate results.

From the regression model with monthly charges, demographic attributes did not have significant impacts on the customer's monthly charge in any customer group, thus only adding additional services were relevant to increase monthly charges. When customers have only internet service, any attribute related to online services can increase their monthly charges, such as online security service and streaming services. Having a one year contract can also slightly increase monthly charge. In the second group when customers have only phone service, adding multiple lines service would be the single attribute to increase monthly charges. Lastly, when customers have both internet and phone services, Fiber optic internet service has the highest weightage followed by streaming TV and Streaming Movies. When we integrate our findings from all groups, customers with certain demographics did not impact monthly charges either, but only the additional services made a difference.

From the Random Forest model, a SHAP Waterfall plot report in Appendix 3.3.3, summarizing the important features that affect Random Forest classifiers' performance, is conducted and indicates that *Tenure, InternetService, MonthlyCharges* and *PaymentMethod* were the most crucial features for our churn prediction model. In addition, we made SHAP Dependence Plots for these variables to identify churned customers' characteristics by observing positive SHAP values. We found that most churned customers only stayed 15 months with the company; use Fiber Optic Internet service; have monthly charges higher than 70 dollars and their payment methods are automatic bank transfer and electronic check. Since *Tenure* and *InternetService* are the most important features from the SHAP Waterfall plot, we suggest companies to focus on customers who stay less than 15 months and own Fiber Optic Service to decrease churn rate.

From the Extension of Model3 results (Appendix 3.4), the average CLV of $542 for non-seniors was higher than that of $91 for senior citizens. The average CLVs of customers who were using automatic bank transfer, automatic credit card, electronic check and mailed check were $834, $542, $134, and $169 respectively. Noticeable difference exists between automatic payment methods and check payment methods. The limitation is that retention rate and margin were assumed to be constant overtime in this model, but the rates may change in real-world situations due to competitors' campaigns or internal operation adjustment.

Recommendation and Managerial Implications

To increase revenue, the company would need to increase its customers' tenure, increase their monthly pay, improve the churn rate and target those with higher CLV. Another reason to improve tenure is because the longer a customer stays with a firm, the less likely he/she will churn due to switching costs. In general, the firm can provide incentives to let the customer sign longer contracts or incentives for month-to-month customers to switch to long-term contracts. For example, it can have promotions which give discounts on longer contract terms. The firm should also encourage customers to use automatic payment methods.

From a customer demographic's perspective, if the firm wants to improve tenure, they can target those who have partners but do not have dependents. Targeting customers who are more likely to stay with the company for a longer period means the company can use its resources more efficiently. Further, for the additional services that the company offered, although they could increase customer's monthly charges, they did not necessarily play an important role in keeping the customers longer. If the company wants to increase its short-term revenue, then it can spend more effort in incentivizing customers to sign up for the additional services. However, if the goal of the company is to increase long-term revenue by increasing tenure or decreasing churn rate, then it can only pay attention to those services that mattered. For example, if a customer only signed up for internet service, then the firm should put the most effort into encouraging the customers to also sign up for the online backup service, and decrease promotions in other additional services.

To increase customer's monthly charge, the firm can target three groups of customers by different strategies: encourages customers to purchase online-related services if they have internet service, promotes multiple lines services for customers who have phone service, and recommends fiber optic internet service for users who have both phone service and internet service. In all online-related services that the firm can promote, streaming services including TV and movies can bring the highest revenue based on our linear model, so they can be prioritized.

Lastly, using the churn prediction model, the firm can find the probability of a customer churning during its time with the company. To retain customers, the firm can send out target retain offers to customers who stayed less than 15 months and have fiber optic Internet service to reduce waste of resources that provides offers to every customer. Additionally, by comparing CLVs of different groups, we could target marketing those who would bring higher CLVs to the firm to leverage resources to maximize profit. For example, compared to senior citizens, the company should target more non-senior citizens since they have a much higher CLV.

Conclusion

In conclusion, the results from the three models brought great insights on how to increase the revenue of the firm from the perspectives of tenure, monthly charges, and churn rate. We would also recommend the company to collect more information about its customers because some other factors such as income level might also affect their CLV with the company. The company can also try to extend the analysis by using other experimental methods such as A/B testing or difference-in-differences methods to validate the analysis results. In conclusion, using the information available, our model would be helpful for the company to make managerial decisions on better allocation of its resources to recruit valuable customers and retain them for a long term.

References

Department, P., & 4, N. (2020, November 04). Customer service: Churn rate by industry U.S. 2018. Retrieved December 16, 2020, from https://www.statista.com/statistics/816735/customer-churn-rate-by-industry-us/

Lechintan, A. (2017). June's smartKPI: $ Telecom subscriber acquisition cost. Retrieved December 16, 2020, from https://www.performancemagazine.org/june-smartkpi-telecom-sub-acquisition/

What Is Customer Lifetime Value (CLV)? // Qualtrics. (2020, November 16). Retrieved December 16, 2020, from https://www.qualtrics.com/experience-management/customer/customer-lifetime-value/

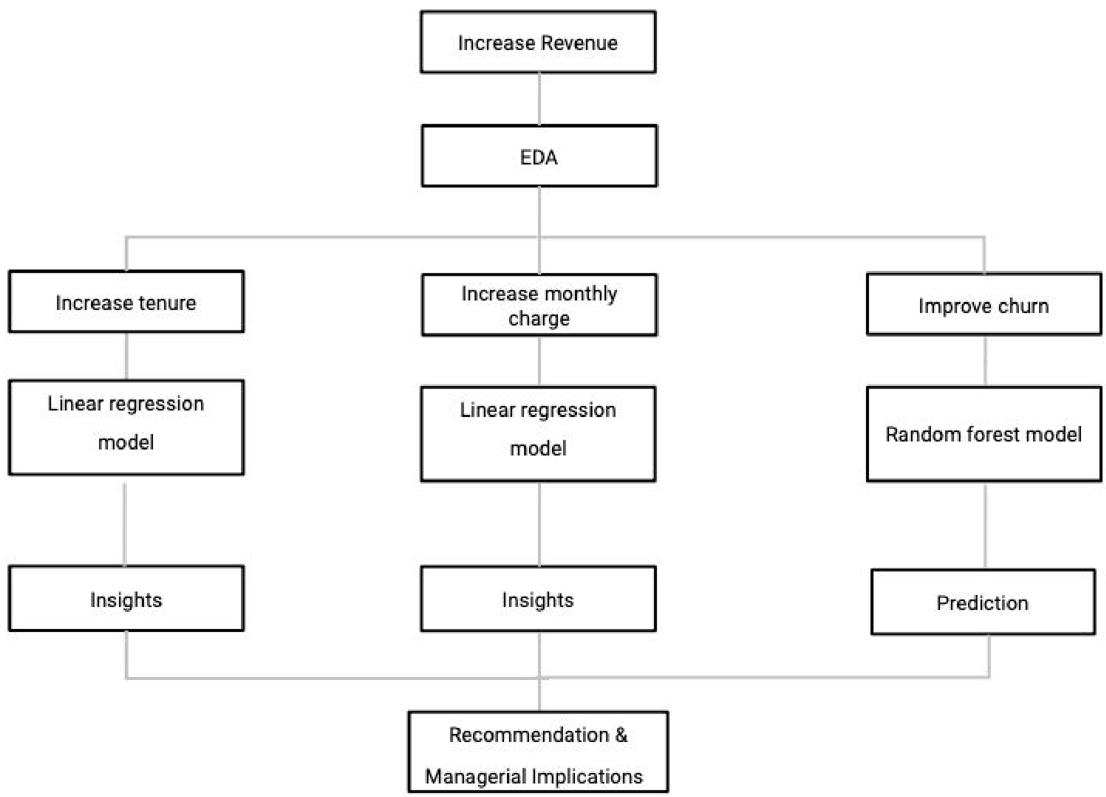
El-Shimy, H., & Dangelo, S. (2018). Build a customer churn predictor using Watson Studio and Jupyter Notebooks. Retrieved December 16, 2020, from https://developer.ibm.com/technologies/data-science/patterns/predict-customer-churn-using-watson-studio-and-jupyter-notebooks/

BlastChar. (2018, February 23). Telco Customer Churn. Retrieved December 16, 2020, from https://www.kaggle.com/blastchar/telco-customer-churn

Communications Services Industry Profitability. (n.d.). Retrieved December 16, 2020, from https://csimarket.com/Industry/industry_Profitability_Ratios.php?ind=905

Appendix

Appendix 1 Scheme of our work



Appendix 2 Exploratory Data Analysis

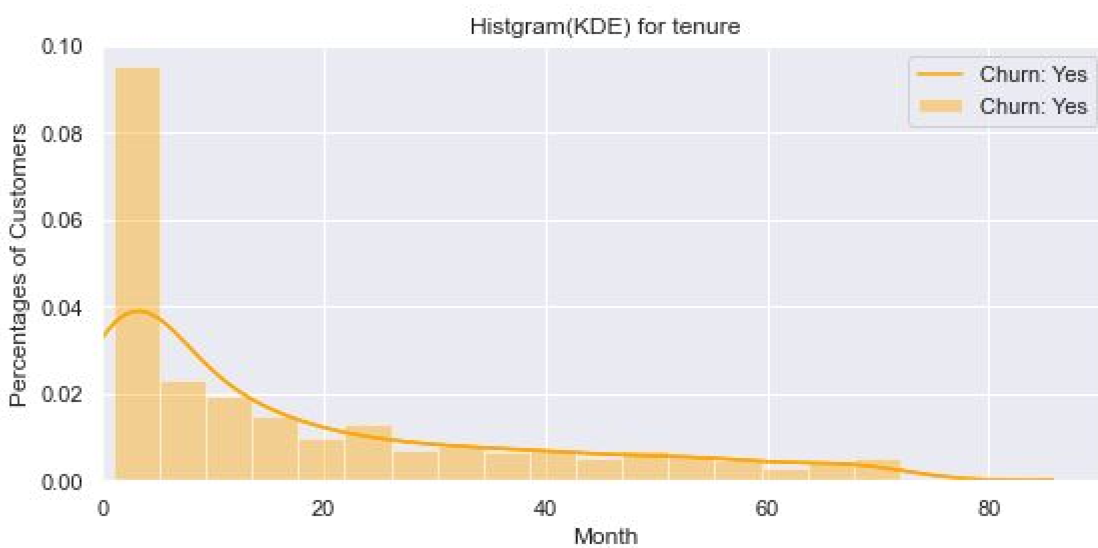
Appendix 2.1 Data Description

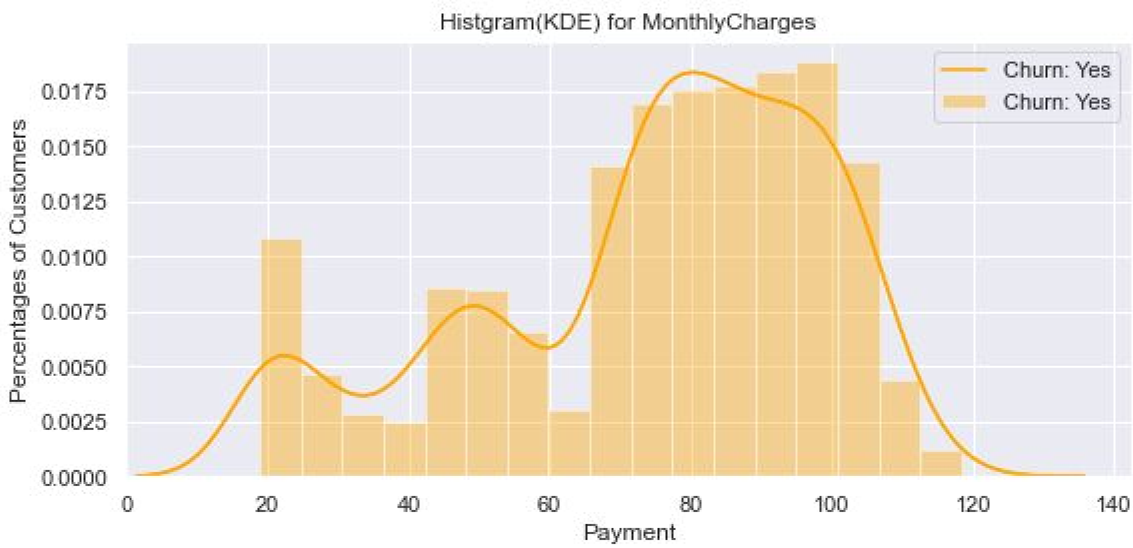
Attribute Name	Description
Gender	Whether the customer was a male or female
SeniorCitizen	Whether the customer was a senior citizen (1) or not (0)
Partner	Whether the customer had a partner or not
Dependents	Whether the customer had dependents or not
Tenure	The number of months a customer had stayed with the company
PhoneService	Whether the customer signed up for the phone service
MultipleLines	Whether the customer signed up for multiple lines of phone service
InternetService	Whether the customer signed up for the internet service and the kind of internet service
OnlineSecurity	Whether the customer signed up for the online security service or not
OnlineBackup	Whether the customer signed up for the online backup service or not
DeviceProtection	Whether the customer signed up for the device protection service or not
TechSupport	Whether the customer signed up for the tech support service or not
StreamingTV	Whether the customer signed up for the streaming TV service or not
StreamingMovies	Whether the customer signed up for the streaming movies service or not
Contract	The contract type of the customer (month-to-month, one year, two year)
PaperlessBilling	Whether the customer signed up for paperless billing or not
PaymentMethod	The current payment method of the customer (automatic bank transfer, automatic credit card payment, electronic check, mailed check)
MonthlyCharges	The monthly charge amount of the customer
TotalCharges	The total charge amount of the customer since the customer signed up for services
Churn	Whether the customer churned within the last month.

Appendix 2.2 Numerical features description

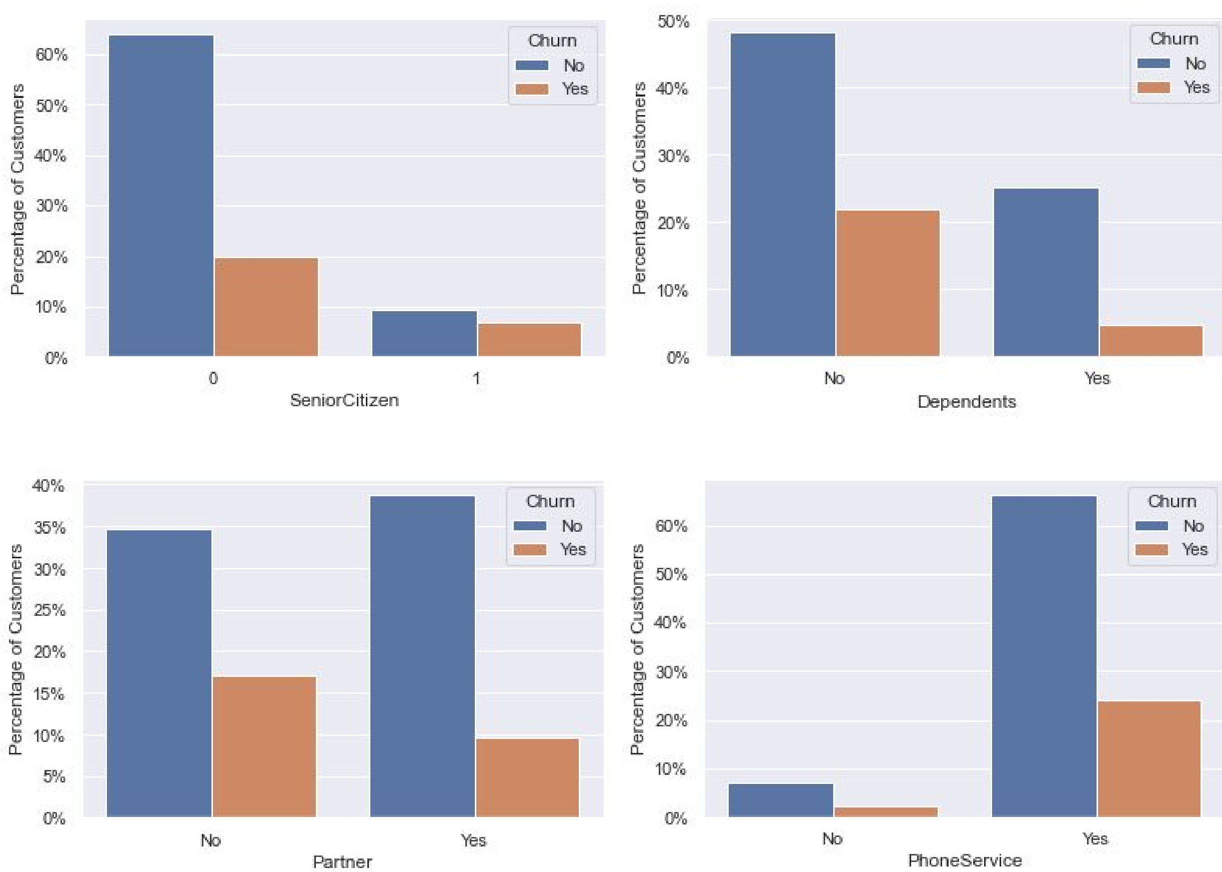
	tenure	MonthlyCharges	TotalCharges
count	7032.000000	7032.000000	7032.000000
mean	32.421786	64.798208	2283.298340
std	24.545260	30.085974	2266.770508
min	1.000000	18.250000	18.799999
25%	9.000000	35.587500	401.449997
50%	29.000000	70.350000	1397.475037
75%	55.000000	89.862500	3794.737488
max	72.000000	118.750000	8684.799805

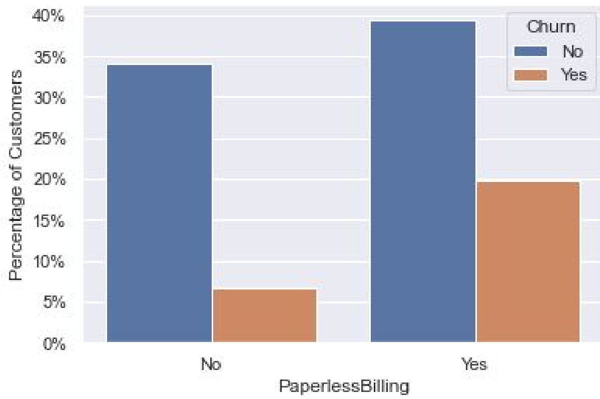
Appendix 2.3 Histograms and KDE plots of numerical features

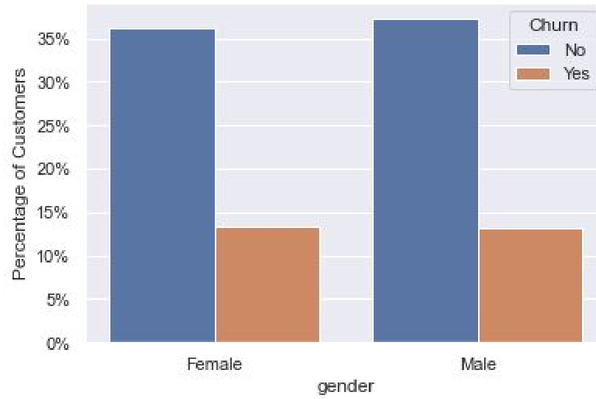




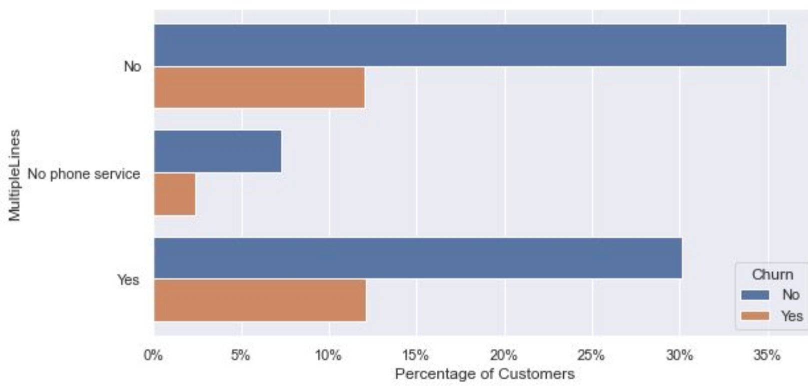
Appendix 2.4 Histograms of categorical features with two binary values

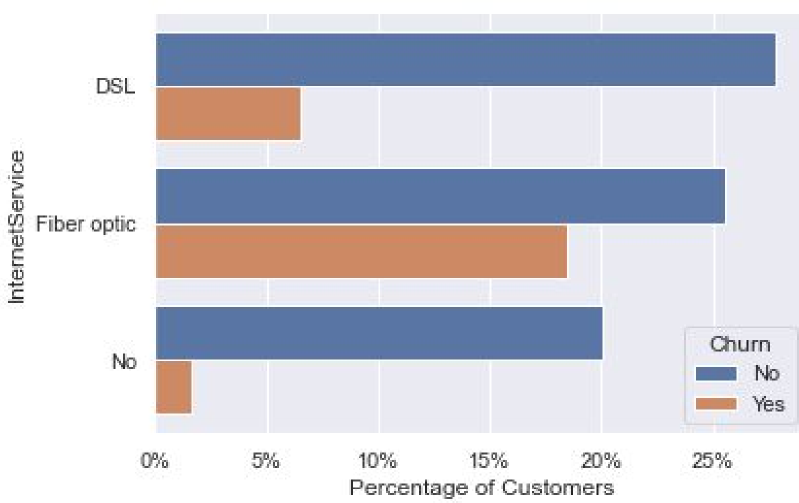


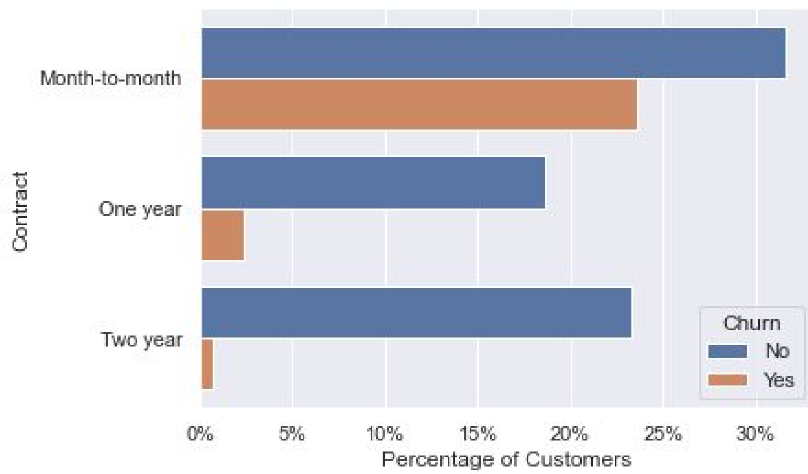




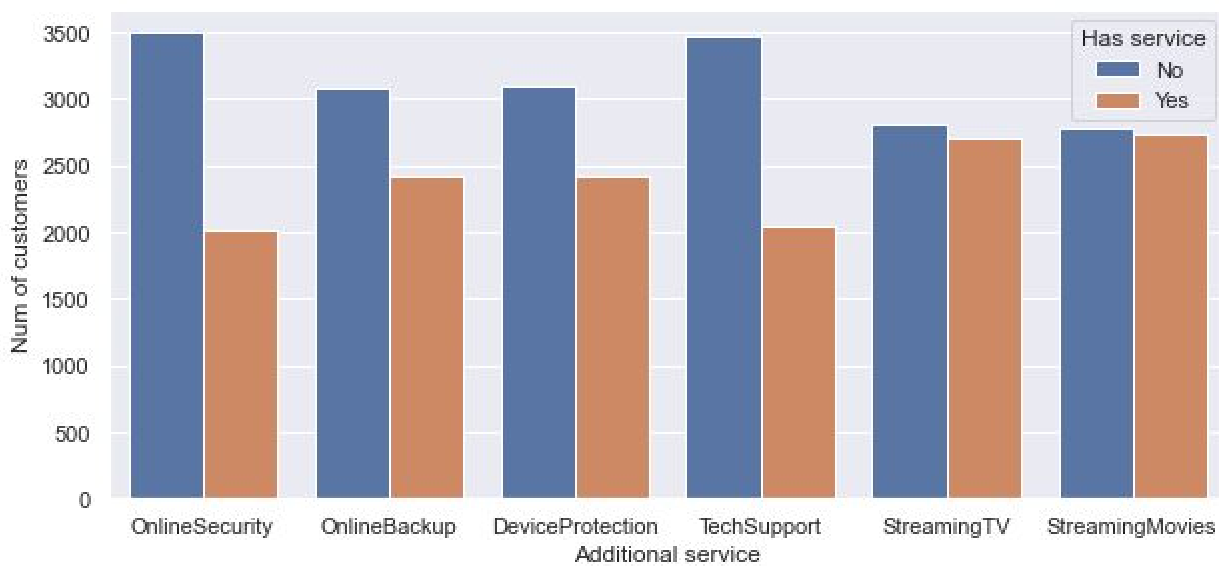
Appendix 2.5 Histograms of categorical features with three unique values

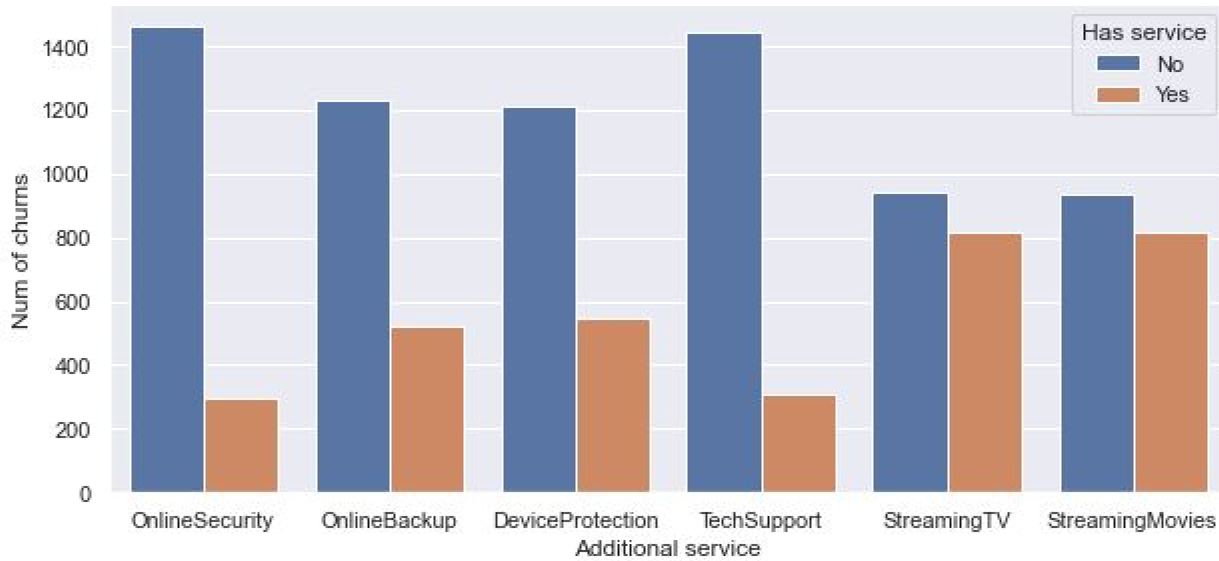




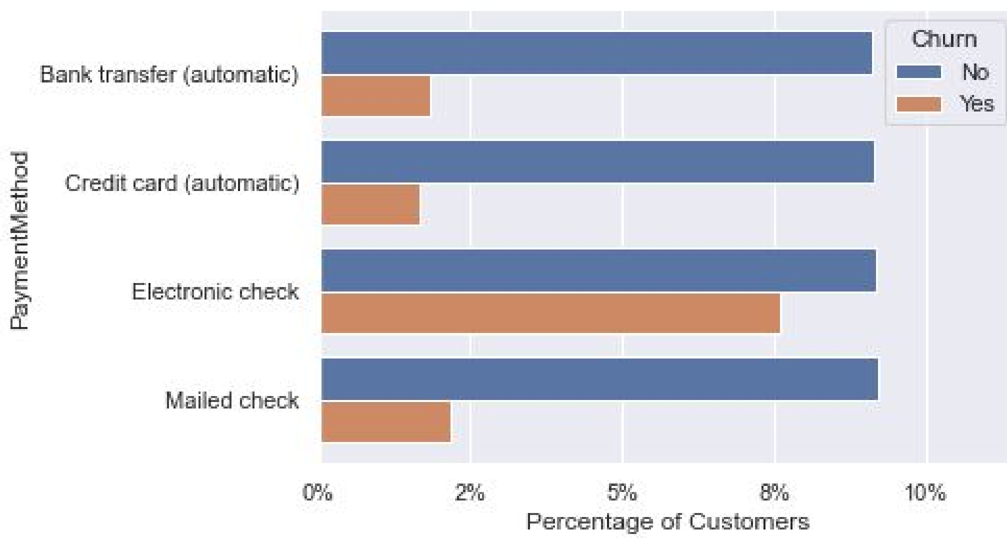


Appendix 2.6 Histograms of additional Internet services

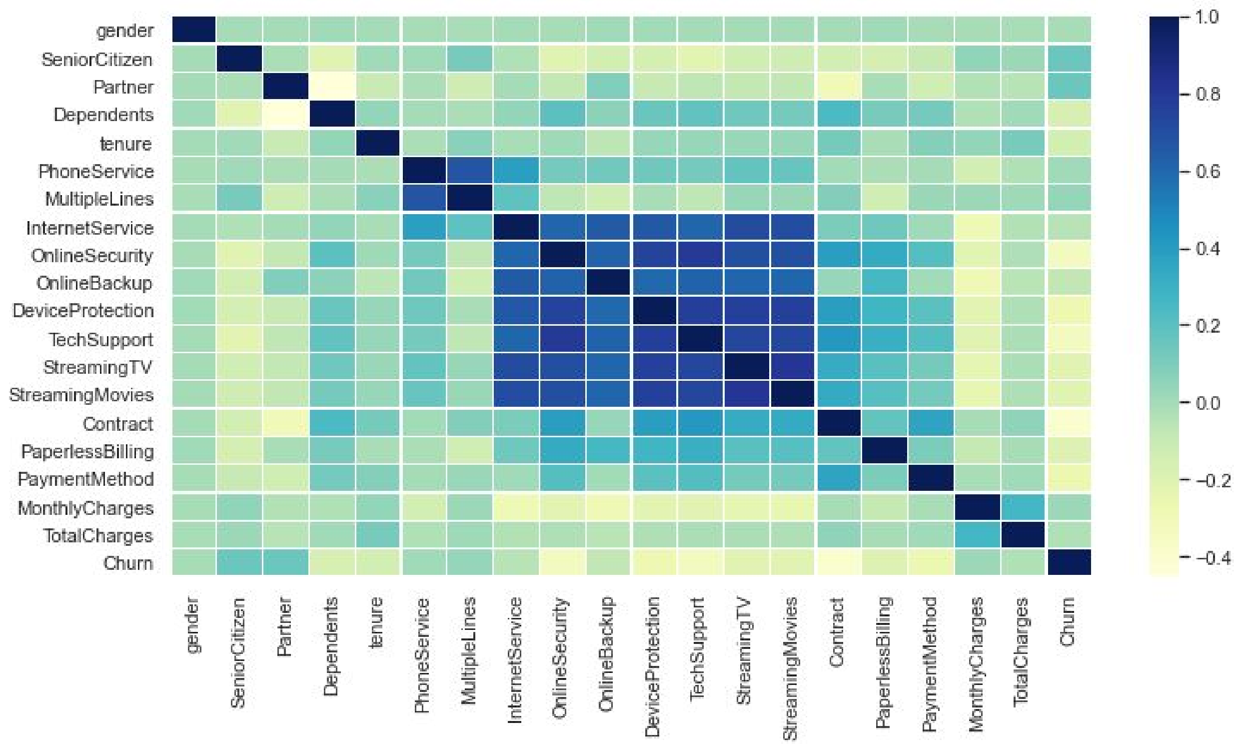




Appendix 2.7 Histogram of categorical features with four unique values



Appendix 2.8 Correlations matrix between different features



Appendix 3 Models

Appendix 3.1 Linear Regression Model on *Tenure*

Appendix 3.1.1 Regression on non-service related attributes

```
Call:
lm(formula = tenure ~ gender + seniorCitizen + partner + dependents +
    contractMtoM + contract1Year + paperlessBilling + paymentMethodElectronicCheck +
    paymentMethodMailedCheck + paymentMethodBankTransfer + monthlyCharges)

Residuals:
    Min      1Q  Median      3Q     Max
-59.755 -11.087  -0.637  10.090  60.485

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  49.190382   0.794636  61.903  < 2e-16 ***
gender                        0.556132   0.372364   1.494   0.1353
seniorCitizen                 2.462547   0.536367   4.591 4.48e-06 ***
partner                      -8.150630   0.439366 -18.551  < 2e-16 ***
dependents                    2.104693   0.476543   4.417 1.02e-05 ***
contractMtoM                -35.515680   0.513764 -69.128  < 2e-16 ***
contract1Year               -13.858497   0.563943 -24.574  < 2e-16 ***
paperlessBilling             -1.003563   0.413126  -2.429   0.0152 *
paymentMethodElectronicCheck -6.003049   0.548151 -10.951  < 2e-16 ***
paymentMethodMailedCheck     -9.237683   0.587715 -15.718  < 2e-16 ***
paymentMethodBankTransfer     0.790743   0.564493   1.401   0.1613
monthlyCharges                0.185991   0.007183  25.894  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 3.1.2 Customers with only phone service

```
Call:
lm(formula = tenure ~ partner + dependents + contractMtoM + contract1Year +
    paymentMethodCheck + monthlyCharges + multipleLinesNo)

Residuals:
    Min      1Q  Median      3Q     Max
-50.921  -9.134  -0.751  11.030  52.033

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         71.5850    18.4558   3.879 0.000109 ***
partner             -7.3448     1.0423  -7.047 2.77e-12 ***
dependents           4.0149     1.0121   3.967 7.62e-05 ***
contractMtoM       -32.3527     1.0662 -30.343  < 2e-16 ***
contract1Year      -15.6266     1.0783 -14.492  < 2e-16 ***
paymentMethodCheck  -9.1616     0.8668 -10.570  < 2e-16 ***
monthlyCharges      -0.5271     0.7379  -0.714 0.475178
multipleLinesNo    -10.8785     3.8452  -2.829 0.004730 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 3.1.3 Customers with only internet service

```
Call:
lm(formula = tenure ~ seniorCitizen + partner + contractMtoM +
    contract1Year + paperlessBilling + paymentMethodCheck + monthlyCharges +
    onlineBackupNo)

Residuals:
    Min      1Q  Median      3Q     Max
-61.891 -11.231  -0.298  10.582  50.218

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         36.09345    4.27895   8.435  < 2e-16 ***
seniorCitizen        4.37851    1.79271   2.442  0.01485 *
partner             -5.79300    1.31892  -4.392 1.30e-05 ***
contractMtoM       -23.55328    1.94240 -12.126  < 2e-16 ***
contract1Year       -9.30706    1.93490  -4.810 1.86e-06 ***
paperlessBilling    -3.95125    1.31717  -3.000  0.00280 **
paymentMethodCheck  -8.87126    1.35352  -6.554 1.11e-10 ***
monthlyCharges       0.53076    0.06982   7.602 9.86e-14 ***
onlineBackupNo      -4.31316    1.36597  -3.158  0.00166 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 3.1.4 Customers with both internet and phone service

```
Call:
lm(formula = tenure ~ seniorCitizen + partner + dependents +
    contractMtoM + contract1Year + paymentMethodCheck + monthlyCharges +
    multipleLinesNo + onlineSecurityNo + onlineBackupNo + deviceProtectionNo +
    techSupportNo)

Residuals:
    Min      1Q  Median      3Q     Max
-58.357  -9.061  -1.130   8.430  53.217

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         55.64349    1.46845  37.893  < 2e-16 ***
seniorCitizen        2.14119    0.52897   4.048 5.25e-05 ***
partner             -7.01451    0.47064 -14.904  < 2e-16 ***
dependents           1.23598    0.52653   2.347  0.01895 *
contractMtoM       -27.60889    0.70657 -39.074  < 2e-16 ***
contract1Year       -8.98667    0.68201 -13.177  < 2e-16 ***
paymentMethodCheck  -5.54072    0.44241 -12.524  < 2e-16 ***
monthlyCharges       0.17428    0.01388  12.553  < 2e-16 ***
multipleLinesNo     -7.26061    0.46194 -15.717  < 2e-16 ***
onlineSecurityNo    -4.52935    0.46728  -9.693  < 2e-16 ***
onlineBackupNo      -6.78185    0.44743 -15.157  < 2e-16 ***
deviceProtectionNo  -3.35413    0.47273  -7.095 1.48e-12 ***
techSupportNo       -1.28619    0.48368  -2.659  0.00786 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix 3.2 Linear Regression Model on *MonthlyCharges*

Appendix 3.2.1 Customers with both internet and phone services

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         MonthlyCharges   R-squared:                       0.996
Model:                            OLS   Adj. R-squared:                  0.996
Method:                 Least Squares   F-statistic:                 4.625e+04
Date:                Tue, 15 Dec 2020   Prob (F-statistic):               0.00
Time:                        14:59:44   Log-Likelihood:                -5237.1
No. Observations:                3384   AIC:                         1.051e+04
Df Residuals:                    3364   BIC:                         1.064e+04
Df Model:                          19
Covariance Type:            nonrobust
=========================================================================================================
                                            coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------
gender                                   -0.0356      0.039     -0.905      0.366      -0.113       0.042
SeniorCitizen                             0.0108      0.052      0.207      0.836      -0.091       0.113
Partner                                  -0.0179      0.047     -0.380      0.704      -0.110       0.074
Dependents                               -0.0215      0.051     -0.421      0.674      -0.122       0.079
tenure                                   -0.0005      0.001     -0.373      0.709      -0.003       0.002
PhoneService                             45.0934      0.074    611.547      0.000      44.949      45.238
MultipleLines                             5.0558      0.044    113.700      0.000       4.969       5.143
OnlineSecurity                            5.0079      0.047    106.385      0.000       4.916       5.100
OnlineBackup                              4.9879      0.044    113.344      0.000       4.902       5.074
DeviceProtection                          5.0758      0.045    112.022      0.000       4.987       5.165
TechSupport                               4.9658      0.048    103.823      0.000       4.872       5.060
StreamingTV                               9.9418      0.046    215.435      0.000       9.851      10.032
StreamingMovies                           9.9636      0.046    216.455      0.000       9.873      10.054
PaperlessBilling                         -0.0456      0.045     -1.024      0.306      -0.133       0.042
InternetService_Fiber optic              24.9766      0.048    515.254      0.000      24.882      25.072
Contract_One year                        -0.0392      0.063     -0.623      0.533      -0.162       0.084
Contract_Two year                         0.0186      0.080      0.231      0.818      -0.139       0.176
PaymentMethod_Credit card (automatic)    -0.0498      0.059     -0.840      0.401      -0.166       0.066
PaymentMethod_Electronic check           -0.0496      0.055     -0.897      0.370      -0.158       0.059
PaymentMethod_Mailed check                0.0031      0.069      0.044      0.965      -0.132       0.138
==============================================================================
Omnibus:                        8.080   Durbin-Watson:                   2.035
Prob(Omnibus):                  0.018   Jarque-Bera (JB):                9.644
Skew:                          -0.024   Prob(JB):                      0.00805
Kurtosis:                       3.257   Cond. No.                         217.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Appendix 3.2.2 Customers with only phone service

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         MonthlyCharges   R-squared:                       0.931
Model:                            OLS   Adj. R-squared:                  0.930
Method:                 Least Squares   F-statistic:                     1187.
Date:                Tue, 15 Dec 2020   Prob (F-statistic):               0.00
Time:                        14:26:37   Log-Likelihood:                -889.46
No. Observations:                1068   AIC:                             1805.
Df Residuals:                    1055   BIC:                             1870.
Df Model:                          12
Covariance Type:            nonrobust
===========================================================================================================
                                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------------
gender                                    0.0007      0.035      0.020      0.984      -0.067       0.069
SeniorCitizen                             0.0218      0.094      0.232      0.817      -0.163       0.206
Partner                                  -0.0213      0.045     -0.472      0.637      -0.110       0.067
Dependents                                0.0290      0.043      0.670      0.503      -0.056       0.114
tenure                                   -0.0014      0.001     -1.322      0.187      -0.004       0.001
PhoneService                             20.0252      0.057    349.586      0.000      19.913      20.138
MultipleLines                             5.0062      0.045    110.861      0.000       4.918       5.095
PaperlessBilling                          0.0135      0.038      0.354      0.724      -0.061       0.088
Contract_One year                        -0.0374      0.050     -0.748      0.455      -0.136       0.061
Contract_Two year                         0.0383      0.057      0.670      0.503      -0.074       0.150
PaymentMethod_Credit card (automatic)    -0.0395      0.052     -0.757      0.449      -0.142       0.063
PaymentMethod_Electronic check           -0.0748      0.074     -1.011      0.312      -0.220       0.070
PaymentMethod_Mailed check               -0.0445      0.048     -0.932      0.351      -0.138       0.049
==============================================================================
Omnibus:                        1.256   Durbin-Watson:                   1.916
Prob(Omnibus):                  0.534   Jarque-Bera (JB):                1.134
Skew:                          -0.043   Prob(JB):                        0.567
Kurtosis:                       3.135   Cond. No.                         217.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

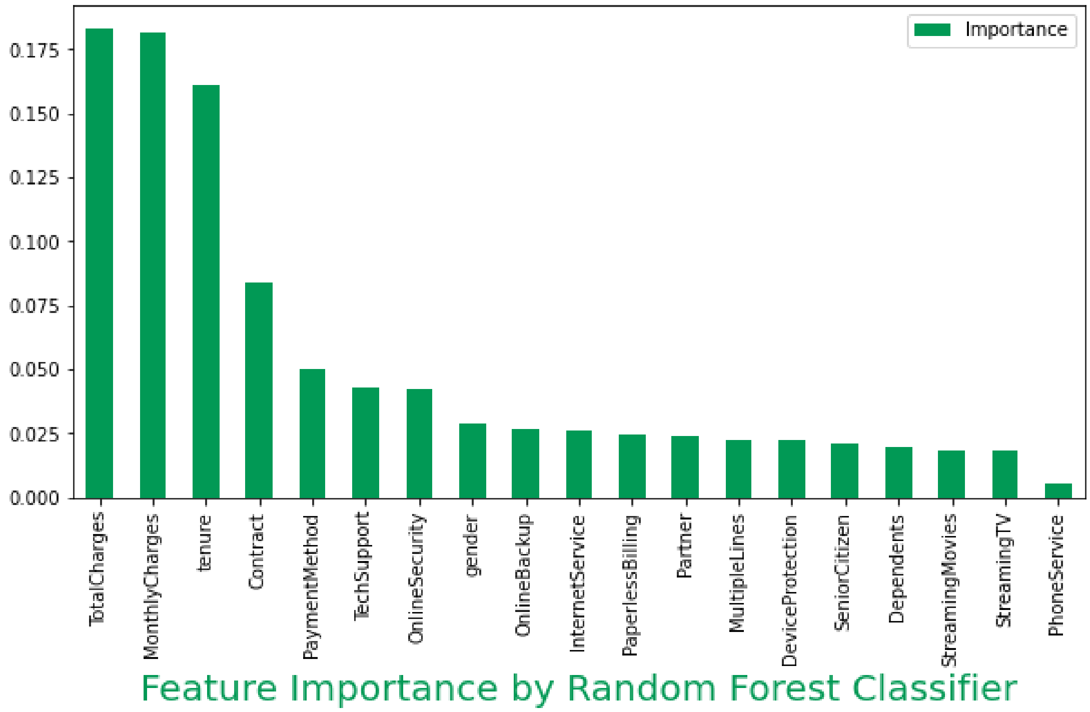
Appendix 3.2.3 Customers with only internet service

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         MonthlyCharges   R-squared:                       0.994
Model:                            OLS   Adj. R-squared:                  0.994
Method:                 Least Squares   F-statistic:                     4341.
Date:                Tue, 15 Dec 2020   Prob (F-statistic):               0.00
Time:                        14:58:32   Log-Likelihood:                -626.33
No. Observations:                 477   AIC:                             1289.
Df Residuals:                     459   BIC:                             1364.
Df Model:                          17
Covariance Type:            nonrobust
===========================================================================================================
                                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------------
const                                    24.8864      0.155    160.063      0.000      24.581      25.192
gender                                   -0.0656      0.086     -0.761      0.447      -0.235       0.104
SeniorCitizen                             0.1549      0.127      1.216      0.224      -0.095       0.405
Partner                                   0.0203      0.099      0.204      0.838      -0.175       0.215
Dependents                                0.0418      0.107      0.392      0.695      -0.168       0.251
tenure                                   -0.0007      0.003     -0.267      0.790      -0.006       0.005
PhoneService                           2.626e-15    6.6e-17     39.797      0.000     2.5e-15    2.76e-15
OnlineSecurity                            4.8778      0.100     48.661      0.000       4.681       5.075
OnlineBackup                              5.0327      0.091     55.414      0.000       4.854       5.211
DeviceProtection                          4.8602      0.101     47.923      0.000       4.661       5.059
TechSupport                               5.1612      0.101     51.201      0.000       4.963       5.359
StreamingTV                               9.9241      0.100     99.572      0.000       9.728      10.120
StreamingMovies                           9.9459      0.101     98.493      0.000       9.747      10.144
PaperlessBilling                          0.1371      0.091      1.500      0.134      -0.043       0.317
Contract_One year                         0.2936      0.128      2.290      0.022       0.042       0.546
Contract_Two year                         0.1565      0.163      0.963      0.336      -0.163       0.476
PaymentMethod_Credit card (automatic)     0.1143      0.131      0.869      0.385      -0.144       0.373
PaymentMethod_Electronic check            0.0079      0.131      0.060      0.952      -0.250       0.266
PaymentMethod_Mailed check               -0.0420      0.136     -0.308      0.758      -0.310       0.226
==============================================================================
Omnibus:                        7.689   Durbin-Watson:                   1.876
Prob(Omnibus):                  0.021   Jarque-Bera (JB):                8.961
Skew:                           0.195   Prob(JB):                       0.0113
Kurtosis:                       3.546   Cond. No.                     2.67e+18
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

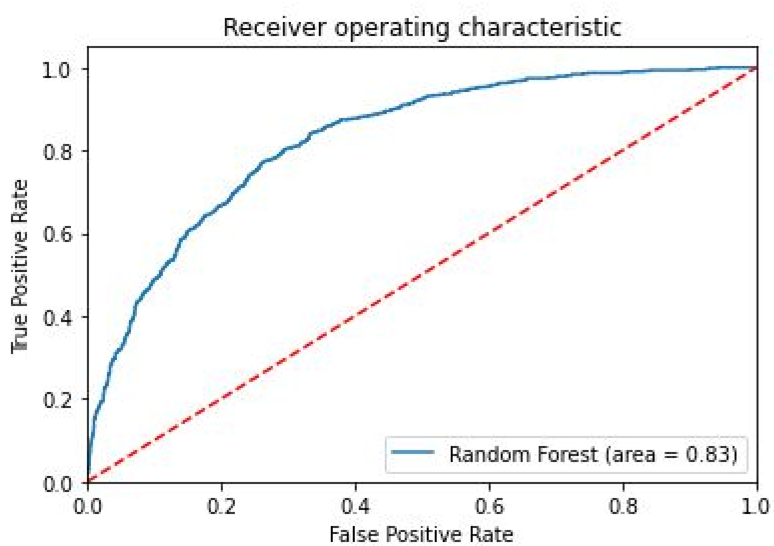
Appendix 3.3 Random Forest Model on *Churn*

Appendix 3.3.1 Overall feature importance plot



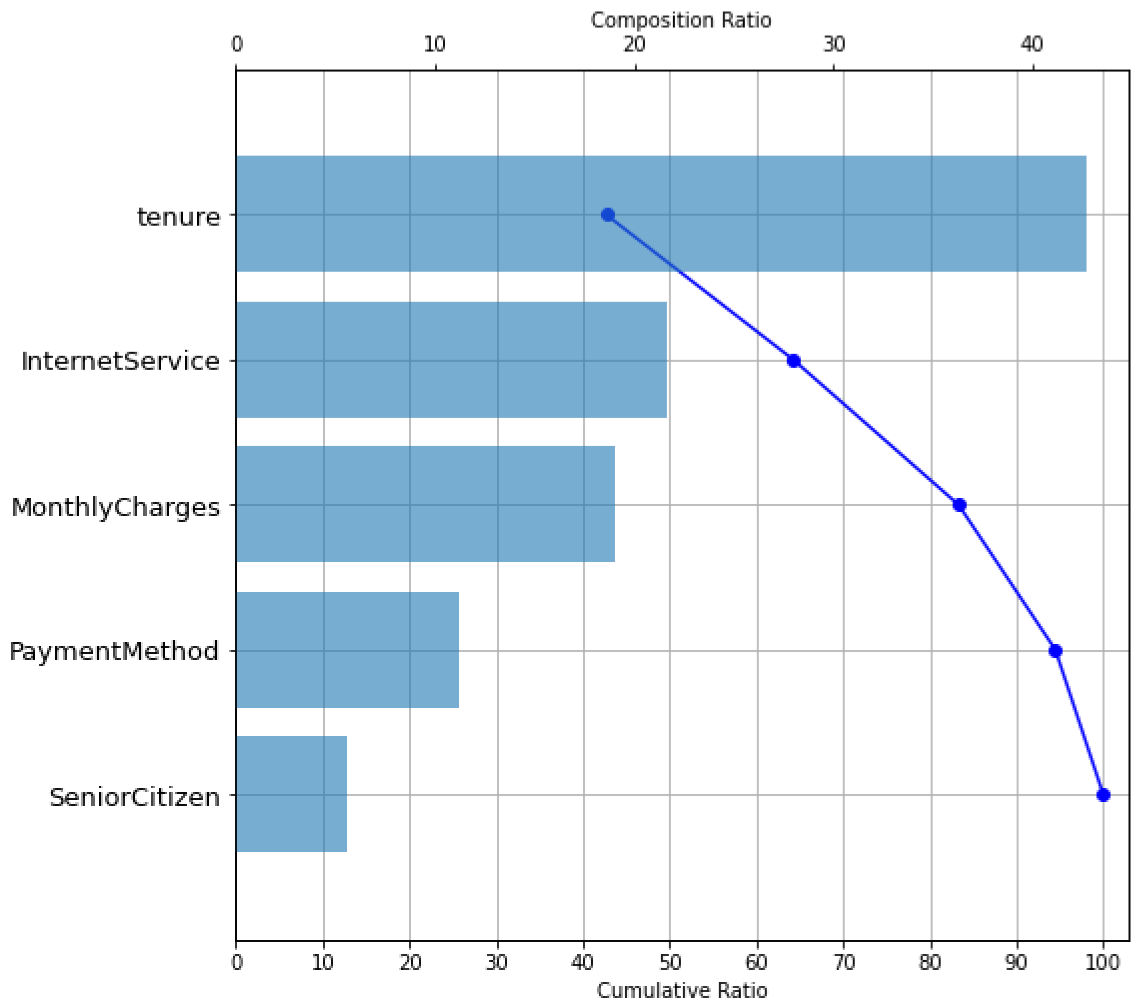
Appendix 3.3.2 ROC curve, AUC score and accuracy for test and train data

```
=============ROC AUC======================
```

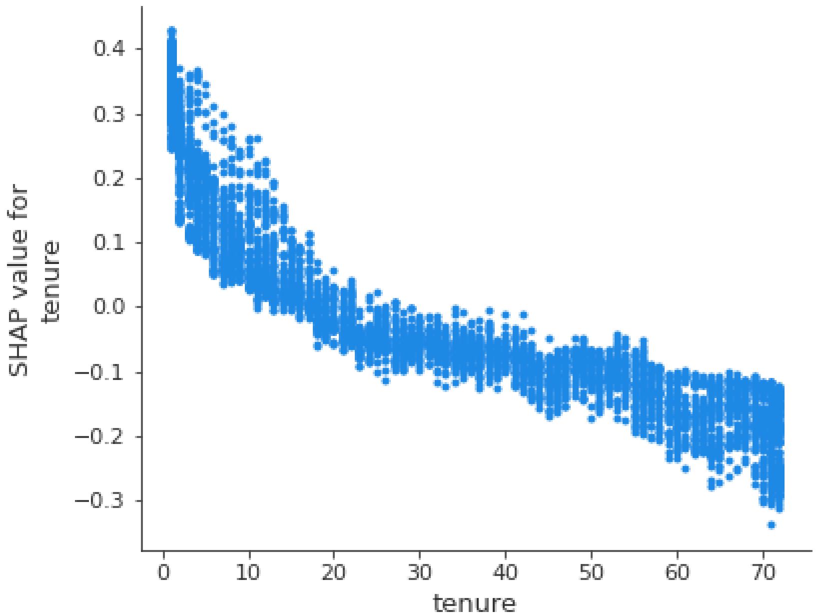


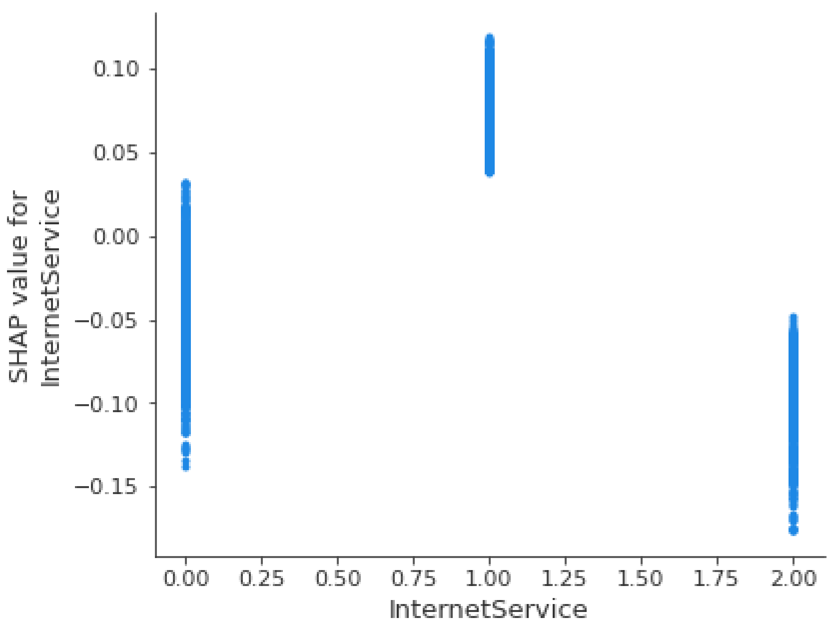

```
AUC: 0.8268628258807523
Accuracy score for test data : 0.7910383455407152
Accuracy score for train data : 0.8367650180428784
```

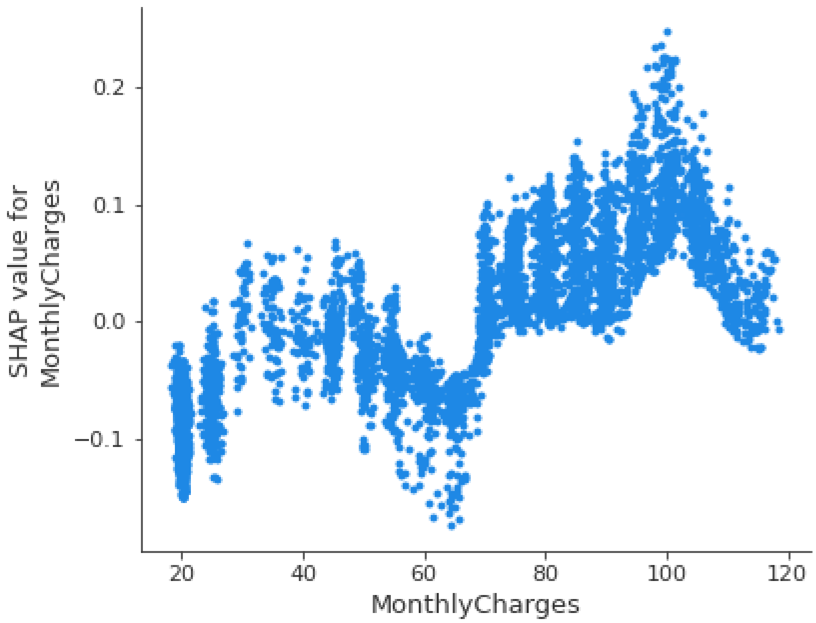
Appendix 3.3.3 SHAP waterfall plot for feature importance

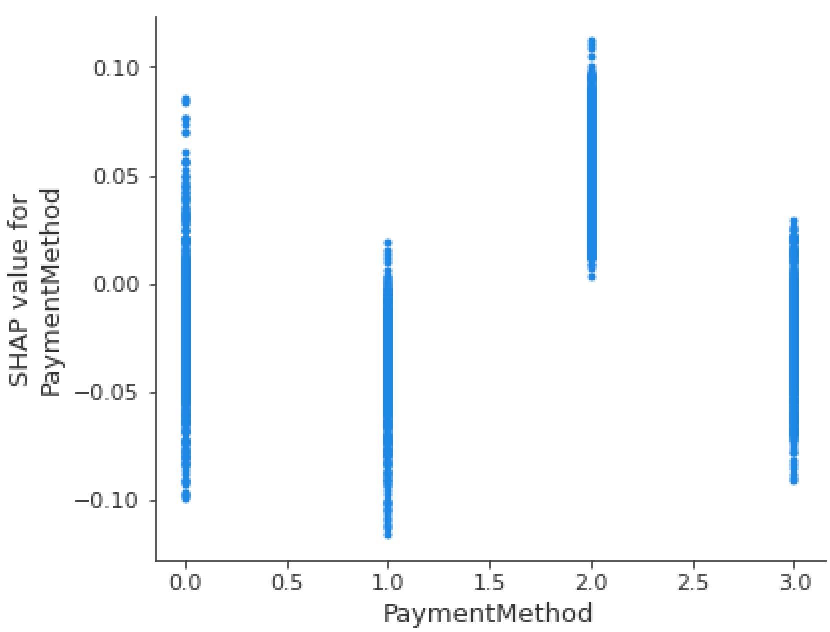


Appendix 3.3.4 SHAP dependence plots for important features






SHAP value for
MonthlyCharges
0.2
0.1
0.0
−0.1
20
40
60
80
100
120
MonthlyCharges


SHAP value for
PaymentMethod
0.10
0.05
0.00
−0.05
−0.10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
PaymentMethod

Appendix 3.4 Results of CLV calculation

```
> notseniorCLV
[1] 541.506
> seniorCLV
[1] 90.56
> # automatic payment
> pay0CLV
[1] 834.3015
> pay1CLV
[1] 541.506
> pay2CLV
[1] 133.779
> pay3CLV
[1] 168.6358
```