Executive Summary Report
Brandon Arias

Introduction:

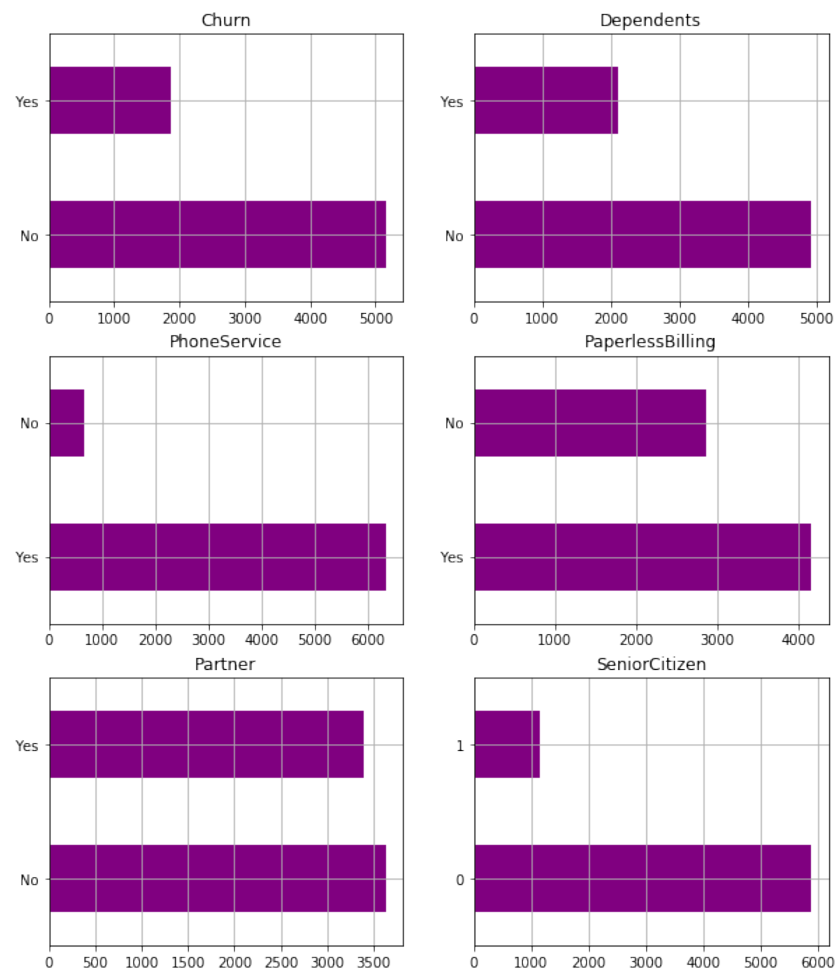
Phone companies provides telecommunications services such as telephony and data communications access to their patrons. However, to provide the best quality of service requires understanding their audience and their patronage. For this Telco company, it offers internet p

In this report, we want to explore their audience and come up with an algorithm that best explains what variables predict churn

Data Analysis:

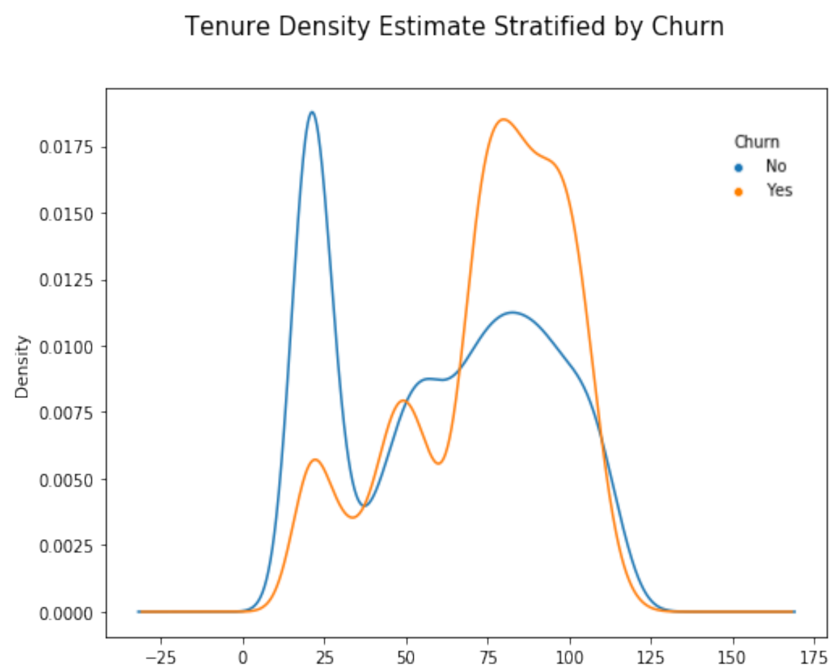
From doing a simple exploration of the data, I noticed the following:

- ~ 1/3 people churn
- ~ 1/3 people have dependents
- ~ 1/8 people are not subscribed for phone service
- ~ Paperless Billing and Partner distributions are roughly equal
- 1/12 customers are considered senior citizens

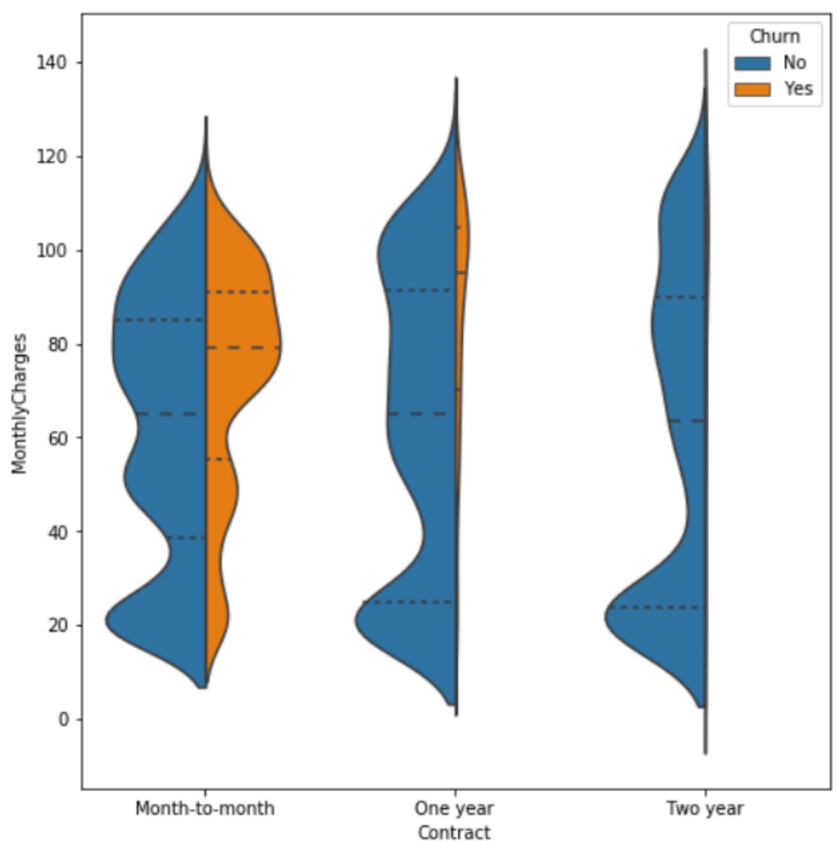


Other distribution of other features can be found in the appendix.

When looking at the average monthly charges and broken down if they churned or not, we can note that as the charges get higher, there seems to be a higher density of people that will churn.

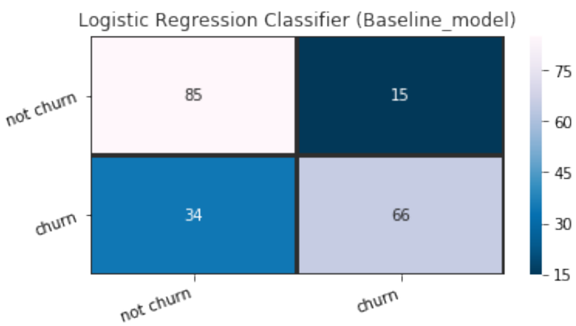


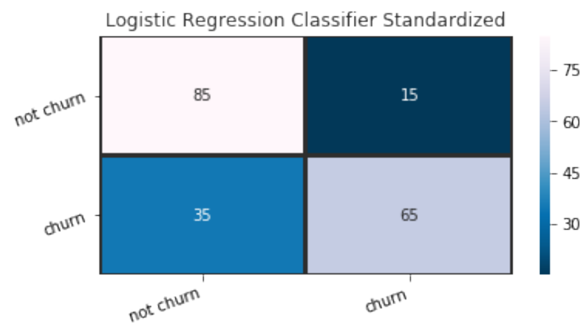
Exploring more on churn, I explored the type of contracts and the monthly charges broken down by churn. From this approach, I found that those in the month-to-month contract found themselves to churn more and especially when they were charge more by month, particularly where the charge is more than 60 dollars. Those that were in the one-year contract and two-year contract generally stayed as a patron.

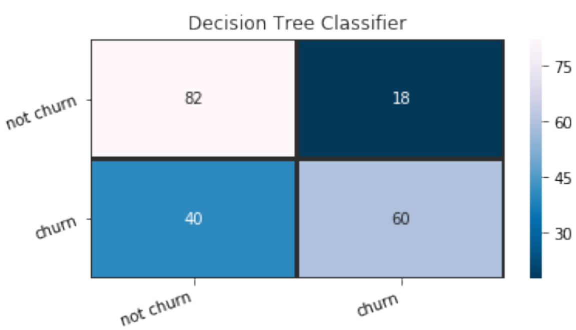


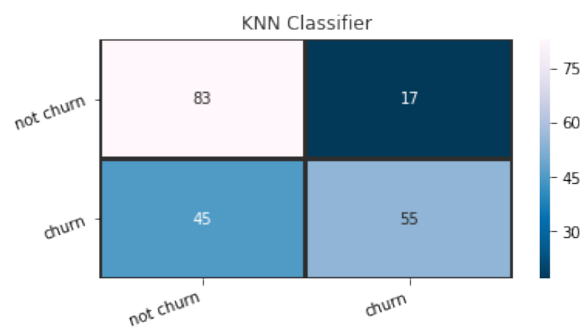
<u>Modeling:</u>

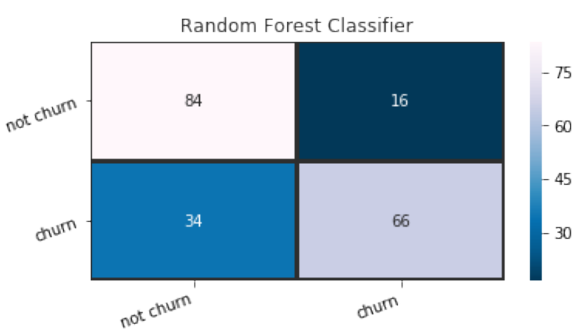
In utilizing various classification models, I found that logistics regression classifier (standardized or not standardize), Random Forest Classifier, and NeuralNets Classifier works best in classifying churns.

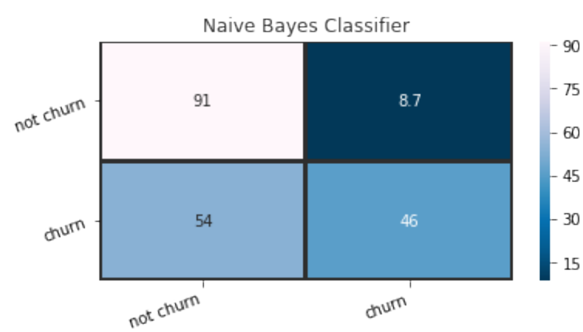


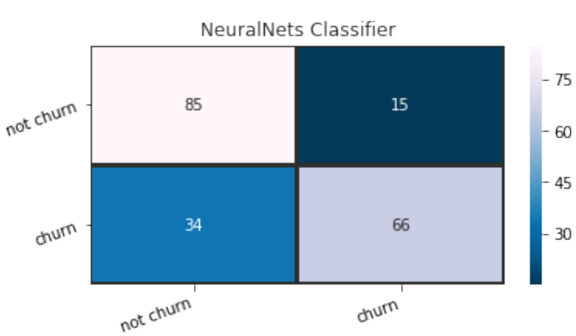




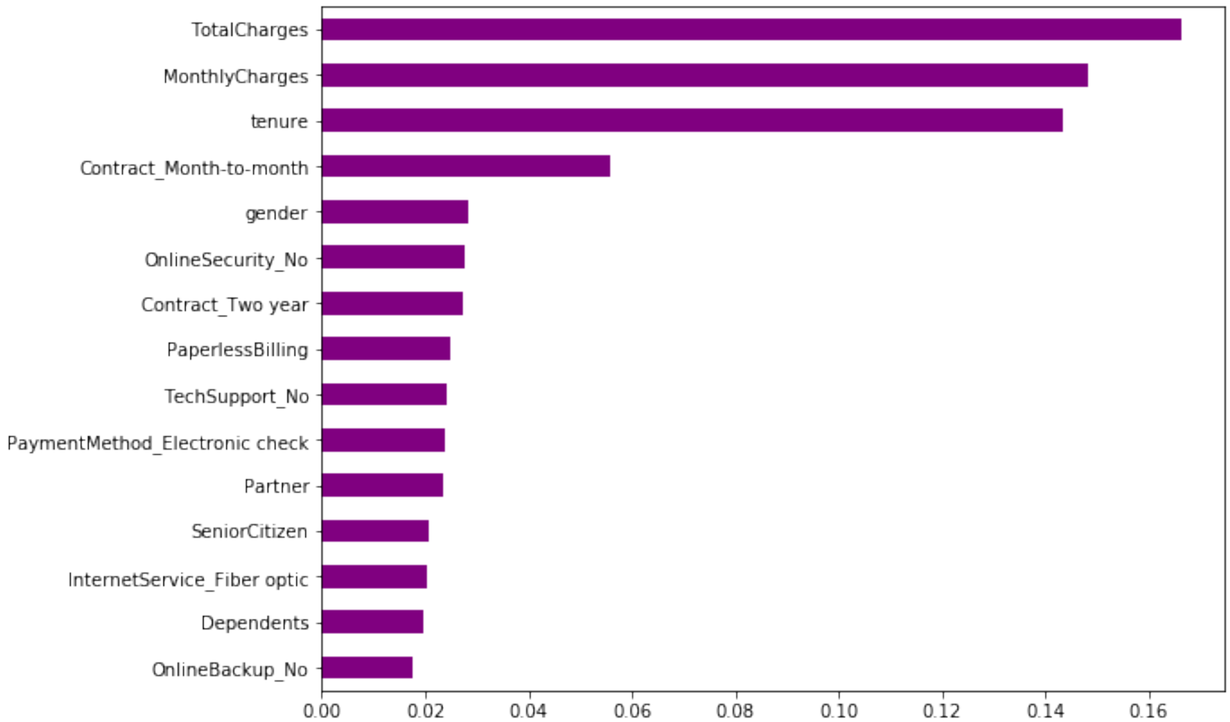








For the variables that helped explain for whether a customer would churn, I utilized Random Forest feature importance since Random Forest is more robust than logistic regression and very quick to train. When looking at the feature variables, we can note that total charges, monthly charges, tenure, and contract that is month-to-month are very important in predicting churn.

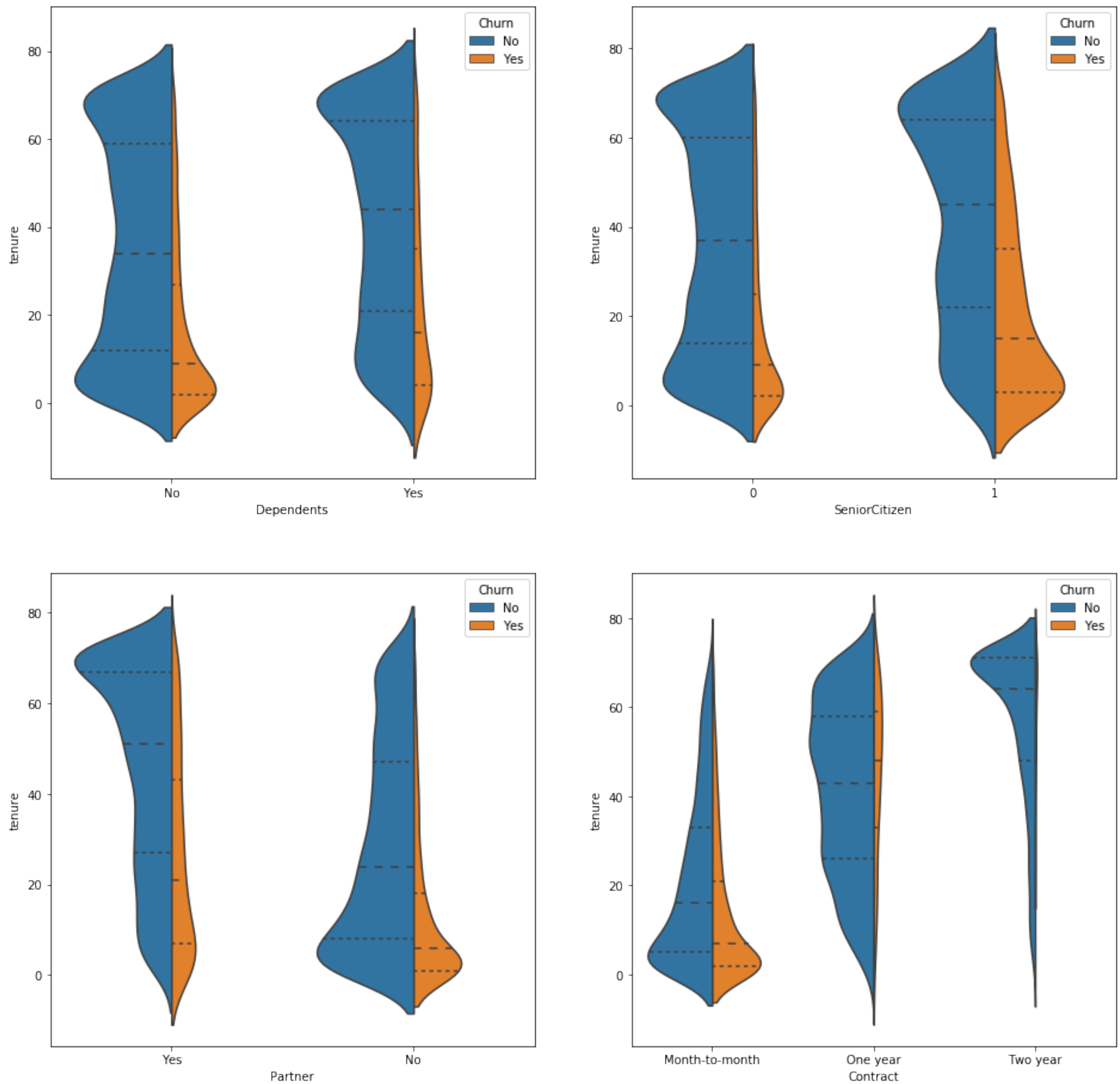


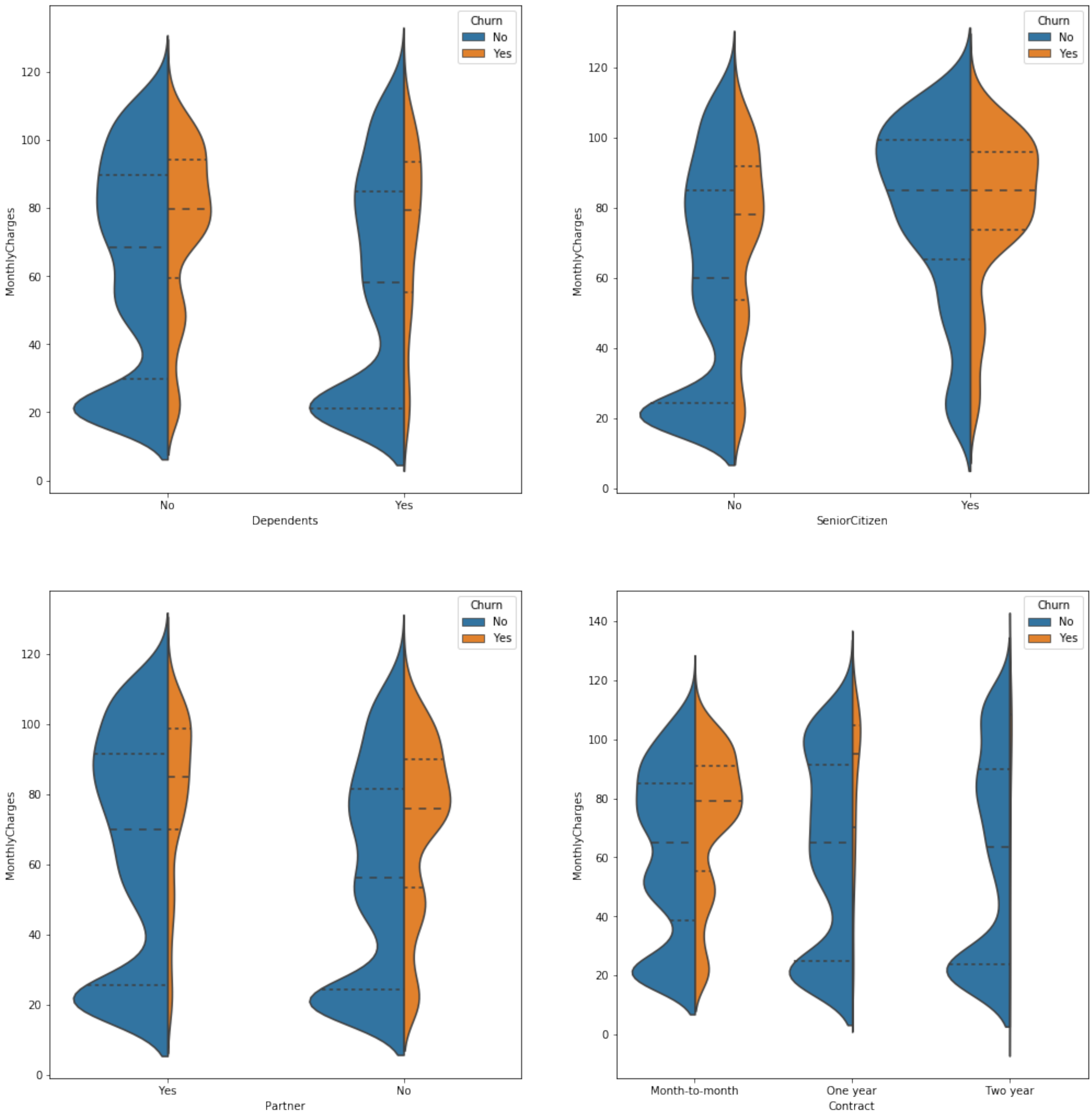
Conclusion:

Many of the data analysis visualization did add up when looking at the feature variables of the best performing models. This is a good sign! The next step for this would be finding a way to explore more on understanding more about these individuals that fall in these three four categories.

Appendix:


TechSupport Status
No internet service
Yes
No
0 500 1000 1500 2000 2500 3000 3500
Contract Status
One year
Two year
Month-to-month
0 500 1000 1500 2000 2500 3000 3500 4000
StreamingTV Status
No internet service
Yes
No
0 500 1000 1500 2000 2500
MultipleLines
No phone service
Yes
No
0 500 1000 1500 2000 2500 3000 3500
InternetService
No
DSL
Fiber optic
0 500 1000 1500 2000 2500 3000
OnlineSecurity
No internet service
Yes
No
0 500 1000 1500 2000 2500 3000 3500
OnlineBackup
No internet service
Yes
No
0 500 1000 1500 2000 2500 3000
DeviceProtection
No internet service
Yes
No
0 500 1000 1500 2000 2500 3000


Churn
No
Yes
tenure
80
60
40
20
0
No
Yes
Dependents
SeniorCitizen
0
1
Partner
Contract
Month-to-month
One year
Two year



Churn
No
Yes
MonthlyCharges
Dependents
SeniorCitizen
Partner
Contract
Month-to-month
One year
Two year
0
20
40
60
80
100
120
140