Nolan Long

Telco Customer Churn Final Report

Problem Statement

Telco is an Internet Service Provider with a large customer base. The company would like to lower the percentage of people who cancel; known as the "churn rate". The business decision makers would like me to create a model that can predict customers who are likely to leave soon in order for the company to effectively tailor retention initiatives.

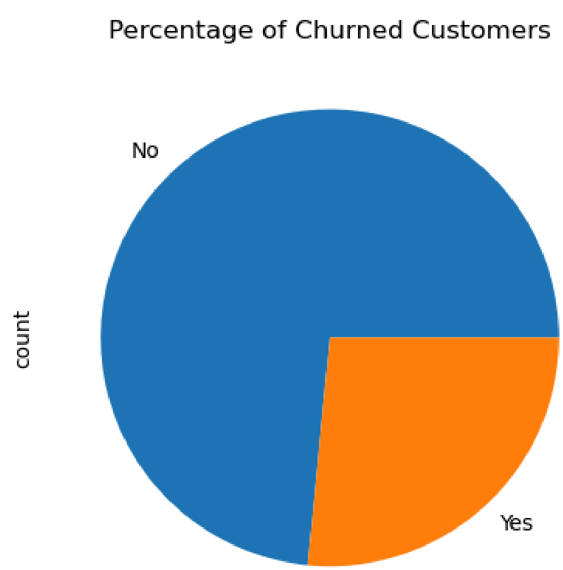
Data Wrangling

I have been given a CSV of 7,043 past and current customers with 20 columns describing their types of service and whether they decided to cancel or not. The list of features are below. Most of these features are categorical with the linear features being tenure, monthly charges, and total charges.

```
['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn']
```

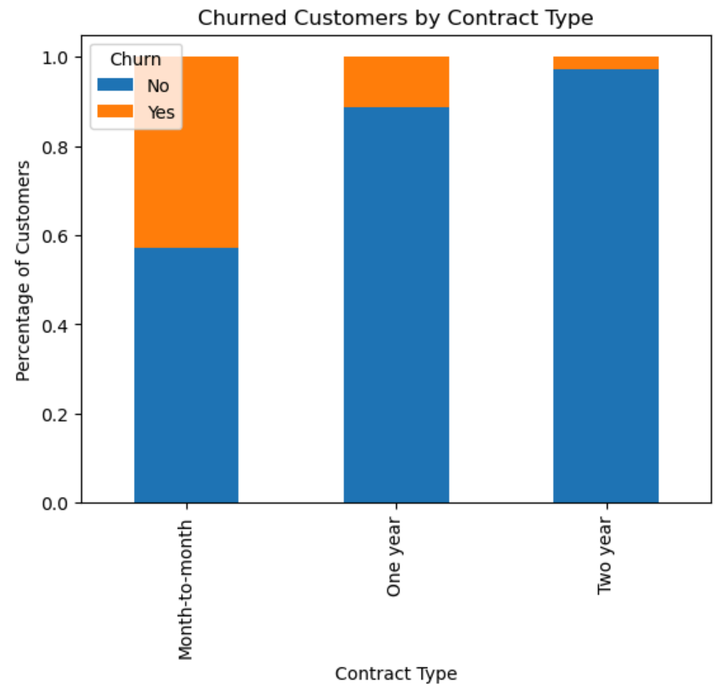
Exploratory Data Analysis

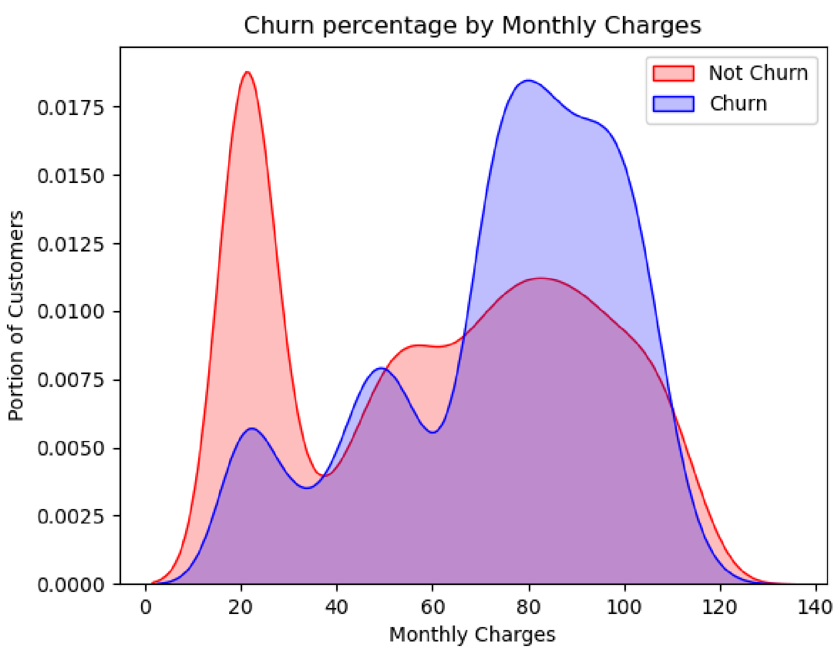
The first thing I wanted to find in this dataset is how many examples of churned customers were given. This being the dependent variable I will be modeling on it is important that I had enough and that they were not biased.

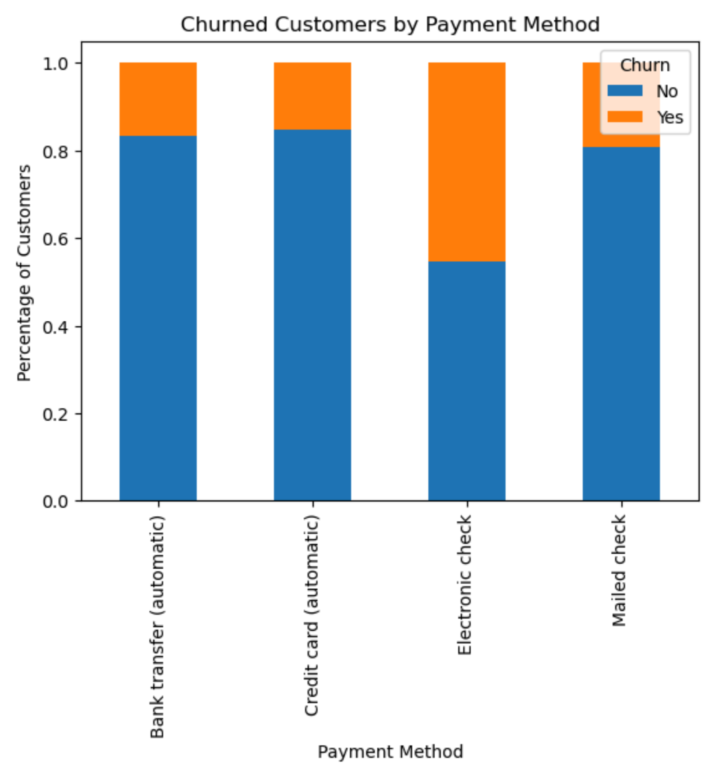


Nolan Long

Next I wanted to look at some features that intuitively would seem like they would lead to higher churn rate. Analyzing these could prove some valuable business insight for the decision makers. Looking at contract types we can see that the longer the contract length the less likely they are to cancel. Another interesting piece I found during EDA was that customers who paid by electronic check were much more likely to churn than the other three categories. Finally we can see the distribution of churn at different price points.







Nolan Long

Preprocessing

After evaluating all the features in this dataset I felt they would all be beneficial to the end model aside from the Customer_ID, which I removed. Most of the features were categorical with a low number of options so I used one hot encoding to create a data frame that showed 1 or 0 for each category within each column. For the linear features I took these out, regularized them using a standard scaler and then merged them into the other dataframe. Finally, I created a training and testing set with 25% for testing.

Modeling

The evaluation metric I chose for these models was accuracy. I felt this would be the best metric because there are no harsh penalties for false positives or false negatives.

Logistic Regression

This baseline model with no grid search hyperparameter tuning would help me evaluate the other models.

ACCURACY SCORE = .790

	Predicted Not Churned	Predicted Churn
Not Churned	1152	148
Churned	222	236

Decision Tree Classifier

For this model I used a grid search for hyperparameter tuning with the following options `{'max_depth': np.arange(3, 10), 'criterion': ['gini', 'entropy'], 'splitter': ['best', 'random'], 'min_samples_split': np.arange(2, 10), 'min_samples_leaf': np.arange(1, 10)}`

The best hyperparameters were

`{'criterion': 'entropy', 'max_depth': 6, 'min_samples_leaf': 1, 'min_samples_split': 3, 'splitter': 'random'}`

ACCURACY SCORE = .779

	Predicted Not Churned	Predicted Churn
Not Churned	1135	177
Churned	177	269

Gradient Boosting Classifier

I again used a grid search for hyperparameter tuning.

```
{'n_neighbors': array([1, 2, 3, 4, 5, 6, 7, 8, 9]), 'weights': ['uniform',
'distance']}
```

The best hyperparameters were

```
{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 50}
```

This was the best performing model and will be the best to predict customers who are likely to cancel.

ACCURACY SCORE = .814

	Predicted Not Churned	Predicted Churn
Not Churned	1190	122
Churned	205	241

Outcomes and Final Thoughts

In conclusion, I have developed a machine learning model for the company that can significantly enhance customer retention strategies within the organization. This model not only provides valuable foresight into potential churn risks, but also allows Telco to tailor retention initiatives more effectively, ultimately fostering stronger relationships with their customers. It will be important for the data team at Telco to continue updating and improving the model as customer behavior adapts over time. As this continues to move forward. Integrating this predictive capability will be able to sustain long-term business growth and ensure customer satisfaction for Telco.