10Pearls Shine Data Science Internship

Project Report

Purpose of the Project

The goal was to identify patterns which were precursors for customer churn with the help of the “Telco Customer Churn” dataset. In addition to data analysis, I would also need to develop a machine learning model for professional use in predicting whether a specific customer would churn.

Customer Information for Analysis

As previously mentioned, we had to make use of the “Telco Customer Churn” dataset. It mentions:

1. customerID
2. Gender
3. Senior Citizen
4. Partner
5. Dependents
6. Tenure
7. Phone Service
8. Multiple Lines
9. Internet Service
10. Online Security
11. Online Backup
12. Device Protection

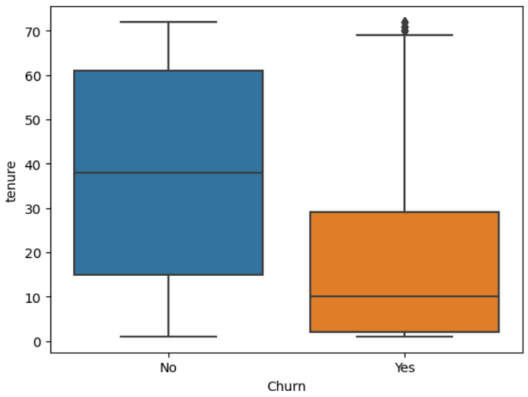
13. Tech Support
14. Streaming TV
15. Streaming Movies
16. Contract
17. Paperless Billing
18. Payment Method
19. Monthly Charges
20. Total Charges
21. Churn

Of these I can discard customer id as it is not a valid predictor of churn. For the remaining data, after cleaning the data set, I conducted exploratory data analysis (EDA) and data visualization. The findings will be mentioned in detail in the next section.

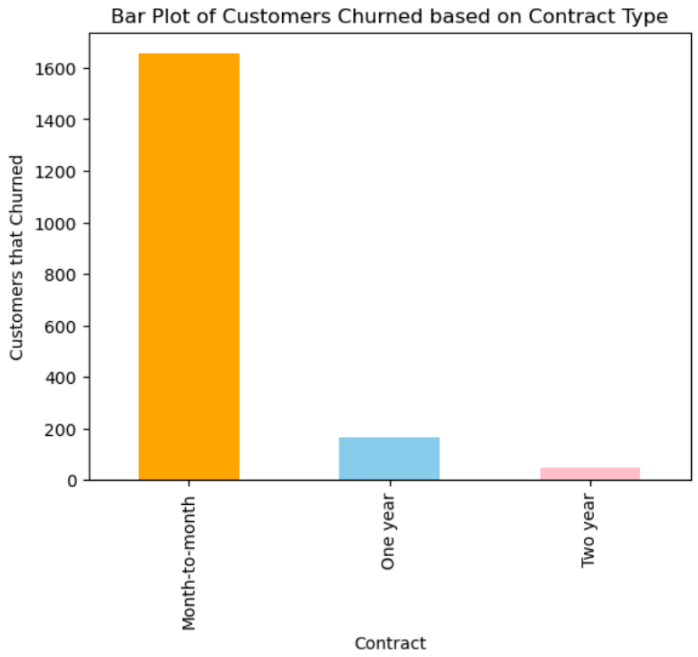
In addition to the prior mentioned EDA I also stored the data and model predictions in an SQLite database and conducted further investigation using SQL. The results of that investigation are also mentioned separately later in this report.

Findings from EDA

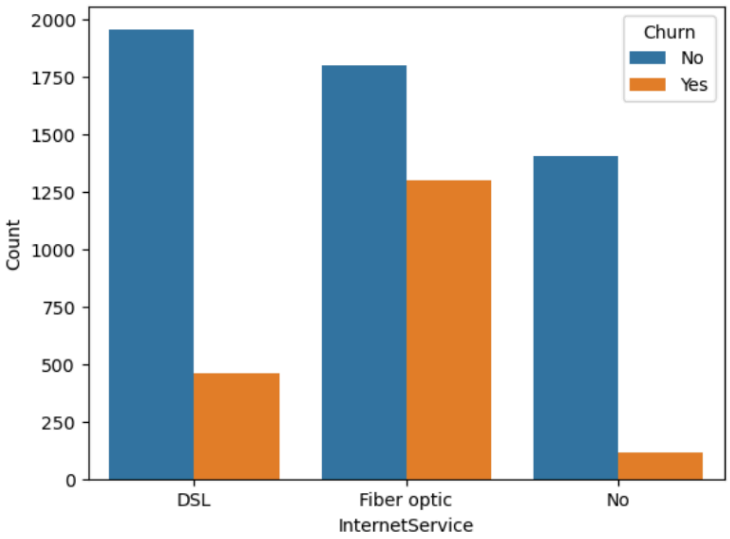
1. It appears customers were more likely to churn if they had been with the service for approx 30 months (2.5 years) or less.



2. Customers with higher monthly charges were more likely to churn.
3. A strong majority of customers that churned were those who purchased a month-to-month contract.

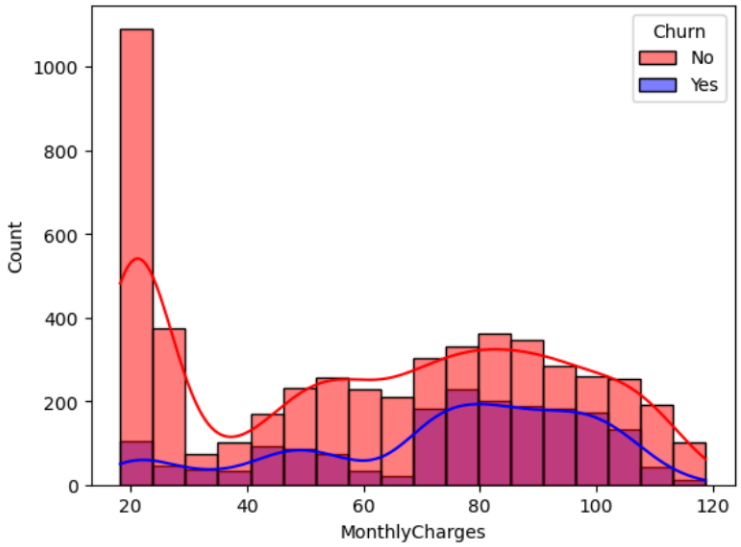


4. Customers using fiber optic for the internet service were more likely to churn than those using DSL.



5. Having a partner or dependents makes you less likely to churn.

6. Increased customer retention with lower monthly charges.



Findings using SQL

1. Most customers joined a service using a month-to-month contract. Those customers that decided to stay with the company far longer preferred 2 year contracts over 1 year contracts.
2. Our logistic regression model correctly predicted 4683 customers that were retained and 1016 customers that churned.
3. Customers that churned appeared to have paid higher monthly charges on average.
4. Most customers chose to opt for both a phone service and a fiber optic internet service (3096 customers)
5. More customers who churned were using both the phone service and fiber optic.
6. Customers had an average tenure of 32.4 months.

Suggestions for the Future

Based on our findings above, we can take a number of steps to lowering the likelihood of churn:

1. Reducing monthly charges or offering discount offers on an alternate monthly basis.
2. Promoting more towards larger families.

3. Given lower average market costs for DSL over Fiber, it is preferred to promote DSL in smaller communities and Fiber for larger communities where higher internet speeds and bandwidth may be required.