Telco Customer Churn Analysis

Analysis and Prediction

Problem

Every business, including Telecommunication companies, exist to grow and maintain their customer base.

Therefore understanding the causes of customer churn is critical, since it would allow for prevention strategies.

Telco business generates numerous data types in its interaction with its customers, and identifying key factors to customer churn is very challenging without advanced techniques in data science.

What factors might affect customer churn?

- Household demographics
- Plan types
- Product SKUs
- Customer Tenure
- Billing Methods
- Monthly Charge amount

Not included in this analysis are:

- Competitors' promotions
- Macroeconomic variables
- Other external factors

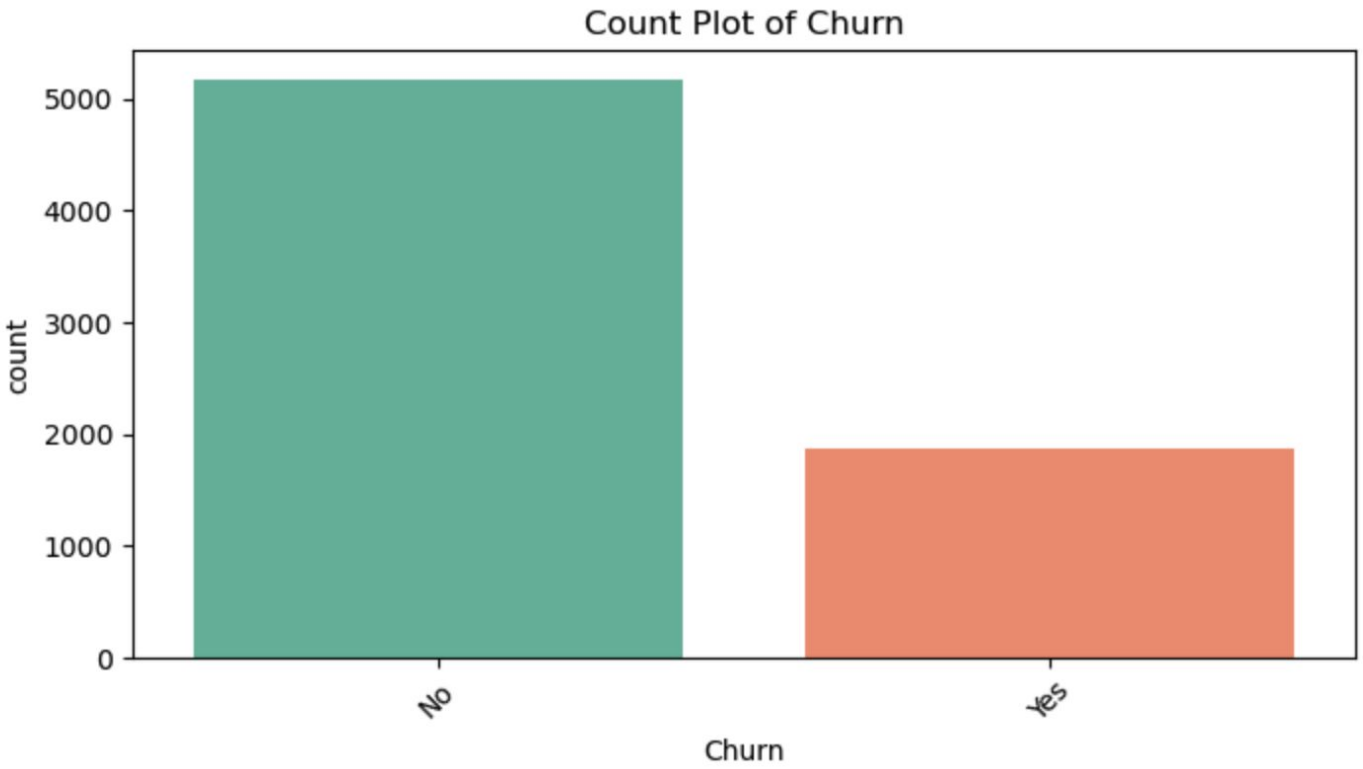
Data Information

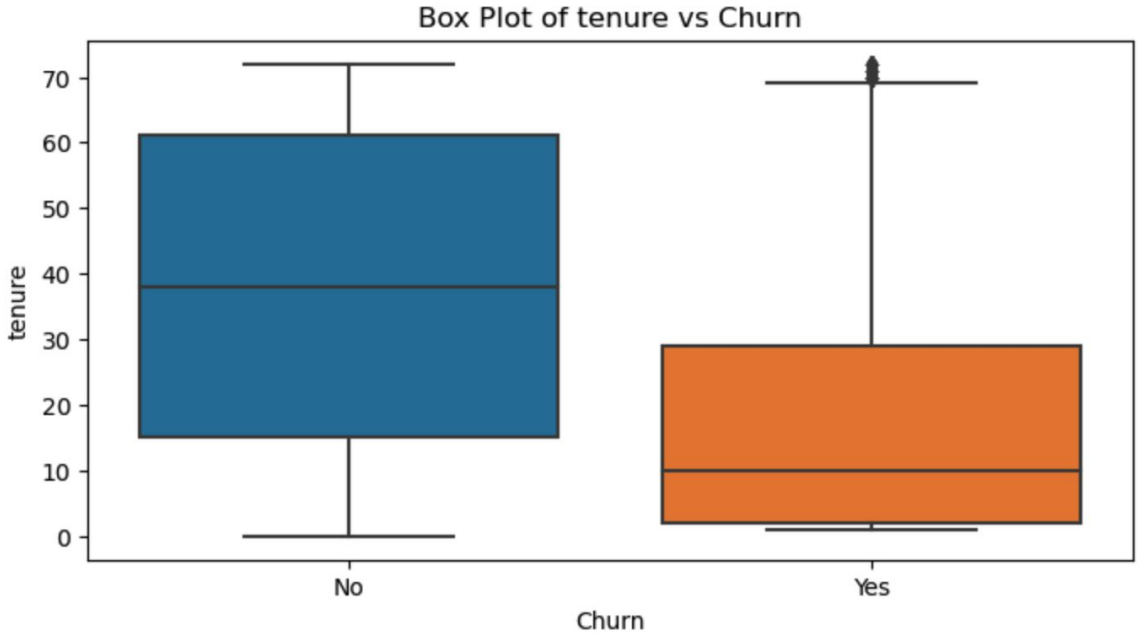
- Used a dataset found on Kaggle (https://www.kaggle.com/datasets/blastchar/telco-customer-churn)
- Contains 7043 rows (customers) and 21 features.

The data set includes information about:

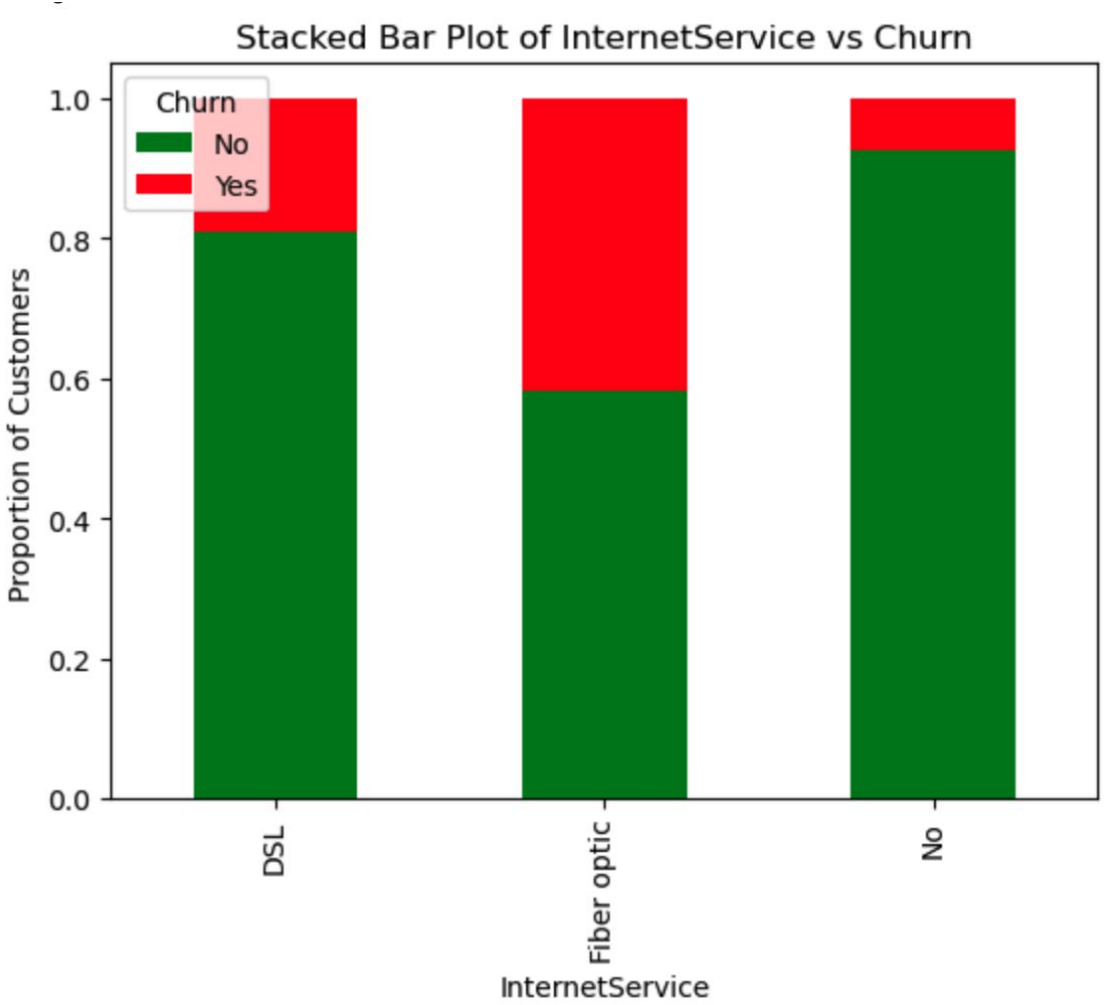
- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Target value distribution

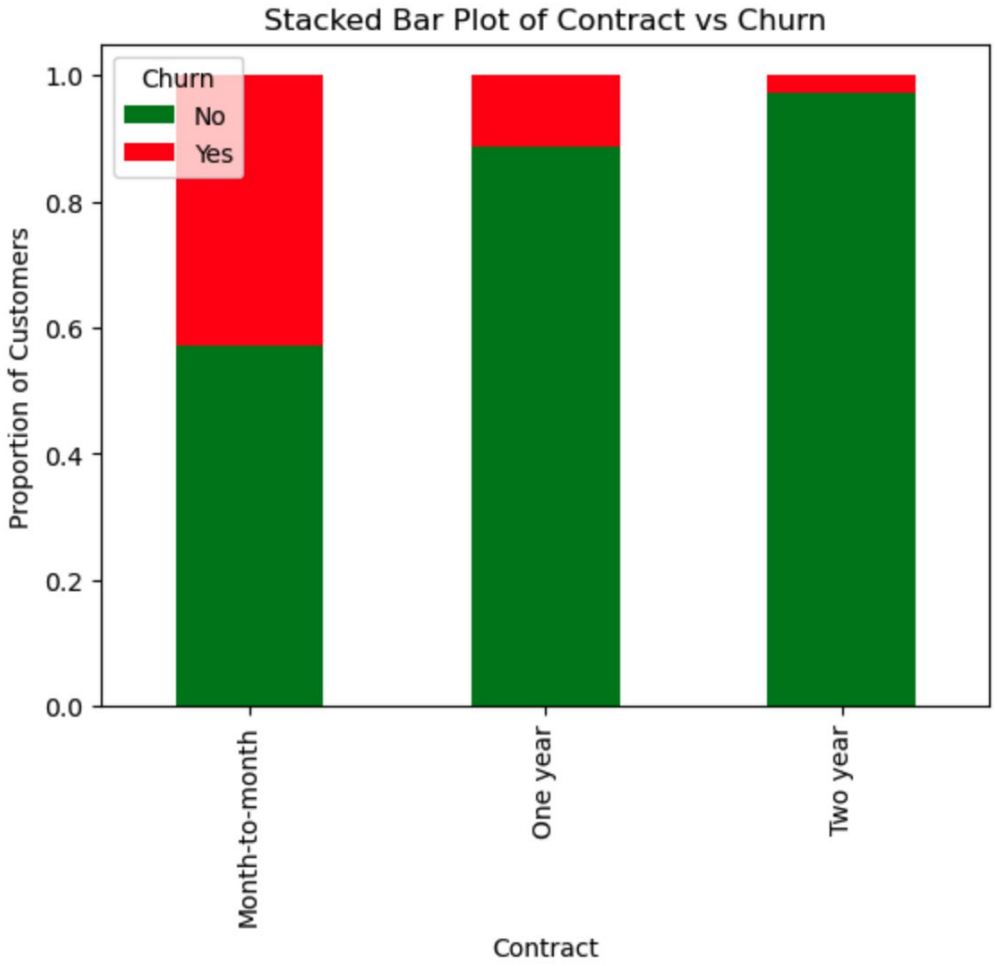




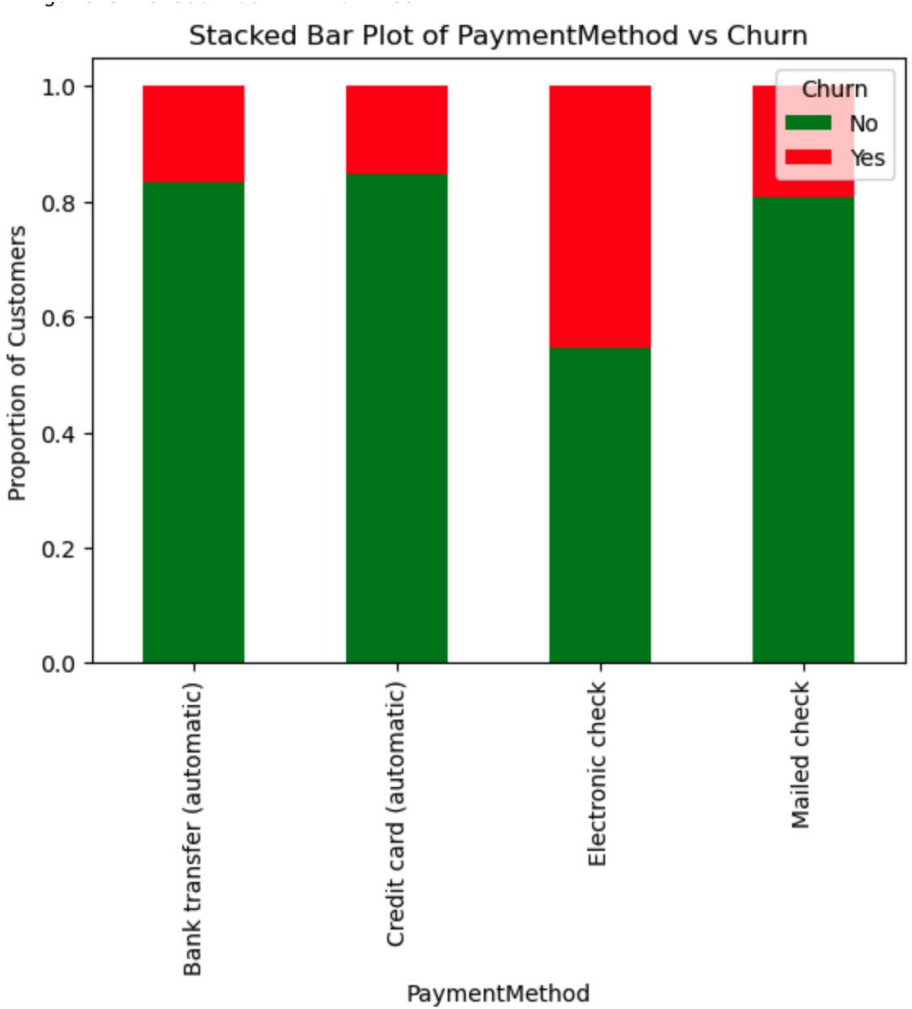
Churn seems to occur more with customers with shorter history with this Telco.



Churn seems to occurring at higher rate for customers on the internet service



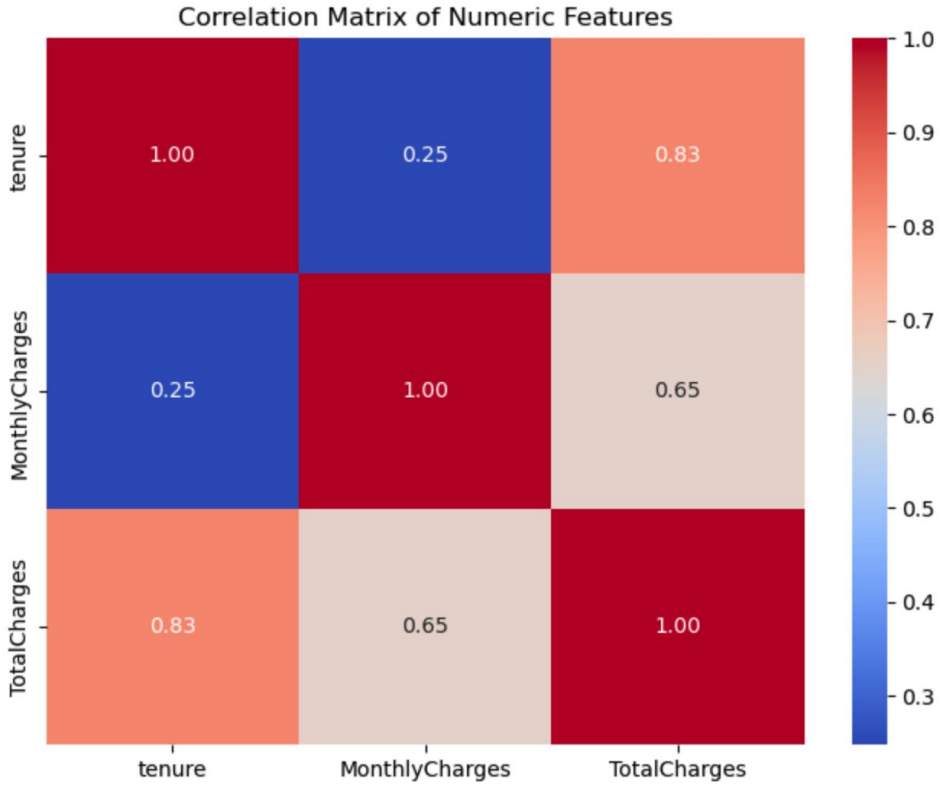
Significantly higher rate of churn among month-to-month customers



While automated payment options show less churn rate. Among the manual payment group, Electronic Check show much bigger churn rate compared to Mailed Check.

Perhaps indicating digital friendly customers eagerness to find better deals and change providers.

Correlation Heatmap



Data Preprocessing

Numeric Conversion

'TotalCharges' were converted to numeric values, with billions indicated by 'B' transformed into corresponding numeric values.

Handling Missing Values

Missing values in the 'TotalCharges' column were filled with zeros, because when 'TotalCharges' was missing, 'tenure' was always 0, indicating that the customers were still in their 1st month with the contract, which meant no charge has accrued.

Categorical Data Consistency

Checked for unique values in each categorical column to ensure consistency.

Check for Outliers

Checked 'MonthlyCharges' for outliers and found none.

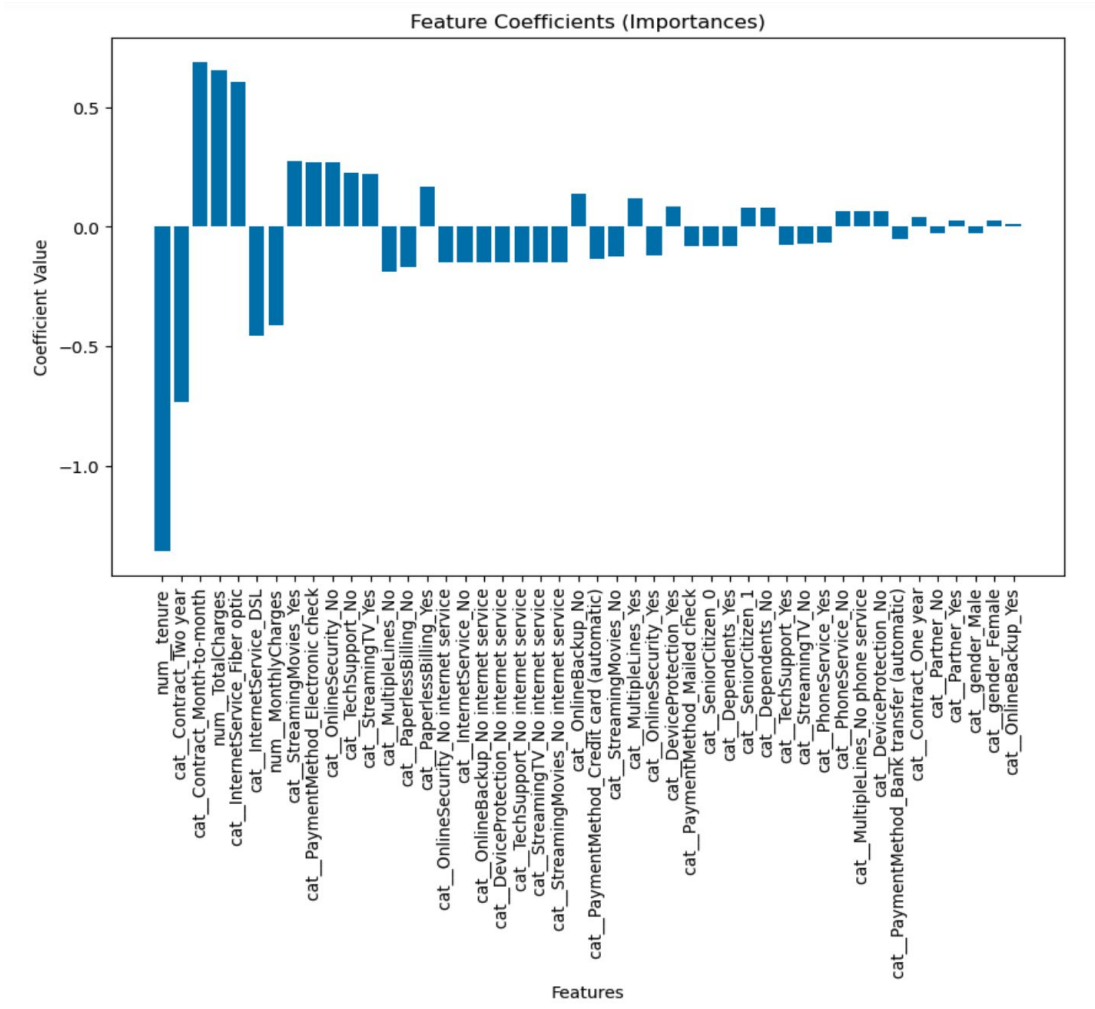
Model Selection and Evaluation

Initial benchmark model was derived using LogisticRegression model. It yielded 0.82 Accuracy and 0.86 in AUC ROC, which were solid results.

Machine Learning models often do even better, so I tried Random Forest and GBM as well.

- **Random Forest Regressor**: Initial try with 100 trees resulted in 0.79 Accuracy, which was lower that LogRegression. Tried RandomizedSearch with a limited set of 5 folds for 10 candidates, raised the accuracy to 0.81 and 0.86 AUC ROC. This was still less accurate than the LogRegression model, but showed promised that with higher compute power, we could perform more thorough GridSearch for better performance.

- **Gradient Boosted Models**: is typically known to be effective for classification problem like churn, but the initial run with 100 trees also yielded, ~0.81 in accuracy, which was lower tha LogRegression. Similarly to the RF model, I tried a RandommizedSearch with 5 folds for 10 candidates (50 fits), but wasn't able to raise the accuracy in this case. Again, it's worth retrying this with higher compute power.


Feature Coefficients (Importances)
Coefficient Value
0.5
0.0
−0.5
−1.0
num__tenure
cat__Contract_Two year
cat__Contract_Month-to-month
num__TotalCharges
cat__InternetService_Fiber optic
cat__InternetService_DSL
num__MonthlyCharges
cat__StreamingMovies_Yes
cat__PaymentMethod_Electronic check
cat__OnlineSecurity_No
cat__TechSupport_No
cat__StreamingTV_Yes
cat__MultipleLines_No
cat__PaperlessBilling_No
cat__PaperlessBilling_Yes
cat__OnlineSecurity_No internet service
cat__InternetService_No
cat__OnlineBackup_No internet service
cat__DeviceProtection_No internet service
cat__TechSupport_No internet service
cat__StreamingTV_No internet service
cat__StreamingMovies_No internet service
cat__OnlineBackup_No
cat__PaymentMethod_Credit card (automatic)
cat__StreamingMovies_No
cat__MultipleLines_Yes
cat__OnlineSecurity_Yes
cat__DeviceProtection_Yes
cat__PaymentMethod_Mailed check
cat__SeniorCitizen_0
cat__Dependents_Yes
cat__SeniorCitizen_1
cat__Dependents_No
cat__TechSupport_Yes
cat__StreamingTV_No
cat__PhoneService_Yes
cat__PhoneService_No
cat__MultipleLines_No phone service
cat__DeviceProtection_No
cat__PaymentMethod_Bank transfer (automatic)
cat__Contract_One year
cat__Partner_No
cat__Partner_Yes
cat__gender_Male
cat__gender_Female
cat__OnlineBackup_Yes
Features

Findings and Insights

Long customer tenure and 2 year contracts show strong negative correlation to churn, while month-to-month is strongly correlated to churn.

Also customers with DSL and FiberOptic internet services are highly correlated to churn while movie streaming is negatively correlated.

Recommendations

Push Multi-year and Renewal Promotions: This Telco should consider all means it has to encourage long-term commitment from the customer, and move customers away from month-to-month plans as much as possible.

Internet Service Products: This firm also should reevaluate its competitiveness as an ISP, as customers using its internet service are churning more.

Movie Streaming: This service seems to keep customers to this Telco. Consider promoting this service more strongly.