Customer Churn Analysis Report

Introduction

Customer retention is a top priority for telecom companies due to the high cost of acquiring new customers. The goal of this project is to analyze customer behavior and service patterns to identify what drives churn (the voluntary departure of customers) and to build a predictive model that can help the business act proactively.

Using the Telco Customer Churn dataset, I explored trends, built a churn prediction model using logistic regression, and uncovered actionable insights that can inform customer success strategies and reduce attrition.

Project Objectives

- ✓ Understand customer attributes that correlate with churn.
- ✓ Explore trends and patterns in customer behavior.
- ✓ Build a predictive model to identify customers likely to churn.
- ✓ Deliver actionable recommendations to improve customer retention.

Dataset Overview

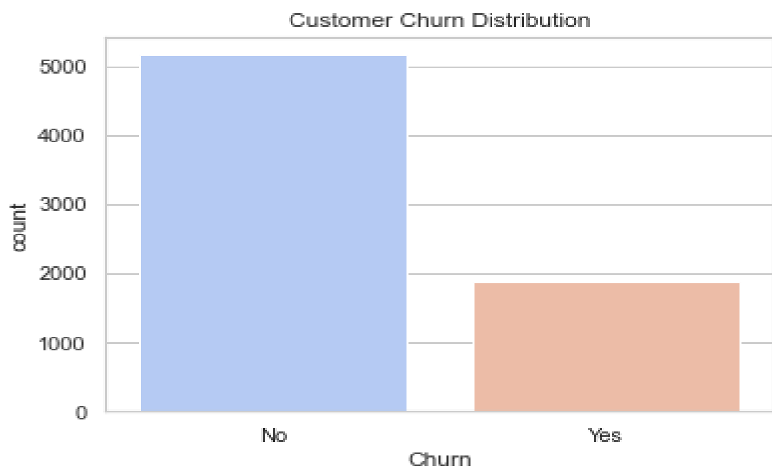
Dataset: Telco Customer Churn

- ✓ Rows: 7,043 customers
- ✓ Target: Churn (Yes/No)
- ✓ Features: Demographic info, services used, contract type, billing details

Exploratory Data Analysis (EDA)

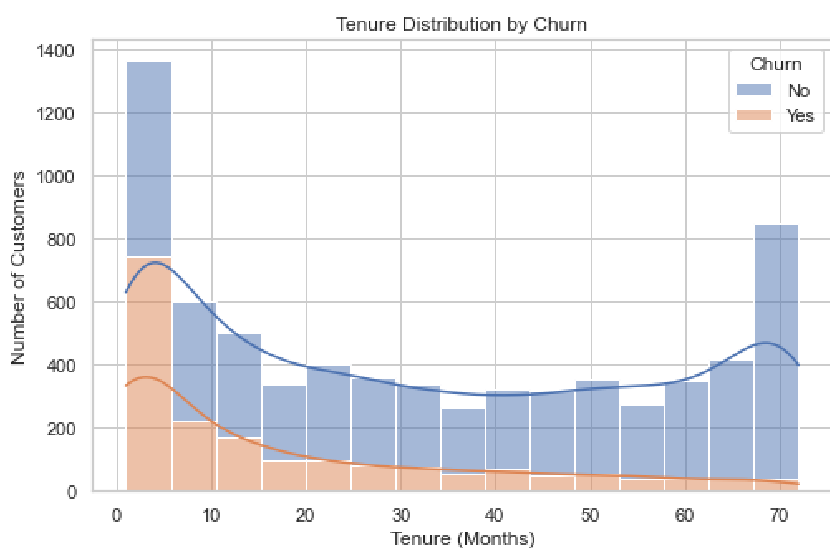
The EDA phase helped uncover patterns and relationships between features and churn behavior. Here's what was explored and discovered:

a. Churn Distribution



Insight: About 26.5% of customers in the dataset churned.

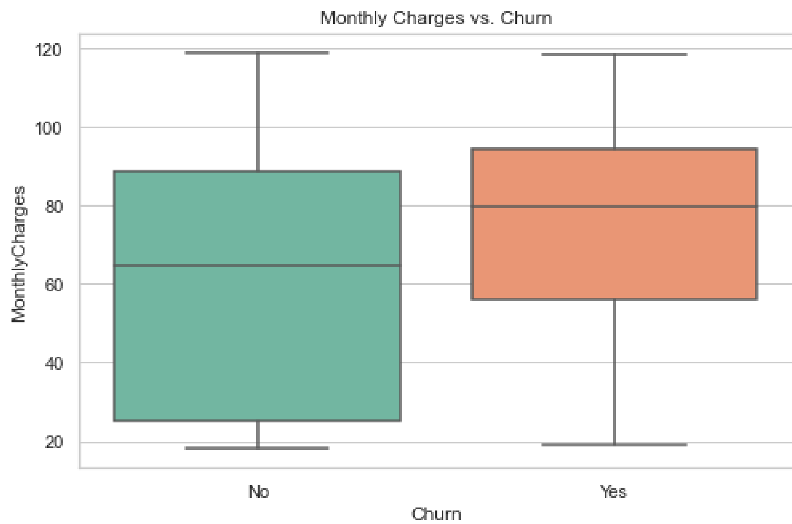
b. Tenure and Churn



Insight:

- ✓ Customers with short tenure (<12 months) are far more likely to churn.
- ✓ Churn rate significantly drops as tenure increases, showing the importance of early customer engagement.

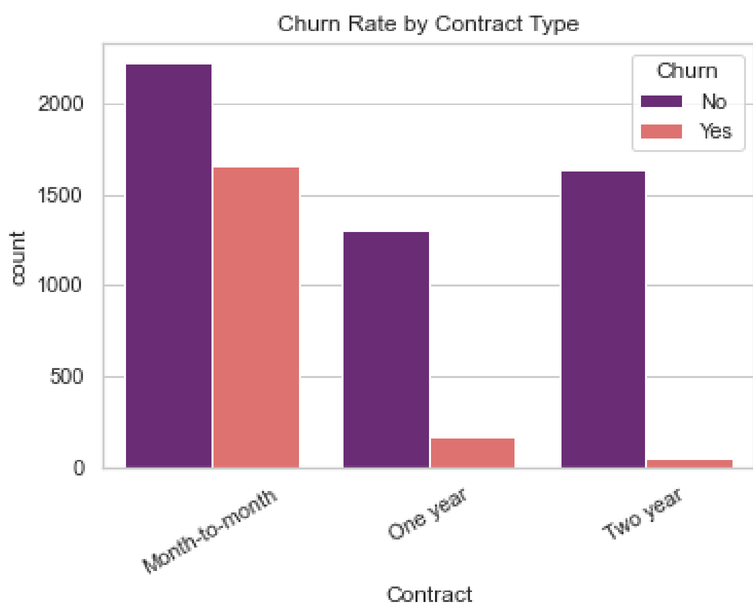
c. Monthly Charges and Churn



Insight:

- ✓ Churned customers have higher monthly charges on average.
- ✓ Pricing models may be contributing to dissatisfaction for high-paying users.

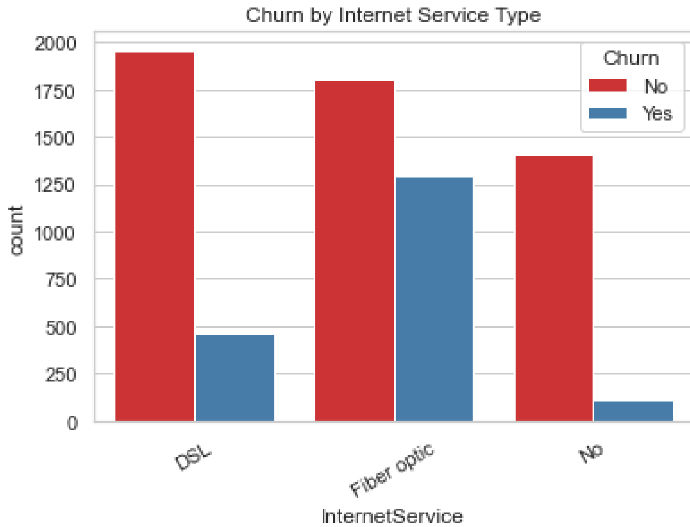
d. Contract Type and Churn



Insight:

- ✓ Customers with month-to-month contracts churn much more than those on one- or two-year contracts.
- ✓ Long-term contracts appear to stabilize retention.

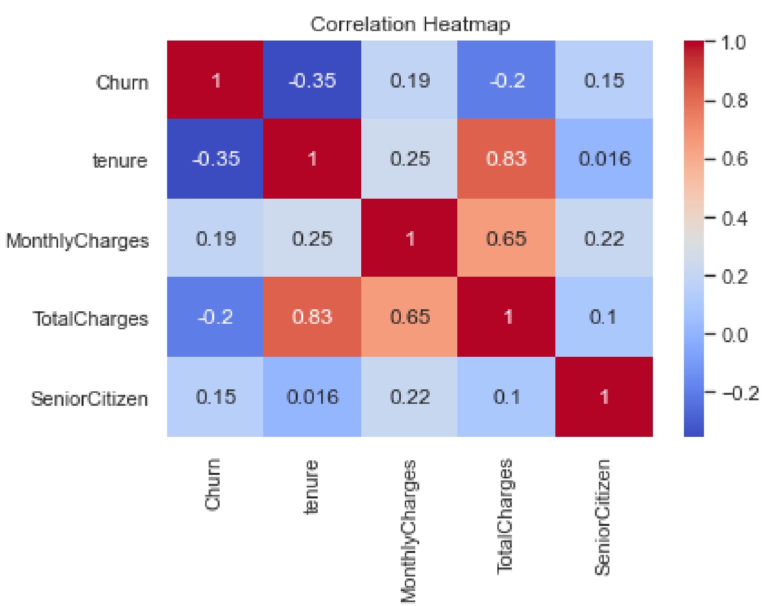
e. Internet Service Type and Churn



Insight:

- ✓ Customers using fiber optic internet show higher churn than DSL users.
- ✓ Service reliability or pricing could be factors here.

f. Correlation Matrix



Insight:

- ✓ Tenure has a strong negative correlation with churn.
- ✓ MonthlyCharges and TotalCharges show weaker but positive correlations.

Model Development

a. Model Used: Logistic Regression
   - ✓ Trained on one-hot encoded data
   - ✓ Scaled numerical features using StandardScaler
   - ✓ Split dataset: 70% train, 30% test

b. Model Performance

Metric	**Score**
Accuracy	~78%
Precision	~72%
Recall	~67%
F1-Score	~69%
AUC Score	~83%

> Note: High recall is vital (better to flag potential churners even if some are false positives).

c. Model Optimization

Used GridSearchCV to tune C (regularization strength) and penalty type.

Applied L1 (Lasso) regularization to reduce feature noise and improve interpretability.

TASK: Optimize the Model and Interpret Feature Importance

Objective:

Improve your churn prediction model's performance using hyperparameter tuning and identify the most impactful features driving churn.

1. Import required libraries

In [2]:
```
import pandas as pd
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report
```

2. Load your data

In [3]:
```
df = pd.read_csv('telco_cleaned_for_analysis.csv')
```

3. Encode the target

In [4]:
```
df['Churn'] = df['Churn'].map({'Yes': 1, 'No': 0})
```

4. Prepare X and y

In [5]:
```
X = df.drop(columns=['customerID', 'Churn'], errors='ignore')
y = df['Churn']
```

5. One-hot encode X

In [6]:
```
X = pd.get_dummies(X, drop_first=True)
```

6. Split BEFORE scaling

In [7]:
```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42, stratify=y
)

```

7. Scale AFTER split

In [8]:
```
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

8. Confirm shapes match

In [9]:
```
print("X_train_scaled shape:", X_train_scaled.shape)
print("y_train shape:", y_train.shape)


```

```
X_train_scaled shape: (4922, 30)
y_train shape: (4922,)
```

9. Set up and run GridSearchCV

In [10]:
```
model = LogisticRegression()
params = {
    'C': [0.01, 0.1, 1, 10, 100],
    'penalty': ['l1', 'l2'],
    'solver': ['liblinear']
}

grid = GridSearchCV(model, param_grid=params, cv=5, scoring='f1', n_jobs=
grid.fit(X_train_scaled, y_train)
```

Out[10]:
```
GridSearchCV(cv=5, estimator=LogisticRegression(), n_jobs=-1,
             param_grid={'C': [0.01, 0.1, 1, 10, 100], 'penalty': ['l1', 'l
2'],
                         'solver': ['liblinear']},
             scoring='f1')
```

10. Evaluate

In [12]:
```
best_model = grid.best_estimator_
print("Best parameters:", grid.best_params_)

y_pred = best_model.predict(X_test_scaled)
print(classification_report(y_test, y_pred))

```

```
Best parameters: {'C': 0.01, 'penalty': 'l2', 'solver': 'liblinear'}
              precision    recall  f1-score   support

           0       0.85      0.88      0.87      1549
           1       0.64      0.58      0.61       561

    accuracy                           0.80      2110
   macro avg       0.75      0.73      0.74      2110
```

Feature Importance

Top Predictive Features (by absolute coefficient size):

1. Contract_Month-to-month — Strong positive impact on churn
2. TechSupport_No — High risk of churn
3. tenure — Longer tenure reduces churn
4. OnlineSecurity_No — Higher risk without this service
5. PaperlessBilling_Yes — Slightly increases churn likelihood

8. Interpretation Coefficient

In [10]:
```python
# Get feature importance
coefficients = pd.Series(model.coef_[0], index=x.columns)
important_features = coefficients.sort_values(key=abs, ascending=False)
print("Top Predictive Features:")
print(important_features.head(10))
```

```
Top Predictive Features:
Contract_Two year                  -1.561469
Contract_One year                  -0.887144
InternetService_Fiber optic         0.821503
PhoneService_Yes                   -0.647363
TechSupport_Yes                    -0.474001
MultipleLines_No phone service      0.466385
PaymentMethod_Electronic check      0.445537
OnlineSecurity_Yes                 -0.421360
MultipleLines_Yes                   0.347389
Dependents_Yes                     -0.307221
dtype: float64
```

Business Recommendations

Based on the analysis and model findings, here's what I recommend:

1. Incentivize long-term contracts:
   Reduce churn by promoting yearly contracts through discounts or loyalty perks.
2. Target new customers aggressively:
   Focus on engagement within the first 12 months of service.
3. Upsell security & support services:
   Promote tech support and online security to high-risk groups.
4. Address high-billing customers' pain points:
   Investigate churn in customers with high monthly charges, especially on fiber plans.
5. Monitor electronic billing churn patterns:
   Consider A/B testing billing format (paperless vs. mailed invoices).

Conclusion

This project successfully explored customer churn behavior, developed a predictive model, and offered business strategies for reducing churn. With a recall-focused approach, the model can help proactively identify at-risk customers and prioritize retention efforts where they matter most.