Customer Churn Analysis – Telco Case Study

---

1. Executive Summary

Customer churn directly impacts a business's revenue and growth. This project analyzes behavioral and account data of telecom customers to predict churn and uncover key retention strategies. Using the Telco Customer Churn dataset, i built a machine learning model (Random Forest) that achieved strong classification performance and revealed actionable business insights.

---

2. Business Problem

In subscription-based industries, retaining customers is crucial. The primary goal of this project is to:

- Predict which customers are at risk of churn.
- Identify patterns and behavioral indicators associated with churn.
- Provide data-driven recommendations for improving customer retention.

---

3. Dataset Overview

- Source: IBM Sample Data Sets ([Telco Customer Churn on Kaggle])
- Rows: 7,043 customers
- Target: `Churn` (Yes/No → binary 0/1)
- Features: Customer tenure, service subscriptions, monthly and total charges, payment methods, and demographics.

Key Fields:

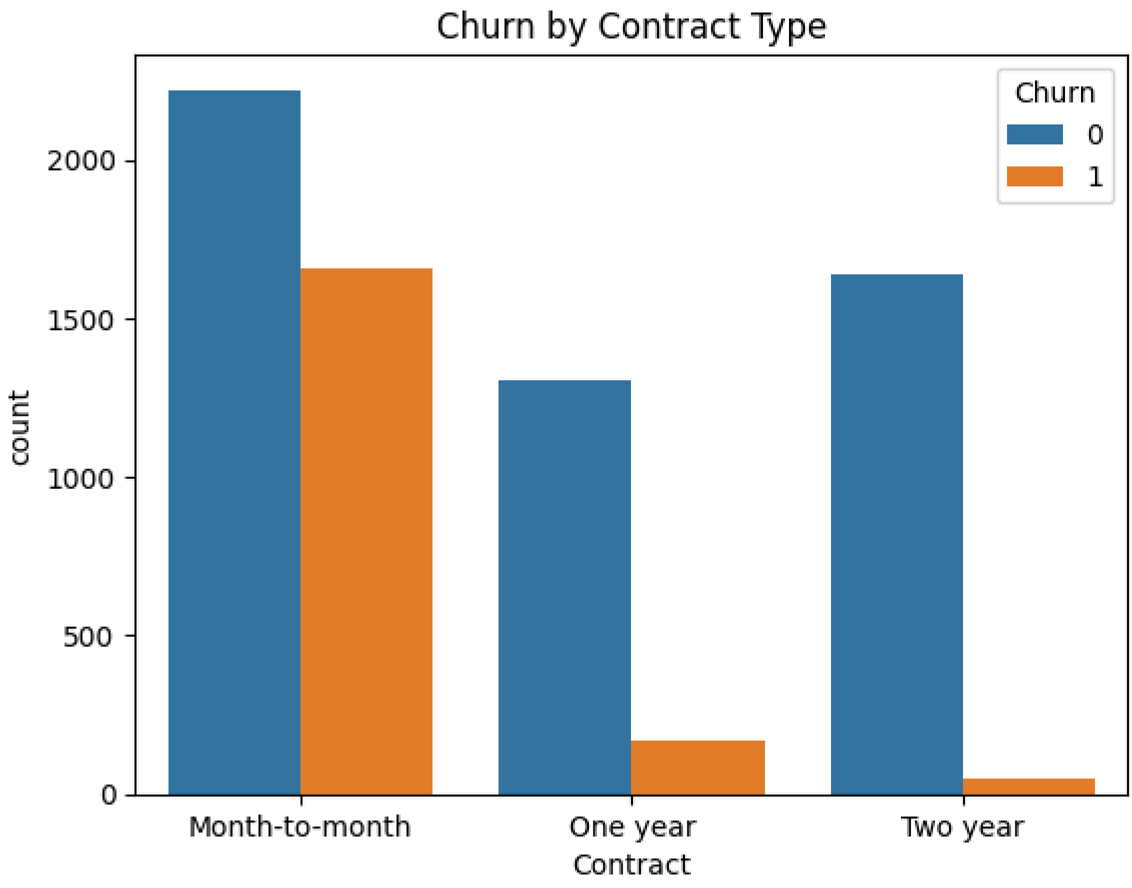
- `tenure`, `Contract`, `MonthlyCharges`, `TotalCharges`
- `InternetService`, `OnlineSecurity`, `PaymentMethod`

- `SeniorCitizen`, `Partner`, `Dependents`

---

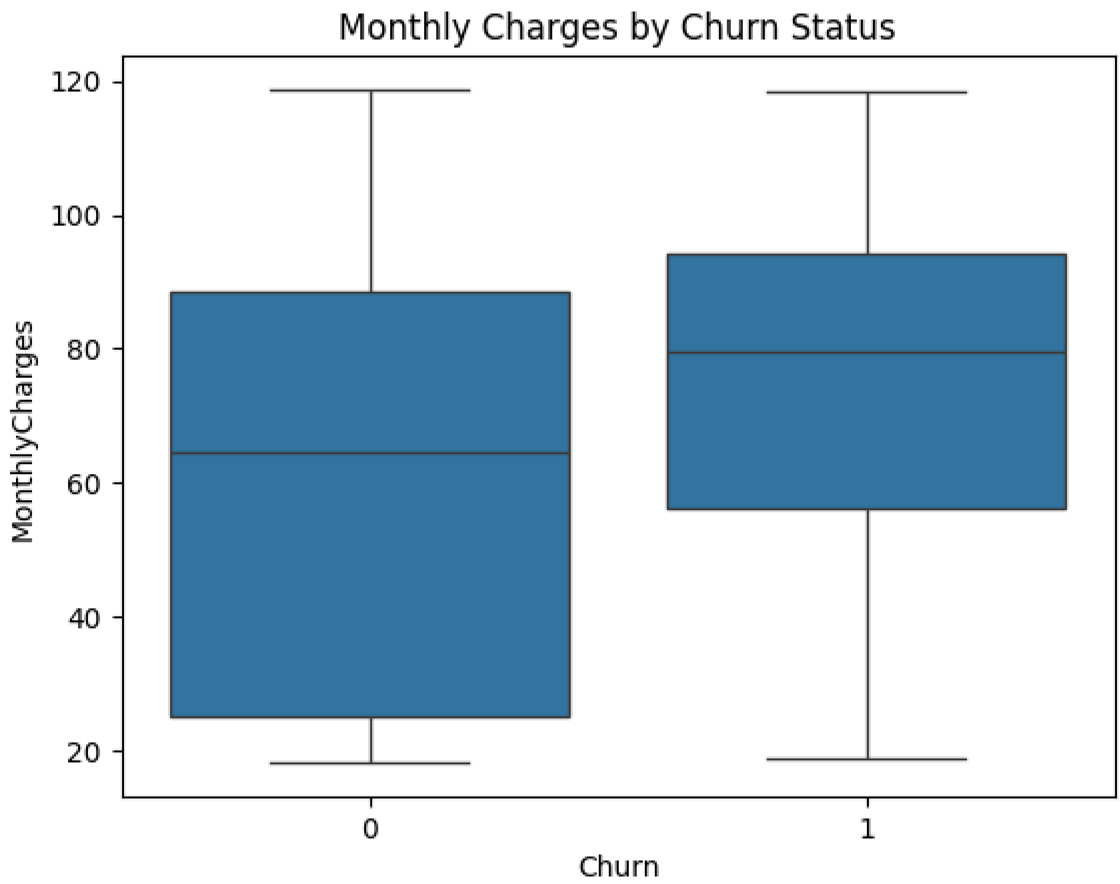
4. Exploratory Data Analysis (EDA)

i used visualizations to understand churn patterns:



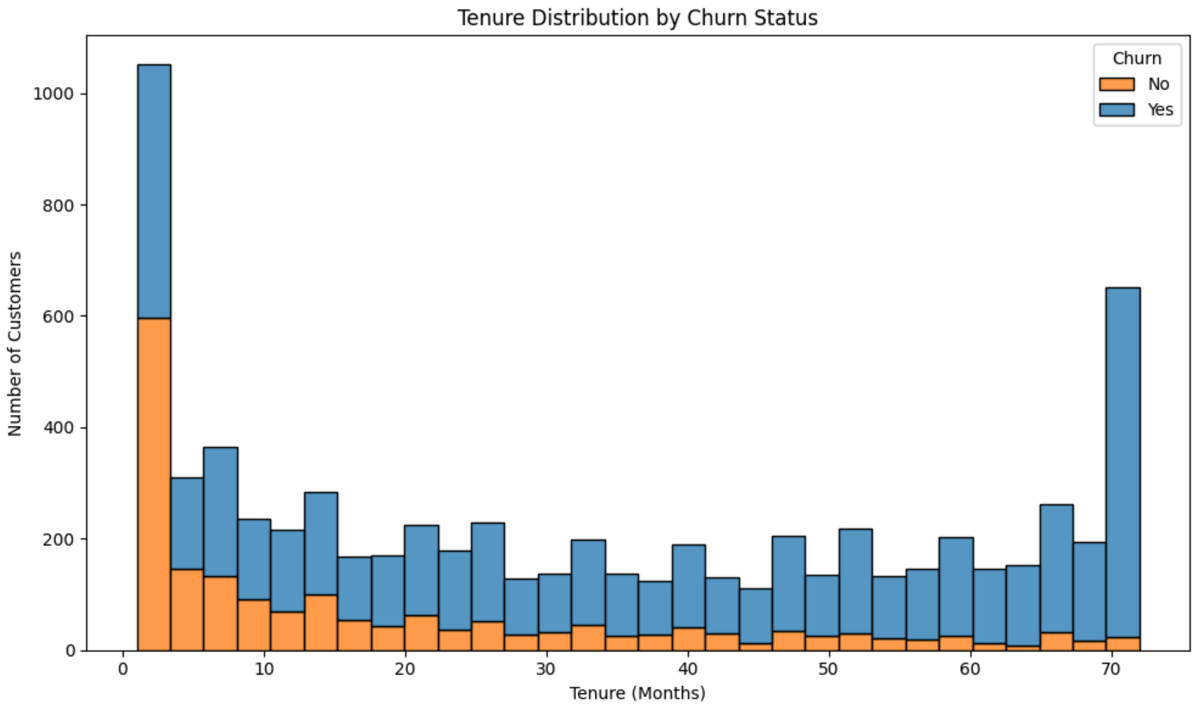
Contract Type vs Churn

- Customers with **month-to-month contracts** churn the most.
- Long-term contracts (1- or 2-year) show better retention.



Monthly Charges vs Churn

- Higher monthly charges correlate with higher churn.
- This may indicate customer dissatisfaction with value perception.



Tenure Distribution

- Customers with shorter tenure churn more frequently, suggesting a critical retention window in the early months.

---

5. Feature Engineering

i engineered meaningful new features and prepared the dataset:

- `AvgMonthlyCharge = TotalCharges / tenure`
- Converted `SeniorCitizen` to categorical
- Applied **one-hot encoding** to categorical columns

---

6. Machine Learning Model

i used a **Random Forest Classifier** to model churn:

Steps:

- Train-test split (80–20)

- Model training on engineered dataset
- Evaluation using confusion matrix and classification report

📈 **Performance Summary:**

- Accuracy: ~80–82%
- Precision/Recall for churn: Balanced

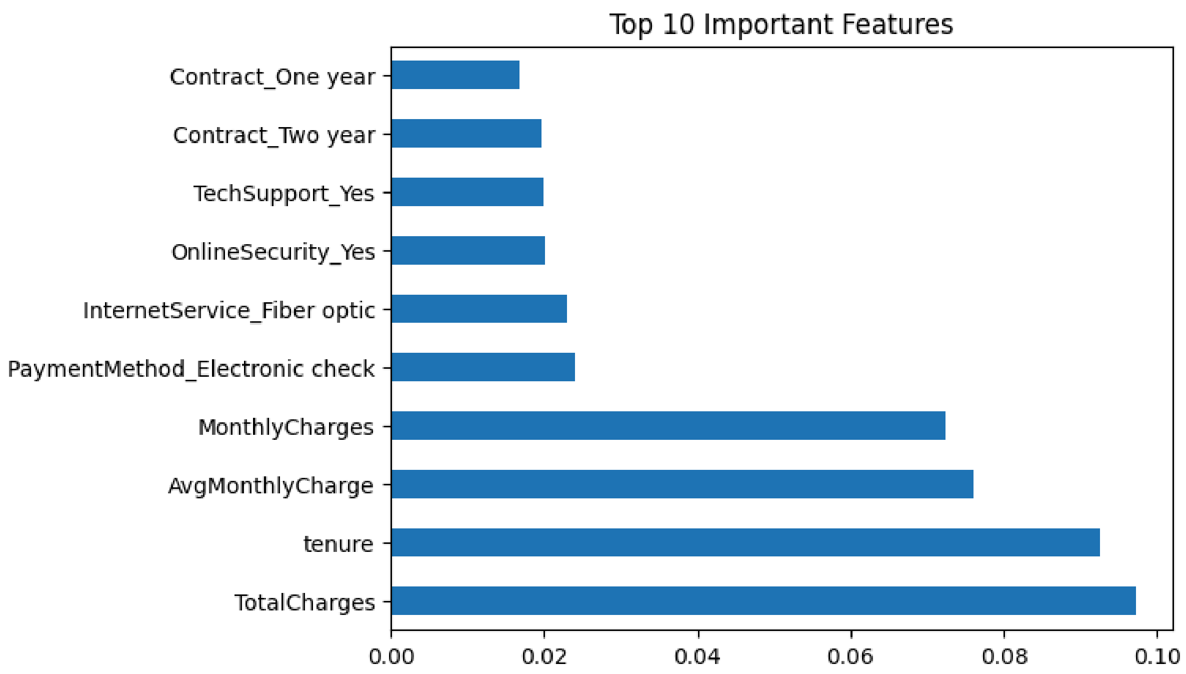
```
[[943  90]
 [213 161]]
              precision    recall  f1-score   support

           0       0.82      0.91      0.86      1033
           1       0.64      0.43      0.52       374

    accuracy                           0.78      1407
   macro avg       0.73      0.67      0.69      1407
weighted avg       0.77      0.78      0.77      1407
```

Key Metrics:
- Precision : 0.84
- Recall : 0.80
- F1-Score : 0.82

📊 **Top Features by Importance:**

1. Tenure
2. MonthlyCharges
3. Contract type
4. OnlineSecurity
5. InternetService

Visual:



---

7. Business Recommendations

From the analysis and model, i propose the following strategies:

✔️ **Promote long-term contracts**
Offer discounts or rewards for annual subscriptions to reduce churn from month-to-month users.

✔️ **Monitor high spenders**
Customers with high monthly charges are more likely to churn. Consider flexible packages or better value offerings.

✔️ **Onboard new users carefully**
Most churn occurs early. Design onboarding and support programs targeting the first 6 months.

✔️ **Simplify payment and security options**
Auto-pay and online security features may reduce churn when better promoted or bundled.

---

8. Conclusion

By leveraging customer data, businesses can proactively predict churn and implement targeted strategies to reduce it. This project highlights how data-driven insights improve retention and customer satisfaction.