Customer Churn Prediction Using Machine Learning

Telco Customer Churn Analysis with Logistic Regression, Random Forest, and XGBoost

Gafar Mustopha

Data Analyst

August 2025

Problem Statement

Business Question:

- Identify factors influencing customer churn.
- Predict which customers are likely to leave.

Why It's Important:

- Losing customers increases acquisition costs.
- Retention improves revenue and customer lifetime value.

Dataset Overview

Rows: 7,043.   **Columns:** 21.   **Source:** Kaggle – Telco Customer Churn dataset

Key Feature Groups:

- ○ Demographics (gender, senior citizen, dependents)
- ○ Service details (Internet type, security, tech support)
- ○ Account info (contract, payment method, paperless billing)
- ○ Financials (monthly & total charges)
- ○ Target: Churn (Yes/No)

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	Str
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No	

5 rows × 21 columns

Data Cleaning & Preparation

Handled missing values (TotalCharges column)

- Identified missing value in Total charges column
- Dropped the missing values
- Converted SeniorCitizen from (0/1) to Yes/No
- Encoded categorical variables for modeling
- Split into **80% training / 20% testing**

Exploratory Data Analysis (EDA)

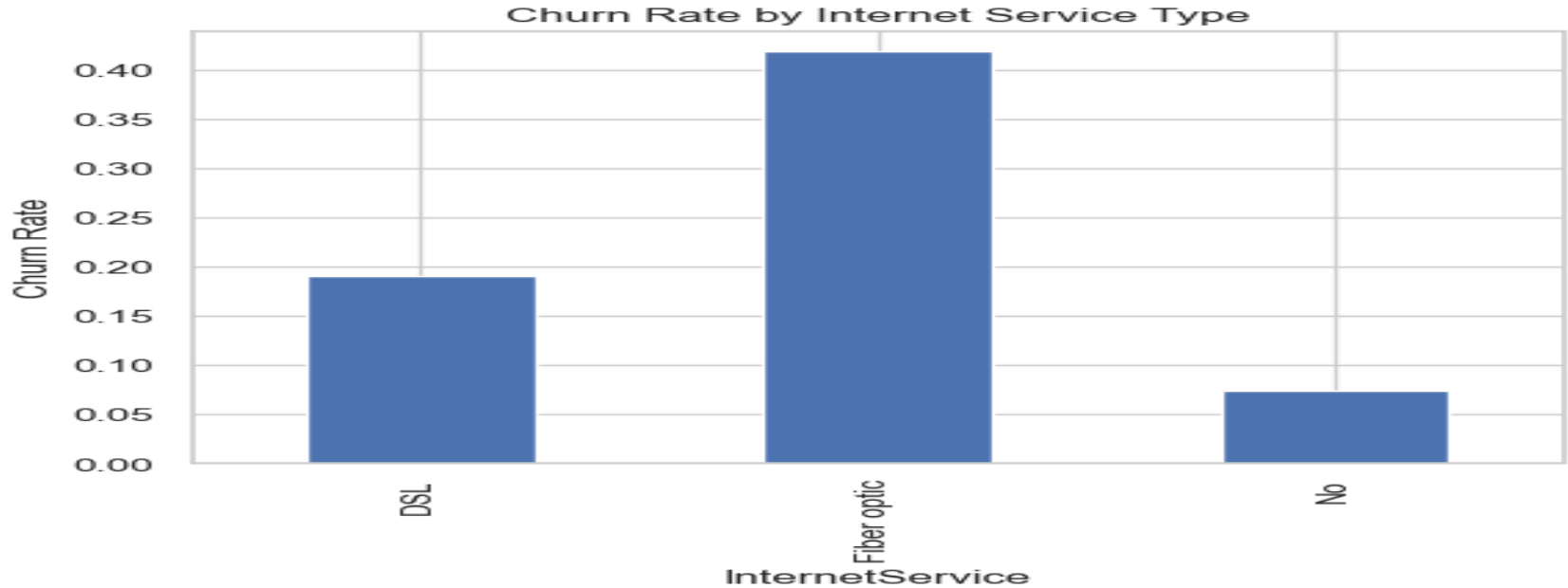**Churn Rate:** 26% – 27% → 1 in 4 customers leave annually

Business Impact: High churn increases costs due to expensive customer acquisition

Churn Rate by Contract Type
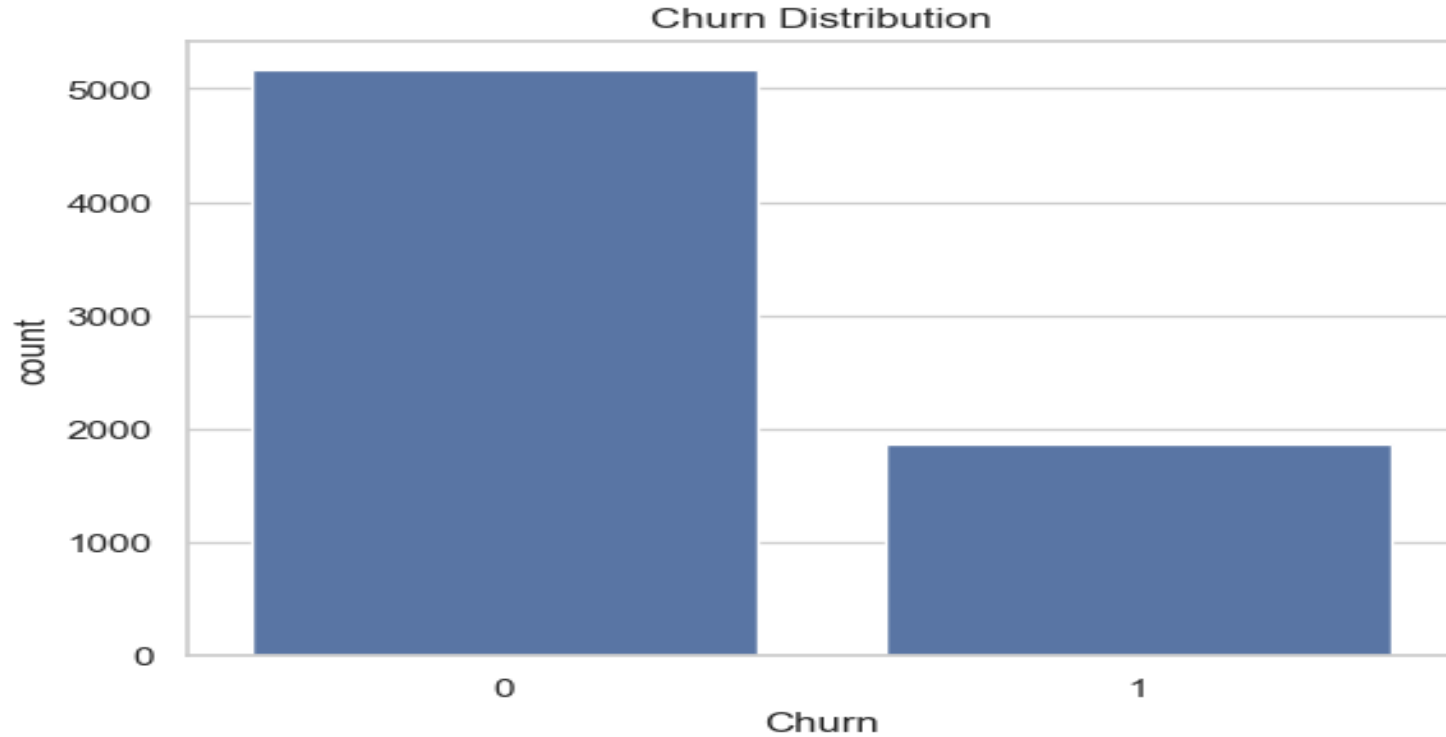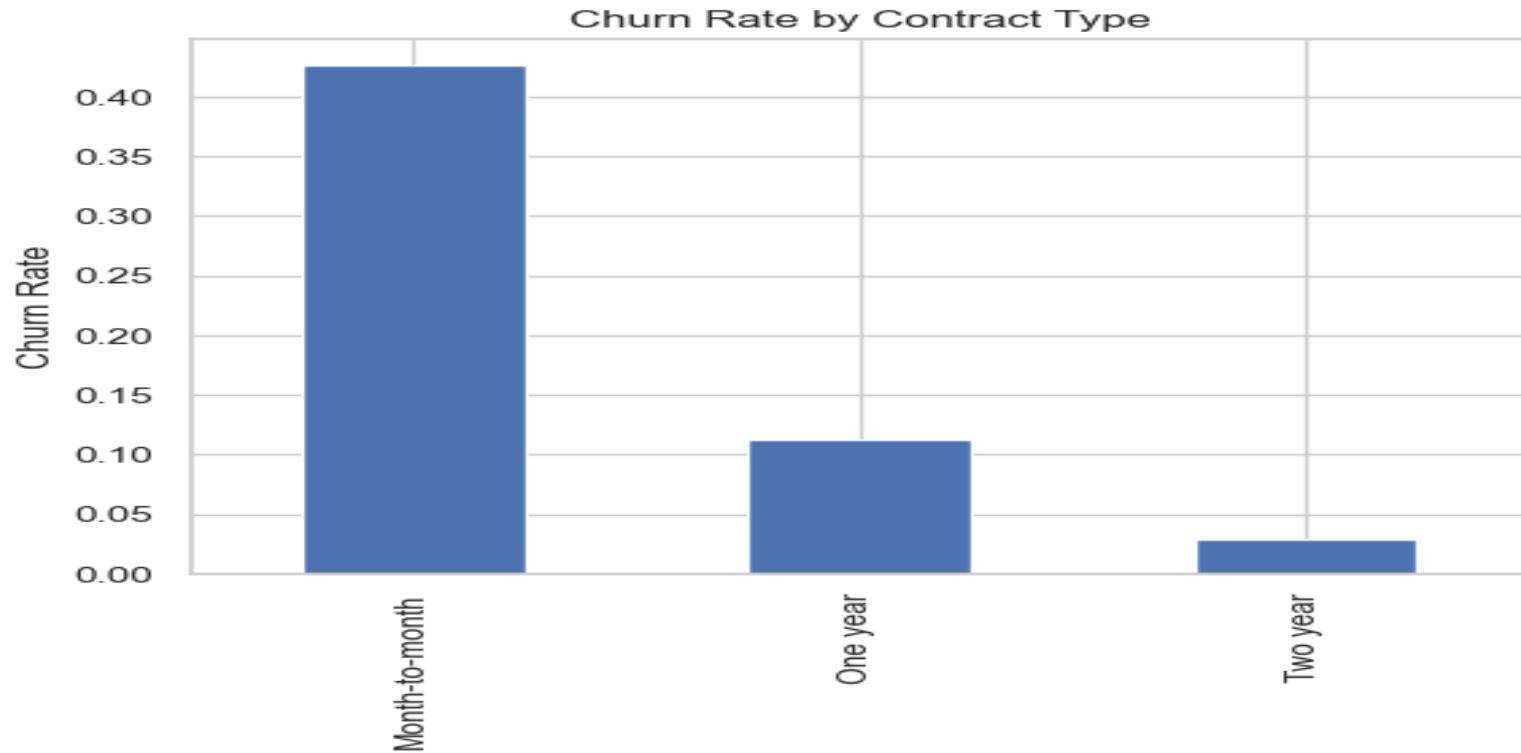
- Month-to-month → **Highest churn** (>40%)
- Long-term contracts → **Lower churn** (<15%)
- *Recommendation:* Offer incentives for long-term plans

Internet Service Type:

Churn Rate by Internet Service Type

Fiber optic users churn more than DSL users

*Possible reasons:* Price sensitivity, competition, service issues

Model Comparison

	Model	Accuracy	Precision	Recall	F1-score
0	Random Forest	0.785359	0.632353	0.459893	0.532508
1	XGBoost	0.737029	0.503984	0.676471	0.577626
2	XGBoost	0.737029	0.503984	0.676471	0.577626
3	XGBoost	0.737029	0.503984	0.676471	0.577626
4	XGBoost	0.737029	0.503984	0.676471	0.577626
5	Logistic Regression	0.786780	0.619355	0.513369	0.561404
6	Logistic Regression	0.786780	0.619355	0.513369	0.561404
7	XGBoost	0.737029	0.503984	0.676471	0.577626

XGBoost chosen for highest recall → best at catching churners.

SHAP Feature Impact (XGBoost)

SHAP Summary Plot: XGBoost Churn Prediction
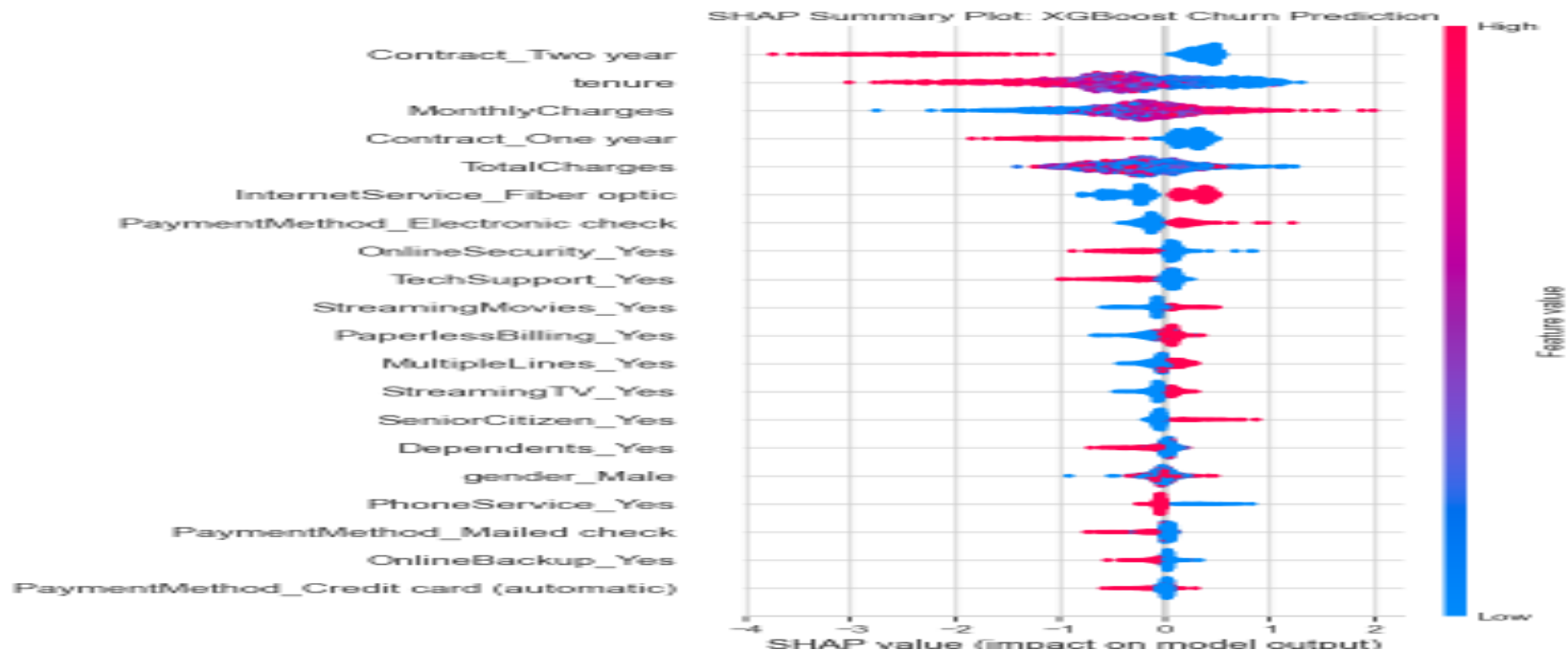
- Two-year contracts & long tenure → reduce churn risk
- High monthly charges & fiber optic internet → increase churn risk
- Paying by electronic check → increases churn risk
- Support services (online security, tech support) → reduce churn risk

Recommendations

- **Promote Long-Term Contracts:** Incentives for annual or two-year plans.

- **Target New Customers:** Focus onboarding efforts on first 6 months.

- **Address High-Bill Customers:** Offer loyalty discounts for high monthly charges.

- **Improve Fiber Optic Retention:** Service quality improvements or bundle deals.

- **Encourage Support Services:** Market online security and tech support packages.

Conclusion

- Churn rate ~26.6% → a significant business challenge.

- Key churn drivers: contract type, tenure, charges, and service type.

- Predictive model (XGBoost) enables proactive churn prevention.

- Data-driven retention strategies can **reduce churn and increase profitability**.

THANKS FOR LISTENING