Telco Customer Churn Analysis

Customer churn is a major issue in the telecom industry. This project uses machine learning and data analysis techniques to understand which factors lead to customer churn and build a predictive model to identify high-risk customers.

We used Python for data preprocessing, EDA, and modeling, and evaluated the model using accuracy, F1 score, and ROC-AUC metrics.

📁 Dataset Information

The dataset was sourced from IBM Sample Datasets and contains 7,043 customer records with the following features:

- `CustomerID`: Unique identifier (dropped in analysis)

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

- `Tenure`: Number of months with the company

- `PhoneService`, `MultipleLines`, `InternetService`, `OnlineSecurity`, etc.

- `Contract`, `PaymentMethod`, `MonthlyCharges`, `TotalCharges`

- `Churn`: Target variable (Yes/No)

We performed data cleaning, handled missing values, and converted categorical variables using one-hot encoding.

✅ Short Summary (Bullet Points)

--> 📊 Built a data-driven solution to predict customer churn for a telecom company.

--> 🔍 Applied data cleaning, exploratory analysis, and feature engineering.

--> 🧠 Trained and evaluated 3 models: Logistic Regression, Random Forest, and XGBoost.

--> 🎯 Achieved 80% accuracy and 0.84 ROC-AUC with Logistic Regression.

--> 📁 Delivered insights through visuals and saved all artifacts in a reusable format.

📊 **Auto-Generated Insights from Plots**

📌 1. Churn Distribution

About **26.6% of customers have churned**, showing a significant proportion of at-risk users. This confirms the business concern and validates the need for churn prediction.

📌 2. Contract Type vs Churn

Customers on **month-to-month contracts churn the most**, while those on **one or two-year contracts** are much more stable. Offering contract-based incentives may reduce churn.

📌 3. Tenure Distribution by Churn

Churn is **heavily concentrated in customers with tenure < 12 months**. Early-stage retention strategies such as onboarding programs or loyalty discounts can significantly reduce churn in this segment.

📌 4. Monthly Charges by Churn Status

Customers with **monthly charges over ₹90–₹100** are more likely to churn. Higher-cost plans may be causing dissatisfaction unless they deliver exceptional value.

📌 5. Internet Service Type vs Churn

Churn is **highest among Fiber optic users,** while customers with DSL or no internet churn less. This may reflect issues in service quality or expectations from high-speed users.

📌 6. Payment Method vs Churn

Customers who pay via **Electronic check** churn far more than those using **credit card or bank transfers**. This may signal payment flexibility or engagement differences.

🧠 Bonus: Model Insight

The **Logistic Regression model performed best**, achieving **80.2% accuracy** and **0.84 ROC-AUC**, making it the most reliable model for deployment or dashboard integration.

📌 7. Tech Support vs Churn

Customers **without tech support** are more likely to churn. This suggests dissatisfaction when help isn't readily available — offering bundled support services may improve retention.

📌 8. Online Security & Device Protection

Users **without online security or device protection** churn more often. Customers may leave due to concerns over safety or lack of add-on value.

📌 9. Streaming Services

Streaming TV and movies don't show strong correlation with churn, indicating they're **not primary drivers** of retention or dissatisfaction.

📌 10. Paperless Billing

Slightly higher churn observed in customers using **paperless billing**, possibly reflecting a more modern, digital-savvy user base who are quicker to switch providers.

📌 11. Senior Citizens

Senior citizens **do not churn more than younger users**, which may go against typical assumptions. The impact of age on churn appears minimal when other factors are considered.

📌 12. Total Charges vs Churn

Users with **lower total lifetime charges** churn more, likely because they're newer or lower-value customers. High lifetime value users are more stable and loyal.

📌 13. Multiple Services = Lower Churn

Customers subscribed to **multiple services** (phone, internet, security, backup) are less likely to churn, indicating bundled offerings improve stickiness.

✅ Bonus: Model Insight

- ✅ **Logistic Regression** delivered the best results:

- Accuracy: `80.2%`

- F1 Score: `0.605`

- ROC-AUC: `0.836`

- Simple, interpretable, and ready for production deployment or dashboard integration.

✅ Key Takeaways (Bullet Points)

--> ⚠️ Month-to-month contracts have the highest churn rate.

--> 💰 Customers with higher monthly charges are more likely to leave.

--> ⏳ Longer-tenure customers are more loyal and less likely to churn.

--> 🌐 Fiber-optic internet users showed a higher likelihood of churning.

--> 🎯 Early identification of high-risk customers enables targeted retention strategies.