Customer Churn Prediction Report

**1. Introduction**

Customer churn is a critical business challenge, especially in subscription-based industries like telecom, banking, and SaaS.  
Churn occurs when a customer **stops using a service**. High churn rates impact revenue and customer acquisition costs.

**Objective of this project**:

* Predict which customers are likely to churn.
* Identify key drivers of churn.
* Provide actionable insights to reduce churn and improve retention.

**2. Dataset Overview**

* **Source**: Telco Customer Churn dataset (Kaggle)
* **Rows**: 7,043 customers
* **Columns**: 21 features + target variable

**Key Columns**

* **Demographics**: gender, SeniorCitizen, Partner, Dependents
* **Services**: PhoneService, InternetService, StreamingTV, TechSupport
* **Billing**: Contract, PaymentMethod, MonthlyCharges, TotalCharges
* **Tenure**: tenure (number of months with company)
* **Target**: Churn (Yes/No)

**3. Data Quality Assessment**

**Missing Values**

* TotalCharges has missing values (to be imputed).
* No other major missing fields.

**Data Types**

* Numerical: tenure, MonthlyCharges, TotalCharges
* Categorical: gender, Contract, PaymentMethod, etc.
* Target: Churn → binary (Yes = churned, No = retained)

**Duplicates & Noise**

* Duplicates: checked and removed.
* Errors: TotalCharges stored as text (needs conversion).

**4. Initial Observations**

* Churn distribution appears **imbalanced** (roughly ~26–30% churners, 70–74% non-churners).
* Certain categories (like **month-to-month contracts**) seem more prone to churn.
* Tenure (length of stay) might be a **strong churn predictor**.

**5. Planned Methodology**

1. **Data Cleaning**
   * Handle missing values (TotalCharges)
   * Encode categorical variables
   * Create tenure groups
2. **Exploratory Data Analysis (EDA)**
   * Visualize churn distribution
   * Analyze churn by demographic, contract, billing, and services
3. **Feature Engineering**
   * Group tenure into buckets
   * Create binary flags for bundled services
4. **Modeling**
   * Logistic Regression (baseline)
   * Random Forest
   * XGBoost
5. **Evaluation Metrics**
   * Accuracy, Precision, Recall, F1, ROC-AUC
6. **Business Insights**
   * Identify key churn drivers
   * Provide recommendations

**6. Expected Deliverables**

* 📊 Interactive notebook with full workflow
* 📈 Visualizations highlighting churn patterns
* 🤖 Machine learning models with performance comparison
* 📝 Business report with actionable recommendations

A green and red pie chart

AI-generated content may be incorrect.

A graph of a number of different colored bars

AI-generated content may be incorrect.

A graph of different colored columns

AI-generated content may be incorrect.

A graph with green and red lines

AI-generated content may be incorrect.

A graph of a logistic regression

AI-generated content may be incorrect.

A graph of a forest

AI-generated content may be incorrect.

A graph of a blue and white box

AI-generated content may be incorrect.

A graph of a number of colored bars

AI-generated content may be incorrect.

A graph with different colored bars

AI-generated content may be incorrect.

A chart with red and green squares

AI-generated content may be incorrect.

A green and red rectangular bars

AI-generated content may be incorrect.