# CUSTOMER CHURN PREDICTION PROJECT REPORT

## 1. Introduction

Customer churn is a major challenge for subscription-based businesses, especially in the telecommunications sector. Retaining customers is significantly more cost-efficient than acquiring new ones, which makes churn prediction an essential analytical function.
This project focuses on analyzing customer behavior, identifying churn patterns, and building a machine learning model to predict whether a customer is likely to leave the service.

---

## 2. Business Problem Statement

The organization experiences a steady loss of customers each month, impacting revenue and long-term growth. To reduce churn, the company needs a reliable prediction system and a deeper understanding of churn drivers.

**The goal of this project is to:**

- Identify customer segments with high churn risk.
- Understand the factors that influence churn using detailed EDA.
- Build a predictive model to classify customers as “Churn” or “Not Churn.”
- Suggest actionable strategies to improve customer retention.

---

## 3. Objectives of the Project

- Perform detailed exploratory data analysis to understand patterns.
- Identify key factors influencing churn.
- Build and evaluate a machine learning model.
- Develop a user-friendly interface for churn prediction (Streamlit App).
- Provide business recommendations based on findings.

---

## 4. Dataset Overview

The dataset contains customer-level information including demographics, contract details, billing patterns, services subscribed, and churn status.

### Feature Groups

- **Demographic:** Gender, Partner, Dependents, SeniorCitizen
- **Services:** PhoneService, InternetService, Security, Backup, Streaming, Tech Support
- **Account Info:** Contract Type, Payment Method, Paperless Billing
- **Billing:** MonthlyCharges, TotalCharges
- **Target Variable:** Churn (Yes/No)
- Total Records: **7043**
- Total Features: **21**

---

## 5. Exploratory Data Analysis

This section provides an in-depth analysis of the customer dataset to identify patterns, trends, and relationships that influence churn.

---

### 5.1 Initial Observations

- MonthlyCharges range widely indicating diverse service packages.
- Tenure varies from new customers (0 months) to highly loyal ones (72 months).
- TotalCharges has zeros for new subscribers.
- Churn rate is **approximately 26.5%**, showing significant customer loss.

---

### 5.2 Summary Statistics

#### Numerical Feature Summary

| Feature | Min | Max | Mean | Comments |
|---|---|---|---|---|
| Tenure | 0 | 72 | 32.37 | Highly variable – clear separation of new vs loyal customers |
| MonthlyCharges | 18.25 | 118.75 | 64.76 | Higher charges often linked to churn |
| TotalCharges | 0 | 8684.80 | 2283.30 | Zero values show first-month subscribers |

#### Key Insights

- Significant portion of customers leave within the first few months.
- Higher billing customers tend to churn faster.
- Tenure is highly correlated with loyalty.

---

### 5.3 Distribution of Target Variable

**Described Visual:**
“A bar chart shows that approximately one-fourth of customers have churned. This imbalance indicates a need for careful model evaluation.”

---

### 5.4 Univariate Analysis (Feature-by-Feature)

#### Tenure

“A histogram reveals a large concentration of customers in the first 0–6 months, indicating onboarding issues.”

#### Monthly Charges

“Density plots show two major groups: low billers ($20–$45) and high billers ($70–$100). The higher billing group has a higher churn rate.”

#### Total Charges

“Customers with lower TotalCharges have higher churn rates, especially those below $500.”

---

### 5.5 Bivariate Analysis (Churn vs Features)

#### Churn vs Gender

“Both genders show almost identical churn rates, indicating gender does not influence churn."

#### Churn vs Senior Citizen

“Bar charts show senior citizens have significantly higher churn, possibly due to pricing issues or service complexity.”

#### Churn vs Partner & Dependents

“Customers with families (partners/dependents) churn less.”

#### Churn vs Contract Type

“One of the clearest patterns:

- Month-to-Month → Highest churn
- One-Year → Moderate
- Two-Year → Lowest”

**Churn vs Payment Method**

“Electronic check users show very high churn, while automatic payments (credit card/bank transfer) reduce churn.”

**Churn vs Paperless Billing**

“Paperless billing users churn more, strongly linked with flexible contracts.”

**Churn vs Internet Service**

“Fiber optic users churn more, suggesting dissatisfaction with network reliability.”

**Churn vs Security and Tech Services**

“For all services (TechSupport, OnlineSecurity, Backup, Protection), customers who opt in show reduced churn.”

---

**5.6 Multivariate Analysis**

**Correlation Heatmap (Described)**

“A heatmap showed:

- Strong positive correlation: Tenure ↔ TotalCharges
- Weak but positive: MonthlyCharges ↔ Churn
- Negative: Tenure ↔ Churn
- Contract influences churn indirectly through tenure”

**Monthly Charges + Tenure Interaction**

“3D scatter plot shows customers with high charges AND low tenure churn the most.”

**Contract + Billing + Churn**

“Visuals confirm: high monthly charges + month-to-month contract = churn hotspot.”

---

**5.7 Hidden Patterns Identified**

- New customers are most vulnerable to churn.
- Fiber optic customers churn more despite being premium service users.
- Add-on services (Tech, Security) significantly reduce churn.
- Families (partners/dependents) show more loyalty.
- High billing with no long-term contract = highest churn group.

---

### 5.8 Summary Table of EDA Findings

| Feature | Finding |
|---|---|
| Tenure | Strong negative correlation with churn |
| Contract Type | Strong influence on churn |
| MonthlyCharges | Higher charges → higher churn |
| Payment Method | Electronic check users churn most |
| Additional Services | Reduce churn |
| Internet Type | Fiber customers churn more |

## 6. Data Preprocessing

- Converted TotalCharges to numeric
- Handled blank values
- Encoded categorical features
- Standardized inputs
- Split data into train-test sets
- Saved model and encoders for deployment

---

## 7. Machine Learning Model

### Algorithms Tested

- Logistic Regression
- Random Forest Classifier
- SVM
- Gradient Boosting

**Best Performer:** Random Forest

---

### Model Results (Text Format)

- **Accuracy:** 82.4%
- **Precision (Churn Class):** 0.69
- **Recall (Churn Class):** 0.71
- **F1 Score:** 0.70

“Model shows strong recall, meaning it catches most churners — useful for retention strategies.”

---

## 8. Feature Importance

Described Visual:
“A bar plot ranked Contract, Tenure, MonthlyCharges, OnlineSecurity, and TechSupport as the most important predictors.”

---

## 9. Streamlit App

The application includes:

**Input Form Section**

- Dropdowns for categorical features
- Sliders for numerical values
- Clean layout for entering customer details

**Prediction Section**

- Displays churn probability
- Shows whether customer is “Likely to Churn”

**Visualization Section**

Described visuals included:

- Tenure distribution chart
- Churn by contract type
- Billing analysis

---

## 10. Business Recommendations

- Improve early customer onboarding to reduce first-month churn.
- Encourage long-term contracts with incentives.
- Review fiber optic service performance.
- Promote security/tech support add-ons.
- Target high billing customers with discounts.
- Introduce loyalty benefits to reward tenure.
- Focus on customers who use electronic checks — introduce easier payment options.

---