# Telco Customer Churn Prediction Report

## 1. Project Overview

Customer churn is a major challenge for telecom companies as it directly affects revenue and brand reputation.
This project aims to develop a predictive model that identifies customers at high risk of leaving the company (churn) and provides actionable strategies for customer retention.

## 2. Business Problem

A large telecom company experiencing high customer churn (around 25% annually). Losing customers directly impacts revenue, and senior management is under pressure to reduce churn. They want a predictive system to identify customers most at risk of leaving so the retention team can proactively intervene. They also want actionable recommendations on how to retain these customers, including example personalized communication strategies they can deploy.

**Objectives:**

- Clean and preprocess customer data (demographics, tenure, billing, complaints).
- Perform EDA to uncover key churn drivers with clear visuals.
- Build and evaluate a classification model to predict churn risk.
- Analyze feature importance to identify actionable insights.
- Propose retention strategies with example personalized messages.

## 3. Data Requirements

The project uses customer-level data containing:

- Demographics: Gender, SeniorCitizen, Partner, Dependents
- Account details: Tenure, Contract, PaymentMethod, MonthlyCharges, TotalCharges
- Services: InternetService, OnlineSecurity, TechSupport, StreamingTV, etc.
- Target: Churn (Yes/No)

Dataset Source: Kaggle - Telco Customer Churn Dataset.

## 4. Data Collection

The dataset was downloaded from Kaggle and contains 7,043 rows and 21 columns representing customer attributes related to demographics, service usage, and billing.

## 5. Data Validation

- Verified data types using df.dtypes and df.info()
- Confirmed no duplicate records using df.duplicated().sum().
- Checked missing values using df.isnull().sum().
- Converted TotalCharges from object to numeric.
- Used IQR method to check for outliers (only SeniorCitizen flagged, but invalid).

## 6. Data Profiling

Profiled dataset structure and key statistics using:
df.head(), df.tail(), df.sample(), df.shape, df.columns, df.nunique(), df.describe(), df.count(), df.isnull().sum()
This helped understand data quality, value distributions, and missing values.

## 7. Data Cleaning and Preprocessing

Cleaning Steps:

- Removed extra spaces in column names.
- Converted TotalCharges to numeric and handled missing values.
- Verified no duplicate records.

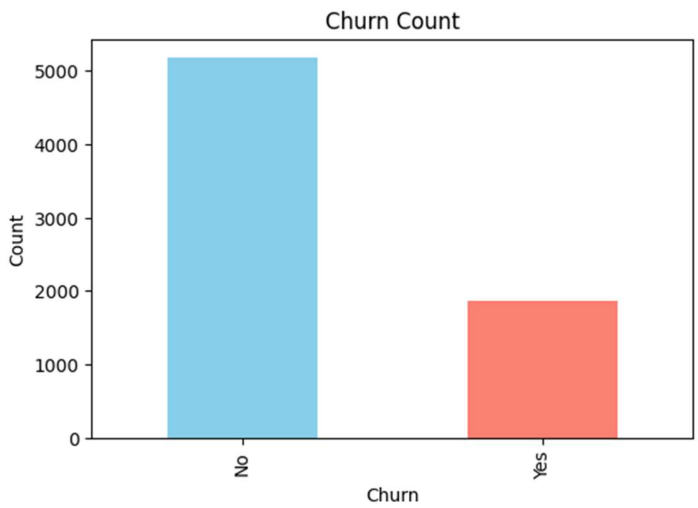
Preprocessing Steps:

- Encoded categorical columns using One-Hot Encoding.
- Scaled numerical columns (tenure, MonthlyCharges, TotalCharges) using StandardScaler.
- Created tenure buckets (0–12, 13–24, etc.) to capture customer loyalty.

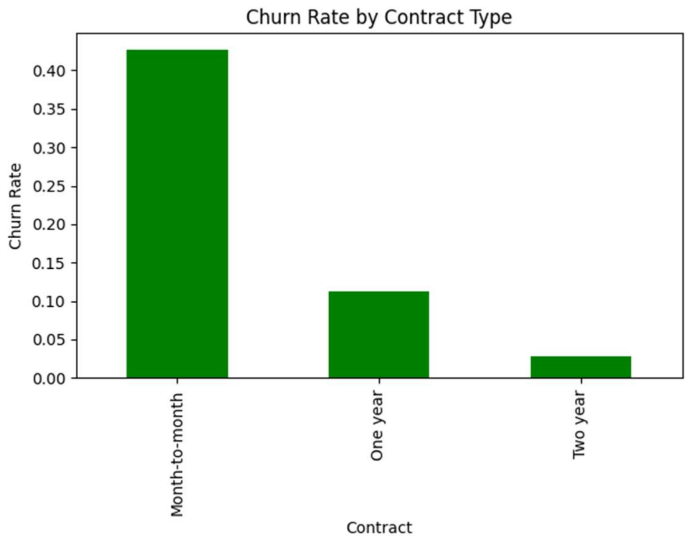
## 8. Exploratory Data Analysis (EDA)

Key Insights:

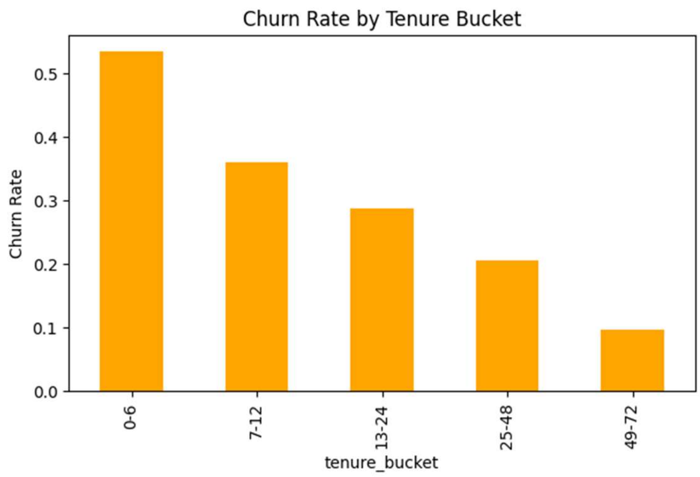
- 26.5% customers churned.



- Month-to-month contracts show highest churn rate.




- Customers with short tenure and high monthly charges are more likely to churn.



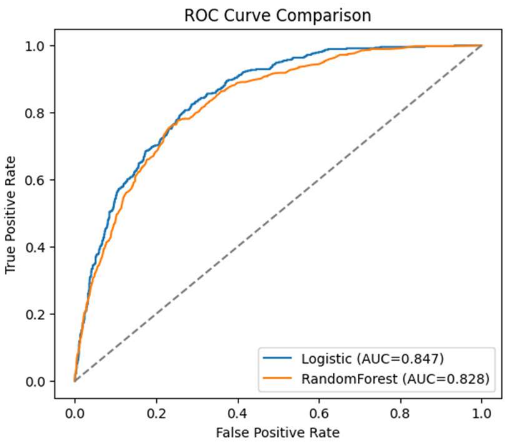

- Lack of online security or tech support strongly correlates with churn.

## 9. Model Building and Evaluation

Two models were built:
1. Logistic Regression (after scaling)
   - Accuracy: 0.8075
   - ROC AUC: 0.8467

2. Random Forest
- Accuracy: 0.7888
- ROC AUC: 0.8284

Logistic Regression performed slightly better and was chosen as the final model.

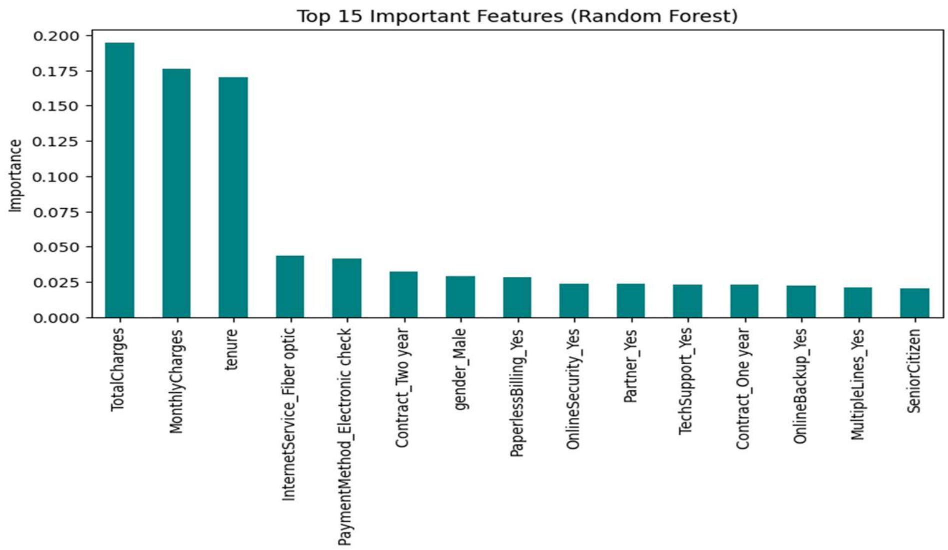



## 10.Feature Importance Analysis

After training the models, feature importance was analyzed to understand which factors most strongly influence customer churn.

- In the Random Forest model, the top features were:
- Contract type – Month-to-month contracts had the highest impact on churn.
- Tenure – Customers with shorter tenure were more likely to leave.
- MonthlyCharges – Higher monthly charges increased churn risk.
- OnlineSecurity and TechSupport – Lack of these services significantly contributed to customer loss.

This analysis helps the company understand which features to focus on when planning retention strategies.




## 11.Identifying High-Risk Customers

The trained model was then used to predict the probability of churn for each customer. Customers with a churn probability above 0.6 were considered high-risk.
By identifying these customers early, the retention team can take targeted actions such as:

- Offering discounts for contract renewal.
- Providing personalized assistance for billing or service issues.
- Encouraging the use of value-added services like online security.

This approach ensures resources are used efficiently by focusing only on customers who are most likely to leave.

| | customerID | gender | SeniorCitizen | tenure | Contract | MonthlyCharges | TotalCharges | Churn_Probability |
|---|---|---|---|---|---|---|---|---|
| **6623** | 9248-OJYKK | Male | 1 | -1.277445 | Month-to-month | 0.388472 | -0.972051 | 1.000 |
| **809** | 1820-TQVEV | Male | 0 | -1.277445 | Month-to-month | 0.159144 | -0.975096 | 1.000 |
| **6488** | 0488-GSLFR | Female | 0 | -1.277445 | Month-to-month | 0.157482 | -0.975118 | 1.000 |
| **1731** | 8375-DKEBR | Female | 1 | -1.277445 | Month-to-month | 0.160806 | -0.975074 | 1.000 |
| **6495** | 7254-IQWOZ | Male | 0 | -1.277445 | Month-to-month | 0.162467 | -0.975051 | 1.000 |
| **1739** | 9804-ICWBG | Male | 0 | -1.277445 | Month-to-month | 0.170776 | -0.974941 | 0.995 |
| **2194** | 2514-GINMM | Male | 0 | -1.277445 | Month-to-month | 0.489842 | -0.970706 | 0.995 |
| **2927** | 5542-TBBWB | Male | 0 | -1.277445 | Month-to-month | 0.170776 | -0.974941 | 0.995 |
| **4585** | 1069-XAIEM | Female | 1 | -1.277445 | Month-to-month | 0.674301 | -0.968257 | 0.985 |
| **3682** | 3716-BDVDB | Male | 0 | -1.277445 | Month-to-month | 0.144188 | -0.975294 | 0.980 |

## 12. Insights

- Month-to-month contracts and electronic check payments are key churn indicators.

- Senior citizens and customers with higher monthly charges show higher churn tendency.
- Providing online security and tech support can significantly reduce churn.

## 13. Retention Strategies

1. Offer discounts for annual contracts.
2. Provide personalized offers to high-risk customers.
3. Implement loyalty programs for early-tenure customers.
4. Improve support experience for customers without tech support.
5. Introduce bundle services to increase engagement.

## 14. Conclusion

The project successfully built a predictive churn model using Logistic Regression with 80.75% accuracy.
Insights from EDA and feature importance guided targeted retention strategies, helping the business reduce churn and increase customer satisfaction.
Future work can involve adding real-time customer feedback and advanced algorithms like XGBoost or Neural Networks.

## 15. Tools and Technologies Used

Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
Environment: Google Colab
Evaluation Metrics: Accuracy, ROC AUC

## 16. References

Kaggle: Telco Customer Churn Dataset
Scikit-learn Documentation