

CUSTOMER CHURN ANALYSIS AND PREDICTION

Project Overview

This project focuses on understanding and predicting customer churn in the telecom industry.

It combines Power BI dashboards for descriptive analysis with Python machine learning models for predictive insights.

Problem Statement

Customer churn is one of the most pressing challenges for subscription-based businesses. In the telecom industry, churn often occurs due to factors like high monthly charges, short-term contracts, poor service quality, or billing preferences.

This project aims to:

- 1. Analyze churn patterns using Power BI dashboards.**
- 2. Predict churn likelihood using machine learning models in Python.**

By blending business intelligence and predictive analytics, this project provides both insights and actionable recommendations for customer retention.

Dataset

- **Source:** Telco Customer Churn Dataset (Kaggle)
 - **Records:** 7043 customers
 - **Features:** Demographics, account information, services subscribed, billing, and churn label.
-

Tools & Technologies

- **Data Analysis & Visualization:** Power BI
 - **Programming & ML:** Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
 - **Version Control:** Git & GitHub
-

Approach

1. Exploratory Data Analysis (EDA)

- Data cleaning (handling missing values, encoding categorical features)
- Feature exploration (tenure, contract type, monthly charges, internet service)

2. Churn Analysis (Power BI)

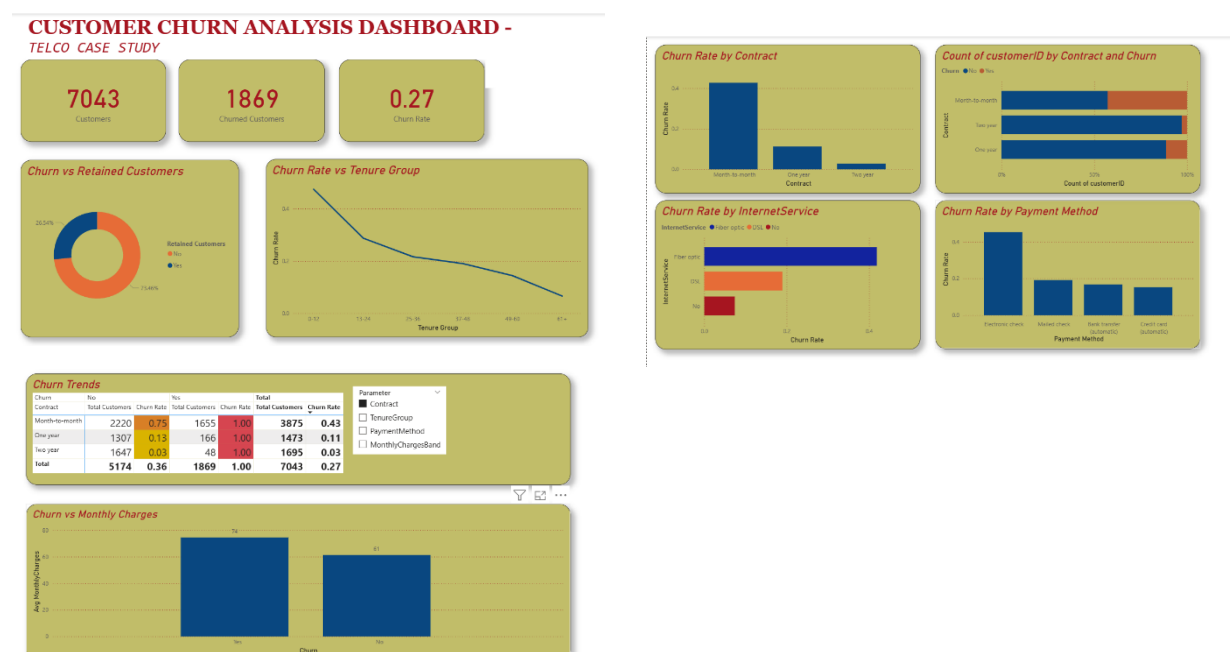
Dashboards created:

- **Dashboard 1 – Overview:** Total Customers, Churned Customers, Churn Rate, Donut Charts
- **Dashboard 2 – Demographics & Services:** Churn by Contract Type, Churn by Internet Service, Churn by Tenure Group
- **Dashboard 3 – Churn Drivers:** Monthly Charges vs Churn, Matrix view, Key Insights

3. Churn Prediction (Python)

- Data preprocessing: Label encoding, feature scaling
- Models applied: Logistic Regression, Random Forest, XGBoost
- Model evaluation: Accuracy, Precision, Recall, F1-score, ROC-AUC

Dashboards (Power BI)



Machine Learning Results

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.75	0.53	0.84	0.65	0.86
Random Forest	0.81	0.68	0.54	0.60	0.85
XGBoost	0.79	0.64	0.52	0.57	0.83

Key Insights

- **Month-to-Month contracts** show the highest churn.
- **Fiber optic users** have higher churn compared to DSL.
- **Higher monthly charges** lead to higher churn rates.
- Customers with **automatic payment methods** churn less.
- Retention improves significantly after the first 12 months of tenure.

Conclusion

This project demonstrates how data analytics + machine learning can be used together to solve real-world business problems.

- **Power BI dashboards help visualize churn patterns and key drivers.**
- **Python ML models allow businesses to predict churn likelihood and target at-risk customers.**