📉 ***Customer Churn Analysis & Prediction (Python)***

📌 Project Overview

This project focuses on analyzing and predicting customer churn using a realistic, raw telecom-style customer dataset.
The objective is to identify churn drivers, predict customers at high risk of leaving, and convert analytical results into actionable business deliverables for retention teams.

The project follows a complete **end-to-end data analytics and machine learning workflow**, starting from raw, uncleaned data and ending with exportable churn-risk customer lists.

🎯 Objectives

- ★ Analyze customer behavior and service usage patterns
- ★ Identify key factors contributing to customer churn
- ★ Build predictive churn models with interpretable results
- ★ Segment customers based on churn risk
- ★ Provide business-ready churn risk exports for retention strategies

---

🛠️ Tools & Technologies

- ★ **Language:** Python
- ★ **Environment:** Google Colab (Cloud-based Jupyter Notebook)
- ★ **Libraries:**
  - pandas
  - numpy
  - matplotlib
  - seaborn
  - scikit-learn

📊 Dataset

★ **Type:** Synthetic but realistic telecom customer dataset
★ **Records:** ~7,000 customer-level records
★ **Key Characteristics :**
  - Mixed categorical and numerical features
  - Raw and uncleaned fields (object-typed numeric columns)
  - Missing and inconsistent values
  - Realistic churn distribution
★ **Key Fields**
  - customerID
  - tenure
  - MonthlyCharges
  - TotalCharges
  - Contract
  - PaymentMethod
  - Service subscriptions
  - Churn (target variable)

The dataset was intentionally designed to mimic real-world raw business data, requiring thorough data cleaning before analysis and modeling.

🔍 Data Cleaning & Preprocessing

Key cleaning steps included:

★ Converting `TotalCharges` from object to numeric
★ Handling blank and invalid values using business logic
★ Encoding churn target variable
★ Standardizing categorical values
★ Removing duplicate records

These steps ensured the dataset was fully analysis- and model-ready.

📊 Exploratory Data Analysis (EDA)

Key insights from EDA:

❖ Approximately one-fourth of customers churned

- ❖ Customers with shorter tenure churn significantly more
- ❖ Higher monthly charges are associated with increased churn risk
- ❖ Month-to-month contracts show the highest churn
- ❖ Auto-payment and long-term contracts reduce churn

Visualizations included:

- ❖ Churn distribution plots
- ❖ Tenure vs churn analysis
- ❖ Monthly charges vs churn (box & violin plots)
- ❖ Contract and payment method churn comparisons
- ❖ Correlation heatmaps

🧠 Advanced Feature Engineering

To improve predictive performance and business interpretability, the following features were engineered:

- ★ **Tenure groups** representing customer lifecycle stages
- ★ **Service count** as a proxy for customer engagement
- ★ **Average monthly spend** to normalize revenue across tenure
- ★ **Long-term contract flag** to capture commitment
- ★ **Auto-payment indicator** reflecting billing behavior

🤖 Modeling & Evaluation

Two baseline models were developed:

1. Logistic Regression
   - ❖ Used as the primary model due to high interpretability
   - ❖ Enabled identification of key churn drivers
   - ❖ Balanced performance between precision and recall

2. Decision Tree
   - ❖ Used for comparison
   - ❖ Captured non-linear relationships
   - ❖ Limited depth to reduce overfitting

Model evaluation focused on:

- ❖ Recall for churned customers

- ❖ Confusion matrix analysis
- ❖ Business interpretability over pure accuracy

📌 Churn Risk Segmentation

Using predicted churn probabilities, customers were segmented into:

- ★ **High Risk**
- ★ **Medium Risk**
- ★ **Low Risk**

This segmentation enables targeted retention actions and efficient allocation of marketing resources.

📤 Business Deliverables

A **marketing-ready churn risk dataset** was exported, including:

- ★ Customer ID
- ★ Churn probability
- ★ Churn risk segment
- ★ Tenure
- ★ Monthly charges
- ★ Contract type
- ★ Payment method
- ★ Engagement indicators

This dataset can be directly used by marketing and customer success teams for retention campaigns.

📌 Conclusion

This project demonstrates how raw customer data can be transformed into actionable churn intelligence using a structured analytics and machine learning workflow.

By combining exploratory analysis, feature engineering, interpretable modeling, and business-focused deliverables, the project highlights how predictive analytics can support proactive customer retention strategies and reduce revenue loss.

🚀 Future Enhancements

★ Hyperparameter tuning and advanced models (XGBoost, Random Forest)
★ Integration with RFM segments for deeper customer intelligence
★ Time-based churn analysis
★ Retention campaign impact measurement
★ Interactive dashboards (Power BI / Tableau)

☁️ Execution Environment

The entire project was developed and executed using **Google Colab**, ensuring cloud-based accessibility, reproducibility, and ease of collaboration.

🧒🏻 Author

Anirudha Das
Aspiring Data Analyst / Business Intelligence Professional
📍Bolpur, WB, India
