# Telco Customer Churn Prediction

## Project Overview

This project analyzes customer data from a telecommunications company to predict which customers are likely to "churn" (cancel their service). I cleaned the data, performed Exploratory Data Analysis (EDA), and built a Logistic Regression model to identify at-risk customers.

## Technologies Used

* **Python:** Data manipulation and analysis.
* **Pandas:** Data cleaning and converting "object" types to numeric.
* **Seaborn/Matplotlib:** Visualizing churn distribution.
* **Scikit-Learn:** Building the Logistic Regression model.

## Key Challenges Solved

### 1. Handling Dirty Data

The TotalCharges column contained hidden blank strings (" ") which prevented analysis. I used pd.to\_numeric with errors='coerce' to identify and remove these corrupt records.

### 2. Feature Engineering

I converted the categorical target variable Churn (Yes/No) into numeric values (1/0) to make it compatible with the Logistic Regression algorithm.

## Model Results

* **Model Used:** Logistic Regression
* **Features:** Tenure, Monthly Charges, Total Charges
* **Accuracy:** **78%**

## Code Snippet (Data Cleaning)

# Force converting string numbers to floats and handling errors  
df['TotalCharges'] = pd.to\_numeric(df['TotalCharges'], errors='coerce')  
  
# Dropping the rows with missing values created by the conversion  
df.dropna(inplace=True)

## How to Run

1. Clone the repository.
2. Install dependencies:  
   pip install pandas seaborn scikit-learn matplotlib
3. Run the notebook churn\_analysis.ipynb.