# Task-Four:

# Telco Customer Churn Prediction - Logistic Regression vs XGBoost

## Introduction

This project analyzes the Telco Customer Churn dataset to predict customer churn.  
Two models were trained and compared: **Logistic Regression** and **XGBoost**.  
The objective was to determine which model performs better in predicting churn.

## Dataset Details

The dataset contains details about Telco customers including demographics, account information, and service usage.  
The target variable is **Churn**, which indicates whether a customer has left the service.

## Data Preprocessing

* Dropped customerID column.
* Converted TotalCharges to numeric and replaced missing values with the median.
* Applied Label Encoding to binary categorical variables.
* Applied One-Hot Encoding to multi-category categorical variables.
* Split the dataset into training (80%) and testing (20%) sets using stratified sampling.
* Scaled numerical features (tenure, MonthlyCharges, TotalCharges) using StandardScaler.
* Applied median imputation to handle any remaining missing values.

## Model Training

Two models were trained on the processed dataset:

1. **Logistic Regression** (max\_iter=1000).
2. **XGBoost Classifier** (n\_estimators=200, learning\_rate=0.1, max\_depth=6).

## Results

The models were evaluated using Accuracy, Precision, Recall, and F1-Score.

**Logistic Regression:**

* Accuracy: **80.6%**
* Precision: **65.8%**
* Recall: **55.6%**
* F1-Score: **60.3%**

**XGBoost:**

* Accuracy: **78.6%**
* Precision: **61.4%**
* Recall: **52.7%**
* F1-Score: **56.7%**

## Interpretation

* Logistic Regression outperformed XGBoost across all evaluation metrics.
* Logistic Regression achieved higher Accuracy, Precision, Recall, and F1-score.
* Recall is especially important in churn prediction as it measures the ability to correctly identify churners.
* XGBoost, though powerful, may require hyperparameter tuning to improve performance.

## Conclusion

For this dataset, **Logistic Regression** is the stronger model.  
It provides better Recall and F1-score, making it more suitable for identifying churners accurately.  
This is critical for the business as correctly identifying churners helps reduce customer loss.

XGBoost may outperform Logistic Regression with further **hyperparameter tuning**, but Logistic Regression is currently the best choice.