Task 3 - Customer Churn Prediction

Internship at Nexus AI Digital — Data Science & Machine Learning
Author: Syed Huzaifa Bin Khamis

Objective:

GlobalMart wants to **predict customer churn** so that they can offer incentives to customers at risk of leaving.

In this task, I built a **Logistic Regression classifier** on the Telco Customer Churn dataset.

Steps Taken:

1. Preprocessing

- Removed unnecessary IDs.
- Converted `TotalCharges` to numeric and handled missing values.
- Encoded categorical variables with Label Encoding.

2. Train-Test Split:

- 80% training, 20% testing (stratified to keep churn ratio balanced).

3. Model Training:

- Logistic Regression (`sklearn`).

4. Evaluation:

- Accuracy Score.
- Confusion Matrix for error analysis.

Results:

- Accuracy: 80% (varies slightly depending on random seed).
- Confusion Matrix Interpretation:
  - True Negatives (correctly predicted non-churners).
  - True Positives (correctly predicted churners).
  - False Positives (predicted churn but actually stayed → unnecessary retention cost).
  - False Negatives (predicted stay but actually churned → most dangerous, lost customers).

Key Insight:

While Logistic Regression gives a decent baseline, it struggles with **false negatives**.In business, these are critical because they represent customers we failed to retain. Next steps could include testing more advanced models like Random Forests or XGBoost.

How to Run:
1. Install dependencies:
   ```bash
   pip install pandas numpy matplotlib seaborn scikit-learn