**Problem Statement – Telecom Industry: SIM Swap & Usage Fraud Detection**

**Goal:**  
Build a simple fraud detection pipeline to find suspicious mobile users who might be involved in SIM swap or fake subscription fraud.

**Tasks:**

1. **Data Preparation:**
   * Use or simulate a dataset with columns like customer\_id, plan\_type, calls\_per\_day, data\_usage\_gb, device\_change\_count, location\_change\_count, and fraud\_label.
   * Clean missing values and scale numeric features.
2. **Feature Engineering:**
   * Create new features such as:
     + Average data usage per day
     + Ratio of location changes to call count
     + Number of device changes in a week
3. **Model Building:**
   * Train a model (e.g., Random Forest or XGBoost) to classify users as **fraudulent** or **normal**.
   * Handle imbalance using oversampling or class weights.
4. **Fraud Risk Scoring:**
   * Assign a fraud probability (0–1) to each customer.
   * Label users as:
     + Low Risk: score < 0.3
     + Medium Risk: 0.3–0.7
     + High Risk: > 0.7
5. **Evaluation:**
   * Check model performance using accuracy, precision, recall, and ROC-AUC.
   * Plot a simple bar chart showing fraud vs. normal user count.

**Expected Output:**  
A working pipeline that reads telecom data, trains a model, and outputs fraud risk scores for each customer.