Task 2

**IBM HR Analytics Employee Attrition & Performance Dataset**:  
<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

**Task Steps:**

1. **Understand the Problem**
   * What is "attrition"? Why would a company care about this?
   * Define the target variable: Attrition
2. **Data Preprocessing**
   * Handle categorical variables using label/one-hot encoding
   * Check for class imbalance (Yes/No in Attrition)
   * Optional: Apply SMOTE if the imbalance is serious
3. **EDA**
   * Use visualizations to understand:
     + Which departments have high attrition?
     + Does job satisfaction or age correlate with leaving?
4. **Model Building**
   * Build a Random Forest Classifier
   * Compare XGBoost or any other preferred classifier
   * Use a stratified train/test split
5. **Evaluation**
   * Accuracy, precision, recall, F1-score
   * Confusion matrix
   * ROC Curve
6. **Explainability Task (Key Focus)**
   * Explain:
     + Which features most influence the attrition prediction?
   * Create bar plots to visualize key factors
7. **Business Insight Report**
   * + What factors are causing employees to leave?
     + What departments should they focus on?
     + Simple, actionable suggestions based on model output

**Deliverables (Deadline: EOD or Next Day):**

* Clean Jupyter Notebook (code)
* GitHub push with structured README
* Business Insight PDF:
  + Summary of model performance
  + Top 5 features influencing attrition
  + Actionable advice