Problem Statement: Employee Attrition Prediction

Objective:

As a data analyst at a company, your task is to understand the factors contributing to employee attrition (employees leaving the company). You need to perform Exploratory Data Analysis (EDA) to uncover trends and patterns, and then train a machine learning model to predict whether an employee is likely to leave or not.

Dataset Link:

Download from Kaggle: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

Dataset Overview:

Each row in the dataset represents an employee with details like:

- Age, Gender, Department, MonthlyIncome, DistanceFromHome
- JobSatisfaction, OverTime, YearsAtCompany, JobRole, etc.
- Attrition (Target Variable): Yes or No

Part 1: Exploratory Data Analysis (EDA)

1. How many employees are there in the dataset? How many left the company?
2. What is the average age of employees who left vs. stayed?
3. What is the attrition rate across departments?
4. Is there any correlation between distance from home and attrition?
5. What is the relationship between overtime and attrition?
6. How does monthly income impact attrition?
7. Which features have the most missing values (if any)?

Part 2: Visualizations

1. Bar chart: Count of employees by Attrition.
2. Boxplot: MonthlyIncome vs Attrition.
3. Histogram: Age distribution by Attrition.
4. Stacked bar chart: Department vs Attrition.

5. Countplot: JobSatisfaction by Attrition.

6. Heatmap: Correlation between numerical variables.

Part 3: Machine Learning Model Training

Step-by-step tasks:

1. Encode categorical variables (LabelEncoding / OneHotEncoding).

2. Prepare features (X) and target (y).

3. Train/Test Split (80/20).

4. Train at least two classification models:

   - Logistic Regression

   - Random Forest / Decision Tree / KNN (your choice)

5. Evaluate using:

   - Accuracy

   - Confusion Matrix

   - Classification Report

   - ROC Curve

Final Task:

- Submit your work in the form of a Jupyter Notebook (.ipynb).

- Ensure your notebook includes:

  - Answers to all EDA questions

  - All visualizations

  - Model code and evaluation