HR Data Analysis

Objective

The goal of this assignment is to create a dataset for an HR scenario, perform basic operations using Python libraries like Pandas and NumPy, and visualize the data using a visualization library (Matplotlib/Seaborn).

This assignment will help to understand data creation, manipulation, and basic visualization techniques in the context of HR analytics.

Scenario:

You are working as a Data Analyst in the HR department of a company. The HR manager wants to understand employee data to identify trends and make informed decisions. Your task is to:

1. **Create a Dataset**
2. **Perform Basic Data Analysis**
3. **Visualize Key Insights**

Assignment Steps

Step 1: Dataset Creation

Create an employee dataset with the following columns:

- **Employee ID**: A unique ID for each employee (e.g., 101, 102, …).
- **Name**: Employee names.
- **Age**: Employee age.
- **Department**: Department they work in (e.g., `HR`, `IT`, `Finance`, `Marketing`).
- **Salary**: Monthly salary in USD.
- **Joining Year**: The year the employee joined the company.
- **Performance Rating**: A rating between 1 to 5 (1 = poor, 5 = excellent).
- **Leaves Taken**: Number of leaves taken in the current year.

Example Data (at least 15 rows):

Employee ID	Name	Age	Department	Salary	Joining Year	Performance Rating	Leaves Taken
101	Alice	28	IT	5500	2020	4	8
102	Bob	35	HR	4800	2018	3	12
103	Charlie	30	Marketing	6200	2019	5	5

Step 2: Perform Basic Analysis

1. **Load the Dataset**
   - Use Pandas to create and load the dataset.
2. **Basic Insights**
   - Display the first 5 rows of the dataset.
   - Show the dataset information (data types, non-null values, etc.).
   - Calculate the average salary and average performance rating.
   - Count the number of employees in each department.
3. **Data Manipulation**
   - Add a new column `Experience` calculated as `2024 - Joining Year`.
   - Create a filtered dataset of employees with a `Performance Rating` greater than 3.
   - Find the highest-paid employee and their department.

Step 3: Visualization

Use Matplotlib or Seaborn to create the following visualizations:

1. **Bar Chart**: Show the count of employees in each department.
2. **Histogram**: Plot the distribution of employee salaries.
3. **Box Plot**: Compare salaries across different departments.
4. **Scatter Plot**: Show the relationship between `Age` and `Salary`.

Deliverables

- Python script or Jupyter Notebook containing:

- Dataset creation code.
- Pandas and NumPy operations for analysis.
- Visualizations with clear labels and titles.

Expected Learning Outcomes

- Understand how to create a real-world dataset.
- Perform essential data manipulation and analysis using Pandas and NumPy.
- Visualize data to identify patterns and insights.

Optional Challenge

- Identify the department with the highest average performance rating.
- Analyze if there's a relationship between `Performance Rating` and `Leaves Taken`.