Employee Performance Data Analysis

🎯 Objective:

1. Clean raw employee performance data

Real-world data often contains issues like missing values, incorrect formats (e.g., dates as strings), or inconsistent entries.
✅ In this step, the goal is to:

- Correct data types (e.g., convert dates to datetime)

- Ensure numerical columns like salary and performance rating are properly formatted

- Handle or fill missing values appropriately

This ensures the dataset is **reliable and ready for analysis**.

---

2. Engineer new insights (e.g., tenure, salary category)

Feature engineering means **creating new columns or insights** from existing data to reveal patterns.
✅ In your project:

- Tenure is derived from the employee's joining year (2025 - JoinYear)

- SalaryCategory is introduced to classify employees as **Low**, **Medium**, or **High** based on salary

These new features allow for **more meaningful group-wise analysis**.

---

3. Generate department-wise summaries

To make informed decisions, it's important to look at trends **within each department**.
✅ This involves:

- Calculating the **average salary** per department

- Counting the **number of employees by gender** in each department

- Analyzing the **average performance rating** across departments

- Identifying departments with **low-performing employees**

This helps highlight which departments are excelling and which may need support.

---

4. Visualize key metrics

Data visualizations make complex information easy to understand.
✅ In your case:

- **Pie chart** shows the distribution of salary categories

- **Bar chart** shows average salary per department

These charts allow stakeholders to **quickly grasp trends** and make decisions backed by data.

---

5. Export results to multi-sheet Excel file

Instead of just viewing results in code, it's useful to export cleaned data and analysis to Excel for sharing or further use.
✅ The Excel file contains **multiple sheets**, such as:

- Cleaned data

- Salary analysis

- Gender count

- Performance summaries

This makes the report **organized, portable, and easy to review** by non-technical users too.

Tasks Performed:

1. Data Cleaning

- Converted JoinDate to proper datetime format (dd-mm-yyyy)

- Converted Salary and PerformanceRating columns to numeric types

- Handled missing values to ensure data consistency

2. Feature Engineering

- Created a new column Tenure calculated as: 2025 - JoinYear

- Added a SalaryCategory column to classify salaries:

  - Low: Less than ₹50,000

 - Medium: ₹50,000 – ₹90,000

  - High: More than ₹90,000

3. Aggregated Analysis

- Calculated average salary by department

- Counted gender distribution across departments

- Computed average performance rating per department

- Identified low performers (Performance Rating ≤ 2)

- Generated overall summary statistics:

  - Total employees

  - Average salary

  - Average performance rating

Visualizations:

- Bar Chart – Average Salary by Department
Displays the average salary across different departments to identify compensation trends.
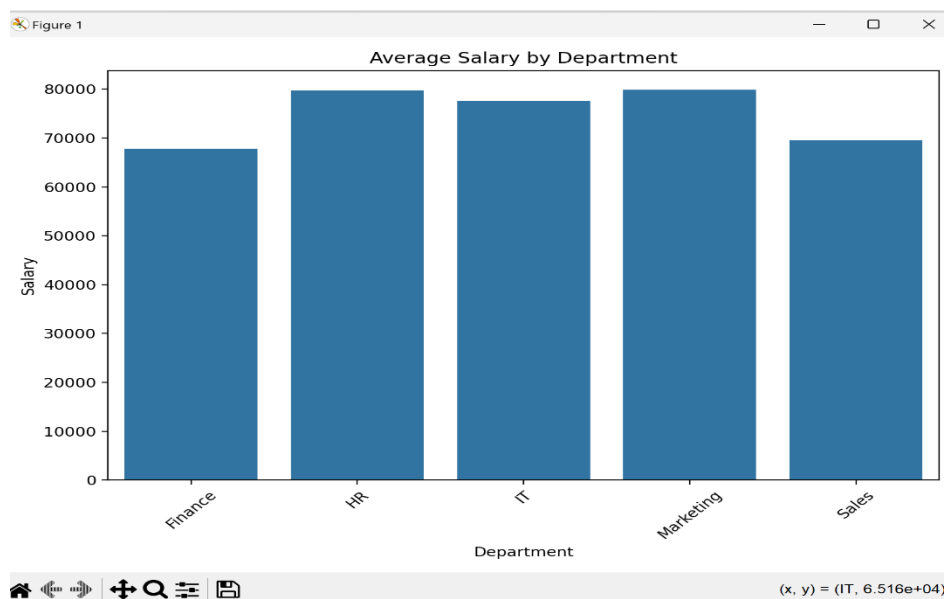
- Pie Chart – Salary Category Distribution
Illustrates the percentage distribution of employees across salary categories (Low, Medium, High).

Salary Category Distribution
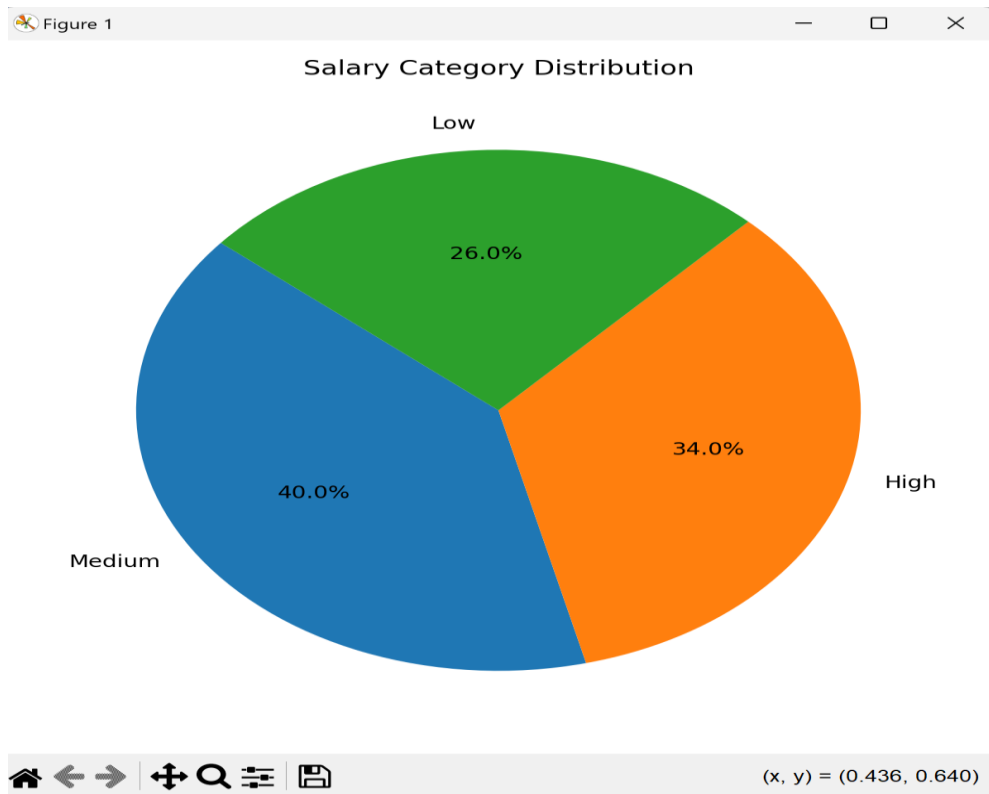
Tools Used:

- Python 3.x
- Pandas
- Matplotlib
- Seaborn
- Openpyxl

📝 Summary:

This project analyzed employee performance data by cleaning and transforming raw inputs, creating useful features, and generating department-wise insights. Key metrics were visualized through charts, and all results were exported to a structured Excel file for easy review and decision-making.