# Employee Salary Data Cleaning and Aggregation for HR Insights

## Problem Statement

The HR department of a company maintains employee records, including salaries, departments, and job titles. However, the dataset contains duplicates, missing values, and inconsistent data formats. The goal is to clean the employee dataset using Pandas, perform aggregation to extract useful HR insights, and export the results for reporting.

## Objectives

1. Clean the dataset:  
- Handle missing values in salaries and job titles.  
- Remove duplicate employee records.  
- Standardize department names and job titles.

2. Perform Aggregations:  
- Calculate average salary per department.  
- Identify highest-paid employee in each department.  
- Find total salary expenditure per department.  
- Group employees by job titles and compute their average salaries.  
- Count employees per department.

3. Export cleaned and aggregated results into CSV files.

## Dataset Structure (employees.csv)

EmpID, Name, Department, JobTitle, Salary, JoiningDate  
Example rows:  
E001, Alice Wong, HR, HR Manager, 60000, 2019-03-15  
E002, Bob Smith, IT, Software Engineer, 75000, 2020-06-20  
E003, Carol Lee, Finance, Accountant, 55000, 2018-11-12

## Full Dataset (50 Rows, CSV Format)

EmpID,Name,Department,JobTitle,Salary,JoiningDate  
E001,Alice Wong,HR,HR Manager,60000,2019-03-15  
E002,Bob Smith,IT,Software Engineer,75000,2020-06-20  
E003,Carol Lee,Finance,Accountant,55000,2018-11-12  
E004,David Kim,IT,Data Scientist,95000,2021-02-10  
E005,Eva Brown,Sales,Sales Executive,50000,2019-07-01  
...  
(50 rows in total)

## Tasks to Perform (Using Pandas Only)

1. Load dataset (pd.read\_csv).

2. Inspect dataset (info(), describe()).

3. Remove duplicates (if any).

4. Handle missing salaries/job titles (if we simulate).

5. Standardize department and job titles (e.g., 'Software Engg' → 'Software Engineer').

6. Add derived columns (e.g., YearsOfService = CurrentDate – JoiningDate).

7. Group by Department:  
 - Average salary.  
 - Total salary cost.  
 - Employee count.

8. Group by JobTitle:  
 - Average salary.  
 - Highest-paid employee.

9. Export summaries to CSV (dept\_summary.csv, job\_summary.csv).