Project Summary
This project involves cleaning and analyzing employee performance data using Python. The raw dataset contains information about employees' hours worked, tasks completed, overtime hours, and efficiency ratings. The goal is to clean the data, correct errors, and derive meaningful insights through analysis.

Objectives

1. Clean the raw dataset to address issues such as missing values, inconsistent formatting, and duplicate rows.
2. Conduct analysis to:
    - Visualize workload distribution by department.
    - Identify employees with excessive overtime.
    - Analyze productivity correlations.

Steps Executed
1. Data Cleaning

The data_cleaning.py script performed the following tasks:

- Removed duplicate rows to eliminate redundancy.
- Standardized department names to ensure consistency.
- Handled invalid and missing values:
    - Replaced negative overtime hours with None and filled missing values with 0.
    - Imputed missing Hours_Worked and Tasks_Completed using the median.
    - Imputed missing Efficiency_Rating using the mean.

2. Data Analysis

The analysis.py script provided the following insights:

- **Workload Distribution by Department**: Visualized average hours worked per department using a bar chart.
- **Excessive Overtime**: Identified employees working more than 10 overtime hours.
- **Productivity Analysis**:
    - Created a scatter plot to show the relationship between hours worked and tasks completed.
    - Visualized efficiency rating distribution using a histogram.