EDA Insights - Employee Performance

This report summarizes the Exploratory Data Analysis steps to prepare employee performance data for modeling. Run the supplied Jupyter Notebook (EDA_employee_performance.ipynb) after placing your dataset file in the same folder. The notebook includes code for data loading, cleaning, missing value treatment, duplicates removal, count plots for categorical variables, boxplots for numeric variables, and a correlation heatmap.

Key Findings (to look for after you run the notebook):

1. Distribution of employees across Departments and Job Roles - identify imbalances.

2. Presence of missing values in features like JobSatisfaction, NumCompaniesWorked - treat appropriately.

3. Outliers in MonthlyIncome and YearsAtCompany - investigate seniority or data entry errors.

4. Correlations: check whether YearsAtCompany, JobSatisfaction, WorkLifeBalance correlate with PerformanceRating.

5. Overtime & Performance: inspect whether employees with overtime have higher ratings (also consider burnout risk).

Deliverables:

1. EDA_employee_performance.ipynb (Notebook with code & visualizations)

2. clean_employee_performance.csv (generated by notebook)

3. EDA_Insights_Report.pdf (this file)

Instructions: Place 'employee_performance.csv' in this folder alongside the notebook. Open the Jupyter Notebook and run all cells. After running, 'clean_employee_performance.csv' will be created. Export graphs/pictures to include in your final report or PPT.