This project focuses on analyzing an HR dataset to understand the relationship between employee performance scores and absences. The analysis provides insights into department and manager performance, identifies outliers, and explores correlations to help improve organizational strategies.

Project Overview

- **Objective:**
  - Investigate employee performance and absenteeism patterns.
  - Identify high-performing departments and managers.
  - Detect outliers and understand their potential causes.
- **Dataset:**
  - The dataset used in this project is the `HRDataset_v14.csv` file.
  - Contains information on employee performance, absences, department, and manager details.

Steps and Methodology

1. **Data Preparation:**
   - Checked and converted column types (e.g., `PerformanceScore` and `Absences`).
   - Handled missing data through appropriate cleaning steps.
2. **Descriptive Statistics:**
   - Computed key statistics such as mean, median, mode, variance, and standard deviation for performance scores and absences.
3. **Data Visualization:**
   - Histograms to analyze performance score distribution.
   - Boxplots to visualize absences and detect outliers.
   - Bar charts to compare average performance by department and manager.
4. **Outlier Analysis:**
   - Applied the 1.5 * IQR rule to detect outliers in performance scores and absences.
5. **Correlation Exploration:**
   - Examined the relationship between absences and performance scores using scatter plots and correlation coefficients.
6. **Department and Manager Analysis:**
   - Ranked departments and managers by average performance score to identify the best and worst performers.

Results

- **Correlation:** A weak/moderate correlation (depending on the dataset) was found between absences and performance scores.

- **Outliers:** Significant outliers in absences and performance scores were identified, which may indicate underlying issues requiring further investigation.
- **Best and Worst Performers:**
    - Top-performing department: `<Department Name>`.
    - Top-performing manager: `<Manager Name>`.
    - Departments or managers with poor performance were also highlighted for improvement.

Visualizations

The project includes the following visualizations:

1. Distribution of performance scores.
2. Boxplots for absences and performance scores to detect outliers.
3. Scatter plot showing the relationship between absences and performance scores.
4. Bar charts comparing average performance by department and manager.

Usage

To run this project:

1. Clone this repository:

   ```
   git clone https://github.com/<your-username>/HR-Performance-Analysis.git
   ```

2. Navigate to the project directory:

   ```
   cd HR-Performance-Analysis
   ```

3. Open the notebook in your preferred environment (e.g., Jupyter, Kaggle, Colab).

Dependencies

The project uses the following Python libraries:

- `pandas`
- `numpy`
- `matplotlib`
- `seaborn`

Install them using:

```
pip install pandas numpy matplotlib seaborn
```

License

This project is licensed under the MIT License. See the LICENSE file for details.