The dataset I used, employee\_performance\_dataset.csv, contains information on 201 employees across different departments. Each row represents one employee and includes details such as Employee ID, Name, Department, Hire Date, Salary, Attendance Rate, Projects Completed, Performance Score, and Manager. The Attendance Rate column shows the proportion of days the employee attended work, while Performance Score represents the employee’s rating on a numeric scale.

The goal of this analysis was to explore whether there is a relationship between employee attendance and performance, essentially asking: Do employees who attend work more consistently tend to perform better?

To start, I loaded the dataset into Python using pandas:

import pandas as pd

df = pd.read\_csv("employee\_performance\_dataset.csv", header=0)

Next, I renamed the columns for clarity:

df.columns = [

"Employee\_ID", "Name", "Department", "Hire\_Date",

"Salary", "Attendance\_Rate", "Projects\_Completed",

"Performance\_Score", "Manager"

]

Then, I converted the numeric columns into proper numeric types and removed rows with missing values to prepare the data for plotting:

df['Attendance\_Rate'] = pd.to\_numeric(df['Attendance\_Rate'], errors='coerce')

df['Performance\_Score'] = pd.to\_numeric(df['Performance\_Score'], errors='coerce')

df = df.dropna(subset=['Attendance\_Rate', 'Performance\_Score'])

To visualize the relationship between attendance and performance, I created a scatter plot using matplotlib and seaborn:

import matplotlib.pyplot as plt

import seaborn as sns

plt.figure(figsize=(8,6))

sns.regplot(x='Attendance\_Rate', y='Performance\_Score', data=df)

plt.title("Attendance Rate vs Performance Score")

plt.xlabel("Attendance Rate")

plt.ylabel("Performance Score")

plt.show()

In this scatter plot, each point represents an employee, with the x-axis showing Attendance Rate and the y-axis showing Performance Score. A regression line shows the overall trend. The plot revealed a positive correlation, indicating that employees with higher attendance generally tend to have higher performance scores. The average attendance rate is approximately 80%, and the average performance score is around 3.5. A few outliers were visible, showing employees with unusually high or low performance relative to their attendance.

From this analysis, the insight is that attendance can be a useful predictor of performance, which could help managers assess employee effectiveness. Additional enhancements could include adding average lines to the scatter plot, exploring correlations with other variables like projects completed or department, and generating summary tables for top performers. Overall, this project demonstrates the ability to clean, analyse, and visualize data, as well as to extract meaningful insights that answer the key question: How does attendance relate to performance?