1 EMPLOYEE SALARY ANALYSIS REPORT

PREPARED BY : EISHA

BRANCH : CSE(AI)

SECTION : B

UNIVERSITY ROLL NO. : 202401100300109

INTRODUCTION

This report presents an analysis of employee salary data using Python. The analysis includes statistical calculations, identification of highest and lowest salaries, and visual representation of salary distribution.

In today's fast-paced and competitive business environment, understanding the strengths, weaknesses, and potential of an organization's workforce is critical for long-term success. The Employee Analysis Program is designed to provide organizations with valuable insights into employee performance, engagement, skills, and development needs.

By leveraging data-driven analytics, this program helps companies make informed decisions about talent management, optimize workforce allocation, enhance productivity, and identify areas for growth and improvement. Whether through performance reviews, feedback systems, or data analytics, the program aims to create a more transparent, accountable, and supportive work culture.

This program is not only a tool for assessing individual employee performance but also a strategic resource for fostering employee development, improving job satisfaction, and aligning team goals with organizational objectives. The result is a stronger, more efficient workforce that contributes meaningfully to the company's growth and success.

METHODOLOGY

The dataset used in this analysis is stored in 'employee_data.csv'. The methodology involves the following steps:

1. Loading the dataset using pandas.
2. Calculating key salary statistics (mean, median, min, max).
3. Identifying the highest and lowest-paid employees.
4. Visualizing salary distribution using matplotlib.

CODE IMPLEMENTATION

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = "employee_data.csv"
df = pd.read_csv(file_path)

# Display basic information and first few rows
print("Dataset Overview:")
print(df.info())
print(df.head())

# Basic salary statistics
print("\nSalary Statistics:")
print("Mean Salary:", df['Salary'].mean())
print("Median Salary:", df['Salary'].median())
print("Minimum Salary:", df['Salary'].min())
print("Maximum Salary:", df['Salary'].max())

# Highest and lowest-paid employees
highest_paid = df.loc[df['Salary'].idxmax()]
lowest_paid = df.loc[df['Salary'].idxmin()]
print("\nHighest Paid Employee:")
print(highest_paid)
print("\nLowest Paid Employee:")
print(lowest_paid)

# Salary distribution plot
plt.figure(figsize=(8, 5))
plt.hist(df['Salary'], bins=20, color='skyblue', edgecolor='black')
plt.xlabel('Salary')
plt.ylabel('Frequency')
plt.title('Salary Distribution')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

1.1   4. SCREENSHOTS & OUTPUT

Insert screenshots of the output and salary distribution graph here.

```
Highest Paid Employee:
EmployeeID              8
Age                    46
Department        Finance
Experience             34
Salary             144637
Name: 7, dtype: object

Lowest Paid Employee:
EmployeeID              2
Age                    28
Department        Finance
Experience              2
Salary              41742
Name: 1, dtype: object
```

```
Dataset Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   EmployeeID  20 non-null     int64
 1   Age         20 non-null     int64
 2   Department  20 non-null     object
 3   Experience  20 non-null     int64
 4   Salary      20 non-null     int64
dtypes: int64(4), object(1)
memory usage: 932.0+ bytes
None
   EmployeeID  Age Department  Experience  Salary
0           1   23    Finance           8   93563
1           2   28    Finance           2   41742
2           3   37         HR           8   56905
3           4   23         HR          23  138397
4           5   55         IT          29   96879

Salary Statistics:
Mean Salary: 102503.15
Median Salary: 101315.5
Minimum Salary: 41742
Maximum Salary: 144637
```

Salary Distribution