Department of CSE AI and AIML

Introduction to AI

REPORT

EMPLOYEE SALARY ANALYSIS

NAME: DIVYANSH YADAV

PROGRAM: B.TECH

ROLL NO.: 20240110040085

BRANCH: CSE(AIML)

SECTION: B

SEMESTER: II

SUBMITTED TO: MR. ABHISHEK SHUKLA SIR

INTRODUCTION

This report presents an analysis of employee salary data to explore salary distribution, job roles, years of experience, and their correlation. The goal is to gain insights into salary trends and the impact of experience on earnings. Various statistical techniques and visualizations are employed to better understand the dataset and derive meaningful conclusions.

METHODOLOGY

The analysis follows these steps:

Data Collection

A dataset containing employee names, job titles, salaries, and years of experience was created. The data was stored in a **Pandas Dataframe** for easy manipulation and analysis.

Data Processing

- **Exploratory Data Analysis (EDA)** was performed to understand the dataset structure.
- **Descriptive statistics** were generated to summarize salary and experience distributions.
- **Data visualization techniques** such as histograms, box plots, scatter plots, and heatmaps were used to analyze salary trends.

CODE

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


# Sample dataset

employee_data = {

   'Employee': ['Divyansh', 'Dhruv', 'Disha', 'Dev', 'Rohit', 'Babloo', 'Gitansh',
'Harsh', 'Inaya', 'Jitesh'],

   'Job Title': ['Manager', 'Engineer', 'Technician', 'Manager', 'Engineer',
'Technician', 'Manager', 'Engineer', 'Technician', 'Manager'],

   'Salary': [85000, 70000, 50000, 90000, 72000, 48000, 88000, 73000, 51000,
87000],

   'Years of Experience': [10, 5, 3, 12, 6, 2, 11, 7, 4, 10]

}


# Create DataFrame

df = pd.DataFrame(employee_data)


# Display the first few rows

print(df.head())
```

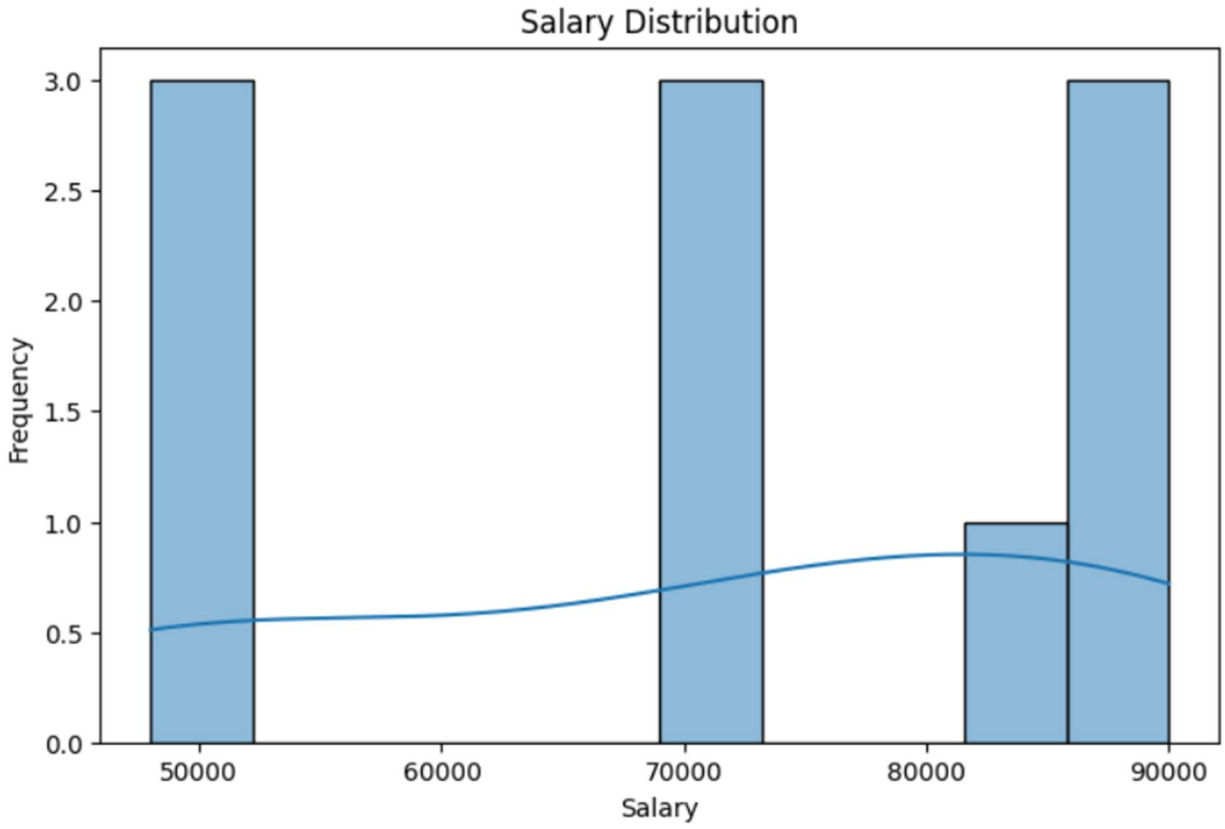
```
# Basic statistics
print(df.describe())

# Salary distribution
plt.figure(figsize=(8, 5))
sns.histplot(df['Salary'], bins=10, kde=True)
plt.title('Salary Distribution')
plt.xlabel('Salary')
plt.ylabel('Frequency')
plt.show()

# Box plot of salary by job title
plt.figure(figsize=(8, 5))
sns.boxplot(x='Job Title', y='Salary', data=df)
plt.xticks(rotation=45)
plt.title('Salary Distribution by Job Title')
plt.show()

# Scatter plot: Experience vs. Salary
plt.figure(figsize=(8, 5))
sns.scatterplot(x='Years of Experience', y='Salary', data=df)
```

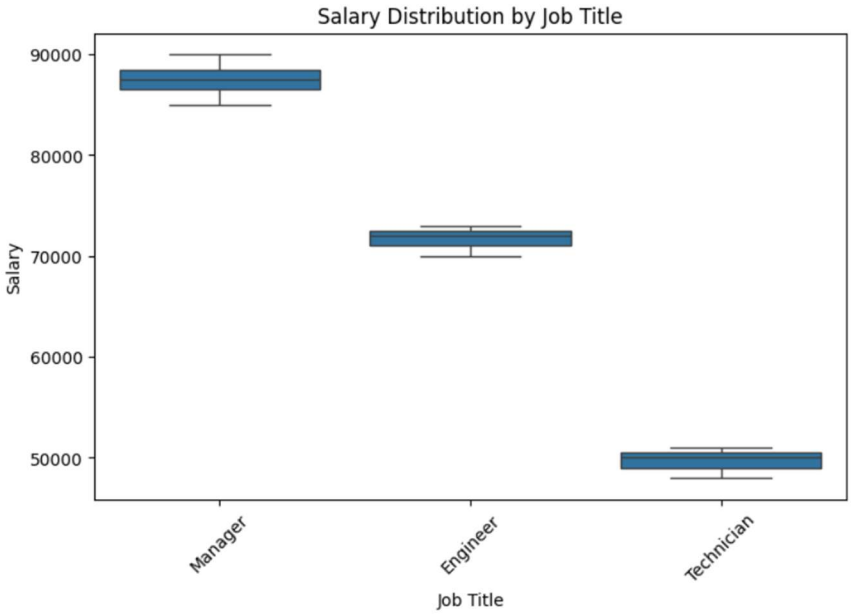
```
plt.title('Years of Experience vs Salary')

plt.xlabel('Years of Experience')

plt.ylabel('Salary')

plt.show()


# Correlation matrix

plt.figure(figsize=(6, 5))

sns.heatmap(df[['Salary', 'Years of Experience']].corr(), annot=True,
cmap='coolwarm', fmt=".2f")

plt.title('Correlation Matrix')

plt.show()


# Grouping by Job Title to check average salary

avg_salary_by_title = df.groupby('Job Title')['Salary'].mean().sort_values()

print(avg_salary_by_title)


# Checking correlation between Salary and Experience

correlation = df[['Salary', 'Years of Experience']].corr()

print("\nCorrelation between Salary and Years of Experience:\n",correlation)
```

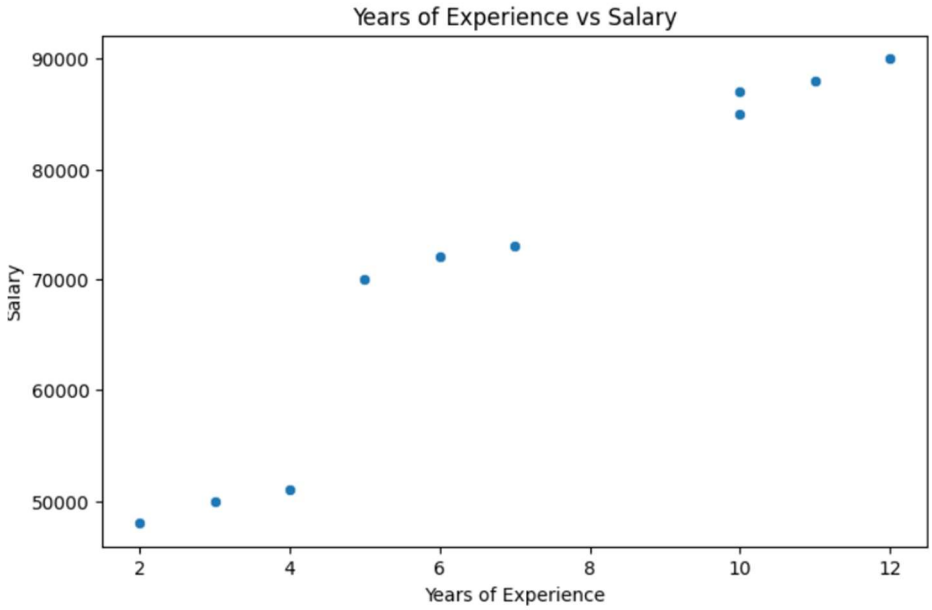
OUTPUT:

	Employee	Job Title	Salary	Years of Experience
0	Divyansh	Manager	85000	10
1	Dhruv	Engineer	70000	5
2	Disha	Technician	50000	3
3	Dev	Manager	90000	12
4	Rohit	Engineer	72000	6

	Salary	Years of Experience
count	10.000000	10.000000
mean	71400.000000	7.000000
std	16587.813465	3.559026
min	48000.000000	2.000000
25%	55750.000000	4.250000
50%	72500.000000	6.500000
75%	86500.000000	10.000000
max	90000.000000	12.000000




Salary Distribution by Job Title
90000
80000
70000
60000
50000
Salary
Manager
Engineer
Technician
Job Title


Years of Experience vs Salary
90000
80000
70000
60000
50000
Salary
2
4
6
8
10
12
Years of Experience


```
Job Title
Technician    49666.666667
Engineer      71666.666667
Manager       87500.000000
Name: Salary, dtype: float64

Correlation between Salary and Years of Experience:
                     Salary  Years of Experience
Salary             1.000000             0.965505
Years of Experience  0.965505             1.000000
```

REFERENCE:

Python libraries: pandas, matplotlib, seaborn.

Basic statistical analysis and visualization techniques