Title Page

Problem Statement:

Employee Salary Analysis using Python

Name: Vansh karnwal
Roll No.: 2024011100300274
Course: INTRODUCTION TO AI
Institution: KIET

Introduction

Employee salary analysis is crucial for organizations to ensure fair compensation, identify salary trends, and make data-driven decisions. This project aims to analyze employee salary data using Python to gain insights into salary distribution, trends, and influencing factors such as experience, education, and job role. By leveraging Python libraries such as Pandas, NumPy, and Matplotlib, we will perform data cleaning, visualization, and statistical analysis to extract meaningful insights.

The results of this analysis can help businesses optimize salary structures, promote pay equity, and improve employee satisfaction. Additionally, it can assist job seekers in understanding industry salary trends and making informed career decisions.

Methodology

Brief Description of Code

This Python script analyzes employee salary data using Pandas, Matplotlib, and Seaborn. The following steps are performed:

1. **Loading the Dataset**: The CSV file containing employee salary data is loaded using Pandas.
2. **Descriptive Statistics**: The script prints summary statistics of the salary column, including mean, standard deviation, min/max, and quartiles.
3. **Histogram Visualization**: A histogram is plotted to display the salary distribution, including a kernel density estimate.
4. **Boxplot for Outliers**: A boxplot is generated to identify outliers in the salary data.
5. **Experience vs. Salary Correlation**: If an 'Experience' column exists, a scatter plot is created to visualize the relationship between experience and salary, and the correlation coefficient is calculated.

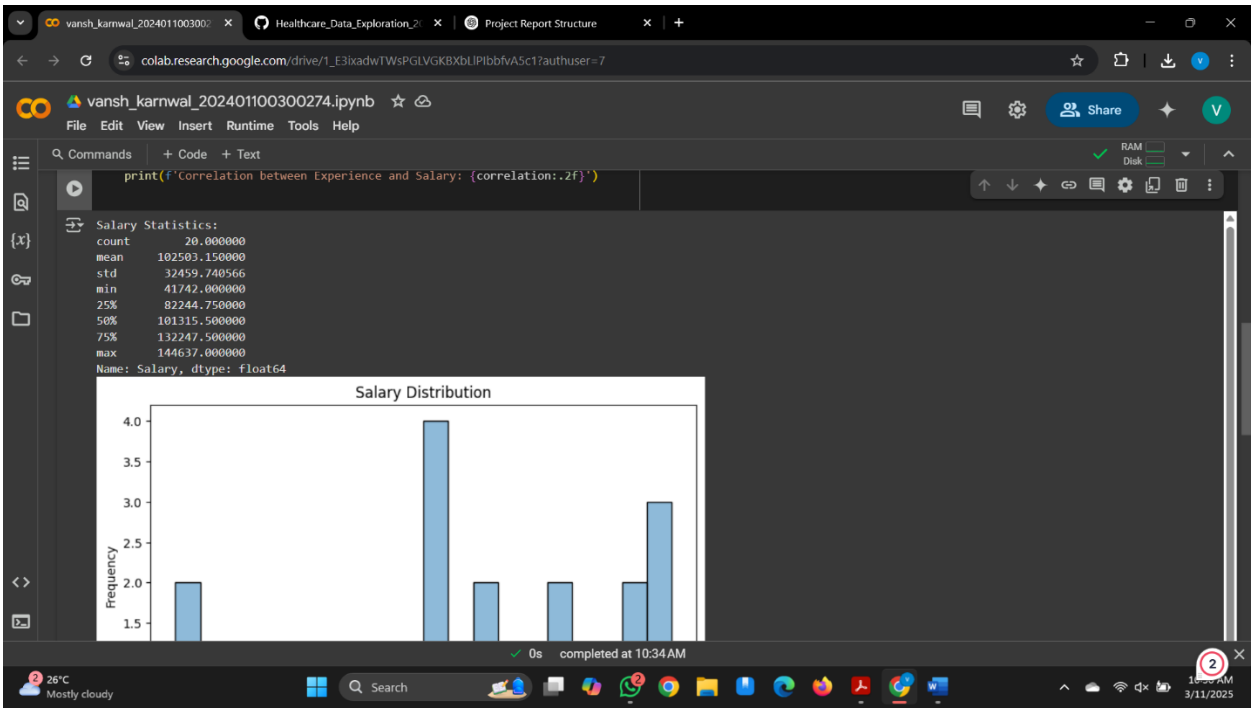
These steps help in understanding salary trends, distribution, and potential influencing factors.

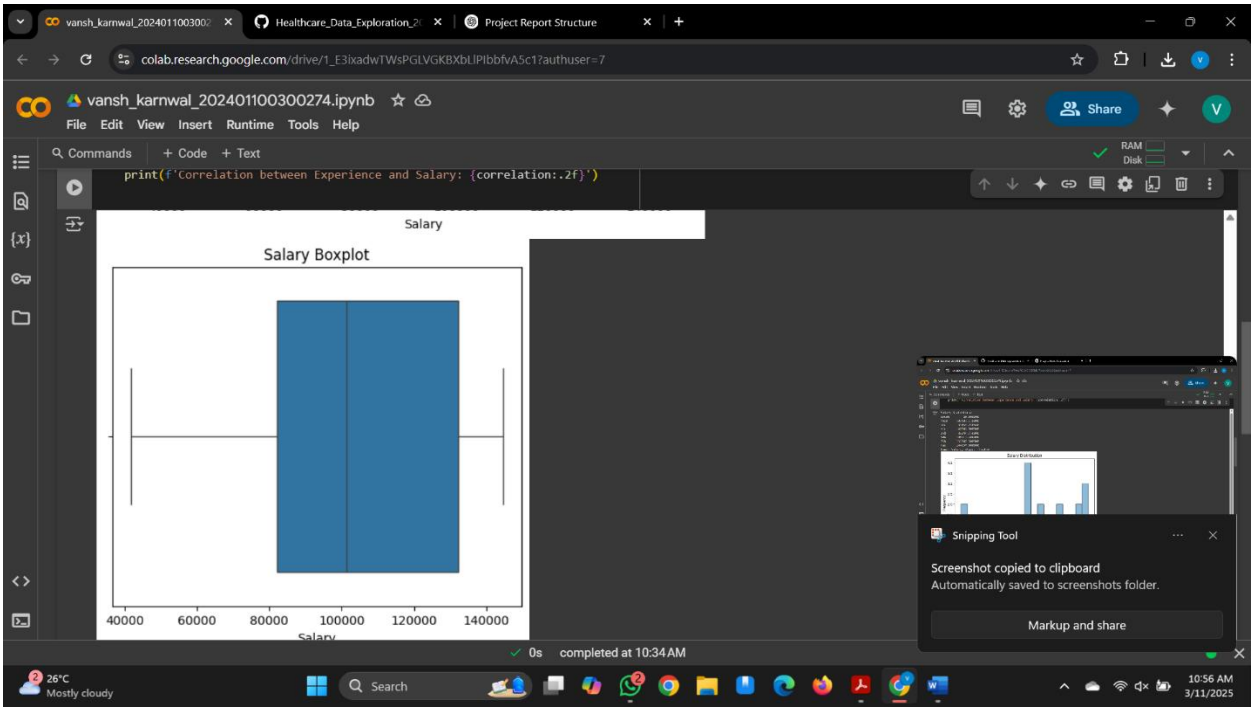
Code

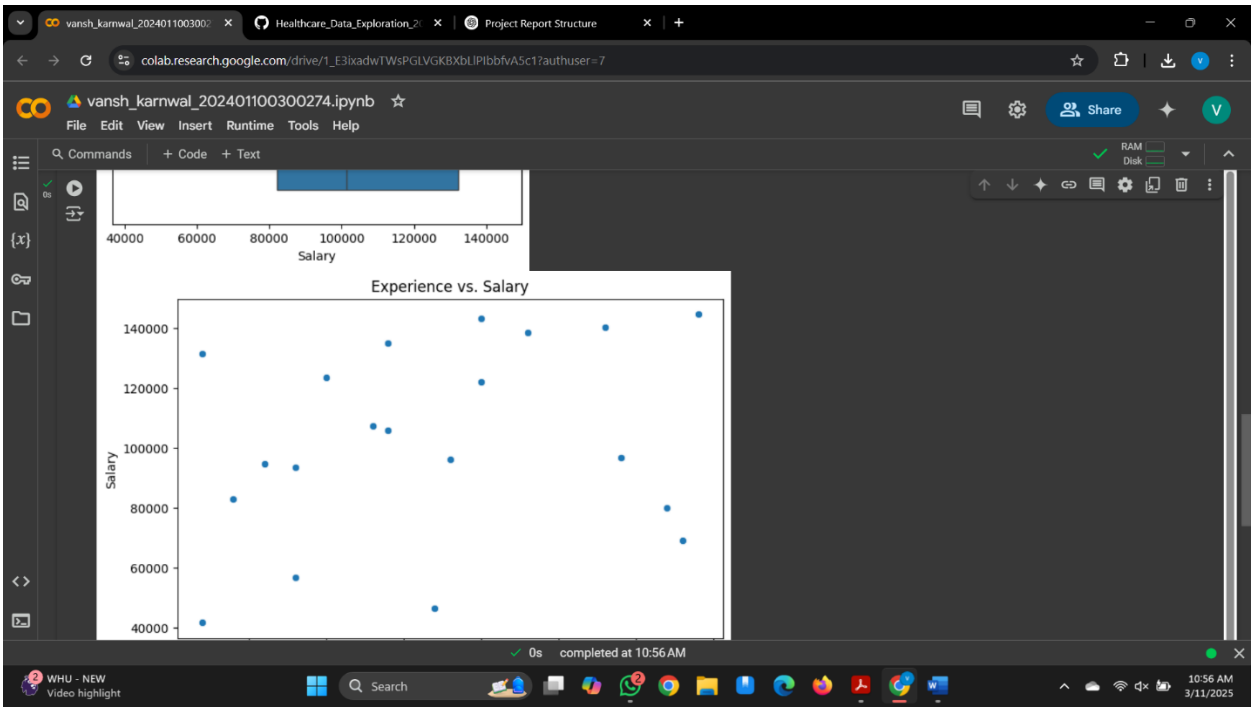
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset (Replace 'employee_salaries.csv' with your actual file)
df = pd.read_csv('/content/employee_data.csv')

# Display basic salary statistics
print("Salary Statistics:")
print(df['Salary'].describe())  # Provides count, mean, std, min, max, and quartiles

# Histogram of salary distribution
plt.figure(figsize=(8,5))
sns.histplot(df['Salary'], bins=20, kde=True)  # kde=True adds a kernel density estimate
plt.title('Salary Distribution')
plt.xlabel('Salary')
plt.ylabel('Frequency')
plt.show()
```

```
# Boxplot to detect outliers in salary data
plt.figure(figsize=(6,5))
sns.boxplot(x=df['Salary'])
plt.title('Salary Boxplot')
plt.show()

# Check correlation with experience (if the column exists in the dataset)
if 'Experience' in df.columns:
    plt.figure(figsize=(8,5))
    sns.scatterplot(x=df['Experience'], y=df['Salary'])  # Scatter plot for Experience vs. Salary
    plt.title('Experience vs. Salary')
    plt.xlabel('Years of Experience')
    plt.ylabel('Salary')
    plt.show()

    # Calculate and display correlation coefficient between Experience and Salary
    correlation = df[['Experience', 'Salary']].corr().iloc[0,1]
    print(f'Correlation between Experience and Salary: {correlation:.2f}')
```

OUTPUT


vansh_karnwal_202401100300274.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
print(f'Correlation between Experience and Salary: {correlation:.2f}')
Salary Statistics:
count 20.000000
mean 102503.150000
std 32459.740566
min 41742.000000
25% 82244.750000
50% 101315.500000
75% 132247.500000
max 144637.000000
Name: Salary, dtype: float64
Salary Distribution
Frequency
4.0
3.5
3.0
2.5
2.0
1.5
0s completed at 10:34 AM
26°C
Mostly cloudy
Search
3/11/2025


vansh_karnwal_202401100300274.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
print(f'Correlation between Experience and Salary: {correlation:.2f}')
Salary
Salary Boxplot
40000
60000
80000
100000
120000
140000
Salary
Snipping Tool
Screenshot copied to clipboard
Automatically saved to screenshots folder.
Markup and share
0s completed at 10:34 AM
26°C
Mostly cloudy
Search
10:56 AM
3/11/2025


vansh_karnwal_202401100300274.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text
RAM
Disk
40000 60000 80000 100000 120000 140000
Salary
Experience vs. Salary
Salary
140000
120000
100000
80000
60000
40000
0s completed at 10:56 AM
WHU - NEW
Video highlight
Search
10:56 AM
3/11/2025