K.I.E.T. Group of Institutions

Ghaziabad

Name: Prathvi Raj Chauhan

Branch: CSE(AI&ML) - B

Roll No: 72

Date: 10/03/2025

Project Report On - Employee Salary Analysis Visualization

Introduction

In this project we analyze employee salaries based on various factors such as experience years, age, and gender. Using data visualization techniques, we explore trends and relationships in the dataset to gain insights into salary distribution.

Methodology

1. **Data Collection:** The dataset consists of employee details, including ID, experience in years, age, gender, and salary.
2. **Data Processing:** The dataset is cleaned by handling missing values and converting salary values to numerical format.
3. **Visualization Techniques:**
    - Histograms for salary distribution.
    - Boxplots to compare salaries by experience.
    - Correlation heatmaps to identify relationships between variables.
    - Scatter plots to visualize salary trends over experience.
    - Bar charts to analyze salary differences by gender.
4. **Tools Used:** Python, Pandas, Matplotlib, and Seaborn.

Code

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files


# Upload the dataset
print("Please upload the dataset (CSV file)")
uploaded = files.upload()
# Get the filename
filename = list(uploaded.keys())[0]
# Load the dataset
df = pd.read_csv(filename)
# Ensure correct column names by stripping whitespace
df.columns = df.columns.str.strip()
# Ensure Salary column is numeric
df['Salary'] = pd.to_numeric(df['Salary'], errors='coerce')
# Visualizing Salary Distribution
plt.figure(figsize=(10, 5))
sns.histplot(df['Salary'].dropna(), bins=30, kde=True)
plt.title('Salary Distribution')
plt.xlabel('Salary')
plt.ylabel('Frequency')
plt.show()
# Spacer for readability
print("\n----------------------------------------\n")
# Boxplot of Salaries by Experience Years
```

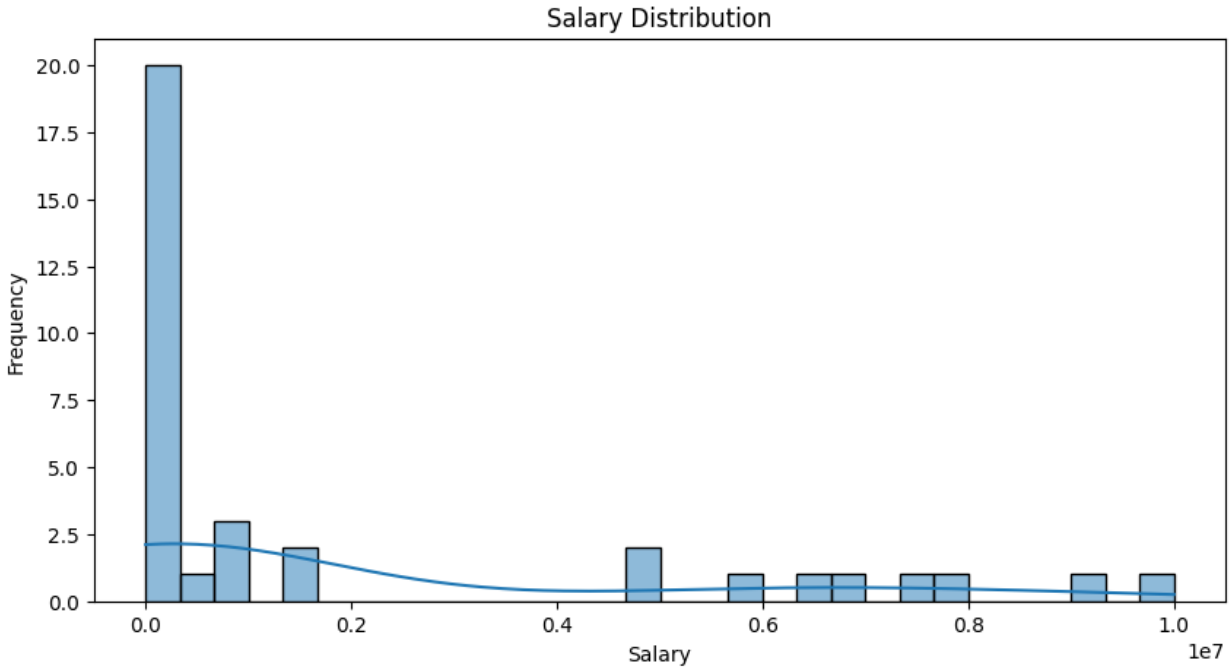
```
plt.figure(figsize=(12, 6))
sns.boxplot(x='Experience_Years', y='Salary', data=df)
plt.xticks(rotation=45)
plt.title('Salary Distribution by Experience Years')
plt.show()
# Spacer for readability
print("\n----------------------------------------\n")
# Correlation Heatmap (excluding non-numeric columns)
numeric_df = df.select_dtypes(include=['number'])
plt.figure(figsize=(8, 5))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
# Spacer for readability
print("\n----------------------------------------\n")
# Bar chart of Average Salary by Gender
plt.figure(figsize=(8, 5))
if 'Gender' in df.columns:
    avg_salary = df.groupby('Gender')['Salary'].mean().sort_values(ascending=False)
    sns.barplot(x=avg_salary.index, y=avg_salary.values)
    plt.title('Average Salary by Gender')
    plt.ylabel('Average Salary')
    plt.xlabel('Gender')
    plt.show()
else:
    print("Column 'Gender' not found in dataset")
# Spacer for readability
```

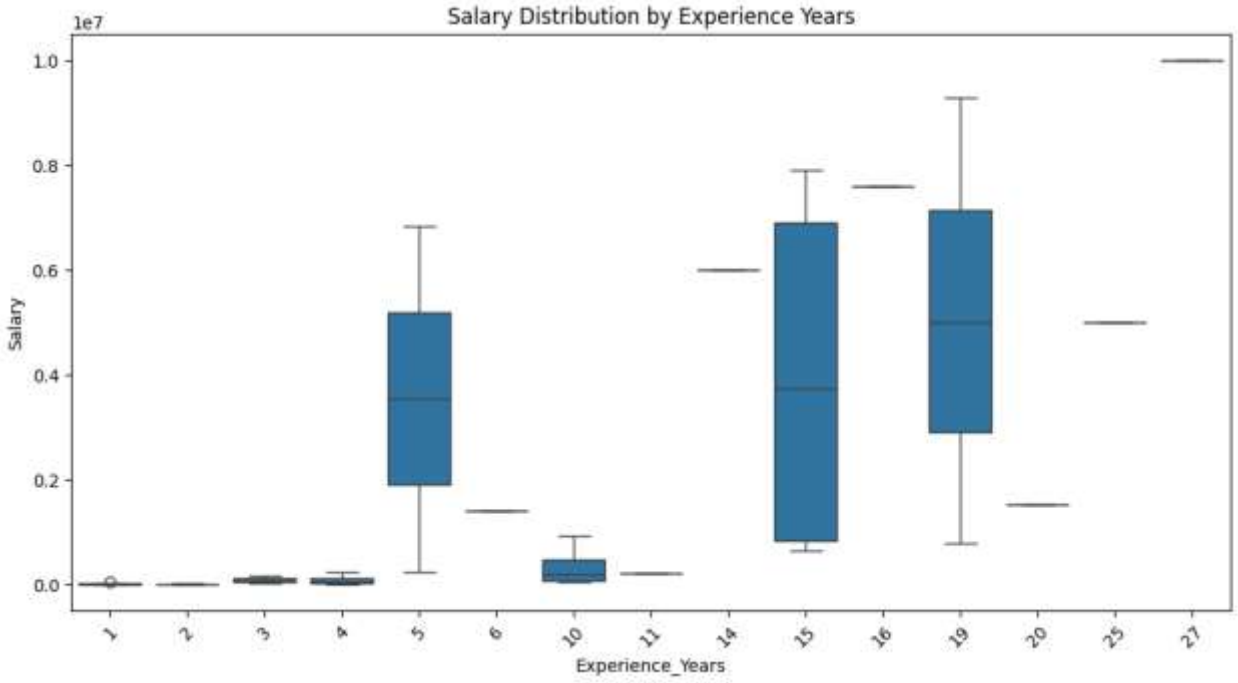
```
print("\n----------------------------------------------\n")

# Scatter plot of Salary vs. Experience Years

plt.figure(figsize=(10, 5))

if 'Experience_Years' in df.columns:

  sns.scatterplot(x='Experience_Years', y='Salary', hue=df['Gender'] if 'Gender' in df.columns else None, data=df)

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

  plt.xlabel('Years of Experience')

  plt.ylabel('Salary')

  plt.show()

else:

  print("Column 'Experience_Years' not found in dataset")

print("\nAnalysis Complete! Thank you for using the Employee Salary Analysis tool.")
```

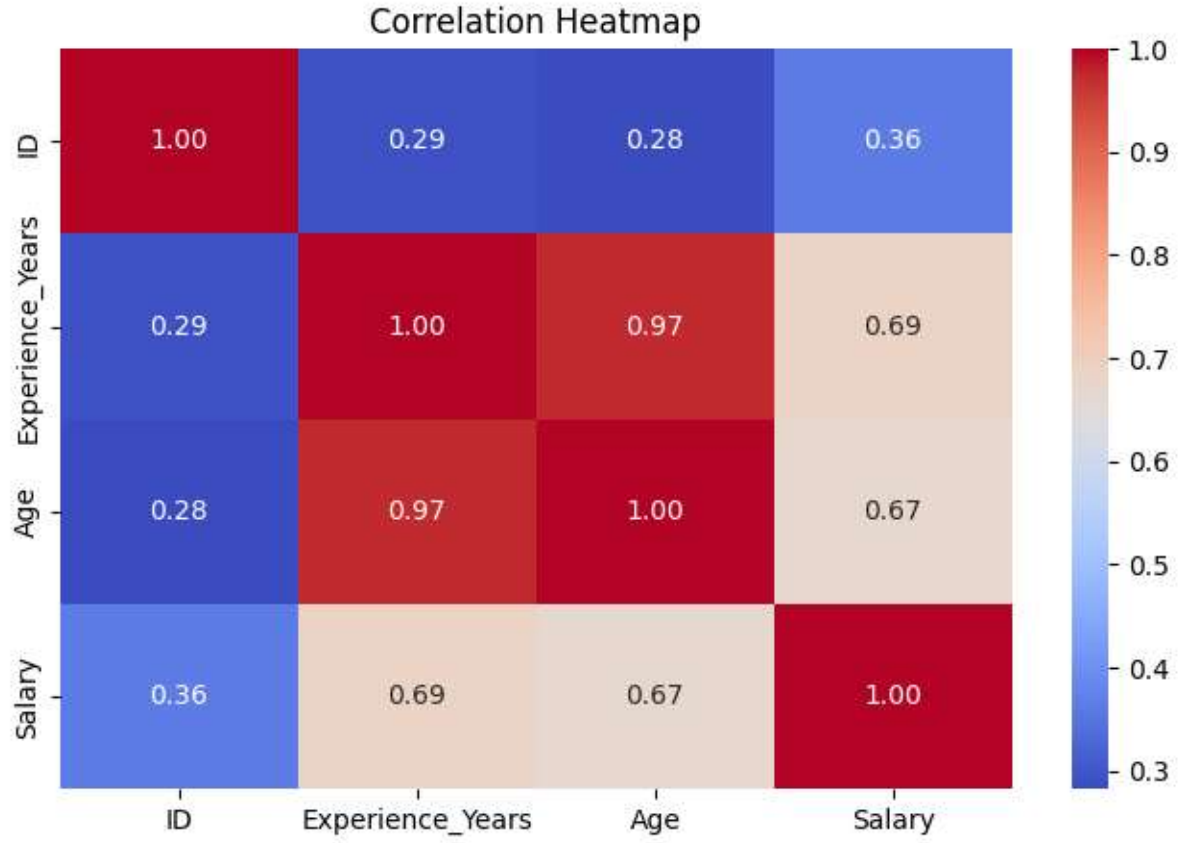
Output/Result

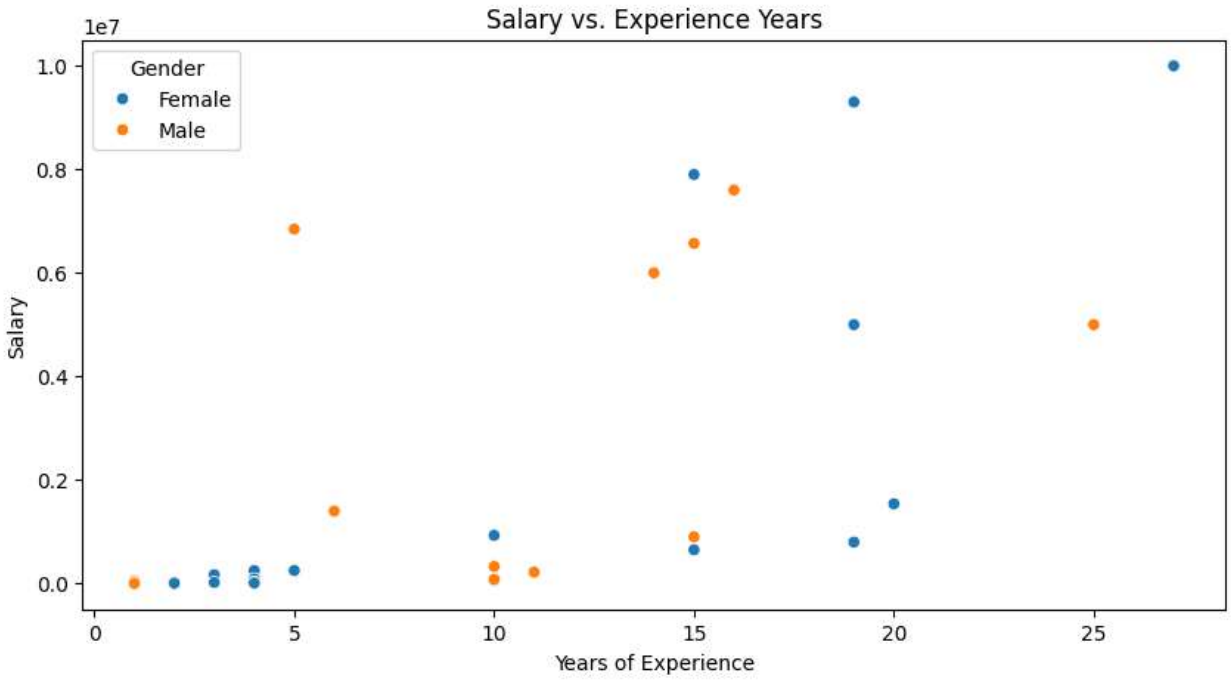
Observation from the analysis is given below:

- Salaries increase with experience, but outliers exist.
- Some high salaries distort the average salary range.
- Gender-based salary disparities can be visualized using bar charts.
- Correlation analysis highlights the relationship between experience and salary.




Salary Distribution by Experience Years
1e7
1.0
0.8
0.6
0.4
0.2
0.0
Salary
1
2
3
4
5
6
10
11
14
15
16
19
20
25
27
Experience_Years


Correlation Heatmap
ID
Experience_Years
Age
Salary
1.00
0.29
0.28
0.36
0.29
1.00
0.97
0.69
0.28
0.97
1.00
0.67
0.36
0.69
0.67
1.00
ID
Experience_Years
Age
Salary
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3



References/Credits:

- Dataset: [source: Kaggle]
- Libraries: Pandas, Matplotlib, Seaborn