

Employee Salary Analysis

Report

Problem Statement: Analyze employee salary data to explore correlations between salary, years of experience, and job position using data visualization techniques.

Student Name: Aditi Narang

Class: AIML A

Roll No.: 202401100400012

1. Introduction

Employee salary analysis helps in understanding salary trends, identifying discrepancies, and making informed business decisions regarding compensation structures. This report explores correlations between employee salaries and factors such as job position, years of experience, and department.

2. Objectives

- Analyze the relationship between salary and experience.
- Identify salary trends across different positions and departments.
- Detect any anomalies or outliers in salary distributions.
- Provide insights for optimizing salary structures.

3. Methodology

3.1 Data Collection

The dataset consists of 30 employee records with the following attributes:

- **Employee ID:** Unique identifier for each employee.
- **Name:** Placeholder names for employees.
- **Position:** Job role of the employee.
- **Department:** Department in which the employee works.
- **Years of Experience:** Total years of professional experience.
- **Salary:** Annual salary of the employee.

3.2 Data Processing & Cleaning

- Removed any missing or duplicate values.
- Converted categorical data where necessary for visualization.
- Filtered outliers using statistical methods to ensure accurate analysis.

4. Data Analysis & Visualizations

4.1 Correlation Analysis

A correlation matrix was generated using a heatmap to identify relationships between numerical variables.

- **Findings:**
 - A positive correlation between **years of experience** and **salary** was observed.
 - Other factors such as department and position influenced salary but were not directly represented in the correlation matrix.

4.2 Salary Distribution

A histogram of salaries showed the distribution of employee salaries across different ranges.

- **Findings:**
 - Salaries ranged between **₹30,000** to **₹1,50,000**.
 - The distribution was slightly skewed towards higher salaries, indicating a concentration of employees in mid-to-high salary ranges.

4.3 Salary by Position

A boxplot analysis helped visualize salary variations across job positions.

- **Findings:**
 - Higher-level positions such as **Product Manager** and **Finance Manager** had significantly higher salaries compared to entry-level roles like **IT Support** and **Sales Executive**.
 - There were some outliers indicating exceptionally high or low salaries in certain positions.

4.4 Experience vs Salary

A scatter plot showed the relationship between **years of experience** and **salary**.

- **Findings:**
 - Employees with more years of experience generally had higher salaries.

- However, some mid-experienced employees earned more than highly experienced ones, suggesting that factors like performance and role also impact salaries.

5. Key Insights & Recommendations

- **Strong correlation between experience and salary**, but job role also plays a crucial part in salary determination.
- **Significant salary variations across positions**, with managerial roles earning the highest.
- **Salary outliers** should be further investigated to ensure fair compensation.
- **Recommendation:** Consider revising salary structures to maintain equity across departments and positions.

6. Conclusion

This analysis provides valuable insights into salary distributions, experience-based trends, and department-wise salary variations. Further investigation with a larger dataset can help refine these findings for better salary structuring and employee retention strategies.

Here's the code :

1. Importing the libraries and loading the dataset

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# Load dataset

df = pd.read_csv("employee_salaries.csv")
```

2. Taking the basic info about the dataset

```
print(df.info())
```

```
print(df.describe())
```

Handling missing values (if any)

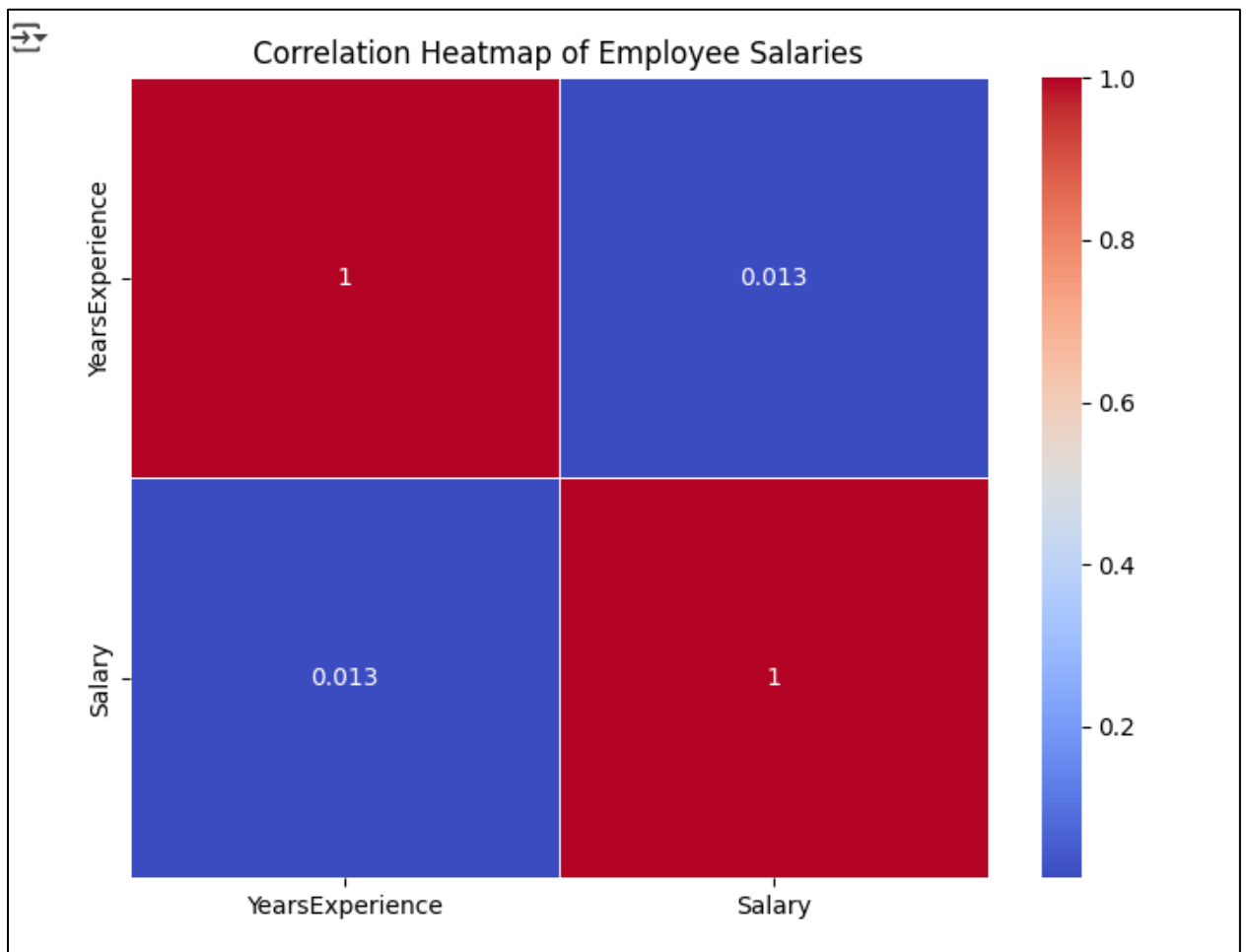
```
df.dropna(inplace=True)
```

```
➡ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Employee_ID           30 non-null    object
1   Name                  30 non-null    object
2   Position              30 non-null    object
3   Department            30 non-null    object
4   YearsExperience       30 non-null    int64
5   Salary               30 non-null    int64
dtypes: int64(2), object(4)
memory usage: 1.5+ KB
None
```

	YearsExperience	Salary
count	30.000000	30.000000
mean	9.633333	99052.266667
std	5.041095	36100.221920
min	1.000000	30206.000000
25%	7.000000	69480.500000
50%	9.000000	109442.000000
75%	13.000000	125011.000000
max	19.000000	149176.000000

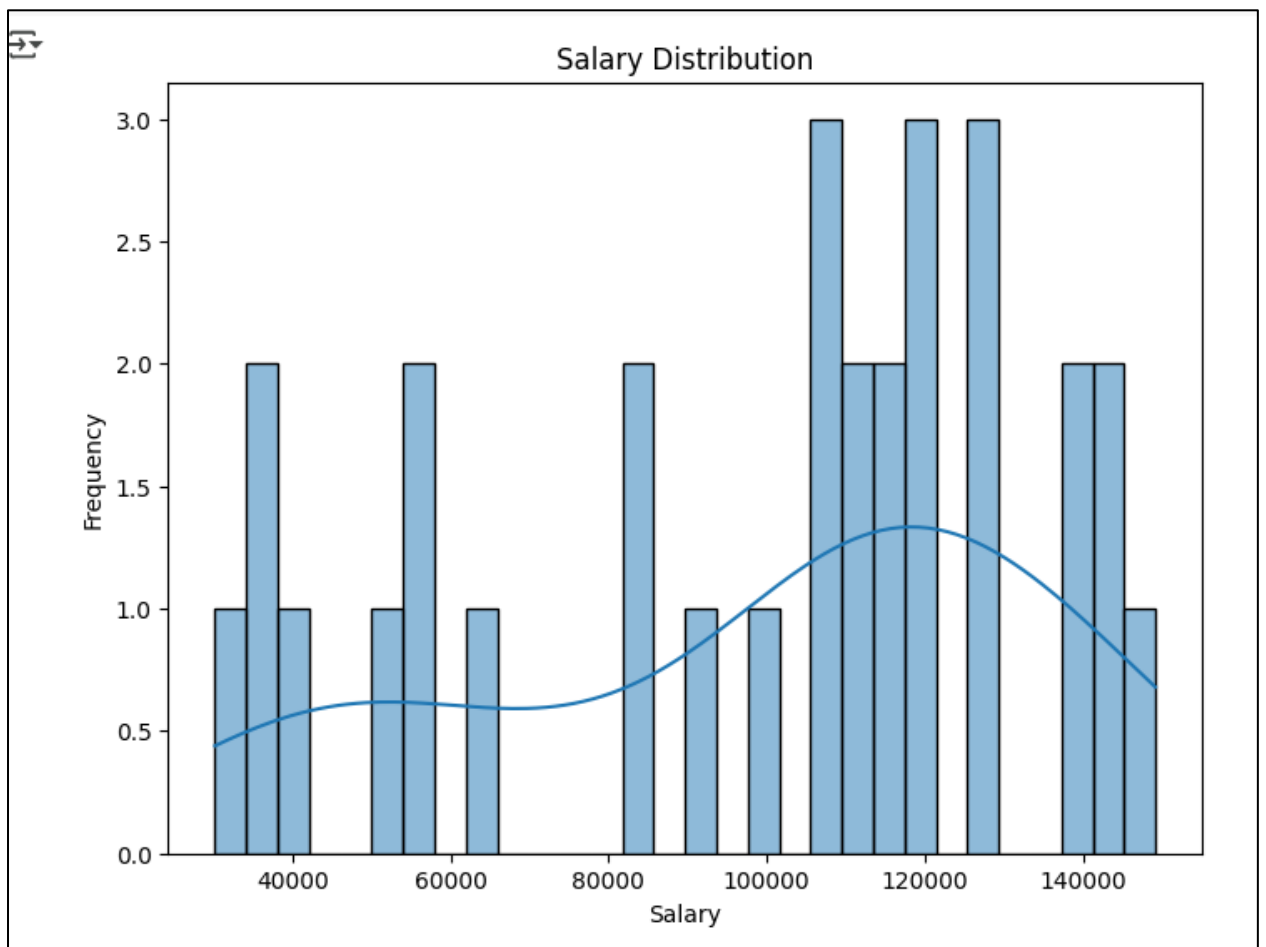
3. Correlation matrix

```
plt.figure(figsize=(8, 6))  
sns.heatmap(df.corr(), annot=True, cmap="coolwarm",  
linewidths=0.5)  
plt.title("Correlation Heatmap of Employee Salaries")  
plt.show()
```



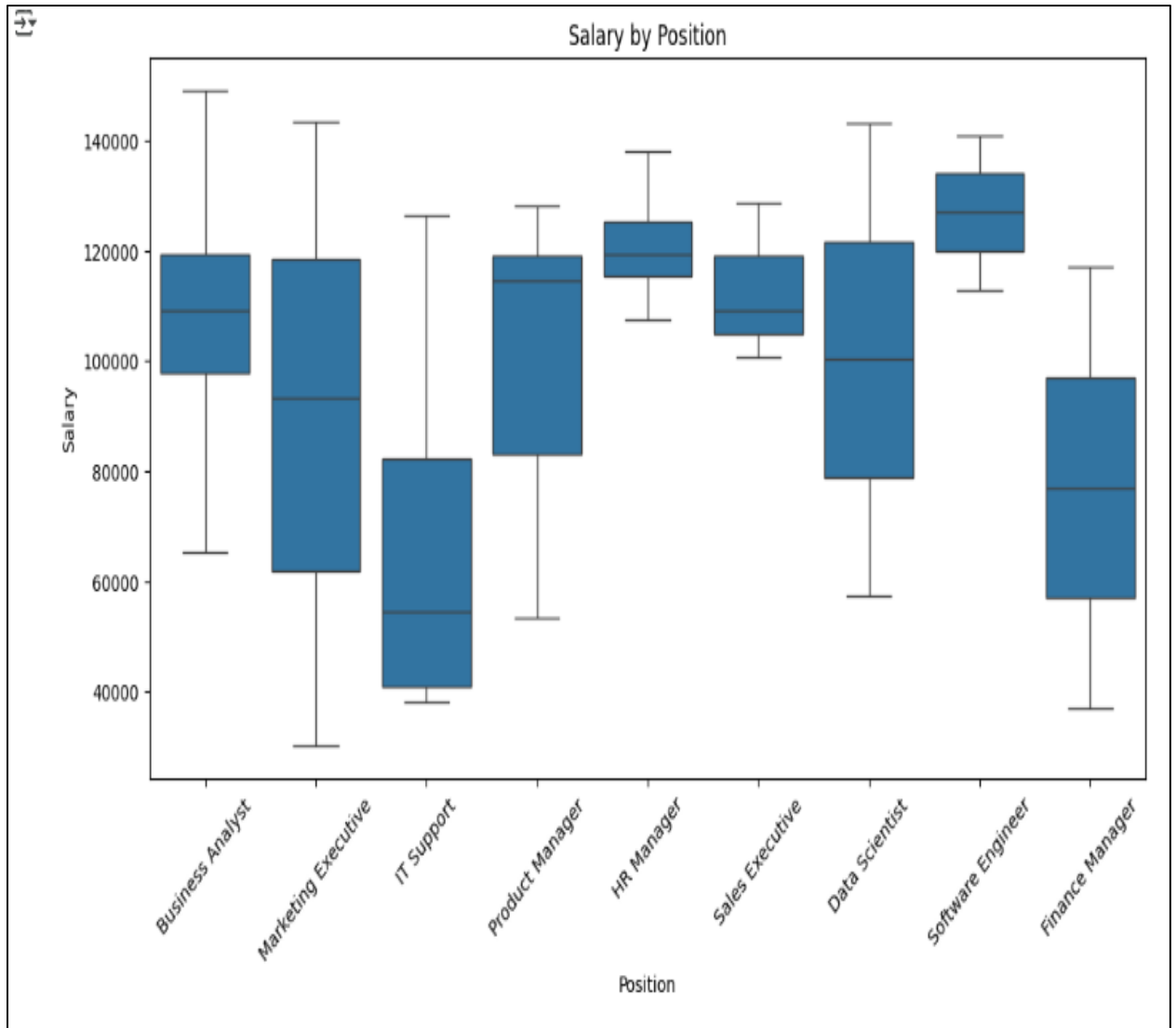
4. Salary Distribution

```
plt.figure(figsize=(8, 6))  
sns.histplot(df['Salary'], bins=30, kde=True)  
plt.title("Salary Distribution")  
plt.xlabel("Salary")  
plt.ylabel("Frequency")  
plt.show()
```



5. Salary by position

```
plt.figure(figsize=(12, 6))  
sns.boxplot(x='Position', y='Salary', data=df)  
plt.xticks(rotation=45)  
plt.title("Salary by Position")  
plt.show()
```



6. Scatter Plot

```
plt.figure(figsize=(8, 6))  
sns.scatterplot(x='YearsExperience', y='Salary', hue='Position',  
data=df)  
plt.title("Experience vs Salary")  
plt.xlabel("Years of Experience")  
plt.ylabel("Salary")  
plt.show()
```

