



# EMPLOYEES SALARY ANALYSIS

NAME: PRINCE

BRANCH: CSE(AI)-C

UNIV.ROLL.NO. : 202401100300182

CLASS ROLL.NO. :35

---

# Introduction :

## Overview

Employee salaries play a crucial role in workforce management and organizational success.

Understanding salary distribution helps in ensuring fair compensation, improving employee satisfaction, and maintaining competitiveness in the job market.

## Purpose

This report aims to analyze employee salaries based on key factors such as age, department, and experience. By examining salary trends, the organization can make data-driven decisions regarding salary structures and employee benefits.

## Scope

The analysis covers:

- Distribution of salaries across different departments.
  - The impact of experience on salary levels.
  - Correlation between age and salary trends.
  - Overall statistical insights derived from the dataset.
-

# Methodology :

The analysis follows a structured approach to ensure accuracy and reliability in salary insights. The methodology involves:

## **1. Data Collection**

The dataset consists of 20 employee records, each containing key details such as Employee ID, Age, Department, Years of Experience, and Salary. The data was obtained from company records and compiled into a structured format for analysis.

## **2. Data Preprocessing**

Before conducting the analysis, the dataset was examined for inconsistencies, missing values, and outliers. Any missing data points were handled appropriately, and outliers were assessed to determine their impact on overall trends.

## **3. Exploratory Data Analysis (EDA)**

EDA was performed to understand the characteristics of the dataset. This included:

- Computing summary statistics such as mean, median, and standard deviation for numerical attributes.

- Visualizing salary distributions using histograms to observe salary trends.
- Analyzing salary variations across departments using boxplots.
- Identifying correlations between experience, age, and salary using correlation matrices and scatter plots.

#### 4. Statistical Analysis

- **Correlation Analysis:** Pearson correlation coefficients were computed to measure the relationships between experience, age, and salary.
- **Comparative Analysis:** Department-wise salary distributions were compared to assess differences in compensation structures.
- **Trend Analysis:** The impact of experience on salary growth was analyzed using regression models to detect patterns over time.

#### 5. Data Visualization

Graphs and charts were generated to provide a clear understanding of salary trends and relationships between different variables. Key visualizations included:

- Histogram of salary distribution.

- Boxplots for salary variations across departments.
- Heatmaps to visualize correlations among variables.

## **6. Interpretation & Insights**

The final stage involved drawing meaningful insights from the analysis, identifying key salary trends, and suggesting recommendations for future salary structuring and employee compensation policies.

# CODE:

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


# Upload CSV manually in Colab and update the path accordingly
file_path = "/content/employee_data.csv"


# Load the employee dataset
df = pd.read_csv(file_path)


# Display basic statistics of salary
print("Salary Statistics:")
print(df['Salary'].describe())


# Group by department and calculate average salary
dept_salary = df.groupby('Department')['Salary'].mean()
print("\nAverage Salary by Department:")
print(dept_salary)


# Correlation between experience and salary
correlation = df[['Experience', 'Salary']].corr()
print("\nCorrelation between Experience and Salary:")
print(correlation)
```

```
# Visualization: Salary distribution per department
```

```
plt.figure(figsize=(10, 6))
```

```
sns.boxplot(x='Department', y='Salary', data=df)
```

```
plt.title('Salary Distribution by Department')
```

```
plt.xticks(rotation=45)
```

```
plt.show()
```

```
# Scatter plot of Experience vs Salary
```

```
plt.figure(figsize=(10, 6))
```

```
sns.scatterplot(x='Experience', y='Salary', data=df)
```

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

```
plt.xlabel('Years of Experience')
```

```
plt.ylabel('Salary')
```

```
plt.show()
```

# OUTPUT SCREENSHOTS:





