EMPLOYEE SALARY ANALYSIS

By Ayushman Bhowmik

INTRODUCTION

- Analysis tool developed to reveal compensation patterns across departments, age groups, and experience levels
- Identifies salary inequities and experience-to-compensation relationships
- Supports data-driven decisions for HR planning and retention strategies
- Makes complex salary patterns accessible through statistics and visualization

METHODOLOGY

- Analyzed data from 20 employees across Finance, HR, IT, and Sales
- Segmented by department, age brackets (20-30, 31-40, etc.), and experience levels
- Applied statistical analysis: means, medians, standard deviations, correlations
- Developed custom metrics: experience-to-age ratio and salary per year of experience
- Created visualizations showing distributions, trends, and comparisons

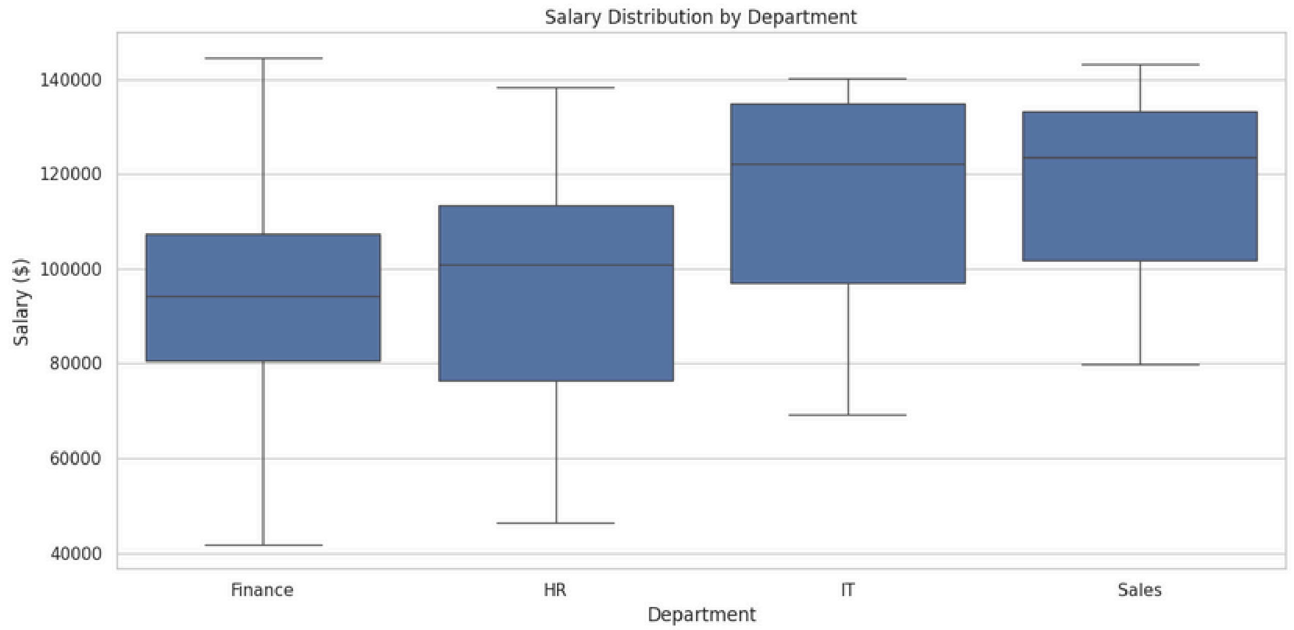
CODE

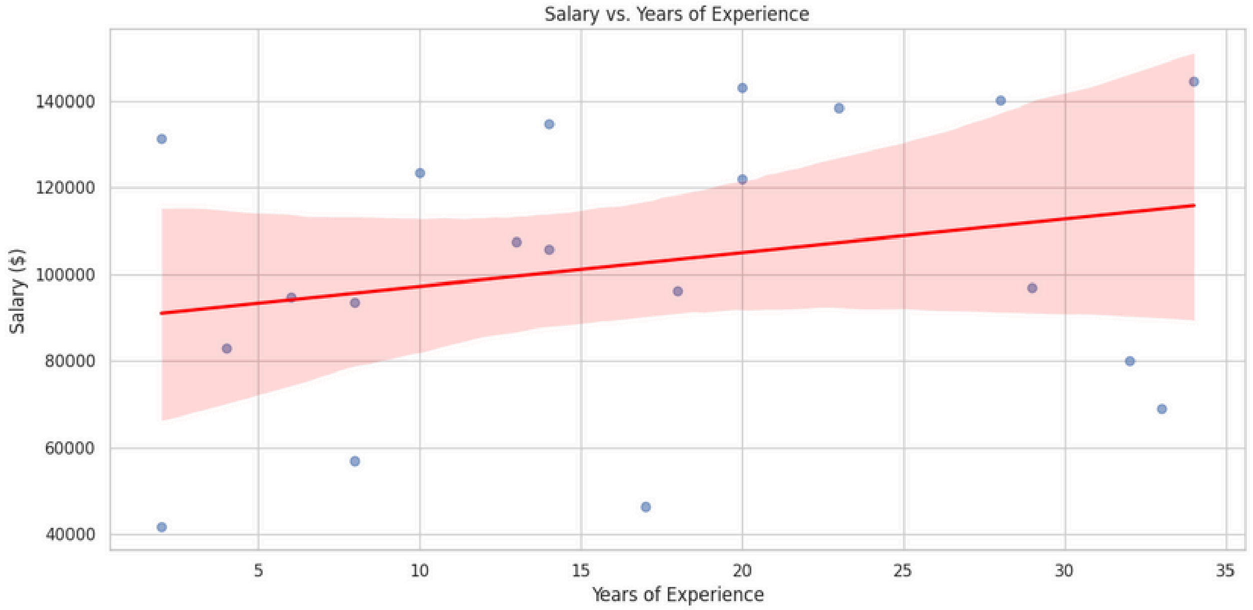
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

data = {
    'EmployeeID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20],
    'Age': [23, 28, 37, 23, 55, 32, 58, 46, 53, 58, 29, 46, 49, 57, 53, 57, 43, 29, 23, 50],
    'Department': ['Finance', 'Finance', 'HR', 'HR', 'IT', 'Sales', 'Finance', 'Finance',
                  'HR', 'HR', 'HR', 'HR', 'IT', 'Sales', 'IT', 'HR', 'HR', 'Sales', 'IT', 'IT'],
    'Experience': [8, 2, 8, 23, 29, 10, 6, 34, 2, 17, 13, 14, 20, 32, 33, 4, 18, 20, 14, 28],
    'Salary': [93563, 41742, 56905, 138397, 96879, 123436, 94781, 144637, 131361,
              46377, 107468, 105752, 122125, 79949, 69121, 83010, 96227, 143220, 134907, 140206]
}

df = pd.DataFrame(data)

def analyze_employee_data(df):
    print("Employee Salary Analysis\n" + "="*25 + "\n")
print("1. Basic Statistics:")
    print(f"Total number of employees: {len(df)}")
    print(f"Average salary: ${df['Salary'].mean():.2f}")
    print(f"Median salary: ${df['Salary'].median():.2f}")
    print(f"Highest salary: ${df['Salary'].max():.2f}")
    print(f"Lowest salary: ${df['Salary'].min():.2f}")
    print(f"Salary standard deviation: ${df['Salary'].std():.2f}\n")

    print("2. Department Analysis:")
    dept_analysis = df.groupby('Department').agg({
        'EmployeeID': 'count',
        'Salary': ['mean', 'median', 'min', 'max', 'std']
    }).round(2)

    dept_analysis.columns = ['Count', 'Avg Salary', 'Median Salary', 'Min Salary', 'Max Salary', 'Std Dev']
    print(dept_analysis)
    print()

    print("3. Age Group Analysis:")
    df['Age Group'] = pd.cut(df['Age'], bins=[20, 30, 40, 50, 60],
                            labels=['20-30', '31-40', '41-50', '51-60'])

    age_analysis = df.groupby('Age Group').agg({
        'EmployeeID': 'count',
        'Salary': ['mean', 'median']
    }).round(2)
```

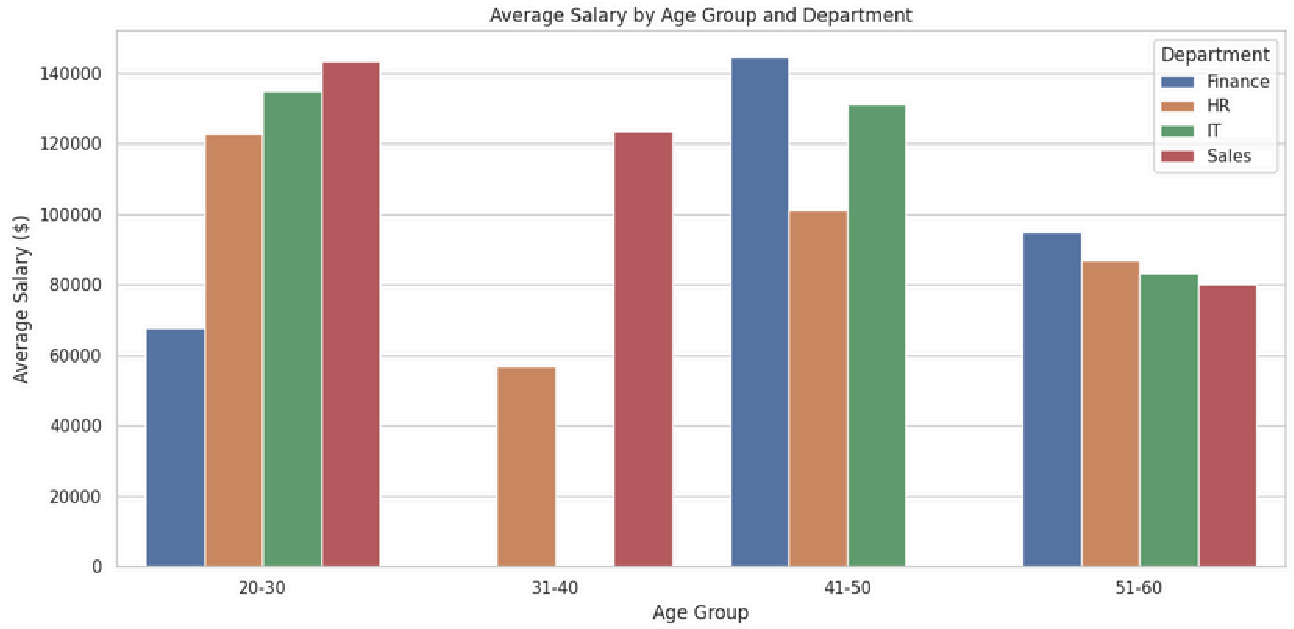
```
print("4. Experience vs Salary Correlation:")
    correlation = df['Experience'].corr(df['Salary'])
    print(f"Correlation coefficient: {correlation:.2f}")
    print(f"This indicates a {'strong' if abs(correlation) > 0.7 else 'moderate' if abs(correlation) > 0.3 else 'weak'} "
          f"{'positive' if correlation > 0 else 'negative'} correlation between experience and salary.\n")

    print("5. Experience Group Analysis:")
    df['Experience Group'] = pd.cut(df['Experience'], bins=[0, 5, 10, 20, 35],
                                  labels=['0-5 yrs', '6-10 yrs', '11-20 yrs', '21-35 yrs'])

    exp_analysis = df.groupby('Experience Group').agg({
        'EmployeeID': 'count',
        'Salary': ['mean', 'median', 'min', 'max']
    }).round(2)

    exp_analysis.columns = ['Count', 'Avg Salary', 'Median Salary', 'Min Salary', 'Max Salary']
    print(exp_analysis)
    print()

    print("6. Top 5 Highest Paid Employees:")
    top_earners = df.sort_values(by='Salary', ascending=False).head(5)
    print(top_earners[['EmployeeID', 'Age', 'Department', 'Experience', 'Salary']])
    print()
```

```
def create_visualizations(df):
    sns.set(style="whitegrid")

    plt.figure(figsize=(12, 6))
    sns.boxplot(x='Department', y='Salary', data=df)
    plt.title('Salary Distribution by Department')
    plt.ylabel('Salary ($)')
    plt.tight_layout()
    plt.savefig('salary_by_department.png')

    plt.figure(figsize=(12, 6))
    sns.regplot(x='Experience', y='Salary', data=df, scatter_kws={'alpha':0.6}, line_kws={'color':'red'})
    plt.title('Salary vs. Years of Experience')
    plt.xlabel('Years of Experience')
    plt.ylabel('Salary ($)')
    plt.tight_layout()
    plt.savefig('salary_vs_experience.png')

    plt.figure(figsize=(12, 6))
    sns.barplot(x='Age Group', y='Salary', hue='Department', data=df, errorbar=None)
    plt.title('Average Salary by Age Group and Department')
    plt.ylabel('Average Salary ($)')
    plt.tight_layout()
    plt.savefig('salary_by_age_department.png')
```

OUTPUTS


Salary Distribution by Department
Salary ($)
140000
120000
100000
80000
60000
40000
Finance
HR
IT
Sales
Department


Salary vs. Years of Experience
Salary ($)
140000
120000
100000
80000
60000
40000
5
10
15
20
25
30
35
Years of Experience


Average Salary by Age Group and Department
Average Salary ($)
140000
120000
100000
80000
60000
40000
20000
0
20-30
31-40
41-50
51-60
Age Group
Department
Finance
HR
IT
Sales

THANK YOU