

Salary Data Analytics Report

1. Introduction:

- The purpose of this project is to analyze a salary dataset using **PySpark**. By applying distributed data processing techniques, we explored workforce demographics, education-based salary differences, gender-wise compensation trends, and the impact of work experience on salary growth.
- The results are visualized through scatter plots, histograms, bar charts, and boxplots to derive insights into salary distribution and compensation dynamics.

2. Dataset Overview:

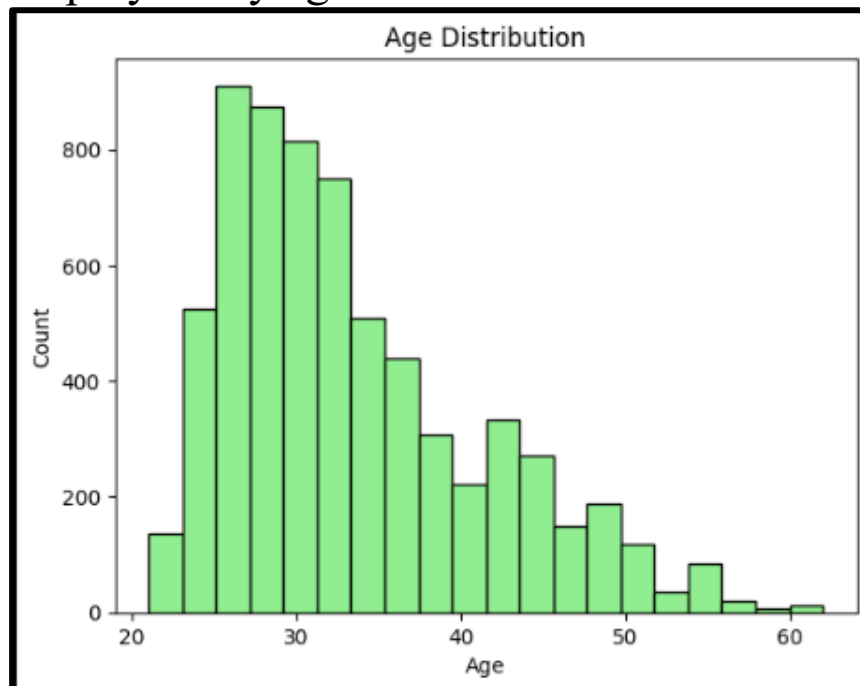
- Entries: ~1000 employee salary records
- Columns:
 - **Age** – Age of employee
 - **Gender** – Gender of employee
 - **Education Level** – Education background (e.g., Bachelor's, Master's, PhD)
 - **Job Title** – Designation or role of employee
 - **Years of Experience** – Total work experience in years
 - **Salary** – Annual salary in USD
- The dataset is clean and complete, with no missing values, making it reliable for analysis.

3. Key Findings:

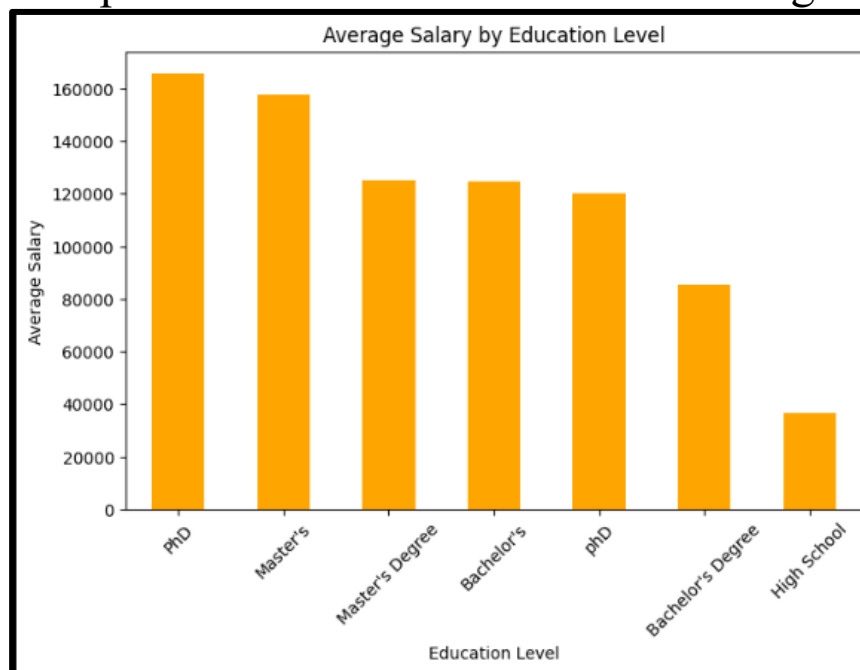
- **Workforce Composition:**
 - Employees are distributed across various education levels, with the majority having Bachelor's and Master's degrees.
 - Job titles show a diverse workforce, including Developers, Managers, Analysts, and Testers.
- **Salary and Education Analysis:**
 - Employees with higher education levels (Master's, PhD) earn significantly higher salaries.
 - Bachelor's degree holders earn comparatively less, highlighting the role of education in career growth.
- **Gender and Salary Trends:**
 - Male and female employees show differences in salary distribution.
 - Gender-based salary gaps exist, with male employees generally earning more on average.
- **Experience and Compensation:**
 - A strong positive correlation exists between Years of Experience and Salary.
 - Salaries increase steadily with experience up to mid-career, then plateau after ~15–20 years.
- **Age and Compensation:**
 - Younger employees cluster at lower salary ranges, while older employees dominate higher salary levels.
 - Salary growth is visible through the 20s and 30s, stabilizing in the 40s and beyond.

4. Data Visualization (DV):

- Using **PySpark with Matplotlib/Seaborn**, the following charts were generated:
 - **Age Distribution (Histogram):** Shows spread of employees by age.



-
- **Average Salary by Education Level (Bar Chart):** Compares salaries across education backgrounds.



○

- **Salary Distribution by Gender (Boxplot):**
Highlights compensation spread across genders.



-
- **Salary vs Years of Experience (Scatter Plot):**
Demonstrates the positive correlation between experience and salary.



5. Conclusion:

- Education level and work experience are strong predictors of salary growth.
- Gender-based salary gaps remain visible, requiring organizational attention.
- Younger employees earn lower salaries, while mid-career professionals achieve higher compensation.
- Average salary increases steadily with age and experience but plateaus after mid-career.
- The dataset highlights the importance of **education, experience, and age** in shaping compensation trends.