B. Sushma Sri Durga

Salary Data Analysis Report

1. Dataset Description

1.1 Source:

The dataset used for this project is Salary_Data.csv, containing information about 1,000 employees.

It captures key details required for salary analysis.

1.2 Columns:

- **Age:** Age of the employee
- **Gender:** Male or Female
- **Education_Level:** Qualification of the employee (e.g., Bachelor's, Master's etc.)
- **Job_Title:** Role/designation held by the employee
- **Years_of_Experience:** Employee Work experience in years
- **Salary:** Annual salary of the employee

1.3 Data Quality:

- The dataset is clean.
- It has no missing values.
- Data is balanced with variation across gender, education level, and job titles.

2. Operations Performed

2.1 Data Loading and Exploration

- Created a Spark session and loaded the dataset.
- Verified the schema and displayed the first 10 rows using .show().
- Counted total rows and columns, confirming 1,000 records.

2.2 Data Inspection

- Selected specific columns (age, salary) to focus on salary-related attributes.
- Checked salary distribution and basic statistics.

2.3 Filtering Operations

- Filtered employees with salary > 50,000 to identify high earners.
- Filtered female employees with more than 5 years of experience.

2.4 Aggregation and Summary Statistics

- Calculated average, minimum, and maximum salary using PySpark aggregate functions.
- Computed average salary by gender to study pay differences.
- Calculated average salary by education level to assess the impact of qualification on salary.

3. Key Insights

3.1 Salary Distribution

- Salaries vary widely across the dataset, showing diversity in roles and experience levels.
- Average salary is around the mid-level range, with notable outliers at both ends.

3.2 Gender-Based Insights

- Average salary differs by gender, indicating possible pay gaps or differences in job roles.
- Male employees tend to have slightly higher average salaries than female employees.

3.3 Education-Level Insights

- Higher education levels are associated with higher average salaries.
- Postgraduates and specialized degree holders show significantly better salary outcomes compared to undergraduates.

3.4 Experience Insights

- Employees with more years of experience generally have higher salaries, confirming the expected positive correlation.

4. Recommendations

4.1 Compensation Strategy

- Conduct deeper analysis into gender-based salary differences to ensure pay equity.
- Align salary bands closely with experience and education levels to encourage skill development.

4.2 Workforce Planning

- Encourage continuous learning programs to help employees upgrade qualifications and improve salary outcomes.
- Use salary data to identify top talent and plan retention strategies.

4.3 Future Analytics Opportunities

- Implement predictive models to forecast salary growth trends based on experience and education.
- Use clustering analysis to group employees by salary ranges for more targeted HR strategies.

5. Conclusion

The analysis of the dataset highlights that salary is strongly influenced by experience, education, and job role. More experienced and highly qualified employees tend to earn higher pay, while senior or specialized positions command the top salaries. The dataset's clean structure and diverse workforce make it ideal for understanding salary trends, identifying key pay drivers, and building predictive models.

Overall, these insights provide a clear picture of the company's compensation patterns and can help guide HR decisions, salary planning, and career development strategies.