

Data Analysis Report

Name: Ashok kumar meena

Enrollment No.: 23114011

1.Introduction

This report presents an in-depth exploratory data analysis (EDA) of the given dataset, which contains information about employees, including their ID, Name, Age, Gender, Salary, Department, Experience, and Performance Score. The goal of this analysis is to understand relationships between different variables and uncover patterns that may explain salary distribution, experience trends, and performance insights.

2.Dataset Overview

The dataset contains 29 entries and 8 columns, including:

- ID: Unique identifier for each employee.
- Name: Name of the employee.
- Age: Age of the employee.
- Gender: Male or Female.
- Salary: Monthly salary of the employee.
- Department: The department where the employee works.
- Experience: Number of years of experience.
- Performance Score: Employee performance rating (Scale: 1-10).

3.Data Cleaning and Preprocessing

1. Handling Missing Values:

- The Salary column had 1 missing value, which was imputed using the median salary to preserve data consistency.

2. Identifying & Removing Duplicates:

- Duplicate records were found for employee Alice and removed.

3. Outlier Removal:

- Interquartile Range (IQR) method was used to identify and remove extreme salary values.

4. Standardization of Categorical Data:

- Converted all Department and Gender values to lowercase for consistency.

5. Encoding Categorical Data:

- Categorical columns Gender and Department were encoded using Label Encoding to facilitate numerical analysis.

4.Univariate Analysis

Numerical Variables:

- Salary: Mean = 5830, Median = 5500, Variance = 650000, Skewness = 1.18
- Experience: Mean = 7.2, Median = 6, Variance = 14.6, Skewness = 0.79

Categorical Variables:

- Gender: Male (15), Female (14)
- Department: IT (7), Finance (7), HR (7), Marketing (8)
- Visualizations:
 - Histograms revealed a right-skewed salary distribution, indicating that most employees earn below the highest salaries.
 - Box plots highlighted outliers in Salary and Experience.
 - Bar charts showed a balanced gender and department distribution in the dataset.

5.Bivariate Analysis

Correlation Analysis:

- Salary vs Experience: Positive correlation ($r = 0.75$), indicating that employees with more experience tend to earn higher salaries.
- Salary vs Performance Score: Weak correlation ($r = 0.32$), suggesting performance does not strongly impact salary.

Salary Distribution by Gender:

- Males had slightly higher median salaries than females.
- No extreme gender salary gaps were detected in this dataset.

Salary Distribution by Department:

- Finance and IT departments had the highest median salaries.
- Marketing had the lowest salary distribution.

Visualizations:

- Scatter Plot: Showed an upward trend between Experience and Salary.
- Box Plot: Compared salary distributions across Departments and Gender.
- Bar Plot: Showed that the IT and Finance departments had the highest average salaries.

6. Multivariate Analysis

1. Pair Plot:

- Helped visualize relationships between Salary, Experience, and Performance Score.
- Confirmed the positive correlation between Salary and Experience.

2. Heatmap:

- Highlighted a strong correlation ($r = 0.75$) between Salary and Experience.
- Showed minimal correlation between Salary and Performance Score ($r = 0.32$).

3. Grouped Comparisons:

- Box Plots: Revealed how salary distributions varied across different departments and genders.
- Pivot Table: Showed the average salary per department grouped by gender.

7. Key Findings & Interpretations

1. Salary Trends:

- Employees with higher experience tend to have higher salaries, as seen in the correlation analysis.

- Finance and IT departments pay the highest salaries, while Marketing has the lowest salary distribution.

2. Gender Pay Insights:

- No significant gender-based salary disparity was observed.

3. Performance vs Salary:

- Performance Score does not significantly impact Salary, meaning high performers do not necessarily earn more.

4. Outlier Impact:

- Certain extreme salaries were removed as outliers, ensuring more accurate analysis.

End Of Report