Exploratory Data Analysis (EDA) Report

Project: Intelligent Payroll System using Machine Learning

This report presents a comprehensive Exploratory Data Analysis (EDA) performed on payroll data. The objective is to understand salary patterns, employee attendance, experience, attrition trends, and detect anomalies before building machine learning models.

1. Dataset Overview

The payroll dataset consists of employee demographic details, attendance records, salary components, and employment-related attributes. The dataset is used to support salary prediction, attrition analysis, and payroll anomaly detection.

Key Attributes Analyzed:

Salary / Gross Pay
Department
Employee Experience
Attendance Days
Leave Days
Working Days
Attrition Status
Date-related payroll information

2. Salary Distribution Analysis

Salary distribution analysis helps identify typical pay ranges and extreme values. Most employees fall within a standard salary band, while a few outliers represent senior roles or potential payroll inconsistencies.

3. Experience vs Salary Relationship

A positive relationship between employee experience and salary is observed. As experience increases, salary generally rises, validating the fairness of the payroll structure. Any deviation from this pattern may indicate data quality issues or exceptional compensation cases.

4. Attendance & Leave Pattern Analysis

Attendance days and leave days directly impact payroll calculations. Employees with unusually low attendance but high salary are flagged for further investigation. Negative or very low working days indicate attendance or data entry errors.

5. Attrition Analysis

Attrition analysis highlights employees who have left or may leave the organization. Understanding attrition trends helps HR teams design retention strategies and improve workforce stability.

6. Correlation Analysis

Correlation analysis reveals how different features relate to salary and attrition. Strong correlations with salary include experience, attendance, and role-related attributes. These insights guided feature selection for machine learning models.

7. Payroll Anomaly Detection

Anomaly detection was performed using an unsupervised learning approach. Records flagged as anomalies represent unusual salary values, incorrect attendance data, or potential payroll fraud. Early detection prevents financial loss and compliance issues.

8. Business Insights & Impact

- Improves payroll accuracy and transparency
- Identifies salary inconsistencies and fraud
- Supports HR decision-making with data-driven insights
- Enhances employee retention strategies
- Reduces manual payroll auditing effort

Conclusion

The Exploratory Data Analysis provided critical insights into payroll data quality, salary structure, and workforce behavior. These findings laid a strong foundation for building reliable machine learning models for salary prediction, attrition forecasting, and anomaly detection.