HR Attrition Analysis & Prediction Report

Abstract

Employee attrition is a critical challenge for organizations as it directly impacts productivity, cost, and workforce stability. This project analyses employee attrition data to identify key factors influencing employee turnover and to predict attrition risk using data analytics and machine learning techniques. Python is used for data cleaning, exploratory data analysis, and predictive modeling, while Power BI is used to design an interactive dashboard for business insights. The project helps HR teams proactively identify high-risk employees and design effective retention strategies.

Introduction

Employee attrition refers to the gradual reduction of employees due to resignation, retirement, or other reasons. High attrition rates can lead to increased hiring costs, loss of skilled talent, and reduced organizational performance.
The objective of this project is to analyse historical HR data to understand attrition patterns across departments, job roles, and salary bands, and to predict the likelihood of employee attrition. The insights generated from this analysis can support data-driven HR decision-making.

Tools Used

- **Python** – Data analysis and machine learning
  - Pandas, NumPy (data manipulation)
  - Matplotlib, Seaborn (data visualization)
  - Scikit-learn (Logistic Regression model)
- **Jupyter Notebook** – Analysis and model development
- **Power BI** – Interactive dashboard and visualization
- **CSV Dataset** – HR employee data

Steps Involved in Building the Project

1. **Data Collection**
   - Imported HR dataset containing employee demographics, job roles, salary details, and attrition information.
2. **Data Cleaning & Pre-processing**

    - Checked for missing values and data types.
    - Encoded categorical variables.
    - Scaled numerical features where required.
3. **Exploratory Data Analysis (EDA)**
    - Analysed overall attrition rate.
    - Studied attrition by department, job role, and salary band.
    - Visualized attrition probability distribution and correlations.
4. **Predictive Modeling**
    - Built a **Logistic Regression model** to predict employee attrition.
    - Generated attrition probability scores.
    - Classified employees into **Low, Medium, and High Risk** bands.
5. **Power BI Dashboard Development**
    - Created KPI cards for total employees, attrition count, and attrition rate.
    - Designed charts for department-wise, role-wise, and salary-based attrition.
    - Added slicers for department, job role, and risk band.
    - Included a dedicated **Key Insights** page.

Conclusion

The analysis reveals an overall attrition rate of **13.27%**, indicating moderate employee turnover. Certain departments and job roles exhibit higher attrition, and employees in **low and medium salary bands** are more likely to leave the organization. Predictive analysis highlights a large **medium-risk employee group**, enabling early HR intervention.
By combining Python-based analytics with Power BI visualization, this project demonstrates how data-driven approaches can help organizations reduce attrition and improve employee retention strategies.