

# **Summary (IBM Employee Attrition Prediction)**

**Objective:** The project aims to analyze the factors that lead to employee attrition within a company and predict which employees are likely to leave based on these factors. The goal is to help the organization understand the reasons behind employee turnover and take measures to enhance workforce retention and organizational performance.

**Context:** Employee attrition refers to the process of employees leaving the company, either voluntarily or involuntarily, and not being replaced immediately. High employee turnover is a significant challenge for businesses as it negatively impacts productivity and the ability to meet corporate objectives. Understanding the factors influencing attrition is crucial for organizations to reduce turnover rates and maintain a competitive edge.

## **Data Description:**

The dataset used for the analysis contains information on IBM employees and includes features such as:

- **Demographics:** Age, Department, Education, Education Field, Marital Status
- **Work-related Factors:** Job Satisfaction, Work-Life Balance, Monthly Income, Years at Company, Number of Companies Worked
- **Other Factors:** Distance from Home, Environmental Satisfaction

## **Solution Overview:**

To achieve this objective, a logistic regression model was implemented using various employee-related features. The following steps were undertaken to develop the solution:

## **1. Data Analysis and Preprocessing:**

- Explored the dataset, which included factors such as age, department, education, marital status, monthly income, job satisfaction, work-life balance, and more.
- Categorical variables were converted into numerical formats for compatibility with machine learning algorithms.
- Visualizations, including histograms and bar charts, were used to analyze distributions and relationships among factors affecting attrition.

## **2. Model Development:**

- A logistic regression model was built to predict the probability of employee attrition.
- Training and testing datasets were created, and the model was trained using features such as age, department, distance from home, education, education field, and years at the company.
- The model's accuracy on the training set was approximately 84%, indicating its effectiveness in predicting attrition.

## **3. Model Evaluation:**

- The model's performance was evaluated using metrics like accuracy, precision, recall, and F1-score.
- Confusion matrix analysis showed that the model accurately predicted most cases of employees who would not leave but had limitations in predicting those who would.
- ROC-AUC scores indicated a moderate capability of distinguishing between employees who stayed and those who left.

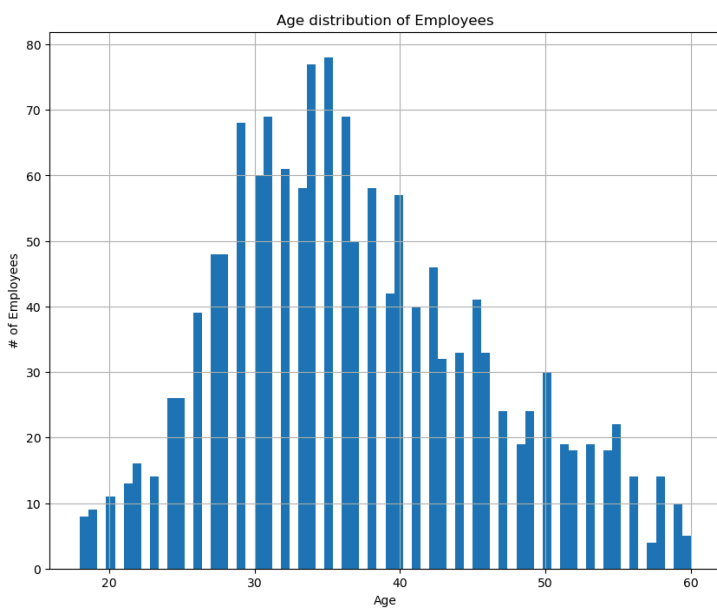
## **4. Key Findings and Recommendations:**

- Employees with lower job satisfaction, longer commute distances, and lower monthly incomes were more likely to leave the organization.
- The model's insights suggested the importance of addressing work-life balance, job satisfaction, and competitive compensation to reduce attrition rates.

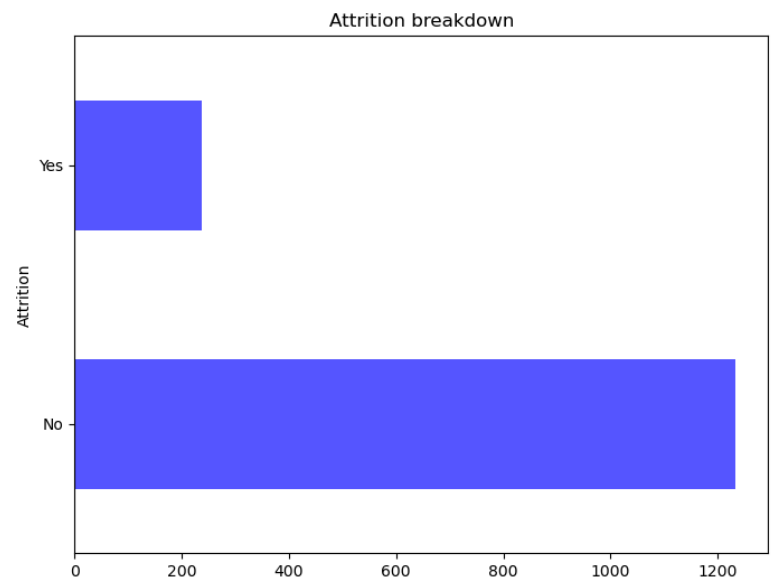
## **Conclusion**

The model successfully identified the probability of attrition for each employee, helping the organization to proactively address issues related to workforce stability. Insights gained from the analysis can be used to implement targeted retention strategies and improve employee engagement.

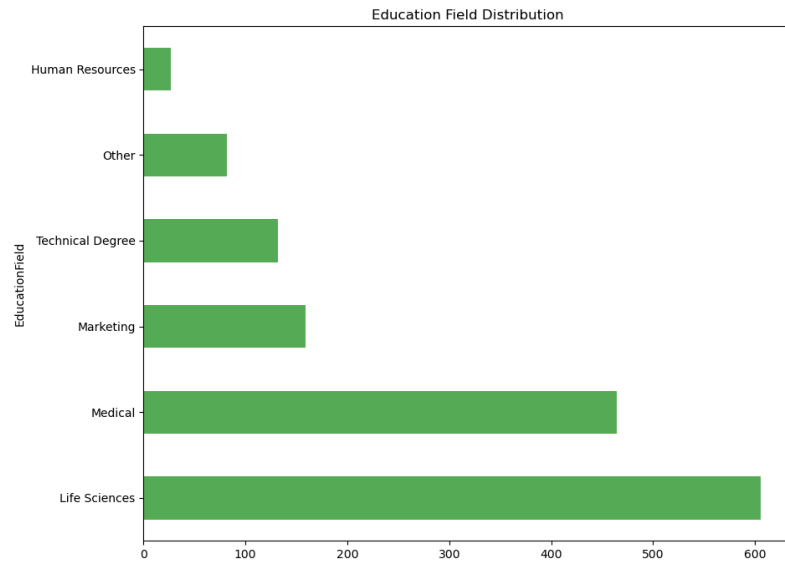
## **Visualization of Data (Output)**



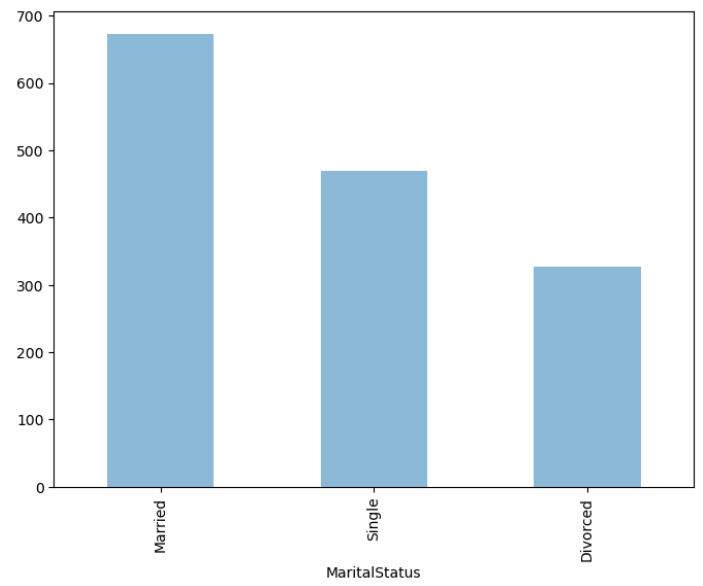
**Age Distribution of Employees**



**Attrition Breakdown**



**Education Field Distribution**



**Marital Status**