

# Employee Attrition Analysis

## Introduction

Employee attrition, or turnover, is a critical challenge for organizations. High attrition rates can lead to increased hiring costs, loss of knowledge, and decreased team morale. This project aims to identify key factors contributing to attrition using data-driven methods and help the HR department make informed decisions to reduce turnover.

## Abstract

The goal of this project is to build a predictive model to identify employees at risk of attrition and understand the key factors driving attrition. We applied exploratory data analysis (EDA), feature engineering, and built a classification model (Logistic Regression and/or Random Forest) using the IBM HR Analytics Employee Attrition dataset. The outcome of this model is evaluated using a confusion matrix and other classification metrics. Visualizations and summaries are prepared for further interpretation via tools like Power BI.

## Tools Used

- Programming Language: Python (Jupyter Notebook)
- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- Data Visualization: Power BI and Matplotlib

## Steps Involved in Building the Project

1. Data Collection:
  - IBM HR Analytics Attrition Dataset was used.
2. Data Cleaning & Preprocessing:
  - Handled missing values, encoded categorical variables, and created new features like AgeGroup, IncomeGroup.
3. Exploratory Data Analysis (EDA):
  - Visualizations to identify attrition trends by department, job role, income, and age.
4. Model Building:

# Employee Attrition Analysis

- Trained classification models using Logistic Regression and Random Forest.

## 5. Model Interpretation:

- Identified important features (e.g., Job Role, Work-Life Balance, Overtime).

## 6. Report & Visualization:

- Summary tables, confusion matrix, and Power BI dashboards generated.

## Conclusion

This project demonstrates how machine learning and analytics can help identify key risk factors leading to employee attrition. By understanding these patterns, organizations can design proactive retention strategies. Power BI dashboards can provide dynamic insights for HR executives, while the classification model can help flag high-risk employees in real-time.