# HR Analytics - Predict Employee Attrition

## INTRODUCTION :

Employee attrition is one of the critical challenges faced by organizations as it increases hiring costs, reduces productivity, and impacts business continuity. To address this, organizations need data-driven strategies that help identify key attrition factors and predict employees at risk of leaving.

This project uses the IBM HR Analytics Dataset to analyze attrition patterns and build predictive models. The study integrates Python-based machine learning and Power BI visualizations to uncover insights and suggest actionable recommendations for retention.

---

## ABSTRACT :

The primary objective of this project is to analyze employee attrition and identify the major factors influencing it. Exploratory Data Analysis (EDA) revealed trends across departments, job roles, compensation, and tenure. Predictive models such as Logistic Regression and Decision Tree Classifier were implemented to classify employees based on attrition risk.

Additionally, Power BI dashboards were designed to provide interactive visualization of attrition by department, salary levels, work-life balance, and years at the company. These insights can guide HR teams in making informed decisions to reduce attrition, enhance employee engagement, and improve organizational performance.

---

## TOOLS USED :

- **Python** → Data preprocessing, model building, evaluation (Logistic Regression, Decision Tree)
- **Libraries** → pandas, numpy, matplotlib, seaborn, scikit-learn, SHAP
- **Power BI** → Interactive dashboards for visualizing attrition patterns
- **Dataset** → IBM HR Analytics Employee Attrition Dataset

---

## STEPS INVOLVED IN BUILDING THE PROJECT :

1. **Dataset Loading & Cleaning**

   - Loaded `IBM HR.csv` dataset (1,470 records, 35 features).
   - Dropped irrelevant columns (`EmployeeNumber`, `EmployeeCount`, `Over18`, `StandardHours`).
   - Encoded categorical variables using Label Encoding.

2. **Exploratory Data Analysis (EDA)**

   - Analyzed attrition distribution across departments and job roles.
   - Compared attrition with monthly income, work-life balance, and years at the company.
   - Found that younger employees, low-income groups, and sales/research departments face higher attrition.

3. **Model Building**

- Split dataset into train/test (80:20).
- Trained **Logistic Regression** and **Decision Tree** models.
- Evaluated models using Accuracy, Confusion Matrix, Precision, Recall, and F1-score.
- Logistic Regression achieved higher accuracy compared to Decision Tree.

4. **Feature Importance & Explainability**

- Decision Tree feature importance highlighted Monthly Income, Job Role, and Years at Company as key predictors.
- SHAP analysis provided interpretability of predictions.

5. **Power BI Dashboard**
- Created KPIs: Total Employees, Employees Left, Attrition Rate.
- Built visualizations: Attrition by Department, Job Role, Salary, Age, and Work-Life Balance.
- Added slicers for Gender, Department, and Job Role.

---

# CONCLUSION :

The project successfully identified key drivers of attrition such as compensation, job role, work-life balance, and career progression. Predictive modeling demonstrated that machine learning can provide reliable estimates of attrition risk, while Power BI dashboards offered HR teams actionable insights in an interactive manner.

By implementing targeted strategies like salary restructuring, improved onboarding, mentorship for new employees, and flexible work policies, organizations can significantly reduce attrition rates and improve workforce stability.