# HR Analytics — Predicting Employee Attrition

**Introduction**:

This project analyses HR records to identify drivers of employee attrition and to build a predictive model that flags high-risk employees. The aim is to support targeted retention actions and reduce turnover costs.

**Abstract**:

Using Python (Pandas, Seaborn) for EDA, Power BI for visualization, and scikit-learn for modelling, we evaluated department-wise attrition, salary band impacts, and promotion effects. A classification model (Logistic Regression / Decision Tree) was trained to predict attrition and SHAP analysis was used to interpret model decisions

**Tools Used:**

- Python (Pandas, Seaborn) — data cleaning & EDA
- scikit-learn — Logistic Regression / Decision Tree modelling
- SHAP — model interpretability- Power BI — interactive dashboards and stakeholder reporting

**Steps in Project:**

1. Data cleaning & preprocessing (handled missing values, encoded categoricals).

2. EDA: department-wise attrition, salary band analysis, promotion (years since last promotion).

3. Feature engineering (e.g., Attrition_numeric, SalarySlab categories, AgeGroup).

4. Model training and evaluation (train/test split, scaling, imputation).

5. Explainability with SHAP and dashboarding in Power BI

## Key Findings (EDA):

- **Attrition rate:** 16% overall (higher than industry avg ~12%). Total attritions: 237 / 1,233 employees.
- **Top drivers:** low salary, high workload, lack of career growth; frequent travel and low job satisfaction are linked to higher attrition.
- **Department & role:** Sales (21%), R&D; (19%), HR (14%) show higher attrition; Sales Representatives (40%), Laboratory Technicians (24%), HR Specialists (23%) are high-risk roles.
- **Demographics:** Majority aged 26–35 and 36–45; gender ratio skewed (68% one gender); Education field dominated by Life Sciences & Medical (>50%).
- **Compensation:** Average salary: $6,879; highest averages in Managers and Research Directors.

## Power BI Dashboard