# HR Analytics: Predicting Employee Attrition

**1. Introduction**

Employee attrition, or employee turnover, is a critical challenge for organizations aiming to retain talent and minimize recruitment costs. Understanding why employees leave helps HR departments formulate better engagement and retention strategies. This project aims to identify key drivers of employee attrition and predict future attrition using machine learning.

**2. Abstract**

In this project, we analyze synthetic HR data to predict employee attrition using classification models. By conducting EDA, building machine learning models (Logistic Regression and Decision Tree), and visualizing the results with Power BI and Plotly, we aim to offer both predictive capability and strategic insights. The goal is to help companies proactively manage retention by focusing on high-risk segments.

**3. Dataset Description**

The dataset contains 500 employee records with the following features:  
- Demographics: Age, Gender  
- Job Info: Department, JobRole, YearsAtCompany, YearsSinceLastPromotion  
- Behavioral: JobSatisfaction, EnvironmentSatisfaction, OverTime  
- Target Variable: Attrition (Yes/No)

**4. Tools Used**

- Python (Pandas, Seaborn, Scikit-learn): Data analysis, modeling  
- Power BI: Dashboard visualizations  
- Plotly (in Colab): Interactive plots  
- Jupyter Notebook: Development environment

**5. Steps Involved in Building the Project**

Step 1: Exploratory Data Analysis (EDA)  
We found that employees working overtime had a 65% higher attrition rate. Attrition was more common among Sales and HR departments, and among employees with lower monthly income or job satisfaction.  
  
Step 2: Data Preprocessing  
Categorical variables were label-encoded, unnecessary columns were dropped, and the data was split into training and test sets.  
  
Step 3: Model Building & Evaluation  
- Logistic Regression: Accuracy ~81%  
- Decision Tree: Accuracy ~87%  
Models were evaluated using confusion matrix and classification metrics.  
  
Step 4: Visualization (Power BI & Plotly)  
Visuals include attrition ratio, department-wise attrition, income distributions, and interactive filters.

**6. Key Insights & Recommendations**

Insights:  
- Overtime is the most significant factor affecting attrition  
- Low job satisfaction and no promotions increase the risk  
- Employees under 3 years at a company are more likely to leave  
  
Recommendations:  
- Limit overtime and promote wellness  
- Offer performance-based incentives  
- Improve early employee engagement

**7. Conclusion**

Combining machine learning with BI dashboards provides a strong approach to analyzing and preventing attrition. Organizations can use these findings to make evidence-based HR decisions.

**8. Deliverables**

- HR\_Attrition\_Analysis.ipynb: Python code  
- Synthetic\_HR\_Attrition\_Dataset.csv: Raw data  
- Cleaned\_HR\_Attrition.csv: Dataset for dashboard  
- HR\_Attrition\_Dashboard.pbix: Power BI dashboard  
- Attrition\_Report.docx: Final report