

# HR Employee Attrition Analysis – Project Report

## INTRODUCTION

Employee attrition is a significant challenge for many organizations, as it directly impacts productivity, morale, and financial resources. Understanding the reasons behind employee exits and identifying the attributes of at-risk employees is crucial. This project focuses on analyzing HR data to uncover patterns associated with employee attrition using SQL and Python-based machine learning techniques.

## ABSTRACT

The project aims to analyze HR data to identify key factors contributing to employee attrition. SQL is used to perform descriptive analytics and extract meaningful insights from the dataset. Python is then used for data preprocessing, exploratory data analysis (EDA), classification modeling, and SHAP-based model explainability. The ultimate goal is to help the HR department understand attrition trends and take preventive measures.

## TOOLS USED

- **SQL:** Data exploration and aggregation (e.g., attrition rate, income, satisfaction).
- **Python:** Data preprocessing, machine learning modeling, and interpretability.
- **Libraries:** Pandas, Seaborn, Matplotlib, Numpy
- **Jupyter Notebook:** Python development environment.
- **Microsoft Word/Excel:** Documenting questions and output.
- **Power BI/Tableau (optional):** For visualization if extended.

## STEPS INVOLVED IN BUILDING THE PROJECT

### 1. Data Collection & Understanding:

- Used `HR-Employee-Attrition.csv` dataset with 35 columns and 1470 records.

### 2. SQL Analysis:

- Formulated 10 core SQL queries, including attrition counts, department-level breakdowns, top earners, and promotion eligibility.

### 3. Data Preprocessing (Python):

- Encoded categorical variables.
- Checked for null values and cleaned the dataset.

### 4. Exploratory Data Analysis (EDA):

- Identified relationships between attrition and variables like overtime, income, satisfaction.

### 5. Model Building:

- Used classification models (e.g., Random Forest, Logistic Regression).
- Evaluated performance using accuracy, precision, recall.

### 6. SHAP Analysis:

- Applied SHAP values to interpret feature importance and understand key drivers of attrition.

## CONCLUSION

This project provided a comprehensive understanding of employee attrition using a blend of SQL-based data querying and Python-based machine learning. Key factors such as overtime, job satisfaction, and years since last promotion were identified as strong indicators of attrition. These insights can help HR teams proactively manage talent and reduce turnover. Future enhancements could include deploying an interactive dashboard or implementing real-time attrition risk prediction.