**Problem Statement:**

Our client for this project is the HR Department at a software company. They want to try a new initiative to retain employees. The idea is to use data to predict whether an employee is likely to leave. Once these employees are identified, HR can be more proactive in reaching out to them before it's too late. They only want to deal with the data that is related to permanent employees. Current Practice Once an employee leaves, he or she is taken an interview with the name “exit interview” and shares reasons for leaving. The HR Department then tries and learns insights from the interview and makes changes accordingly. This suffers from the following problems: This approach is that it's too haphazard. The quality of insight gained from an interview depends heavily on the skill of the interviewer. The second problem is these insights can't be aggregated and interlaced across all employees who have left. The third is that it is too late by the time the proposed policy changes take effect. The HR department has hired us as data science consultants. They want to supplement their exit interviews with a more proactive approach.

**Data Description:**

The Business Intelligence Analysts of the Company provided us three datasets that contain information about past employees and their status (still employed or already left).

department\_data

This dataset contains information about each department. The schema of the dataset is as follows:

dept\_id – Unique Department Code

dept\_name – Name of the Department

dept\_head – Name of the Head of the Department

employee\_details\_data

This dataset consists of Employee ID, their Age, Gender and Marital Status. The schema of this dataset is as follows:

employee\_id – Unique ID Number for each employee

age – Age of the employee

gender – Gender of the employee

marital\_status – Marital Status of the employee

employee\_data

This dataset consists of each employee’s Administrative Information, Workload Information, Mutual Evaluation Information and Status.

Target variable

status – Current employment status (Employed / Left)

Administrative information

department – Department to which the employees belong(ed) to

salary – Salary level with respect to rest of their department

tenure – Number of years at the company

recently\_promoted – Was the employee promoted in the last 3 years?

employee\_id – Unique ID Number for each employee

Workload information

n\_projects – Number of projects employee has worked on

avg\_monthly\_hrs – Average number of hours worked per month

Mutual evaluation information

satisfaction – Score for employee’s satisfaction with the company (higher is better)

last\_evaluation – Score for most recent evaluation of employee (higher is better)

filed\_complaint – Has the employee filed a formal complaint in the last 3 years?