**EDA, Statistics and Machine Learning Case Study.**

*To predict the likelihood of a candidate leaving the company.*

A major problem in high employee attrition is its cost to an organization. Job postings, hiring processes, paperwork and new hire training are some of the common expenses of losing employees and replacing them. Additionally, regular employee turnover prohibits your organization from increasing its collective knowledge base and experience over time. This is especially concerning if your business is customer facing, as customers often prefer to interact with familiar people. Errors and issues are more likely if you constantly have new workers.

Data Dictionary:

satisfaction\_level: Employee satisfaction level

last\_evaluation: last evaluation score

number\_of\_projects: Number of projects for each employee

average\_monthly\_hours: monthly working hours of employee

years\_at\_company: year at company

work accident: any accident on work

left: left or stayed

promotion\_last\_5years: received promotion or not

department: department of employee

salary: salary of an employee

**Exploratory Data Analysis and Statistics: 50 Marks**

1. Identify the missing values, outliers if any and treat them according. – 5 Marks
2. Come up with appropriate results and visualization for the following:
   1. Analysis of percentage left the organization. – 5 Marks
   2. Perform Bivariate analysis and document appropriate results/inference in the markdown cell. – 20 Marks
3. Perform any one t-test and one chi-square test relevant hypothesis testing w.r.t target variable (t, chi-Square) – 20 Marks (split-up below)
   1. Stating hypothesis statements – 5 Marks
   2. Performing hypothesis testing – 10 Marks
   3. Interpreting results – 5 Marks

**Predictive Analysis/ Machine Learning: 50 Marks**

1. Split the data into training and testing in the ratio of (80:20). - 5Marks
2. Build Logistic Regression and Decision tree model on the data. – 15 Marks
3. Compare both the model’s metrics. – 10Marks
4. Justify the best model out of the two models. – 5 Marks
5. Improving the model performance using hyperparameter tuning or any other method – 15 Marks.