**Company Background**

The dataset provided has been sourced from **L&T Corp**, a leading multinational corporation. The dataset captures various factors influencing employee attrition. L&T Corp aims to understand these factors better to strategize employee retention policies and reduce turnover rates.

**Objective**

The goal is to analyze the dataset to gain insights into the key drivers of employee attrition. This includes data preprocessing, cleaning, encoding, standardization, missing value imputation, exploratory data analysis (EDA), and building a logistic regression model to predict employee attrition.

**Data Dictionary**

| **Column Name** | **Description** |
| --- | --- |
| **EmployeeID** | Unique identifier for each employee |
| **Age** | Age of the employee |
| **Gender** | Gender of the employee (Male/Female) |
| **EducationLevel** | Highest education level attained by the employee |
| **JobRole** | Role of the employee within the company |
| **MonthlyIncome** | Monthly income of the employee in USD |
| **YearsAtCompany** | Number of years the employee has been with the company |
| **WorkLifeBalance** | Employee's work-life balance (Bad, Good, Better, Best) |
| **JobSatisfaction** | Employee's job satisfaction on a scale of 1 to 4 |
| **OverTime** | Whether the employee works overtime (Yes/No) |
| **DistanceFromHome** | Distance from home to workplace in kilometers |
| **PerformanceRating** | Employee's performance rating on a scale of 1 to 4 |
| **Attrition** | Whether the employee has left the company (Yes/No) (Target) |

**Steps to be Performed**

1. **Data Preprocessing (15 marks)**
   * Load the dataset.
   * Identify and handle missing values appropriately.
   * Encode categorical variables using suitable encoding techniques.
2. **Data Cleaning (10 marks)**
   * Remove or correct any inconsistencies or errors in the data.
   * Handle any outliers in the dataset.
3. **Data Standardization (10 marks)**
   * Standardize numerical columns to ensure they have a mean of 0 and a standard deviation of 1.
4. **Exploratory Data Analysis (EDA) (20 marks)**
   * Perform univariate and bivariate analysis.
   * Visualize the data using plots (e.g., histograms, box plots, scatter plots).
   * Summarize findings and insights from the data.
5. **Feature Engineering (10 marks)**
   * Create any new features that could be useful for the analysis.
6. **Model Building (Logistic Regression) (20 marks)**
   * Split the data into training and testing sets.
   * Build a logistic regression model to predict employee attrition.
   * Evaluate the model using appropriate metrics.
7. **Interpretation and Reporting (15 marks)**
   * Interpret the results of the logistic regression model.
   * Prepare a report summarizing the findings and providing recommendations for L&T Corp.

**Total: 100 marks**

Dataset:- <https://docs.google.com/spreadsheets/d/1hnpf4cbYjKuFn2nb2snH2AT51MXgmqfH/edit?usp=sharing&ouid=117552746394951055076&rtpof=true&sd=true>