**HR Project**

You are given data of HR records about the employees of a company. There are number of attributes with a binary class label defining whether or not the employee left the company.

Data Columns:

* Satisfaction\_level: A numerical value of employee satisfaction based on multiple metrics.
* Last\_evaluation: Numerical evaluation score of the employee by the company.
* Number\_projects: The number of projects employee worked on
* Average\_monthly\_hours: Average monthly number of hours the employee worked during the stay with the company.
* Time\_spend\_company: Total years spent in the company.
* Work\_accident: Whether or not the employee had an accident during the tenure with the company.
* Promotion\_last\_5years: Binary attribute indicating if the employee got a promotion during last year.
* Department: The department of the employee.
* Salary
* Left: Binary attribute indicating whether or not the employee left the company.

The management has made this data available to their data scientists and expects to have meaningful visualizations and information which can help employees retention and facilitate managerial decisions. **Please note that the management is always interested in claims backed by visualizations and graphs rather than plain textual reports. You are expected to extract knowledge from the data and write professional reports which would be presented to the management.**

This project would have 3 deliverables. Keep on adding the deliverables to the same document. This way, final report would be a comprehensive report.

1. **Exploratory Data Analysis (EDA)**

For this deliverable, focus on initial data exploration and graphs. This could include but not limited to:

1. Finding missing values and replacing them by an appropriate strategy.
2. Calculating and visualization of summary statistics (consider boxplots)
3. Finding correlation between attributes.
4. Finding the attribute that has highest influence on employee retention.
5. Finding out dependence between categorical and numerical attributes.
6. **Cluster and outlier Analysis**
7. Perform cluster analysis (Use your own imagination for how many clusters should be made. It depends on the objective you are trying to achieve. So first think of what insight you can get and then go for actual implementation).
8. Perform outlier analysis and report your findings.
9. **Recommendations Based on Insights/Final Report**

For this deliverable, focus on following points. Try to apply all the concepts learned during the course. You would not be told which techniques to apply but rather what objectives to achieve. You should be able to identify the approaches and techniques. Your conclusions, recommendations, reasoning, and findings should be supported by the relevant graphs/visualizations. This could include but not limited to the following.

1. What factors cause the employees to leave?
2. Which factor has most influence on employee retention?
3. Does work related accidents cause employees to leave? Why and why not?
4. Do employees leave the company even after they got promotion? If yes, what is the cause?
5. Provide your recommendations to the management to retain the employees.
6. Which department has highest rate of retention?

**Above questions act as a guidelines for you. Do not limit yourself to these questions and avoid “Yes/No” answers. Your reports should center around reasons.**