LESSON 2 : EXPLORATORY DATA ANALYSIS

**- Deadline : *Giảng viên / Mentor set deadline và điền vào mục này trước khi gửi cho học viên***

**- Phương thức nộp bài : *Nêu rõ học viên sẽ gửi bài tập trực tiếp cho mentors hay gửi vào 1 folder đã được tạo sẵn. Nếu là folder tạo sẵn, vui lòng đính kèm link nộp bài vào đây !***

# Mục đích

*Bài tập này sẽ giúp các bạn củng cố kiến thức đã học, sau khi hoàn thiện các bài tập, các bạn có có thể đạt được :*

**Kỹ năng :**

* Thành thạo kỹ năng EDA trước khi thực hiện phân tích

**Kiến thức :**

Các định nghĩa, khái niệm sẽ được củng cố thông qua phần bài tập này, bao gồm :

* Imbalance Data
* Exploratory Data Analysis
* Missing Value
* Visualization Data With Matplotlib - Seaborn

# Bài tập

## **Bài tập 1** :

## Business Statement:

Every year, around 5% of its employees have been promoted in the company. So, we want to predict how many employees will be promoted this year to come up with a treatment plan.

**Target/Objective:**

To solve the above problem, you will have to use machine learning models to solve it. Machine learning models will be learned in the following lessons, however, before building a machine learning model. You must perform pre-building model steps such as EDA, Cleaning Data, Feature Engineering.

With the data set of employees who have been promoted in the past, you will perform the pre-building model steps mentioned above.

**Data:** [**Link**](https://drive.google.com/file/d/12AQMi007QX_m5DqOdI-OFMwaxDz0SfTI/view?usp=drive_link)

**Data Dictionary in the page below.**

**Data Dictionary:**

| **Column name** | Description |
| --- | --- |
| **Employee\_id** | Employee identification |
| **Department** | Department of Employee |
| **Region** | region of company branch |
| **Education** | highest education of employee |
| **Gender** | Employee’s gender |
| **Recruitment\_channel** | recruitment channel to which the employee applied |
| **No\_of\_trainings** | Number of training days for employees |
| **age previous\_year\_rating** | Health assessment of the previous year |
| **Length\_of\_service** | number of working days |
| **KPIs\_met >80%** | Whether the KPI is over 80$% or not |
| **Awards\_won?** | Whether the award of company is over 80$% or not |
| **Avg\_training\_score** | scoring of employee when training |
| **Is\_promoted** | label for promoted of employee, 0 for not, 1 for promoted |

**HẾT !**