**Name: Khalid Alnahed**

Design: The idea is to predict the number of activated students and find which feature affects it the most.

Dataset Description:

Name: students\_activity\_data

The data set contains 14 columns and 10k rows

Columns description:

* Grade: The grade of a user.
* Active Subject: How many subjects did the user visit.
* Activated Subject: How many subjects did the user spent more than 30mins on.
* Live\_time\_spent: How much time did a user spent on live session.
* Replay\_time\_spent: How much time did a user spent on replay session.
* Total\_time\_spent: How much time did a user spent on all activities on the platform.
* Live\_sessions: How many live sessions did a user join.
* Replay\_sessions: How many replay sessions did a user join.
* Competition: How many competitions sessions did a user join.
* Breakout: How many breakouts sessions did a user join.
* Avg\_Ranking: The average ranking of a user in a breakout session.
* Is\_Converted: It represents wither the user got converted into a paying student (1) or not (0).
* Is\_Activated: It represents wither the user spent more than 30 mins (1) or not (0).
* Paid\_amount: How much did this user pay

Each row represents one student.

Algorithm: logistic regression.

Tools:

* python, python libraries such as (pandas, numpy, matplotlib and sklearn) for EDA and prediction.
* Tableau for visualization.