**SQL – Student Streaks Analysis**

**Objective**

In this SQL project, the objective is to analyse student engagement by leveraging streak data, defined as the number of consecutive days of interaction on an online education platform. By working with real-world data from the 365 Data Science website, the project aims to identify the most engaged students based on their learning streaks. This analysis will provide key insights into user commitment and behaviour patterns, enabling the identification of top users for further feedback and testimonials since the results may be used to inform strategies for improving the platform’s offerings and enhancing its unique selling proposition.

**Setup**

In this specific project, the data that we will work with is already cleaned as it is part of the actual 365 Data Science database, so, there is no need to clean or pre-organize any data from this set.

The dataset used for the project was provided by the learning platform, it just needs to be loaded as any other schema and make sure it has been saved correctly in our session.

This dataset [part of it], is comprised of the fields/columns:

streak\_id | user\_id | streak\_active | streak\_frozen | sreak\_platform | streak\_created

As mentioned before, we will use this fields to calculate the top students according to their streak’s length.

**Analysis & Data Exploration**

***[Refer to the SQL code document attached in the portfolio]***

Before starting, we can take a quick look at the dataset, more specifically, the table we will be working with, this can be done briefly with the DESCRIBE command.

Next, a couple of variables are pre-defined to work with the newly created values of *streak\_count, prev\_user\_id and prev\_streak\_activeI .*

These values are used to calculate the streak’s lengths along with a CASE-WHEN statement and the whole streaks count query is stored in a temporary table.

Finally, the temporary table is used to obtain the students with a streak length higher or equal to 30.

Additionally, a small query was made to obtain the total number of distinct users from the database, in order to compare to the number of top students.

**Key Findings**

1. The main idea behind finding the key metric of the ‘top’ students, is to be able to get feedback from the users, whether it is directly asking them about their experience with the platform or checking their account’s data to see what other insights we can obtain from this data.
2. Another small but important metric that we can observe out of the calculation of the total number of users is that amongst the 8255 users in the database, 66 are top students, this means only 0,8% accounts for top students, since we are interested in the students having a higher engagement with the learning platform we may need to see what is breaking the majority of the users streaks.

(It is important, to take into account other viable metric apart from the strict concept of a streak, since there may be a lot of users who are healthily engaging but with different schedules to use the platform, a simple example would be the fact that a lot of people may not be studying on weekends, but these users may be active for long periods of time from Monday to Friday)