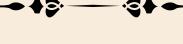# Student Streaks Analysis with SQL

Dataset Source: 365 Data Science
Tools: MySQL.

Prepared by :

Payal Gupta

https://www.linkedin.com/in/payalgupta01/

Contact No: +91- 8054433189

PROJECT OVERVIEW

This project **analyzes user streaks** on an **EdTech learning platform** (**31 days of activity logs**, **~27.5K records, ~8.2K users**).
 Streaks measure consecutive learning days — a key engagement metric for online education. The **goal** was to:

1. **Explore and understand streak data.**
2. **Calculate each student's longest active streak.**
3. **Examine streak distribution in buckets to find engagement drop-off points.**
4. **Identify top performers (≥30 days).**

KEY FINDINGS

-**Dataset Overview**

- **Total records: 27,548**
- **Unique users: 8,255**
- Time window: **31 days** (Dec 1 – Jan 1)
- Active streaks: 27,160 (majority of records)
- Frozen streaks: 459 (rarely used)
- **Streak length range: 1 – 32 days**

-*Longest Streaks Distribution*

streak_bucket	num_users
1-5 days	7083
6-10 days	580
11-15 days	268
16-20 days	133
21-25 days	53
26-30 days	76
31-32 days	62

- **Almost 86%** of users **drop off within 5 days**. Only 1% (87 users) sustain 30+ days.

- *Top Performers*
1. **Max streak** achieved: **32 days**
2. **87 users reached ≥30 days** (ranked as top performers).
3. These users **showcase strong commitment and could be role models/ambassadors**

While only a few students maintain long streaks (≥30 days), **improving early-stage engagement (days 1–10) could significantly boost long-term learning consistency.**

Github Url for complete project detail & files:
https://github.com/payalgupta02/Student-Streaks-Analysis-with-SQL

PROCESS

Dataset downloaded *as zip folder 'project-files-student-streaks-analysis-with-sql' from 365DataScience*

↓

extracted *the folder*

↓

imported *the .sql file* **to MySQL Workbench & ran the script**

↓

Schema *created : One Database & One table*

DB: streaks

Table: user_streaks_sql

↓

Explored the complete data *by using count( ), sum( ), distinct, where clause etc*

↓

Calculated Streak Lengths
- *Used SQL* **variables (@streak_count, @prev_user_id, etc.)**
- *Created* **temporary table streaks_new**
- *Assigned* **streak_length = consecutive days of activity** *(excluding frozen days)*

↓

Created buckets *(1–5, 6–10, … 31–32)* **& Counted users in each bucket**

↓

Identifed the Top Performers *(streaks ≥ 30 days) & Ranked them using RANK( )*

SQL PREPARATION

- ***Calculating Longest Streak:***

```sql
SET @streak_count = 0, @prev_user_id = NULL, @prev_streak_active = NULL;
CREATE TEMPORARY TABLE streaks_new AS
SELECT
    user_id,
    streak_created,
    streak_active,
    @streak_count := CASE
        WHEN @prev_user_id = user_id
            AND @prev_streak_active = 1
            AND streak_active = 1
            AND IFNULL(streak_frozen,0) = 0
        THEN @streak_count + 1
        ELSE 1
    END AS streak_length,
    @prev_user_id := user_id,
    @prev_streak_active := streak_active
FROM user_streaks_sql
ORDER BY user_id, streak_created;
```

- ***Creating Longest_Streak bucket:***

```sql
create table Longest_streak_bucket
as
SELECT
    CASE
        WHEN longest_streak BETWEEN 1 AND 5 THEN '1-5 days'
        WHEN longest_streak BETWEEN 6 AND 10 THEN '6-10 days'
        WHEN longest_streak BETWEEN 11 AND 15 THEN '11-15 days'
        WHEN longest_streak BETWEEN 16 AND 20 THEN '16-20 days'
        WHEN longest_streak BETWEEN 21 AND 25 THEN '21-25 days'
        WHEN longest_streak BETWEEN 26 AND 30 THEN '26-30 days'
        ELSE '31-32 days'
    END AS streak_bucket,
    COUNT(user_id) AS num_users
FROM (
    SELECT user_id, MAX(streak_length) AS longest_streak
    FROM streaks_new
    GROUP BY user_id
) t
GROUP BY streak_bucket
ORDER BY MIN(longest_streak);
```

- **Ranking Top users** *(longest_streak >= 30)*

```sql
Create table streak_new_length_30orMore
As SELECT
    user_id,
    MAX(streak_length) AS longest_streak,
    RANK() OVER (ORDER BY MAX(streak_length) DESC) AS streak_rank
FROM streaks_new
GROUP BY user_id
HAVING longest_streak >= 30
ORDER BY longest_streak DESC;
```

RECOMMENDATIONS

- *Reduce Early Drop-Offs*

Since most users quit in first 5 days, **introduce onboarding nudges** *(emails, notifications, gamified challenges) within this window.*

- *Build Mid-Streak Motivation*

Reward milestones at 7, 14, and 21 days *(badges, certificates, discounts) to encourage continuation.*

- *Celebrate Top Performers*

Showcase students with 30+ day streaks in a "Leaderboard" *or community highlights.*
Introduce monthly streak challenges *to motivate others.*

- *Leverage Frozen Streak Feature*

Only 459 frozen streaks recorded → many users may be unaware. **Educating students about freezing streaks could reduce unnecessary streak breaks.**