

Student Engagement Analysis – SQL Project

Prepared By: Akshaya Mekala

Date:08-09-2025

1.Introduction

The **Online Course Platform Analysis** project focuses on understanding student engagement, course completion, and dropout patterns on an e-learning platform. With the growing importance of online education, analysing how students interact with courses is crucial for improving learning outcomes, designing better courses, and enhancing user experience.

This project uses data from three main tables:

1. **Users** – containing details of registered students, including their demographics and signup information.
2. **Courses** – containing course details, including course names and difficulty levels.
3. **Enrollments** – tracking student progress and completion percentages for each course.

The main objectives of this analysis are:

- To identify patterns in **student engagement** and completion across courses of different difficulty levels.
- To understand **trends in user signups** based on demographics such as country and gender.
- To calculate **key performance indicators (KPIs)** such as completion rates, dropout rates, and top-performing students.
- To provide **actionable insights** that can help the platform enhance course effectiveness, reduce dropouts, and improve student engagement.

The analysis is divided into two main parts:

1. **Exploratory Data Analysis (EDA)** – to understand the structure, distribution, and patterns in the data.

2. **Key Performance Indicators (KPIs) & Business Insights** – to evaluate critical metrics for platform performance and student engagement.

This documentation presents the SQL queries used, results obtained, and the insights derived from them, providing a complete overview of student behaviour and course performance on the platform.

2.Dataset Overview

The dataset for this project is designed to represent a real-world online course learning platform. It consists of three key tables: Users, Courses, and Enrollments. These tables are interconnected and provide a complete picture of student demographics, course offerings, and engagement behaviour.

1. **Users Table** – Contains details of students such as their unique ID, name, email, gender, age, and signup date. This table helps in analysing demographic trends and sign-up patterns.
2. **Courses Table** – Contains information about each course offered on the platform, including course ID, course name, and difficulty level (Beginner, Intermediate, Advanced). This table is useful for analysing course catalog performance and difficulty distribution.
3. **Enrollments Table** – Acts as the fact table linking users and courses. It captures enrollment details such as enrollment ID, user ID, course ID, progress percentage, and enrolment status (Completed, In Progress, Dropped). This table is the foundation for measuring student engagement, course completion, and dropout rates.

Together, these three tables provide a comprehensive dataset that allows us to explore questions related to:

- Student sign-ups and demographics
- Course distribution and popularity
- Student engagement, performance, and dropout behaviour

This dataset forms the backbone of the Student Engagement Analysis Project, enabling us to derive meaningful insights into how students interact with online courses.

3.Data Cleaning

Before beginning the analysis, the dataset was carefully reviewed to ensure accuracy, consistency, and reliability of results. The following steps were undertaken as part of the data cleaning process:

1. Handling Missing Values

- Checked for null or missing entries across all tables.
- No significant missing values were found, ensuring the dataset was complete.

2. Removing Duplicates

- Verified the uniqueness of key identifiers such as user_id, course_id, and enrollment_id.
- No duplicate records were detected.

3. Standardizing Data Formats

- Ensured date fields (e.g., signup_date) were properly formatted to support time-based analysis.
- Confirmed that categorical fields such as gender and course_level were consistently labeled (e.g., “Male/Female” instead of mixed variations).

4. Ensuring Logical Consistency

- Verified that every enrollment record was linked to a valid user and a valid course.
- Checked that progress_percent values were within the valid range of 0–100.

After these checks, the dataset was found to be **clean, consistent, and analysis-ready**. This provided a strong foundation for the subsequent **Exploratory Data Analysis (EDA)** and KPI evaluation.

4.Exploratory Data Analysis (EDA)

After preparing a clean dataset, the next step was to perform Exploratory Data Analysis (EDA). The goal of EDA is to understand the structure of the data, identify patterns, and generate insights that will guide KPI analysis and decision-making.

The following exploratory queries were performed:

1. Total Users – Count of registered users on the platform.

Results		Messages
	total_users	
1	200	

2. Country-wise Signups – Distribution of users across different countries.

Results		Messages
	country	signups
1	USA	41
2	India	38
3	Germany	36
4	Australia	30
5	Canada	30
6	UK	25

3. Gender-wise Count – Analysis of platform usage by gender.

Results		Messages
	gender	user_count
1	Female	63
2	Male	60
3	Other	77

4. Month-wise Signups – Trends in user registration over time.

Results		Messages
	signup_month	signups
1	2022-08	4
2	2022-09	7
3	2022-10	7
4	2022-11	10
5	2022-12	6
6	2023-01	5
7	2023-02	7
8	2023-03	5
9	2023-04	9
10	2023-05	2
11	2023-06	5
12	2023-07	6
13	2023-08	5
14	2023-09	6
15	2023-10	2
16	2023-11	3
17	2023-12	5
18	2024-01	8
19	2024-02	5
20	2024-03	2
21	2024-04	5
22	2024-05	3
23	2024-06	8
24	2024-07	12
25	2024-08	5
26	2024-09	3
27	2024-10	7
28	2024-11	6
29	2024-12	11
30	2025-01	3
31	2025-02	4
32	2025-03	8
33	2025-04	5
34	2025-06	2
35	2025-07	8
36	2025-08	1

5. Number of Users in Each Group – Categorizing students based on their completion status (Completed, In Progress, Dropped).

Results		Messages
	age_group	user_count
1	18-25	46
2	26-35	71
3	36-45	53
4	46+	30

6. Course Catalog Analysis – Number of courses offered across different difficulty levels.

Results		Messages
	category	No_of_courses
1	Data Science	8
2	Web Development	5
3	Cloud Computing	5
4	Cybersecurity	1
5	AI & ML	1

7. Number of Enrollments per Course – Identifying which courses attract the most enrollments.

Results Messages			
	course_id	course_name	No_of_enrollments
1	15	Cross-platform radical alliance	43
2	16	Organized radical service-desk	38
3	8	Public-key homogeneous emulation	34
4	9	Fully-configurable logistical hardware	33
5	18	User-centric 6thgeneration initiative	32
6	14	Front-line demand-driven architecture	30
7	10	Distributed even-keeled core	30
8	3	Public-key hybrid moderator	30
9	11	Up-sized well-modulated budgetary management	29
10	19	Organic discrete moratorium	29
11	20	Extended web-enabled time-frame	27
12	13	Seamless zero administration adapter	27
13	1	Quality-focused client-driven projection	23
14	4	De-engineered upward-trending definition	23
15	5	Adaptive web-enabled adapter	23
16	6	Innovative maximized service-desk	22
17	2	Monitored bottom-line conglomeration	22
18	17	Triple-buffered actuating encryption	20
19	7	Up-sized needs-based interface	20
20	12	Automated stable instruction set	17

Each of these analyses provided a deeper understanding of the dataset and student behavior. The results highlighted user demographics, platform growth patterns, and course distribution, laying the foundation for measuring performance through defined KPIs.

5.KPI and Insights

Based on the business problem, a set of Key Performance Indicators (KPIs) were defined to measure student engagement, course performance, and overall platform effectiveness. The following analyses were carried out:

1. Overall Course Completion Rate

- Measures the percentage of students who successfully completed their enrolled courses.
- *Insight:* The overall completion rate provides a benchmark of platform effectiveness in driving students toward finishing their learning journey.

Results Messages	
	course_completion_rate_percent
1	0.180000000000

2. Number of Completed Courses

- Identifies how many students have completed at least one course.
- Insight:* Helps in understanding the success rate of learners across the platform.

	name	course_name	completion_status
1	Steven Flynn	Monitored bottom-line conglomeration	Completed

3. Completion by Difficulty Level

- Compares completion rates between Beginner, Intermediate, and Advanced courses.
- Insight:* Beginner-level courses recorded higher completion rates compared to advanced ones, indicating that course difficulty impacts engagement.

Results Messages		
	Course_Level	Number_Of_Completed_Courses
1	Intermediate	61
2	Beginner	44
3	Advanced	37

4. Top 5 Most Active Students

- Highlights the top-performing students with the highest progress percentages.

- *Insight:* Shows highly engaged learners who can be used as success stories or case studies for the platform.

Results Messages		
	Top_5_Most_Active_Students	Progress_percentage
1	Steven Flynn	100
2	Sarah Wagner	99.8600006103516
3	Bary Hensley	99.7200012207031
4	Jennifer Zavala	99.5899963378906
5	Whitney Peters	99.5100021362305

5. Top 5 Dropped-Out Courses

- Identifies the courses with the highest dropout counts.
- *Insight:* These courses may require redesign or improvement in content delivery to reduce dropout rates.

100 %		
Results Messages		
	Course_Name	Dropout_Count
1	Cross-platform radical alliance	17
2	Public-key homogeneous emulation	11
3	Up-sized well-modulated budgetary management	10
4	Distributed even-keeled core	9
5	Organized radical service-desk	9

6. Percentage of Each Completion Status

- Breaks down enrollments into Completed, In Progress, and Dropped.
- *Insight:* Provides a clear view of how students are progressing overall, highlighting areas where learners disengage.

Results		Messages	
	completion_status	status_count	percentage
1	In Progress	319	57.790000000000
2	Just Started	147	26.630000000000
3	Almost Done	85	15.400000000000
4	Completed	1	0.180000000000

Each KPI was supported with SQL queries (available in the GitHub repository), and screenshots of outputs are included in this document. These insights collectively help understand student behavior, course effectiveness, and areas for improvement on the platform.

6.Final Insights

The analysis of the online course platform dataset provided several valuable findings about student engagement and course performance:

1. Student Demographics & Growth

- The platform has built a diverse user base with steady growth in signups across months and countries.
- Gender distribution showed balanced participation, highlighting inclusivity in learning.

2. Course Engagement & Completion

- Intermediate-level courses had the highest completion rates, while advanced-level courses faced higher dropout levels.
- Overall course completion was moderate, suggesting opportunities to improve support for learners.

3. Dropouts & Challenges

- A few courses showed consistently high dropout rates, signaling a need to review their design, difficulty, or delivery methods.

4. Active Learners

- A group of highly active students stood out with strong progress rates, reflecting the platform's potential to nurture committed learners.

5. Platform Effectiveness

- While the platform has successfully attracted users, the challenge lies in improving long-term engagement and ensuring more learners reach course completion.

7. Conclusion

This project showcased how SQL can be effectively used to analyze student engagement on an online learning platform. Starting with data cleaning, the dataset was prepared to ensure accuracy and consistency. Through Exploratory Data Analysis (EDA), key patterns were identified in user demographics, signups, and course distribution. Finally, KPI-driven analysis provided measurable insights into course completion, dropout trends, and active learner behavior.

The findings revealed that while the platform has been successful in attracting a diverse learner base, challenges remain in improving engagement and increasing completion rates, particularly for advanced-level courses. At the same time, the presence of highly active learners highlights the platform's potential to create success stories and strengthen long-term retention strategies.

Overall, this project demonstrates a structured approach to data analysis — from raw data to actionable insights — and highlights the importance of using SQL as a powerful tool for business decision-making in the education domain.