

User Engagement and Course Performance Analysis using SQL

(SQL Data Analysis)
- Harsh Kumar Gupta

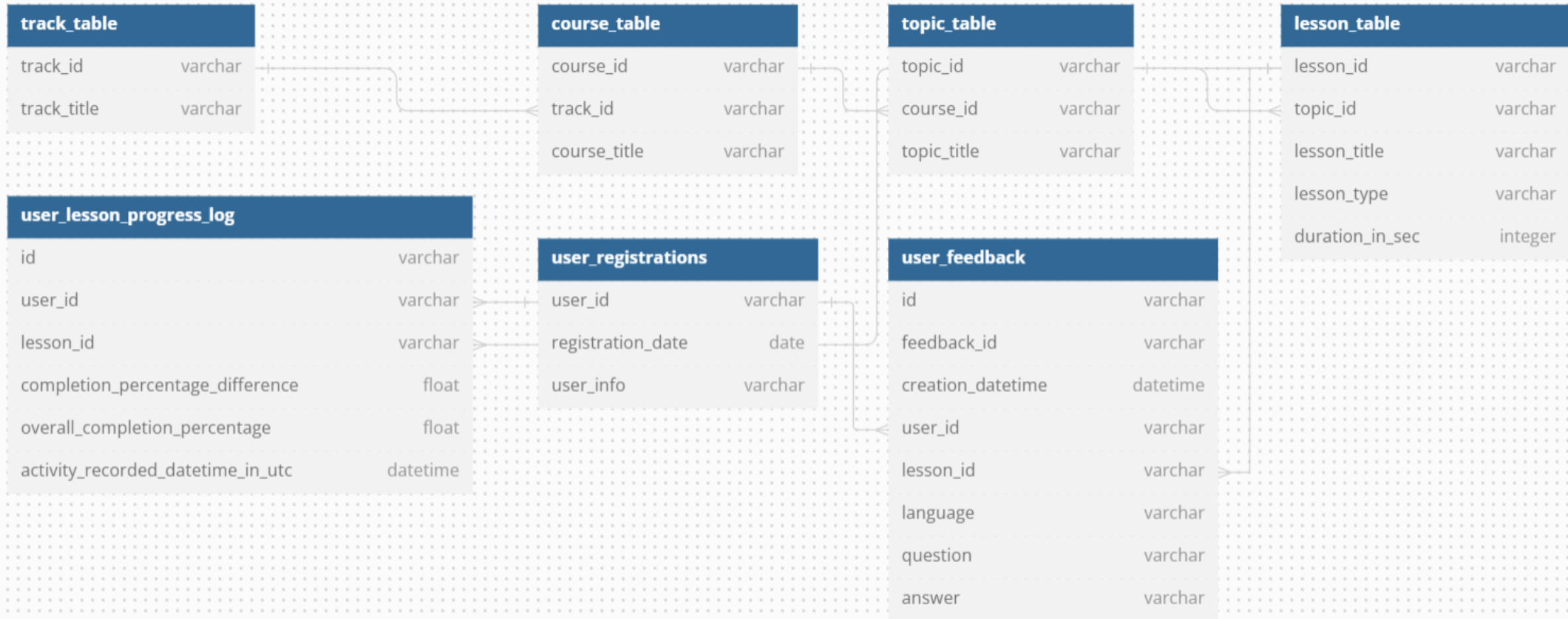
Understanding Dataset

Source : [kaggle](#)

Tables used:

1. Tracks
2. Course
3. Topic
4. Lesson
5. User_lesson_progress
6. Registrations
7. User_feedback

Database Schema




How many users have logged progress in any lesson within the last 30 days?

QUERY

```
SELECT COUNT(DISTINCT user_id) AS active_users
FROM user_lesson_progress
WHERE activity_recorded_datetime_in_utc >=
      ((select max(activity_recorded_datetime_in_utc) from user_lesson_progress) - INTERVAL '30 days');
```

OUTPUT

	active_users bigint 
1	3

How many courses are available for each track?

QUERY

```
SELECT t1.track_id,t1.Track_title,count (distinct t2.course_id) number_of_courses
FROM track t1 left join course t2 on t1.track_id=t2.track_id
GROUP BY t1.track_id,t1.Track_title;
```

OUTPUT

track_id [PK] character varying	track_title character varying	number_of_courses bigint
TRCK101	Tech Foundations	2
TRCK102	Competitive Programming	2
TRCK103	Frontend Developer	2

What is the average duration of lessons in each course?

QUERY

```
SELECT COUNT(DISTINCT user_id) AS active_users
FROM user_lesson_progress
WHERE activity_recorded_datetime_in_utc >=
      ((select max(activity_recorded_datetime_in_utc) from user_lesson_progress) - INTERVAL '30 days');
```

OUTPUT

course_title	avg_duration_in_sec
character varying	numeric
Practical Python	3900.00
Elementary Data Structures and Algorithms	3750.00
React JS - Getting started	3000.00
Developer Foundations	3150.00
Data Structures	3450.00

What is the overall completion percentage for each user on 'Google Colab and Numpy' lesson?

QUERY

```
SELECT user_id, overall_completion_percentage
FROM user_lesson_progress t1
     left join lesson t2 on t1.lesson_id=t2.lesson_id
and t2.lesson_id = 'Google Colab and Numpy'
limit 10;
```

OUTPUT

user_id	overall_completion_percentage
character varying	double precision
CCBP1001USR	5
CCBP1001USR	10
CCBP1001USR	35
CCBP1001USR	50
CCBP1001USR	70
CCBP1016USR	25
CCBP1016USR	35
CCBP1016USR	50
CCBP1016USR	75
CCBP1016USR	100

What is the average completion percentage of lessons across all users?

QUERY

```
SELECT l.lesson_title,  
       ROUND(AVG(ulp.overall_completion_percentage)::numeric,2) AS avg_completion  
FROM lesson l  
JOIN user_lesson_progress ulp ON l.lesson_id = ulp.lesson_id  
GROUP BY l.lesson_title  
ORDER BY 2 Desc;
```

OUTPUT

lesson_title	avg_completion
character varying	numeric
Coding Exam - 1	69.44
Working with Files - Coding Practice - 2	65.54
Jobby App Exam	63.86
Priority Queues & Heaps - Coding Contest - 1	57.78
Binary Search Overview	52.60
Git Branches	52.40
React Popup Reading Material	51.89
Git Commands Hands-on	51.40
Setting Up WSL2	51.32
Best, Worst & Average Case of an Algorithm Slides	51.32

Pandas Coding Assignment - 3	49.92
React Popup Exam	48.20
MCQ Practice	48.10
Operations on Binary Search Trees Assignement - 1	47.88
Broadcasting and Ufuncs	47.88
Coding Practice	47.69
Working with the Ccommand Line Introduction	46.80
Sorting Algorithms Assignmnet	44.83
Library Functions- Sort (Python)	43.24
Command Line Hands-on	43.10
Staging Area & Commits Cheat Sheet	42.56
Google Colab and Numpy	38.60

How many feedback entries have been submitted for each lesson?

QUERY

```
SELECT l.lesson_title, COUNT(uf.id) AS feedback_count
FROM lesson l
JOIN user_feedback uf ON l.lesson_id = uf.lesson_id
GROUP BY l.lesson_title
ORDER BY 2 DESC;
```

OUTPUT

lesson_title	feedback_count
character varying	bigint
Operations on Binary Search Trees Assignment ...	17
Library Functions- Sort (Python)	14
Coding Exam - 1	14
Setting Up WSL2	14
Best, Worst & Average Case of an Algorithm Slides	13
Staging Area & Commits Cheat Sheet	12
Coding Practice	12
Jobby App Exam	12

Working with Files - Coding Practice - 2	11
MCQ Practice	11
Google Colab and Numpy	10
React Popup Reading Material	10
Pandas Coding Assignment - 3	8
Priority Queues & Heaps - Coding Contest - 1	7
Broadcasting and Ufuncs	7

What are the most common aspects mentioned in user feedback for the category "LIKED"?

QUERY

```
SELECT liked, COUNT(*) AS frequency
FROM (
  SELECT unnest(string_to_array(answer, ', ')) AS liked
  FROM user_feedback
  WHERE question = 'LIKED'
) AS liked_feedback
GROUP BY liked
ORDER BY frequency DESC
LIMIT 10;
```

OUTPUT

liked text	frequency bigint
Content	8
Doubts Clarification	5
Questions Depth	5
Problem Solving Approach	3
Explanation	3
Code Explanation	3
Doubts clarification	2

What are the most common aspects mentioned in user feedback for the category "BETTER_TO_IMPROVE"?

QUERY

```
SELECT better_to_improve, COUNT(*) AS frequency
FROM (
  SELECT unnest(string_to_array(answer, ', ')) AS better_to_improve
  FROM user_feedback
  WHERE question = 'BETTER_TO_IMPROVE'
) AS improve_feedback
GROUP BY better_to_improve
ORDER BY frequency DESC
LIMIT 10;
```

OUTPUT


better_to_improve	frequency
text	bigint
Nice	8
No Need	6
Everything is good	6
Good	5
Very Fast	2

How many users have logged progress in any lesson within the last 30 days?

QUERY

```
SELECT COUNT(DISTINCT user_id) AS active_users
FROM user_lesson_progress
WHERE activity_recorded_datetime_in_utc >=
      ((select max(activity_recorded_datetime_in_utc) from user_lesson_progress) - INTERVAL '30 days');
```

OUTPUT

	active_users bigint 
1	3

THANK YOU



harshkumargupta49



HarshKumarGuptagit