SQL QUERY Data Retrieval

EdTech COMPANY

QUESTIONS

1. Find the average completion percentage of units for each user.

2. Calculate the total number of units completed in LEARNING_SET by users residing in HYDERABAD location.

3. Find the programs with the highest completion rates from each city.

4. Count the number of units of each type within each course.

5. Identify users who have completed the most units within the last month.

6. Determine the percentage of users belonging to each educational qualification category.

Source: Kaggle [click here]

Loading the dataset

```
Query    Query History
1  CREATE TABLE day_wise_user_activity (
2      activity_datetime TIMESTAMP,
3      user_id VARCHAR(50),
4      unit_id VARCHAR(50),
5      unit_type VARCHAR(50),
6      day_completion_percentage FLOAT,
7      overall_completion_percentage FLOAT
8  );
9  COPY day_wise_user_activity FROM 'D:\PROJECTS\Courses - SQL\day_wise_user_activity.csv' DELIMITER ',' CSV HEADER;
10
11 CREATE TABLE user_basic_details (
12     user_id VARCHAR(10),
13     gender VARCHAR(10),
14     current_city VARCHAR(50),
15     batch_start_datetime TIMESTAMP,
16     referral_source VARCHAR(50),
17     highest_qualification VARCHAR(30)
18 );
19 COPY user_basic_details FROM 'D:\PROJECTS\Courses - SQL\users_basic_details.csv' DELIMITER ',' CSV HEADER;
20
21 CREATE TABLE learning_resource_details (
22     program_id  VARCHAR(10),
23     program_title  VARCHAR(50),
24     course_id  VARCHAR(10),
25     course_title  VARCHAR(30),
26     topic_id  VARCHAR(10),
27     unit_id  VARCHAR(10),
28     unit_type VARCHAR(30),
29     unit_duration_in_mins INT
30 );
31 COPY learning_resource_details FROM 'D:\PROJECTS\Courses - SQL\learning_resource_details.csv' DELIMITER ',' CSV HEADER;
32
33
```

Table Description

users_basic_details:

- user_id: unique id of the user [string]
- gender: gender of the enrolled user [string]
- current_city: city of residence of the user [string]
- batch_start_datetime: start datetime of the batch, for which the user is enrolled [datetime]
- referral_source: referral channel of the user [string]
- highest_qualification: highest qualification (education details) of the enrolled user [string]

day_wise_user_activity:

- activity_datetime: date and time of learning of the user [datetime]
- user_id: unique id of the user [string]
- unit_id: unique id of the unit [string]
- unit_type: type of the unit. It can be "LEARNING_SET", "QUESTION_SET", "PRACTICE", "EXAM" or "PROJECT" [string]
- day_completion_percentage: percent of the unit completed by the user on a particular day (out of 100%) [float]
- overall_completion_percentage: overall completion percentage of the unit till date by the user (out of 100%) [float]

learning_resource_details:

- program_id: unique id of the program [string]
- program_title: name of the program [string]
- course_id: unique id of the course [string]
- course_title: name of the course [string]
- topic_id: unique id of the topics [string]
- unit_type: type of the unit. It can be "LEARNING_SET", "QUESTION_SET", "PRACTICE", "EXAM" or "PROJECT" [string]
- unit_id: unique id of the unit [string]
- unit_duration_in_mins: duration of the unit in mins [integer]

PROBLEM 1:

Find the average completion percentage of units for each user.

SQL QUERY

```
SELECT
        t1.user_id,

        ROUND(AVG(t2.overall_completion_percentage)::NUMERIC,2) as
Average_completion

FROM
        user_basic_details t1 JOIN

        day_wise_user_activity t2 ON t1.user_id=t2.user_id

GROUP BY 1

ORDER BY 2 DESC;
```

RESULT SET

user_id	average_completion	user_id	average_completion
user_8	100	user_37	94.32
user_14	100	user_38	93.7
user_6	99.19	user_12	93.67
user_9	98.97	user_26	93.28
user_11	98.5	user_42	93.25
user_45	98.47	user_43	93.06
user_46	98.22	user_25	92.49
user_44	97.62	user_24	92.21
user_18	97.15	user_50	91.8
user_39	97.14	user_33	91.56
user_34	97.02	user_10	90.94
user_32	97	user_16	90.77
user_41	96.37	user_13	90.51
user_35	96.29	user_4	89.93
user_15	95.93	user_20	89.45
user_17	95.86	user_27	89.4
user_30	95.6	user_49	89.13
user_7	95.44	user_48	88.91
user_22	95.03	user_47	87.56
user_31	94.85	user_36	87.06
user_29	94.79	user_5	87
user_1	94.72	user_23	82.42
user_19	94.71	user_21	81.58
user_28	94.48	user_40	80.12

PROBLEM 2:

Calculate the total number of units completed in LEARNING_SET by users residing in HYDERABAD location.

SQL QUERY

```
WITH cte AS(
            SELECT
            t1.USER_ID,
             t2.UNIT_ID,
            t2.OVERALL_COMPLETION_PERCENTAGE,t1.CURRENT_CITY
FROM
        user_basic_details t1 JOIN
        day_wise_user_activity t2 ON t1.user_id=t2.user_id
        AND t2.overall_completion_percentage=100
        AND t1.CURRENT_CITY ='Hyderabad'
        AND t2.unit_type ='LEARNING_SET')
SELECT user_id,COUNT(unit_id) units_completed
FROM cte
GROUP BY 1
ORDER BY COUNT(unit_id) DESC
```

RESULT SET

user_id	units_completed
user_5	249
user_35	107
user_31	77
user_45	52
user_26	33
user_15	27
user_38	23
user_48	12
user_20	11
user_1	11

PROBLEM 3:

Find the programs with the highest completion rates from each city.

SQL QUERY

```
WITH cte AS
        (SELECT
                t3.current_city,t1.program_title,t2.overall_completion_percentage
        FROM
                learning_resource_details t1
                JOIN day_wise_user_activity t2 on t1.unit_id=t2.unit_id
                JOIN user_basic_details t3 on t2.user_id=t3.user_id
),
cte2 as(
        SELECT current_city,program_title, AVG(overall_completion_percentage) OVER (PARTITION
by current_city,program_title ) as average_completion
        FROM cte),
cte3 as
        (select
                current_city,program_title,average_completion,ROW_NUMBER() OVER (PARTITION
BY current_city ORDER BY average_completion desc) as rn
        from cte2)
SELECT current_city as city,program_title from cte3
where rn=1;
```

RESULT SET

city	program_title
Bengaluru	Masterclass with Srividya Pranavi
Delhi	Self Assessment
Hyderabad	Masterclass with Srividya Pranavi
Kochi	AI-ML
Mumbai	CCBP 4.0 Academy - Intro
Visakhapatnam	Instruction Flows

PROBLEM 4:

Count the number of units of each type within each course.

SQL QUERY

```
SELECT

        course_title,

        SUM(CASE WHEN unit_type='QUESTION_SET' THEN 1 ELSE 0 END) as Question_set,

        SUM(CASE WHEN unit_type='PRACTISE' THEN 1 ELSE 0 END) as PRACTISE,

        SUM(CASE WHEN unit_type='LEARNING_SET' THEN 1 ELSE 0 END) as LEARNING_TYPE,

        SUM(CASE WHEN unit_type='PROJECT' THEN 1 ELSE 0 END) as PROJECT,

        SUM(CASE WHEN unit_type='EXAM' THEN 1 ELSE 0 END) as EXAM

FROM learning_resource_details

GROUP BY course_title;
```

RESULT SET

course_title	question_set	practise	learning_type	project	exam
My Past & Future	0	0	11	0	3
Test Based Learning	0	0	20	0	18
Instruction Flows	0	0	21	0	8
Introduction	0	0	18	0	0
Node JS	12	0	32	0	3
Logical Reasoning	0	0	6	0	0
My Daily routine	0	0	11	0	4
About Me	0	0	9	0	10
Setting Priorities	0	0	7	0	15
Verbal Ability	0	0	0	0	3
CCBP 4.0 Introduction	0	0	9	0	0
Fundamentals of ML	7	0	115	4	52
Practical Python	3	0	24	0	7
Test Your Knowledge	3	0	0	0	8
My Family	0	0	9	0	4
Srividya Pranavi	0	0	1	0	0
Self - talk	0	0	21	0	0
Developer Foundations	0	0	14	0	0
Data Structures MASTERCLASS SERIES	66	0	106	0	5
	0	0	1	0	0
Input - Output	0	0	8	0	0

PROBLEM 5:

Identify users who have completed the most units within the last month.

SQL QUERY

```
WITH cte AS
        (SELECT
                EXTRACT ( DAYS from ((SELECT MAX(activity_datetime) FROM
                day_wise_user_activity)-t2.activity_datetime) ) as days,
                t1.user_id,t2.unit_id,t2.overall_completion_percentage
        FROM
                user_basic_details t1
                JOIN day_wise_user_activity t2 ON t2.user_id=t1.user_id
                AND t2.overall_completion_percentage=100
                AND t2.activity_datetime >= (SELECT MAX(activity_datetime) FROM
day_wise_user_activity) - INTERVAL '1 month')

SELECT user_id,COUNT(DISTINCT unit_id) AS number_of_units_completed
FROM cte
GROUP BY 1
ORDER BY 2 DESC
limit 5;
```

RESULT SET

user_id	number_of_units_completed	user_id	number_of_units_completed
user_5	399	user_44	52
user_4	135	user_43	50
user_16	113	user_7	50
user_49	103	user_40	48
user_37	98	user_24	46
user_19	95	user_42	45
user_18	92	user_31	44
user_35	89	user_45	42
user_11	85	user_9	38
user_25	81	user_46	38
user_22	80	user_26	33
user_39	69	user_21	28
user_34	67	user_15	26
user_6	64	user_20	23
user_28	63	user_36	22
user_30	63	user_33	17
user_47	61	user_14	16
user_13	61	user_38	16
user_50	57	user_23	16
user_10	56	user_32	16
user_17	55	user_12	14
user_27	54	user_1	14
user_29	53	user_48	10
user_41	52	user_8	1

PROBLEM 6:

Determine the percentage of users belonging to each educational qualification category.

SQL QUERY

```
with cte as
        (select highest_qualification,count(user_id) as no_of_users
        from user_basic_details
        group by 1)
select
        highest_qualification,
        concat(round((no_of_users/(select sum(no_of_users) FROM cte)*100),1),' %') as percentage
from cte
order by no_of_users desc;
```

RESULT SET

highest_qualification	percentage
B.Tech	34.00%
Intermediate	28.30%
Degree	24.50%
10th Completed	11.30%
M.Tech	1.90%