

Part 2: SQL

To identify patterns in patient behavior with respect to exercises and find the driving factors for their exercise completion rates, answered the following questions using SQL.

1.How many users completed an exercise in their first month per monthly cohort?

Assume you have two tables in our company's database:

- 'users' table, with columns 'user_id', 'created_at'
- 'exercises' table, with columns 'exercise_id', 'user_id', 'exercise_completion_date'

Write a single SQL query that breaks up the users based on the month that they signed up (their cohort month), and determines the percentage of users that have a completed exercise in their first month for each monthly cohort (e.g., the 2018 January cohort has x% of users completing an exercise in their first month, 2018 February cohort has x% of users completing an exercise in their first month, etc.).

In []:

Solution Q1:

```
SELECT YEAR(created_at) AS user_created_year, MONTH(created_at) AS user_created_month, (COUNT(users.user_id)*100/(SELECT COUNT(user_id) FROM users)) AS complition_first_month FROM users
INNER JOIN exercises ON users.user_id=exercises.user_id AND YEAR(users.created_at)=YEAR(exercises.exercise_completion_date) AND MONTH(users.created_at)= MONTH(exercises.exercise_completion_date)
GROUP BY YEAR(users.created_at), MONTH(users.created_at)
ORDER BY YEAR(users.created_at), MONTH(users.created_at);
```

2. How many users completed a given amount of exercises?

Assume you have two tables in our company's database:

- 'users' table, with columns 'user_id', 'created_at'
- 'exercises' table, with columns 'exercise_id', 'user_id', 'exercise_completion_date'

Write a single SQL query that returns a frequency distribution of the number of activities each user completed. (Ex: 1000 users completed 1 activity, 500 completed 10 activities, 100 completed 100 activities, etc...)

In []:

Solution Q2:

```
SELECT user_id, COUNT(exercise_id) AS frequency_exercises FROM exercises
GROUP BY user_id
ORDER BY COUNT(exercise_id) DESC;
```

3. Which organizations have the most severe patient population?

Assume you have two tables in our company's database:

- 'Providers' table that contains 'provider_id', 'organization_id', and 'organization_name'
- 'Phq9' table that contains 'patient_id','provider_id', 'score','datetime_created'

For context, A phq9 score ranges from 0-27 and anything 20 or above is considered severe. Write a single query that finds the top five organizations that have the highest average phq9 score per patient.

In []:

Solution Q3:

```
SELECT organization_name FROM Providers
INNER JOIN Phq9 ON Providers.provider_id = Phq9.provider_id
GROUP BY organization_id
ORDER BY AVG(score) DESC
LIMIT 5;
```