

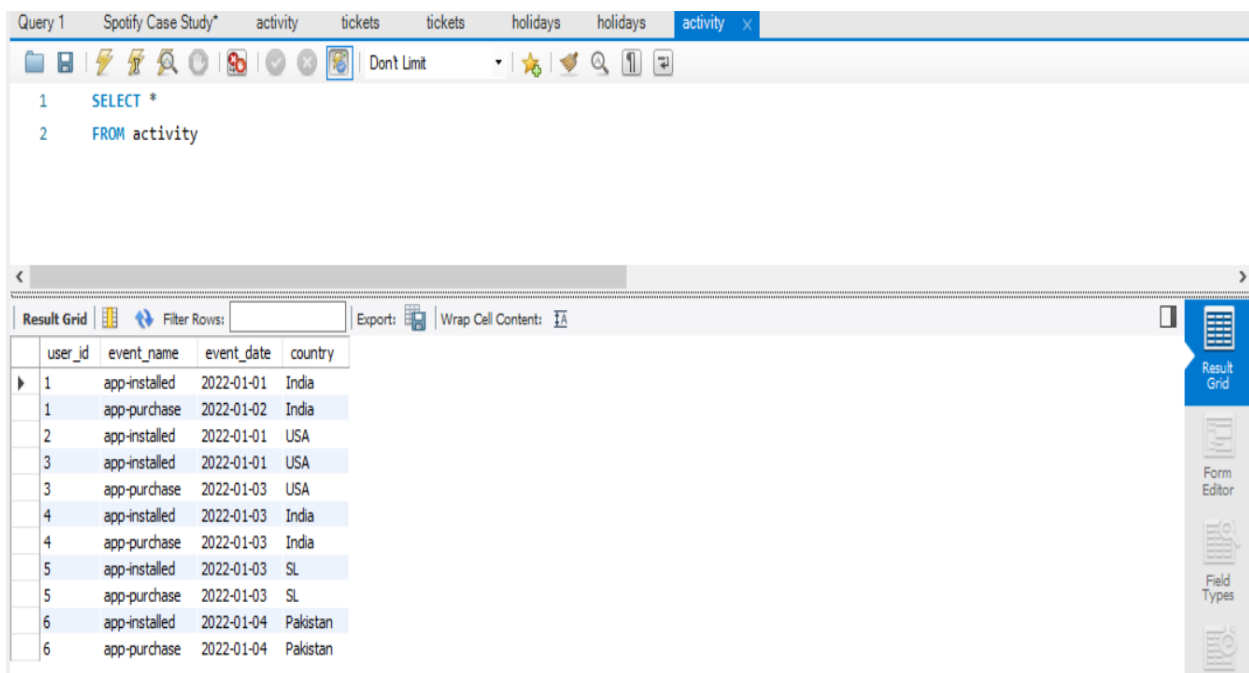
# Spotify Case Study

In this case study, I tried to do an ad\_hoc analysis on spotify user data to generate valuable insights. It comprises of 5 different queries and their code snippets as well as query results.

While doing so, I am going to assume the questions as the business requirement and as an Analyst I have to solve the queries for further decision making.

## Studying the schema

Its necessary to understand what the data is like. So we will spend good amount of time on understanding the schema. Though in our case we are going to work on only one table 'activity' given by stakeholder. So it's not going to take much of our time. Let's understand it.



The screenshot shows a data analysis tool interface. At the top, there are tabs for 'Query 1', 'Spotify Case Study\*', 'activity', 'tickets', 'tickets', 'holidays', 'holidays', and 'activity x'. Below the tabs is a toolbar with various icons. The main area displays a SQL query:

```
1 SELECT *
2 FROM activity
```

Below the query, there is a 'Result Grid' section. It includes a 'Filter Rows' input field, an 'Export' button, and a 'Wrap Cell Contents' checkbox. The 'Result Grid' contains a table with the following data:

	user_id	event_name	event_date	country
1		app-installed	2022-01-01	India
1		app-purchase	2022-01-02	India
2		app-installed	2022-01-01	USA
3		app-installed	2022-01-01	USA
3		app-purchase	2022-01-03	USA
4		app-installed	2022-01-03	India
4		app-purchase	2022-01-03	India
5		app-installed	2022-01-03	SL
5		app-purchase	2022-01-03	SL
6		app-installed	2022-01-04	Pakistan
6		app-purchase	2022-01-04	Pakistan

On the right side of the interface, there are buttons for 'Result Grid', 'Form Editor', and 'Field Types'.

1. What is total count of active users by each day?

Query 1 Spotify Case Study\* x activity tickets tickets holidays holidays activity

1 -- 1. Find the total active users each day.  
2  
3 • SELECT event\_date,  
4 COUNT(DISTINCT(user\_id)) as user\_count  
5 FROM activity  
6 GROUP BY event\_date;  
7

Result Grid Filter Rows: Export: Wrap Cell Content: IA

	event_date	user_count
▶	2022-01-01	3
	2022-01-02	1
	2022-01-03	3
	2022-01-04	1

Result Grid  
Form Editor

2. Find the total active users by each week of a year.

Query 1 Spotify Case Study\* x activity tickets tickets holidays holidays activity

Don't Limit

```
9 -- 2. find the total active users each week.
10 WITH cte as(
11     SELECT *,
12         week(event_date) as week
13     FROM activity)
14
15 SELECT week, COUNT(DISTINCT(user_id)) AS user_id
16 FROM cte
17 GROUP BY week;
```

Result Grid Filter Rows: Export: Wrap Cell Content:

	week	user_id
▶	0	3
	1	5

Result Grid Form

3. Find the total number of users who 'installed' and did 'app purchase' on the same day.

Query 1 Spotify Case Study activity tickets tickets holidays holidays activity

Don't Limit

```
-- 3. Date wise number of users who purchased the app on same day they installed the app.
20
21
22 • SELECT a.event_date,
23       COUNT(*) AS user_count
24 FROM activity as a
25 JOIN activity as b
26 ON a.user_id = b.user_id
27 WHERE a.event_date = b.event_date
28       AND a.event_name = 'app-purchase' AND b.event_name = 'app-installed'
29 GROUP BY a.event_date;
```

Result Grid

	event_date	user_count
▶	2022-01-03	2
	2022-01-04	1

Result Grid

4. Find the percentage of users who did app purchase of total user purchase. Make bin as 'Others' for countries except 'India' and 'USA'.

```
Query 1 Spotify Case Study activity tickets tickets holidays holidays activity
-- 4. find the percentage of users who purchased the subscription and group them in new category 'Others' except country 'India' and 'USA'
32
33
34 WITH cte AS (
35     SELECT country,
36            count(*) purchase_count
37     FROM activity
38     WHERE event_name = 'app-purchase'
39     GROUP BY country),
40
41 cte2 AS ( SELECT sum(purchase_count) AS total
42           FROM cte),
43
44 cte3 AS (
45     SELECT CASE WHEN country IN ('India', 'USA') THEN country ELSE 'Others' END AS new_col,
46            SUM(purchase_count) AS premium_users, total
47     FROM cte, cte2
48     GROUP BY (CASE WHEN country IN ('India', 'USA') THEN country ELSE 'Others' END))
49
50 SELECT new_col ,
51        ROUND(premium_users / total *100) AS percentage
52 FROM cte3;
```

## Result

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
new_col	percentage			
India	40			
USA	20			
Others	40			

5. Find the number of users who bought the subscription on very next day of app installation.

Query 1 Spotify Case Study activity tickets tickets holidays holidays activity

55 -- 5. Among all the users who installed the app, how many of them purchased the subscription on very next day?

56

57 WITH cte as(  
58     SELECT \*,  
59         LAG(event\_name) OVER(PARTITION BY user\_id ORDER BY event\_date) AS pre\_event,  
60  
61         LAG(event\_date) OVER(PARTITION BY user\_id) AS previous\_date  
62     FROM activity)  
63  
64     SELECT event\_date,  
65         COUNT(DISTINCT(user\_id)) AS count FROM cte  
66     WHERE event\_name = 'app-purchase' AND  
67         pre\_event = 'app-installed' AND  
68         datediff(event\_date, previous\_date) = 1

Result Grid Filter Rows: Export: Wrap Cell Content:

event_date	count
2022-01-02	1

Result Grid