OPERATION ANALYTICS AND INVESTGATING METRIC SPIKE

PROJECT DESCRIPTION:

Operation Analytics is the process of analysing a company's overall operations from beginning to end to identify areas for improvement. As a Data Analyst, help them understand the data they collect and use it to make better decisions. The analysis helps the company to predict its growth or decline and optimize its workflows. Investigating metric spikes is also a critical part of this analysis to understand why certain metrics may have gone up or down.

APPROACH;

The process of gathering information i.e by the reading the description, understanding the dataset and tasks to do. By using an IDE(DBeaver) further continued to write queries for the tasks. Created the table with SQL Query, performed the tasks, faced many errors when working, took a lot of time clearing those errors.

TECH USED:

DBeaver – An IDE of SQL. The User Interface is easy to understand. No complications. Easy to handle for beginners.

INSIGHTS:

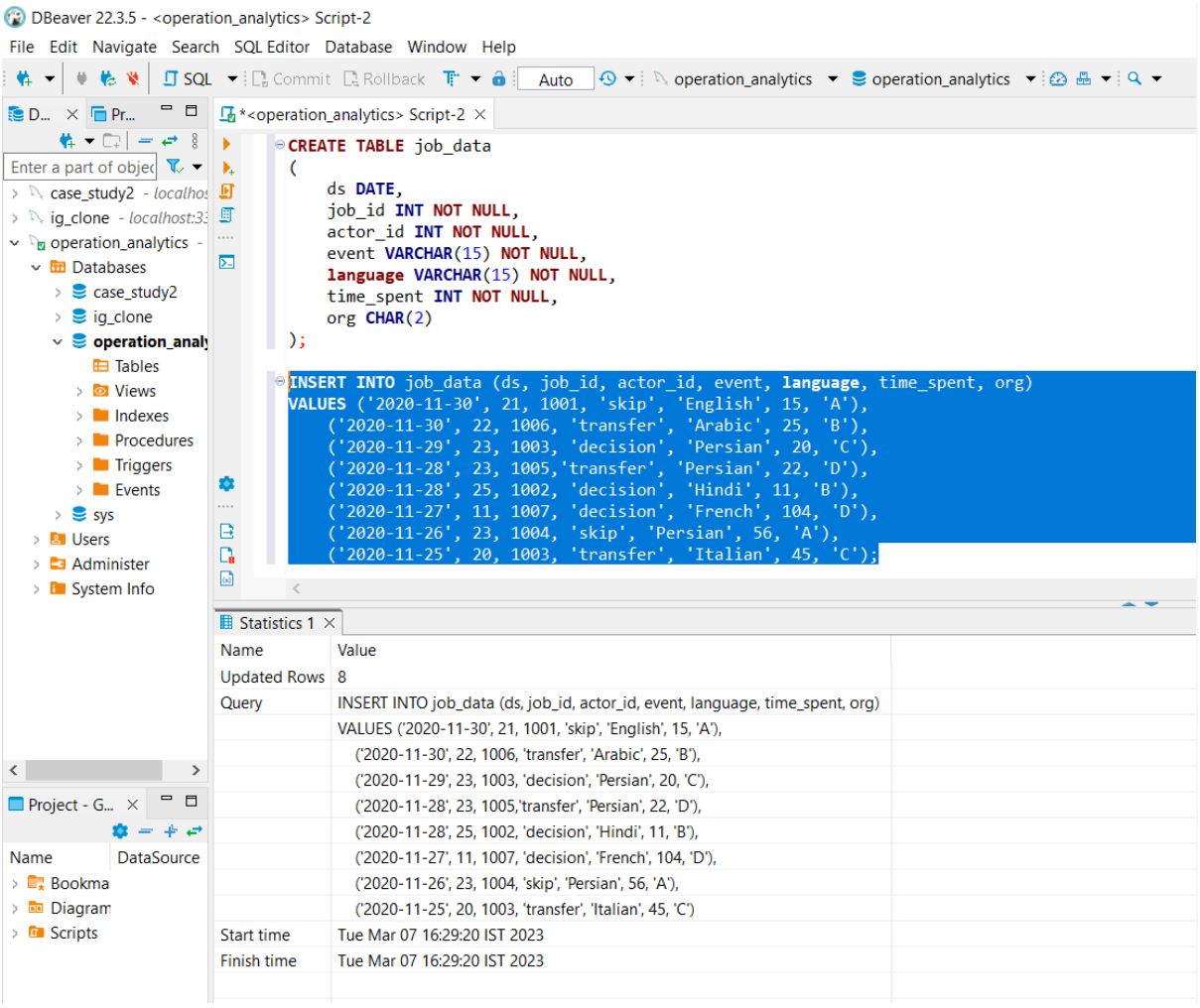
Considering the first task of this task which was "Calculate the number of jobs reviewed per hour per day for November 2020?". First I calculated the total hours of time spent for reviewing jobs for November. Then further lead to the calculation of per day. Then per hour, it took a lot of my thinking power to complete this project.

RESULT:

How to proceed for a task, how to clear errors, how to check whether the output is correct. These were the things learned.

QUERIES:

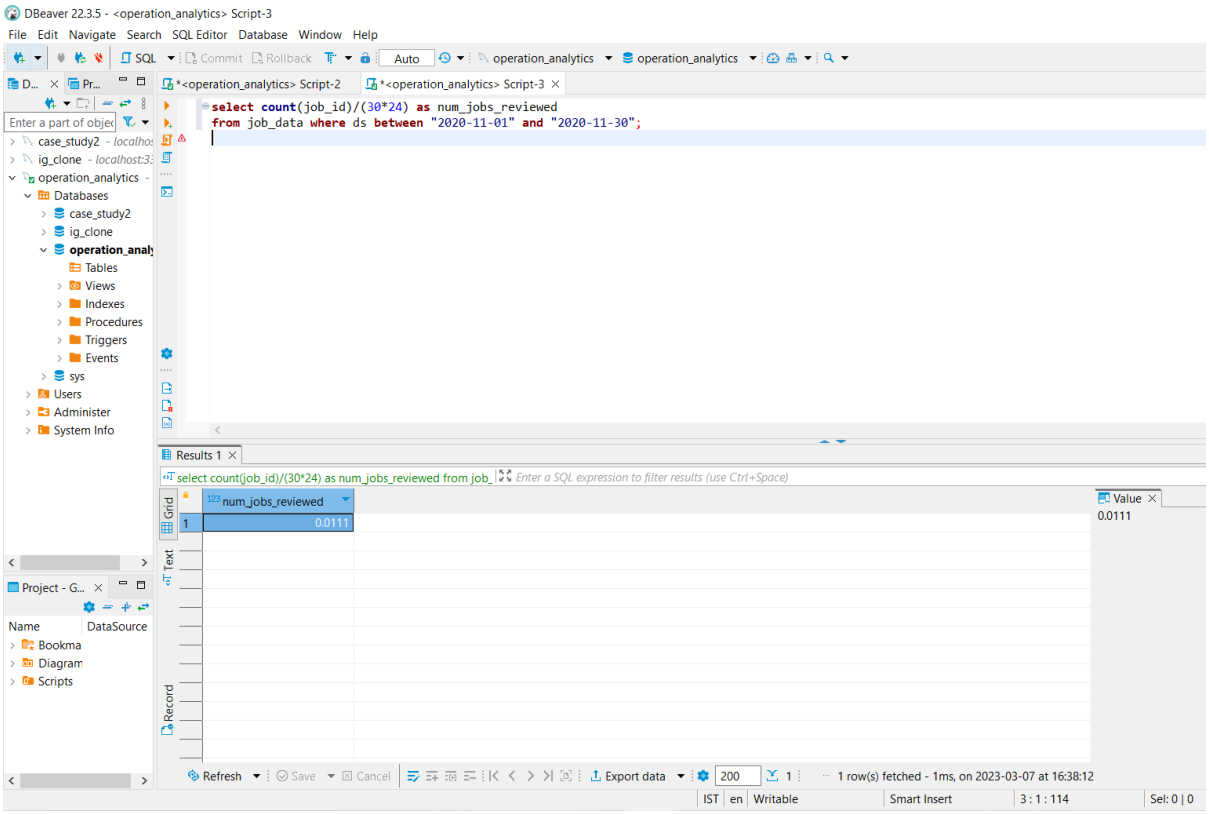
CREATING TABLES:



CASE STUDY-1 (JOB DATA)

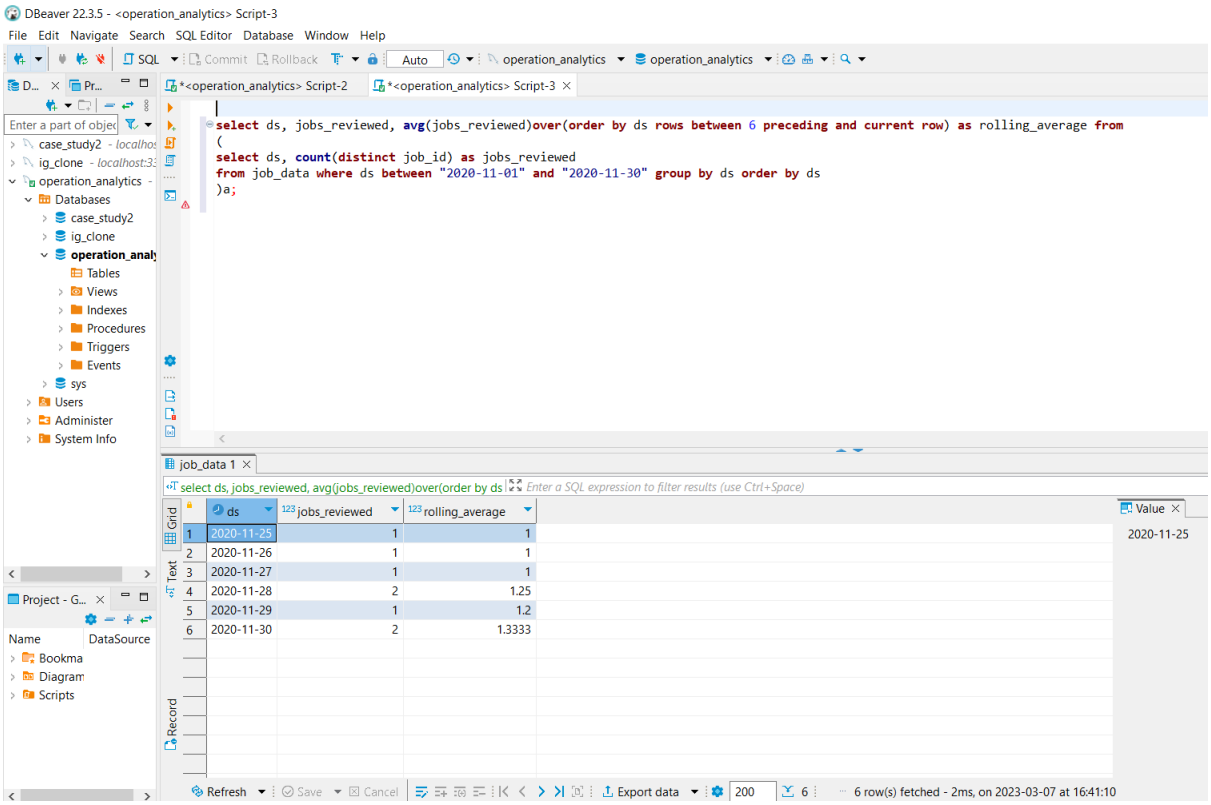
1.Calculate the number of jobs reviewed per hour per day for November 2020?

```
select count(job_id)/(30*24) as num_jobs_reviewed

from job_data where ds between "2020-11-01" and "2020-11-30";
```



2.Let's say the above metric is called throughput. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

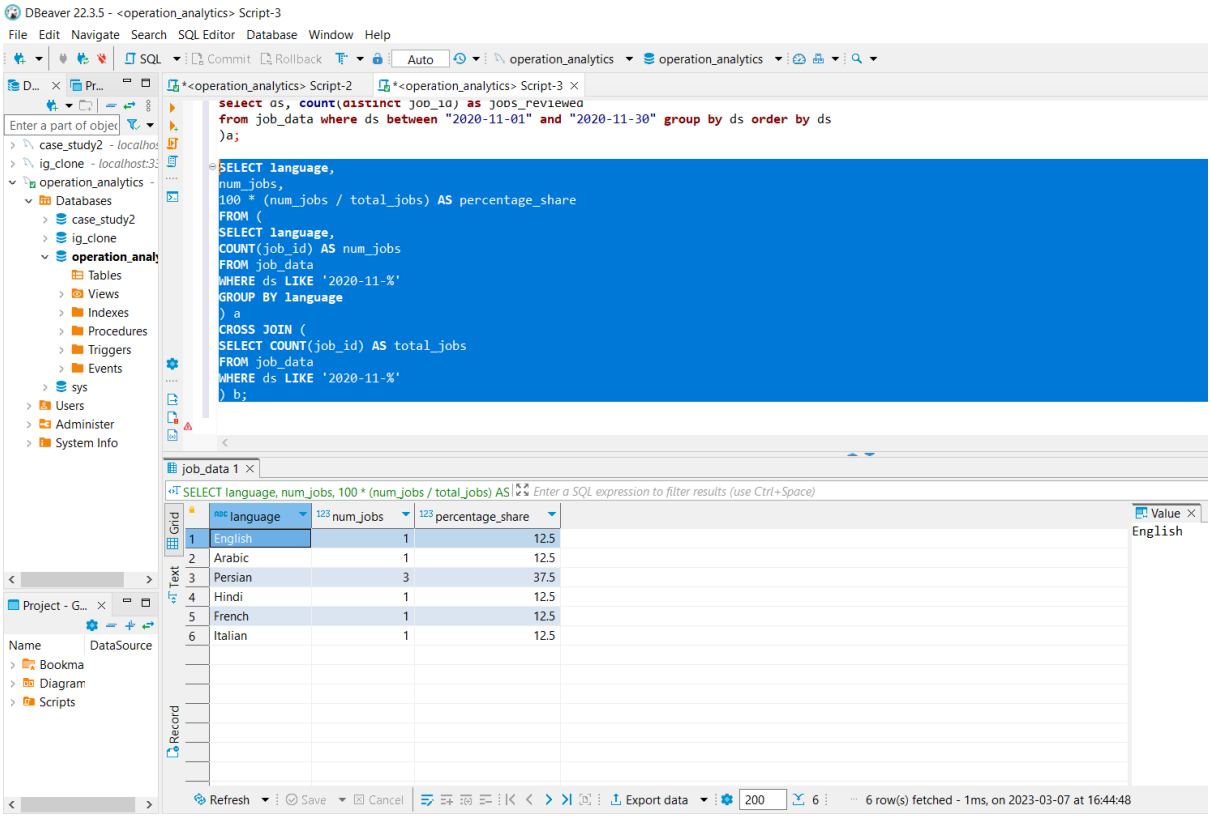
```
select ds, jobs_reviewed, avg(jobs_reviewed)over(order by ds rows between 6
preceding and current row) as rolling_average from

(

select ds, count(distinct job_id) as jobs_reviewed

from job_data where ds between "2020-11-01" and "2020-11-30" group by ds
order by ds

)a;
```



3.Calculate the percentage share of each language in the last 30 days?

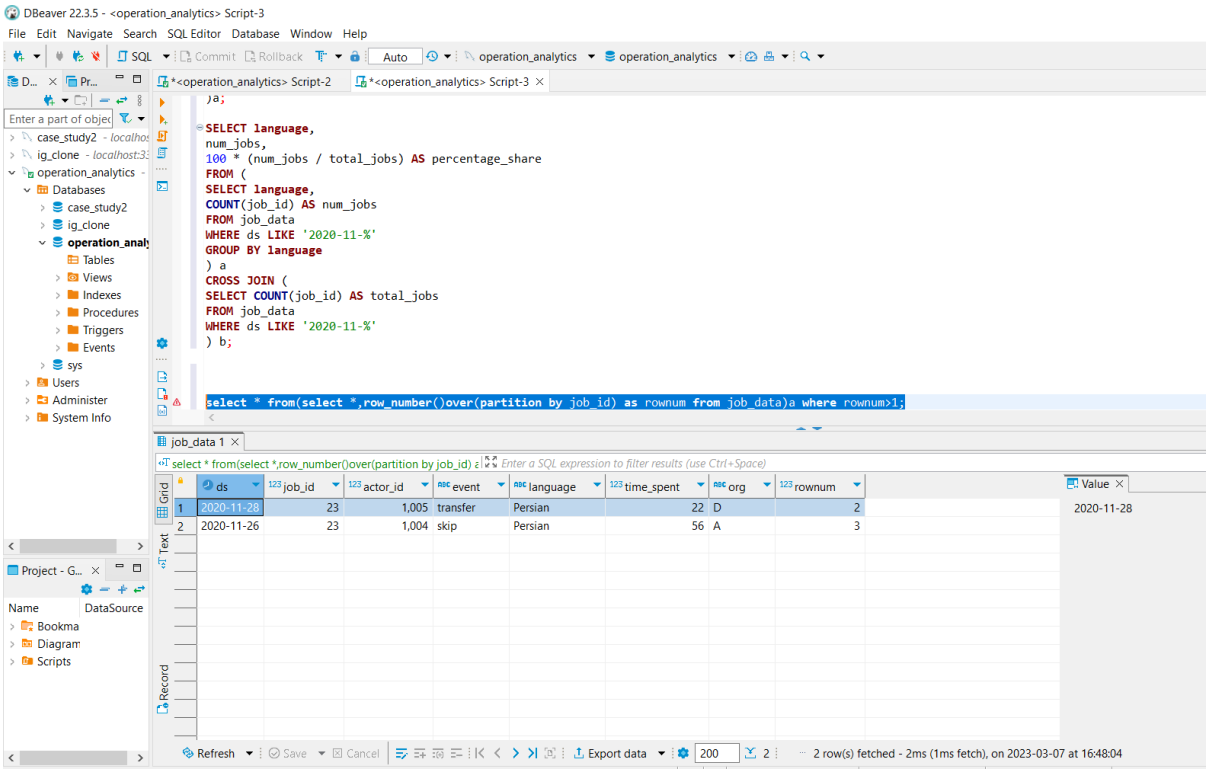
```
SELECT language,

num_jobs,

100 * (num_jobs / total_jobs) AS percentage_share

FROM (

SELECT language,

COUNT(job_id) AS num_jobs

FROM job_data

WHERE ds LIKE '2020-11-%'

GROUP BY language

) a

CROSS JOIN (

SELECT COUNT(job_id) AS total_jobs

FROM job_data

WHERE ds LIKE '2020-11-%'
```

```
) b;
```



4.Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

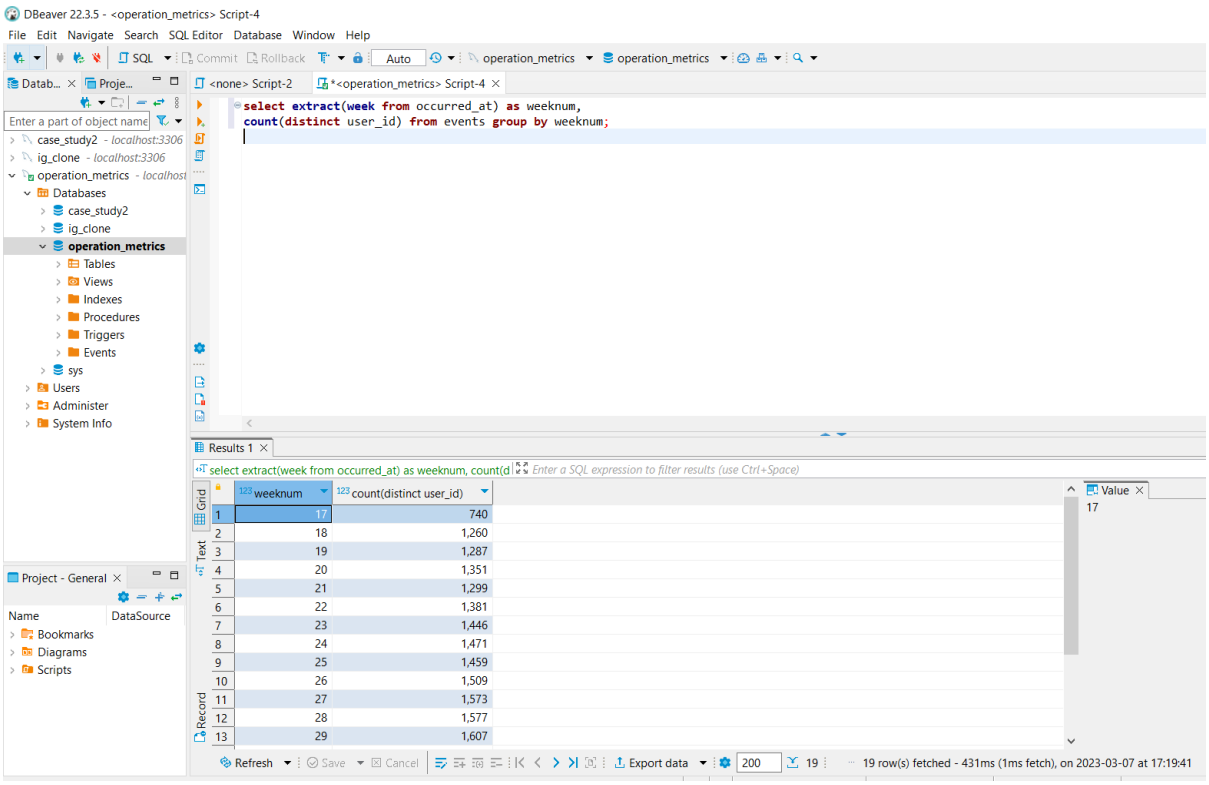
```
select * from(select *,row_number()over(partition by job_id) as rownum from job_data)a where rownum>1;
```


DBeaver 22.3.5 - <operation_analytics> Script-3
File Edit Navigate Search SQL Editor Database Window Help
*<operation_analytics> Script-2
*<operation_analytics> Script-3
)a;
SELECT language,
num_jobs,
100 * (num_jobs / total_jobs) AS percentage_share
FROM (
SELECT language,
COUNT(job_id) AS num_jobs
FROM job_data
WHERE ds LIKE '2020-11-%'
GROUP BY language
) a
CROSS JOIN (
SELECT COUNT(job_id) AS total_jobs
FROM job_data
WHERE ds LIKE '2020-11-%'
) b;
select * from(select *,row_number()over(partition by job_id) as rownum from job_data)a where rownum>1;
job_data 1
Enter a SQL expression to filter results (use Ctrl+Space)
ds job_id actor_id event language time_spent org rownum
1 2020-11-28 23 1,005 transfer Persian 22 D 2
2 2020-11-26 23 1,004 skip Persian 56 A 3
Value
2020-11-28
Refresh Save Cancel Export data
2 row(s) fetched - 2ms (1ms fetch), on 2023-03-07 at 16:48:04

CASE STUDY 2 (INVESTIGATING METRIC SPIKE)

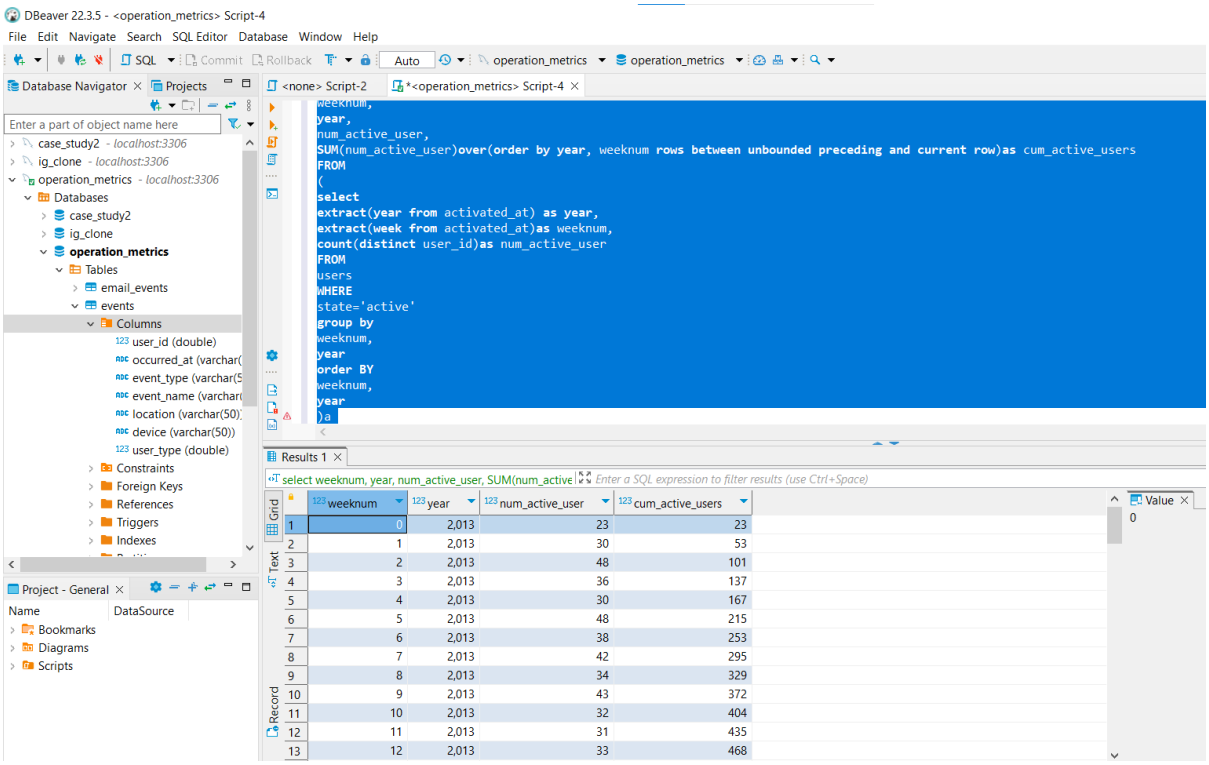
1.Calculate the weekly user engagement?

```
select extract(week from occurred_at) as weeknum,

count(distinct user_id) from events group by weeknum;
```



2.Calculate the user growth for product

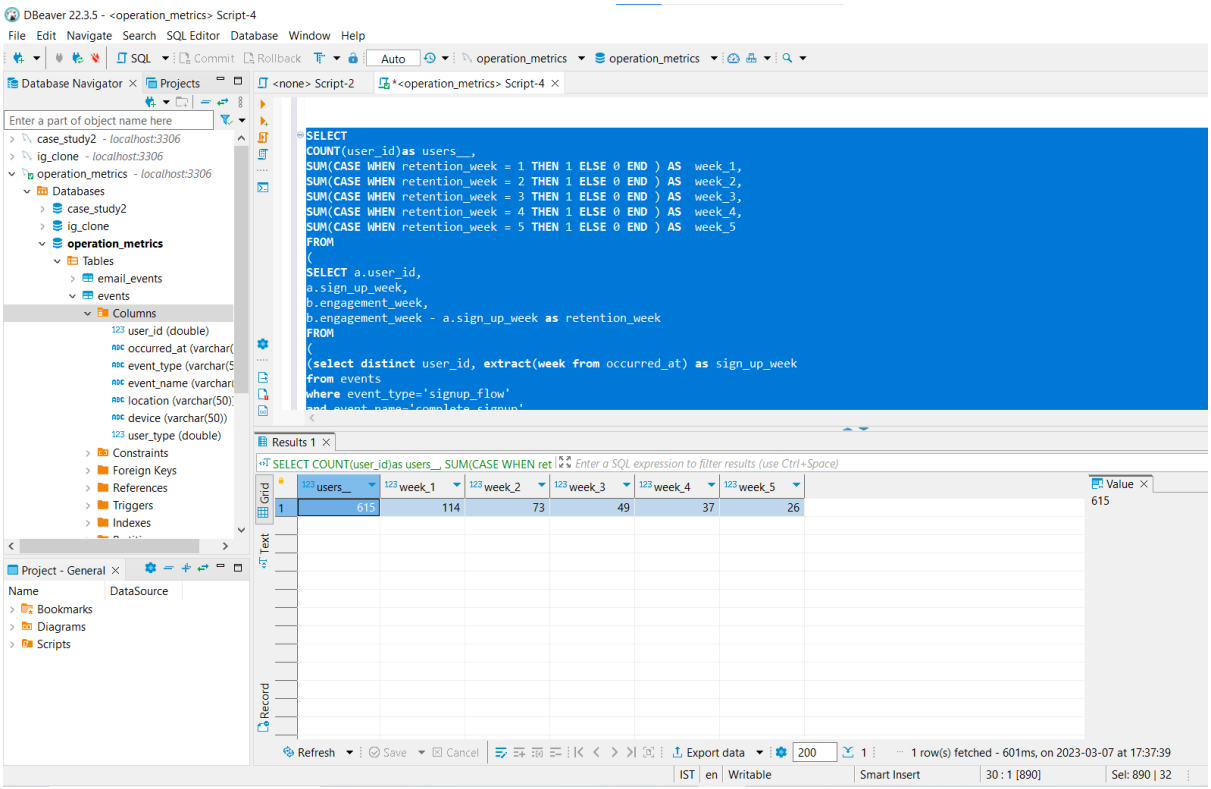
```
select year,weeknum,num_active_user,sum(num_active_user) over(order by
year,weeknum rows between unbounded preceding and current row) as
cum_active_users

from(select extract(year from activated_at) as year,extract(week from
activated_at)as weeknum,count(distinct user_id) as num_active_user

from users a where state="active" group by year,weeknum order by
year,weeknum)a;
```



3 Calculate the weekly retention of users-sign up cohort?

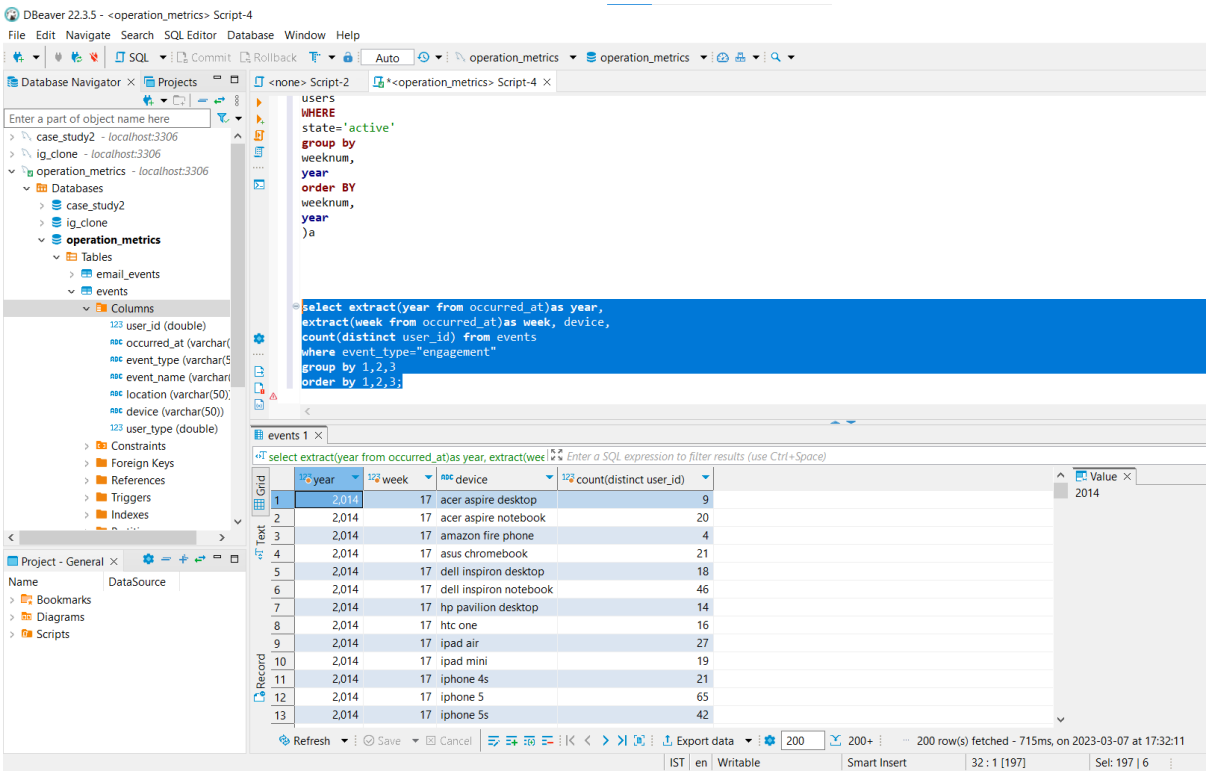
```
SELECT

COUNT(user_id)as users__,

SUM(CASE WHEN retention_week = 1 THEN 1 ELSE 0 END ) AS week_1,

SUM(CASE WHEN retention_week = 2 THEN 1 ELSE 0 END ) AS week_2,

SUM(CASE WHEN retention_week = 3 THEN 1 ELSE 0 END ) AS week_3,

SUM(CASE WHEN retention_week = 4 THEN 1 ELSE 0 END ) AS week_4,

SUM(CASE WHEN retention_week = 5 THEN 1 ELSE 0 END ) AS week_5

FROM

(

SELECT a.user_id,

a.sign_up_week,

b.engagement_week,

b.engagement_week - a.sign_up_week as retention_week

FROM

(
```

```sql
(select distinct user_id, extract(week from occurred_at) as sign_up_week
from events
where event_type='signup_flow'
and event_name='complete_signup'
and extract(week from occurred_at)=18
)a
left JOIN
(
select distinct user_id,
extract(week from occurred_at) as engagement_week
from events
where event_type='engagement'
)b
on a.user_id=b.user_id
)
order by
a.user_id )a
```



4. Calculate the weekly engagement per device?

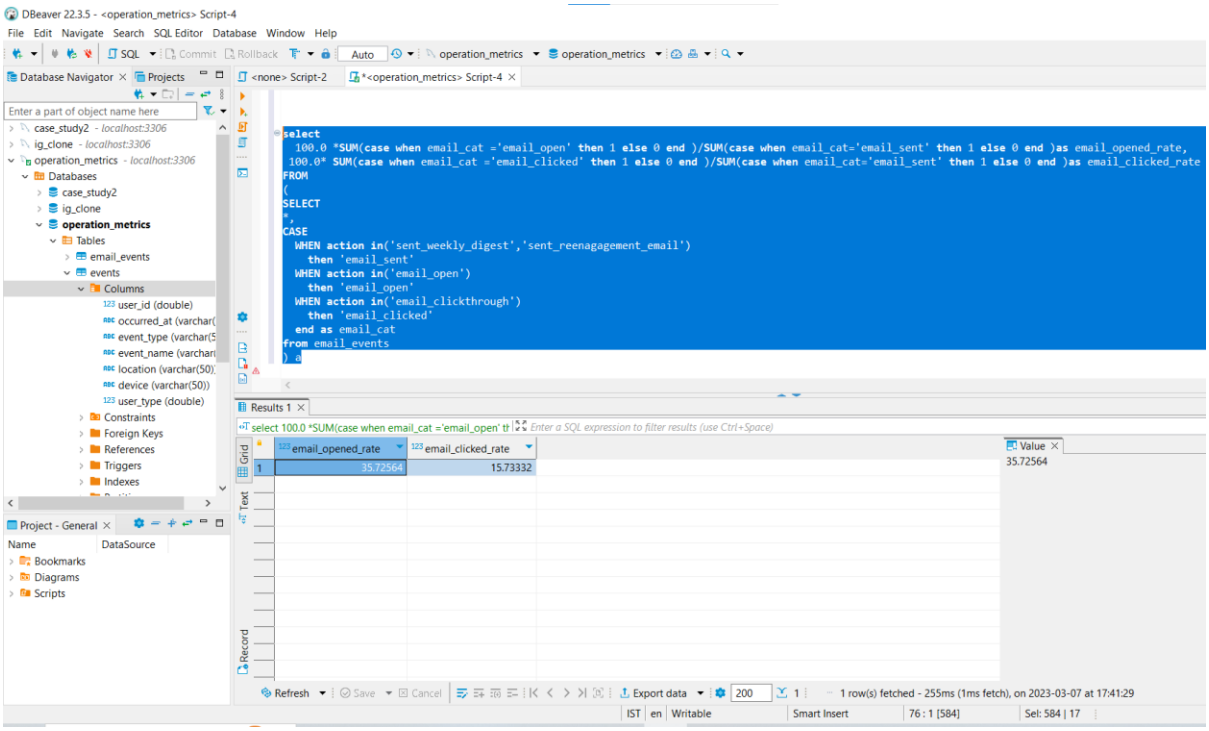
```
select extract(year from occurred_at)as year,

extract(week from occurred_at)as week, device,

count(distinct user_id) from events

where event_type="engagement"

group by 1,2,3

order by 1,2,3;
```



5.Calculate the email engagement metrics?

```
select

100.0 *SUM(case when email_cat ='email_open' then 1 else 0 end )/SUM(case
when email_cat='email_sent' then 1 else 0 end )as email_opened_rate,

100.0* SUM(case when email_cat ='email_clicked' then 1 else 0 end
)/SUM(case when email_cat='email_sent' then 1 else 0 end )as
email_clicked_rate

FROM

(

SELECT

*,

CASE

WHEN action in('sent_weekly_digest','sent_reenagagement_email')

then 'email_sent'
```

```
WHEN action in('email_open')

then 'email_open'

WHEN action in('email_clickthrough')

then 'email_clicked'

end as email_cat

from email_events

) a
```



ALL THE QUERIES ARE EXECUTED SUCCESSFULLY AND GIVES A VALID OUTPUT!!!