TRAINITY

**PROJECT – 3**

**Operation and Metric Analytics**

Submitted By

**BOJJA JAHNAVI**

Data Analytics Trainee

Submitted On:- 26-10-2023

**Operation and Metric Analytics**

**Project Description:**

Operation analytics is the process in which a company’s insights is identified by analysing them. An organization contains different departments to run the business, so as a data analyst my role is to find the various operations in the organization to know the changes in the business units.

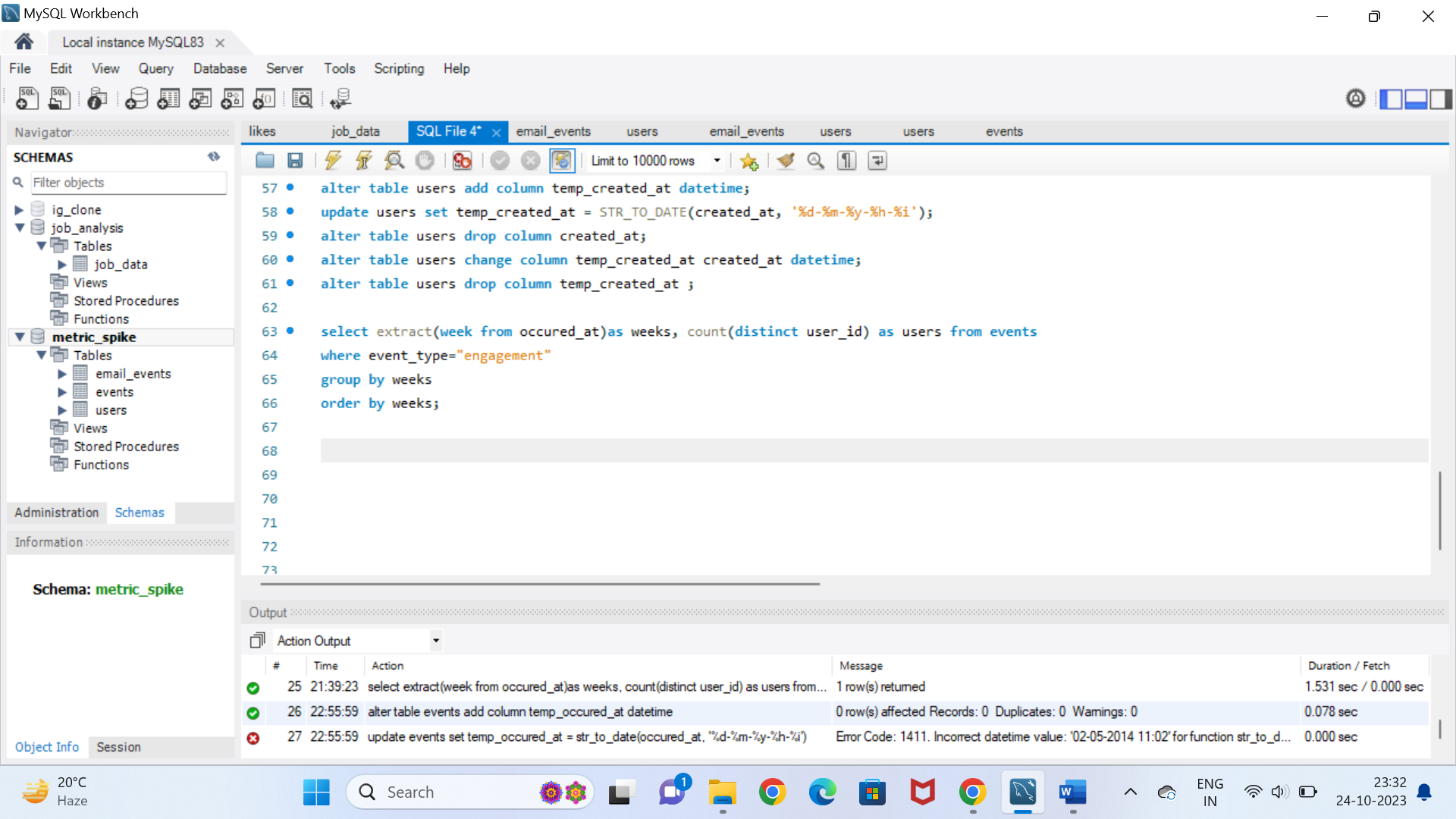
In this project we have 2 case studies to solve the business operations and metrics. Advanced Sql functions are used to investigate the case studies.

Case Study 1: Job Data Analysis

Case Study 2: Investing Metric Spike

**Approach:**

This Project is intiated by creating the databases for the case studies. As we have two case studies let us create 2 databases named job\_anlysis and metric\_spike. For each database create tables in their respective schema. Create a single table job\_data for the job\_analysis and create three tables events, email\_events, users for the metric\_spike.



The following picture shows the databases and the tables existed in them.

**Case Study 1: Job Data Analysis:**

1. **Jobs Reviewed Over Time: Number of jobs reviewed per hour per day in November 2020 are allowed to find.**

**Jobs reviewed per hour:**

**Query:**

select avg(t) as jobs\_reviewed\_per\_hour, avg(p) as jobs\_reviewed\_per\_second

from

(select ds,

((count(job\_id)\*3600)/sum(time\_spent)) as t,

((count(job\_id))/sum(time\_spent)) as p

from

job\_data

where month(ds)=11

group by ds)a;

**Output:**

Average jobs reviewed per hour per day are 126.18048333

1. Throughput Analysis: The average number of events per 7 days are identified. Then average of 7 days and per day is identified and given the best results.

Average throughput analysis:

**Query:**

select round(count(event)/sum(time\_spent),2) as weekly\_throughput

from

job\_data;

select ds as dates, round(count(event)/sum(time\_spent),2) as daily\_metric

from

job\_data

group by ds

order by ds;

**Output:**

Weekly throughput analysis is 0.03 and daily metrics is 0.02, 0.01, 0.05, 0.06. I would prefer the average of weekly throughput anlysis as it takes same when calculated on daily bases.

1. Language Share Analysis: The percentage share of each language for 30 days is analyzed.

Percentage share of each language:

**Query:**

select language, round(100\*count(\*)/total,2) as percentage, sub.total

from job\_data

cross join (select count(\*) as total from job\_data) as sub

group by language, sub.total;

**Output:**

Language Percentage Total

English 12.50 8

Arabic 12.50 8

Persian 37.50 8

Hindi 12.50 8

French 12.50 8

Italian 12.50 8

The percentage of all languages except Persain is 12.50 for 30 days. The percentage share of Parsian language is 37.50

1. Duplicate Rows Detection: From the job\_data table there are many duplicants found in the rows. The duplicants are to be removed.

Duplicate rows:

**Query:**

select actor\_id, count(\*) as duplicates

from

job\_data

group by actor\_id

having count(\*)>1;

**Output:**

There are 2 duplicants found in the actor\_id – 1003.

**Case Study 2: Investigating Metric Spike:**

1. Weekly User Engagement: active users are to be identified on weekly basis. This is found from the events table.

Active weekly users:

**Query:**

select extract(week from occured\_at)as weeks, count(distinct user\_id) as users

from events

where event\_type="engagement"

group by weeks

order by weeks;

**Output:**

Weeks Users

17 663

Weeks Users

18 1068

19 1113

20 1154

21 1121

22 1186

23 1232

24 1275

25 1264

26 1302

27 1372

28 1365

1. User Growth Analysis: A single product is analyzed and its user growth is estimated over a period of time.

User growth:

**Query:**

select year, week\_num, active\_users,

sum(active\_users) over(order by year, week\_num rows between unbounded preceding and current row) as cum\_users

from

(select extract(week from activated\_at) as week\_num,

extract(year from activated\_at) as year,

count(distinct user\_id) as active\_users from users

where state = "active"

group by year, week\_num

order by year, week\_num)a;

**Output:**

The user growth is highest in 33rd week and lowest in 35th week.

1. Weekly Retention Analysis: The control of users is found based on weekly basis on a product after signing off.
2. Weekly Engagement Per Device: The activeness of users is found based on the weekly basis per device.

Activeness of users:

**Query:**

select extract(week from occured\_at) as weeknum, device, count(distinct user\_id) as usercnt

from events

where event\_type = "engagement"

group by weeknum, device

order by weeknum;

**Output:**

31 and 32 weeks of 2014 have the highest users i.e 317 who used the device MacBook Pro.

1. Email Engagement Analysis: Email engagement is found that how users are engaging with the email services.

Users in email services:

**Query:**

select 100\*sum(case when email\_category='email\_open' then 1 else 0 end)/

sum(case when email\_category='email\_sent' then 1 else 0 end)as email\_open\_rate,

100\*sum(case when email\_category='email\_clicked' then 1 else 0 end)/

sum(case when email\_category='email\_sent' then 1 else 0 end) as email\_click\_rate

from

(

select \*,

case

when action in ('sent\_weekly\_digest', 'sent\_reengagement\_ email') then 'email\_sent'

when action in ('email\_open') then 'email\_open'

when action in ('email\_clickthrough') then 'email\_clicked'

end as email\_category

from email\_events) a;

**Output:**

The email open rate is 35.72 and email click rate is 15.73

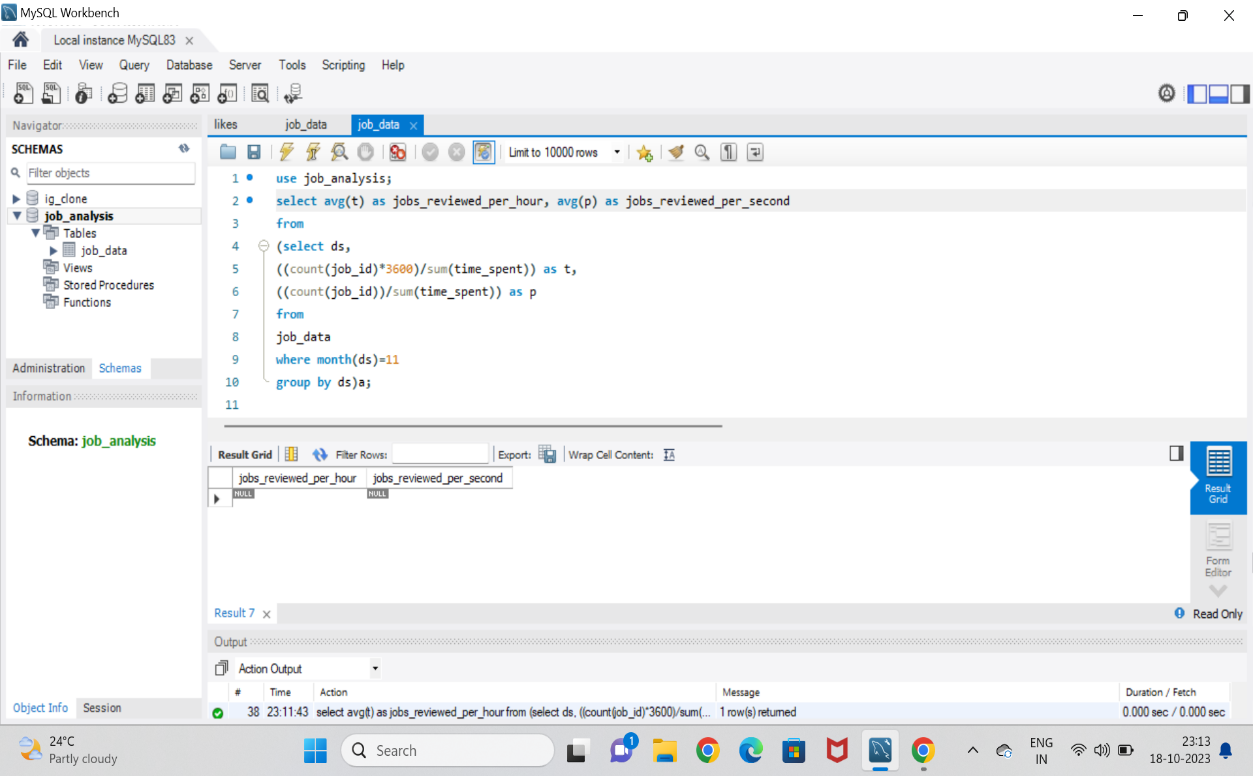
**Tech-Stack Used:**

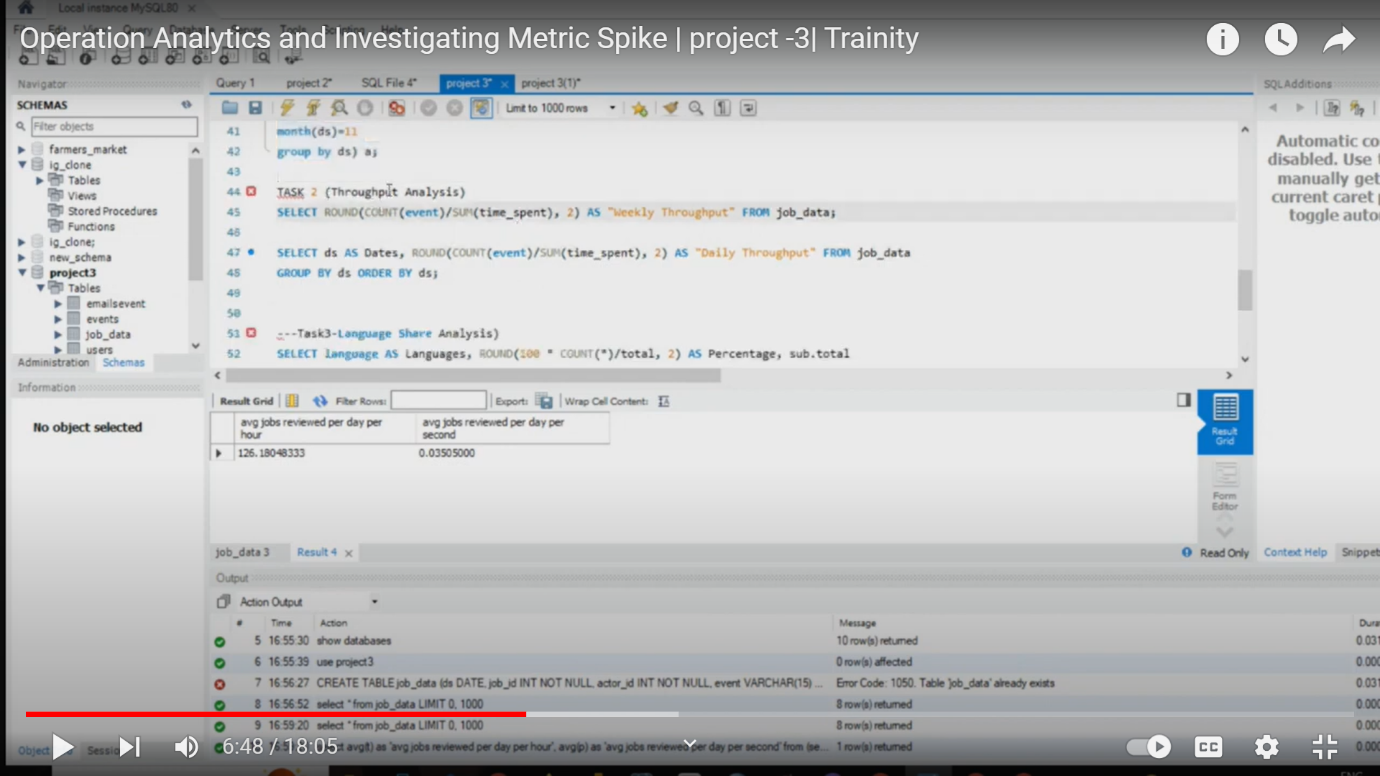
MySQL Workbench with 8.0.34 version. The software is easy to write the queries and Used execute the modified solutions in the form of outputs. The Inputs are given in the form of sql queries which are very flexible to execute.

**Insights:**

**Case Study 1: Job Data Analysis:**

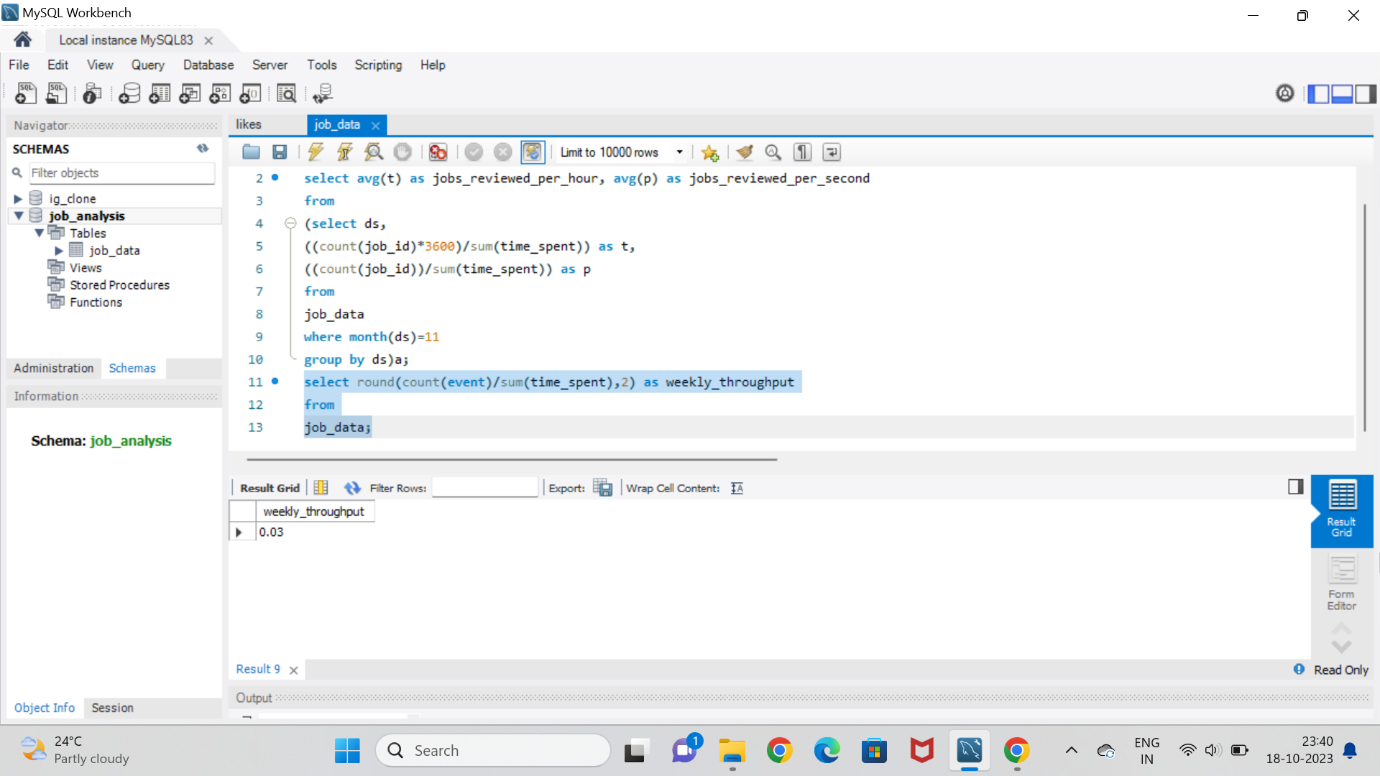
1. **Jobs Reviewed Over Time:**

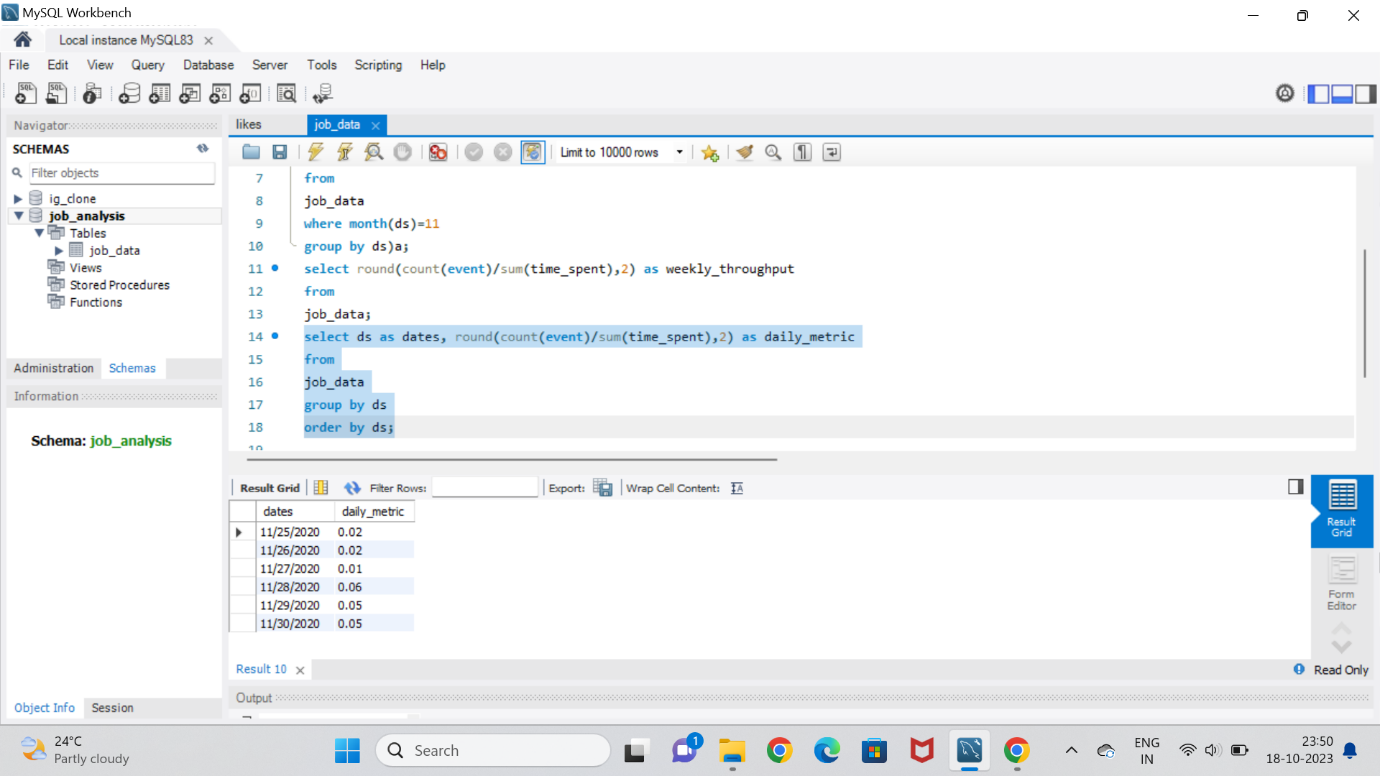
****

****

Jobs reviewed per hour per day is **126.18**

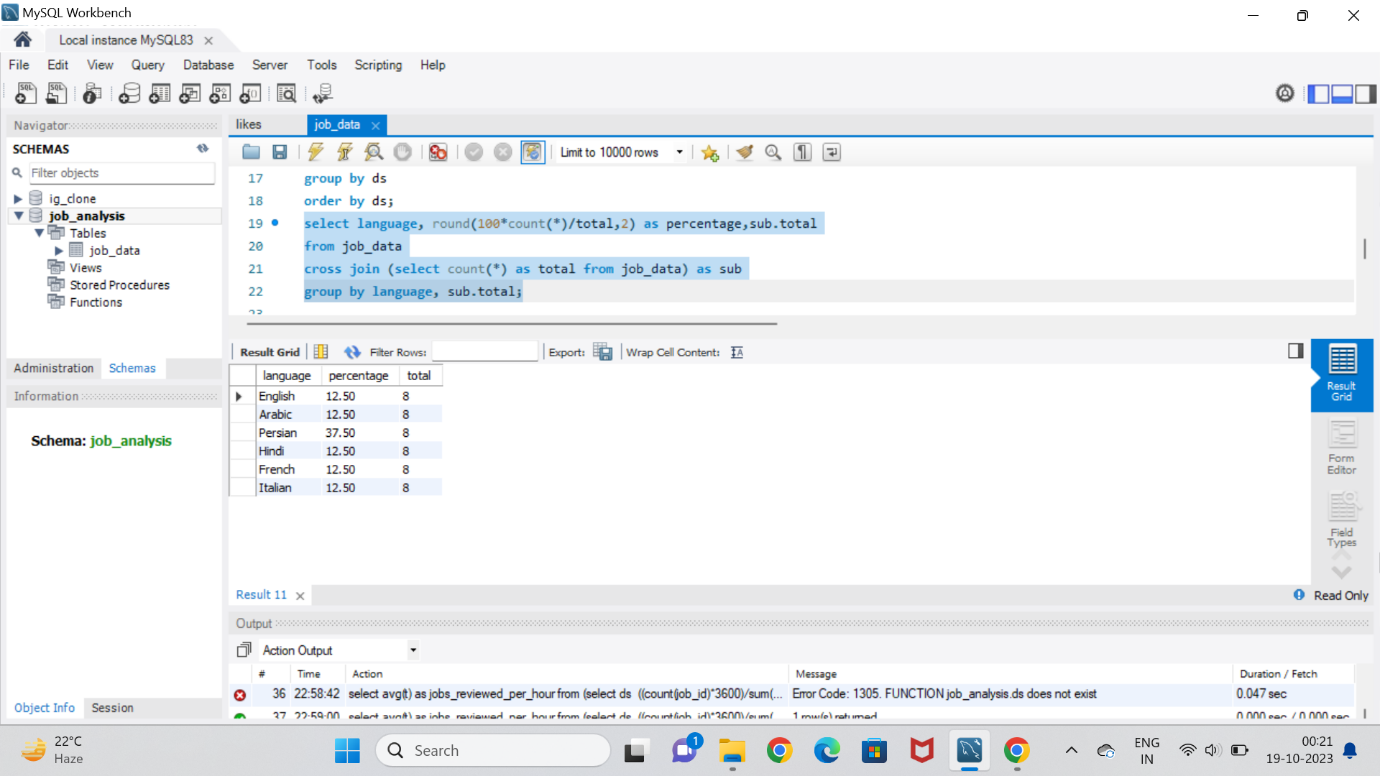
1. **Throughput Analysis:**

****

****

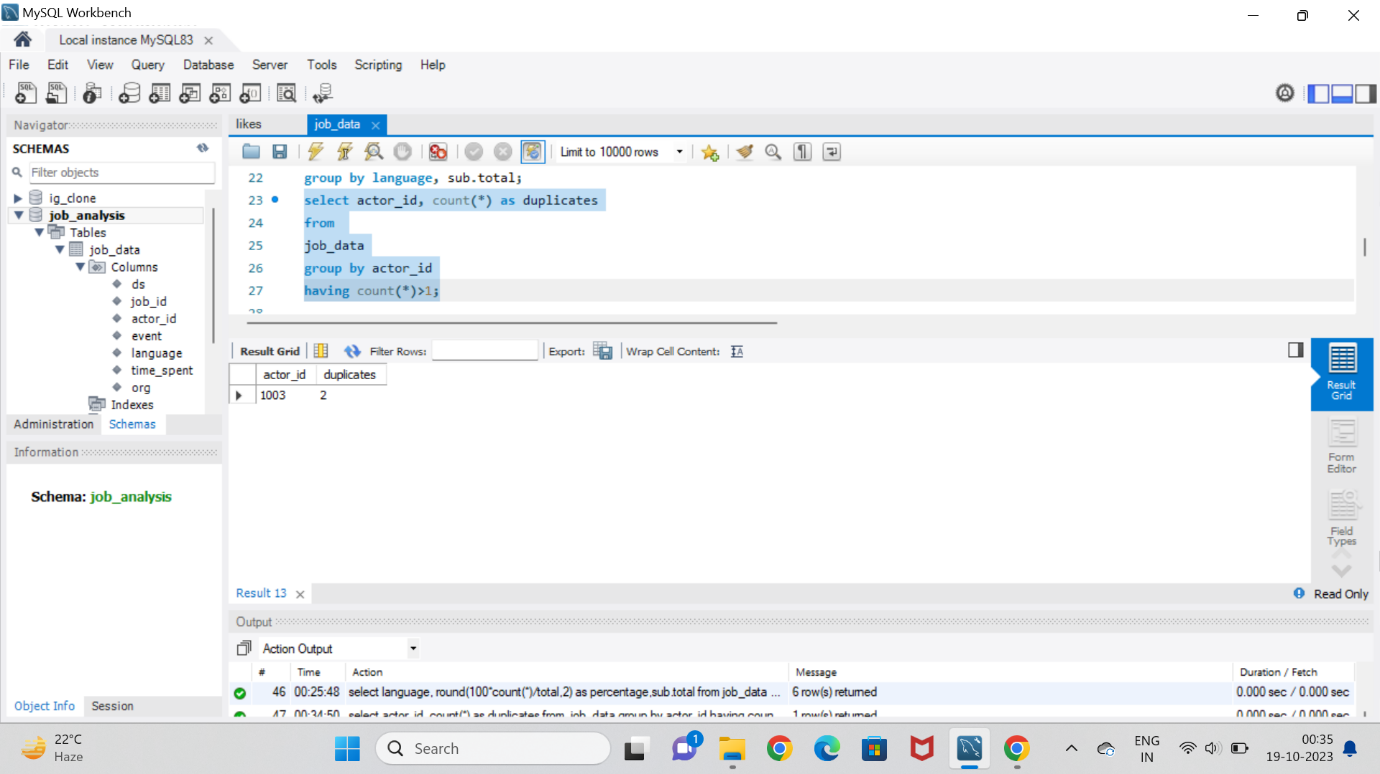
**Weekly** throughput analysis is better than daily metric analysis i.e **0.03**

1. **Language Share Analysis:**

****

The percentage analysis for **Persian** language is higher than other languages.

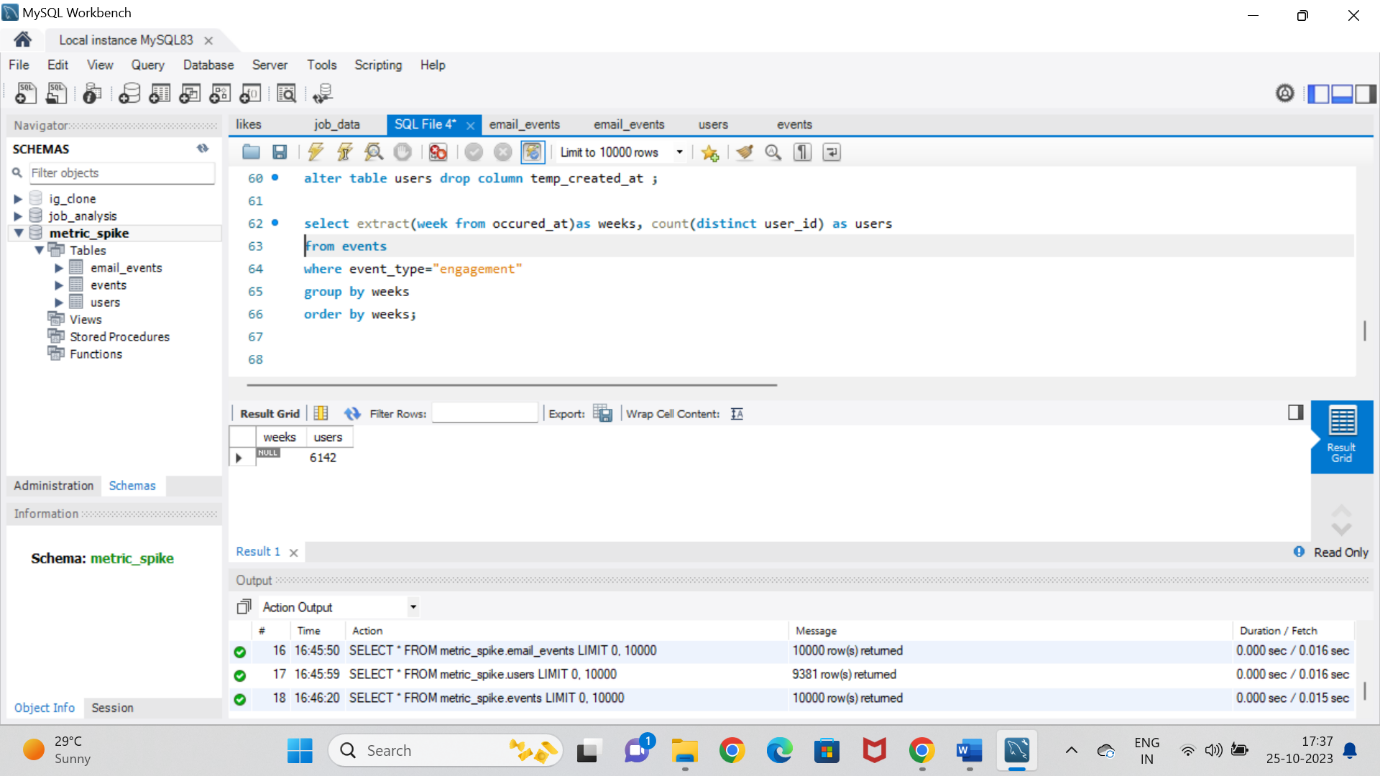
1. **Duplicate Rows Detection:**

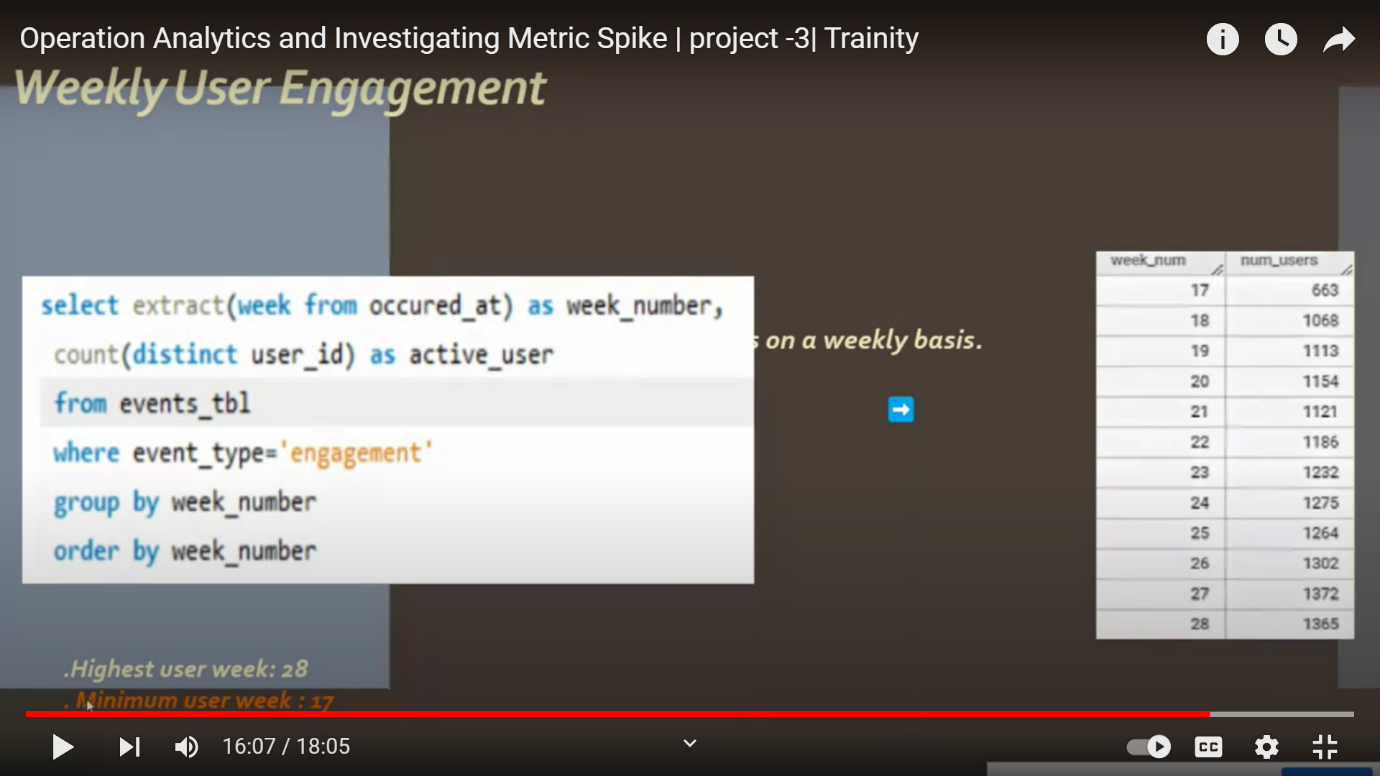
****

There are total **2 duplicate rows** of the **actor\_id 1003**.

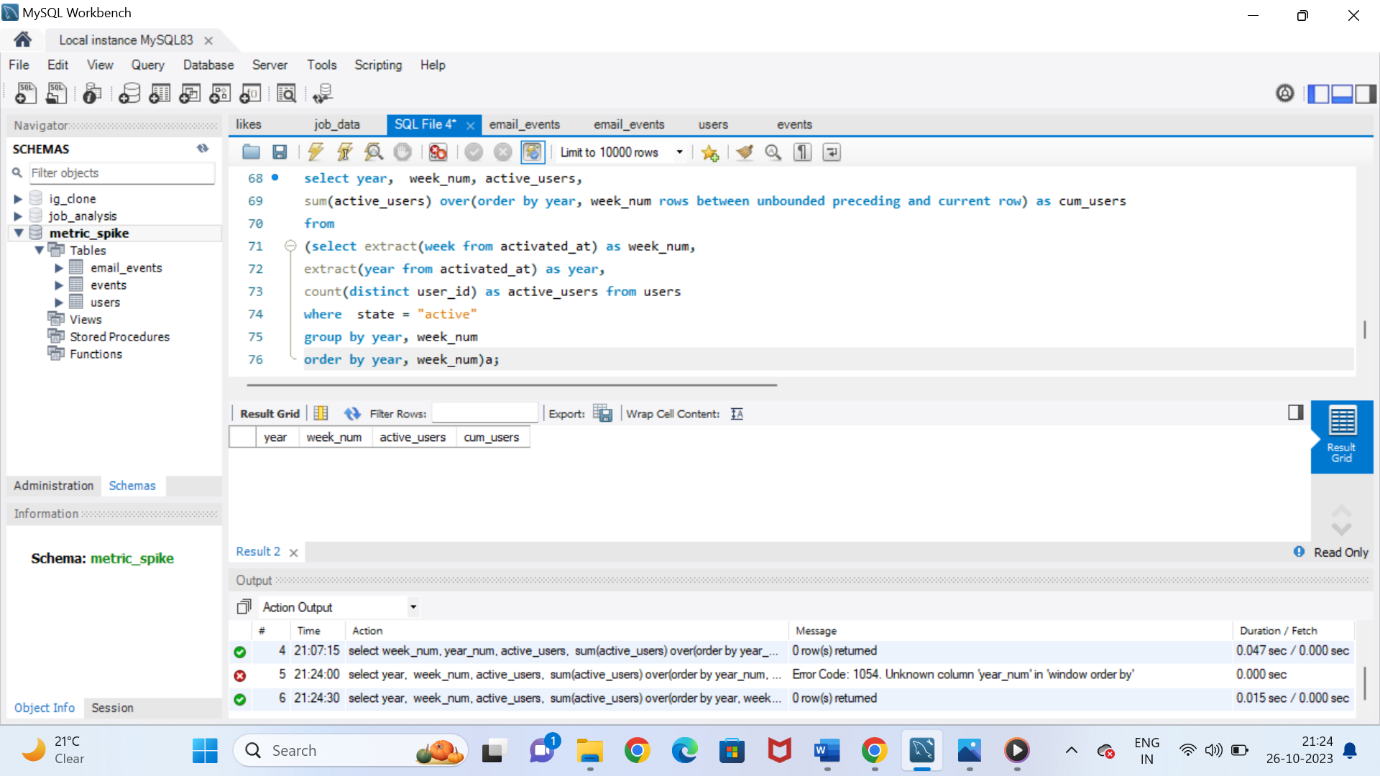
**Case Study 2: Investigating Metric Spike:**

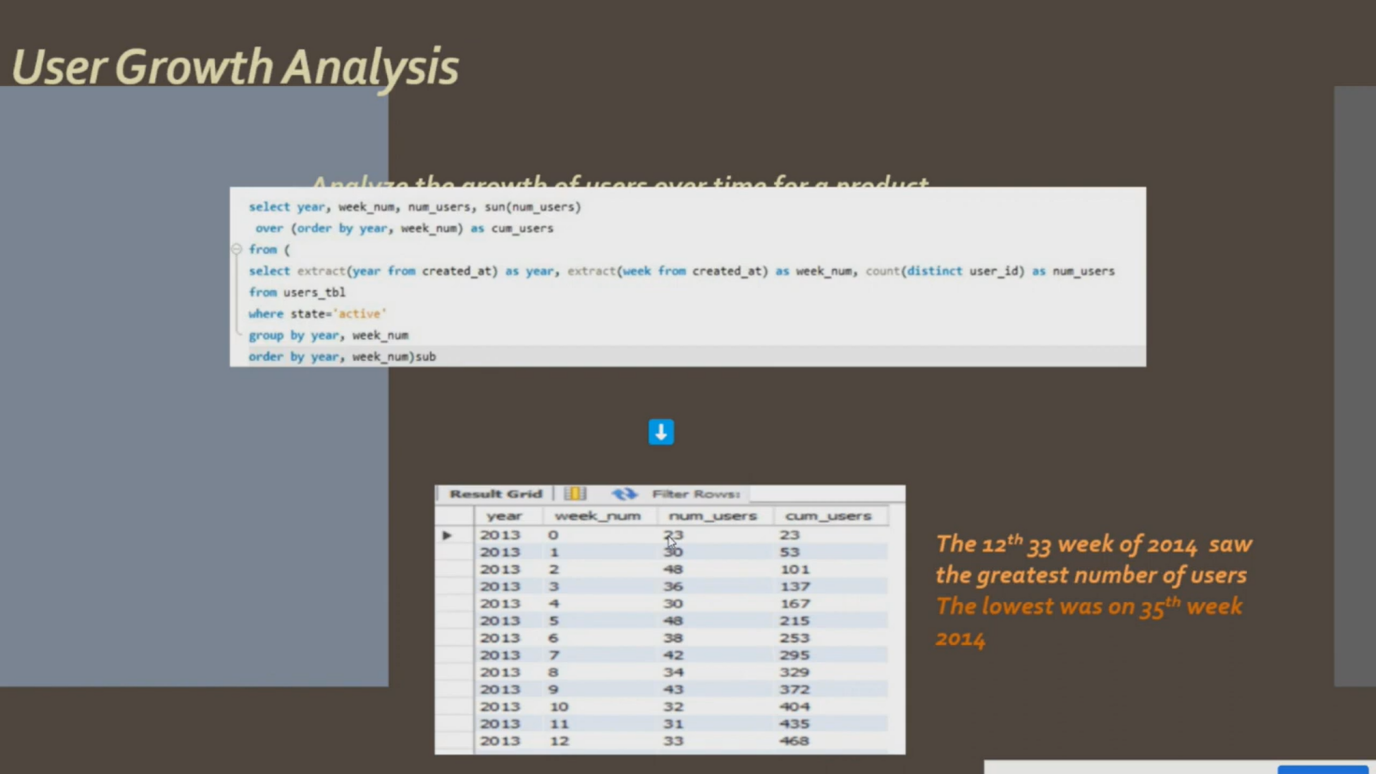
1. **Weekly User Engagement:**





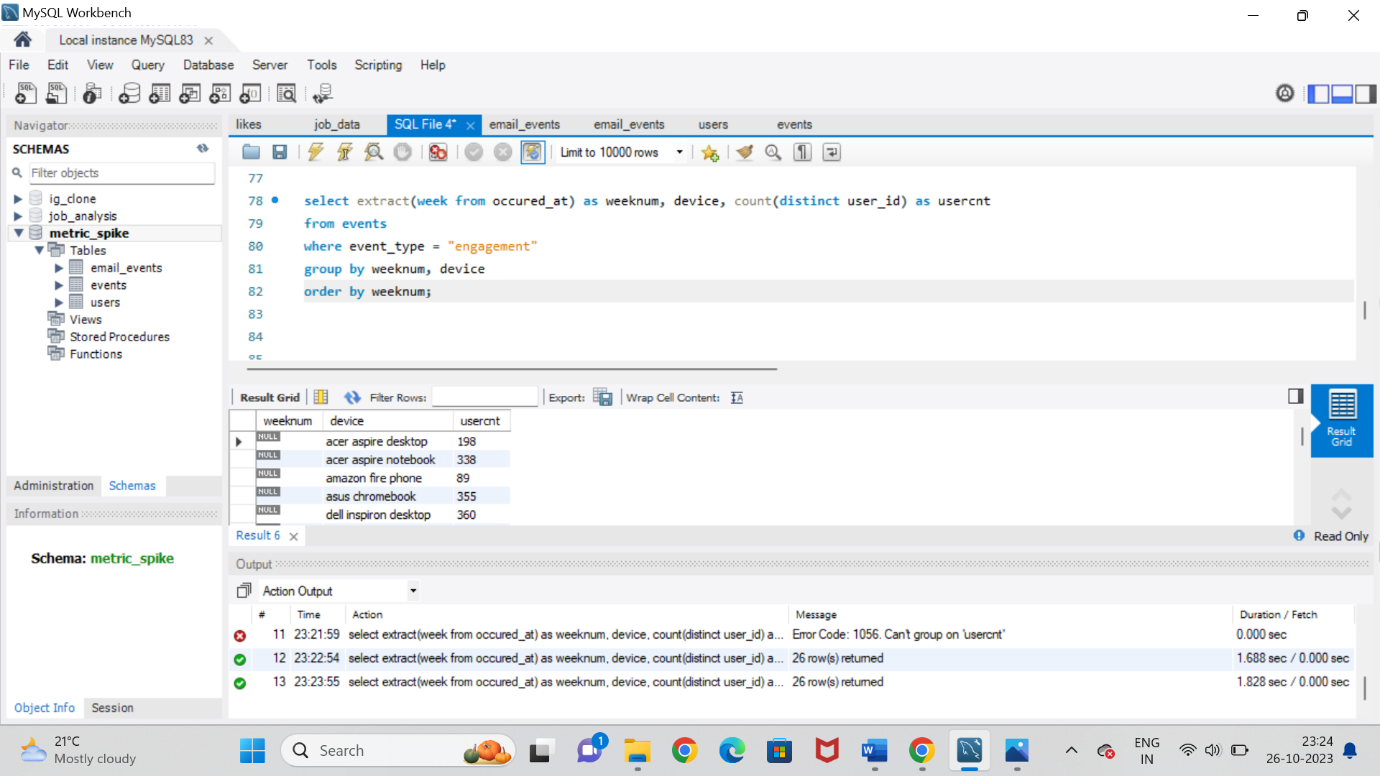
**B** **User Growth Analysis:**

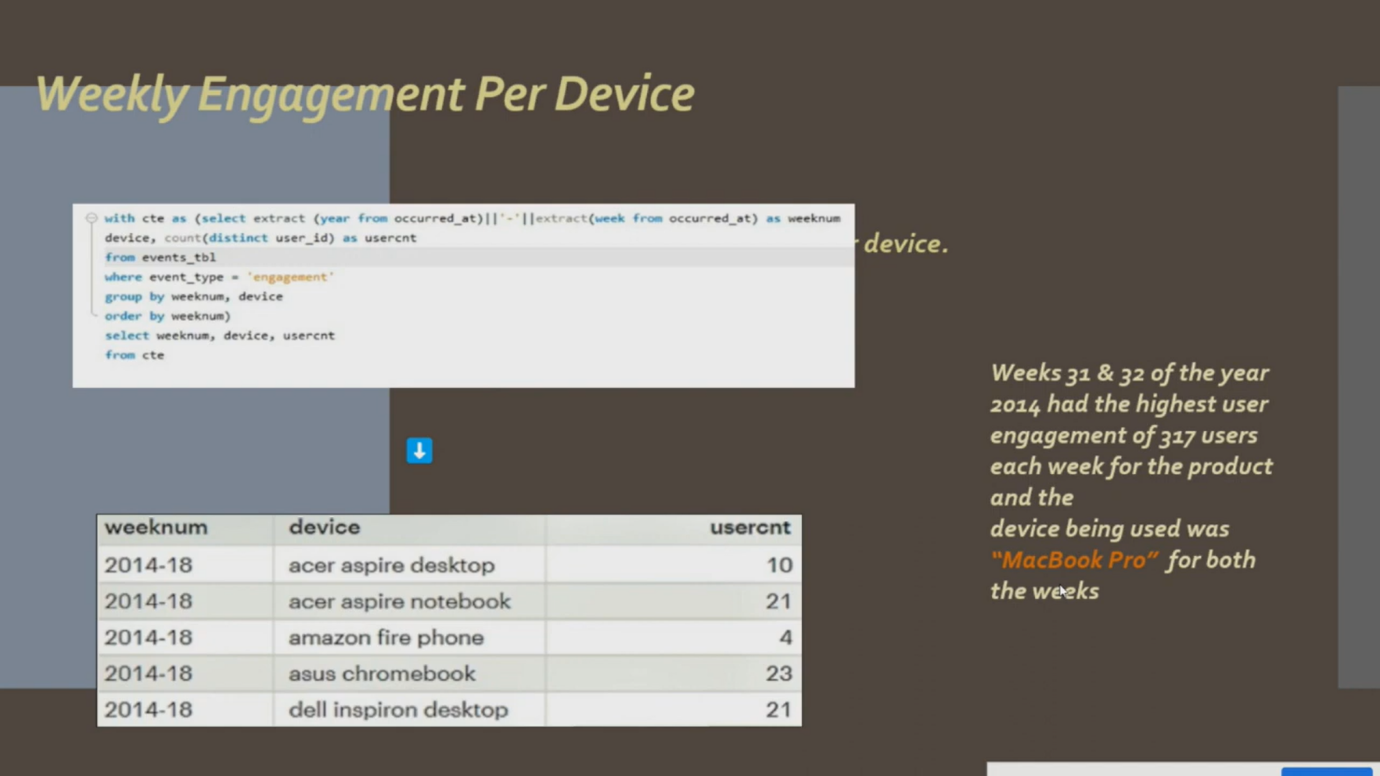




The users growth is higher in **33rd week**

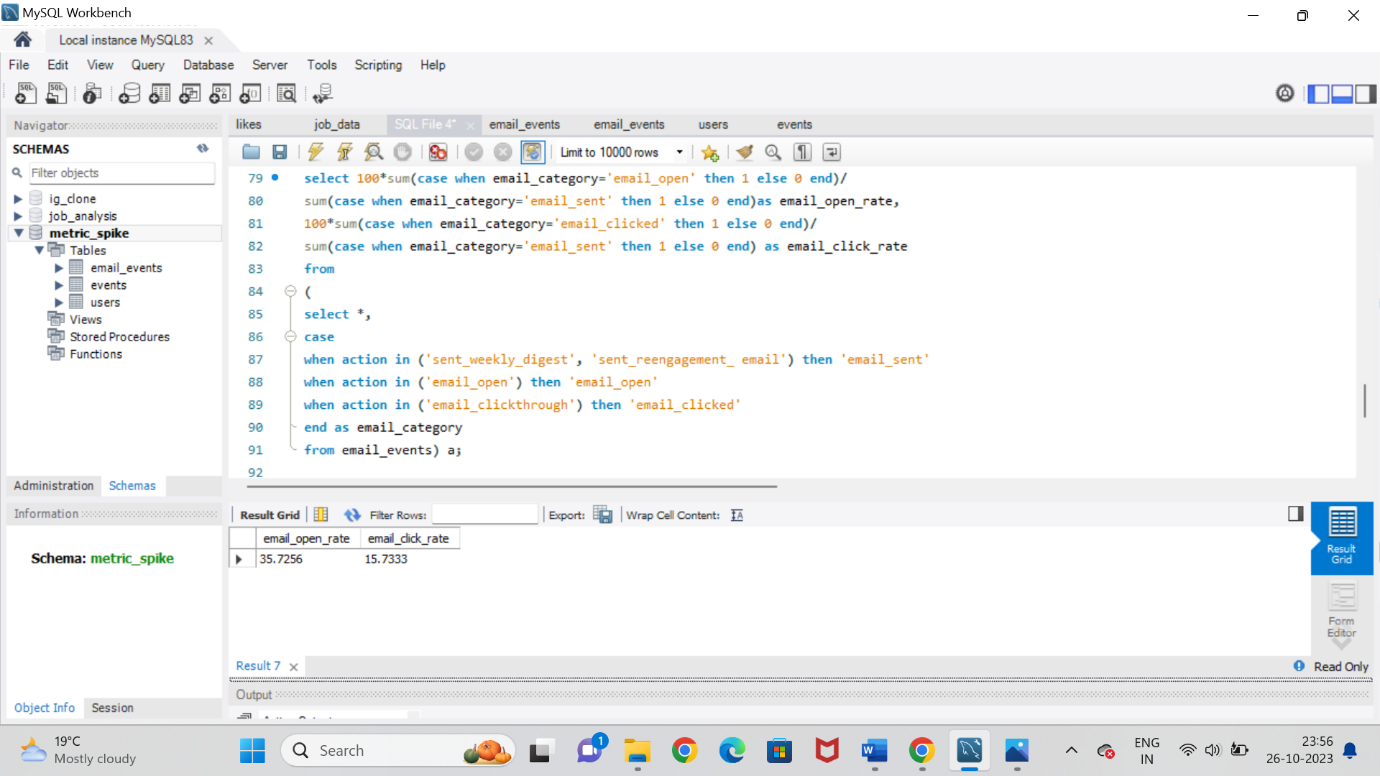
**D** **Weekly Engagement Per Device:**





**31st and 32nd weeks** of 2014 has highest users for **MacBookPro**.

1. **Email Engagement Analysis:**

****

Email open rate is 35.72 and click rate is 15.73

These are the insights provided for the job analysis and investigating metrics.

**Results:**

From the project “Operation and Metric Analytics” we are able to derive the information using sql. I have learnt and clearly understood the process of executing the queries and getting the accurate outputs. This gave me hands on experience on advanced sql by analyzing the data and providing valuable insights. So I conclude that my knowledge on SQL has improved.