PROJECT 2: Operation Analytics and Investigating Metric Spike

Advanced SQL

NAME: K. Yaswanth
GMAIL: kunayaswanth123@gmail.com

To find best reports for the below two case studies operation analytics and investigating metric spike.

SUBMITTING TO TRAINITY :)

SPECIAL THANKS TO TRAINITY TEAM

Project Description:

CASE STUDY 1:
The first case study involves analysing job_data to improve operational efficiency. The project includes finding various metrics such as throughput, productivity and providing recommendations for users to improving overall efficiency.

CASE STUDY 2:
The second case study involves analysing the data to identify patterns and trends such as User growth, user engagement, retention, email metrics

APPROACH: -

First, the data prepared and analysed using SQL. Various SQL functions like SELECT, WHERE, GROUP BY, JOIN, etc. are used to extract meaningful information from the dataset. Using proper window functions like over() at the required part to get better result.

TECH STACK USED:

MySQL, mode.com (used the datasets yammer.events, yammer.email, yammer.users)

(Taken the help of power labs for case study 2in retention problem and cohert analysis)

Case Study 1 :- Operation Analytics :-

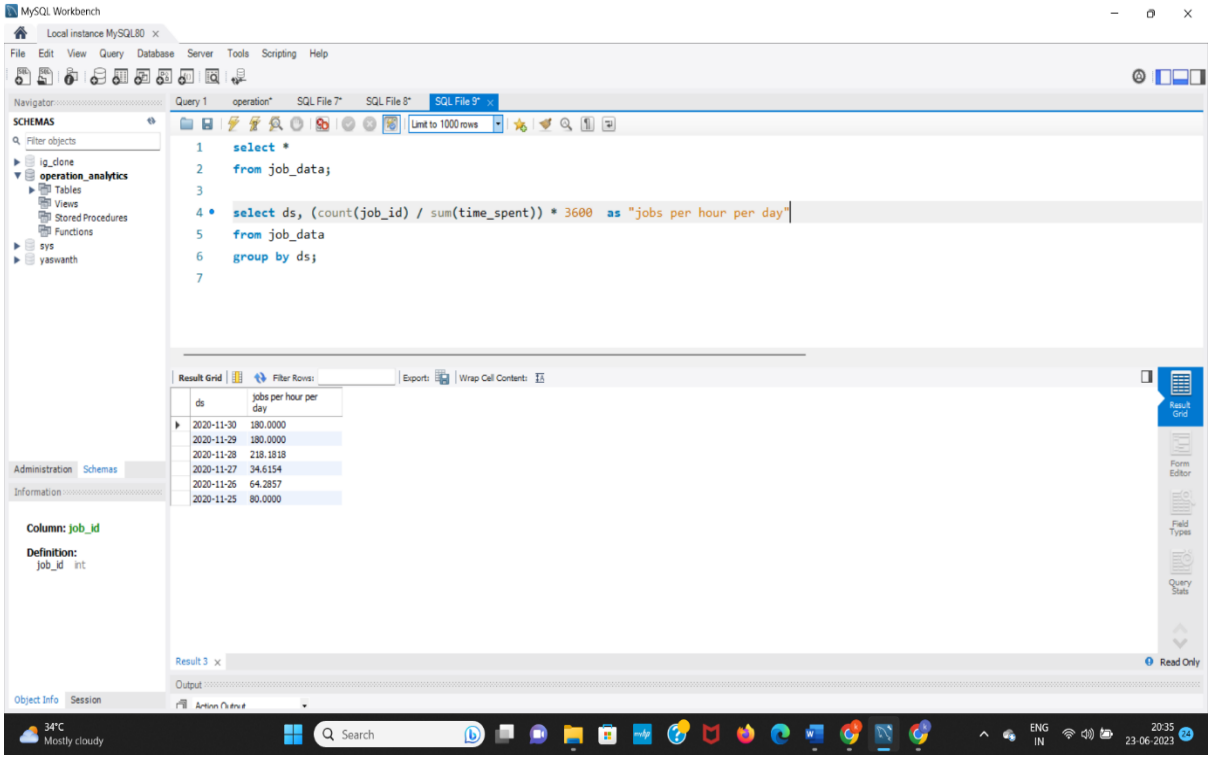
1. **Number of jobs reviewed:** Amount of jobs reviewed over time.
   Your task: Calculate the number of jobs reviewed per hour per day for November 2020?

First We create the table job_data

```
create table job_data(
job_id int,
actors_id int,
event varchar(255),
language varchar(255),
time_spent int,
org varchar(255),
ds date);
```

```
INSERT INTO job_data
(ds, job_id, actors_id, event, language, time_spent, org)
VALUES
('2020-11-30', 21, 1001, 'skip', 'English', 15, 'A'),
('2020-11-30', 22, 1006, 'transfer', 'Arabic', 25, 'B'),
('2020-11-29', 23, 1003, 'decision', 'Persian', 20, 'C'),
('2020-11-28', 23, 1005,'transfer', 'Persian', 22, 'D'),
('2020-11-28', 25, 1002, 'decision', 'Hindi', 11, 'B'),
('2020-11-27', 11, 1007, 'decision', 'French', 104, 'D'),
('2020-11-26', 23, 1004, 'skip', 'Persian', 56, 'A'),
('2020-11-25', 20, 1004, 'transfer', 'Italian', 45, 'C');
```

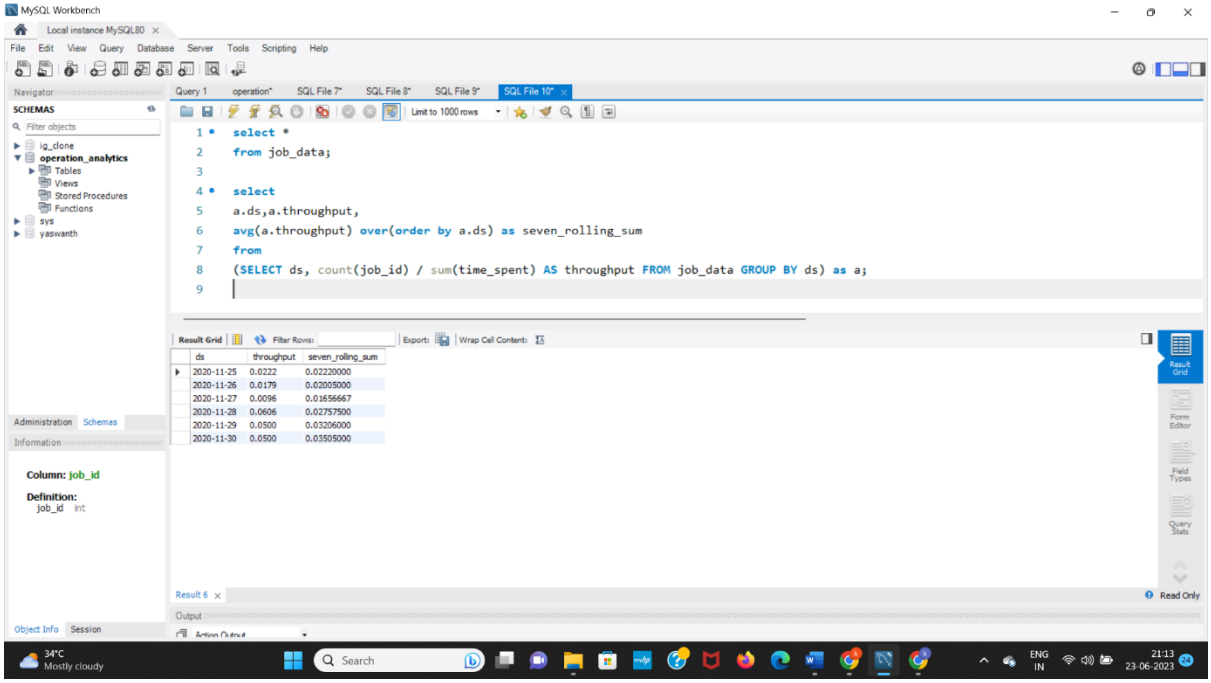
Ans:-



Explanation :- To calculate no of jobs we use count(job_id), for that we calculate need hour from seconds so we multiply by 3600 , we group it with date because they asked per day in question.

2. **Throughput:** It is the no. of events happening per second.
   Your task: 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?

Ans:-



Explaination:

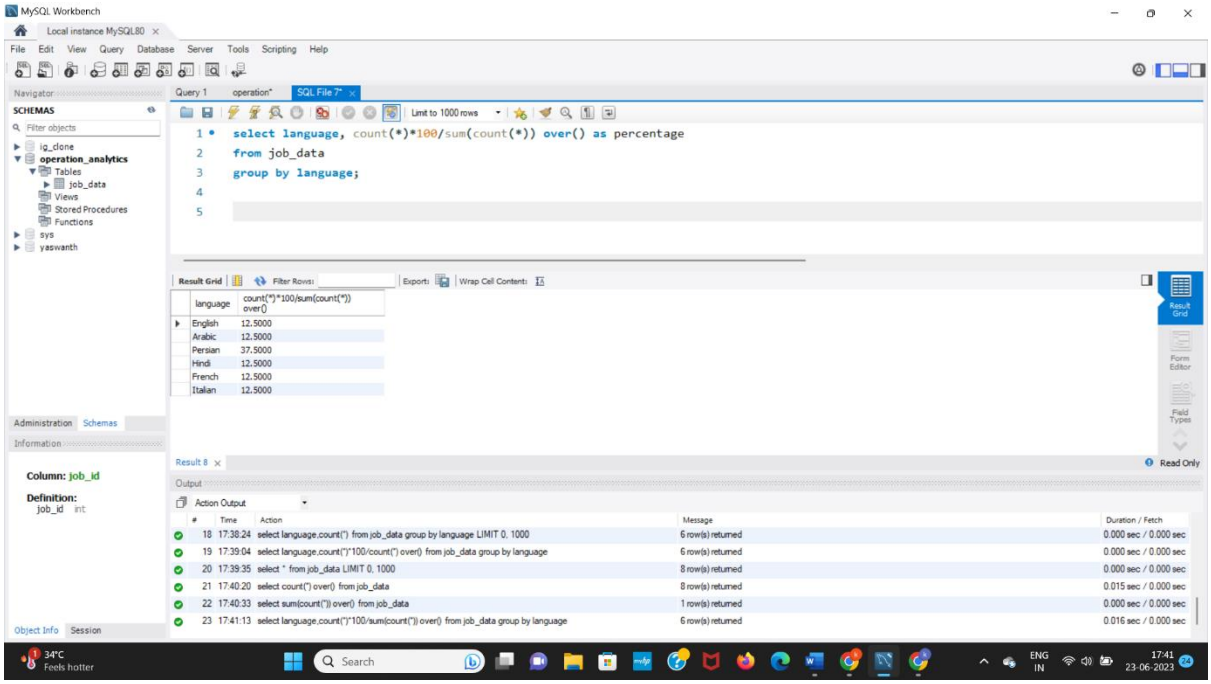
Daily throughput vs 7 day rolling throughput

To find daily throughput: it is no of events per sec. So, we count(job_id) to the total seconds group by date. Then, we get daily throughput.

Then seven day rolling throughput :-
We use over() window function in that order by then we find cumulative average in step which is rolling average

3. **Percentage share of each language:** Share of each language for different contents.
   Your task: Calculate the percentage share of each language in the last 30 days?

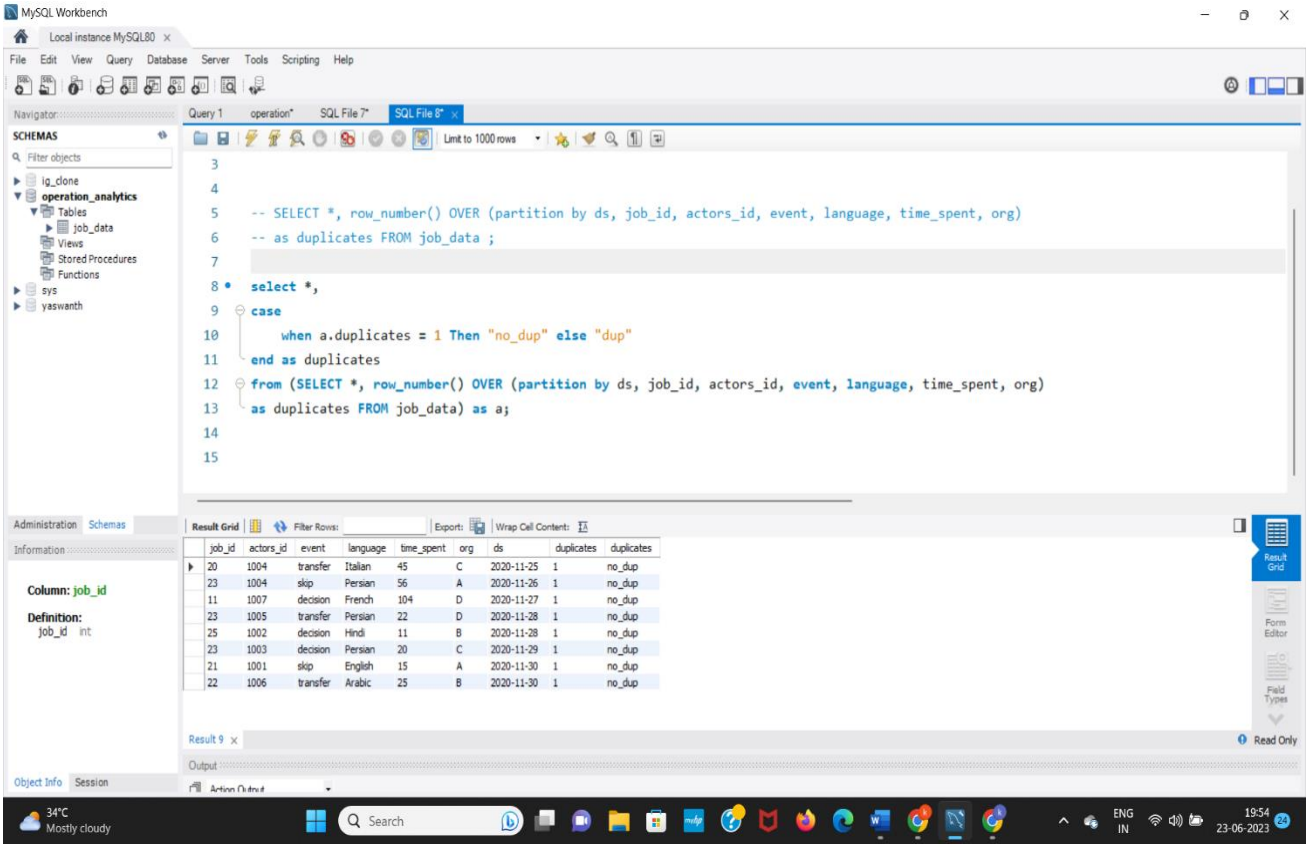
Ans:-



Explaination :-
Percentage share of each language is multiply by 100. Total contents is 8, we find it by using over() window function(total inputs = total outputs) then group by each language and we aggregate the language share.

4. **Duplicate rows:** Rows that have the same value present in them.
   Your task: Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

Ans:



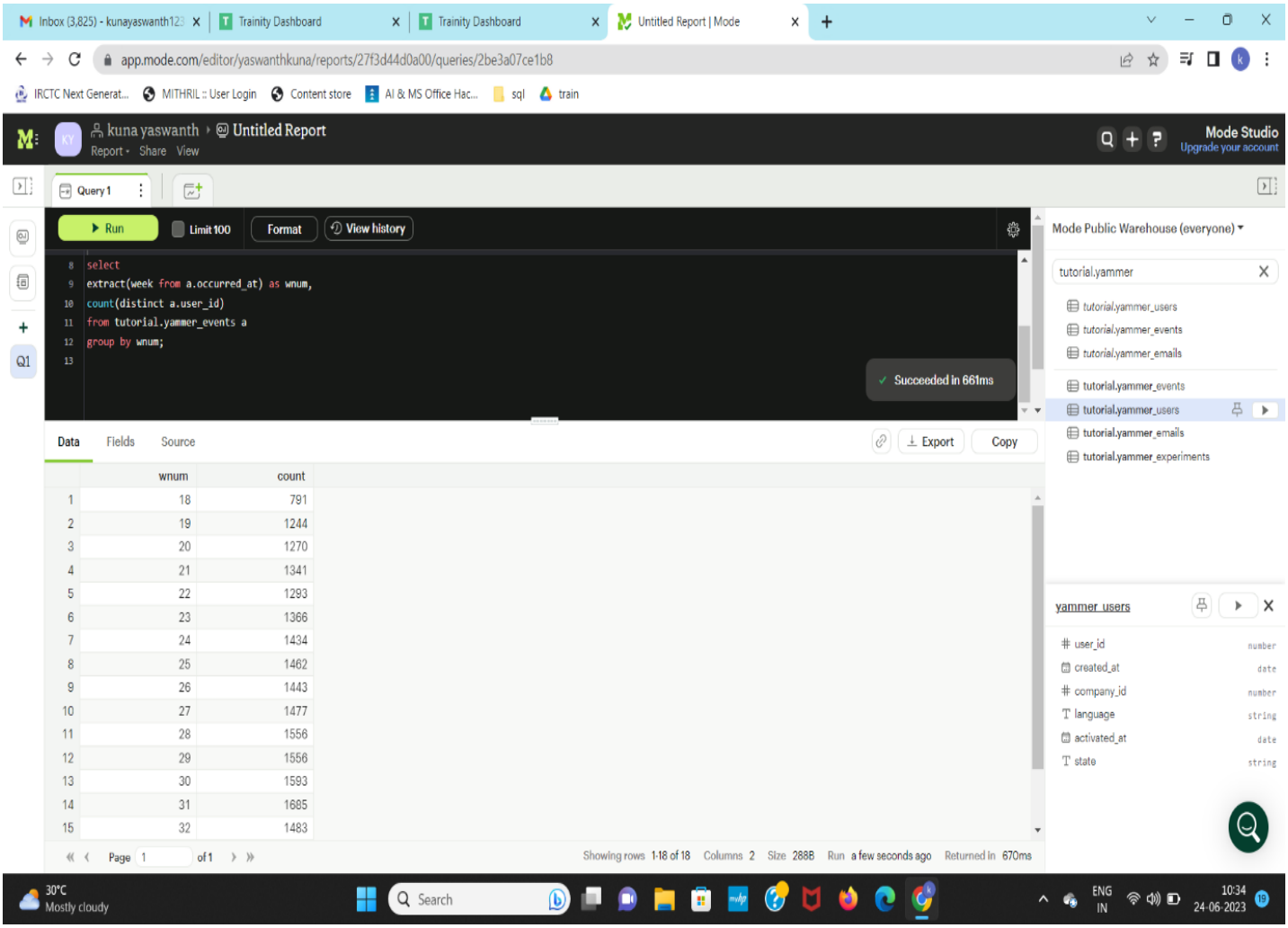
Explanation:-

1. We use row_number() with over() window function to assign same row numbers and in that part with date,job_id,actors_id,event,language ,time_spent,org because we are in need of finding duplicates.
2. When we get duplicates greater than 2 then it is dup
3. When we get duplicates = 1 then there is no dup (no duplicates)
4. Then we select duplicates finding > 1

CASE STUDY 2: INVESTIGATING METRIC SPIKE

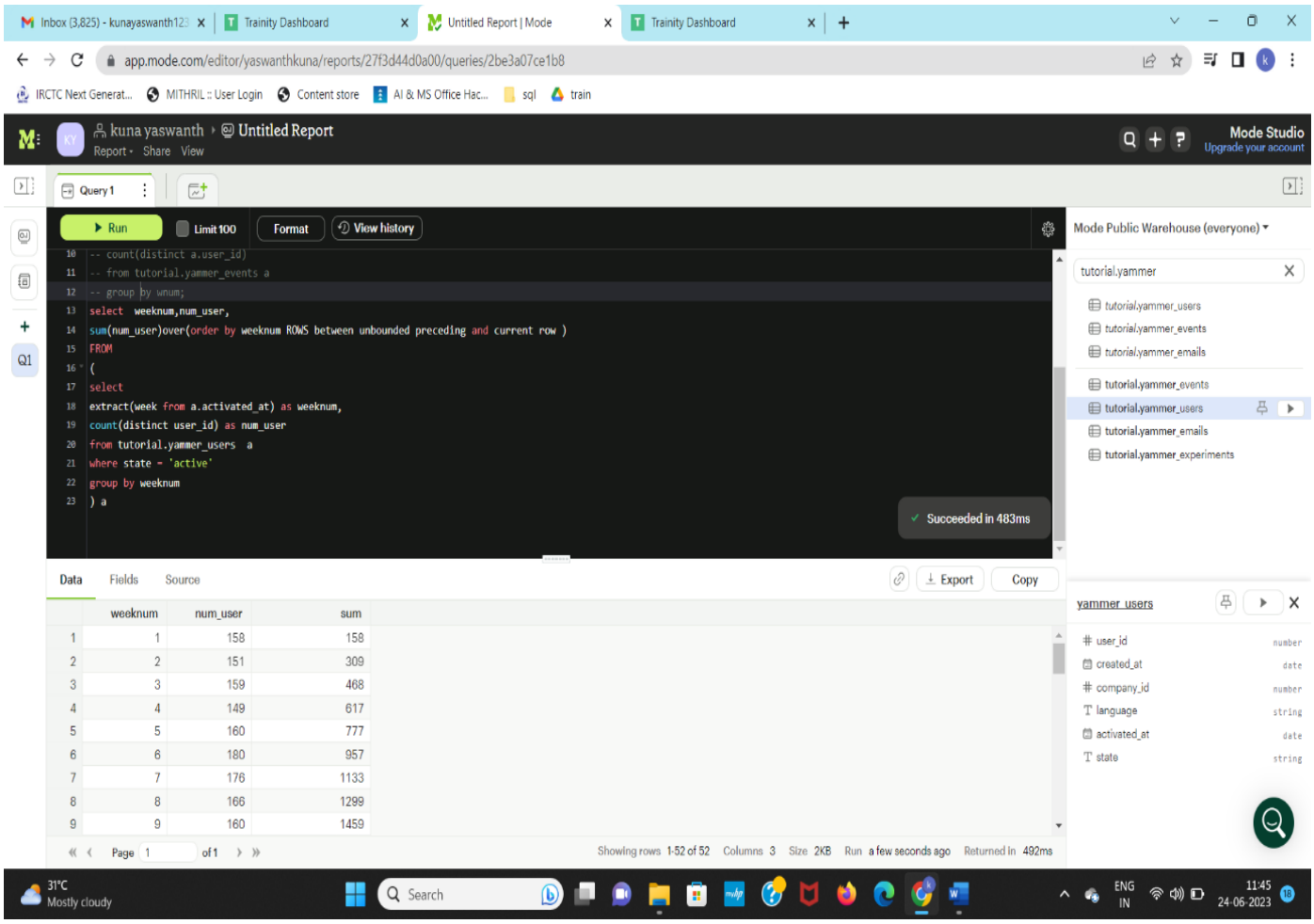
1. **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.
   Your task: Calculate the weekly user engagement?

Ans:



Explanation: - To find weekly user engagement.
Weekly – so group by week, Engagement – use events table
Aggregating by counting different users.

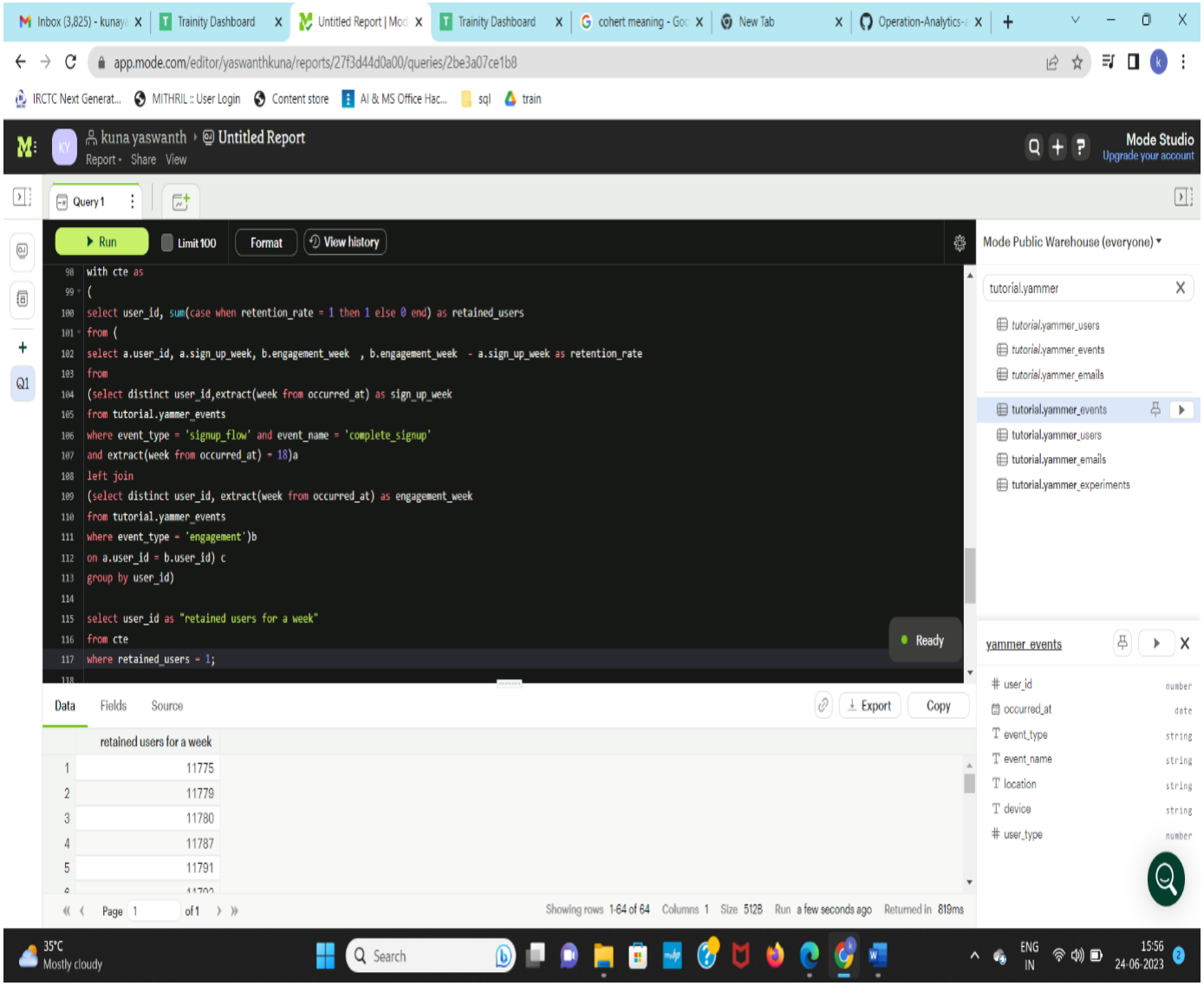
2. **User Growth:** Amount of users growing over time for a product.
   Your task: Calculate the user growth for product?



Explanation :- amount of users growing over time
To find total users for every week we group by week and find count of different users whose state is active
Then to find growth we find the cumulative of users using window function.

3. **Weekly Retention:** Users getting retained weekly after signing-up for a product.
Your task: Calculate the weekly retention of users-sign up cohort?

Ans:-

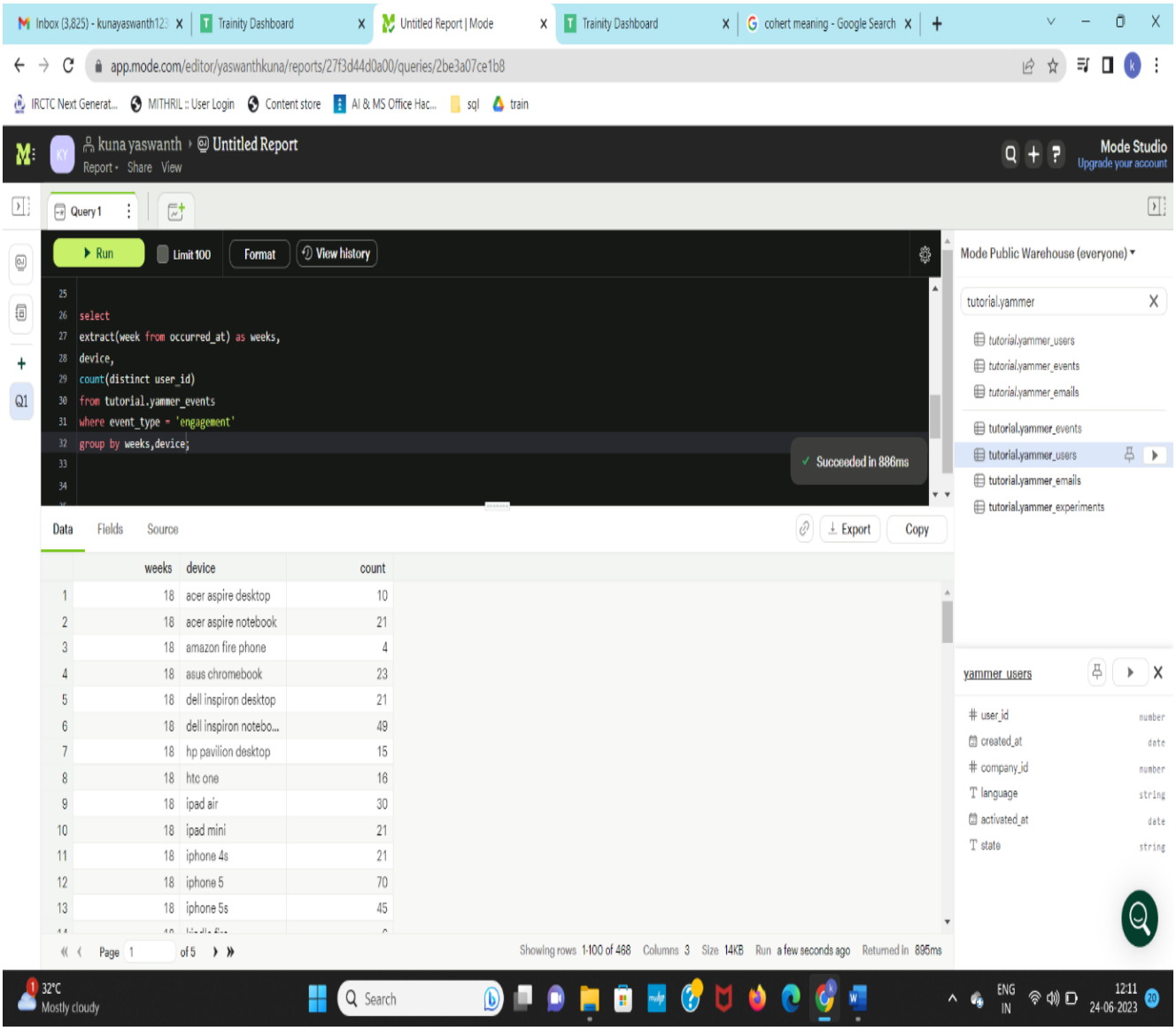


Explanation: retention = next_day – today

1. Find people signup in first week left join
2. Find people engage in the next weeks with user_id is common in both tables
3. From this we select a.user_id, a.sign_up_week, b.engagement_week, b.engagement_week – a.sign_up_week = retention_rate
4. From this we assign retention rate value 1 with 1 other with 0 using case statement we make this table as common table expression.
5. From this common table expression we select users having retained_rate as 1
6. Finally, we get all the retained users.

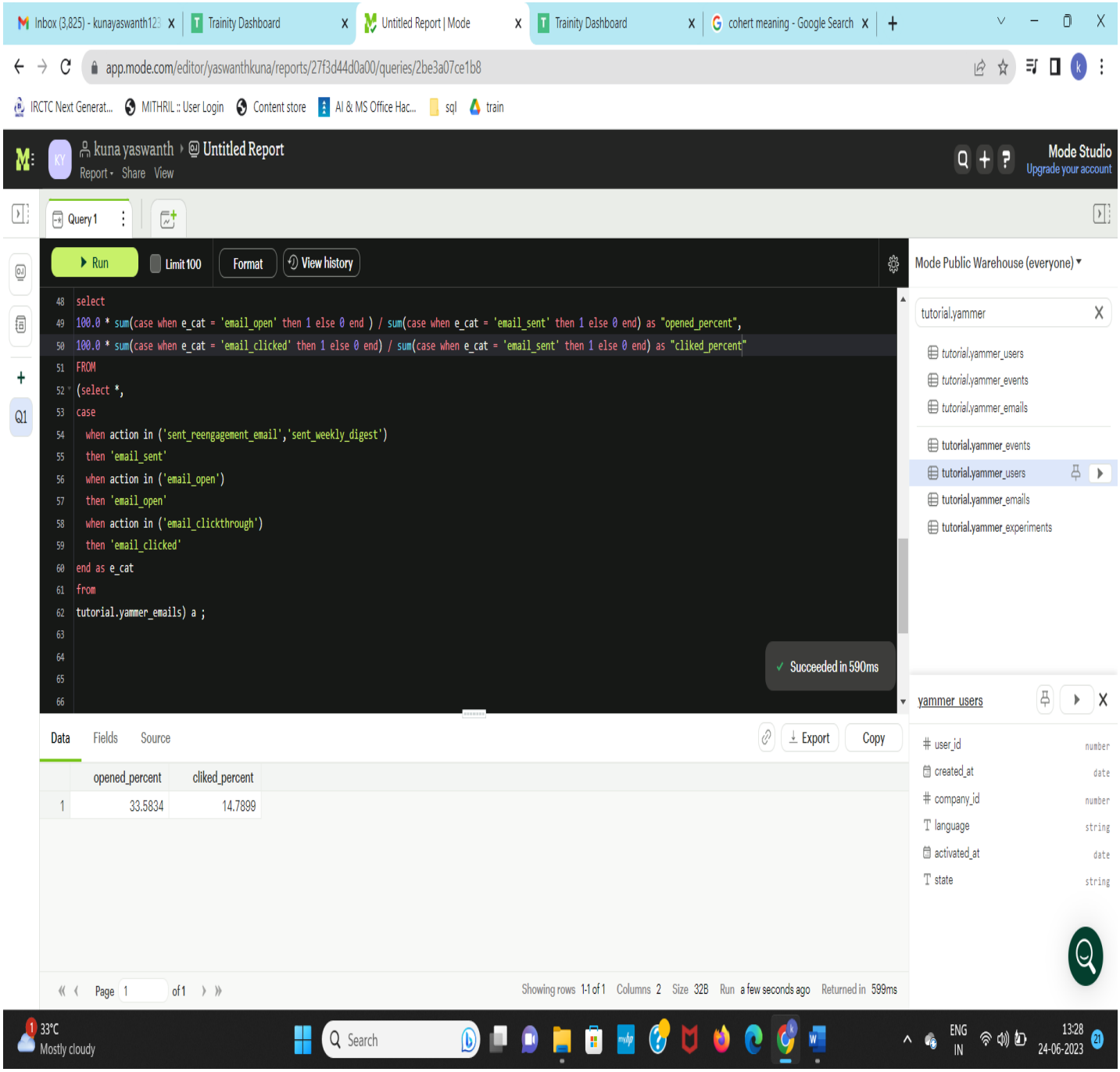
4. Weekly Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
Your task: Calculate the weekly engagement per device?

Ans:-



Explanation:- From events table, weekly engagement per device. So, group by week, device. And event_type = “engagement”, Counting the users.

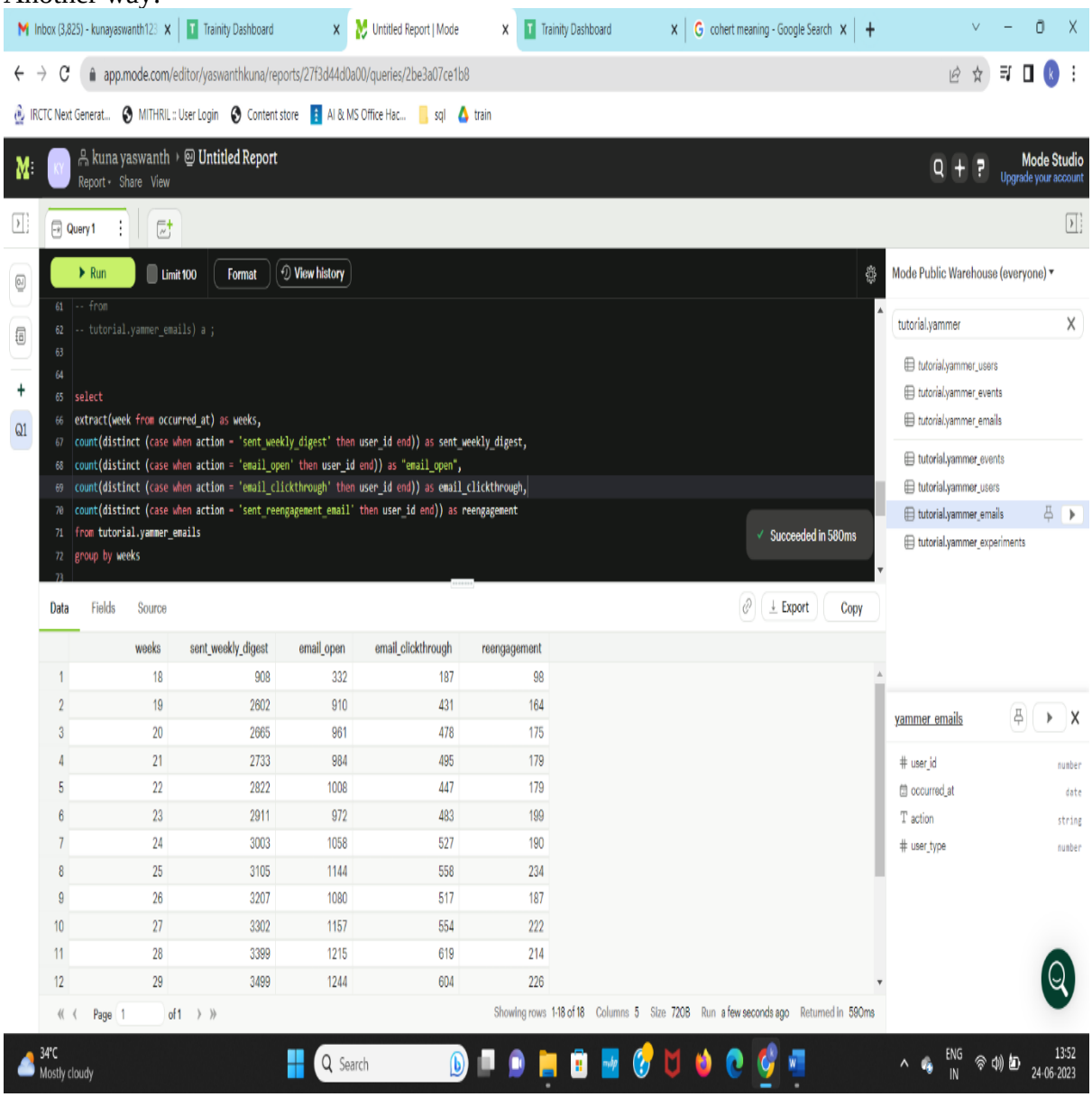
5. **Email Engagement:** Users engaging with the email service.
   Your task: Calculate the email engagement metrics?



Explanation :-

1. Here, we have distinct action such as email_open, sent_reengagement_email ,sent_weekly_digest ,email_sent
2. sent_reengaement_email as sent using case, sent_weekly_digest as sent using case email_open as open all these as e_cat
3. finally, we select opened percent as (count(open)) / count(sent) as open
4. then we select clicked percent as (count(click)) / count(sent) as clicked
5. we get these two things as email engagement.

Another way:-



Explanation: email engagement metrics.
users engaging in distinct action over 7 days means we group by week and then count all the distinct actions
(all distinct actions means count(sent_weekly_digest), count(sent_reengagement), count(email_clickthrough), count(email_open))

We get the actions of the mail done by users in 7 days.

Insights:-

1. With the help of this project, I gained hands on practical experience with window functions
2. With the help of this project, I gained at most knowledge on subqueries and usage of them in the practical way including hands on
3. Solved the complex problems like retention with better way
4. Through project labs learnt some good knowledge on retention and cohert analysis.

Result: -

It was a great learning opportunity for me as I was able to get hands on experience, and also improve my skills in MySQL. This project also helped me understand and complex queries, retention, cohert analysis. The guidance and resources offered by the team made learning easier and the project a success. Project labs helped me to solve retention problems. I' m grateful and thankful for the wonderful opportunity.

Thank you trainity 😊