

Operation Analytics & Investigating Metric Spike

Trainity Project Report Rohit Kumar rohitk.ug20.ce@nitp.ac.in

Description

The study carried out for the entire end-to-end operations of a corporation is covered in this report's discussion of Operation Analytics and Investigating Metric Spike. This helps the business identify the areas where it needs to make improvements. The VP, top officials, operational team, support staff, marketing team, sales team, etc. can all benefit from this report.

Being one of the most crucial components of a business, this form of analysis is also utilized to forecast the general upward or downward trend in a company's fortune. Better automation, improved communication among cross-functional teams, and more efficient workflows are the results.

This report is based on the "Job data" dataset, which was used in the previous case study's operation analytics. Case Study 2 will use the "Yammer dataset" to investigate metric spikes and provide answers to queries about user engagement, user growth, email engagement metrics, and other topics.

Approach

This report will be divided up primarily into two sections.

1) Operation Analytics

- **I.** Number of jobs reviewed To number of jobs reviewed per hour per day for November 2020?
- II. Throughput will calculate on weekly and daily basis
- III. Percentage share of each language -To find percentage share of each language in the last 30 days?
- IV. Duplicate rows To highlight the duplicate rows

Approach

2) <u>Investigating Metric Spike</u>

- I. User Engagement User engagement basis on activeness of user on weekly basis
- II. User Growth Growth of active user on weekly basis
- III. Weekly Retention Applying cohort analysis to calculate the weekly retention of users-sign up.
- IV. Weekly Engagement calculation of weekly engagement per device
- V. Email Engagement Email Engagement metrics

Tech-Stack Used

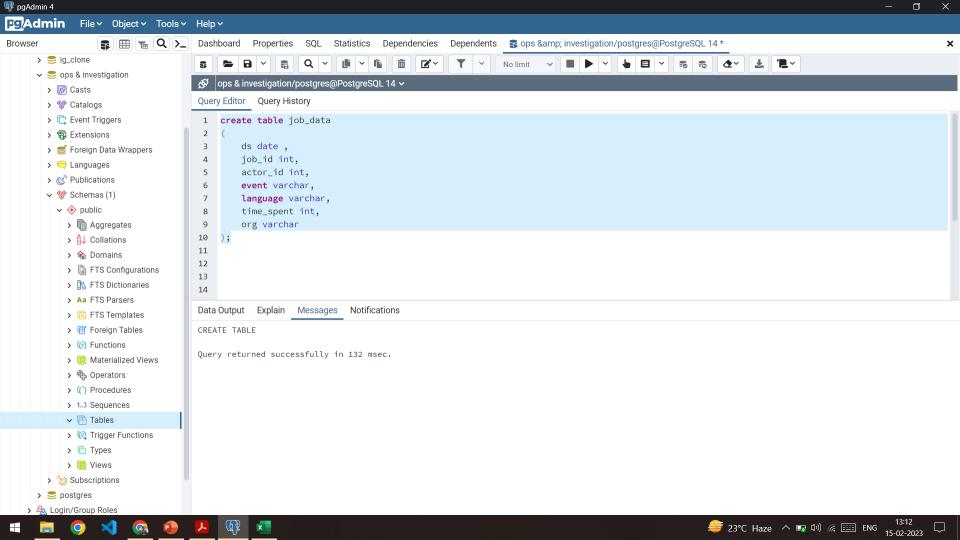
For this Project I used PostgreSQL PG Admin 4 version 6.8

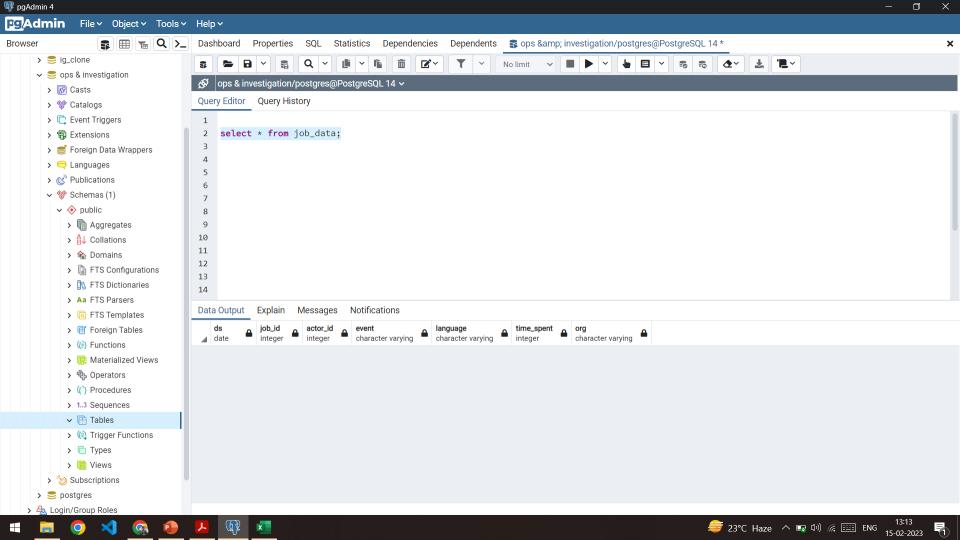
About pgAdmin 4									
Version	6.8								
Application Mode	Desktop								
Current User	pgadmin4@pgadmin.org								
NW.js Version	0.55.0								
Browser	Chromium 92.0.4515.107								
Operating System	Windows-10-10.0.19045-SP0								

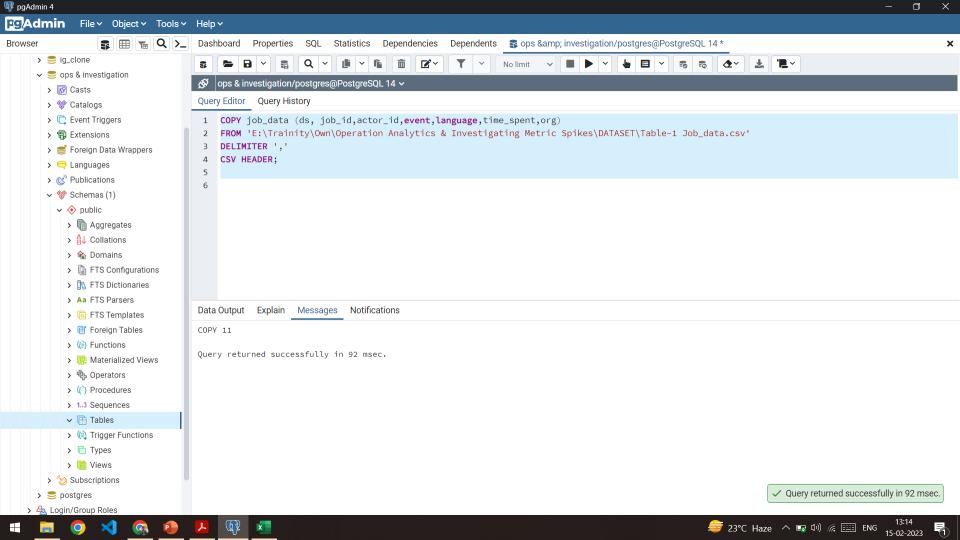
This project is based on the provided "Case_Study" dataset.

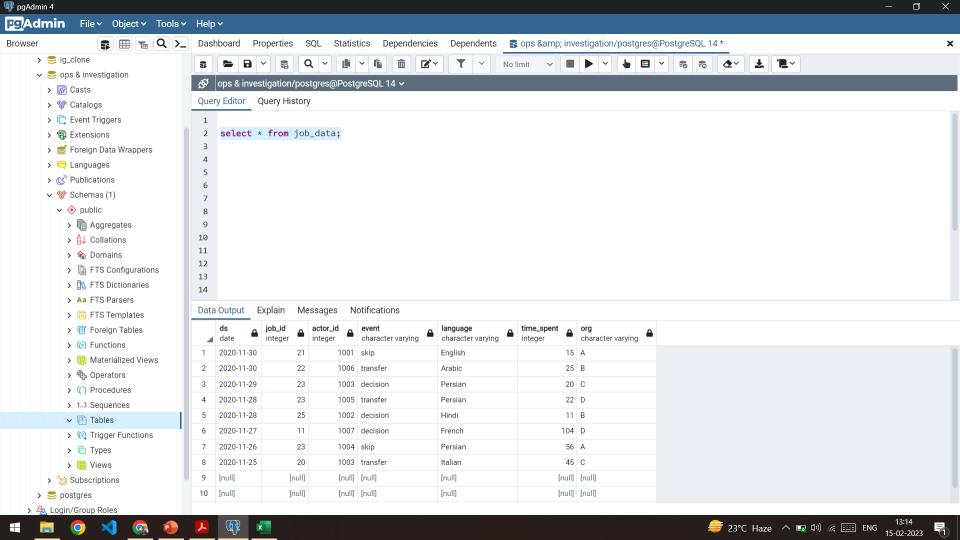
CASE_STUDY_Dataset]













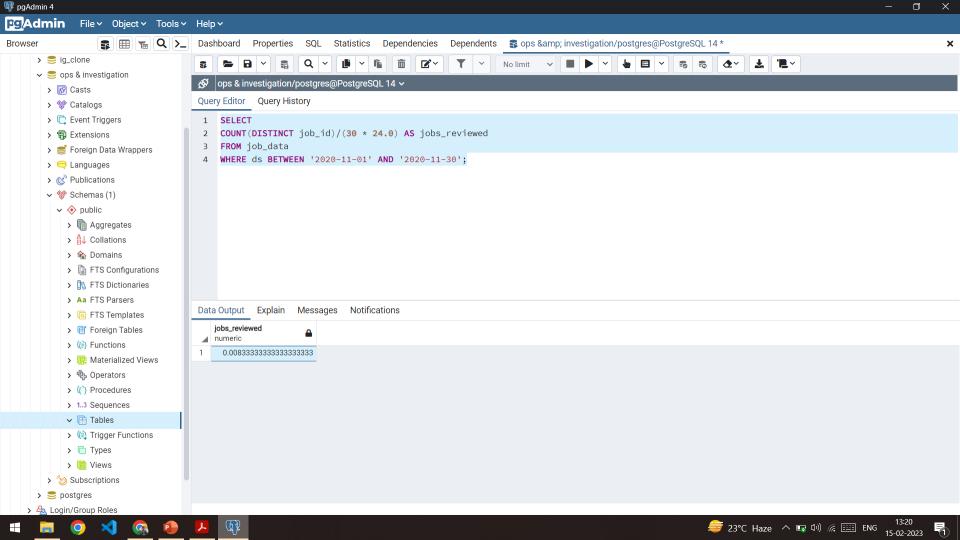
1 (I) Numbers of Job Reviewed

Number of jobs reviewed per hour per day for November 2020

- 0.0083 jobs reviewed per hour per day in November 2020

```
Query:-
```

```
SELECT
COUNT(DISTINCT job_id)/(30 * 24.0) AS jobs_reviewed
FROM job_data
WHERE ds BETWEEN '2020-11-01' AND '2020-11-30';
```



1 (II) Throughput

- The weekly throughput is 0.03.
- On date <u>11/28/2020</u> the throughput is highest 0.06

Weekly Throughput

weekly_throughput

0.03

Daily Throughput

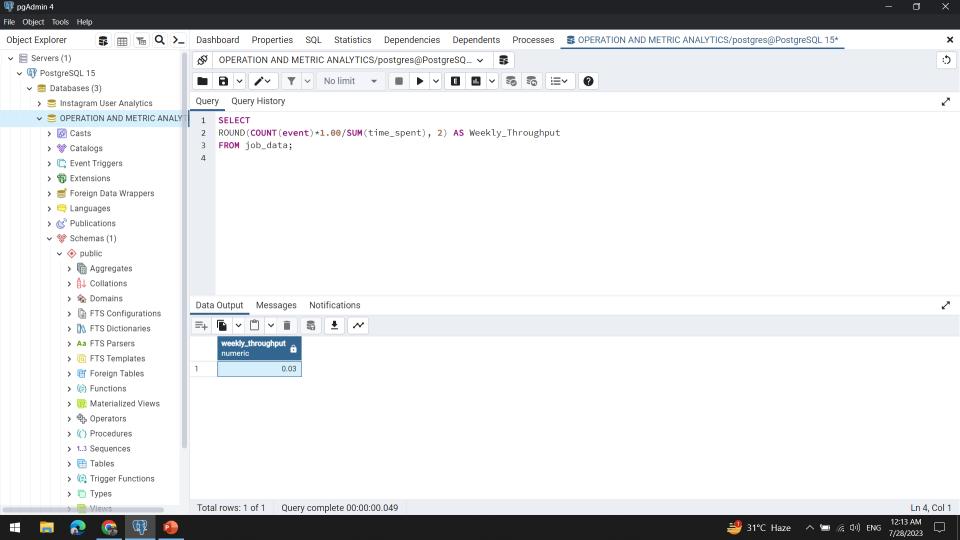
dates	daily_throughput
11/25/2020	0.02
11/26/2020	0.02
11/27/2020	0.01
11/28/2020	0.06
11/29/2020	0.05
11/30/2020	0.05

1 (II) Throughput

Query:-

Weekly Throughput

SELECT ROUND(COUNT(event)*1.00/SUM(time_spent), 2) AS Weekly_Throughput **FROM** job_data;

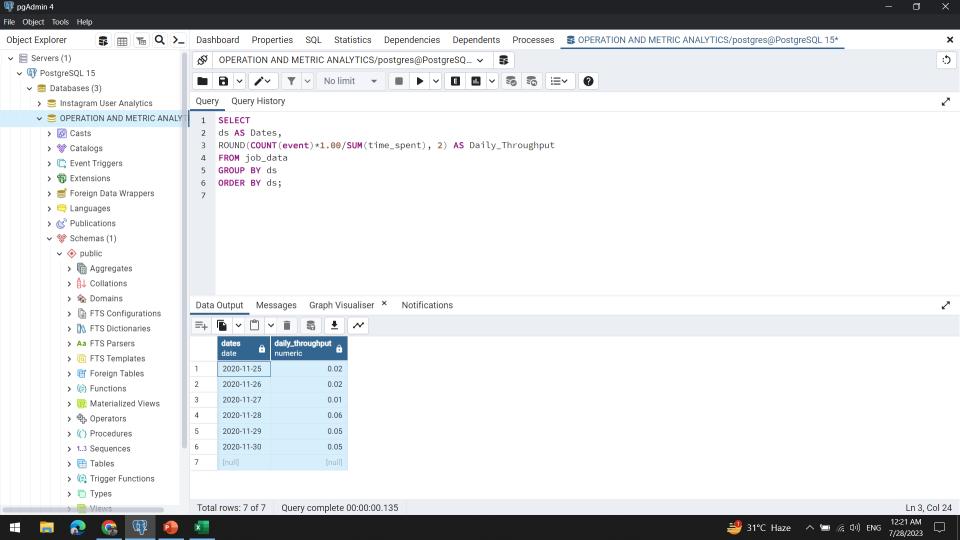


1 (II) Throughput

Query:-

Daily Throughput

```
SELECT
ds AS Dates,
ROUND(COUNT(event)*1.00/SUM(time_spent), 2) AS Daily_Throughput
FROM job_data
GROUP BY ds
ORDER BY ds;
```



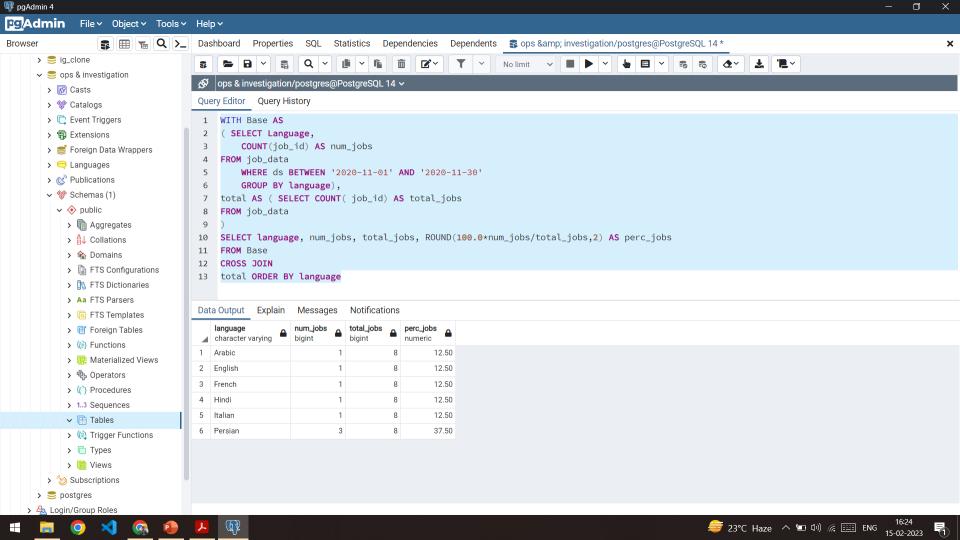
1 (III) Percentage share of each language

• find percentage share of each language in the last 30 days?

Language	num_jobs	total_jobs	perc_jobs		
Arabic	1	8	12.50		
English	1	8	12.50		
French	1	8	12.50		
Hindi	1	8	12.50		
Italian	1	8	12.50		
Persian	3	8	37.50		

1 (III) Percentage share of each language

```
Query:-
WITH Base AS
( SELECT Language,
          COUNT(job_id) AS num_jobs
FROM job_data
          WHERE ds BETWEEN '2020-11-01' AND '2020-11-30'
          GROUP BY language),
total AS ( SELECT COUNT( job_id) AS total_jobs
FROM job_data
SELECT language, num_jobs, total_jobs, ROUND(100.0*num_jobs/total_jobs,2) AS perc_jobs
FROM Base
CROSS JOIN
total ORDER BY language
```



1 (IV) Duplicate rows

Highlight the duplicate rows:-

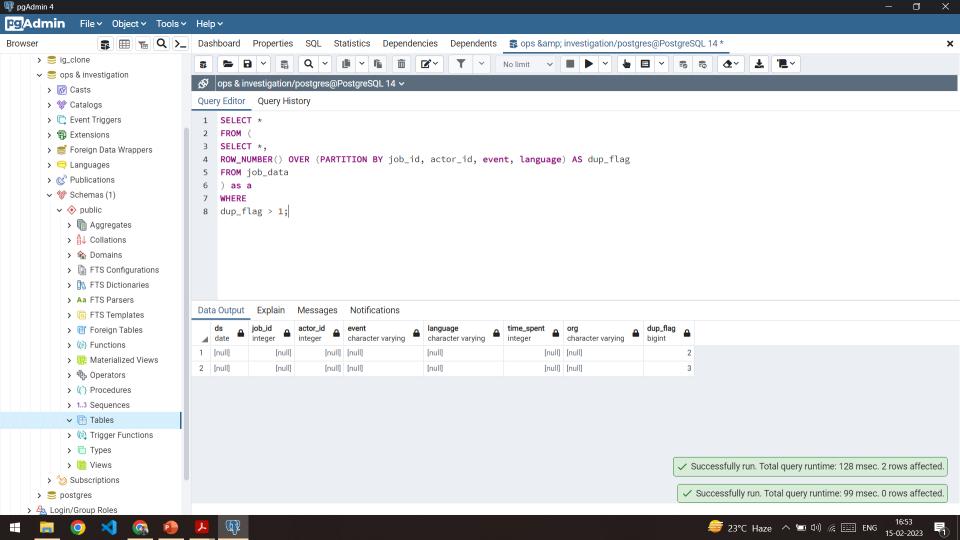
- Here we can highlight duplicate rows basis on their row number ('dup_flag' column) which is based on their occurrence repetition.
- All the rows have dup_flag higher than 1 are duplicate rows.

Note: Here all key column are consider for duplication entries There is no duplicate row to show present in dataset

1 (IV) Duplicate rows

```
Query:-

SELECT *
FROM (
SELECT *, ROW_NUMBER() OVER (PARTITION BY job_id, actor_id, event, language)
AS dup_flag
FROM job_data ) AS a
WHERE dup_flag > 1;
```



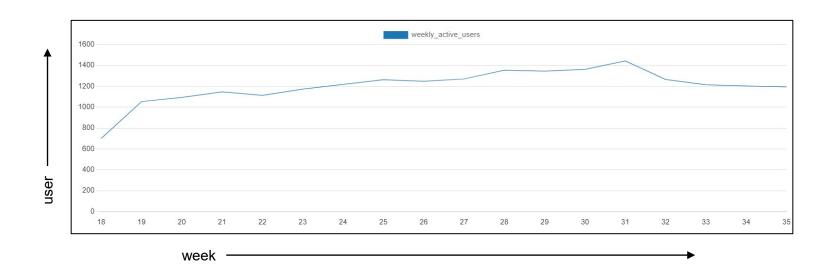
Insights

- i. This is very less amount of data to take insights from however,
- ii. We can conclude there is less quantity of work for each actor i.e getting 1 or 2 job per day
- iii. Dataset was clean and there is no single duplicate entry
- iv. We are getting equal quantity of work base on language and have stable work load for each day.



2 (I) User Engagement

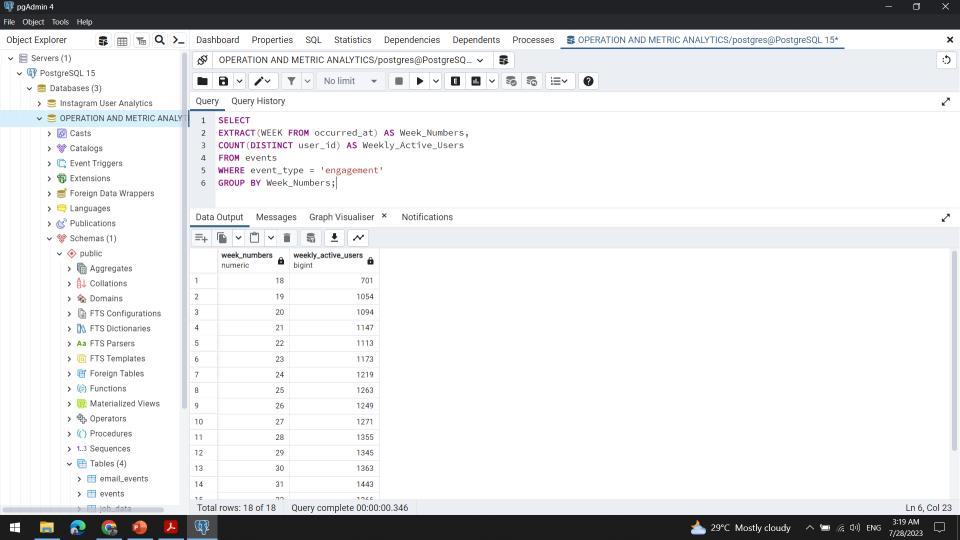
• User engagement basis on activeness of user on weekly basis



2 (I) User Engagement

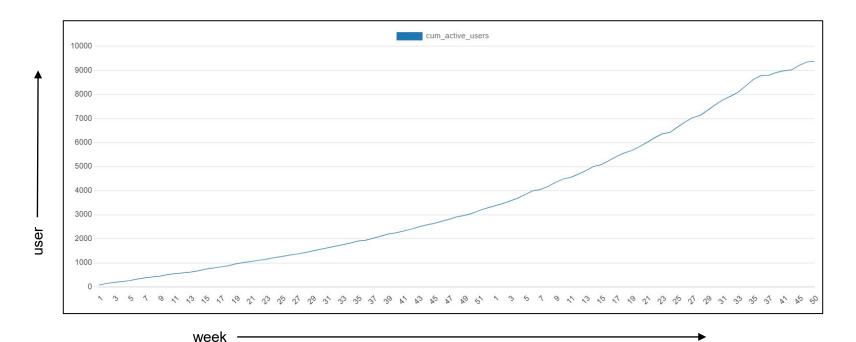
```
Query:-
```

```
SELECT
EXTRACT(WEEK FROM occurred_at) AS Week_Numbers,
COUNT(DISTINCT user_id) AS Weekly_Active_Users
FROM events
WHERE event_type = 'engagement'
GROUP BY Week_Numbers;
```



2 (II) User Growth

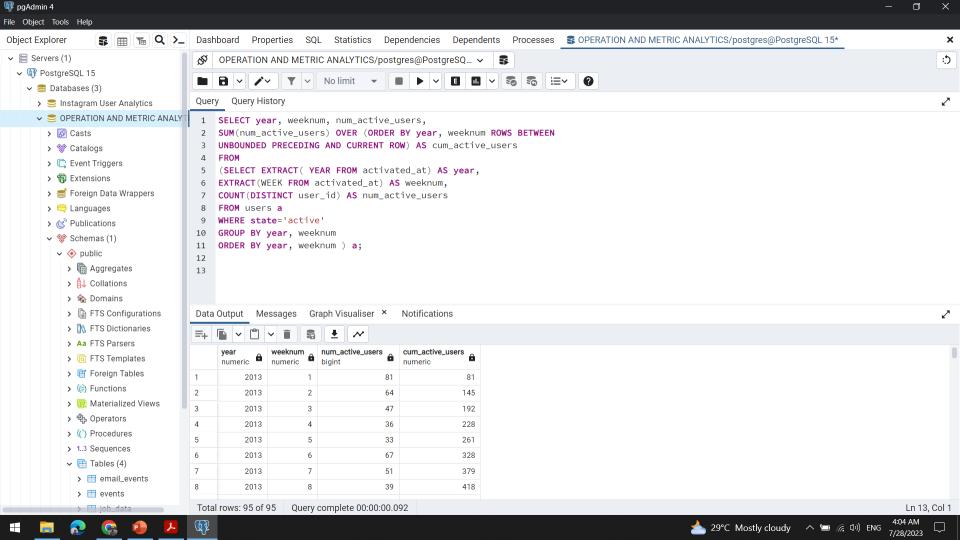
Growth of active user on weekly basis



2 (II) User Growth

Query:-

```
SELECT year, weeknum, num_active_users,
SUM(num_active_users) 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_users
FROM users a
WHERE state='active'
GROUP BY year, weeknum
ORDER BY year, weeknum ) a:
```



2 (III) Weekly Retention

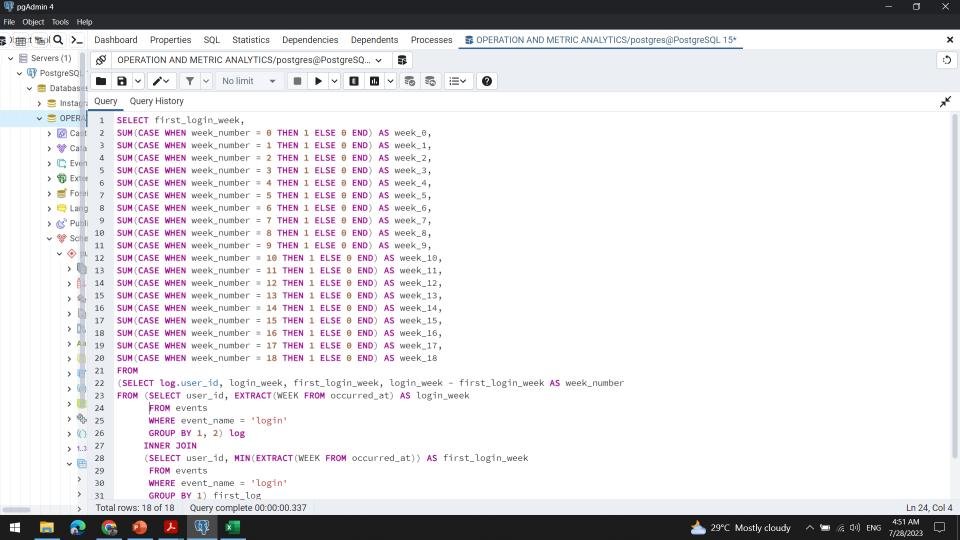
Query:-

```
SELECT first_login_week,
SUM(CASE WHEN week_number = 0 THEN 1 ELSE 0 END) AS week_0,
SUM(CASE WHEN week_number = 1 THEN 1 ELSE 0 END) AS week_1,
SUM(CASE WHEN week_number = 2 THEN 1 ELSE 0 END) AS week_2,
SUM(CASE WHEN week_number = 3 THEN 1 ELSE 0 END) AS week_3,
SUM(CASE WHEN week_number = 4 THEN 1 ELSE 0 END) AS week_4,
SUM(CASE WHEN week_number = 5 THEN 1 ELSE 0 END) AS week_5,
SUM(CASE WHEN week_number = 6 THEN 1 ELSE 0 END) AS week_6.
SUM(CASE WHEN week_number = 7 THEN 1 ELSE 0 END) AS week_7,
SUM(CASE WHEN week_number = 8 THEN 1 ELSE 0 END) AS week_8,
SUM(CASE WHEN week_number = 9 THEN 1 ELSE 0 END) AS week_9,
SUM(CASE WHEN week number = 10 THEN 1 ELSE 0 END) AS week 10.
SUM(CASE WHEN week_number = 11 THEN 1 ELSE 0 END) AS week_11,
SUM(CASE WHEN week_number = 12 THEN 1 ELSE 0 END) AS week_12,
SUM(CASE WHEN week_number = 13 THEN 1 ELSE 0 END) AS week_13,
SUM(CASE WHEN week_number = 14 THEN 1 ELSE 0 END) AS week_14,
SUM(CASE WHEN week number = 15 THEN 1 ELSE 0 END) AS week 15.
SUM(CASE WHEN week_number = 16 THEN 1 ELSE 0 END) AS week_16,
SUM(CASE WHEN week_number = 17 THEN 1 ELSE 0 END) AS week_17,
SUM(CASE WHEN week_number = 18 THEN 1 ELSE 0 END) AS week_18
```

2 (III) Weekly Retention

Query (Continue):-

first_login_	week	week_																	
week	_0	_1	_2	_3	_4	_5	_6	7	_8	_9	10	11	12	13	14	15	16	17	18
18	701	487	332	256	214	192	169	149	151	146	139	135	132	146	122	94	84	82	0
19	567	336	246	194	160	142	135	121	104	113	98	112	120	102	91	80	64	0	0
20	426	287	174	156	112	94	87	81	94	81	68	64	62	41	50	49	0	0	0
21	358	211	159	118	92	70	62	67	64	67	65	43	41	34	39	0	0	0	0
22	320	179	133	92	75	65	76	72	56	45	43	37	33	27	0	0	0	0	0
23	327	215	150	110	88	77	63	62	54	47	40	39	32	0	0	0	0	0	0
24	330	218	138	96	83	76	69	60	55	48	35	29	0	0	0	0	0	0	0
25	332	189	141	100	81	62	63	60	37	36	30	0	0	0	0	0	0	0	0
26	308	202	142	101	79	64	51	44	38	35	0	0	0	0	0	0	0	0	0
27	286	171	112	83	71	52	47	44	27	0	0	0	0	0	0	0	0	0	0
28	293	188	119	110	68	53	38	36	0	0	0	0	0	0	0	0	0	0	0
29	274	177	111	65	44	31	29	0	0	0	0	0	0	0	0	0	0	0	0
30	277	179	102	71	48	38	0	0	0	0	0	0	0	0	0	0	0	0	0
31	290	183	120	76	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	211	135	76	53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	272	177	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	284	187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	286	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

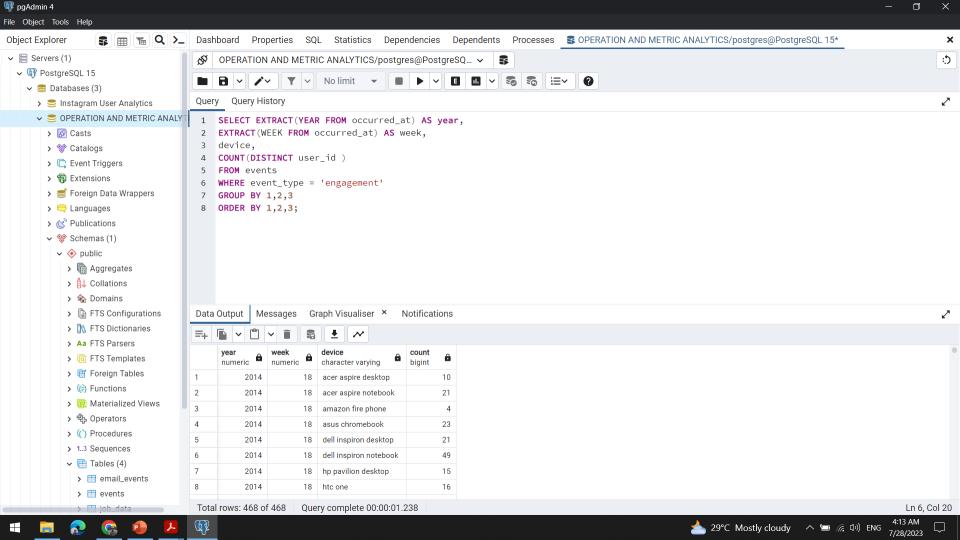


2 (IV) Weekly Engagement

calculation of weekly engagement per device

Query:-

```
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;
```



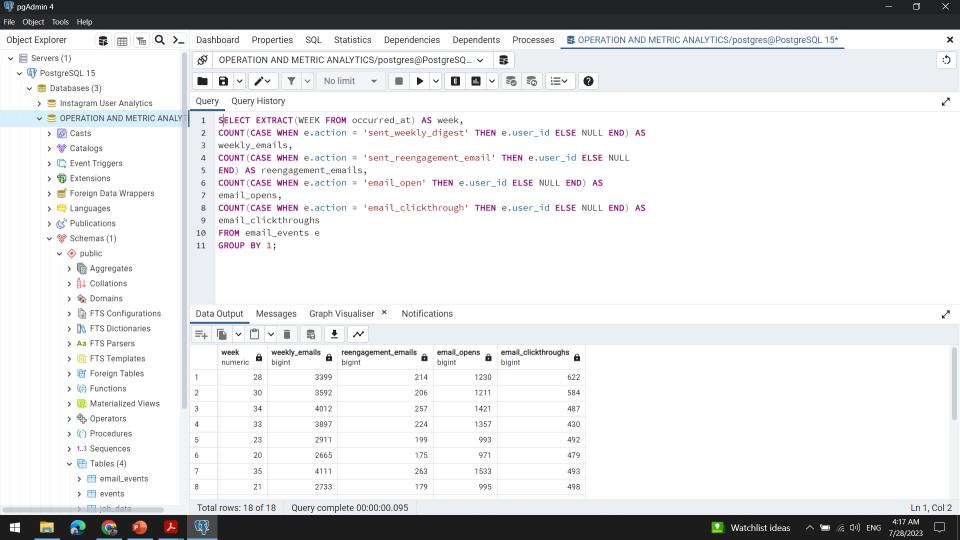
2 (V) Email Engagement

week	weekly_emails	reengagement_emails	email_opens	email_clickthroughs
28	3399	214	1230	622
30	3592	206	1211	584
34	4012	257	1421	487
33	3897	224	1357	430
23	2911	199	993	492
20	2665	175	971	479
35	4111	263	1533	493
21	2733	179	995	498
31	3706	230	1386	633
24	3003	190	1070	533
27	3302	222	1168	559
26	3207	187	1090	524
25	3105	234	1161	563
18	908	98	332	187
19	2602	164	919	434
22	2822	179	1026	453
32	3793	206	1336	432
29	3499	226	1260	607

2 (V) Email Engagement

Query:-

```
SELECT
EXTRACT(WEEK FROM occurred_at) AS week,
COUNT(CASE WHEN e.action = 'sent_weekly_digest' THEN e.user_id ELSE NULL END) AS weekly_emails,
COUNT(CASE WHEN e.action = 'sent_reengagement_email' THEN e.user_id ELSE NULL END) AS reengagement_emails,
COUNT(CASE WHEN e.action = 'email_open' THEN e.user_id ELSE NULL END) AS email_opens,
COUNT(CASE WHEN e.action = 'email_clickthrough' THEN e.user_id ELSE NULL END) AS email_clickthroughs
FROM email_events e
GROUP BY 1:
```



Insights

- I. From week 17 to 30 there is up down trend of user engagement with overall upward trend and after 30 there is downward trend in both activeness of users and activity done by user
- II. On the above metric we can say there is direct relation between activeness of users and activity done by user
- III. There is tremendous growth on after week 17 and then have stable up & down trend while activeness of users has up and down trend
- IV. Till the week 30 user signup have positive growth afterward there is downfall while there is retention rate is good
- V. most of users using our product with macbook or laptops, and there is less no. Of users with cell phone device using our product.
- VI. Out of 100 sent emails there is only 33 emails get open and out of 33 emails of 14-15 emails getting encourage people for clicks, we have to improve our email standard.

Results & Conclusions

- a) This project was very helpful to understand and strengthen key advance concept of sql
- b) Here i learn use of windows function, lead & lag function, diff date function, case function and application of it.
- c) It was helpful toward practicing of basic to intermediate sql quarries, and it was fun to write them, getting stuck and resolve error.

