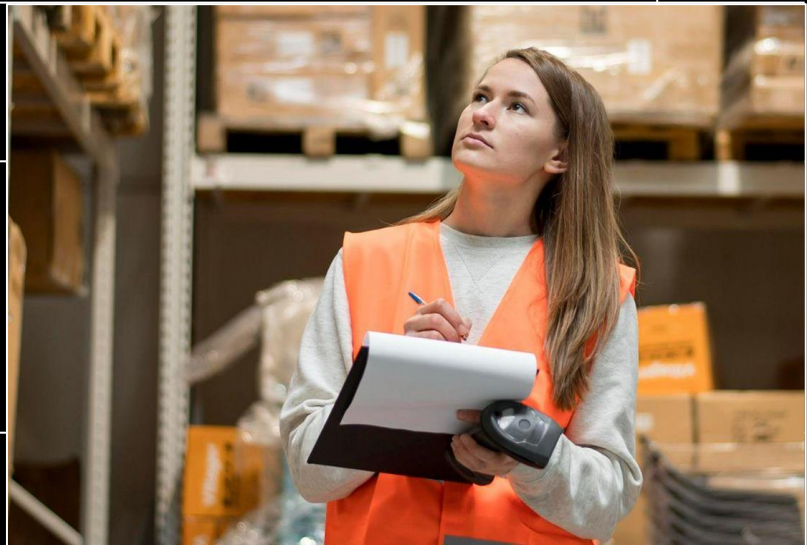
OPERATION & METRIC ANALYSIS

Operation Analytics and Investigating Metric Spike

Presented by- Ankita Yadav



Advanced SQL

Table of contents

01 PROJECT DESCRIPTION

This project is about OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKES. The project is based on ADVANCED SQL concepts. Operation Analytics is done for end to end operations of the company to to find the areas for improvement. It predicts the overall growth or decline of companies fortune. Investigating for metric spikes is very important part of this to answer various questions on daily bases for companies functions.

Being the Lead Data Analyst I was provided with different Data sets, tables from which I am supoosed to derive certain insights and answer the questions asked by different teams.

02

TECH STACK USED

MY SQL WORK BENCH

APPROACH

03

Case Study 1 – JOB DATA

- Job Data Table

	A	B	C	D	E	F	G	H
1	ds	job_id	actor_id	event	language	time_spent	org	
2	11/30/2020	21	1001	skip	English	15	A	
3	11/30/2020	22	1006	transfer	Arabic	25	B	
4	11/29/2020	23	1003	decision	Persian	20	C	
5	11/28/2020	23	1005	transfer	Persian	22	D	
6	11/28/2020	25	1002	decision	Hindi	11	B	
7	11/27/2020	11	1007	decision	French	104	D	
8	11/26/2020	23	1004	skip	Persian	56	A	
9	11/25/2020	20	1003	transfer	Italian	45	C	
10								

Case study 1

- In the given case study we have used the concept of advanced SQL like Windows Function to draw the insights for the given problems-

1. Number of Jobs Reviewed
2. Throughput
3. Percentage share of each problems
4. Duplicate Rows

Number of Jobs Reviewed

- **Your task:** Calculate the number of jobs reviewed per hour per day for November 2020?

c1p1 c1p3 c1p4 c1p2 SQL File 9* SQL Fi

Limit to 1000 rows

```sql
-- Number of jobs reviewd
SELECT
count(DISTINCT job_id)/(24*30) AS jb_cnt
FROM sql_table
WHERE ds BETWEEN '01-11-2020' AND '30-11-2020';
```

Result Grid Filter

jb_cnt
0.0083

Throughput

- **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?

c1p1 c1p3 c1p4 c1p2 SQL File 9* SQL File 10* SQL File 11* SQL File 12* SQL File

Limit to 1000 rows

```sql
-- Throughput

SELECT ds, jb_cnt,
(AVG(jb_cnt) OVER(ORDER BY ds ROWS BETWEEN 6 PRECEDING AND CURRENT ROW ) ) as Rolling_avg
From (
 SELECT ds,
 Count(Distinct job_id) AS jb_cnt
FROM sql_table
WHERE ds BETWEEN '2020-11-01' AND '2020-11-30'
GROUP BY ds)a;
```

Result Grid Filter Rows:

ds	jb_cnt	Rolling_avg
2020-11-25	1	1.0000
2020-11-26	1	1.0000
2020-11-27	1	1.0000
2020-11-28	2	1.2500
2020-11-29	1	1.2000
2020-11-30	2	1.3333

Percentage share of each problems

- **Your task:** Calculate the percentage share of each language in the last 30 days?

c1p1 | c1p3 | c1p4 | c1p2 | SQL File 9* | SQL File 10*

Limit to 1000 rows

```
-- Percentage Share of each Language

SELECT language, count(*) * 100.0/ sum(count(*)) over()
FROM sql_table
GROUP BY language;
```

Result Grid | Filter Rows:

language	count(*) * 100.0/ sum(count(*)) over()
English	12.50000
Arabic	12.50000
Persian	37.50000
Hindi	12.50000
French	12.50000
Italian	12.50000

Duplicate Rows

- **Your task:** Let’s say you see some duplicate rows in the data. How will you display duplicates from the table?

c1p1 c1p3 c1p4 c1p2 SQL File 9* SQL File 10*

Limit to 1000 rows

```sql
-- Duplicate Rows

SELECT ds,job_id,actor_id,event,language,time_spent,org,
COUNT(*) AS cnt
FROM sql_table
GROUP BY ds,job_id,actor_id,event,language,time_spent,org
HAVING COUNT(*) >1;
```

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

	ds	job_id	actor_id	event	language	time_spent	org	cnt

Case Study 2- INVESTIGATING METRIC SPIKES

- Events Table

	A	B	C	D	E	F	G
1	user_id	occurred_	event_type	event_name	location	device	user_type
2	10522	########	engagement	login	Japan	dell inspiron noteboc	3
3	10522	########	engagement	home_page	Japan	dell inspiron noteboc	3
4	10522	########	engagement	like_message	Japan	dell inspiron noteboc	3
5	10522	########	engagement	view_inbox	Japan	dell inspiron noteboc	3
6	10522	########	engagement	search_run	Japan	dell inspiron noteboc	3
7	10522	########	engagement	search_run	Japan	dell inspiron noteboc	3
8	10612	########	engagement	login	Netherlands	iphone 5	1
9	10612	########	engagement	like_message	Netherlands	iphone 5	1
10	10612	########	engagement	send_message	Netherlands	iphone 5	1
11	10612	########	engagement	home_page	Netherlands	iphone 5	1
12	10612	########	engagement	like_message	Netherlands	iphone 5	1
13	10612	########	engagement	home_page	Netherlands	iphone 5	1
14	10612	########	engagement	view_inbox	Netherlands	iphone 5	1
15	10612	########	engagement	like_message	Netherlands	iphone 5	1
16	10612	########	engagement	home_page	Netherlands	iphone 5	1
17	10612	########	engagement	send_message	Netherlands	iphone 5	1
18	10612	########	engagement	like_message	Netherlands	iphone 5	1
19	10612	########	engagement	send_message	Netherlands	iphone 5	1
20	10736	########	engagement	login	Austria	iphone 4s	2
21	10736	########	engagement	like_message	Austria	iphone 4s	2
22	10736	########	engagement	send_message	Austria	iphone 4s	2

Case Study 2- INVESTIGATING METRIC SPIKES

- User Table

	A	B	C	D	E	F	G
1	user_id	created_at	company_	language	activated_at	state	
2	0	1/1/2013 20:59	5737	english	1/1/2013 21:01	active	
3	1	1/1/2013 13:07	28	english		pending	
4	2	1/1/2013 10:59	51	english		pending	
5	3	1/1/2013 18:40	2800	german	1/1/2013 18:42	active	
6	4	1/1/2013 14:37	5110	indian	1/1/2013 14:39	active	
7	5	1/1/2013 13:39	2463	spanish		pending	
8	6	1/1/2013 18:37	11699	english	1/1/2013 18:38	active	
9	7	1/1/2013 16:19	4765	french	1/1/2013 16:20	active	
10	8	1/1/2013 4:38	2698	french	1/1/2013 4:40	active	
11	9	1/1/2013 8:04	1	french		pending	
12	10	1/1/2013 9:36	10	arabic		pending	
13	11	1/1/2013 8:07	3745	english	1/1/2013 8:09	active	
14	12	1/1/2013 18:05	903	english		pending	
15	13	1/2/2013 12:27	4025	english	1/2/2013 12:29	active	
16	14	1/2/2013 17:17	5077	portugese		pending	
17	15	1/2/2013 15:39	4259	english	1/2/2013 15:41	active	
18	16	1/2/2013 17:45	10408	german		pending	
19	17	1/2/2013 10:56	5025	japanese	1/2/2013 10:57	active	
20	18	1/2/2013 8:18	3	japanese		pending	

Case Study 2- INVESTIGATING METRIC SPIKES

- Email_event Table

	A	B	C	D	E
1	user_id	occurred_at	action	user_type	
2	0	5/6/2014 9:30	sent_weekly_digest	1	
3	0	5/13/2014 9:30	sent_weekly_digest	1	
4	0	5/20/2014 9:30	sent_weekly_digest	1	
5	0	5/27/2014 9:30	sent_weekly_digest	1	
6	0	6/3/2014 9:30	sent_weekly_digest	1	
7	0	6/3/2014 9:30	email_open	1	
8	0	6/10/2014 9:30	sent_weekly_digest	1	
9	0	6/10/2014 9:30	email_open	1	
10	0	6/17/2014 9:30	sent_weekly_digest	1	
11	0	6/17/2014 9:30	email_open	1	
12	0	6/24/2014 9:30	sent_weekly_digest	1	
13	0	7/1/2014 9:30	sent_weekly_digest	1	
14	0	7/8/2014 9:30	sent_weekly_digest	1	
15	0	7/15/2014 9:30	sent_weekly_digest	1	
16	0	7/22/2014 9:30	sent_weekly_digest	1	
17	0	7/29/2014 9:30	sent_weekly_digest	1	
18	0	7/29/2014 9:30	email_open	1	
19	0	8/5/2014 9:30	sent_weekly_digest	1	
20	0	8/12/2014 9:30	sent_weekly_digest	1	
21	0	8/19/2014 9:30	sent_weekly_digest	1	

Case study 2

- In the given case study we have used the concept of advanced SQL like Windows Function, Date and Time concept to draw the insights for the given problems-

1) User Engagement
2) User Growth
3) Weekly Retention
4) Weekly Engagement
5) Email Engagement

User Engagement

- **Your task:** Calculate the weekly user engagement?

c1p1 c1p3 c1p4 c1p2 c2p1 c2p2 c2p3*

Limit to 50000 rows

```
-- user engagement rate
SELECT
event_type, event_name,
EXTRACT(month from occurred_at)as week,
COUNT(DISTINCT user_id) AS weekly_active_users
FROM events
GROUP BY event_name, event_type, week ;
```

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

event_type	event_name	week	weekly_active_users
signup_flow	complete_signup	5	779
signup_flow	complete_signup	6	873
signup_flow	complete_signup	7	272
engagement	home_page	5	729
engagement	home_page	6	808
engagement	home_page	7	282
engagement	home_page	8	10
engagement	like_message	5	590
engagement	like_message	6	660
engagement	like_message	7	228
engagement	like_message	8	7
engagement	login	5	824

Result 1

User Growth

- **Your task:** Calculate the user growth for product?

c1p1 c1p3 c1p4 c1p2 c2p1 c2p2* c2p3* c2p4 c2p5

Limit to 50000 rows

```sql
-- user growth
SELECT
EXTRACT(day FROM created_at) AS day,
COUNT(*) AS all_users,
COUNT(CASE WHEN activated_at IS NOT NULL THEN user_id ELSE NULL END) AS activated_users
FROM users
GROUP BY 1
ORDER BY 1;
```

Result Grid Filter Rows:

day	all_users	activated_users
1	250	250
2	273	273
3	246	246
4	261	261
5	241	241
6	266	266
7	262	262
8	270	270
9	246	246
10	252	252
11	260	260

Result 1

Weekly Retention

- **Your task:** Calculate the weekly retention of users-sign up cohort?

c1p1 c1p3 c1p4 c1p2 c2p1 c2p2 c2p3* c2p4 c2p5

Limit to 50000 rows

```sql
-- weekly retension
with signup_cohort AS
(SELECT company_id, date(created_at) AS signup_week
FROM users)
SELECT signup_week,
COUNT(DISTINCT signup_cohort.company_id) AS total_signups,
COUNT(DISTINCT events.user_id) as retained_users,
COUNT(DISTINCT events.user_id)/COUNT(DISTINCT signup_cohort.company_id) as retension_rate
FROM signup_cohort
left JOIN events  on events.user_id=signup_cohort.company_id
GROUP BY 1
ORDER BY 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

signup_week	total_signups	retained_users	retension_rate
2013-01-01	13	0	0.0000
2013-01-02	11	0	0.0000
2013-01-03	13	1	0.0769
2013-01-04	11	0	0.0000
2013-01-05	3	1	0.3333
2013-01-06	4	1	0.2500
2013-01-07	13	0	0.0000
2013-01-08	12	0	0.0000
2013-01-09	10	1	0.1000
2013-01-10	11	0	0.0000

Result 10

Weekly Engagement

- **Your task:** Calculate the weekly engagement per device?

c1p1 c1p3 c1p4 c1p2 c2p1 c2p2 c2p3*

Limit to 50000 rows

```sql
-- weekly engagement
With weekly_engagement as
(SELECT  user_id, device,
EXTRACT(week from occurred_at)as week,
COUNT(*) as engagement
FROM events
GROUP BY user_id, device, week
ORDER BY user_id, device, week)

SELECT device, week, SUM(engagement) as weekly_engagement
FROM weekly_engagement
GROUP BY device, week
```

Result Grid | Filter Rows: | Export: | Wrap Cel

device	week	weekly_engagement
acer aspire desktop	17	11
acer aspire desktop	18	25
acer aspire desktop	20	8
acer aspire desktop	21	23
acer aspire desktop	23	54
acer aspire desktop	24	50
acer aspire desktop	25	13
acer aspire desktop	26	28
acer aspire desktop	27	32
acer aspire desktop	33	5

Result 2

Email Engagement

- **Your task:** Calculate the email engagement metrics?

c1p1 c1p3 c1p4 c1p2 c2p1 c2p2 c2p3* c2p4 c2p5

Limit to 50000 rows

```sql
-- email engagement
SELECT
count(distinct user_id) as user,
Date_format("occurred_at", "%U") AS week,
COUNT(CASE WHEN action = 'sent_weekly_digest' THEN user_id ELSE NULL END) AS weekly_emails,
count(CASE WHEN action = 'sent_reengagement_email' THEN user_id ELSE NULL END) AS reengagement_emails,
count(CASE WHEN action = 'email_open' THEN user_id ELSE NULL END) AS email_opens,
count(CASE WHEN action = 'email_clickthrough' THEN user_id ELSE NULL END) AS email_clickthroughs
FROM email_events
GROUP BY action ;
```

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

user	week	weekly_emails	reengagement_emails	email_opens	email_clickthroughs
5277	NULL	0	0	0	9010
5927	NULL	0	0	20459	0
3653	NULL	0	3653	0	0
4111	NULL	57267	0	0	0

04

INSIGHTS

Case Study 1

Number of jobs reviewed -

Job count of 0.0083 was obtained.

Throughput -

7-day rolling was prefered to calculate the rolling average.

Percentage share of each language -

Persian has the highest percentage share of all the languages.

Duplicate Rows -

None row having all similar values as any other was obtained.

Case Study 2

User Engagement

Number of users that were active on week basis were calculated.

User Growth

Amount of users growing over time were measured.

Weekly Retention

The retention rate was measured using total signup and retained users.

Weekly Engagement

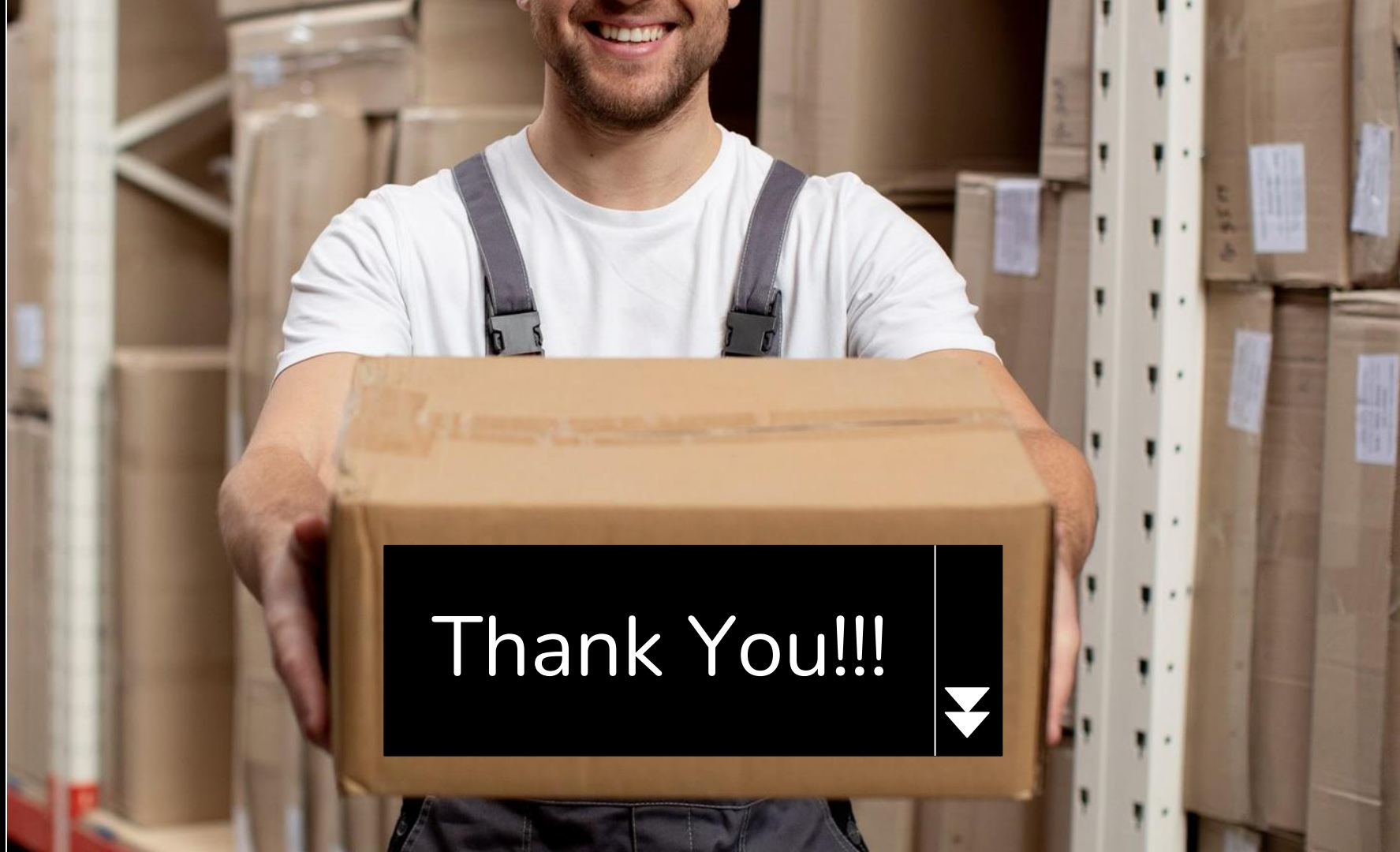
Engagement of users based on different devices was found.

Email Engagement

Users engaging in different email services were calculated.

RESULTS

While working for the project I gained quite a knowledge about ADVANCED SQL concepts. I learnt to apply windows function and applications of various Date and Time functions and gained a practical insight of application of advanced sql.


Thank You!!!