Operation Analytics and Investigating Metric Spike

Shristy Pandey

Project Description

- Operation Analytics is the analysis done for the complete end to end operations of a company.
- By analyzing the Data company then finds the areas on which it must improve upon.
- Need to help different departments like- Ops team, support team, marketing team, etc help them derive insights out of the data they collect.
- This kind of analysis is further used to predict the overall growth or decline of a company's fortune.
- As a data analyst I need to understand the requirement of data and also able to give the answer to the other departments for their better understanding.

Approach

- Firstly I have go through the description to understand the data which is given there and see the requirements.
- Then I take a raw data which Is provided in the form of tables.
- By using MYSQL workbench I imported the file in new data base and I read the question carefully and started to write the query to get a result in the table form.
- I have executed the queries and If any error is occurs then I have try to modify it and run again. And get a result.
- After completed all the question which is asked I have cross check the query. Than I take the screenshot to attached them on the ppt as result.

Tech-Stack Used

Insight and Result

Case Study 1 (Job Data)

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

Query 1 job_data job_data

Limit to 1000 rows

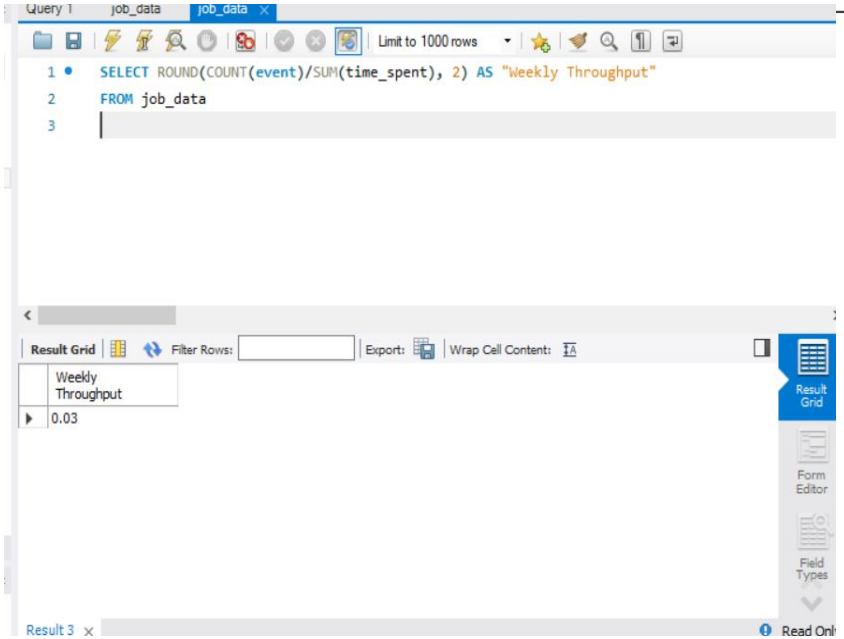
```
SELECT ds AS DATES, ROUND((COUNT(job_id)/SUM(time_spent))*3600)AS "Jobs Reviewed per Hours
FROM job_data
WHERE ds BETWEEN '2020-11-01' AND '2020-11-30'
GROUP BY ds;
```

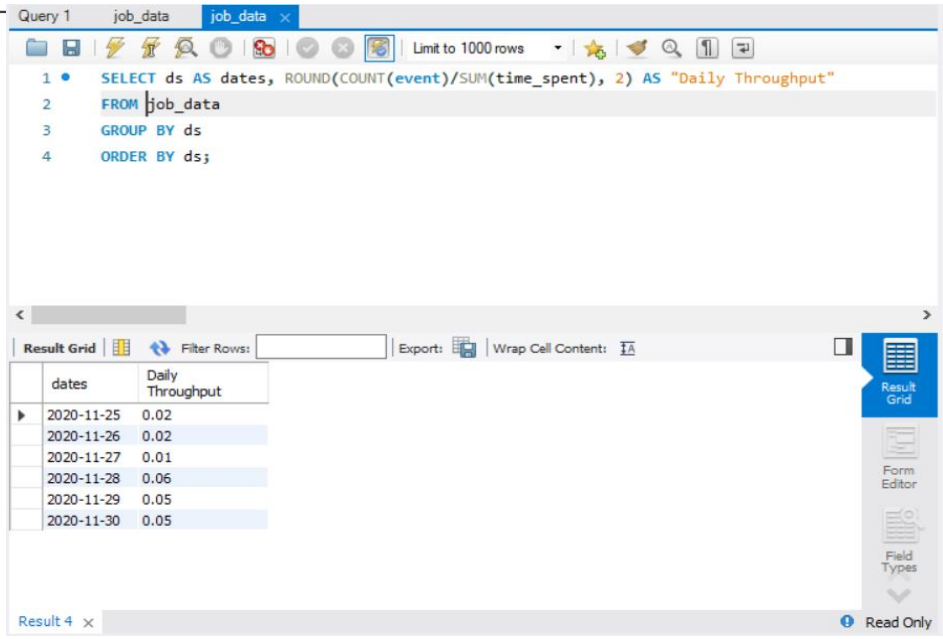
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

DATES	Jobs Reviewed per Hours per day
2020-11-30	180
2020-11-29	180
2020-11-28	218
2020-11-27	35
2020-11-26	64
2020-11-25	80

Result Grid
Form Editor
Field Types

2. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?





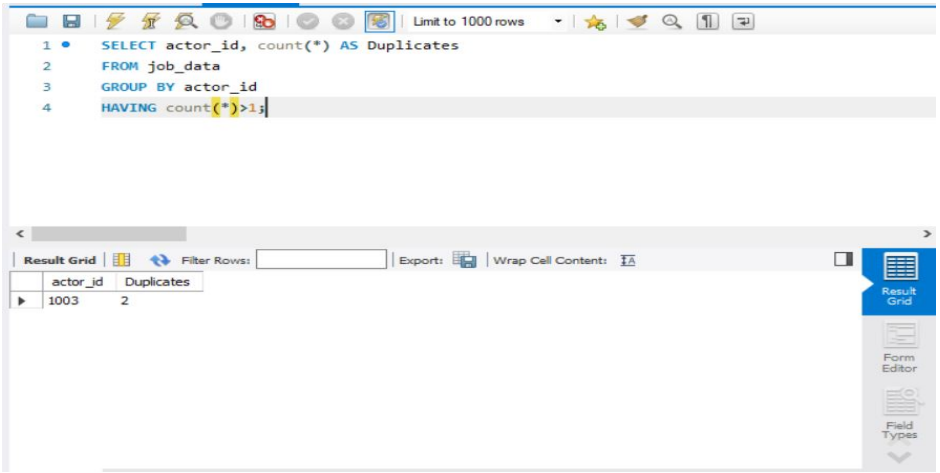
3. Calculate the percentage share of each language in the last 30 days?

```
SELECT language AS Languages, ROUND(100*COUNT(*)/total, 2) AS Percentage
FROM job_data
CROSS JOIN (SELECT COUNT(*) AS total FROM job_data) sub
GROUP BY language;
```

Languages	Percentage
English	12.50
Arabic	12.50
Persian	37.50
Hindi	12.50
French	12.50
Italian	12.50

4. Let's say you see some duplicate rows in the d

How will you display duplicates from the table?



Case Study 2 (Investigating metric spike)

1. Calculate the weekly user engagement?

```
SELECT EXTRACT( week from e.occurred_at) AS weeks,
COUNT(DISTINCT e.user_id) AS weekly_active_users
FROM events e
WHERE e.event_type = 'engagement'
AND e.event_name = 'login'
GROUP BY 1
ORDER BY 1 ;
```

week_num	num_users
NULL	6142

2. CALCULATE THE USER GROWTH FOR PRODUCT?

```
SELECT EXTRACT( WEEK FROM U.CREATED_AT) AS
DAY, COUNT(*) AS ALL_USERS,
COUNT(CASE WHEN U.ACTIVATED_AT IS NOT NULL
THEN U.USER_ID ELSE NULL END) AS
ACTIVATED_USERS
FROM USERS U
WHERE U.CREATED_AT >= '2013-01-01' AND U.
CREATED_AT < '2014-09-01'
GROUP BY 1
ORDER BY 1;
```

Months	Users	Growth in %
1	712	NULL
2	685	-3.79
3	765	11.68
4	907	18.56
5	993	9.48
6	1086	9.37
7	1281	17.96
8	1347	5.15
9	330	-75.50
10	390	18.18
11	399	2.31
12	486	21.80

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

```sql
SELECT EXTRACT(week FROM z.occurred_at) AS "week", AVG(z.age_at_event) AS "Average age du
COUNT(DISTINCT CASE WHEN z.user_age > 70 THEN z.user_id ELSE NULL END) AS "10+ weeks",
COUNT(DISTINCT CASE WHEN z.user_age < 70 AND z.user_age >=63 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 63 AND z.user_age >=56 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 56 AND z.user_age >=49 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 49 AND z.user_age >=42 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 42 AND z.user_age >=35 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 35 AND z.user_age >=28 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 28 AND z.user_age >=21 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 21 AND z.user_age >=14 THEN z.user_id ELSE NULL END
COUNT(DISTINCT CASE WHEN z.user_age < 14 AND z.user_age >=7 THEN z.user_id ELSE NULL END)
COUNT(DISTINCT CASE WHEN z.user_age < 7 AND z.user_age >=0 THEN z.user_id ELSE NULL END)
FROM (SELECT e.occurred_at, u.user_id, EXTRACT(week from u.activated_at) AS activation_we
EXTRACT(DAY FROM e.occurred_at - u.activated_at) AS age_at_event,
DATEDIFF('2014-09-01',u.activated_at) AS user_age
FROM users u JOIN events e
ON e.user_id = u.user_id AND e.event_type = 'engagement'
AND e.event_name= 'login' AND e.occurred_at >= '2014-05-01' AND e.occurred_at < '2014-09-
WHERE u.activated_at IS NOT NULL ) z
GROUP BY 1
ORDER BY 1
```

year	week	device	count(distinct user_id)
NULL	NULL	acer aspire desktop	198
NULL	NULL	acer aspire notebook	338
NULL	NULL	amazon fire phone	89
NULL	NULL	asus chromebook	355
NULL	NULL	dell inspiron desktop	360
NULL	NULL	dell inspiron notebook	677
NULL	NULL	hp pavilion desktop	339
NULL	NULL	htc one	196
NULL	NULL	ipad air	478
NULL	NULL	ipad mini	292
NULL	NULL	iphone 4s	409
NULL	NULL	iphone 5	1025
NULL	NULL	iphone 5s	626
NULL	NULL	kindle fire	205
NULL	NULL	lenovo thinkpad	1309
NULL	NULL	mac mini	150
NULL	NULL	macbook air	950
NULL	NULL	macbook pro	1952
NULL	NULL	nexus 10	273
NULL	NULL	nexus 5	621
NULL	NULL	nexus 7	355
NULL	NULL	nokia lumia 635	211
NULL	NULL	samsumg galaxy tablet	107
NULL	NULL	samsung galaxy note	119
NULL	NULL	samsung galaxy s4	803
NULL	NULL	windows surface	182

4. Calculate the weekly engagement per device?

```sql
SELECT EXTRACT(week FROM occurred_at) AS week,
COUNT(DISTINCT e.user_id) AS weekly_active_users,
COUNT(DISTINCT CASE WHEN e.device
IN('macbook pro','lenovo thinkpad','macbook air','dell inspiron notebook','asus chromebook'
THEN e.user_id ELSE NULL END) AS computer,
COUNT(DISTINCT CASE WHEN e.device
IN('iphone 5','samsung galaxy s4','nexus 5','iphone 5s','iphone 4s','nokia lumia 635','htc
THEN e.user_id ELSE NULL END) AS phone,
COUNT(DISTINCT CASE WHEN e.device
IN('ipad air','nexus 7','ipad mini','nexus 10','kindle fire','windows surface','samsung gal
THEN e.user_id ELSE NULL END) AS  tablet
FROM events e
WHERE e.event_type = 'engagement' AND e.event_name = 'login'
GROUP BY 1
ORDER BY 1
LIMIT 100;
```

year	week_num	num_active_user	cum_active_users
NULL	NULL	9381	9381

5. Calculate the email engagement metrics?

```
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_ema
COUNT(CASE WHEN e.action = 'sent_reengagement_email' THEN e.user_id ELSE NULL END) AS reeng
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_cli
FROM email_events e
GROUP BY 1;
```

Week	Weekly Digest Rate	Email Open Rate	Email Clickthrough Rate	Reengagement Email Rate
17	62.32	21.28	11.39	5.01
18	63.45	22.24	10.49	3.83
19	62.16	22.67	11.13	4.04
20	61.62	22.64	11.43	4.31
21	63.52	22.82	9.97	3.69
22	63.59	21.56	10.66	4.19
23	62.39	22.34	11.18	4.09
24	61.61	22.92	10.99	4.48
25	63.77	21.79	10.54	3.90
26	62.99	22.22	10.61	4.18
27	62.24	22.49	11.37	3.90
28	62.92	22.48	10.77	3.83
29	63.98	21.71	10.51	3.79
30	62.29	23.24	10.59	3.88
31	65.27	23.25	7.66	3.82
32	66.59	22.85	7.14	3.42
33	64.73	23.10	7.91	4.26
34	64.33	23.91	7.67	4.08
35	0.00	32.28	29.92	37.80

Result

How this project helped me- This project make me to understand the importance of operational analytics. By doing this project I am able to understand how the companies use this secret weapon.

Challenges that I faced in this project- The main challenge of this project is when I started the case study 2 than I realise that the data is huge in amount. So I have use different queries to insert the data into the table. But after done some modifications I was able to insert the data into the table.

Conclusion- Operational analysis can achieve a significant positive effect on our general public and world everywhere and increment the general efficiency of specific areas.