

Operation Analytics and Investigating Metric Spike

Project Description:

Opera on Analytics is a comprehensive analysis of a company's end-to-end operations. Its purpose is to identify areas for improvement and enable be er cross-functi00onal collaboration and more effective workflows. As a Data Analyst Lead at Microsoft, I work with various data sets and tables to derive valuable insights that can answer the questions posed by different departments.

One critical aspect of operational analytics is investigating metric spikes. This involves understanding why there is a sudden dip in daily engagement, sales, or other metrics. Addressing these questions is crucial for effective decision-making and ensuring the company's continued success.

Given the importance of operational analytics, I leverage my expertise to provide the necessary insights to improve the company's performance.

By analysing the data provided to me, I work collaboratively with cross-functional teams to develop actionable strategies that drive business growth and improve customer satisfaction.

Approach:

To begin my analysis, I dedicated me to thoroughly understanding the data and tables provided. I needed to comprehend the meaning of the various fields such as job_id, actor_id, and event, and determine what factors needed to be taken into account when reviewing the data.

Leveraging my SQL skills, I was able to extract valuable insights from the dataset provided by the management team. I first created a database named "operation_analytics" and then designed the tables using the structure and links provided by the team. This enabled me to organize the data and make it more accessible for analysis.

Once the tables were set up, I conducted a thorough analysis to generate valuable insights for the company. By applying various SQL queries, I was able to derive meaningful information that would help inform decision-making across the organization. The insights I uncovered would provide the company with the knowledge needed to improve its operations and drive business growth.

Once the tables were set up, I conducted a thorough analysis to generate valuable insights for the company. By applying various SQL queries, I was able to derive meaningful information that would help inform decision-making across the organization. The insights I uncovered would provide the company with the knowledge needed to improve its operations and drive business growth.

Tech Stack:

MySQL 8.0 is used to harness the expected results/insights as per requirements given in the project description.

Insights

Case Study 1 (Job Data): -

A. Number of jobs reviewed: - Number of jobs reviewed over me.

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

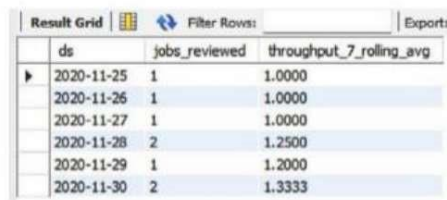
```
SELECT COUNT
(DISTINCT job_id) / 720
AS num_jobs_reviewed
FROM job_data1
WHERE ds LIKE '2020-11%';
```

Result Grid		Filter Row
	num_jobs_reviewed	
	0.0083	

The number of jobs reviewed per hour per day for November 2020 are 0.008

B. Throughput: - It is the no. of events happening per second.

```
SELECT ds, COUNT
(DISTINCT job_id)
AS jobs_reviewed,
AVG (COUNT (DISTINCT job_id))
OVER (ORDER BY ds ROWS BETWEEN 6 PRECEDING AND CURRENT ROW)
AS throughput_7_rolling_avg
FROM job_data1
WHERE ds BETWEEN '2020-11-01' AND '2020-11-30'
GROUP BY ds ORDER BY ds;
```



The screenshot shows a database query result grid with three columns: 'ds', 'jobs_reviewed', and 'throughput_7_rolling_avg'. The data is as follows:

ds	jobs_reviewed	throughput_7_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

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?

The 7_day_rolling average of throughput is about 1.0000 -1.3333, we have used 7_day_rolling due to its simplicity.

C. 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?

```

SELECT language, COUNT
(DISTINCT job_id)
AS num_jobs, 100.0 * COUNT
(DISTINCT job_id) / (SELECT COUNT
(DISTINCT job_id)
FROM job_data1)
AS pct_share_jobs
FROM job_data1 GROUP BY language;

```

	language	num_jobs	pct_share_jobs
▶	Arabic	1	16.66667
	English	1	16.66667
	French	1	16.66667
	Hindi	1	16.66667
	Italian	1	16.66667
	Persian	1	16.66667

The percentage share of each language in the last 30 days is 16.66667.

D. 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?

```

SELECT * FROM ( SELECT *, ROW_NUMBER()
OVER (PARTITION BY job_id)
AS rownum FROM job_data1 ) a
WHERE rownum > 1;

```

	ds	job_id	actor_id	event	language	time_spent	org	rownum
▶	2020-11-28	23	1005	transfer	Persian	22	D	2
	2020-11-26	23	1004	skip	Persian	56	A	3

Above is the method to display the duplicate rows.

Case Study 2 (Investigating metric spike)

A. 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?

```

SELECT EXTRACT(WEEK FROM occurred_at)
AS week_number,
COUNT(DISTINCT user_id)
AS distinct_users
FROM events GROUP BY week_number;

```

week_number	distinct_users
17	85
18	194
19	208
20	195
21	208
22	230
23	224
24	252
25	245
26	230
27	249
28	240
29	253
30	74
31	25
32	5
33	5
34	1

- B. User Growth: - Number of users growing over me for a product.
 Your task: Calculate the user growth for product.

```

SELECT year, num_week, num_active_users,
SUM(num_active_users)
OVER(ORDER BY year, num_week
ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)
AS cumm_active_users
FROM (SELECT EXTRACT(year FROM activated_at)
AS year, EXTRACT(week FROM activated_at)
AS num_week, COUNT(DISTINCT user_id)
AS num_active_users
FROM users
WHERE state='active')
GROUP BY year, num_week
ORDER BY year, num_week ) a;

```

	year	num_week	num_active_users	cumm_active_users
▶	2013	0	23	23
	2013	1	30	53
	2013	2	48	101
	2013	3	36	137
	2013	4	30	167
	2013	5	48	215
	2013	6	38	253
	2013	7	42	295
	2013	8	34	329
	2013	9	43	372
	2013	10	32	404
	2013	11	31	435
	2013	12	33	468
	2013	13	39	507
	2013	14	35	542
	2013	15	43	585
	2013	16	46	631
	2013	17	49	680
	2013	18	44	724
	2013	19	57	781

C. Weekly Retention: - Users getting retained weekly after signing- up for a product.

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

```

SELECT COUNT(user_id),
SUM(CASE WHEN retention_week = 1 THEN 1 ELSE 0 END)
AS per_week_retention
FROM (SELECT a.user_id, a.sign_up_week, b.engagement_week, b.engagement_week - a.sign_up_week
AS retention_week
FROM (SELECT DISTINCT user_id, EXTRACT(week FROM occurred_at)
AS sign_up_week
FROM events
WHERE event_type = 'signup_flow'
AND event_name = 'complete_signup'
AND EXTRACT(week FROM occurred_at) = 18 ) a
LEFT JOIN ( SELECT DISTINCT user_id, EXTRACT(week FROM occurred_at)
AS engagement_week
FROM events
WHERE event_type = 'engagement') b
ON a.user_id = b.user_id ) sub
GROUP BY user_id ORDER BY user_id;

```

	COUNT(user_id)	per_week_retention
1	0	
1	0	
1	0	
1	0	
1	0	
1	0	
1	0	
1	0	
1	0	
1	0	
3	1	
1	0	
1	0	
1	0	

Above are the weekly retention of users-sign up cohort.

D. 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?

```

SELECT EXTRACT(YEAR FROM occurred_at)
AS year_num, EXTRACT(WEEK FROM occurred_at)
AS week_num, device, COUNT(DISTINCT user_id)
AS no_of_users
FROM events WHERE event_type = 'engagement'
GROUP BY 1, 2, 3
ORDER BY 1, 2, 3

```

Result Grid		 Filter Rows	 Export	
	year_num	week_num	device	no_of_users
▶	2014	17	acer aspire desktop	2
	2014	17	acer aspire notebook	2
	2014	17	amazon fire phone	1
	2014	17	asus chromebook	3
	2014	17	dell inspiron desktop	1
	2014	17	dell inspiron notebook	4
	2014	17	hp pavilion desktop	2
	2014	17	htc one	2
	2014	17	ipad air	1
	2014	17	ipad mini	3
	2014	17	iphone 4s	3
	2014	17	iphone 5	11
	2014	17	iphone 5s	5
	2014	17	lenovo thinkpad	8
	2014	17	mac mini	1
	2014	17	macbook air	4
	2014	17	macbook pro	13
	2014	17	nexus 5	4
	2014	17	nexus 7	4
	2014	17	nokia lumia 635	2
	2014	17	samsung galaxy tablet	2
	2014	17	samsung galaxy note	1

Above are some results displayed showing expected weekly engagement per user.

E. Email Engagement: - Users engaging with the email service.

Your task: Calculate the email engagement metrics?

```

SELECT 100.0 * SUM(CASE WHEN action = 'email_open' THEN 1 ELSE 0 END) /
SUM(CASE WHEN action IN ('sent_weekly_digest', 'sent_reengagement_email') THEN 1 ELSE 0 END)
AS email_opening_rate, 100.0 * SUM(CASE WHEN action = 'email_clickthrough' THEN 1 ELSE 0 END) /
SUM(CASE WHEN action IN ('sent_weekly_digest', 'sent_reengagement_email') THEN 1 ELSE 0 END)
AS email_clicking_rate
FROM email_events;

```

Result Grid	Filter Rows
email_opening_rate	email_clicking_rate
33.58339	14.78989

The email engagement metrics is shown above in the row.

Result:

How this project helped me: - This project helps me to understand the importance of operation analytics. Through this project I am able to understand how the companies use metric spike as a secret weapon.

With an informed and proactive approach, they can leverage insights to make data-backed decisions that optimize their strategy and boost ROI.

Challenges that I faced in this project: - The challenge here is that the data in case study 2 is very huge, as the huge amount of data SQL

Workbench is very slow to import it. To tackle this situation, I have to use LOAD DATA statements. Now, there is another problem arises in the column user type in events table that has datatype int which is stopping the process of importing. First, I need to change its datatype to text then restart the process of loading the data into events table.

Conclusion: - Operational Analytics tackles the problem by

synchronizing real- me data. Operational Analytics has the capability to aggregate data from multiple data sources into a cumulative, organized, actionable solution capable of delivering analytical models in real- me to create individual customer profiles and a holistic view of operations for a company. This guarantees that your operational routines and systems are used efficiently. Whenever utilized correctly, operational analytics can achieve a significant positive effect on our general public and world everywhere and increment the general efficiency of specific areas.