Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Description

As a Data Analyst Lead at Microsoft, I focus on Operation Analytics, analyzing end-to-ecompany operations to identify areas for improvement. Working closely with various teams like operations, support, and marketing, I derive insights from collected data to enhance automation, team collaboration, and workflows. This analysis is crucial for predicting overall company growth or decline. Investigating metric spikes is a key aspect, addressing questions like daily engagement dips or sales declines. Using diverse data sets and tables, I provide insights and answers to questions posed by different departments, contributing to the company's success.

Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. As a Data Analyst, you'll work closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.

One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales. As a Data Analyst, you'll need to answer these questions daily, making it crucial to understand how to investigate these metric spikes.

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Problem

Case Study 1 (Job Data)

☐ 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?

☐ 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?

☐ 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?

☐ 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?

.

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Problem

Case Study 2 (Investigating metric spike)

➢ 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?

➢ User Growth: Amount of users growing over time for a product.

➢ Your task: Calculate the user growth for product?

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

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

➢ 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?

➢ Email Engagement: Users engaging with the email service.

➢ Your task: Calculate the email engagement metrics?

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Design

Steps taken to load the data into the data base.

- Using the 'create db' function of MySQL create a data base.
- Then add tables and column names.
- Then add the values into them using the 'insert into' function of MySQL.
- By using the 'select' command we can query the desired output.

Software used for querying the results:-

MySQL Workbench 8.0 CE

Software used for analyzing using Bar plots:-

Microsoft Excel

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Job Data

Findings - 1

To find the number of jobs reviewed per hour per day of November 2020
1. We will use the data from job_id columns of the job_data table.

2. Then we will divide the total count of job_id (distinct and nondistinct) by (30 days * 24 hours)for finding the number of jobs reviewed per day.

Output/Result

number_of_jobs_reviewed_per_day_non_distinct
0.0111

number_of_jobs_reviewed_per_day_distinct
0.0083

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Job Data

Findings - 2

For calculating the 7-day rolling daily metric average of throughput:-

1. We will be first taking the count of job_id(distinct and non-distinct) and ordering them w.r.t ds (date of interview).

2. Then by using the ROW function we will be considering the rows between 6 preceding rows and the current row.

3. Then we will be taking the average of the jobs_reviewed.

Output /Result

date_of_review	jobs_reviewed	throughput_7_rolling_average
25-11-2020	1	1
26-11-2020	1	1
27-11-2020	1	1
28-11-2020	2	1.25
29-11-2020	1	1.2
30-11-2020	2	1.3333

date_of_review	jobs_reviewed	throughput_7_rolling_average_non_distinct_job_id
25-11-2020	1	1
26-11-2020	1	1
27-11-2020	1	1
28-11-2020	2	1.25
29-11-2020	1	1.2
30-11-2020	2	1.3333

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Job Data

Findings - 3

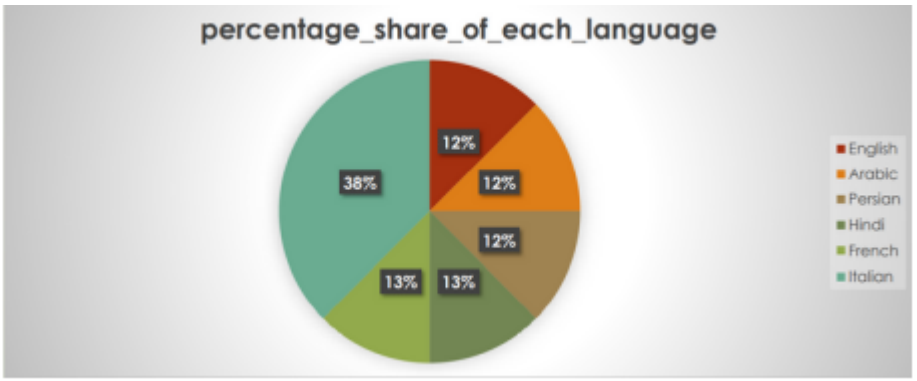
To Calculate the percentage share of each language (distinct and nondistinct):-

1. We will first divide the total number of languages (distinct/non-distinct) by the total number of rows presents in the table.

2. Then we will do the grouping based on the languages..

Output /Result

job_id	language	total_of_each_language	percentage_share_of_each_language
21	English	1	12.5
22	Arabic	1	12.5
23	Persian	3	37.5
25	Hindi	1	12.5
11	French	1	12.5
20	Italian	1	12.5

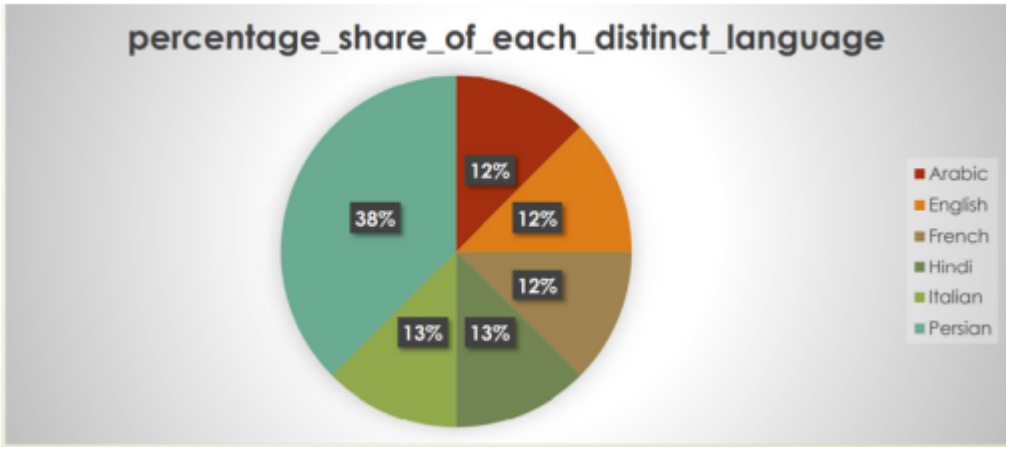


Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Output/result

job_id	language	total_of_each_language	percentage_share_of_each_distinct_language
22	Arabic	1	12.5
21	English	1	12.5
11	French	1	12.5
25	Hindi	1	12.5
20	Italian	1	12.5
23	Persian	1	37.5



Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Job Data

Findings - 4

To view the duplicate rows having the same value we will:-
1. First decide in which do we need to find the duplicate row values

2. After deciding the column(parameter) we will use the ROW_NUMBER function to find the row numbers having the same value

3. Then we will portioning the ROW_NUMBER function over the column (parameter) that we decided i.e. job_id

4. Then using the WHERE function we will find the row_num having value greater than 1 i.e. row_num > 1 based on the occurrence of the job_id in the table

Output /Result

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

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Investigating Metric Spike

Findings - 1

To find the weekly user engagement:-

1. We will extract the week from the occurred_at column of the events table using the EXTRACT function and WEEK function

2. Then we will be counting the number of distinct user_id from the events table

3. Then we will use the GROUP BY function to group the output w.r.t week from occurred_at

Output /Result

week_number	number_of_users
18	791
19	1244
20	1270
21	1341
22	1293
23	1366
24	1434
25	1462
26	1443
27	1477
28	1556
29	1556
30	1593
31	1685
32	1483
33	1438
34	1412
35	1442

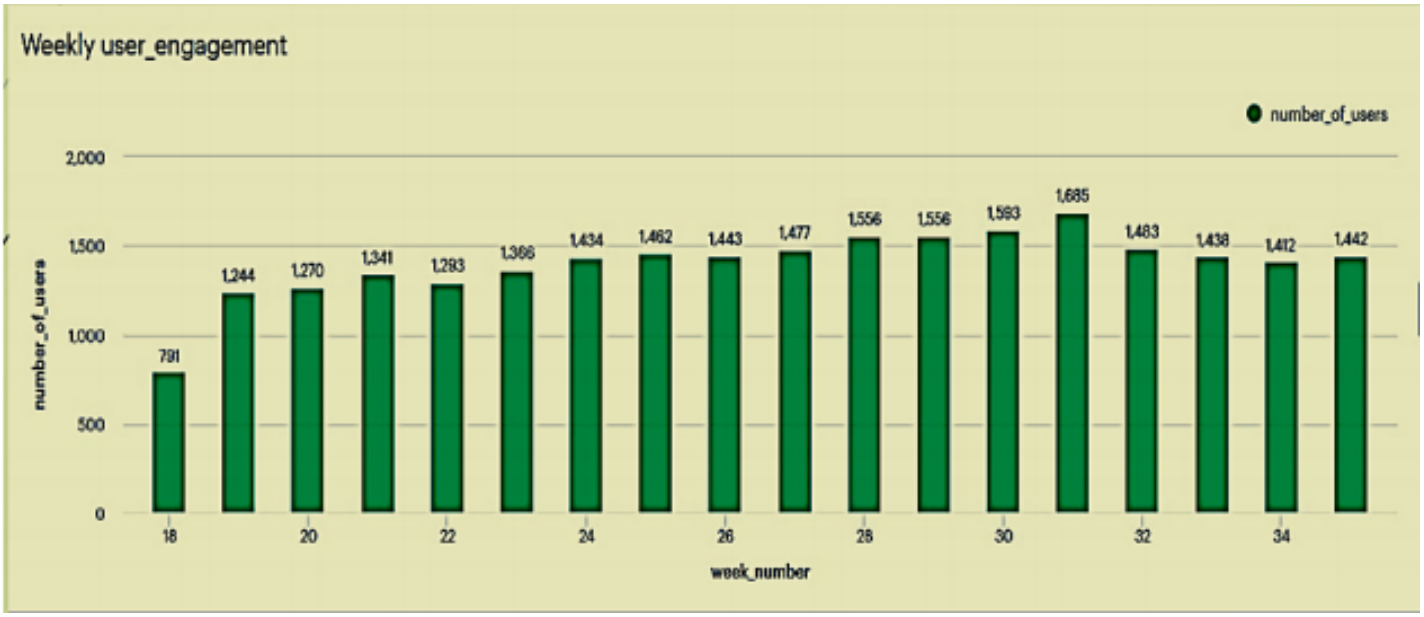
Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Investigating Metric Spike

Findings - I

Output /Result



Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Investigating Metric Spike

Findings - 2

To find the user growth (number of active users per week):-

1. First we will the extract the year and week for the occurred_at column of the users table using the extract, year and week functions

2. Then we will group the extracted week and year on the basis of year and week number

3. Then we ordered the result on the basis of year and week number

4. Then we will find the cumm_active_users using the SUM, OVER and ROW function between unbounded preceding and current row

Output /Result

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

year_num	week_num	num_active_users	cum_active_users	year_num	week_num	num_active_users	cum_active_users
2013	1	67	67	2013	45	97	2564
2013	2	29	96	2013	46	94	2658
2013	3	47	143	2013	47	82	2740
2013	4	36	179	2013	48	103	2843
2013	5	30	209	2013	49	96	2939
2013	6	48	257	2013	50	117	3056
2013	7	41	298	2013	51	123	3179
2013	8	39	337	2013	52	104	3283
2013	9	33	370	2014	1	91	3374
2013	10	43	413	2014	2	122	3496
2013	11	33	446	2014	3	112	3608
2013	12	32	478	2014	4	113	3721
2013	13	33	511	2014	5	130	3851
2013	14	40	551	2014	6	132	3983
2013	15	35	586	2014	7	135	4118
2013	16	42	628	2014	8	127	4245
2013	17	48	676	2014	9	127	4372
2013	18	48	724	2014	10	135	4507
2013	19	45	769	2014	11	152	4659
2013	20	55	824	2014	12	132	4791
2013	21	41	865	2014	13	151	4942
2013	22	49	914	2014	14	161	5103
2013	23	51	965	2014	15	166	5269
2013	24	51	1016	2014	16	165	5434
2013	25	46	1062	2014	17	176	5610
2013	26	57	1119	2014	18	172	5782
2013	27	57	1176	2014	19	160	5942
2013	28	52	1228	2014	20	186	6128
2013	29	71	1299	2014	21	177	6305
2013	30	66	1365	2014	22	186	6491
2013	31	69	1434	2014	23	197	6688
2013	32	66	1500	2014	24	198	6886
2013	33	73	1573	2014	25	222	7108
2013	34	70	1643	2014	26	210	7318
2013	35	80	1723	2014	27	199	7517
2013	36	65	1788	2014	28	223	7740
2013	37	71	1859	2014	29	215	7955
2013	38	84	1943	2014	30	228	8183
2013	39	92	2035	2014	31	234	8417
2013	40	81	2116	2014	32	189	8606
2013	41	88	2204	2014	33	250	8856
2013	42	74	2278	2014	34	259	9115
2013	43	97	2375	2014	35	266	9381
2013	44	92	2467				

count
9381

Investigating Metric Spike

Findings - 3

The weekly retention of users-sign up cohort can be calculated by two means i.e. either by specifying the week number (18 to 35) or for the entire column of occurred_at of the events table.

1. Firstly we will extract the week from occurred_at column using the extract, week functions

2. Then, we will select out those rows in which event_type = 'signup_flow' and event_name = 'complete_signup'

3. If finding for a specific week we will specify the week number using the extract function

4. Then using the left join we will join the two tables on the basis of user_id where event_type = 'engagement'

5. Then we will use the Group By function to group the output table on the basis of user_id

6. Then we will use the Order By function to order the result table on the basis of user_id

Output /Result

Google Drive Link for saved result

https://drive.google.com/file/d/1X7ErT9nyXAT6YRLz8l-Sq9RJTzzueneH/view?usp=drive_link

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Output /Result

Google Drive Link for saved result (specifying week number as 18)

https://drive.google.com/file/d/1saTYQ3CX6H3tZf_2gsC52IvKt7Ieaw2h/view?usp=sharing

Investigating Metric Spike

Findings - 4

To find the weekly user engagement per device:-

1. Firstly we will extract the year_num and week_num from the occurred_at column of the events table using the extract, year and week function

2. Then we will select those rows where event_type = 'engagement' using the WHERE clause

3. Then by using the Group By and Order By function we will group and order the result on the basis of year_num, week_num and device

Output /Result

Google Drive link for saved result

Weekly user engagement per device

https://drive.google.com/file/d/1p6nQ64n_fA0oicNPJ0KOLDSiG7zpozm0/view?usp=sharing

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Investigating Metric Spike

Findings - 4

To find the email engagement metrics(rate) of users:-

1. We will first categorize the action on the basis of email_sent, email_opened and email_clicked using the CASE, WHEN, THEN functions

2. Then we select the sum of category of email_opened divide by the sum of the category of email_sent and multiply the result by 100.0 and name is as email_opening_rate

3. Then we select the sum of category of email_clicked divide by the sum of the category of email_sent and multiply the result by 100.0 and name is as email_clicking_rate

4. email_sent = ('sent_weekly_digest','sent_reengagement_email')

5. email_opened = 'email_open'

6. email_clicked = 'email_clickthrough'

Output /Result

email_opening_rate	email_clicking_rate
33.58338805	14.78988838

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Analysis

From the tables and Bar plot I have infer the following:-

- number of distinct job reviewed per day is 0.0083
- number of non-distinct jobs reviewed per day is 0.0111
-
- 7 day rolling average throughput for 25, 26, 27, 28, 29 and 30 Nov 2020 are 1, 1, 1, 1.25, 1.2 and 1.3333 respectively(for both distinct and non-distinct)
- Percentage Share of each language i.e. Arabic, English, French, Hindi, Italian and Persian are 12.5, 12.5, 12.5, 12.5, 12.5 and 37.5 respectively(for both distinct and nondistinct)
- There are 2 duplicates values/rows having job_id = 23 and language = Persian in both the rows

Using the Why's approach I am trying to find more insights

- Why there is a difference of values between the number of distinct jobs reviewed per day and number of non-distinct jobs reviewed per day?
- May be due to repeated values in two or more rows or the dataset consisted of duplicate rows Why one shall use 7 day rolling average for calculating throughput and not daily metric average?
- For calculating the throughput we will be using the 7-day rolling because 7-day rolling gives us the average for all the days right from day 1 to day 7 Whereas daily metric gives us average for only that particular day itself.
- Why is it that percentage share of all other languages is 12.5% but that of language = 'Persian' is 37.5?
- In such cases there are two chances i.e. either there were duplicate rows having language as 'Persian' or there were really two or more unique people who were

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

speaking in Persian language.

- Why do we need to look for duplicate rows in an dataset?
- Duplicates have a direct influence of the Analysis going wrong and may led to wrong Business Decision leading to loss to the company or any entity; so to avoid these one must look for duplicates and remove them where necessary.

From the tables and Bar plot I have infer the following:-

- The weekly user engagement is the highest for week 31 i.e. 1685.
- There are in total 9381 active users from 1st week of 2013 to the 35th week of 2014.
- The email_opening_rate is 33.5833 and email_clicking_rate is 14.78988.

I have used the Why's approach to gain few more insights:-

- Why is Email Engagement plays an important role?
  Email Engagement helps the firms to decide the discounts and offers on specific products. In this case the email_opening_rate is 33.58 i.e. out of the 100 mails send only 34 mails were opened and the email_clicking_rate is 14.789 i.e. out of 100 mails opened only 15 mails were clicked for more details regarding the discount/product details. This means that the current firm needs to have some more catchy line for mails also the firm needs to do rigorous planning and deciding content before sending the mails.
- Why is weekly engagement per device plays an important role?
  Based on the reviews from users weekly engagement per device helps the firms on which devices they must focus more and which devices need more improvements so they also get a good review in users weekly engagement per device

Operation Analytics and Investigating Metric Spike

Prepared by: Tushar Bhagwat

Conclusion

In conclusion, I emphasize the critical importance of Operation Analytics and Investigating Metric Spikes in a comprehensive business strategy. These activities should be conducted diligently on a daily, weekly, monthly, quarterly, or yearly basis, aligning with the specific needs and objectives of the firm.

Moreover, an unwavering focus on Email Engagement with customers is paramount. Utilizing compelling subject lines, coupled with judiciously applied discounts and coupons, can significantly contribute to expanding the existing customer base. This strategic approach not only enhances customer loyalty but also attracts new clientele, fostering sustained growth.

Furthermore, it is advisable for any firm or entity to consider establishing a dedicated department, if feasible, to address the concerns of visitors who abandoned the sign-up process. By actively engaging with these individuals and providing guidance, the firm can effectively nurture them from mere visitors into valuable customers. This proactive approach not only demonstrates a commitment to customer satisfaction but also has the potential to convert lost opportunities into successful business relationships.