OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE
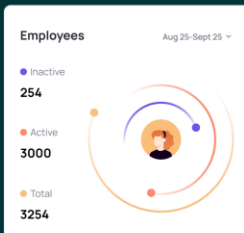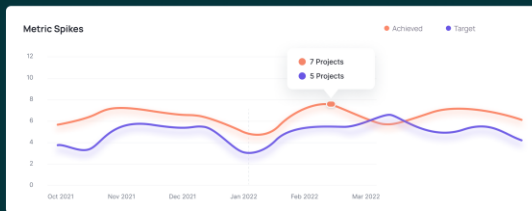
BY ANANWITA SARKAR

Project Agenda

1. **Project Description:** The project's objective is to leverage operational analytics for end-to-end analysis of a company's operations, identify areas for improvement, and provide actionable insights to cross-functional teams. As the Data Analyst Lead at Microsoft, I will collaborate closely with departments like operations, support, and marketing to derive valuable insights from their data. The primary focus is on optimizing workflows, enhancing automation, and predicting the company's overall growth or decline. Additionally, the project will involve analyzing metric spikes in daily engagement and sales, investigating their causes, and providing insights to address any dips in these metrics.

2. **Approach:** This project involves gathering data sets and tables from various departments and sources within the company, ensuring data quality and normalization during the import into an SQL database. Through SQL queries and aggregations, the collected data will be explored and analyzed to uncover patterns, trends, and anomalies. Key performance metrics, such as daily engagement, sales, and customer satisfaction, will be identified in collaboration with cross-functional teams and tracked using SQL queries. Metric spikes will be investigated using SQL subqueries, analyzing historical data and collaborating with departments for contextual understanding. Insights and recommendations derived from SQL analysis will drive process improvements, workflow optimization, and automation.

3. **Tech-Stack Used:** Used MySQL workbench 8.0 community server version 8.0.33 which is owned by oracle.

Task 1: Calculate the number of jobs reviewed per hour per day in November 2020

Output:

date	jobs viewed per hour per day
30-11-2020	180
29-11-2020	180
28-11-2020	218
27-11-2020	35
26-11-2020	64
25-11-2020	80

Insights: The number of jobs reviewed per hour per day in November 2020 varies, with higher activity on some days and lower activity on others.

Investigate if there were any technical issues or other factors affecting user engagement.

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

Output:

weekly_throughput
0.03

date	daily_metric
30-11-2020	0.05
29-11-2020	0.05
28-11-2020	0.06
27-11-2020	0.01
26-11-2020	0.02
25-11-2020	0.02

Insights: The 7-day rolling average of throughput provides a smoothed view of the data, allowing you to observe trends over time without being affected by daily fluctuations.

Continue using the 7-day rolling average for throughput analysis, as it provides a more stable representation of performance trends. This can help in identifying long-term patterns and making more informed decisions.

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

Output:

language	share_of_lang
English	12.50
Arabic	12.50
Persian	37.50
Hindi	12.50
French	12.50
Italian	12.50

Insights: The language distribution in the last 30 days is relatively balanced, with Persian having the highest share.

Consider investing in language-specific content or features to enhance user engagement in languages with lower shares.

Task 4: Display duplicate row count from the table?

Output:

actor_id	tot_count
1003	2

Insights: There is one duplicate row in the data based on the actor_id column.

Implement data validation mechanisms to prevent such duplicates in the future.

Task 5: Calculate the weekly user engagement?

Output:

weeks	no_of_users
17	663
18	1068
19	1113
20	1154
21	1121
22	1186
23	1232
24	1275
25	1264
26	1302
27	1372
28	1365
29	1376
30	1467
31	1299
32	1225
33	1225
34	1204
35	104

Insights: User engagement seems to have peaked around week 30 and has shown some fluctuations over the observed period.

Look for patterns related to content updates, marketing campaigns, or any external events that might have influenced user behavior. Use these insights to plan future engagement strategies.

Task 6: Calculate the user growth for product

Output:

week_num	year_num	cumulative_sum
0	2013	23
0	2014	106
1	2013	136
1	2014	262
2	2013	310
2	2014	419
3	2013	455
3	2014	568
4	2013	598
4	2014	728
5	2013	776
5	2014	909
6	2013	947
6	2014	1082
7	2013	1124
7	2014	1249
8	2013	1283
8	2014	1412
9	2013	1455
9	2014	1588
10	2013	1620
10	2014	1774
11	2013	1805
11	2014	1935

12	2013	1968
12	2014	2116
13	2013	2155
13	2014	2322
14	2013	2357
14	2014	2519
15	2013	2562
15	2014	2726
16	2013	2772
16	2014	2951
17	2013	3000
17	2014	3170
18	2013	3214
18	2014	3377
19	2013	3434
19	2014	3619
20	2013	3658
20	2014	3834
21	2013	3883
21	2014	4066
22	2013	4120
22	2014	4316
23	2013	4366
23	2014	4562

24	2013	4607
24	2014	4836
25	2013	4893
25	2014	5100
26	2013	5156
26	2014	5357
27	2013	5409
27	2014	5631
28	2013	5703
28	2014	5918
29	2013	5985
29	2014	6206
30	2013	6273
30	2014	6511
31	2013	6578
31	2014	6771
32	2013	6842
32	2014	7087
33	2013	7160
33	2014	7421
34	2013	7499
34	2014	7758
35	2013	7821
35	2014	7839

36	2013	7911
37	2013	7996
38	2013	8086
39	2013	8170
40	2013	8257
41	2013	8330
42	2013	8429
43	2013	8518
44	2013	8614
45	2013	8705
46	2013	8793
47	2013	8895
48	2013	8992
49	2013	9108
50	2013	9232
51	2013	9334
52	2013	9381

Insights: User growth has generally been positive over time, with some fluctuations.

While user growth is positive, it's important to understand the factors driving this growth. Analyze the periods of higher growth and see if they align with specific product updates, marketing efforts, or market trends. This can help in replicating successful strategies to sustain and accelerate user growth.

Task 7: Calculate the weekly retention of users-sign up cohort?

Output:

weeks	no_of_users
17	72
18	163
19	185
20	176
21	183
22	196
23	196
24	229
25	207
26	201
27	222
28	215
29	221
30	238
31	193
32	245
33	261
34	259
35	18

Insights: Weekly user retention shows a gradual decline over time.

Focus on improving user retention strategies. Identify key touchpoints in the user journey where users might be dropping off and work on enhancing user experience, engagement, and value during those stages.

Task 8: Calculate the weekly engagement per device?

Output: *

device	weeks	no_of_users
acer aspire desktop	17	9
acer aspire notebook	17	20
amazon fire phone	17	4
asus chromebook	17	21
dell inspiron desktop	17	18
dell inspiron notebook	17	46
hp pavilion desktop	17	14
htc one	17	16
ipad air	17	27
ipad mini	17	19
iphone 4s	17	21
iphone 5	17	65
iphone 5s	17	42
kindle fire	17	6
lenovo thinkpad	17	86
mac mini	17	6
macbook air	17	54
macbook pro	17	143
nexus 10	17	16
nexus 5	17	40
nexus 7	17	18
nokia lumia 635	17	17

Insights: Engagement varies across different devices and weeks. Some devices consistently show higher engagement than others.

Consider optimizing the user experience for devices that show lower engagement. Additionally, monitor device trends over time to adapt your strategies and prioritize user engagement on devices with the highest potential.

* this is just a sample output of only 22 rows. There are 491 rows returned from the above query which could not be accommodated in a single page.

Task 9: Calculate the email engagement metrics?

Output:

open_rate	click_rate
33.5834	14.7899

Insights: The email engagement metrics show an open rate of approximately 33.58% and a click rate of about 14.79%.

Compare these metrics with industry benchmarks to determine how your email engagement measures up. If your rates are below average, consider refining your email content, subject lines, and targeting to increase engagement. Regularly test and optimize your email campaigns to improve these metrics.

Result:

Working on this project has allowed me to delve into the intricacies of operational analytics, gaining a deeper understanding of its principles and methodologies. Through data integration, I have learned how to effectively merge and normalize diverse datasets, ensuring accurate and reliable analysis. Employing SQL analysis techniques has sharpened my ability to query, aggregate, and uncover insights from complex datasets. Furthermore, collaborating closely with cross-functional teams has enhanced my communication and teamwork skills, enabling me to effectively translate data findings into actionable recommendations for process optimization, thus contributing to the overall success of the project.

THANK YOU