

OPERATION AND INVESTIGATING METRIC ANALYTICS

PROJECT BY ABHISHEK BHAKAT

Insights & Results



Project Description



Approach



Tech-Stack Used



Job Data Analysis



Investigating Metric Spike



Content



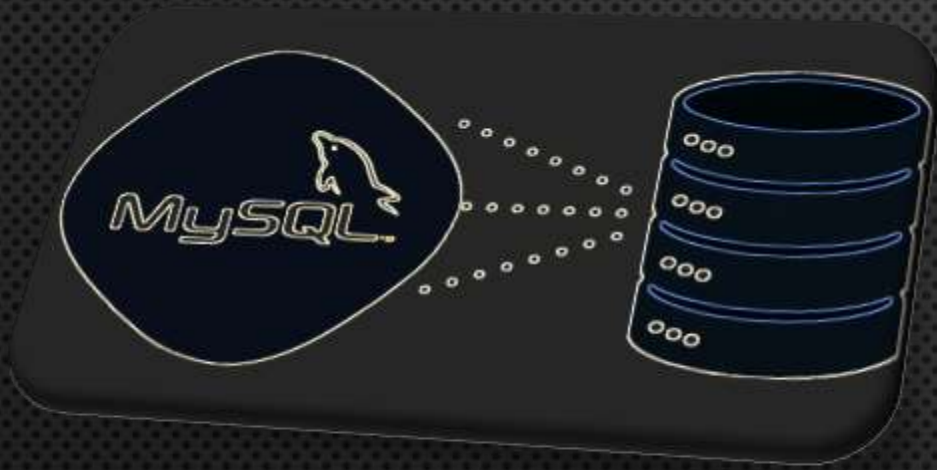
PROJECT DESCRIPTION



In this Project we are supposed to use our advanced SQL skills to analyse the data and provided valuable insights that can help to improve company operations and understand sudden changes in key metrics.

Through this projects we perform job Data Analysis and Investigating Metric Spike Analysis.

APPROACH



1. Database Creation : Database is created in MySQL workbench by importing the dataset provided by the company.

2. Insights Extraction : After creating the database, required insights are generated from database tables by running queries in MySQL workbench.

TECH-STACK USED



To perform the analysis in MySQL workbench version 8.0.41 is used as it is free and open source Relational Database Management System (RDMS) that used SQL.

JOB DATA ANALYSIS

Task 1 : Jobs Reviewed Over Time

Calculate the number of jobs reviewed per hour for each day in November 2020.

```
SELECT
  ds AS date,
  COUNT(job_id) AS no_job_id,
  sum(time_spent)/3600 as hours
FROM
  job_data
WHERE
  ds >= '2020-11-01'
  AND ds <= '2020-11-30'
GROUP BY ds
ORDER BY ds;
```

date	no_job_id	hours
2020-11-25	1	0.0125
2020-11-26	1	0.0156
2020-11-27	1	0.0289
2020-11-28	2	0.0086
2020-11-29	1	0.0056
2020-11-30	2	0.0111

- During November 2020 the job reviewing activity was very limited, on an average 1-2 jobs were reviewed per day, with less than 0.03 hours (about 1-2 minutes) per job reviewing.

JOB DATA ANALYSIS

Task 2 : Throughput Analysis

Calculate the 7 Day rolling average of throughput (number of events per second).

```
SELECT
  ds AS date,
  ROUND(COUNT(event) / SUM(time_spent), 2) AS daily_throughput,
  (SELECT
    ROUND(COUNT(event) / SUM(time_spent), 2)
  FROM
    job_data) AS 7_day_rolling_avg
FROM
  job_data
GROUP BY ds
ORDER BY ds;
```

date	daily_throughput	7_day_rolling_avg
2020-11-25	0.02	0.03
2020-11-26	0.02	0.03
2020-11-27	0.01	0.03
2020-11-28	0.06	0.03
2020-11-29	0.05	0.03
2020-11-30	0.05	0.03

On 2020-11-28, daily throughput jumped to 0.06 suggesting a temporary increase in processing or an influx of events

JOB DATA ANALYSIS

Task 3 : Language Share Analysis

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

```
SELECT
    language,
    SUM(time_spent) / (SELECT
        SUM(time_spent)
        FROM
            job_data) * 100 AS percentage
FROM
    job_data
GROUP BY language
ORDER BY percentage DESC;
```

language	percentage
French	35.1351
Persian	32.4324
Italian	15.2027
Arabic	8.4459
English	5.0676
Hindi	3.7162

French Language has the highest percentage share 35%.

JOB DATA ANALYSIS

Task 4 : Duplicate Rows Detection

Identify the duplicate rows from the data.

```
SELECT
    actor_id, COUNT(actor_id) AS No_of_Duplicates
FROM
    job_data
GROUP BY actor_id
HAVING No_of_Duplicates > 1;
```

actor_id	No_of_Duplicates
1003	2

Actor ID 1003 has duplicate rows
In data.

INVESTIGATING METRIC SPIKE

Task 1 : Weekly User Engagement

Measure the activeness of users on a weekly basis.

```
SELECT
    EXTRACT(WEEK FROM occurred_at) AS week_no,
    COUNT(DISTINCT user_id) AS no_of_users
FROM
    events
WHERE
    event_type = 'engagement'
GROUP BY week_no
ORDER BY week_no;
```

week_no	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

Week 30 has maximum active users while week 35 has the minimum active users.

INVESTIGATING METRIC SPIKE

Task 2 : User Growth Analysis

Analyse the growth of users over time for a product.

```
SELECT
    year, week_no, no_of_users, SUM(no_of_users)
    OVER (order by year, week_no ) AS user_growth
FROM
    ( SELECT
        EXTRACT(YEAR FROM created_at) AS year,
        EXTRACT(WEEK FROM created_at) AS week_no,
        COUNT(DISTINCT user_id) AS no_of_users
    FROM
        users
    GROUP BY year , week_no
    ORDER BY year , week_no) as temp_table;
```

year	week_no	no_of_users	user_growth
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
2013	20	39	820
2013	21	49	869
2013	22	54	923

year	week_no	no_of_users	user_growth
2014	13	167	5098
2014	14	162	5260
2014	15	164	5424
2014	16	179	5603
2014	17	170	5773
2014	18	163	5936
2014	19	185	6121
2014	20	176	6297
2014	21	183	6480
2014	22	196	6676
2014	23	196	6872
2014	24	229	7101
2014	25	207	7308
2014	26	201	7509
2014	27	222	7731
2014	28	215	7946
2014	29	221	8167
2014	30	238	8405
2014	31	193	8598
2014	32	245	8843
2014	33	261	9104
2014	34	259	9363
2014	35	18	9381

In 2014 week 33 has the highest user growth while 2014 week 35 has the lowest User growth.

INVESTIGATING METRIC SPIKE

Task 3 : Weekly Retention Analysis

Analyse the retention of users on a weekly basis after signing up for a product.

```
with cte1 as (  
  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) = 1),  
cte2 as (SELECT DISTINCT  
  user_id, EXTRACT(WEEK FROM occurred_at) AS engagement_week  
  FROM  
    events  
  where event_type = 'engagement')  
select count(user_id) as total_engaged_user,  
sum(case when retention_week = 1 then 1 else 0 end) as retained_users  
from (  
  SELECT  
    a.user_id,  
    a.sign_up_week,  
    b.engagement_week,  
    b.engagement_week - a.sign_up_week AS retention_week  
  FROM  
    cte1 a  
  left join cte2 b  
  on a.user_id = b.user_id  
  order by a.user_id) sub;
```

Weekly user retention shows a gradual decline over Time.

total_engaged_user	retained_users
615	114

INVESTIGATING METRIC SPIKE

Task 4 : Weekly Engagement Per Device

Calculate the activeness of users on a weekly basis per device.

```
SELECT
    EXTRACT(WEEK FROM occurred_at) AS week_no,
    device,
    COUNT(DISTINCT user_id) AS no_of_users
FROM
    events
GROUP BY week_no, device
ORDER BY week_no;
```

week_no	device	no_of_users
17	acer aspire desktop	9
17	acer aspire notebook	20
17	amazon fire phone	4
17	asus chromebook	21
17	dell inspiron desktop	18
17	dell inspiron notebook	46
17	hp pavilion desktop	14
17	htc one	16
17	ipad air	27
17	ipad mini	19
17	iphone 4s	21
17	iphone 5	65
17	iphone 5s	42
17	kindle fire	6
17	lenovo thinkpad	86
17	mac mini	6
17	macbook air	54
17	macbook pro	143
17	nexus 10	16
17	nexus 5	40
17	nexus 7	18
17	nokia lumia 635	17
17	samsung galaxy tablet	8
17	samsung galaxy note	7

Engagement varies across different device and weeks. Some devices consistently shows higher engagement than others.

INVESTIGATING METRIC SPIKE

Task 5 : Email Engagement Analysis

Analyse how users are engaging with the email service.

```
SELECT
  WEEK(occurred_at) AS week_no,
  COUNT(DISTINCT (CASE
    WHEN action = 'sent_weekly_digest' THEN user_id
  END)) AS sent_weekly_digest,
  COUNT(DISTINCT (CASE
    WHEN action = 'email_open' THEN user_id
  END)) AS email_open,
  COUNT(DISTINCT (CASE
    WHEN action = 'email_clickthrough' THEN user_id
  END)) AS email_clickthrough,
  COUNT(DISTINCT (CASE
    WHEN action = 'sent_reengagement_email' THEN user_id
  END)) AS sent_reengagement_email
FROM
  email_events
GROUP BY week_no
ORDER BY week_no;
```

Week_no	sent_weekly_digest	email_open	email_clickthrough	sent_reengagement_email
17	908	310	166	73
18	2602	900	425	157
19	2665	961	476	173
20	2733	989	501	191
21	2822	996	436	164
22	2911	965	478	192
23	3003	1057	529	197
24	3105	1136	549	226
25	3207	1084	524	196
26	3302	1149	550	219
27	3399	1207	613	213
28	3499	1228	594	213
29	3592	1201	583	213
30	3706	1363	625	231
31	3793	1338	444	222
32	3897	1318	416	200
33	4012	1417	490	264
34	4111	1502	481	261
35	0	41	78	48

Week 34 has the highest engagement using email services while week 35 has the lowest email engagement

INSIGHTS

JOB DATA ANALYSIS

- During November 2020 the job reviewing activity was very limited, on an average 1-2 jobs were reviewed per day, with less than 0.03 hours (about 1-2 minutes) per job reviewing.
- On 2020-11-28, daily throughput jumped to 006 suggesting a temporary increase in processing or an influx of events.
- French Language has the highest percentage share (35%) among others language.
- Actor ID 1003 has duplicate rows In data.

INVESTIGATING METRIC SPIKE

- Week 30 has the maximum active user while week 35 has the minimum active users.
- In 2014 week 33 has the highest user growth while 2014 week 35 has the lowest User growth.
- Weekly user retention shows gradual decline over time.
- Engagement varies across different device and weeks. Some devices consistently shows higher engagement than others.
- Week 34 has the highest engagement using email services while week 35 has the lowest email engagement.

RESULTS

Working on this project has allowed me to delve into the intricacies of operational analysis, gaining a deeper understanding of its principle and methodologies. Through the data integration, I have learnt how to effectively merge and normalize diverse dataset, ensuring accurate and reliable analysis. Employing SQL analysis techniques has sharpen my ability to query, aggregate, and uncover insight 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 process optimization, thus contributing to the overall success of the project.

THANK YOU