Operation Analytics and Investigating Metric Spike

Sourav Pattanayak

Table of contents

Project description

Objectives

Approach

Methodology and Tech-Stack Used

Insights

Key findings

Achievements

Project description

A designated data analyst has an important role in evaluating the processes and progresses of organizations, with the help of given datasets a data analyst performs few steps to reach the insights.

In this project I have gathered the insights on end to end operations of my organization and inspected the metric spikes that will be helpful in further growth and improvement of the organization.

Objectives

➤ To Calculate the number of jobs reviewed.

➤ Calculating the throughput.

➤ To find details on percentage shares of each language.

➤ Finding the duplicate rows in the given dataset.

➤ Calculating the user engagement and user growth over time.

➤ To inspect and find weekly retention, weekly engagement and Email engagement.

Approach

I have used relational database software to inspect and answer queries I was asked, business analytics tool for visualization of the insights, gathered the information and loopholes and jotted them down.

Understand

Plan
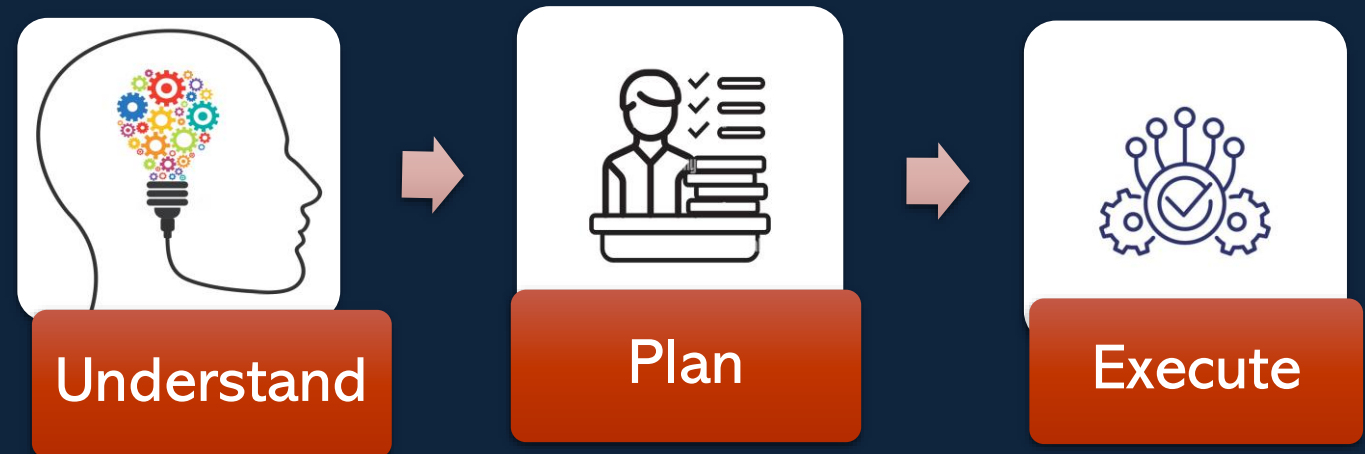
Execute

Methodology and Tech-Stack Used

➢ **My SQL version 8.0 was used in this project.**

➢ **The MySQL workbench is used to write and execute queries.**

➢ **The entire dataset is separated by two different approaches, one is provided to analyse the end to end operation of the organization, case study-1 and the other one is provided with information to investigate the metric spike, case study-2.**

➢ **A number of SQL functions and queries are used in this project.**

➢ **Microsoft Power BI was used to convert and visualize the output tables for better understanding and prediction.**

➢ **After jotting down the required insights , Microsoft Powerpoint was used to present the project.**
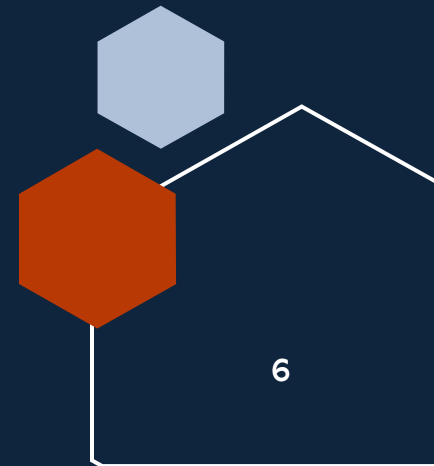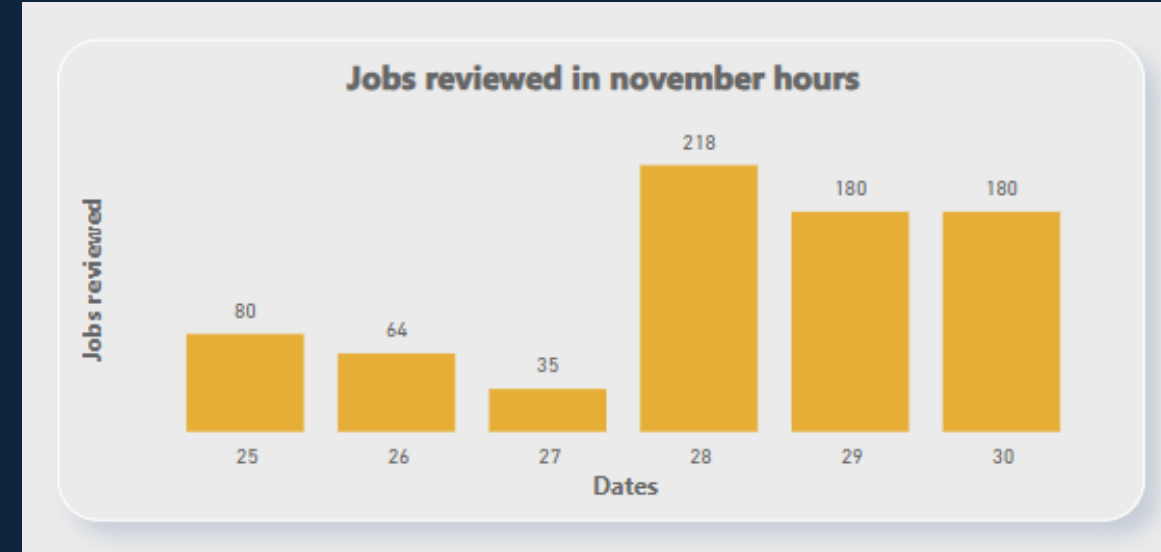
Insights(Case study-1)

Query 1: Number of jobs reviewed per hour per day for November 2020?

Code:

```
select ds as day,
round(count(job_id)/sum(time_spent)*3600) as
jobs_reviewed_perhour
from job_data
where ds between '2020-11-01' AND '2020-11-30'
group by ds;
```
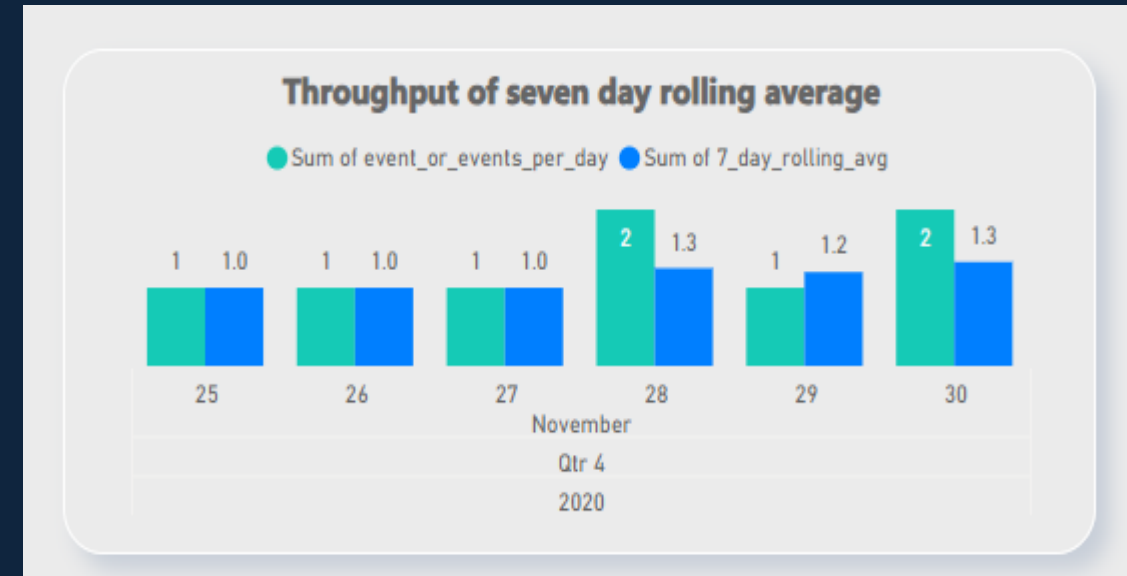
Insights(Case study-1)

Query 2: 7 day rolling average of Throughput

Code:

```
select ds, event_or_events_per_day,
round(avg(event_or_events_per_day)
over(order by ds rows between 6 preceding
and current row),2) as 7_day_rolling_avg
from (select ds, count(distinct event) as
event_or_events_per_day from
job_datagroup by ds) as temptable;
```
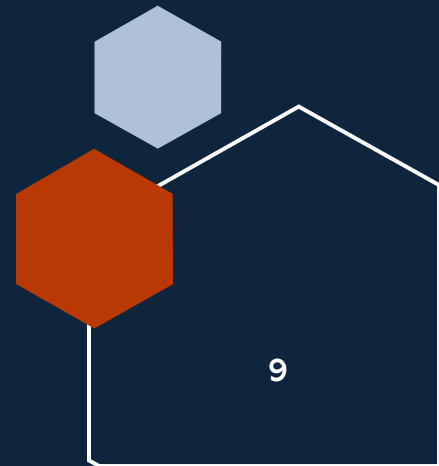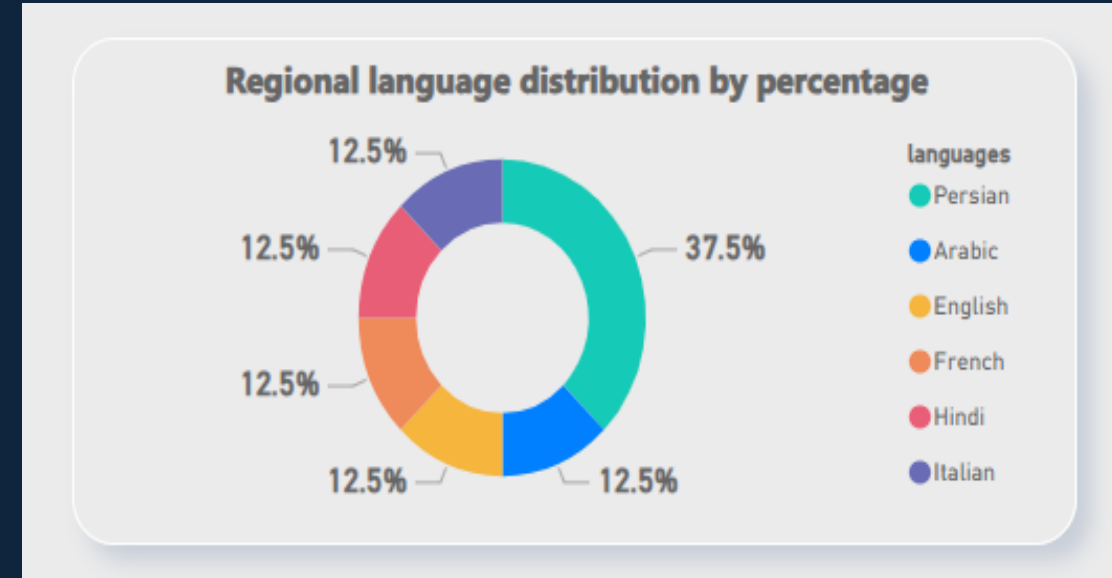
Insights(Case study-1)

Query 3: Percentage share of each language

Code:

```
select language as languages,
concat(round(count(*)*100/(select
count(*)from job_data),2),'%') as
percentage_share
from job_data
group by language;
```

Insights(Case study-1)

ds	no_of_duplicate
2020-11-30	2
2020-11-28	2

Query 4: Displaying duplicate rows

Code:

```
select ds, COUNT(ds) as no_of_duplicate
from operation_analytics.job_data
group by ds
having no_of_duplicate > 1;
```

In this code ,I have fetched the duplicate rows in 'ds' column, put more column names in select statement to filter more duplicate rows.
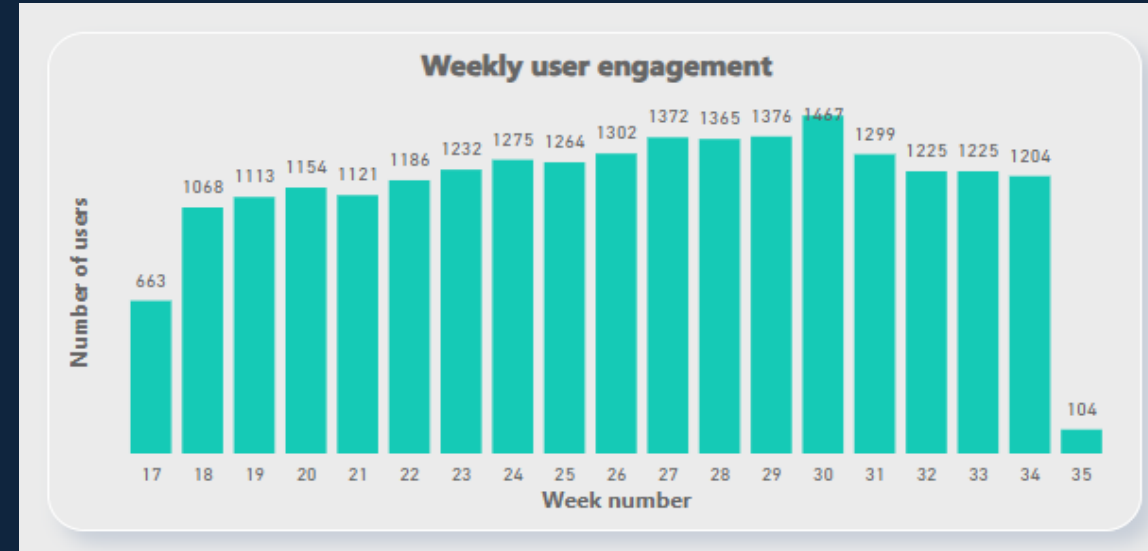
10

Insights(Case study-2)

Query 1: Calculating weekly user engagement

Code:

```
select week(occurred_at) as
week,count(distinct user_id) as
weekly_user_engagement
from events
where event_type='engagement'
group by week(occurred_at)
order by week(occurred_at);
```
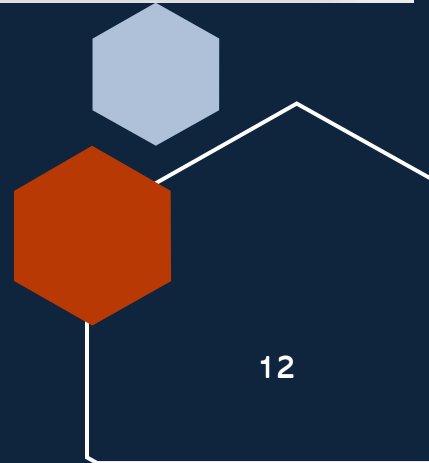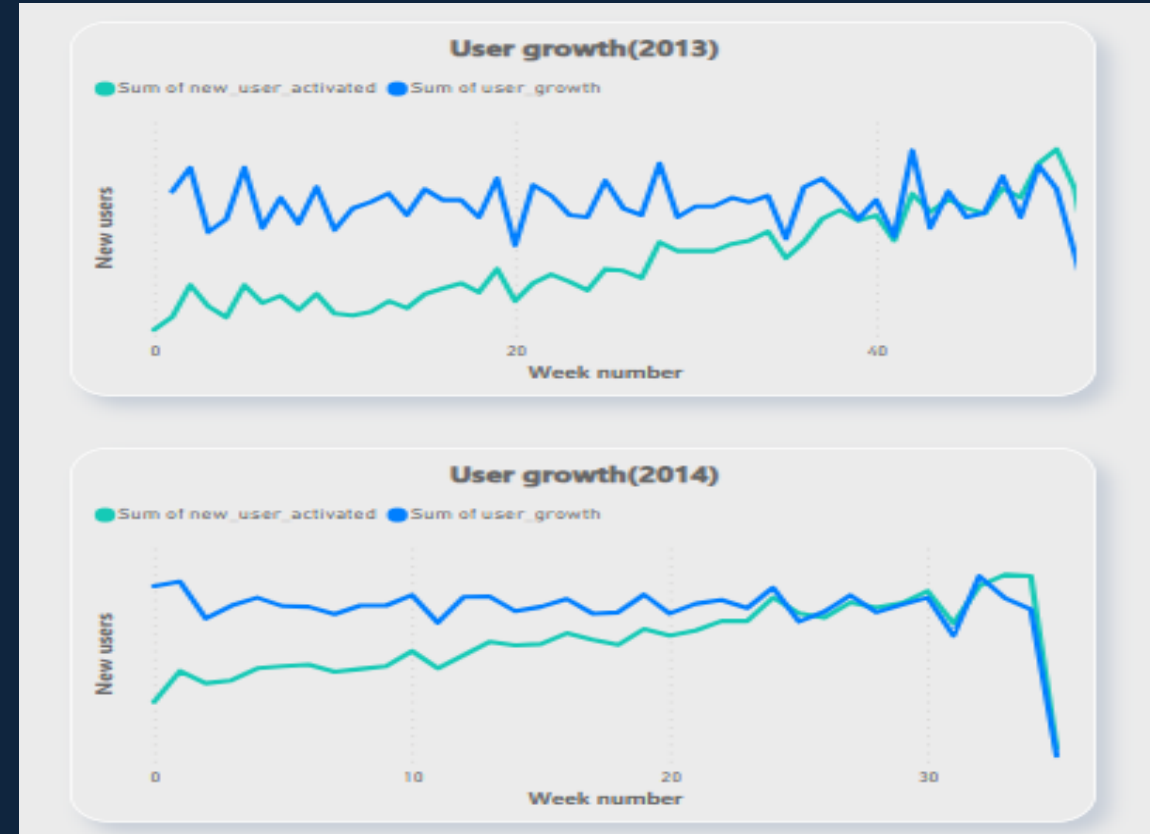
Insights(Case study-2)

User growth(2013)

User growth(2014)

Query 2: Calculating user growth over time.

Code:

```
select year,week_num,
new_user_activated,new_user_activated-
lag(new_user_activated) over( order by year,week_num )
as user_growth
from(select year(activated_at) as year,week(activated_at)
as week_num,count(user_id) as new_user_activated
from users
where activated_at is not null and state='active'
group by year,week_num
order by year,week_num) as temptable;
```
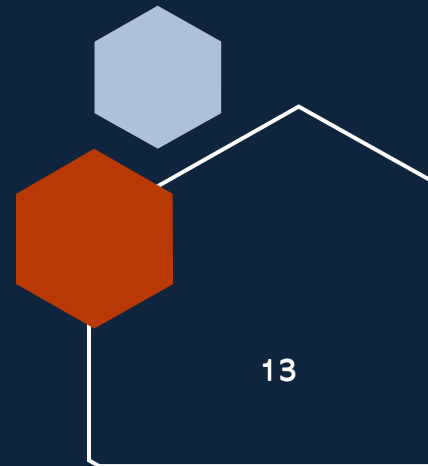
Insights(Case study-2)

Query 3: Calculating weekly retention of users-sign up cohort?

Code:

```
select t1.week_num,(t2.old_users - t1.new_users)as
Retained_users
from(select week(occurred_at) as
week_num,count(distinct user_id) as new_users
from events
where event_type = "signup_flow"
group by week_num) as t1
Join(select week(occurred_at) as
week_num,count(distinct user_id) as old_users
from events
where event_type = "engagement"
group by week_num) as t2 on t1.week_num =
t2.week_num;
```

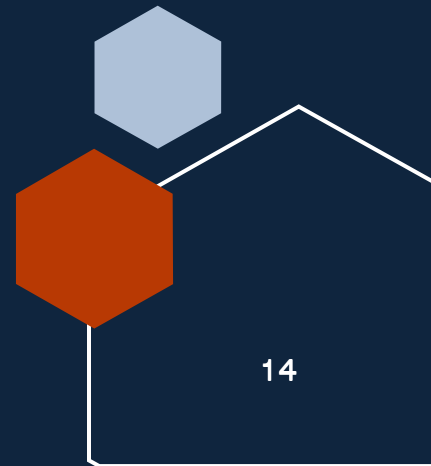# Insights(Case study-2)

weeks	dell inspiron n	ipad air	iphone 4s	iphone 5	iphone 5s	lenovo thinkpad	macbook air	macbook pro	nexus 5	samsung galaxy
35	10	1	6	4	4	18	10	22	4	8
17	48	29	27	69	47	94	61	154	45	66
18	84	62	51	130	76	174	136	280	89	108
21	88	55	49	147	86	193	131	268	99	117
19	93	57	48	123	88	204	123	296	99	115
20	93	63	63	142	86	201	134	289	110	125
33	118	50	38	122	73	208	162	354	81	108
32	116	53	48	134	75	205	145	353	78	109
22	100	67	52	135	81	199	159	280	111	139
34	115	46	57	117	81	224	160	332	79	122
23	112	49	62	163	85	201	137	308	100	133
24	113	62	62	156	95	186	171	281	99	133
25	116	67	44	152	94	220	141	311	96	121
26	101	63	58	170	97	217	150	303	100	139
31	123	59	58	159	79	233	166	360	80	124
27	105	64	73	185	95	217	154	332	92	144
28	113	57	68	164	102	240	168	335	98	143
29	123	59	71	158	102	226	173	334	94	159
30	131	75	73	176	119	244	177	358	95	138
Total	1902	1038	1008	2606	1565	3704	2658	5550	1649	2251

Query 4: Measuring weekly user engagement.

Code:

```
select week(occurred_at) as weeks, device,
count(distinct user_id) as device_engagement
from events
group by device, week(occurred_at)
order by week(occurred_at);
```

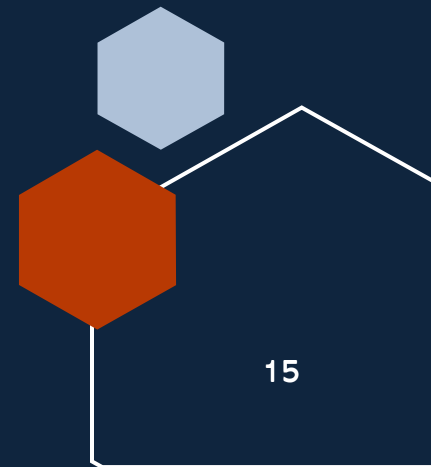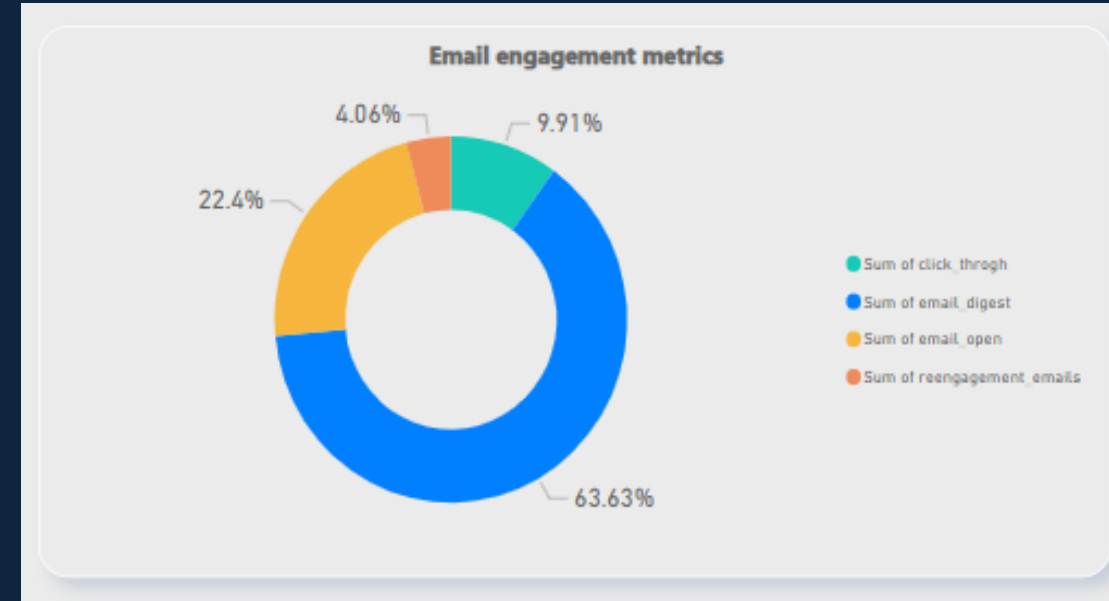*Top 10 device engagement is shown in the table.*

14

Insights(Case study-2)

Query 5: Calculating E-Mail engagement metrics.

Code:

```
select distinct week(occurred_at) as
week_num,count(distinct case when action =
'sent_weekly_digest' then user_id end) as
email_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
click_throgh,
count(distinct case when
action='sent_reengagement_email' then user_id
end) as reengagement_emails
from email_events
group by week(occurred_at);
```


Email engagement metrics

4.06%   9.91%

22.4%

63.63%

- Sum of click_throgh
- Sum of email_digest
- Sum of email_open
- Sum of reengagement_emails

Key findings

- In the given date range, the highest number of jobs were reviewed on 28th November,2020 which is 218.

- It seems that the difference between events happening each day and the throughput is not so big so we can prefer 7 day rolling over daily metrics.

- 37.5% of the language share is taken by persian language which is the highest in the distribution.

- 28th and 30th November,2020 has two duplicate rows in the dataset.

- 30th week holds the highest user engagement,1467.

- In 2013,highest user growth was in 42th week and highest number of total new users were seen in 50th week whereas the same calculation comes for 32nd and 34th week in 2014,there is sharp drop in user growth on 20th week(2013) and 35th week(2014).

- With some small decrease, the overall number of retained users gradually increased till 30th week and then started decreasing, the last week (i.e : 35th week) has seen the lowest number of retained users, from 34th week to 35th week, there is a sharp drop.

- Samsung galaxy is the most used device and on 29th week it has seen highest use.

- Most number of email engagements were acted upon email weekly digest that holds almost 64% of total email engagements.

Achievements

1. Data cleaning: this entire project has helped me learning the approaches to convert raw data into clean data.

2. Practical knowledge: experienced a practical exposure with different SQL commands and their uses in real life industries.

3. Uses of Power BI as business tool: Learned practical exposure and vast uses of strong tools like MS Power BI in business.

4. Uploading Large dataset: Learned the tricks and methods to upload large datasets into MySQL.

5. Business optimization methods: gained an idea on how the industries and organizations optimize their business problems, calculate the metric spike and perform operation analytics that helps in the growth of an industry or organization.

Thank you