# Operation and Metric Analytics

**Project Description:** Using Data Analytics to calculate the metrics of given datasets on jobs, users, events occurred, email events occurred to extract useful information like amount of jobs reviewed over time, no of events happening per second, share of each language and duplicate rows in case study 1 and activeness of users, user growth, weekly retention, weekly engagement, email metrics in case study 2.

**Tech-Stack Used**: MYSQL workbench 8.0 CE

**Approach**:
➢ Create the database in the workbench using the data provided

## Case study 1:

A. **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?

➢ To calculate number of jobs reviewed per hour per day of November 2020

Query:

```
select (count(distinct job_id)*3600)/(sum(time_spent)*count(distinct ds)) as answer from table1
```

Output:


➢ To calculate number of jobs reviewed per hour per each day of November 2020

Query:

```
select ds, (count(distinct job_id)*3600)/(sum(time_spent)*count(distinct ds)) as answer

from table1
```

group by ds

Output:

Output

| ds | answer |
|---|---|
| 2020-11-25 | 80 |
| 2020-11-26 | 64 |
| 2020-11-27 | 34 |
| 2020-11-28 | 218 |
| 2020-11-29 | 180 |
| 2020-11-30 | 180 |

B. **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?

Query:

```
select sum(answer)/7
from(
  select ds, (count(distinct job_id)*3600)/(sum(time_spent)*count(distinct ds)) as answer
  from table1
  group by ds
)
```

Output:

Output

| sum(answer)/7 |
|---|
| 108 |

For throughput it is better to use daily metric rather than 7-day rolling as the values are not closer to each other.

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?

Query:

```
select language, count(language)*100/(select count(language) from table1) as percentage

from table1

group by language
```

Output:

Output

| language | percentage |
|---|---|
| Arabic | 12 |
| English | 12 |
| French | 12 |
| Hindi | 12 |
| Italian | 12 |
| Persian | 37 |

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?

Query:

```
select ds, job_id, actor_id, event, language, time_spent, org, count(actor_id)

from table1

group by actor_id, job_id, ds, event, language, time_spent, org having count(actor_id)>1
```

Output:

Output

SQL query successfully executed. However, the result set is empty.

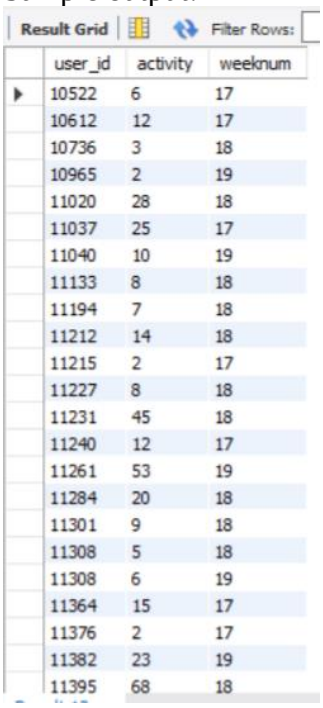
# Case study 2:

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?

Query:

```
SELECT user_id, count(occurred_at), weeknum

FROM (select user_id, occurred_at, week(occurred_at) as weeknum from events where event_type = 'engagement') ot

GROUP BY user_id, weeknum;
```

Output:
Sample output:



Result Grid | Filter Rows:

| user_id | activity | weeknum |
|---|---|---|
| 10522 | 6 | 17 |
| 10612 | 12 | 17 |
| 10736 | 3 | 18 |
| 10965 | 2 | 19 |
| 11020 | 28 | 18 |
| 11037 | 25 | 17 |
| 11040 | 10 | 19 |
| 11133 | 8 | 18 |
| 11194 | 7 | 18 |
| 11212 | 14 | 18 |
| 11215 | 2 | 17 |
| 11227 | 8 | 18 |
| 11231 | 45 | 18 |
| 11240 | 12 | 17 |
| 11261 | 53 | 19 |
| 11284 | 20 | 18 |
| 11301 | 9 | 18 |
| 11308 | 5 | 18 |
| 11308 | 6 | 19 |
| 11364 | 15 | 17 |
| 11376 | 2 | 17 |
| 11382 | 23 | 19 |
| 11395 | 68 | 18 |

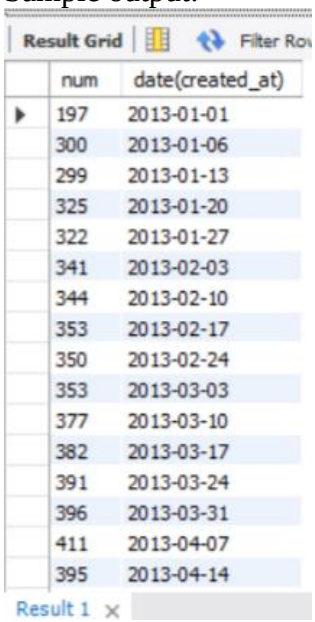
Link:
https://drive.google.com/file/d/1RrKW2vf5KUn4y4IcRc1-xcP2l9XI7Cht/view?usp=sharing

B. **User Growth:** Amount of users growing over time for a product.
**Your task:** Calculate the user growth for product?

Query:

```
select count(user_id) as num, date(created_at) from users

group by week(date(created_at));
```

Output:
Sample output:



Result Grid | Filter Rov

| num | date(created_at) |
|---|---|
| 197 | 2013-01-01 |
| 300 | 2013-01-06 |
| 299 | 2013-01-13 |
| 325 | 2013-01-20 |
| 322 | 2013-01-27 |
| 341 | 2013-02-03 |
| 344 | 2013-02-10 |
| 353 | 2013-02-17 |
| 350 | 2013-02-24 |
| 353 | 2013-03-03 |
| 377 | 2013-03-10 |
| 382 | 2013-03-17 |
| 391 | 2013-03-24 |
| 396 | 2013-03-31 |
| 411 | 2013-04-07 |
| 395 | 2013-04-14 |

Result 1

Link:
https://drive.google.com/file/d/1ZnOcysirgKHtXGcNK504aexwjMMtDact/view?usp=sharing

C. **Weekly Retention:** Users getting retained weekly after signing-up for a product.
**Your task:** Calculate the weekly retention of users-sign up cohort?

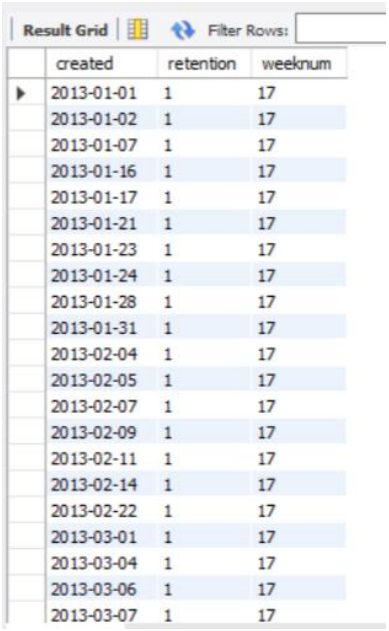
Query:

```
select created, count(distinct u1) as retention, week(occurred) as weeknum from(

select date(occurred_at) as occurred, u1, u2, date(created_at) as created from

(select user_id as u1, occurred_at, event_name from events

where event_type='engagement' order by user_id)t1 left join (select user_id as u2, created_at
from users where state='active')t2
```

on t1.u1=t2.u2 group by occurred, created )t3

group by week(occurred), created

Output:
Sample output:



Result Grid | Filter Rows:

| created | retention | weeknum |
|---|---|---|
| 2013-01-01 | 1 | 17 |
| 2013-01-02 | 1 | 17 |
| 2013-01-07 | 1 | 17 |
| 2013-01-16 | 1 | 17 |
| 2013-01-17 | 1 | 17 |
| 2013-01-21 | 1 | 17 |
| 2013-01-23 | 1 | 17 |
| 2013-01-24 | 1 | 17 |
| 2013-01-28 | 1 | 17 |
| 2013-01-31 | 1 | 17 |
| 2013-02-04 | 1 | 17 |
| 2013-02-05 | 1 | 17 |
| 2013-02-07 | 1 | 17 |
| 2013-02-09 | 1 | 17 |
| 2013-02-11 | 1 | 17 |
| 2013-02-14 | 1 | 17 |
| 2013-02-22 | 1 | 17 |
| 2013-03-01 | 1 | 17 |
| 2013-03-04 | 1 | 17 |
| 2013-03-06 | 1 | 17 |
| 2013-03-07 | 1 | 17 |

Link:

https://drive.google.com/file/d/1ZTN9V2bnjrJWcXuXY2qHDmsyr-eP6qbm/view?usp=sharing

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?

- To calculate the sum of total number of times the users use a particular type of device in a week

Query:

SELECT device, count(user_id), weeknum

FROM (select user_id, week(occurred_at) as weeknum, device from events where event_type = 'engagement') ot

GROUP BY device, weeknum;

Output:
Sample output:

Result Grid | Filter Rows:

| device | count(user_id) | weeknum |
|---|---|---|
| dell inspiron notebook | 503 | 17 |
| iphone 5 | 706 | 17 |
| iphone 4s | 448 | 18 |
| windows surface | 163 | 19 |
| macbook air | 1604 | 18 |
| iphone 4s | 217 | 17 |
| iphone 5s | 964 | 19 |
| macbook pro | 3301 | 18 |
| kindle fire | 265 | 18 |
| ipad mini | 309 | 18 |
| nexus 7 | 252 | 18 |
| nexus 5 | 382 | 17 |
| samsung galaxy s4 | 1130 | 18 |
| lenovo thinkpad | 1732 | 18 |
| samsumg galaxy tablet | 70 | 17 |
| acer aspire notebook | 406 | 19 |
| iphone 5 | 1328 | 18 |
| dell inspiron notebook | 953 | 18 |
| dell inspiron notebook | 1193 | 19 |
| iphone 5s | 473 | 17 |

Link:

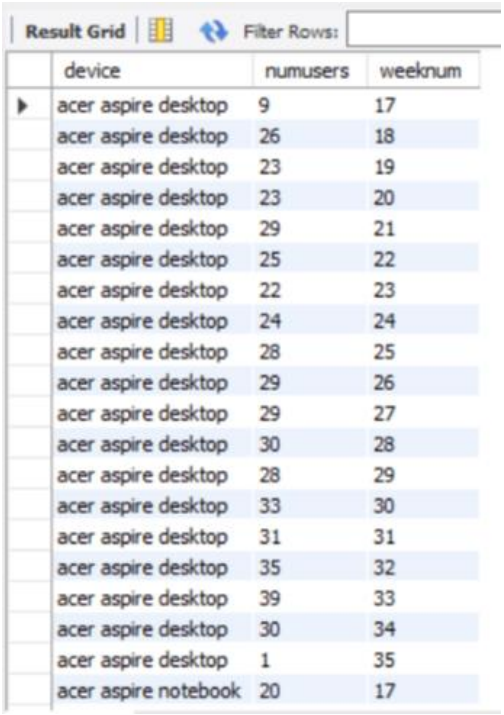
https://drive.google.com/file/d/1QLLQbhAPQRcra7o09_8YS_LXAOpnHD4u/view?usp=sharing

- To calculate the total number of distinct users that use a particular type of device in a week

Query:

```
SELECT device, count( distinct user_id) as numusers, weeknum

FROM (select user_id, week(occurred_at) as weeknum, device from events where event_type = 'engagement') ot

GROUP BY device, weeknum;
```

Output:
Sample output:



Result Grid | Filter Rows:

| device | numusers | weeknum |
|---|---|---|
| acer aspire desktop | 9 | 17 |
| acer aspire desktop | 26 | 18 |
| acer aspire desktop | 23 | 19 |
| acer aspire desktop | 23 | 20 |
| acer aspire desktop | 29 | 21 |
| acer aspire desktop | 25 | 22 |
| acer aspire desktop | 22 | 23 |
| acer aspire desktop | 24 | 24 |
| acer aspire desktop | 28 | 25 |
| acer aspire desktop | 29 | 26 |
| acer aspire desktop | 29 | 27 |
| acer aspire desktop | 30 | 28 |
| acer aspire desktop | 28 | 29 |
| acer aspire desktop | 33 | 30 |
| acer aspire desktop | 31 | 31 |
| acer aspire desktop | 35 | 32 |
| acer aspire desktop | 39 | 33 |
| acer aspire desktop | 30 | 34 |
| acer aspire desktop | 1 | 35 |
| acer aspire notebook | 20 | 17 |

Link:

https://drive.google.com/file/d/18J2QiYmgNX8rHxNVXEyPTR2nfT2dKAwj/view?usp=sharing

E. **Email Engagement:** Users engaging with the email service.
**Your task:** Calculate the email engagement metrics?

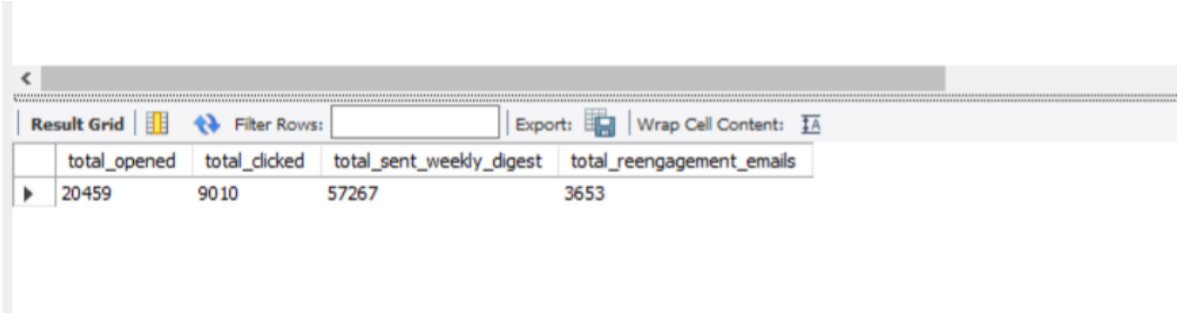
- To calculate the total number of weekly digests sent, emails opened, emails clicked through and re engagement emails sent.

Query:

```
SELECT

  SUM(CASE WHEN action = 'email_open' THEN 1 ELSE 0 END) AS total_opened,

  SUM(CASE WHEN action = 'email_clickthrough' THEN 1 ELSE 0 END) AS total_clicked,

  SUM(CASE WHEN action = 'sent_weekly_digest' THEN 1 ELSE 0 END) AS
total_sent_weekly_digest,

  SUM(CASE WHEN action = 'sent_reengagement_email' THEN 1 ELSE 0 END) AS
total_reengagement_emails

FROM
```

email_events;

Output:



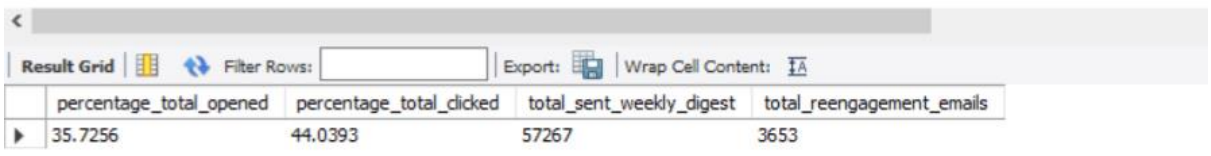
| total_opened | total_clicked | total_sent_weekly_digest | total_reengagement_emails |
|---|---|---|---|
| 20459 | 9010 | 57267 | 3653 |

- To calculate the total number of weekly digests sent, re engagement emails sent, percentage of emails opened, percentage of emails clicked through after opening.

Query:

```
SELECT

  (SUM(CASE WHEN action = 'email_open' THEN 1 ELSE 0 END)/SUM(CASE WHEN
action = 'sent_weekly_digest' THEN 1 ELSE 0 END)*100 )AS percentage_total_opened,

  (SUM(CASE WHEN action = 'email_clickthrough' THEN 1 ELSE 0 END)/SUM(CASE
WHEN action = 'email_open' THEN 1 ELSE 0 END)*100) AS percentage_total_clicked,

  SUM(CASE WHEN action = 'sent_weekly_digest' THEN 1 ELSE 0 END) AS
total_sent_weekly_digest,

  SUM(CASE WHEN action = 'sent_reengagement_email' THEN 1 ELSE 0 END) AS
total_reengagement_emails

FROM

  email_events;
```

Output:



| percentage_total_opened | percentage_total_clicked | total_sent_weekly_digest | total_reengagement_emails |
|---|---|---|---|
| 35.7256 | 44.0393 | 57267 | 3653 |

**Result:**

The project helped me in getting a clarity of the concepts that I learned and in learning new concepts that were required for the project and it helped me understand how the operation analytics work in a company.