# Operation Analytics and Investigating Metric Spike

**Project Description**

This is a detailed report for operation analytics and investigation metric spike for given two operation case studies, deriving certain insights from it.

Following are the operations, to whom we have to provide a detailed report:

**Case Study 1 (Job Data)**

1. **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?
2. **Throughput:** It is the no. of events happening per second.
   **Your task:** Let's say the above metric is called throughput. Calculate 7 days rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling, and why?
3. **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?
4. **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?

**Case Study 2 (Investigating metric spike on Users, Emails, and Email_Events table):**

1. **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?
2. **User Growth:** Amount of users growing over time for a product.
   **Your task:** Calculate the user growth for product?
3. **Weekly Retention:** Users getting retained weekly after signing-up for a product.
   **Your task:** Calculate the weekly retention of users-sign up cohort?
4. **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?
5. **Email Engagement:** Users engaging with the email service.
   **Your task:** Calculate the email engagement metrics?

**Approach**
Since the dataset provided is in CSV format, hence here we can freely explore the dataset via different analysis tools to get insights and a better understanding of its schema.
I selected MSSMS for understanding the structure of the tables and how they are connected with each other.

**Tech-Stack Used**
The software used for the analysis is *Microsoft SQL Server Management Studio 18*.
MSSMS is quite easy to execute the queries. Apart from this, it is easy to learn and beginner friendly, making it my first choice for analysis. Also, its recommendation system comes a  lot of handy while writing complex queries.

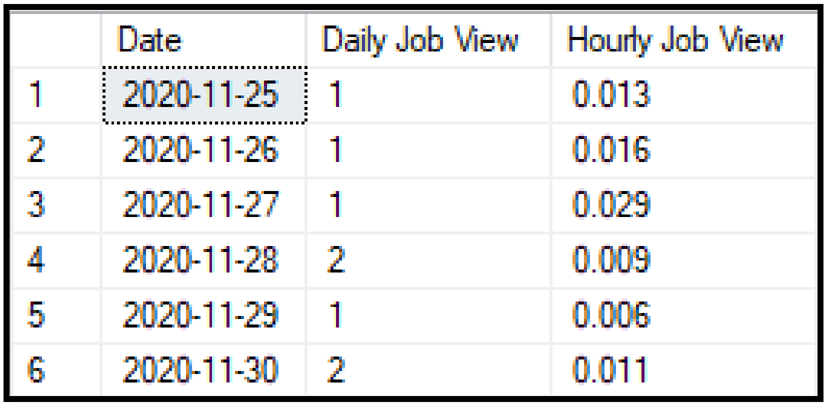
**Insights**

Following are all the insights that we got from the dataset after executing different queries.

**Case Study 1 (Job Data)**

1. **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?

```
select distinct CONVERT(date, ds) as 'Date',COUNT(job_id)as 'Daily Job View',
round(sum(time_spent)/(60*60),3) as 'Hourly Job View'  from job_data
where MONTH(ds) = 11
group by ds
```

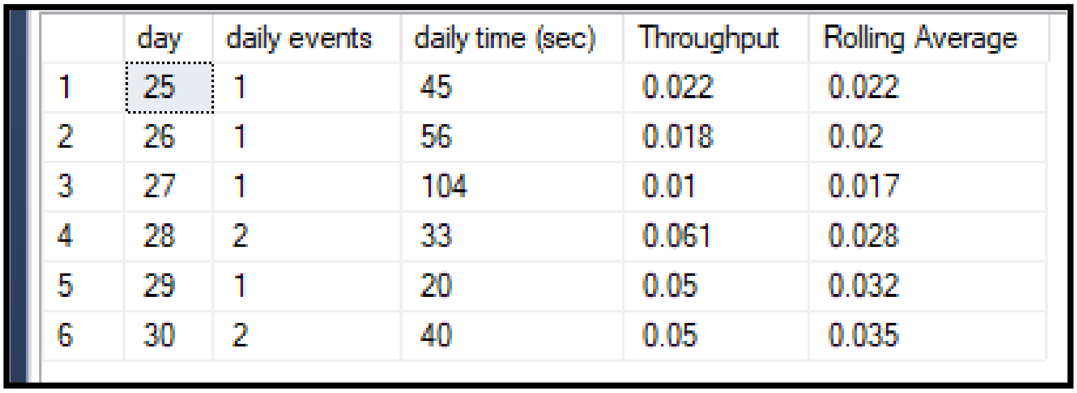


| | Date | Daily Job View | Hourly Job View |
|---|---|---|---|
| 1 | 2020-11-25 | 1 | 0.013 |
| 2 | 2020-11-26 | 1 | 0.016 |
| 3 | 2020-11-27 | 1 | 0.029 |
| 4 | 2020-11-28 | 2 | 0.009 |
| 5 | 2020-11-29 | 1 | 0.006 |
| 6 | 2020-11-30 | 2 | 0.011 |

2. **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?

**(NOTE:** Since the dataset contains days less than a weekday (i.e 7), hence I have calculated the throughput on Daily Metric).

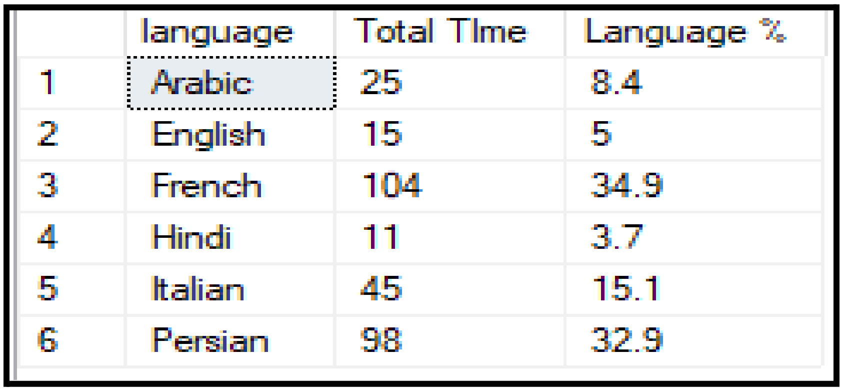
```
select *, round(avg(Throughput) over(order by day),3) AS 'Rolling Average'
from (
        select *,  round([daily events]/[daily time (sec)],3) as 'Throughput'
        from (
                select day(ds) as day,
                count(event) as 'daily events',sum(time_spent) as 'daily time (sec)'
                from job_data
                where MONTH(ds)=11
                group by day(ds)
        )table_1
)table_2
```



|  | day | daily events | daily time (sec) | Throughput | Rolling Average |
|---|---|---|---|---|---|
| 1 | 25 | 1 | 45 | 0.022 | 0.022 |
| 2 | 26 | 1 | 56 | 0.018 | 0.02 |
| 3 | 27 | 1 | 104 | 0.01 | 0.017 |
| 4 | 28 | 2 | 33 | 0.061 | 0.028 |
| 5 | 29 | 1 | 20 | 0.05 | 0.032 |
| 6 | 30 | 2 | 40 | 0.05 | 0.035 |

3. **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?

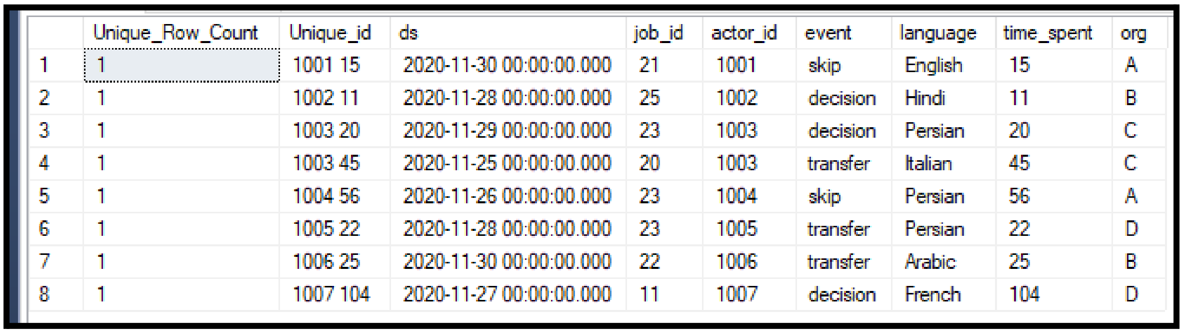
```
select distinct language,SUM(time_spent) as 'Total TIme',
round((SUM(time_spent)) / (select SUM(time_spent) from job_data),3)*100 as 'Language %'
from job_data
where ds between (select CONVERT(date, (select DATEADD(DAY,-30, (select MAX(ds) from
job_data))))) and
(select CONVERT(date, (select MAX(ds) from job_data)))
group by language;
```



| | language | Total TIme | Language % |
|---|---|---|---|
| 1 | Arabic | 25 | 8.4 |
| 2 | English | 15 | 5 |
| 3 | French | 104 | 34.9 |
| 4 | Hindi | 11 | 3.7 |
| 5 | Italian | 45 | 15.1 |
| 6 | Persian | 98 | 32.9 |

4. **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?

```
select ROW_NUMBER() over (Partition by CONCAT( actor_id,' ',time_spent) order by CONCAT(
actor_id,' ',time_spent)) as 'Unique_Row_Count' ,
CONCAT( actor_id,' ',time_spent) as 'Unique_id',*
from job_data
```

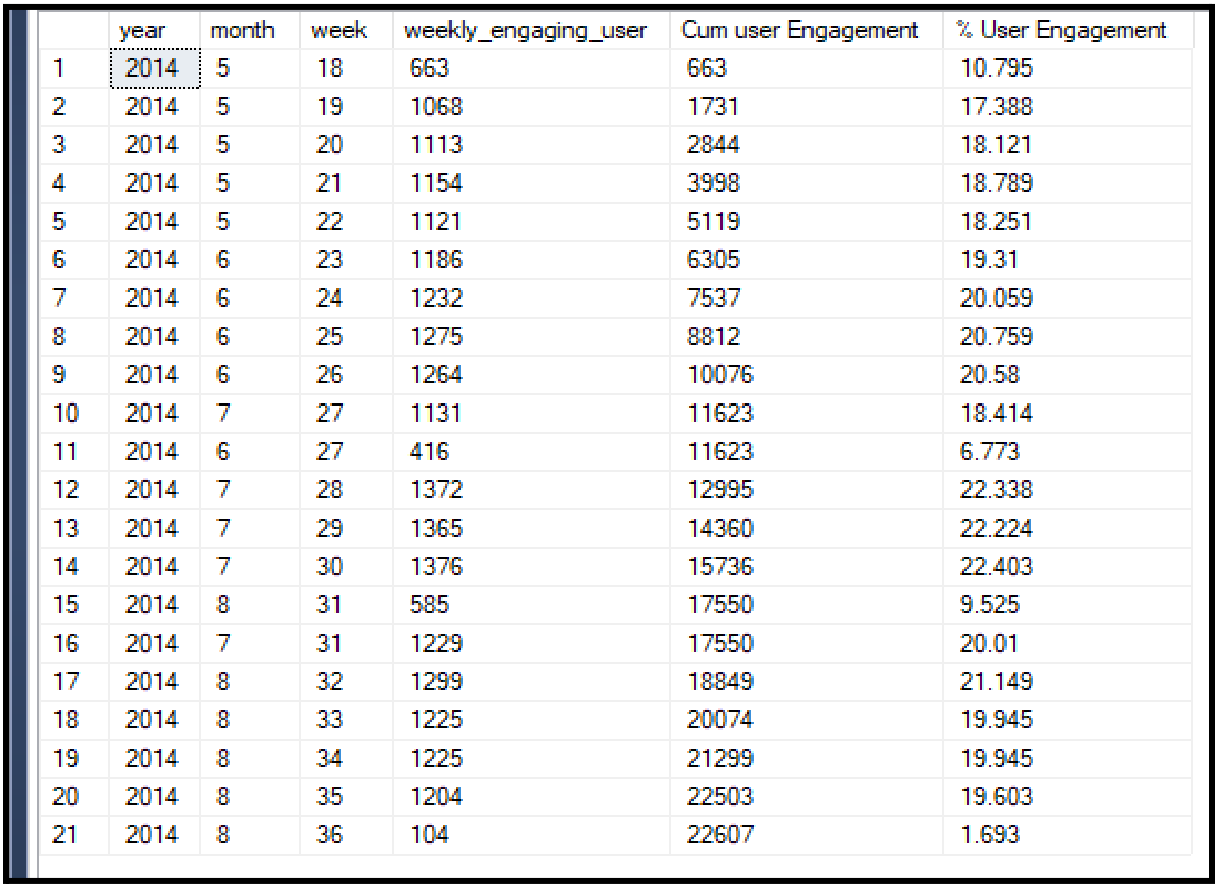


| | Unique_Row_Count | Unique_id | ds | job_id | actor_id | event | language | time_spent | org |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1001 15 | 2020-11-30 00:00:00.000 | 21 | 1001 | skip | English | 15 | A |
| 2 | 1 | 1002 11 | 2020-11-28 00:00:00.000 | 25 | 1002 | decision | Hindi | 11 | B |
| 3 | 1 | 1003 20 | 2020-11-29 00:00:00.000 | 23 | 1003 | decision | Persian | 20 | C |
| 4 | 1 | 1003 45 | 2020-11-25 00:00:00.000 | 20 | 1003 | transfer | Italian | 45 | C |
| 5 | 1 | 1004 56 | 2020-11-26 00:00:00.000 | 23 | 1004 | skip | Persian | 56 | A |
| 6 | 1 | 1005 22 | 2020-11-28 00:00:00.000 | 23 | 1005 | transfer | Persian | 22 | D |
| 7 | 1 | 1006 25 | 2020-11-30 00:00:00.000 | 22 | 1006 | transfer | Arabic | 25 | B |
| 8 | 1 | 1007 104 | 2020-11-27 00:00:00.000 | 11 | 1007 | decision | French | 104 | D |

## Case Study 2 (Investigating metric spike on Users, Emails, and Email_Events table):

1. **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?

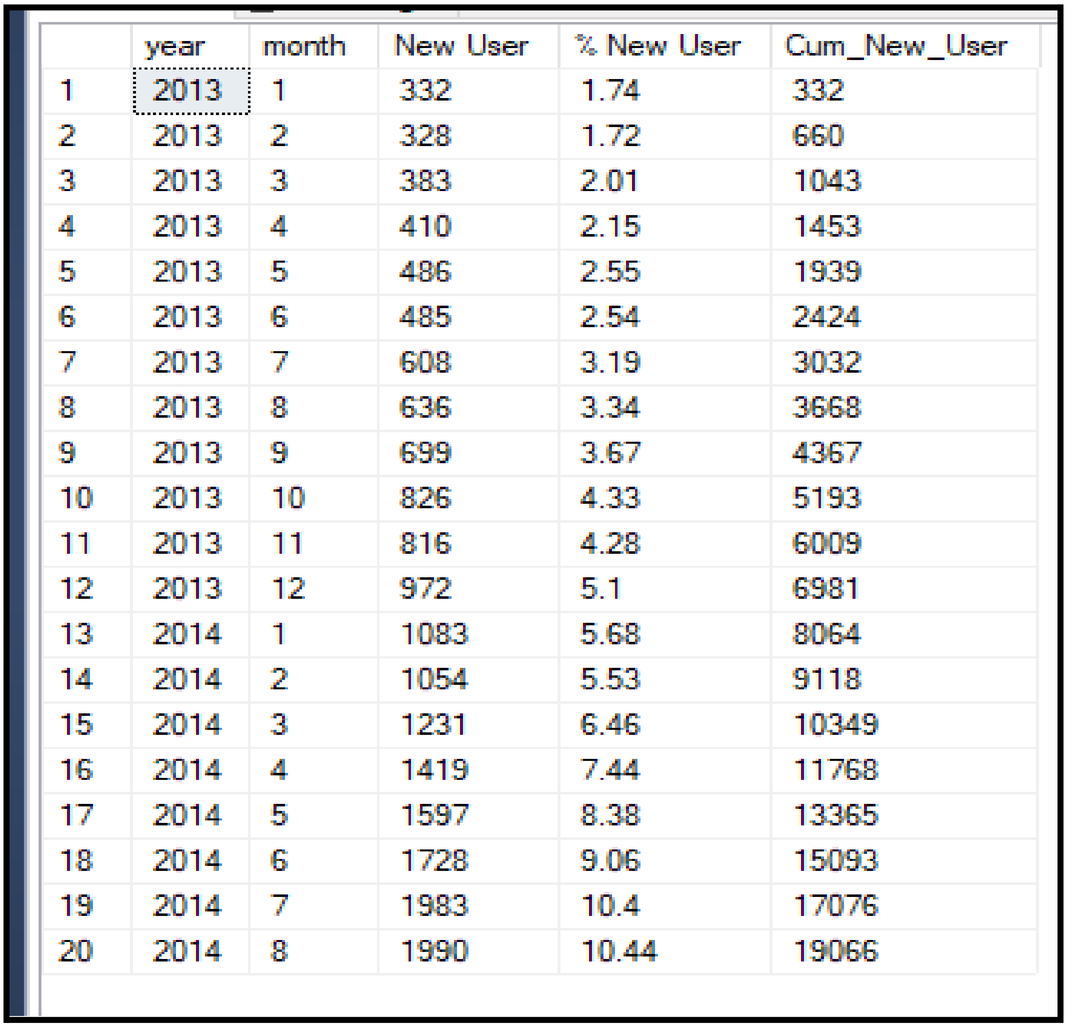
```
select *,
sum([weekly_engaging_user])over(order by week) as 'Cum user Engagement',
 round( cast(  ([weekly_engaging_user]*100.0) /
             (select count(distinct user_id) from events where event_type='engagement')
 as varchar(30)) ,3)   as '% User Engagement'
from (
          select year(occurred_at) as year ,month(occurred_at) as month
,datepart(week,occurred_at) as week,
          count(distinct user_id) as weekly_engaging_user
          from events
          where event_type='engagement'
          group by year(occurred_at),month(occurred_at),datepart(week,occurred_at)
)table_1
Order by week;
```



| | year | month | week | weekly_engaging_user | Cum user Engagement | % User Engagement |
|---|---|---|---|---|---|---|
| 1 | 2014 | 5 | 18 | 663 | 663 | 10.795 |
| 2 | 2014 | 5 | 19 | 1068 | 1731 | 17.388 |
| 3 | 2014 | 5 | 20 | 1113 | 2844 | 18.121 |
| 4 | 2014 | 5 | 21 | 1154 | 3998 | 18.789 |
| 5 | 2014 | 5 | 22 | 1121 | 5119 | 18.251 |
| 6 | 2014 | 6 | 23 | 1186 | 6305 | 19.31 |
| 7 | 2014 | 6 | 24 | 1232 | 7537 | 20.059 |
| 8 | 2014 | 6 | 25 | 1275 | 8812 | 20.759 |
| 9 | 2014 | 6 | 26 | 1264 | 10076 | 20.58 |
| 10 | 2014 | 7 | 27 | 1131 | 11623 | 18.414 |
| 11 | 2014 | 6 | 27 | 416 | 11623 | 6.773 |
| 12 | 2014 | 7 | 28 | 1372 | 12995 | 22.338 |
| 13 | 2014 | 7 | 29 | 1365 | 14360 | 22.224 |
| 14 | 2014 | 7 | 30 | 1376 | 15736 | 22.403 |
| 15 | 2014 | 8 | 31 | 585 | 17550 | 9.525 |
| 16 | 2014 | 7 | 31 | 1229 | 17550 | 20.01 |
| 17 | 2014 | 8 | 32 | 1299 | 18849 | 21.149 |
| 18 | 2014 | 8 | 33 | 1225 | 20074 | 19.945 |
| 19 | 2014 | 8 | 34 | 1225 | 21299 | 19.945 |
| 20 | 2014 | 8 | 35 | 1204 | 22503 | 19.603 |
| 21 | 2014 | 8 | 36 | 104 | 22607 | 1.693 |

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

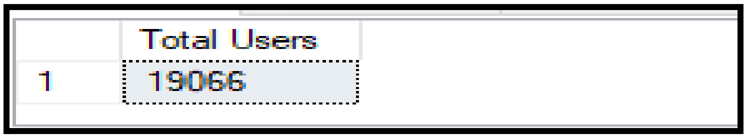
```
select *, SUM([New User]) over(order by year,month) as 'Cum_New_User'
from (
select  YEAR(created_at) as year , month(created_at) as month, count(user_id) as 'New User',
round( cast(  (count(created_at)*100.)/(select count(user_id)from users)  as varchar(30)),2)   as
'% New User'
from users
group by YEAR(created_at),month(created_at)
) table_1
order by year, month;
```



| | year | month | New User | % New User | Cum_New_User |
|---|---|---|---|---|---|
| 1 | 2013 | 1 | 332 | 1.74 | 332 |
| 2 | 2013 | 2 | 328 | 1.72 | 660 |
| 3 | 2013 | 3 | 383 | 2.01 | 1043 |
| 4 | 2013 | 4 | 410 | 2.15 | 1453 |
| 5 | 2013 | 5 | 486 | 2.55 | 1939 |
| 6 | 2013 | 6 | 485 | 2.54 | 2424 |
| 7 | 2013 | 7 | 608 | 3.19 | 3032 |
| 8 | 2013 | 8 | 636 | 3.34 | 3668 |
| 9 | 2013 | 9 | 699 | 3.67 | 4367 |
| 10 | 2013 | 10 | 826 | 4.33 | 5193 |
| 11 | 2013 | 11 | 816 | 4.28 | 6009 |
| 12 | 2013 | 12 | 972 | 5.1 | 6981 |
| 13 | 2014 | 1 | 1083 | 5.68 | 8064 |
| 14 | 2014 | 2 | 1054 | 5.53 | 9118 |
| 15 | 2014 | 3 | 1231 | 6.46 | 10349 |
| 16 | 2014 | 4 | 1419 | 7.44 | 11768 |
| 17 | 2014 | 5 | 1597 | 8.38 | 13365 |
| 18 | 2014 | 6 | 1728 | 9.06 | 15093 |
| 19 | 2014 | 7 | 1983 | 10.4 | 17076 |
| 20 | 2014 | 8 | 1990 | 10.44 | 19066 |

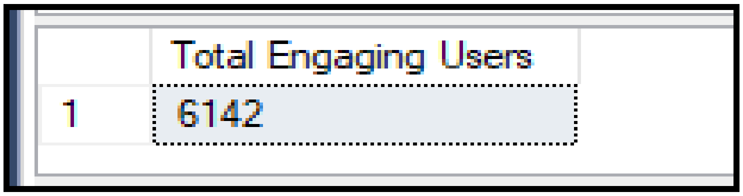
3. **Weekly Retention:** Users getting retained weekly after signing-up for a product.
   **Your task:** Calculate the weekly retention of users sign up cohort?

```
select COUNT(*)as 'Total Users' from users
```



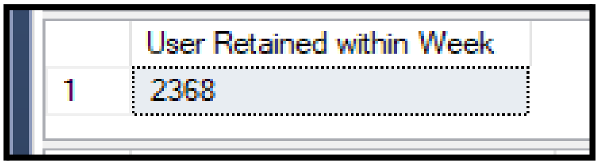
| | Total Users |
|---|---|
| 1 | 19066 |

```
select count(distinct user_id) as 'Total Engaging Users' from events
where event_type='engagement'
```



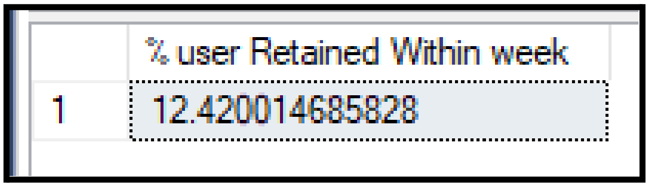
| | Total Engaging Users |
|---|---|
| 1 | 6142 |

```
select COUNT(*) as 'User Retained within Week'
from(
        select A.user_id, DATEADD(DAY,7,convert(date,created_at)) as '1 week later',
        min(convert(date, occurred_at)) as occur_date
        from users A
        right join  events B
        on A.user_id=B.user_id
        where event_type like 'engagement'
        group by A.user_id, DATEADD(DAY,7,convert(date,created_at)) ,B.user_id
        having  DATEADD(DAY,7,convert(date,created_at)) < min(convert(date, occurred_at))
)table1
```



| | User Retained within Week |
|---|---|
| 1 | 2368 |

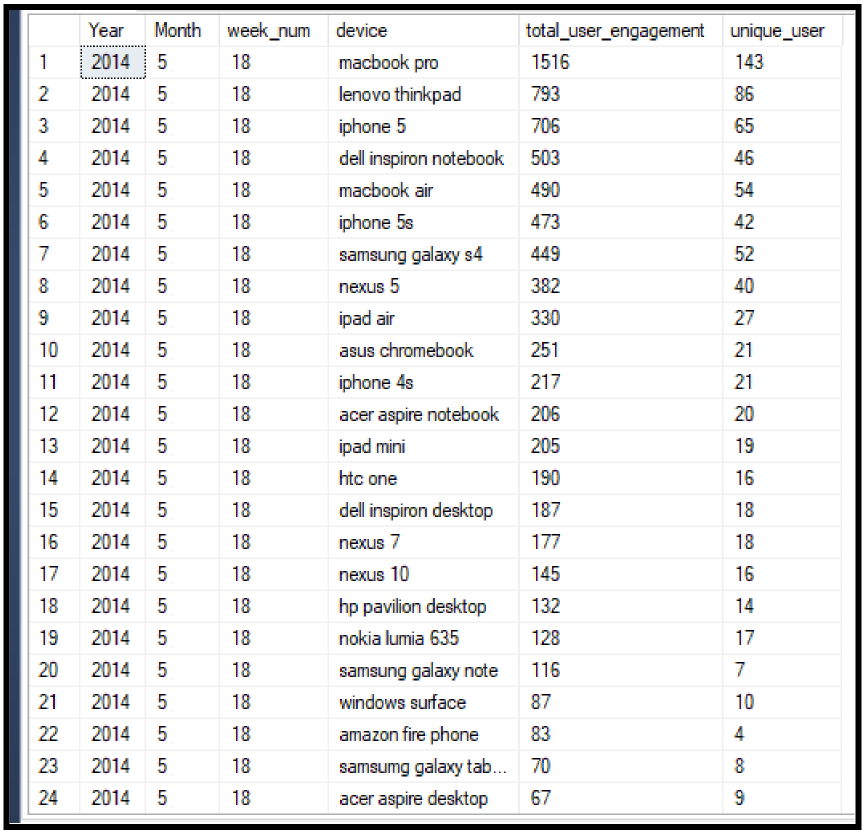
```
select(
        (
        (select COUNT(*) as 'User Retained within Week' from(
        select A.user_id, DATEADD(DAY,7,convert(date,created_at)) as '1 week later',
        min(convert(date, occurred_at)) as occur_date
        from users A
        right join  events B
        on A.user_id=B.user_id
        where event_type like 'engagement'
        group by A.user_id, DATEADD(DAY,7,convert(date,created_at)) ,B.user_id
        having  DATEADD(DAY,7,convert(date,created_at)) < min(convert(date, occurred_at))
        --order by A.user_id
        )temp
        )*100.0)
        / (select COUNT(*)as 'Total Users' from users )
) as '% user Retained Within week'
```



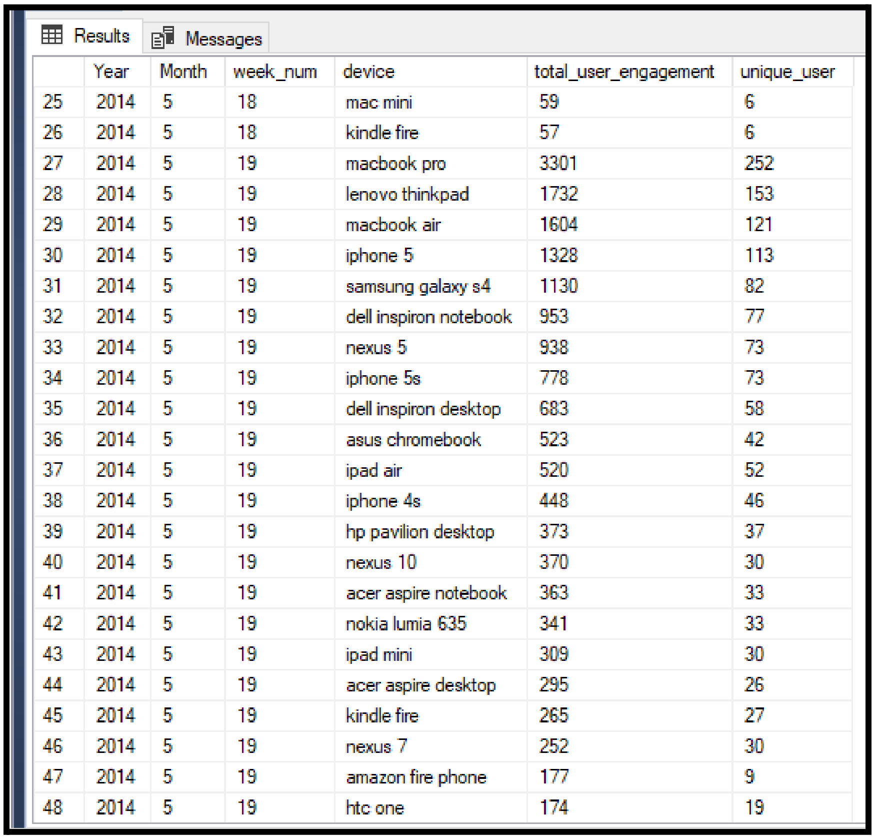
| | % user Retained Within week |
|---|---|
| 1 | 12.420014685828 |

4. **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?

```
select   YEAR(occurred_at) as 'Year',Month(occurred_at) as
'Month',DATEPART(WEEK,occurred_at) as week_num,device,
COUNT(user_id)as 'total_user',
COUNT(distinct user_id)as 'unique_user'
from events
where event_type='engagement'
group by Year(occurred_at),month(occurred_at),DATEPART(WEEK,occurred_at), device
order by week_num,total_user desc
```

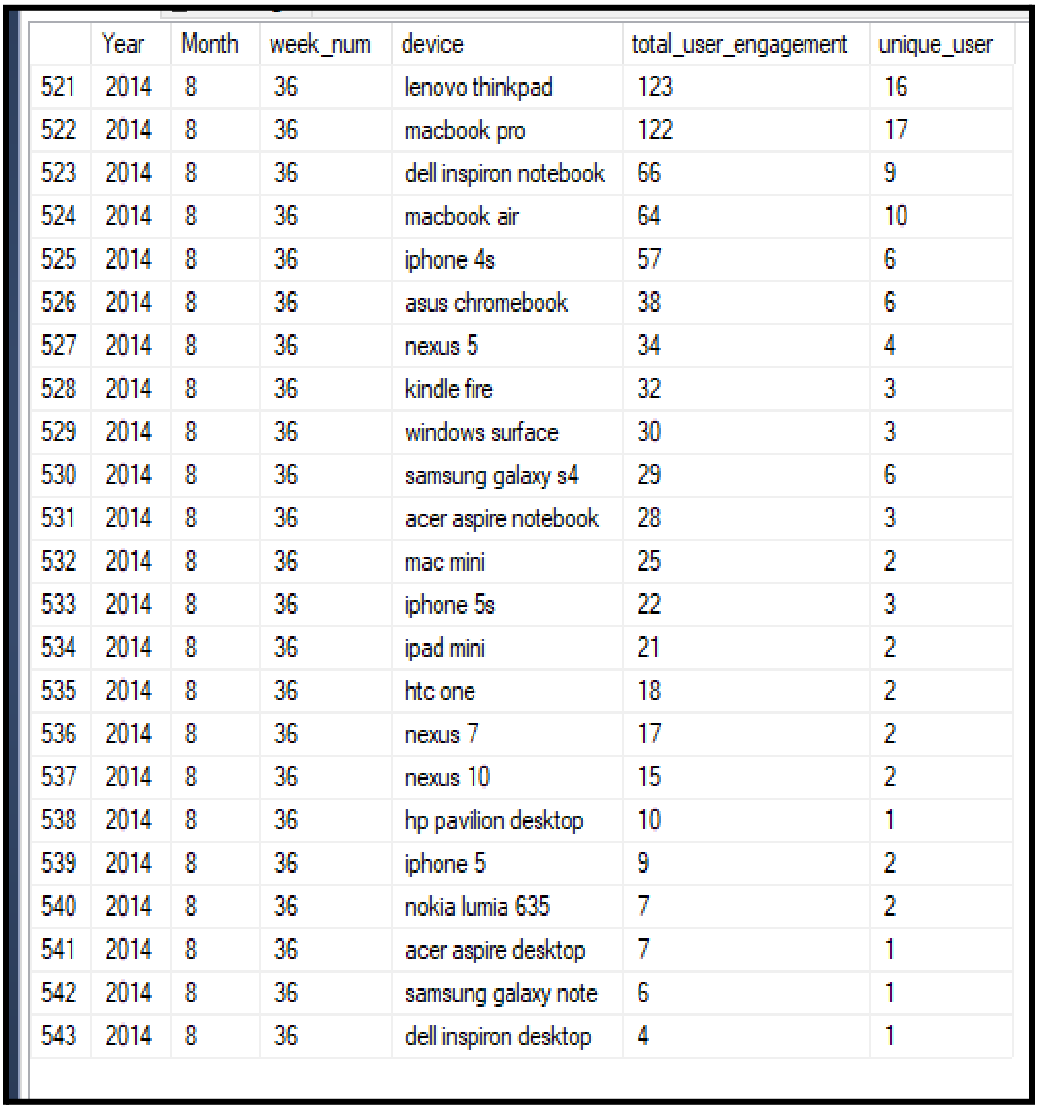


| | Year | Month | week_num | device | total_user_engagement | unique_user |
|---|---|---|---|---|---|---|
| 1 | 2014 | 5 | 18 | macbook pro | 1516 | 143 |
| 2 | 2014 | 5 | 18 | lenovo thinkpad | 793 | 86 |
| 3 | 2014 | 5 | 18 | iphone 5 | 706 | 65 |
| 4 | 2014 | 5 | 18 | dell inspiron notebook | 503 | 46 |
| 5 | 2014 | 5 | 18 | macbook air | 490 | 54 |
| 6 | 2014 | 5 | 18 | iphone 5s | 473 | 42 |
| 7 | 2014 | 5 | 18 | samsung galaxy s4 | 449 | 52 |
| 8 | 2014 | 5 | 18 | nexus 5 | 382 | 40 |
| 9 | 2014 | 5 | 18 | ipad air | 330 | 27 |
| 10 | 2014 | 5 | 18 | asus chromebook | 251 | 21 |
| 11 | 2014 | 5 | 18 | iphone 4s | 217 | 21 |
| 12 | 2014 | 5 | 18 | acer aspire notebook | 206 | 20 |
| 13 | 2014 | 5 | 18 | ipad mini | 205 | 19 |
| 14 | 2014 | 5 | 18 | htc one | 190 | 16 |
| 15 | 2014 | 5 | 18 | dell inspiron desktop | 187 | 18 |
| 16 | 2014 | 5 | 18 | nexus 7 | 177 | 18 |
| 17 | 2014 | 5 | 18 | nexus 10 | 145 | 16 |
| 18 | 2014 | 5 | 18 | hp pavilion desktop | 132 | 14 |
| 19 | 2014 | 5 | 18 | nokia lumia 635 | 128 | 17 |
| 20 | 2014 | 5 | 18 | samsung galaxy note | 116 | 7 |
| 21 | 2014 | 5 | 18 | windows surface | 87 | 10 |
| 22 | 2014 | 5 | 18 | amazon fire phone | 83 | 4 |
| 23 | 2014 | 5 | 18 | samsumg galaxy tab... | 70 | 8 |
| 24 | 2014 | 5 | 18 | acer aspire desktop | 67 | 9 |



| | Year | Month | week_num | device | total_user_engagement | unique_user |
|---|---|---|---|---|---|---|
| 25 | 2014 | 5 | 18 | mac mini | 59 | 6 |
| 26 | 2014 | 5 | 18 | kindle fire | 57 | 6 |
| 27 | 2014 | 5 | 19 | macbook pro | 3301 | 252 |
| 28 | 2014 | 5 | 19 | lenovo thinkpad | 1732 | 153 |
| 29 | 2014 | 5 | 19 | macbook air | 1604 | 121 |
| 30 | 2014 | 5 | 19 | iphone 5 | 1328 | 113 |
| 31 | 2014 | 5 | 19 | samsung galaxy s4 | 1130 | 82 |
| 32 | 2014 | 5 | 19 | dell inspiron notebook | 953 | 77 |
| 33 | 2014 | 5 | 19 | nexus 5 | 938 | 73 |
| 34 | 2014 | 5 | 19 | iphone 5s | 778 | 73 |
| 35 | 2014 | 5 | 19 | dell inspiron desktop | 683 | 58 |
| 36 | 2014 | 5 | 19 | asus chromebook | 523 | 42 |
| 37 | 2014 | 5 | 19 | ipad air | 520 | 52 |
| 38 | 2014 | 5 | 19 | iphone 4s | 448 | 46 |
| 39 | 2014 | 5 | 19 | hp pavilion desktop | 373 | 37 |
| 40 | 2014 | 5 | 19 | nexus 10 | 370 | 30 |
| 41 | 2014 | 5 | 19 | acer aspire notebook | 363 | 33 |
| 42 | 2014 | 5 | 19 | nokia lumia 635 | 341 | 33 |
| 43 | 2014 | 5 | 19 | ipad mini | 309 | 30 |
| 44 | 2014 | 5 | 19 | acer aspire desktop | 295 | 26 |
| 45 | 2014 | 5 | 19 | kindle fire | 265 | 27 |
| 46 | 2014 | 5 | 19 | nexus 7 | 252 | 30 |
| 47 | 2014 | 5 | 19 | amazon fire phone | 177 | 9 |
| 48 | 2014 | 5 | 19 | htc one | 174 | 19 |

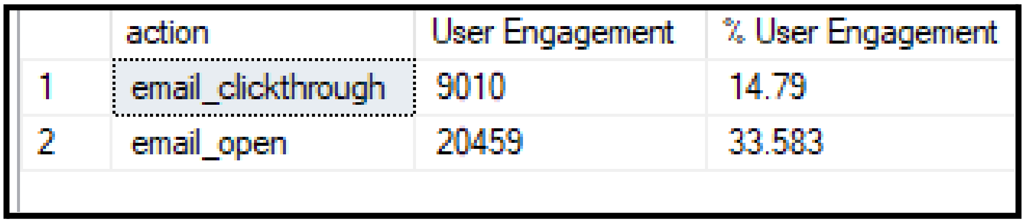
(...continuing to 543 rows. The last window picture is attached below)



| | Year | Month | week_num | device | total_user_engagement | unique_user |
|---|---|---|---|---|---|---|
| 521 | 2014 | 8 | 36 | lenovo thinkpad | 123 | 16 |
| 522 | 2014 | 8 | 36 | macbook pro | 122 | 17 |
| 523 | 2014 | 8 | 36 | dell inspiron notebook | 66 | 9 |
| 524 | 2014 | 8 | 36 | macbook air | 64 | 10 |
| 525 | 2014 | 8 | 36 | iphone 4s | 57 | 6 |
| 526 | 2014 | 8 | 36 | asus chromebook | 38 | 6 |
| 527 | 2014 | 8 | 36 | nexus 5 | 34 | 4 |
| 528 | 2014 | 8 | 36 | kindle fire | 32 | 3 |
| 529 | 2014 | 8 | 36 | windows surface | 30 | 3 |
| 530 | 2014 | 8 | 36 | samsung galaxy s4 | 29 | 6 |
| 531 | 2014 | 8 | 36 | acer aspire notebook | 28 | 3 |
| 532 | 2014 | 8 | 36 | mac mini | 25 | 2 |
| 533 | 2014 | 8 | 36 | iphone 5s | 22 | 3 |
| 534 | 2014 | 8 | 36 | ipad mini | 21 | 2 |
| 535 | 2014 | 8 | 36 | htc one | 18 | 2 |
| 536 | 2014 | 8 | 36 | nexus 7 | 17 | 2 |
| 537 | 2014 | 8 | 36 | nexus 10 | 15 | 2 |
| 538 | 2014 | 8 | 36 | hp pavilion desktop | 10 | 1 |
| 539 | 2014 | 8 | 36 | iphone 5 | 9 | 2 |
| 540 | 2014 | 8 | 36 | nokia lumia 635 | 7 | 2 |
| 541 | 2014 | 8 | 36 | acer aspire desktop | 7 | 1 |
| 542 | 2014 | 8 | 36 | samsung galaxy note | 6 | 1 |
| 543 | 2014 | 8 | 36 | dell inspiron desktop | 4 | 1 |

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

```
select action,Email_cat as 'User Engagement',
round( cast(((Email_cat*100.0)/
(select COUNT( action ) from email_events where action in
('sent_reengagement_email','sent_weekly_digest')) )as varchar(20)),3) as '% User
Engagement'
from (
        select action,COUNT(action)as Email_cat  from email_events
        where action in ('email_clickthrough','email_open')
        group by action
)temp
group by Email_cat,action;
```



|  | action | User Engagement | % User Engagement |
|---|---|---|---|
| 1 | email_clickthrough | 9010 | 14.79 |
| 2 | email_open | 20459 | 33.583 |

**Result:**
By completing the project, I am feeling a lot more confident in my SQL knowledge. It really helped me to brush up on my concepts related to Window Functions, Sub-queries, Joins, Views, and Aggregate functions.
It also helped me to understand how much important Operation Analytics and Metrics insights have for any company.

**Name: Umesh Sugara**
**Task: 3**