**Operation analytics and investigating metrics spike**

**❖ Project Description:**

**➢ The given project consists of 2 case studies:-**

* **First is regarding Operation Analytics where job data is provided and the number of jobs reviewed, a 7-day rolling average of throughput, percentage share of language used, and duplicates are found.**
* **Second is Investigating Metric Spikes where user engagement, user growth, weekly retention, and email engagement are determined.**
* **The following information is found with the help of SQL queries.**

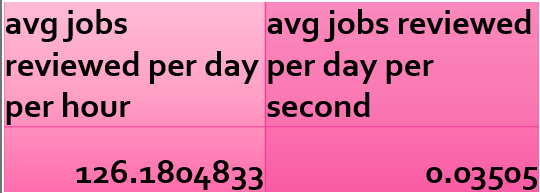
**❖ Approach:-**

**The required information was determined via SQL queries where the database was created first in SQL and moreover for the second case study due to the size of the data excel was used to make charts for better visualization.**

**operation analytics:-**

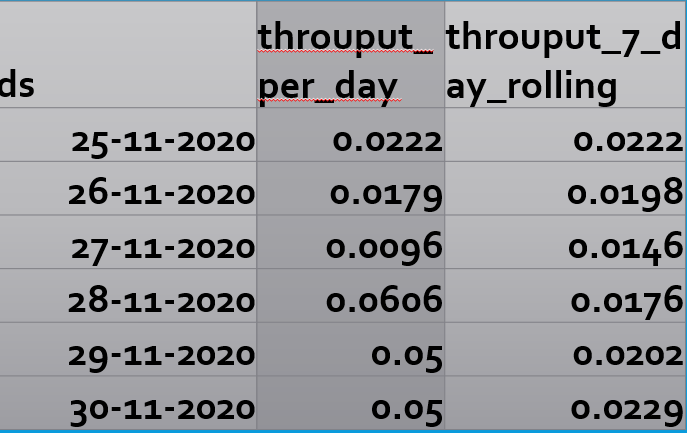
**1.Number of jobs reviewed:-**

**select avg(t) as 'avg jobs reviewed per day per hour’,avg(p) as 'avg jobs reviewed per day per second’From(select ds,((count(job\_id)\*3600)/sum(time\_spent)) as t,((count(job\_id))/sum(time\_spent)) as p from job\_data where month(ds)=11 group by ds) a;**

****

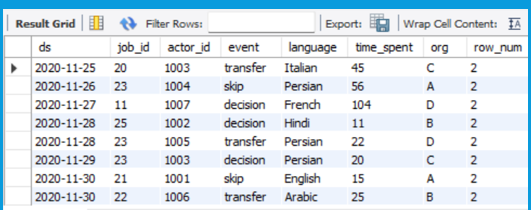
**2.Throughput:-**

**Select ds, c/t as throuput\_per\_day, c7/s7 as throuput\_7\_day\_rolling From (select ds, count(job\_id) as c,sum(time\_spent) as t, count(job\_id) over(order by ds rows between 6 preceding and current row) as c7,sum(time\_spent) over(order by ds rows between 6 preceding and current row) as s7 from job\_data where month(ds)=11 group by ds) a;**

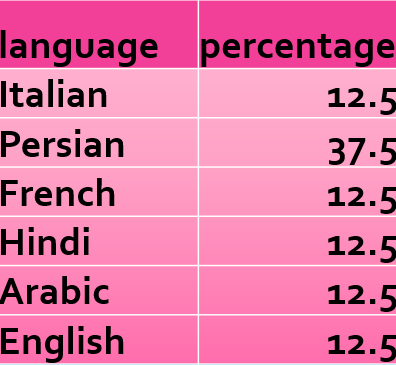
****

**3.Finding duplicates:-**

**select \* from(select \*,row\_number() over(partition by ds,actor\_id,job\_id) as row\_num from job\_data) a where row\_num>1;**

****

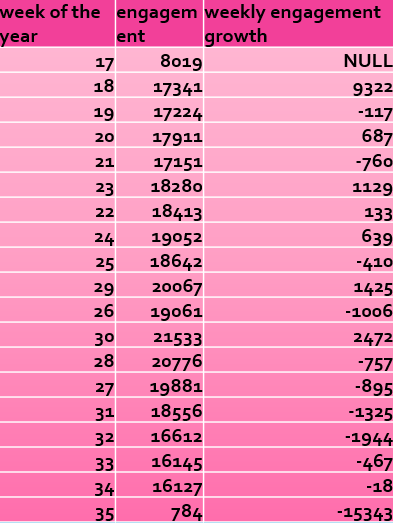
**4.Percentage share of language used in last 30 days:-**

**select language , num\_jobs, OperatiOn analytics and investigating metrics spike 100.0\*num\_jobs/total\_jobs as pct\_share\_jobs From ( Select language,count(distinct job\_id)as num\_jobs From job\_data Group by language)a Cross join (select count(distinct job\_id) as total\_jobs From job\_data)b; **

* **Investigating Metric Spike Approach:-**

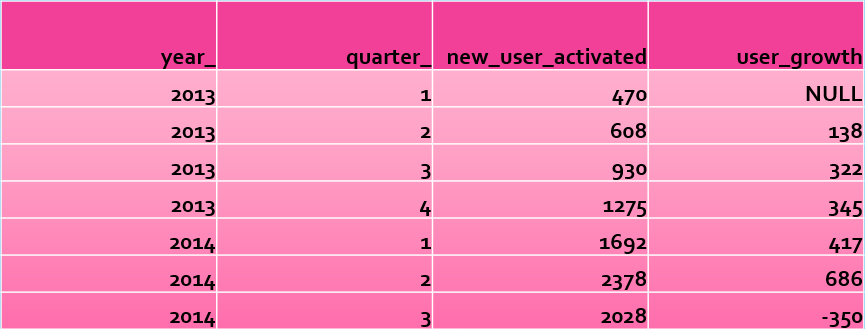
**1.Weekly user engagement:-**

**select \*, engagement-lag(engagement) over(partition by'week of the year') as 'weekly engagement growth’ From (select week(occurred\_at) as 'week of the year’, count(event\_name) as 'engagement’ from events where event\_type!='signup\_flow’ group by week(occurred\_at))a;**

****

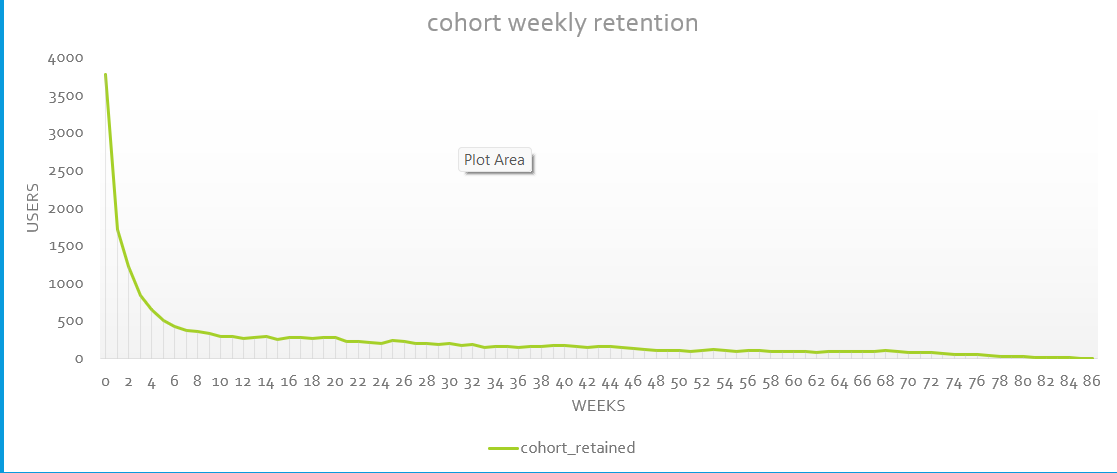
**2.User growth:-**

**select \*, new\_user\_activated-lag(new\_user\_activated) over( order by year\_,quarter\_ ) as user\_growth from(select year(created\_at) as year\_,quarter(created\_at) as quarter\_,count(user\_id) as new\_user\_activated from users where activated\_at is not null and state='active’ group by 1,2)a ;**

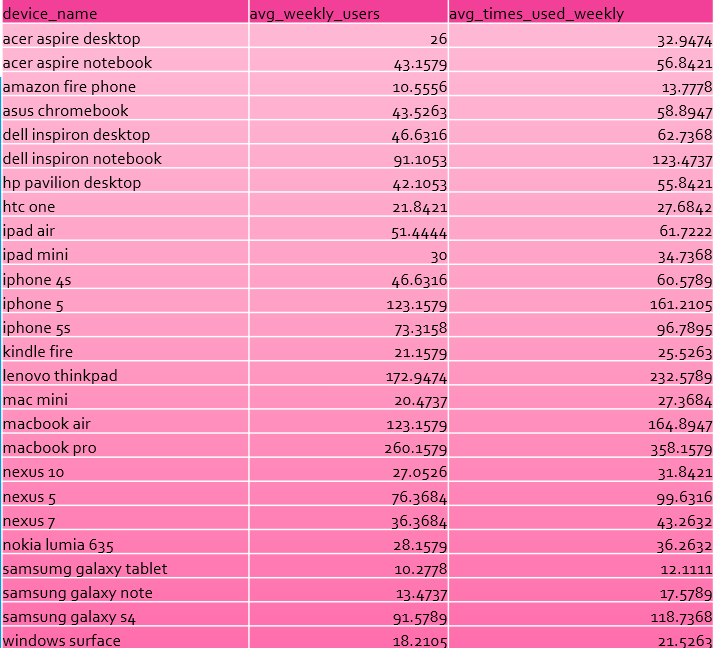
****

**3.Weekly retention cohort analysis:-**

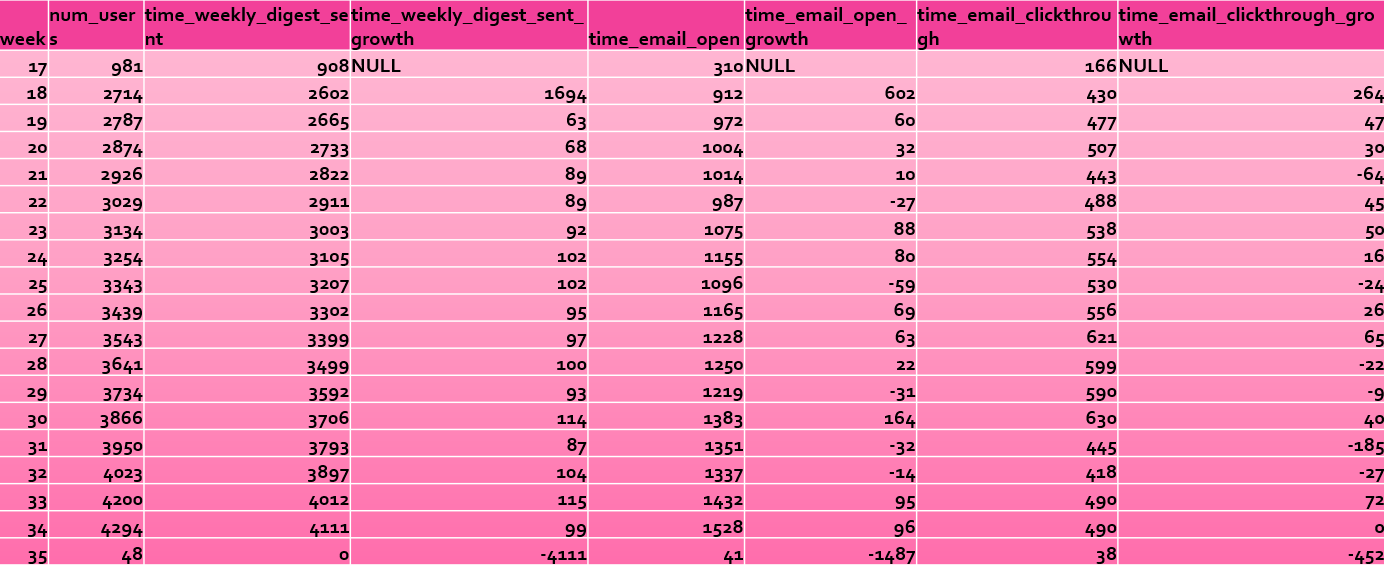
**Select week\_period, first\_value(cohort\_retained) over (order by week\_period) as cohort\_size, cohort\_retained, cohort\_retained / first\_value(cohort\_retained) over (order by week\_period) as pct\_retained From (select timestampdiff(week,a.activated\_at,b.occurred\_at) as week\_period, count(distinct a.user\_id) as cohort\_retained From (select user\_id, activated\_at from users where state='active'group by 1) a inner join (select user\_id,occurred\_at from events )b on a.user\_id=b.user\_id group by 1) c;**

****

**4.Weekly engagement per device:-**

**Select device\_name, avg(num\_users\_using\_device) as avg\_weekly\_users, avg(times\_device\_use\_current\_week) as avg\_times\_used\_weekly From (select week(occurred\_at) as week, device as device\_name , count(distinct user\_id) as num\_users\_using\_device, count(device) as times\_device\_use\_current\_week from events where event\_name='login’ group by 1,2 OperatiOn analytics and investigating metrics spike order by 1) a group by 1; **

**5.E-mail engagement metric:-**

**Select week, num\_users, time\_weekly\_digest\_sent, time\_weekly\_digest\_sent-lag(time\_weekly\_digest\_sent) over(order by week) as time\_weekly\_digest\_sent\_growth, time\_email\_open,time\_email\_open-lag(time\_email\_open) over(order by week) as time\_email\_open\_growth, time\_email\_clickthrough,time\_email\_clickthrough-lag(time\_email\_clickthrough) over(order by week) as time\_email\_clickthrough\_growth From (select week(occurred\_at)as week, count(distinct user\_id) as num\_users, sum(if(action='sent\_weekly\_digest',1,0)) as time\_weekly\_digest\_sent, sum(if(action='email\_open',1,0)) as time\_email\_open, sum(if(action='email\_clickthrough',1,0)) as time\_email\_clickthrough from email group by 1 order by 1) a; **

* **Insights:-**
* **Case Study 1 (job Data):-**

**▪ In November 2020, the number of different jobs reviewed per hour per day was 83%.**

**▪ We use a 7-day moving average throughput as it provides an average for all days from day 1 to day 7, whereas the daily metric only provides an average for a particular day itself.**

**▪ Persian has the highest rate (37.5%). Operation analytics and investigating metrics spike**

**▪ If we split the job\_id, there are two duplicate rows. But if we look at all columns, each row is unique.**

* **Case Study 2 (investigated Metrics spike ):**

**▪ All data corresponding to Day 35 should be disregarded, as this is only the first day of the week.**

* **A total of 9381 active users from the first week of 2013 to the 35th week of 2014.**
* **Average retention rate after first week 72.5%**
* **MacBook users and Samsung galaxy tablet users contribute with more than one device per week.**
* **E-mail open rate is 1117, e-mail click-through rate is around 493. People use email services that are essential for the growth of the company.**
* **Tech stack used:-**
* **MySQL was used to run the queries.**
* **The language was selected because of comfort and experience in the same.**
* **MS Excel was used in the second case study for better visualization.**
* **As I am currently learning this tool, it was utilized to get more hands-on experience.**
* **Result:-**

**Really engaging project, the difficulty of the project makes it more fulfilling to execute. Learned a lot of new things like rolling over and get cohort retention analysis.**