**BY**

**Operation Analytics and Investigating Metric Spike**

**Rahul Patil**

**Project Description:**

* Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon.
* This kind of analysis is further used to predict the overall growth or decline of a company’s fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows.
* In this project we are going to analyze company operation and user engagement with product and services, growth of product and increase or decrease in user engagement which will help to find the area which required improvement which further help to generate more revenue.

**Approach:**

* To successfully carry out this project we are going to use **SIX STEP** of Data Analysis Process i.e (Ask, Prepare, Process, Analyze, Share, Act)
* Ask step include asking right set of question which justify goal and give motivation to carry out analysis
* We have following set of question (reasons) to justify goal of this project.
  + Calculate the number of jobs reviewed per hour per day for November 2020?
  + Calculate 7-day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?
  + Calculate the percentage share of each language in the last 30 days?
  + How will you display duplicates from the table?
  + Calculate the weekly user engagement?
  + Calculate the user growth for product?
  + Calculate the weekly retention of users-sign up cohort?
  + Calculate the weekly engagement per device?
  + Calculate the email engagement metrics?
* Prepare: We have users and event data store in RDBMS in different table like users, events, email\_events, job\_data.
* This step includes selecting right data, tools, data source to make project successful
* Process: Data we have is already in process format i.e data store in RDBMS is process, clean and  useful data.
* Analyze: We are using SQL to analyze data to draw insight.
* Share: we are showing data obtain from analysis in the form of row and column as well as chart wherever required for better and easy understanding.
* Act: Step include taking decision based on insight opt from this project.

**Tech-Stack Used:**

* Data is stored in RDBMS (Relational Database Management System) in different table
* We could use jupyter notebook and programming language to carry out this project. Which requires programming skills as well further extra step such as connecting to database, loading data from DBMS to notebook.
* While RDBMS come with inbuilt data query language called SQL (Structure Query Language) which can successfully carry out entire analysis for this project.
* Tech we are going to use for this project is PostgreSQL which is RDBMS and excel to visualize output using charts and graph.

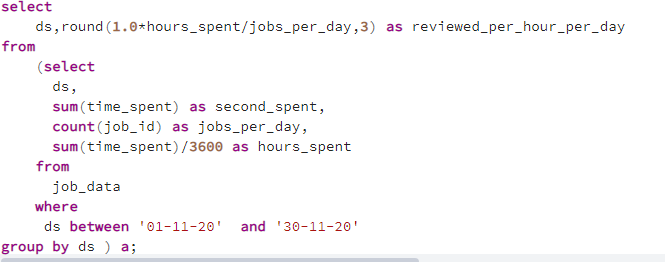
**Insights:**

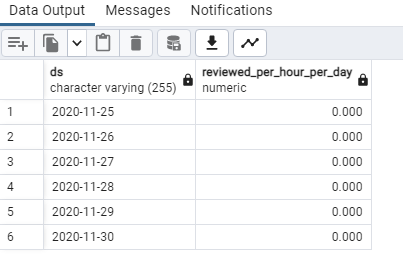
**Case Study 1 (Job Data):**

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

select count(\*) from job\_data

Total number of record are very few it will difficult to do analysis and take decision when data is limited.

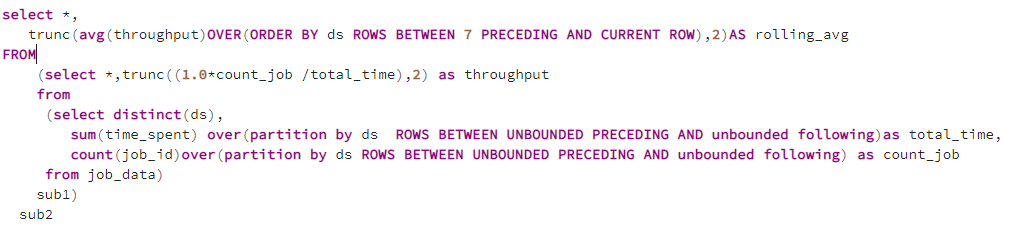


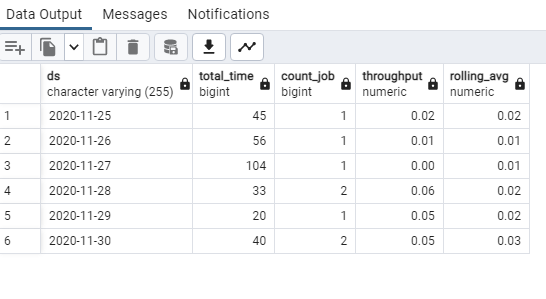


**most job review in seconds, hardly 1 minute spend that’s why jobs reviewed per hour per day is zero**

2.Calculate 7-day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

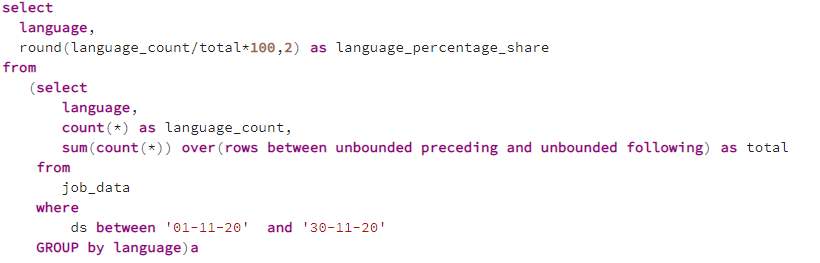
Throughput: It is the no. of events happening per second

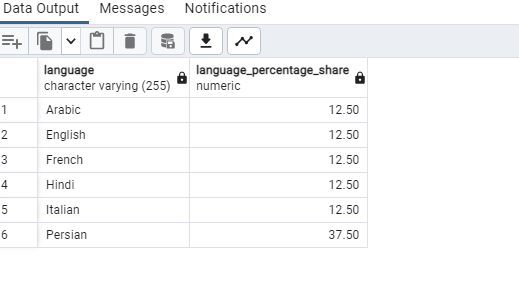




Data is limited so we can’t conclude whether daily metric or 7-day rolling average should we prefer, we required more data to make conclusion.

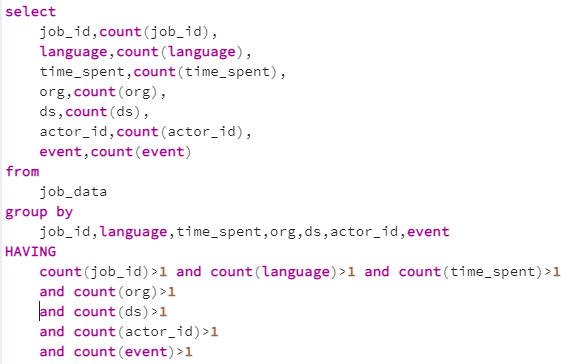
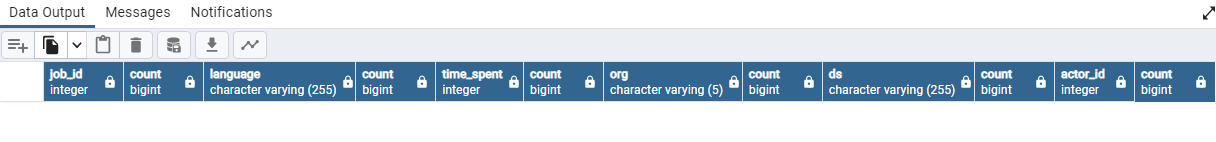
3.Calculate the percentage share of each language in the last 30 days?





**Each language has equal weightage of 12.50% except Persian language having higher weightage of 37.50%**

4.How will you display duplicates from the table?

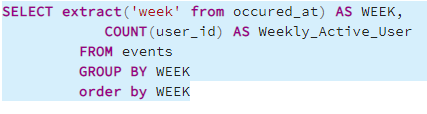


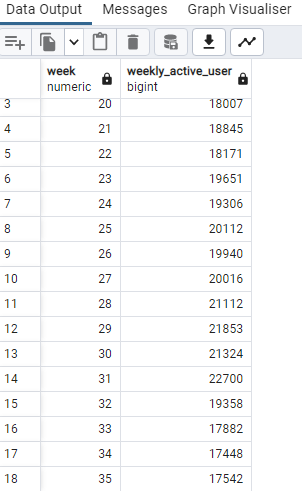
**No duplicate records found for given data**

**Insights:**

**Case Study 2 (Investigating metric spike)**

1.Calculate the weekly user engagement?

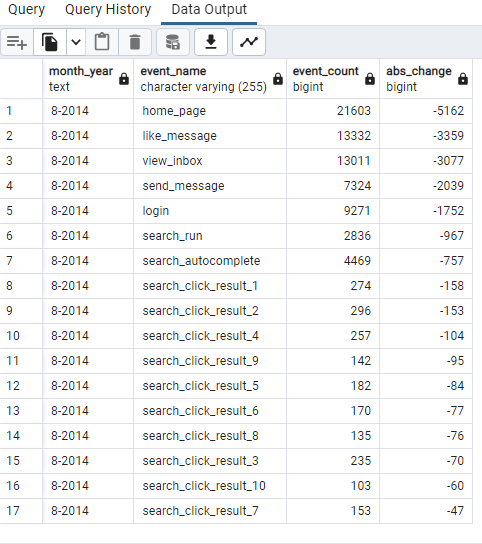




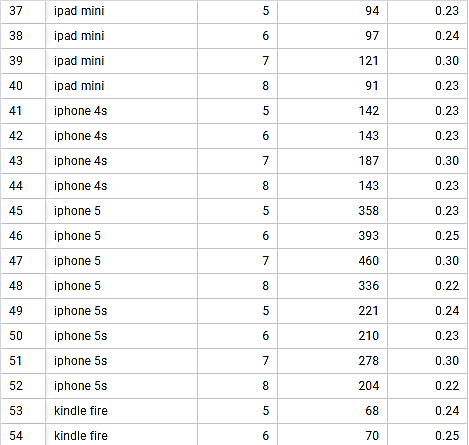
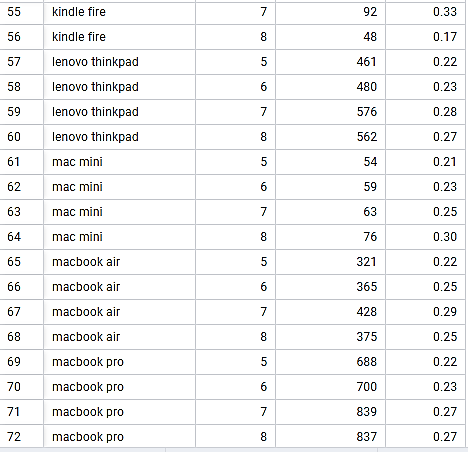
**From above query its clearly visible that there is decline in user engagement from week 32.**

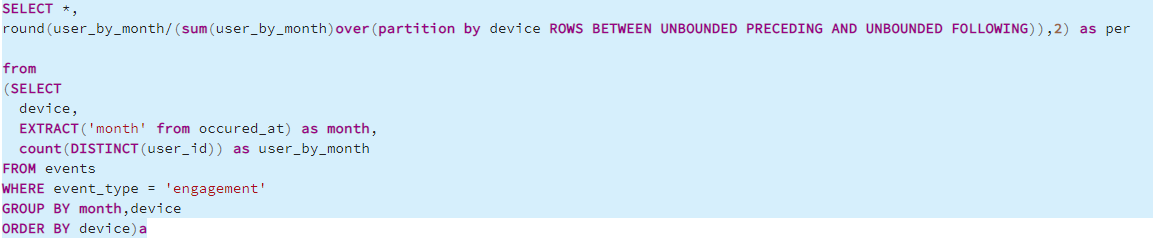
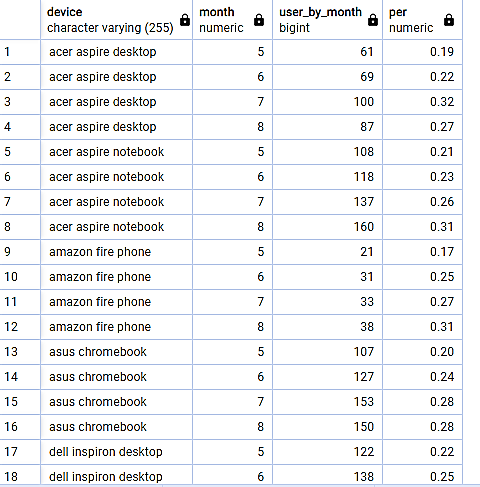
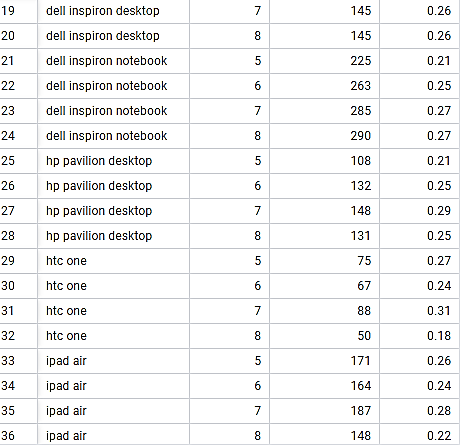
**The next thing I wanted to do was count the number of occurrences of each ‘engagement’ event month over month as decline in user in between July and August**

**to see if there’s any significant event(s) that caused the dip after 32 week**

****

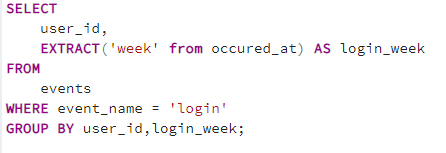
**The dip in engagement was largely attributed to home\_page, like\_message\_ view\_inbox, send\_message, and login. It seems like the drop in all of these events are simply related to the fact that users are loging in less.**

2.Calculate the user growth for product?

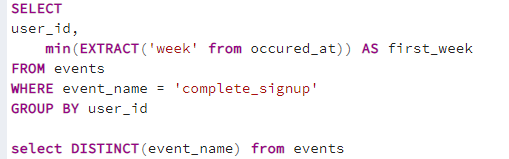


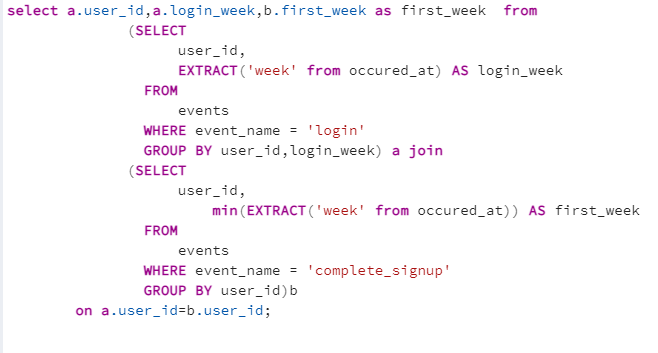
**Number of user per device is decline in month of August mostly for tablet and mobiles.**

3.Calculate the weekly retention of users-sign up cohort?

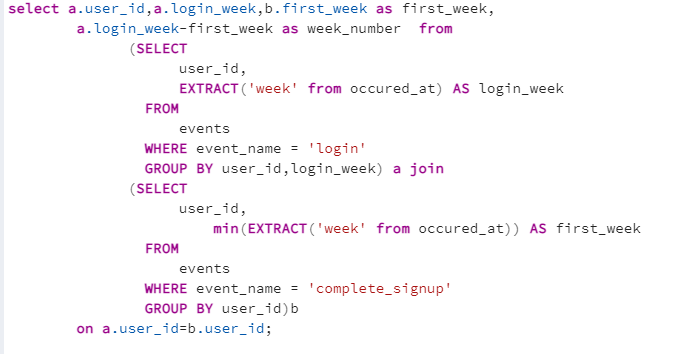
**first calculate user login by week**

**calculate user who completed sign up we use this user whether this user login on weekly basis or not**

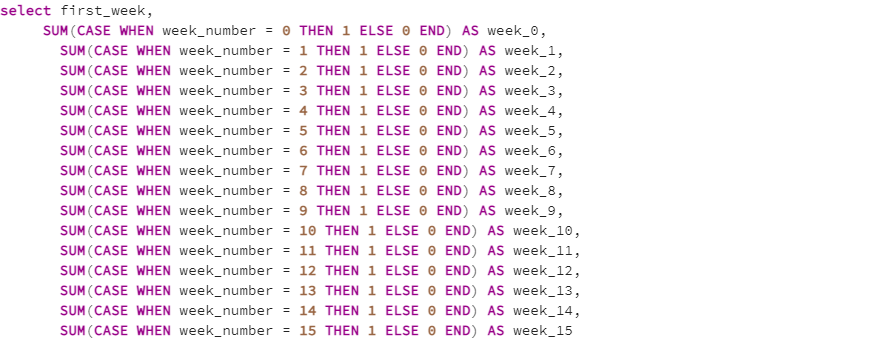
****

**we get login\_week and first\_week side by side for each user using the query below, with an INNER JOIN**

**calculate the difference between login\_week and first\_week to calculate week\_number (number of week)**

****

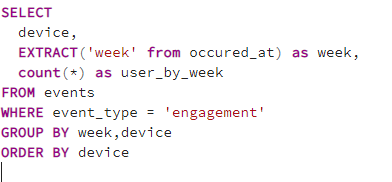
**Combining all above query**

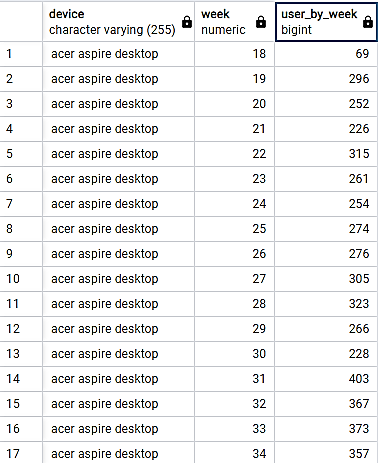
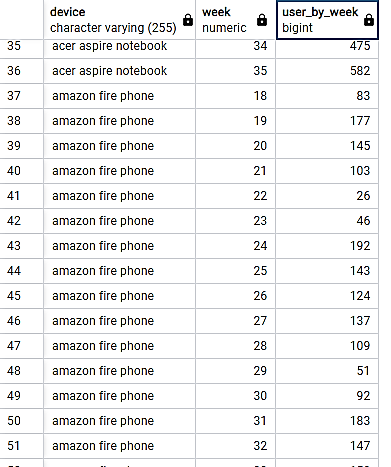
****

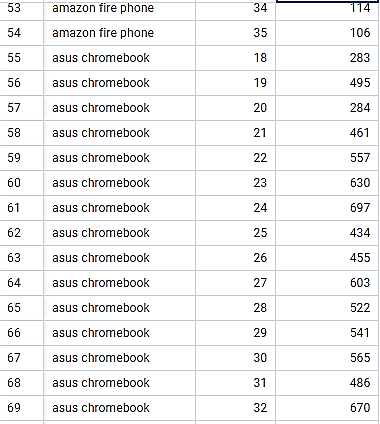
****

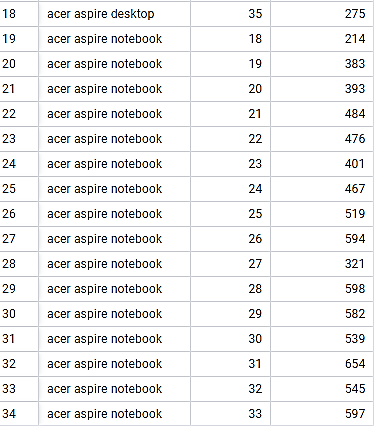
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **first\_week** | **week\_0** | **week\_1** | **week\_2** | **week\_3** | **week\_4** | **week\_5** | **week\_6** | **week\_7** | **week\_8** | **week\_9** | **week\_10** | **week\_11** | **week\_12** | **week\_13** | **week\_14** | **week\_15** |
| 18 | 81 | 64 | 27 | 19 | 15 | 19 | 13 | 10 | 7 | 9 | 9 | 9 | 8 | 9 | 7 | 5 |
| 19 | 160 | 104 | 67 | 45 | 34 | 22 | 18 | 24 | 11 | 17 | 12 | 11 | 14 | 10 | 9 | 9 |
| 20 | 186 | 144 | 77 | 64 | 40 | 26 | 20 | 19 | 23 | 18 | 15 | 15 | 13 | 11 | 8 | 10 |
| 21 | 177 | 121 | 79 | 50 | 40 | 28 | 21 | 32 | 23 | 23 | 23 | 18 | 18 | 10 | 9 | 0 |
| 22 | 186 | 117 | 78 | 53 | 34 | 24 | 33 | 30 | 18 | 18 | 13 | 14 | 15 | 9 | 0 | 0 |
| 23 | 197 | 133 | 83 | 60 | 49 | 42 | 29 | 25 | 26 | 17 | 17 | 12 | 7 | 0 | 0 | 0 |
| 24 | 198 | 146 | 85 | 56 | 44 | 41 | 34 | 28 | 22 | 21 | 14 | 10 | 0 | 0 | 0 | 0 |
| 25 | 222 | 135 | 89 | 57 | 41 | 31 | 28 | 25 | 15 | 17 | 12 | 0 | 0 | 0 | 0 | 0 |
| 26 | 210 | 151 | 100 | 62 | 44 | 30 | 24 | 19 | 15 | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27 | 199 | 130 | 82 | 60 | 43 | 34 | 33 | 26 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28 | 223 | 152 | 95 | 83 | 52 | 39 | 26 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | 215 | 144 | 91 | 52 | 33 | 19 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30 | 228 | 155 | 82 | 59 | 40 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 31 | 234 | 154 | 94 | 64 | 47 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 32 | 189 | 126 | 69 | 48 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 33 | 250 | 163 | 82 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 34 | 259 | 173 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 35 | 266 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

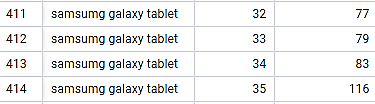
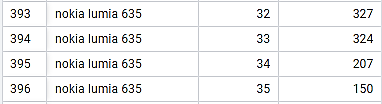
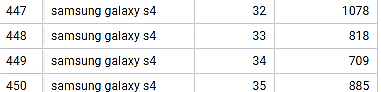
**From 2nd week user retention is almost half of the 1st week**

4.Calculate the weekly engagement per device?

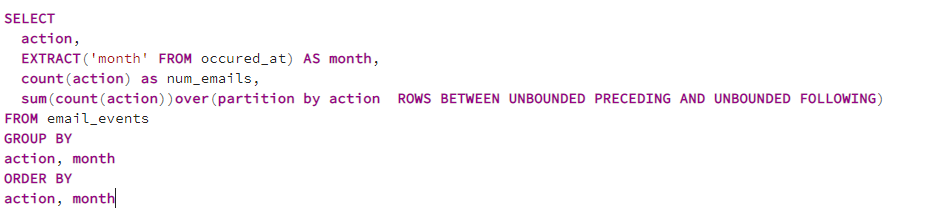
****

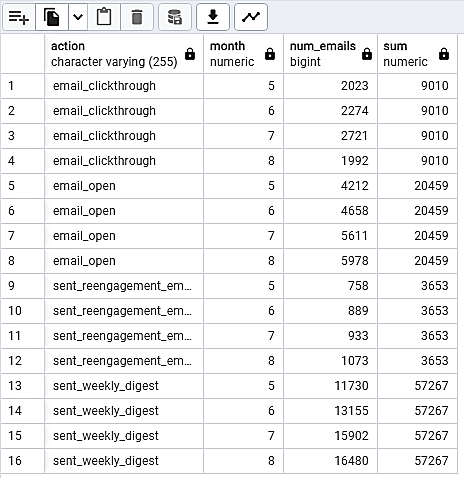
****

****

****

**Number of user per device per week decline from 32 Week for mobile and tablet**

5.Calculate the email engagement metrics?

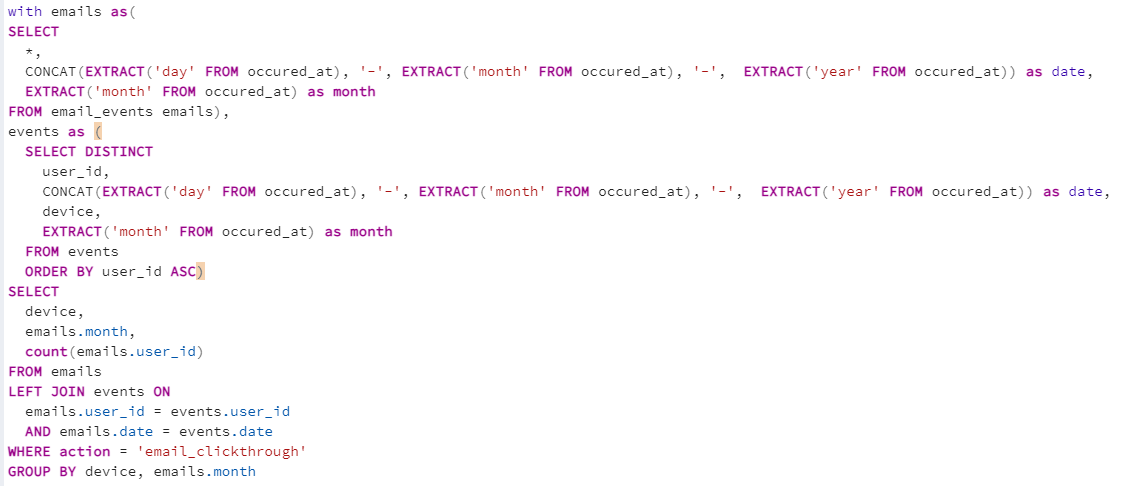


**there was a steady increase in the number of emails sent**

**(weekly digest and re-engagement) and emails opened**

**but but there was a decrease in clickthrough rates. From July to August,**

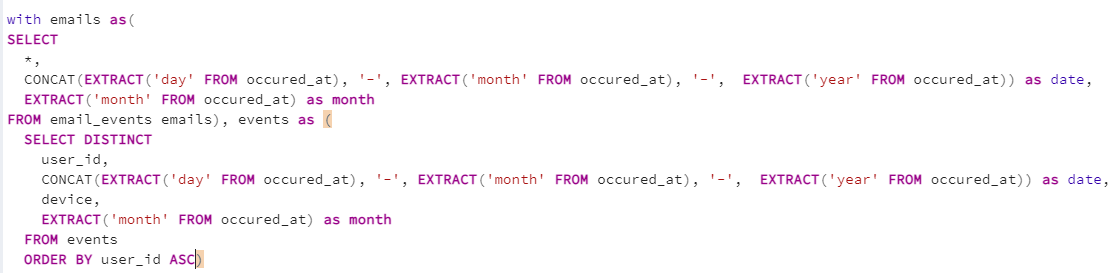
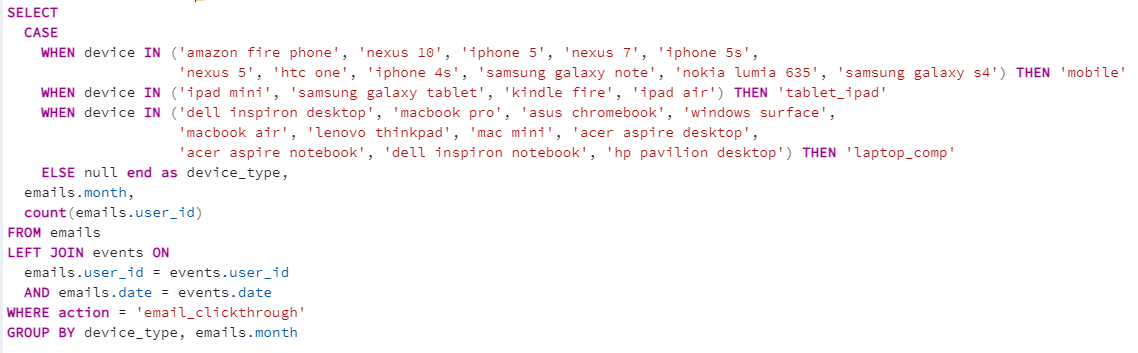
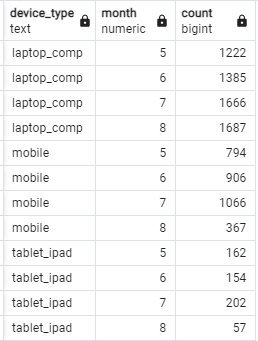
**there was a 6.5% increase in emails open but a 27% decline in clickthrough rates.**

****

****

**Using the query above, I noticed that the clickthrough rates on laptops and computers were stable from July to August,**

**but not the tablets and cellphones**

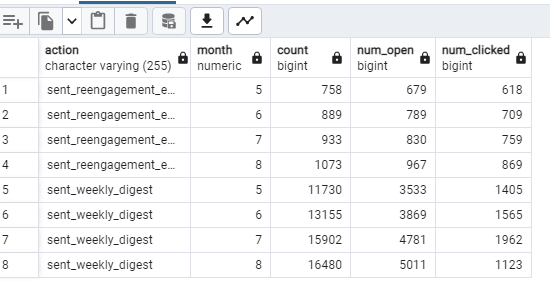
****

**it seems to be the case that the drop in clickthrough rates was attributed specifically to mobile devices and tablets.**

**lack of engagement is due to a decrease in email clickthrough rates from July to August. To gather more information,**

**Let’s see if we can narrow the problem even further by email type.**

****

****

**Result:**

* In this project we understand how to use SIX step process of Data analysis.
* Project help to understand how to use SQL joins as data store in different table for simple data analytics job.
* Project help to understand how to use date and window functions
* Key project insight are:
  + Due to lack of data first case is difficult to analyze still we are able to find out number of share for each language
  + Persian language has higher per share of 37.5% while other languages have 12.5%
  + The drop in engagement was mainly attributed to a drop in five engagement events (home\_page, like\_message\_ view\_inbox, send\_message, and login)
  + Decrease in events was caused by a reduction in total active users from August, as well as a decrease in engagement per user.
  + After I took a close look at the emails table and I noticed that there was a significant decrease in click-through rates from July to August even though there was an increase in the number of emails opened.
  + By segmenting the clickthrough rates by device type (mobile, tablet, laptop), I noticed that the drop in clickthrough rates was attributed to mobile and tablet devices.
  + The decline in click-through rates is attributed to the weekly digest email and not the re-engagement email.
  + It’s possible that there’s a technical problem, making it difficult for users to click the email or simply a UX problem, where the content and layout of the email are not enticing users to click. A good first step would be to see what changes have been made from July to August and working backward.