Operation Analytics & Investigating Metric Spike

Trainity-3

Description

- The given project consists of 2 case studies:-
- First is regarding Operation Analytics where job data is provided and number of jobs reviewed , 7day rolling average of throughput, percentage share of language used and duplicates are found out.
- Second is Investigating Metric Spike where user engagement, user growth, weekly retention, weekly engagement and email engagement is determined.
- The following information is found with the help of SQL queries.

Approach

- At first I imported the files that has been given into the MYSQL workbench.
- After making a Schema I started writing queries for the questions
- I imported the Tables into Tableau and made visualizations
- Also exported all the result tables for the provided question to include in the ppt presentation.

Tech-Stack Used

MYSQL WORKBENCH

- This tool is used to create the data base and store records. It is also used to carry out the required analysis by writing SQL queries.

Tableau Public

- This tool is used to create graphical representation of the results and to understand the result set better.

Case study 1

- **Below is the structure of the table with the definition of each column that I work on:**
- **Table-1:** job_data

1. **job_id:** unique identifier of jobs
2. **actor_id:** unique identifier of actor
3. **event:** decision/skip/transfer
4. **language:** language of the content
5. **time_spent:** time spent to review the job in seconds
6. **org:** organization of the actor
7. **ds:** date in the yyyy/mm/dd format. It is stored in the form of text and we use presto to run. no need for date function
8.

Number of jobs reviewed

- select avg(t) as 'jobs reviewed per hour' from (select ds,(count(job_id)*3600)/sum(time_spent)as t from job_data where month(ds)=11 group by ds)two;

jobs reviewed per hour
126.18048333

Throughput

- **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?

ds	throuput_per_day	throughput_7_rolling_avg
2020-11-25	0.0222	0.02220000
2020-11-26	0.0179	0.02005000
2020-11-27	0.0096	0.01656667
2020-11-28	0.0606	0.02757500
2020-11-29	0.0500	0.03206000
2020-11-30	0.0500	0.03505000

Percentage share of each language

- **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?
-

language	percentage
Italian	12.5000
Persian	37.5000
French	12.5000
Hindi	12.5000
Arabic	12.5000
English	12.5000

Duplicate rows

- **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?

NO DUPLICATE ROWS

CASE STUDY - 2

CASE STUDY - 2

Case Study 2 (Investigating metric spike)

The structure of the table with the definition of each column that I work on

- **Table-1:** users
  This table includes one row per user, with descriptive information about that user's account.
- **Table-2:** events
  This table includes one row per event, where an event is an action that a user has taken. These events include login events, messaging events, search events, events logged as users progress through a signup funnel, events around received emails.
- **Table-3:** email_events
  This table contains events specific to the sending of emails. It is similar in structure to the events table above.

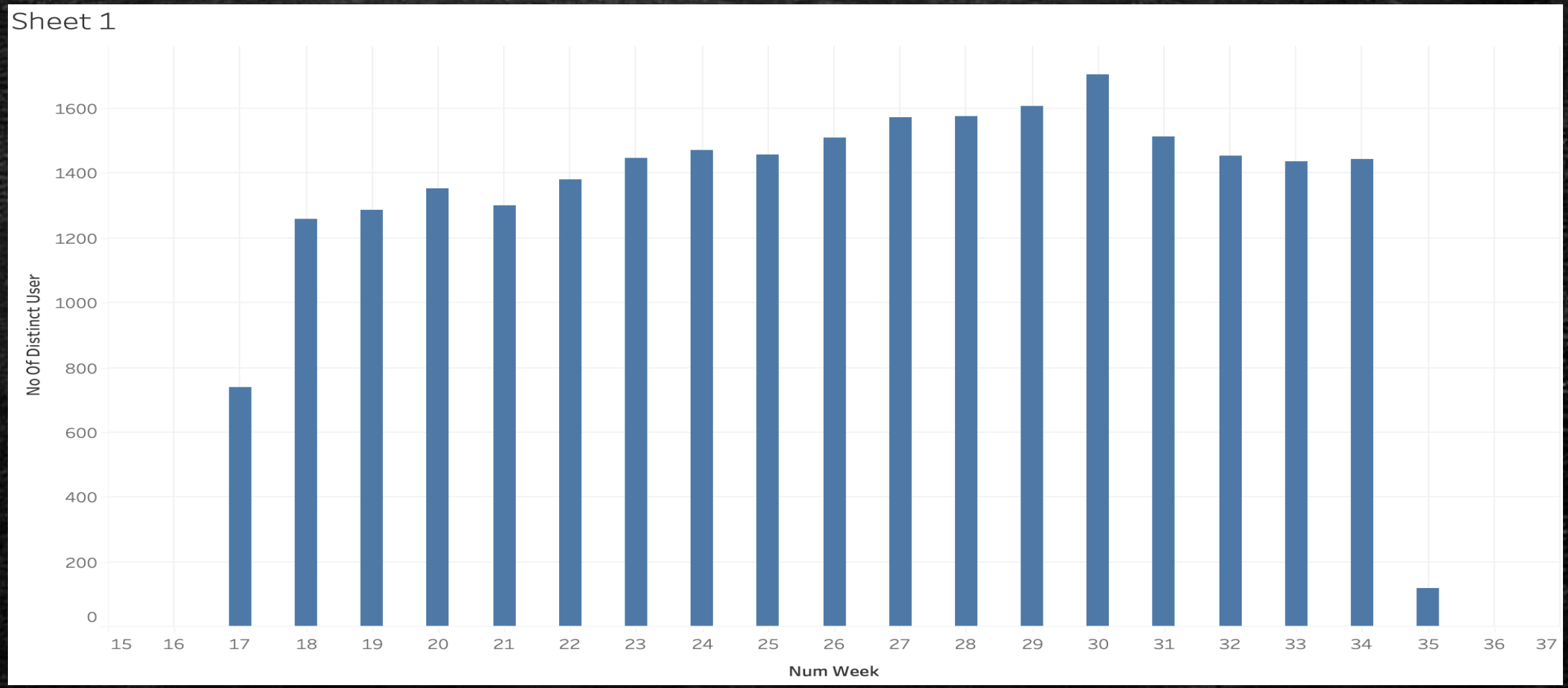
User Engagement

- To measure the activeness of a user. Measuring if the user finds quality in a product/service.
- Your task: Calculate the weekly user engagement?

num_week	no_of_distinct_user
30	1706
31	1514
32	1454
33	1438
34	1443
35	118

num_week	no_of_distinct_user
17	740
18	1260
19	1287
20	1351
21	1299
22	1381
23	1446
24	1471
25	1459
26	1509
27	1573
28	1577
29	1607

USER ENGAGEMENT



USER GROWTH

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

year	quarter	num_active_users	user_growth
2013	1	470	NULL
2013	2	608	138
2013	3	930	322
2013	4	1275	345
2014	1	1692	417
2014	2	2378	686
2014	3	2028	-350

Weekly Retention

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

per_week_retention	count
0	49
1	114

Above data is retention of people from 18th week to 19 week
0=people not retained
1=people are retained
Nearly 70 percent of people retained for next week

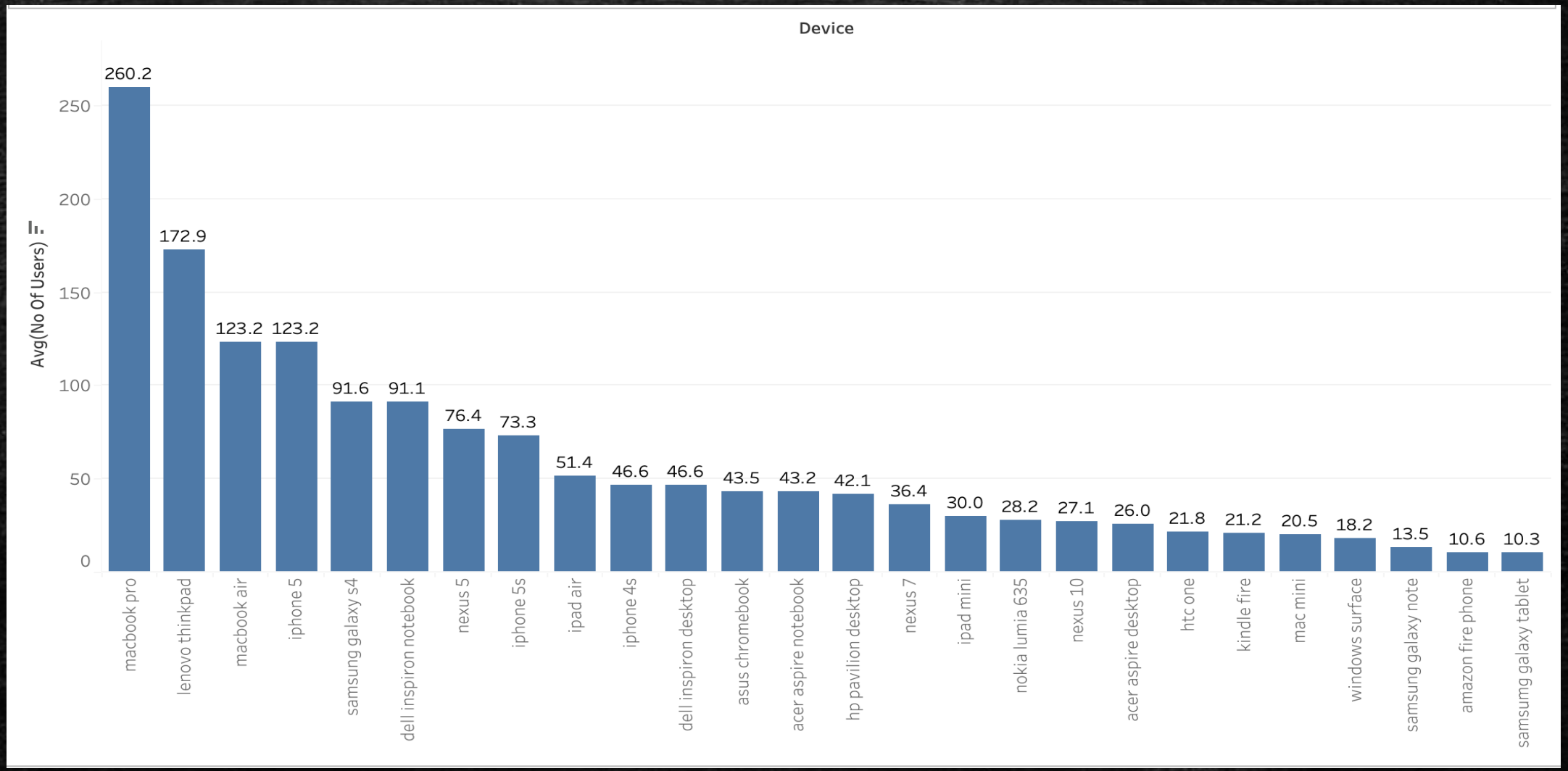
Weekly Engagement

- **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?

THIS IS THE AVERAGE OF ALL WEEKS

```
select device, avg(no_of_users) from
(select
extract(year from occurred_at) as year_num,
extract(week from occurred_at) as week_num,
device,
count(distinct user_id) as no_of_users
from events
where event_type = 'engagement'
group by 1,2,3
order by 3)a
group by a.device ;
```

WEEKLY ENGAGEMENT



Email Engagement

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

email_opening_rate	email_clicking_rate
33.58339	14.78989

RESULT

- I have no knowledge about Tableau prior to this now I know little bit of Tableau.
- This is the advance SQL project little challenging for me, I learned lot about SQL by this project
- I am waiting for upcoming projects I hope they are going to me more complex so I can test my limits.
- THANK YOU.