TRAINITY

Name:Attulaya
Email: attulaya@gmail.com

Operation Analytics and Investigating Metric Spike

Attulaya, Jr. Data Analyst
attulaya@gmail.com

CONTENTS

Project description

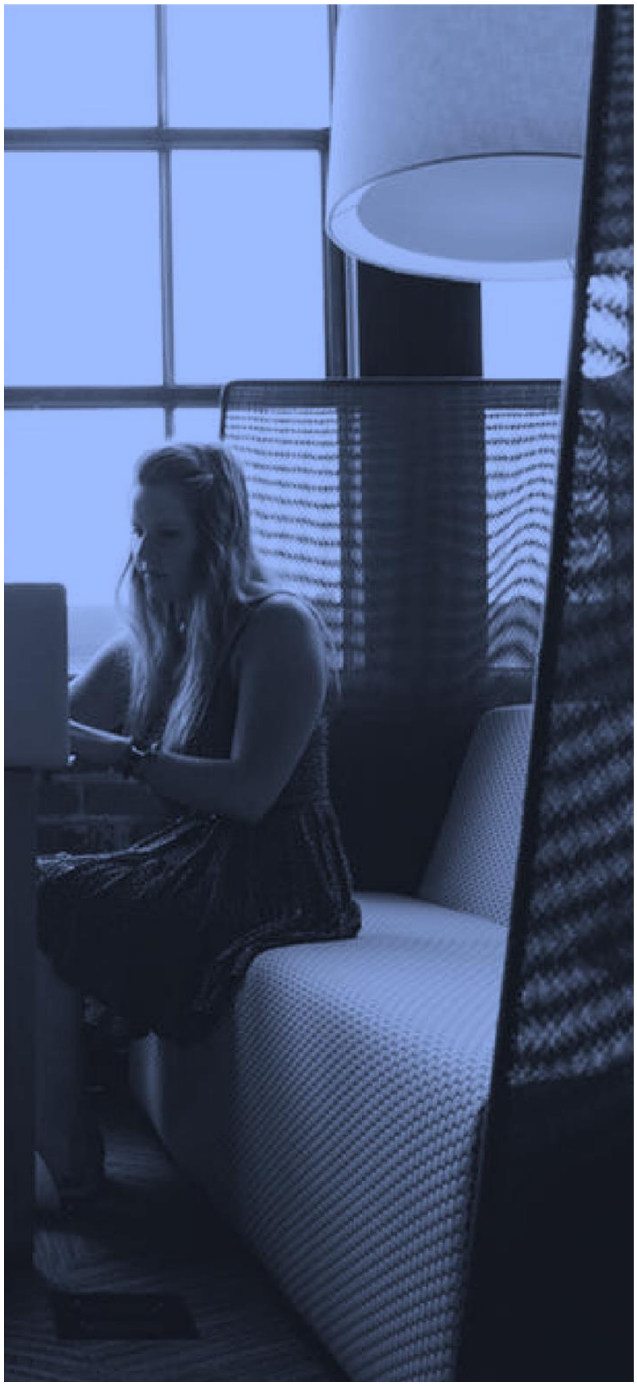
This project includes the implementation of performance measures for a large company. As a Data Manager, my job is to work with different teams to collect and analyze end-to-end performance-related data, predict growth or decline, and control metric increases. Through performance analysis, I provide insights that help teams make informed decisions and drive the company to success. The aim of the project is to improve the company's performance, reduce costs and increase overall efficiency. Working with diverse teams and providing valuable insights from a variety of data and charts, I help companies achieve their goals and solve problems before they become problems. The project aims to facilitate better teamwork and communication within the team, thus reducing silos and increasing organizational effectiveness. Overall, the project aims to ensure the company's performance and achieve its goals through performance evaluation.

Approach:

Here in this project I have used the direct approach of analyzing the topic at first I understood the question and then I thought of which technologies I will be using. I extracted the data using SQL then I used excel and tableau for further analysis and visualization.

Tech-Stack used

1. MY SQL workbench (query)
2. Big Query(query)
3. MS excel (visualization)
4. Tableau (visualization)

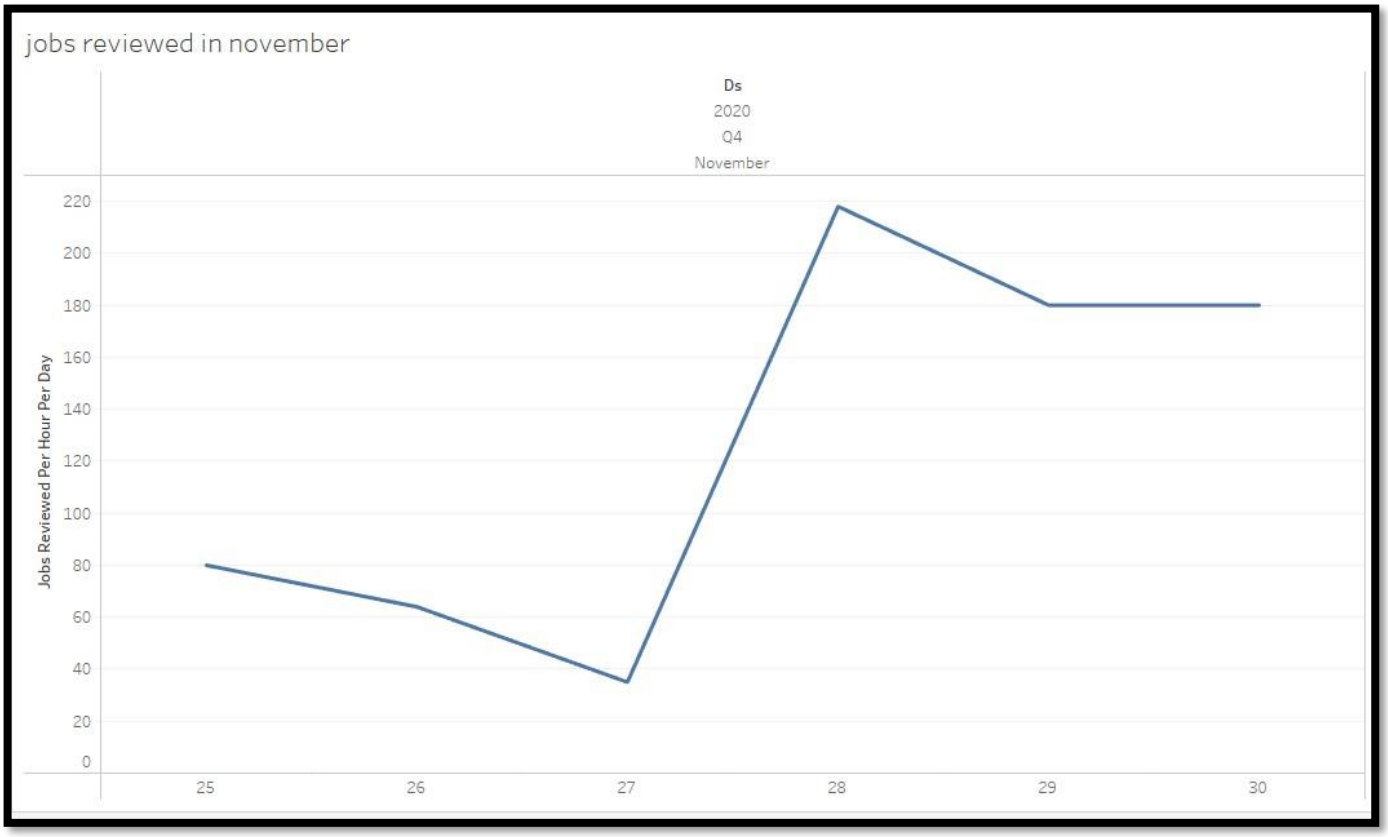


Case Study -1(JOB DATA)

1. Number of jobs reviewed:

```
select
ds,
round((count(job_id)/sum(time_spent))*3600) as "jobs reviewed per hour per day"
from  `sql project-1 table - sheet1`
group by ds;
```

ds	jobs reviewed per hour per day
30-11-2020	180
29-11-2020	180
28-11-2020	218
27-11-2020	35
26-11-2020	64
25-11-2020	80



Insights:

We get to know that the maximum jobs were reviewed on 28th November the company reviewed 218 jobs per hour and the least was on 27th November i.e 35 jobs per hour

2. THROUGHPUT:

Weekly Throughput:

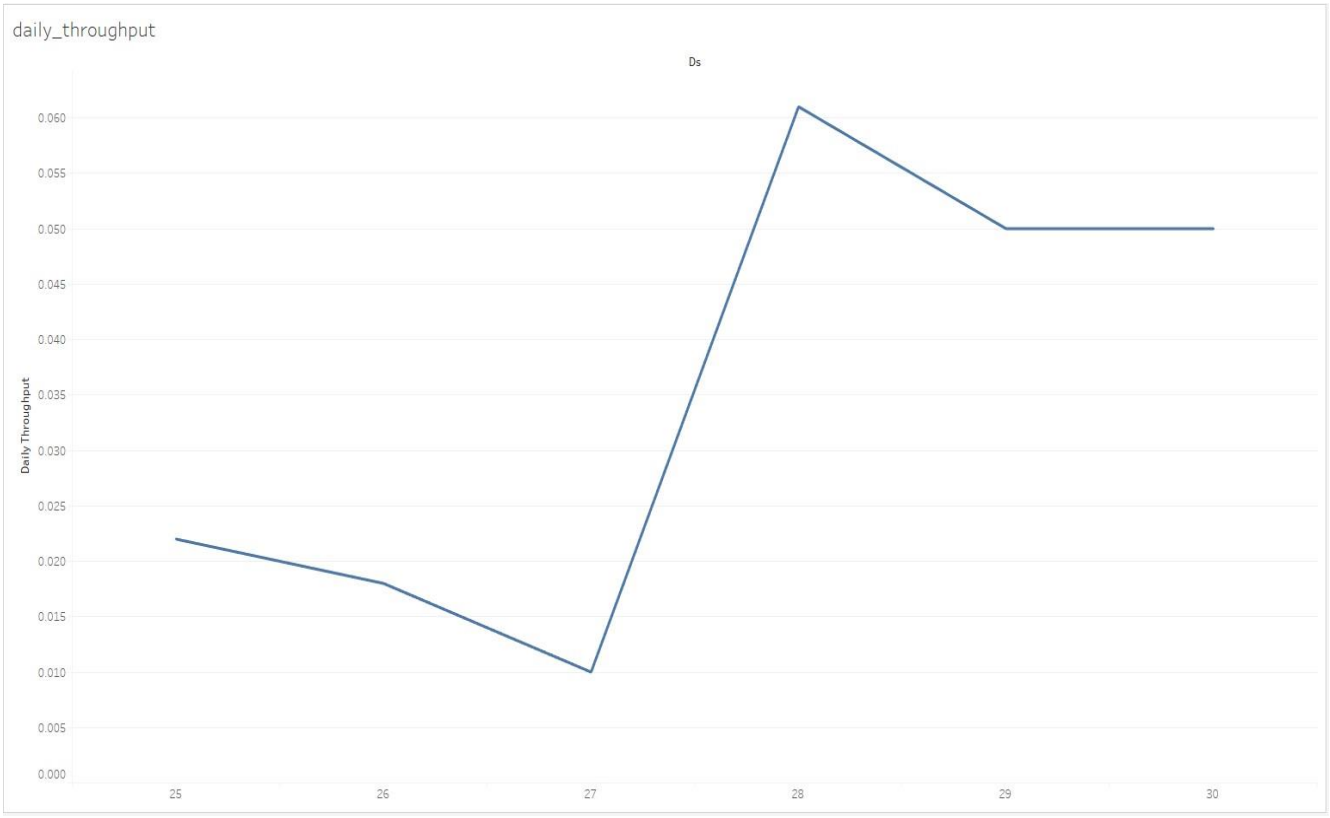
```
select round(count(event)/sum(time_spent),3) as "weekly throughput"
from `sql project-1 table - sheet1`;
```

Result:0.027

Daily Throughput:

```
select ds,
 round(count(event)/sum(time_spent),3) as "daily throughput"
from `sql project-1 table - sheet1`
group by ds
order by ds;
```

ds	daily throughput
25-11-2020	0.022
26-11-2020	0.018
27-11-2020	0.01
28-11-2020	0.061
29-11-2020	0.05
30-11-2020	0.05



Insights:

It is the no. of events happening per second. We would prefer daily throughput over weekly throughput as it gives more detailed information than the weekly throughput.

3. PERCENTAGE SHARE OF EACH LANGUAGE:

```
select
   language,
   round(100*count(*)/(select count(*) from `sql project-1 table - sheet1`),2)as percentage
   from `sql project-1 table - sheet1`
   group by 1;
```

language	percentage
English	12.5
Arabic	12.5
Persian	37.5
Hindi	12.5
French	12.5
Italian	12.5

Insights:

From this we get to know that the Persian language has the highest share and rest others have equal share.

4. Duplicate rows:

```
select actor_id as duplicate_id,count(*) as number_of_duplicates
  from`sql project-1 table - sheet1`
  group by actor_id
  having count(*)>1;
```

duplicate_id	number_of_duplicates
1003	2

Insight:

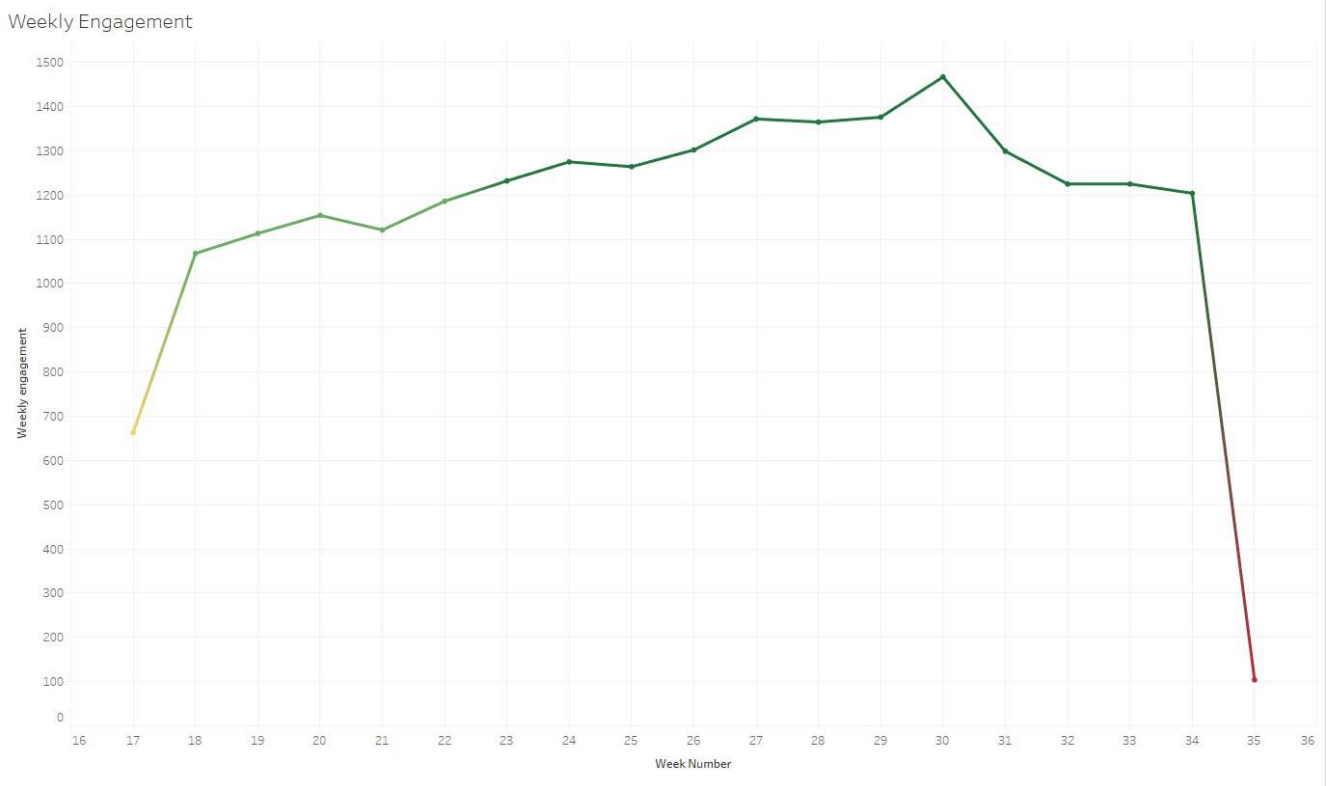
We found that the id_no 1003 has 2 duplicate entries hence one of them should be removed

Case Study -2 (Investigating metric spike)

1. USER ENGAGEMENT:

```
select  extract(week from occurred_at)as week_number,count(distinct
user_id)as Weekly_engagement
from `operation-metrics.Investigating_metric_spike.events`
where event_type='engagement'
group by 1
order by 1;
```

week_number	Weekly_engagement
17	663
18	1068
19	1113
20	1154
21	1121
22	1186
23	1232
24	1275
25	1264
26	1302
27	1372
28	1365
29	1376
30	1467
31	1299
32	1225
33	1225
34	1204
35	104



Insights:

We come to know that the highest engagement was found to be in week 30

2. User Growth:

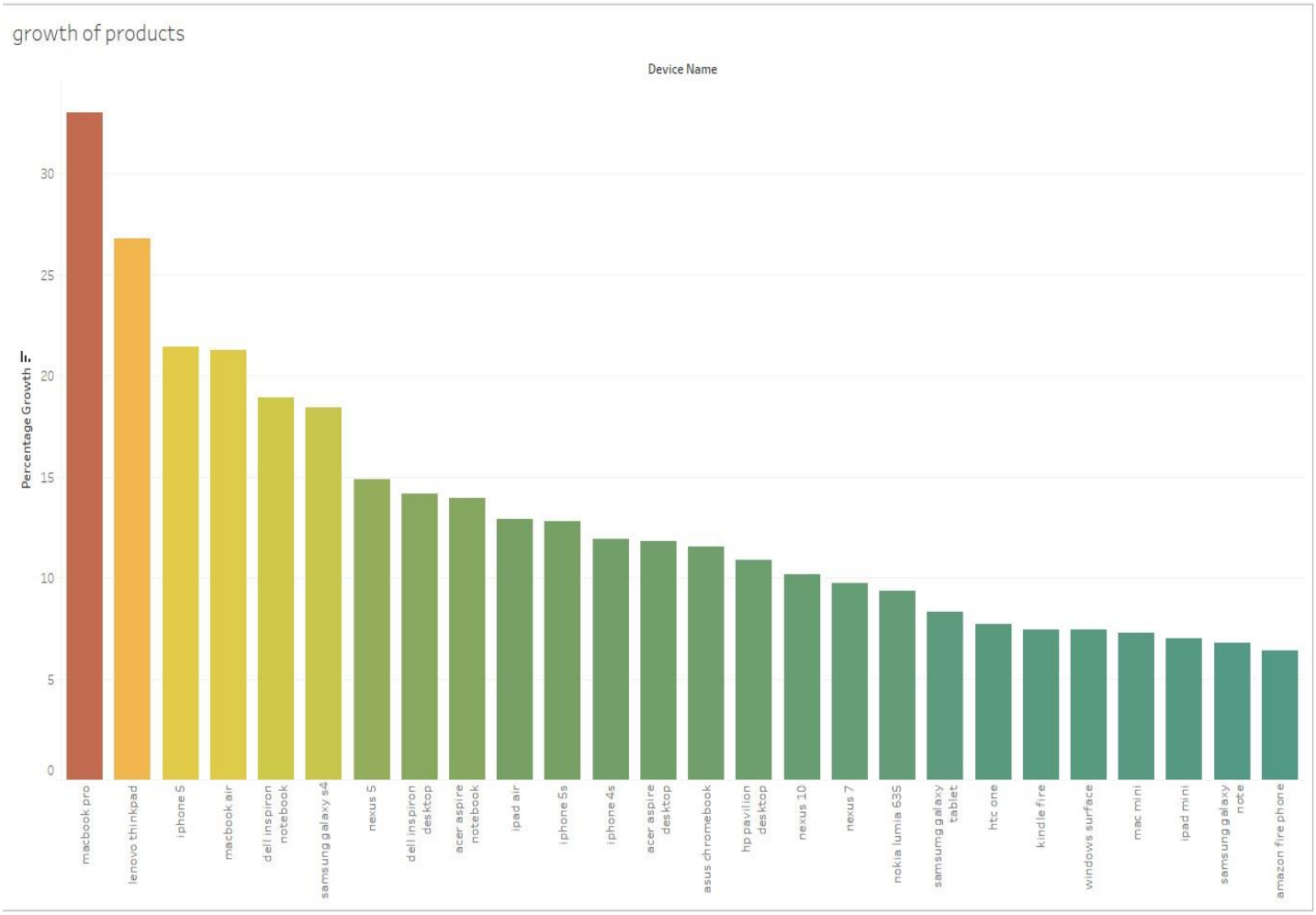
```
SELECT
  distinct device,
  EXTRACT(month FROM occurred_at) AS year_month,
  COUNT(DISTINCT user_id) AS user_count,
  AVG(COUNT(DISTINCT user_id)) OVER (PARTITION BY device,
EXTRACT(MONTH FROM occurred_at)) AS avg_user_growth

FROM
  `operation-metrics.Investigating_metric_spike.events`
WHERE
  device IS NOT NULL
GROUP BY
  device,
  occurred_at

ORDER BY
  device,
  year_month,
  user_count DESC;
```

amazon fire phone	6.374501992
asus chromebook	11.51919866
dell inspiron desktop	14.16666667
dell inspiron notebook	18.91776736
hp pavilion desktop	10.90604027
htc one	7.704918033
ipad air	12.90577989
ipad mini	7.00152207
iphone 4s	11.9412942
iphone 5	21.43712575
iphone 5s	12.79775806
kindle fire	7.450980392
lenovo thinkpad	26.81000439
mac mini	7.281553398
macbook air	21.24809741
macbook pro	33.02060647
nexus 10	10.18922853
nexus 5	14.85714286
nexus 7	9.726775956
nokia lumia 635	9.354120267
samsumg galaxy tablet	8.333333333
samsung galaxy note	6.788511749
samsung galaxy s4	18.42617842
windows surface	7.438016529

Note: I have used excel to get the percentage as aggregation of aggregation was not possible



Insights: from the analysis we get to know that MacBook pro has the best growth rate i.e of 33.02%

3. Weekly Retention:

```
SELECT first AS Week_Numbers,
SUM(CASE WHEN week_number = 0 THEN 1 ELSE 0 END) AS Week_0,
SUM(CASE WHEN week_number = 1 THEN 1 ELSE 0 END) AS Week_1,
SUM(CASE WHEN week_number = 2 THEN 1 ELSE 0 END) AS Week_2,
SUM(CASE WHEN week_number = 3 THEN 1 ELSE 0 END) AS Week_3,
SUM(CASE WHEN week_number = 4 THEN 1 ELSE 0 END) AS Week_4,
SUM(CASE WHEN week_number = 5 THEN 1 ELSE 0 END) AS Week_5,
SUM(CASE WHEN week_number = 6 THEN 1 ELSE 0 END) AS Week_6,
SUM(CASE WHEN week_number = 7 THEN 1 ELSE 0 END) AS Week_7,
SUM(CASE WHEN week_number = 8 THEN 1 ELSE 0 END) AS Week_8,
SUM(CASE WHEN week_number = 9 THEN 1 ELSE 0 END) AS Week_9,
SUM(CASE WHEN week_number = 10 THEN 1 ELSE 0 END) AS Weck_10,
SUM(CASE WHEN week_number = 11 THEN 1 ELSE 0 END) AS Week_11,
SUM(CASE WHEN week_number = 12 THEN 1 ELSE 0 END) AS Week12,
SUM(CASE WHEN week_number = 13 THEN 1 ELSE 0 END) AS Week13,
SUM(CASE WHEN week_number = 14 THEN 1 ELSE 0 END) AS Week14,
SUM(CASE WHEN week_number = 15 THEN 1 ELSE 0 END) AS Week15,
SUM(CASE WHEN week_number = 16 THEN 1 ELSE 0 END) AS Week16,
SUM(CASE WHEN week_number = 17 THEN 1 ELSE 0 END) AS Week17,
SUM(CASE WHEN week_number = 18 THEN 1 ELSE 0 END) AS Week18
FROM
(
SELECT m.user_id,m.login_week,n.first, m.login_week-first as week_number
FROM
(SELECT user_id, EXTRACT(WEEK FROM occurred_at) AS login_week FROM `operation-metrics.Investigating_metric_spike.events`
GROUP BY 1, 2) m,
(SELECT user_id, MIN(EXTRACT(WEEK FROM occurred_at)) AS first FROM `operation-metrics.Investigating_metric_spike.events`
GROUP BY 1)n
WHERE m.user_id= n.user_id
)sub
GROUP BY first
ORDER BY first;
```

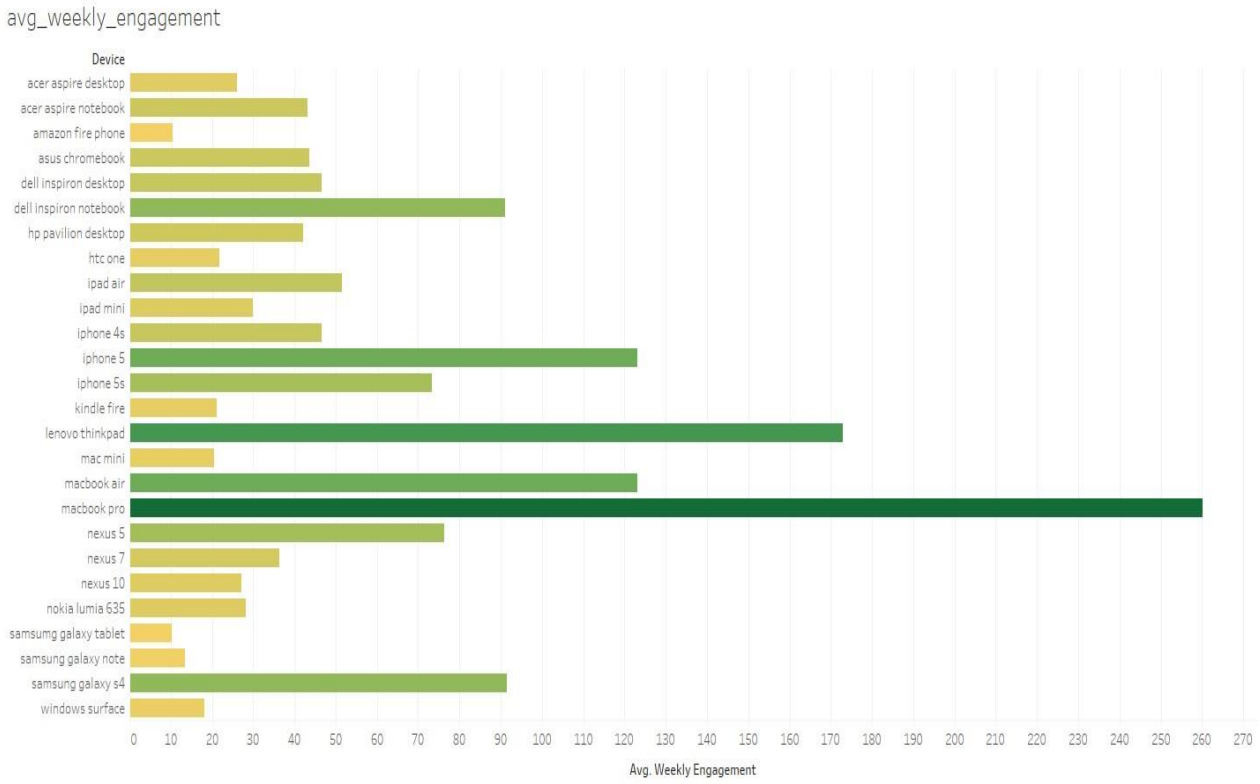
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Week_Numbers	Week_0	Week_1	Week_2	Week_3	Week_4	Week_5	Week_6	Week_7	Week_8	Week_9	Week_10	Week_11	Week12	Week13	Week14	Week15	Week16	Week17	Week18
17	740	472	324	251	205	187	167	146	145	145	136	131	132	143	116	91	82	77	5
18	788	362	261	203	168	147	144	127	113	122	106	118	127	110	97	85	67	4	0
19	601	284	173	153	114	95	91	81	95	82	68	65	63	42	51	49	2	0	0
20	555	223	165	121	91	72	63	67	63	65	67	41	40	33	40	0	0	0	0
21	495	187	131	91	74	63	75	72	58	48	45	39	35	28	2	0	0	0	0
22	521	224	150	107	87	73	63	60	55	48	41	39	31	1	0	0	0	0	0
23	542	219	138	101	90	79	69	61	54	47	35	30	0	0	0	0	0	0	0
24	535	205	143	102	81	63	65	61	38	39	29	0	0	0	0	0	0	0	0
25	500	218	139	101	75	63	50	46	38	35	2	0	0	0	0	0	0	0	0
26	495	181	114	83	73	55	47	43	29	0	0	0	0	0	0	0	0	0	0
27	493	199	121	106	68	53	40	36	1	0	0	0	0	0	0	0	0	0	0
28	486	194	114	69	46	30	28	3	0	0	0	0	0	0	0	0	0	0	0
29	501	186	102	65	47	40	1	0	0	0	0	0	0	0	0	0	0	0	0
30	533	202	121	78	53	3	0	0	0	0	0	0	0	0	0	0	0	0	0
31	430	145	76	57	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	496	188	94	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	499	202	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	518	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

We got the above weekly signup cohort

4. WEEKLY ENGAGEMENT:

```
SELECT
   device,
    EXTRACT(WEEK FROM occurred_at) AS week,
   COUNT(distinct user_id) AS weekly_engagement
FROM
   `operation-metrics.Investigating_metric_spike.events`
WHERE
   event_type = 'engagement'
GROUP BY
   device,
   week
ORDER BY
   device,
   week;
```

device	week	weekly_engagement
acer aspire desktop	17	9
acer aspire desktop	18	26
acer aspire desktop	19	23
acer aspire desktop	20	23
acer aspire desktop	21	29
acer aspire desktop	22	25
acer aspire desktop	23	22
acer aspire desktop	24	24
acer aspire desktop	25	28
acer aspire desktop	26	29
acer aspire desktop	27	29
acer aspire desktop	28	30
acer aspire desktop	29	28
acer aspire desktop	30	33
acer aspire desktop	31	31
acer aspire desktop	32	35
acer aspire desktop	33	39



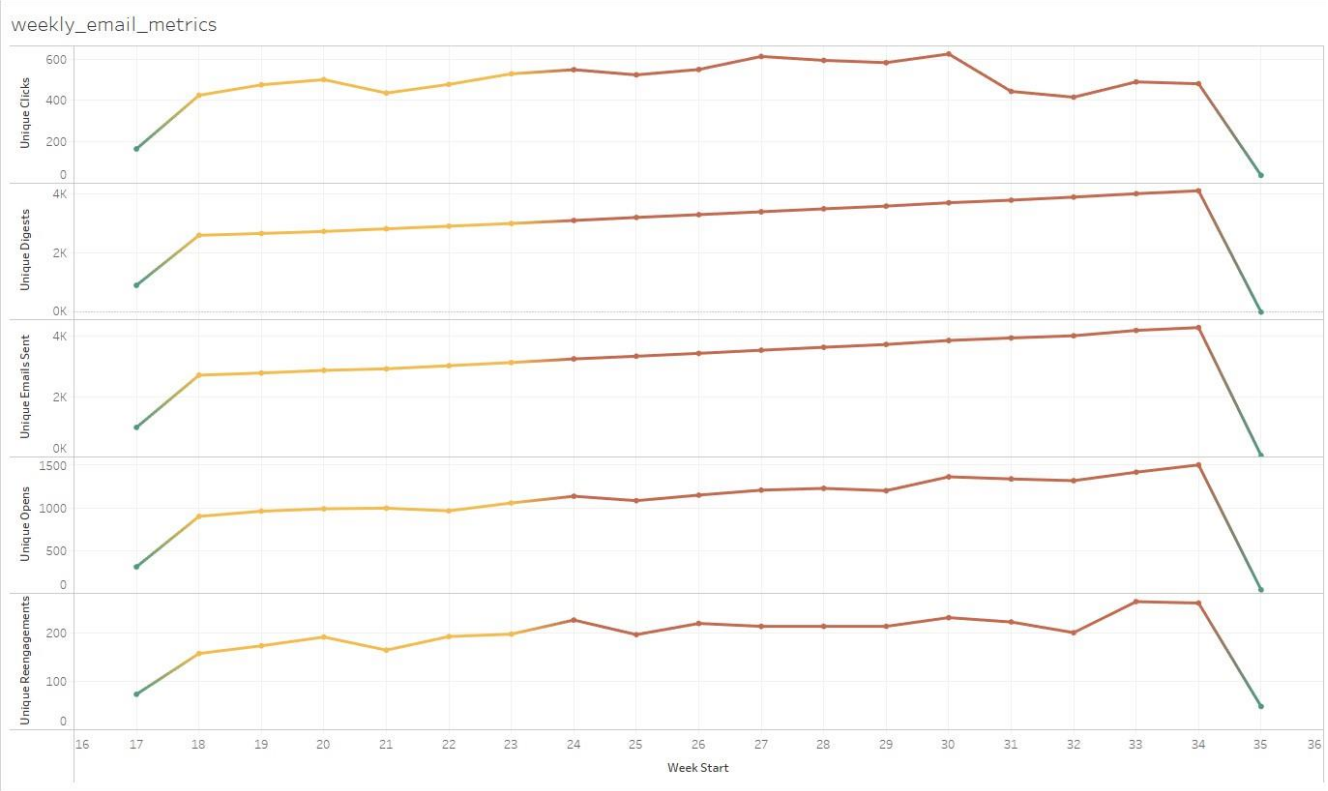
Insight:

The best avg. weekly engagement was found on MacBook pro

5. EMAIL ENGAGEMENT:

```
SELECT
  EXTRACT(WEEK FROM occurred_at) AS week,
  COUNT(DISTINCT CASE WHEN action = 'email_open' THEN user_id END) AS
unique_opens,
  COUNT(DISTINCT CASE WHEN action = 'email_clickthrough' THEN user_id
END) AS unique_clicks,
  COUNT(DISTINCT CASE WHEN action = 'sent_reengagement_email' THEN
user_id END) AS unique_reengagements,
  COUNT(DISTINCT CASE WHEN action = 'sent_weekly_digest' THEN user_id
END) AS unique_digests,
  COUNT(DISTINCT user_id) AS unique_emails_sent
FROM `operation-metrics.Investigating_metric_spike.email_events`
GROUP BY week
ORDER BY week;
```

week_start	unique_opens	unique_clicks	unique_reengagements	unique_digests	unique_emails_sent
17	310	166	73	908	981
18	900	425	157	2602	2714
19	961	476	173	2665	2787
20	989	501	191	2733	2874
21	996	436	164	2822	2926
22	965	478	192	2911	3029
23	1057	529	197	3003	3134
24	1136	549	226	3105	3254
25	1084	524	196	3207	3343
26	1149	550	219	3302	3439
27	1207	613	213	3399	3543
28	1228	594	213	3499	3641
29	1201	583	213	3592	3734
30	1363	625	231	3706	3866
31	1338	444	222	3793	3950
32	1318	416	200	3897	4023
33	1417	490	264	4012	4200
34	1502	481	261	4111	4294
35	41	38	48	0	48



Insights: there is a sudden drop in the email engagement in week 35 though there has been a lot of activities in week 34. Maybe the company has gone through a layoff session because of the sudden drop in the email engagement.