Operational Analysis And Investigating Metric Spike ↗
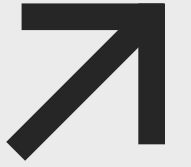
Project Description

This project focuses on analyzing operational and user engagement data to uncover trends, identify inefficiencies, and investigate metric anomalies.

The findings will help improve operations, understand user behavior, and support data-driven decision-making.

Approach

Database Setup
Imported the database into MySQL Workbench and verified the structure for accuracy.

Query Development
Wrote SQL queries to answer key business questions, focusing on marketing and operational metrics.

Data Analysis
Validated query outputs and extracted actionable insights aligned with business goals.

Results Summary
Consolidated findings to support strategic decisions in product, marketing, and development.

Tech-Stack Used

Database:

MySQL Workbench (8.0.39): User-friendly and efficient for managing databases and running SQL queries.

Programming:

SQL: Essential for extracting, analyzing, and interpreting data.

Other Tools:

PowerPoint/Word: Used to create a clear and professional report/presentation.

Why These Tools?

Selected for their ease of use, compatibility, and ability to efficiently perform data analysis and reporting.

JOB DATA ANALYSIS

Case Study - 1

Agenda

JOB DATA ANALYSIS

1. Frequency of jobs reviewed

2. Throughput Analysis

3. Language Share Analysis

4. Duplicate rows Detection

1. Frequency of Jobs reviewed

Task : Calculate the number of jobs reviewed per hour for each day in November 2020

Query:

```
SELECT
ds, ROUND(SUM(time_spent)/3600,2) AS
total_hrs_pday,
ROUND((COUNT(job_id)/(SUM(time_spent)/36
00)),2) as job_reviewed_phr_pday
FROM job_data
GROUP BY ds
ORDER BY ds;
```

ds	total_hrs_pday	job_reviewed_phr_pday
2020-11-25	0.01	80.00
2020-11-26	0.02	64.29
2020-11-27	0.03	34.62
2020-11-28	0.01	218.18
2020-11-29	0.01	180.00
2020-11-30	0.01	180.00

Insights

- The total_hrs_pday metric reveals the total hours spent reviewing jobs each day.

- The job_reviewed_phr_pday metric highlights the number of jobs that can be reviewed per hour.

- Lower no. of total_hrs_pday points out that less time was spent reviewing jobs that day

2. Throughput Analysis

Task : Calculate the 7-day rolling average of throughput (number of events per second).
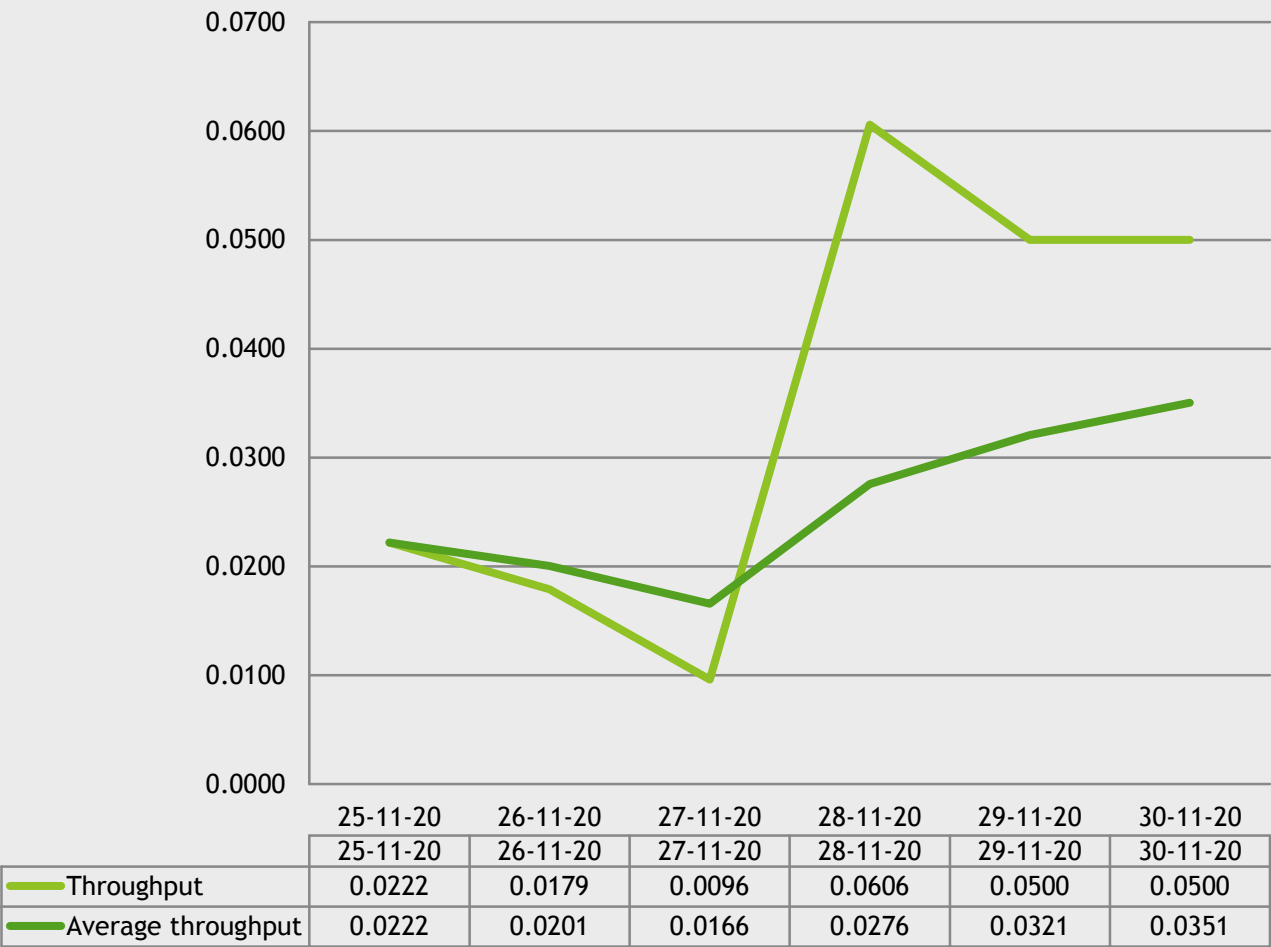
Query:

```
SELECT
ds,throughput,AVG(throughput) OVER(ORDER
BY ds ROWS BETWEEN 6 PRECEDING AND
CURRENT ROW)  as daily_throughput
FROM(
 SELECT
ds,COUNT(`event`)/SUM(time_spent) AS
throughput FROM job_data
WHERE ds BETWEEN '2020-11-24' AND '2020-
11-30' GROUP BY DS ORDER BY ds
) as sq1;
```

| ds	throughput	Daily_throughput
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

Insights

Calculating a 7-day rolling average can be considered as a better option in comparison with the daily metric as 7-day rolling average helps identify trends by smoothing out short term fluctuations in data as shown in the graph.

	25-11-20	26-11-20	27-11-20	28-11-20	29-11-20	30-11-20
Throughput	0.0222	0.0179	0.0096	0.0606	0.0500	0.0500
Average throughput	0.0222	0.0201	0.0166	0.0276	0.0321	0.0351

3. Language Share Analysis

Task :
Calculate the percentage share of each language in the last 30 days.
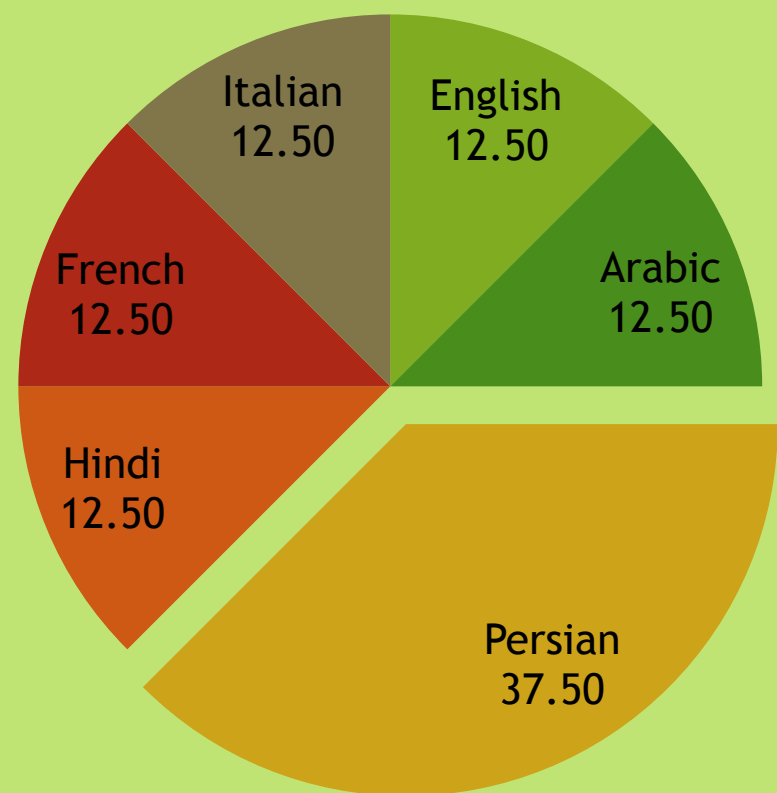
Query:
```
SELECT `language`,COUNT(job_id) as jobs,
ROUND(100 * COUNT(job_id) / (SELECT
COUNT(job_id) FROM job_data),2)
AS percentage_share
FROM job_data
GROUP BY `language`;
```


language	jobs	percentage_share
English	1	12.50
Arabic	1	12.50
Persian	3	37.50
Hindi	1	12.50
French	1	12.50
Italian	1	12.50

Insights

- Persian language is used by millions of people living in Middle East and Central part of Asia.

- User base from Middle East and Central Asia are actively engaging with job postings.

- Investing in more language support can be a good choice if the organisation plans to expand the market

4. Duplicate Row Detection

Task :
Write an SQL query to display duplicate rows from the job_data table.

Query:
```
SELECT * FROM (
SELECT * , CASE WHEN COUNT(*) > 1 THEN
'Duplicate' Else 'UNIQUE' END AS
record_status
FROM job_data
ORDER BY
ds,job_id,actor_id,`event`,`language`,time_
spent,org
) as sq1
WHERE record_status = 'Duplicate';
```

ds	job_id	actor_id	event	language	time_spent	org	record_status

Insights

- No Duplicate records were found in the data.

- Not a single user interacted with the same job posting more than once

Summary

Metrics Analysis:

- **total_hrs_pday**: Represents the total hours spent reviewing jobs daily; lower values indicate less time spent.
- **job_reviewed_phr_pday**: Measures the number of jobs reviewed per hour.
- A 7-day rolling average is recommended for trend analysis, as it smooths short-term fluctuations better than daily metrics.

User Engagement:

- Persian is widely spoken in the Middle East and Central Asia, regions with active engagement in job postings.
- Expanding language support could enhance user experience and market reach.

Data Integrity:

- No duplicate records were found.
- Users did not interact with the same job posting more than once.

INVESTIGATING METRIC SPIKE

Case Study - 2

Agenda

INVESTIGATING METRIC SPIKE

1. Weekly User Engagement

2. User Growth Analysis

3. Weekly Retention Analysis

4. Weekly Engagement on each device

5. Email Engagement Analysis

1. Weekly User Engagement

Task :
**Measure the activeness of users on a weekly basis.**

Query:

```
SELECT EXTRACT(WEEK FROM
occurred_at) AS `WEEK`,
COUNT(DISTINCT user_id) AS
users_engaged
FROM events WHERE event_type =
'engagement'
GROUP BY `WEEK`
ORDER BY `WEEK`;
```

| WEEK	users_engaged
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

- Sudden increase in user engagement in week 18

- User engagement started to decline after week 30

- User engagement was lowest in week 35

- Organising Virtual Contests during this period can help with engagement

Weekly User Engagement

Legend: no_of_users
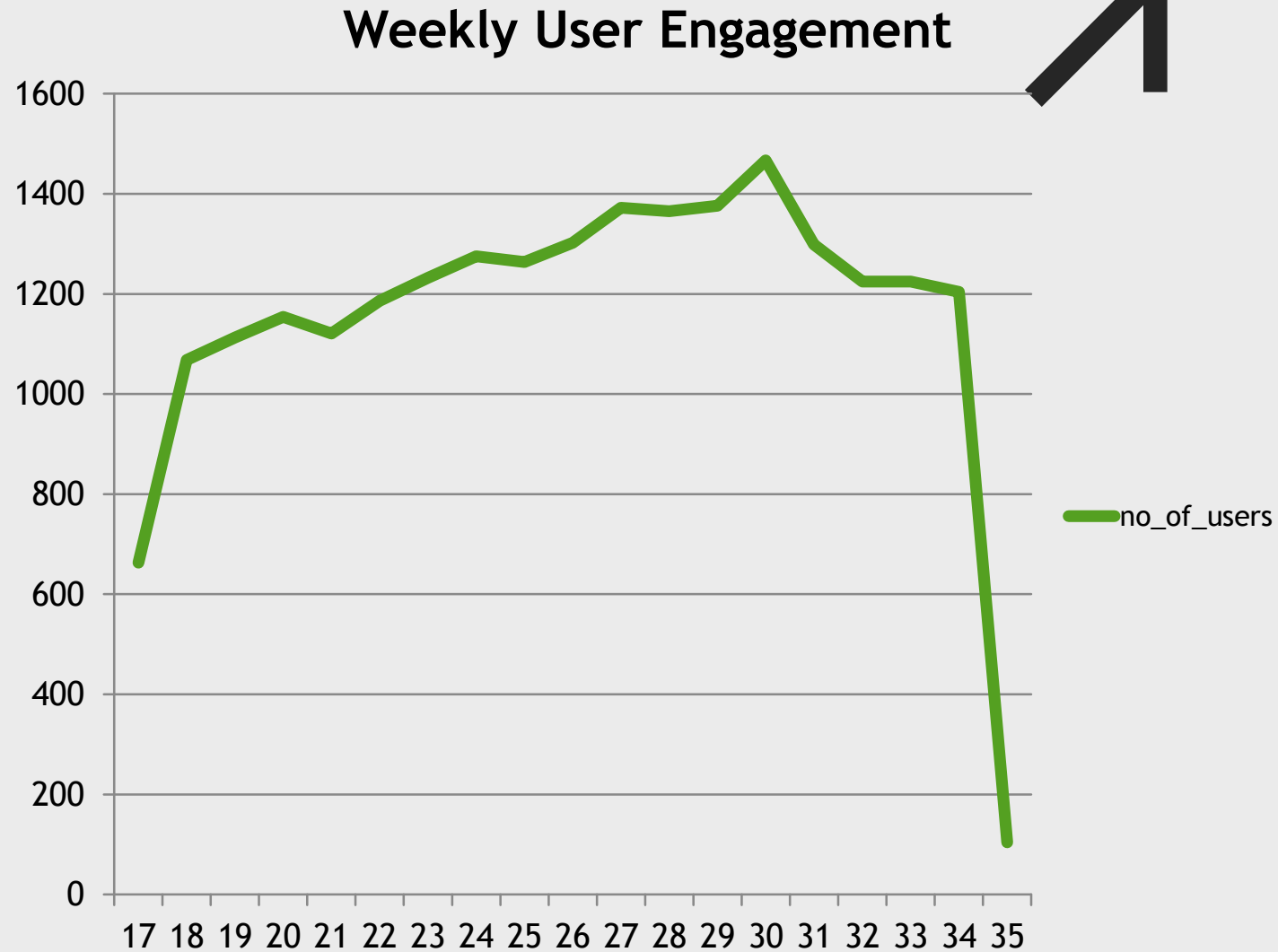
2. User Growth Analysis

Task :
**Analyze the growth of users over time for a product**

Total Users
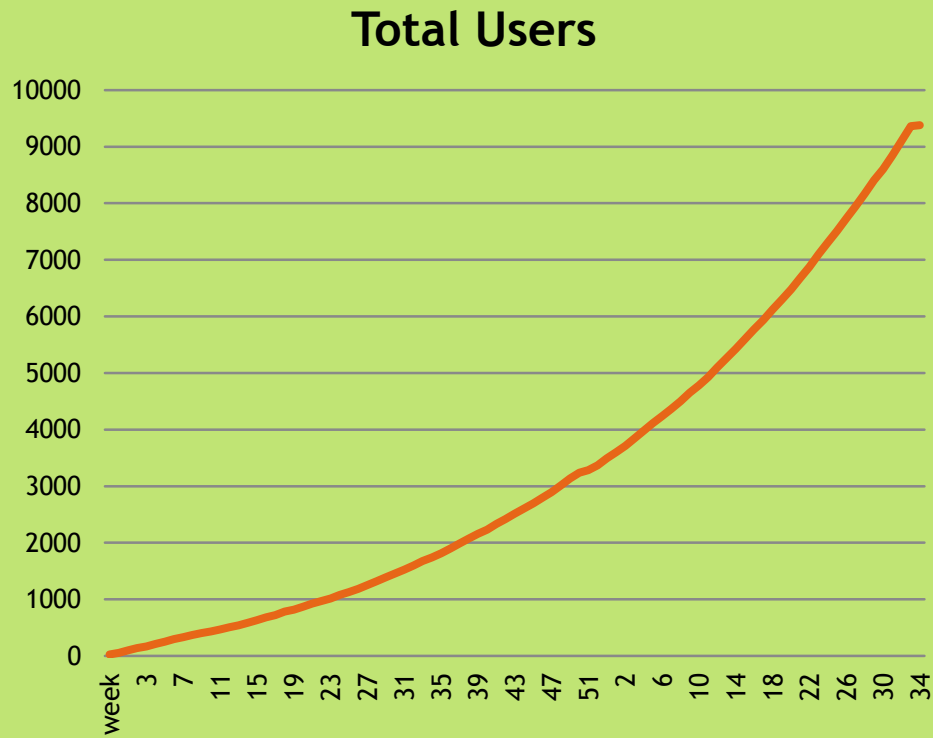
Query:
```
SELECT
`year`,`week`,active_users,SUM(active_users)
OVER(ORDER BY `year`,`week` ROWS BETWEEN
UNBOUNDED PRECEDING AND CURRENT ROW) AS
total_active_users
FROM ( SELECT EXTRACT(YEAR FROM activated_at)
as `year`,EXTRACT(WEEK FROM activated_at) as
`week`,COUNT(DISTINCT user_id) as active_users
FROM users
GROUP BY `year`,`week`
ORDER BY `year`,`week`) as sq1;
```

Weekly Active Users
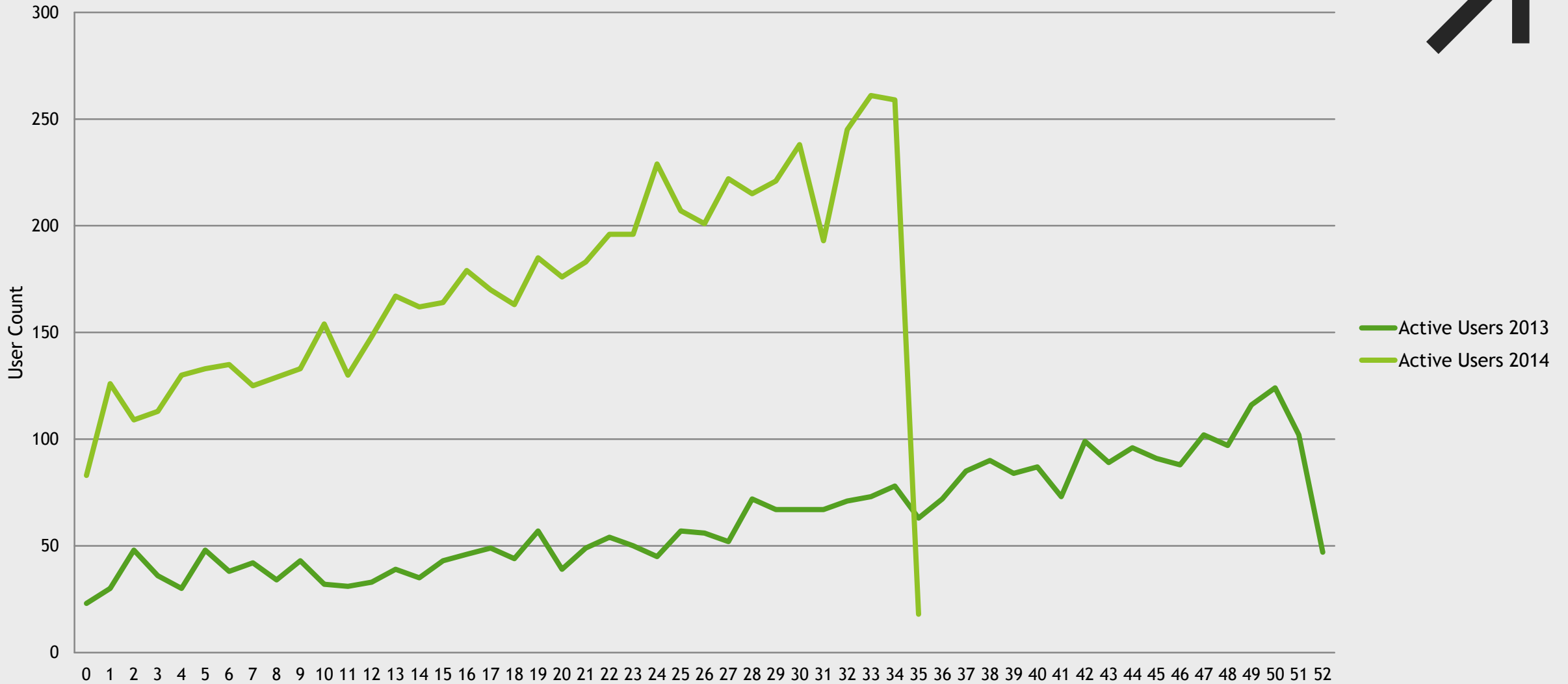
Insights

year	week	active_users	total_active_users
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167
2013	5	48	215
2013	6	38	253
2013	7	42	295
2013	8	34	329
2013	9	43	372
2013	10	32	404
2013	11	31	435
2013	12	33	468
2013	13	39	507
2013	14	35	542
2013	15	43	585
2013	16	46	631
2013	17	49	680
2013	18	44	724
2013	19	57	781
2013	20	39	820
2013	21	49	869
2013	22	54	923
2013	23	50	973
2013	24	45	1018

- More no. of users were active per week in 2014 as compared to 2013.

- No decline in total users on the platform

- **Seasonal patterns** or **campaign effectiveness** can be understood by observing spikes or drops in new users.

3. Weekly Retention Analysis

Task :
Analyze the retention of users on a weekly basis after signing up for a product

start_date	year_week	active_users	total_users	retention_rate(%)
2013-01-01	201417	1	7	14.2857
2013-01-01	201418	1	7	14.2857
2013-01-01	201419	2	7	28.5714
2013-01-01	201420	2	7	28.5714
2013-01-01	201421	1	7	14.2857
2013-01-01	201422	1	7	14.2857
2013-01-01	201423	1	7	14.2857
2013-01-01	201424	2	7	28.5714
2013-01-01	201425	2	7	28.5714
2013-01-01	201426	1	7	14.2857
2013-01-01	201427	1	7	14.2857
2013-01-01	201430	2	7	28.5714
2013-01-01	201431	1	7	14.2857
2013-01-02	201417	1	7	14.2857
2013-01-02	201418	2	7	28.5714
2013-01-02	201419	1	7	14.2857
2013-01-02	201420	1	7	14.2857
2013-01-02	201421	1	7	14.2857
2013-01-02	201422	2	7	28.5714
2013-01-02	201423	2	7	28.5714

Query:

```
SELECT
sq1.start_date,sq2.year_week,sq2.active_users,
sq1.total_users,sq2.active_users/sq1.total_users*100
AS `retention_rate(%)`
FROM ( SELECT DATE(activated_at) AS
start_date,COUNT(*) AS total_users FROM users
GROUP BY 1    ) as sq1,
( SELECT DATE(u.activated_at) AS
start_date,YEARWEEK(e.occurred_at) AS
year_week,COUNT(DISTINCT e.user_id) AS
active_users FROM events e    JOIN users u ON
e.user_id = u.user_id    WHERE e.event_type =
'engagement'    GROUP BY start_date,year_week) AS
sq2WHERE sq1.start_date  = sq2.start_date;
```
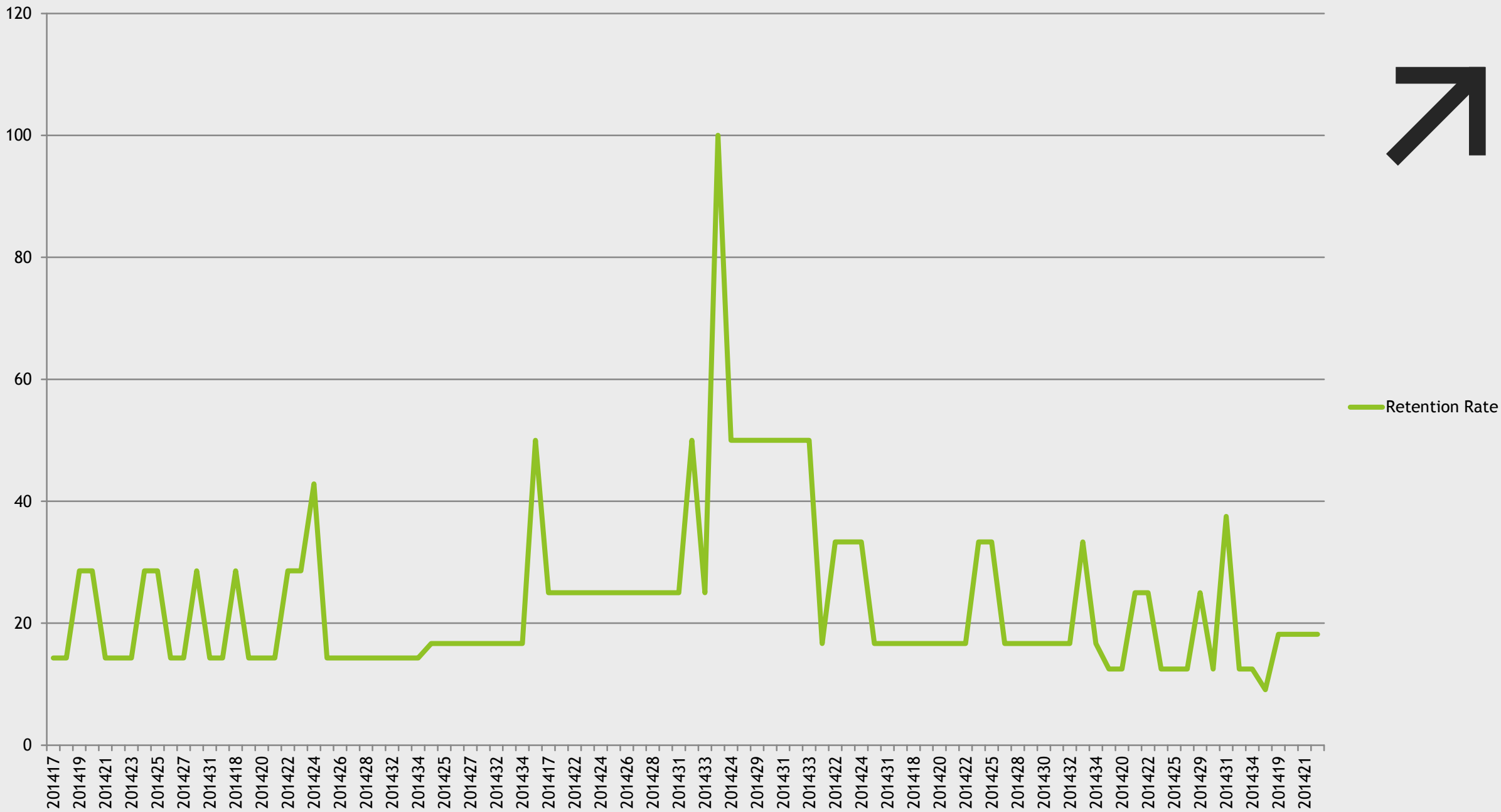
Insights

- Retention refers to the percentage of users who performed some action (e.g., logging in, using a feature, making a purchase) in subsequent weeks after signing up.

- If a feature was introduced during a specific week, measure its impact by comparing retention rates before and after its release.

- Steeper declines suggests that the product struggles to maintain engagement.

- Weeks with high retention rate an indicate successful campaigns

4. Weekly Engagement Per Device

Task :
Measure the activeness of users on a weekly basis per device

Query:

SELECT EXTRACT(WEEK FROM occurred_at) AS `WEEK`,device, COUNT(DISTINCT user_id) as users_engagedFROM events WHERE event_type = 'engagement'
GROUP BY `WEEK`,device
ORDER BY `WEEK`,device;

WEEK	device	users_engaged
17	acer aspire desktop	9
17	acer aspire notebook	20
17	amazon fire phone	4
17	asus chromebook	21
17	dell inspiron desktop	18
17	dell inspiron notebook	46
17	hp pavilion desktop	14
17	htc one	16
17	ipad air	27
17	ipad mini	19
17	iphone 4s	21
17	iphone 5	65
17	iphone 5s	42
17	kindle fire	6
17	lenovo thinkpad	86
17	mac mini	6
17	macbook air	54
17	macbook pro	143
17	nexus 10	16
17	nexus 5	40
17	nexus 7	18
17	nokia lumia 635	17
17	samsumg galaxy tablet	8
17	samsung galaxy note	7
17	samsung galaxy s4	52
17	windows surface	10
18	acer aspire desktop	26

Devices

Week 17
Week 18
Week 19
Week 20
Week 21
Week 22
Week 23
Week 24
Week 25
Week 26
Week 27
Week 28
Week 29
Week 30
Week 31
Week 32
Week 33
Week 34
Week 35

Insights

- Device-based activeness data can reveal where users may face difficulties.

- Lower activeness on mobile phones as compared to desktops might highlight the need for app-specific features or improved mobile interfaces.

- Most no. of users in desktop category use Macbook Pro.

- Most no. of users in mobile category use Iphone 5.

5. Email Engagement Analysis

Task :
Analyze how users are engaging with the email service

Query:

SELECT
action,COUNT(DISTINCT user_id) as
users_engaged
FROM email_events
GROUP BY action;

action ▲	users_engaged
email_clickthrough	5277
email_open	5927
sent_reengagement_email	3653
sent_weekly_digest	4111

Users Engaged

■ Users Engaged

Email Clickthrough	Email Open	Sent re-engagement Email	Sent weekly digest
5277	5927	3653	4111
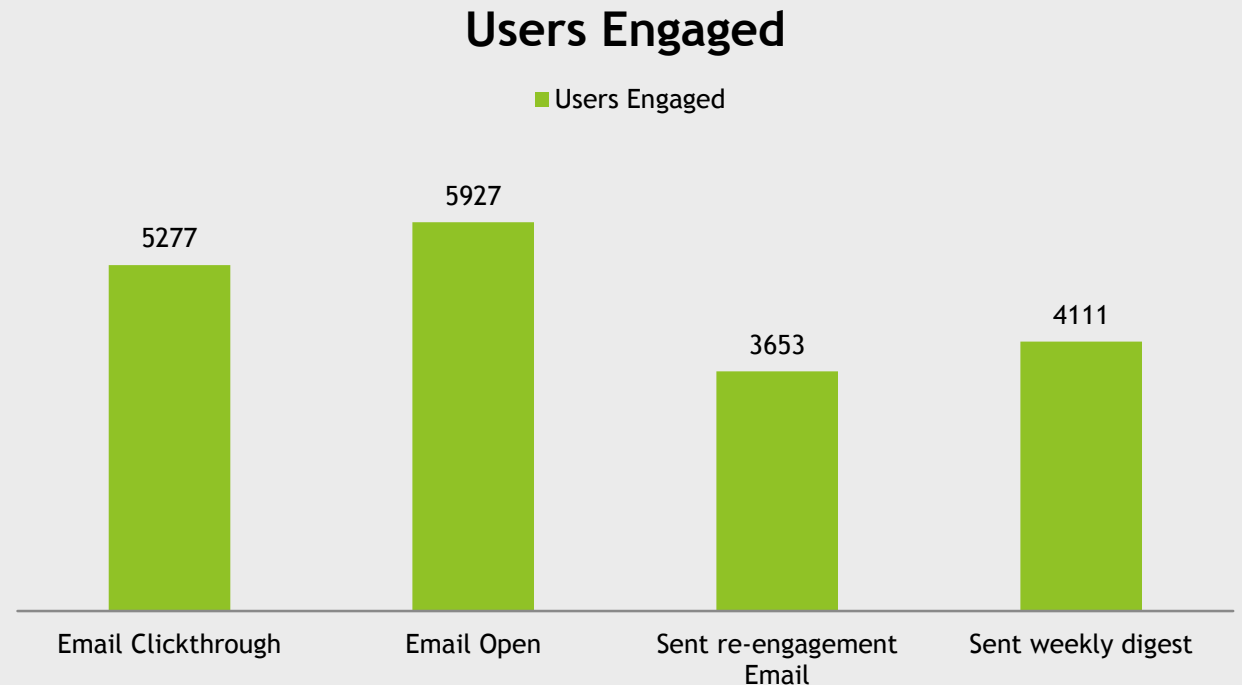
Insights

- Email_open action was performed more than sent_reengagement_mail action, this suggests that a significant number of users are active.

- Around 4000 people have mailed weekly digests so far thus having a part in increasing the no. of email_open action.

Summary

User Engagement Trends:
- Sharp increase in user engagement observed in week 18, followed by a decline after week 30.
- Engagement hit its lowest point in week 35.
- Virtual contests during low-engagement periods could help improve activity.
- Weekly active users were higher in 2014 compared to 2013.
- Despite fluctuations, the total user base on the platform has remained stable.

Seasonal Patterns and Campaign Analysis:
- Spikes or drops in new users can indicate seasonal trends or campaign effectiveness.
- Retention measures the percentage of users who stay active after signing up.
- Retention analysis before and after feature releases can reveal their impact.
- High retention rates correlate with successful campaigns, while steep declines indicate challenges in sustaining engagement.

Device Usage and Performance:
- Mobile activeness is lower than desktop, suggesting the need for mobile-specific improvements.
- Most desktop users prefer MacBook Pro; most mobile users use iPhone 5.\

Email Engagement:
- The **email_open** action is significantly higher than **sent_reengagement_mail**, indicating active user interest.
- Weekly digests have been mailed by around 4,000 users, contributing to increased email engagement.

Reflections

- SQL Skills: Gained experience in writing efficient queries and analyzing complex data.

- Analytical Growth: Improved ability to interpret data and provide actionable insights.

- Challenges Overcome: Navigated complex table relationships.

  Optimized query performance through iteration.

- Key Takeaways: Strengthened problem-solving and reinforced the importance of data-driven decisions.

Thank you

By: Aditya. C. Mohite
Date: 20/01/2025