

Operation Analytics - Case Study 2

Project Description

Operational Analytics is a critical process aimed at analyzing a company's end-to-end operations to identify areas for improvement. As a Lead Data Analyst, the objective was to leverage advanced SQL skills to analyze various datasets and tables, providing valuable insights to different departments within the company. The focus was on understanding and explaining sudden changes in key metrics, particularly in user engagement.

Approach

Utilized three tables: `users`, `events`, and `email_events`. Complex queries were written and executed to perform the tasks for the project showcasing DML commands skills. Additional features were extracted from the data using SQL functions to uncover hidden patterns in the data.

Tech-Stack Used

SQL queries were executed in MySQL Workbench 8.0 CE for database creation and Beekeeper Studio 4.0.3 CE for query editing, chosen for its user-friendly interface and quality-of-life features.

Insights

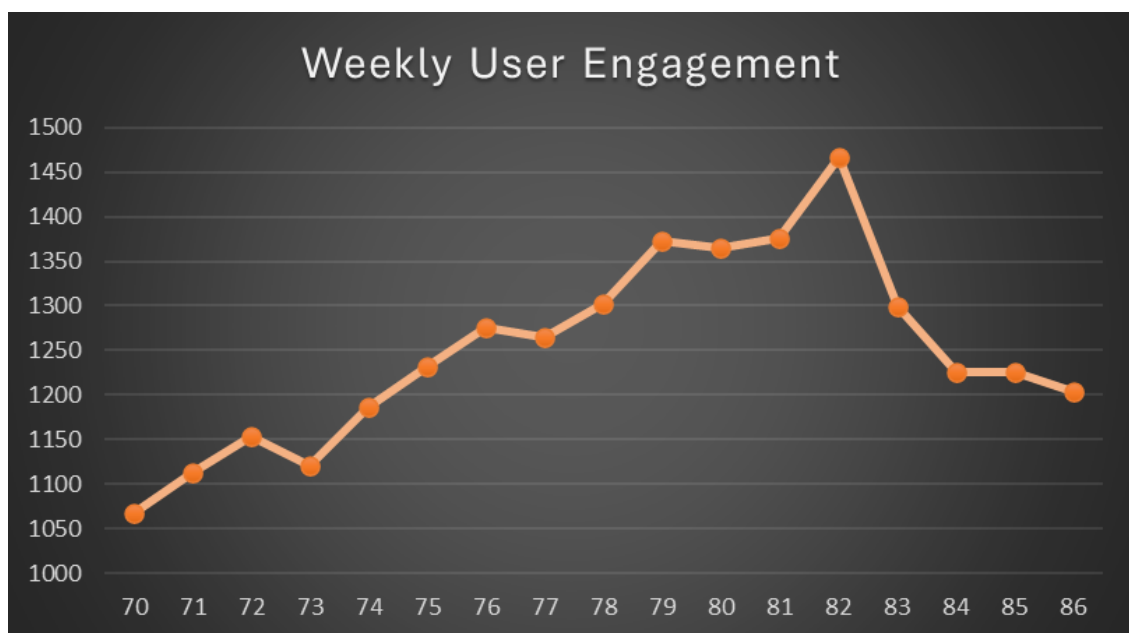
Weekly User Engagement

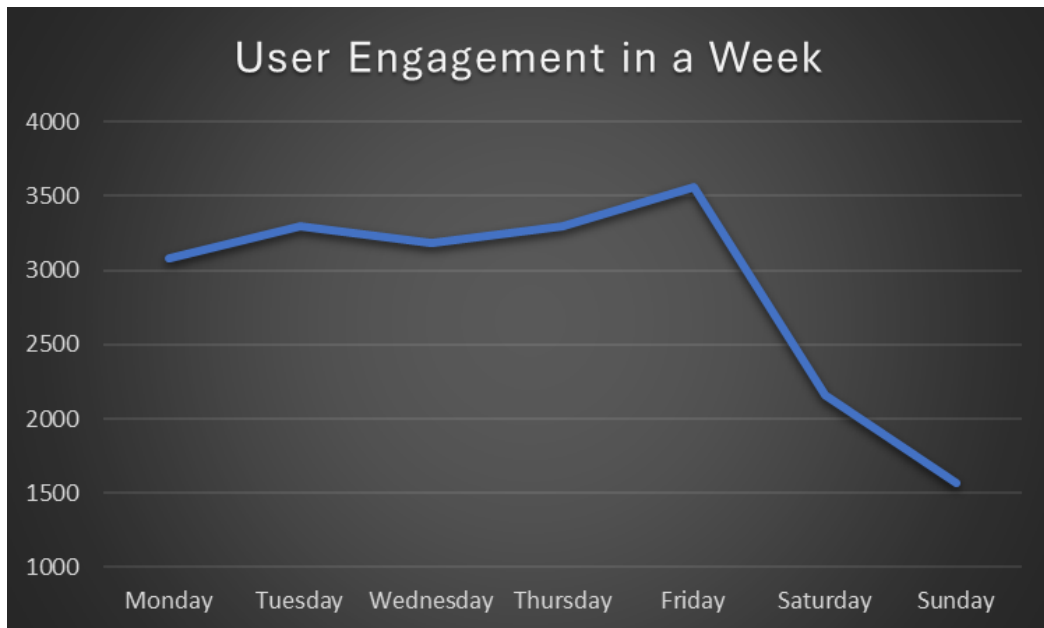
```
# Weekly user engagement
SELECT
    FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7) AS week, # assigning unique week number
    COUNT(DISTINCT user_id) AS weekly_active_users
FROM
    events
WHERE
    event_type = 'engagement'
GROUP BY
    FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7)
ORDER BY
    week;

# Weekday user engagement
SELECT
    DAYNAME(occurred_at) AS weekday,
    WEEKDAY(occurred_at) AS num_weekday,
    COUNT(DISTINCT user_id) AS active_users
FROM
    events
WHERE
    event_type = 'engagement'
GROUP BY
    DAYNAME(occurred_at),
    WEEKDAY(occurred_at)
ORDER BY
    2;
```

week	weekly_active_users
69	663
70	1068
71	1113
72	1154
73	1121
74	1186
75	1232
76	1275
77	1264
78	1302
79	1372
80	1365
81	1376
82	1467
83	1299
84	1225
85	1225
86	1204
87	104

weekday	num_weekday	active_users
Monday	0	3077
Tuesday	1	3300
Wednesday	2	3187
Thursday	3	3298
Friday	4	3561
Saturday	5	2159
Sunday	6	1566





- Engagement data becomes available from the 69th week onward.
- A consistent increasing trend in weekly active users is evident over time.
- Minor decreases occurred between weeks 76-78, 83-84, and 85-86.
- A notable spike of approximately 6.6% occurred between weeks 81 and 82, indicating potential successful marketing campaigns or product updates.
- Another significant growth of about 5.3% was observed between weeks 78 and 79.
- A decrease of nearly 5% from weeks 82 to 83 warrants investigation to understand the factors contributing to the decline.
- Weeks 83-86 show stable but slightly declining weekly active users.
- Monitoring this pattern is crucial for timely intervention if further declines occur.
- Engagement drops 39% on Saturdays and 56% on Sundays as compared to Friday peak.

While overall user engagement exhibits positive growth, the focus should be on maintaining steady growth rates without drastic drops. Analyzing the reasons behind both increases and decreases will inform future decision-making processes and ensure sustained user engagement.

User Growth Analysis

```
# Weekly User Growth
WITH
  new_weekly_users AS (
    SELECT
      FLOOR(DATEDIFF(created_at, '2012-12-30') / 7) AS week,
      COUNT(DISTINCT user_id) AS weekly_new_users,
      COUNT(DISTINCT user_id) - LAG(COUNT(DISTINCT user_id)) OVER (
        ORDER BY
          FLOOR(DATEDIFF(created_at, '2012-12-30') / 7)
      ) AS weekly_incremental_new_users
    FROM
      users
  )
GROUP BY
```

```

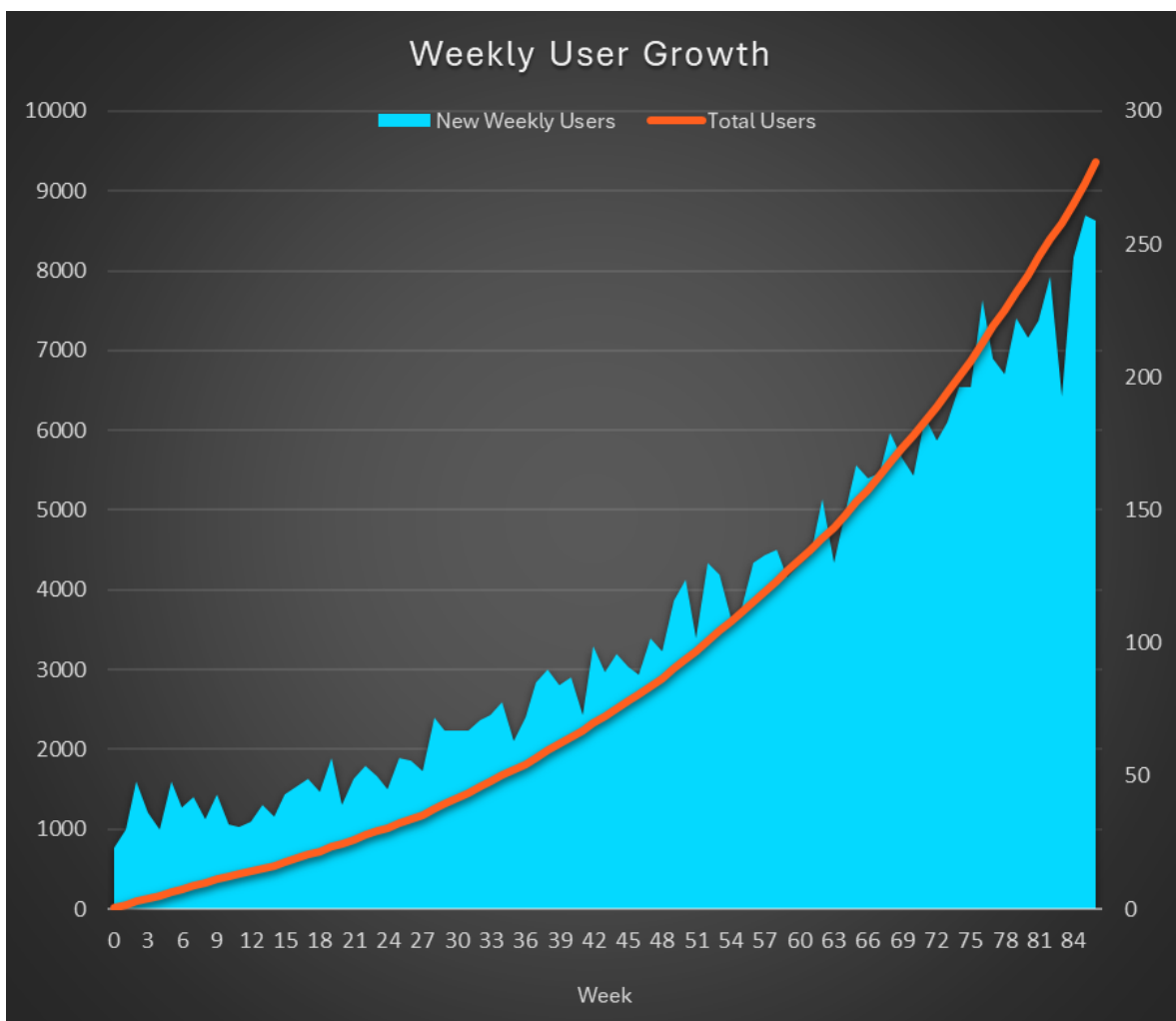
        FLOOR(DATEDIFF(created_at, '2012-12-30') / 7)
    )
SELECT
    week,
    weekly_new_users,
    weekly_incremental_new_users,
    SUM(weekly_new_users) over (
        ORDER BY
            week ROWS BETWEEN UNBOUNDED PRECEDING
            AND CURRENT ROW
    ) AS cumulative_users
FROM
    new_weekly_users
ORDER BY
    week;

```

week	weekly_new_users	weekly_incremental_new_users	cumulative_users
0	23	NULL	23
1	30	7	53
2	48	18	101
3	36	-12	137
4	30	-6	167
5	48	18	215
6	38	-10	253
7	42	4	295
8	34	-8	329
9	43	9	372
10	32	-11	404
11	31	-1	435
12	33	2	468
13	39	6	507
14	35	-4	542
15	43	8	585
16	46	3	631
17	49	3	680
18	44	-5	724
19	57	13	781
20	39	-18	820
21	49	10	869
22	54	5	923
23	50	-4	973
24	45	-5	1018
25	57	12	1075
26	56	-1	1131
27	52	-4	1183
28	72	20	1255
29	67	-5	1322
30	67	0	1389

week	weekly_new_users	weekly_incremental_new_users	cumulative_users
31	67	0	1456
32	71	4	1527
33	73	2	1600
34	78	5	1678
35	63	-15	1741
36	72	9	1813
37	85	13	1898
38	90	5	1988
39	84	-6	2072
40	87	3	2159
41	73	-14	2232
42	99	26	2331
43	89	-10	2420
44	96	7	2516
45	91	-5	2607
46	88	-3	2695
47	102	14	2797
48	97	-5	2894
49	116	19	3010
50	124	8	3134
51	102	-22	3236
52	130	28	3366
53	126	-4	3492
54	109	-17	3601
55	113	4	3714
56	130	17	3844
57	133	3	3977
58	135	2	4112
59	125	-10	4237
60	129	4	4366
61	133	4	4499
62	154	21	4653
63	130	-24	4783
64	148	18	4931
65	167	19	5098
66	162	-5	5260
67	164	2	5424
68	179	15	5603
69	170	-9	5773
70	163	-7	5936
71	185	22	6121
72	176	-9	6297
73	183	7	6480
74	196	13	6676
75	196	0	6872

week	weekly_new_users	weekly_incremental_new_users	cumulative_users
76	229	33	7101
77	207	-22	7308
78	201	-6	7509
79	222	21	7731
80	215	-7	7946
81	221	6	8167
82	238	17	8405
83	193	-45	8598
84	245	52	8843
85	261	16	9104
86	259	-2	9363
87	18	-241	9381





- A general increasing trend in weekly new users with fluctuations, averaging ~107 new users per week.
- Peak weeks at 76 (229) and 84 (245) indicate successful efforts to attract new users.
- A significant increase of approximately 17.3% occurred between weeks 75 (196) and 76 (229), suggesting effective promotions or features. Identifying the factors behind this jump can guide future strategies.
- A significant drop of about 18.5% between weeks 82 (238) and 83 (193) in additional new users warrants investigation to prevent recurrences.
- Smaller decreases, like those between weeks 3 and 4 (-12.5%) or weeks 40 and 41 (-14.6%), offer useful insights into user acquisition efficiency.
- Weeks 83 and 84 exhibit large fluctuations with -45 and 52 additional new users, requiring comprehensive analysis.
- Normal weekly fluctuations are in the range of +/- 10 additional users, with an average of ~3 additional new users added every week.
- Despite setbacks, the overall direction demonstrates gradual and consistent expansion, particularly evident towards the end of the analyzed period.

Analyzing both positive and negative spikes uncovers drivers of exceptional customer acquisition or churn weeks. Comparing periods of steady additional gains provides context on normal variability. No seasonality is evident, and the focus should be on unexpected large swings in either direction vs. the standard weekly fluctuation.

Weekly Retention Analysis

```
# Weekly Rentention by Sign-up Cohort
CREATE VIEW signup_engagement AS (
  WITH cte AS (
    SELECT
```

```

        DISTINCT FLOOR(DATEDIFF(activated_at, '2012-12-30') / 7) AS signup_week,
        FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7) AS engagement_week,
        u.user_id
    FROM
        users u
        LEFT JOIN events e ON u.user_id = e.user_id
    )
    SELECT
        engagement_week - signup_week AS weeks_after_signup,
        engagement_week,
        signup_week,
        user_id
    FROM
        cte
);
WITH signups_per_week AS (
    SELECT
        FLOOR(DATEDIFF(activated_at, '2012-12-30') / 7) AS week,
        COUNT(user_id) AS num_users
    FROM
        users
    GROUP BY
        FLOOR(DATEDIFF(activated_at, '2012-12-30') / 7)
),
weekly_retention AS (
    SELECT
        DISTINCT signup_week,
        IFNULL(weeks_after_signup, -1) AS weeks_after_signup, #-1 for no engagement
        COUNT(user_id) OVER (
            PARTITION BY signup_week,
            weeks_after_signup
            ORDER BY
                signup_week
        ) AS engagement_user_count
    FROM
        signup_engagement
)
SELECT
    signup_week,
    SUM(engagement_user_count / num_users) / (
        (
            SELECT
                MAX(engagement_week)
            FROM
                signup_engagement
        ) - signup_week + 1
    ) AS avg_retention # dividing by total engagement weeks since and including signup
FROM
    weekly_retention AS wr
    JOIN signups_per_week AS spw ON wr.signup_week = spw.week
WHERE
    signup_week <> (
        SELECT
            MAX(signup_week)
        FROM

```



```

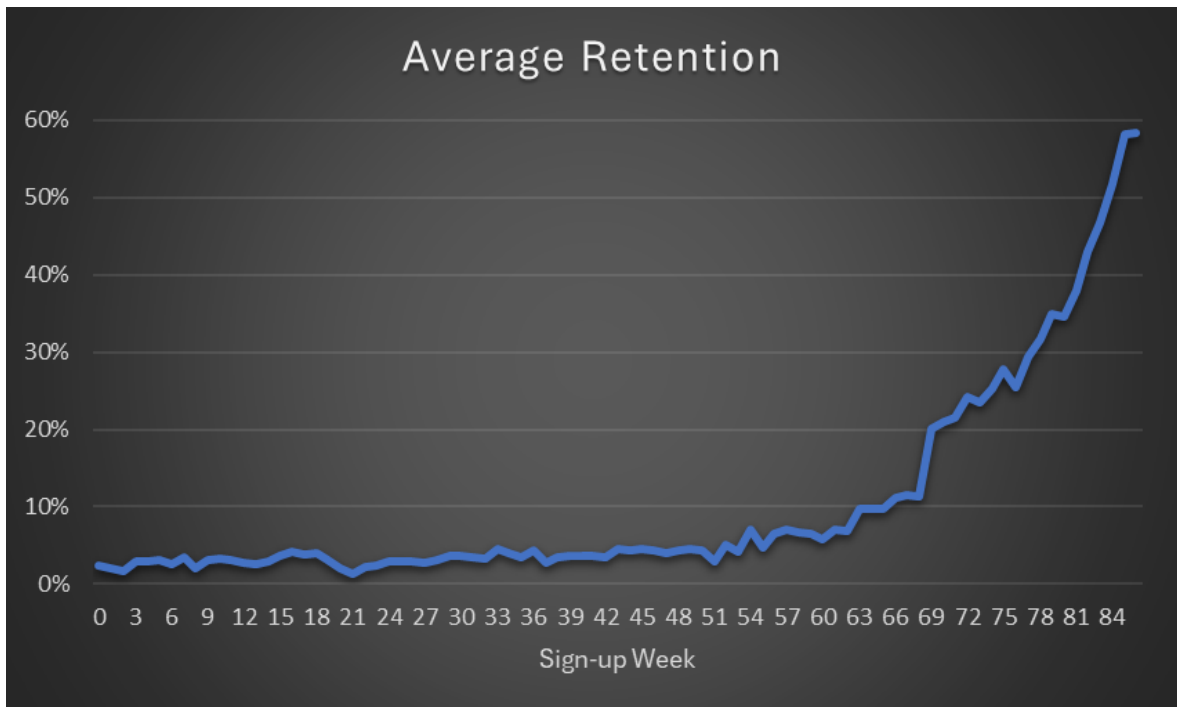
        signup_engagement
    ) # excluding last signup week from the calculation
    AND weeks_after_signup <> -1 # excluding no engagement
GROUP BY
    signup_week
ORDER BY
    1;

```

signup_week	avg_retention
0	0.02371591
1	0.01953908
2	0.01720000
3	0.02941294
4	0.02936667
5	0.03112289
6	0.02503293
7	0.03527531
8	0.02021750
9	0.03031772
10	0.03245897
11	0.03100000
12	0.02750921
13	0.02495733
14	0.03011757
15	0.03567534
16	0.04196944
17	0.03794507
18	0.03928714
19	0.03152899
20	0.01960588
21	0.01309254
22	0.02245000
23	0.02430769
24	0.03020781
25	0.02952381
26	0.02995484
27	0.02679836
28	0.03124500
29	0.03719153
30	0.03680345
31	0.03482456
32	0.03345179
33	0.04608182
34	0.04036111
35	0.03534151
36	0.04406923
37	0.02698627

signup_week	avg_retention
38	0.03511200
39	0.03619592
40	0.03711667
41	0.03701915
42	0.03535000
43	0.04594889
44	0.04356591
45	0.04472558
46	0.04328810
47	0.04016341
48	0.04381250
49	0.04464872
50	0.04392368
51	0.02913514
52	0.05150000
53	0.04171714
54	0.06961765
55	0.04691818
56	0.06587188
57	0.06960968
58	0.06617000
59	0.06537931
60	0.05814286
61	0.06989259
62	0.06818462
63	0.09692000
64	0.09740833
65	0.09789130
66	0.11110455
67	0.11587619
68	0.11424000
69	0.20154211
70	0.20962222
71	0.21525882
72	0.24218750
73	0.23460667
74	0.25291429
75	0.27746923
76	0.25472500
77	0.29469091
78	0.31642000
79	0.34884444
80	0.34651250
81	0.38008571
82	0.42998333

signup_week	avg_retention
83	0.46736000
84	0.51837500
85	0.58240000
86	0.58300000



- Recent cohorts (week 84-86) have 1-2 weeks of data, while week 0 cohorts have ~86 weeks averaged.
- High retention averages for recent cohorts are based on limited data, likely not representing a true improvement over the full timeframe.
- Gradual decline in average retention for earlier sign-up weeks, reaching a lower plateau, then a notable increase for recent sign-up weeks.
- Focus on cohorts with a more extended engagement history (e.g., week 0) for assessing long-term stability and trends.
- Recent sharp rise can be attributed to the fact that the majority of user engagement started from week 69 onwards.

While recent cohorts show positive trends, caution is needed due to the limited data period. Focus on cohorts with an extended engagement history for more reliable insights into user retention and assess whether the observed rates are sustainable over a more extended period.

Weekly Engagement Per Device

```
# Weekly engagement per device
CREATE TABLE
  devices (device VARCHAR(50), type VARCHAR(10));

INSERT INTO # table to classify devices
  devices (device, type)
VALUES
  ('dell inspiron notebook', 'laptop'),
  ('iphone 5', 'phone'),
```

```

('iphone 4s', 'phone'),
('windows surface', 'tablet'),
('macbook air', 'laptop'),
('iphone 5s', 'phone'),
('macbook pro', 'laptop'),
('kindle fire', 'tablet'),
('ipad mini', 'tablet'),
('nexus 7', 'tablet'),
('nexus 5', 'phone'),
('samsung galaxy s4', 'phone'),
('lenovo thinkpad', 'laptop'),
('samsung galaxy tablet', 'tablet'),
('acer aspire notebook', 'laptop'),
('asus chromebook', 'laptop'),
('samsung galaxy note', 'phone'),
('mac mini', 'desktop'),
('hp pavilion desktop', 'desktop'),
('ipad air', 'tablet'),
('htc one', 'phone'),
('dell inspiron desktop', 'desktop'),
('amazon fire phone', 'phone'),
('acer aspire desktop', 'desktop'),
('nokia lumia 635', 'phone'),
('nexus 10', 'tablet');

WITH cte AS (
    SELECT
        DISTINCT FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7) AS week,
        e.device,
        type,
        user_id
    FROM
        events AS e
        JOIN devices AS d ON e.device = d.device
    WHERE
        event_type = 'engagement'
)
SELECT
    week,
    SUM(
        CASE
            WHEN type = 'desktop' THEN 1
            ELSE 0
        END
    ) AS desktop,
    SUM(
        CASE
            WHEN type = 'laptop' THEN 1
            ELSE 0
        END
    ) AS laptop,
    SUM(
        CASE
            WHEN type = 'tablet' THEN 1
            ELSE 0
        END
    ) AS tablet

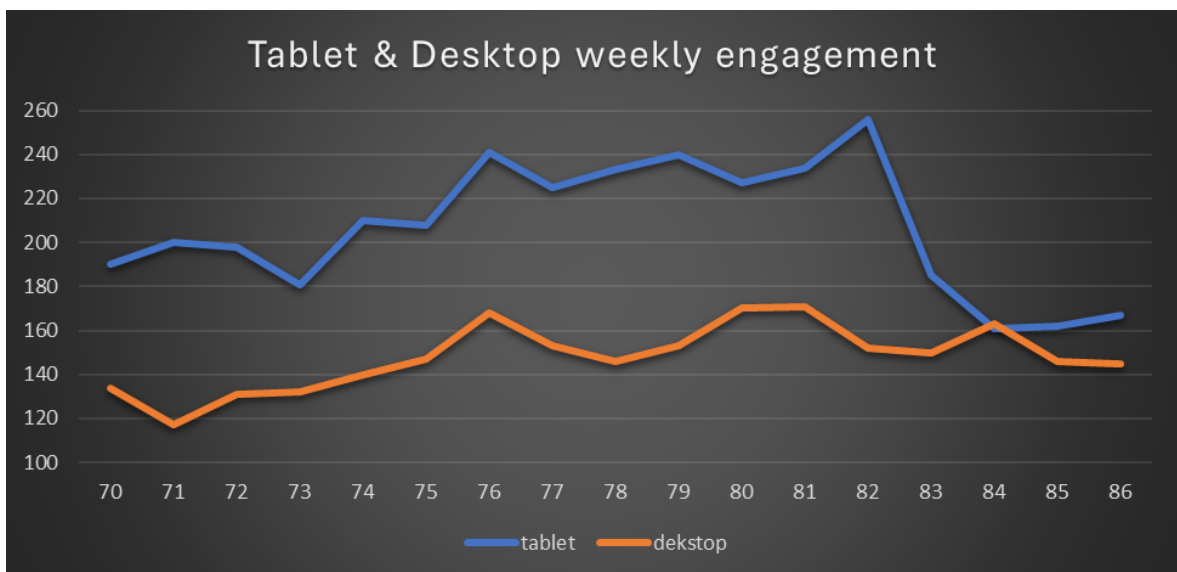
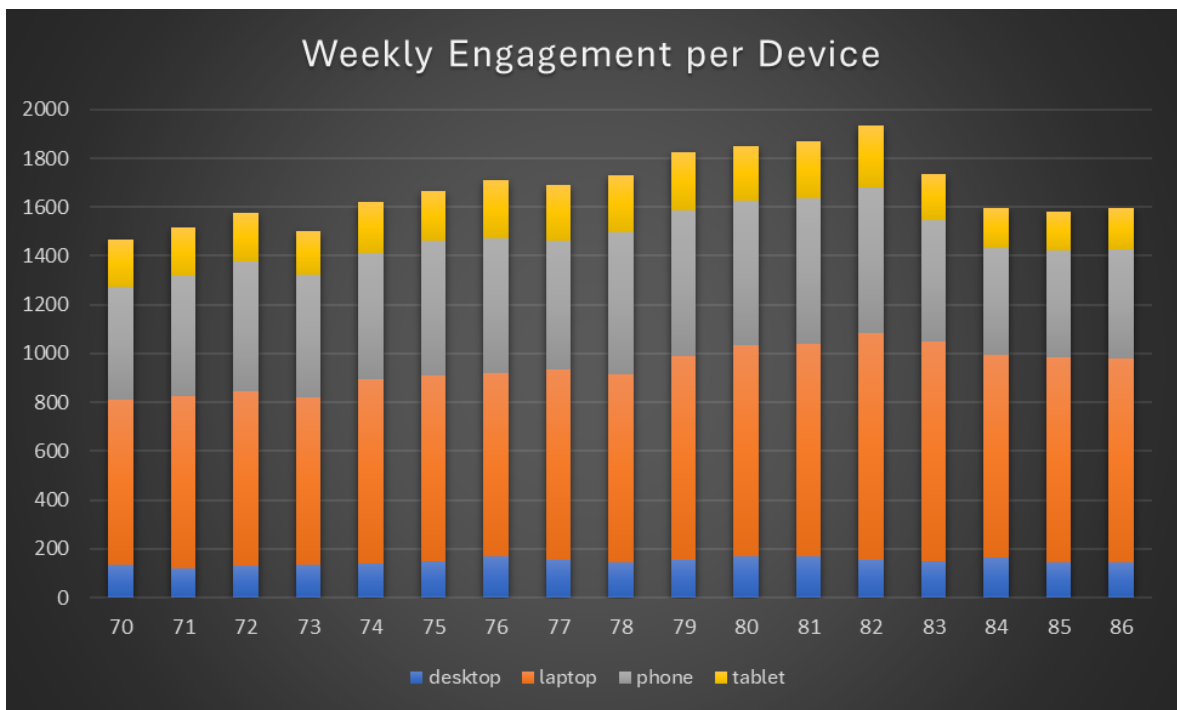
```

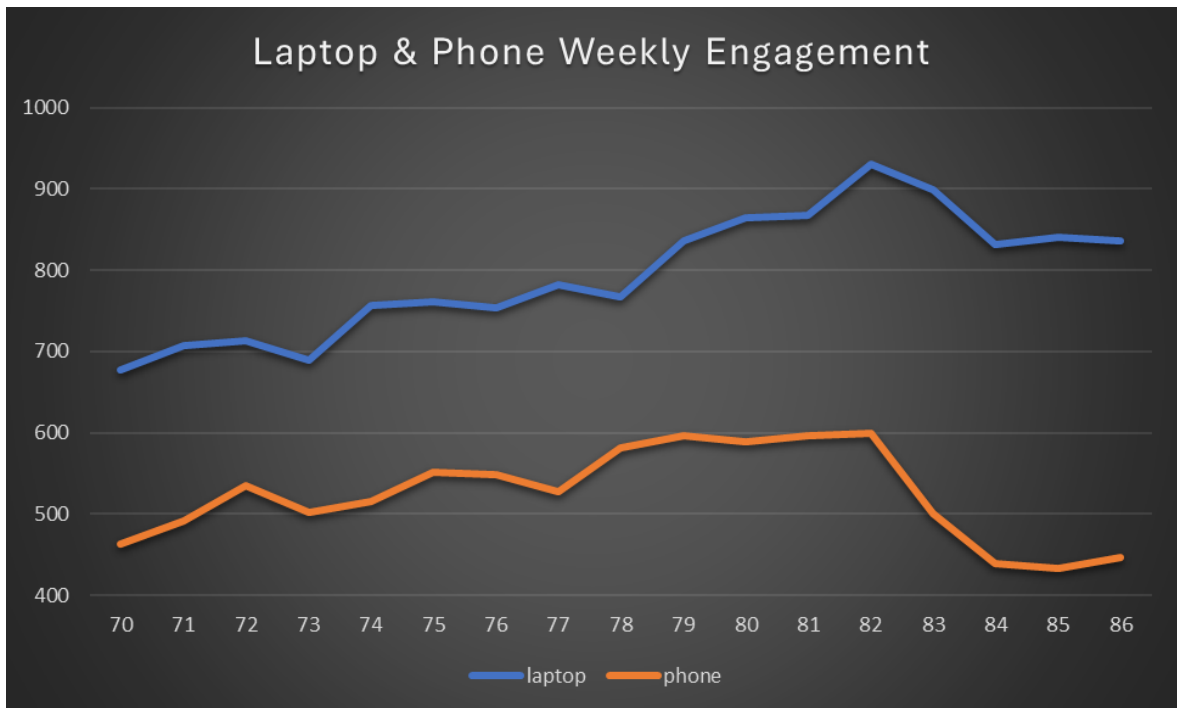
```

        END
    ) AS tablet,
    SUM(
        CASE
            WHEN type = 'phone' THEN 1
            ELSE 0
        END
    ) AS phone
FROM
    cte
GROUP BY
    week;

```

week	desktop	laptop	tablet	phone
69	47	370	104	264
70	134	678	190	463
71	117	707	200	492
72	131	713	198	535
73	132	689	181	502
75	147	761	208	552
74	140	757	210	515
76	168	754	241	548
77	153	783	225	528
81	171	867	234	596
78	146	768	233	582
82	152	930	256	599
80	170	865	227	589
79	153	836	240	596
83	150	899	185	500
84	163	832	161	439
85	146	841	162	433
86	145	836	167	447
87	5	61	12	26
Average (excluding week 87)	148.12	795.06	524.47	206.94





- Laptop consistently has the highest engagement, followed by phone, while tablet and desktop lag behind. Laptop engagement shows a steadily increasing trend over time, reaching 930 by week 82.
- Tablet engagement remains in the 180-250 range, indicating a small but steady user base. Desktop engagement shows a general upward trend, peaking at 171 in week 81. Phone engagement is more variable but peaks in week 82.
- Laptop dominance increases over time, accounting for 52% of total engagement by week 86. Phone share declines from 32% to 28%.
- In recent weeks, variance in engagement share is higher, with laptops taking a disproportionate share. Early weeks show a more equal distribution of engagement share.

While there isn't a universal trend affecting all devices, individual analyses suggest potential areas for improvement. Focusing on enhancing the user experience for desktops and laptops due to their rising engagement trends could prove beneficial. Additionally, investigating reasons behind inconsistent tablet engagement and capitalizing on phone engagement peaks could further drive overall user interactions.

Email Engagement Analysis

```
# Weekly Email Engagement
SELECT
    FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7) AS week,
    SUM(
        CASE
            WHEN action LIKE 'sent%' THEN 1
            ELSE 0
        END
    ) AS deliveries,
    SUM(
        CASE
            WHEN action = 'email_clickthrough' THEN 1
            ELSE 0
        END
    ) AS clicks,
```

```

SUM(
    CASE
        WHEN action = 'email_open' THEN 1
        ELSE 0
    END
) AS opens,
SUM(
    CASE
        WHEN action = 'email_clickthrough' THEN 1
        ELSE 0
    END
) / SUM(
    CASE
        WHEN action LIKE 'sent%' THEN 1
        ELSE 0
    END
) AS CTR,
SUM(
    CASE
        WHEN action = 'email_clickthrough' THEN 1
        ELSE 0
    END
) / SUM(
    CASE
        WHEN action = 'email_open' THEN 1
        ELSE 0
    END
) AS CTOR
FROM
    email_events
GROUP BY
    FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7)
ORDER BY
    week;

# Weekday email engagement
SELECT
    DAYNAME(occurred_at) AS weekday,
    WEEKDAY(occurred_at) AS num_weekday,
    SUM(
        CASE
            WHEN action LIKE 'sent%' THEN 1
            ELSE 0
        END
    ) AS deliveries,
    SUM(
        CASE
            WHEN action = 'email_clickthrough' THEN 1
            ELSE 0
        END
    ) AS clicks,
    SUM(
        CASE
            WHEN action = 'email_open' THEN 1
            ELSE 0

```



```

        END
    ) AS opens,
    SUM(
        CASE
            WHEN action = 'email_clickthrough' THEN 1
            ELSE 0
        END
    ) / SUM(
        CASE
            WHEN action LIKE 'sent%' THEN 1
            ELSE 0
        END
    ) AS CTR,
    SUM(
        CASE
            WHEN action = 'email_clickthrough' THEN 1
            ELSE 0
        END
    ) / SUM(
        CASE
            WHEN action = 'email_open' THEN 1
            ELSE 0
        END
    ) AS CTOR
    SUM(
        CASE
            WHEN action = 'email_open' THEN 1
            ELSE 0
        END
    ) / SUM(
        CASE
            WHEN action LIKE 'snet%' THEN 1
            ELSE 0
        END
    ) AS `OR`
FROM
    email_events
GROUP BY
    DAYNAME(occurred_at),
    WEEKDAY(occurred_at)
ORDER BY
    2;

# Weekly re - engagement metrics
SELECT
    FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7) AS week,
    COUNT(*) AS re_enagements
FROM
    email_events
WHERE
    action = 'sent_reengagement_email'
GROUP BY
    FLOOR(DATEDIFF(occurred_at, '2012-12-30') / 7);

# Weekday re - enagement metrics

```

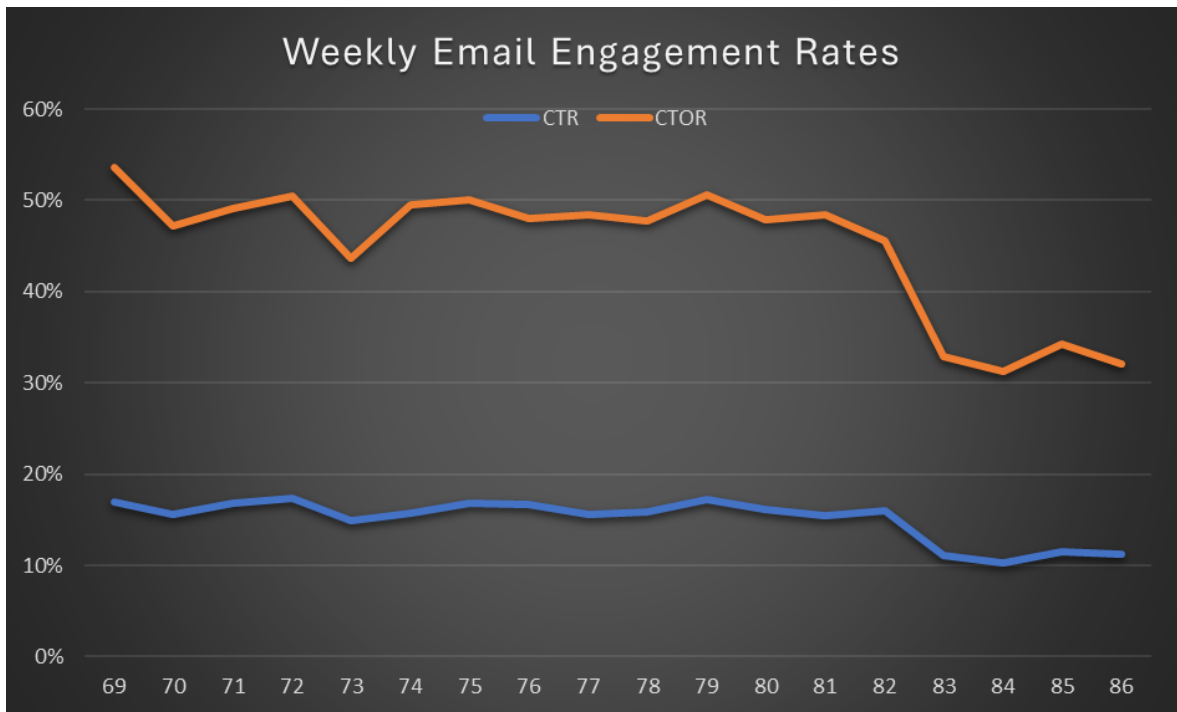
```

SELECT
    dayname(occurred_at) AS weekday,
    weekday(occurred_at) AS num_weekday,
    COUNT(*) AS re_engagements
FROM
    email_events
WHERE
    action = 'sent_reengagement_email'
GROUP BY
    dayname(occurred_at),
    weekday(occurred_at)
ORDER BY
    2;

```

Weekly Analysis

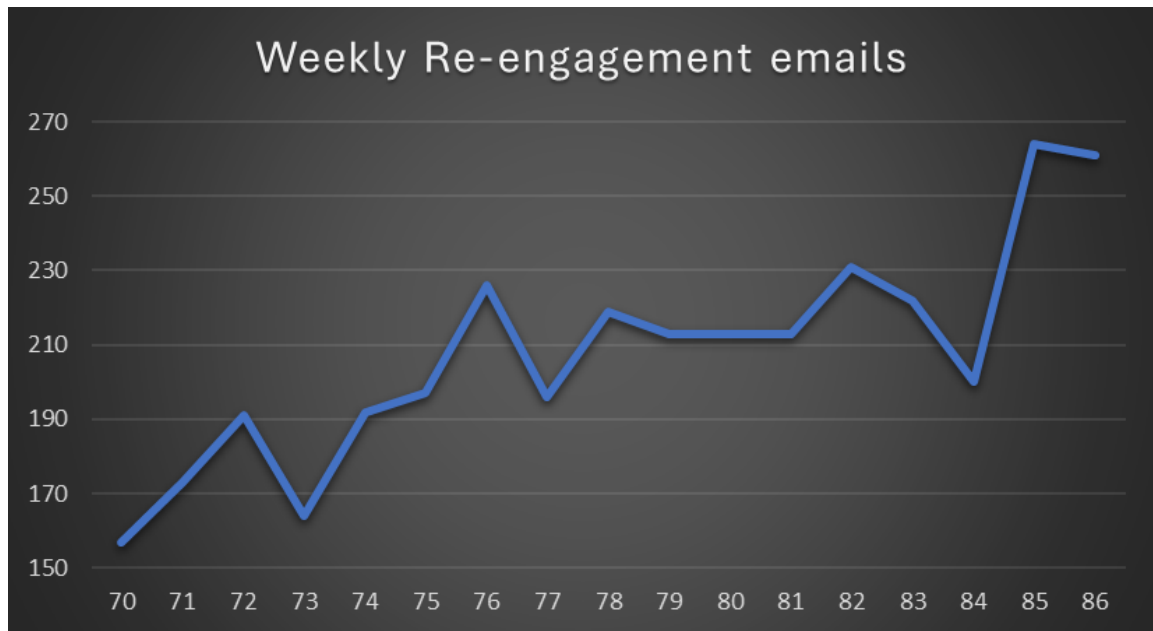
week	deliveries	clicks	opens	CTR	CTOR
69	981	166	310	0.1692	0.5355
70	2759	430	912	0.1559	0.4715
71	2838	477	972	0.1681	0.4907
72	2924	507	1004	0.1734	0.5050
73	2986	443	1014	0.1484	0.4369
74	3103	488	987	0.1573	0.4944
75	3200	538	1075	0.1681	0.5005
76	3331	554	1155	0.1663	0.4797
77	3403	530	1096	0.1557	0.4836
78	3521	556	1165	0.1579	0.4773
79	3612	621	1228	0.1719	0.5057
80	3712	599	1250	0.1614	0.4792
81	3805	590	1219	0.1551	0.4840
82	3937	630	1383	0.1600	0.4555
83	4015	445	1351	0.1108	0.3294
84	4097	418	1337	0.1020	0.3126
85	4276	490	1432	0.1146	0.3422
86	4372	490	1528	0.1121	0.3207
87	48	38	41	0.7917	0.9268



- Deliveries show an upward trend over time as the user base grows, increasing 4.5x from 981 in week 69 to 4,372 in week 86.
- Opens and clicks also grow over time but with some fluctuations week to week. Generally opens are 33% of deliveries and clicks are 15-17% of deliveries.
- Clickthrough rate (CTR) averages around 15% most weeks, ranging from 10.2% to 17.3%. There is no clear trend in CTR over time.
- Open rates (CTOR) average 45-50% typically. The low is 31.3% in week 84. The highs are 53.6% in week 69 and 50.6% in week 79.
- Week 69 saw high engagement with CTR at 16.9% and CTOR at 53.6%. This could be an initial honeymoon period/lower delivery volume.
- Weeks 83-84 saw a dip in both CTR (10.2%) and CTOR (31.3%), which could be investigated further.
- Aside from those outlier weeks, engagement metrics are relatively stable week to week. No major chemistry is evident.

week	re_engagements
69	73
70	157
71	173
72	191
73	164
74	192
75	197
76	226
77	196
78	219
79	213
80	213
81	213

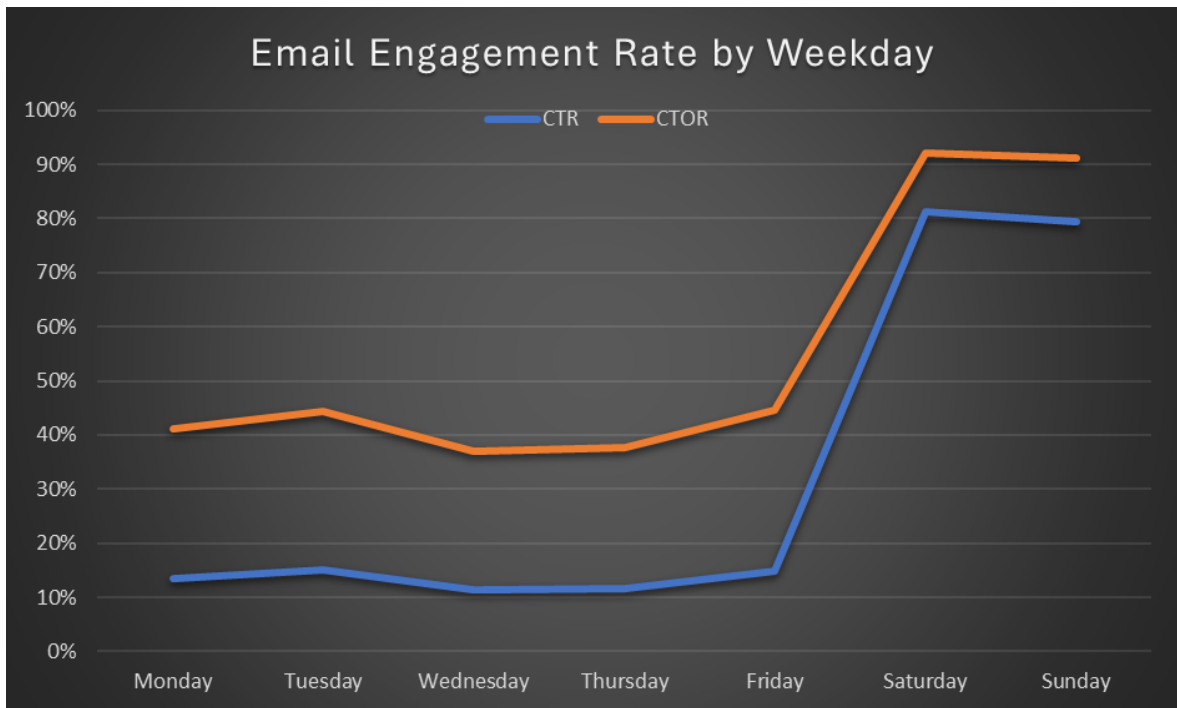
week	re_engagements
82	231
83	222
84	200
85	264
86	261
87	48



- There is a general increasing trend in the number of re-engagement emails sent over the weeks, with occasional fluctuations. This suggests a proactive approach to re-engage users.
- A notable spike is observed in re-engagement emails sent during weeks 85 and 86. This could be a strategic effort to re-engage users during a specific period, possibly in response to observed user behavior or external factors.
- The number of re-engagement emails has been relatively consistent in recent weeks (weeks 78-81), indicating a stable approach to user re-engagement during this period.

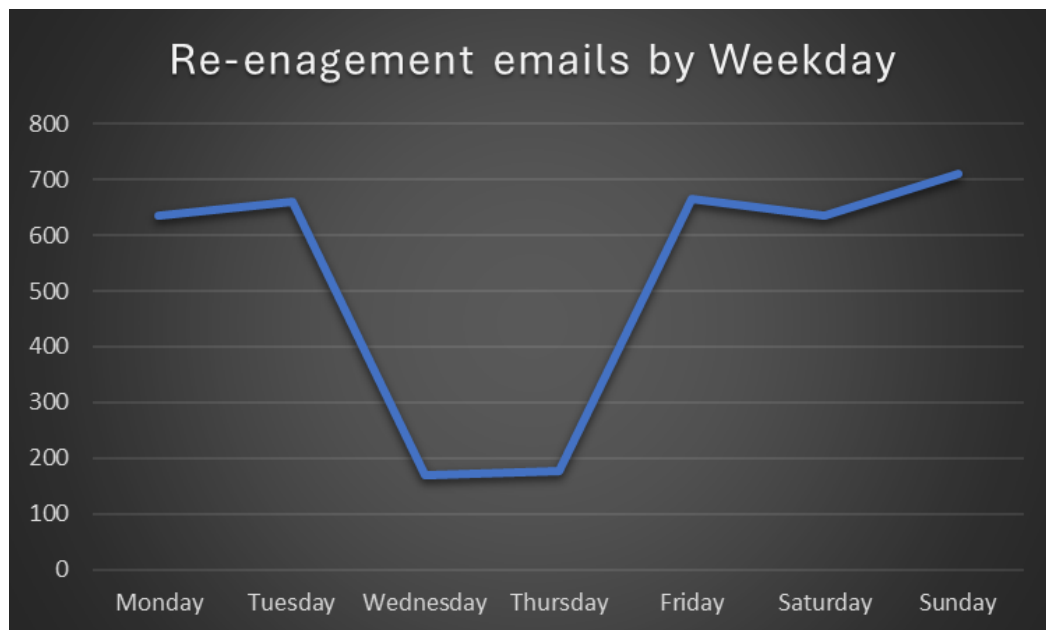
Weekday Analysis

weekday	num_weekday	deliveries	clicks	opens	CTR	CTOR
Monday	0	16562	2230	5431	0.1346	0.4106
Tuesday	1	10983	1649	3711	0.1501	0.4444
Wednesday	2	9789	1110	2996	0.1134	0.3705
Thursday	3	11060	1289	3430	0.1165	0.3758
Friday	4	11180	1651	3711	0.1477	0.4449
Saturday	5	635	516	561	0.8126	0.9198
Sunday	6	711	565	619	0.7947	0.9128



- Weekday deliveries are highest on Mondays with 16,562, followed by Thursdays and Fridays. Weekday engagement remains relatively stable with some fluctuations.
- Saturday and Sunday have significantly lower deliveries (only re-engagement emails), but make up for it with very high engagement rates. These high engagement rates can be attributed to:
 - With lower delivery volumes on weekends, each click and open has a more pronounced effect on the overall engagement rates.
 - Since re-engagement emails are specifically crafted to prompt user actions (clicks and opens), the observed high engagement rates on weekends align with the intended purpose of these emails.
- Clickthrough rate (CTR) ranges from 11-15% on weekdays, but jumps to 79-81% on weekends. Similarly, open rates (CTOR) are 41-45% on weekdays and 91-92% on weekends.
- Wednesday has the lowest CTR and CTOR of the weekdays. Focusing efforts on improving midweek engagement could have an impact.
- Of the weekdays, Tuesday sees the highest CTR at 15% while Friday has the highest CTOR at 44.5%.

weekday	num_weekday	re_engagements
Monday	0	635
Tuesday	1	660
Wednesday	2	169
Thursday	3	177
Friday	4	666
Saturday	5	635
Sunday	6	711



- There is a notable number of re-engagement emails sent on Saturday and Sunday. This is interesting, especially considering that weekends typically see lower user engagement as shown in an earlier section. The decision to target users on weekends might be based on the necessity to drive engagement on the weekends.
- The decline in re-engagement emails on Wednesday and Thursday could be a deliberate strategy or an area for improvement. Wednesdays and Thursdays may pose challenges for re-engagement, possibly due to midweek fatigue or lower user responsiveness. Given the fact that user engagement peaks on Fridays, operational resources can be re-allocated towards Wednesday/Thursday to drive overall engagement.

Results

This project deepened my grasp of operational analytics, emphasizing the vital role of advanced SQL skills in interpreting user engagement trends and growth strategies. The insights directly inform strategic decisions, offering a clear path for optimal feature rollouts, targeted marketing, and user acquisition. Addressing duplicate data and exploring language dynamics improves data quality, while throughput analysis enhances operational preparedness. Insights on weekly engagement per device guide targeted UI/UX improvements, enhancing overall user satisfaction and engagement. Identifying peak days for re-engagement emails and addressing midweek declines optimize email campaign effectiveness. The project not only yielded immediate insights but also contributed to skill enhancement, positioning me effectively for future challenges.