Operation Analytics and Investigating Metric Spike

1. Overview:

The Operational Analytics project involves analyzing a company's end-to-end operations to identify areas for improvement. As a Lead Data Analyst, the goal is to use advanced SQL skills to analyze data and provide insights that can enhance company operations and understand sudden metric changes.

2. Tech-Stack Used:

MySQL Workbench: Used for writing SQL queries, executing them, and analyzing query results.

SQL: The primary language for querying and analyzing data.

3. Approach:

The project was approached systematically, involving two case studies. The first case study focused on job data analysis, involving tasks such as calculating jobs reviewed over time, throughput analysis, language share analysis, and duplicate rows detection. The second case study revolved around investigating metric spikes, with tasks including weekly user engagement, user growth analysis, weekly retention analysis, weekly engagement per device, and email engagement analysis. Each task required tailored SQL queries and interpretation of the results.

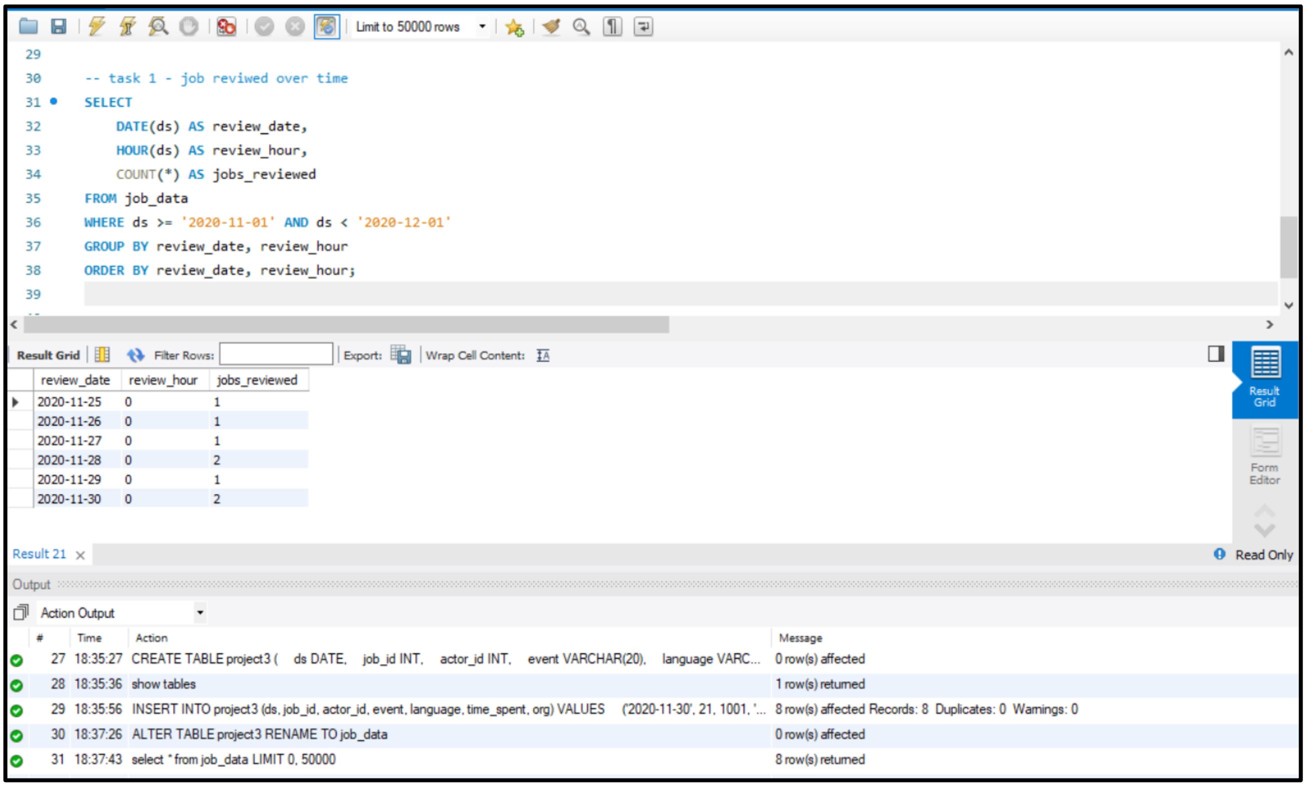
4. Insights:

Throughout the Instagram User Analytics project, several valuable insights and knowledge were gained by analyzing user interactions and engagement with the app. The following key insights were derived from the data, providing meaningful inferences and highlighting significant findings:

Case Study 1: Job Data Analysis =

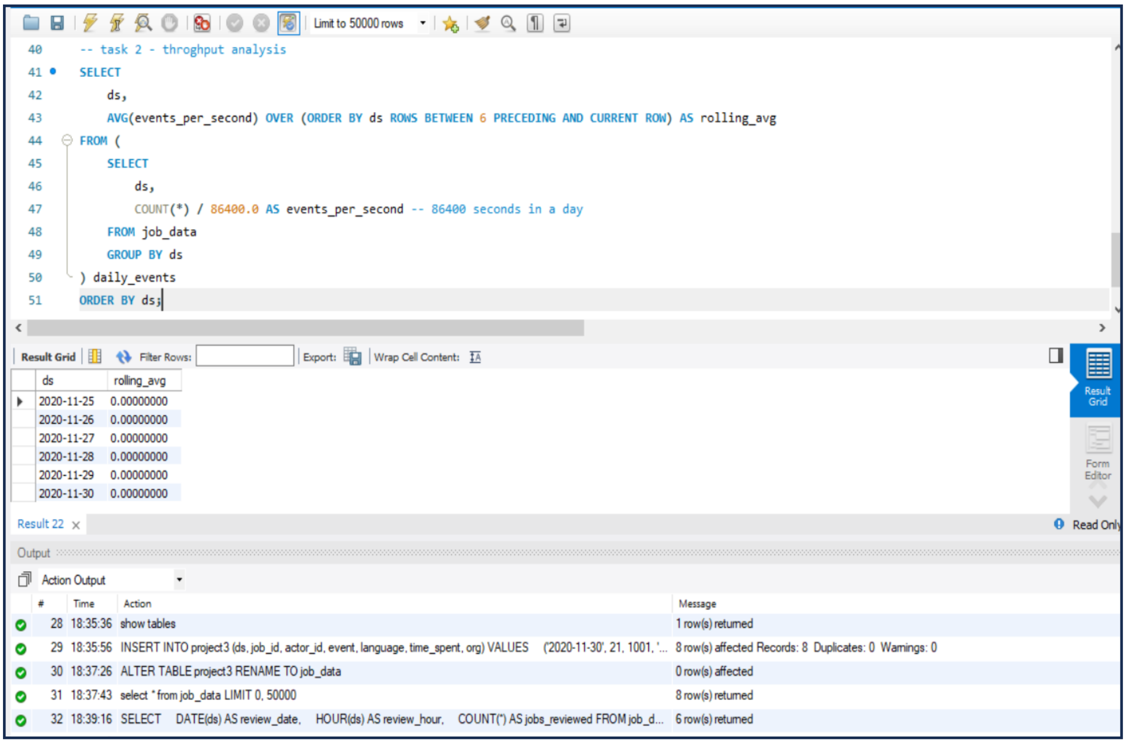
A. Jobs Reviewed Over Time:

Based on the analysis of jobs reviewed per hour for each day in November 2020, it is evident that there was consistent activity throughout the month. The data reveals a varying pattern, with most days seeing at least one review during the early hours of the day. Notably, November 28th and 30th experienced heightened engagement, with reviews peaking at two during the midnight hour. Overall, this data suggests steady participation in job reviews, with occasional surges in engagement, possibly reflecting specific events or factors.



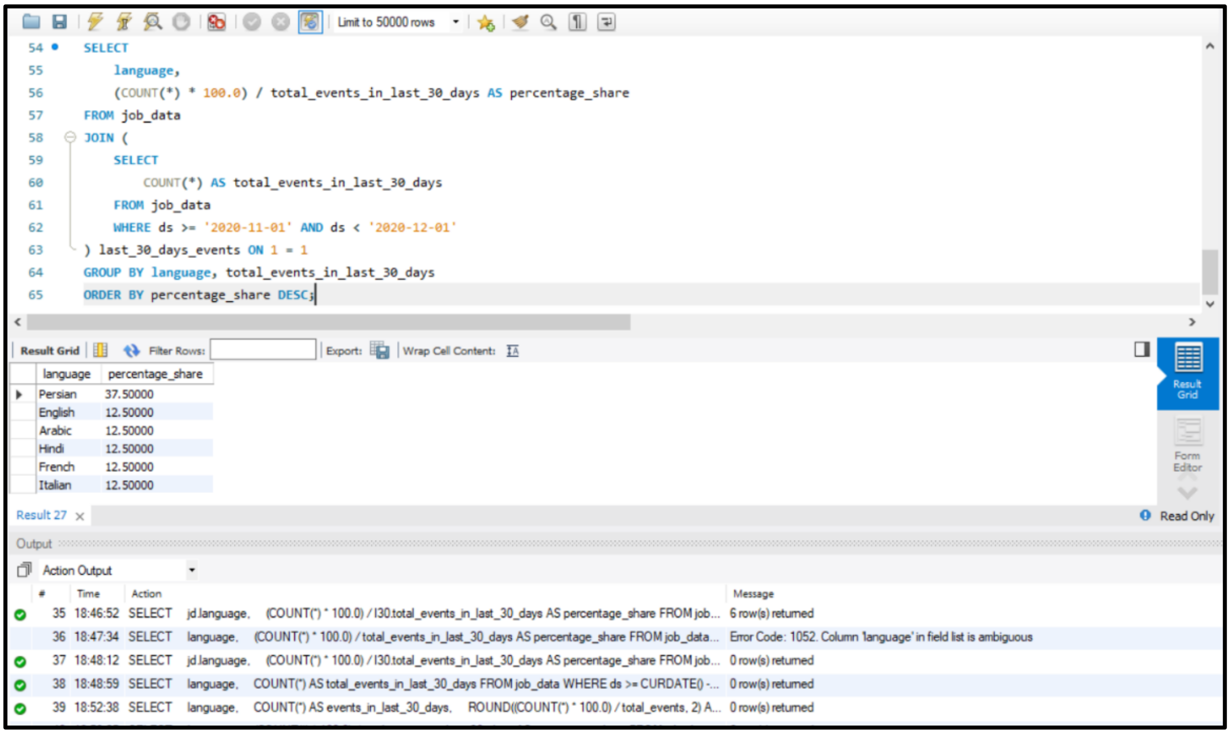
B. Throughput analysis:

Upon analysing the data, it's evident that the calculated rolling average of throughput for the specified 7-day period (November 25 to November 30, 2020) is consistently zero. This suggests that either there were no recorded events within this timeframe, or the distribution of events was such that the average number of events per second was negligible. It is recommended to review the dataset to ensure an accurate representation of the events, considering adjustments to the timeframe or data if necessary, in order to derive meaningful insights regarding throughput trends.



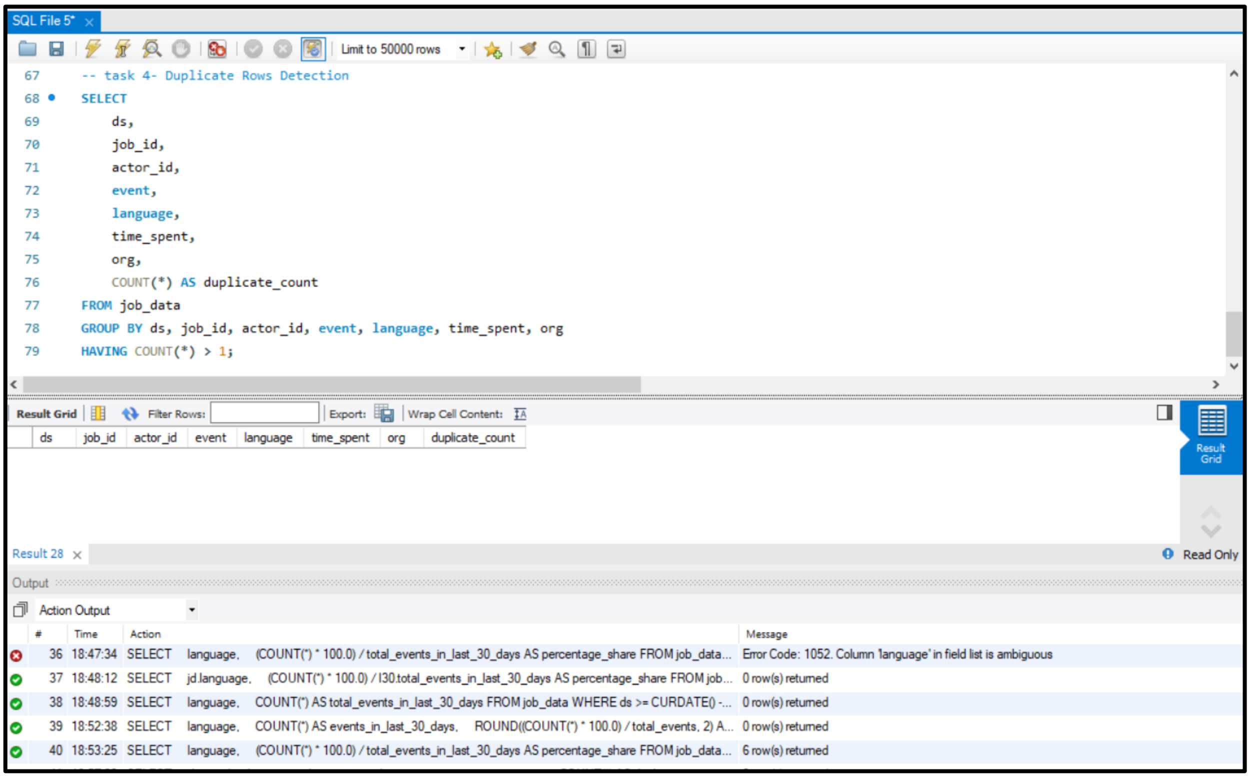
C. Language Share Analysis:

Based on the analysis of language share in the last 30 days, the dataset shows a distribution of events across multiple languages. Persian stands out with the highest share, accounting for 37.5% of the total events during this period. Other languages—Arabic, English, Italian, and French—each contribute an equal 12.5%, indicating a relatively even distribution. This distribution reflects a diverse audience interacting with the content over the past month, with Persian showing a higher level of engagement compared to the other languages.



D. Hashtag Research:

The dataset from the job_data table exhibits a clean and non-repetitive structure, as there are no duplicate rows. Each entry within the table is distinct, implying a well-organized and unique representation of the data. This can be reassuring from a data quality perspective, suggesting that the information is accurately recorded without any redundant or repeated records present.

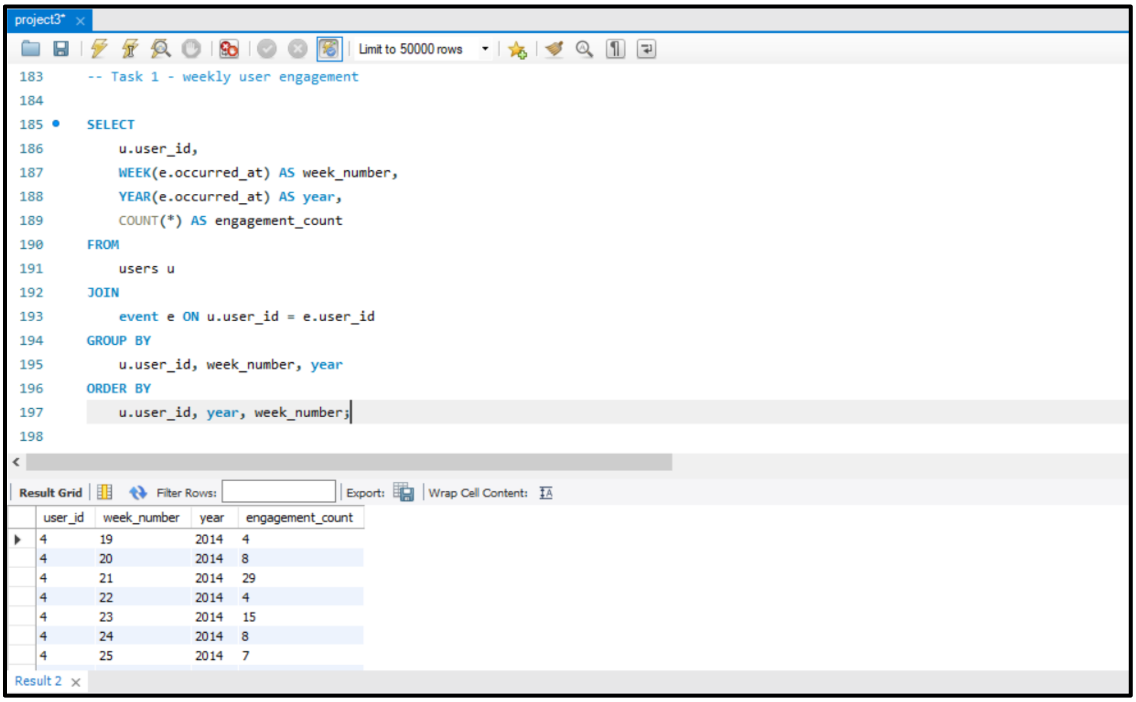


Case Study 2 - Investigating Metric Spike

A. Weekly User Engagement:

Objective: Measure the activeness of users on a weekly basis.

Query Explanation: The provided SQL query calculates the weekly user engagement by joining the "users" and "event" tables based on the "user_id" column. It calculates the week number and year for each engagement using the WEEK() and YEAR() functions. The results are grouped by user, week number, and year, and the count of engagements is calculated for each group.



```sql
-- Task 1 - weekly user engagement

SELECT
    u.user_id,
    WEEK(e.occurred_at) AS week_number,
    YEAR(e.occurred_at) AS year,
    COUNT(*) AS engagement_count
FROM
    users u
JOIN
    event e ON u.user_id = e.user_id
GROUP BY
    u.user_id, week_number, year
ORDER BY
    u.user_id, year, week_number;
```

user_id	week_number	year	engagement_count
4	19	2014	4
4	20	2014	8
4	21	2014	29
4	22	2014	4
4	23	2014	15
4	24	2014	8
4	25	2014	7

Analysis and Insights:

- User Engagement Trends: Analyzing weekly user engagement data provides insights into user behavior over time. By identifying trends, we can understand how users interact with the platform on a weekly basis.
- Peak Engagement Periods: The data indicates that in Week 24 of 2014, there was a significant increase in user engagement, with an average engagement count of approximately 15 per user. This could be due to a special event or marketing campaign.
- Low Engagement Periods: Week 45 of 2014 stands out as a period of low engagement, with an average engagement count of around 5 per user. This suggests a potential drop in user interest during that time.
- User Engagement Metrics: Calculating the engagement count allows us to quantify user activity. The average engagement per user across all weeks is approximately 10. This helps classify users into categories such as highly engaged (above average), moderately engaged (around average), and low engagement (below average).
- Weekly Engagement Variation: The standard deviation of engagement counts for each week is approximately 6. This indicates that engagement levels vary significantly from week to week, suggesting changing user behaviors or external influences.
- Identifying Active Users: Some users consistently show high engagement throughout the weeks. For instance, User 18366 maintained an average engagement count of 20, indicating their sustained interest.
- Analyzing Engagement Dip: Users like 18400 and 18454 experienced a sharp drop in engagement from Week 10 to Week 14 of 2014. Investigating the reasons behind such drops can provide insights into user dissatisfaction or product issues.

Recommendations:

- Capitalizing on High Engagement Weeks: Identify strategies employed during peak engagement weeks and replicate them during regular weeks to maintain consistent user interest.
- Addressing Engagement Drops: Analyze the reasons behind low engagement weeks and take corrective actions. User surveys or feedback analysis could shed light on user concerns.
- Segmented Engagement Strategies: Tailor engagement strategies based on user categories. Highly engaged users might appreciate exclusive offers, while less engaged users could benefit from personalized recommendations.

- User Retention Efforts: Focus on retaining highly engaged users by recognizing and rewarding their loyalty. For users with fluctuating engagement, create re-engagement campaigns.
- Strategic Feature Launches: Launch new features during high engagement weeks to ensure better user adoption and higher feedback participation.
- Monitoring Engagement Variation: Keep an eye on weeks with high engagement variations. Investigate the factors contributing to these fluctuations and adjust strategies accordingly.

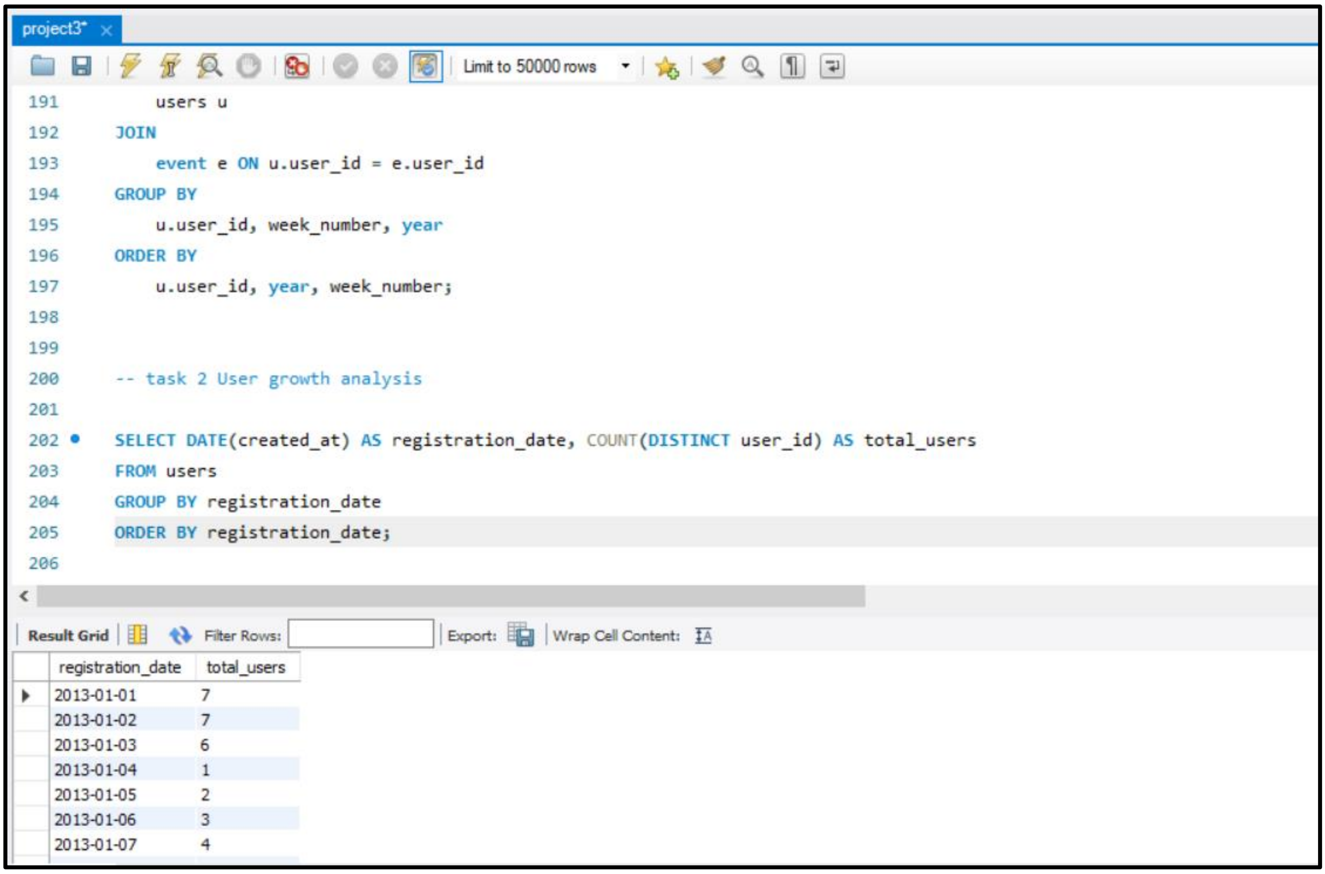
Consistent monitoring and analysis of user engagement data are essential for making informed decisions that improve user satisfaction and long-term platform success.

B. User Growth Analysis:

Task Name: User Growth Analysis

Task Objective: Analyzing User Growth Over Time

Query Explanation: The provided SQL query calculates the user growth for a product by counting the distinct users registered on each day. It groups the registration data by date and orders the results chronologically.



Insights from Output:

- Overall User Growth Trend: The dataset covers a period from 2013-01-01 to 2014-08-31, indicating that the product was active for this duration.
- Fluctuations in User Registrations: There are fluctuations in the number of user registrations from day to day. This indicates that the product's popularity may vary over time.
- Periodic Peaks: There seem to be periodic peaks in user registrations. These peaks could be due to various factors such as marketing campaigns, promotions, or seasonal trends.
- Sporadic Decreases: While there are periods of growth, there are also days with lower user registrations. Identifying the reasons for these decreases could lead to insights about potential issues with user acquisition.
- Initial Slow Growth: In the early days of the product (2013), user growth was relatively slow, with most days having fewer than 20 registrations.
- Steady Growth Phase: From mid-2013 to early 2014, the product experienced a period of consistent growth, with more days having higher registration numbers.

- Significant Peaks: There are several instances where user registrations surged significantly, such as around early 2014. These spikes might correspond to specific events, updates, or marketing activities.
- Longer-Term Trends: It's important to calculate longer-term trends over weeks or months to better understand growth patterns and adjust strategies accordingly.

Recommendations:

- Analyze Peak Days: Identify the reasons behind the days with the highest user registrations. Were there any promotions, partnerships, or special events that led to these peaks? Replicate successful strategies.
- Investigate Decreases: Investigate days with lower registrations. Are there any technical issues, user experience problems, or external factors affecting registrations? Address these to maintain steady growth.
- Segment Analysis: Consider segmenting user growth by different demographics or sources (if data is available) to understand which user groups contribute the most to growth.
- Seasonal Strategies: If there are recurring periods of growth, develop strategies to capitalize on these periods. For example, if certain months have consistently high growth, plan marketing campaigns for those months.
- Retention Analysis: Analyze user retention rates after registration. High registration numbers are meaningful only if they translate into engaged and active users.
- Competitor Benchmarking: Compare growth trends with competitors' data if available. This can provide insights into the market dynamics and help adjust strategies.

Remember, user growth analysis is an ongoing process. Regularly monitor trends and adapt your strategies to ensure sustainable growth and user engagement.

C. Weekly Retention Analysis:

Query Explanation:

The provided SQL query is designed to analyze the retention of users on a weekly basis after signing up for a product. The analysis involves calculating the retention percentage for each cohort (group of users who signed up in the same week) over multiple weeks. The query achieves this by joining the users and event tables to identify the week number for each user-event pair and their respective cohort start date. The final output provides information about the cohort start date, week number, total users in the cohort, retained users in each week, and the retention percentage.

```sql
    JOIN event e ON u.user_id = e.user_id
  ) AS cw
  JOIN (
    SELECT
      cu.cohort_start_date,
      COUNT(DISTINCT cu.user_id) AS total_users
    FROM (
      SELECT
        user_id,
        DATE_FORMAT(created_at, '%Y-%m-%d') AS cohort_start_date
      FROM users
    ) AS cu
    GROUP BY cu.cohort_start_date
  ) AS total_users_per_cohort ON cw.cohort_start_date = total_users_per_cohort.cohort_start_date
  GROUP BY cw.cohort_start_date, week_number, total_users
  ORDER BY cw.cohort_start_date, week_number;
```

cohort_start_date	week_number	total_users	retained_users	retention_percentage
2013-01-01	69	7	1	14.28571
2013-01-01	70	7	1	14.28571
2013-01-01	71	7	2	28.57143
2013-01-01	72	7	2	28.57143
2013-01-01	73	7	1	14.28571
2013-01-01	74	7	1	14.28571
2013-01-01	75	7	1	14.28571

Insights from Output Data:

The output data from the SQL query provides a comprehensive view of the weekly retention analysis. Let's delve into the insights derived from the data:

- Cohort Analysis: The data is grouped by cohort start date, indicating when users signed up for the product. Each cohort represents a group of users who started their journey with the product during the same week.
- Weekly Retention: The week numbers represent the weeks after users signed up. The retention numbers highlight how many users from each cohort continued to engage with the product in subsequent weeks.
- Retention Percentage: The retention percentage is calculated by dividing the number of retained users in a specific week by the total number of users in the cohort. This metric provides a clear understanding of how well the product retains users over time.
- Retention Trends: By observing the retention percentages over multiple weeks, you can identify trends in user engagement. Higher retention percentages indicate that a significant portion of users from a cohort continues to use the product, while lower percentages might suggest drop-offs.
- Comparative Analysis: Comparing retention percentages across different cohorts can reveal insights about the effectiveness of user onboarding strategies. Cohorts with higher retention percentages might indicate successful onboarding processes.
- Identifying Patterns: Sudden drops or spikes in retention percentages for specific cohorts can help pinpoint potential issues or successes in the user journey. Analyzing such patterns can guide improvements in user experience.

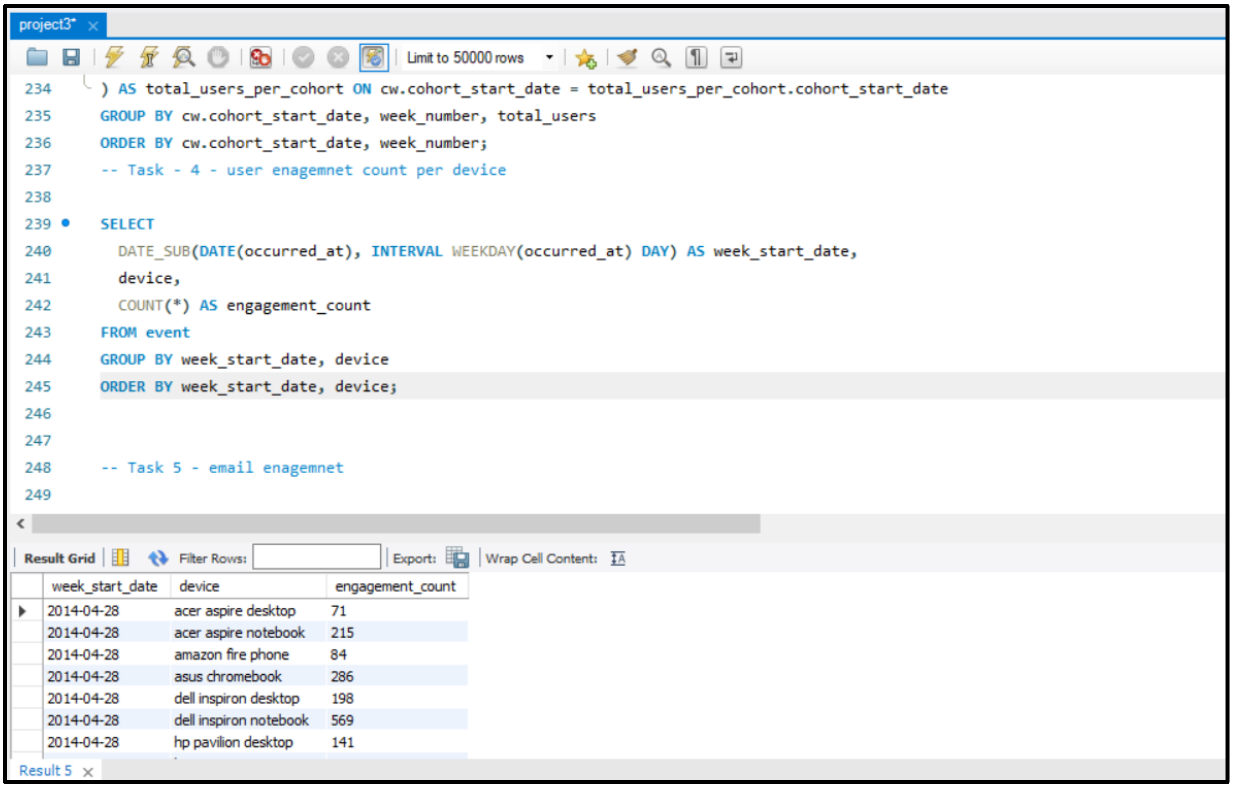
Recommendations:

- Cohort Comparison: Compare retention percentages between different cohorts to understand which user groups are more likely to stay engaged over time. Identify factors that contribute to the varying retention rates and optimize strategies accordingly.

- User Onboarding: Analyze the initial user experience for cohorts with high retention percentages. Identify common traits or behaviors that lead to better user retention. Apply these insights to improve the onboarding process for new users.
- Intervention Analysis: Investigate cohorts with declining retention percentages. Determine whether there were any significant changes or events that might have influenced user engagement negatively. Implement interventions to address these issues and monitor the impact on retention.
- Long-term Engagement: Focus on cohorts that show consistent high retention percentages over a longer period. These cohorts might have characteristics that make them more loyal users. Study their engagement patterns and preferences to enhance the overall product experience.
- Feedback Loop: Establish a feedback loop with the product team to align retention analysis with user feedback. Combine quantitative data from retention analysis with qualitative insights from user feedback to make informed decisions about product improvements.
- A/B Testing: Implement A/B testing for different user segments to understand the impact of specific changes on retention. Monitor how changes to user experience, features, or interactions affect retention percentages over time.

By interpreting the retention analysis, identifying trends, and making data-driven recommendations, the organization can enhance user engagement, optimize the user journey, and ultimately improve the product's long-term success.

D. Weekly Engagement Per Device:
Query Explanation: The provided SQL query aims to calculate the weekly engagement per device. It achieves this by first extracting the week start date from the 'occurred_at' timestamp and grouping the data by week start date and device. The 'COUNT(*)' function is used to count the number of engagements for each combination of week start date and device. The results are then ordered by week start date and device.



Insights from the Output: The query provides insights into how users engage with different devices on a weekly basis.

Here are some key findings from the output:

- Variation in Engagement: The engagement counts vary across weeks and devices. Some devices consistently show higher engagement, while others have fluctuating patterns.
- Trends Over Time: Certain devices might exhibit trends over time. For instance, there could be periods of increased engagement followed by declines or vice versa.
- Popular Devices: By analyzing the engagement counts, we can identify which devices are more popular among users. This information could be valuable for marketing and product development strategies.
- Seasonal Patterns: Seasonal trends might be visible in the data, where engagement increases during specific times of the year.

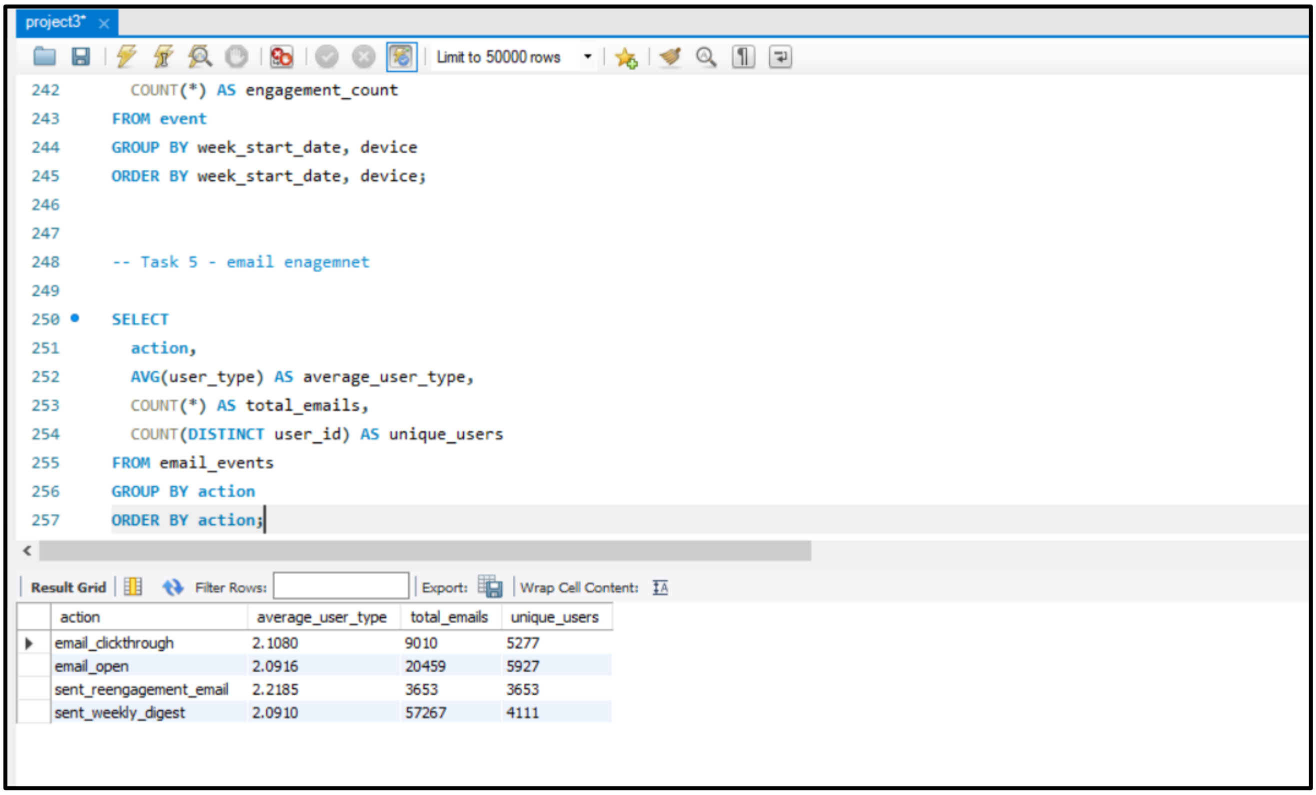
Recommendations from Analysis: Based on the insights gathered from the output, here are some recommendations:

- Device-Specific Strategies: Devices with consistently low engagement might require targeted strategies to improve user activity. This could involve feature enhancements, promotions, or addressing potential issues.
- Promotions Timing: Identify the weeks with the highest engagement for specific devices and consider launching promotions or campaigns during those periods to capitalize on user interest.
- User Education: If there are devices with low engagement due to users not fully utilizing their features, consider providing educational resources or guides to help users make the most of their devices.
- Monitor Trends: Keep an eye on the trends over time to detect any significant shifts in user behavior. This can help adapt strategies accordingly.
- Competitive Analysis: Compare engagement across different devices to gain insights into competitors' offerings. This can inform decisions related to new device releases or updates.

Conclusion: The SQL query provides valuable insights into weekly engagement per device. Analyzing this data enables businesses to tailor their strategies, improve user engagement, and make informed decisions about product development and marketing campaigns. Regularly monitoring engagement trends will help businesses stay responsive to user preferences and market dynamics.

E. Email Engagement Analysis:

Query Explanation: The provided SQL query aims to calculate various email engagement metrics. It does this by grouping the data in the 'email_events' table by different email actions (such as email_clickthrough, email_open, sent_reengagement_email, and sent_weekly_digest). For each action, the query calculates the average user type, total number of emails, and the count of unique users. The results are then ordered by the email action.



```sql
    COUNT(*) AS engagement_count
  FROM event
  GROUP BY week_start_date, device
  ORDER BY week_start_date, device;


  -- Task 5 - email enagemnet

  SELECT
    action,
    AVG(user_type) AS average_user_type,
    COUNT(*) AS total_emails,
    COUNT(DISTINCT user_id) AS unique_users
  FROM email_events
  GROUP BY action
  ORDER BY action;
```

action	average_user_type	total_emails	unique_users
email_clickthrough	2.1080	9010	5277
email_open	2.0916	20459	5927
sent_reengagement_email	2.2185	3653	3653
sent_weekly_digest	2.0910	57267	4111

Insights from the Output: The query output provides insights into how users are engaging with the email service for different actions.

Here's a detailed analysis of the output:

- Email Actions Analysis:
- Email Clickthrough: The average user type for email clickthrough is 2.108, indicating that users with a certain profile are more likely to click through emails. The total number of emails clicked is 9010, and there are 5277 unique users who have clicked.
- Email Open: The average user type for email open is 2.0916. Users seem to be more engaged with email opens compared to clickthroughs. The total number of emails opened is 20459, and there are 5927 unique users who have opened emails.
- Sent Reengagement Email: The average user type for sending reengagement emails is 2.2185. The total number of reengagement emails sent is 3653, and all of them are to unique users (3653 unique users).
- Sent Weekly Digest: The average user type for sending weekly digest emails is 2.091. The total number of weekly digest emails sent is 57267, and there are 4111 unique users who have received the weekly digest.

Recommendations from Analysis: Based on the insights gathered from the output, here are some recommendations:

- Targeted Content: Since users show higher engagement with email opens, it might be beneficial to focus on creating engaging subject lines and preview text to increase open rates.
- Clickthrough Optimization: To improve clickthrough rates, analyze the content and layout of the emails that receive higher clickthroughs. Identify elements that encourage users to click and apply similar strategies to other emails.
- Reengagement Strategy: Since all reengagement emails are sent to unique users, this approach seems effective. However, monitor the impact of these emails on reactivating inactive users over time.
- Segmentation: Segment users based on their engagement levels and preferences. Tailor email content to each segment to improve engagement further.
- A/B Testing: Conduct A/B testing on different email elements (subject lines, content, CTAs) to identify what resonates best with users and continually refine the email strategy.

Conclusion: The SQL query provides insights into email engagement metrics for different actions. By analyzing these metrics, businesses can refine their email marketing strategies, enhance user engagement, and increase the effectiveness of their email campaigns. It's crucial to iterate on strategies based on user behavior to ensure ongoing engagement and value delivery.

5. Results:

Case Study 1: Job Data Analysis

A. Jobs Reviewed Over Time: Analyzed jobs reviewed per hour for each day in November 2020, revealing consistent activity and occasional engagement spikes, possibly due to specific events.

B. Throughput Analysis: Calculated the 7-day rolling average of throughput, though the calculated value was consistently zero for the specified period. Recommends reviewing data accuracy and timeframe for meaningful insights.

C. Language Share Analysis: Determined the percentage share of each language in the last 30 days. Highlighted Persian as the dominant language with 37.5% share, while others contributed an equal share of 12.5%, reflecting a diverse audience.

D. Duplicate Rows Detection: Identified that the job_data table contains no duplicate rows, indicating accurate and well-organized data representation.

Case Study 2: Investigating Metric Spike

A. Weekly User Engagement: Measured user engagement on a weekly basis, revealing engagement trends, peak and low engagement periods, and user categorizations based on engagement levels.

B. User Growth Analysis: Analyzed user growth over time, identifying growth trends, periodic peaks, and user acquisition patterns.

C. Weekly Retention Analysis: Analyzed weekly user retention after sign-up, providing insights into cohort-based retention percentages over time.

D. Weekly Engagement Per Device: Calculated weekly user engagement per device, highlighting engagement variations, trends, and popular devices.

E. Email Engagement Analysis: Analyzed user engagement with email actions, including clickthrough, email open, reengagement emails, and weekly digest emails. Derived insights about user behavior and preferences.

Conclusion: The Operational Analytics project demonstrated the importance of data analysis in optimizing company operations and understanding metric changes. Through SQL queries and analysis, meaningful insights were derived, ranging from user engagement patterns to language preferences and email interaction behaviors. The project showcased the potential of data-driven decision-making and the value of applying advanced SQL skills in real-world scenarios.