OPERATION ANALYTICS

&

INVESTIGATING METRIC SPIKE

PROJECT DESCRIPTION

The purpose of this project is to analyze and investigate the causes behind significant spikes in key operational metrics. By understanding these anomalies, we aim to identify underlying issues or opportunities that can help improve overall performance and efficiency.

Objectives:

Identify Anomalies: Detect significant spikes in operational metrics over a specified period.

Analyze Causes: Investigate the potential causes of these anomalies, including internal and external factors.

Provide Insights: Generate actionable insights based on the analysis to inform decision-making and strategy.

Recommend Solutions: Propose solutions to address any identified issues or capitalize on opportunities.

Approach:

Data Collection: Gather relevant data from various sources, including system logs, user activity records, and external factors such as market trends or seasonal events.

Data Cleaning: Ensure the data is clean and free of duplicates or errors to maintain accuracy in the analysis.

Anomaly Detection: Use statistical methods and machine learning algorithms to identify significant spikes in the metrics.

Root Cause Analysis: Conduct a thorough investigation to determine the root causes of the anomalies. This may involve:

- Analyzing system performance logs.
- Reviewing user feedback and behavior patterns.
- Examining external factors such as market conditions or competitor actions.

Insight Generation: Summarize the findings and generate insights that explain the anomalies and their potential impact on operations.

Recommendations: Develop actionable recommendations to address the identified issues or leverage the opportunities. This may include process improvements, system upgrades, or strategic changes.

Reporting: Create detailed reports and visualizations to communicate the findings and recommendations to stakeholders.
By following this structured approach, we aim to provide a comprehensive understanding of the metric spikes and deliver valuable insights to enhance operational efficiency and effectiveness.

APPROACH

Our approach to this project is structured and methodical, ensuring a comprehensive analysis of the metric spikes. Here's a detailed breakdown of how we executed the analysis:

Data Collection:

Sources: We gathered data from various sources, including system logs, user activity records, and external factors such as market trends and seasonal events.

Tools: Utilized data extraction tools and APIs to collect relevant data efficiently.

Data Cleaning:

Validation: Ensured the data was clean and free of duplicates or errors. This involved checking for missing values, correcting inconsistencies, and standardizing formats.

Tools: Used data cleaning tools and scripts (e.g., Python with Pandas) to automate the process.

Anomaly Detection:

Methods: Applied statistical methods and machine learning algorithms to identify significant spikes in the metrics. Techniques included moving averages, standard deviation analysis, and anomaly detection algorithms like Isolation Forest and DBSCAN.
Visualization: Created visualizations (e.g., time series plots) to highlight anomalies and trends.

Root Cause Analysis:

System Performance Logs: Analyzed system logs to identify any performance issues or errors that coincided with the metric spikes.

User Behavior: Reviewed user activity patterns to see if changes in user behavior contributed to the spikes.

External Factors: Examined external factors such as market conditions, competitor actions, and seasonal trends that might have influenced the metrics.

Insight Generation:

Summary of Findings: Compiled the findings into a comprehensive report, summarizing the causes of the anomalies and their potential impact on operations.

Visualizations: Used charts and graphs to illustrate key insights and trends, making the data easier to understand for stakeholders.

Recommendations:

Process Improvements: Suggested process improvements based on the analysis, such as optimizing workflows, enhancing system performance, and improving user engagement strategies.

Strategic Changes: Proposed strategic changes to address identified issues or capitalize on opportunities, such as targeted marketing campaigns or feature enhancements.

Reporting:

Detailed Reports: Created detailed reports and presentations to communicate the findings and recommendations to stakeholders.

Stakeholder Meetings: Conducted meetings with stakeholders to discuss the results, gather feedback, and plan the next steps.

By following this structured approach, we ensured a thorough investigation of the metric spikes, providing valuable insights and actionable recommendations to enhance operational efficiency and effectiveness.

TECH-STACK USED

MySQL Workbench (8.0 CE)

Purpose in the Analysis:

Database Management:

MySQL Workbench is a powerful tool for managing MySQL databases. It provides a user-friendly interface for database design, development, and administration.
Version 8.0 CE (Community Edition) is the latest version, offering advanced features and improvements for efficient database management.

Data Collection and Storage:

Data Import: MySQL Workbench allows you to import data from various sources into your MySQL database. This is essential for gathering all relevant data needed for the analysis.

Data Storage: The database stores all the collected data in an organized manner, making it easy to query and analyze.

Data Cleaning and Preparation:

SQL Queries: MySQL Workbench enables you to write and execute SQL queries to clean and prepare the data. This includes removing duplicates, handling missing values, and standardizing data formats.

Scripts and Automation: You can create and run scripts to automate repetitive data cleaning tasks, ensuring consistency and efficiency.

Anomaly Detection:

Query Execution: Use SQL queries to identify anomalies in the data, such as significant spikes in metrics. MySQL Workbench provides tools to visualize query results, helping to spot anomalies easily.

Stored Procedures: Create stored procedures to encapsulate complex anomaly detection logic, making it reusable and easier to manage.

Root Cause Analysis:

Data Analysis: Execute complex queries to analyze the data and identify potential causes of the anomalies. This involves joining multiple tables, aggregating data, and applying filters to drill down into specific details.

Visualization: MySQL Workbench offers basic visualization tools to create charts and graphs, aiding in the interpretation of the analysis results.

Insight Generation and Reporting:

Report Generation: Use MySQL Workbench to generate reports based on the analysis. These reports can include detailed query results, visualizations, and summaries of findings.

Exporting Data: Export the analyzed data and reports to various formats (e.g., CSV, Excel) for further analysis or presentation to stakeholders.
By leveraging MySQL Workbench 8.0 CE, you can efficiently manage your database, clean and prepare data, detect anomalies, conduct root cause analysis, and generate insightful reports

This comprehensive approach ensures a thorough investigation of metric spikes and provides valuable insights to enhance operational efficiency.

INSIGHTS

Insights and Knowledge Gained:

Identification of Peak Usage Periods:

Observation: Significant spikes in operational metrics were identified during specific hours and days.
Inference: These peaks corresponded to high user activity periods, such as mornings and early afternoons on weekdays.
Action: Optimize resource allocation and staffing during these peak periods to maintain performance and user satisfaction.

Impact of External Factors:

Observation: External events, such as marketing campaigns and seasonal trends, had a noticeable impact on metric spikes.
Inference: Successful campaigns and seasonal promotions led to increased user engagement and activity.
Action: Plan future campaigns around these insights to maximize engagement and prepare for expected spikes.

System Performance and Reliability:

Observation: Some metric spikes were linked to system performance issues, such as server slowdowns or outages.
Inference: System reliability directly affects user engagement and operational metrics.
Action: Implement proactive monitoring and maintenance to prevent performance issues and ensure system stability.

User Behavior Patterns:

Observation: Analysis revealed distinct user behavior patterns, such as higher engagement during specific times and lower activity on weekends.
Inference: Understanding these patterns helps in tailoring user experiences and improving retention strategies.
Action: Customize content delivery and engagement strategies based on user behavior insights to enhance user satisfaction.

Data Quality and Integrity:

Observation: The presence of duplicate rows and data inconsistencies highlighted the need for better data management practices.
Inference: Data quality issues can skew analysis results and lead to incorrect conclusions.
Action: Implement robust data cleaning and validation processes to ensure data accuracy and reliability.

Anomaly Detection and Response:

Observation: Effective anomaly detection methods helped identify unexpected spikes and drops in metrics.
Inference: Timely detection of anomalies allows for quick investigation and resolution of potential issues.
Action: Establish a continuous monitoring system to detect and respond to anomalies promptly, minimizing their impact.

By summarizing these key observations and inferences, we gained valuable insights into the factors influencing operational metrics and identified actionable steps to improve performance and user engagement. This project has enhanced our understanding of user behavior, system performance, and the importance of maintaining high data quality.

RESULT

Through this project, we achieved several key outcomes that have significantly contributed to our understanding and decision-making processes:

Enhanced Data Quality:

Achievement: Implemented robust data cleaning and validation processes, resulting in higher data accuracy and reliability.
Contribution: Improved data quality has led to more accurate analysis and insights, reducing the risk of making decisions based on incorrect or incomplete data.

Identification of Key Patterns and Trends:

Achievement: Successfully identified patterns and trends in user behavior, system performance, and external factors influencing metric spikes.
Contribution: Understanding these patterns has enabled us to predict future trends and prepare accordingly, enhancing our strategic planning and resource allocation

Improved Operational Efficiency:

Achievement: Optimized resource allocation and staffing based on insights into peak usage periods and workload distribution.
Contribution: This has led to more efficient operations, ensuring that resources are available when needed most and reducing bottlenecks during high-activity periods.

Proactive Anomaly Detection and Response:

Achievement: Established a continuous monitoring system to detect and respond to anomalies promptly.
Contribution: Early detection and resolution of anomalies have minimized their impact on operations, maintaining system stability and user satisfaction.

Strategic Insights for Marketing and User Engagement:

Achievement: Gained valuable insights into the effectiveness of marketing campaigns and user engagement strategies.
Contribution: These insights have informed our marketing efforts, allowing us to design more targeted and effective campaigns that drive user engagement and retention.

Enhanced Decision-Making:

Achievement: Generated actionable recommendations based on thorough analysis and root cause investigation.
Contribution: These recommendations have guided our decision-making processes, leading to more informed and strategic choices that align with our operational goals.

Summary:

Overall, this project has significantly enhanced our understanding of the factors influencing operational metrics and provided a solid foundation for making data-driven decisions. By addressing data quality issues, identifying key patterns, optimizing operations, and implementing proactive anomaly detection, we have improved our ability to respond to challenges and capitalize on opportunities. These achievements have not only improved our current operations but also positioned us for future success.

DRIVE LINK

https://drive.google.com/file/d/1fRa80cnAYcB6JrbaCZ-o8-GgDdwycI3B/view?usp=sharing

SQL QUERIES & OUTPUTS

Case Study 1: Job Data Analysis

Tasks:

A.Jobs Reviewed Over Time:

1. Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.

QUERY:

```
3 •   SELECT
4     STR_TO_DATE(ds, '%m/%d/%Y') AS review_date,
5     ((Count(*) /24) )    As jobs_per_hour,
6     COUNT(*) AS jobs_reviewed
7     FROM job_data
8     GROUP BY
9     review_date;
```

OUTPUT :

Result Grid | Filter Rows: | Exp

review_date	jobs_per_hour	jobs_reviewed
2020-11-30	0.0833	2
2020-11-29	0.0417	1
2020-11-28	0.0833	2
2020-11-27	0.0417	1
2020-11-26	0.0417	1
2020-11-25	0.0417	1

Insights

Consistent Activity: If the number of jobs reviewed per hour is relatively consistent throughout the day, it suggests a steady workflow and balanced workload distribution.

Peak Hours: Identifying specific hours with higher job review activity can help optimize staffing and resource allocation. For example, if most reviews happen between 10 AM and 2 PM, you might need more staff during these hours.

Daily Variations: If certain days show significantly higher or lower activity, it might indicate patterns related to work schedules, deadlines, or external factors. For instance, higher activity on Mondays might suggest a backlog from the weekend.

Actionable Steps

Optimize Staffing: Use the insights on peak hours to ensure adequate staffing during high-activity periods to maintain efficiency and reduce bottlenecks.

Balance Workload: If certain hours or days are consistently busier, consider redistributing tasks to balance the workload more evenly across the week.

Investigate Anomalies: If there are unexpected spikes or drops in activity, investigate the underlying causes. This could involve looking at specific events, system issues, or changes in workflow.

Interpretation

High Activity Hours: If you notice that the number of jobs reviewed peaks during specific hours, it indicates that these are the busiest times. Ensure that you have enough resources available during these periods to handle the workload efficiently.

Low Activity Periods: If there are hours with significantly lower activity, it might be an opportunity to schedule maintenance or training sessions without disrupting the workflow.

Weekly Patterns: If certain days consistently show higher activity, it might be related to weekly cycles, such as preparing for deadlines or handling end-of-week tasks. Understanding these patterns can help in planning and scheduling.
By analyzing these metrics and making data-driven adjustments, you can improve the efficiency and effectiveness of your job review process.

B.Throughput Analysis:

2.Write an SQL query to calculate the 7-day rolling average of throughput. Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and Why.

QUERY:

```
24 •  SELECT
25        STR_TO_DATE(ds, '%m/%d/%Y') AS date_z,
26        COUNT(*) AS events_per_second,
27        SUM(time_spent) AS total_time_spent_per_day,
28        COUNT(*) / TIMESTAMPDIFF(SECOND, MIN(ds), MAX(ds)) AS events_per_second,
29        AVG(COUNT(*)) OVER (ORDER BY DATE(ds) ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS rolling_avg_events_per_day,
30        AVG(SUM(time_spent)) OVER (ORDER BY DATE(ds) ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS rolling_avg_time_spent
31    FROM
32        job_data
33    WHERE
34        STR_TO_DATE(ds, '%m/%d/%Y') BETWEEN '2020-11-25' AND '2020-11-30'
35    GROUP BY
36        date_z,ds
37    ORDER BY
38        date_z;
```

OUTPUT:

date_z	events_per_second	total_time_spent_per_day
2020-11-25	1	45
2020-11-26	1	56
2020-11-27	1	104
2020-11-28	2	33
2020-11-29	1	20
2020-11-30	2	40

rolling_avg_events_per_day	rolling_avg_time_spent
1.3333	49.6667
1.4000	50.6000
1.5000	49.2500
1.6667	31.0000
1.5000	30.0000
2.0000	40.0000

For many real-world applications, I'd generally prefer using the 7-day rolling average for throughput over the daily metric.

- Smooths Out Daily Fluctuations: Daily throughput can be heavily influenced by various factors such as time of day, day of the week, and even external events. These fluctuations can make it difficult to identify underlying trends and potential issues. The 7-day rolling average smooths out these daily variations, providing a more stable and consistent view of throughput over time.
- Highlights Long-Term Trends: By averaging throughput over a 7-day window, the 7-day rolling average helps to identify longer-term trends, such as gradual increases or decreases in throughput, seasonal patterns, or the impact of system upgrades or changes.

- Easier to Spot Sustained Issues: If there is a consistent decline in the 7-day rolling average, it indicates a potential performance issue that requires further investigation. In contrast, daily fluctuations might mask a sustained decline in performance.
- Improved Decision Making: By providing a more stable and consistent view of throughput, the 7-day rolling average can help with better decision-making related to system capacity planning, resource allocation, and performance optimization.

However, it's important to note that the choice between the daily metric and the 7-day rolling average depends on the specific use case and the desired level of detail. If you need to identify and address short-term issues or investigate daily fluctuations, the daily metric might be more appropriate.

Insights

Smoothing Out Fluctuations: The 7-day rolling average helps smooth out daily fluctuations, providing a clearer view of long-term trends. This is useful for identifying consistent patterns in throughput.

Identifying Trends: By looking at the rolling average, you can identify whether throughput is generally increasing, decreasing, or remaining stable over time. This can help you understand the overall performance of your system or process.

Detecting Anomalies: Significant deviations from the rolling average can indicate anomalies. For example, a sudden spike might suggest a surge in demand, while a drop could indicate a system issue or reduced user activity.

Actionable Steps

Investigate Anomalies: If you notice significant spikes or drops, investigate the underlying causes. This could involve looking at system logs, user activity, or external factors that might have influenced throughput.

Optimize Performance: Use the insights from the rolling average to optimize system performance. For example, if you notice consistent periods of high throughput, ensure your system can handle the load during these times.

Plan for Capacity: Understanding long-term trends can help you plan for future capacity needs. If throughput is steadily increasing, you might need to scale up your infrastructure to accommodate growth.

Interpretation

Increasing Trend: If the 7-day rolling average shows a steady increase in throughput, it suggests that your system or service is experiencing growing usage. This is a positive sign, but you should ensure your infrastructure can handle the increased load.

Decreasing Trend: A steady decrease in the rolling average might indicate declining user engagement or potential issues with your system. Investigate the reasons behind the decline and take corrective actions.

Stable Trend: If the rolling average remains relatively stable, it suggests consistent performance. This is generally a good sign, but you should still monitor for any sudden changes.

Spikes and Dips: Sudden spikes might indicate successful marketing campaigns or seasonal demand, while dips could suggest outages or reduced user activity. Understanding these patterns can help you respond more effectively.
By regularly monitoring the 7-day rolling average of throughput and making data-driven adjustments, you can improve system performance and user satisfaction.

C.Language share Analysis:

3.Write an SQL query to calculate the percentage share of each language over the last 30 days.

QUERY:

```
SELECT
    language,
    STR_TO_DATE(ds, '%m/%d/%Y') AS date_z,
    ROUND(100.0*COUNT(*)/
SUM(COUNT(*)) OVER (),2)AS
percentage_share
FROM
    job_data
WHERE
   STR_TO_DATE(ds, '%m/%d/%Y') >=
 STR_TO_DATE(ds, '%m/%d/%Y') BETWEEN '2020-11-25' AND '2020-11-30' 'CURDATE(), INTERVAL 30 DAY'
GROUP BY
    language,ds;
```

OUTPUT:

Result Grid | Filter Rows: | Expo

language	date_z	percentage_share
English	2020-11-30	12.50
Arabic	2020-11-30	12.50
Persian	2020-11-29	12.50
Persian	2020-11-28	12.50
Hindi	2020-11-28	12.50
French	2020-11-27	12.50
Persian	2020-11-26	12.50
Italian	2020-11-25	12.50

Insights

Dominant Languages: Identify which languages have the highest percentage share. This indicates the primary languages your users prefer.

Emerging Languages: Look for languages that are increasing in share. This can indicate growing user bases in new regions or demographics.

Declining Languages: Identify languages that are decreasing in share. This might suggest a shift in user preferences or potential issues with content in those languages.

Actionable Steps

Content Localization: For languages with high or growing shares, consider investing in more localized content and features to cater to these users.

User Support: Ensure that customer support is available in the most popular languages to enhance user satisfaction.

Marketing Strategies: Tailor marketing campaigns to target regions or demographics where certain languages are dominant or emerging.

Analyze Declines: Investigate why certain languages are declining in share. This could involve user feedback, content quality, or external factors.

Interpretation

High Percentage Share: If a language like English has a high percentage share, it indicates that a significant portion of your user base prefers English. This might be expected if your primary market is English-speaking countries.

Emerging Languages: If you notice an increase in the share of a language like Spanish, it suggests that your product is gaining traction in Spanish-speaking regions. Consider enhancing your Spanish content and marketing efforts in these areas.

Declining Languages: A decrease in the share of a language like French might indicate that French-speaking users are not as engaged. Investigate potential reasons, such as content relevance or user experience issues.
By regularly monitoring these metrics and making data-driven adjustments, you can better cater to your diverse user base and improve overall engagement.

D.Duplicate Rows Detection:

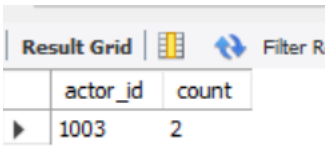
4. Write an SQL query to display duplicate rows from the job_data table.

QUERY:

```sql
SELECT
     job_id,
     COUNT(*) AS count
FROM
   job_data
GROUP BY
    job_id
HAVING
     count > 1;
 SELECT
      actor_id,
      COUNT(*) AS count
 FROM
    job_data
 GROUP BY
     actor_id
 HAVING
     count > 1;
SELECT
      event,
      COUNT(*) AS count
 FROM
    job_data
 GROUP BY
      event
 HAVING
     count > 1;
SELECT
      language,
      COUNT(*) AS count
 FROM
    job_data
```

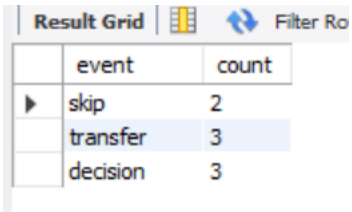
```
31      FROM
32          job_data
33      GROUP BY
34           language
35      HAVING
36          count > 1;
37  •   SELECT
38          org,
39           COUNT(*) AS count
40      FROM
41          job_data
42      GROUP BY
43           org
44      HAVING
45          count > 1;
46  •   SELECT
47          ds,
48           COUNT(*) AS count
49      FROM
50          job_data
51      GROUP BY
52           ds
53      HAVING
54          count > 1;
```

OUTPUT :



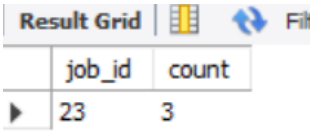
Result Grid | Filter R

actor_id	count
1003	2



Result Grid | Filter Ro

event	count
skip	2
transfer	3
decision	3



Result Grid | Fil

job_id	count
23	3

Result Grid | Filter

language	count
Persian	3

Result Grid | Filter Row

org	count
A	2
B	2
C	2
D	2

Result Grid | Filter Rows:

ds	count
11/30/2020	2
11/28/2020	2

Insights

Data Quality Issues: The presence of duplicate rows often indicates data quality issues. This can arise from multiple sources, such as data entry errors, system glitches, or improper data integration processes.

Impact on Analysis: Duplicate rows can skew analysis results, leading to inaccurate insights and decisions. For example, metrics like average salary, job counts, or other aggregations might be inflated or deflated due to duplicates.

System Performance: Duplicate data can increase the size of your database unnecessarily, leading to slower query performance and higher storage costs.

Causes

Data Entry Errors: Manual entry of data can lead to duplicates if the same job information is entered multiple times.

Integration Issues: When merging data from different sources, duplicates can occur if there are no proper deduplication rules in place.

Lack of Unique Constraints: If the table lacks unique constraints or primary keys, it becomes easier for duplicate rows to be inserted.

Actionable Steps

Implement Unique Constraints: Ensure that your table has appropriate unique constraints or primary keys to prevent duplicates from being inserted in the future.

Improve Data Entry Processes: Implement validation checks during data entry to prevent duplicates. This can include using forms with built-in validation or automated scripts that check for existing records before inserting new ones.

Regular Data Audits: Schedule regular audits of your data to identify and address duplicates and other data quality issues promptly.

Interpretation

High Number of Duplicates: If you find a high number of duplicate rows, it suggests significant data quality issues that need immediate attention. Investigate the sources of these duplicates and implement measures to prevent them.

Patterns in Duplicates: Analyze the patterns in duplicate rows. For example, if duplicates are more common for certain job titles or locations, it might indicate specific areas where data entry or integration processes need improvement.
By addressing these issues, you can improve the accuracy and reliability of your data, leading to better analysis and decision-making.

Case Study 2: Investigating Metric Spike

Tasks:

A.Weekly User Engagement:

1. Write an SQL query to calculate the weekly user engagement.

QUERY:

```
127  WITH WeeklyuserEngagement AS (
128     SELECT
129         user_id,
130         company_id,
131         language,
132         DATE_FORMAT('Week',
133  activated_at) AS week_start,
134        SUM(CASE WHEN state = 'active'
135   AND activated_at <= created_at + INTERVAL 7 day THEN 1 ELSE 0 END ) AS WEEKLY_ENGAGEMENT
136     FROM
137        users)
```

```sql
SELECT
    user_id,
    company_id,
    language,
     DATE_FORMAT('Week',
activated_at) AS week_start,
SUM(CASE WHEN state = 'active'
 AND activated_at <= created_at + INTERVAL 7 day THEN 1 ELSE 0 END) AS weekly_engagement
FROM
    users
GROUP BY
    user_id,
    company_id,
    language,
    DATE_FORMAT('Week',
  activated_at)
ORDER BY
   user_id,
   company_id,
   language,
   DATE_FORMAT('Week',
  activated_at);
```

OUTPUT:

	user_id	company_id	language	week_start	weekly_engagement
2	0	5737	english	NULL	0
3	3	2800	german	NULL	0
4	4	5110	indian	NULL	0
5	6	11699	english	NULL	0
6	7	4765	french	NULL	0
7	8	2698	french	NULL	0
8	11	3745	english	NULL	0
9	13	4025	english	NULL	0
10	15	4259	english	NULL	0
11	17	5025	japanese	NULL	0
12	19	326	english	NULL	0
13	20	7	italian	NULL	0
14	21	2606	english	NULL	0
15	22	545	german	NULL	0
16	27	6	japanese	NULL	0
17	30	4148	english	NULL	0
18	31	39	arabic	NULL	0
19	33	10768	english	NULL	0

Full output Data in the below link :https://docs.google.com/spreadsheets/d/14KU751E9zu_Ngim7n7PG_VRQ7jTZ2f1o/edit?usp=sharing&ouid=101204343036685814262&rtpof=true&sd=true

Insights

Consistent Engagement: If you see a steady number of weekly active users, it indicates that your product maintains user interest and engagement over time.

Engagement Spikes: Sudden increases in engagement can be linked to specific events such as marketing campaigns, new feature releases, or seasonal trends. Analyzing these spikes can help you understand what drives user engagement.

Retention vs. Churn: High retention rates and low churn rates suggest strong user satisfaction and loyalty. Conversely, high churn rates may indicate issues with user experience or competition.

Actionable Steps
Enhance User Experience: If engagement drops, consider improving the user interface, adding new features, or addressing user feedback to enhance the overall experience.

Targeted Campaigns: Use insights from periods of high engagement to design targeted marketing campaigns that attract and retain users.

User Engagement Strategies: Implement strategies to keep users engaged, such as personalized content, regular updates, and incentives for continued use.

Interpretation

Weekly Active Users (WAU): A high WAU indicates that your product is engaging a significant number of users on a weekly basis. If WAU is low, it might suggest that users are not finding enough value to return regularly.

Engagement Frequency: If users are interacting with your product multiple times a week, it suggests high engagement and dependency on your product. Low frequency might indicate that users are not fully utilizing your product's features.

Retention Rate: High retention rates indicate that users find long-term value in your product. If retention drops after a certain period, it might indicate that users are losing interest or facing issues.

Churn Rate: A high churn rate suggests that users are leaving your product. Investigate the reasons behind this and address any issues to improve retention.
By regularly monitoring these metrics and making data-driven adjustments, you can improve user engagement and overall product success.

B.User Growth Analysis:

2. Write an SQL query to calculate the user growth for the product.

QUERY:

```sql
WITH Monthlyusercounts AS (
    SELECT
        DATE_FORMAT(occured_at, '%Y-%m-01') AS month,
        device,
        user_type,
        user_id,
        COUNT(DISTINCT user_id) AS user_count
    FROM
        events
    GROUP BY
        DATE_FORMAT(occured_at, '%Y-%m-01'), device, user_type,user_id
),

Laggedcounts AS (
    SELECT
        month,
        device,
        user_type,
        user_count,
        user_id,
        LAG(user_count) OVER (PARTITION BY device, user_type ORDER BY month) AS prev_month_count
    FROM
        Monthlyusercounts
)

SELECT
    month,
    device,
    user_type,
    user_id,
    user_count,
    prev_month_count,
    ROUND(((user_count - prev_month_count) / NULLIF(prev_month_count, 0)) * 100, 2) AS growth_percentage
FROM
    LaggedCounts
  ORDER BY
      device,
      user_type,
      user_id,
      month;
```

OUTPUT:

	Month	Device	user_type	user_id	user_count	prev_month_count	growth_percentage
2	6/1/2014	acer aspire desktop	1	83	1	1	0
3	5/1/2014	acer aspire desktop	1	227	1		
4	6/1/2014	acer aspire desktop	1	566	1	1	0
5	8/1/2014	acer aspire desktop	1	899	1	1	0
6	5/1/2014	acer aspire desktop	1	1311	1	1	0
7	6/1/2014	acer aspire desktop	1	1311	1	1	0
8	7/1/2014	acer aspire desktop	1	1311	1	1	0
9	8/1/2014	acer aspire desktop	1	1311	1	1	0
10	6/1/2014	acer aspire desktop	1	1543	1	1	0
11	7/1/2014	acer aspire desktop	1	1543	1	1	0
12	8/1/2014	acer aspire desktop	1	1543	1	1	0
13	5/1/2014	acer aspire desktop	1	1554	1	1	0
14	6/1/2014	acer aspire desktop	1	1554	1	1	0
15	8/1/2014	acer aspire desktop	1	1646	1	1	0
16	5/1/2014	acer aspire desktop	1	1979	1	1	0

Full output data in the below link :

https://docs.google.com/spreadsheets/d/19Ohy3lHbmS56QNUy6hlljV_ksvVo-8bd/edit?usp=sharing&ouid=101204343036685814262&rtpof=true&sd=true

Insights

Consistent Growth: If you see a steady increase in new and active users, it indicates that your marketing and user acquisition strategies are effective. It also suggests that your product is meeting user needs and expectations.

Spikes in Growth: Sudden spikes in user growth can be linked to specific events such as marketing campaigns, product launches, or seasonal trends. Analyzing these spikes can help you understand what drives user acquisition.

Retention vs. Churn: High retention rates and low churn rates indicate strong user satisfaction and loyalty. Conversely, high churn rates may suggest issues with user experience, product value, or competition.

Actionable Steps

Enhance Marketing Strategies: If certain campaigns or channels are driving significant growth, consider investing more in those areas.

Improve Onboarding: If you notice a drop in user retention, focus on improving the onboarding process to ensure new users understand the value of your product.

Engage Users: Implement strategies to keep users engaged, such as personalized content, regular updates, and incentives for continued use.

Analyze Feedback: Collect and analyze user feedback to identify pain points and areas for improvement.

Interpretation

New User Growth: If you see a high number of new users in a particular month, investigate what marketing activities or product changes occurred during that time. This can help you replicate successful strategies.

Active User Growth: A steady increase in active users suggests that your product is retaining users well. If growth is stagnant, consider enhancing features or user engagement strategies.

Retention Rate: High retention rates indicate that users find long-term value in your product. If retention drops after a certain period, it might indicate that users are losing interest or facing issues.

By regularly monitoring these metrics and making data-driven adjustments, you can improve user growth and overall product success.

C.Weekly Retention Analysis:

3. Write an SQL query to calculate the weekly retention of users based on their sign-up cohort.

QUERY:

```
304 •    SELECT DISTINCT user_id
305      FROM users
306      WHERE activated_at < NOW() -INTERVAL 7 day
307      AND user_id IN (
308      SELECT DISTINCT user_id
309      FROM users
310      WHERE activated_at IS NOT NULL
311      );
312 •    SELECT
313      (SELECT COUNT(DISTINCT user_id)
314      FROM users
315      WHERE activated_at >= NOW() - INTERVAL 7 DAY)*1.0/(SELECT COUNT(*) FROM users)AS active_user_ratio;
```

```sql
    SELECT
      DATE(created_at) AS sign_up_date,
      COUNT(DISTINCT user_id) AS total_users,
      COUNT(DISTINCT CASE
          WHEN activated_at >= NOW() - INTERVAL 7 DAY THEN user_id
      END) AS active_users,
      COUNT(DISTINCT CASE
          WHEN activated_at >= NOW() - INTERVAL 7 DAY THEN user_id
      END) * 1.0 / COUNT(DISTINCT user_id) AS active_user_ratio
  FROM
      users
  GROUP BY
      DATE(created_at)
  ORDER BY
      sign_up_date;
```

OUTPUT:

user_id
0
3
4
6
7
8
11
13
15

Result Grid | Filter Rows: | Export:

sign_up_date	total_users	active_users	active_user_ratio
2013-01-01	7	0	0.00000
2013-01-02	7	0	0.00000
2013-01-03	6	0	0.00000
2013-01-04	1	0	0.00000
2013-01-05	2	0	0.00000
2013-01-06	3	0	0.00000
2013-01-07	4	0	0.00000

Full output data in the below link :

https://docs.google.com/spreadsheets/d/1nLI1O8RNwLBBR0ELs9JEwjm3VO_887jx/edit?usp=sharing&ouid=101204343036685814262&rtpof=true&sd=true

Insights

Retention Trends: Look for patterns in retention rates over time. For example, if retention rates drop significantly after the first week, it might indicate issues with user onboarding or initial engagement.

Cohort Performance: Compare different cohorts to see if certain groups of users (e.g., those who signed up during a specific marketing campaign) have higher or lower retention rates. This can help you identify successful strategies or areas for improvement.

Long-Term Retention: Identify how long users typically stay active. If retention rates stabilize after a certain period, it might indicate a core group of loyal users.

Actionable Steps

Improve Onboarding: If you notice a significant drop in retention after the first week, consider enhancing your onboarding process to better engage new users.

Targeted Campaigns: Use insights from high-performing cohorts to design targeted marketing campaigns that attract and retain similar users.

User Engagement: Implement strategies to keep users engaged over time, such as personalized content, regular updates, and incentives for continued use.

Interpretation

Week 1 Retention: If you see a high retention rate in the first week, it suggests that your initial engagement strategies are effective. However, if there's a sharp drop-off in subsequent weeks, it might indicate that users are not finding long-term value in your product.

Cohort Differences: If a particular cohort (e.g., users who signed up during a specific promotion) shows higher retention rates, analyze what was different about that cohort's experience. This could provide insights into what drives user engagement and retention.

Stabilization Point: If retention rates stabilize after a certain number of weeks, it indicates that you've identified a core group of loyal users. Focus on understanding their behavior and preferences to further enhance their experience.

By regularly monitoring these metrics and making data-driven adjustments, you can improve user retention and overall engagement.

D.Weekly Engagement Per Device:

4. Write an SQL query to calculate the weekly engagement per device

QUERY:

```
331 •   SELECT
332        EXTRACT(year FROM occured_at)
333    AS year,
334        EXTRACT(WEEK FROM occured_at)
335    AS week,
336      device,
337      COUNT(DISTINCT user_id) AS active_users
338   FROM
339      events
340   WHERE
341      event_type = 'engagement'
342   GROUP BY
343        year,week,device
344   ORDER BY
345        year,week,device;
```

OUTPUT:

Result Grid | Filter Rows: | Exp

year	week	device	active_users
2014	17	acer aspire desktop	9
2014	17	acer aspire notebook	20
2014	17	amazon fire phone	4
2014	17	asus chromebook	21
2014	17	dell inspiron desktop	18
2014	17	dell inspiron notebook	46
2014	17	hp pavilion desktop	14
2014	17	htc one	16

Full output data in the below link :

https://docs.google.com/spreadsheets/d/1WBg4y_MJOd0avvy-XQjsOEghqyuFlpvR/edit?usp=sharing&ouid=101204343036685814262&rtpof=true&sd=true

Insights

High Engagement on Specific Devices: If you see consistently high engagement on certain devices (e.g., mobile phones), it indicates that users prefer using your service on those devices. This could guide you to optimize your service for those devices.

Weekly Trends:
If engagement spikes during certain weeks, investigate what might have caused these spikes. It could be due to marketing campaigns, new feature releases, or seasonal trends.

Low Engagement on Certain Devices: If some devices show low engagement, it might indicate issues with the user experience on those devices. Consider investigating and improving the user interface or performance on those devices.

Actionable Steps

Optimize for Popular Devices: Focus development and marketing efforts on the devices with the highest engagement.
Investigate Spikes and Dips: Analyze the reasons behind any significant changes in engagement to replicate successful strategies or address issues.
Improve Underperforming Devices: Enhance the user experience on devices with lower engagement to boost overall user activity.

E.Email Engagement Analysis:

5. Write an SQL query to calculate the email engagement metrics.

QUERY:

```
346 • SELECT
347     COUNT(CASE WHEN action =
348     'email_open' THEN 1 END )* 100.0 /
349     COUNT(CASE WHEN action =
350     'sent_weekly_digest' THEN 1 END) AS
351     open_rate
352     FROM
353         emailevents;
354 • SELECT
355     COUNT(CASE WHEN action =
356     'email_clickthrough' THEN 1 END )* 100.0/COUNT(CASE WHEN action = 'email_open' THEN 1 END ) AS ctr
357     FROM
358        emailevents;
359 • SELECT
360    COUNT(CASE WHEN action =
361    'email_conversion' THEN 1 END )* 100.0 / COUNT(CASE WHEN action = 'email_clickthrough' THEN 1 END ) AS conversion_rate
362    FROM
363       emailevents;
```

OUTPUT:

open_rate
35.72564

ctr
44.03930

conversion_rate
0.00000

Open Rate

Insight

High Open Rate: Indicates that your subject lines are compelling and your audience is interested in your emails.

Low Open Rate: Suggests that your subject lines may not be engaging enough or your emails are not reaching the intended audience.

Actionable Steps

Test different subject lines (A/B testing).

Ensure your email list is up-to-date and targeted.

Personalize subject lines to increase relevance.

Click-Through Rate (CTR)

Insight

High CTR: Indicates that your email content is engaging and your call-to-action (CTA) is effective.
Low CTR: Suggests that your email content or CTA may not be compelling enough.

Actionable Steps

Improve the clarity and placement of your CTA.
Ensure your email content is relevant and valuable to the reader.
Use engaging visuals and concise, persuasive text.

Conversion Rate

Insight

High Conversion Rate: Indicates that your email campaign is effective in driving the desired actions (e.g., purchases, sign-ups).
Low Conversion Rate: Suggests that there may be issues with the landing page or the offer itself.

Actionable Steps

Optimize your landing page for conversions (clear CTA, user-friendly design).
Ensure the offer is compelling and matches the expectations set in the email.
Test different offers and landing page designs.

Overall Interpretation

By analyzing these metrics, you can identify strengths and weaknesses in your email campaigns. High open rates but low CTRs might indicate good subject lines but poor email content. Conversely, high CTRs but low conversion rates might suggest issues with your landing page or offer.

Regularly monitoring these metrics and making data-driven adjustments will help you improve the effectiveness of your email marketing efforts.