**Operation Analytics and Investigating Metric Spike SQL**

**Introduction**

This project focuses on analyzing the data which is provided by the company. My task is to derive insights and answer the questions asked by different departments. So that these insights are then used by **opsteam**, **support team**, **marketing team**, etc to predict the overall growth or decline of a company’sfortune. It means better automation, and better understanding between cross-functional effective workflows.

In case study 1 there is **job\_data**table while in case study 2 there are **users**, **events**and **email\_events** tables

**Table of Contents**

* [Project Approach](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#project-nbsp-approach)
* [Tech-Stack Used](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#tech-stack-nbsp-used)
* [Insights](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#insights)
  + [Case Study 1 (Job Data)](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#case-nbsp-study-nbsp-1-nbsp-job-nbsp-data)
    - [1. Calculate the number of jobs reviewed per hour per day for November 2020?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-1-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-the-strong-strong-nbsp-strong-strong-number-strong-strong-nbsp-strong-strong-of-strong-strong-nbsp-strong-strong-jobs-strong-strong-nbsp-strong-strong-reviewed-strong-strong-nbsp-stron)
    - [2. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7- day rolling and why?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-2-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-7-strong-strong-nbsp-strong-strong-day-strong-strong-nbsp-strong-strong-rolling-strong-strong-nbsp-strong-strong-average-strong-strong-nbsp-strong-strong-of-strong-strong-nbsp-strong-s)
    - [3. Calculate the percentage share of each language in the last 30 days?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-3-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-the-strong-strong-nbsp-strong-strong-percentage-strong-strong-nbsp-strong-strong-share-strong-strong-nbsp-strong-strong-of-strong-strong-nbsp-strong-strong-each-strong-strong-nbsp-stro)
    - [4. Let’s say you see some duplicate rows in the data. How will you display duplicates from the table?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-4-nbsp-strong-strong-lets-strong-strong-nbsp-strong-strong-say-strong-strong-nbsp-strong-strong-you-strong-strong-nbsp-strong-strong-see-strong-strong-nbsp-strong-strong-some-strong-strong-nbsp-strong-strong-duplicate-strong-strong-nbsp-strong-stro)
  + [Case Study 2 (Investigating Metric Spike)](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#case-nbsp-study-nbsp-2-investigating-nbsp-metric-nbsp-spike)
    - [1. Calculate the weekly user engagement?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-1-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-the-strong-strong-nbsp-strong-strong-weekly-user-engagement-strong-strong-strong)
    - [2. Calculate the user growth for product?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-2-nbsp-strong-strong-calculate-the-strong-strong-nbsp-strong-strong-user-strong-strong-nbsp-strong-strong-growth-strong-strong-nbsp-strong-strong-for-strong-strong-nbsp-strong-strong-product-strong-strong-strong)
    - [3. Calculate the weekly retention of users-sign up cohort?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-3-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-the-strong-strong-nbsp-strong-strong-weekly-strong-strong-nbsp-strong-strong-retention-of-users-sign-strong-strong-nbsp-strong-strong-up-strong-strong-nbsp-strong-strong-cohort-strong-)
    - [4. Calculate the weekly engagement per device?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-4-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-the-strong-strong-nbsp-strong-strong-weekly-engagement-per-strong-strong-nbsp-strong-strong-device-strong-strong-strong)
    - [5. Calculate the email engagement metrics?](https://tech.pinkybe.in/operation-analytics-and-investigating-metric-spike-sql/#strong-5-nbsp-strong-strong-calculate-strong-strong-nbsp-strong-strong-the-strong-strong-nbsp-strong-strong-email-strong-strong-nbsp-strong-strong-engagement-strong-strong-nbsp-strong-strong-metrics-strong-strong-strong)

**Project Approach**

This project is developed using SQL Workbench. First I need to create database by using dataset file which was provided by the company. Next step load the data into SQL Workben h then performed analysis and find the information that will the help the **ops team**, **support team**, **marketing team**, etc to understand questions like – Why is there a dip in daily engagement? Why have sales taken a dip? Etc. Questions like these must be answered daily and for that it’s very important to investigate metric spike.

**Tech-Stack Used**

MySQL Workbench is a visual editor that unifies data modeling, SQL development, and database administration in one interface.

MySQL Workbench is widely used to handle structured data. It is an open-source Relational Database Management System (RDBMS) developed by Oracle Corporation, Sun Microsystems that uses Structured Query language (SQL) to interact with databases.

MySQL Workbench offers database migration options, making it easier to move data to and from the Microsoft SQL Server, Microsoft Access and other RDBMS tables.

**Insights**

Case Study 1 (Job Data)

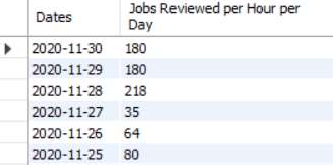
**1. Calculate the number of jobs reviewed per hour per day for November 2020?**

SELECT ds AS Dates, ROUND((COUNT(job\_id)/SUM(time\_spent))\*3600) AS “Jobs Reviewed per Hour per Day”

FROM job\_data

WHERE ds BETWEEN ‘2020-11-01’ AND ‘2020-11-30’ GROUP BY ds;

On date 2020-11-28 there is maximum number of jobs reviewed that is **218**.



**2. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7- day rolling and why?**

SELECT ROUND(COUNT(event)/SUM(time\_spent), 2) AS “Weekly Throughput” FROM job\_data;

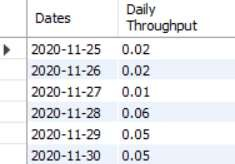
The weekly throughput is **0.03**.

https://tech.pinkybe.in/wp-content/uploads/2023/05/Picture4.png

SELECT ds AS Dates, ROUND(COUNT(event)/SUM(time\_spent), 2) AS “Daily Throughput” FROM job\_data

GROUP BY ds ORDER BY ds;

On date 2020-11-28 the throughput is highest **0.06**.



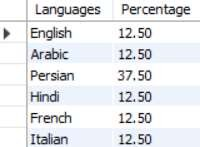
Metrics will always go up and down on a weekly and daily basis. You’ll get numbers faster every day or minute if you want. As a result, rolling metrics are superb at showing if your metrics are trending up or down on a daily level.

**3. Calculate the percentage share of each language in the last 30 days?**

SELECT language AS Languages, ROUND(100 \* COUNT(\*)/total, 2) AS Percentage FROM job\_data

CROSS JOIN (SELECT COUNT(\*) AS total FROM job\_data) sub GROUP BY language;

Persian language is highest with **37.5%**total.



**4. Let’s say you see some duplicate rows in the data. How will you display duplicates from the table?**

SELECT actor\_id, COUNT(\*) AS Duplicates FROM job\_data

GROUP BY actor\_id HAVING COUNT(\*) > 1;

Actor ID **1003**has duplicate rows.

https://tech.pinkybe.in/wp-content/uploads/2023/05/Picture7.png

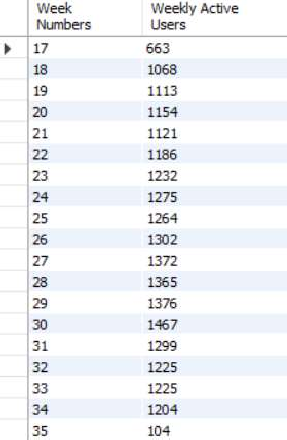
Case Study 2 (Investigating Metric Spike)

**1. Calculate the weekly user engagement?**

SELECT EXTRACT(WEEK FROM occurred\_at) AS “Week Numbers”, COUNT(DISTINCT user\_id) AS “Weekly Active Users”

FROM events

WHERE event\_type = ‘engagement’ GROUP BY 1;



**2. Calculate the user growth for product?**

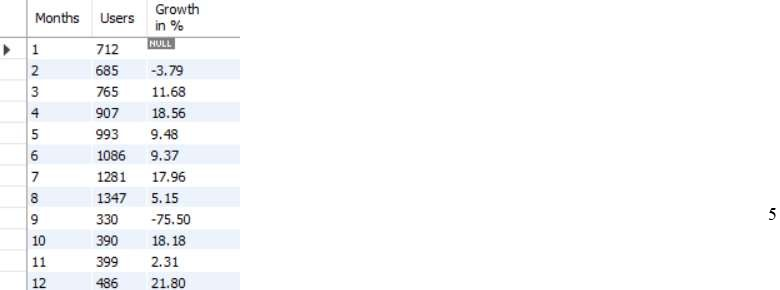
SELECT Months, Users, ROUND(((Users/LAG(Users, 1) OVER (ORDER BY Months) –

1)\*100), 2) AS “Growth in %” FROM

(SELECT EXTRACT(MONTH FROM created\_at) AS Months, COUNT(activated\_at) AS Users FROM users

WHERE activated\_at NOT IN(“”) GROUP BY 1

ORDER BY 1) sub;



**3. Calculate the weekly retention of users-sign up cohort?**

SELECT first AS “Week Numbers”,

SUM(CASE WHEN week\_number = 0 THEN 1 ELSE 0 END) AS “Week 0”,

SUM(CASE WHEN week\_number = 1 THEN 1 ELSE 0 END) AS “Week 1”,

SUM(CASE WHEN week\_number = 2 THEN 1 ELSE 0 END) AS “Week 2”,

SUM(CASE WHEN week\_number = 3 THEN 1 ELSE 0 END) AS “Week 3”,

SUM(CASE WHEN week\_number = 4 THEN 1 ELSE 0 END) AS “Week 4”,

SUM(CASE WHEN week\_number = 5 THEN 1 ELSE 0 END) AS “Week 5”,

SUM(CASE WHEN week\_number = 6 THEN 1 ELSE 0 END) AS “Week 6”,

SUM(CASE WHEN week\_number = 7 THEN 1 ELSE 0 END) AS “Week 7”,

SUM(CASE WHEN week\_number = 8 THEN 1 ELSE 0 END) AS “Week 8”,

SUM(CASE WHEN week\_number = 9 THEN 1 ELSE 0 END) AS “Week 9”,

SUM(CASE WHEN week\_number = 10 THEN 1 ELSE 0 END) AS “Week 10”,

SUM(CASE WHEN week\_number = 11 THEN 1 ELSE 0 END) AS “Week 11”,

SUM(CASE WHEN week\_number = 12 THEN 1 ELSE 0 END) AS “Week 12”,

SUM(CASE WHEN week\_number = 13 THEN 1 ELSE 0 END) AS “Week 13”,

SUM(CASE WHEN week\_number = 14 THEN 1 ELSE 0 END) AS “Week 14”,

SUM(CASE WHEN week\_number = 15 THEN 1 ELSE 0 END) AS “Week 15”,

SUM(CASE WHEN week\_number = 16 THEN 1 ELSE 0 END) AS “Week 16”,

SUM(CASE WHEN week\_number = 17 THEN 1 ELSE 0 END) AS “Week 17”,

SUM(CASE WHEN week\_number = 18 THEN 1 ELSE 0 END) AS “Week 18”, FROM

(

SELECT m.user\_id, m.login\_week, n.first, m.login\_week – first AS week\_number FROM

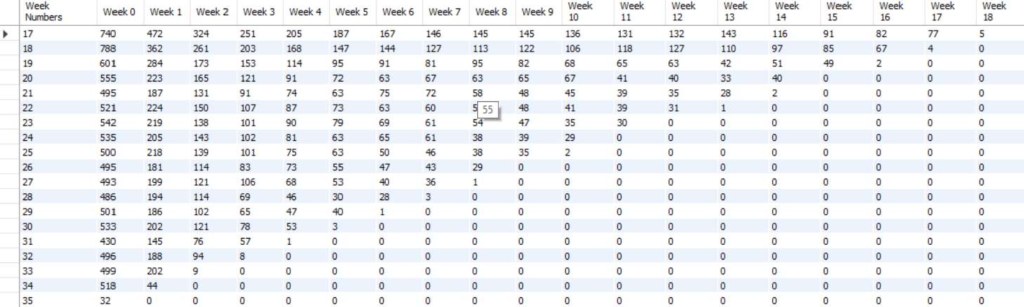
(SELECT user\_id, EXTRACT(WEEK FROM occurred\_at) AS login\_week FROM events GROUP BY 1, 2) m,

(SELECT user\_id, MIN(EXTRACT(WEEK FROM occurred\_at)) AS first FROM events GROUP BY 1) n

WHERE m.user\_id = n.user\_id

) sub

GROUP BY first ORDER BY first;



6

**4. Calculate the weekly engagement per device?**

SELECT EXTRACT(WEEK FROM occurred\_at) AS “Week Numbers”,

COUNT(DISTINCT CASE WHEN device IN(‘dell inspiron notebook’) THEN user\_id ELSE NULL END) AS “Dell Inspiron Notebook”,

COUNT(DISTINCT CASE WHEN device IN(‘iphone 5’) THEN user\_id ELSE NULL END) AS

“iPhone 5”,

COUNT(DISTINCT CASE WHEN device IN(‘iphone 4s’) THEN user\_id ELSE NULL END) AS

“iPhone 4S”,

COUNT(DISTINCT CASE WHEN device IN(‘windows surface’) THEN user\_id ELSE NULL END) AS “Windows Surface”,

COUNT(DISTINCT CASE WHEN device IN(‘macbook air’) THEN user\_id ELSE NULL END) AS “Macbook Air”,

COUNT(DISTINCT CASE WHEN device IN(‘iphone 5s’) THEN user\_id ELSE NULL END) AS

“iPhone 5S”,

COUNT(DISTINCT CASE WHEN device IN(‘macbook pro’) THEN user\_id ELSE NULL END) AS “Macbook Pro”,

COUNT(DISTINCT CASE WHEN device IN(‘kindle fire’) THEN user\_id ELSE NULL END) AS “Kindle Fire”,

COUNT(DISTINCT CASE WHEN device IN(‘ipad mini’) THEN user\_id ELSE NULL END) AS “iPad Mini”,

COUNT(DISTINCT CASE WHEN device IN(‘nexus 7’) THEN user\_id ELSE NULL END) AS

“Nexus 7”,

COUNT(DISTINCT CASE WHEN device IN(‘nexus 5’) THEN user\_id ELSE NULL END) AS

“Nexus 5”,

COUNT(DISTINCT CASE WHEN device IN(‘samsung galaxy s4’) THEN user\_id ELSE NULL END) AS “Samsung Galaxy S4”,

COUNT(DISTINCT CASE WHEN device IN(‘lenovo thinkpad’) THEN user\_id ELSE NULL END) AS “Lenovo Thinkpad”,

COUNT(DISTINCT CASE WHEN device IN(‘samsumg galaxy tablet’) THEN user\_id ELSE NULL END) AS “Samsumg Galaxy Tablet”,

COUNT(DISTINCT CASE WHEN device IN(‘acer aspire notebook’) THEN user\_id ELSE NULL END) AS “Acer Aspire Notebook”,

COUNT(DISTINCT CASE WHEN device IN(‘asus chromebook’) THEN user\_id ELSE NULL END) AS “Asus Chromebook”,

COUNT(DISTINCT CASE WHEN device IN(‘htc one’) THEN user\_id ELSE NULL END) AS “HTC One”,

COUNT(DISTINCT CASE WHEN device IN(‘nokia lumia 635’) THEN user\_id ELSE NULL END) AS “Nokia Lumia 635”,

COUNT(DISTINCT CASE WHEN device IN(‘samsung galaxy note’) THEN user\_id ELSE NULL END) AS “Samsung Galaxy Note”,

COUNT(DISTINCT CASE WHEN device IN(‘acer aspire desktop’) THEN user\_id ELSE NULL END) AS “Acer Aspire Desktop”,

COUNT(DISTINCT CASE WHEN device IN(‘mac mini’) THEN user\_id ELSE NULL END) AS “Mac Mini”,

COUNT(DISTINCT CASE WHEN device IN(‘hp pavilion desktop’) THEN user\_id ELSE NULL END) AS “HP Pavilion Desktop”,

COUNT(DISTINCT CASE WHEN device IN(‘dell inspiron desktop’) THEN user\_id ELSE NULL END) AS “Dell Inspiron Desktop”,

COUNT(DISTINCT CASE WHEN device IN(‘ipad air’) THEN user\_id ELSE NULL END) AS “iPad Air”,

7

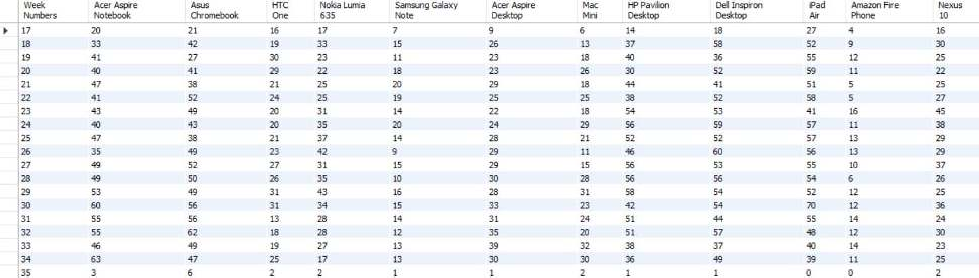
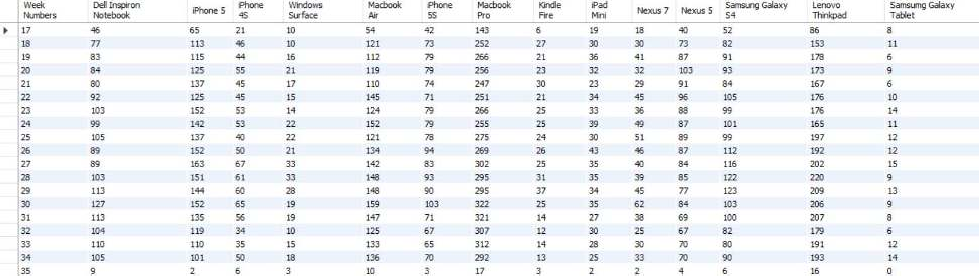
COUNT(DISTINCT CASE WHEN device IN(‘amazon fire phone’) THEN user\_id ELSE NULL END) AS “Amazon Fire Phone”,

COUNT(DISTINCT CASE WHEN device IN(‘nexus 10’) THEN user\_id ELSE NULL END) AS

“Nexus 10” FROM events

WHERE event\_type = ‘engagement’ GROUP BY 1

ORDER BY 1;



8

**5. Calculate the email engagement metrics?**

SELECT Week,

ROUND((weekly\_digest/total\*100),2) AS “Weekly Digest Rate”, ROUND((email\_opens/total\*100),2) AS “Email Open Rate”, ROUND((email\_clickthroughs/total\*100),2) AS “Email Clickthrough Rate”, ROUND((reengagement\_emails/total\*100),2) AS “Reengagement Email Rate” FROM

(SELECT EXTRACT(WEEK FROM occurred\_at) AS Week,

COUNT(CASE WHEN action = ‘sent\_weekly\_digest’ THEN user\_id ELSE NULL END) AS weekly\_digest,

COUNT(CASE WHEN

email\_opens,

action = ’email\_open’ THEN user\_id   ELSE

NULL END)   AS

COUNT(CASE WHEN action = ’email\_clickthrough’ THEN user\_id ELSE NULL END) AS email\_clickthroughs,

COUNT(CASE WHEN action = ‘sent\_reengagement\_email’ THEN user\_id ELSE NULL END) AS reengagement\_emails,

COUNT(user\_id) AS total FROM email\_events GROUP BY 1

) sub

GROUP BY 1

ORDER BY 1;

