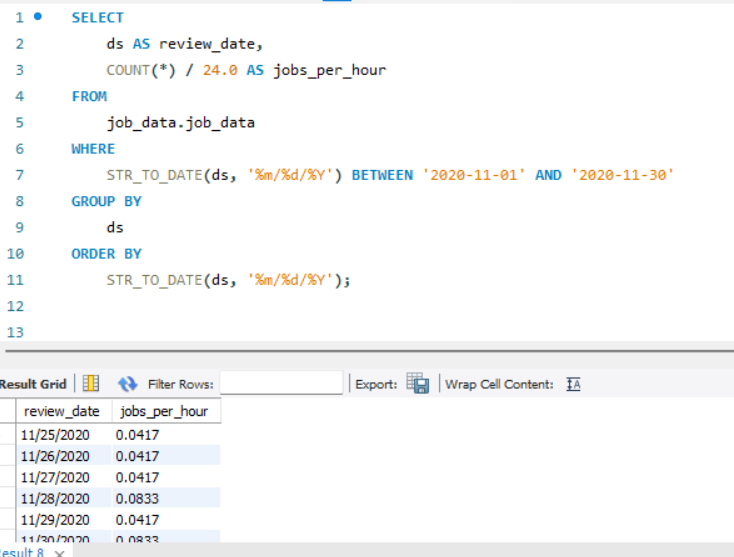
**Insights:**

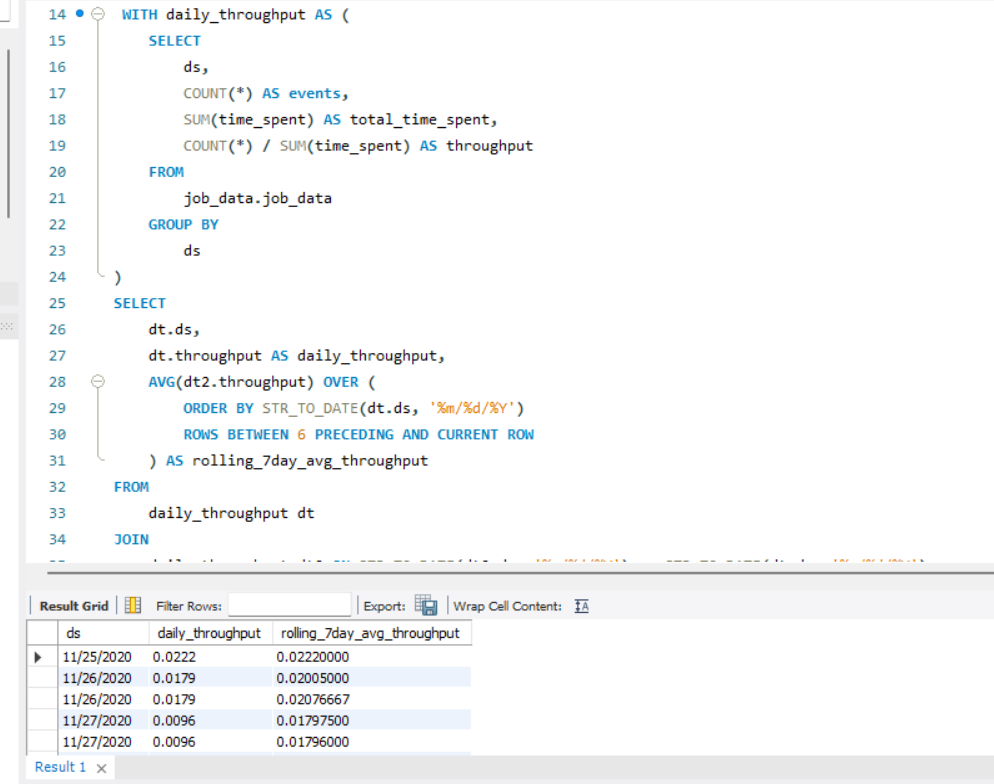
**Case Study 1: Job Data Analysis**

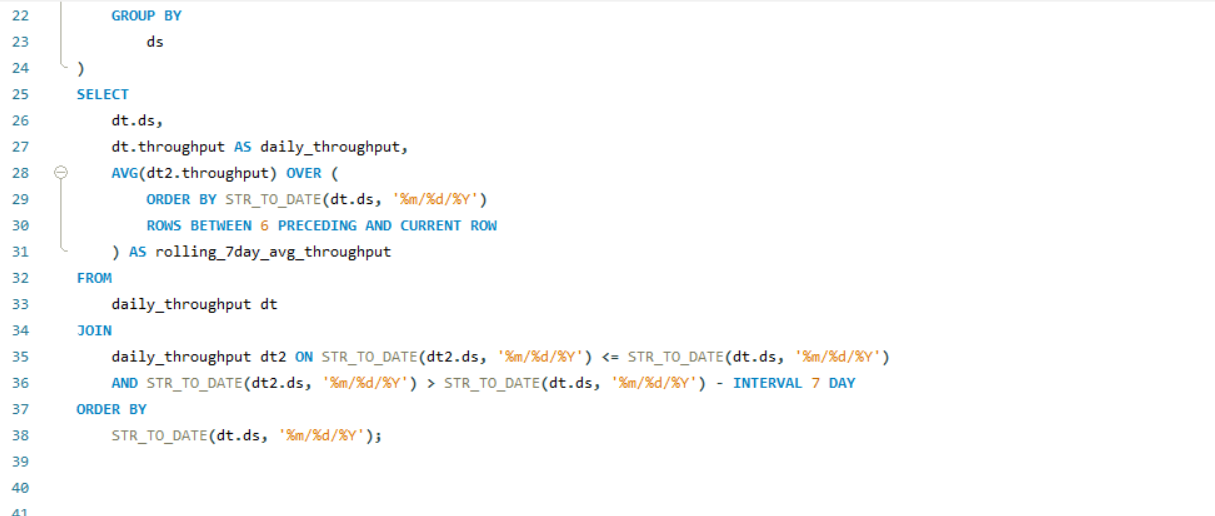
1. **Jobs Reviewed over time: In this, the number of jobs reviewed per hour is identified and calculated.**

****

**In this, the ds column is used. All rows are counted and divided by 24.0 to get the hourly rate. The WHERE clause is prominent as it filters for that specific date, in this case, it is November 2020. STR\_TO\_DATE is used to convert string dates to a compatible format. Finally, the output is ordered by date.**

1. **Throughput Analysis: In this, the 7-day rolling average of throughput which is events per second is calculated. It is done to smooth out any fluctuations.**

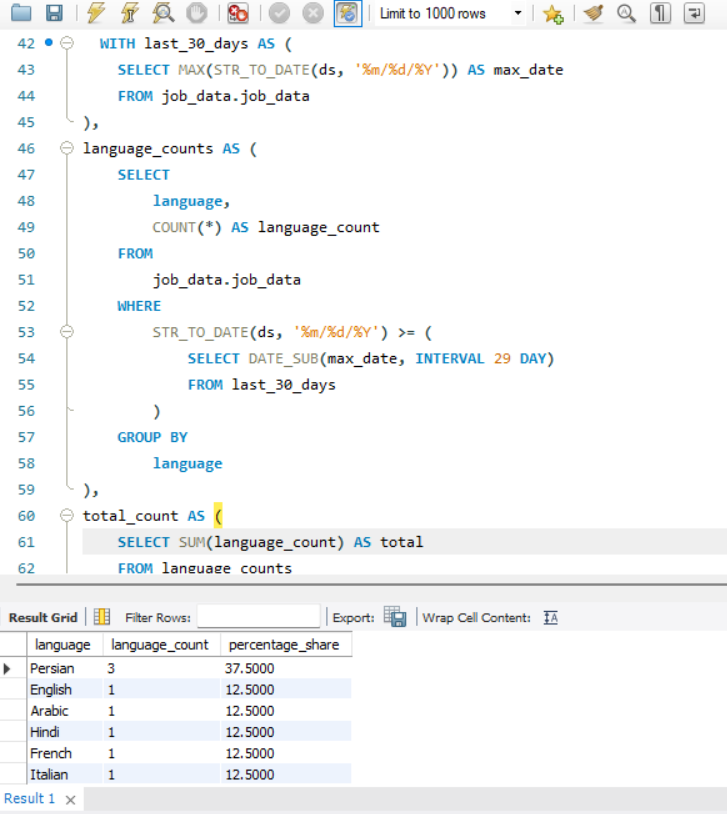
****

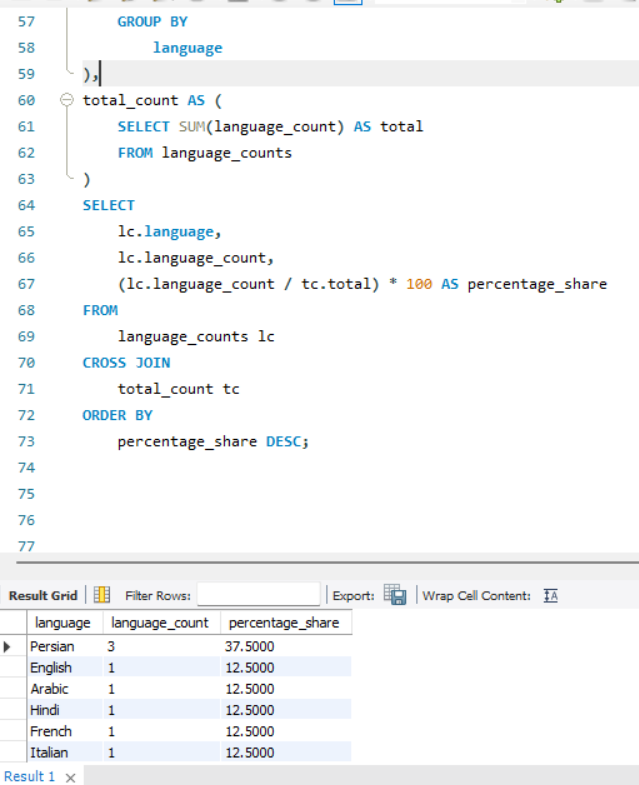
****

**First of all, daily throughput is calculated then with the help of the window function 7-day rolling average is calculated. Then STR\_TO\_DATE is used to change the date’s format.**

**I would prefer a 7-day rolling average for throughput because:**

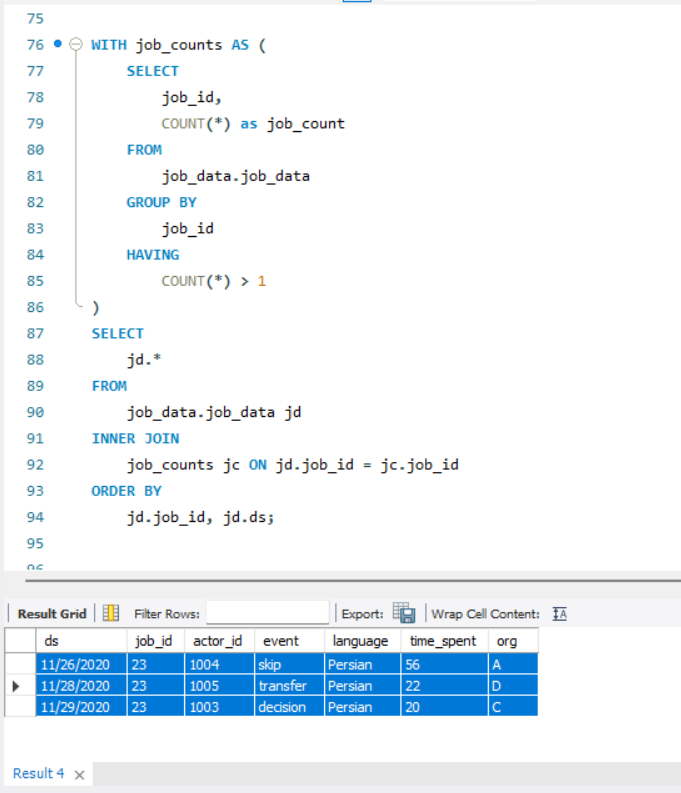
1. **It helps to provide a more stable view of the trend.**
2. **Unusual spikes have less impact on it.**
3. **It is compatible with weekly cycles of the company’s operations.**
4. **It is easier to point out trends and changes over time.**
5. **It is preferred for making informed and strategic decisions.**
6. **Language Share Analysis: In this, the percentage share of each language is being found in the last 30 days.**

****

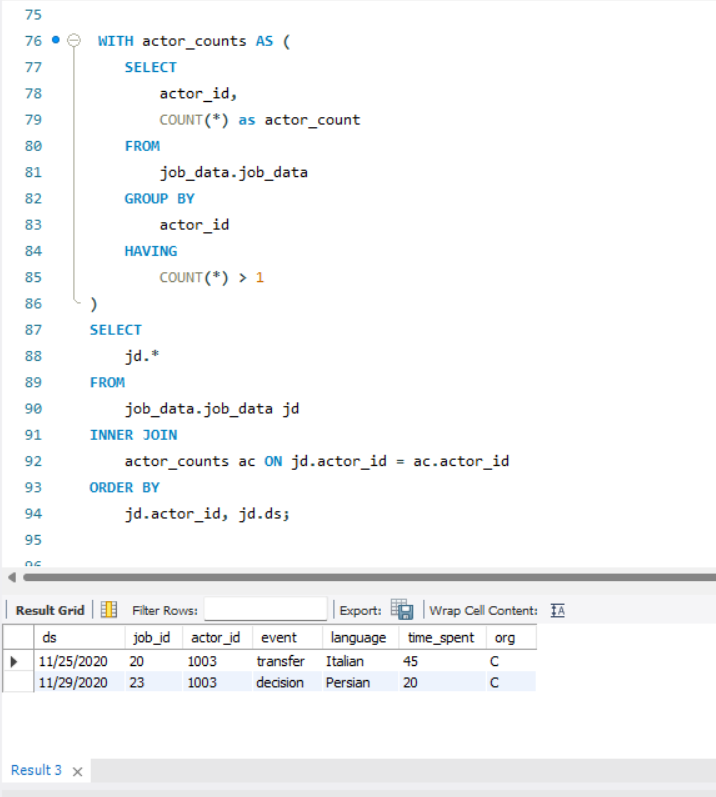
****

**With the help of the last\_30 \_days the most recent date is found. The language\_counts count the occurrence of each language in the last 30 days. The Date\_Sub is used to get a date 29 days before the max date. The total\_count is used to find the total number of entries in the last 30 days. Finally, language\_counts are joined with total\_count and percentage shares are found for each language. Then the output is ordered in descending order.**

1. **Duplicate Rows Detection: In this any duplicate rows is detected and displayed. Here I have taken two cases where I could find duplicate rows, first taking job\_id as the main row and second taking actor\_id as the main row.**

****

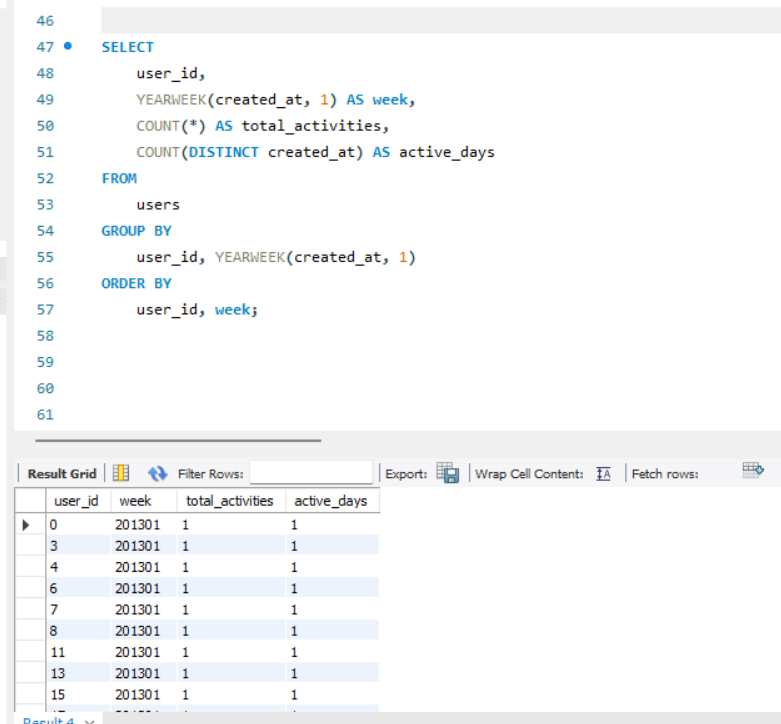
**In this job\_counts identifies all job\_id occurring more than once, after this main query joins its output to original table to get all rows for duplicate job\_id(s). The output is ordered by job\_id and ds.**

****

**The actor\_counts identifies actor\_id appearing more than once in the table after this main query joins its output to the original table to get all rows for duplicate actor\_id(s). The output is ordered by actor\_id and ds.**

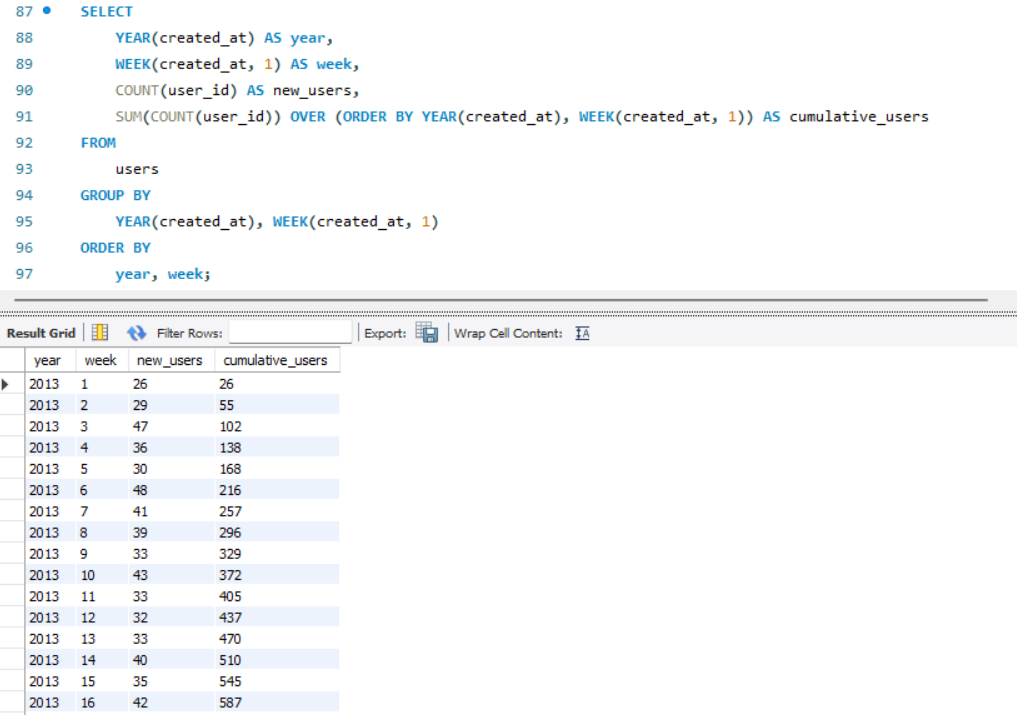
**Case Study 2: Investigating Metric Spike**

1. **Weekly User Engagement: In this, the activeness of users is measured on weekly basis.**

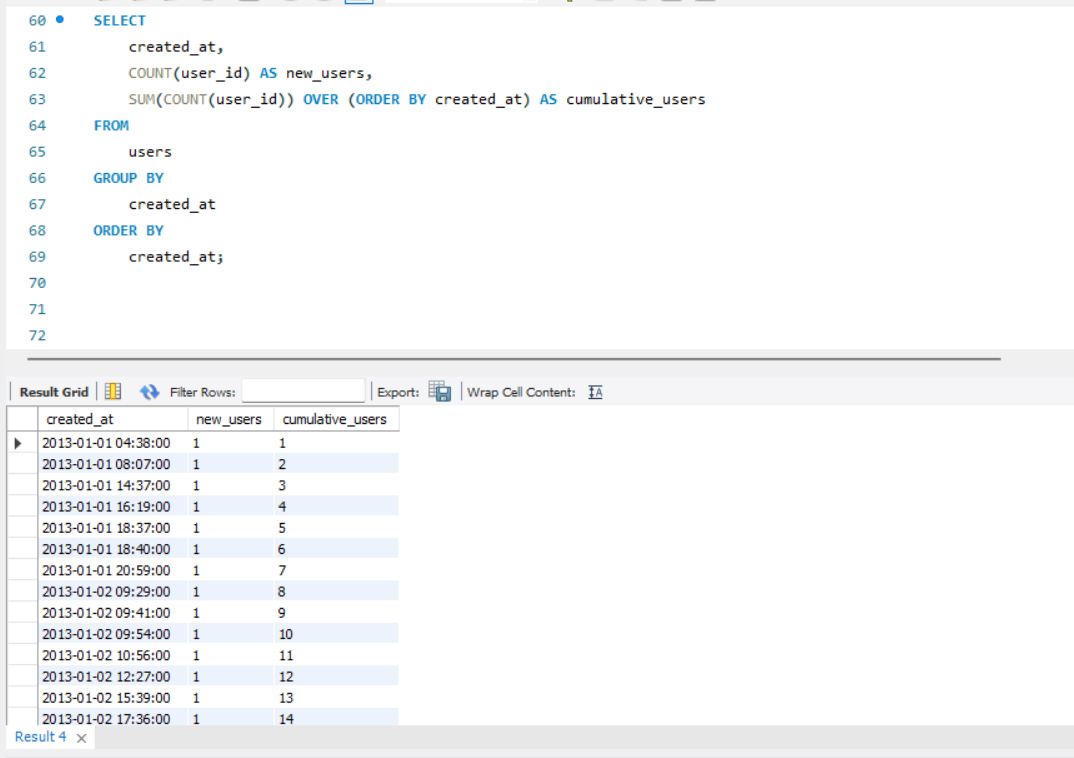
****

**In this, YEARWEEK is used to calculate the week number where 1 shows that the week starts on Monday. COUNT \* is being used to count all activities within a week. Whereas, the COUNT (DISTINCT created\_at) counts the unique days user was active. It is grouped by user\_id and YEARWEEK and ordered by user\_id and week to show each user’s activity week by week.**

1. **User Growth Analysis:**
2. **Weekly Analysis**

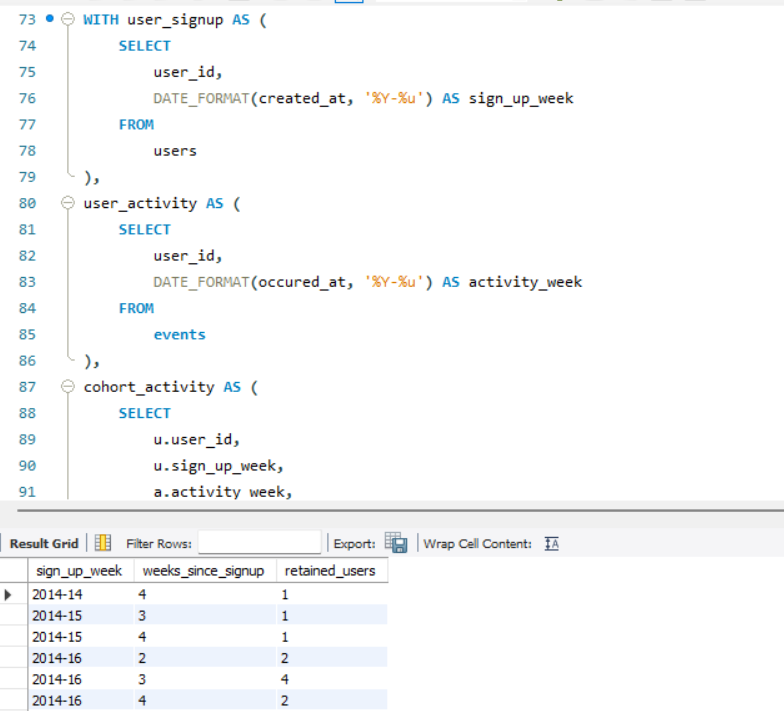
****

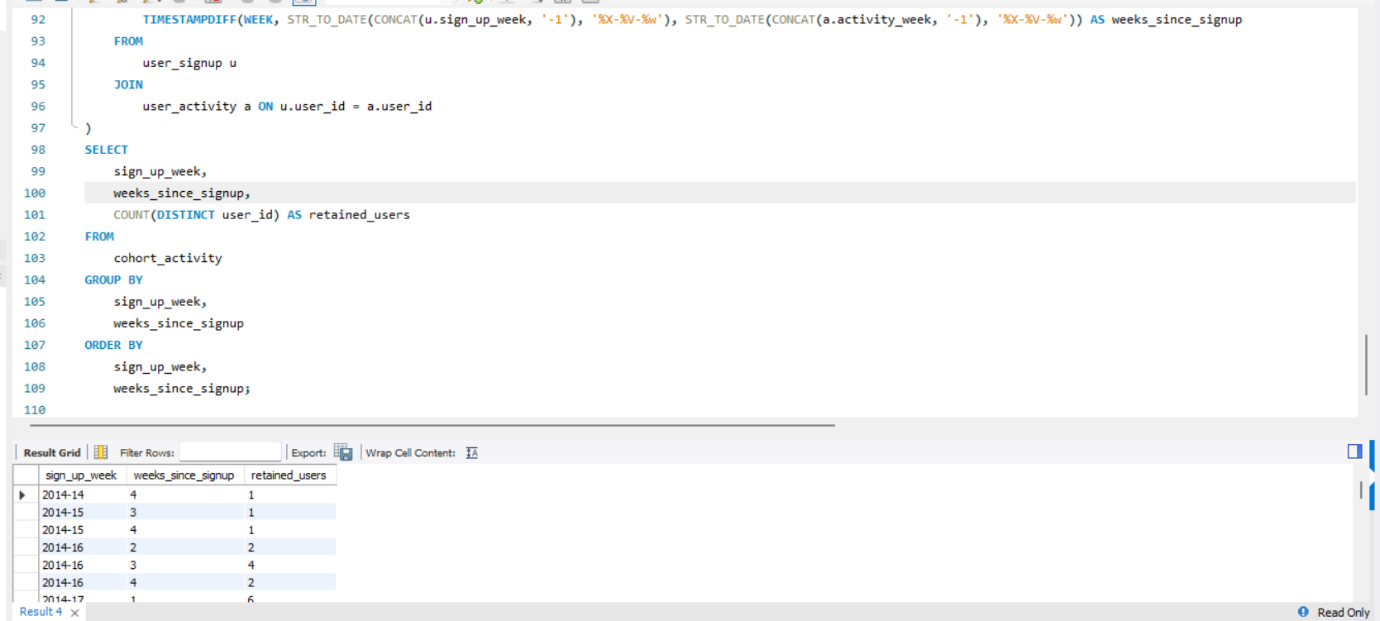
1. **Daily Analysis**

****

**In this, the data on the growth of users over time for a specific product is obtained. In this the user growth are being dealt with. The select clause has YEAR, WEEK, COUNT(user\_id), and SUM(COUNT(user\_id)) which calculates the time, users created in each week. The created\_at is ultimately used by both to be grouped and finally is ordered by the same.**

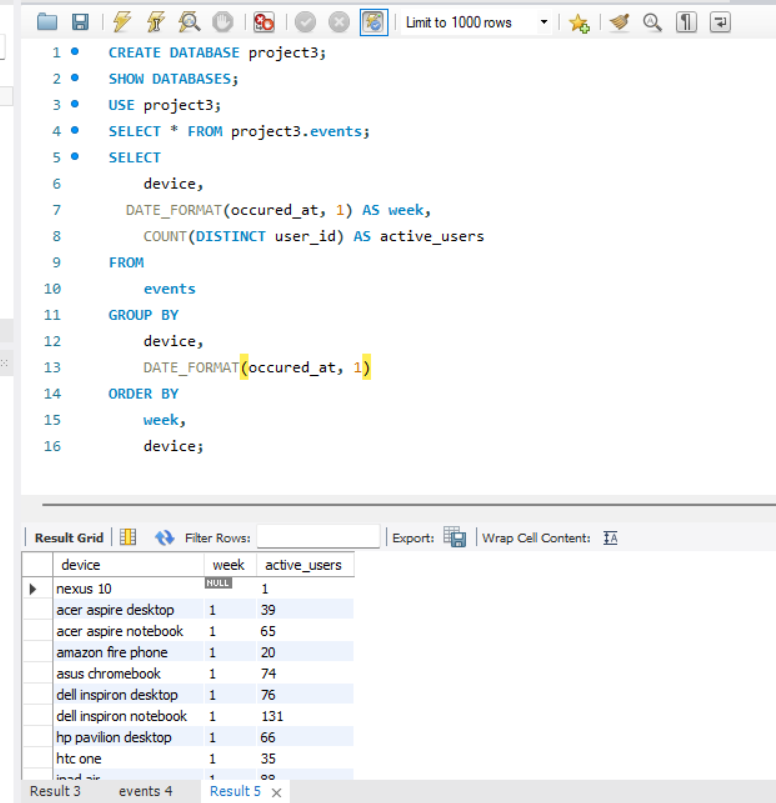
1. **Weekly Retention Analysis:**

****

****

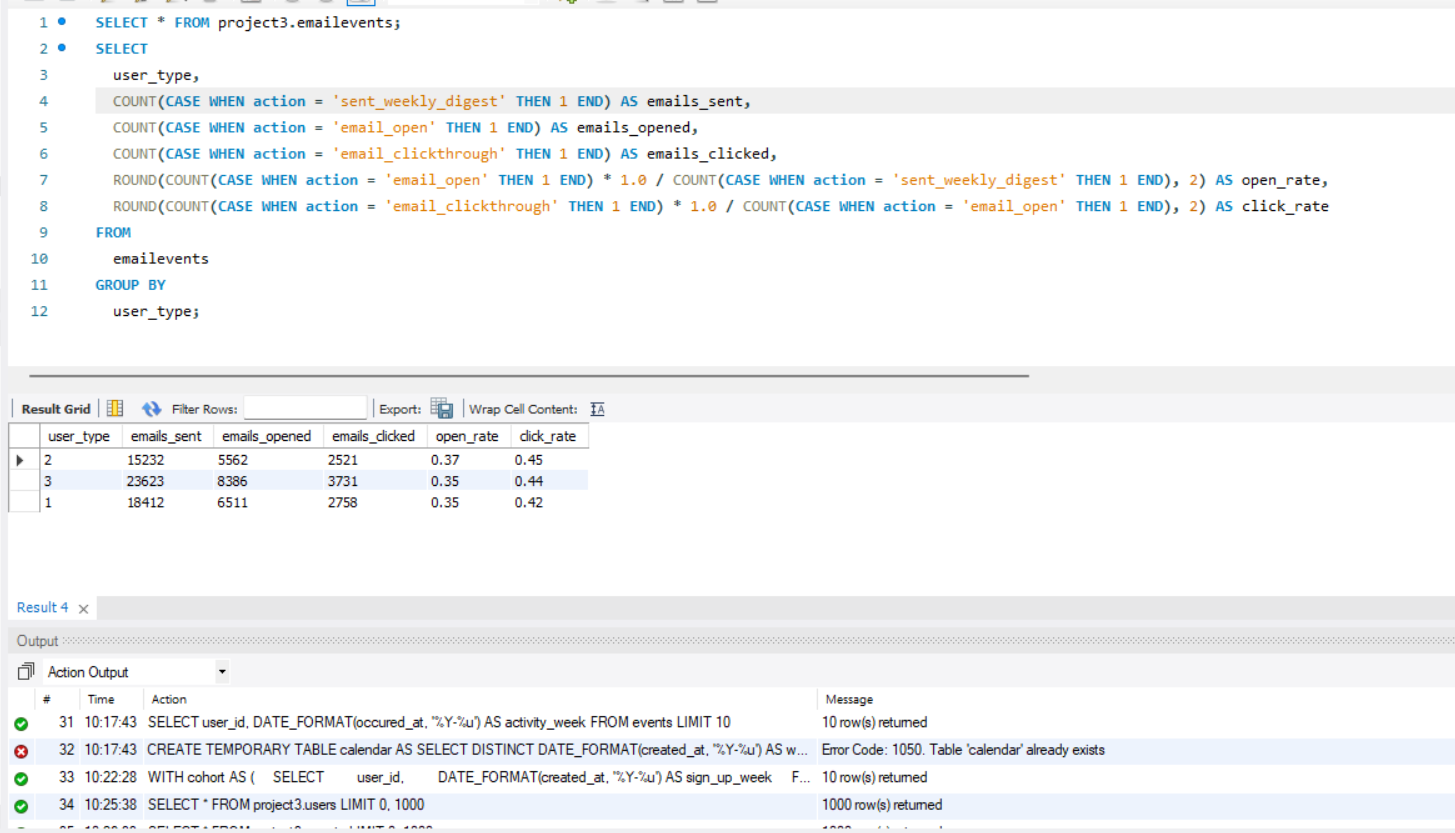
**In this user retention on a weekly basis based on their sign-up activity is done. The WITH clause states the sign-in data of user. In this the time formatted into year and week format. The user\_activity displays the result of users with their activity in a week. In the cohort\_activity section data of weeks since signup is obtained. In this the output is users with their activity in a week is displayed.**

1. **Weekly Engagement Per Device:**

****

**Here, again the activity of users on a specific device is being monitored. It will give an insight into the weekly engagement of a user on a particular device. In this, DATE\_FORMAT(occurred\_at,1) AS week makes the timestamp in week format where the week starts from Monday. Wheres, the COUNT(DISTINCT user\_id)counts number of users for each device.**

1. **Email Engagement Analysis:**

****

**In this, the data involving email services being used by users is involved. In this the email management metrics is being dealt with. The ROUND(COUNT) case of open rate calculates open-rate for users by dividing by 1.0 and rounding it off, whereas the click rate case calculates click-to-open rate.**

**Result:**

**Achievements:**

1. **I got to know how to understand and explain the sudden changes in sales, ser engagement, how to identify duplicate data, user growth, engagement per device, email engagement etc.**
2. **I also learned how to import a dataset (CSV file) into MySQL workshop. I got to know how to create a database for a project and create tables with the help of table structures.**
3. **Along with understanding the impact of metrics spikes I also understood the effectiveness of email communications and optimizing email marketing strategies.**
4. **I got to learn how to use SQL queries for monitoring and analysis of user engagement and job data.**

**Insights gained:**

1. **With the help of this project, I gained deeper insights as to how valuable insights are collected from data to help in a company’s end-to-end operations.**
2. **I also got to know how to derive useful data for marketing, operations, support and development teams, and investors thus ultimately helping in the platform’s growth.**
3. **Last but not least it also helped me gain profound skills in SQL, how to work with MYSQL workbench software, and how to clean and prepare the data. This all helped me build my skills in data analysis.**