**Operation Analytics and Investigating Metric Spike**

**https://github.com/AllThingsDataWithAngelina/Data-Analyst-Projects/blob/main/cohort\_rentention\_analysis.sql**

**Advanced SQL**

Difficulty Level:

**Description:**

Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. As a Data Analyst, you'll work closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.  
  
One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales. As a Data Analyst, you'll need to answer these questions daily, making it crucial to understand how to investigate these metric spikes.  
  
In this project, you'll take on the role of a Lead Data Analyst at a company like Microsoft. You'll be provided with various datasets and tables, and your task will be to derive insights from this data to answer questions posed by different departments within the company. Your goal is to use your advanced SQL skills to analyze the data and provide valuable insights that can help improve the company's operations and understand sudden changes in key metrics.

**Case Study 1: Job Data Analysis**

**You will be working with a table named job\_data with the following columns:**

* **job\_id:**Unique identifier of jobs
* **actor\_id:**Unique identifier of actor
* **event:**The type of event (decision/skip/transfer).
* **language:**The Language of the content
* **time\_spent:**Time spent to review the job in seconds.
* **org:**The Organization of the actor
* **ds:**The date in the format yyyy/mm/dd (stored as text).

**Tasks:**

1. **Jobs Reviewed Over Time:**
   * Objective: Calculate the number of jobs reviewed per hour for each day in November 2020.
   * Your Task: Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.
2. **Throughput Analysis:**
   * Objective: Calculate the 7-day rolling average of throughput (number of events per second).
   * Your Task: 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.
3. **Language Share Analysis:**
   * Objective: Calculate the percentage share of each language in the last 30 days.
   * Your Task: Write an SQL query to calculate the percentage share of each language over the last 30 days.
4. **Duplicate Rows Detection:**
   * Objective: Identify duplicate rows in the data.
   * Your Task: Write an SQL query to display duplicate rows from the job\_data table.

**Case Study 2: Investigating Metric Spike**

**You will be working with three tables:**

* **users**: Contains one row per user, with descriptive information about that user’s account.
* **events**: Contains one row per event, where an event is an action that a user has taken (e.g., login, messaging, search).
* **email\_events**: Contains events specific to the sending of emails.

**Tasks:**

1. **Weekly User Engagement:**
   * Objective: Measure the activeness of users on a weekly basis.
   * Your Task: Write an SQL query to calculate the weekly user engagement.
2. **User Growth Analysis:**
   * Objective: Analyze the growth of users over time for a product.
   * Your Task: Write an SQL query to calculate the user growth for the product.
3. **Weekly Retention Analysis:**
   * Objective: Analyze the retention of users on a weekly basis after signing up for a product.
   * Your Task: Write an SQL query to calculate the weekly retention of users based on their sign-up cohort.
4. **Weekly Engagement Per Device:**
   * Objective: Measure the activeness of users on a weekly basis per device.
   * Your Task: Write an SQL query to calculate the weekly engagement per device.
5. **Email Engagement Analysis:**
   * Objective: Analyze how users are engaging with the email service.
   * Your Task: Write an SQL query to calculate the email engagement metrics.

Please note that for each task, you should also provide insights and interpretations of the results obtained from your queries.

**How to do this Project?**

1. **Create the Database and Tables:**Start by creating a database for the project and then create the necessary tables using the provided table structures and links. You need to import the csv file into MySQL Workbench.
2. **Perform Analysis:**Utilize SQL to perform the analysis and answer the questions mentioned in the case studies. Ensure that you understand the table structures and the meaning of various columns such as event types and what to consider for reviewing.
3. **Submit a Report:**Prepare a report (PDF/PPT) to present your findings to the leadership team. The report should include the following details:

**Project Description**

Provide a brief overview of the project, explaining its purpose and how you plan to handle the analysis.

**Approach**

Describe your approach towards the project and explain how you executed the analysis.

**Tech-Stack Used**

Mention the software and versions used for the project, such as MySQL Workbench, and explain their purpose in the analysis.

**Insights**

Summarize the insights and knowledge you gained during the project. Highlight key observations and inferences you made from the data, keeping them concise and to the point.

**Result**

Discuss what you have achieved through the project and explain how it has contributed to your understanding and decision-making.

**Drive Link**

Save your report as a PDF file and upload it to your Google Drive. Include the shareable link (with public visibility) in your report. Check the detailed steps below for submission.

**Note: Make sure you mention the SQL Queries along with the outputs in the report. You can take snapshots of the SQL Queries and the outputs. It is advised to do the project in MySQL Workbench.**

**How to Submit your Project:**

**Step 1** Do your project in PPT/PDF. It’s your choice. For PDF, write your project in Ms Word, then save the file in PDF format. You will find the option in the ‘Save As’ Box.  
  
**Step 2** Now, upload the file i.e., PPT/PDF on your google drive. It would be wise to create a Single Folder and Name it “Trainity Assignments”. Inside that folder, create another Folder name “Project Name”. Upload your assignment there.  
  
**Step 3** Right Click on the File, and select ‘Get Link’ under ‘Share’. A Dialog Box will Pop up. Inside the Box, Under General Access, Change ‘Restricted’ to ‘Anyone with the Link’. No need to change anything else.  
  
**Step 4** Click on ‘Copy Link’.  
  
**Step 5** Now paste the link in the Drive Link box when submitting the project.

**Judgement Criteria:**

**SQL Understanding**

The SQL queries used should provide correct outputs and be easily understandable. Ensure that the queries are accurate, efficient, and well-structured.

**Case study completion**

All the questions presented in the case studies must be answered completely with accurate responses. Address each question thoroughly, demonstrating your understanding of the analysis required.

**Insights**

Use your own imagination and creativity to answer the case study questions, providing insightful and meaningful observations and recommendations. Think beyond the obvious and showcase your ability to derive valuable insights from the data.

**Plagiarism**

The project should be your own original work and should not be copied from the internet or any other source. Ensure that the analysis, interpretations, and visualizations are your own, demonstrating your skills as a data analyst.

By following these guidelines and demonstrating your SQL understanding, completing the case studies, providing insightful observations, and avoiding plagiarism, you will showcase your expertise and deliver a comprehensive and original SQL project.