Operation Analytics and Investigating Metric Spike

Advanced SQL

Difficulty Level: 4/5

Description:

Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. You work closely with the ops team, support team, marketing team, etc and help them derive insights out of the data they collect.

Being one of the most important parts of a company, this kind of analysis is further used to predict the overall growth or decline of a company's fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows.

Investigating metric spike is also an important part of operation analytics as being a Data Analyst you must be able to understand or make other teams 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 its very important to investigate metric spike.

You are working for a company like Microsoft designated as Data Analyst Lead and is provided with different data sets, tables from which you must derive certain insights out of it and answer the questions asked by different departments.

You are required to provide a detailed report for the below two operations mentioning the answers for the related questions:

Case Study 1 (Job Data)

Below is the structure of the table with the definition of each column that you must work on:

- **Table-1:** job_data

    - **job_id:** unique identifier of jobs
    - **actor_id:** unique identifier of actor
    - **event:** decision/skip/transfer
    - **language:** language of the content
    - **time_spent:** time spent to review the job in seconds
    - **org:** organization of the actor
    - **ds:** date in the yyyy/mm/dd format. It is stored in the form of text and we use presto to run. no need for date function

Use the dataset attached in the Dataset section below the project images then answer the questions that follows

    A. **Number of jobs reviewed:** Amount of jobs reviewed over time.
        Your task: Calculate the number of jobs reviewed per hour per day for November 2020?
    B. **Throughput:** It is the no. of events happening per second.
        Your task: Let's say the above metric is called throughput. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?
    C. **Percentage share of each language:** Share of each language for different contents.
        Your task: Calculate the percentage share of each language in the last 30 days?
    D. **Duplicate rows:** Rows that have the same value present in them.
        Your task: Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

Case Study 2 (Investigating metric spike)

The structure of the table with the definition of each column that you must work on is present in the project image

- **Table-1:** users
  This table includes one row per user, with descriptive information about that user's account.
- **Table-2:** events
  This table includes one row per event, where an event is an action that a user has taken. These events include login events, messaging events, search events, events logged as users progress through a signup funnel, events around received emails.
- **Table-3:** email_events
  This table contains events specific to the sending of emails. It is similar in structure to the events table above.

Use the dataset attached in the Dataset section below the project images then answer the questions that follows

    A. **User Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service.
        Your task: Calculate the weekly user engagement?
    B. **User Growth:** Amount of users growing over time for a product.
        Your task: Calculate the user growth for product?
    C. **Weekly Retention:** Users getting retained weekly after signing-up for a product.
        Your task: Calculate the weekly retention of users-sign up cohort?
    D. **Weekly Engagement:** To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
        Your task: Calculate the weekly engagement per device?

E. **Email Engagement:** Users engaging with the email service.
   Your task: Calculate the email engagement metrics?

How to do this Project?

NOTE:- You have to spend some time understanding the table given. What does the event mean? What to consider for reviewing? should be some of the questions you have to be clear with in your mind. Use the database that you will create to answer the questions above.

1. **Create the Database and Tables:** You are supposed to create a database and then the tables using the structure and links provided.
2. **Perform Analysis:** Use SQL to perform your entire analysis answering the questions asked above.
3. **Submit a Report:** Make a report (PDF/PPT) to be presented to the leadership team. The report should/can contain the following details:

   Project Description
   Give a brief about your project description i.e. what is this project about, how are you going to handle the things and what are the things that you are going to find out through the project.

   Approach
   Write a short paragraph about your approach towards the project and how you have executed it.

   Tech-Stack Used
   Do mention the software and the version used while making the project (For Eg. Jupyter Notebook, etc) and mention the purpose of using it.

   Insights
   Jot down the insights and the knowledge you gained while making the project. You need to write that what do you infer about the things. Make sure its brief and up to the point only. For Eg. If you got a graph then what do you understand by the graph, what changes can you make or what can you derive from the graph.

   Result
   Mention what have you achieved while making the project and how do you think it has helped you.

   Drive Link
   Save your file as a ".pdf" file and upload it to your Google Drive. Mention the sharable link **(link visibility should be set to public)** in your pdf file which you will be uploading. Do not directly upload your project.

Judgement Criteria:

SQL Understanding
   The queries run should give correct output as well as should be easily understandable.

Case study completion

All the questions present must be answered completely having correct answers.

Insights

You need to use your own imagination to answer the case study while improvising it as well.

Plagiarism

Project submitted should not be copied from the internet or anywhere else, it should be your own work.