Data Analyst Technical Assignment

Introduction

We would like to thank you again for applying for this position.

In this assignment, we will assess the following:

- If you can write appropriate SQL queries for data analysis
- If you have basic knowledge of algorithms and complexity analysis

Note: There are 2 (!) assignments, don't miss the second one which is at the end of the document!

Submission details

Please submit all answers in a PDF. The format (slides, documents, etc.) is up to you.

Assignment 1 - SQL

Purpose

In Assignment 1, the following points will be assessed:

- If you can write appropriate SQL queries for data analysis
- If you can visualize your analysis results appropriately

Important notes

- In your queries, use /*comments*/ or -- comments to describe the intent of the operation as much as possible.
- For readable queries, please use "with" clauses as much as you can. ● If you need to, you can search SQL functions and formulas during the test. ● You can use any tools for visualization.

Environment settings

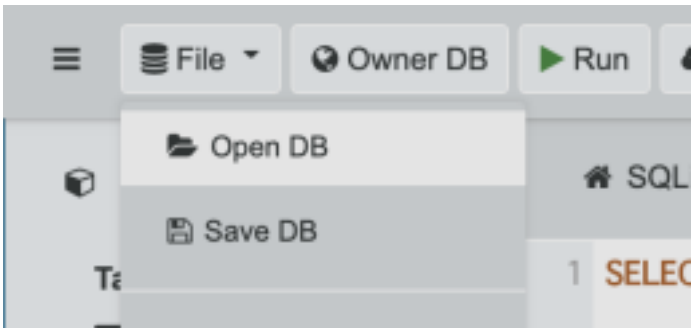
In this assignment, we will ask you to execute queries in an online SQL execution environment. Therefore, please start by preparing and setting up the following.

1. Prepare the data set

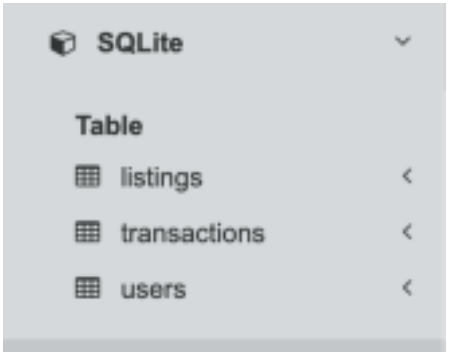
Please prepare the technical_assignment.db file that was sent with this document.

2. SQL execution environment settings

Use SQLite Online as an execution environment.
Open the file prepared in step 1 from File > Open DB in the upper left corner to load the data set required for this assignment.



When loaded correctly, “listings”, ”transactions”, and “users” will be listed under Table on the left.



3. Query execution

Write your query in the center editor section and click Run to execute the query.

4. Exporting the results

By clicking Export, you can export the results to CSV or other formats.

Table descriptions

When you load technical_assignment.db, you will see three different tables. These tables and their columns are described below.

These three tables contain data about a virtual online marketplace service launched on January 1, 2019. All of these data are dummy data and are not actual data from the Mercari service. The data covers a period of one year, from January 2019 to December 2019.

In this service:

- Users can register for the service.
- After registration, users can list items.
- After registration, users can purchase items listed by other users.

In addition, for simplicity, we have set the following conditions:

- Once an item is listed, the price cannot be changed.
- No cancellation will occur after the transaction has occurred
- Since this is a free-market service, there shall be one inventory for every item.

Table 1: users

This table records user registrations to the service.

- Columns
    - user_id: ID uniquely determined for each user
    - created: Time the user registered

Table 2: listings

This table records information about items listed by users.

- Columns
    - item_id: ID uniquely determined for each item
    - created: Time the user listed the item
    - lister_id: User ID of the user who listed the item
    - category_id: ID of the category to which the listed item belongs
    - category_name: The name of the category to which the listed item belongs ○ price: Price of the listed item

Table 3: transactions

This table records information about transactions between users. A transaction is defined as the purchase of an item from one user by another user.

- Columns
    - transaction_id: ID that is uniquely determined for each transaction
    - created: Time the transaction occurred (= time the buyer made the purchase)
    - item_id: Item ID of the purchased item
    - buyer_id: User ID of the user who made the purchase
    - seller_id: User ID of the user who sold the item

What to submit

For 1-1 and 1-2 below, please submit the calculated results and the query used for the calculation. For 1-3, please submit the answer visualized in a way that is easy to understand and the query used for the calculations.

Questions

1-1

Get the name of the category with the highest total transaction value in September 2019, along with that total transaction value.
(Total transaction value = sum of the transaction prices in yen)

1-2

Of users who listed items two or more times, calculate the percentage (to one decimal place) of users who listed their first and second items in the same category.

1-3

For each category, calculate the average value per transaction (in whole numbers) for each month.

Assignment 2 - Algorithm

Provide meta-code that implements a double-ended queue with two stacks.

What is the runtime complexity of inserting and removing items (worst case). Discuss the runtime complexity according to workload.