

Data Platform Craft Demo

Overview

Thanks for taking the time to complete this assessment as a part of the interview process. There are two parts to this assignment: the SQL assessment and the Python assessment. These parts are separate and do not depend on the output from the other, and you may begin with whichever task you'd like.

Please try to keep the total amount of time spent on the entire assignment to around 2 hours or less, even if you are not able to fully complete the tasks. Even partial answers will allow us to continue the conversation during the technical interview.

This assessment is meant to be very straightforward i.e. no gotchas, tricks, etc. If you think there's an obvious solution, you're probably right, and we encourage you to spend more time implementing it than questioning if it's right.

When returning the assessment, please include the following:

- SQL query
- .CSV output of the SQL query
- Python code
- .CSV output of the Python code

Background

Attached you will find the three .CSV files of sample data described below. These files will be used for both portions of the assessment.

- **date_dimension:** a table that contains different time periods
- **customer_dimension:** a table that contains customer IDs and corresponding attributes
- **emails_sent_fact:** a table with information about emails sent by customers on a given day. It contains metrics about how many times a given email was opened and clicked

SQL Assessment

Using the tables described above, **please write an SQL query that will create a new table that contains the following:**

- Year
- Month name
- Month number (e.g. '12' for December)
- Country name
- Maximum number of emails opened
- Running total of email opens for each country (from the beginning of time to now)
- Number of customers that sent at least 1 email

The grain of the data in this table is defined by the columns: **year, month name, month number, and country name.**

Your query should be runnable, and you should also attach your query's output in CSV format.

Feel free to use any DB + tools you are comfortable with. We suggest using the free online tool <https://sqliteonline.com/>, as it allows you to import CSV files, query those tables, and export results back into CSV. If using this tool, make sure to set the "Column name" option to "First line" when importing and exporting.

Link to CSV Files -

<https://drive.google.com/file/d/1ZRvOvCZ9I8fCPOKAUTbz2rBMvI1PERAE/view?usp=sharing>

Python Assessment

There is a legacy system that does not recognize full country names, but instead accepts abbreviations that follow the rules below:

- Single-word country names
 - 1st letter, 2nd letter, and last letter, all capitalized
 - For example: **France** -> **FRE**
- Multiple-word country names
 - 1st letter of each separate word
 - For example: **United States of America** -> **USOA**

Using the `customer_dimension` table, please write a Python function that, given a country name string, returns a country abbreviation string that follows the above rules.

Then, please implement this function into a short Python program/script to create a new column “`country_abbreviation`” in your final dataset.

We suggest using the Python standard library **csv** as part of your program/script, but you are not required to do so. You may also use any other **Python standard** libraries.

This exercise is meant to give us a basic idea of your coding ability. As such, it is meant to be very straightforward, and you should **avoid using third-party libraries/packages** e.g. pandas, requests, etc.

When submitting your completed assessment, please include your Python code and .CSV output.