Congratulations!

We've been impressed with your background so far and would like you to tackle a realistic example of the type of service Neoway delivers to its customers.

Let's suppose that Neoway is working with a large vodka producing business, Bob's Vodka, which is looking to find new sales leads to pursue in New York state. Neoway has identified the New York State Liquor Authority website as a relevant data source to identify businesses with active liquor licenses - this is valuable intelligence to deliver to Bob's Vodka. Your mission is to complete the tasks outlined below to deliver this to Bob's Vodka for use in their sales efforts.

Open and download the spreadsheet linked below:

https://docs.google.com/spreadsheets/d/1kWzXgN2oy4ZrDC73JUKEL0Mg5iHkAg8Trib-Xur5Mjs/edit?usp=sharing

The "Liquor License" tab is the liquor license data, while the "Main database" tab is a sample of Neoway's existing business registry data.

To begin the task, you will need to set up a database, either locally or in the cloud (Example: AWS, Azure).

Using SQL, complete the following tasks:

1. Define a function (example: PL/pgsql on a postgres server) for cleansing names and addresses
2. Match liquor license rows with the corresponding Neoway row based on name, address, and zip code.
3. Refine your code to achieve the best matching rate and accuracy you can.
4. Create a new table of the matched rows with all columns from both tables

Final Deliverables:

1. Link to your GitHub repository.

    Store and document everything in GitHub, including the experience of setting up a database.

2. Report the matching rate of the final version of your matching code.
3. The resulting table of the matching process (see item 3 of the tasks above)

A few tips:

■ This may seem like a lot but work to complete as much as you can
■ For both completed tasks and uncompleted tasks (if any), we want to understand your logic and approach
■ If you need more time for any reason, let your Neoway contact know

You will be evaluated on:

● Utilizing development best practices (nomenclature, functions, etc.)
● Utilizing existing functions in SQL for maximum efficiency

- Organizational logic and documentation (readme, comments, etc.)
- Level of understanding of the data and the problem at hand
- Level of completeness of tasks

Non-required bonus ideas below. We also welcome other surprises you would like to showcase!

- Develop a bot to capture liquor license data, only for the zip codes contained in the sample file. The New York State Liquor Authority website is linked here:

  https://www.tran.sla.ny.gov/JSP/query/PublicQueryAdvanceSearchPage.jsp