Dear all,

First of all, I would like to thank you for the opportunity of doing this exercise. Some years have passed since I worked with coding and my familiarity with SQL was relatively small. Therefore, I faced the exercise as a reminder of how much I like coding and want to work with it going forward and as a good opportunity to study and improve in SQL.

I am deeply interested in the opportunity and hopeful that my results are satisfactory. I am certain that, if I have a chance to work with The Company, I will put a lot of effort to quickly catch up, overcome any challenge due my non-engineering background and, hopefully, contribute to team with my economics and consultancy experience.

**The exercise**

Please find below the main logical steps I performed in order to execute the task:

I started the task by looking at the two datasets with the objective of identify patterns. I quickly built some pivot tables to locate duplications and to get a sense of number of variables (cities, zip codes, etc). I have noticed the need for treating the data to help the match. For example, eliminating apostrophes and standardizing words like North and N, Avenue and Ave, etc.  I also noted several duplications of the same retailer for old and inactive licenses.

Based on this initial assessment, I implemented the following steps in SQL:

1. Imported the database using SQL Server Tools;
2. Created a stored procedure for cleansing names and addresses;
3. Created a stored procedure to remove duplicates by:
   1. Concatenating:
      * 10 first characters of retailer’s name;
      * 10 first characters of retailer’s address; and
      * Zip code
   2. Eliminating old inactive licenses:
      * The objective was to delete old inactive rows and just maintain in the database one line for each retailer;
      * Such line should be the one of the most recent license.
4. Created a function to match rows and to create a new table of the matched rows with all columns from the original tables
5. Exported the new table of matched rows to excel through SQL Server Tools.

**Results**

I founded 56 matched rows. After removing duplicates, the liquor license sheet contained 243 rows. Therefore, my matching rate is 23.0%

Best,

Renata Michalani

917-714-8779

renatamichalani@gmail.com