**Dear Sprocket Central’s team,**

Based on the 3 attached datasets sent early on in the day, there are many issues that need to be addressed before coming into phase two.

In the ‘Customer Demographic’ sheet, there are problems including:

* The ‘gender’ column consists of more than 2 genders and some genders are mistyped, e.g. fem for female, femal for female. Therefore, you need to give an unique abbreviation for each of the genders, for example: F/Female for female gender, and M/Male for male gender. If the business insists on having more than 2 genders, then we can register for more gender abbreviations for different genders.
* In the ‘job\_industry\_category’ column, there are n/a cells appeared within the columns. Instead of writing n/a, we could leave it blank as stand for new industry if that was what you intended. Secondly, when filtered, many ‘job\_title’ cells have their corresponded ‘job\_industry\_category’ classified as n/a and vice versa, many customers do not have a ‘job\_title’, but their corresponding ‘job\_industry\_category’ appear. In my assumption, this situation divides into 2 cases, it is either they have a job but the job is yet to have an official ‘job\_industry\_category’ name in the industry and in the database, or their data had been mistaken. Either ways, you need to recheck the columns.
* In the ‘default’ columns, the data contain unrecognizable values, this might cause it some issues when doing data analysis on your next phase, since this data is typed with special embedded code.
* In the ‘last\_name’ column, customer\_id 1787 has his ‘last\_name’ value typed as a URL link, we need to remove the URL link.

Move on to the ‘Customer Address’ table:

* The state column has 5 values instead of 3. The value ‘New South Wales’ should be replaced with ‘NSW’ which abbreviated for New South Wales and the value ‘Victoria’ should be replaced with ‘VIC’ instead.

With the ‘NewCustomerList’ table:

* As similar to ‘gender’ column in the CustomerDemographic table, each gender should correspond with an abbreviation to reduce complexity. It should have F/Female for female and M/Male for male, or more abbreviattions for genders if you preferred.
* The ‘DOB’ column need to have the same date format for all cells. For example, if a cell has a date of 25 March 2022 then the date format should be in 25 – 03 – 2022 or 25/03/2022 for all the values.
* There are similar issues regarding the ‘job\_industry\_category’ and ‘job\_title ‘ as mentioned above in the ‘CustomerDemographic’table.

We move on to the ‘Transactions’ table:

* There are blank sections in the online section. Since I assume there are two options only, True or False, there should be no third option available.
* On the transaction id number 8708, there is an incorrectly formatted value in the ‘standard\_cost’ column. Instead of writing 312.7350159 as in General Number format, we need to change it to the Currency format; as same as for the transaction id number 17469.
* The ‘product\_first\_sold\_date’ values are not written in date format but in General number format. You should fix this column.

After every problems have been addressed, you can go to the ‘NewCustomerList’ table and insert unique key values to each row of the table. This will help identify the customer distinctiveness. From then you can move on to your next phase of analysis.

Warm Regards,

*Quang Dong Nguyen*