Respectfully, Sprocket Central Pty Ltd.

We appreciate you sharing the Sprocket dataset with us. Following is a summary of the data quality issues with the dataset that we identified after reviewing it. We have also discussed how we have dealt with these data quality issues and laid out a strategy for continuing the data cleaning.

Sheet Name	Data Quality Issue
Transactions	Completeness & Relevancy
New Customer List	Completeness & Consistency
Customer Demographic	Completeness, Consistency & Relevancy

A few data quality issues with the Sprocket Central Pty Ltd dataset are shown in the table above. In order to prevent these data quality issues from reoccurring, we have taken the necessary steps to identify them and have provided the following recommendations.

In sheet name Transactions, we Identified null values for columns "online_order" and "brand". The column for "product_first_sold_date was converted into a date-time format.

- We found a number of null values in the columns mentioned above; it is crucial to remove these null values from the dataset as they raise questions about the dataset's completeness and may produce inaccurate modeling results.
- The "product_first_sold_date" column was organized into a date-time format that is simple to understand; however, this issue might occur when exporting data from a third party that converts date values to integers. However, they are difficult to interpret, so switching to date-time format makes the data easier to interpret.

Where we found null values on sheet name New Customer List, there were also inconsistent values for gender.

- As previously mentioned, the sheet's "last_name" column contained null values, but this is unimportant because we can simply use first names in its place. Because of this, it is not as significant, but there were still null values. The columns "job_title" and "job_industry" then contained more blanks and null values.
- There is inconsistency in the gender column, which is a categorical variable. that are misspellings of the gender female. as well as abbreviations in some rows. Columns M and F now stand for male and female, respectively. The column also contained the irrelevant variable "U," which was eliminated. However, it would be great if more detail could be provided on this; otherwise, the column cannot address it at this time.

In sheet name Customer Demographic which has inconsistency for gender, there were missing values and irrelevant field called "default"

- The gender field was changed to M and F, and the worksheet was handled similarly to the sheet for the new customer list. The value "U," which was unnecessary, was eliminated from the field.
- The fields "job title" and "job industry" no longer contain null values.
- A useless field called default was eliminated because it had no bearing on the data.

The team will continue cleaning and transforming data as needed for modeling in the future. Along the way, concerns will be expressed, and assumptions will be separately recorded. To make sure all of our assumptions are in line with the comprehension of Sprocket Central Pty Ltd, it would be great to spend some time with your data SME.

Kind regards.
Ramisa Farha
Junior Consultant