Dear Sprocket Central Pty Ltd,

Thank you for providing us with the dataset from Sprocket, we have reviewed the dataset and summarized the following data quality issue with the dataset. The plan to move forward with cleaning the data has been outlined in our comments regarding how we have dealt with these data quality issues.

Worksheet Name	Data Quality Issue
Transaction	Completeness & Relevancy
New Customer List	Completeness & Consistency
Customer Demographic	Completeness, Consistency & Relevancy

As can be seen in the table above, there are a few data quality issues associated with the Sprocket central Pty Ltd dataset. Below are recommendations to avoid these data quality issues from recurring in the future.

1. Worksheet name transaction where we identified blank values for columns "online_order" and "brand". The column for "product_first_sold_date" was converted into a date/time format.

a. Various blank values were identified in the columns mentioned above. Blank values should be removed from the dataset since they raise data quality issues and may cause inaccurate modeling results.

b. The column for "product_first_sold_date" was converted into a date/time format which is easy to interpret, this problem may arise when exporting data from third- party which may convert date value to integer, however they are not easy to interpret therefore changing it to date/time format makes it easier to interpret data.

2. The worksheet named New Customer List had blank values as well as inconsistent values for gender.

a. As mentioned above, blank values were discovered in the sheet for the column "second_name" however it is not an issue as we may only use first name instead therefore it is not as important, however there were still blank values. These were followed by more blank and null values in columns "job_title" and "industry".
b. The column for "gender" which is a categorical variable has inconsistency, there are spelling errors for female, and some rows had abbreviations. This was changed to the columns being M for male and F for female. The column also consisted of an irrelevant variable "U" which was

discarded from the column. However, if more clarity could be provided on this it would be great or else for now it is irrelevant for the column.

3. Worksheet name customer demographic which has inconsistency for gender, there were missing values and irrelevant field called "default"

a. This worksheet was dealt with in a similar way to the worksheet for new customer list, the field for gender was changed to M/F. "U" which was an irrelevant value in the field was removed from the field.

b. Null Values were removed from "job_title" and "job_industry" e. Irrelevant field called default was removed as it had no relationship to the data.

As part of the modeling process, the team will continue to clean and transform data. Questions will be raised along the way and assumptions will be documented separately. It would be great to spend some time with your data SME, to ensure all our assumptions are in line with the Sprocket Central Pty Ltd understanding.

Kind Regards

Ayush Kumar

Junior Data Consultant