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. We have further given our comments about how we have tackled these data quality issue and laid out a plan to move forward with the data cleaning.

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

The table above, outlines a few data quality issues with Sprocket central Pty Ltd dataset. We have taken relevant steps to identify these issues and given recommendation below to avoid these data quality issue from arising again.

1. Worksheet ‘Transaction’ where we identified blank values for columns “online\_orders” and “brand”. The column for “product\_first\_sold\_date” was converted into a date/time format which is easy to interpret.
2. Worksheet ‘NewCustomerList’ where we identified and removed blank values from below columns
   * 1. second\_name,
     2. job\_title
     3. industry
   1. Inconsistent values for gender and has spelling error for female which was corrected
      1. Male was replaced to M
      2. Female was replaced F
      3. Column also has ‘U’ which was discarded
         1. If more clarity could be provided on this it would be great or else for now it is irrelevant for column.
3. Worksheet ‘CustomerDemographic’ which had inconsistency for gender, there were missing values and
   1. The worksheet was dealt with similar way to the previous worksheet ‘NewCustomerList’, the field for gender was changed to M/F.
   2. Column also has ‘U’ which was discarded
      1. If more clarity could be provided on this it would be great or else for now it is irrelevant for column.
   3. Null Values from column ‘job\_title’ & ‘Job\_industry’ was removed.
   4. Irrelevant field called default was removed as ti had no relationship to the data.

Moving forward the team will continue the data cleaning and data transformation process for modelling. Questions will be raised along the way and assumption 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

Sandeep Kumar Rai

Data Analyst