Dear Sprocket Central Pty Ltd,

Thank you for providing with the dataset from Sprocket. I have gone throught the dataset and summarised the following data quality issue with the dataset. I have further described about how I 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 |
| NewCustomerList | Completeness & Consistency |
| CustomerDemographic | Completeness, Relevancy & Consistency |

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

1. **Transaction Worksheet :**

* **Removed blank values in columns “online\_order” and “brand”** for completeness of dataset and may lead to inaccurate results while building a model.
* **Converted** **“product\_first\_sold\_date” column into Date/Time format** because its easy to interpret. The problem may arise while exporting data from third-party which may convert date value to integer hence difficulty in getting proper output.

1. **NewCustomerList Worksheet :**

* **Removed blank values in columns “Last\_name”,”job\_title” and “job\_industry\_category”** for completeness of dataset and may lead to inaccurate results while building a model.
* **Corrected spellings and removed irrelevant “U” field from the column “Gender”.** There were spelling error for Female and some rows had abbreviations. So, It was changed to M for Male and F for Female. “U” field was discarded from the column to give more clarity as it didn’t seem relevant for building model.

1. **CustomerDemographic :**

* **Removed Null values from “job\_title” and “job\_industry\_category”.**
* **Irrelevant column called “Default” was removed** as it had no relationship to the data.
* This worksheet was similar with NewCustomerList worksheet**, the field for gender was changed to M/F. Value “U” int the column was removed** which was an irrelevant value in the field.

1. **CustomerAddress** seemed to have correct, relevant and consistent data so no changes made.

Moving forward, the team will continue with the data cleaning and data transformation process for modelling. 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,

Ankita Patnaik

Junior Data Consultant