Dear Sprocket Central Pty Ltd,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below table highlights the summary statistics from the three datasets received. Please let us know if the concepts are not aligned with your understanding.

|  |  |
| --- | --- |
| **Worksheet Name** | **Data Quality Issue** |
| Transactions | Completeness & Relevancy |
| New Customers List | Completeness, Relevancy & Consistency |
| Customer Demographics | Completeness, Relevancy & Consistency |
| Customer Address | Consistency |

In the next few words, I will detail the data quality issues in the form of points, and each point will contain an issue and the tables that contain these problems and some recommendations have been provided to avoid the reoccurrence of data quality issues and improve the accuracy of the underlying data used to drive business decisions.

* **Completeness:** In the data set there are many missing data that can greatly affect the process of data analysis and its results and increase the time of the data cleaning process, and this is shown in the “Transactions” table, “New Customer List” table and “Customer Demographics” table.
* **Relevancy:** Data types should be dealt with more accurately and made more uniqueness, it seems that there are some problems in databases where data types have been modified such as “product\_first\_sold\_date” column, I converted from “int” to “Date Format” and “Postcode” column I converted from “Text” to “Number Format”.
* **Consistency:** There are many data entry problems you also have, as there are many spelling errors and many of the values are the same, so you must check the data entry process or the company's program, as there is an unknown value in the "Gender" column and some spelling errors such as the "State" column where "Victoria" have the same value as 'Vick' and “New South Wales” have the same value as “NSW”, so work must be done to clarify the data values ​​more.

Moving forward, the team will continue with the data cleaning, standardization, and transformation process for the purpose of model analysis. Questions will be raised along the way and assumptions documented. After we have completed this, it would be great to spend some time with your data SME to ensure that all assumptions are aligned with Sprocket Central’s understanding.

Kind regards,

Talaat Mohamed Mohamed

Junior Data Analyst