Dear Consumer,

Thank you for providing us with the four datasets “Transactions”, “New Customer List”, “ Customer Demographic”, “Customer Address” from Sprocket Central Pty Ltd. The following write-up highlights the summary statistics from the four datasets received. Please let us know if the observation is not to your understanding. We found some noteworthy data quality issues in the above mentioned datasets and have actively mitigated the errors that could rise from it. They are as follows:

* Various columns, such as the brand of a purchase, job title, gender, etc., have empty values in certain records.

Solution: These missing records have been filtered out of the dataset.

* Inconsistent values for the same attribute.

(e.g. Victoria being represented as “V”, “Vic” and “Victoria”)

Solution: The data has been cleaned to retain uniformity in its values. Rows with ‘U’ in the gender section have been replaced from the training set on the basis of individual name.

* Inconsistent data type for the same attribute.

Solution: The numeric values are string for some fields and vice versa. They have been replaced and corrected as per the data representation.

* Incoherent fields.

Solution: The default field in “Customer demographic” has been removed owning to incomprehensible data input. Some numeric values like the date field have also been converted into a format that is readable by the user.

* Additional field inputs for better understanding and calculation.

Solution: Some extra fields will be added for better data visualization in the future modules.

Regards,

[Shreyanta Kar]