

Dear Kathleen,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below table highlights the data quality issue encountered in the provided dataset. Please let us know if you have any queries surrounding the issues presented.

	Accuracy	<u>Completeness</u>	<u>Consistency</u>	<u>Currency</u>	<u>Relevancy</u>	<u>Validity</u>	<u>Uniqueness</u>
Customer Demographic	<i>DOB</i> <i>Age missing</i>	<i>DOB</i> <i>Job Title</i>	<i>Gender</i>		<i>Default</i>	<i>Deceased</i>	
Customer Address		<i>Customer_id</i>	<i>State</i>				
Transaction Data	<i>Profit missing</i>	<i>Customer_id</i> <i>Online Order</i> <i>Brand</i>		<i>Product sold date format</i> <i>List price format</i>	<i>Cancelled order filtered out</i>		

Notable data quality issues that were encountered and the methods used to mitigate the identified data inconsistencies are as follows. Furthermore, 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.

- Customer Demographics

The Gender column contains inconsistent values. We have replaced all values of "M", "F", and "femal" with "Male" and "Female". The value of "U" has been left unchanged.

The Dob column contains some blank values and one outlier value. We have calculated the age of each customer based on their Dob value and formatted the age as a number. We have also removed the blank values from the Dob column.

The Job_title column contains some blank values. We have removed these blank values for simplicity.

The Default column has been deleted as it is not relevant to our analysis.

The Wealth_segment column does not have any issues.

The Deceased column has been deleted as it is not relevant to our analysis.

The Owns_car and tenure columns do not have any issues.

- Customer Address

The States column contains inconsistent values. We have replaced the values of "New South Wale" and "Victoria" with "NSW" and "VIC" respectively.

The Customer_id column contains some inconsistencies. We have corrected these inconsistencies.

- Transactions

The Customer_id column is missing the value of 5034 in the master sheet.

The Online_order column contains some blank values. We have removed these blank values.

The Brand column contains some blank values. We have removed these blank values.

The Product_first_sold_date column is not formatted as a date. We have formatted this column as a date.

The Profit column has been calculated as the difference between the List price and the Standard_deviation.

- New Customer

The Past_3 column should be a number. We have converted this column to a number.

The Dob column and the Job_ind column contain some blank values. We have removed these blank values.

The Property_value column is followed by multiple columns that are not relevant to our analysis. We have removed these columns.

The Rank column has been deleted.

The Value column has been formatted as a number.

Moving forward, the team will continue with the data cleaning, standardisation 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,
Arif Shaikh