

Hi There,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The summary table below highlights key quality issues that we discovered within given datasets. Please let us know if you any queries surrounding the issues presented.

	Number of Records	Distinct Customer_IDs	Date Data Received
Customer Demographic	4002	4002	13/11/2022
Customer Address	4001	4001	13/11/2022
Transactions_Dataset	20002	3495	13/11/2022

Data Quality Assessment parameter	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographics (Mitigation Techniques)	DOB: Inaccurate (Filtered) Age: Missing (Added)	Job title: Blank (Added) Customer id: Missing (Filtered)	Gender: Inconsistency (Added)	Deceased Customer: Present (Filtered Out)	Default Column: Unnecessarily (Deleted)	–
Customer Address (Mitigation techniques)	–	Customer id: Missing (Filtered)	States: Inconsistency (Added)	–	–	–
Transactions (Mitigation Techniques)	Profit: Missing (Added)	Customer id: Missing (Filtered) Online order: Blank (Added) Brand: Blank	–	–	Cancelled status order: Filter out	List price: Format Product sold date: format

Below are more in-depth descriptions of data quality issues discovered and methods of mitigation used. Recommendations and explanations have also been included to avoid further data quality issues in the future.

Accuracy Issues:

- DOB was inaccurate for “Customer Demographics” and missing an Age column. Missing Profit Column in “Transactions”
 - Mitigation: *Filter out an outlier in DOB*
 - Recommendation: *Creating Additional columns for Age and Profit will allow us for easier identification for errors. Profit columns will assist in further analysis.*

Completeness Issues:

- Additional Customer_Ids were inconsistent among “Customer Demographics,” “Customer Address,” and “Transactions”
 - Mitigation: *filter all Customer Ids From 1 to 3500*
 - Recommendations: *Ensure tables are up to date, with incomplete data the analysis results may be skewed.*
- Blanks in the job_titles for “Customer Demographics,” in online_order and brand_column for the “Transactions”
 - Mitigation: *Filter out blanks for job_title, online_order, and brand_column*
 - Recommendations: *Provide Dropdown for such columns. Blanks are treated as incomplete data and can be skewed further analysis.*

Consistency Issues:

- Inconsistency in Gender for “Customer Demographics” and “Customer Address”
 - Mitigation: *Filter all the M under ‘Male’ Category, All F and Femal under ‘Female’ Category. Filter all ‘New south wales’ under ‘NSW’ and ‘Victoria’ to ‘VIC’ for states.*
 - Recommendations: *Creating a dropdown option for such categorical data minimizes the manual entry and human errors.*

Currency Issues:

- People that are ‘Y’ in deceased_indicator are not current customer for “Customer Demographics”
 - Mitigation: *Filter out Customer Checked as ‘Y’ in deceased indicator*
 - Recommendations: *Making record up to date.*

Relevancy:

- Lack of relevancy in default_Column for “Customer demographic” and order_status for “Transactions”
 - Mitigation: *Deleted the Default_column. Filter out “cancelled order_status.”*
 - Recommendation: *Check for incompressible metadata and delete to make comprehensible.*

Validity:

- Format of list_price, product_sale_date for “Transactions”
 - *Mitigation: format product_sale_date to short date format, format list_price to currency.*
 - *Recommendations: Set up columns so that formats such as price and decimals are already in place. Allowable data will make data to be interpreted more easily.*

That summarises all the data quality issues discovered through the first stage of the data quality analysis.

Please let us know if you have questions regarding mitigation or any data quality issues.

Kind regards,

Mohit Kamble

Data Analyst