

Dear (Client to contact),

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

Summary Table:

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Transactions	Profit: missing	Customer id: incomplete Online order: blank Brand: blank			Cancelled status order: filter out	List price: format Product sold data: format
Customer Address		Customer id: incomplete	States: inconsistency			
Customer Demographic	DOB: inaccurate Age: missing	Job title: Blank Customer id: incomplete	Gender: inconsistency	Deceased customer: filter out	Default column: delete	

Below are more in depth descriptions of data quality issues discovered and method of mitigation used. Recommendations and explanations are provided to avoid future data quality issues. Following recommendations will improve accuracy of data used to influence business decisions of Sprocket Central Pty Ltd in future.

Accuracy Issues

- **Missing a profit column in “Transactions”; DOB is inaccurate in “Customer Demographic” and missing Age column.**
Mitigation: Filter out outlier in DOB
Recommendations: Create profit column in “Transactions” to check accuracy of sales.
Create Age column to check errors more accurately in “Customer Demographic”.

Completeness Issues

- **Additional customer ids were inconsistent in “Transaction”, “Customer Address” and “Customer Demographic”.**
Mitigation: Filter all customer ids
Recommendations: Ensure tables are up to date. For our model, only customer ids which have complete information will be used.

The data received may not be in sync across all spreadsheets, with incomplete data, the analysis results maybe skyed. This is a “completeness” issue, to prevent future occurrences, it is encouraged to cross check the spreadsheets and sync data.

- **Blanks in online_order and Brand column in “Transactions”; blanks in job_titles in “Customer Demographic”**
Mitigation: Filter out blanks from online_order, brand and job_titles
Recommendations: Provide dropdown options for online_order and brand column. Provide dropdown option for job_title.

Consistency Issues

- **Inconsistency in gender for “Customer Address” and “Customer Demographic” respectively.**

Mitigation: Filter all ‘M’ under ‘Male’ and all ‘Femal’ and ‘F’ under ‘Female’ for gender. Filter all ‘New South Whale’ to ‘NSW’ and ‘Victoria’ to ‘VIC’ for states.

Recommendations: Create dropdown options for ‘Male’, ‘Female’ and ‘U’ for gender. Create dropdown option for all state abbreviations.

Currency

- **People that are ‘Y’ in deceased_indicator for “Customer Demographic” are not current customers.**

Mitigation: filter out customers which are checked as ‘Y’ in “Customer Demographic”

Recommendations: Once this information is received, one must update data accordingly.

Deceased customers are not current customers, removing them from data will increase currency of data and will result in more accurate estimates in future analysis.

Relevancy

- **Lack of relevancy in default column for “Customer Demographic” and order status for “Transactions”**

Mitigation: Filter out cancelled orders_status. Deleted metadata in default_column

Recommendations: Check for incomprehensible Metadata and delete or format to make comprehensible.

“Cancelled” order_status is irrelevant information for future analysis as it can skew data – for example total number of customers per annum will be an overestimate.

Validity

- **Format of list_price; product_sold_date for “Transaction”**

Mitigations: Format product_sold_date to short data format; format list_price to currency

Recommendations: Set up columns so that formats such as price and decimals are already in place when entering new data.

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,

Aarti Nagpure