

Dear Client,

Thank you for providing us with three datasets form Sprocket Central Pty Ltd. The following table highlights the key quality issues in that were discovered in the three datasets. Kindly let us know if you have any queries regarding issues presented.

### **Summary Table:**

Sheet	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	<ul style="list-style-type: none"><li>• DOB: inaccurate</li><li>• Age: missing</li></ul>	<ul style="list-style-type: none"><li>• Customer_id: incomplete</li><li>• Job_title: nulls</li><li>• DOB: nulls</li></ul>	<ul style="list-style-type: none"><li>• Gender: inconsistency</li><li>• Tenure: format</li></ul>	<ul style="list-style-type: none"><li>• Deceased_indicator: filter out</li></ul>	<ul style="list-style-type: none"><li>• Default column: delete</li></ul>	
Customer Address		<ul style="list-style-type: none"><li>• Customer_id: incomplete</li></ul>	<ul style="list-style-type: none"><li>• States: inconsistency</li><li>• Customer_id: format</li><li>• Postcode: format</li><li>• Property valuation: format</li></ul>			
Transactions	<ul style="list-style-type: none"><li>• Profit: missing</li></ul>	<ul style="list-style-type: none"><li>• Product_id: incomplete</li><li>• Online_order: nulls</li><li>• Brand: nulls</li></ul>	<ul style="list-style-type: none"><li>• Transaction id: format</li><li>• Product id: format</li><li>• Customer id: format</li></ul>		<ul style="list-style-type: none"><li>• Order_status: filter out</li></ul>	<ul style="list-style-type: none"><li>• Transaction_date: format</li><li>• List_price: format</li><li>• Standard_price: format</li><li>• Product_first_sold: format</li></ul>

More in-depth descriptions of data quality issues discovered and methods of mitigation used are mentioned below. Recommendations and explanations have been included to avoid further data quality issues in future. Following recommendations will improve accuracy of data used to influence business decisions of Sprocket Central Pty Ltd in the future.

### **Accuracy Issues:**

***DOB was inaccurate for “Customer Demographic”, DOB was having null values and an age column was missing; profit column was missing for “Transactions”***

Mitigation: Filter out outlier and remove null values in DOB.

Recommendation: Create an age column in “Customer Demographic” for more comprehensive view and it would be easier to check for errors. Create a profit column in “Transactions” to check accuracy of sales and will assist for future monetary analysis.

## **Completeness:**

***Customer\_id was incomplete, Job\_title and DOB columns were having null values in “Customer Demographics”; customer\_id was incomplete in “Customer address”; product\_id was incomplete, online\_order and brand columns were having null values.***

Mitigation: Remove incomplete and null values in the mentioned columns across the three spreadsheets.

Recommendation: Ensure tables are up to date not having null values or incomplete data across all spreadsheets since it will cause skewed results in our analysis.

## **Consistency:**

- ***Gender was having inconsistent values in “Customer Demographics”; state column was having inconsistent values in “Customer Address”.***

Mitigation: filter all ‘M’ under Male, filter all ‘F’ under Female, filter all ‘Femal’ under Female, filter all ‘U’ under Unspecified for gender column for “Customer Demographics”. Filter all ‘New South Wales’ under NSW, filter all ‘Victoria’ under VIC in state column for “Customer Address”

Recommendation: For entering the data in spreadsheet create a dropdown menu for gender and state columns in order to reduce human error and increase consistency of terminology.

- ***Tenure for “Customer Demographics”; customer\_id, postcode, property\_valuation for “Customer Address”; transaction\_id, product\_id, customer\_id for “Transactions” are having inconsistent format.***

Mitigation: these columns are having values in general format, convert them to numeric format in order to make these consistent.

Recommendation: setup columns such that format like numeric and number of decimal places are already in place when entering new data.

## **Currency:**

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

Mitigation: Filter out customers checked ‘Y’ in deceased\_indicator.

Recommendation: Ensure to deal with up-to-date data by considering the current customers. Removing the deceased customers will increase currency of data and will lead to more accurate estimates in future analysis.

## **Relevancy:**

***Lack of relevancy or comprehensibility in default column for “Customer Demographic” and order\_status for “Transactions”.***

Mitigation: Delete the default column having metadata, filter out cancelled order\_status.

Recommendation: check for incomprehensible metadata and delete or format to make it comprehensive. The cancelled order\_status is irrelevant information will skew our estimations for future analysis.

## **Validity:**

***Format of transaction\_date, list\_price, standard\_price, product\_first\_sold for “Transactions”.***

Mitigation: Format transaction\_date and product\_first\_sold to short date. Format list\_price and standard\_price to currency.

Recommendation: setup columns such that formats like price and dates are already in place when entering new data.

That summarizes the data quality issues discovered through first stage of data quality analysis. The mitigation strategies suggested are simple and effective ways of improving data quality for future analysis. These will not only improve the output of analysis that one could perform within the company but will increase the level of analysis that can be performed by KPMG and other hired analysis teams.

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

Kind Regards,

Haseeb Ahmad