

Dear Jhon ,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below table highlights the summary statistics from the three datasets received. Please let us know if the figures are not aligned with your understanding.

## Sammary table

Table name	Accuracy	Completeness	Consistency	Currency	Relevancy
Customer Demographic	<ul style="list-style-type: none"><li>• DOB: Inaccurate</li><li>• Age: missing</li></ul>	<ul style="list-style-type: none"><li>• Job titles: Blanks</li><li>• Customer id: incomplete</li></ul>	<ul style="list-style-type: none"><li>• Gender: inconsistent</li></ul>	<ul style="list-style-type: none"><li>• Deceased customers: filtered out</li></ul>	<ul style="list-style-type: none"><li>• Default Deleted</li></ul>
Customer Address		<ul style="list-style-type: none"><li>• Customer id: incomplete</li></ul>	<ul style="list-style-type: none"><li>• State: inconsistence</li></ul>		
Transaction Data	<ul style="list-style-type: none"><li>• Profit: missing</li></ul>	<ul style="list-style-type: none"><li>• Customer id: incomplete</li><li>• Online Orders: blanks</li><li>• Brand: blanks</li></ul>			<ul style="list-style-type: none"><li>• Cnceled status order: Filter out</li></ul>

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. 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" and missing an age\_column;

missing a profit column for "Transactions"

Mitigation: Filter out outlier in DOB.

Recommendation: Create an age\_column, allowing for more comprehensible data and easier to check for errors. Create a profit column in "Transactions" to check accuracy of sales.

Creating additional columns for age and profit will allow for easier identification of errors. The

profit column will assist in future monetary analysis.

#### Completeness

- ✓ Additional customer\_ids were inconsistent among "Customer Demographic,"

"Customer Address," and "Transactions"

Mitigation: Filter all customer\_ids from 1 to 3500

Recommendation: Ensure tables are up to date (from the same time period). For our model, only customer\_ids from 1 to 3500 will be used as they have complete data.

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

- ✓ Blanks in job\_title for "Customer Demographic," "in online order and

brand\_column for "Transactions"

Mitigation: Filter out 'blanks' for job\_title, online\_order, and brand\_column.

Recommendation: Simplify job\_title to another category such as industry\_industry or provide dropdown options for job\_title. Provide dropdown options for online order and brand\_column.

Blanks are treated as incomplete data and can skew further analysis results. The addition of dropdown options will allow to have more complete data and will result in more accurate analysis.

- **Consistency**

Inconsistency in gender for "Customer Demographic" and "Customer Address" respectively

Mitigation: Filter all 'M' under category of 'Male,' filter all 'Femal' and 'F' under 'Female' for gender. Filter all New South Wales to 'NSW' and 'Victoria' to 'VIC' for states.

Recommendation: Create dropdown options for 'Male,' 'Femal,' and 'U' in gender. Create dropdown options for all state abbreviations.

Dropdown options, minimizes manual entry and human error. Allows for increase of consistency of terminology. Gender identity can be a sensitive topic, proceed with caution when creating options.

- **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: Can be difficult to check for deceased customers, but once this information is received one should 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 or comprehensibility in default\_column for "Customer Demographic" and order\_status for "Transactions"

Mitigation: Deleted Metadata in default\_column. Filter out 'Cancelled' order\_status.

Recommendation: 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\_sale\_date for "Transactions"

Mitigation: Format product\_sale\_date to short date format, format list price to currency.

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

Allowable values will make data to be interpreted more easily. Formatting into price and allowing for either 2 or 3 decimals placed consistently will increase readability. This will reflect positively on speed

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

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

Kind regards,

AISSAME BOUKHALLAD