

To Whom It May Concern:

Thank you for providing us with data for Sprocket Central Pty Ltd and giving us the opportunity to provide a data quality assessment. Below is a summary of our team's assessment according to our Standard Data Quality Dimensions which we used evaluate the provided datasets.

Dataset	No. of Records	Distinct Customer IDs	Date Data Received
Customer Demographic	4000	4000	August 10, 2021
Customer Address	3999	3999	August 10, 2021
Transaction Data	20000	3494	August 10, 2021

	Customer Demographic	Customer Address	Transactions
Accuracy	<ul style="list-style-type: none">- Inaccurate DOB- Age column recommended		<ul style="list-style-type: none">- Profit column recommended
Completeness	<p>Columns with blanks:</p> <ul style="list-style-type: none">- Last Name (125)- DOB (87)- Job Title (506)- Job Industry Category (656)- Tenure (87) <p>(#) means amount of missing values in the respective column</p>		<p>Columns with blanks:</p> <ul style="list-style-type: none">- Online Order (360)- Brand (197)- Product Line (197)- Product Class (197)- Product Size (197)- Standard Cost (197)- Product First Sold Date (197) <p>(#) means amount of missing values in the respective column</p>
Consistency	<ul style="list-style-type: none">- Customer ID across all Datasets inconsistent- Gender Column inconsistencies	<ul style="list-style-type: none">- Customer ID across all Datasets inconsistent- States inconsistencies	<ul style="list-style-type: none">- Customer ID across all Datasets inconsistent
Currency	<ul style="list-style-type: none">- Deceased customers are included		

Relevancy	- Default Column can be deleted		- Cancelled orders can be removed
Validity			- List price does not mention currency - Product Sold Date are not valid dates
Uniqueness	- No duplicates found	- No duplicates found	- No duplicates found

Below are detailed descriptions of the issues discovered in our analyses. Included are our recommendations at resolving these issues as well as strategies to avoid these issues in the future. Suggested strategies will improve the accuracy of data used to influence business decisions of Sprocket Central Pty Ltd.

ACCURACY

Customer Demographic Dataset:

- The max age was found to be 177 using the given DOB. An age column would help filter out errors.

Transactions Dataset:

- Adding a profit column will help with future monetary analysis.

COMPLETENESS:

Customer Demographic Dataset:

- The following columns had blank values, Last Name (125), DOB (87), Job Title (506), Job Industry Category (656), and Tenure (87).

Transactions Dataset:

- The following columns had blank values, Online Order (360), Brand (197), Product Line (197), Product Class (197), Product Size (197), Standard Cost (197), Product First Sold Date (197)

For our analyses we believe the best solution is to filter out records containing missing data. This is also due to inconsistencies seen with the customer ID column. Regularly updating the datasets will help with having complete data.

CONSISTENCY:

Customer Demographic Dataset:

- The gender field did not have a uniform method for classifying gender. Some used an acronym for Male and Female while others has misspellings of male and female.

Customer Address:

- The states field did not have a uniform method for classifying state. Some used an acronym for example NSW compared to other cases with New South Wales.

Customer IDs were inconsistent across datasets as they may not have been in sync across the datasets. Having a uniform naming system will result in greater consistency. If customers are responsible for typing in information a dropdown menu with options will help with consistency.

CURRENCY:

Customer Demographic Dataset:

- Deceased customers are included in the dataset. Removal of deceased customers will help with the analyses of current customers.

RELEVANCY:

Customer Demographic Dataset:

- Default field was incomprehensible.

Transactions Dataset:

- Cancelled orders can be removed as they could lead to incorrect analyses for current profit.

Ensure that data is formatted correctly, and irrelevant information is removed.

Validity:

Transactions Dataset:

- List price was not in a form of currency. Product sale date was also in an invalid date format.

Monetary values in their given currency can be compared effectively to other prices. Ensure that data is formatted correctly.

These were all the issues we discovered through the first stage of our analysis. Please let us know if you have any questions.

Sincerely,
Sunny Patel