Subject: Data Quality Issues in the Datasets for Sprocket Central Pty Ltd

Dear Associate Director,

I trust this email finds you in good health. I have conducted a thorough review of the three datasets that were recently provided for data quality assessment. Below, I have outlined the data issues that require immediate attention to prevent inconsistencies and errors in future data analyses:

Data Accuracy:

In the 'Customer Demographic' dataset, there is an entry indicating a customer's age as 174 years, which appears to be inaccurate.

The same dataset contains a 'default' column with inconsistent and erroneous data that needs correction.

Data Completeness:

The following columns in each dataset exhibit missing or incorrect data:

Transaction File: 'online\_order,' 'brand\_column,' 'product\_column,' 'product\_class,' 'product\_size,' 'standard\_cost,' 'product\_first\_sold\_date,' and all date columns.

New Customer List: 'gender\_column' (consider renaming 'U' to 'Unknown'), 'job\_title' (missing values), 'property\_valuation' (consider correcting values 6.00, 7.00, and 8.00 to 6, 7, and 8), and 'value\_columns' (round all numbers to two decimal places for consistency).

Customer Demographic: 'last\_name' column with missing/blank values, mixed values for 'gender\_categories' (Male/Female and M/F), and a 'default' column filled with missing and broken data.

Data Consistency: Some tables have inconsistent data types. Consider implementing data validation for columns such as 'gender,' 'product\_line,' 'product\_class,' and 'product\_size.'

Data Currency: The datasets cover transactions from October to December 2017, indicating no issues with data currency.

Data Validity: An example of data validity issues is a record in the 'Customer Demographic' dataset, which suggests a customer is 174 years old.

Data Uniqueness: No duplicate data was found within the datasets, ensuring data uniqueness.

To mitigate future errors, I recommend the following actions:

* Conduct regular data audits to remove invalid records and ensure data integrity.

Instead of leaving data fields blank, consider populating them with 'unknown' or 'n/a' for clarity when others view the data.

* Establish clear data value guidelines.
* Periodically verify data types in the dataset columns.

Should you have any questions or require further clarification, please do not hesitate to reach out. Your prompt attention to these data quality issues is greatly appreciated. Wishing you a productive day.

Sincerely,

Caroline Natasia Cahyadi