Dear Client,

I hope this email finds you well. I have thoroughly reviewed the datasets provided by Sprocket Central Pty Ltd and would like to share my findings and recommendations regarding data quality for our upcoming analysis in phase two.

\*\*Transaction Sheet:\*\*

- The Product ID column ranges from 0 to 100, which seems inconsistent. Please clarify the intended range.

- The "Online Order" column contains empty cells.

- Empty cells are observed in the "Brand," "Product Line," "Product Class," and "Product Size" columns.

- The "Product First Sold Date" column has data-type inconsistencies and contains empty cells, particularly where related columns such as the "Brand," "Product Line," "Product Class," and "Product Size" also have missing.

- The "List Price" column's data-type should be standardized to reflect currency accurately.

\*\*Customer Demographic Sheet:\*\*

- Empty cells are found in the "Last Name" column.

- The "Gender" column has inconsistent values such as "femal," "F," and "M."

- The "DOB" column contains blank cells and a data entry error with a year of 1843.

- "Job Industry Category" has null values.

- The "Property Valuation" column should be reformatted to numeric.

- The purpose of the "Default" column is unclear; it contains errors, abnormalities, and blank cells.

- The "Tenure" column contains blank cells.

\*\*Customer Address Sheet:\*\*

- The "State" column has disparities, including variations like "VIC" and "Victoria," "NSW" and "New South Wales." Standardization is needed.

Moreover, there is a discrepancy in the customer count between the "Customer Demographic" sheet (4000) and the "Customer Address" sheet (4003). The state column in the "Customer Address" sheet needs standardization to ensure consistency.

To address these data quality concerns, I recommend implementing the following strategies:

- \*\*Accuracy:\*\* Review data entry procedures to prevent errors and inconsistencies.

- \*\*Completeness:\*\* Ensure all required fields are consistently filled, and address the issue of missing data.

- \*\*Consistency:\*\* Standardize values across relevant columns (e.g., "State" column) for consistency.

- \*\*Currency:\*\* Correct data-type inconsistencies, particularly for currency-related columns.

- \*\*Relevancy:\*\* Verify the purpose and necessity of the "Value" and "Default" columns.

- \*\*Validity:\*\* Validate data entries against predefined criteria to remove errors.

- \*\*Uniqueness:\*\* Identify and rectify duplicate entries across datasets.

If you require further clarification or would like to discuss these findings and recommendations, please feel free to reach out. I'm committed to ensuring that the data is prepared to meet the highest standards for our analysis.

Thank you for your attention to this matter.

Best regards,

Jamiu Arogundade.