

TASK 1: Data Quality Assessment and Recommendations

Dear Sprocket Central Pty Ltd

I hope this message finds you well. We would like to express our gratitude for entrusting us with your data for analysis. After a thorough review of the datasets you provided—Customer Demographic, Customer Addresses, and Transaction data for the past three months—we have identified several data quality issues that may impact the accuracy and reliability of our analysis. Below, we outline these issues and provide recommendations to mitigate them, all in accordance with our Data Quality Framework Table.

Data Quality Issues:

Completeness: We observed missing values in key fields across all datasets. This incompleteness may lead to biased or incomplete insights. To address this, we recommend implementing data validation checks during data entry and regularly auditing data for completeness.

Accuracy: Inconsistencies and inaccuracies were detected in the Customer Demographic dataset, such as inconsistent date formats and incorrect gender entries. We suggest implementing data validation rules to ensure data accuracy during collection and updating.

Consistency: The Transaction data exhibited inconsistencies in product names and codes, which could affect our analysis. We recommend establishing data standards and enforcing data consistency across your systems.

Timestamp Missing: The Transaction data lacks a clear timestamp, making it challenging to ascertain the recency of the information. Incorporating a timestamp for each transaction would improve the data's currency and enhance its usability.

Relevance: Some fields in the datasets may not be directly relevant to the analysis goals. We recommend a thorough review of the data attributes to determine their relevance, potentially removing redundant or unnecessary fields.

Validity: We also noticed instances where data values in certain fields do not conform to expected formats or ranges. To improve data validity, consider implementing data validation rules to restrict the entry of incorrect or out-of-range values.

Uniqueness: Duplicate records were found in all three datasets. Identifying and removing duplicate records will ensure the uniqueness of data entries, reducing the risk of overestimating the importance of specific data points.

Recommendations:

Data Validation Rules: Develop and implement data validation rules at the data entry point to ensure completeness, accuracy, and validity.

Data Standardization: Establish standardized data formats, units, and codes to ensure consistency.

Timestamps: Incorporate timestamp attributes in the Transaction data to indicate when each transaction occurred.

Data Governance: Implement data governance practices to oversee data quality, including regular audits, validation checks, and the removal of duplicate records.

Data Relevance Review: Periodically assess the relevance of data attributes, removing redundant or obsolete fields.

We believe that addressing these data quality issues and implementing the recommended strategies will enhance the reliability and usability of your data for our analysis in phase two. Our team is committed to assisting you in achieving the highest data quality standards.

Please feel free to reach out if you have any questions or require further clarification regarding these findings and recommendations. We are excited to continue working together to unlock valuable insights from your data.

Thank you for your trust in our team..

Warm regards,

Onome Erigha.

Data Analyst

09160029243