

Subject: Data Quality Review for Phase Two Analysis - Sprocket Central Pty Ltd Data

Dear [Client Name],

Thank you for providing the customer demographic, customer address, and transaction data for the past three months. We've conducted a thorough review of the data quality as per your request to ensure it's ready for in-depth analysis in phase two.

Data Quality Framework Analysis

We evaluated the data using the "Data Quality Framework Table" you provided, focusing on key dimensions that impact the usability of the information for our analysis. Here's a breakdown of our findings:

Accuracy: We identified potential inconsistencies in data points such as dates, formats for numbers, addresses, or gender. Ensuring data accuracy is crucial for reliable analysis, and we may need to discuss data validation techniques to mitigate these issues.

Completeness: There might be missing data points in fields. Incomplete data can lead to skewed results, and we recommend exploring data imputation techniques like mean/median imputation, k-Nearest Neighbors to address these gaps.

Consistency: We observed variations in how data is formatted, for example, addresses might use different abbreviations for state names (e.g., CA vs. California) or dates might be entered in different formats (e.g., DD/MM/YYYY vs. YYYY-MM-DD). Inconsistent formatting can hinder analysis, and we suggest implementing standard data formats for all fields to ensure consistency.

Currency: While the transaction data covers the past three months, it would be beneficial to understand if there's historical data available. Having a more comprehensive timeframe can provide richer context for analysis. Please let us know if historical data can be incorporated.

Relevancy: We'll need to confirm that the data provided aligns with the specific goals of our analysis in phase two. A discussion to ensure the data reflects the key metrics we need to examine is crucial.

Validity: We will verify that the data adheres to established data quality standards and expectations. This might involve checking for outliers or values outside the expected range for specific fields. Data validation processes can help ensure the data is within acceptable parameters.

Uniqueness: There might be duplicate records within or across datasets, potentially inflating results. Identifying and removing duplicate records will ensure data integrity and prevent skewed analysis.

Next Steps and Recommendations

To ensure a smooth transition to phase two and achieve the best possible results, we recommend a collaborative discussion to address the following:

Clarify any assumptions we made during the data review.

Discuss the feasibility of implementing the suggested data improvement strategies, such as data cleaning, imputation, and standardization.

Determine the best approach for handling missing data and duplicate records.

Confirm the availability of historical data, if applicable.

Refine the scope of the analysis to ensure the data aligns with our specific goals (data relevancy).

Establish data validation processes to maintain data quality throughout the analysis.

Please let us know your preferred method for scheduling a call to discuss these findings and next steps. We're confident that by working together, we can ensure the data is high-quality and ready to yield valuable insights in phase two.

Thank you for your cooperation.

Sincerely,

Junior Consultant, Angel May