

To Whom it May Concern:

Thank you for providing the datasets from Sprocket Central Pty Ltd for review. I have thoroughly examined the data quality, and I am pleased to share my findings, assumptions, issues, and recommendations with you.

Customer Demographic:

Blank Cells: I noticed the presence of blank cells in some columns of the Customer Demographic dataset. To address this, we recommend considering imputation techniques such as mean, median, or mode to fill in the missing values. However, if there is a systematic pattern or a significant number of missing values, we may need further clarification from you or consider excluding those records from our analysis.

Special Characters: Certain columns in the Customer Demographic dataset contain special characters. These characters might indicate data entry errors or inconsistencies. It would be helpful to identify the specific special characters and determine their significance to the analysis. Depending on the context, we suggest removing or replacing these special characters to ensure accurate and consistent data.

First and Last Name in Separate Columns: The dataset shows that the first and last names of customers are stored in separate columns. To enhance data usability and streamline analysis, we recommend combining these columns into a single column for the customer's full name.

Customer Addresses:

Data Entry Errors: We noticed discrepancies and potential data entry errors in the customer addresses dataset, such as misspelled street names, incorrect postal codes, and inconsistent formatting. To address these issues, we recommend performing data cleansing and standardization procedures to ensure accurate and consistent address information.

Transaction Data in the Past Three Months:

Decimals on Costs: The cost values in the transaction data contain decimals. It would be helpful to clarify whether these decimal values represent fractions of a currency unit or if they are data entry errors. If they are fractions, we suggest keeping the decimal values. However, if they are errors, we recommend rounding the values to the appropriate decimal places or removing the decimals altogether.

Wrong Age: We found instances where the recorded age for some customers is incorrect. To address this discrepancy, we recommend investigating the source of the error and determining the appropriate age values. Depending on the circumstances, we may need your guidance or consider using other demographic variables or techniques (e.g., date of birth) to estimate the customers' correct ages.

Based on these findings, we recommend the following actions to mitigate the current data quality concerns:

Consultation: We kindly request your clarification on any ambiguities, guidance on addressing missing values, and discussion of the identified data quality issues.

Data Cleaning and Standardization: We will apply data cleaning techniques to address missing values, remove or replace special characters, and rectify data entry errors in the datasets. Additionally, we will standardize address formats to ensure consistency.

Feature Engineering: As part of our analysis, we propose combining the first and last name columns in the Customer Demographic dataset to create a single column for full names, enhancing data usability and analysis convenience.

Data Validation and Verification: To ensure data accuracy and resolve any inconsistencies or discrepancies, we recommend cross validating the datasets with external sources or consulting with you to verify the accuracy of the data.

We will proceed with these recommendations upon receiving your confirmation and further guidance. If you have any additional questions or specific areas you would like us to focus on, please let us know.

Thank you for your attention to these matters, and we look forward to your response.

Best regards,

Uche Charles Oji

Data Analyst

KPMG Consulting