

Dear Customer Sprocket Central Pty Ltd,

I hope this email finds you well. I have reviewed the three data sets provided by Sprocket Central Pty Ltd: customer demographics, customer addresses, and transaction data for the past three months. After assessing the quality of the data using our Data Quality Framework, I have identified several data quality issues that may affect our analysis going forward.

In the Transactions dataset, there are missing values in some columns and the product_first_sold_date column has float64 data but should be a date. In the Customer Demographic dataset, there are missing values in some columns, the gender column has inconsistent values and the default column has "metadata" data. Also, the DOB column is missing too many values, making it difficult to analyze the age of customers. Also, the deceased_indicator column has 2 "Y" values indicating that the customer is deceased. In the Customer Address dataset, the status column has inconsistent values.

We recommend applying data validation rules to ensure that future data entries are accurate and consistent.

To mitigate these issues, we will clean up the underlying data by filling in missing values, correcting inconsistent values, and converting the data to the correct format, as well as removing what is necessary to have complete and consistent information.

If you have any questions or concerns, please contact us. We will improve the quality of your data to avoid biases and be able to perform a good analysis of the data to find business-relevant insights.

Best regards,
Alejandro Thompson, Data Analyst KPMG

Dataset	Accuracy	Completeness	Consistency	Relevancy	Uniqueness	Validity	Currency
Transactions	Yes	No, some columns have missing values.	Yes	Yes	Yes, no duplicates.	No, the product_first_sold_date column has float64 data but should be a date.	Yes
Customer Demographic	No, the gender column has inconsistent values.	No, some columns have missing values.	No, the default column has “metadata” data.	No, the deceased_indicator column is not relevant	Yes, no duplicates.	No, the deceased_indicator column has 2 “Y” values indicating that the customer is deceased.	No
Customer Address	Yes	Yes, all columns have non-null values.	No, the state column has inconsistent values.	No, the country column is not relevant, because we only work in Australia for now	Yes, no duplicates.	Yes	Yes