

Dear Client,

This is Salomi Mpazi from the KPMG Data Analytics (Virtual Internship) Team. We are grateful for the datasets you provided. However, during our review of all datasets, we identified and explained certain problems with data quality.

- **Accuracy issues**
 - *Customer demographics table* - in the “DOB” column there is a DOB of 1843-12-21 which is inaccurate and the data type is inconsistent.
- **Completeness issues**
 - *Customer demographics table* - There are blank fields in the “Last name” column, and “Job_title” column.
 - *Transaction table* - There are blank fields in the “online_order”, “brand”, “product_line”, “product_size”, “standard_cost”, “product_first_sold_date” column.
 - Additional Customer IDs in the Customer Address and Transaction table are inconsistent with Customer demographics customer ID’s. This suggests that the data received might not be in sync with one another.
- **Consistency issues**
 - *Customer demographics table* - “Gender” column, “Female” being represented as “F” and “Femal”
 - *Customer Address table* - “State” column, “Victoria” is represented as “VIC”
- **Currency issues**
 - *Customer demographics table* - “deceased” column has “Y” values which shows some customers are not current.
- **Relevancy issues**
 - *Customer demographics table* - “Past_3_years_bike_related_purchases” and “Default” columns are irrelevant to this table
 -
- **Validity issues**
 - *Transaction table* - The format is inconsistent in the “list_price”, “standard_cost”, and “product_first_sold_date” columns.

Additionally, we provided a few suggestions for eliminating these issues as well as improving the accuracy of the underlying data that serves as the basis for business decisions.

- Ensure data types are defined before capturing data in columns.
- Ensure that all tables are from the same period.
- Ensure the core fields are assigned values use a defined list of values to fill in missing values, if they are only a small number of rows that are optional, filter them out.
- Use option sets/dropdowns for fields such as “Gender” so that the users cannot make any mistakes.

We are moving forward with further processes like data cleaning and standardization.

Thank you,

Salomi Mpazi