

Dear Client,

We would like to thank you for providing us the datasets from Sprocket Central Pty Ltd. The below table contains statistical details from all the datasets.

Table Name	No. of Records	Unique Customer IDs	Date Data Received
Transactions	20000	3495	02-07-2023
Customer Demographic	4000	3999	02-07-2023
Customer Address	4000	3999	02-07-2023

The datasets contain several data quality issues that can be rectified effectively. Here are the issues found in the datasets:

- 1) Duplicate customer ID present in the Customer Address table.
- 2) Inconsistent values with the same attributes, such as different representations for the same values.
 - a) New South Wales = NSW
 - b) Victoria = VIC
 - c) Female = Femal = F
 - d) Male = M
- 3) Unwanted string in default row of customer demographic table
- 4) Inconsistent data types for the same attribute, where numerical values are replaced with string values.
- 5) Rows with empty values in certain records for attributes such as job_title, DOB, Tenure, job_industry_category, product_first_sold_date, standard_cost, product_size, product_class, product_line, brand, and online_order.
- 6) Duplicate customer IDs and a transaction ID in the Transaction table due to customers making multiple transactions on different days.
- 7) A false record in the DOB attribute, indicating a customer who is 175 years old.
- 8) product_first_sold_date row as values integer datatype rather being in Date format

These issues can be addressed to improve the data quality and ensure accurate analysis. We will clean, standardize, and transform the data for analysis. Questions will arise and document assumptions. Collaboration with your data SME will ensure alignment with Sprocket Central's understanding.

Kind regards,

Doni Satish Velmurugan