

Dear Sir/Mam,

Thank you for providing us with the datasets from Sprocket Central Pty Ltd. The below are the highlights of the summary statistics from the three datasets received. Please let us know if the figures are not aligned with your understanding.

Transaction Table

- Record Shape : (20000, 13)
- There are missing values in 7 Columns.They can be dropped or treated according to the nature of analysis.
- The Values in the product_first_sold_date columns are not correct as it shows everything happening the same day at different times.
- There are no duplicate records, so data is unique in this table.

NewCustomerList Table

- Record Shape : (1000,23)
- There are missing values in 4 columns.They can be dropped or treated according to the nature of analysis.
- Inconsistent Data type for columns.
- Title(Column name) is not defined to 5 columns - we have to drop that columns.

Customer Demographic

- Record Shape : (4000,13)
- There are missing values in 6 columns.They can be dropped or treated according to the nature of analysis.
- We have to fix the values in gender columns - Values - (Female, Femal, F, Male, M, U)
- Recommendation: Enforce a drop-down list for the user entering the data rather than a free text field

Customer Address

- Record Shape : (3999,6)
- There is no missing value.
- Postcode column having float data type.Change it to integer datatype.
- In state columns Victoria is also mentioned as VIC and V. so we have to convert all the values to victoria.
- There are no duplicate records, so data is unique in this table.

Moving forward, the team will continue with the data cleaning, standardization and transformation process for the purpose of model analysis. Questions will be raised along the way and assumptions documented. After we have completed this, it would be great to spend some time with your data SME to ensure that all assumptions are aligned with Sprocket Central's understanding.

Kind regards,

Omkar Sanjay Barge