

Dear Client,

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

Table Name	No. Of Records	Distinct Customer IDs	Date Data Recieved
Transactions	(20000,13)	3494	9-june-2023
NewCustomerList	(1000,23)	No customer IDs	9-june-2023
CustomerDemographic	(4000,13)	4000	9-june-2023
CustomerAddress	(3999,6)	3999	9-june-2023

Notable data quality issues that were encountered and the methods used to mitigate the identified data inconsistencies are as follows.

- **Transactions sheet** :- In transaction sheet, we encountered with the blanks cells in the rows presented in the Online_order and brand column.

To mitigate it we remove the blanks rows from the sheet.

- **NewCustomerList**:- In NewCustomerList sheet, we encountered with the blanks cells in the rows presented in the last_name , DOB and job_title column.

To mitigate it we remove the blanks rows from the sheet.

Next we encountered with the wrong spelling of male and female in the gender column.

To mitigate it we put right spelling in the gender column.

- **CustomerDemographic**:- In CustomerDemographic sheet, we encountered with the blanks cells in the rows presented in the last_name , DOB and job_title column.

To mitigate it we remove the blanks rows from the sheet.

Next, we encountered with the one row in which the person DOB is 1843.

As we know it must be an outliers so for mitigate it we delete this row.

- **CustomerAddress**:- In CustomerAddress sheet, we encountered with the wrong spelling of New South Wales and Victoria in the state column.

To mitigate it we put right spelling in the gender column.

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,

Ashu Kumar

