

Dear [Client\_Name],

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below table highlights the summary statistics from the three datasets received.

Table_name	No of records	Distinct Customer_ID	Date Data received
Transaction Table	20000	3494	2020
Customer Address	3999	3999	2020
Customer Demographic	4000	4000	2020

By seeing the table, there are issues that we have encountered and methods used to remove the data inconsistencies are as follows:

1. Additional Customer ID in the 'Transactions Table' and 'Customer Address table' but not in 'Customer Demographic table'.

**Method to resolve-**

Please ensure that all tables are from the same period. Only customers in the Customer Master list will be used as a training set for our model.

This indicates that the data received may not be in sync with each other which may effect the analysis results if there are missing data records.

2. Various columns, such as the brand of a purchase, or job title, have no/empty values in certain records.

**Method to resolve-**

If there is only small no of records that are empty, remove the record entirely from the training set for

Prediction of the model. Else, if it is a important field, impute based on distribution in the training dataset.

3. Different values for the same attribute.

(e.g. New South Wales as 'NSW')

**Method to resolve-**

Use regular expression to ensure same/consistency across the data values.

**Recommendation:**

Make use of a drop down list for the user entering the data rather than a text field/area.

4. Inconsistent data type for the same attribute across tables.

**Method to resolve-**

1. Change the selected records in characters to numeric.

2. Remove the non-numeric characters from string.

**Recommendation:**

Remember to have constraint for database table data types. Because having different data type for the same field makes it difficult to interpret results.

5. Irrelevant data field 'default' in 'customer demographic table'.

**Method to resolve:**

In the customer demographic table, the default field had been removed due to irrelevant data.

**Recommendation:**

Ensure to have constraint on each data field for the tables.

Our team will continue to clean the data, standardize it for analysis purpose.

Kinds Regards

[Consultant name]