

Dear Client,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. Our team looked and assessed the quality of the datasets you provided us with. We were able to identify some issues with the datasets received which are described below. Please let us know if the datasets are not aligned with your understanding. Few recommendations have also been made to prevent the data quality issues and to improve the datasets used to drive future business decisions.

The 3 datasets attached from Sprocket Central Pty Ltd were:

- Customer Demographic
- Customer Addresses
- Transaction data in the past three months

We reviewed the datasets quality to ensure that it's ready for analysis in phase two. We found notable data quality issues that were dealt with and the methods used to mitigate the identified data issues are as follows.

• **Inconsistency in the count of customer_ids in the given three datasets**

Please ensure that all information of customers is available as missing records may lead to skewed analysis results which are not recommended to be used for data driven decisions.

Mitigation: Only customers in the Transaction list to be used as a training dataset for our analysis model.

• **Various columns, such as the brand of a purchase, online order, job title have empty values in certain records**

For key datasets, such as transactions, only few transactions have missing values which have been removed from the training dataset.

Mitigation: If it is a core field, impute based on mean values in the training dataset. If only a small number of rows are empty, delete the rows.

• **Inconsistent values for the same field (e.g. Victoria being represented as "V", "Vic" and "Victoria", Female being represented as "F", "Femal")**

Data has been cleaned using the find and replace function to ensure consistency of representations of values.

Mitigation: Use data validation techniques like drop-down lists to ensure consistency across fields.

• **Inconsistent data type for the same attribute (e.g. different date formats)**

Data was transformed by changing and maintaining consistent data types in fields such as DOB, list price.

Mitigation: ensure there is consistency among the data types in fields by using predefined constraints.

Our team will continue with the data cleaning and transformation process for the purpose of model analysis. After this, we would like to ensure that all assumptions are aligned with Sprocket Central's understanding.

with regards,
Srijan Chitranshi