

Dear Tony Smith,

Thank you for providing the datasets needed in order to carry out the analysis. Below is an overview of the data we received. Kindly verify whether the values align with what you provided from your end.

Dataset	Unique ID's	Number
Customer Demographics	Customer_id	4000
Customer Addresses	Customer_id	3999
Transactions	Transaction_id	20000

Assumptions

Moving forward, the following assumptions were made while cleaning the data. Feel free to point out any incorrect assumptions so that we can come up with the best possible recommendations from these records.

Assumption 1: Sprocket Central Pty operates using both a physical location as well as an online platform.

Assumption 2: The currency used is Australian Dollars (AUD).

Assumption 3: The datasets are of transactions that happened before 12/30/2017.

Data Quality Issues, mitigations and recommendations

A number of key data quality issues encountered are as follows, as well as ways to mitigate them and recommendations to avoid them in future.

Completeness – A number of blank records were present in both the customer demographics and transaction tables.

Mitigation: If the blank entries compose only a small percentage of the total data, filter them out and exclude them from the data to be analyzed.

Recommendation: Put measures in place to ensure that all records are entered when the data is being input to reduce or even completely eliminate the occurrence of blank entries.

Timeliness - The Transactions table contained data from the entire year of 2017 while we only required data from the last three months.

Mitigation: Provide only the data needed at a given time. We deleted all records that were older than three months from the Transactions table.

Validity - Entries in columns such as customer_id, date of birth and list_price had the wrong data types thus interfering with the accuracy of the data.

Mitigation: Ensure that data provided is in the correct format for example, D.O.B. should be in date format.

Recommendation: Put in place data constraints and drop down inputs where possible when data is being captured so that it can be stored in the correct format.

Consistency - Inconsistencies present in the data include representation of the same item in different ways such as Femal, F, Female and New South Wales.

Mitigation: Ensure that data representing the same thing is recorded in the same way across the entire data set.

Recommendation: Limit the number of options when the data is being entered by using elements such as a drop down menu or list to capture the data hence reducing inconsistencies.

Relevance - Additional customer_id's were present in the transactions and customer_address and not present in the master customer_demographics table.

Mitigation: Only data contained in the customer_demographics table will be used to train the data.

Our team will continue to clean and transform the data in order to be used for the model analysis. Kindly have your team go through the assumptions that we are working with to ensure that they align with Sprocket Central's business operations.

Kind Regards,

Winnie Gaturu.