

Dear Client,

Thank you for providing our firm with the three datasets from Sprocket Central Pty Ltd. I have created a summary table below which highlights the primary quality issues that we found within the data sets provided. If you have any questions about the issues presented, please let us know.

**Summary Table:**

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
<b>Customer Demographic</b>	-DOB format is inaccurate for some records -Age: missing	-Job title: (Blanks) -Customer Id: incomplete -Last Names: (Blanks)	- Gender: Inconsistency in values	- (Y)Diseased Customers: Filter Out	Default Column: Delete	
<b>Customer Address</b>		-Customer ID: Incomplete	-States: Inconsistent Values			
<b>Transactions</b>	-Profit: Missing	-Customer ID: Incomplete -Brand: (Blanks) -Online Order: (Blanks)	Property valuation: Inconsistent values and duplicate		-Cancelled Status Order: Filter Out	-Product Sold Format - List price format should be currency

Below we have gone more in depth into the data quality issues and have provided mitigation methods that can be used. Follow these mitigation recommendations and the data sets will improve in accuracy and in turn provide better insight for Sprocket Central Pty Ltd to make more effective business decisions.

**Accuracy Issues:**

- **DOB was inaccurate for the “Customer Demographic” dataset, missing age\_column**
- **Profit\_Column Missing for “Transactions” dataset**

Mitigation: Filter out the outliers for DOB

Recommendation: Create a profit column to get more comprehensive information and to check accuracy of sales, also create age column to have more comprehensive data and to view outliers and errors much quicker.

**Completeness:**

- **Customer\_ID has missing values for some records amongst all three datasets**
- **Missing Job Titles**

- **Last Names are missing for 125 records in “Customer Demographic” data set**
- **Brands and Online\_Order Columns are missing values (Blanks)**

Mitigation: Filter out the IDs from 1-3500

Recommendation: Cross check spreadsheets to make sure they are all up to date, this will ensure more accurate data.

## **Consistency:**

- **Gender\_Column contains inconsistent values for both “Customer Demographic” and “Customer Address”**
- **States have inconsistent values in the “Customer Address” dataset**
- **Property\_valuation has inconsistent values and duplicate in the “New Customer List” data set**
- Mitigation: Filter all “Femal” and “F” under “Female”. Filter all “M” under “Male”. Filter all “New South Wales” to “NSW” and “Victoria” under “VIC” for States\_Column.

Recommendation: Create drop-down options for “Females” and “Male” and “U” and create another drop down option for States\_Column values.

## **Currency:**

- **People with the “Y” value for Disease\_Indicator in the “Customer Demographic” dataset are not current customers therefore the data is not up to date.**

Mitigation: Filter out all customers with the “Y” value

Recommendation: Deleting the Customers that are diseased will ensure for more accurate up-to-date data which will lead for better forecasting and analysis.

## **Relevancy:**

- **No relevancy for default\_column in the “Customer Demographic”, and no relevancy for “Cancelled” value in order\_status.**

Mitigation: Remove default\_Column. Filter out all “Cancelled” statuses

Recommendation: Format metadata to make comprehensible or delete it entirely to ensure clearer data. And cancelled orders are not needed when trying to forecast or do future analysis.

## **Validity:**

- **List\_price and Product\_sale\_date are not valid formats for “Transaction” dataset**

Mitigation: Format the List\_price to currency and format the Product\_sale\_date to short date format.

Recommendation: Set up column formats so price and decimals are automatically inputted when entering new data.

Moving forward, the team will continue with the data cleanup procedures, standardization and transformation processes for the purpose of model analysis. After we have completed this, it would be great to discuss with your data SME to ensure that all the information is aligned with Sprocket Central's understanding.

Kind regards,

Waleed Hassanien

Junior Consultant