

To,
Sprocket Central Pty Ltd

Subject - Data Quality issues and strategies to mitigate these issues.

Dear Sprocket,

Below are the identified data quality issues and strategies to mitigate these issues.

Sheet Name	Unique column	No. of Rows (Before cleaning)	No. of Rows (After cleaning)
Transactions	transaction_id	20000	19946
NewCustomerList	NA (<i>since it is the target sheet</i>)	1000	736
CustomerDemographic	customer_id	4000	2863
CustomerAddress	customer_id	4000	4000

1. Completeness (Fields with missing values):

Transactions sheet → *online_order, brand, product_line, product_class, product_size, standard cost, product first sold date* contain null values.

NewCustomerList → *DOB, job title, job industry* contains null values.

CustomerDemographic → *DOB, job title, job industry* contains null values.

Mitigate Steps -

The Data Handling team can try to tally the old data or relevant records to fill in some null values.

2. Relevancy (Un-relevant values)

NewCustomerList → *gender* ⇒ Deleted value = "U" because don't know whether it is male or female

CustomerDemographics → *gender* ⇒ Deleted value = "U" because don't know whether it is male or female,
deceased_indicator ⇒ Deleted value = "Y" because want only alive customers, *default* ⇒ Deleted column since it is not decodable

Mitigate Steps -

The Data Handling team should tally the record of "U" in gender to find out the correct values.

3. Consistency (Up to date values)

CustomerDemographics → *gender* ⇒ Changed value = "F" to "Female" and "M" to "Male" CustomerAddress → *state* ⇒ Changed value = "New South Wales" to "NSW", "Victoria" to "VIC"

Mitigate Steps -

The Data Handling team should check the latest values used for fields to keep up to datedatabase.

4. Accuracy (Values not in the correct format)

Transactions → *product first sold date* ⇒ Changed data type from number to date.

Mitigate Steps -

The Data Handling team should Check the data type of columns to mitigate this type of issue.