**Dear John Doe,**

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below Summary describes about the key quality issues discovered in the three datasets. Please let us know if you have any queries about the issues presented.

**Summary:**

**Transactions Table:**

**Accuracy:**

* profit column is missing

**Completeness:**

* online\_order column contains blanks
* brand column contains blanks
* product\_line column contains blanks
* product\_class column contains blanks
* product\_size column contains blanks
* ptandard\_cost column contains blanks
* product\_first\_sold\_date column contains blanks

**Validity:**

* product\_first\_sold\_date column format

**CustomerDemographic Table**:

**Accuracy:**

* DOB is inaccurate
* Age column is missing

**Completeness:**

* Job\_title column contains blanks
* tenure column contains blanks

**Consistency:**

* gender column is inconsistent

**Relevancy:**

* Default column: delete

**CustomerAddress table:**

**Completeness:**

* Customer\_id: incomplete

**Consistency:**

* state column is inconsistent

Below are more in-depth descriptions of the issues and the methods of mitigations used, Recommendations are also included to avoid the further data quality issues in future.

**Accuracy issues:**

1. **DOB is inaccurate and age column is missing in “CustomerDemographic” table.**
2. **Profit column is missing in “Transaction” table.**

**Mitigation**: Filter out outlier in DOB.

**Recommendation**: Create age column allowing to easily check errors.

Create profit column allowing to check the accuracy of sales.

**Completeness issues:**

1. **customer\_id was inconsistent throughout the all three tables:**

**Mitigation**: Filter all customer\_ids from 1 to 3500

The data received may be not sync across the all spreadsheets, with incomplete data results will be skewed which is a completeness issue, please sync the data for better results:

1. **Blanks in job\_title in “CustomerDemographic” table and blanks in brand in “transaction table”**

**Mitigation**: filter out the blanks in job\_title and brand columns

**Recommendation**: provide a dropdown options for the job\_title and the brand

**Consistency issues:**

**1.gender column inconsistency in “CustomerDemographic” and state inconsistency in “CustomerAddress”**

**Mitigation**: Filter all ‘M’ values in gender to male and ‘F’ to female in gender column

Filer all ‘New south wales’ to NSW and ‘Victoria’ to VIC in the states column

**Recommendation**: Create dropdown options for the both state and gender column which reduces the effects of manual entry

**Relevancy issues:**

**1.Lack of relevancy in default column in “CustomerDemogrphic”**

**Mitigation**: Delete the default column in customer demographic

**Recommendation**: check for the incomprehensible metadata and delete that column

**Validity issues:**

**1.Format of product\_first\_sold\_date in “Transactions”**

**Mitigation**: format the product\_first\_sold\_date in transactions to short date format.

**Recommendations**: set up the columns in the required format so that when data entered it converts into the required format.

That summarises the all data quality issues discovered through the first stage of the data quality analysis. The mitigation methods are very simple and will improve data quality.

Please let us know if you have questions regarding mitigation methods or data quality issues.

Thank You,

Phaneendra