Dear Tony Smith,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The summary table below presents the quality issues that we discovered within the data sets.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** | **Uniqueness** |
| **Customer Demographic** | DOB: inaccurate  Age: missing | Customer\_ID: blanks  Job title: blanks | Gender: no consistency | Deceased Customers: Filter out | Default: Delete |  |  |
| **Customer Addresses** |  | Customer ID: blanks | States: no consistency |  |  |  |  |
| **Transactions** | Profit: missing | Online Order: blanks  Brand: blanks  Transaction ID: blanks |  |  | Cancelled Status Order: delete | List price: format  Product sold date: format to short date | Customer ID: duplicated values |

Below are more descriptions on the data quality issues and how to mitigate them with recommendations. These recommendations will help improve the accuracy of data and business decisions of Sprocket Central Pty Ltd.

**Accuracy Issues**

* DOB was inaccurate for *Customer Demographic* and missing an age column.
* Also missing a profit column for *Transactions.*

Recommendation : Filter out outliers in DOB and create age and profit columns. These would help check for errors and check accuracy of sales.

**Completeness**

* Customer ID and Job title had blanks for *Customer Demographic* and missing an age column.
* Customer ID and Job title had blanks for *Customer Address.*
* Online Order, brand and transaction ID had blanks for *Customer Demographic.*

Recommendation: Filter all data for blanks and uncheck them. Ensure tables are up to date (from the same time period).

**Consistency**

* Inconsistency in gender for *Customer Demographic* and States for *Customer Address*

Recommendation: Filter all M and F under ‘Male’ and ‘Female’ respectively. Do the same with the states.

This will minimize human error and increase consistency.

**Currency**

* People that are ‘Y’ in deceased\_indicator are not current.

Recommendation: Filter out customers checked ‘Y’. This will increase currency of data and have less errors.

**Relevancy**

* Lack of relevancy in default\_column and order\_status.

Recommendation : Delete data in default\_column and filter for ‘cancelled’ in order\_status. This is not important information for the data set.

**Validity**

* Format list\_price and product\_sale for *Transactions.*

Recommendation : Format product\_sale\_date to short date format and list\_price to currency.

This will make that the data is interpreted more easily.

**Uniqueness**

* Filter Customer\_ID for Unique values for *Transactions.*

Recommendation : Filter Customer ID and only leave the unique values. This will avoid repetition and miss confusion in the data.

This summarizes all data quality issues that we have observed in the first stage of quality analysis. Please let us know if you have any questions regarding the presented issues.

Best regards,

Eva Beyebach