Dear Mike Ross,

Thank you for providing us with four data sets from Sprocket Central Pty Ltd. Below is a summary table that highlights the key quality issues that were discovered in these data sets. Please let us know if there are any queries regarding these issues discovered.

**Summary Table:**

|  | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| --- | --- | --- | --- | --- | --- | --- |
| **Customer Address** |  |  | -States: inconsistency |  |  |  |
| **Customer Demographic** | -DOB: inaccurate  -Age: missing | -JobTitle: blanks  -LastName: blanks  -Tenure:  blanks | -Gender: inconsistency | -Deceased customers: filtered out | -Default column:  deleted |  |
| **NewCustomerList** |  | -Last Name: blanks  -DOB: blanks  -JobTitles: blanks |  |  |  | -3 Hidden Nameless columns: formatted |
| **Transactions** | -Profit: missing | -OnlineOrder: blanks  -Brand: blanks |  |  | -Cancelled status order:  Filtered out | -List price: formatted  -Standard cost: formatted  -Product sold date: formatted |

Below are the in-depth descriptions of the data quality issues that were discovered and mitigations used. We have also included recommendations and explanations to avoid future data quality issues. These recommendations will help in increasing the quality and accuracy of the data which is used by Sprocket Central Pty Ltd to make business decisions.

**Accuracy Issues:**

* **DOB was inaccurate for “Customer Demographic” and missing an age column; missing profit column for “Transactions”**

*Mitigation: Filter out outliers in DOB.*

*Recommendation:* Create an age column that calculates age to be stored as a number and makes it easier to find errors in the data. Create a “Profit” column in the Transactions data to check the accuracy of the sales data provided.

**Completeness:**

* **Job Titles, LastName, Tenure had missing data in “Customer Demographic”; LastName, DOB, Job Titles had blanks in the “New Customer List”; Online Order, Brand had missing data from the “Transactions”**

*Mitigation: Filter out these missing values from the sheets.*

*Recommendation:* Ensure tables are up to date. These inconsistencies can be a result of sync error across various tables.

**Consistency:**

* **Inconsistency in the demographic metric gender in “CustomerDemographics” and “CustomerAddress”; States inconsistency in Customer Address**

*Mitigation: Filtered all ‘M’ as the category Male, filtered all ‘Femal’, ’F’ as ‘Female’ for gender. Filter ‘New South Wales’ as ‘NSW’ and ‘Victoria’ as ‘VIC’ for states.*

*Recommendation:* Creation of dropdown menu for entering the data for states and gender will help in minimizing the human error and will allow only these options to be entered in the database.

**Currency:**

* **Customers that have ‘Y’ in the deceased aren’t the present customers in ‘CustomerDemographic’**

*Mitigation: Filter out the deceased customers from the ‘CustomerDemographics’*

*Recommendation:* It can be difficult to catch hold of the data about deceased status of the customers but once found it should be updated immediately in the database.

By not including the current customers of the firm it will enable us to make better future decisions.

**Relevancy:**

* **Lack of relevancy/ comprehensibility in the default column of the “Customer Demographic” and Order\_status for “Transactions”**

*Mitigation: Deleted the metadata column. Deleted the transactions with the order status as cancelled.*

*Recommendation:* By keeping the orders with status cancelled it will create an overestimation in our profit and other financial calculations.

**Validity:**

* **Format of list\_price, product\_sale\_date for “Transactions”**

*Mitigation: Format dates in their proper format, format list\_price as currency.*

*Recommendation:* Set up columns with proper formats while entering the data will allow us to overcome these errors automatically while entering the data.

This summarises all the data quality issues that were discovered through the first stage of the data quality analysis. The mitigation strategies that are suggested are simple and effective for improving the data quality for future data analysis.

Please let us know if you have any questions regarding these mitigations and any data quality issues that were identified by us.

Regards,

Tanmay Phalke