Dear Sir,

Thank you for making provisions for three (3) sets of data from Sprocket Central Pty Ltd Data Set, below is a summary and findings made for the data quality.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sheet** | **Completeness** | **Accuracy** | **Consistency** | **Currency** | **Relevance** | **Validity** |
| Customer Demographics | **Job Title:**  Blanks  **Customer ID**:  Incomplete | **DOB:** Inaccurate  **Age:**  Omitted | **Gender:**  Inconsistent | **Deceased Customer:**  Filtered out | **Default Column:**  Deleted |  |
| Customers’ Address | **Customer ID:**  Incomplete |  | **State:**  Inconsistent |  |  |  |
| Transaction | **Customer ID:**  Incomplete  **Online Order:** Blank  **Brand:**  Blank | **Profit:**  Omitted |  |  | **Canceled status Order:**  Filtered Out | **List Price:**  Formatted  **Product sold Date:**  Formatted |

**Data Quality Explanation and Resolution**

1. **Completeness:**

* The Customer Demographics and Customer’s Address worksheet had extra inconsistent data for the Customer ID column and had blank spaces for the job title column,
* The transaction worksheet witnessed inconsistencies in the customer ID column also accompanied by blank spaces in the Online order and Brand column

Recommended Mitigations

* Filter out the customer ID column (to rid of inconsistent data)
* Make sure tables are up to date (within the same period under consideration)
* Filter out the blanks in the job title column in the customer demographics worksheet,
* Filter out the blanks in the online order and brand columns respectively in the transaction worksheet.

1. **Accuracy (Correct Values):**

* The customer Demographics Worksheet contained a measure of inaccuracy in values in the DOB column and do not have a column for the Ages of the customers (Omission)
* The Transaction Worksheet does not have a column for Profit (which can be derived from the difference between the list price column and the standard cost column).

Recommended Mitigations

* Filter out outliers from the DOB column and possible blank spaces in Customer Demographics Worksheet
* Create a column for the customer’s age (this will reduce error and improve readability) in Customer Demographics Worksheet.
* Create a column for profit (This will aid future financial analysis)

1. **Consistency:**

* The customer Demographics worksheet contains some inconsistencies in the gender column.
* The customer address worksheets witnessed inconsistency in the state columns.

Recommended Mitigations

* Filter out all the options for the gender column and rename them to “Male” and “Female” in the Customer Demographics Worksheet.
* Filter out the options for the state column and rename them to “Queensland”, “New South Wales” and “Victoria”

1. **Currency:**

The Customer Demographics worksheet is a combination of living and deceased customers, the deceased customers are no longer current customers and hence should be removed.

Recommended Mitigations

Filter out customers checked “Y” (Indicating affirmation for Deceased customers)

1. **Relevance**

The Customer Demographics worksheet contains a default column that is not relevant to the use of the data.

The Transaction worksheet have status orders that were canceled (hence are not relevant)

Recommended mitigations

* Delete the default column
* Filter out the canceled status order

1. Validity

The Transaction worksheet has the list price column in a text format instead of numbers, also the product sold Date column is in a text format instead of a date format(short date). This implies that the data types used in storing the data in both columns are not appropriate and hence invalid.

Recommended Mitigations

* Format both columns into the proper data type which is numeric (numbers).

That summarizes the data quality issues in the first stage of the data quality analysis. The mitigation strategies employed are simple and straight to the point as they will improve the data analysis output and visualization without hampering the integrity of the data. Kindly revert if there are other issues not covered by this analysis.

Kind regards,

Egbujua Michael C.