Dear Client,

Thank you for providing us the datasets from sprocket central Pty Ltd .below are identified quality issues that we discovered within your three dataset during our analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Sheet Names | Number of records | Unique customer IDS | Unique column |
| Transactions | 2000 | 3494 | Transaction id |
| New Customer List | 1001 |  | None |
| Customer Demographic | 4000 | 4000 | Customer id |
| Customer Address | 3999 | 3999 | Customer id |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sheet name | Accuracy | Completeness | consistency | currency | Relevancy | validity |  |
| Transactions | Add profit column it is missing. | Missing values: Online order, product id, brand, product line, product class, standard cost |  | List price in wrong data type |  | Product first sold date: format (change text to date) |  |
| New Customer List |  | Missing values: last name, job title, job industry. |  |  |  |  |  |
| Customer Demographic | Inaccurate DOB. Age column is missing | Missing values: DOB, job titles.  Customer id has blank cells | Gender inconsistency | Deceased indicator “y” filters out | Delete default column.  Filter out the cancelled order |  |  |
| Customer Address |  | Customer id incomplete | State inconsistency |  |  |  |  |

Below are more in-depth description of data quality issues and ways to solve these issues:

**Accuracy issues:**

* DOB from customer demographics was inaccurate and age column is missing.
* Column for profit is missing in transaction table

Solution:

* filter out the outliners in DOB.
* create age column for checking errors easily.
* create a profit column to check accuracy in sales.

**completeness:**

* Missing values: Online order, product id, brand, product line, product class, standard cost in “transactions table”
* Missing values: last name, job title, job industry.
* Missing values: DOB, job titles in “customer demographics”.

Solution:

* Since only a few of the record has missing values filter out the blanks.

**Currency:**

* People with “y” in deceased \_indicator column are not current customer of the company.

Solution:

* Filter out customer with “y” indicator

**Relevancy:**

* They is no relevancy in default and order status in customer demographics

Solution:

* filter out cancelled orders so that it would not lead to overestimate of orders per year.

**Validity**

* First product sold date and List price is in the wrong data type in “transaction “.
* Solution: change the format of the first product sold date to “short date” ang list price to” currency”.

**Consistency:**

* Gender inconsistency in “customer demographic” and for state in “customer addresses”.

Solution:

* Find and replace all ‘M’ by male and all ‘F’ by female for gender.
* Find and replace “new south Wales” by NSW in state .

The team will continue with data cleaning, standardization and transforming it into a valuable insight. Please let me know if you have any question or queries regarding the issues and solutions.

Best regards

Miss Oloruntoba Oluwatoyin (intern)

In Behalf of the KPMG Data Analytics Team