Dear Client,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The table below highlights key quality issues that we discovered within the three data sets. Please let us know if you have any queries surrounding the issues presented.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **ACCURACY** | **COMPLETENESS** | **CONSISTENCY** | **CURRENCY** | **RELEVANCY** | **VALIDITY** |
| **TRANSACTIONS** |  | There are missing values in 7 columns.  customer\_id: not in synchronization |  |  |  | product\_first\_sold\_date:  format |
| **CUSTOMER ADDRESS** |  | customer\_id: not in synchronization |  |  |  |  |
| **CUSTOMER**  **DEMOGRAPHIC** | DOB: Inaccurate | There are missing values in 5 columns.  customer\_id: not in synchronization | Categories are not correctly named in gender column.  The values are inconsistent in default column. | Deceased customers: filter out | Default column: delete | DOB: format |

Below are more in depth descriptions of data quality issues discovered and methods of mitigation used.

**ACCURACY ISSUES:**

DOB was inaccurate for Customer Demographic dataset.

Mitigations: Filter out extreme outliers from DOB column

**COMPLETENESS:**

online\_order, brand, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date fields in Transactions dataset contains missing values/blanks.

last\_name, DOB, job\_title, default, tenure fields in Customer Demographic dataset contains missing values/blanks.

customer\_id column were inconsistent among Customer Demographic, Customer Address and Transactions datasets.

Mitigations: Filter out records containing incomplete data.

**CONSISTENCY:**

Inconsistency with spellings of categorical values in gender field for Customer Demographic and Customer Address datasets respectively.

The values are inconsistent in default column for Customer Demographic dataset.

Mitigations: Renaming the categorical values to ensure consistency of terminologies across datasets for categorical fields. Dropping default column from Customer Demographic dataset.

**CURRENCY:**

The Customer Demographics dataset contains records related to deceased customers.

Mitigations: Filter out Customer Demographic dataset for deceased customers by checking the deceased\_indicator field.

**RELEVANCY:**

The values are inconsistent in default column for Customer Demographic dataset.

Mitigations: Dropped the default column which contained unrelated/unclear symbols.

**VALIDITY:**

product\_first\_sold\_date column in Transactions dataset is incorrectly formatted.

DOB column in Customer Demographic dataset is incorrectly formatted.

Mitigations: Convert the datatype of product\_first\_sold\_date column from numeric to date format in Transactions dataset. Convert the datatype of DOB column from numeric to date format in Customer Demographic dataset.

Please let us know if you have questions regarding any data quality issues identified.

Thank you,

Yours sincerely,

Srushti Desai