Dear Sprocket Central ltd.

Your dataset was well received and below in Table I is the descriptive statistics of the three dataset. Please let us know if the figures doesn’t align with your own.

Table I: Table Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table Name** | **No. of records** | **No. of columns** | **Data types of the columns** | **Distinct Customer ID** | **Date received** |
| Transaction Data | 20,000 | 13 | Numeric: 7  Date: 1  Category: 5 | 3,494 | 27-08-2022 |
| Customer Demographic | 4,000 | 13 | Numeric: 3  Date: 1  Category: 9 | 4,000 | 27-08-2022 |
| Customer Address | 3,999 | 6 | Numeric: 3  Category: 3 | 3,999 | 27-08-2022 |
| New Customer List | 1,000 | 23 | Numeric: 11  Category: 11  Date: 1 | Nil | 27-08-2022 |

Data quality dimensions like completeness, accuracy, consistency, validity, uniqueness, relevancy, and currency (values up to date) was used to assess the data qualities issues in the three datasets, you can find the summary in Table II

**Table IIa: Transaction Sheet**

|  |  |  |
| --- | --- | --- |
| **Column** | **Issue** | **Dimension** |
| product\_first\_sold\_date | the column name is misleading, it should be product\_first\_sold | Validity |
| online\_order, brand, product\_line, product\_sales, product\_size, list\_price, standard\_cost and product\_first\_sold\_date | Missing values in online\_order, brand, product\_line, product\_sales, product\_size, list\_price, standard\_cost and product\_first\_sold\_date columns | completeness |
| Product\_id | product\_id with value 0 | Accuracy |
| Online\_order | online\_order value contain a float number | Validity |

**Table IIb: Customer Demographic Sheet**

|  |  |  |
| --- | --- | --- |
| **Column** | **Issue** | **Dimension** |
| last\_name, DOB, job\_title, job\_industry\_category, default, tenure | Missing values in last\_name, DOB, job\_title, job\_industry\_category, default, tenure | Completeness |
| gender | inconsistent values | Consistency |
| DOB | 1. Name of the column is inconsistent with the other columns 2. datatype issue | Validity |
| Deceased\_indicator | Inconsistent values compare with other columns | Consistency |
| default | Doesn’t carry meaningful values | Accuracy |

**Table IIc: Customer Address Sheet**

|  |  |  |
| --- | --- | --- |
| **Column** | **Issue** | **Dimension** |
| state | inconsistence values | Consistency |
| country | Unique value | Uniqueness |

**Table IId: New Customer Sheet**

|  |  |  |
| --- | --- | --- |
| **Column** | **Issue** | **Dimension** |
|  | Four columns doesnt have name description | Validity |
| past\_3\_years\_bike\_related\_purchases | column name is too long | Validity |
| Deceased\_indicator | Inconsistent value as compare to other columns | Consistency |
| last\_name, job\_title, DOB, and job\_industry\_category | missing values in last\_name, job\_title, DOB, and job\_industry\_category | Completeness |
| Gender | invalid value 'U' | Validity |
| DOB | date format | Validity |
| country | Unique value | Uniqueness |

Above are the issues detected, take note it for future occurrences, attached will be the quality assessment notebook with the details of each operations.

Regards,

Dauda Sani Abdullahi.