Dear Client,

Thank you for providing us with the datasets from Sprocket Central Pty Ltd. The table below highlights the summary statistics of the datasets received. Please let us know if there is any discrepancy in the figures.

|  |  |  |  |
| --- | --- | --- | --- |
| WORKSHEET NAME | NUMBER OF RECORDS | UNIQUE CUSTOMER ID | DATE DATA RECIEVED |
| CUSTOMER DEMOGRAPHIC | 4,001 | 4,001 | 10/03/2021 |
| TRANSACTION | 20,000 | 3,495 | 10/03/2021 |
| CUSTOMER ADDRESS | 4,000 | 4,000 | 10/03/2021 |

The data quality issues encountered and the methods used to rectify the identified inconsistencies are listed below.

* Some columns have empty values in certain records e.g job title, brand of purchase. If the missing or null values are not significant, then we can filter them out completely from our data set . Otherwise, we would input based on the distribution in the data set.
* Additional customer ids are present in the Transaction and Customer address worksheets which can not be found in the Customer Demographic sheet. The data provided are not in sync with each other which could be as a result of missing data or data is not up to date.
* Data set set contains inconsistent values for the same attribute. For example, Victoria is written as V, Vic, Victoria. We replaced such values to ensure consistency in the data set.
* Data types are inconsistent for the same attribute. These data types were converted to ensure consistency through out the data sets.

Regards,

Ajiboye-Femi Faith