**quality assessment**

Dear sprocket central marketing team,

Thank you for providing us with the three datasets from sprocket central Pty ltd. Our analytics team here at KPMG has made a thorough assessment of the data sets you have provided, Please let us know if the figures are not aligned with your understanding.

|  |  |  |  |
| --- | --- | --- | --- |
| Table name | No.of records | Distinct customer ID’s | Date data received |
| Customer demographic | 50956 | 4000 |  |
| Customer address | 23995 | 4000 |  |
| Transaction data | 258460 | 20000 | 19999 |

The following notable issues with the quality of the data were detected. Their respective mitigations have also been included.

* Additional customer Id’s in the ‘Transactions table’ and ‘Customer Address table’ but not in ‘Customer Master (Customer Demographic)’. Please ensure that all tables are from the same period. Only customers in the Customer Master list will be used as a training set for our model. This indicates that the data received may not be in sync with each other which may skew the analysis results if there are missing data records. Please refer to excel file ‘data\_outliers.xlsx’ for the list of outliers between tables.
* The customer demographic sheet contains a “default” column which is unclear as its contents are also not displayed. Please use a clearer title for the column and correct the format of the content to display actual data.
* Please note that there are missing values throughout the transaction sheet, we kindly ask that you complete the sheet.
* Please note that there are also missing values in the customer demographics sheet and the new customer’s sheet.
* Please correct the format of your first sold column on the transaction sheet.
* Please refrain from including N/A in the dataset. The term “other” would be better on the new customers sheet- job industry category column for working on the data more smoothly.
* The address column in the new customers sheet has inconsistent formats. Please list them in the same format.
* The property validation column in the new customers sheet has inconsistent values, and is also inaccurately and/or inappropriately depicted in a numerical value.
* The value column in the new customers list contains inaccurately and unclearly placed numerical values.
* Inconsistent data type for the same attribute (e.g. numeric values for some fields and strings for others). Please Convert the selected records in characters to numeric. Remove non-numeric characters from string. Recommendation: Ensure that fact tables in the given database have constraints on data types. Having different data types for a given field makes it difficult to interpret results at the later stage. Therefore, appropriate data transformations are made to ensure consistent data types for a given field.

Best regards,

Kaleb Alemayehu