Dear Client,

Thank you for providing us with the three datasets from Sprocket Central Pty Ltd. The below table highlights the summary statistics from the three datasets received.

|  |  |  |  |
| --- | --- | --- | --- |
| Table | No. of Records | No. of Distinct Customer IDs | Date Data Received |
| Customer Address | 4000 | 3999 | 15/05/2023 |
| Customer Demographic | 4000 | 4000 | 15/05/2023 |
| Transaction Data | 20000 | 3494 | 15/05/2023 |

The following are notable data issues that were encountered during the data quality assessment and methods that were applied to mitigate them.

* **Inconsistent Data type**

The columns with inconsistent data type includes

**Customer Demographic data**: **DOB, Tenure, past\_3\_years\_bike\_related\_purchases**

**Transaction data**: **list\_price, standard\_cost**

**Mitigation:** Convert selected records to appropriate data types, e. g. **DOB** column to date type

* **New Columns**

Performed some calculations to get new columns for the following tables

**Transaction data: Profit, and Profit margin**

**Customer Demographic data: age**

* **Inconsistent Values**

**Examples: Female being represented as “F” and “Femal”, Male as “M”, New South Wales as “NSW”**

**Mitigation:** Use a regular expression to ensure consistency.

* **Columns with Empty Values**

**Transaction data**: **order online** (360 /20000), **standard cost** (197/20000)

**Customer Demographic data: job title (**506/4000), **tenure** (87/4000),

Mitigation: If only a small number of rows are empty, filter out the record entirely from the training.

To achieve the said goals, the team will continue with the data cleaning, standardisation and transformation process for the purpose of model analysis.

Thank you for your time.

Regards,

Louis