

Dear SME ,

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. Please let us know if the figures are not aligned with your understanding.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sheet Name** | **Column Name** | **Data Quality Issue** | **Action Needed** |
| Transaction | Online order | Blank (Null values) | Remove blank |
|  | Brand | Blank (Null values) | Remove blank |
|  | Product first sold date | Relevancy | To convert from integer to date time format |
|  |  |  |  |
| New Customer List | Last name | Blank (Null values) | Remove blank |
|  | Gender | Consistency | 1. which is undefined |
|  | Job title | Blank (Null values) | Remove blank |
|  | JOB industry category | Blank (Null values) | Remove blank |
|  |  |  |  |
| Customer Demographic | Last name | Blank (Null values) | Remove blank |
|  | Gender | Spelling error,  ‘U’ - Undefined value |  |
|  | Job title | Blank (Null values) | Remove blank |
|  | JOB industry category | Blank (Null values) | Remove blank |
|  | deceased\_indicator | Remove unnecessary ‘Y’ from column as we only want customer that are still arrive |  |
|  | Default | Relevancy | Removed the column |
| Customer Address | States | Consistency | Replace Victoria by VIC |

Furthermore, recommendations have been provided to avoid the re occurrence of data quality issues and improve the accuracy of the underlying data used to drive business decisions.

**Various columns, such as the brand of a purchase, or job title, have empty values in certain records**

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

**Inconsistent values for the same attribute (e.g. Victoria being represented as VIC ”)**

*Mitigation:* Replaced extended values into abbreviations to ensure consistency across addresses.

Additionally in gender records where ‘U’ have been removed

**Inconsistent data type for the same attribute (e.g. numeric values for some fields and strings for others)**

*Mitigation:*Change data type for columns.Therefore, appropriate data transformations are made to ensure consistent data types for a given field.

Moving forward, the team will continue with the data cleaning, standardization and transformation process for the purpose of model analysis. Questions will be raised along the way and assumptions documented. After we have completed this, it would be great to spend some time with your data SME to ensure that all assumptions are aligned with Sprocket Central’s understanding.

Kind regards,

Samir Mendhe