Dear Kathleen,

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity |
| Customer Demographic | DOB: inaccurate  Age: Missing | Job Title: blanks  Customer Id: Incomplete | Gender: Inconsistency | Deceased customers: filter out | Default column: delete |  |
| Customer Address |  |  | States: inconsistency |  |  |  |
| Transactions | Profit: missing | Customer id: incomplete  Online order: blanks  Brand: blanks |  |  | Cancelled status order: filter out | List price: format,  Product sold date: format |

Notable data quality issues that were encountered and the methods used to mitigate the identified data inconsistencies are as follows. Furthermore, recommendations have been provided to avoid the reoccurrence of data quality issues and improve the accuracy of the underlying data used to drive business decisions.

* **Accuracy issues:**

‘DOB’ was inaccurate and ‘age’ column was missing for ‘Customer Demographic’ sheet. A profit column was missing for ‘Transaction’ sheet.

Mitigation: Filter out outlier in DOB.

Recommendation: Create an age column for more comprehensible data and easier to check the errors. Create a profit column in Transaction sheet to check accuracy of sales.

* **Completeness:**

Customer ids were incomplete among ‘Customer demographic’ and ‘Transactions’ sheet. Blanks in job title for customer demographic sheet; blanks in online order and Brand column for ‘Transaction’

Mitigation: Filter customer id column, job title for ‘Customer demographic’ sheet and filter out online order and brand column for ‘Transactions’ sheet.

Recommendation: Provide drop down options for job title, online order, and brand columns.

* **Consistency:**

Inconsistency in gender column for ‘Customer Demographic’ sheet and inconsistency in states column for customer address.

Mitigation: Change all ‘M’ value under the category of ‘Male’, all ‘Femal’ and ‘F’ under the category of ‘Female’. Filter all ‘New South Wales’ to ‘NSW’ and ‘Victoria’ to ‘VIC’ for states.

Recommendation: Create dropdown options for ‘Male’, ‘Female’, and ‘U’ in gender. Create dropdown options for all states.

* **Currency:**

Customers that are ‘Y’ in deceased\_indicator are not current customers for ‘Customer Demographic’.

Mitigation: Filter out customers checked ‘Y’ in deceased indicator.

* **Relevancy:**

Lack of relevancy in default column for ‘Customer Demographic’ and Cancelled order status in Transactions.

Mitigation: Deleted the default column and filter out ‘Cancelled order status’.

Recommendation: Check for irrelevant Metadata and filter it to make comprehensible.

* **Validity:**

Format of ‘List Price’ and ‘Product sold date’ in ‘Transaction’.

Mitigation: Format ‘product sold date’ to short date format, Format ‘list price’ to currency.

Recommendation: set up columns to current format such as price and decimals are already in place when entering new data.

Moving forward, the team will continue with the data cleaning, standardisation, 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,

Agnishwar Das