Dear Customer,

Thank You for providing the datasets for Sprocket Central Pty Ltd. The table below is a summary of the key data quality issues that were found on the 3 datasets. Recommendations to mitigate the issues are provided after the Summary table

**Summary Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Completeness** | **Accuracy** | **Consistency** | **Currency** | **Validity** |
| **Customer**  **Demographic** | **-** Job title:Blanks | **-** DOB : Inaccurate  **-** Age: column missing | **-** Gender:  inconsistent | **-** Deceased Customers: Filtered |  |
| **Transactions** | **-** online order: blanks  **-** brand: blanks |  |  |  | **-** list\_price  :format  **-** Product sold date: format |
| **Customer**  **Address** |  |  | **-** States: inconsistent |  |  |

Below is detailed explanations for the quality issues and mitigation applied and further recommendations to improve the data quality. Recommendations would improve the accuracy for further process in analysis for the data.

**Completeness:**

- *job\_title in “Customer Demographic”*: Few data cells were blanks in the the column

*Mitigation*: Filtered out the missing data

*Recommendation*: To avoid the missing data add drop down with all the job titles so that a

value has to be chosen for that column and there will not be blanks.

-*online\_order in “Transactions*” : Data cells were having blanks

*Mitigation*: Filtered out missing data

*Recommendation*: Provide option to select the status of the order for example True or False.

-*brand in “Transactions”*: Data cells were having blanks

*Mitigation*: Filtered out missing data

*Recommendation*: Provide drop-downs for all available brands to select in the data.

**Accuracy:**

- *DOB in “Customer Demographic”*: Some dates were inaccurate

*Mitigation*: Filtered out the outlier by adding Age column.

*Recommendation*: To avoid the inaccurate data Age column is added to check if the

customer have relevant DOB or not which can be then filtered by

comparing with Age column.

**Consistency:**

- *Gender in “Customer Demographic*”: Specific genders were entered using various ways and

were inconsistent

*Mitigation*: Male gender was entered using M and male.Female was entered using F,Femal

and Femalee which were then converted to Female by replacing inconsistent

values.

*Recommendation*: To avoid further inconsistency Option to select a particular gender must

be provided.

- *States in “Customer Address”*: Specific States were entered using various ways and were

inconsistent

*Mitigation*: New South Wales were replaced with NSW and Victoria with VIC .

*Recommendation*: To avoid further inconsistency Option to select a particular state must be

provided.

**Currency:**

- *deceased indicator in “Customer Demographic”*: Information of deceased customer are

used.

*Mitigation*: No need for deceased customer so filtered those values.

*Recommendation*: The data has to be checked frequently and customer which are deceased

should be updated.

**Validity:**

- *list\_price and product\_first\_sold\_date in “Transactions”*: Price and date were not in a proper format

*Mitigation*: Price format changed as dollar and date was changed into Short date format.

*Recommendation*: Set up columns with specific format before filling the cells of data.

That gives all the data quality issues their mitigation which are an immediate solutions and the recommendations which would be necessary to further improve the data quality. Let me know if you have any questions.

Kind Regards,

Vivek Mudaliar