Dear Client,

Thank you for providing us with these Dataset from Sprocket Central Pty Ltd. The summary table below highlights key quality issues we discovered within the three data sets.

Summary Table:-

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Customer Demographic** | **Customer Address** | **Transactions** |
| **Accuracy** | * DOB: inaccurate | * Property Value : Different format |  |
| **Completeness** | * Job Title : Blank |  | * Profit Missing * Online Order blank, * Brand blank |
| **Consistency** | * Gender Inconsistency | * State Inconsistency |  |
| **Currency** | * Deceased Customers filter out |  |  |
| **Relevancy** |  |  | * Cancelled Status :Filter Out |
| **Validity** |  |  | * List Price : Format * Product date : Format |
| **Uniqueness** | * Customer ID |  |  |

Following are in dept descriptions of data quality issues discovered and the methods of mitigation used.

* Accuracy : -
  + DOB was inaccurate for ‘Customer Demographic’ and Property Value for ‘Customer Address’ were of different format.

**Mitigation:** Filter Out redundant data in DOB.

* Completeness :-
  + Additional Customer ids were inconsistent among ‘Customer Demographic’.

**Mitigation:** Filter all Customer ids from 1 to 3500.

* + Job Title was also blank on some occasion.

**Mitigation:** Filter Out Blank data.

* + Online Order and Brand column were blank in Transactions.

**Mitigation:** Filter Out Blank data.

* Consistency :-
  + Inconsistency in gender for ‘Customer Demographic’ and ‘Customer Address’ respectively.

**Mitigation:** Changed ‘M’ and ‘F’ into Male and Female for gender. Changsed ‘New South Wales’ and ‘Victoria’ to ‘NSW’ and ‘VIC’ for states.

* Currency :-
  + People 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 or comprehensibility in default \_column for ‘Customer Demographic’ and Order\_status for ‘Transactions’.

**Mitigation:** Filter out ‘Cancelled’ Order\_status.

* Validity :-
  + Format of list\_price, product\_sale\_date for ‘Transactions’.

**Mitigation:** Format product\_sale\_date to short date format, format list\_price to currency.

**That summarizes all the data quality issues discovered through the data quality analysis.**

**Please let us know if you have queries regarding mitigation or any data quality issues identified.**

**Kind Regards,**

**Faheem Shaikh.**