Dear Sir,

Thank You for providing us with three dataset from Sprocket Central Pty Ltd. The Summary table below highlights key point issues that we discovered within these given datasets. Please let us know if you have any queries surrounding the issues presented.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Completeness | Consistency | Relevance | Validity |
| Customer  Demographic | DOB: Inaccurate  Age: Missing | Job title: Blank  Customer Id: Incomplete | Gender: Inconsistent | Default column:  Deleted |  |
| Customer Address |  | Customer Id: Incomplete | States: Inconsistent |  |  |
| Transactions | Profit: Missing | Customer Id: Incomplete  Online Order & Brand: Blanks |  |  | List price:  Format  Product sold date: format |

Below are in detailed description of data quality issues found and mitigation used.

* Customer Demographic:

1. DOB was inaccurate in customer demographic data set.

Mitigation: Filtered outlier in DOB. Created additional column of age for easier identification of error.

1. Additional Customer ids were inconsistent

Mitigation: Filtered customer ids from 1 to 3500.

1. Inconsistency in gender

Mitigation: Filter all ‘M’ under category of ‘Male’ and filter all ‘f’ and ‘femal’ under category of ‘Female’.

* Customer address:

1. Additional Customer ids were inconsistent

Mitigation: Filtered customer ids from 1 to 3500.

1. Filter all ‘New South Wales’ under category of ‘NSW’ and Filter all ‘Victoria’ under category of ‘VIC’ for states.

* Transactions:

1. Missing Profit column in transaction dataset.

Mitigation: Created additional column of Profit for easier identification of error.

1. Incomplete online order and blank brands

Mitigation: Blanks are consider as incomplete data hence omitted.

1. Format of list\_price, Product\_sale\_date

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

This summaries all data quality issues through this stage of the data quality analysis. The suggested mitigations are simple and effective way to improve data quality for future analysis.

Please let us know if you have any questions regarding this.

Kind Regards,

Anushka Mhaskar