Dear Sir/Madam,

Thank you for providing datasets from Sprocket Central Pty Ltd. The summary table below highlights key quality issues that we identified within the datasets. Please let me know if you have any issues.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Transactions** | **Customer Demographic** | **Customer Address** |
| **Accuracy** | * Profit missing | * DOB inaccurate * Age missing |  |
| **Completeness** | * Missing data in brand, online order. * Customer Id incomplete | * Missing data in Last name, job title, tenure * Customer Id incomplete | * Customer id incomplete |
| **Consistency** |  | * Gender is inconsistent | * State is inconsistent |
| **Currency** | * Remove deceased customers |  |  |
| **Relevancy** | * Remove cancelled status orders | * Delete default |  |
| **Validity** | * List price and Product first sold date format | * DOB format |  |

Below are a few recommendations to mitigate current data quality issues to improve accuracy of data to influence better business decisions of Sprocket Central Pty Ltd.

**Accuracy Issues:**

Date of Birth is inaccurate in customer demographic. Missing Age and Profit column in customer demographic and Transaction.

*Mitigation:* filter out anomalies in date of birth

*Recommendation:* Creating new columns of such as Age would help in identifying anomalies and creating a profit column would help in better analysis.

**Completeness Issues:**

Missing data in Customer demographic and Transaction. Customer id are not identical in three datasets

*Mitigation:* We need to filter out customer id which are above 3500

*Recommendation:* Data needs to be updated constantly to avoid such issues. Customer id within 3500 have complete data and will be used for the analysis.

**Consistency Issues:**

Gender and state columns are not consistent in customer demographic and customer address

*Mitigation:* Gender and state should be corrected so general form

*Recommendation:* Drop down boxes should be used during data collection as it minimizes these types of errors.

**Currency Issues:**

Data of deceased users are present in the datasets

*Mitigation:* filter out deceased users’ data.

*Recommendation:* These data should be avoided and make sure that deceased users’ data is automatically removed while updating. For our analysis we are not concerned with deceased users.

**Relevancy Issues:**

Default column in customer demographic and cancelled order status contains lack of relevancy

*Mitigation:* default column containing meta data can be removed and cancelled order status can also be omitted

*Recommendation:* Review the data before incorporating data into data sets.

**Validity Issues:**

Format of list price, product first sold date and few rows in date of birth are not in a correct format

*Mitigation:* product first sold date needs to be converted into dates. List price into currencies and need to correct the data format of few unfamiliar rows in date of birth.

*Recommendation:* Formatting should be specified while collecting the data.

These are all the issues identified in the first stage of data quality analysis. The mitigation and recommendation suggested will help in improving the data quality for better business analysis.

Please let me know if you have questions regarding the mitigation or any issues regarding the data quality issues identified.

Kind regards,

Nikhil Thapa