Hi there,

Thank you for giving us the opportunity to analyse and identify the data quality issues and strategies to mitigate these issues. Below is a list of the Data Quality issues which we discovered from datasets.

Data Quality Framework Table

Datasets	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	Inaccurate D.O.B Age Missing	Job title: blanks	Gender: inconsistency	Filter out deceased customers	Delete: default column	
Customer Address			States: inconsistency			
Transactions	Profit: missing	Online order: blanks Brand: blanks			Cancelled order status: filter out	List price: format Product first sold date: format

Detailed description of data quality issues and there are recommendations to mitigate data quality issues.

Accuracy Issues

- DOB was inaccurate for the “Customer Demographic” dataset.
- Profit column and age column are missing in the “Transactions” dataset.
- **Mitigate:** Filter out DOB.
- **Recommendations:** Add the age column in the dataset. Add a profit or profit margin column in the “Transactions” dataset for better insight into sales data.

Completeness

- Blanks in job title for the “Customer Demographic” dataset.
- Blanks in online order and brand column in the “Transactions” dataset.
- **Mitigate:** Filter out job title; brand; and online order columns.

- **Recommendations:** Blank values in columns affect data analysis in a negative way, so avoid null values.

Consistency

- Inconsistency in the gender column for the “Customer Demographic” dataset.
- Inconsistency in the states column for the “Customer Address” dataset
- **Mitigate:** Filter out ‘M’ as ‘Male’ and ‘F’ as ‘Female’ in gender column. Filter out ‘New South Wales’ as ‘NSW’ and ‘Victoria’ as ‘VIC’ in the status column.

Currency

- The deceased indicator column having the value ‘Y’ in the “Customer Demographic” dataset is not a current customer.
- **Mitigate:** Filter out deceased indicator column having value ‘Y’ in the dataset.
- **Recommendations:** Up-to-date data on deceased customers will increase the efficiency of data.

Relevancy

- Lack of relevance in the default column for the “Customer Demographic” dataset and order status column for the “Transactions” dataset.
- **Mitigate:** Delete meta-data in the default column in “Customer Demographic”. Remove the cancelled value from the order status column in “Transactions”.

Validity

- Format of list price; and product first sold date column in “Transactions” dataset.
- **Mitigate:** Change the format of the list price column to currency and product first sold date to short date.
- **Recommendations:** Assign a suitable format for any column to make data more readable and understandable.

The recommendations and mitigation strategies are effective and easy to implement. The recommendations will improve data quality and make it more effective or efficient. Also, these suggestions will improve data analysis output.

Please contact us for any queries and issues regarding the data quality framework table.

Best regards,
Mehul Bisht