

Dear Client,

Firstly, I would like to thank you for providing three datasets from Sprocket Central Pty Ltd. During the Processing of the dataset, I encountered some key quality issues inside the dataset. I have described below with data quality framework table for your ease of use and understanding. Please let me know if you have any queries about the presented issues.

Dimensions	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity	Uniqueness
Transactions	Missing: Profit column	Incomplete: Customer ID Blanks: online order, brand, product line, product class, product size, product first sold date and standard cost			Delete: Default column	Format: list price and product first sold date	
Customer Demographics	Inaccurate: DOB Missing: Age column	Blanks: DOB, job title, and tenure	Inconsistency: Gender				
Customer Address		Incomplete: Customer ID	Inconsistency: state				

Dataset Name	Unique ID	Number of Records	Date
Customer Demographic	4002	4002	3/27/2022
Transactions	3495	20001	3/27/2022
Customer Address	4001	4001	3/27/2022

Here, I have identified some data quality issues and below is the descriptive understanding of issues their mitigating solutions and recommendation with a data quality framework for better results.

1. Accuracy
For the missing values of Transaction and Customer Demographics, create Profit and Age columns respectively to identify the errors.
For DOB, filter out outliers.
2. Completeness
For incomplete data, filter all the customer ids.
For Blanks, filter out blanks, simplify to another category or fill with appropriate mean, or median, and provide the drop-down option.
3. Consistency
For inconsistency data, filter all or change inappropriate words to appropriate ones as changing M to male and F to female from Customer Demographic.
4. Currency
5. Relevancy
For irrelevant data delete the data if they are not in an understandable format.
6. Validity
For price indicating columns format with price or currency and for date format with short date format.
7. Uniqueness

In this way, the first step of data analysis is completed with the identification of possible errors and risks it may encounter. Thus, there are some simple and effective mitigating strategies described above. We are moving forward to steps like data cleaning and visualization. Please let me know if you have any related queries.

Thank you

Kind Regards

Saroj Rimal