

Dear Client,

Thank you for providing us with 3 datasets from Sprocket Central Pty Ltd. The table below emphasizes all the quality issues that we discovered within the three data sets.

Quality Issues	Customer Demographic	Customer Addresses	Transaction in the past 3 months
Accuracy	DOB : inaccurate Age : missing		Profit : missing
Completeness	Job_title : blanks last_name : incomplete		Online_Order : incomplete Brand : incomplete Product_line : incomplete
Consistency	Gender : inconsistency	States : inconsistency	
Currency			
Relevancy	Default column : delete		
Validity			List_price : does not conform to a specific format Product_first_sold_date : doesn't conform to a specific format
Uniqueness			

Strategies to mitigate these issues:

- As mention above DOB was inaccurate and age was missing in “Customer Demographic” data set. Also, the profit section was missing in “Transactions” dataset.
 - One way to solve these issues are to add columns called Age and Profit to the data sets for better accuracy.
 - When we created the Age column, we found one row with age 120 years which is not a accurate data, so we removed that data to make it accurate.
- Blank rows in job_title and last_name columns in “Customer Demographic”, online_order and product_line in “Transaction” data sets are incomplete. we weren't sure whether we will need these data for future analysis or not, so we decided to keep the blank data rows for now.
- For the Inconsistency in gender column in “Customer Demographic” and States column in “Customer Address” datasets, we can create dropdown option for ‘Female’, ‘Male’, ‘U’ in gender and ‘NSW’, ‘VIC’ options for states to solve the quality issue.

- To get rid of the lack of relevancy in default column in “Customer Demographic”, we deleted the metadata in default column.
- To validate the list_price column in “Transactions”, we change the format to “currency” format.
- To validate the product_first_sold_date in “transactions”, we change the format to “short date” format.

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

Regards,

Shruti Shah