

Dear Jack Maxwell,

I am Chaitanya Mundle from the KPMG Data Analytics team. I hope you are doing good.

Thanks a lot for allocating us with three datasets from Sprocket Central Pty Ltd. The tabular representation of the summary specified below emphasizes on some of the key data quality issues that we came across while working on these datasets. Kindly let our team know about any queries in our presented work.

Summary Table

Datasets → Issues ↓	Transactions	Customer Demographics	Customer Address
Accuracy	Profit section missing	Incorrect DOB, with 87 null values and age not displayed	
Completeness	Online Order and Brands have Null values (94 and 48 respectively)	The job title has 505 blanks with no ID for customers	Incomplete customer ID
Consistency		Gender not consistent throughout the column	State not consistent throughout the column
Relevancy	Not filtered Cancelled status order	Missing filter for passed away customers	
Validation	Formatting inconsistency in list price and sold date column and 48 Null values respectively		

- Accuracy Issues:

1. The profit section was missing in the "Transaction Table" and inculcating a profit section would have an advantage and flesh out more clarity about the data.
2. Blank or null values were discovered in the DOB and job title columns of the "Customer demographics table".
3. Eliminating the blanks would always help in focusing more on meaningful data.

- Completeness Issues:

1. Supplementary customer ID's were not complete in the "Transactions" and "Customer Demographic" tables.
2. Due to this the data may be more skeptical and skewed and will result in more amount of errors in the analysis. This completeness issue is pointed out to cross-check the future syncing of the excel dataset.

- Consistency Issues:

1. Inconsistencies in the gender column of "Customer Demographic" and for the "Customer Address" table.
2. This inconsistency includes filters like "M", "Male", "F", "Femal" etc. similarly for the states column too.
3. The solution here would be to just provide electricity via drop-down choices.

- Relevancy Issues:

1. Absence of relevance in the status column of the "Transaction table" and default column for "Customer Demographics" table
2. The recommendation here would be checking for inconsistent data and make it comprehensible.
3. No need for "canceled order status" for the analysis.

- Validation Issues:

1. Improper formatting of List\_price and product sale date in the "Transactions table".
2. The solution here would be to set the price so that it would be already adjusted with the decimal places when new data arrives.

That summarizes all the queries/issues which were discovered during the initial data quality analysis. The approach implemented by our time is fairly simple and will improve the quality of the dataset as a whole for any further analysis and will also abate the complexities in it.

I hope that we would get feedback the same.

Regards,  
Chaitanya Mundle.