

Dear Client,

I hope this email finds you well. I wanted to provide you with an update on our progress regarding the data quality assessment of the datasets you provided from Sprocket Central Pty Ltd. Our team has thoroughly reviewed the datasets, and I would like to share our findings along with some strategies to mitigate the identified data quality issues.

We have analysed the quality of the raw data and found several quality issues in the dataset. we believe that these issues needed to be addressed before we continue this further processing of the data.

Datasets	Accuracy	Completeness	Relevancy	Validity	Consistency
Customer Demographic	Age: Missing DOB: Inaccurate	Job title: Blanks	Default Column: Irrelevant		Gender: Inconsistent
Transaction	Profit: Missing	Online orders: Blanks	Cancelled status: Filtered Out	List price and date product sold: Format	
Customer Address					Status: Inconsistent

Below we’ve suggested ways on how you can mitigate those data quality issues in the future.

1. Accuracy Issues:

Customer Demographic Dataset:

We noticed discrepancies in the date of birth for some customers. There are instances where birth dates appear to be entered incorrectly, which affects the accuracy of the demographic information.

Recommendation: Filter out outliers in the DOB column and double check the age and the profit columns to ensure there are no missing values.

2. Completeness Issue:

Customer Demographic Dataset: Some cells in the job title were blank.

Transaction Dataset: Both the online_order and the brands were blank.

Recommendations: Double checking the columns to ensure that the columns don't have blank spaces.

3. Consistency Issue:

Customer Demographic Dataset: Gender Column were Inconsistent.

Customer Address Dataset: State column were inconsistent.

Recommendations: Using selective option minimize manual entry which in return reduce human error.

4. Relevancy Issue:

Customer Demographic Dataset: Irrelevancy of data values in the default column.

Transaction Dataset: Irrelevancy of data values in the order_status column.

Recommendations: Recheck and delete all the irrelevant data values.

5. Validity:

Transaction Dataset: Wrong data format in both the list_price and product_first_sold_date.

Recommendation: Ensure that the Column data values are in the correct format.

Please look into above-mentioned quality issues along with the recommendation changes. If all the suggestion are matched, we can proceed with further analysis of the data.

If you have any questions or require further clarification, please do not hesitate to reach out. We appreciate your collaboration in ensuring the data's quality and look forward to moving ahead with the analysis phase.

Thank You

Best Regards,

Rudresh Patidar