

Dear Client,

I am writing to you in response to the SP rocket central Pty limited datasets that you shared with our team.

Transaction Dataset

Customer Demographic Dataset

Customer address dataset

We have analyzed the quality of raw data and found several quality issues in the datasets.

We believe that this issues had to be addressed before we further process the data

	Accuracy	Completeness	Relevancy	Validity	Consistency
Customer Demographic	DOB : inaccurate values	Age column : missing DOB : Blanks Customer ID : missing	Default column : useless Deceased customer : filtered out		Gender : Inconsistent
Transaction		Profit column : missing Online order : Blanks Brands :blanks Customer ID : missing		List price : format Product sold date : format	
Customer address		Customer ID : missing			States : Inconsistent

Below, we have shown how you can mitigate those data quality issues in future.

Data Quality Issue and Recommendations

Data Accuracy issues

Incorrect Values in some columns

Recommendation : Check and remove any outlier's or values with different datatype.

Data Completeness issues :

Missing columns : many columns such as age and profit that were required for analysis were missing.

Recommendation : Ensure the data is complete for the required analysis

Blank cells/missing data : many cells were null.

Use data validation check to ensure that no fields are left blank while inputting the data. If there's any blank cells still left then filter those rows out.

Data Validity Issues :

Many columns which were representing currency and date were in wrong format.

Recommendation : After data merging, check the format of each column and keep the format consistent for all the values in a column.

Data Consistency issues :

Inconsistent nominal qualitative data in columns such as gender, states

Recommendation : Use consistent names for same attributes.

Relevancy issues :

Irrelevant data in columns or columns themselves such as default column.

Recommendation : Recheck and delete all irrelevant data . Also, filter out all irrelevant data from a column

Please, look into above mentioned quality issues along with the recommended changes.

Regards,

Aashish Telgote