

Dear Client,

I am writing to you in response to the Sprocket Central Pty Ltd Datasets that you shared with our team.

- Transaction Dataset
- Customer Demographic Dataset
- Customer Addresses Dataset

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

Datasets	Total Number of Records	Unique ID Records
Transaction	20000	3427
Customer Demographic	4000	4000
Customer Addresses	4003	4003

We believe that these issues needed to be addressed before we continue this further processing of the data.

	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 Brands: 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.

**a. Accuracy Issues:**

- **Customer Demographic Dataset**

Some cells in the DOB column had inaccurate and some of the cells in the age columns had missing values.

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

**b. Completeness Issues:**

- **Customer Demographic Dataset**

Some cells in the job\_title were blank.

- **Transaction Dataset**

Both the online\_order and the brands were blank.

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

**c. Consistency Issues:**

- **Customer Demographic Dataset**  
Both the Female and the Male genders were inconsistent.
- **Customer Addresses Dataset**  
Both the South New Wales and the Victoria states in the state's column were inconsistent.

**Recommendation:** Using selective options minimizes manual entry which in return reduces human error.

**d. Relevancy Issue:**

- **Customer Demographic Dataset**  
Irrelevancy of data values in the default column.
- **Transaction Dataset**  
Irrelevancy of data values in the order\_status column.

**Recommendation:** Recheck and delete all the irrelevant data values.

**e. 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 the above-mentioned quality issues along with the recommendation changes. If all the suggestions are matched we can proceed with further analysis of the data.

Thanks & Regards,

Rajdeep Dalal