Hello,

This is in regards to datasets provided by Sprocket Central Pty Ltd for the data analysis purpose.

Thank you for providing the datasets for the analysis.

In the first step towards the analysis, we have performed some data cleaning on the datasets and came up with details. The summary table below highlights few key abnormalities we discovered in the data. Please have a look and let us know if you have any further question for us.

Data Count:

Table	Total records
Transaction	Total – 19273 Total Record without blank job titles: 3491 Unique Customer ID - 4000
Demographics	Unique Customer ID - 3997
Customer Address	Total-records - 3999

Summary:

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customers Demographics	Age: Missing DOB: Inaccurate data type	Job Title: blanks	Gender: Inconsistency	Deceased customers filtered out	Column with irrelevant information removed	
Transaction	Profit: missing Transaction month: missing	Brands: blank Online Order: blank	Price and date format: incorrect		Order_status: cancelled filtered out	Sold_date : format Price: format
Customer Address			States: inconsistence			

Changes in the dataset:

- Converting columns to their specific data types.
  - Dates to short dates
  - Text to numbers where needed
- Removing few blank values in brand, job titles,
- Creating age, profit and transaction months columns for analysis
- Cleaning gender, states columns
- Deceased customers filtered
- Date and price columns format with correct data types
- Gender column cleaned with two values

This summarized the data quality issues found in the first stage of data analysis. Above steps will not only improve the analysis but will also help in finding more accurate insights and output.

Please let us know if you have any further questions for us regarding data quality.

Thank you.

Regards,
Siddharth badal