Dear Sprocket Central Pty Ltd,

Thank you for providing us with the dataset from Sprocket, we have reviewed the dataset and summarised the following data quality issue with the dataset. We have further given our comments about how we have tackled these data quality issue and laid out a plan to move forward with the data cleaning.

Summary Table:-

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	• DOB: Inaccurate • Age Missing	• Job Title: Blank • Customer ID: Incomplete	• Gender: Inconsistency	• Deceased customers: Filter out	• Default Column: Delete	
Customer Address		• Customer ID: Incomplete	• States: Inconsistency			
Transactions	• Profile Missing	• Customer ID: Incomplete • Online order: Blanks • Brand: Blanks			• Cancelled Status Order: Filter out	• List price: Format • Product sold data: Format

As per the preliminary task, in the above-mentioned list, we have analysed the quality of the raw data and we found multiple quality issues that need to be addressed. Also, we have suggested recommendations to mitigate the quality issues and improve the effectiveness of the data.

Accuracy issues:-

• DOB was inaccurate for "Customer demographic" and "Transactions" was missing an age_column and a profit column.

Mitigation:- Remove anomalies / Outliers from the DOB

Completeness:-

• The "Customer Address", "Customer Demographic", and "Transactions" fields had additional customer_ids that fluctuated.

 Mitigation:- Filter all customer_ids between 1 and 3500

• The fields "Customer demographic" and "Customer Address" in job_title are both null.

Mitigation:- For gender, filter all "M"s under the "Male" category and all "Female"s and "F"s under the "Female" category. Change all instances of "New South Wales" to "NSW" and all occurrences of "Victoria" to "VIC" for States.

Currency:-

• Individuals with the 'Y' prefix in deceased_indicator are no longer "Customer Demographic" customers.

Mitigation:- Remove consumers who ticked "Y" in the deceased_indicator.

Relevancy:-

• Order_status for "transactions" and default_column for "customer demographic" are not relevant or understandable.

Mitigation:- Deleted metadata in default_column. Filter out "canceled" order_status.

Validity:-

• List_price and product_sale_date formats for "Transactions"

Mitigation:- Format list_price to currency and product_sale_date to start date format.

That sums up all the data quality issues discovered through the first stage of the data quality analysis. Please look into the above-mentioned quality issues along with the recommended changes to ensure the consistent quality of the dataset across all the tables. If all the suggestions are matched we can proceed with further analysis of the data to find some suitable insights for the company.

Please let me know if you've queries regarding mitigation or any data quality issues identified.

Kind regards,
Amey Desai