Hi Sprocket Central Pty Ltd,

Thanks for providing the dataset. We have reviewed the dataset and summarized the following data quality issues with the dataset.
PFB table highlighting the key data issues

Sheet Names	Completeness	Consistency	Relevancy	Validity
Transactions	➢ online_order: blanks ➢ brand: blanks	➢ list_price: format ➢ standard_cost: Inconsistent ➢ product_first_sold_date: format		
New Customer List	➢ DOB: blanks ➢ job_title: blanks ➢ job_industry_ category: n/a	➢ property_valuation: Inconsistent ➢ past_3_years_bike_related_purchases: format ➢ postcode: format		
Customer Demographic	➢ DOB: blanks ➢ job_title: blanks ➢ job_industry_ category: n/a	➢ gender: spelling	➢ default: delete	➢ deceased_ indicator: filter
Customer Address		➢ state: spelling		

PFB in depth descriptions of data quality issues discovered and the methods used to mitigate them. Recommendations and explanations have also been included to avoid further data quality issues in the future.

Following recommendations will improve the accuracy of the data used to influence the business decision of Sprocket Central Pty Ltd in the future.

Completeness issues

- Blanks in the following columns: DOB, job_title, job_industry_category, online_order, brands
- N/A in job_industry_category

Mitigate

Filter out the 'Blanks' and 'N/A' from the above-mentioned columns.
Recommendation: Provide drop-down option using data validation for the mentioned columns to easily filter.

Blanks, n/a are treated as incomplete data and can skew analysis results. The addition of drop-down option will allow to have more complete data
and will result in more accurate results.

Consistency issues

- Format issues in the following columns: list_price, product_first_sold_date, past_3_years_bike_related_purchases
- Inconsistency in decimal places in standard_cost and property_valuation columns
- Inconsistent spelling for gender and state

Mitigate

Convert list_price to Currency format.
Convert past_3_years_bike_related_purchases from Text to Number format
Convert product_first_sold_date to short date format
Convert property_valuation from Text to Number format and make decimal places consistent.
Convert standard_cost to Currency format and make decimal places consistent.
Replace 'F', 'Femal' with 'Female' and
'M' with 'Male'
for gender
Replace 'New South Wales' with 'NSW' and
'Victoria' with 'VIC'
for state

Recommendations

Set up columns so that formats such as price, date, decimal are already in place while entering the data
Create drop-down options for 'Male', 'Female' and 'U' using Data Validation
Create drop-down options for all state abbreviations using Data Validation

Allowable values will make the data to be interpreted more easily. Formatting into price and allowing for either 2 or 3 decimal places consistently will increase readability. This will reflect positively on speed and accuracy of analysis for business decisions Use Data Validation for creating dropdown options which minimizes manual entry and human error. Allows for increased consistency of terminology.
Gender identity can be a sensitive topic, proceed with caution while creating options.

Relevancy issues

- Junk data in default column not relevant for analysis

Mitigate

Deleted column named default from Customer Demographic worksheet.
Recommendation:
Check for incomprehensible metadata and delete or format to make it comprehensible.

Columns having junk data which is not related and not useful for our analysis can be dropped.

Validity issues

- People that are ‘Y’ in the deceased indicator are not current customers for Customer Demographic

Mitigate

Filter out customers checked ‘Y’ in deceased indicator
Recommendation:
Can be difficult to check for deceased customer, but once this information is received the data should be updated accordingly.

Deceased customers are not current customers, removing them from the list increase data validity and will result in more accurate estimates in the future

---

That summarizes all data quality issues discovered through the first stage of data analysis. The mitigation strategies are suggested are simple yet effective ways of improving data quality for the future analysis. This will improve the analysis output. Please let me know if you have any questions regarding mitigation or any data quality issues identified

Thanks and Regards,
Vishnu K