Dear Manager,

Firstly, thankyou for providing the three datasets for, Sprocket Central Pty Ltd. This email comprises the key data issues that were discovered within the respective datasets. Kindly refer to the below table that gives a summary on the data quality issues.

Summary Table:

Table name	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity	Uniqueness
Transactions	Missing Profit	**Blanks removed from:** Transaction_date, online_order,order_status and Brand				**Product_first_sold_date:** Converted to Date format from number. **List_price:** Converted to currency	
NewCustomerList		**Blanks removed:** Last_name, DOB, Job_title				**Values are converted to numbers:** Past_3_year_bike_related_purchase, Property_valuation	
Customer Demographic		**Blanks removed:** DOB, Job-title	**Gender:** Inconsistency	**DOB:** Error Data	**Default column:** Irrelevant		
Customer Address							Renaming Duplicate state entries

Detailed Data Quality issues and Strategies to mitigate the issues:

Data Table - Transactions
1. Accuracy:
   We can find the Standard_cost and list_price for the respective products, but we don't have the profit for further analysis. Curtail Standard_cost from List_price to calculate the profit.
2. Completeness:
   Using filters, We can remove the blank rows in the Transaction_date and online, order_status and Brand columns to get a complete dataset.

3. Validity:

        In the column, Product_first_sold_date We can note that the data is in the number format. Which can be converted into date using the format tool.

Data Table- NewCustomerList

1. Consistency:

        Using filters, We can remove the blank rows in the columns Last_name, DOB, Job_title to get a complete dataset.

2. Validity:

        To calculate the Total and average sales of the column past_3_year_bike_related_purchase, we need the attributes to be numbers, instead of values. So, those values are converted to numbers.

Data Table - Customer Demographic

1. Completeness:

        Using filters, We can remove the blank rows in the columns DOB, Job-title.

2. Consistency:

        Incorrect gender attributes like M and F instead of Male and female. We have to replace those values using Find and Replace comments.

3. Relevancy:

        In the column named Default, the data is irrelevant for the analysis. We can delete the irrelevant data.

4. Currency:

        Incorrect DOB that is an outlier of the dataset. This data can be removed from the dataset.

Data Table - Customer Address

1. Uniqueness:

        Converting the New South Wales and Victoria in the state column of the dataset into their respective acronym to maintain the distinct values of the data.

This gives you a brief about the Data Quality issues present in the given dataset. It also includes the strategies that will mitigate the data in a simple and efficient way. These strategies not only resolve issues of this dataset, which can also be used with various datasets. Please let me know, if you need further assistance with the mitigation and Data Quality issues identified.

Regards,
Sowmiya Muralidharan.