SPROCKET CENTRAL CUSTOMERS ANALYSIS

Quality Assessment of Sprocket Central Pty Ltd.'s Data.

Below are my findings and recommendations based on the Data quality dimension used for assessing the data;

Table Name	Number Of Records	Distinct Customers ID	Date Data Received	outlier
Transaction	20,000	3494	20,000	1
Customers Demography	4,000	4,000	3912	507
Customers Address	3999	3999	none	3

1. **Different data:** There are more Customer IDs in the Customers Demography table than in the Transaction table although customer IDs appear in the transaction table but are not in the customer's Demography table. The analysis will be based on the customer Id from the customer's demography table.

   3,493 customers id from the customer demography table match **19,997** customers Id from the transaction table.

   Customer Id (5034) is not in the customers' demography table and was removed

   507 customer Ids in the customer demography table that have no information in the transaction table were removed.

2. **Correct Values:** in the new customers' list table, the use of 'U' as a type of gender is not explicit because one can find it hard to understand what it is used to represent. Also, the Date of the birth column in customers' demography shows that someone was born in "1843" which is likely to be an input error. The correct value needs to be provided

3. **Completeness:** there seem to be a lot of blank rows across all the tables in the data while some rows might be removed from the data, it might be useful to contact some customers and request their Last name, date of birth, and job title as this may be useful in the analysis.

4. **Consistency:** the gender column in the customers' demography table, was represented with F, Female, Male, M, and U. this shows inconsistency in the column. A consistent format needs to be adopted throughout the column

5. **Currency:** in the transaction table, the "standard cost" column was formatted as currency but the "list price" column was not formatted.

6. **Relevancy:** in terms of relevance, the "default" column in the customers' demography table is unnecessary and needs to be removed. Information about some deceased customers is also not needed for the analysis.

7. **Validity:** the data validity needs to be performed on the "product_first_sold_date" in the transaction table to allow for only date input into the column. Also, the "Postal code" and "Property Valuation" columns were formatted as text instead of general and number respectively.

8. **Uniqueness:** the transaction table contains some duplicated records which need to be removed to be able to get the real correct insights during the analysis.

Data Cleaning and Transformation

The data was loaded to Excel Powe Query and the following transformation was made

Transaction table

Null values from the product brand column totaling 197 were removed from the transaction table

Customers Demography table

- Removed default column
- Merged first and last name as full name
- Replaced the gender column with the correct values
- Null values in the Job title column were replaced with N/A
- Replaced N with No and Yes with Yes in the Deceased column

Created a new Table by merging Customers' demography and customers' address tables.

Created a general calendar table

The cleaned data was loaded into the Data model and a connection only was created to the queries.

Data Modelling in (Power Pivot)

- Created a relationship between the Customers Table and the Transaction table using the customers' ID keys
- Created a relationship between the Transaction table and Calendar table using the date as keys

Data Analysis Expression (DAX)

- Dax was written to create new measures such as the Total Orders, Average Orders, Total Revenue, Profits, and Bike Brands
- Age column was added to the customers' table to calculate the current age of the customers
- Customer's Ages were grouped into a new column

Dashboarding

Pivot tables were used to summarise the data

Various charts were also used to represent the insights.

Insights

- The top 10 customers with the highest revenue purchased bikes ranging from 89 - 99 bikes each in the past 3 years.
- Female Customers account for more than 50% of the total orders while 2.3% of customers' gender is unknown.
- The mass customer segment contributed 50% of the total orders
- The highest Orders were made by customers between the Age group 35 and 60 years old
- Solex is the most purchased brand of bike accounting for 21.5% of the totals while Norco Bicycles is the least ordered bike brand.
- A larger number of the customers reside in New South Wales.

Recommendations

- New Customers who have purchased more than 80 bikes in the past 3 years should be prioritized.
- More focus should be on new customers that reside in New South Wales as they are likely to drive more sales for the company.
- new customers between the ages of 35 and 60 should be targeted because old customers of that age category contributed more than half of the total orders.
- New Customers from the manufacturing, Financial Services, Health, and Retail Industries should also be targeted as data showed that the old customers from these industries made the highest orders.