Hi Sprocket Central Pty Ltd,

Looking at the dataset that was provided, There were noticeable problems within that need to be resolved before we can go any further.

From the transactions spreadsheet, there are many records of transaction which are blank for product line, class and size. While this may be attributed to the actual product, these values shouldn’t be blank.

Additionally, under the online\_order column, there are also many blank cells, but this column should only be either YES or NO as an order is either an online order or it isn’t. There will be a need to check your records for all of these transactions.

Another problem noticed was that under the column product\_first\_sold\_date, there were values which were blanks. There seems to be correlations that other columns were also blank (standard\_cost, brand, product\_line, product\_class and product\_size) and these transactions have product\_id of 0.

In CustomerDemographics and NewCustomerList, there are values for DOB which are left blank or with the incorrect data type. These values are normally a datetime data type, and can be changed easily. However, for blank values, there is a need to check the records you have for those people and update the DOB manually.

Regarding people’s jobs and industries, there were many instances where people’s job title was left blank, yet their industry had a value. When job industry was indicated as n/a when the person has a job title. This is inconsistent as any job should belong to a particular industry.

The values in the gender column in CustomerDemographics are not named consistently. For example, you have F, Femal and Female all for the same gender. Inconsistent data types are present too. The past 3 years bike purchases should be a number type instead of text, postcode is written as both number and text types and property value has numbers with decimals and text. There is a need to streamline all the data values to the same data type to make it simpler to perform the data analysis.

The CustomerAddress spreadsheet is the last spreadsheet given. In the customer\_id column, some value were skipped (e.g. 3, 10, 22, 23, etc.), which begs to ask why these values were removed. The address column indicates that some people have a street number of 0, which is not possible. Finally, the state column has values which mean the same thing but are denoted differently (e.g. NSW and New South Wales), which can be easily fixed if the data values are streamlined, and there is an agreed format.

The aforementioned are the issues that have been uncovered in this spreadsheet, and the possible recommendations to mitigate these issues.

Sincerely,

Daniel.