Dear Jane Blow,

Please, read through the report I made regarding the data that provided for quality check from your company, Sprocket Central Pty Ltd. I found some errors, which I pointed out and gave ways on how I handled them. I received three datasets being Customer Demographic, Customer Address, and Transactions.

**Accuracy:**

**Customer Demographic –**

* *Problem:* The *DOB* column had some outliers. An example is the customer born in the year 1843.
* *Action:* Filter out the outliers and create another column, which shows the actual age of the customers to make analysis easier since we cannot tell by just looking at their DOB.

**Transaction –**

* *Problem:* There was profit column to show the company’s revenue over time.
* *Action:* Subtracting the *list\_price* from the *standard\_cost* to get the profit.

**Completeness:**

**Customer Demographic –**

* *Problem:* The *DOB*, *age* (which I created), *job title, job category* columns had many blanks.
* *Action:* Get rid of the blanks.

**Customer Address –**

* *Problem:* *Customer\_id* column was up to 4003
* *Action:* Filter it out

**Transaction –**

* *Problem:* The customer\_id column had a figure 5034, which was out of place. *Online\_order* column had blanks, same with *brand, product line, product class, product size, standard cost, product\_first\_sold\_date* columns.
* *Action:* Filter it out the *customer\_id* with 5034 and as well filter out the blanks.

*Note:* The Customer Demographic *customer\_id* had a max of 4000, so we will be using it as our benchmark for analysis.

**Consistency:**

**Customer Demographic –**

* *Problem:* The *gender* column was inconsistent. It had M, F, Femal, U.
* *Action:* Find and replace them with Male, Female, and Undecided.

**Customer Address –**

* *Problem:* The *state* column was inconsistent.
* *Action:* Find and replace New South Wales with NSW and Victoria with VIC.

**Currency:**

**Customer Demographic –**

* *Problem:* The *deceased\_indicator* column had just one Yes.
* *Action:* Filter it out since it is just one customer and it as well indicates the person is not a current customer.

**Relevancy:**

**Customer Demographic –**

* *Problem:* The *default* column had different characters in it and no sense could be made out of it.
* *Action:* Get rid of it.

**Validity**

**Transaction –**

* *Problem:* *list\_price* was not in currency format. The *first\_product\_sold\_date was not in date format*.
* *Action:* Format the *list\_price* column to currency so it will not just look like a random integer. Format for the latter to a date format.

As you can see, these are the issues and ways of resolving them. Carrying out these actions will be of tremendous help while running an analysis.

Please, let me know if you have any questions regarding this. I am here to be of assistance.

Regards,

Judith Jude