Dear [Client Name from Sprocket Central Pty Ltd],

We have assessed the quality of the data you have shared with us. The data includes some minor problems that can be addressed to move further on analyzing the data. Below, I have given the problems we have observed in the data:

## Missing Values

In transaction value data set, by column total of missing values is given below. It is better to address missing values in transaction for next steps. One easy way to address to problem is to drop any rows with missing values but another approach can be to replace these missing rows with a mean value or an expected value. It can be observed a huge portion of missing values are in correlation with the value “197”. These group of missing rows includes value of list price but no other description.

transaction\_id 0

product\_id 0

customer\_id 0

transaction\_date 0

online\_order 360

order\_status 0

brand 197

product\_line 197

product\_class 197

product\_size 197

list\_price 0

standard\_cost 197

product\_first\_sold\_date 197

Customer list also includes missing data, however that is to be expected in a data list with customer information, this missing data should not be removed or replaced with a guessed value.

## Standardizing Values

In customer demographic, it can be noted gender column includes the following values “F, Femal, Female, M, Male, U”. Here it is needed to replace “F” and “Femal” values with “Female”, “M” values with “Male” to standardize all the values to same format.

In similar case, the state values in Customer Address data set can be addressed in the same manner.

## Invalid Values

One of our Day-of-Birth value in Customer Demographic includes the value of “1843-12-21”. This value means there is a person with age over 150 which is not possible. It is advised to replace this date data with the mean of date data set.

In addition, in Customer Demographic, the column “default” includes values that are either corrupted or encoded. With current values under this column, the data is not useable.

We have already addressed these issues and can share with you the updated data set in excel or csv format.

Happy to be working with you,

Best regards,

Metin Ünlü

KPGM Data Analytics Team Member