Hi there,

Regarding the data sets provided by Sprocket Central Pty Ltd, it is necessary to first clean the data so that we can analyse the data properly and can have accurate results.

Looking and analysing the data properly in Excel I found the following data quality issues that can be treated to get accurate values.

1. Transactions – Here, we can delete the inappropriate values from each column so that our can be more accurate.

* Online Order – Under this column, there are False and blanks values present that should be removed first. They are providing us with incomplete data.
* Order Status- Removing the order status of the products which got cancelled would help to get clean data.
* Profit- Calculate the profit by using the values of list\_price and standard\_cost.

1. Customer Demographic – There are data values with different names so we can find and replace those words.

* Gender- It has been seen that Female are often written as “Femal”, “F” and Male is written as “M” in some rows so find and replace these words with Female and Male keywords.
* Age- As we are given with DOB only, so we can calculate the age of every customer and can delete the outliers. There is one ID with age 120 that is not practically possible so remove this outlier.
* Job Titles- check the unique job titles and delete the blank values.
* Default- The column with the name default can be removed, as is not providing us any such appropriate information.

1. Customer Address- The same error discovered in this sheet also.

* State- New South Wales and Victoria are also written as NSW and VIC. So replace these with NSW and VIC.

I have provided the excel sheet with the changes, I thought would help me to clean the data. Kindly see to it.

Thank You