Dear Sprocket Central Team,

After we analyzed your data in the KPMG team, we have identified multiple issues in your data in terms of quality:

In the transactions table, we had multiple empty rows in the online order, brand, product class, list price, standard cost & product first sold date columns & we had no currency for the list price value We also identified that the data format for the column product first sold date was wrong as it was in numbers format.

In the Customer Demographic table, we had empty rows in the last name for some clients, in the gender column, the values where not well organized as we have 6 different values being female F and female that indicates he female gender, M & Male that indicated the male gender and U which is an outlier. In the DOB column we had empty values and an outlier value which is someone born in 1843 and it is not possible for him to have an order in 2017 & in the job title we also had empty values.

In the last table we have Some unorganised data in the state column as we had NSW and New South Wales as separate values & VIC and Victoria as separate values.

We had no duplicates in the 3 tables.

To fix these problems we have the suggestions below:

For the transaction table, we can filter out the empty data as they are not values that can be replaced, we can assume that the currency for list price is $ because the currency for standard cost was in $, so we can add it to the list price column. We can also change the format of the column product first sold date to Date instead of numbers.

In the Customer Demographic table, we can leave the empty rows in the last name since it is not crucial data, we should combine the genders in the gender column to female and male only and remove the U gender type since it is an outlier, in the DOB we can remove the outlier, we can keep the empty rows since this also isn’t a very crucial data, in the job title we had empty rows but we can consider them unemployed.

In the customer address table, we can combine the NSW and New South Wales values in the state table and the VIC and Victoria values.

Regards,

Mohamad Ali Ghaddar,