**Dear Sprocket Central Pty Ltd,**

I’m Gurushiya SP, data analyst from KPMG. I would like to share the data quality issues that we have encountered in the data set that you’ve provided and the strategies to mitigate those issues.

**First worksheet: Transactions**

* The shape of the data was 20000 x 13.
* The fields “Brand”, “Product line”, “Product class”, “Online order”, “Product size”, “standard cost”, “Product first sold date” had missing values.
* The “Online order” field is filled with values ‘True’, ‘False’. Instead of that we can give make two tables one fully consists of online orders and the other for off-line orders.
* Changing all the optional fields to mandatory, is the best way to avoid null values.
* After deleting the missing values and segregating the online and offline orders in two separate sheets, the shape of the data is 9739 x 12 and 9706 x 12.

**Second worksheet: New customer list**

* The shape of the data was 1000 x 18.
* The fields “Last name”, “DOB”, “Job Title” had null values. After deleting the missing values, the shape of the data is 983 x 18.
* Make the above-mentioned fields mandatory to avoid such issues.

**Third worksheet: Customer demographic**

* The shape of the was 4000 x 13
* The fields “Last name”, “DOB”, “Job Title”, “Default”, “Tenure” had null values. After deleting the missing values, the shape of the data is 3912 x 12.
* Make the above-mentioned fields mandatory to avoid such issues.

**Fourth worksheet: Customer Address**

* The shape of the data was 3999 x 6
* There were no null values but we had data inconsistency in the field “state”. For example (New south wales is defined as NSW in some places)
* Creating a drop-down for the list of states can help avoiding these issues.

***Thanks***

*Regards*

*Gurushiya SP*

*Data Analyst at KPMG*