**Sprocket Central Pty Ltd**

*Having done a detailed review of the dataset sent across for analysis, I discovered the following problems with the data quality. Kindly Take a look at it and make necessary adjustments.*

Firstly, there are lots of missing data. Try to fill up as much as you can with the appropriate values. Missing data could be handled either by dropping rows or columns that contain them, or filling the gap with guess values, mean of data for numerical data, mode of data for categorical data and many other approaches.

However, this would reduce the accuracy of the analysis and as such, filling with the correct values is preferable.

**For Transaction Sheet;**

* The following columns has missing data:

1. 360 missing entries for the column "online\_order"
2. 197 missing entries for the column "brand"
3. 197 missing entries for the column "product\_line"
4. 197 missing entries for the column "product\_class"
5. 197 missing entries for the column "product\_size"
6. 197 missing entries for the column "standard\_cost"
7. 197 missing entries for the column "product\_first\_sold\_date"

* Luckily, for the data set in this sheet, all missing entry for columns 'G', 'H', 'I', 'J', 'L', and 'M' intersect and they are a total of 197 entries. Since the entire dataset contains up to 20,000 columns, we can easily drop this data. However, we require an update for the entries in the column "online\_order".

**For CustomerDemographic;**

* + The following columns has missing data:

1. 125 missing entries for the column "last\_name"
2. 87 missing entries for the column "DOB"
3. 506 missing entries for the column "job\_title"
4. 656 missing entries for the column "job\_industry\_category"
5. 240 missing entries for the column "default"
6. 87 missing entries for the column "tenure"
   * All missing entries for the column "tenure" intersects with that of the column "DOB"
   * The entry format ('F', 'Femal', 'M', 'U', before 'Male' and 'Female') for the column "gender" is inconsistent and needs editing.
   * There's a total error for entry format "default". The data supplied can hardly be used for analysis.

**For NewCustomerList;**

* The following columns has missing data:

1. 29 missing entries for the column "last\_name"
2. 17 missing entries for the column "DOB"
3. 106 missing entries for the column "job\_title"
4. 165 missing entries for the column "job\_industry" (filled in as "n/a")

* The format of entries for the column "state" needs to be clearly stated for better analysis. ('VIC', 'QLD' and 'NSW')
* The column has some entries "U" which I don't understand.
* For columns 'Q', 'R', 'S', 'T', and "Value", every entry ought to be floats and if possible, have same decimal place.
* The column "property\_valuation" does not need any data entry of type "float" (e.g. 6.00, 7.00, 8.00 and 11.00). They are all integers
* The columns 'Q', 'R', 'S', 'T' and 'U' are hidden and they contain data entries with no header. The values of columns 'Q', 'R', 'S' and 'T' interchange at odd clicks while I used the filter function.

**For CustomerAddress;**

* + The format of inputs in the column "state" in inconsistent and needs editing before analysis

*I’d be expecting an adjustment in the data set so I commence analysis. Cheers!*

Onyeasigbulem Arinze.