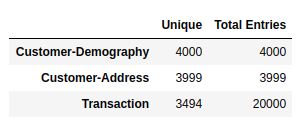
Dear Sprocket Central Pty Ltd,

We received your data sets for the customer demography, customer address, transaction and new customer. The Table below shows the summary of the data in the files. First column shows the unique entries and second column shows the total entries and the rows define the name of the files.



There were several quality issues that we came across and we will point out the methods we used to identify these issues and how they can be avoided in the future. Furthermore, we will also provide recommendations on how to improve it so the data can be gathered more accurately in the future.

We co-joined all the three data sets on the basis of customer id column, and there were total of 32 columns. The default column was found to be unnecessary with unidentified datatype and was removed from the main set.

Some Initial Assessment about the data set:

- There are total of 32 columns out of which 18 have ‘NaN’ columns and 13 have Fully filled columns.

- The DOB column has been converted to Age and 1 outlier exists which is 176.

- Two entries of Name repair in the name column

- Inconsistent data type for some attributes.

Recommendations:

- Ensure that same type of data is saved in each column

- If any entry has a lot of missing entries it should be avoided in the data set.

- Drop down options should be available to the user to avoid missing values including default values

- All tables should have uniform columns and entry options

We will further cleanse the dataset, for further analysis we have made a Jupyter notebook of the data set analysis which can be viewed on any standard browser and has been saved as HTML.

Kind regards,

Sarmast Bilawal Khuhro