Dear Client (Sprocket Central Ltd),

In the three datasets CustomerAddress, Transactions, CutomerDemographic and across these datasets we have come across some data quality issues.

The central part of this data quality lies around the CustomerDemographic data. The customer\_id’s in this sheet are subjected and data analysis plays on this dataset.

Issue:

Converted the product\_first\_sold\_date to datatime type attribute from float data type.

Recommendation:

Fix the attribute constraint data types while filling the table

Issue:

There are some columns where null values are stored, for now we taken out those null values and filled with the next immediate index values.

Recommendation:

For each and every column fix the default value and calculate the default value from the findings so that even if any value misses it doesn’t give much impact to the dataset.

Issue:

In gender regarding female, f, femal everything we have taken them as Female and male as male and U as unspecified.

Recommendation:

For the columns where it takes only some couple of inputs, for those columns fix a drop down list so that the value can be entered easily.

In the next step, we move forward taking the updated dataset from this step and we perform the three phases: Data Exploration, Model Development, and Interpretation. If there are any doubts regarding this, please get back to us.

Thanking you,

Thirumal Reddy Konatham

konathty@mail.uc.edu