Hi,

After receiving the three datasets from Sprocket Central Pty Ltd, the

exploration and data analysis are performed by our data analytics team. The quality of the data is assessed using various parameters. From the insights developed, the issues dealt with the respective datasets are identified and the strategies are found out.

The issues and insights of the dataset includes:

* There are null values present in the transaction, new customer list, customer demographic datasets. So, it is necessary to fill the missing values with 0, or mode, mean, or drop the columns to make the dataset complete.
* The date type column in the transaction dataset is of unstandardized and irrelevant format of datetime. So, it is necessary to convert the value type to the required format.
* There is a “default” column in Customer Demographic dataset which has garbage data and data with special characters. It is necessary to drop the column to make the dataset meaningful.
* Also, there are different categorical values in gender column of the Customer Demographic dataset. So, the values should be replaced with some standard consistent value.
* There are few unnamed columns in the new customer list dataset which shows the inaccuracy and incompleteness of data. These columns need to be replaced with meaningful column names or drop those columns.
* Many columns are filled with NaN values. So the meaningful informations are lost and may distrupt the correct interpretation of data.

Thanks and Regards,

Honey Sharma

Data Analytics team