**Vaahan Bima ‘CLTV’ Prediction**

**Approach :**

**Data Visualization:**

Data was derived from given sources and visualize. It is found that there is skewness present in the data. There is no direct linear relationship found in any independent variable and target variable. Pairplot and scatterplots are plotted to visualize the data.

**Model 1: Linear regression**

All the variables are included in building model. However, I got the very less accuracy of 16% on train data and 15% on test data. It might be due to skewness present in the data.

**Model 2: Linear regression with Non-linear transformation**

Different combinations were tried for nonlinear transformation. Boxcox, log transformation are tried for independent variables. Due to skewness in target variable, log transformation is also applied to target variable. Accuracy increased to 31% for training data.

**Model 3: Decision tree regressor**

Decision tree regressor is great way to handle nonlinearities in data. DT model is built with log transformation on target variable. There is significant increase in accuracy is observed.

**Model 4: Random Forest regressor**

Random forest is ensemble technique built of number of decision trees. Random forest model is built with nonlinear transformation of dependent variable. Accuracy of 36% is obtained on training data and on test data it is 35 %. As data is skewed, accuracy can’t be increased much. So this model is ran on unseen data and accuracy of 14% is obtained.