# Approach

**Task-**

To build a high performance and interpretable machine learning model to predict the CLTV based on the user and policy data.

**Steps for the model building**-

- Here we use linear regression model for prediction
- Import the csv file into jupyter server
- Check the null position(values)
- Check the data type of the columns and try to convert necessary column into integer
- Create dummy variables
- Identify the outlier in the claim amount column using boxplot
- Build linear regression model

- Intercept is [72663.2736867])
- Accuracy of the model is 43.27%
- r 2 is 0.1522