Linear Regression:

Here in this file we are going to perform Linear regression on our dataset. With the help of this we are going to find the attributes which are contributing towards the opening or closing of business.

So first let us import all necessery modules. Let us Load the dataset using pandas and see the columns in the dataset.

Let us see how many busnesses are open and close using countplot.

Now we will take *business_lattitude, business_longitude, business_stars, buness_review_count* as independant variable. i.e, we are going to predict how much these attributes contribute towards opening and closinf of businesses. *business_open* is our dependant variable.

Now let us split our data for training and testing using *train_test_split* module leaving 30% of the dataset for testing purpose. SMOTE is use for sampling purpose.

Now let us fit our training set for sampling. Now let us make all our training and testing dataset as a DataFrame.

Now let us apply Linear Regression for our training dataset and do the testing for our testing dataset. So our model will predict whether the business is open or close by taking all the independant variables as it's input. By doing that the accuracy we got is 56.35% using Linear Regression. Let us do classification report which will give the information about precsion recall and f1 score. Precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned.

We can also plot a confusion matrix for the model which gives True Positive, True Negetive, False Positive and False Negetive values.

The last but not the least is the contribution of each attribute for the model. By seeing this we can say *business_stars* has more impact i.e, 10.5% in predicting the openining and closing of the business. Location of the business is not imapcting the opening/closing of business that much.