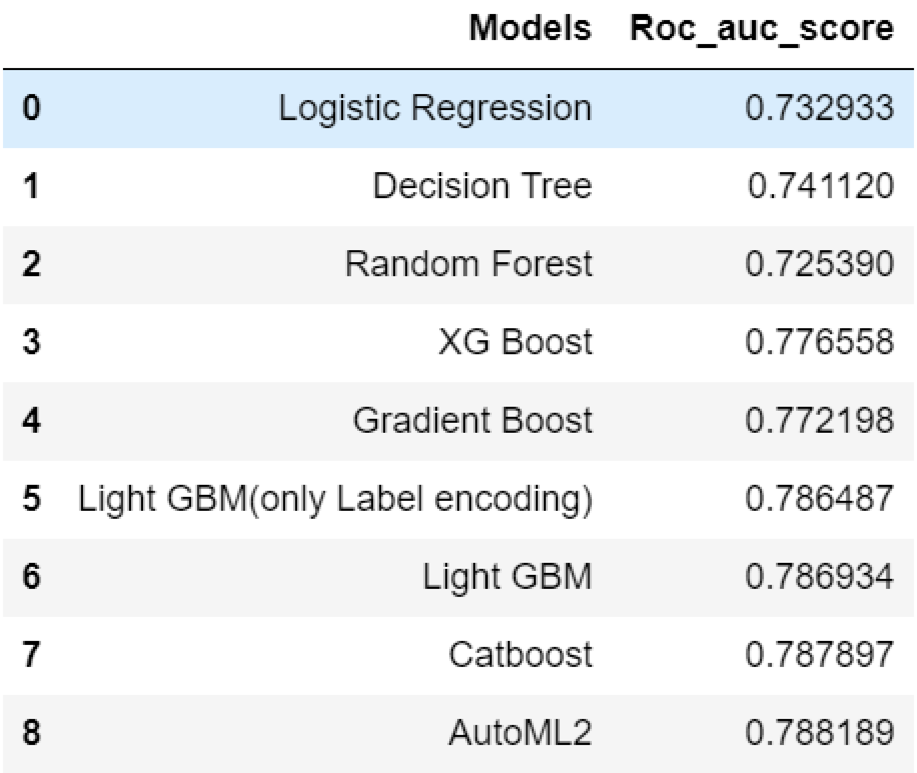
The Steps I have taken for classifying or finding the probabilities of customers who would turn on for credit card purchase from the Happy Bank are:

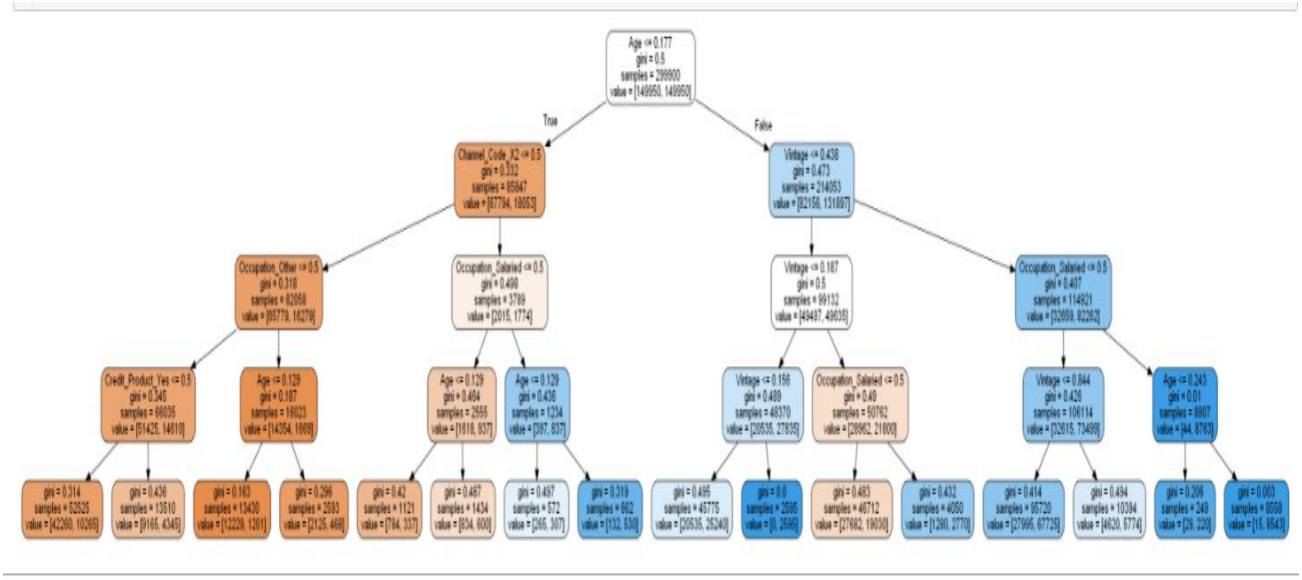
1. Firstly, I have viewed the datasets given in the train and test set to work on it
2. I have then checked for missing values in the dataset i,.e, Missing value analysis. In this step I have filled the missing values of categorical columns with mode. There are not many missing values. The percentage of missing values are very less can say as negligible amount. There are null values in only one column in the dataset and it comprises to 11% of the data.
3. After missing values imputation, I started cleaning the data before I proceeded to exploratory data analysis. In data cleaning I found the whether the datatypes are correct to the data or not. Next have gone through outlier analysis using IQR method.
4. In EDA, I have found the categorical and numerical features with their statistics interpretation using describe method and pandas profiling
5. Next, I have plotted some graphs in data visualization to find the insights from the data.
6. Next comes the feature engineering and feature selection out the new data formed from the raw data. Here I have done some normalization and one hot encoded (created dummies) the data.
7. Split the data into train and validation sets for training purpose and evaluating it. Also tried different ways here like balanced the data for some models and used unencoded values for some models according the possibilities present in the ML models.
8. Build the models using various ML algorithms like Logistic, Decision tree, Random Forest, XG Boost, Gradient Boosting, Cat Boost, Stacking Classifier, Microsoft AutoML.
9. Evaluated the models and performed hyperparameter tuning to improve the model performance.
10. Final model was decided based on roc_auc_score of various models.
11. Made predictions on the test set.
12. Submission file was created.

RESULTS:

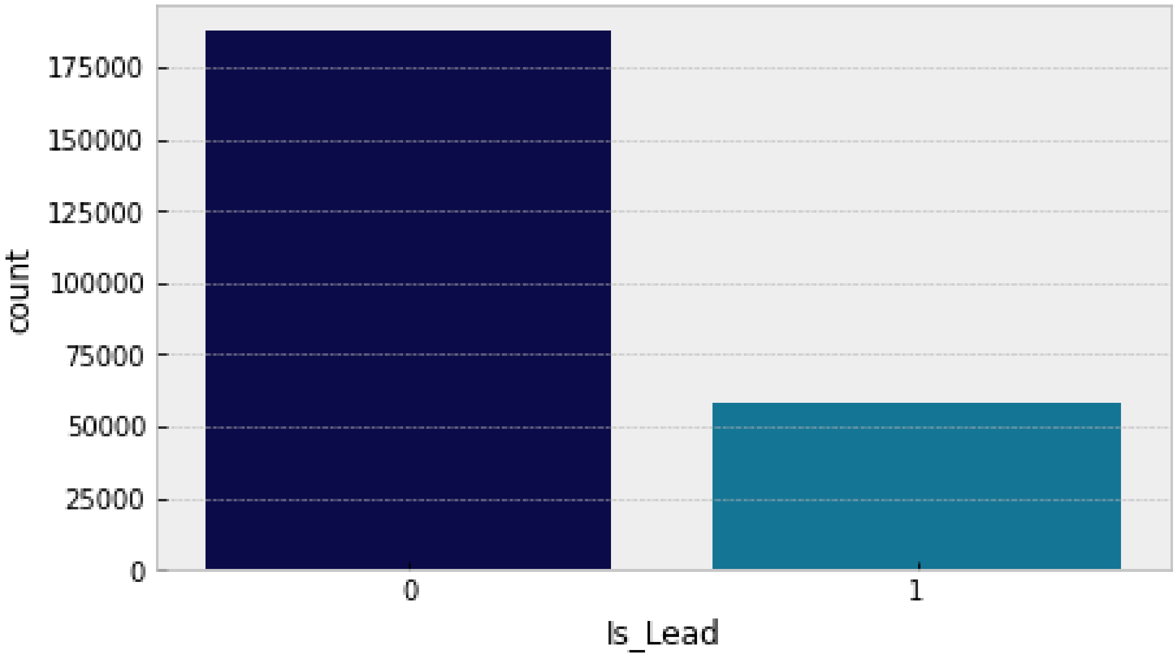


	Models	Roc_auc_score
0	Logistic Regression	0.732933
1	Decision Tree	0.741120
2	Random Forest	0.725390
3	XG Boost	0.776558
4	Gradient Boost	0.772198
5	Light GBM(only Label encoding)	0.786487
6	Light GBM	0.786934
7	Catboost	0.787897
8	AutoML2	0.788189

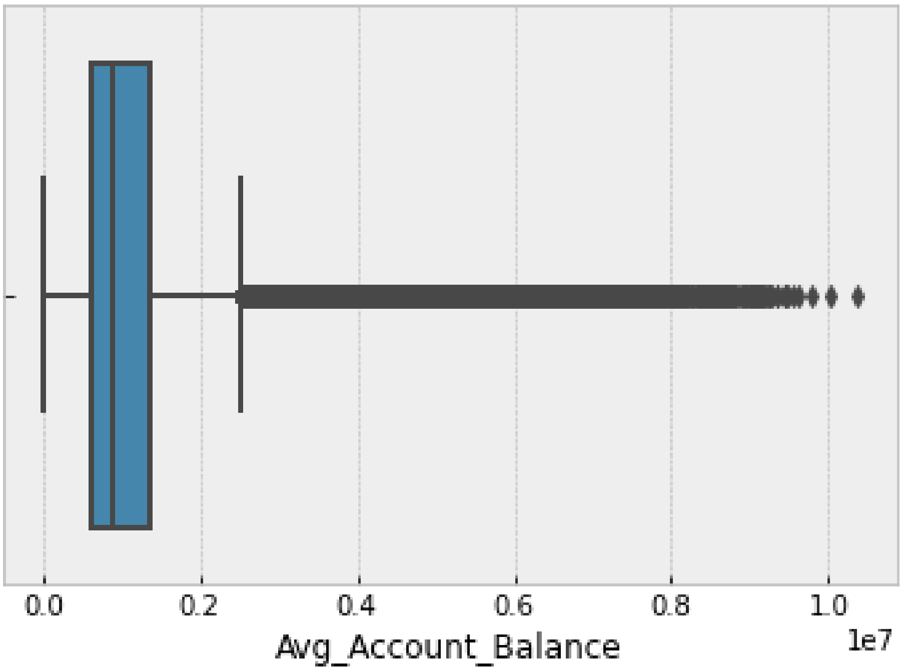
SOME DATA VISUALIZATION FROM THE DATASET:


Age <= 0.177 gini = 0.5 samples = 299900 value = [149950, 149950]
True
False
Channel_Code_X2 <= 0.5 gini = 0.332 samples = 85847 value = [67794, 18053]
Vintage <= 0.438 gini = 0.473 samples = 214053 value = [82156, 131897]
Occupation_Other <= 0.5 gini = 0.318 samples = 82058 value = [65779, 16279]
Occupation_Salaried <= 0.5 gini = 0.498 samples = 3789 value = [2015, 1774]
Vintage <= 0.187 gini = 0.5 samples = 99132 value = [49497, 49635]
Occupation_Salaried <= 0.5 gini = 0.407 samples = 114921 value = [32659, 82262]
Credit_Product_Yes <= 0.5 gini = 0.345 samples = 66035 value = [51425, 14610]
Age <= 0.129 gini = 0.187 samples = 16023 value = [14354, 1669]
Age <= 0.129 gini = 0.484 samples = 2555 value = [1618, 937]
Age <= 0.129 gini = 0.436 samples = 1234 value = [397, 837]
Vintage <= 0.156 gini = 0.489 samples = 48370 value = [20535, 27835]
Occupation_Salaried <= 0.5 gini = 0.49 samples = 50762 value = [28962, 21800]
Vintage <= 0.844 gini = 0.428 samples = 106114 value = [32615, 73499]
Age <= 0.243 gini = 0.01 samples = 8807 value = [44, 8763]
gini = 0.314 samples = 52525 value = [42260, 10265]
gini = 0.436 samples = 13510 value = [9165, 4345]
gini = 0.163 samples = 13430 value = [12229, 1201]
gini = 0.296 samples = 2593 value = [2125, 468]
gini = 0.42 samples = 1121 value = [784, 337]
gini = 0.487 samples = 1434 value = [834, 600]
gini = 0.497 samples = 572 value = [265, 307]
gini = 0.319 samples = 662 value = [132, 530]
gini = 0.495 samples = 45775 value = [20535, 25240]
gini = 0.0 samples = 2595 value = [0, 2595]
gini = 0.483 samples = 46712 value = [27682, 19030]
gini = 0.432 samples = 4050 value = [1280, 2770]
gini = 0.414 samples = 95720 value = [27995, 67725]
gini = 0.494 samples = 10394 value = [4620, 5774]
gini = 0.206 samples = 249 value = [29, 220]
gini = 0.003 samples = 8558 value = [15, 8543]

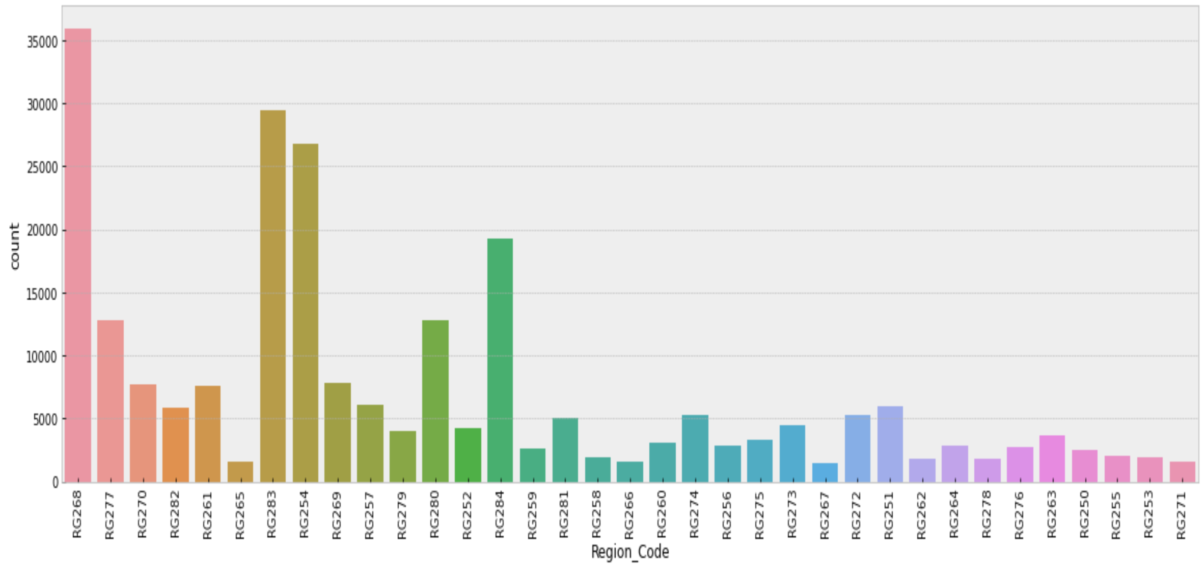
- This is Decision tree graph obtained from decision tree modelling



- This is the target imbalance present in the dataset which can we can overcome using oversampling , undersampling and some boosting models which can deal with such type of data.



- From this it can be observed that this is right tail distribution where there are many customers with balance in their accounts is greater than average balance. Also can say as there are more high income customers.



- These are the various region codes present in the dataset which are in the encoded format. It can also be observed that there are more customer or high number in RG268 code region.