The problem is to help the bank in identifying customers that show higher intent towards a recommended credit card.

On Train Set, there were 6 categorical features and 4 numerical features. On Test Set, there were 6 categorical features and 3 numerical features. While inspecting the data null values were present under Credit_Product feature. As this feature was of categorical type 'mode' was used to fill the missing values. By doing so it will lead to obtaining correct prediction.

During data exploration, under Credit_Product feature it was observed that majority of the customers does not have any active credit product. Under Occupation feature it was observed that majority of the customers were self employed and entrepreneur were having the least.

Do build a model for the prediction, ID feature from both the sets were removed as they weren't of much use for predicting. The Train Set were used for training the model. The target is Is_Lead.

All features except Is_Lead were copied to x and only Is_Lead was copied to y. Before splitting the data for training, all the categorical features were on hot encoded and dummies were created for them which is later merged with the original data set. This updated data set is then Train test split with test size 0.2 and random state 70. XGBclassifer was used instead of Random forest classifier as the XGBclassifer has got less prediction error than random forest. Also roc_auc_score obtained for XGBclassifer was higher than that obtained for random forest.

An roc_auc_score 0.6177 was obtained.