Code Approach:

1. Read the data from train and test data.
2. Checked for the null values in both train and test data.
3. Later I tried filling the **Credit_Product** null values with mode but later realised that it might not be the correct approach to fill with mode as there might be the possibility that the customer didn't take or the data is not updated from our side correctly. So finally filled with **Other** in both train and test data.
4. If we give NA in place of Other then the order of priority in Label encoding is NA, NO, YES i.e giving High priority to NA.
5. We can see that they are outliers in **Avg_Account_Balance**. So removed the values greater than 0.99 quartiles in the train data
6. Since **Avg_Account_Balance** has a wide range of continuous values divided into 10 bins for both train and test data.
7. Performed the label encoding for relevant variables for both test and train data.
8. Plotting the heatmap for verifying the correlation.
9. Removed the ID column from the train data.
10. Divided the data into train and test of size 0.7 and 0.3 respectively.
11. Later performed different decision tree models, gradient boosting and XGBoost models with different parameters using GridSearchCV and finally got the best AUC value by **XGBoost** with **learning_rate=0.1, max_depth=7, n_estimators=50, random_state=42.**
12. Now predicting the results on the test data and separated the ID and Is_Lead columns to separate CSV file i.e '**submission.csv**'