Problem Statement

The Portuguese Bank had run a telemarketing campaign in the past, making sales calls for a term-deposit product. Whether a pros pect had bought the product or not is mentioned in the column named 'response'.

The marketing team wants to launch another campaign, and they want to learn from the past one. You, as an analyst, decide to build a supervised model in Python and achieve the following goals:

- **Reduce the marketing cost by X% and acquire Y% of the prospects (compared to random calling), where X and Y are to be maximized**

- **Present the financial benefit of this project to the marketing team Data**

Analysis Outcomes:

I.----> Addressing to first goal of this problem the conclusive approach can be:

1. **Use of Recall as metrics:**
   Recall should be emphasized as well for both training and evaluating model as telemarketing campaign tries to identify more clients who will subscribe.

   Recall = True Positive / (True Positive + False Negative). It measures how many clients will be predicted accurately to

subscribe term deposit among all the clients who actually subscribe. Higher Recall value is preferred for a good model.

II.---> Addressing to second goal of this problem about the financial benefits of this project to the marketing team Data it can:

1. **Firstly there will be cost reduction in resource to be used for marketing campaign due to highly target marketing strategy.**

2. **Since marketing is done to target customer only there is high chance of retention means which leads to growth of business.**

3. **The gain and lift model metrics helps in finding the customer retention in the sample percent of total population which balances the X % and Y% which saves the marketing cost while increasing the percentage of business prospects.**

[I also have done some visualization within the data given.]

Model Summary/ Classification Report:

```
              precision    recall  f1-score   support

          no       0.96      0.94      0.95      7317
         yes       0.58      0.65      0.61       921

   micro avg       0.91      0.91      0.91      8238
   macro avg       0.77      0.80      0.78      8238
weighted avg       0.91      0.91      0.91      8238


array([[6887,  430],
       [ 323,  598]])
```

***** Due to time constraint I have not done further analysis and some of the task I have thought of for future improvements are:**

1. **Using the Cumulative Gain and Lift metrics for measuring the model performance which is very useful for this classification task in real world scenario.**

Gain : Measures the effectiveness of classification model
→ It's ratio between results with or without model
→ It shows performance in a portion of a population

Lift:
→ Shows how much more likely we are to receive positive responses than if we contact a random sample of customer.

2. Further visualization on the basis of prediction model and some in depth marketing , business domain statistical analysis