**Problem Statement**

**Business Use Case**

There has been a revenue decline for a Portuguese bank and they would like to know what actions to take. After investigation, they found out that the root cause is that their clients are not depositing as frequently as before. Knowing that term deposits allow banks to hold onto a deposit for a specific amount of time, so banks can invest in higher gain financial products to make a profit. In addition, banks also hold better chance to persuade term deposit clients into buying other products such as funds or insurance to further increase their revenues. As a result, the Portuguese bank would like to identify existing clients that have higher chance to subscribe for a term deposit and focus marketing efforts on such clients.

**Data Science Problem Statement**

Predict if the client will subscribe to a term deposit based on the analysis of the marketing campaigns the bank performed.

**Evaluation Metric**

We will be using [roc\_auc\_score](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html" \t "_blank) for evaluation.

**Objective of this notebook**

The main objective of this notebook is to take you through the entire working pipeline of a typical data science workflow while solving a problem.

We will be writing modular code in form of functions, the major advantages of this are

* Functions are reusable and at a later stage, the same task can be performed again by just importing the function without explicitly having to write the same piece of code again.
* This can save a good amount of time not only at work places but also at hackathons.
* So make sure you assemble all your lines of code within a function.

**The tasks performed below should serve as a good guide regarding the steps that you should go about while solving a problem. But kindly do not restrict yourself to only the tasks that have been performed in this notebook and feel free to bring your ideas,skills and strategies and implement them as well.**

**Word of caution**

This template is just an example of a data-science pipeline, every data science problem is unique and there are multiple ways to tackle them. Go through this template and try to leverage the information in this while solving your hackathon problems but you may not be able to use all the functions created here.

**Understanding the dataset**

**Data Set Information**

The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be subscribed ('yes') or not ('no') subscribed.

There are two datasets: train.csv with all examples (32950) and 21 inputs including the target feature, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014]

test.csv which is the test data that consists of 8238 observations and 20 features without the target feature

Goal:- The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).