**Proposal for capstone 2**

1. Banking markets are ultimately interested profits brought in by deposits coming from potential clients. However it is not easy to trace down the factors that impact the decision of deposits. As a data scientist I propose to the bankers that I build models to do machine learning using data collected from clients.
2. Therefore my clients are those managers in financial industry. I will provide them with tuned models in which important factors identified and bankers will steer these importance factors to better profits.
3. Fortunately there is very complete dataset available online collected by Portuguese banking institute. They basically acquired data from phone calls and this dataset include 41188 observations and 21 variables. Of course the original dataset needs cleaning up and wrangled prior to do feature selection.
4. There are basically 6 steps involved in this project.

1) That is, downloaded data is cleaned and duplicate observations are deleted. Missing data will be deleted or kept as a new class based on the contingency testing.

2) cleaned data is then used to do exploratory data analysis to find the relationship between the variable of interest and the target response, i.e., decision on deposit (binary, ‘yes’ or ‘or’).

3) the correlation between features is to be studied to find the features are strongly correlated and keep only one of those closed correlated features.

4) ML model including Logistic Regression, Random Forest and/or Decision Tree are to be tuned on the cleaned data after the data is split into train and test datasets.

5) metrics for each ML model is studied and to be compared and optimal model is to be chosen to real data.

6) make recommendations to banker such as important features to be tweak to achieve better profits.

1. The project is to be tackled by coding with Python. After debug and a good coding program is obtained, a report about the findings and all related figures and tables is to be prepared and a powerpoint slides deck is to be accomplished.