

# ACQUISITION ANALYTICS

## 1. Data Exploration and preparation

- All the code was done in Python 3.
- Dataset contains 20 features for 41000 clients.
- Dataset contained categorical and numerical. Categorical features were explored first.
- The young and the old were most likely to buy an investment product offered by bank.
- When we saw from the job perspective, student and retired were the ones who have responded positively. Again, falling into the bucket of young and old.
- There is a slight uptick in response rate for single customers; this can be attributed in part to the fact that people under the age of 20 have a very high response rate.
- Also, we notice that many prospects have never been contacted.
- Data is highly imbalanced with only 11% positive response.

## 2. Modelling- Logistic Regression

- Response variable is categorical, hence logistic regression is applied.
- Since machine learning algorithms only take numerical values, all five categorical variables are transformed into dummy variables.
- The dataset was split into train and test data.
- We take PCA variable as 16 for logistic regression.
- The sensitivity, specificity was determined to be 0.88 and 0.85 respectively.
- The AUC score was 0.93.
- The ROC curve area was 0.93.
- Lift chart was created.
- Total cost of acquisition was also counted.

## 3. Conclusion

- The most responsive customers possess these features:
- Feature 1: age < 30 or age > 60
- Feature 2: students or retired people
- The bank should re-evaluate the content and design of its current campaign, making it more appealing to its target customers.
- Telephone seems to be the most preferred mode of communication.
- Although the variable duration was not used for predictions, we can see that the higher the duration of the call, more is the chances of subscription.
- Also, the marketing team should reach out to people whom they have already reached out, as there is a higher chance of positive rate.