**Data Exploration**

The [dataset](https://archive.ics.uci.edu/ml/datasets/Bank+Marketing) comes from the UCI Machine Learning repository, and it is related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict whether the client will subscribe (1/0) to a term deposit (variable y). The dataset can be downloaded from here (<https://raw.githubusercontent.com/madmashup/targeted-marketing-predictive-engine/master/banking.csv)>.

**Brief of the dataset**

***Input Variables***[***¶***](http://localhost:8888/notebooks/A%20Logistic%20Regression_Python.ipynb#Input-Variables)

1. age (numeric)
2. job : type of job (categorical: “admin”, “blue-collar”, “entrepreneur”, “housemaid”, “management”, “retired”, “self-employed”, “services”, “student”, “technician”, “unemployed”, “unknown”)
3. marital : marital status (categorical: “divorced”, “married”, “single”, “unknown”)
4. education (categorical: “basic.4y”, “basic.6y”, “basic.9y”, “high.school”, “illiterate”, “professional.course”, “university.degree”, “unknown”)
5. default: has credit in default? (categorical: “no”, “yes”, “unknown”)
6. housing: has housing loan? (categorical: “no”, “yes”, “unknown”)
7. loan: has personal loan? (categorical: “no”, “yes”, “unknown”)
8. contact: contact communication type (categorical: “cellular”, “telephone”)
9. month: last contact month of year (categorical: “jan”, “feb”, “mar”, …, “nov”, “dec”)
10. day\_of\_week: last contact day of the week (categorical: “mon”, “tue”, “wed”, “thu”, “fri”)
11. duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y=’no’). The duration is not known before a call is performed, also, after the end of the call, y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model
12. campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
13. pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
14. previous: number of contacts performed before this campaign and for this client (numeric)
15. poutcome: outcome of the previous marketing campaign (categorical: “failure”, “nonexistent”, “success”)
16. emp.var.rate: employment variation rate — (numeric)
17. cons.price.idx: consumer price index — (numeric)
18. cons.conf.idx: consumer confidence index — (numeric)
19. euribor3m: euribor 3 month rate — (numeric)
20. nr.employed: number of employees — (numeric)

##### *Predict variable (desired target)*[*¶*](http://localhost:8888/notebooks/A%20Logistic%20Regression_Python.ipynb#Predict-variable-(desired-target))

y — has the client subscribed a term deposit? (binary: “1”, means “Yes”, “0” means “No”)

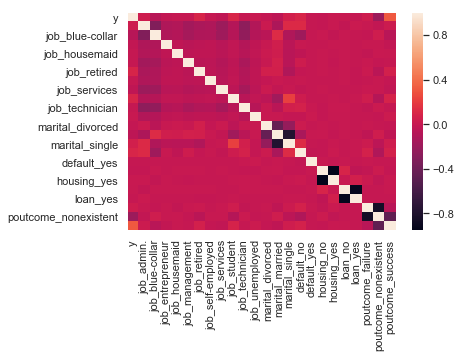
**Algorithms Used**

1. Standard Scaler: Used this to scale the variables in standard normal scale.
2. PCA: Used PCA to extract features from the dataset.
3. Logistic Regression: Used this binary classifier to classify whether the client will subscribe (1/0) to a term deposit (variable y).

**Metrices Used**

1. Confusion Matrix:
2. Classification Report:

**Correlation Heatmap between variables:**



**Tools Used**

* Spyder IDE
* Microsoft Excel
* Jupyter Notebook