# BU7142 Group Assignment

This group assignment is a data analysis project. You are given a data set about bank marketing. The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. This dataset is public available for research. More details can be found in below two publications.

S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology.

S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS.

Below is the information about the data set for this group assignment, and the data set is available on Blackboard in Excel format. The original data set in CSV format can be found at: https://archive.ics.uci.edu/ml/datasets/bank+marketing

**Number of records: 4521**

**Number of variables: 17.**

**List of variables:**

  1 - age (numeric)

  2 - job : type of job

(categorical:"admin.","unknown","unemployed","management","housemaid","entrepreneur","student", "blue-collar","self-employed","retired","technician","services")

  3 - marital : marital status (categorical: "married","divorced","single"; note: "divorced" means divorced or widowed)

  4 - education (categorical: "unknown","secondary","primary","tertiary")

  5 - default: has credit in default? (binary: "yes","no")

  6 - balance: average yearly balance, in euros (numeric)

  7 - housing: has housing loan? (binary: "yes","no")

  8 - loan: has personal loan? (binary: "yes","no")

 # related with the last contact of the current campaign:

  9 - contact: contact communication type (categorical: "unknown","telephone","cellular")

 10 - day: last contact day of the month (numeric)

 11 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

 12 - duration: last contact duration, in seconds (numeric)

  13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

  14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)

15 - previous: number of contacts performed before this campaign and for this client (numeric)

16 - poutcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")

17 - y - has the client subscribed a term deposit? (binary: "yes","no")

As a business analyst, you can freely explore the data set and find insights through data analysis. You need to complete a report to summarize all insights supported by data analysis and discussions. In your report, you should study and answer **at least three different research questions** with a word limit of **3000** (excluding figures, tables, references and appendix).

A typical report should include:

- **Introduction:** Briefly introduce the questions you are studying in the report with justifications.
- **Analysis:** Details of the analysis tools you use, new variables you generate, process for analysis and the results
- **Discussion/Conclusion:** discuss the insights from your analysis and the limitations
- **Reference:** list any reference you used in the report (using Harvard style)
- **Appendix (optional):** anything else that you would like to include to support your analysis

Below is a list of sample questions for your reference, please feel free to explore other questions that you find interesting:

- The factors affecting a client's decision to subscribe a term deposit
- The factors affecting a client's balance
- Impact of age, marital status or education on loan, deposit or previous marketing campaign outcome