File ​**bank-additional-full.csv** contains records relevant to a direct marketing campaign of a Portuguese banking institution. The marketing campaign was executed through phone calls. Often, more than one call needs to be made to a single client before they either decline or agree to a term deposit subscription. The classification goal is to predict if the client will subscribe (yes/no) to the term deposit (variable y).

Source: https://www.kaggle.com/brckalo/bank-marketing/data

**Variables:**

Bank client data

* age (numeric)
* job : type of job (categorical)
* marital : marital status (categorical)
* education (categorical)
* default: has credit in default? (categorical)
* housing: has housing loan? (categorical)
* loan: has personal loan? (categorical)

Related with the last contact of the current campaign

* contact: contact communication type (categorical)
* month: last contact month of year (categorical)
* day\_of\_week: last contact day of the week (categorical)
* duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, 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.

Other attributes

* campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
* 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)
* previous: number of contacts performed before this campaign and for this client (numeric)
* poutcome: outcome of the previous marketing campaign (categorical)

Social and economic context attributes

* emp.var.rate: employment variation rate - quarterly indicator (numeric)
* cons.price.idx: consumer price index - monthly indicator (numeric)
* cons.conf.idx: consumer confidence index - monthly indicator (numeric)
* euribor3m: euribor 3 month rate - daily indicator (numeric)
* nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target)

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

Please consider the following tasks.

* Build a binary logistic regression model. Dependent variable is “y”. Select not less than 6 predictors. Out of them at least 3 predictors should be categorical.
* Formulate all possible research hypotheses which will be verified during the data analysis.
* Specify the regression equation (the linear part of it).
* Assess the goodness-of-fit of the model.
* Describe the relationships between each predictor and dependent variable in detail. This is very important point of the task. We should understand clearly from the analysis which predictors influence the choice of the clients and how.
* Perform the model diagnostics. You should interpret all the diagnostic statistics discussed at the lecture.
* Make final conclusions about the results of analysis.

Please send the Jupyter Notebook with comments and answers to amelikyan@hse. The task should be done **individually**.The deadline for submitting the task is **11 May 14:00**.