Take-Home Technical Assessment

A Portuguese banking institution ran direct marketing campaigns to assess if clients would subscribe to a product (bank term deposit). The marketing campaigns were based on phone calls and more than one contact to the same client was required at times.

The *bank-direct-marketing.csv* dataset contains 41188 examples and 21 different variables, ordered by date (from May 2008 to November 2010). A description of these variables is provided in Table 1.

You are required to analyse these data and prepare a short presentation (15 minutes) explaining your analysis and conclusions. This might include any quantitative (exploratory data analysis or summary statistics) or qualitative (interpretation and context) insights gained from the data. Your analysis must include, but need not be limited to, the following points:

1) Develop a model (*Model-1*) that, given variables 1-20, predicts the probability of a client subscribing for the product;
2) Identify which variable is the most predictive of a client's subscription;
3) Identify any highly correlated pairs of variables;
4) Develop a second model (*Model-2*) that uses only 5 of the provided variables – explain your choice;
5) Compare *Model-1* and *Model-2* – explain your choice of metrics;

You are required to submit the Powerpoint document presenting your results as well as all the files containing your workings (Excel, R, Python, Stata, Matlab etc… - choose whichever you prefer) by email by **5pm on Monday** [redacted] to [redacted] on [redacted]

Please note this is a real dataset and so may require some cleaning. You don't need to describe the steps you have taken to do this in your presentation, but they need to be present in the files you submit.

You'll be asked to present your results during the first 15 minutes of your face-to-face interview.

Table 1 – Variables Description

#	name	description	type
Bank client Data			
1	age	-	numeric
2	job	type of job	categorical
3	marital	marital status	categorical
4	education	-	categorical
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"

Variables related to the last contact of the current campaign			
8	contact	contact communication type	categorical: "cellular","telephone"
9	month	last contact month of year	categorical
10	day_of_week	last contact day of the week	categorical: "mon","tue","wed","thu","fri"
11	duration	last contact duration, in seconds	numeric
Other attributes related to the campaign(s)			
12	campaign	number of contacts performed during this campaign and for this client	numeric
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"
Social and economic context attributes			
16	emp.var.rate	employment variation rate - quarterly indicator	numeric
17	cons.price.idx	consumer price index - monthly indicator	numeric
18	cons.conf.idx	consumer confidence index - monthly indicator	numeric
19	euribor3m	euribor 3 month rate - daily indicator	numeric
20	nr.employed	number of employees - quarterly indicator	numeric
Variables related to the product			
21	subscribed	has the client subscribed a term deposit?	binary: "yes","no"