Project I

This data relates to telemarketing phone calls to sell long-term deposits. Within a campaign, the agents make phone calls to a list of clients to sell the product (outbound) or, if meanwhile the client calls the contact-center for any other reason, he is asked to subscribe the product (inbound). Thus, the result is a binary unsuccessful or successful contact.

This study considers real data collected from one of the retail bank, from May 2008 to June 2010, in a total of 39883 phone contacts.  Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

You are going to work with part of the data as indicated by the variable CODE. You have to use only the data referring to your final digit in your registration number

The purpose of your project is to examine which variables contribute to a successful contact (the client subscribes to the product). The rest variables are potential candidates for examining the variable under study. Things that you need to consider are the following (but clearly not restricted to them)

* Which variables are important?
* Do we need to transform them (create new ones)?
* How good is the model?
* Are there any assumptions that need to be check carefully?
* In the current setting we care only for a model that explains the data. Your model does not need to be predictive. Comment on whether the model can be used for prediction or not

You have to write a detailed enough report to explain to your boss about that, explaining the model you used together with sufficient technical details on what you have done.   Use tables and plots that really contribute to your story. Explain what other information you may need. it is important that the report is self-contained and if somebody reads it could follow the arguments even with limited statistical knowledge.  
  
You have to use R.  The code needs to be attached. You must upload:  
  
a) a word or pdf file with your report,  
  
b) a txt file with the code used,  
  
c) a 5 slides powepoint presentation with your main findings as you have to present this to your boss.  
  
If you prefer zip them altogether.

Deadline Friday 28/2/2020, 23.55.  After that date you will have -2 point for every day of delay

About the  Data:

bank-additional-full.csv

Input variables:  
# bank client data:  
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'; note: 'divorced' means divorced or widowed)  
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')

# related with the last contact of the current campaign:  
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).

# other attributes:  
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')  
# 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)  
  
Output variable (desired target):  
21 - SUBSCRIBED - has the client subscribed a term deposit? (binary: 'yes','no')