Data Science Test – Advanced Loyalty Modelling, WolliesX

Assume you have been hired as a Data Scientist by a Bank and you are asked to help them to optimise the following marketing campaign.

The marketing team of the Bank wants run a telemarketing campaign, in which they tried to sell a Credit Card product to their existing customers by contacting them via phone calls with some offer giving a discounted interest rate. In this experiment, first they selected a random sample of customers who were eligible for this product (group A). Then out of these customers they targeted a random subset of customers (Group B) with this initial campaign. At the end of the experiment period, for all the selected customers (Group A), they recorded whether the customer took up this product or not. For the customers who naturally applied for this product (without being targeted by this campaign) no discount on the interest rate was offered.

They have formulated a dataset of the customers in Group A containing the variables given below, which has been given to you:

Variables in the dataset:

Bank client data:
1 - age
2 - job : type of job
3 - marital : marital status
4 - education
5 - default: has credit in default?
6 - housing: has housing loan?
7 - loan: has personal loan?

Related with the contact of the current campaign:

8. – campaign: targeted with the campaign or not
9 - contact: contact communication type
10 - month: contact month of year
11 - day_of_week: contact day of the week
12. - duration: duration of the phone call

Other attributes:
13 - previous: number of contacts performed before this campaign and for this client
14 - poutcome: outcome of the previous marketing campaign
Social and economic context attributes:
15 - cons.price.idx: consumer price index - monthly indicator
16 - cons.conf.idx: consumer confidence index - monthly indicator

#Response variable
17 - y - has the client taken the credit card product or not?

Now the marketing team wants to roll-out this campaign to the entire customer population who are eligible for this product. For this they want your help on analysing the above dataset they have created to optimise output of their telemarketing campaign by maximise the conversion rate and reducing the cost (call centre cost and offer cost – interest rate discount).

1. Describe your solution to the problem with any assumptions you may make
2. Implement your solution in Python with a well-commented out code, by emphasising the following points:
   - If you select to build a statistical model:
     - Specify reasoning behind the selection of your algorithm
     - From the given dataset, which variables you use to build such a model/solution
     - How would you do the hyper-parameter selection of the selected algorithm
     - What are the most important variables in the model/solution you built?
     - Any other considerations you may have to make given the characteristics of the dataset
     - How would you validate your model/solution
3. How would you use the proposed model in the actual campaign?
4. Now assume that the Bank decides to change the eligibility criteria of this product so that they could include more eligible customers in the campaign they want to run using your solution, where some of these customers could look different from the customers they ran the random campaign on:
   a. Do you think whether you would be able to apply the same solution/model you build above as it is with this new extended customer set?
   b. If you think you'll have to do any adjustments to your solution, specify why you would have to do that and what those adjustments are (no need of implementation)

Please email back your solutions (as a pdf/MS Word file), source code and any results/plots you may produce.