***Background:***

Because of recent financial problems and increased competition, European banks are under pressure to increase financial assets. This case study considers the telemarketing efforts for a targeted campaign by a Portuguese bank to increase deposit subscriptions. The promotions offered long-terms with good interest rates mainly through telephone calls between May 2008 and November 2010 with a total of 41,188 customer contacts. Many customers were contacted more than once. The average success rate was about 8%. The goal is to predict which customers are more likely to subscribe to increase the effectiveness of future campaigns.

***Instructions:***

The questions for this exercise are based upon the assigned reading for this class: Moro et al (2011) "Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology".  Please complete this reading before answering these questions.  In this exercise you want to design a model that will predict bank customer response to a cross-selling campaign.

Predictive bank customer response to a cross-selling campaign:

Links for the dataset and R script to perform the logistic regression are provided here:

[bank.R](https://canvas.cmu.edu/courses/9526/files/3508767/download?wrap=1)

[bank-full.csvPreview the document](https://canvas.cmu.edu/courses/9526/files/3508766/download?wrap=1)

An R script with an extended analysis that does both logistic regression and tree is provided here:

[bank\_ExtendedAnalysis.R](https://canvas.cmu.edu/courses/9526/files/3508765/download?wrap=1)

**This exercise is to be completed individually (not as a group).**

Please provide a clear, concise, and well organized essay organized in the form of a memo to Marketing Director charged with increasing bank deposit subscriptions.  Make sure that you at least addresses the required questions, but you are free to address other issues in the case as well.  The intent of the assignment is to have you think critically about the business problem faced in the case and how it can be solved through data mining.  Analyze the quantitative material in the case to support your answers.  Spend most of your time in defining and defending your recommendation for what should be done.

Good answers may require assumptions of facts that may not be presented in the case.  You are welcome to make these assumptions, but please state these assumptions and briefly justify why that are reasonable.  Also, you may use whatever resources you can locate to provide further information about this industry or the web in general.  Please reference your sources.

Your response must be typed, double spaced, with one-inch margins, and a 10 to 12 point font size, and must be a PDF file.  This writeup must not exceed 3 pages in length.  You may attach exhibits, tables, and/or graphs to support your arguments.  These supporting materials must be referenced in the text and do not count toward the 3 page limit.

The purpose of this memo is to persuade the Marketing Director about your recommendations.  You should assume that the Director is not a data scientist, but it a knowledgeable manager and is interested in using the data that is available to make better decisions.  In your memo you should carefully lay out your analysis and recommendations, but do it in a way that will appeal to a manager who is not trained in data mining.  This means that you need to give enough background about the data and your analysis to help convince the manager that your recommendation is a good one.  My learning goal for this exercise is not for you to provide me tables and output from R, but to use your analysis and model to interpret the data and tell a story that persuades the reader about motivating users to go “from free to fee”.

To help you write a good memo think about your memo as not a series of tables and analytical models, but a story.  This viewpoint is well illustrated in the following graphic from the blog by Michael Walker entitled “Data Scientists Draw Pictures and Tell Short Stories” ([http://www.datasciencecentral.com/profiles/blogs/data-scientists-draw-pictures-and-tell-short-stories (Links to an external site.)Links to an external site.](http://www.datasciencecentral.com/profiles/blogs/data-scientists-draw-pictures-and-tell-short-stories)).  Although this graphic is about social media metrics, it is still pertinent for our data scientists.  Carefully consider the four main points: capture, analyze, context, and package.  Michael’s suggestion is to “1) be brief, 2) be blunt, and 3) be gone”.

Another good resource for understanding how data scientists can create stories is “Tell a Meaningful Story with Data”

[https://www.thinkwithgoogle.com/marketing-resources/data-measurement/tell-meaningful-stories-with-data/ (Links to an external site.)Links to an external site.](https://www.thinkwithgoogle.com/marketing-resources/data-measurement/tell-meaningful-stories-with-data/)

Your memo will be evaluated on the following four dimensions:

Content:          How well did your memo answer one of the questions posed?

Analysis:          Did you support your recommendation?

Concept:          Did you demonstrate good knowledge of data science principles and data mining concepts?

Overall:           Is your recommendation persuasive?

***Required:***

Make sure you address at least these three points in your essay:

1. Explain the marketing problem that the bank faces in attracting new customers.  Name and discuss at least two alternative approaches it use to attract new customers?  Is a data-based approach a good one?
2. Using the data provided estimate a model that predicts whether a customer will subscribe to a new deposit subscription.  Make sure to support your choice of model.  What do you learn from the model about a good campaign?  How would you convince a non-technical manager that this is a good model?
3. Design a new cross-selling campaign using your model, be as specific as possible. What type of profit gain do you think you could hope to achieve?

***Dataset Information:***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | # | Variable | Description | Potential Values |
| Bank Client Data | 1 | age | Age of customer | 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 | 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 |
| Other Attributes | 13 | campaign | Number of contacts performed during this campaign and for this client | numeric, includes last contact |
| 14 | pday | 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 | numeric |
| Output variable | 17 | y | Has the client subscribed a term deposit | binary: “yes” or “no” |

**Rubric**

Case Rubric

| Case Rubric | | |
| --- | --- | --- |
| **Criteria** | **Ratings** | **Pts** |
| This criterion is linked to a Learning Outcome Content  Were all questions answered? |  | 10.0 pts |
| This criterion is linked to a Learning Outcome Analysis  Was an appropriate analysis (either quantitative or qualitative) conducted to support the answer? |  | 10.0 pts |
| This criterion is linked to a Learning Outcome Concept  Were the marketing concepts applied consistently? |  | 10.0 pts |
| This criterion is linked to a Learning Outcome Overall  Is the recommendation provided consistent with the answer in a compelling manner? |  | 10.0 pts |
| Total Points: 40.0 | | | |