## Scenario



You are part of a research and modeling team at National City Bank. You team has been asked to create a customer propensity model for a new product, specifically a line of credit against a household’s *used* car. Since the line of credit product is only in pilot, you are asked to identify the next 100 customers from a prospective customer list to contact. Bankers will call and direct mail will be sent to households your model identifies with the greatest probability of accepting the offer. Once your team has modeled and identified the customers, you must present your findings to the bank’s chief product officer. Once she/he feels comfortable with your proposal, marketing will begin its process.

**You are asked to examine the historical data from 4000 previous calls and mailings for the line of credit offer. Using this historical data, and any supplemental data, create a propensity model, evaluate it and identify by uniqueID the top 100 households to contact from the prospective customer list. Additionally, bank executives are eager to learn more about the customer profile for historical and top prospective customers. As a result, variable importance and sound EDA will aid the presentation. Your will need to turn in code and PowerPoint slides, a written supplemental, and a CSV of the top 100 customers by probability from your model(s).**

## Data

Source: <https://www.kaggle.com/kondla/carinsurance>

Supplemental data represents fictitious 3rd party data that the bank would purchase to improve the model’s accuracy.

## Example *Abridged* Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| HHuniqueID | Communication | LastContactDay | LastContactMonth | CallStart | … | Y\_AccetpedOffer |
| HHd4d0af8c72 | telephone | 28 | jan | 13:45:20 | … | 0 |
| HH8d3e87c164 | NA | 26 | may | 14:49:03 | … | 0 |
| HHdd53ef1db6 | cellular | 3 | jun | 16:30:24 | … | 1 |
| HH6fa0de6516 | cellular | 11 | may | 12:06:43 | … | 1 |
| HHeb436ca7cf | cellular | 3 | jun | 14:35:44 | … | 0 |
| HH5119beb3cd | cellular | 22 | may | 14:58:08 | … | 1 |

## Criteria for Success

|  |
| --- |
| Organization – Was the presentation well organized? |
| Delivery – Was the content delivered clearly and persuasively with the audience in mind? |
| Code Documentation – Was the data mined to support the conclusion? |
| Written Supplemental – Is it grammatically acceptable, organized and error free.  -Is the data supported clearly and coincides with the data, and narration while being contextualized with external information? |
| Data Mining Process – Overall, as a complete portfolio of work, is the topic interesting, organized, researched, supported and delivered effectively? Was CRISP-DM, SEMMA, or a similar workflow followed to organize the work if appropriate? |

## Another resource may be a public kaggle kernel

*Keep in mind this may not be helpful but code can be examined for additional ideas.*

<https://www.kaggle.com/kondla/simple-random-forest-on-insurance-call-forecast/code>

**Overall guidance:**

You are expected to submit R code, a voiceover narration of a live business presentation (can be a standalone video file, or inserted within the powerpoint) , a set of slides, and a written supplemental. If you submit documents with links to youtube videos or cloud drive files, you must ensure all links are accessible. Links set to private which are not viewable or downloadable will automatically result in 0 for that section of the rubric. You also need to turn in a CSV of the top prospective customers identified by your model. Remember as a pilot the audience wants more context about the model behavior so additional EDA on the identified prospective customers is encouraged.

**Delivery and Narration Guidance**

You are not allowed to use an ai avatar, or speech creation for narration. While this is certainly useful technology in many instances, the purpose of the business case presentation is to improve *your* presentation skills. In a business setting you will still be expected to articulate your findings and not send an avatar for this type of business meeting. As a result services like <https://elevenlabs.io/> or <https://www.heygen.com/> are not permitted. Your “boss” or “audience” in the case will not accept these technologies.

**Written Supplemental Guidance**

Submit a document to represent the entirety of your presentation including the data, process, findings, and implications in a business setting. Thus it’s a professional report, anything less than a professionally written and organized report will be considered sub-optimal. Amazon for example doesn’t use PowerPoint and instead uses “6 pagers” to make business recommendations, as such some organizations prefer written information over presentations. The use of external and verifiable sources is expected to add context and support any component of the paper. **The minimum is 2 pages maximum is 5**. **Double spaced and 12 point font.**

Helpful tips:

Markdown is not encouraged. It is to be a professional report similar to amazon's 6 pagers though not as long. In light of these expectations and changes to the rigor of the assignment, I would dissuade you from using a lot of code screenshots and instead describe your intentions/problem statement, data aspects, and results/findings/implications in an organized manner. Lastly, outlines and bullets alone will not earn you robust marks.