## 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 team will need to turn in code and PowerPoint slides.**

## 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 the information clear and supported in narration and code? Did the information satisfy the case problem? Were external and trustworthy sources used?

**Data Mining & Modeling 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?

## 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>