SVG Analytics - Junior Data Scientist Project

The purpose of this project is to evaluate your critical thinking, analysis, and data science skills. There is no one right way to approach the problem presented; we are more interested in your process than the result. As much as possible, please share your thought process and work.

You will have several days to complete it, but the project should take around 4 hours to complete. If you find yourself tempted to go over that deadline, we encourage you to find a natural stopping point and summarize the additional steps you would like to take given more time. Python is the preferred language of SVG Analytics, but regardless of the tools used, please send us any code/notebooks/worksheets/etc along with your analysis.

Problem description

One aspect of our business involves calling potential customers (leads) who inquire about our products. In the best-case scenario, these customers work with our licensed sales agents to submit an application for a new policy (we consider this a sale). The typical structure this process takes is:

1. We create a prediction score for all available *uncontacted* leads at the time of a match.
2. The agent will match with the "best" *uncontacted* lead and call that lead.
3. If the customer is not available, we put the lead back in the inventory of available leads for later consideration.
4. If the customer is available, the agent engages in a conversation; this is a contact.
5. Agent assesses the customer's interest and eligibility in our products.
6. If applicable, the agent provides custom pricing information; this is a quote.
7. If the customer is interested, the agent helps the customer submit an application.

Problem: *How do we use available data to improve our prediction score (1) to maximize sales?*

Assumptions

- Each observation in the provided data set is a single match event
- Leads may be matched multiple times over their lifecycle
- If a lead is contacted, it is removed from the inventory of available leads

Request

1. Using the provided dataset, analyze the data and share your insights.
2. Design a solution for the problem above. This could be a predictive model, heuristic, or logic tree, e.g. Evaluate your solution and explain how it attempts to solve the problem.
    - Describe how you will measure the success of your solution. If there is a similar model already in production, how do we compare it to the new solution?
    - Record assumptions you make and questions that come up as you work.
3. List additional steps you would take if you had more time or resources.

A note about AI:

We are aware of the usefulness of AI coding tools, and we use them ourselves at SVG. We encourage you to use AI ***responsibly*** *throughout this project – it will benefit you to think deeply about the problem presented. Please explain any usage of AI in your analysis, as this will help us evaluate your development process.*