Windfall Data Science Challenge
PROBLEM STATEMENT

One of our customers is having a fundraising drive. They are interested in building a Propensity to Give model that targets major donors. A Propensity model is used to predict how likely someone is to make a contribution to your organization. For this fundraising drive, a major donor is defined as any donor who is likely to give at a level of at least $20,000 over a 5 year period. There are approximately 130,000 potential donors in this database, some who have donated significant amounts previously and others who have donated nothing at all. Our customer's fundraising team has a limited capacity for targeting donors, so it is important to properly identify only those who are most likely to meet the targeted donation amount. Using the attached historical donation data in donations.csv and Windfall's unique wealth features in windfall_features.csv, train a model that is able to identify which constituents are likely to give at a level of > $20,000 in total over the next 5 years.

TRAINING DATA

- Donations.csv is the same dataset used in your Data Preparation interview. This dataset includes the date and amount of all the past donations. If using this data for features, please keep in mind the timeframe used to construct the major donor labels.

- Major_donor_labels.csv is a dataset containing major donor labels for each candidate in the Donations.csv dataset. In this file, a candidate is defined as a major donor if they have donated at least $20,000 in total between August 1, 2016 and July 31, 2021.

- Windfall_features.csv is a new dataset that includes 22 features for the majority of candidates.
    - There are different types of donation-based features in this dataset such as "ClassA", "ClassB", etc. ("isClassADonor", "sumClassADonation") that represent different sources of donation data. Note that these are just categories, and how they are named ("ClassA" versus "ClassB") is not necessarily indicative of their predictive quality. Those that contain "CauseA", "CauseB", etc ("sumCauseADonations") represent charity groupings - e.g. environmental non-profits.
    - The other donation-related variables in windfall_features.csv are independent from "amount" and can be safely used as features in your model.
    - The target population for this model is all the candidates in donations.csv for whom the model should be able to predict a propensity score. Please note that some of these individuals do not have any records in windfall_features.csv.

PROJECT SUBMISSION

Once you have developed your model, please provide your source code and summarize in a document that includes the following:

- Instructions for running your training script
- Summary of data preprocessing and feature engineering
- Explanation of your model algorithm choice and chosen model parameters
- Metrics/figures of your model's performance
- Advice for how to use the model for decision making to achieve the business objective

Keep in mind that we are more interested in the overall approach you take and the model you build than we are in the actual accuracy of the model.

Your submission is due one week from the time you receive it.

Thanks for your interest in Windfall Data.

Best of luck,
Windfall Data Science Team