What is the problem you want to solve?

To analyze the donors' behavior and how they respond to different project proposals.

Who is your client and why do they care about this problem?

The client is DonorsChoose.

The analysis will help the organization to better understand the preference and difference among various donor groups, which in turn can help improve its donor targeting and engagement and increase effectiveness of fundraising campaigns.

What data are you using? How will you acquire the data?

The data was collected from DonorsChoose open data platform: https://research.donorschoose.org/t/download-opendata/33. It was read directly into Python and saved as csv files. The datasets I will use include the one that contains project data from year 2002 to 2016 and the one that contains all donation data that are associated with each project. The two files will be merged together for the purpose of data wrangling and machine learning models.

Variables include:

IDs - school ID, project ID, teacher ID, donor's ID

Geographic information - school location, school district, donor location, etc.

Project related information - project type, project focus, funding amount, targeting students, donation time, donation status, etc.

Briefly outline how you'll solve this problem.

I will first do some exploratory data analysis to find the relationship among different variables. This step will help me to get acquainted with the data and form a sense of the correlation among the variables and help to determine which features will be used to build the machine learning model. A few examples include:

- Geographic distribution of donors
- School districts vs. funding needs
- Poverty level vs. funding needs
- Project type vs. funding status
- Payment method popularity
- Project types that received the most donation
- Donation completion duration distribution
- Fundraising success rate, campaign-driven vs. teacher-driven

- Donors and donations (highest donation, most frequent donation, etc.)
- Donation amount by year, month
- Highest donations among US states
- Most frequent project types/categories
- Funding status by year, month
- Days of week the donations are made

The donor's behavior analysis will be done by k-means clustering to segment donors into different clusters. The features used by the model will be determined after the initial EDA.

What are your deliverables? Typically, this includes code, a paper, or a slide deck.

- Codes (Jupyter Notebook) for all data wrangling steps and machine learning
- Data visualization with Seaborn and Matplotlib
- Final project write-up
- A client facing slide deck to explain the problems and suggest actions