Donor Analysis Report: Project Excitement and Donor Segmentation

Author: Yuxia (Claudia) He

1. Problem Statement

The charity is currently focused on identifying "exciting" projects that effectively encourage more donations to enhance student education. Additionally, the organization seeks a deeper understanding of its donor base to improve the overall user experience and foster long-term, repeat contributions.

2. Data Quality Issues and Solutions

To ensure the integrity of the analysis, several data cleaning steps were performed:

- **Redundant variables and target variables**
  We remove some variables that have high cardinality, like donationid, projected
- **Character Variables**
  We dummy-encoded some character variables into 1s and 0s. For example, if the donor is a teacher, he will be classify as 1; if not, then 0.
- **Outliers**
  We removed some outliers by using filter() function. For example, donation_to_project (Amount to project excluding optional tip).
- **Missing Values**
  We imputed median to NA values of variables such as teacher_refer_count that have high completeness.

3. Exploratory Data Analysis

- **Explore Target Variable**
  Our Target variable is ‘is exciting’, which explain whether or not the project is considered as exciting program. By analyzing the variable, we found out that the exciting programs take up 11% of the total programs, which is good for us to do analysis on since we don’t have to worry about the number of projects being too small.
- **Explore Numeric Variable**
  1) We converted the date a project went live on the site to the days the project went live from now.
  2) Boxplots Analysis: we found out some variables play a important role in determining the exciting programs. For example, the exciting projects have a higher proportion of unique comments on a page than not-exciting projects. Besides, we also find out exciting projects have a shorter days that have gone live.

3) Bar Chart: we found out that all of the numeric variables in the donation datasets is left skewed, for example, donation to projects, which means that most of people donates money like $10.

- **Explore Categorical Variables**
  We found some interesting things through exploring the categorical variables. First, public charter schools have a higher proportion of exciting projects. Besides, few doctors donate money.

4. First Goal: Identifying Exciting Projects

- **Final Analytical Solution**
  We used three predictive models to identify exciting projects, among which **random forest** was chosen because it had the highest precision and recall rate and the highest AUC value of the both training and testing dataset that is close to 1, which indicates that the model is good at separating two classes. Therefore, this model is the best to identify the exciting projects.
- **Key Factors for Identification**

Based on the results of the random forest variables importance chart, we found 10 variables that will impact the identification of whether or not the projects will be classified as exciting projects. The top 3 variables are: the number of donors that were teacher referred, the proportion of unique comments on a page, and the time gap between the days the went live and today. We found out that if the posts went live recently, it is more likely to be classified as exciting projects. Besides, when the proportion of unique comments exceeds 60%, their proportion of classifying as exciting projects will increase. Additionally, the average donors that were teacher referred is two, meaning that having a teacher referred is important to determine whether or not it is a exciting project. The reason behind such circumstances probably due to the fact that teacher is usually seen as a reliable people.

5. Second Goal: Understanding Donors

- **Final Analytical Solution**
  We used K-means Clustering to cluster our donors into 5 groups. By analyzing the donation activities and donation behaviors, we found out the characteristics of each group, so as to figure out solutions to better their user experience and encourage donations.
- **Key Characteristics of Donors**

  Cluster 1
  - Has the majority amount of donors
  - Donation amounts are either under $10 or $10-$100.
  - Likes using cash for payment
  - Likely to donate using PayPal, credit card, amazon
  - Don’t like making a donation included corporate sponsored gift card.

  Cluster 2
  - Likes using credit card, amazon for payment.
  - Least likely to using cash for payment
  - Most Likely to make donations that was matched 1-1 with corporate funds
  - Has a high portion of donation that was given through a giving / campaign page

  Cluster 3

- Has the least amount of donors
- Seldom make any donations

Cluster 4

- Has a portion of a donation used accounts credit redemption
- Has the highest proportion of people that donate more than $100
- Has the lowest proportion of people that donate under $100
- Has a high portion of donation that was given through a giving / campaign page

- **Recommendation**

Cluster 1

- **Collaborate with payment institution.** Put more advertising or related information by working with Paypal, credit card, amazon, since Cluster 1 prefer to using above payment methods for donations.
- **Hold more onsite events for donations.** Since this cluster takes up the majority amount of people and are more likely to donate cash, they are probably know more about the donations through onsite activities.

Cluster 2

- **Use gift cards or hold campaign frequently.** Since this cluster has the highest likelihood of making donations that was matched 1-1 with funds, they are the group thar is prudent. Besides, among all payment methods, they are least likely to donate through cash, which means they are not the active donors that are willing to make a donation and they are largely driven by incentives provided by gift cards.
- **Put more advertising on campaign page.** Since they like donating through pages, this cluster is likely people that are young and like go through website often.

Cluster 3

- **Explore or Drop.** Since we don't have too much information about this cluster, and the number of people is relatively small, further exploration is needed.

Cluster 4 Wealthy and Generous

- **Give credit as incentive.** Since the donor with this cluster is mostly likely to use accounts credit redemption, which means that they care about their credit, giving them credit as incentives will encourage their donations.
- **Send follow-up emails to them.** Since this cluster are most likely to donate money above $100, it is good to send them follow-up emails to engage them and better the user experience of visiting the website