Springboard Data Science Career Track

Capstone Project 1 - DonorsChoose

Irene Yao
December 12, 2018

Introduction

DonorsChoose.org is a non-profit education crowdfunding platform that allows individuals to donate directly to public school classroom projects.

- Inception: 2000
- 1,200,000+ projects
- 31,000,000+ students reached

Project Summary:

- Exploration of factors that may affect the project's success rate
- Machine learning: donor clustering using K-means

Data Source & Data Cleansing

Data Source

- Source: https://research.donorschoose.org/t/download-opendata/33
- Tables: Project data & Donation data

Data Cleansing

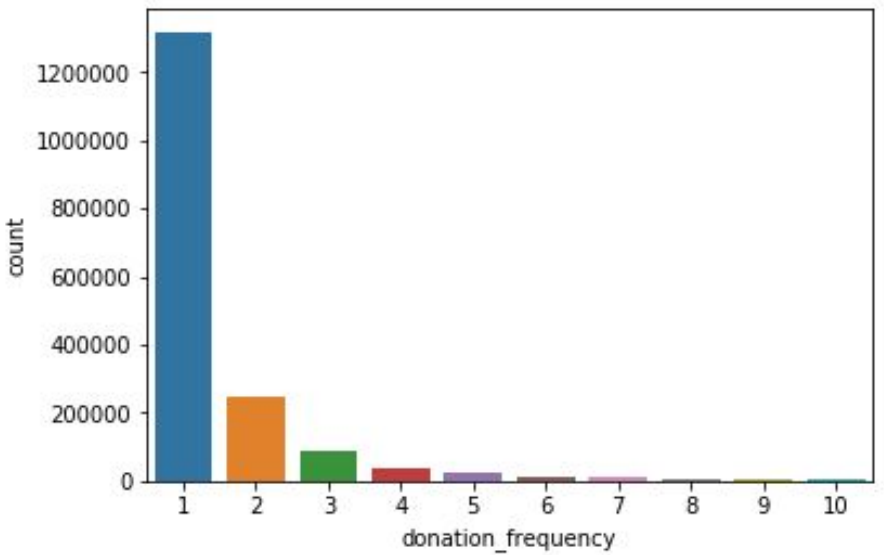
- Feature selection
- Memory optimization
- Missing values

EXPLORATORY DATA ANALYSIS & STATISTICAL ANALYSIS

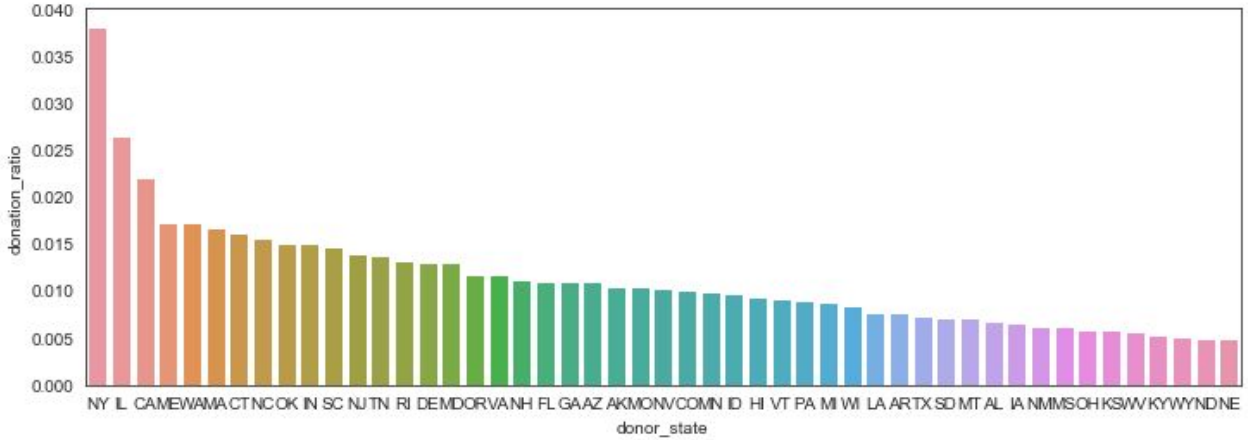
Donation Analysis

Donors & Donations

- Total 1,786,756 donors
- 1,318,268 single-time donors
- 26% frequent donors



Donation Distribution - States with most donations

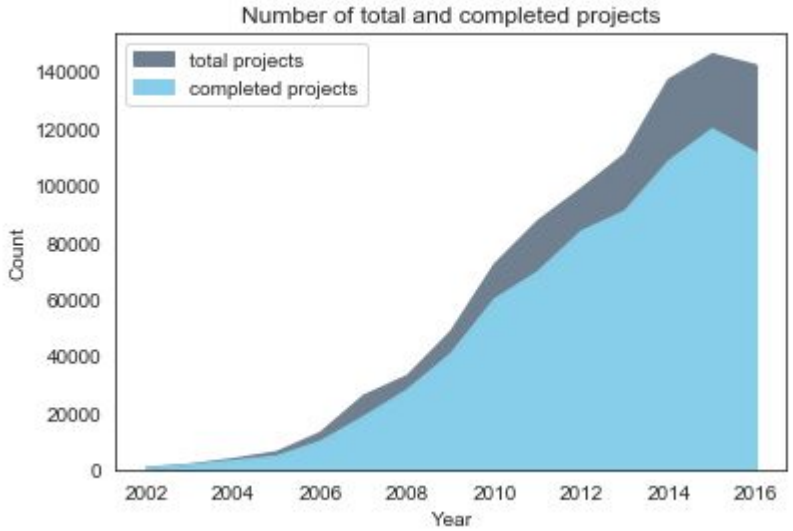


Top 3 states with highest per capita donation: New York, Illinois, California

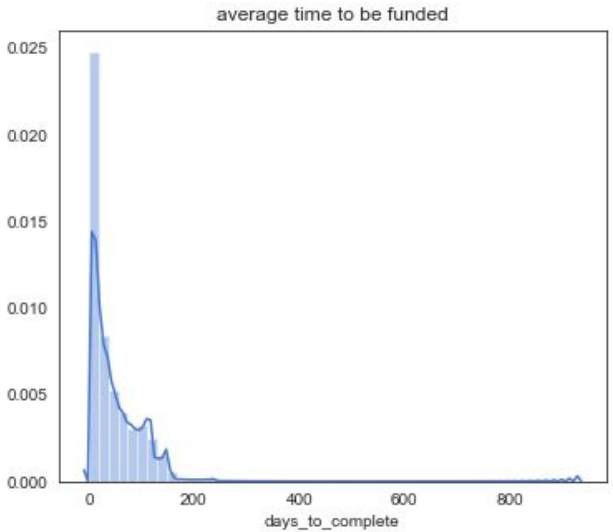
Project Analysis

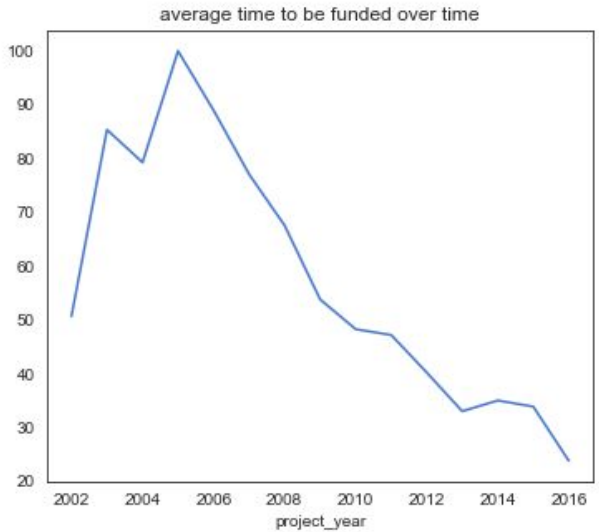
Count & Amount

The project flow started to pick up in a fast pace starting around year 2006. The organization has been given Charity Navigator's highest rating every year since 2005.



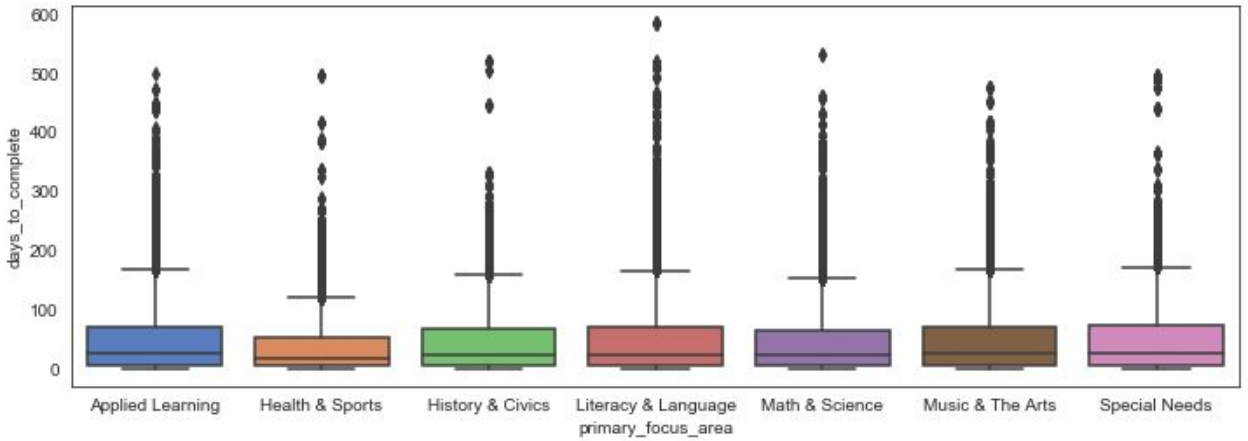
Project Completion Time





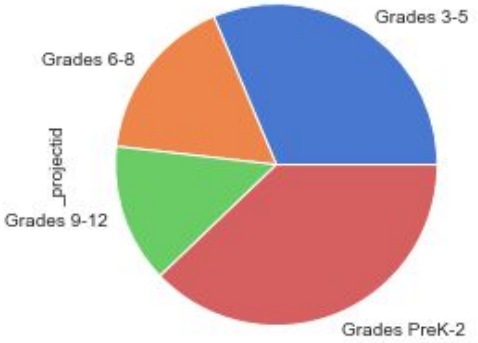
- Majority of the projects are fully-donated within 100 days
- Recent projects were completed within around a month

Project Completion Time vs. Primary Focus Area

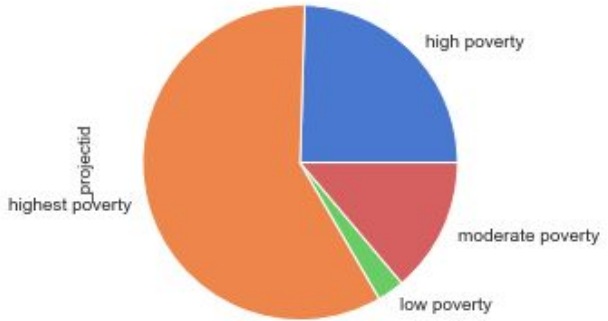


Health & Sports has the lowest mean and the narrowest interquartile range. It is the category that requires the least time to collect donation.

Projects vs. Grade Level

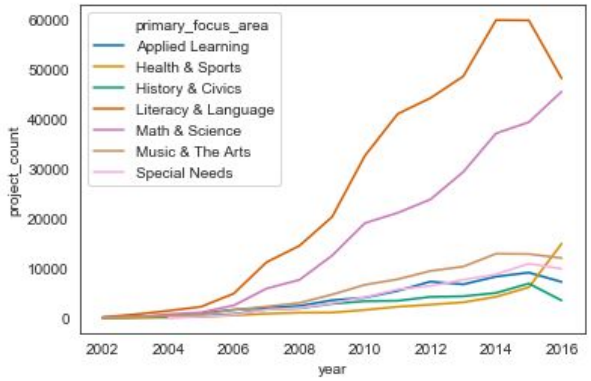


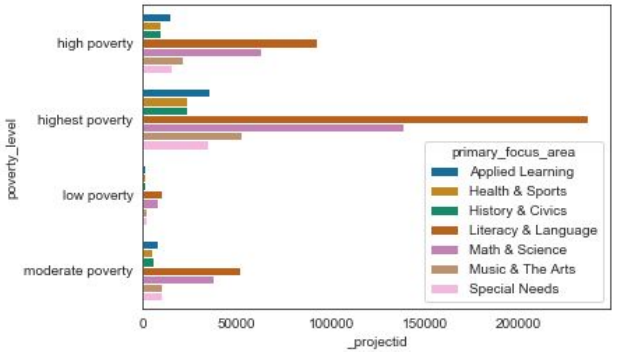
Projects vs. Poverty Level



Projects vs. Focus Area

- Teachers are requesting more support in Literacy & Language and Math & Science categories
- Demand in Math & Science and Literacy & Language have grown more rapidly than the rest
- This holds true across all regions of different wealth levels

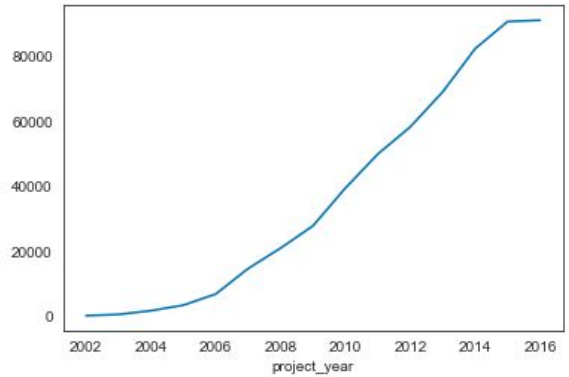




Teacher Analysis

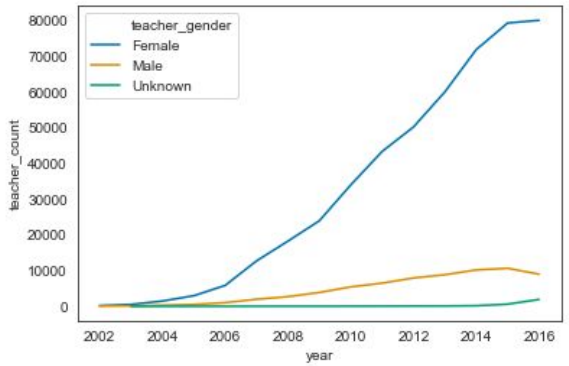
Trends

Overall Trend



Teacher users have more than doubled from year 2010 to 2016

Female vs. Male Teachers



Number of male teachers grew very little and majority of the projects are posted by female teachers

Observation from EDA

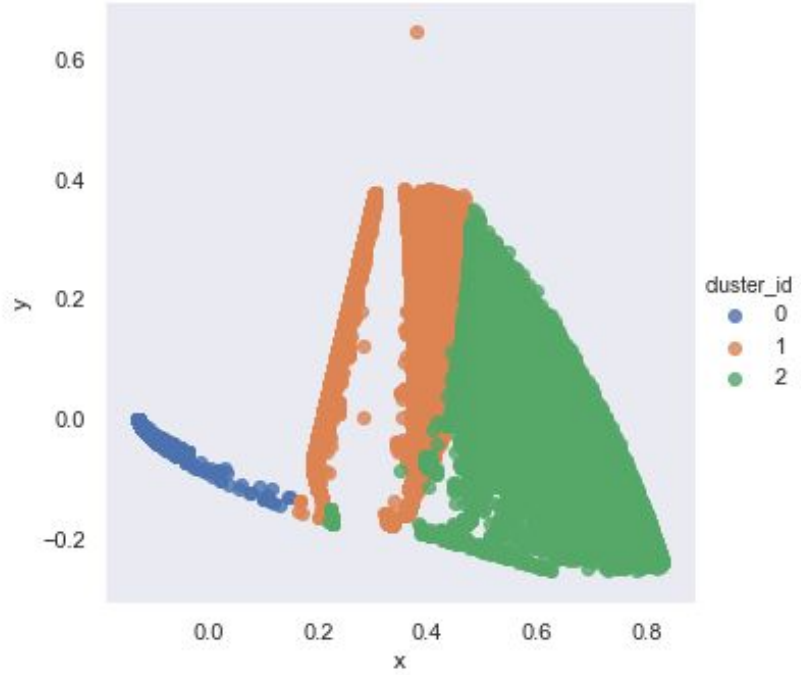
Observation 1	Out of the massive number of donors online, only 26% has donated more than once.
Observation 2	NY, IL, and CA are the top 3 states with the highest per capita donation.
Observation 3	Project volume has grown rapidly in the past 10 years and the time to collect donation has decreased drastically.
Observation 4	Health & Sports tends to raise money more quickly compared to other categories.
Observation 5	Lower grades in elementary schools need the most financing support, especially in schools that are located in high-poverty regions.
Observation 6	Despite the poverty level, Literacy & Language and Math & Science are the categories that need the most funding.
Observation 7	Among all the projects posted, most of them are led by female teachers. The completion rate of the projects is around 83%.

MACHINE LEARNING - K-MEANS CLUSTERING

Donor Clustering

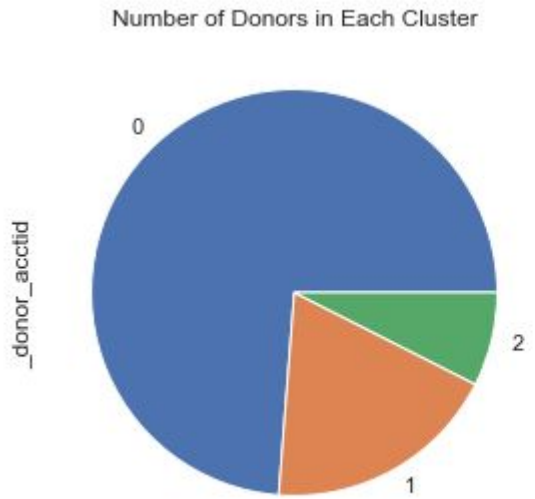
The Clusters

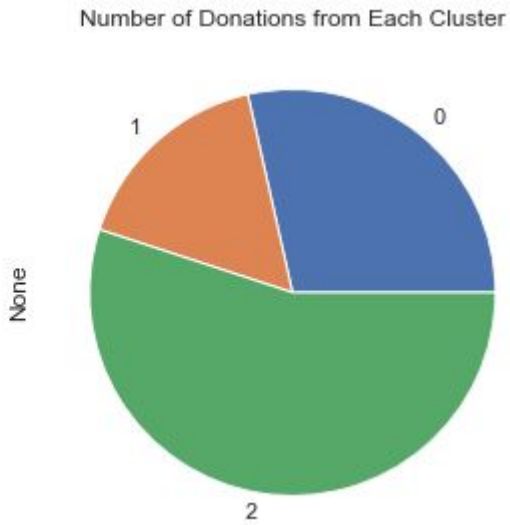
Machine learning using k-means divided the donors into 3 distinct groups where the donors in each group share some similar characteristics



Cluster Analysis

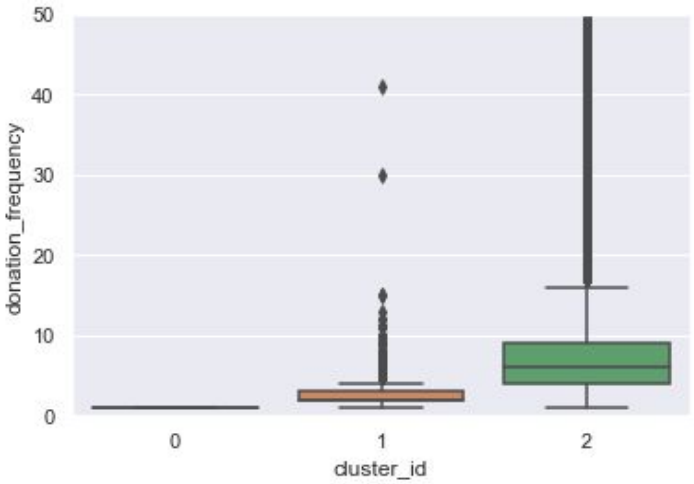
Number of Donors & Donations in Each Cluster





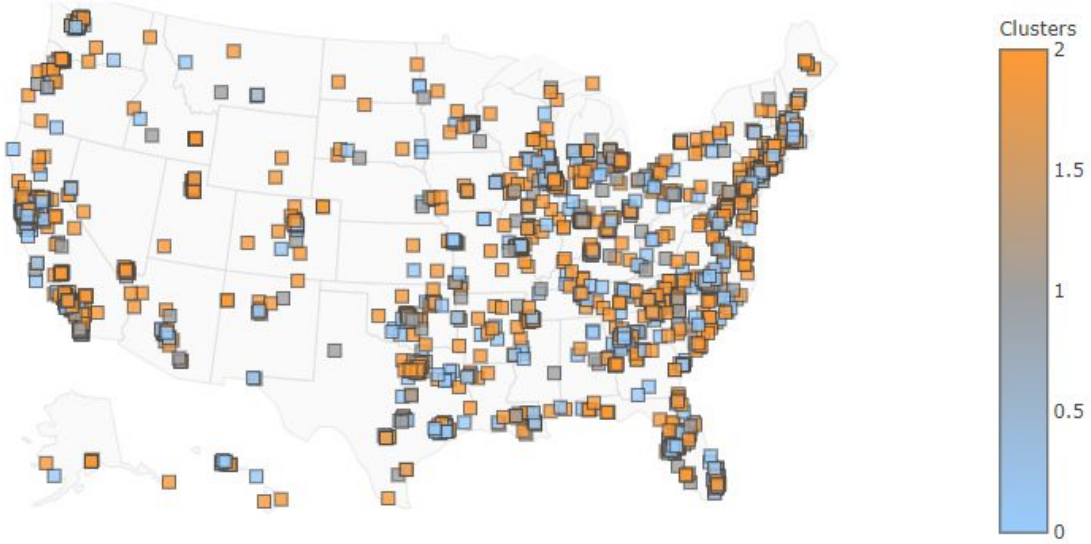
Donation Frequency

	cluster_id	min	max	mean	median
0	0	1	1	1.000000	1
1	1	1	41	2.267197	2
2	2	1	117096	19.055990	6



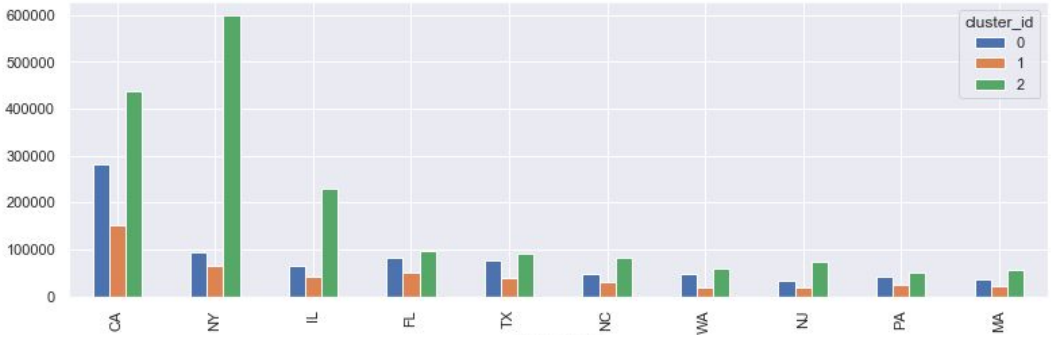
- ➔ Cluster 0 are single-time donors
- ➔ Donors in cluster 2 on average donated 19 times each person
- ➔ Majority of donors in cluster 1 donated between 2-4 times

Donors by State

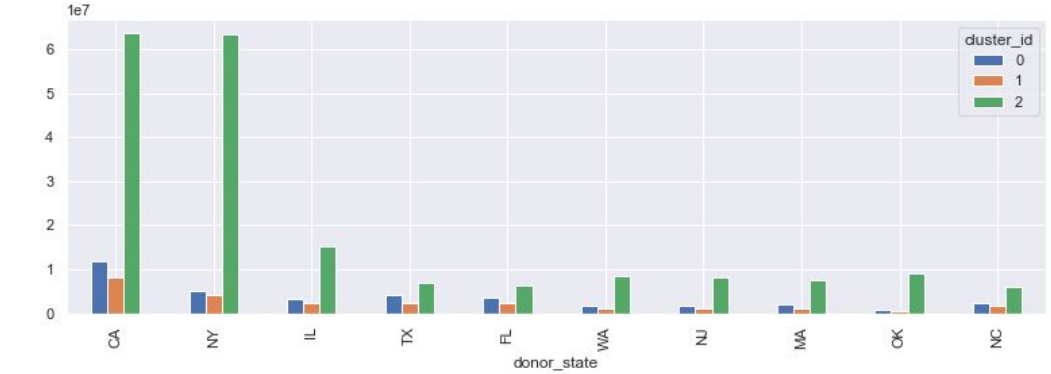


Donation Counts & Amount by State

Count

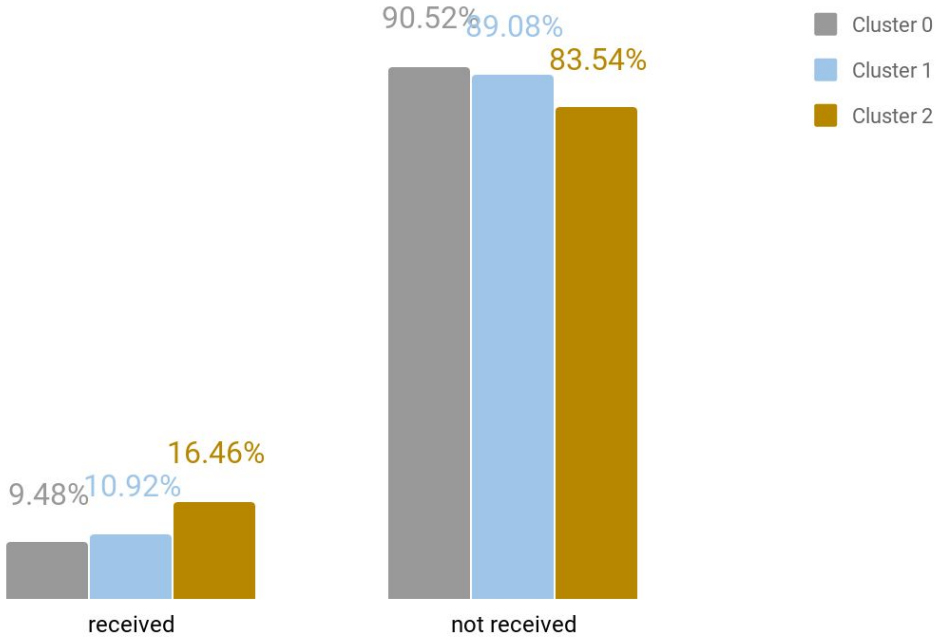


Amount



Donations vs. Thank You Packets

Group 2 received the most percentage of thank you cards for their donations



Observation from Clustering

Observation 1	Group 0 has the largest number of donors and Group 2 has the least number of donors.
Observation 2	Even though Group 2 has the least number of donors, they contributed to the largest percentage of total number of donations.
Observation 3	Group 0 are all single-time donors and Group 2 consists of many frequent donors.
Observation 4	Geographically, donors in all 3 groups scatter across US. Most of the donors come from east coast and California and Washington on the west coast.
Observation 5	New York has the largest proportions of donations from Group 2 compared to other states.
Observation 6	Group 2 received the most percentage of thank you letters amongst all three groups.

RECOMMENDATIONS

Incentivize Group 2 Donors

Create recognition system
Donor leaderboard Special volunteering opportunities

Motivate Single-time Donors

Encourage teachers and students to send more thank you packets and show appreciation

Expand Geographically

More market campaigns in underserved regions
Partnership with local media and organization

FUTURE WORK:

- Application of different machine learning algorithms
- Analysis of description of projects
- A/B testing on different campaign methods