



Donorschoose.org

Capstone Project #1

Reena Turak
11/13/2018

Introduction

Founded in 2000 by a Bronx history teacher, DonorsChoose.org has raised \$685 million for America's classrooms. Teachers at three-quarters of all the public schools in the U.S. have come to DonorsChoose.org to request what their students need, making DonorsChoose.org the leading platform for supporting public education.

To date, 3 million people and partners have funded 1.1 million DonorsChoose.org projects. But teachers still spend more than a billion dollars of their own money on classroom materials. To get students what they need to learn, the team at DonorsChoose.org needs to be able to connect donors with the projects that most inspire them.

In the second Kaggle Data Science for Good challenge, DonorsChoose.org, in partnership with Google.org, is inviting the community to help them pair up donors to the classroom requests that will most motivate them to make an additional gift. To support this challenge, DonorsChoose.org has supplied anonymized data on donor giving from the past five years. The winning methods will be implemented in DonorsChoose.org email marketing campaigns.

Problem Statement

DonorsChoose.org has funded over 1.1 million classroom requests through the support of 3 million donors, the majority of whom were making their first-ever donation to a public school. If DonorsChoose.org can motivate even a fraction of those donors to make another donation, that could have a huge impact on the number of classroom requests fulfilled.

A good solution will enable DonorsChoose.org to build targeted email campaigns recommending specific classroom requests to prior donors. Part of the challenge is to assess the needs of the organization, uncover insights from the data available, and build the right

solution for this problem. Submissions will be evaluated on the following criteria:

- Performance - How well does the solution match donors to project requests to which they would be motivated to donate? DonorsChoose.org will not be able to live test every submission, so a strong entry will clearly articulate why it will be effective at motivating repeat donations.
- Adaptable - The DonorsChoose.org team wants to put the winning submissions to work, quickly. Therefore a good entry will be easy to implement in production.
- Intelligible - A good entry should be easily understood by the DonorsChoose.org team should it need to be updated in the future to accommodate a changing marketplace.

Client

Client is DonorsChoose.org. DonorsChoose.org will use this analysis to email marketing campaigns and get donations for the projects.

Data:

Data source is Kaggle where Donorschoose.org provided dataset for Kaggle competition “ Data Science for Good: DonorsChoose.org” -Help DonorsChoose.org connect donors with projects they care about.

Link: <https://www.kaggle.com/donorschoose/io>

Data set description:

- Dataset includes 6 tables:
 - Donors
 - Donations
 - Projects
 - Schools

-
- Resources
 - Teachers
 - Data cleaning
 - Exported data from CSV to pandas dataframe
 - For each table, sum of null values for all column are found and analysed.
 - Dropped all the rows with any null values for table Donors as the null values were 10% of the total data.
 - Dropped all the rows with any null values for table Schools as the null values were 2-3% of the total data.
 - Dropped all the rows with any null values for table Teachers as the null values were just 28 records
 - In Projects table in column Project Fully Funded Date there are 283253 records with null value so let it be as it is as this particular column might not be required though we may need to consider other columns for insights.
 - Donations table do not have any null values.
 - Resources table do have high % of null values:

Project ID	0
Resource Item Name	561
Resource Quantity	24299
Resource Unit Price	24310
Resource Vendor Name	81994

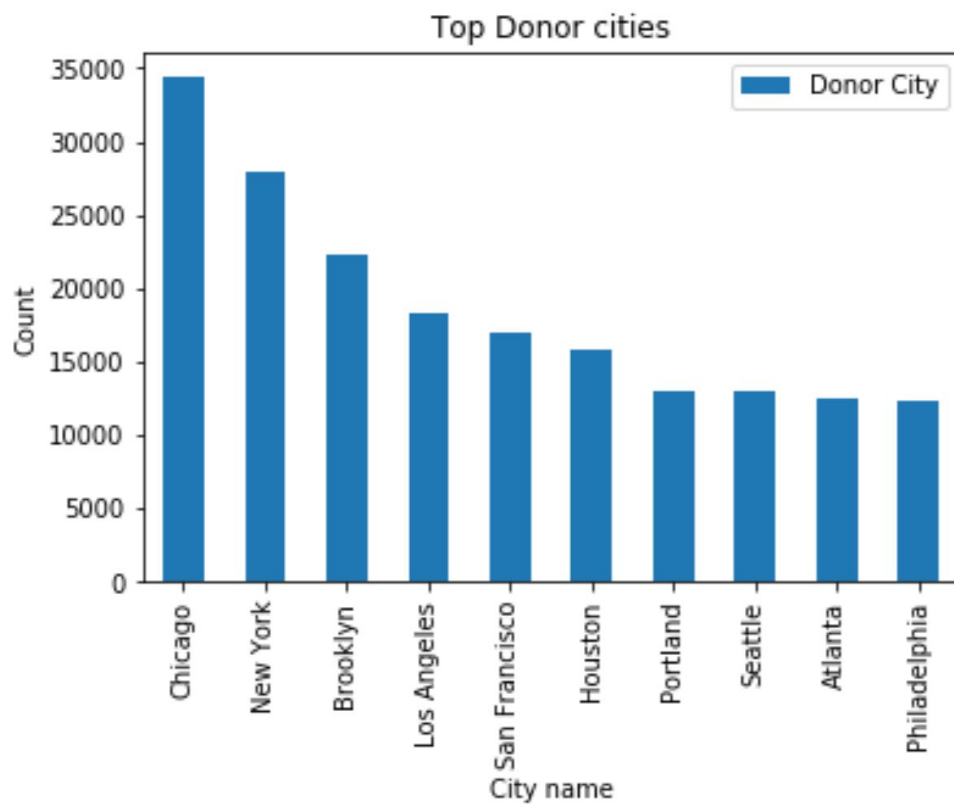
So for now just keeping this as it is.

As the missing values were not in high percentage and its not very easy to replace them so dropped the records with null value in the few tables as stated above.

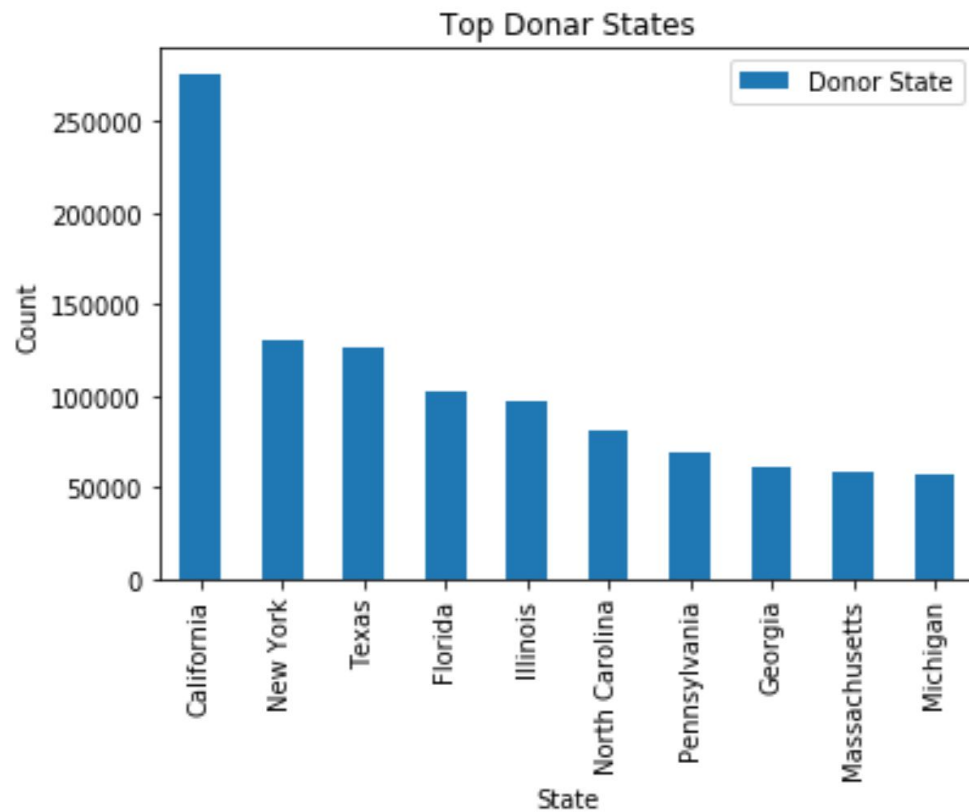
Exploratory Data Analysis:

Summary

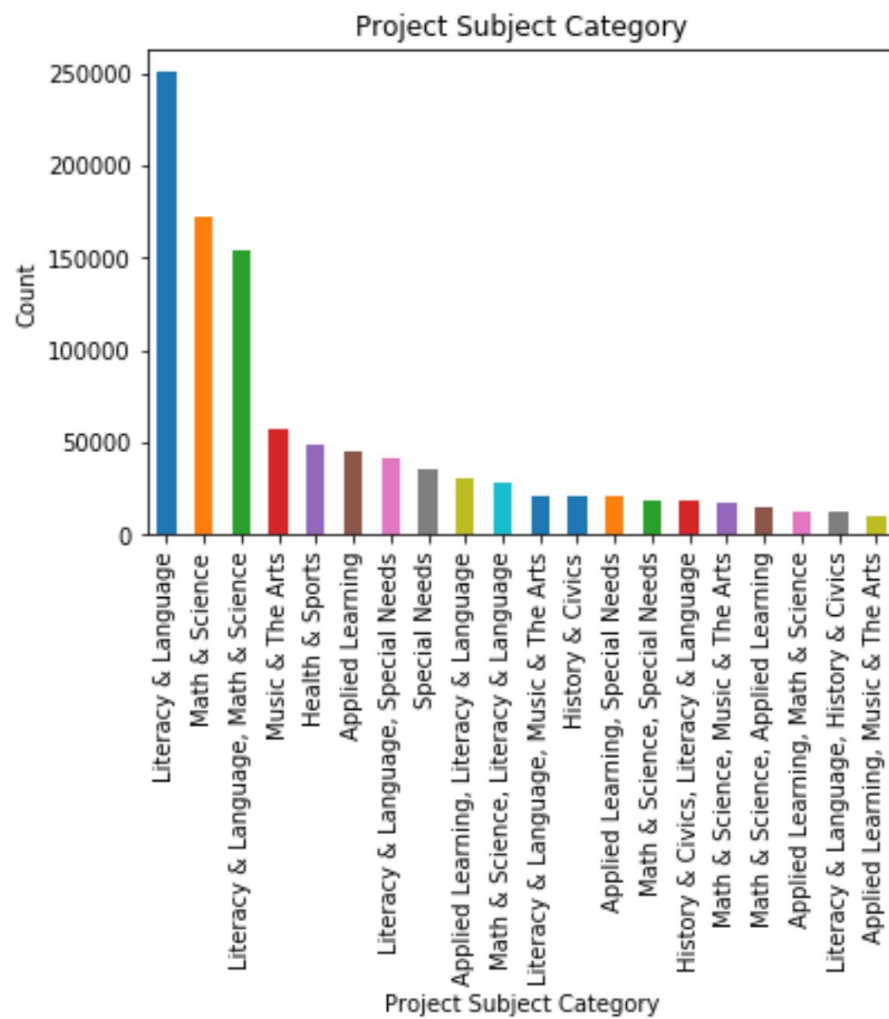
1. Total donation amount raised by Donorchoose.org is \$284408243.28
2. Minimum Donation amount is USD 0.01 , Mean donation amount is USD 60.67 and Maximum donation amount is USD 60000
3. Top Donor cities include Chicago, New York, Brooklyn, Los Angeles and San Francisco



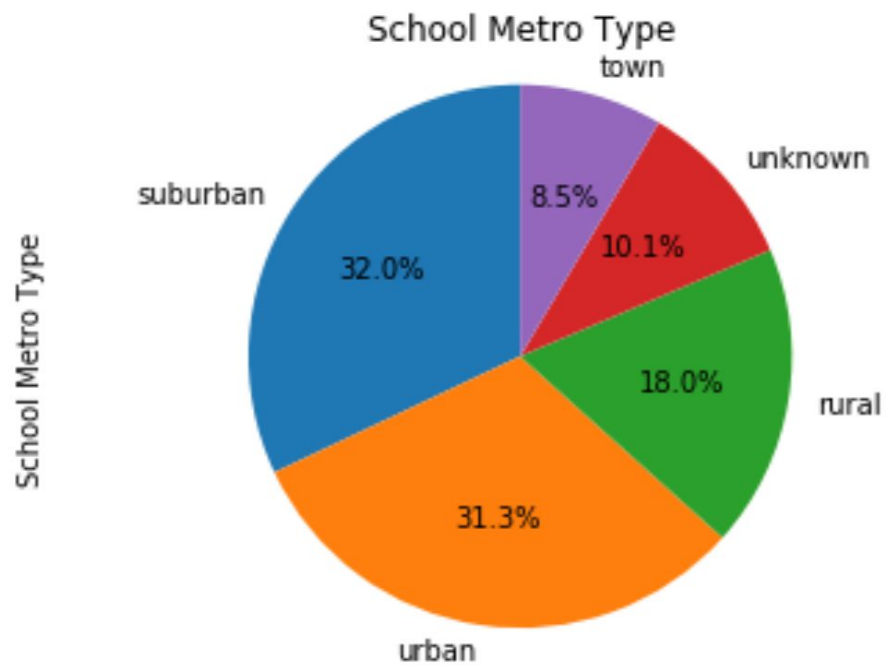
4. Top donor state is California, New York, Texas Florida and Illinois



5. 9.5 % Donors are teacher and 90.5% are non teacher.
6. Yearly Donations trend show clearly that No of Donors are increasing
7. donations are mostly received on working days like tuesday and wednesdays.
8. Top Project sub-categories are : Literacy & Language , Math & Science , Music & Arts.



9. Top Project subject Sub-categories are : Literacy , Mathematics & Writing.
10. Most of the projects are led by Teacher and very small amount of projects are Professional Development and Student Led.
11. Metro categories



- Suburban - having 31.5 % schools
- Urban - having 31.2 % schools
- Rural - having 17.8 % schools
- Town - having 8.38 % schools
- Unknown - 11.1 % schools

12. 86.4% Teacher who posted the projects are females

13. Top weekdays when teachers posted their first project : Sunday - Approx. 73 K

Saturday - Approx. 66 K Monday - Approx. 61 K

14. Project Status:

- Fully Funded - 74.5%
- Expired - 21.5%
- Live - 3.8%

