Donors Choose

Attracting and Retaining Recurring Donors

Kelli Peluso
DSC Capstone 2

Background

The Problem

DonorsChoose.org funds classroom requests through the support of over 2 million donors. This non-profit company is unique because it allows the donors to fund specific projects based upon many different criteria.

Donors can be attracted to projects for different reasons - maybe they're an alumni of the school, or know the teacher who has posted the project. Maybe it's your child's classroom.

Donors typically interact with the platform to make one donation,. In order to continue supporting and funding projects in public schools, Donors Choose must work to attract and retain recurring donors.

The Data Set

- The dataset was obtained from https://www.kaggle.com/donorschoose/io. This dataset had been shared during a previously held contest. The data was initially broken up into 6 different .csv files (donations, donors, projects, teachers, resources and schools). In order to understand the data better, all data was initially examined separately.
- I initially cleaned and began exploring this dataset in Jupyter notebook, primarily using pandas. I quickly discovered that using my current toolset in Jupyter notebook was not going to be effective to process the large amount of data that I was working with. I moved my work to Databricks and EMR in AWS in order to begin using PySpark.

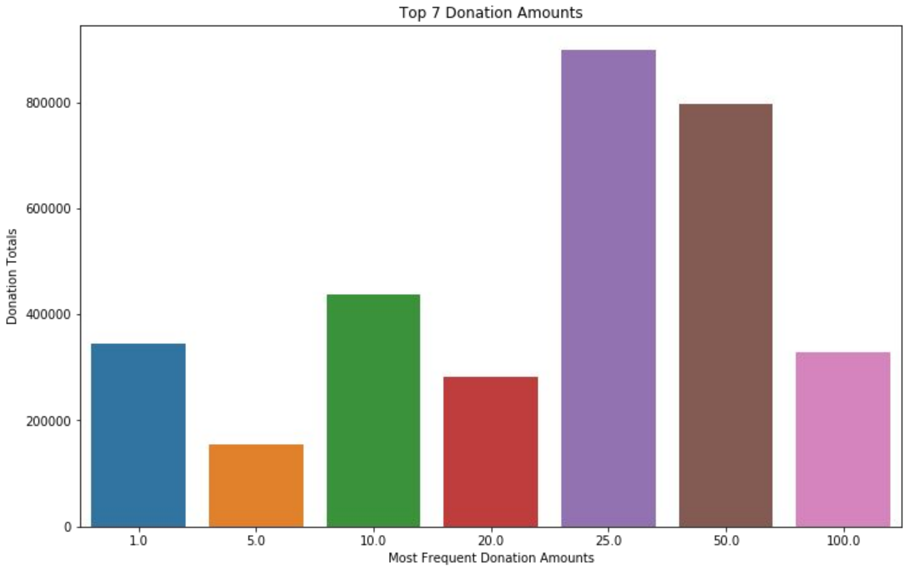
Exploratory Data Analysis

Data Cleaning

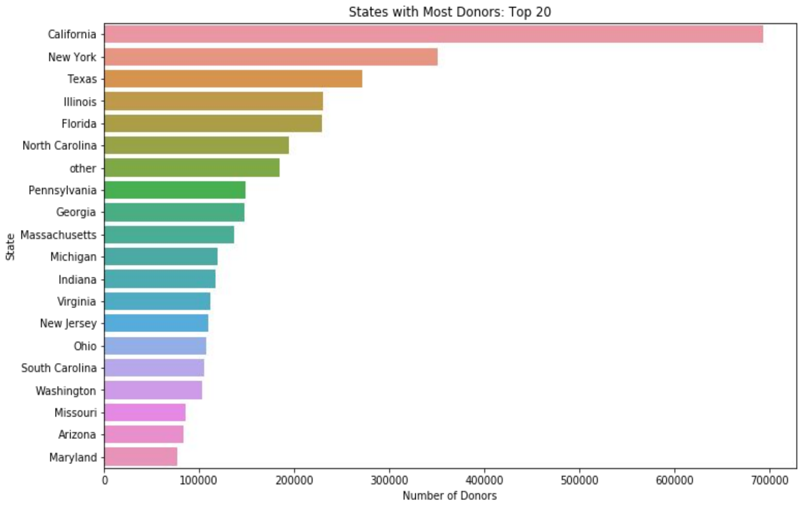
- Databricks
- AWS S3 Bucket storage and mount
- Schema creation for projects.csv
- Backfill null values for Washington DC
- Create new columns (datetime, splitting up main categories and sub categories)

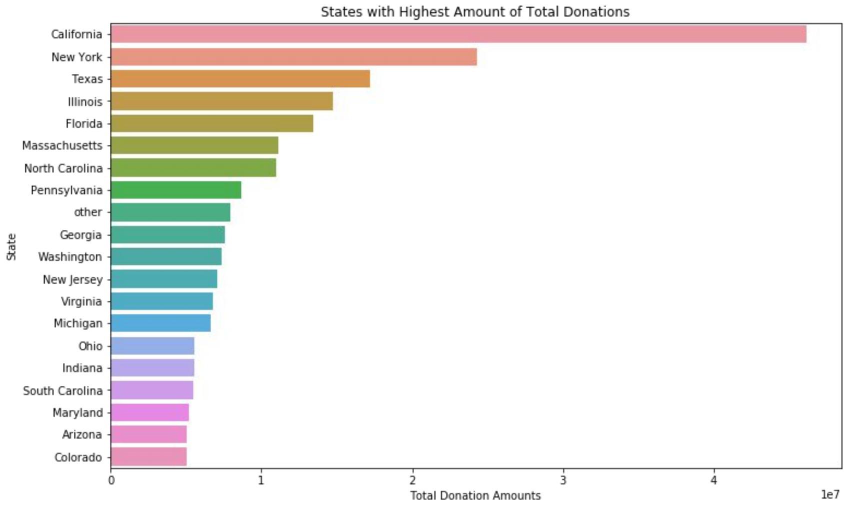
Initial Findings: Donors & Donations

- Total of 2, 122, 640 unique donations
- 2,003,188 unique donors
- 552,941 recurring donors
    - More than 2 donations: 278,039
    - More than 5 donations: 98,487
    - More than 10 donations: 40,299

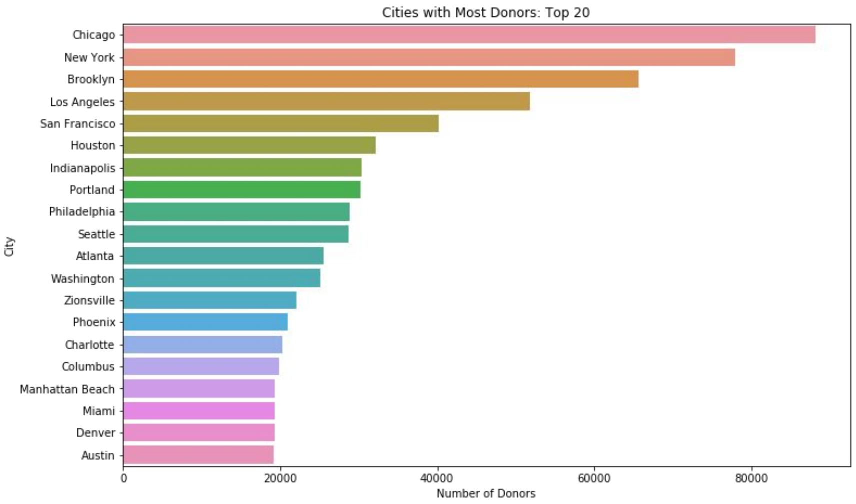


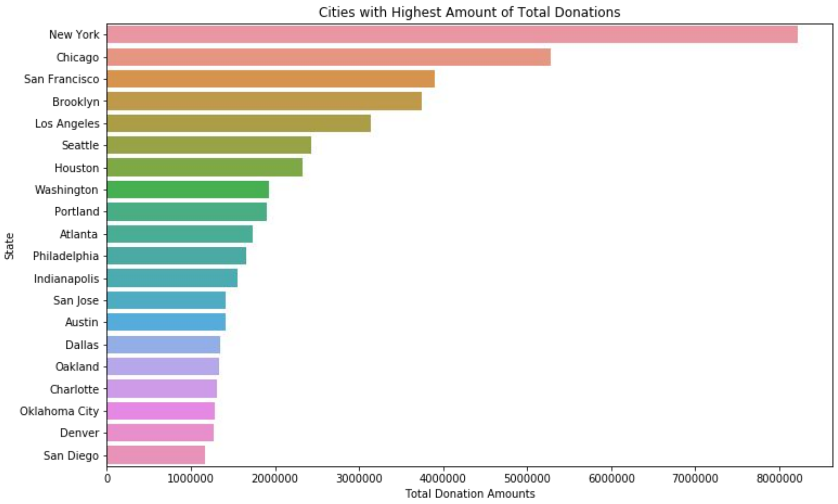
Donors & Donations - States



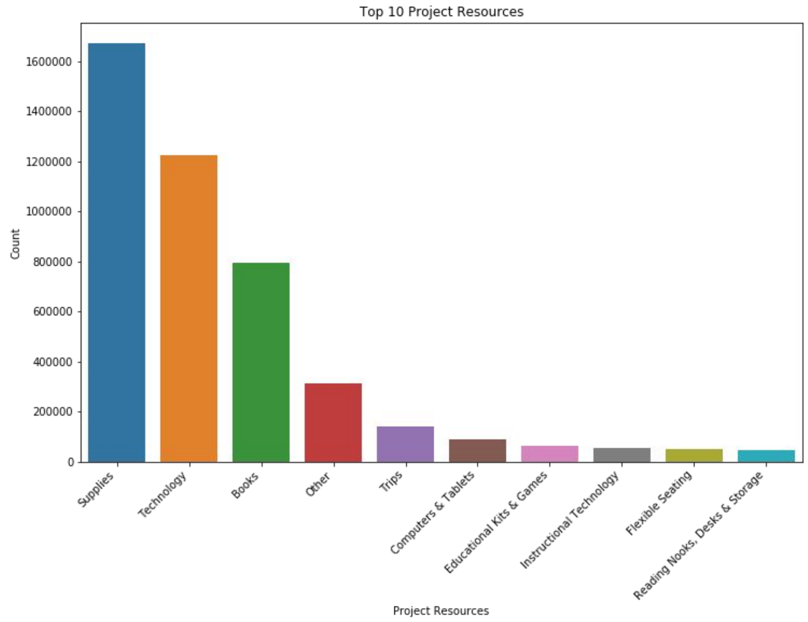


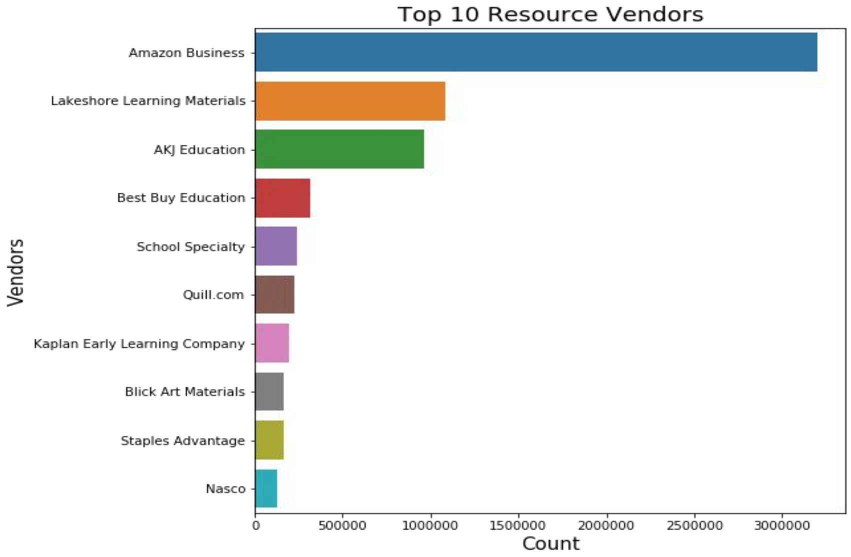
Donors & Donations: Cities



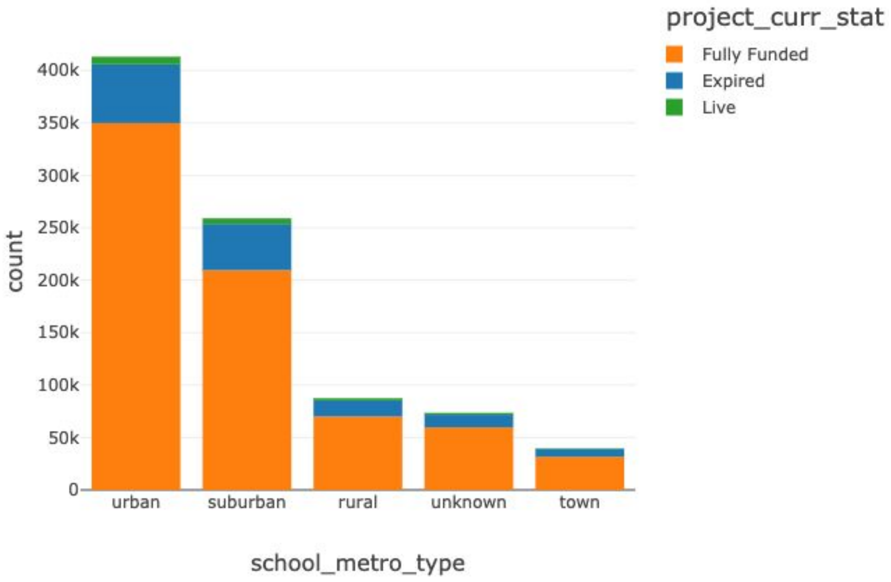


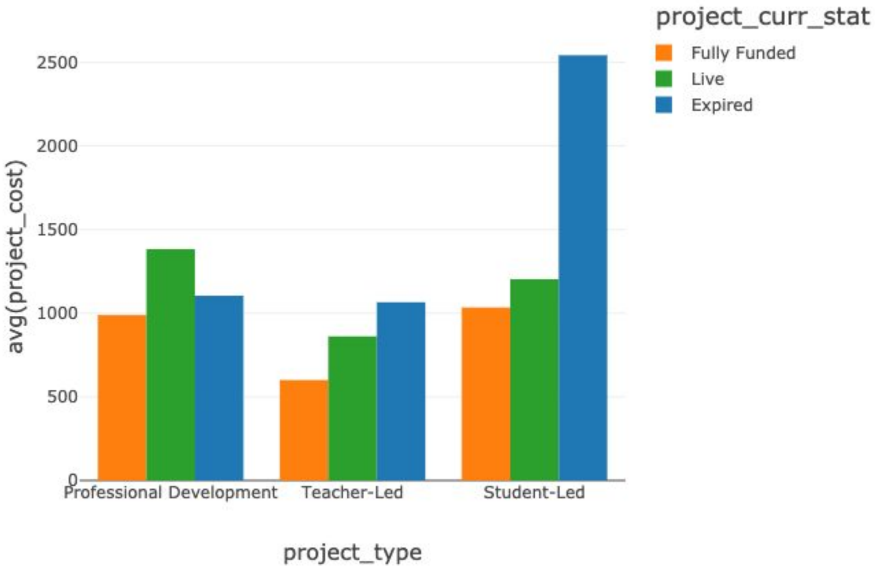
Projects & Resources





Project Status





Modeling

Recommendation Model Types

- **Collaborative Filtering:** This method is used to make automatic predictions about project preferences that a donor may have. Based on projects that they've donated to in the past, this method will predict future behavior based on similar projects that other donors have donated to.
- **Content Based Filtering:** Alternatively, content based filtering uses the information from the description, project essay and project need statement of previous projects that a donor has donated to. The content based filtering method then predicts other projects to these donors based upon their content similarities.
- **Hybrid (mix between Content Based and Collaborative Filtering):** A hybrid approach combines collaborative filtering methods and content based filtering methods. By using a combination of project content and prior project donations, recommendations can be made for future projects.

Alternating Least Squares (ALS)

- Collaborative Filtering
- A matrix factorization algorithm that runs in a parallel fashion, built for large scale collaborative filtering problems.
- Simple, scales well to large data sets
- Explicit Ratings vs. Implicit Ratings

ALS Model #1 - Total Donation Amount

- Uses total donation amount by a donor to a unique project ID as an implicit rating
- The idea is that donors will donate more money, more frequently, to projects that interest them more
- Performance Evaluator: Root Mean Square Error (RMSE)
- Base Model RMSE: 1.9427716349485609
- Optimized Model RMSE: 1.8975567342104507

ALS Model #2 - Project Categories

- Uses number of times a unique donor has donated to a project category
- The idea being that donors could be attracted to certain subjects (main categories) and continue to donate to them.
- Based off of the Million Songs Data Set
- Performance Evaluator - Rank Ordering Error Metric (ROEM)
  - Adapted from this function
- Base model ROEM: 0.18879302687557417
- Optimized Model ROEM: TBD

Conclusions & Next Steps

Recurring Donors & Recommendation Engines

- Recommendation Engines can certainly help Donors Choose attract and retain recurring donors, powering their projects with more capital and generating more interest.
- While the amount of money that a donor provides to a particular project is important, it did not appear to be the most efficient and accurate predictor of projects that the donor would be interested in donating to in the future.
- Using features that donors interact with frequently, such as the main category of the project, appears to be a more efficient and accurate predictor of projects that the donor would be interested in donating to in the future
- With additional time and resources, all project features should be explored to determine which features are most important to donors