

What factors influence charity funding for public school projects?

Analysing the DonorsChoose dataset

Introduction

With rising wealth inequality comes an obligation to ensure adequate educational opportunity for all children to enable fairer outcomes in the future. With cuts in education funding, charitable donations for public schools, though not an optimal solution, have become an important pathway to ensure that public schools receive enough funding to empower children to receive better education and facilities.

This report analyses an open-source dataset by a crowd-sourcing charity organisation named DonorsChoose. DonorsChoose has funded over a million projects and attracted over 3.5 million donors. The dataset consists of data on projects, donors, donations, schools and teachers.

The research objective is to analyse which attributes of the projects make them more likely to receive full funding. This helps examine the reasons why certain school projects successfully obtain funding while others don't. The report uses a multivariable logistic regression model to answer the research question.

Research Question: What factors influence charity funding for public school projects?

Data

The DonorsChoose dataset is available on Kaggle at:
<https://www.kaggle.com/donorschoose/io>

The dataset consists of six tables:

- 1) Donations.csv: A table of individual donations with dollar amounts, and dates.
- 2) Donors.csv: A table of names, cities, states, and zip codes of individual donors.
- 3) Projects.csv: A table of project names, resource categories, subject categories, project titles, project need statements, project essays.
- 4) Resources.csv: Resources requested with details on unit price, quantity and description
- 5) Schools.csv: Schools with zip code, city, state and percentage of students receiving free lunch
- 6) Teachers.csv: Details on gender and state of residence for teachers.

This study using the Projects file. It uses the non-text attribute data on this file for analysis.

Method

The data consists of categorical variables, continuous variables and text data. For the purpose of this project, only the first two are utilized.

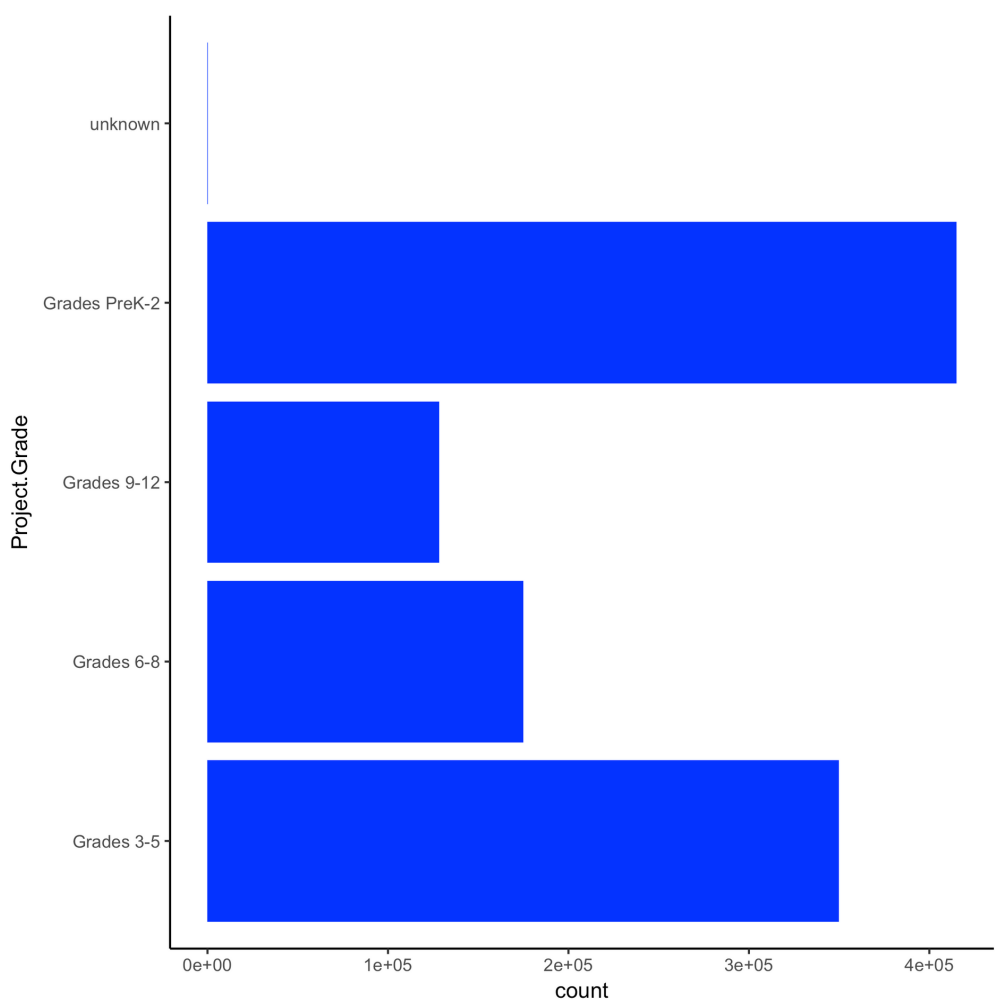
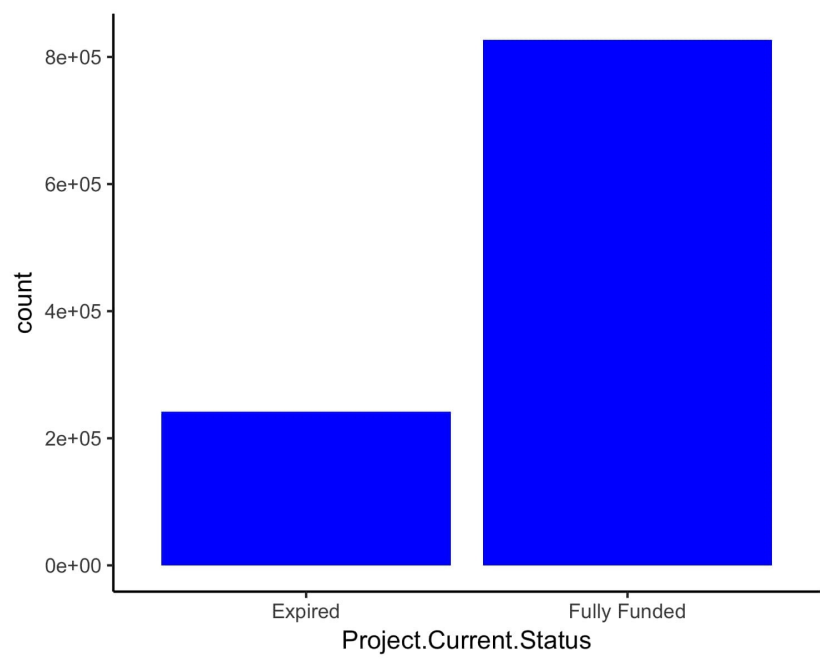
Dependent Variable:

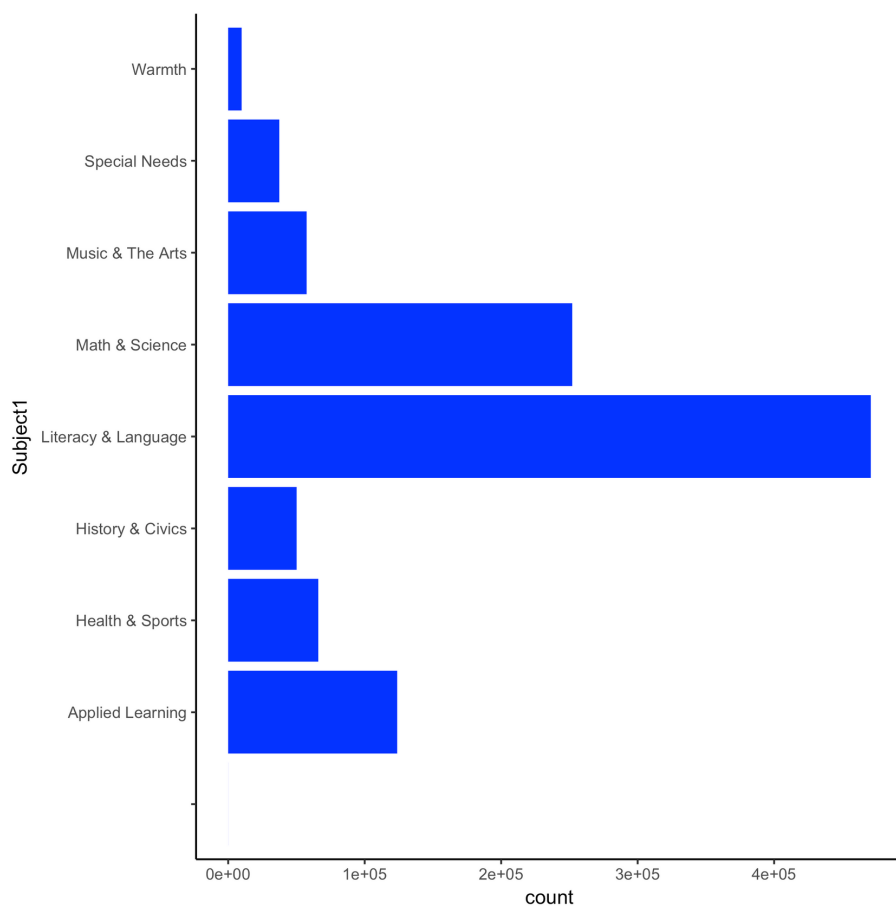
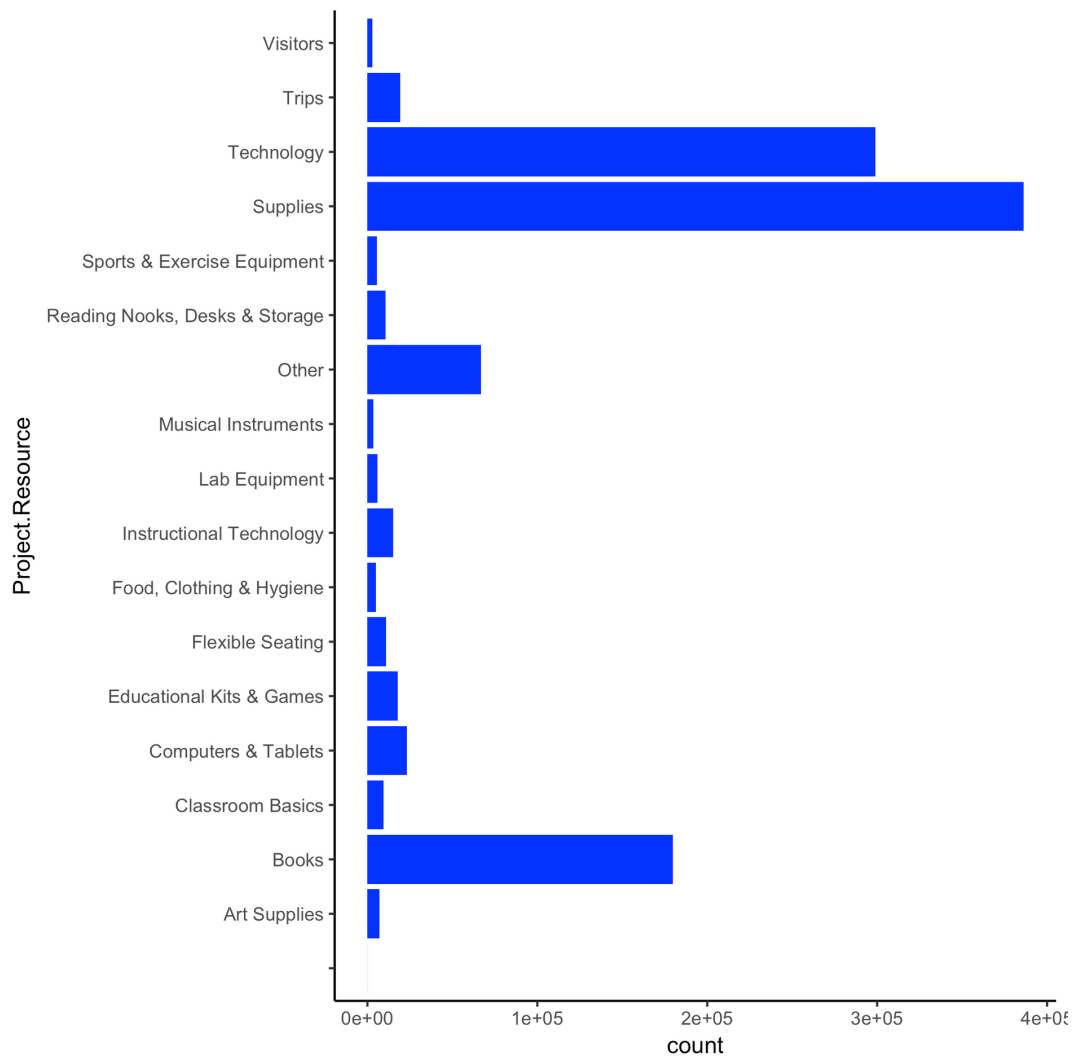
The dependent variable was chosen to be a binary categorical variable: Project.Current.Status answering the question: "Was the project fully funded?" The Project.Current.Status variable in the projects table records the status of the project as either 'Expired', 'Fully Funded' or 'Live'. The rows with 'Live' status were excluded from analysis. The goal of the study is to analyse which factors make it more likely that a project is fully funded.

Independent Variables:

- 1) Project Cost (Continuous variable): The estimated cost of the project
- 2) Grade (Categorical variable): Which grades the project targets. Consists of 5 levels: Grades 3-5, Grades 6-8, Grades 9-12, Grades PreK-2, Grade unknown
- 3) Subject 1 (Categorical variable): The first subject of the project: Applied Learning, Health & Sports, History & Civics , Literacy & Language, Math & Science, Music & The Arts, Special Needs, Warmth
- 4) Subject 2 (Categorical variable): The second subject of the project: Applied Learning, Health & Sports, History & Civics , Literacy & Language, Math & Science, Music & The Arts, Special Needs, Warmth, NA

5) Resource (Categorical variable): The type of resource for which the project seeks funding. Categories: Art Supplies, Books, Classroom Basics, Computers & Tablets, Educational Kits & Games, Flexible Seating, Clothing & Hygiene, Instructional Technology, Lab Equipment, Musical Instruments, Other, Reading Nooks, Desks & Storage, Sports & Exercise Equipment, Supplies, Technology, Trips, Visitors





Analysis

The logistic regression was fitted for the independent variables: Project Cost, Project Grade, Subject 1, and Subject 2

Variable	Category	Model 1		Model 2	
		Coefficient	Standard Error	Coefficient	Standard Error
Intercept		1.81E+00	(0.02102)***	1.25E+00	(0.3745)***
Cost		-6.36E-04	(0.000005415)***	-5.29E-04	(0.000003627)***
Grades	Grades 3-5	base			
	Grades 6-8	1.53E-02	(0.01226)-	5.46E-02	(0.007238)***
	Grades 9-12	1.18E-01	(0.01495)***	1.89E-01	(0.008249)***
	Grades PreK-2	4.93E-02	(0.008266)***	1.47E-02	(0.005685)**
	Grade unknown	-2.89E-01	(0.4624)-	-1.27E-01	(0.315)-
Subject 1	Applied Learning	base			
	Health & Sports	6.24E-02	(0.02056)**		
	History & Civics	9.08E-02	(0.01742)***		
	Literacy & Language	-1.67E-01	(0.01198)***		
	Math & Science	1.10E-01	(0.01283)***		
	Music & The Arts	2.88E-01	(0.05737)***		
	Special Needs	-1.56E-01	(0.04467)***		
	Warmth	1.50E+00	(0.0507)***		
Subject 2	Applied Learning	base			
	Health & Sports	7.95E-02	(0.02992)**		
	History & Civics	1.00E-02	(0.02333)-		
	Literacy & Language	-1.82E-01	(0.01935)***		
	Math & Science	-1.57E-01	(0.0187)***		
	Music & The Arts	1.69E-01	(0.0201)***		
	Special Needs	-1.17E-01	(0.01863)***		
	Warmth	1.19E-01	(0.1794)-		
Resource	Desks and Storage			base	
	Arts and Supplies			1.94E+00	(0.3785)***
	Books			6.82E-01	(0.3745).
	Classroom Basics			2.05E+00	(0.3779)***
	Computers & Tablets			1.56E+00	(0.3752)***
	Educational kits & Games			2.02E+00	(0.3762)***
	Flexible Seating			1.53E+00	(0.3762)***
	Food, Clothing & Hygiene			2.79E+00	(0.3871)***
	Instructional Technology			1.78E+00	(0.376)***
	Lab Equipment			2.06E+00	(0.3795)***
	Musical Instruments			2.39E+00	(0.3851)***
	Other			2.86E-01	(0.3746)-
	Reading Nooks			1.78E+00	(0.3766)***
	Sports and Exercise Equipment			2.43E+00	(0.3812)***
	Supplies			2.86E-01	(0.3745)-
	Technology			-5.03E-02	(0.3745)-
	Trips			7.58E-01	(0.375)*
	Visitors			5.00E-01	(0.3772)-

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Models

A model with both 'subjects' and 'resources' as predictors produces insignificant coefficients for the resource dummy categories because both variables are correlated. Therefore, two models are displayed above.

Results

Cost

A one dollar increase in cost results in a 1.000636 decrease in the odds of the project being fully funded. This is a small but significant coefficient. However, the histogram for cost shows that costs are highly skewed, and the distribution is not normal (Logistic regression does not make normality assumptions so this is acceptable). The coefficient could be very low because some projects have substantial costs.

Grade

The coefficients for grades in both models are similar. Taking model 1, if a project is for Grades 6-8 or Grades 9-12 or Grades PreK - 2, the odds of it being funded are 1.0154, 1.125, and 1.051 over it being funded if it was for Grades 3-5. The coefficient for an unknown grade is insignificant because of the low number of observations. All other categories are highly significant. Therefore, high school projects are most likely to receive funding while Grades 6-8 and pre-school are the other categories that receive a good amount of funding.

Subjects

The coefficients for subjects are similar whether they are primary or secondary, apart from Warmth which is significant as a primary subject, but insignificant as a secondary subject due to a lack of observations. 'History and Civics' is also insignificant as a secondary subject. Taking model 1, below are the ratios (exponents of the logit model coefficients) by which the odds of being funded for the following categories are over the base category Applied Learning, in the case of the primary subject. Warmth is the most important category, with an odds ratio of 4.47 over Applied Learning in securing full funding, Music & Arts, and Math & Science are the next more important. Literacy & Language and Special Needs fare worse than Applied Learning.

Health & Sports	1.0644023
History & Civics	1.0950277
Literacy & Language	0.8458836
Math & Science	1.1167390
Music & The Arts	1.3342212
Special Needs	0.8557089
Warmth	4.4669314

Resources

In model 2, below are the ratios (exponents of the logit model coefficients) by which the odds of being funded for the following categories are over the base category Desks and Storage.

Art & Supplies	6.9628852
Books	1.9778141
Classroom Basics	7.7680076
Computers & Tablets	4.7710778
Educational Kits & Games	7.571135
Flexible Seating	4.6115491
Food, Clothing & Hygiene	16.2707657
Instructional Technology	5.9105096
Lab Equipment	7.8520545
Musical Instruments	10.9445628
Other	1.3306309
Reading Nooks	5.938636
Sports & Exercise Equipment	11.3587571
Supplies	1.3305462
Technology	0.9509718
Trips	2.1336299
Visitors	1.6489407

All categories, except for Technology perform better than Desks and Storage. Food, Clothing and hygiene is the most important category, with an odds ratio of 16.27 over Desks and Storage in securing full funding. Sports Equipment, and Musical Instruments are the next more important. Supplies and Books fare badly.

Caveat

The dataset is imbalanced towards certain classes such as the Resource Categories: Supplies, Technology and Books, and the dependent variable class Fully Funded. This could mean that such categories could be overrepresented in the 'Expired' class.