**BAN 6025 Final Project**

For your final project, you will be analyzing data from Donors Choose. DonorsChoose.org (<https://www.donorschoose.org/> ) is an online charity that makes it easy to help students in need through school donations. At any time, thousands of teachers in K-12 schools propose projects requesting materials to enhance the education of their students. When a project reaches its funding goal, they ship the materials to the school.

While all projects on the site fulfill some kind of need, certain projects have a quality above and beyond what is typical. These projects are designated as “exciting projects” and they have the following characteristics:

* Fully funded
* Had at least one teacher acquired donor
* Had a higher than average percentage of donors leaving an original comment
* Had at least one “green” donation
* Had donations from non-teacher referred donors or had one non-teacher referred donation of at least $100 or had a donation from a “thoughtful” donor

By identifying and recommending such projects early, Donors Choose hopes to improve funding outcomes, better the user experience, and help more students receive the materials they need to learn. In other words, Donors Choose would like to be able to predict which projects are most likely to be exciting projects before the donation period has closed.

Donors Choose would also like to gain a better understanding of its donors in order to help them improve the overall user experience for donors who visit their website and in order to encourage repeat donations.

In order to help Donors Choose meet their goals (as stated above), you have been provided with two data sets:

The file *DonorMerge\_Final.csv* contains information on thousands of projects that have been posted on DonorsChoose.org. For each project, there is an indicator of whether or not it was identified as an “exciting project.” There is also information on the teacher and school requesting the funding, the subject area for the funding, and the cost of the project.

The file *Donations.csv* contains information on individual donors to DonorsChoose.org. For each donor, there is information on the amount of the donation made, the method of donation, and whether or not the donation was made through an individual campaign page.

Data dictionaries for each of these files are included at the end of this document.

**Requirements for your submission:**

For this project, you should think of yourself as a consultant who has been retained to help DonorsChoose.org achieve the two goals highlighted on the previous page. It is up to you to determine how to best prepare the data / explore the data / analyze the data in order to provide them with valuable insight that will help them to achieve their goals. That said, your analysis should:

* take advantage of both of the data sets provided (although they do not need to be combined and can be considered separately in order to address the business goals)
* use at least three of the analytical methods that we have learned in BAN 6025 this semester

You may also use methods learned in other classes as long as your analysis meets the above criteria.

While the majority of your analysis will likely be done in R, you are welcome to use any other software tools that you have available to help you complete this project. You are not limited to R nor are you limited to using only one software package to complete the analysis.

**Your Deliverables:**

**Part 1: Preliminary Analysis Plan**

You will need to submit an analysis plan outlining your plans for preparing, cleaning, and analyzing the data NO LATER THAN Monday, December 5. This document will count for a quiz grade and should include the following:

* Steps you will take for data preparation
  + Variables that need to be removed?
  + Data quality issues that need to be addressed (e.g., missing values, etc)?
  + Variable transformations needed (e.g., dummy variables, standardization, etc)?
  + Other?
* Overview of any exploratory analysis you plan to perform
  + Examine distribution of variables?
  + Exam relationships between variables?
  + Other?
* Analytical methodology for addressing business goals
  + Predicting which projects are exciting
    - What methods are you considering?
    - How will you evaluate whether you have arrived at a good solution?
  + Understanding donors
    - What methods are you considering?
    - How will you evaluate whether you have arrived at a good solution?

**Part 2: Final Report (Due Thursday, Dec 8 at 11:59pm)**

Your final submission will include two parts: a written report and a technical appendix.

The written report should be directed towards the business team at DonorsChoose.org. You can assume that they have a thorough understanding of the data and the business problem at hand, but they have limited understanding of analytical methodology. The report should include:

* Brief introduction and overview of the business problem
* High level discussion of any data quality issues that were addressed
* Presentation of final solution addressing the first goal of identifying exciting projects
  + Business focused presentation of the final analytical solution
  + Discussion of key factors that can be used to identify exciting projects
* Presentation of final solution addressing the second goal of understanding donors
  + Business focused overview of the final analytical solution
  + Discussion of key characteristics of the different types of donors
* Summary / Conclusion
  + Wrap-up and high level recommendation for next steps that DonorsChoose should take based on your findings (note that these could be recommended actions to increase donations or identify exciting projects OR they could be recommendations around what DonorsChoose should do in order to help improve the results of these types of analyses in the future)

NOTE: The written report should NOT exceed 3 pages EXCLUDING any tables or figures that you might choose to include.

For the technical appendix you can either

* Knit your R-code and results into a well-documented html file for submission, or
* Summarize your modeling process and results into a short powerpoint presentation that includes all relevant steps in your analysis and full output for the final solutions that you chose

**Data Dictionaries**

**DonorMerge\_Final.csv**

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| Is\_exciting | Response Variable: TRUE = Project was classified as exciting, FALSE = Project was not classified as exciting |
| Projectid | Project’s unique identifier |
| Teacher\_acctid | Unique identifier for teacher that created the project |
| Schooid | School’s unique identifier |
| School\_ncesid | Public national center for education statistics id |
| School\_latitutde | Latitiude |
| School\_longitude | Longitude |
| School\_city | City |
| School\_state | State |
| School\_zip | Zipcode |
| School\_metro | Metro area |
| School\_district | School district |
| One\_non\_teacher\_referred\_donor\_g | Non-teacher referred is a donor that landed on the site by means other than a teacher referral / link |
| Great\_messages\_proportion | Proportion of unique comments on a page |
| Teacher\_referred\_count | Number of donors that were teacher referred |
| School\_county | County |
| School\_charter | TRUE = school is a public charter school |
| School\_magnet | TRUE = school is a public magnet school |
| School\_year\_round | TRUE = school is a year-round school |
| School\_nlns | TRUE = school is a new leader school |
| School\_kipp | TRUE = school is a college prep school |
| School\_charter\_ready\_promise | TRUE = school is a ready promise school |
| Teacher\_prefix | Teacher’s preferred prefix |
| Teacher\_teach\_for\_america | TRUE = teach for America teacher |
| Teacher\_ny\_teaching\_fellow | TRUE = teacher is a NY teaching fellow |
| Primary\_focus\_subject | Main subject for which materials are intended |
| Primary\_focus\_area | Main area for which materials are intended |
| Secondary\_focus\_subject | Secondary subject |
| Secondary\_focus\_area | Secondary area |
| Resource\_type | Main type of resources requested by a project |
| Poverty\_level | School’s poverty level: highest (65%+ receive free lunch), high (40 – 64% free lunch), moderate (10-39% free lunch), low (0 – 9% free lunch) |
| Grade\_level | Grade level for which project materials are intended |
| Fulfillment\_labor\_materials | Cost of fulfillment |
| Total\_price\_excluding\_optional\_s | Project cost excluding tip that donors give for overhead costs |
| Total\_price\_including\_optional\_s | Project cost including tip |
| Students\_reached | Number of students impacted by a project if funded |
| Eligible\_double\_your\_impact\_matc | Project was eligible for a 50% off offer by a corporate partner (logo appears on project like Starbucks or Disney) |
| Elibigle\_almost\_home\_match | Project was eligible for a $100 boost offer by a corporate partner |
| Date\_posted | Date a project went live on the site |

**Donations.csv**

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| Donationid | Unique donation identifier |
| Projected | Unique project identifier (project that received the donation) |
| Donor\_acctid | Unique id for donor that made the donation |
| Donor\_city | City |
| Donor\_state | State |
| Donor\_zip | Zipcode |
| Is\_teacher\_acct | TRUE = donor is also a teacher |
| Donation\_timestamp | Time donation was made |
| Donation\_to\_project | Amount to project excluding optional tip |
| Donation\_optional\_support | Amount of optional tip |
| Donation\_total | Total amount donated |
| Dollar\_amt | Donated amount in US dollars |
| Donation\_included\_optional\_support | TRUE = optional support was included |
| Payment\_method | What card / payment option was used |
| Payment\_included\_acct\_credit | TRUE: a portion of a donation used accounts credit redemption |
| Payment\_included\_campaign\_gift\_card | TRUE: a portion of a donation included corporate sponsored gift card |
| Payment\_included\_web\_purchased\_gift\_card | TRUE: a portion of a donation included a gift card purchased by donor |
| Payment\_was\_promo\_matched | TRUE: donation was matched 1-1 with corporate funds |
| Via\_giving\_page | TRUE: donation was given through a giving / campaign page |
| For\_honoree | TRUE: donation was made for an honoree |
| Donation\_message | Donation comment / message |