



Completed • \$2,000 • 472 teams

## KDD Cup 2014 - Predicting Excitement at DonorsChoose.org

Thu 15 May 2014 – Tue 15 Jul 2014 (20 months ago)

## Dashboard

## Home

Data

Make a submission

## Information

Description

Evaluation

Rules

Prizes

Timeline

Winners

## Forum

## Leaderboard

Public

Private

## My Team

GitHub

## My Submissions

## Private Leaderboard

1. 'STRAYA
2. DataRobot
3. ChaoticExperiments (KIRAN R)
4. dkay & bmax & James King
5. Triskelion, Yan, KazAnova & Shize
6. Giulio, orchid, Luca & Ben
7. :-)
8. 柳景明
9. 三^ . . ^ SPN
10. Abhishek & Silogram

## Forum (62 topics)

R code for Beating the Benchmark and Getting 0.58431 on Leaderboard  
17 months ago

Congrats to STRAYA, NCCU and adamaconguli!  
20 months ago

projects.csv has been updated  
20 months ago

Is the LB final?  
20 months ago

Private Public Split  
20 months ago

Are team mergers allowed in the last week or day?

Competition Details » Get the Data » Make a submission

## Data Files

File Name	Available Formats
donations.csv	<a href="#">.zip (252.91 mb)</a>
outcomes.csv	<a href="#">.zip (11.18 mb)</a>
sampleSubmission.csv	<a href="#">.zip (793.65 kb)</a>
resources.csv	<a href="#">.zip (193.83 mb)</a>
essays.csv	<a href="#">.zip (402.27 mb)</a>
projects.csv	<a href="#">.zip (64.93 mb)</a>

The data is provided in a relational format and split by dates. Any project posted prior to 2014-01-01 is in the training set (along with its funding outcomes). Any project posted after is in the test set. Some projects in the test set may still be live and are ignored in the scoring. We do not disclose which projects are still live to avoid leakage regarding the funding status.

## File descriptions

- **donations.csv** - contains information about the donations to each project. This is only provided for projects in the training set.
- **essays.csv** - contains project text posted by the teachers. This is provided for both the training and test set.
- **projects.csv** - contains information about each project. This is provided for both the training and test set.
- **resources.csv** - contains information about the resources requested for each project. This is provided for both the training and test set.
- **outcomes.csv** - contains information about the outcomes of projects in the training set.
- **sampleSubmission.csv** - contains the project ids of the test set and shows the submission format for the competition.

## "Exciting" Projects

Exciting projects meet a number of requirements specified by DonorsChoose.org. Note that the term "exciting" is meant as a business construct and does not imply that non-exciting projects are not compelling to teachers/students/donors! To be exciting, a project must meet **all** of the following five criteria. The name in parentheses indicates the field containing each feature in the data set.

teams

players

entries

Data - KDD Cup 2014 - Predicting Excitement at DonorsChoose.org | Kaggle

- was fully funded (`fully_funded`)
- had at least one teacher-acquired donor (`at_least_1_teacher_referred_donor`)
- has a higher than average percentage of donors leaving an original message (`great_chat`)
- has at least one "green" donation (`at_least_1_green_donation`)
- has one or more of:
  - donations from three or more non teacher-acquired donors (`three_or_more_non_teacher_referred_donors`)
  - one non teacher-acquired donor gave more than \$100 (`one_non_teacher_referred_donor_giving_100_plus`)
  - the project received a donation from a "thoughtful donor" (`donation_from_thoughtful_donor`)

You will find this information summarized in `outcomes.csv`, including the boolean value for `is_exciting`.

## Data fields

Below is a brief explanation of the provided data fields. Descriptions of self-explanatory names are omitted.

*outcomes.csv*

**is\_exciting** - ground truth of whether a project is exciting from business perspective

**at\_least\_1\_teacher\_referred\_donor** - teacher referred = donor donated because teacher shared a link or publicized their page

**fully\_funded** - project was successfully completed

**at\_least\_1\_green\_donation** - a green donation is a donation made with credit card, PayPal, Amazon or check

**great\_chat** - project has a comment thread with greater than average unique comments

**three\_or\_more\_non\_teacher\_referred\_donors** - non-teacher referred is a donor that landed on the site by means other than a teacher referral link/page

**one\_non\_teacher\_referred\_donor\_giving\_100\_plus** - see above

**donation\_from\_thoughtful\_donor** - a curated list of ~15 donors that are power donors and picky choosers (we trust them selecting great projects)

**great\_messages\_proportion** - how `great_chat` is calculated. proportion of comments on the project page that are unique. If > avg (currently 62%) then `great_chat` = True

**teacher\_referred\_count** - number of donors that were teacher referred (see above)

**non\_teacher\_referred\_count** - number of donors that were non-teacher referred (see above)

*projects.csv*

**projectid** - project's unique identifier

**teacher\_acctid** - teacher's unique identifier (teacher that created a project)

**schoolid** - school's unique identifier (school where teacher works)

**school\_ncesid** - public National Center for Ed Statistics id

**school\_latitude**

**school\_longitude**

**school\_city**

**school\_state**

**school\_zip**

**school\_metro**

**school\_district**

**school\_county**

**school\_charter** - whether a public charter school or not (no private schools in the dataset)

**school\_magnet** - whether a public magnet school or not

**school\_year\_round** - whether a public year round school or not

**school\_nlns** - whether a public nlns school or not  
**school\_kipp** - whether a public kipp school or not  
**school\_charter\_ready\_promise** - whether a public ready promise school or not  
**teacher\_prefix** - teacher's gender  
**teacher\_teach\_for\_america** - Teach for America or not  
**teacher\_ny\_teaching\_fellow** - New York teaching fellow or not  
**primary\_focus\_subject** - main subject for which project materials are intended  
**primary\_focus\_area** - main subject area for which project materials are intended  
**secondary\_focus\_subject** - secondary subject  
**secondary\_focus\_area** - secondary subject area  
**resource\_type** - main type of resources requested by a project  
**poverty\_level** - school's poverty level.  
 highest: 65%+ free of reduced lunch  
 high: 40-64%  
 moderate: 10-39%  
 low: 0-9%

**grade\_level** - grade level for which project materials are intended  
**fulfillment\_labor\_materials** - cost of fulfillment  
**total\_price\_excluding\_optional\_support** - project cost excluding optional tip that donors give to DonorsChoose.org while funding a project  
**total\_price\_including\_optional\_support** - see above  
**students\_reached** - number of students impacted by a project (if funded)  
**eligible\_double\_your\_impact\_match** - project was eligible for a 50% off offer by a corporate partner (logo appears on a project, like Starbucks or Disney)  
**eligible\_almost\_home\_match** - project was eligible for a \$100 boost offer by a corporate partner  
**date\_posted** - data a project went live on the site

*donations.csv*

**donationid** - unique donation identifier  
**projectid** - unique project identifier (project that received the donation)  
**donor\_acctid** - unique donor identifier (donor that made a donation)  
**donor\_city**  
**donor\_state**  
**donor\_zip**  
**is\_teacher\_acct** - donor is also a teacher  
**donation\_timestamp**  
**donation\_to\_project** - amount to project, excluding optional support (tip)  
**donation\_optional\_support** - amount of optional support  
**donation\_total** - donated amount  
**dollar\_amount** - donated amount in US dollars  
**donation\_included\_optional\_support** - whether optional support (tip) was included for DonorsChoose.org  
**payment\_method** - what card/payment option was used  
**payment\_included\_acct\_credit** - whether a portion of a donation used account credits redemption  
**payment\_included\_campaign\_gift\_card** - whether a portion of a donation included corporate sponsored giftcard  
**payment\_included\_web\_purchased\_gift\_card** - whether a portion of a donation included citizen purchased giftcard (ex: friend buy a giftcard for you)  
**payment\_was\_promo\_matched** - whether a donation was matched 1-1 with corporate funds  
**via\_giving\_page** - donation given via a giving / campaign page (example: Mustaches for Kids)  
**for\_honoree** - donation made for an honoree  
**donation\_message** - donation comment/message. Used to calculate great\_chat

*essays.csv*

**projectid** - unique project identifier  
**teacher\_acctid** - teacher id that created a project

**title** - title of the project

**short\_description** - description of a project

**need\_statement** - need statement of a project

**essay** - complete project essay

*resources.csv*

**resourceid** - unique resource id

**projectid** - project id that requested resources for a classroom

**vendorid** - vendor id that supplies resources to a project

**vendor\_name**

**project\_resource\_type** - type of resource

**item\_name** - resource name (ex: ipad 32 GB)

**item\_number** - resource item identifier

**item\_unit\_price** - unit price of the resource

**item\_quantity** - number of a specific item requested by a teacher