**DONORS CHOOSE: PREDICTING EXCITING PROJECTS AND UNDERSTANDING DONOR'S CHARACTERISTICS**

# 

# 

[**BRIEF INTRODUCTION AND OVERVIEW OF THE BUSINESS PROBLEM**](#_lrmpbrcqt9s1) **2**

[**DATA QUALITY ISSUES**](#_5zljkbt40n6h) **3**

[**PREDICTING EXCITING PROJECTS**](#_1wad9tl0yjvd) **5**

[Business discussion of the solution](#_h88uwzyfhlff) 5

[Discussion of key factors that can be used to identify exciting projects](#_ern3ihv6beo9) 6

[**UNDERSTANDING DONORS**](#_rx9x8w2qxisf) **7**

[Business discussion of the analytical focus solution](#_yvpd4d1pv012) 7

[Donation segments and donor profiles](#_8zo14onwrrg4) 7

[**CONCLUSION**](#_xv2ww3bsf3sv) **9**

# 

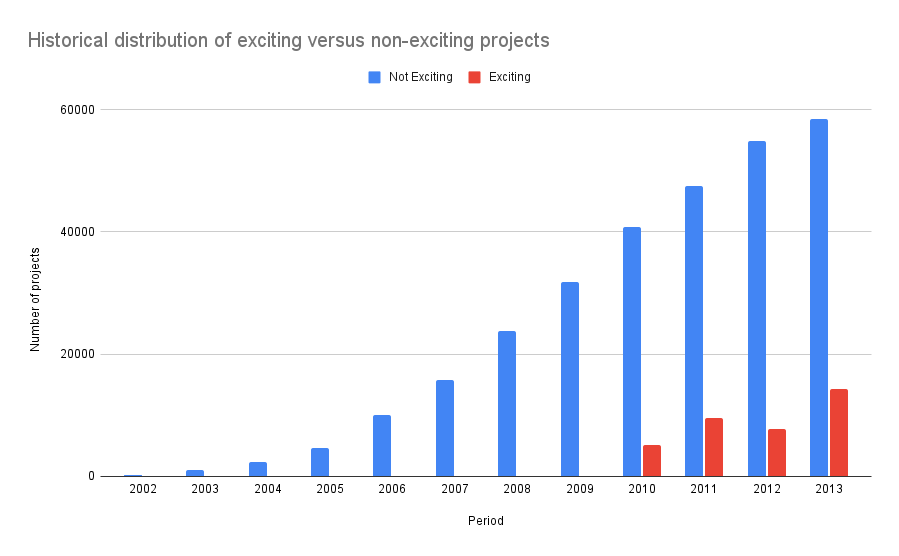
# 

# 

# BRIEF INTRODUCTION AND OVERVIEW OF THE BUSINESS PROBLEM

Our annual rate of exciting project proposals is about 15%. That means, 75 out of every 100 project proposals we get every year are not fully funded and deemed not exciting. Our mission is to help teachers fund projects they need, and if 75% of those projects end up not fully funded, are we achieving our mission? The chart below, showing the historical distribution of exciting projects versus nonexciting projects, visually illustrates the problem we face.

There is a need for us to understand what attributes make some projects exciting and what attributes make some donors repeat donors. We propose to attack this problem using predictive modeling. We will attempt to build a model to predict whether a new project proposal is exciting or not exciting. Lastly, we will attempt to understand what characteristics make our donors repeat donors.



# 

# DATA QUALITY ISSUES

We have collected two datasets for the project: one for the donors, and the other for the projects.

**Data quality issues with the donors' data**

1. We noticed that there was an instance where a donor made a donation to the project which did not reflect in the total donation column. We addressed this issue by recomputing the “donation\_total” column as a summation of the “donation\_to\_project” column” and the "donation\_optional\_support" column.
2. We noticed that there 46 instances where the total donation is zero or negative. In this case, we dropped these instances because in addition to zero all negative total donation amounts, there were null values for the following columns–“payment\_included\_campaign\_gift\_card”, “payment\_included\_web\_purchased\_gift\_card”.
3. We also noticed 8 instances of missing donor account identification. These instances were removed.
4. Lastly, we noticed that the data type for storing the time when a donation was made was inconsistent.

Regarding the foregoing data quality issues, we recommend that Donors Choose write data validation rules as part of its ETL process.

**Data quality issues with the projects’ data**

For the projects’ data, the major data quality issue is missing data for certain columns which we think should not have missing observations. These columns are listed in the table below

| **Attribute or column** | **Number of missing values out of 328018** |
| --- | --- |
| Public national center for education statistics id | 21067 |
| Schoo metro | 40019 |
| School district | 501 |
| Primary focus subject | 23 |
| Primary focus area | 23 |
| Resource type | 26 |
| Grade level | 29 |

Regarding these columns, we recommend that they’re made mandatory fields when a project proposal is submitted.

# PREDICTING EXCITING PROJECTS

## Business discussion of the solution

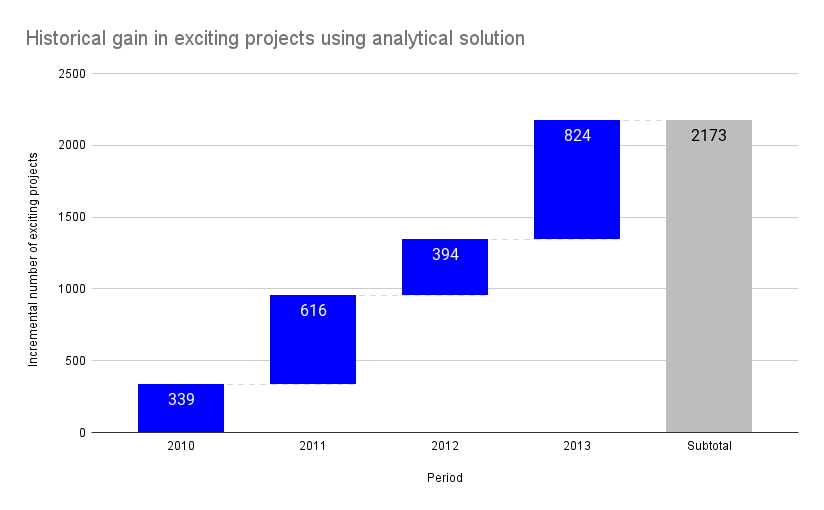
Our proposed analytical solution is to predict ahead of time the probability that a given project is exciting. This gives us the opportunity to increase the visibility of these projects on our “find a class to support page”. Thus, we can rank projects on the “find a class to support page by the magnitude of the predicted probability of excitement of the project. In effect, projects with a higher probability of excitement get seen more and have a greater chance to stand a greater chance of maximizing donation per impression.

Using our analytical solution, we estimated the historical gain if this approach is adopted. The table below shows the historical gain and its associated error rate. The chart following the table is a visual representation of the gain over the historical period from 2010. We’re using 2010 as the starting point because all projects prior to 2010 were labeled not exciting in the data.

**Table:** Incremental gain using the analytical solution to predict exciting projects

| **Period** | **Incremental gain in the number of exciting projects** | **Error Rate** |
| --- | --- | --- |
| 2010 | 339 | 0.76% |
| 2011 | 616 | 1.16% |
| 2012 | 394 | 0.68% |
| 2013 | 824 | 1.17% |

**Chart:** Historical gain in exciting projects using the analytical solution



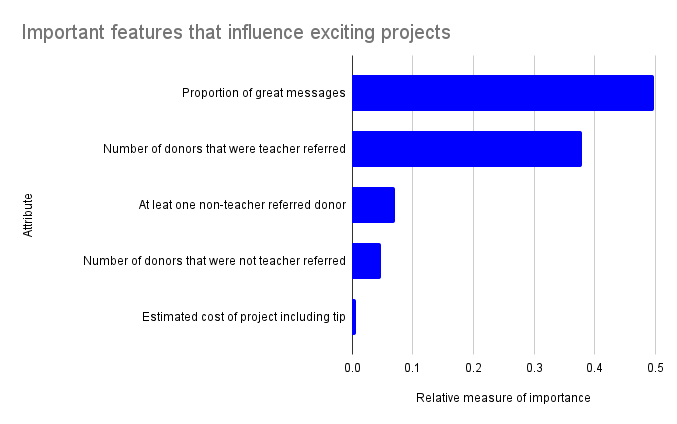
## Discussion of key factors that can be used to identify exciting projects

Chart below shows what features are the most important drivers of exciting projects.

Based on analysis of the data and analytical solution, we observed that exciting projects have at least 25% of great comments on the projects page, at least one non-teacher referred donor, and at least one teacher-referred donor. The relationship between the estimated cost of the project including tip and the exciting project is infinitesimal.

We may use this information to develop an effective guideline for teachers to use to create winning project proposals.

**Chart:** important features influence exciting projects



# UNDERSTANDING DONORS

## Business discussion of the analytical focus solution

To understand donor characteristics, we segmented donations into 6 buckets and explored the donation attributes within each segment. The business value we hope to derive from this solution is to gain a data-driven understanding of donation patterns. We hope Donors Choose will leverage this knowledge to implement targeted donation campaigns, personalize donors’ online experience, and drive repeat donations.

## Donation segments and donor profiles

Donation Segment 1: Credit account-led donations

This segment has about 52,000 unique donors. The total amount of donations within this segment is $2.7m with an average donation of $41. Donations within this segment are mostly done using account credit redemptions. In addition to that, these donations rarely include optional support and they use the donation giving page rarely.

Donation Segment 2: Heavyweight campaign-led donations

This segment is the least populous category with about 28,000 customers but records the highest average amount per donation of $159. The total amount of donations within this segment amounts to $15.1m. Donations within this category use no-cash-received payment methods more than other segments. These donations also do not include optional support donations. Donations within this category use a donation-giving page or a campaign page.

Donation Segment 3: Plain vanilla donations

This is the second most populous segment recording about 157,000 unique customers, and the segment with the highest total amount of donations. Customers within this category use credit-card more than average, often use promo code match promos, rarely include optional support, and rarely use a donation-giving page. The average donation for customers is $90.

Donation Segment 4: Mom & pop donations

Donations within this category are similar to plain vanilla donations except that it has a higher average amount per donation of $98. Donations in this category are made frequently using credit cards and infrequently using a donation-giving page.

Donation Segment 5: Honoree donations

This is the largest donation category by the donor population. It has about 341, 000 customers and an average amount of donation of $35. Donations in this category include optional support donations and are often made for an honoree. These donations rarely use a giving page and account credit.

Donation Segment 6: Featherweight campaign-led donations

Donations in this segment are campaign-led just like those in the heavyweight donations segment. However, the average amount per donation is $81. Also, these donations do not include optional support. Donations in this category use PayPal and Amazon payment methods more relative to other donation segments.

Strategies to attract donors

The most promising donation segments are the heavyweight campaign-led donations, the mom & pop donations, and the honoree donations. Based on the features of these donation segments, we suggest the following marketing strategies to attract more donors;

Donors in the donation campaign-led segments use a donation giving page, hence including them in a mailing list of donation campaigns will get them to make frequent donations.

Mom & pop and plain vanilla donations are characterized by credit card payment methods. Hence, making payments with credit cards seamless for these customers will improve their online experience. more than average.

# CONCLUSION

In conclusion, we highlight the following recommendations mentioned in our discussion of the findings.

1. First, we recommend that Donors Choose rank project proposals on its “find a class to support” page by the magnitude of the predicted probability of excitement of the project. In effect, projects with a higher probability of excitement get seen more and get more donations per impression
2. Second, we recommend that Donors Choose use our findings around features that influence exciting projects to create a guideline for teachers on how to craft winning project proposals. These features are the proportion of great comments on the projects page, at least one non-teacher referred donor, and at least one teacher-referred donor.
3. In addition, we recommend that going forward Donors Choose should only target donors in the campaign-led donations segment with donation campaigns. Also, for donors in the mom & pop as well as plain vanilla donation segments, we recommend that Donors Choose make it easy for them to make donations using credit cards.
4. Lastly, we recommend that going forward, Donors Choose implement data validation rules for its ETL processes to avoid data aggregation errors. In the same light, some data attributes like Schoo metro, school district, primary focus area, and resource type should be made compulsory fields for teachers when they are submitting project proposals.