



Search 

Data Science

Leadership

Professional Development

Prospect Research

Relationship Management

About



DATA SCIENCE

# Predictive Modeling Case Study: From Theory to Application

*By Brian Robertson and Matt Moran and Dr. Lori Thombs | MAY 07, 2020*



*Editor's Note: This article is featured in Best of Connections 2020. Read Apra Editorial Advisory Committee Chair Jessica Balsam's [editor's message](#) to learn more about the top articles of the year.*

---

In February 2018, the prospect development team at the University of Missouri (MU) embarked on a project to review its current affinity scoring system by developing a secondary system. The primary objective was to build an in-house model that solely used predictive modeling techniques on the internal constituent database. The second objective was to use the results to compare any shifts in constituent scores between the current and new affinity scores.

Pursuing a path for building a predictive model with little time, experience and resources was a challenge. This project was an additional responsibility that was performed between other higher priority tasks. As time permitted, ongoing statistical training was performed, meetings with the statistics advisor were coordinated, R Studio software coding skills were fine-tuned, process and results documentation was recorded and implementation plans were outlined. In this article, we would like to share the successes and failures with the larger community of prospect development and data analytic practitioners.

## Process

First, a review of literature was necessary to explore the existing methods used to perform and implement a predictive model. The initial review led us further into another topic relating to the evaluation of historical assumptions: specifically, the comparison of theorized weighting in the existing affinity model versus statistical weighting in the new affinity model. Topics studied included data analytics, nonprofit fundraising, R software coding, statistical techniques, and interpretation for descriptive, predictive and statistical modelling methods. Additionally, case studies from various higher education advancement offices were reviewed. These experiential examples provided important context and direction moving forward with the project.

Second, the data was defined, preprocessed and evaluated before further examination. The original dataset contained 377,606 records or individuals with one dependent *giving* variable and 25 independent *affinity* variables. The dependent giving variable is defined as a continuous variable indicating the amount of dollars each record or individual has given to MU. The 25

independent affinity variables are defined as an identified subset of all the variables available that may describe an individual alum's behavioral characteristics that describe and rank an individual's affinity to MU. These variables were selected based on the literature review and the experience of our advancement colleagues before being imported into the statistical software R Studio for exploration.

Third was the data analysis. Each variable was individually evaluated for errors, reliability and distribution of values. Next, each independent variable was tested against the dependent variable using a correlation test to examine basic relationships between the magnitude and the direction of each variable value. For example, when x went up, did y go up? Or when x went up, did y go down? And if x went up 2% did y go up 10%? This was all part of the pre-processing and evaluation phase to make sure we knew how our data was distributed.

In this process, it became evident that there was a large percentage of non-donor records. This is very common among constituent databases and highlights one of the reasons a secondary affinity scoring system was needed. For example, if approximately 80% of the database were non-donor records (individual people who had zero dollars of giving recorded) and 20% of the database were donor records (individual people who had anything equal to or greater than 0.01 dollars of giving recorded), the non-donor and donor records need to be assessed differently. This is what we expected and adjusted our tests to address the influence of zero giving values among the non-donor records.

Lastly, several statistics resulted from the project. The frequency distributions summarize the variables in terms of which characteristics appeared the most in the dataset. For example, the majority of individuals:

- ➔ were graduates between 1995-2018
- ➔ gave their first gift between 0-14 years after graduation
- ➔ were not involved in any student activities
- ➔ were not involved in a Greek life
- ➔ were not involved with homecoming steering committee
- ➔ were not involved in student government
- ➔ were not Mizzou Alumni Association leaders
- ➔ were living in the Midwest
- ➔ were not phone donors

- ➔ were male
- ➔ were not legacy alums
- ➔ were not parents or past parents
- ➔ had one Mizzou graduate in their immediate family
- ➔ did not have a Mizzou alum spouse
- ➔ did not participate in Mizzou Alumni Association events
- ➔ were not secret honor society members
- ➔ did not have a high grade point average
- ➔ obtained one higher education degree
- ➔ did not have a student loan(s)
- ➔ did not have a student scholarship(s)
- ➔ obtained a bachelor's degree from Mizzou
- ➔ did not give their first gift to athletics

Although this is a very long description, it is accurate to say it represents the majority of the records reviewed in the project. Additionally, the project was able to finalize a linear regression to predict which non-donors with the above behavioral characteristics had a better probability of giving to the university. In other words, the linear regression helps investigate how one thing (predicted giving levels) depends on other things (all the other variables that represent non-donor behavioral characteristics).

## Conclusions

There are many conclusions that can be gleaned from this project. However, the most beneficial were gained by the technology information support group (TISG) and prospect development (PD) teams within the Office of Advancement. Building a predictive model gave the answers and insights to these problems regarding the current affinity scoring system:

- 1. The score was influenced heavily by giving history measures.** The predictive model helped equally weight donors the same as non-donors with the same variables used to describe their behavioral characteristics and likelihood to give. Additionally, it helped represent a “true” affinity based on those characteristics instead of biasing the affinity by their giving history.

- 2. The current affinity score uses theory identified variables versus statistically identified variables.** The predictive model helped to differentiate the importance of the variables used in the scoring calculation. For instance, the initial affinity score has 6 of 16 variables that involve giving measures and the predictive model affinity score has 1 of 9 variables that involve giving measures. Also, the initial affinity score weights the giving measures much higher than the predictive model affinity score. This drastically changed the distribution of affinity scores among donors and fairly represented the affinity of non-donors that shared the same variables or behavioral characteristics.
- 3. The score has not previously been evaluated against another method for validity or accuracy.** The predictive model highlighted the need for systematically testing historical assumptions --some historical perspectives about the definition and process for building the initial affinity scoring system were not well communicated. As a result, unintended assumptions about what the affinity score was *actually* conveying about donor and non-donor giving behaviors was not investigated. Furthermore, it's important to note that the initial affinity score was indeed a benefit to the Advancement office since no affinity score had previously been available. Creating the new affinity score using predictive modeling provided a different perspective and method for testing the validity and accuracy.

## Lessons Learned

Many different outcomes resulted from this project and the biggest were lessons in how to document, review and improve a process. This was realized through many different ways, but the most impactful involved our internal collaboration in TISG and PD as follows:

- 1. Implementing the initial affinity score as a complement with new affinity score.** Now that two different affinity scores were available for review, it was possible to measure what each score was doing, how the measurement was different, and what the individual biases were for each. The initial affinity score relied on giving amounts, giving velocity, giving recency and giving thresholds for bonus points. Non donors could not score anything. The new affinity score relied heavily on the nine independent giving variables explaining the reason giving went up or giving went down with a predicted affinity score. There is always a possibility that a variable was left out or a variable is weighted too much. Individually, the two different affinity scores are helpful, but together help to identify the best prospects by targeting donors who give and who exhibit the behavioral characteristics of more likely donors and non-donors.

- 2. Motivating the need for a team specializing in data analytics.** Specific skills, experience and training are needed to fully incorporate the philosophy and benefits of a predictive model into an Advancement office. Completing this project in approximately 18-24 months helped resituate our team's perspective on data analysis and the need for procuring a specialized team to perform data analytics. Subsequently, the Office of Advancement at the University of Missouri has initiated the development of an Advancement Analytics team. Our team views this as a strategic investment in how the culture and philosophy of data-driven decisions can be incorporated into other teams and departments.
- 3. Reinforcing the documentation of historical and current processes.** Large amounts of information and data are assembled to create the scores and calculations used. The initial affinity score was developed in 2016 and staff changes occurred between then and 2018 when a review was initiated. Thorough documentation on the calculation process was available through computer programming, but less existed for the reasons the variables were selected, the justification for how it aligned with industry best practices and how the affinity should be interpreted. During the development of the new affinity score, careful measures were taken to document the scientific steps for the calculations and the translation of that information into an easy to understand set of handouts. Furthermore, it highlighted the need for a Data Governance committee that could guide the process of defining and organizing the technical information for evaluation and future evaluations.

Overall, the project helped to remove barriers regarding process improvement and investigated unintended bias regarding historical assumptions of non-donor affinity. The realized success was being able to evaluate this information from a different perspective without condemning previous processes. In fact, it opened more opportunities to collaborate and strengthened our team's understanding in how important it is to all be unified when communicating about technical details and projects.



Data Science







## Brian Robertson

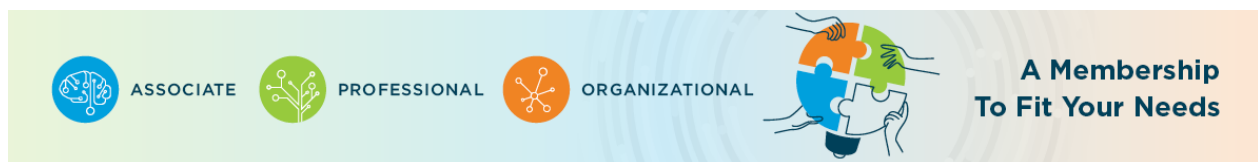
*Prospect Development Analyst, University of Missouri*

**Brian Robertson** is a prospect development analyst at the University of Missouri. He works with fundraising professionals to discover philanthropy that embodies, supports and connects higher education with student development and success. Brian has supported educational opportunities in a variety of experiences from extension outreach, teaching curriculum and prospect research. He has honed his expertise in data analysis throughout this time from publishing research articles in energy conservation, developing curriculum in Precision Agriculture, and implementing predictive modeling techniques that measure constituent affinity in prospect development.

Brian received two bachelor's degrees in Agricultural Systems Management and Agricultural Economics and a master's in Agricultural Systems Management from the University of Missouri. In addition, he completed doctoral courses at the University of Missouri before reallocating his career emphasis on development research and data analytics.

## Matt Moran

## Dr. Lori Thombs



## MORE RECENT ARTICLES FROM CONNECTIONS





## **Apra Fundamentals: An Energizing Way to Build Your Prospect Development Career**

## **Painting the Big Picture: Tips for Prospecting in the Arts**



## **Closing Time: Calculating Risk to Encourage Proposal Velocity**

330 N. WABASH AVE. SUITE 2000, CHICAGO, IL 60611  
312.321.5196 | [INFO@APRAHOME.ORG](mailto:INFO@APRAHOME.ORG)



[LOGIN](#)

