Alumni Scoring System

of Worcester Polytechnic Institute's Alumni Database

A Major Qualifying Project Report

Submitted of the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

_____

Kirsten B. Murphy

_____

Onalie L. Sotak

Date: **APRIL 26, 2007**

Approved:

_____

Professor Arthur C. Heinricher, Advisor

_____

Professor Jon P. Abraham, Advisor

_____

Professor Jayson D. Wilbur, Advisor

Abstract

The primary goal of this project was to construct and evaluate a scoring method for ranking individuals in a database where those more likely to donate receive higher scores. The created spreadsheet takes donor information and generates an assigned score from 1-20. A manual for the spreadsheet was also created, enabling the WPI Office of Development and Alumni Relations to rank selected alumni in order of their likelihood to donate in the future.

Acknowledgments

We would first like to start by thanking our three advisors Professors Arthur Heinricher, Jon Abraham, and Jayson Wilbur for their support, guidance, and tremendous patience with us throughout this year. The different insights provided throughout the project created a challenging, yet overall rewarding experience and we appreciate their continued effort to educate and push us to succeed.

We would like to give a huge thanks to the WPI Office of Development and Alumni Relations for providing assistance and access to the data used in this project. We would especially like to thank Dexter Bailey (Vice President for WPI Office of Development and Alumni Relations) and Lisa Maizite (Assistant Vice President for WPI Office of Development and Alumni Relations) for their cooperation, trust, and continued sincere interest throughout this project was helpful as well as rewarding.

We would like to thank Professor Bogdan Vernescu for initially identifying this project opportunity for us. Huge thanks also go to Yi “Danny” Jin for his work on this database and assistance in answering questions and guidance and conclusions that were used as a basis for a majority of this project. While not in direct contact with us during this project, Professor Joseph Petruccelli provided guidance for Jin and we would like to thank him for his previous work as well as his availability to us during the project if we had needed additional assistance.

We lastly would like to thank our friends and family who supported us throughout the project and provided both insight and comfort during this year.

Table of Contents

Table of Figures

Table of Tables

1. Introduction

Universities spend time and money to collect and organize alumni information and this project aims to reward WPI for its effort in this area. Alumni, as well as friends and family of alumni, are an important source of support for the university, both financially and non-financially. The cost is increased by the task of contacting alumni for donations. While this project focuses specifically on fundraising at an academic institution, this is a problem for any organization that tries to identify likely donators.

This project uses information in the WPI donor database to rank each individual's likelihood of donating. The goal of this project was to build a scoring algorithm to identify likely donors and implement that algorithm in a software application that the WPI Office of Development and Alumni Relations can use to prioritize individuals for fundraising activities.

A spreadsheet was created to implement the *Donor Score System* that could measure an individual's likelihood of donating based on a variety of factors provided in the database that coincide with past donation trends. Advances in technology have made it possible to store and update massive amounts of records easily. More data does not mean more information. Statistical analysis of a dataset builds models that "fit" the data and provide information about the data, and a scoring system allows for the organization of information drawn from a dataset by ranking alumnus. This score is determined from the information provided in the WPI database and the individual scores that are assigned to different recorded information about an individual. Once the donor score system identifies a person as likely to donate, it then assigns that individual a score from 1 to 20.

With this score, the program produces a list of ranked individuals in the database and provides their contact information.

A manual for the spreadsheet was created and was provided to the Alumni Office with instructions for using the donor score system. This manual explains how the spreadsheet works and how it provides a quick and effective means of organizing the database to produce the best prospects to contact first. The donor scoring system provides the WPI Office of Development and Alumni Relations with more time to focus on alumni events and less time sorting through unorganized information or blindly calling alumni with little to no likelihood of donating. With near 25,000 alumni worldwide, WPI can use the donor score system and mine its database information, in order to reduce costs spent on finding and contacting alumnus, and increase donations by using the *Donor Score System* to efficiently identifying and in turn contacting those records in the database who are more likely to donate.

2. Background

This project goal is to extract information from a large data set and use it to help identify donors. Sampling is a method commonly used as a way to reduce the size of a dataset to obtain a manageable and representative set of data. Analysis of the original large datasets relies heavily on computational power, and this type of power is now available to aid in statistical analysis and statistical modeling of data. Computers and software are the tools used to explore large data sets. Statistical analysis usually assumes that variables in a dataset are related in some mathematical way and statistical tools can find these relationships. For example a person's age and lifestyle can be used to predict mortality, and similarly the same type of characteristics can be used to predict donation activity.

Another WPI student, Yi Jin, analyzed the WPI database to identify 24 variables that were related to donation behavior (Jin 2006). His analysis assumed that donation behavior is a function of factors in the database. The following equation 2.1 is a linear regression model with *p* predictors:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon \qquad (2.1)$$

where Y is the value of the dependent variable, β_0 is the intercept, β_i is the coefficient for the i th independent known constant X_i ($i = 1, 2, \ldots p$) and ϵ is the independent random error term (Kutner, Nachtsheim, and Neter, 2005). Using equation (2.1) for a set of data from a donor database, each X_i represents one of the known independent variables in the database. Each β_i would represent the relative affect that variable i has on donation behavior if all other variables were held constant. While no linear regression was done in

this project to determine scores, it was used by Jin to determine the 24 variables that influence donation behavior the most (Jin 2006).

This project relies more heavily on the scoring system methods similar to those developed by Peter B. Wylie (2004). This section will introduce data mining; what it is, what it is used for, and how it relates to this project. It will investigate the scoring systems created by Wylie as well as a modified score system for use on WPI database. Finally, it will describe metrics and how they were used in this project to help rate the scores used in the donor score system.

2.1 Data Mining

Data mining relies heavily on computational power and solves problems by analyzing already present data from databases (Frank 2000). "It [data mining] is not so much a single technique as the idea that there is more knowledge hidden in the data than shows itself on the surface" (Adriaans 1996). For this project, the database being investigated contains information pertaining to WPI's alumni, as well as family and friends of alumni.

The database contains 102 variables for 48,604 individuals. A large part of data mining revolves around not just the access to information but the preparation of the data being analyzed. "One objective of data preparation is to end with a prepared data set that is of maximum use for modeling, in which the natural order of the data is least disturbed, yet that is best enhanced for the particular purposes of the miner" (Pyle 1999). This quote discusses not only the importance of an organized database, but also the importance of data cleaning and preparation. Recognizing that there was hidden information in the database, and then cleaning and preparing the database were introduced through research

on data mining and will be discussed in Chapter 3 when a more detailed summary of the database is discussed.

2.2 Scoring System

A technique known as *list scoring* can be used to rate factors according to their influence on donation behavior, and a model can be created from this information to rank the individuals based on their factor values (Wylie 2004). The score assigned to each factor is guided by patterns in the dataset, and the focus of a score system is to rank individuals based on their donations, not on predicting the amount of donation made by an individual. Therefore a score system essentially rearranges and organizes a dataset based on donation behavior, and *list scoring* is simply a means of organizing the assigned scores into list form.

Some factors will have a relationship with donation behavior, some will not. Wylie identified 3 important factors in his example but some applications may require more. Once these factors are identified, Wylie used them to create a scoring system. A score of 0 was given to the portion of the individual variable that didn't coincide with positive donation behavior, and a 1 was given to value of the variable that did appear to coincide with positive donation behavior. For example, Wylie found that individuals with their e-mail listed donated more than those who did not have their e-mail listed. Wylie therefore would give anyone with their e-mail listed a score of a 1 and anyone who left this variable blank would receive a score of 0.

The scores for individual factors are summed to obtain a total score for each individual in the data set. Wylie's example uses only 3 variables causing a score of a 0 to signify that the individual in question does not fit into any of the positive category three

factors that Wylie had previously identified as coinciding with positive donation behavior listed, while a score of a 3 signifies an individual has all three of the factors that Wylie has previously identified as coinciding with positive donation behavior listed.

With the score system established for the first half of the database, the final step is to apply the score system on the second half of the data. This step may appear redundant seeing as Wylie has access to all the information and it would seem that using all the data when creating the score system would achieve the best conclusions using all available information. However, Wylie explains that "when you do a project like this, it's easy to take advantage of the idiosyncrasies of one sample to generate a scoring formula/segmentation schema that looks great on that particular sample, but turns out to be not so great on another sample. We want to see if the relationship between scores and giving we get in one sample looks as good (or almost as good) on another sample. If it does, then we can be confident we're headed in the right direction" (Wylie 2004).

The set of data that is used to create the score system is commonly called the training data while the second half is called validation or test data. The first half of the data is used as a training set to fit the model, and the remaining 50% is used to assess how accurate the model for the first half of the data fits the second half of the data (Hastie, Friedman, and Tibshirani, 2001). This division of data into two sets is used to help decide between different models on a set of data. A good model would return similar results on the second half of the data as were found for the first half of the data. For this project a good model would not only support the scores for the scoring system, but also identify that while there may be idiosyncrasies in the data they had no influence in determining the scores.

The model for this project is the scoring system and the scores assigned to variables are what are determined through the modeling of the first half of a data set and assessed with the second half of the data set. Sampling is important because "it guides the choice of learning method or model, and gives us a measure of the quality of the ultimately chosen model" and discussion of how data was sampled for this project will be discussed in Chapter 3 (Hastie, Friedman, and Tibshirani, 2001).

To determine important factors Wylie spent a large amount of time organizing data in the database into charts or graphs to become familiar with the data, and then used his familiarity with the database to help him decide on important factors. A frequently asked question in his book revolves around the guidelines for deciding whether or not a difference between two factors is statistically significant. He explained that when analyzing a table or graph, if there is something that immediately jumps out, then it should be studied further in depth to determine its practical significance. He also notes that the factors chosen were primarily found through intuition, and although his example only used three variables, there is no restriction on the number of variables that can be investigated.

Before assigning the individual scores for the variables in the score system, some exploration of the data needs to be performed to first identify good factors. While a factor may appear to be highly correlated with donation behavior; donation activity is extremely sensitive. Sensitivity is due to the occasional large donations made by individuals through money left in a will or a random philanthropist. With donations made to universities generally not being millions of dollars, a random million dollar donation made by a single male mechanical engineer essentially disrupts previous

predictions by suddenly identifying any single donators, any male donators, and any mechanical engineer donators as being extremely likely to donate, when in reality this may not be the case.

While a good scoring system should assign scores to individuals so that an individual with a score of a 1 is donating less than an individual with a score of a 2 and so on, a good score system does not guarantee that the scores will always coincide perfectly with donation behavior. To test how well a score system fits a set of data, metrics are used to evaluate the assigned scores for individual factors and ensure that no variables are incorrectly correlated with donation behavior. While the score system itself determines how good of a chance there is for an individual to donate based off their factors, metrics essentially determine how good is the score system itself in assigning the individual scores.

2.2.1. Donor Score System

The *donor score system* was set up the same way as Wylie's system, with the same objective of assigning scores to records in a database based on their variables. This section will explore some of the differences between the Wylie system and the donor system as well as discuss why a scoring method is such an effective means of determining donation activity. A description of how the score system is used in the Excel spreadsheet is explained in Chapter 4, and the detailed steps for using the Excel spreadsheet are explained in the Users Manual in Appendix G.

The principal difference between the score that Wylie's system assigns and the score that the donor system assigns lies in the assigning of scores not only to donation behavior but also relative to other variables in the score system. For example, if Wylie

identified being married as having a positive association with donation behavior, and being a man as correlated with a positive donation behavior, than each married male would receive a score of 2, and each single female would receive a score of 0. While this still organizes individual's donation behaviors based on their marital status and gender, the donor score system goes an extra step by then investigating how the variables marital status and gender relate to one another.

If married individuals have a high probability of donating, say someone who is married donates 100 times more than someone who is single, but men only donate twice as often as women, then the donor score system takes this into consideration. While it would still be important to give a male a higher score than a female, it would appear that a female who was married is actually more likely to donate than a male who was single because the marital status variable is significantly more influenced by donation behavior than gender. The donor scoring system allows for a variety of scores, therefore instead of limiting the score to either a 0 or 1, a smaller score could be assigned for gender and a larger score could be assigned for marital status.

Another difference lies in the number of variables involved in Wylie's example, which only investigated 3 while the donor system contains 24. The quantity of the variables involved in the system is not as important as the type of variable involved. Wylie's 3 variables are "blank or non-blank" variables, while the donor system deals with both blank and non-blank variables and multiple category variables. Detailed explanations and examples of both blank and non-blank variables and multiple category variables will be discussed further in Chapter 3.

An example of a table that Wylie provided in his book that illustrates a blank or non-blank variable, which Wylie refers to as listed or not listed, can be seen in Table 1. In Table 1 Wylie examined individuals in his database who had provided their e-mail or had no e-mail listed.

Table 1: Distribution of Donation Size by E-Mail Variable

	No E-mail Listed	E-Mail Listed	Total
$0	1,215	523	1,738
$1-$250	1,010	594	1,604
$251 or more	953	705	1,658
Total	**3,178**	**1,822**	**5,000**

He found that approximately one third of the individuals had their e-mail listed and the remaining did not. Table 1 provides information about e-mail listing by donation size. While the variable e-mail in Wylie's example was only investigated based on whether it was listed or not in the database, the variable marital status in the WPI database was investigated according to the details of this specific variable, not just simply whether someone listed their marital status or not. For the donor score system, an example of a table used to illustrate the marital status variable from the WPI database is in Table 2.

Table 2: Distribution of Donation Size by Marital Status Variable

	Married	Single	Other	Blank
Total # Of	12,899	10,260	728	140
Total $ Donated	$66,289,522.44	$2,728,014.51	$7,203,267.70	$4,641,256.02
Total # Donated	9,929	3,797	604	34
Percent Donated	76.97%	37.01%	82.97%	24.29%
Average Donation	**$5,139.12**	**$265.89**	**$9,894.60**	**$33,151.83**

2.2.2. Metrics

The "Metrics" worksheet is setup to help analyze how accurate the scoring system is by using three different techniques: the R-Squared method, Sum of Slopes and the "O.K" method. Each of these techniques has a way of giving a score as to how accurate

the current scoring system is using the current factors. The R-squared method determines how closely the data compares with a best-fit line. The other two methods are there to compare how increasing the values are. A prefect scoring system when comparing the average donation with each score bucket should be increasing as the score is increasing. Meaning that the higher a score a person has, the more likely they are to donate more money. Using these metrics will hopefully help find the best score factors for each of the factors that will maximize each of the metrics.

Metric #1 is the R-squared Technique. The R-squared value is a descriptive measure between 0 and 1. The closer it is to one, the better the model explains the variation in the data. A value of R-squared equal to one implies that the regression provides perfect predictions. The formula for R-squared is $R^2 = 1 - (SS_{Error} / SS_{Total})$. This technique is good to see how much of the donating pattern can be explained by the score "bucket".

Metric #2 is the Sum of Slopes Technique used for the Average Donation. This technique is used to make sure that the function is increasing. It calculates the sum of all slopes and then divides each slope by the number of people in each of the two associated score "buckets", i.e. $_{=1}\Sigma^{max-1}$ $([(k_{i+1} - k_i)(\#k_i + \#k_{i+1})]/600)$. Where k_i is the average donated in the "bucket" where the score is equal to i, and $\#k_i$ is equal to the number of people that are in this i^{th} score "bucket". In the sample population this is finally divided by 300 which is two times the total number people in the population. The optimal for this metric would be for this value to be as high as possible, because the higher the score, the higher the sum of the slopes are, meaning that hopefully the relationship between the score "buckets" and the average donation is an increasing one.

Metric #3, was named the donor technique and is also used to see if the function is a mostly increasing one. The formula for this method is:

For $j = 3 \ldots max$

If $[(k_{j-1} + k_{j-2})/2] < k_j$ → then (+1)(# people in bucket)

If $[(k_{j-1} + k_{j-2})/2] < (k_j)(0.9)$ → then (±0)(# people in bucket)

If $[(k_{j-1} + k_{j-2})/2] > (k_j)(0.9)$ → then (–1)(# people in bucket)

These values are all added up and then divided by the total number of people in the population minus the number of people in score bucket 1 and score bucket 2. The goal of this is to get the final value as close to 1 as possible. This basically is taking the average of the average donation for the previous two scoring "buckets" and comparing it to the following average donation. If the averages are always increasing then the score will be 1.

Metric #4 is also the Sum of Slopes Technique for the percentage of people donating. This technique is used to make sure that the function is increasing. It calculates the sum of all slopes and then divides each slope by the number of people in each of the two associated score "buckets", i.e. $_{=1}\Sigma^{max-1}([(k_{i+1} - k_i)(\#k_i + \#k_{i+1})]/600)$. Where k_i is the percentage of people donating in the "bucket" where the score is equal to i, and $\#k_i$ is equal to the number of people that are in this i^{th} score "bucket". In the sample population this is finally divided by 300 which is two times the total number people in the population. The optimal for this metric would be for this value to be as high as possible, because the higher the score, the higher the sum of the slopes are, meaning that hopefully the relationship between the score "buckets" and the percentage of people donating is an increasing one.

Metric #5 is also used to see if the function of percentage of people donating versus score bucket is a mostly increasing one. The formula for this method is:

For $j = 3 \ldots max$

If $[(k_{j-1} + k_{j-2})/2] < k_j$ → then (+1)(# people in bucket)

If $[(k_{j-1} + k_{j-2})/2] < (k_j)(0.9)$ → then (±0)(# people in bucket)

If $[(k_{j-1} + k_{j-2})/2] > (k_j)(0.9)$ → then (–1)(# people in bucket)

These values are all added up and then divided by the total number of people in the population minus the number of people in score bucket 1 and score bucket 2. The goal of this is to get the final value as close to 1 as possible. This basically is taking the average of the percentage of people donating for the previous two scoring "buckets" and comparing it to the following percentage. If the percentages are always increasing then the score will be 1.

2.3. Summary

In this section, concepts of data mining, scoring systems, and metrics were examined. This section provided some explanation of data mining, and it examined the research of Wylie and his use of a scoring system, as well as an introduction to the donor scoring system which is simply an extension of Wylie's system. Other topics that Wylie has investigated related to alumni data mining were also introduced, and the adaptation of his score system is explained in relation to this project. How to determine scores of individual variables and how metrics determine what the best score for a variable is also explained. Data mining provides an efficient and unbiased means of dealing with large quantities of information, Wylie uses data mining and intuition to create a score system, metrics are used to rate a score system, and the created donor score system organizes members of a database according to their donation behavior.

This project encounters many of the same problems that researchers have to deal with daily when analyzing massive amounts of data. WPI is one of many universities that seek to use its database as a tool to provide information about donation behavior of their donors, and to find a more time and cost efficient means of determining what potential donors of their database are most likely to donate. This section focuses on a score system and how a modified version of Wylie's score system can be implemented with the WPI donor database. With metrics used to rate the scores assigned in a score system, the ranking of donors in the database can provide a reliable means of organizing records of a database based on their likelihood to donate.

3. Exploring the Data

WPI opened in 1865 with its first graduating class in 1871. There have been alumni for the past 135 years however the database provided by our sponsor contains donation activity starting in 1983. There are some individuals in the database who donated before 1983 but the 1983 donation records are the cumulative amount given through 1983, and although there is information for 2007, this year is not complete therefore donation activity it not accurate for the entire fiscal year. This section provides a description of the variables in the data, how they were grouped, and the conclusions that were drawn from the different types of variables. Some conclusions about key parts of the database will also be mentioned, as well as references to the majority of the tables in the Appendices of this report.

The WPI data contains 48,604 individuals and a total of $99,387,742.12 in donations. The first column in the data contained this identification number for each individual. The rest of the database includes 101 additional columns containing an assortment of personal information as well as donation behavior for years 1983 to 2007. Of the 48,604 individuals, 24,204 (49.80%) had made a donation and the average donation for the database was $2,044.85. The average donation for the individuals who did donate was $4,106.25.

While WPI has information about donors on file, especially if the donor attended WPI at any point, the majority of the information found in the real database is self-reported. Self-reported data can be unreliable however for this project it was assumed that any bias in self reported data did not have an effect on conclusions drawn in determining scores (Burstein 1985).

Table 3 provides the data extract key that was given by the alumni office and later modified by Yi Jin (Jin 2006) explaining all 102 personal identification numbers in the database. While there are only 67 variables listed below, row 66 actually contains donation activity per fiscal year from 1983 to 2007 and row 67 contains donations in gift club per fiscal year 1996 to 2007. Each fiscal year for these two variables is allotted its own column in the database, with donation activity recorded in 25 columns and gift club recorded in 12 columns, completing the 102 columns in the database.

Table 3: Modified Data Extract Key

1	PERSON_NUM	Person number for data extract
2	CATEGORY	Constituents best (primary) donor category
3	GENDER	M/F/NA
4	BIRTH_YEAR	4-digit year of birth
5	MARRIED	Married/Single/etc.
6	LEGACY	Yes: the person's admission record indicated a legacy relationship (no details available)
7	GPA [1]	Number for those available, spaces for those unavailable, "N/A" for those not applicable
8	BS_YEAR	WPI B.S. year
9	BS_MAJOR	WPI B.S. major
10	MS_YEAR	WPI M.S. year
11	MS_MAJOR	WPI M.S. major
12	PHD_YEAR	WPI Ph.D. year
13	PHD_MAJOR	WPI Ph.D. major
14	CERT_YEAR	WPI certificate year
15	CERT_MAJOR	WPI certificate major
16	HONOR_YEAR	WPI honorary degree year
17	HONOR_DEG	WPI honorary degree
18	NON_WPI_DEG	value if known (formatted as institution : degree code : year : major)
19	WPI_SPS	Yes: the spouse is a constituent
20	NUM_OF_CHILD	Count of children
21	PREF_CLAS	Preferred class year
22	HAD_SCHOLARSHIP	Yes: had scholarship while at WPI
23	PRES_FND	Yes: a Presidential Founder
24	LIFETIME_PAC	Yes: a lifetime PAC[2] member

25	TRUSTEE	Yes: a trustee of WPI
26	ADM_VOL	Yes: involved in alumni/admissions
27	CLS_AGENT	Yes: involved in solicitation structure
28	REUNION	Yes: constituent attended reunion(s)
29	ALUM_VOLUNTEER	Count of distinct number of activities (involved in/as department advisory board, gold council, …, 42 possibilities)
30	ALUM_CLUB	Count of distinct number of activities (Tech Old Timers, Polyclub, …)
31	ALUM_LEADER	Count of distinct number of activities (involved in/as class officer, trustee search committee, fund board, …, 30 possibilities)
32	FRAT	Name of fraternity/sorority, blank otherwise
33	SPORT_COUNT	Count of varsity sports
34	VARSITY_SPRTS	Concatenated list of varsity sports
35	WPI_AWD	Yes: constituent received this award at WPI
36	TAYLOR_AWD	Yes: constituent received this award at WPI
37	SCHWIEGER_AWD	Yes: constituent received this award at WPI
38	GODDARD_AWD	Yes: constituent received this award at WPI
39	GROGAN_AWD	Yes: constituent received this award at WPI
40	BOYNTON_AWD	Yes: constituent received this award at WPI
41	WASHBURN_AWD	Yes: constituent received this award at WPI
42	RES_CITY	Home city (permanent address)
43	RES_STATE	Home state code
44	RES_ZIP	Home zip code (5 or 9-digit format)
45	RES_COUNTRY	Home country
46	TITLE	Job title if known, blank if unknown
47	WORK_CITY	Work city (business address)
48	WORK_STATE	Work state code
49	WORK_ZIP	Work zip code (5 or 9-digit format)
50	WORK_COUNTRY	Work country
51	STU_CLUB	Count of clubs (Outing Club, Science Fiction, Sport Parachute, …)
52	STU_ARTS	Count of arts and literature organizations (Masque, Pathways, Peddler, …)
53	STU_INTL_CLUB	Count of international clubs (Indian Students Association, …)
54	STU_CLUB_SPORT	Count of club sports (scuba, bowling, autocross, …)
55	STU_PROF_SOC	Count of undergrad professional societies
56	STU_MUSIC	Count of music band: glee club, baker's dozen, …
57	STU_CLS_OFF	Count of class officer (freshman, sophomore, …)
58	STU_SCH_INVOLVE	Count of school involvement (student activities board,

		resident advisor)
59	STU_SPEC_PROG	Count of special programs (undergraduate employment program, exchange, …)
60	STU_INTRAMURAL	Count of intramural sports (basketball, softball, table tennis, …)
61	STU_HONOR_SOC	Count of honor societies (Pershing Rifles, Sigma Mu Epsilon, Skull, …)
62	STU_PROJECT_CTR	Project Center Info (from the student courses)
63	ALU_PROJECT_CTR	Project Center Info (from alumni activities)
64	GRAD_DISTINCTION	H: graduated with high distinction, D: graduated with distinction, and blank
65	ALUM_CONTACTS	Contacts made as an alumnus (phone calls, personal visits, …)
66	FISCAL_YEAR_X (X: 1983~2007)	Total gift and memo for the specific fiscal year [3]
67	GIFT_CLUB_X (X: 1996~2007)	Gift Club designation for the specific fiscal year

[1]. WPI Undergraduates do not have a "true" GPA. Standard "numerical equivalent for passed courses" approved by the faculty was used.

[2]. PAC stands for President's Advisory Council.

[3]. Note the 1983 number is a cumulative amount given up through 1983 when the values were loaded into "Banner". Also note that 2007 data only contains data from the first few months of the fiscal year.

3.1 Focus Population

Danny Yi's research found only 24 of the 101 variables statistically significant when he performed multiple regression analysis. A variable called CATEGORY assigns a record in the database a title that best categorizes their relationship to WPI. There are 18 different categories in the data, including ALUM which refers to a recipient of a Bachelor's Degree, PRNT which refers to a parent of an Alum, FRND which refers to a friend of the institution, or GRAD which refers to a recipient of a Graduate Degree. The CATEGORY variable was divided into the 18 different categories and the donations made in each category. It was found that individuals under the ALUM category had the most complete information in their remaining 100 variables, with 24,027 ALUM in the

database, 49.43% of the entire population in the database donating $80,862,060.67. Also, with the ALUM information in the database having the most complete amount of data in its associated cells, a more accurate score system can be determined because there is a larger quantity of variables that can be used in calculating metrics and calculating scores for individual variables.

3.2 Removing Outliers in Alumni Database

The ALUM database contained 24,027 individuals, with 14,364 (59.78%) donating a total of $80,862,060. The average donation per individual was $3,365, with a median of $45 and a standard deviation of $61,183.

Any individual who donated an exceptionally large amount was removed before testing the score systems because they would have an overwhelming impact metrics. Any set of score factors that happened to capture some of the largest donations would be rated highly by the metrics. Individuals with total donations more than 3 standard deviations above the mean were removed for testing. With this choice, any individual who donated more than $186,916.52 was removed. This group included 62 individuals (0.26% of the database) with total donations $44,899,094; over 45% of the total donations were made by these 62 individuals. For these 62 individuals, the average donation was $724,178 and a standard deviation of $957,396. The remaining 23,965 individuals had an average donation of $1,500 with standard deviation of $8,496. Any score system which captured the outliers would have appeared to be a good score system. Working on the data set with outliers removed gives a better picture of the accuracy of the score system on the general population of alumni.

3.3 Alumni-only Database Summary

The final database used for this project consisted of alumni only with the 62 largest donors removed. For the remainder of this report, "database" will refer to these 23,965 individuals. While donations made per person are recorded in the database by each individual year, for ease of analysis the donation made by each person is actually all their previous donations summed together. Every donation made was increased with inflation; therefore every total amount given by each individual does take into consideration the time value of money.

The final trimmed and organized database contains 46.31% of the original data, with a total donations of $35,962,966 which is approximately 36.18% of original donations in the database. Of the 23,965 individuals, 59.68% (14,302) donated and had an average donation of $1,500, and a median donation of $45. The next few sections explain the types of variables used in this database, as well as individual information about each variable.

3.3.1 Variable Types

Danny Jin identified 24 variables as important factors related to donation activity. Twenty two of these factors were used in the score system (Jin 2006). Of these 22 variables, 18 of them were analyzed based on whether the information on that variable was blank or non-blank. Blank or non-blank refers to whether or not the individual listed in the database either left a certain variable blank, or if there was a data input error on the side of the alumni office, or possibly the individual chose not to fill out the specific variable for other reasons. These 17 blank or non-blank variables can be seen in Appendix A, *Donation Behavior of Blank and Non-Blank Variables*, and will be further

explained in the following section. This blank and non-blank classification was determined based on the information, or lack of information, available about the specific variable, or the variable in question may have previously been established as a "yes or blank" variable in the database.

RES_ZIP was analyzed differently than the other 21 variables because it used zip codes to determine regions in Massachusetts and other locations outside of Massachusetts. Of the 23,965 individuals in the ALUM database, 9,294 (38.78%) are listed as residents of Massachusetts. A listing of all zip codes in Massachusetts, with their coinciding county, and all zip codes outside of Massachusetts were used. Zip code organization for Massachusetts residents only was done to divide the 9,294 individuals in Massachusetts into their appropriate county.

The remaining 5 variables are classified as "multiple category" variables are investigated individually in the next section. They are referred to as multiple category variables because they have multiple significant answers and further conclusions could be drawn from the variables multiple categories. The difference between multiple category variables and blank and non-blank variables is that while blank and non-blank variables assign a score to a variable simply based on whether any information was provided, multiple category variables actually assign scores to the specific information that was provided by that variable.

Appendix A, *Donation Behavior of Blank and Non-Blank Variables*, provides the 18 blank and non-blank variables, and below is an example of some conclusions that can be drawn from the blank and non-blank variable FRAT. It can be seen below in Table 4 that 9,416 individuals said that they were in a fraternity. This means that 39.29% of the

records said they were in a fraternity or sorority, while the remaining 60.71% left this variable blank or chose to not list themselves as being involved in a fraternity or sorority. Then number of people listed as involved in a fraternity is reduced to the number of people who were listed in a fraternity and who donated, this number was found to be 6,959. This means that of the individuals who identified themselves as being in a fraternity, 73.91% donated. The total amount of donations made in this category was $26,566,269.58 and dividing this by the 9,416 individuals listed as being in a fraternity the average donation given by the FRAT variable is $2,821.40, with a median donation of $175.00.

Table 4: Donation Behavior by Fraternity Variable

	# People	**# People Donated**	**% People Donated**	**$ Donated**	**PV $ Donated**	***Difference***
FRAT	9,416	6,959	73.91%	$26,566,269.58	**$21,021,390.98**	*$8,299,469.99*
	Average Donation	**PV Average Donation**	***Difference***	**Median**	**PV Median**	***Difference***
	$2,821.40	**$2,232.52**	*$881.42*	$175.00	**$120.82**	*$77.31*

Donation behavior by state can be seen in Appendix B, *Donation Behavior of States*, and donation behavior for Massachusetts can be seen in Appendix C, D*onation Behavior of Massachusetts Region*. Boston, Eastern Massachusetts and Worcester have a percent donating of around 50%. While Western Massachusetts is slightly higher around 60%, the Cape has the highest percent donating with almost three quarters of residents making a donation. The Cape also has an average donation of almost twice the mean of the database, $1,500.65, while Western Massachusetts is still above average with $1,952 as the average donation, and the other three areas being well below average.

Boston's average is the lowest, it also has the smallest median donation. The Cape has the largest average donation $2,746.50 and a substantially larger median of

$185.00, while Western Massachusetts has an average donation $1,952.91 close to that of the Cape, but with a median of only $50.00. Eastern Massachusetts and Worcester's average donations vary by hundreds of dollars, while their medians only differ by $5.00.

Donation behavior of the regions in Massachusetts allows for additional opportunities to assign scores based on resident location. While an individual may not necessarily be penalized for having such a low percent of donating, such as Boston with only 45.36% and an average donation of only $411.46, it would be important to give an individual who lives on the Cape a better score relative to someone in Boston because Cape residents not only donate almost twice as much as someone in Boston, but they also have an average donation over 5 times more than someone in Boston, and a much larger median as well.

Marital Status

While the majority of the 102 variables were analyzed on a blank/non-blank basis, other variables had multiple options for answers. And example of this can be seen in the MARRIED variable. The answers provided for this variable were married, single, separated, other/partner, divorced, widowed and blank. The two largest groups of people were married and single, while **Widowed** did not contain a large amount of people, it has a huge average donation. If the variable was left **Blank** then it was investigated and the remaining three choices (separated, other/partner, and divorced) were combined into the category **Other**. Table 5 provides an analysis of the Married data similar to how the blank/non-blank data was evaluated. It can be seen that only 136 (0.57%) of the database was left blank in this category, while almost half (53.62%) of the alumni identified themselves as married, and another 42.80% identified themselves as being single. The

category **Other** only had 534 (2.23%) records, but their average donation was much more than any of the other categories, with almost 80% of people in this category donating. It appears that although over 2 million dollars was donated by **Single** people, it has a very small, only 37.00%, percent of people that donate. **Married** appears to not only have a high average donation, but also over three-quarters of **Married** people donate. This type of table can help to conclude that someone listed as **Married** should get a higher score than someone listed as **Single** and that it may also be important to contact individuals in the other categories as well, because even though there are only 534 of them, they're making over 1 million dollars worth of donations and have the second highest percentage of donating.

Table 5: Donation Behavior by Marital Status Variable

	Widowed	Married	Single	Other	Blank
# OF	187	12,850	10,258	534	136
# DONATED	175	9,880	3,795	422	30
% DONATED	93.58%	76.89%	37.00%	79.03%	22.06%
$ DONATED	$2,776,395.61	$29,318,675.87	$2,323,820.12	$1,333,717.48	$210,357.02
AVG. DONATION	$14,847.04	$2,281.61	$226.54	$2,497.60	$1,546.74
MEDIAN DONATION	$3,035.00	$225.00	$0.00	$282.50	$0.00

Gender

Gender was similar to the Married variable in that there were very few blanks; in fact barely 0.01% of alumni left this section blank. M stands for Male, F stands for Female, and N stands for Not Applicable. While males comprised almost all the database (85.63%), the next largest group was females with at 14.34%, then N with 0.02%, and lastly blank with 0.01%. Table 6 provides an analysis of the Gender data similar to how the blank/non-blank data was evaluated. Similar conclusions can be drawn with the

Gender variable as with the Married variable. For example, you can see that regardless of gender, it appears that people tend to donate about 50% of the time, with males donating a little more near 60%. This can be interesting for a number of reasons, first off, this may mean that a score determined by gender may not be necessary, or if one is used, the score between different genders should differ minimally.

Table 6: Donation Behavior by Gender Variable

	M	F	N	Blank
# OF	20,523	3,436	4	2
# DONATED	12,453	1,846	2	1
% DONATED	60.68%	53.73%	50.00%	50.00%
$ DONATED	$34,286,833.81	$1,508,566.27	$20,241.89	$147,324.13
AVG. DONATION	$1,670.65	$439.05	$5,060.47	$73,662.07
MEDIAN DONATION	$50.00	$20.00	$1,810.00	$73,662.07

Bachelor's Degree Major

BS_MAJOR was a variable that had 67 different majors listed, and because the majority of these choices contained a very small number of people in them, the top two majors; Mechanical Engineering (ME) and Electrical Engineering (EE) were analyzed individually and not included in the **Non-Blank** section of Table 7. Only 0.58% of the BS_MAJOR category was left blank, while the two biggest groups, ME and EE, comprise 46.20% of the overall population with Mechanical Engineers at 24.99% and Electrical Engineers at 21.21%. Table 7 provides an analysis of the Bachelor Degree major data similar to how the blank/non-blank data was evaluated. It can be seen that average donation made by each group is relatively equal, and it appears that while overall anyone who listed a major tended to donate over 50% of the time, mechanical engineers and electrical engineers tend to donate closer to 70%. Once again, this is important to

keep in mind when it comes to assigning scores because while it is important to give someone a higher score if the have a major listed rather than if they left is blank, there doesn't need to be as big a score difference assigned to an individual who identifies themselves as a ME or EE major.

Table 7: Donation Behavior by Bachelor's Degree Variable

	Non-Blank (Excludes ME and EE Majors)	Blank	ME	EE
# OF	12,754	138	5,989	5,084
# DONATED	7,160	6	3,781	3,355
% DONATED	56.14%	4.35%	63.13%	65.99%
$ DONATED	$15,505,562.20	$164,356.02	$10,955,177.78	$9,337,870.10
AVG. DONATION	$1,215.74	$1,190.99	$1,829.22	$1,836.72
MEDIAN DONATION	$25.00	$0.00	$74.00	$100.00

Grade Point Average (GPA)

Grade Point Average is different than the other variables because WPI does not calculate a GPA for students when they graduate. In the database there were 14,177 people who had a GPA listed, and there were 9,788 people who had a 0 listed instead of a GPA. **1.0-2.9** means the individual had a GPA ranging from 1.0-2.9 and similarly for 3.0-4.0. Of the people who either choose to calculate their GPA or had the Alumni office calculate their GPA, 46.50% of them had a GPA between a 3.0 and a 4.0, and 37.23% had a GPA between a 1.0 and 2.9. Table 8 provides an analysis of the Grade Point Average data similar to how the blank/non-blank data was evaluated. However in the **0** column and the **Non-Zero** column, it can be seen that people who do not list a GPA tend to donate over 80% of the time, while people who listed a GPA tended to donate only about 40% of the time. This variable has to be handled with a little caution because although in Table 8 it is being evaluated according to numerical GPA WPI doesn't

actually provide a GPA for students. Therefore, conclusions related to the real GPA of an individual are difficult to be drawn, and this should be kept in mind if this variable is used to assign a score that it appears that someone who does not list their GPA should get a slightly higher score than someone who does calculate it or have it calculated for them.

Table 8: Donation Behavior of GPA Variable

	1.0-2.9	3.0-4.0	0	Non- Zero
# OF	4,383	9,794	9,788	14,177
# DONATED	1,632	4,554	8,116	6,186
% DONATED	37.23%	46.50%	82.92%	43.63%
$ DONATED	$510,518.51	$2,330,033.45	$33,122,414.14	$2,840,551.96
AVG. DONATION	$116.48	$237.90	$3,383.98	$200.36
MEDIAN DONATION	$0.00	$0.00	$425.00	$0.00

Preference Class

In the PREF_CLAS column no individual in the database left this column blank or listed something other than a 4-digit year, therefore everyone listed the year that they wish to associate their graduation with. The highest two decades that alumni identified were 1980-1989 with 22.95% and 1990-1999 with 24.70%, and the two lowest decades that alumni identified were anyone before 1930 with 0.14% and 1930-1939 with 0.61%. It can be seen below that each decade's preference class is larger than the previous one and this was expected seeing as WPI's class size has increased over the years, therefore providing more graduates and in turn more selections for preference class. Excluding the **Before 1930** column it can be seen that older preference classes donated around 95%, where in recent years it can be seen that the percentage has decreased drastically. Therefore, it can be concluded that someone whose preference class is more recent should get a much lower score than an individual who graduated a long time ago. Table

9 provides an analysis of the Preference Class data broken down by decade similar to how the blank/non-blank data was evaluated.

Table 9: Donation Behavior of Graduate Classes by Decade

	Before 1930	1930-1939	1940-1949	1950-1959	1960-1969
# OF	33	146	686	1,266	2,370
# DONATED	15	140	654	1,184	1,996
% DONATED	45.45%	95.89%	95.34%	93.52%	84.22%
$ DONATED	$11,925.00	$2,026,011.97	$6,344,607.37	$9,704,131.26	$8,475,135.38
AVG. DONATION	$361.36	$13,876.79	$9,248.70	$7,665.19	$3,576.01
MEDIAN DONATION	$0.00	$3,055.50	$1,690.00	$1,354.50	$500.00

	1970-1979	1980-1989	1990-1999	2000 and Later	Blank/Other
# OF	3,732	5,500	5,919	4,313	0
# DONATED	2,954	3,974	2,440	945	0
% DONATED	79.15%	72.25%	41.22%	21.91%	0.00%
$ DONATED	$5,307,478.10	$3,188,005.12	$837,063.92	$68,607.98	$0.00
AVG. DONATION	$1,422.15	$579.64	$141.42	$15.91	$0.00
MEDIAN DONATION	$295.00	$120.00	$0.00	$0.00	$0.00

3.4 Analysis of Donations per Fiscal Year 1983-2007

The donation columns in the database were evaluated based on individual fiscal year. The average donation and median donation by fiscal year are summarized in Appendix D, *Average Donation and Median Donation by Fiscal Year 1983 to 2007 with Time Value of Money Calculated*. The percent of people who donated per fiscal year are summarized in Appendix E, and the total donations and largest donations made per fiscal year are summarized in Appendix F, *Analysis of Donation Behavior by Fiscal Year 1983 to 2007 with Time Value of Money Calculated*. The 1983 donation records are the cumulative amount given up through 1983 so this number should not be compared to

later years. The largest donation year was 1999 with $2,622,535.69 in total donations. While fiscal year 1998 had the largest single donation in the entire database, $156,000.00, its present value of $192,942.33 is actually less than the present value of 1989's largest single donation of $119,898.68 (present value of $126,130.95). Fiscal year 1983 had the largest number of donors, 7,209, and the next largest number of donors occurred in 1992 with 5,200 individuals. At the bottom of Appendix F, *Analysis of Donation Behavior by Fiscal Year 1983 to 2007 with Time Value of Money Calculated*, the donations made to WPI up to date total $35,962,966.10, which when increased with inflation actually is $28,437,136.78.

While time-value of money was not taken into consideration when determining scores for individual variables, once the spreadsheet was completed inflation was calculated for all donations based off of the historical consumer price index (CPI) data (InflationData.com 2003-2007). Every record in the database has a column for each fiscal year of donations and is either left blank if no donation was made or contains the amount donated. Appendix F contains donation behavior by Fiscal Year 1983 to 2006 and the present value of the total donations made for each fiscal year.

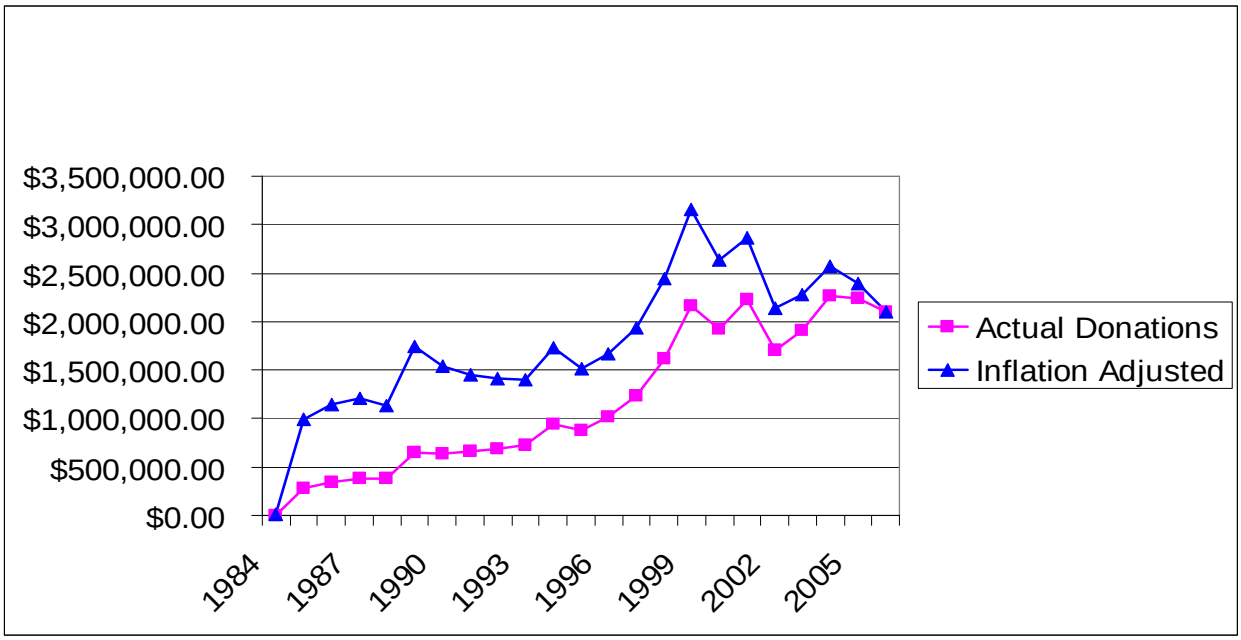


Figure 1: Actual and Inflation Adjusted Donations

3.4.1 Associations between Variables

There are important associations between many individual variables and donation behavior demonstrated in Jin's work (Jin 2006). Some of these positive and negative associations can be misleading when developing a score model. For example, Jin showed that both values, male and female, for the Gender variable had negative association with donating.

Table 10 shows the donation statistics according to both Graduation Year and Gender. Each cell contains the percent donating as well as the total donations and average donation for the group. The first female graduated from WPI in 1972, so no donations appear in the first cell in the "Female" row. Almost 90% of WPI alumni who graduated before 1972 have made a donation. The percent remains high for the 1972—1988 group, with a slightly higher donation rate for Women than Men in this group. The most recent graduates are the least likely to donate, but once again, Woman do donate at a slightly higher rate than Men. It is important to notice that if graduation year is

ignored, Men appear to be better donors than Woman. This difference is explained by the fact that WPI did graduate a Woman until 1972.

Table 10: Donation Behavior by Gender and Graduation Year

	Before 1972	1972 to 1988	1989 to Present	Totals
Men	87.51% donating $27,564,440.45 $5,380.53	74.62% donating $6,014,467.51 $859.09	32.69% donating $707,925.85 $84.29	60.68% donating $34,286,833.81 $1,670.65
Women	**0 students**	79.79% donating $1,185,808.27 $1,109.27	41.95% donating $322,758.00 $136.36	53.73% donating $1,508,566.27 $439.05
Totals	87.51% donating $27,564,440.45 $5,380.53	75.30% donating $7,200,275.78 $892.23	34.73% donating $1,030,683.85 $95.74	59.68% donating $35,795,400.08 $1,494.03

Table 11 shows the distribution of donation statistics according to both Marital Status and Graduation Year. As in Table 10, each cell contains the percent donation, followed by the total donations and the average donation for individuals in the specified category. In both age groups, Married alumni donated more and at a higher rate than Single alumni. This pattern remains when the Graduation Year groups are combined, indication that "Married" should have a positive score factor in the final model.

Table 11: Donation Behavior by Marital Status and Graduation Year

	Before 1972	1972 to present	Totals
Married	90.86% donating $22,932,568.26 $5,619.35	70.38% donating $6,386,107.61 $728.26	76.89% donating $29,318,675.87 $2,281.61
Single	65.45% donating $947,188.70 $1,441.69	35.05% donating $1,376,631.42 $143.38	37.00% donating $2,323,820.12 $226.54
Totals	87.34% donating $23,879,756.96 $5,038.99	51.92% donating $7,762,739.03 $422.58	59.18% donating $31,642,495.99 $1,369.33

3.5 Conclusion

Eighteen of the 22 variables in the database were analyzed based on whether the information on the individual was provided or left blank. While the investigation of these variables was helpful in becoming familiar with the categories of the database, it also helped explain how to assign appropriate scores to different variables. Another five variables, classified as multiple category variables were marital status, gender, bachelor's degree major, student GPA, and preference class. These multiple significant answers are important to the scoring system because different scores may need to be assigned within an individual variable, for example gender may want to assign different scores for men and women, so these variables can actually receive two scores; one for simply having the variable list, and a second score for having one specific value for the variable.

Resident zip code was used to determine donation behavior by geographical location, first by states and then also by regions in Massachusetts. Lastly the donations per fiscal year were organized in Appendix D, Appendix E, and Appendix F, with increase in the value of money due to inflation also calculated. While the alumni database included 101 variables associated with each alumnus, through data cleaning and organization this intimidating amount of data can be analyzed in different categories. This section provided an outline of each which will later be used in creating a scoring system, as well as explains how each type of variable can be examined and a score system can be determined from the available information.

4. Development of the Spreadsheet

The final spreadsheet uses macros to run all of the calculations to assign each individual a score that will in return assist the alumni office in identifying likely donors.

4.1 Factors Used for Real Alumni

The 22 factors that were used in the spreadsheet were based off of the top 24 factors that were found in the project done by Yi "Danny" Jin. All statistically significant factors were used with the exception of BS_MAJOR and WORK_ZIP. Leaving the following 22 variables seen in Table 12:

Table 12: Variables used for Spreadsheet

1. MARRIED	9. PREF_CLAS	17. STU_CLUB_SPORT
2. NON_WPI_DEG	10. HAD_SCHOLARSHIP	18. STU_PROF_SOC
3. FRAT	11. CLS_AGENT	19. STU_MUSIC
4. VARSITY_SPRTS	12. REUNION	20. STU_SCH_INVOLVE
5. RES_ZIP	13.ALUM_VOLUNTEER	21. STU_HONOR_SOC
6. GENDER	14. STU_PROJECT_CTR	22. GRAD_DISTINCTION
7. WPI_SPS	15. STU_GPA	
8. NUM_OF-CHILD	16. STU_INTL_CLUB	

(Jin 2006)

4.2 Donor Score System Spreadsheet

The Donor Score System Spreadsheet has 12 different tabs, each with different functions that either helps calculate an individual's score or help in deciding if the scoring system models the donation behavior of the alumni.

4.2.1 Data Tab

The “Data” tab is where all the data is input in the exact order specified: PERSON_NUM, MARRIED, NON_WPI_DEG, FRAT, VARSITY_SPRTS, RES_ZIP, GENDER, WPI_SPS, NUM_OF_CHILD, PREF_CLAS, HAD_SCHOLARSHIP, CLS_AGENT, REUNION, ALUM_VOLUNTEER, STU_PROJECT_CTR, STU_GPA, STU_INTL_CLUB, STU_CLUB_SPORT, STU_PROF_SOC, STU_MUSIC, STU_SCH_INVOLVE, STU_HONOR_SOC, GRAD_DISTINCTION and TOTAL_DONATION.

This tab is the main tab of the spreadsheet, where the user pastes the data, runs the macro by clicking on the “Calculate Score!” button, and then clears all the data by clicking on the “Clear Data” button.

4.2.2 Top Scorers Tab

This tab lists all individuals with their final resulting score. When the macro is completed, the screen will switch to this tab, for the user. The “Top Scorers” Tab has the individuals ID number with their calculated score sorted in descending order. The individuals with a higher a score are more likely to donate. For example a person with a score of 20 is the most likely to donate money and a person with a score of 1 is probably least likely to donate. The accuracy of the scoring method is analyzed through the metrics defined in Section *** below.

4.2.3 Data 2 Tab

The “Data 2” tab is used for calculation purposes only. This is used only to simplify the given data, so that it can be used for calculations later. It changes the entry

for each of the following 16 blank/nonblank factors to a “Y” if they have a value entered in this field. These factors are: NON_WPI_DEG, FRAT, VARSITY_SPRTS, WPI_SPS, NUM_OF_CHILD, HAD_SCHOLARSHIP, CLS_AGENT, REUNION, ALUM_VOLUNTEER, STU_PROJECT_CTR, STU_INTL_CLUB, STU_CLUB_SPORT, STU_PROF_SOC, STU_MUSIC, STU_SCH_INVOLVE, and STU_HONOR_SOC. This is done by using the following Microsoft Visual Basic (VBA) formula:

=IF (Data!RC="""", """", ""Y"")

This statement looks at the current cell and, if the value in the exact same cell in the Data tab is blank, leave the current cell blank, and if it is not blank to enter a “Y”.

STU_GPA is checked to see if the value entered is a valid GPA value between 0.0 and 4.0. If the GPA listed is between these values, it will be displayed, and if it lies outside of these values, it is left blank. This is done using the following VBA code:

=IF (Data!RC>4, """", IF(Data!RC>0, Data!RC, """"))"

RES_ZIP undergoes the greatest change in this tab of the spreadsheet. Each zip code in the data is given in a text format in excel. Some are 9 digit and some are 5 digit. In order to be able to run calculations with them later, each zip code is truncated to only 5 digits and then converted to a number. This is done with the following code:

= IF(ISERROR(VALUE(LEFT(Data!RC, 5))), """", VALUE (LEFT (Data!RC, 5)))

PERSON_NUM, MARRIED, GENDER, PREF_CLAS, GRAD_DISTINCTION and TOTAL_DONATION are all directly copied from “Data” to “Data 2” with no changes.

4.2.4 Zipcodes Tab

The “ZipCodes” tab is also used for calculation purposes only. It has a list of all possible zip codes used in the U.S. with their corresponding states. Each of these ranges are then assigned a score factor that is used in the score calculation. Alumni that live in a region with low likelihood of donating receive a lower score factor than an alumni that lives in a high donating state.

4.2.5 Score Factors Tab

This tab is used to calculate each individual’s score. The individual scorings that are associated with each piece of information about each person are in this tab. Each factor is listed with the corresponding score for each possibility of the score. The 18 blank/non-blank factors each have a score factor for both if the factor is blank or if the factor is non-blank.

As mentioned above the RES_ZIP score factors are pulled from the “ZipCodes” tab, if the zip code is non-blank. The score factor for a blank zip code is listed in the “Score Factors” tab. The MARRIED factor is listed as “Married”, “Single” and “Other”. This means that “Divorced”, “Widowed”, “Other/Partner”, “Separated”, “NA”, and blanks were all grouped together in the “Other” factor in the spreadsheet. Gender also had three different categories: “M”, “F”, “Other”, saying that “N” and the blanks were group together into one category, when it came to assigning score factors.

For the PREF_CLAS factor, quartiles were calculated for the data. All the preferred graduation classes were looked at and the 25^{th} and 75^{th} percentile were calculated. The 25^{th} percentile is 1974, and the 75^{th} percentile is 1997. Once these quartiles were calculated, the score factors are assigned to:

- Anyone that has PREF_CLAS in the first quartile (before 1974);
- Anyone that has PREF_CLAS in the fourth quartile (after 1997);
- Anyone that has PREF_CLAS between the 25th and 75th percentile (between 1974 and 1997).

An advanced user of the spreadsheet could change the score factors as they wish, to make a better model or to single out specific variables. Also on this tab is the "Corrective Score Factor", used to map all scores to the interval [1, 20]. All these score factors are used on the next tab to calculate each individual's score.

4.2.6 Scoring Tab

The "Scoring" tab is where all the calculations occur. For each individual, it uses the "Data 2" and "Score Factors" tabs to find the correct score factor for each variable.

For the blank/non-blank factors this is done using the following formula

= IF('Data 2'!RC=""Y"" ,'Score Factors'!R9C2, 'Score Factors'!R9C3)

where "Score Factors!R9C2" refers to the score factor when the value is true, and "Score Factors!R9C3" is the score factor when the value is false. For the RES_ZIP, a VLookup was done with the following formula:

= IF('Data 2'!RC="""", 'Score Factors'!R23C3, VLOOKUP('Data 2'!RC, ZipCodes!R2C1:R131C4, 4))

In this case it is checking if the RES_ZIP is blank (which is any individual with a blank or invalid zip code), and if it is giving them the score factor for a blank zip code. If it is not blank it is looking for the zip code in the "ZipCodes" tab, and then giving it the score factor assigned to that zip code.

Factors such as MARRIED, GENDER, and PREF_CLAS were analyzed with Nested IF statements such as the following formula for GENDER:

= IF('Data 2'!RC=""F"", 'Score Factors'!R28C3, IF('Data 2'!RC=""M"",

'Score Factors'!R28C2,'Score Factors'!R28C4))

If GENDER is equal to "F", then the score factor for Female is assigned and if it is not "F", the it checks to see if the value is "M", and then giving it the score factor for Male. If the field is not equal to either, then it gives the field the score factor for "Other".

Each individual is assigned 22 different score factors based on their information, and these are summed to obtain their Total Score. The Total Score is then adjusted so that all scores are in the range from 1 to 20. This is called the "ADJ Score", which is used for all the graphs and is shown on the "Top Scorers" tab. The equation for the "ADJ Score" is shown below:

$$ADJScore = \left(TotalScore * \frac{20}{TotalScoreMax - TotalScoreMin} \right) - \left(\frac{20 * TotalScoreMin}{TotalScoreMax - TotalScoreMax} \right) - \frac{1}{2}$$

Here is an example to see how this calculation is done. In the Donor Score System, the Minimum Total Score is 2.75, and the Maximum Total Score is 23. For a person that for example has a Total Score of 13.25, the following calculation would be done:

(13.25) * (20 / (23 - 2.75)) = (13.25) * (20 / 20.25) = 13.25 * .98765 = 13.08642

(20 * 2.75) / (23 – 2.75) -.5 = (55 / 20.25) - .5 = 2.71605 - .5 = 2.21605

13.08642 – 2.21605 = 10.87037 → Rounded to nearest Number = **11**

This person's Adjusted Score would be 11 on a 1 to 20 scale.

4.2.7 Scoring 2 Tab

The “Scoring 2” tab summarizes the basic statistics on the calculated scores. It has the number of people with each score, the total donated per score, the average donation per score and the percentage of people donating per score. The graphs are drawn using these statistics. This tab also shows the minimum and maximum Total and Adjusted scores. An advanced user could change the maximum ADJ Score if wanted from 20 to any other number, in this tab as well in the field next to “Wanted MAX”.

4.2.8 Metrics Tab

The “Metrics” tab is used to measure the performances of the current scoring system. There are five different calculations done,

- R^2 value on the average donation,
- the sum of slopes of average donation,
- the sum of slopes of percentage of people donating,
- the comparison of the averages for the average donation,
- and the comparison of averages for the percentage of people donating.

The details to the calculations for these metrics are found in *Section 2.2.2. Metrics* Each metric has a numeric value, and the sum of all these values should be maximized. The Maximum value of metrics 1, 3, and 5 is one, and therefore the sum of these three metrics should be as close to 3 as possible.

4.2.9 Graphs

The last 4 tabs are graphs to see how well the scores correlate with the donation behavior. Total People, Total Donation, Average Donation and Percentage of People

Donating for each Score “bucket” are displayed on each of these to see the relationship between the data and the calculated scores. Results of these graphs can be found in *Section 5.3 Graph Results*.

4.3 Summary

Above, each of the different tabs is explained with their individual functions. This spreadsheet calculates the scores for any alumni database that the user may enter, and has the additional tools to be able to analyze the results. With the *Donor Score System*, the Alumni Office of the Worcester Polytechnic Institute will be able to look at the ID numbers of the top scoring people and will be able to target its fundraising activities.

5. Results

In this chapter the main findings from running *Donor Score System* are described. This includes the final score system with the individual score factors that were used, the final results of the metrics and graphs, and the overall findings of the score system.

5.1 Determining the Score Factors

To determine the score factors for each variable, the statistics described Chapter 3 were analyzed to identify variables which are correlated with donation behavior. All of the statistics were compared to the donation statistics for the full population, where the overall average donation among ALUM is $1,500.65, and the percentage of ALUMs donating is 59.68%. Table 13 contains an example of this analysis for the variable FRAT (i.e. those who reported participation in Greek life and those who did not):

Table 13: Statistics for FRAT variable

	# People	# People Donated	% People Donated	$ Donated	Average Donation	Median
Non Blank	9,416	6,959	73.91%	$26,566,269.58	$2,821.40	$175.00
	# People	**# People Donated**	**% People Donated**	**$ Donated**	**Average Donation**	**Median**
Blank	14,549	7,343	*50.47%*	$9,396,696.52	$645.87	*$3.00*

Those with a non-blank field for FRAT had above average frequency and donation size. Both the percentage of individuals donating and the average donation are significantly higher than those that have the FRAT field left blank. In fact, having FRAT blank affects the donation behavior negatively, as the percentage of people donating and the average donation are lower than the overall donation behavior. In this case the score factor for non-blank for FRAT is set to 0.75 and the score factor for a blank FRAT is -0.5.

The same analysis was completed for all blank/non-blank variables. The results are summarized below in Table 14.

Table 14: Score Factors used for Blank/Non-Blank Variables

Factor	Score factor for Non-Blank	Score Factor for Blank
NON_WPI_DEG	**0.75**	**0.0**
FRAT	**1.0**	**-0.5**
VARSITY_SPRTS	**0.75**	**0.0**
RES_ZIP	**See ZipCodes Tab**	**-0.75**
WPI_SPS	**1.0**	**0.0**
NUM_OF_CHILD	**0.5**	**0.0**
HAD_SCHOLARSHIP	**0.25**	**0.5**
CLS_AGENT	**1.0**	**0.0**
REUNION	**1.0**	**-1.0**
ALUM_VOLUNTEER	**1.0**	**-1.0**
STU_PROJECT_CTR	**-0.25**	**0.5**
STU_GPA	**-0.5**	**1.0**
STU_INTL_CLUB	**-0.5**	**0.0**
STU_CLUB_SPORT	**0.25**	**0.0**
STU_PROF_SOC	**0.75**	**0.0**
STU_MUSIC	**0.75**	**0.0**
STU_SCH_INVOLVE	**1.0**	**-0.25**
STU_HONOR_SOC	**-0.5**	**0.25**
GRAD_DISTINCTION	**-0.25**	**0.5**

The score factors for the individual zip codes used for the RES_ZIP factor can be found in Appendix G. where score factors are assigned by state. Table 15 reports the score factors for the MARRIED, GENDER and PREF_CLAS variables.

Table 15: Score Factors used for MARRIED, GENDER and PREF_CLAS

FACTOR			
MARRIED	Score Factor for "Married"	Score Factor for "Single"	Score Factor for Other/Blank
	0.5	**-1.0**	**1.0**
GENDER	Score Factor for "M"	Score Factor for "F"	Score Factor for Other/Blank
	-0.25	**-0.5**	**1.0**
PREF_CLAS	Score Factor for before 25th percentile (before 1974)	Score Factor for after 75th Percentile (after 1997)	Score Factor for 25th to 75th percentile (1974-1997)
	1.0	**-1.0**	**-0.5**

5.2 Metric System

The spreadsheet uses five different metrics as a way of measuring how well the scoring model fits the actual donation behavior. The five metrics are the following:

1. R^2 value on the average donation,
2. the sum of slopes of average donation,
3. the comparison of the averages for the average donation,
4. the sum of slopes of percentage of people donating,
5. the comparison of averages for the percentage of people donating.

The definition for each of these metrics is in Chapter 2.2.2. Metrics. Each of these return a numerical value which can be summed as a comparison for different runs. The sum of all five should be maximized, but the more important value is the sum of metrics #1, #3 and #5. Since the maximum value of these three metrics is 1, the sum of these three should be as close to 3 as possible. If this value was equal to 3, then the R-squared value would be equal to one, meaning that all of the average donations were perfectly explained by the linear regression model (highly unlikely). If the comparison of the averages for both the average donation and percentage of people donating are both equal to one, then this means that their averages are always increasing, meaning the graphs are always increasing, which is the ideal shape of these graphs.

The metrics were used to analyze different combination of score factors. Several possible combinations of score factors were tested on random samples of 1000 individuals and the metric results were compared to see which selection of score factors provided the best fit for the database.

The results for the metrics in a *Donor Score System* run of the 23,965 Alumni are displayed below in Table 16.

Table 16: Metric Values for Final Score System

Metric #1 **R-Squared** **As close to 1 as possible** **R^2 = 0.6405**	**Metric #2** **Sum of Slopes of Average Donations** **Maximize Value** **Sum of Slopes / 1000 = 0.4515**
Metric #3 **Comparison of Averages** **As Close to 1 as Possible** **Score = 0.9626**	**Metric #4** **Sum of Slopes of % Donating** **Maximize Value** **Sum of Slopes *10 = 0.6420**
Metric #5 **Comparison of %Donating Averages** **As Close to 1 as Possible** **Score = 0.9922**	

Table 17 shows the sums of the Metrics used for analysis of the fit of the model.

Table 17: Metric Sums for Final Score System

Sum of All **Sum = 3.6888**	**Sum** As close to 3 (Metrics #1,#3,#5) **Sum = 2.5953**

As seen above the sum of metrics of #1, #3, and #5 is almost 2.6 out of 3, which is considered a very good fit, especially since the R-square value is rarely close to 1.

5.3 Graph Results

5.3.1 Number of Individuals in each Score Bucket

This graph shows the distribution of adjusted scores for the Alumni population. The largest buckets are scores 6, 7, and 8 with more than 2500 individuals in each bucket. There are very few individuals with scores 1 or 2 and very few individuals with scores 19 or 20. The largest group is score 7 but the distribution is slightly skewed to the right. ".

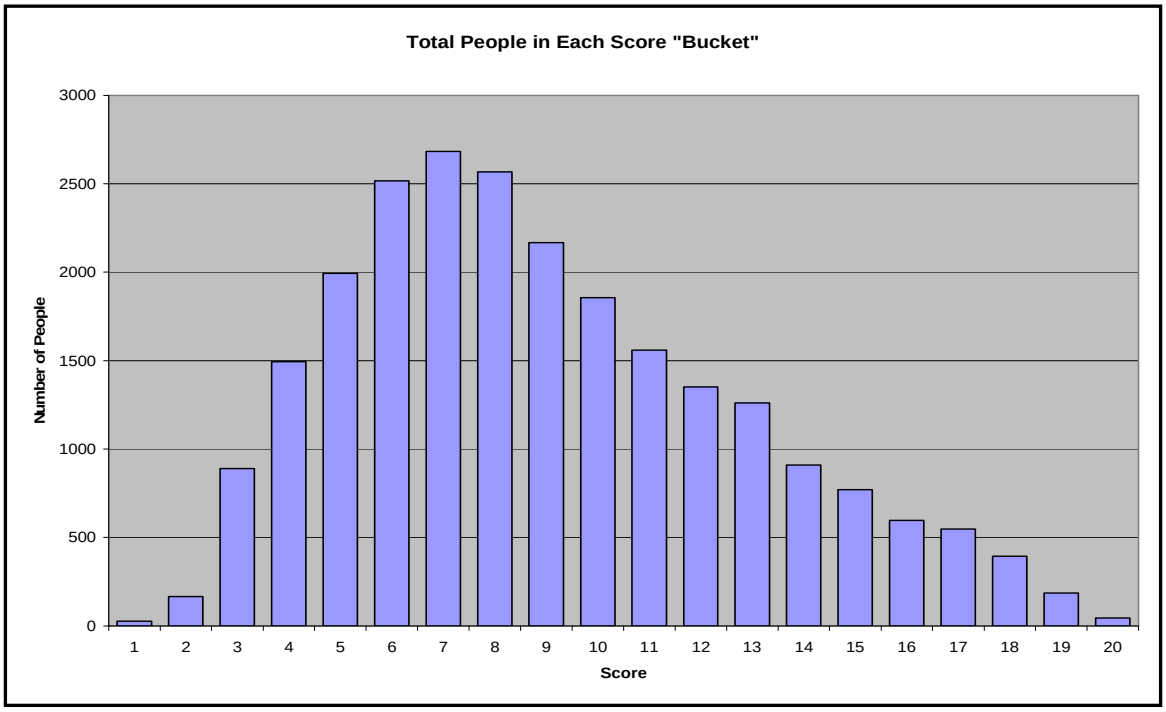


Figure 2: Total People for each Score

5.3.2 Total Donated

The next graph shows the total amount donated for each score bucket. In the graph below (Figure 3) there is a large drop in Total Donated for score buckets 19 and 20. This is explained by the small number of individuals in these buckets; there are only 185 individuals in with a score of 19 and 44 individuals with a score of 20.

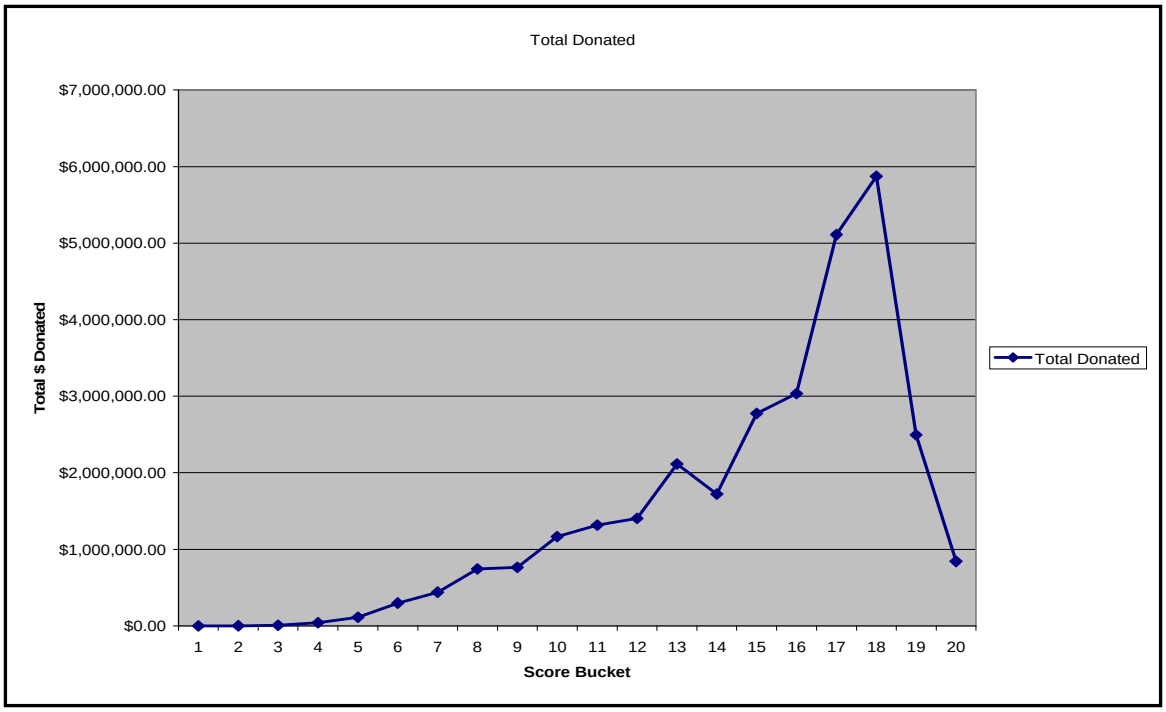


Figure 3: Total Donated for each Score

5.3.3 Average Donation

The "Average Donation" graph displays the average donation for each score bucket. The ideal shape of this graph would be an increasing curve, with a very low value for a score of 1, and a very high value for a score of 20. In Figure 4 the average donation for score bucket 1 is $8.19 and the average donation for score bucket 20 is $19,132.72. There is a small dip in moving from bucket 18 to 19.

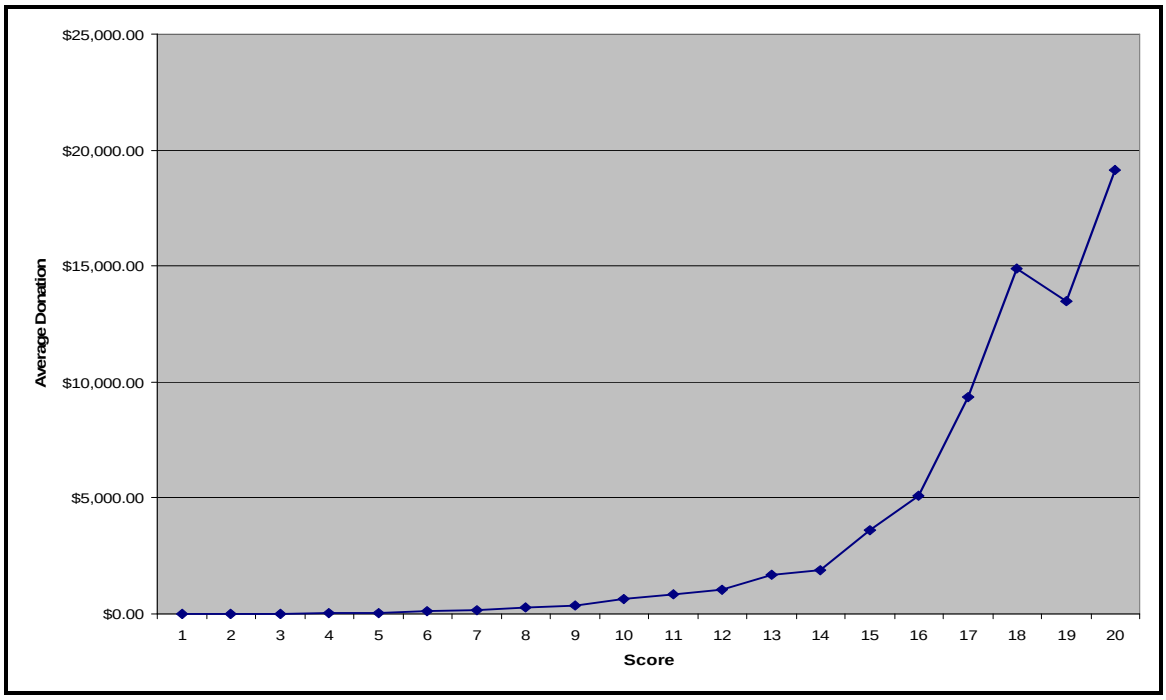


Figure 4: Average Donation for each Score

5.3.4 Percent Donating

This graph shows the percentage of people donating in each score bucket. The red dashed line is the percentage of people donating for all alumni (about 60% donate). Again, with this graph the ideal shape would be an always increasing, with a low percentage for a score of 1, and a high percentage of people donating for the score of 20. The graph below (Figure 5) has an almost perfect shape, as it is increasing throughout the entire graph with the exception at score bucket 19. The percentage of people donating with a score of 1 is 19.23% and the percentage of people donating with a score of 20 is 100.00%. The curve crosses the population average between score bucket 8 and 9. Individuals with a score 9 or higher are more likely than average to donate.

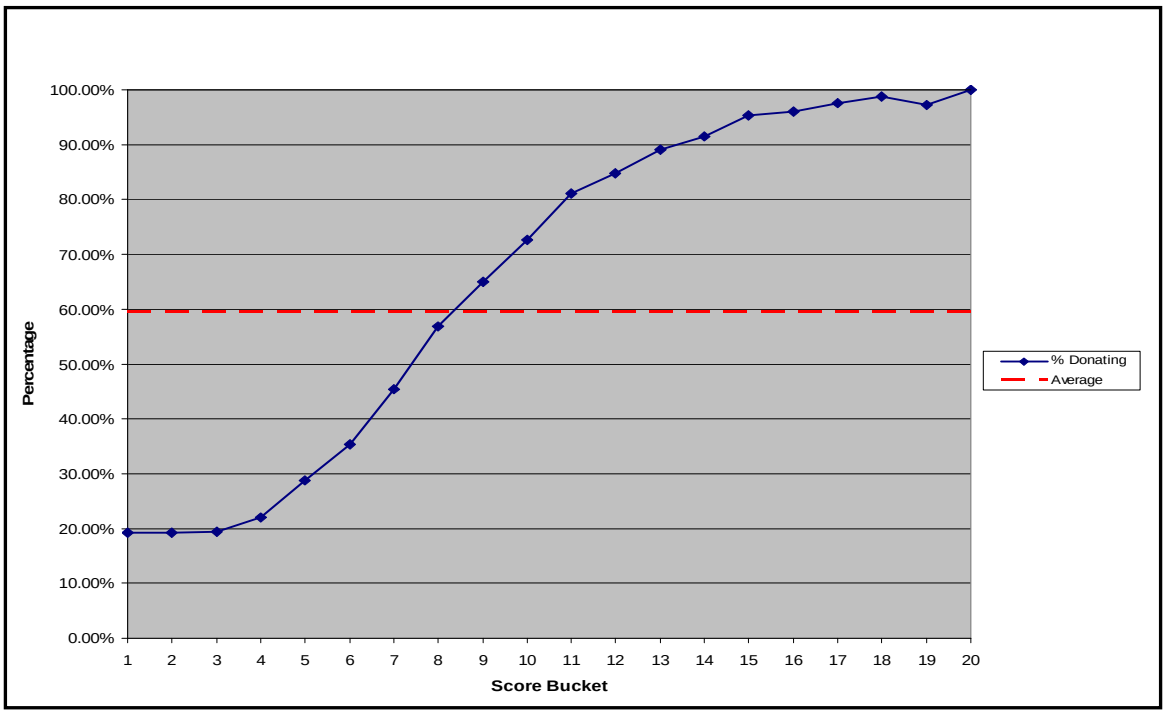


Figure 5: Percentage Donating for each Score

5.4 Summary

The final set of score factors reported above provided a good predictive model for donations in the (trimmed) alumni database. Five different metrics were used to develop a model that the metrics show that the score factors give good information about donation behavior.

6. Conclusions

The goal of this project was to develop a scoring system that would use the insight gained from Danny Jin's statistical analysis (Jin 2006). The score system is a tool that the Office of Development and Alumni Relations can use to organize and explore its donor database. The system does not "predict" donations but it does identify groups of donors that are more (or less) likely to make donations. It is designed to be as simple and flexible as possible so that the Office of Development and Alumni Relations can use the tool on current data and adapt it to future trends in donations.

The score system was tested on the 23,965 alumni in the donor database. Of the 102 variables for each individual, the score system uses 24 variables, assigning a score factor (in the range -1.0 to +1.0) to values for each variable. For 18 of these variables, all that was used was the fact that the value was not blank. For example, it did not matter which fraternity or sorority the individual reported, it only mattered that the individual had been involved in Greek life while at WPI. The score factors for each variable were determined by comparing the effect that each value of the variable had on donations with the donation statistics for the full population.

Three different metrics were developed to test the predictive ability of the score system. A good score system would give larger scores to groups with better donation behavior (frequency and amount). The final score system coded in *Donor Score System.xls* satisfies this criterion.

References:

Adriaans, Pieter; and Zantinge, Dolf (1996). Data Mining. Addison-Wesley Longman Limited, Harlow, England. Pg. 47.

Burstein, Leigh; Freeman, Howard; and Rossi, Peter (1985). Collecting Evaluation Data: Problems and Solutions. Sage Publications, California. (page 51).

Bryman, Alan; and Hardy, Melissa (2004). Handbook of Data Analysis. Sage Publications, California.

Calkins, Keith G. The How and Why of Statistical Sampling. Andrews University, 1998-2005.
<http://www.andrews.edu/~calkins/math/webtexts/stat02.htm> 11/10/2006

Crawford, Jogoda and Frank. Data Mining in a Scientific Environment.
<http://www.csu.edu.au/special/auugwww96/proceedings/crawford/crawford.html > 9/3/2006

Data Description, Inc (2006). Margolis Wylie Associates Mines a Wealth of Information for Higher Education. http://www.datadesk.com/company/profiles/wylie.shtml 11/30/2006

Data Mining Examples & Testimonials, Data Mining Software of data mining solutions.
<http://www.data-mining-software.com/data_mining_examples.htm> 9/3/2006

Davies, Phil. The Magenta Book: Guidance Notes for Policy Evaluation and Analysis. Government Chief Social Researcher's Office, 2004.
<http://www.policyhub.gov.uk/downloads/chapter5.pdf> 11/10/2006

Dictionary.com. "Coefficient of Correlation." American Heritage Dictionary of the English Language, Fourth Edition (2006). Houghton Mifflin Company.
<http://dictionary.reference.com/search?q=coefficient%20of%20correlation> 4/20/2007

Dmoz (2006). "Data Mining"
<http://dmoz.org/Computers/Software/Databases/Data_Mining/> 9/2/2006

Draper, N.R., and Smith, H (1981). Applied Regression Analysis, second edition. John Wiley & Sons, Canada.

Ebecken, Nelson (1998). Data Mining. WIT Press: Southhampton, UK.

Fayyad, Usama. Piatetsky-Shpiro, Gregory. Smyth, Padhraic. From Data Mining to Knowledge Discovery in Databases.
<http://www.aaai.org/AITopics/assets/PDF/AIMag17-03-2-article.pdf> 9/3/2006

Frank, Eibe; Witten, Ian (2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers, San Francisco, California (2000). (Page 3)

Friedman, Jerome; Hastie, Trevor; and Tibshirani, Robert (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer-Verlag New York.

InflationData.com (2003-2007). "Historical CPI-U data from 1913 to the present." Financial Trend Forecaster, Capital Professional Services. <http://inflationdata.com/inflation/Consumer_Price_Index/HistoricalCPI.aspx?rsCPI_currentPage=2> 3/1/2007

Jin, Yi. "Regression Analysis of University Giving Data." MS Project Report. Department of Mathematical Sciences. Worcester Polytechnic Institute, December 2006.

Kutner, Michael H., Neter, John., and Wasserman, William (1990). Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs, third edition. Richard D. Irwin.

Kutner, Michael H., Nachtsheim, Christopher J., Neter, John, and Li, William (2005). Applied Linear Statistical Models, fifth edition. McGraw-Hill.

Linux Information Project (2005). "Metric Definition." <http://www.bellevuelinux.org/metric.html> 3/23/2007

Microsoft Corporation (2006). "Data Mining" MSDN. <http://forums.microsoft.com/msdn/showforum.aspx?forumid=81&siteid=1> 9/3/2006

Pyle, Dorian (1999). Data Preparation for Data Mining. Morgan Kaufmann Publishers, Inc. San Francisco, California.

Saarenvirta, Gary (1998). Mining Customer Data. International Business Machines Corporation. <http://www.db2mag.com/db_area/archives/1998/q3/98fsaar.shtml> 9/3/2006

StatSoft, Inc (2006). "Data Mining Techniques" <http://www.statsoft.com/textbook/stdatmin.html> 9/2/2006

Swinscow, T.D (1997). Statistics at Square One. BMJ Publishing Group Ltd. 9th Edition. <http://www.bmj.com/statsbk/2.dtl> 2/17/2007

Thearling, Kurt (2006). "An Introduction to Data Mining" <http://www.thearling.com/text/dmwhite/dmwhite.htm> 9/3/2006

Thearling, Kurt (2006). "Data Mining"<www.thearling.com> 9/1/2006

United States Postal Service (USPS). "History of the U.S. Postal Service." <http://www.usps.com/history/his2_75.htm#ZIP> 1/16/2007

Wylie, Peter B (2005). A Simple Score. Council for Advancement and Support of Education. <http://www.case.org/files/Bookstore/PDF/Wylie%20White%20Papers/A_Simple_Score.pdf> 9/14/2006

Wylie, Peter B (2004). Data Mining for Fund Raisers: How to use simple statistics to find the gold in your donor database (even if you hate statistics). Council for Advancement and Support of Education: Washington, DC. P. 11-72.

Wylie, Peter B (2005). Deep Pockets: Where the Alumni Money Is. Council for Advancement and Support of Education. <http://www.case.org/files/Bookstore/PDF/Wylie%20White%20Papers/Deep_Pockets_Where_the_Alumni_Money_Is.pdf> 9/14/2006

Appendix A: Donation Behavior of Blank and Non-Blank Variables

	# People	# People Donated	% People Donated	$ Donated
NON_WPI_DEG	6,138	4,807	78.32%	$15,852,739.97
FRAT	9,416	6,959	73.91%	$26,566,269.58
VARSITY_SPRTS	6,239	4,099	65.70%	$15,291,113.29
WPI_SPS	1,512	1,062	70.24%	$3,114,091.76
NUM_OF_CHILD	8,789	7,448	84.74%	$29,145,006.18
HAD_SCHOLARSHIP	13,828	7,511	54.32%	$15,886,715.94
CLS_AGENT	501	499	99.60%	$10,789,259.14
REUNION	3,902	3,549	90.95%	$25,661,838.58
ALUM_VOLUNTEER	4,890	4,534	92.72%	$28,690,606.28
STU_PROJECT_CTR	2,195	1,068	48.66%	$795,295.20
STU_INTL_CLUB	603	174	28.86%	$107,232.99
STU_CLUB_SPORT	6,276	4,119	65.63%	$9,873,910.89
STU_PROF_SOC	7,335	5,577	76.03%	$18,933,647.33
STU_MUSIC	2,013	1,494	74.22%	$5,298,655.71
STU_SCH_INVOLVE	4,172	3,180	76.22%	$11,707,803.02
STU_HONOR_SOC	778	363	46.66%	$336,159.01
GRAD_DISTINCTION	7,203	3,727	51.74%	$3,215,916.59

	Average Donation	Median
NON_WPI_DEG	$2,582.72	$265.00
FRAT	$2,821.40	$175.00
VARSITY_SPRTS	$2,450.89	$100.00
WPI_SPS	$2,059.58	$110.00
NUM_OF_CHILD	$3,316.08	$405.00
HAD_SCHOLARSHIP	$1,148.88	$20.00
CLS_AGENT	$21,535.45	$6,580.00
REUNION	$6,576.59	$1,050.00
ALUM_VOLUNTEER	$5,867.20	$1,035.00
STU_PROJECT_CTR	$362.32	$0.00
STU_INTL_CLUB	$177.83	$0.00

STU_CLUB_SPORT	$1,573.28	$75.00
STU_PROF_SOC	$2,581.27	$225.00
STU_MUSIC	$2,632.22	$170.00
STU_SCH_INVOLVE	$2,806.28	$200.00
STU_HONOR_SOC	$432.08	$0.00
GRAD_DISTINCTION	$446.47	$10.00

Appendix B: Donation Behavior of Individual States

STATE	# People	$ Donated	# People Donated	% People Donated	Average Donation	Median
AL	43	$41,798.00	31	72.09%	$972.05	$135.00
AK	16	$925.00	8	50.00%	$57.81	$5.00
AZ	172	$240,820.00	123	71.51%	$1,400.12	$80.00
AR	14	$21,327.00	8	57.14%	$1,523.36	$25.00
CA	1,084	$1,971,037.40	684	63.10%	$1,818.30	$70.00
CO	260	$406,220.25	174	66.92%	$1,562.39	$100.00
CT	2,159	$3,258,105.15	1,396	64.66%	$1,509.08	$100.00
DE	63	$170,784.14	49	77.78%	$2,710.86	$587.50
DC	36	$10,978.00	18	50.00%	$304.94	$10.00
FL	707	$2,753,005.67	505	71.43%	$3,893.93	$200.00
GA	214	$308,205.92	144	67.29%	$1,440.21	$100.00
GU	1	$75.00	1	100.00%	$75.00	$75.00
HI	48	$13,790.00	25	52.08%	$287.29	$27.50
ID	24	$42,299.08	18	75.00%	$1,762.46	$172.50
IL	187	$599,250.93	134	71.66%	$3,204.55	$150.00
IN	92	$166,587.60	68	73.91%	$1,810.73	$127.00
IA	17	$15,402.00	12	70.59%	$906.00	$315.00
KS	22	$16,964.00	14	63.64%	$771.09	$100.00
KY	43	$73,815.95	33	76.74%	$1,716.65	$113.00
LA	43	$43,577.00	26	60.47%	$1,013.42	$25.00
ME	429	$687,491.14	240	55.94%	$1,602.54	$25.00
MD	429	$772,514.66	327	76.22%	$1,800.73	$200.00
MA	9,294	$10,843,577.81	5,248	56.47%	$1,166.73	$25.00
MI	199	$721,974.05	139	69.85%	$3,628.01	$175.00
MN	81	$127,338.00	48	59.26%	$1,572.07	$85.00
MS	14	$2,914.00	9	64.29%	$208.14	$72.50
MO	48	$75,796.37	36	75.00%	$1,579.09	$232.50
MT	10	$1,024.00	5	50.00%	$102.40	$5.00
NE	11	$13,077.58	8	72.73%	$1,188.87	$225.00

NV	40	$136,175.63	19	47.50%	$3,404.39	$0.00
NH	1,306	$2,028,512.48	757	57.96%	$1,553.23	$25.00
NJ	585	$1,208,724.74	431	73.68%	$2,066.20	$180.00
NM	51	$151,457.40	34	66.67%	$2,969.75	$190.00
NY	1,062	$1,613,834.55	705	66.38%	$1,519.62	$100.00
NC	300	$539,690.49	218	72.67%	$1,798.97	$110.00
ND	5	$300.00	3	60.00%	$60.00	$25.00
OH	262	$783,910.06	204	77.86%	$2,992.02	$232.50
OK	23	$32,933.00	19	82.61%	$1,431.87	$400.00
OR	103	$145,358.65	66	64.08%	$1,411.25	$50.00
PA	527	$1,037,633.07	372	70.59%	$1,968.94	$180.00
PR	33	$173,224.00	19	57.58%	$5,249.21	$35.00
RI	645	$467,406.82	353	54.73%	$724.66	$20.00
SC	135	$372,390.26	102	75.56%	$2,758.45	$225.00
SD	8	$23,453.71	5	62.50%	$2,931.71	$245.00
TN	55	$49,974.00	39	70.91%	$908.62	$125.00
TX	393	$761,140.72	258	65.65%	$1,936.74	$100.00
UT	28	$55,477.43	20	71.43%	$1,981.34	$45.00
VT	208	$488,741.84	138	66.35%	$2,349.72	$95.00
VI	2	$175.00	2	100.00%	$87.50	$87.50
VA	522	$895,343.22	335	64.18%	$1,715.22	$55.00
WA	213	$305,822.00	118	55.40%	$1,435.78	$25.00
WV	14	$54,731.41	10	71.43%	$3,909.39	$760.00
WI	63	$206,589.61	39	61.90%	$3,279.20	$31.00
WY	5	$692.00	4	80.00%	$138.40	$170.00
Blank/Other	1,617	$1,028,598.31	501	30.98%	$636.12	$0.00

Appendix C: Donation Behavior of Massachusetts Region

County/Region	Alumni	Donors	% Donors	Total Giving	Average Giving	Median Giving
NANTUCKET	5	3	60.00%	$7,196.00 ($5,221.51)	$1,439.20 ($1,044.25)	$300.00 ($236.68)
DUKES	12	6	50.00%	$77,883.12 ($66,331.62)	$6,490.26 ($5,527.64)	$1.00 ($0.91)
BARNSTABLE	233	171	73.39%	$601,545.34 ($470,955.75)	$2,581.74 ($2,021.27)	$200.00 ($126.82)
CAPE	**250**	**180**	**72.00%**	**$686,624.46**	**$2,746.50**	**$185.00**
SUFFOLK	291	132	45.36%	$119,734.00	$411.46	$0.00
BOSTON	**291**	**132**	**45.36%**	**$119,734.00**	**$411.46**	**$0.00**
ESSEX	678	408	60.18%	$537,075.51 ($429,684.29)	$792.15 ($633.75)	$40.00 ($25.38)
MIDDLESEX	2,685	1,533	57.09%	$2,411,530.66 ($1,916,900.30)	$898.15 (713.93)	$25.00 ($21.35)
NORFOLK	782	463	59.21%	$1,036,542.90 ($848,644.51)	$1,325.50 ($1,085.22)	$30.00 ($24.70)
BRISTOL	427	235	55.04%	$271,404.11 ($213,718.05)	$635.61 ($500.51)	$20.00 ($17.29)
PLYMOUTH	332	201	60.54%	$481,711.80 ($362,666.06)	$1,450.94 ($1,092.37)	$50.00 ($34.38)
EAST	**4,904**	**2,840**	**57.91%**	**$4,738,264.98**	**$966.20**	**$25.00**
WORCESTER	**3,292**	**1,758**	**53.40%**	**$4,213,135.05**	**$1,279.81**	**$20.00**
FRANKLIN	34	24	70.59%	$51,070.00 ($39,752.63)	$1,502.06 ($1,169.20)	$37.50 ($24.11)
HAMPDEN	326	189	57.98%	$690,662.66 ($554,435.68)	$2,118.60 ($1,700.72)	$47.50 ($31.81)
HAMPSHIRE	118	75	63.56%	$211,023.00 ($163,176.43)	$1,788.33 ($1,394.67)	$75.00 ($72.03)
BERKSHIRE	78	50	64.10%	$133,063.41 ($100,831.29)	$1,705.94 ($1,292.71)	$45.00 ($32.13)
WEST	**556**	**338**	**60.79%**	**$1,085,819.07**	**$1,952.91**	**$50.00**

Appendix D: Average and Median Donation by Fiscal Year 1983 to 2007

YEAR	VALUE OF AVERAGE DONATION AT YEAR DONATED (GIVEN THEY DONATED)	VALUE OF MEDIAN DONATION AT YEAR DONATED
2006	$509.80	$100.00
2005	$579.77	$100.00
2004	$595.13	$100.00
2003	$517.33	$100.00
2002	$492.74	$100.00
2001	$610.84	$100.00
2000	$538.61	$100.00
1999	$577.78	$100.00
1998	$446.31	$100.00
1997	$338.13	$50.00
1996	$290.36	$50.00
1995	$245.01	$50.00
1994	$247.81	$50.00
1993	$198.15	$50.00
1992	$190.65	$50.00
1991	$197.98	$50.00
1990	$193.09	$50.00
1989	$244.22	$50.00
1988	$153.51	$50.00
1987	$152.96	$50.00
1986	$155.10	$50.00
1985	$131.86	$50.00
1984	*$3,046.88*	*$3,046.88*
1983	*$395.87*	*$119.50*

Appendix E: Percent of People Donating by Fiscal Year 1983 to 2007

YEAR	# PEOPLE WHO DONATED	INCREASE IN # OF PEOPLE IN DATABASE	# PEOPLE IN DATABASE AT TIME OF DONATION (year minus 50 years)	% DONATING
2007	*166*	*-111*	*22,341*	*0.74%*
2006	4,109	*609*	22,452	18.30%
2005	3,982	*547*	21,843	18.23%
2004	4,054	*476*	21,296	19.04%
2003	4,040	*469*	20,820	19.40%
2002	3,887	*505*	20,351	19.10%
2001	4,140	*454*	19,846	20.86%
2000	4,180	*479*	19,392	21.56%
1999	4,539	*439*	18,913	24.00%
1998	4,481	*584*	18,474	24.26%
1997	4,603	*440*	17,890	25.73%
1996	4,478	*524*	17,450	25.66%
1995	4,712	*553*	16,926	27.84%
1994	5,186	*586*	16,373	31.67%
1993	5,101	*565*	15,787	32.31%
1992	5,198	*500*	15,222	34.15%
1991	4,934	*545*	14,722	33.51%
1990	5,068	*592*	14,177	35.75%
1989	4,349	*509*	13,585	32.01%
1988	4,309	*574*	13,076	32.95%
1987	4,449	*584*	12,502	35.59%
1986	4,093	*534*	11,918	34.34%
1985	4,014	*578*	11,384	35.26%
1984	*2*	*528*	*10,806*	*0.02%*
1983	*7,207*	*0*	*10,278*	*70.12%*
Total				

Appendix F: Total Donations by Fiscal Year

YEAR	PRESENT VALUE OF DONATION IN 2007	DONATION BY YEAR	DIFFERENCE	PRESENT VALUE OF LARGEST DONATION IN 2007	LARGEST DONATION BY YEAR	DIFFERENCE
2006	$2,094,787.13	$2,094,787.13	*$0.00*	$79,987.95	$79,987.95	*$0.00*
2005	$2,236,513.87	$2,308,659.48	*$72,145.61*	$99,469.54	$102,678.23	*$3,208.69*
2004	$2,260,663.05	$2,412,650.45	*$151,987.40*	$115,794.95	$123,580.00	*$7,785.05*
2003	$1,907,140.55	$2,090,017.04	*$182,876.49*	$91,250.00	$100,000.00	*$8,750.00*
2002	$1,708,920.20	$1,915,267.47	*$206,347.27*	$89,497.57	$100,304.15	*$10,806.58*
2001	$2,221,553.77	$2,528,883.34	*$307,329.57*	$108,903.10	$123,968.75	*$15,065.65*
2000	$1,923,065.05	$2,251,393.23	*$328,328.18*	$95,096.11	$111,332.03	*$16,235.92*
1999	$2,167,234.35	$2,622,535.69	*$455,301.34*	$121,448.92	$146,963.40	*$25,514.48*
1998	$1,617,004.47	$1,999,927.00	*$382,922.53*	$126,130.95	$156,000.00	*$29,869.05*
1997	$1,239,123.15	$1,556,431.32	*$317,308.17*	$61,600.65	$77,375.02	*$15,774.37*
1996	$1,011,942.64	$1,300,239.88	*$288,297.24*	$64,357.54	$82,692.67	*$18,335.13*
1995	$872,747.69	$1,154,500.88	*$281,753.19*	$75,595.24	$100,000.00	*$24,404.76*
1994	$944,735.22	$1,285,145.89	*$340,410.67*	$77,532.78	$105,469.69	*$27,936.91*
1993	$724,480.00	$1,010,762.41	*$286,282.41*	$73,244.51	$102,187.50	*$28,942.99*
1992	$689,681.02	$991,017.06	*$301,336.04*	$58,702.24	$84,350.47	*$25,648.23*
1991	$659,935.92	$976,821.45	*$316,885.53*	$17,270.06	$25,562.73	*$8,292.67*
1990	$634,432.37	$978,588.88	*$344,156.51*	$65,479.66	$101,000.00	*$35,520.34*
1989	$653,282.49	$1,062,110.88	*$408,828.39*	$73,747.20	$119,898.68	*$46,151.48*
1988	$388,159.02	$661,478.10	*$273,319.08*	$14,670.14	$25,000.00	*$10,329.86*
1987	$383,478.56	$680,539.42	*$297,060.86*	$9,861.11	$17,500.00	*$7,638.89*
1986	$345,112.30	$634,805.11	*$289,692.81*	$5,451.95	$10,028.40	*$4,576.45*
1985	$282,497.34	$529,288.70	*$246,791.36*	$13,356.60	$25,025.00	*$11,668.40*
1984	$3,140.58	*$6,093.76*	*$2,953.18*	$2,934.43	*$5,693.76*	*$2,759.33*
1983	$1,409,550.43	*$2,853,065.94*	*$1,443,515.51*	$28,027.32	*$56,730.00*	*$28,702.68*
TOTAL	**$28,437,136.78**	**$35,962,966.10**				

Appendix G: Score Factors for Zip Code Ranges

Zip Min	Zip Max	State	Score Factor
0	1000	NOT IN USE	**-0.75**
1001	2791	MA	**0.5**
2792	2800	NOT IN USE	**-0.75**
2801	2940	RI	**-1.0**
2941	3030	NOT IN USE	**-0.75**
3031	3897	NH	**0.5**
3898	3900	NOT IN USE	**-0.75**
3901	4992	ME	**0.5**
4993	5000	NOT IN USE	**-0.75**
5001	5495	VT	**0.5**
5496	5500	NOT IN USE	**-0.75**
5501	5544	MA	**0.5**
5545	5600	NOT IN USE	**-0.75**
5601	5907	VT	**0.5**
5908	6000	NOT IN USE	**-0.75**
6001	6389	CT	**0.75**
6390	6390	NY	**0.5**
6391	6400	NOT IN USE	**-0.75**
6401	6928	CT	**0.75**
6929	7000	NOT IN USE	**-0.75**
7001	8989	NJ	**1.0**
8990	10000	NOT IN USE	**-0.75**
10001	14975	NY	**0.5**
14976	15000	NOT IN USE	**-0.75**
15001	19640	PA	**1.0**
19641	19700	NOT IN USE	**-0.75**
19701	19980	DE	**1.0**
19981	20000	NOT IN USE	**-0.75**
20001	20039	DC	**-1.0**
20040	20167	VA	**0.5**
20168	20599	DC	**-1.0**
20600	20797	MD	**0.5**
20798	20798	NOT IN USE	**-0.75**
20799	20799	DC	**-1.0**
20800	20811	NOT IN USE	**-0.75**
20812	21930	MD	**0.5**
21931	22000	NOT IN USE	**-0.75**

22001	24658	VA	**0.5**
24659	24700	NOT IN USE	**-0.75**
24701	26886	WV	**0.5**
26887	27005	NOT IN USE	**-0.75**
27006	28909	NC	**0.5**
28910	29000	NOT IN USE	**-0.75**
29001	29948	SC	**0.5**
29949	30000	NOT IN USE	**-0.75**
30001	31999	GA	**0.5**
32000	32003	NOT IN USE	**-0.75**
32004	34997	FL	**1.0**
34998	35003	NOT IN USE	**-0.75**
35004	36925	AL	**0.5**
36926	37009	NOT IN USE	**-0.75**
37010	38589	TN	**0.5**
38590	38600	NOT IN USE	**-0.75**
38601	39776	MS	**0.5**
39777	39900	NOT IN USE	**-0.75**
39901	39901	GA	**0.5**
39902	40002	NOT IN USE	**-0.75**
40003	42788	KY	**0.5**
42789	43000	NOT IN USE	**-0.75**
43001	45999	OH	**1.0**
46000	46000	NOT IN USE	**-0.75**
46001	47997	IN	**0.5**
47998	48000	NOT IN USE	**-0.75**
48001	49971	MI	**0.75**
49972	50000	NOT IN USE	**-0.75**
50001	52809	IA	**0.5**
52810	53000	NOT IN USE	**-0.75**
53001	54990	WI	**0.5**
54991	55000	NOT IN USE	**-0.75**
55001	56763	MN	**0.5**
56764	57000	NOT IN USE	**-0.75**
57001	57799	SD	**0.5**
57800	58000	NOT IN USE	**-0.75**
58001	58856	ND	**0.5**
58857	59000	NOT IN USE	**-0.75**
59001	59937	MT	**-1.0**
59938	60000	NOT IN USE	**-0.75**
60001	62999	IL	**1.0**

63000	63000	NOT IN USE	**-0.75**
63001	65899	MO	**0.5**
65900	66001	NOT IN USE	**-0.75**
66002	67954	KS	**0.5**
67955	68000	NOT IN USE	**-0.75**
68001	68118	NE	**0.5**
68119	68120	IA	**0.5**
68121	68121	NOT IN USE	**-0.75**
68122	69367	NE	**0.5**
69368	70000	NOT IN USE	**-0.75**
70001	71232	LA	**0.5**
71233	71233	MS	**0.5**
71234	71497	LA	**0.5**
71498	71600	NOT IN USE	**-0.75**
71601	72959	AR	**0.5**
72960	73000	NOT IN USE	**-0.75**
73001	73199	OK	**0.5**
73200	73300	NOT IN USE	**-0.75**
73301	73301	TX	**0.5**
73302	73400	NOT IN USE	**-0.75**
73401	74966	OK	**0.5**
74967	75000	NOT IN USE	**-0.75**
75001	75501	TX	**0.5**
75502	75502	AR	**0.5**
75503	79999	TX	**0.5**
80000	80000	NOT IN USE	**-0.75**
80001	81658	CO	**0.5**
81659	82000	NOT IN USE	**-0.75**
82001	83128	WY	**0.5**
83129	83200	NOT IN USE	**-0.75**
83201	83876	ID	**0.5**
83877	84000	NOT IN USE	**-0.75**
84001	84784	UT	**0.5**
84785	85000	NOT IN USE	**-0.75**
85001	86556	AZ	**0.5**
86557	87000	NOT IN USE	**-0.75**
87001	88441	NM	**0.5**
88442	88509	NOT IN USE	**-0.75**
88510	88589	TX	**0.5**
88590	88900	NOT IN USE	**-0.75**
88901	89883	NV	**0.5**

89884	90000	NOT IN USE	**-0.75**
90001	96162	CA	**1.0**
96163	96700	NOT IN USE	**-0.75**
96701	96898	HI	**0.5**
96899	97000	NOT IN USE	**-0.75**
97001	97920	OR	**0.5**
97921	98000	NOT IN USE	**-0.75**
98001	99403	WA	**0.5**
99404	99500	NOT IN USE	**-0.75**
99501	99950	AK	**0.5**
99951	99999	NOT IN USE	**-0.75**

Appendix H: User's Manual provided for Donor Score System

User's manual for *Donor Score System.xls*

This is a user's manual for Donor Score System.xls, an excel spreadsheet which will determine a score to predict the likelihood of a donation for individuals in the Alumni database. This manual includes a detailed description of how the data must be entered, and then will run through an example run of 100 Alumni.

Getting the Data

When using this spreadsheet, it is critical that the data be entered correctly on the spreadsheet. Further, the spreadsheet was developed only for WPI Alumni. This means that the CATEGORY field should be "ALUM" for all those to be entered into the spreadsheet. Using the spreadsheet for other categories will not produce valid results. It is assumed that data will be taken from the Alumni Office database and put into a spreadsheet. The following are the specific items needed. They should be put in an Excel spreadsheet in the order shown:

PERSON_NUMBER	(Column A)
MARRIED	(Column B)
NON_WPI_DEG	(Column C)
FRAT	(Column D)
VARSITY_SPRTS	(Column E)
RES_ZIP	(Column F)
GENDER	(Column G)
WPI_SPS	(Column H)
NUM_OF_CHILD	(Column I)
PREF_CLAS	(Column J)
HAD_SCHOLARSHIP	(Column K)
CLS_AGENT	(Column L)
REUNION	(Column M)
ALUM_VOLUNTEER	(Column N)
STU_PROJECT_CTR	(Column O)
STU_GPA	(Column P)
STU_INTL_CLUB	(Column Q)
STU_CLUB_SPORT	(Column R)
STU_PROF_SOC	(Column S)
STU_MUSIC	(Column T)
STU_SCH_INVOLVE	(Column U)
STU_HONOR_SOC	(Column V)
GRAD_DISTINCTION	(Column W)

TOTAL DONATION (Column X)

The data should look similar to this:

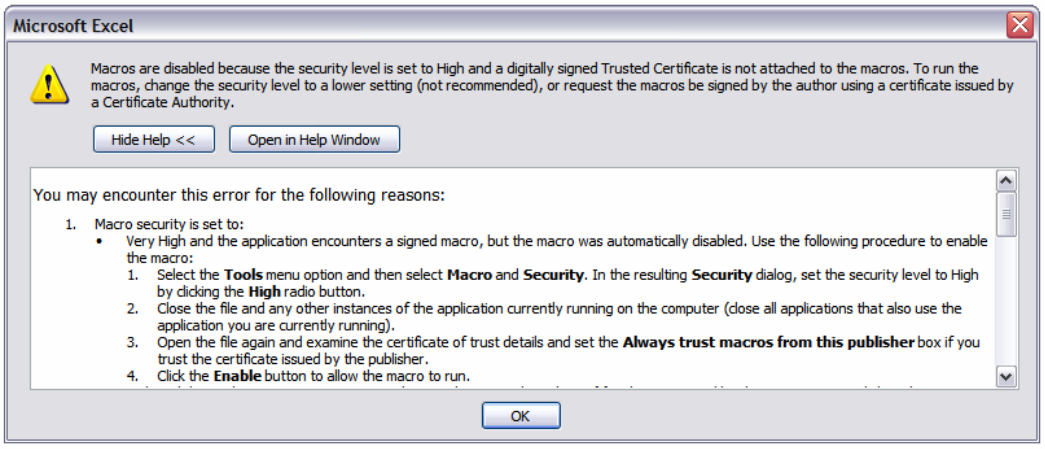
	A	B	C	D	E	F	G	H	I	J	K	L
1	PERSON_NUMBER	MARRIED	NON_WPI_DEG	FRAT	VARSITY_SPRTS	RES_ZIP	GENDER	WPI_SPS	NUM_OF_CHILD	PREF_CLAS	HAD_SCHOLARSHIP	CLS_AGE
2	#######	Married	Hartford State Technical Coll			06415-1025	M		4	1967		
3	#######	Married			BASEBALL	06374-1733	M		3	1983	YES	
4	#######	Married	Brown University:PHD:19	TAU KAPPA EPSILON		08502-4023	M		1	1979	YES	
5	#######	Single			CROSS COUNTRY (men)	02135	M			1999	YES	
6	#######	Married				06109-2350	M			1950		
7	#######	Divorced		SIGMA PHI EPSILON		01520	M		4	1949		
8	#######	Single				01033-9703	M			1974	YES	
9	#######	Married		ALPHA TAU OMEGA		11705-2153	M		1	1994	YES	
10	#######	Single			TENNIS	02116	M			1996	YES	
11	#######	Married			SOCCER	01463-1743	M	YES		1989		
12	#######	Single		LAMBDA CHI ALPHA		30339-4138	M			1997		
13	#######	Widowed		PHI KAPPA THETA	SOCCER	06001-3207	M		2	1958	YES	
14	#######	Married	Univ Texas Dallas:MS:19	LAMBDA CHI ALPHA		02842-7201	M			1981	YES	
15	#######	Single				02138-4646	M			2001	YES	
16	#######	Married				02723	M			1995	YES	
17	#######	Married		LAMBDA CHI ALPHA		66048	M		2	1975		
18	#######	Single		ALPHA GAMMA DELTA		01545-3620	F			1998	YES	
19	#######	Married	Colorado Tech:MS:1992			03304-4012	M			1985		
20	#######	Married		PHI SIGMA KAPPA	BASKETBALL (Men)	23602-7028	M		3	1951	YES	
21	#######	Single		THETA CHI	SWIMMING (Men)	01845-5612	M			1991	YES	
22	#######	Single				07054-5504	F			2004	YES	
23	#######	Single	Mass Institute Technology:MS:1988		GOLF;TENNIS	02780-1057	M			1986		
24	#######	Married	Rensselaer Poly Institute:MS:1982			02351-1420	M		2	1987	YES	
25	#######	Married	Northeastern University:M	SIGMA ALPHA EPSILO	SOCCER;TENNIS	02643-1079	M		2	1953	YES	
26	#######	Single				01520-2106	M			1983		
27	#######	Single				02151-5818	M			2000	YES	
28	#######	Single				01902	M			2000	YES	
29	#######	Single				01906-4533	M			2002	YES	
30	#######	Married	Stonehill College:BA:1993			07848	M			1994		
31	#######	Married	Orange County Cmty College*Trinity College Ct::C		SOCCER		M		2	1976	YES	
32	#######	Single			FOOTBALL;TRACK (MEN)	01606	M			1998	YES	
33	#######	Married				22181-5300	M		1	1995		
34	#######	Single				48154-5314	M			1965	YES	
35	#######	Married				08043-4711	F			1999	YES	
36	#######	Married	Bucknell University:MBA:	PHI GAMMA DELTA	SOCCER	22203	M		2	1966	YES	YES
37	#######	Married	Rensselaer Poly Institute:MS:1990			01534-1260	M		2	1983		
38	#######	Single		PHI KAPPA THETA		03038-1805	M			2003	YES	
39	#######	Married	Univ Mass Amherst:MS			02139-0024	M			1979	YES	
40	#######	Married	Univ Rochester:MBA:199	ALPHA GAMMA DELTA	SOFTBALL	06330	F	YES	2	1991	YES	
41	#######	Married			BASEBALL;FOOTBALL	85249	M			1997	YES	
42	#######	Married				98103-7603	M			1990	YES	
43	#######	Married				01077-9418	M			1973	YES	
44	#######	Single				02723	M			1997	YES	
45	#######	Divorced		ALPHA TAU OMEGA	GOLF	02122-3211	M			1993	YES	

NOTE: The PERSON_NUMBER has been replaced with "######" for confidentiality.

Once you have checked over the data you are ready to open the *Donor Score System* spreadsheet.

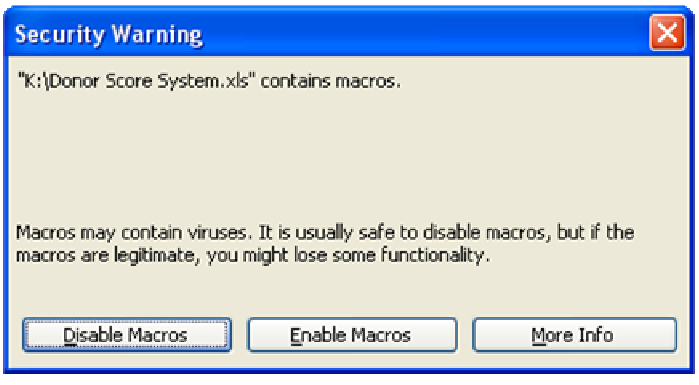
Opening the Spreadsheet

Find the file Donor Score System.xls and open it. You may immediately receive a message explaining that the macros have been disabled due to the security level. This message will look like this:



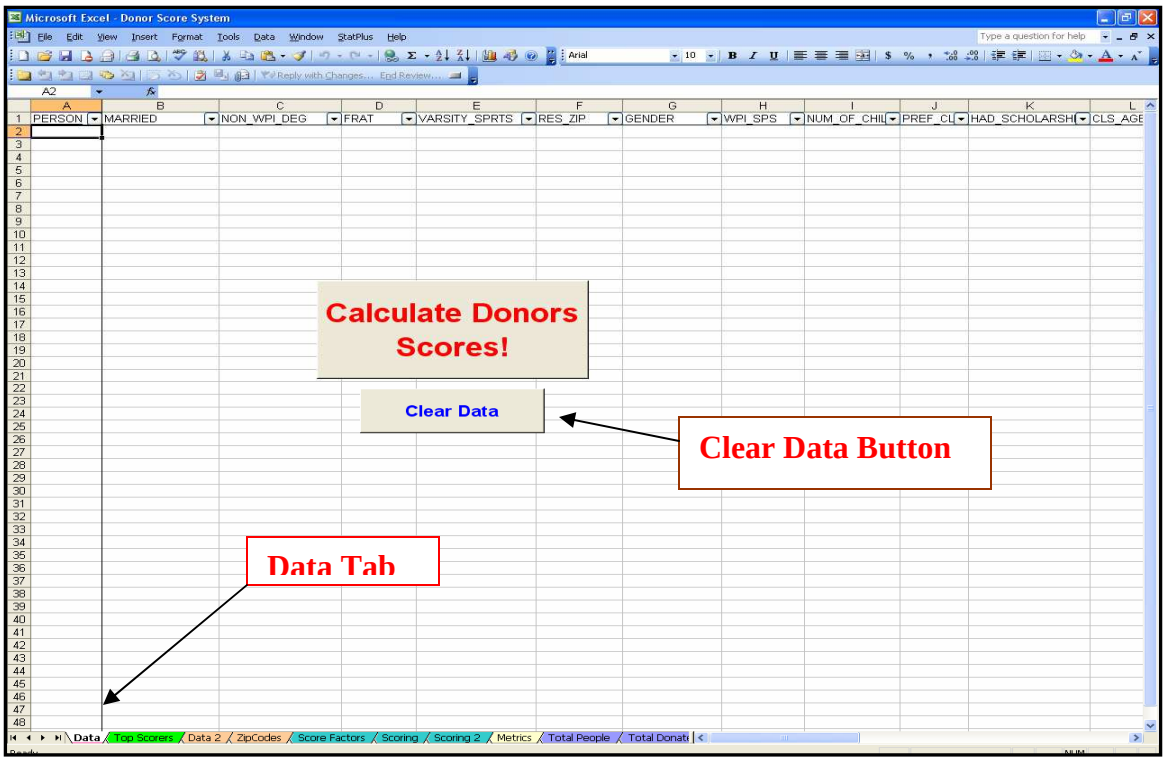
If this happens, you will need to click "**OK**" and then on the menu bar go to the **Tools** menu option and select **Macro** and then **Security**. In the resulting Security dialog, change the security level to medium, then close the spreadsheet and try opening it again.

This time when the excel spreadsheet is opened a security warning will pop-up that looks like this:



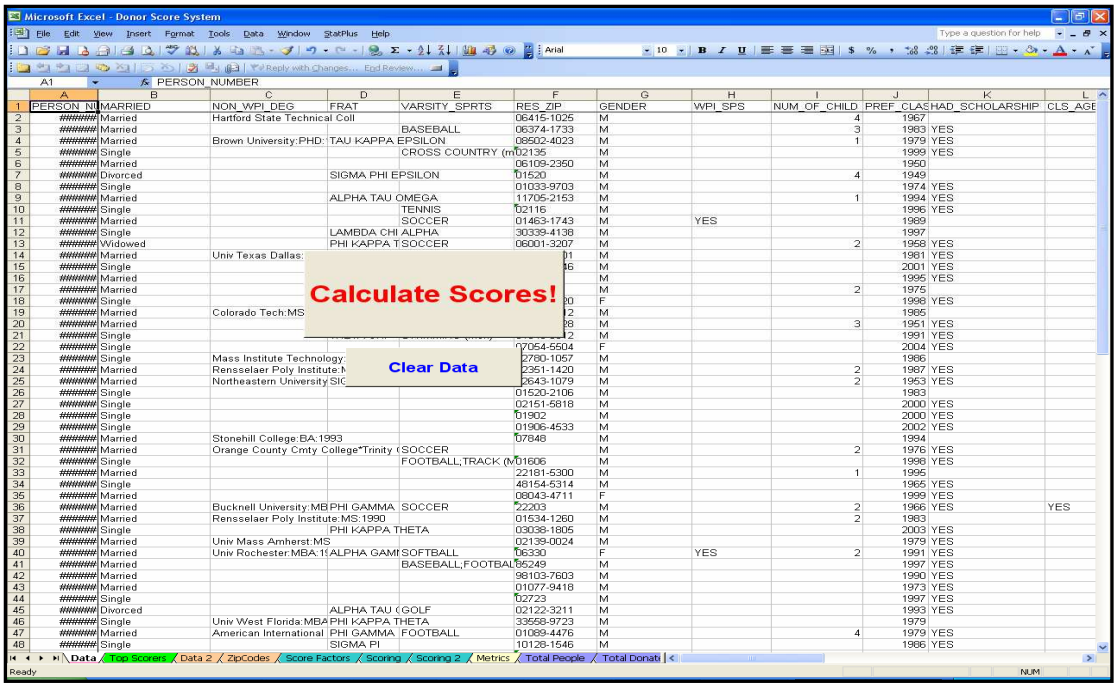
Click on “**Enable Macros**” to open the spreadsheet.

The spreadsheet should open on the “Data” tab. If it does not, click on the first tab in the spreadsheet that is labeled “Data”, as seen below. If the spreadsheet has any data in it already or is not blank, be sure to click on the “Clear Data” button, to clear out any remaining values.



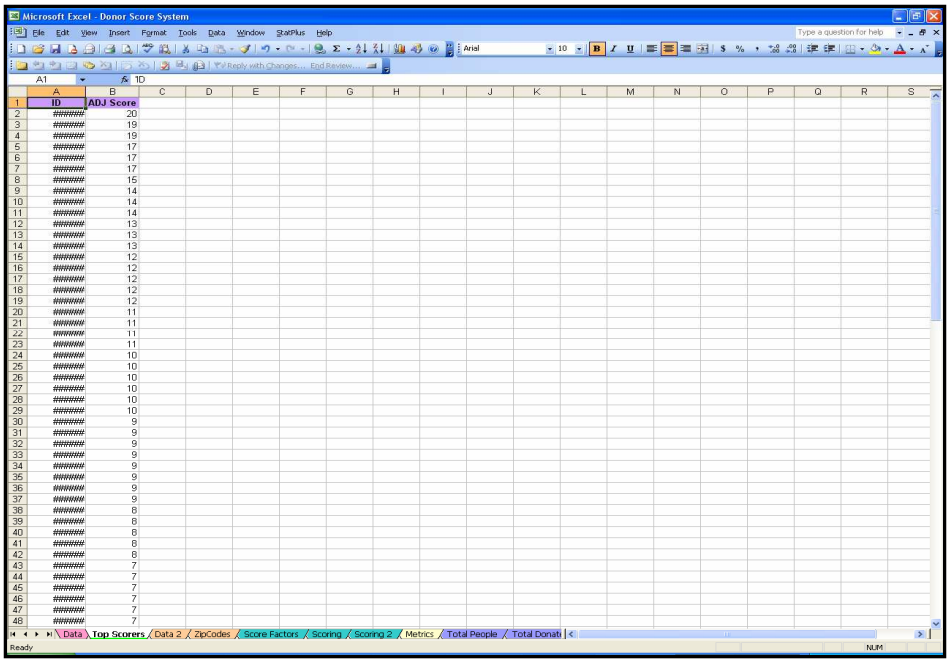
Running the Spreadsheet

At this point return to the Excel File with the Alumni data, where you should copy all 24 factors for each individual. To do this, highlight all the alumni and their factors. Copy all these values by holding down the “Ctrl” key and then typing “c”. Return to the Donor Score System.xls and click on cell **B1,** and paste the data into the spreadsheet by holding down the “Ctrl” key and typing “v”. Double check your data to make sure that the data copied over right, and that all the Alumni factors are lining up with the factors listed on the top row of the “Data” tab. You should now have all the data in the spreadsheet, and it should look like this:

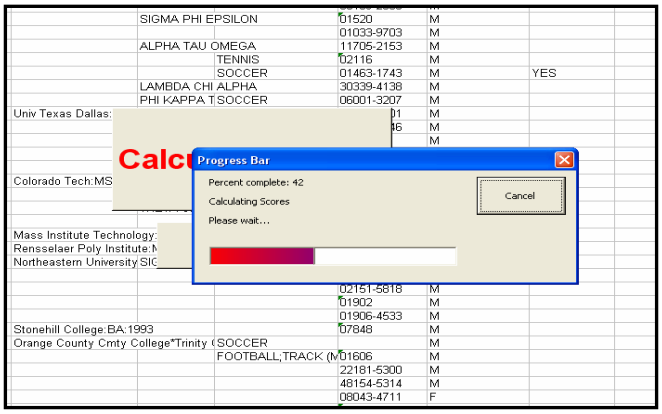


You are now ready to run the spreadsheet; click on the “Calculate Scores!” button.

When the calculations are done, the spreadsheet will automatically move to the "Top Scorers" tab, with a list of all ID numbers and their respective scores in descending order of the score, as seen below:

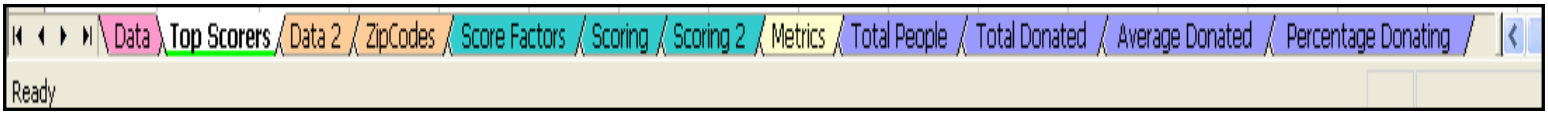


NOTE: This process can take hours for large amounts of data. A progress bar pops up to let you know the spreadsheet is working.



Viewing the Results

Now that the calculations are complete you can review your results. There are a total of 12 different tabs in this spreadsheet. Below is an enlarged image of all the tabs.



- **"Top Scorers" Tab**

  The "Top Scorers" Tab has, as mentioned, each individuals ID number and their calculated score sorted in descending order by score. This is the most important tab, as it ranks each individual as to how likely they are to donate. The higher a score the more likely a person is to donate, and the higher the donation amount will be. For example a person with a score of 20 is the most likely to donate money and a person with a score of 1 is probably least likely to donate. Because these scores were calculated using the factors in the database, this does not always mean that it is completely accurate. It is possible for a person with a score of 1 to give a significant donation, just as it is possible for a person with a score of 20 to not donate at all. Ways to measure if theses scores are accurate can be seen in graphs, and calculations which are described later in this manual.

> **NOTE:** The list of Top Scorers can be copied and moved to another spreadsheet or perhaps written back to the Alumni database for further reference.

- **“Data 2” Tab**

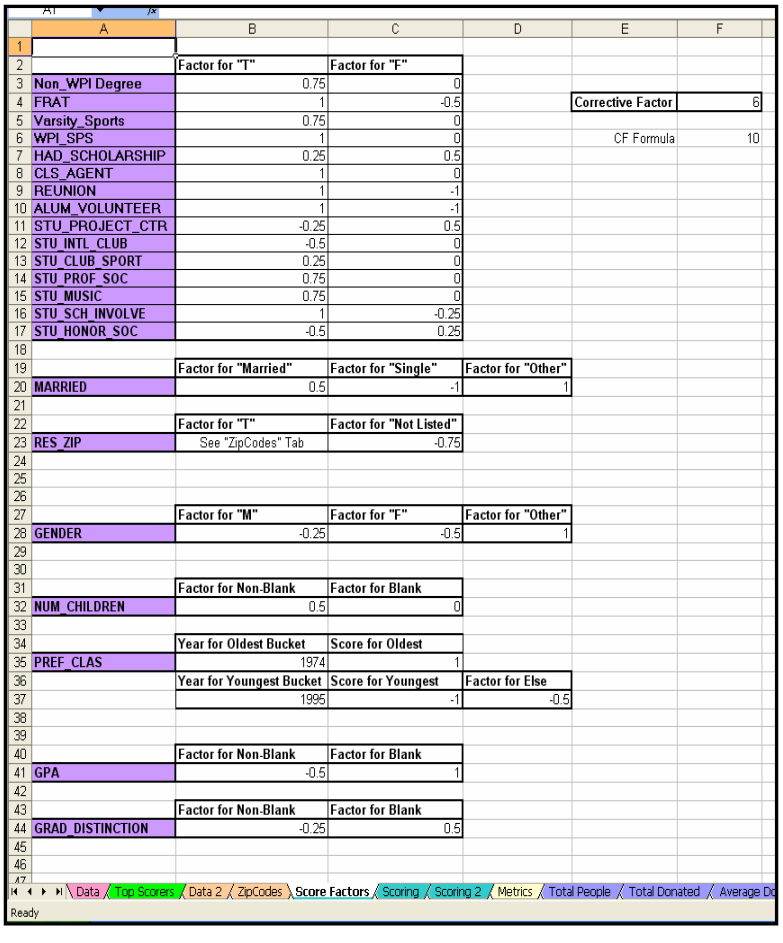
  The “Data 2” tab is used only for calculation purposes only. It simplifies the given data, so that it can be used for calculations later, by turning many of the values into “Y”, so that it can be easier converted to a numeric value later.

- **“ZipCodes” Tab**

  The “ZipCodes” tab is also used for calculation purposes. It has a list of all zip codes currently used in the U.S. with a corresponding score that is used for the calculation of the final score.

- **“Score Factors” Tab**

  This tab is essential in calculating each individual’s score.

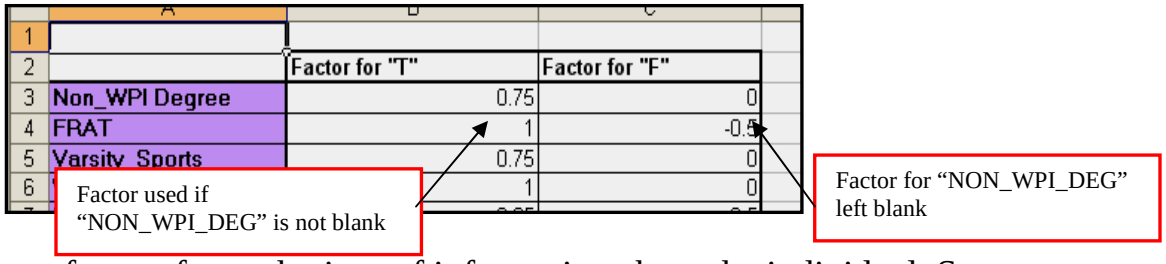


		Factor for "T"	Factor for "F"			
3	Non_WPI Degree	0.75	0			
4	FRAT	1	-0.5		Corrective Factor	6
5	Varsity_Sports	0.75	0			
6	WPI_SPS	1	0		CF Formula	10
7	HAD_SCHOLARSHIP	0.25	0.5			
8	CLS_AGENT	1	0			
9	REUNION	1	-1			
10	ALUM_VOLUNTEER	1	-1			
11	STU_PROJECT_CTR	-0.25	0.5			
12	STU_INTL_CLUB	-0.5	0			
13	STU_CLUB_SPORT	0.25	0			
14	STU_PROF_SOC	0.75	0			
15	STU_MUSIC	0.75	0			
16	STU_SCH_INVOLVE	1	-0.25			
17	STU_HONOR_SOC	-0.5	0.25			
19		Factor for "Married"	Factor for "Single"	Factor for "Other"		
20	MARRIED	0.5	-1	1		
22		Factor for "T"	Factor for "Not Listed"			
23	RES_ZIP	See "ZipCodes" Tab	-0.75			
27		Factor for "M"	Factor for "F"	Factor for "Other"		
28	GENDER	-0.25	-0.5	1		
31		Factor for Non-Blank	Factor for Blank			
32	NUM_CHILDREN	0.5	0			
34		Year for Oldest Bucket	Score for Oldest			
35	PREF_CLAS	1974	1			
36		Year for Youngest Bucket	Score for Youngest	Factor for Else		
37		1995	-1	-0.5		
40		Factor for Non-Blank	Factor for Blank			
41	GPA	-0.5	1			
43		Factor for Non-Blank	Factor for Blank			
44	GRAD_DISTINCTION	-0.25	0.5			

Data | Top Scorers | Data 2 | ZipCodes | Score Factors | Scoring | Scoring 2 | Metrics | Total People | Total Donated | Average Do

Ready

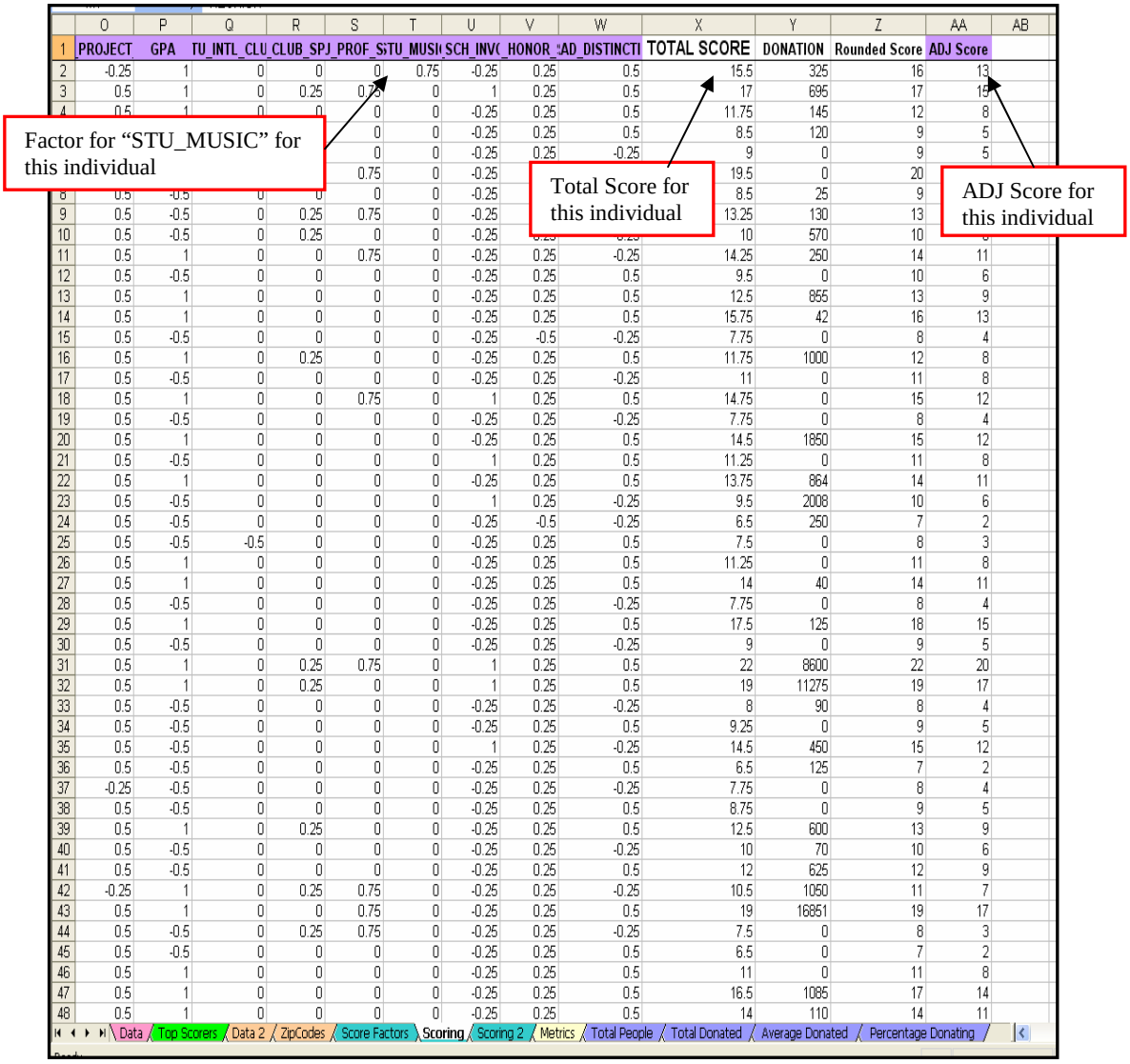
Above are the individual scorings that are associated with each piece of information about each person. For example, in the diagram you can see that for a person that does have a non WPI degree will receive .75 for that individual piece of information, whereas someone who does not have a non WPI degree (i.e. left it blank) will receive a 0.



There are factors for each piece of information about the individual. Some are categorized as blank or non-blank, while others such as "Gender" have different factors for "M", "F", and "Other" (e.g. "N")

- **"Scoring" Tab**

  The "Scoring" tab is where all the calculations occur. For each individual, it looks at the "Data 2" and "Score Factors" tab, and the places the respective factor for each piece of information. As seen in the screenshot below, each individual ends up with 22 different factors based on their information, which then get summed up for their Total Score. From the Total Score a final calculation is done to adjust the scores so that they only range from 1 to 20. This is called the "ADJ Score", which is used for all the graphs and is shown on the "Top Scorers" Tab.



	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	PROJECT	GPA	TU_INTL_CLU	CLUB_SPJ	PROF_S	TU_MUSI	SCH_INV	HONOR_	AD_DISTINCTI	TOTAL SCORE	DONATION	Rounded Score	ADJ Score	
2	-0.25	1	0	0	0	0.75	-0.25	0.25	0.5	15.5	325	16	13	
3	0.5	1	0	0.25	0.75	0	1	0.25	0.5	17	695	17	15	
4	0.5	1	0	0	0	0	-0.25	0.25	0.5	11.75	145	12	8	
5	0.5	1	0	0	0	0	-0.25	0.25	0.5	8.5	120	9	5	
6	0.5	1	0	0	0	0	-0.25	0.25	-0.25	9	0	9	5	
7	0.5	1	0	0	0.75	0	-0.25			19.5	0	20		
8	0.5	-0.5	0	0	0	0	-0.25			8.5	25	9		
9	0.5	-0.5	0	0.25	0.75	0	-0.25			13.25	130	13		
10	0.5	-0.5	0	0.25	0	0	-0.25			10	570	10		
11	0.5	1	0	0	0.75	0	-0.25	0.25	-0.25	14.25	250	14	11	
12	0.5	-0.5	0	0	0	0	-0.25	0.25	0.5	9.5	0	10	6	
13	0.5	1	0	0	0	0	-0.25	0.25	0.5	12.5	855	13	9	
14	0.5	1	0	0	0	0	-0.25	0.25	0.5	15.75	42	16	13	
15	0.5	-0.5	0	0	0	0	-0.25	-0.5	-0.25	7.75	0	8	4	
16	0.5	1	0	0.25	0	0	-0.25	0.25	0.5	11.75	1000	12	8	
17	0.5	-0.5	0	0	0	0	-0.25	0.25	-0.25	11	0	11	8	
18	0.5	1	0	0	0.75	0	1	0.25	0.5	14.75	0	15	12	
19	0.5	-0.5	0	0	0	0	-0.25	0.25	-0.25	7.75	0	8	4	
20	0.5	1	0	0	0	0	-0.25	0.25	0.5	14.5	1850	15	12	
21	0.5	-0.5	0	0	0	0	1	0.25	0.5	11.25	0	11	8	
22	0.5	1	0	0	0	0	-0.25	0.25	0.5	13.75	864	14	11	
23	0.5	-0.5	0	0	0	0	1	0.25	-0.25	9.5	2008	10	6	
24	0.5	-0.5	0	0	0	0	-0.25	-0.5	-0.25	6.5	250	7	2	
25	0.5	-0.5	-0.5	0	0	0	-0.25	0.25	0.5	7.5	0	8	3	
26	0.5	1	0	0	0	0	-0.25	0.25	0.5	11.25	0	11	8	
27	0.5	1	0	0	0	0	-0.25	0.25	0.5	14	40	14	11	
28	0.5	-0.5	0	0	0	0	-0.25	0.25	-0.25	7.75	0	8	4	
29	0.5	1	0	0	0	0	-0.25	0.25	0.5	17.5	125	18	15	
30	0.5	-0.5	0	0	0	0	-0.25	0.25	-0.25	9	0	9	5	
31	0.5	1	0	0.25	0.75	0	1	0.25	0.5	22	8600	22	20	
32	0.5	1	0	0.25	0	0	1	0.25	0.5	19	11275	19	17	
33	0.5	-0.5	0	0	0	0	-0.25	0.25	-0.25	8	90	8	4	
34	0.5	-0.5	0	0	0	0	-0.25	0.25	0.5	9.25	0	9	5	
35	0.5	-0.5	0	0	0	0	1	0.25	-0.25	14.5	450	15	12	
36	0.5	-0.5	0	0	0	0	-0.25	0.25	0.5	6.5	125	7	2	
37	-0.25	-0.5	0	0	0	0	-0.25	0.25	-0.25	7.75	0	8	4	
38	0.5	-0.5	0	0	0	0	-0.25	0.25	0.5	8.75	0	9	5	
39	0.5	1	0	0.25	0	0	-0.25	0.25	0.5	12.5	600	13	9	
40	0.5	-0.5	0	0	0	0	-0.25	0.25	-0.25	10	70	10	6	
41	0.5	-0.5	0	0	0	0	-0.25	0.25	0.5	12	625	12	9	
42	-0.25	1	0	0.25	0.75	0	-0.25	0.25	-0.25	10.5	1050	11	7	
43	0.5	1	0	0	0.75	0	-0.25	0.25	0.5	19	16851	19	17	
44	0.5	-0.5	0	0.25	0.75	0	-0.25	0.25	-0.25	7.5	0	8	3	
45	0.5	-0.5	0	0	0	0	-0.25	0.25	0.5	6.5	0	7	2	
46	0.5	1	0	0	0	0	-0.25	0.25	0.5	11	0	11	8	
47	0.5	1	0	0	0	0	-0.25	0.25	0.5	16.5	1085	17	14	
48	0.5	1	0	0	0	0	-0.25	0.25	0.5	14	110	14	11	

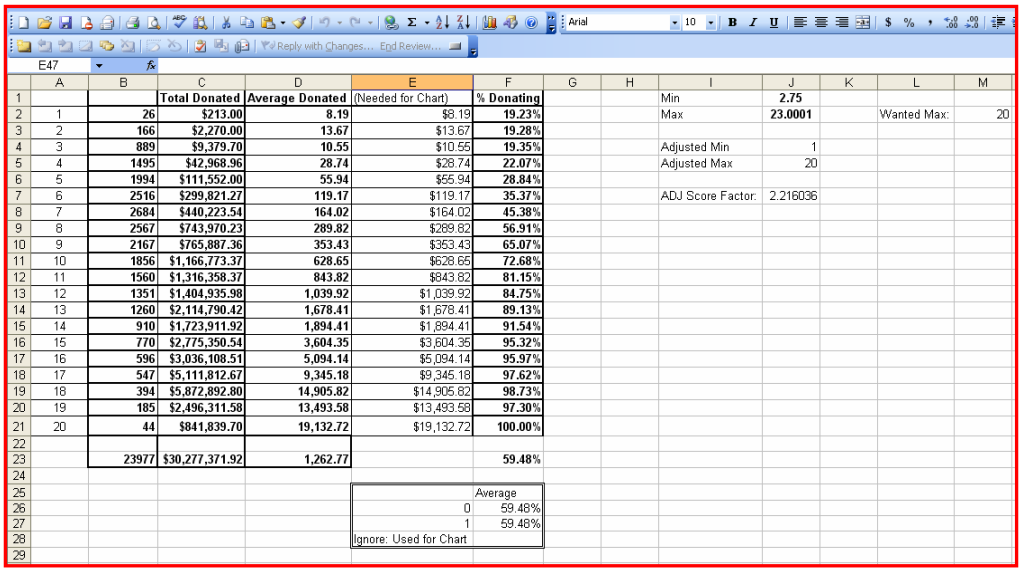
Data | Top Scorers | Data 2 | ZipCodes | Score Factors | Scoring | Scoring 2 | Metrics | Total People | Total Donated | Average Donated | Percentage Donating

NOTE: To portray the most desirable results, for the remainder of this manual all screenshots are based on the 23,977 Alumni that the spreadsheet was run on.

- **"Scoring 2" Tab**

The "Scoring 2" tab summarizes the basic statistics on the calculated scores. It has the number of people with each score, the total donated per score, the average donation per score and the percentage of people donating per score.

The graphs are drawn using these statistics. This tab also shows the minimum and maximum Total and Adjusted scores.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1			Total Donated	Average Donated	(Needed for Chart)	% Donating			Min	2.75			
2	1	26	$213.00	8.19	$8.19	19.23%			Max	23.0001		Wanted Max:	20
3	2	166	$2,270.00	13.67	$13.67	19.28%							
4	3	889	$9,379.70	10.55	$10.55	19.35%			Adjusted Min	1			
5	4	1495	$42,968.96	28.74	$28.74	22.07%			Adjusted Max	20			
6	5	1994	$111,552.00	55.94	$55.94	28.84%							
7	6	2516	$299,821.27	119.17	$119.17	35.37%			ADJ Score Factor:	2.216036			
8	7	2684	$440,223.54	164.02	$164.02	45.38%							
9	8	2567	$743,970.23	289.82	$289.82	56.91%							
10	9	2167	$765,887.36	353.43	$353.43	65.07%							
11	10	1856	$1,166,773.37	628.65	$628.65	72.68%							
12	11	1560	$1,316,358.37	843.82	$843.82	81.15%							
13	12	1351	$1,404,935.98	1,039.92	$1,039.92	84.75%							
14	13	1260	$2,114,790.42	1,678.41	$1,678.41	89.13%							
15	14	910	$1,723,911.92	1,894.41	$1,894.41	91.54%							
16	15	770	$2,775,350.54	3,604.35	$3,604.35	95.32%							
17	16	596	$3,036,108.51	5,094.14	$5,094.14	95.97%							
18	17	547	$5,111,812.67	9,345.18	$9,345.18	97.62%							
19	18	394	$5,872,892.80	14,905.82	$14,905.82	98.73%							
20	19	185	$2,496,311.58	13,493.58	$13,493.58	97.30%							
21	20	44	$841,839.70	19,132.72	$19,132.72	100.00%							
22													
23		23977	$30,277,371.92	1,262.77		59.48%							
24													
25						Average							
26					0	59.48%							
27					1	59.48%							
28					Ignore: Used for Chart								
29													

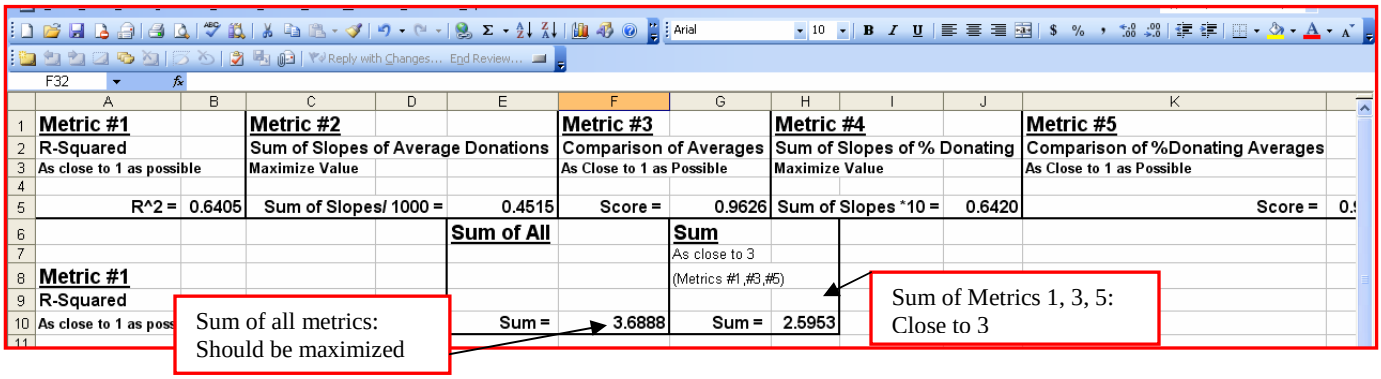
- **“Metrics” Tab**

The “Metrics” tab is used to see how good the scoring system currently is. There are five different calculations;

1. R^2 value on the average donation;
2. the Sum of slopes of average donation;
3. the comparison of averages for the average donation;
4. the Sum of slopes of the percentage of people donating and
5. the comparison of averages for the percentage of people donating.

While this manual does not include details on these calculations, you can find them in the MQP paper. Each metric has a numeric value, and the sum of all these values should be maximized with the sum of metrics 1, 3, and 5 as close to 3 as possible.

The shown example below is based on the 23,900+ database (in a database of only 100, these score and sums would be very low). As you can see, this fit is a good fit, because the sum of Metrics #1, #3 and #5 is almost 2.6.



	A	B	C	D	E	F	G	H	I	J	K
1	**Metric #1**		**Metric #2**			**Metric #3**		**Metric #4**			**Metric #5**
2	R-Squared		Sum of Slopes of Average Donations			Comparison of Averages		Sum of Slopes of % Donating			Comparison of %Donating Averages
3	As close to 1 as possible		Maximize Value			As Close to 1 as Possible		Maximize Value			As Close to 1 as Possible
4											
5	R^2 =	0.6405	Sum of Slopes/ 1000 =		0.4515	Score =	0.9626	Sum of Slopes *10 =		0.6420	Score = 0.9
6					**Sum of All**		**Sum**				
7							As close to 3				
8	**Metric #1**						(Metrics #1,#3,#5)				
9	R-Squared										
10	As close to 1 as poss				Sum =	3.6888	Sum =	2.5953			
11											

Sum of all metrics:
Should be maximized

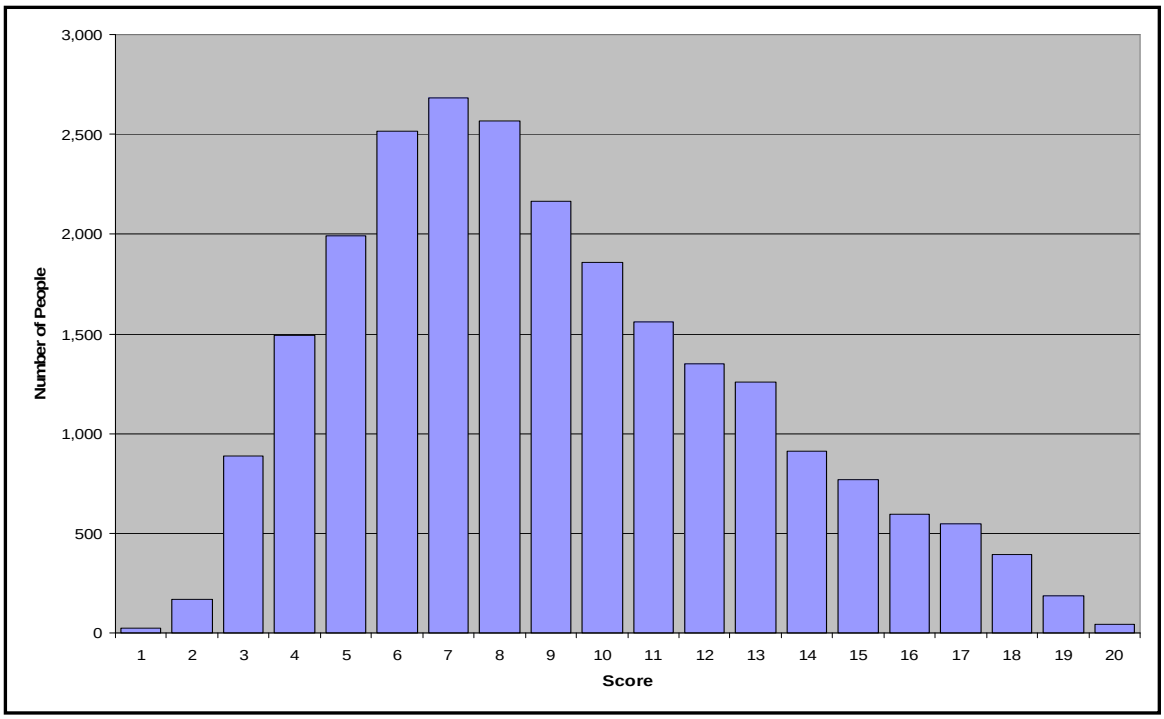
Sum of Metrics 1, 3, 5:
Close to 3

- **Graphs**

  The last 4 tabs are graphs to see how well the scores correlate with the donation behavior.

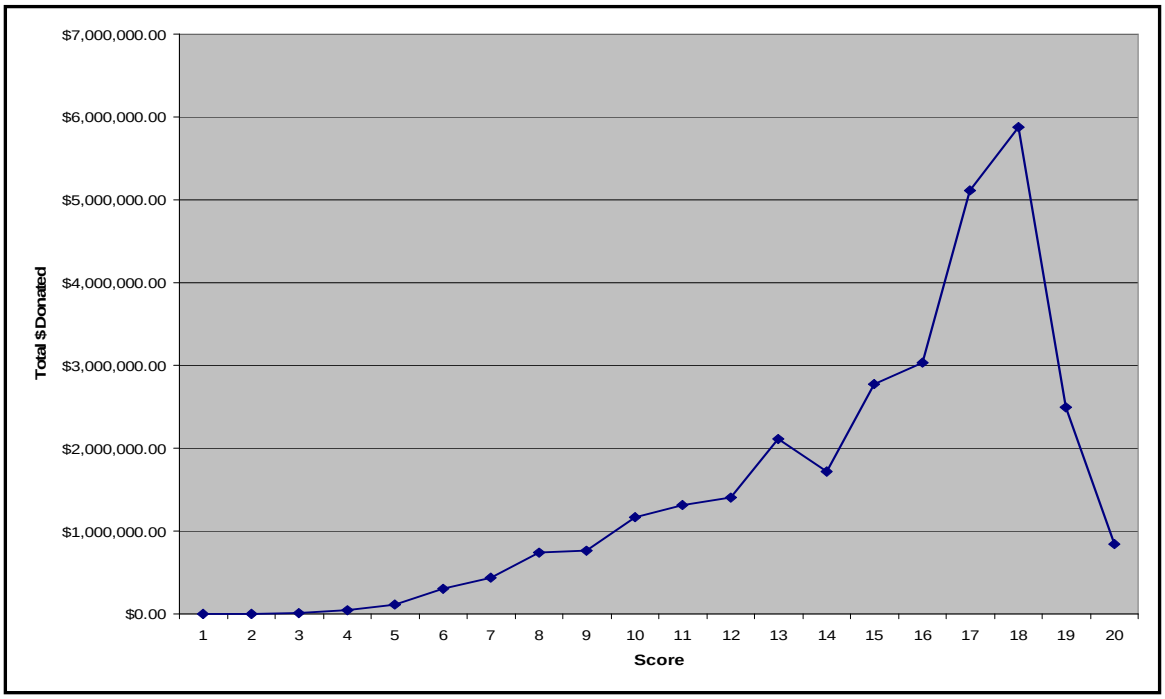
  - **“Total People**

    This graph has number of people with each score portrayed in a bar graph. The ideal shape of this graph would be to have a fairly equal amount in each bar. However it is most likely that it will have lower amounts of people in the higher scoring “buckets”.



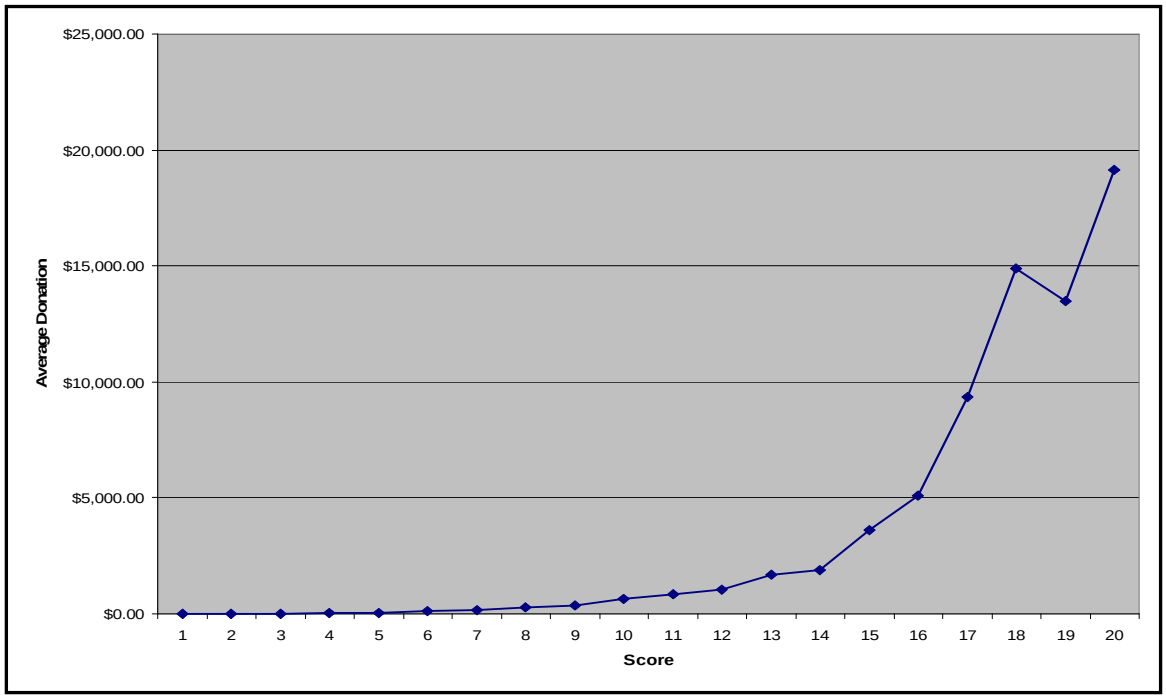
- **“Total Donated”**

  This graph portrays in a line graph the total amount donated per score. Since there are not an equal number of people in each bucket it is hard to say what the ideal shape of this graph would be. Although a low total donation should be associated with the lower scores, for best results.



- **“Average Donation”**

  The “Average Donation” graph displays the average donation with its corresponding score on a line graph. The ideal shape of this graph would be an increasing line as the scores get higher, meaning a very low value for a score of 1, and a very high value for a score of 20.



- **“Percentage Donating”**

  The final graph shows the “Percentage Donating.” This is another line graph that has the percentage of people donating in each score bucket. The red dashed line is the overall percentage of people donating for all individuals. Again, with this graph the ideal shape for this would be an increasing one, with a very low percentage for a score of 1, and a high percentage of people donating for the score of 20.