A Simple Score

By Peter B. Wylie

Over the years I've become an avid mystery fan. Though the subject matter of these "dramas" can be gruesome, I find them wonderfully distracting and more than a little entertaining. What I find especially intriguing about them (whether books or TV shows or movies), is the "mileage" detectives can get out of smidgeons of evidence. Maybe it's a partial fingerprint or a tiny piece of tissue that contains just enough DNA to be linked to a suspect. Maybe it's a worn tire or running shoe impression that matches the findings from a search warrant. Whatever. Those little pieces of evidence can shut down violent criminals whose rampages might otherwise persist for decades.

Okay, I'm not a cop – probably a good thing for both society and criminals. But I do see myself as a detective in the data mining and predictive modeling work I do for higher education. I've gotten pretty good at finding little pieces of information in databases that point to which alums are most likely to give and which are not.

What I want to do here is talk about how (with the help of your IT folks) you can combine five basic pieces of information into a score that will help you save money and generate more revenue -- especially if you work with the annual fund.

Specifically, I'll:

- Describe the five pieces of information
- Show how these pieces of info can be combined into a simple score
- Show how this score works for ten different higher ed institutions
- Talk about how you can use the score in your own school

The Five Pieces of Information

The five pieces of information that go into making up this score are pretty straightforward:

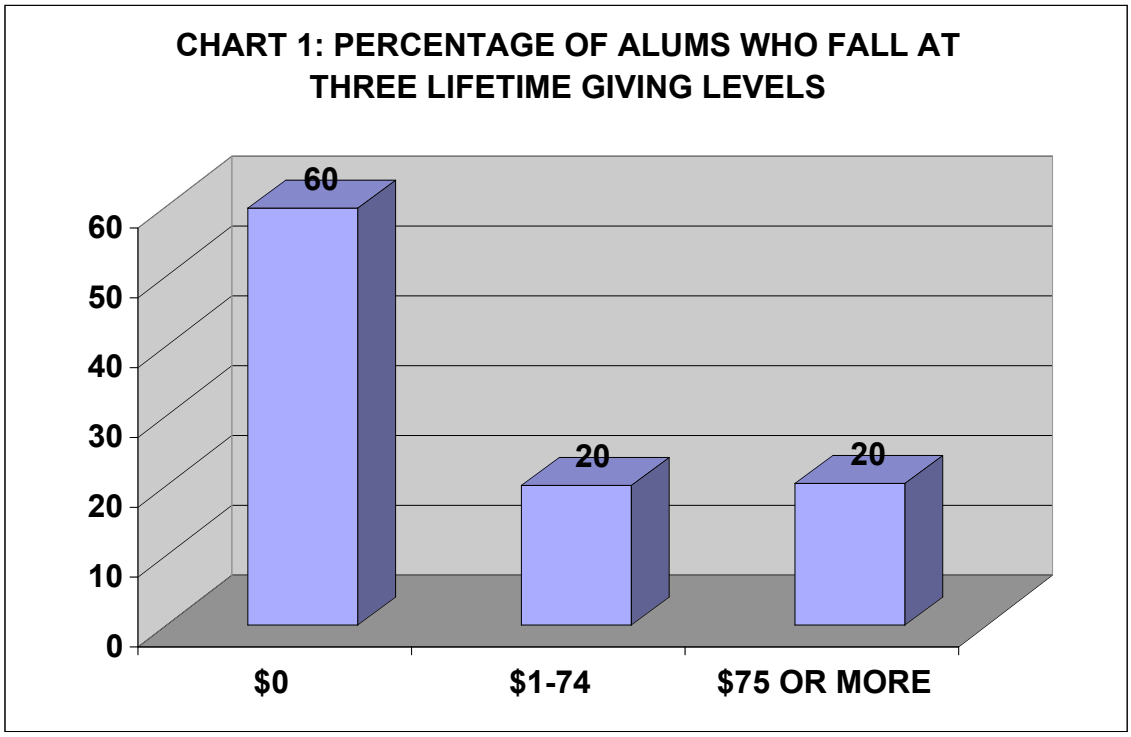
- Whether a **home phone** is listed in the database for a record
- Whether a **business phone** is listed in the database for a record
- Whether an **e-mail address** (either personal or business) is listed in the database for a record
- Whether a record is either in the **oldest or youngest 25% of alums**
- Whether a record is listed as **"missing" or "single" in the marital status field**

Why these pieces of information? What's so special about them? Over the years these things (let's call them variables) began to pop up as I foraged through more and more alumni databases for good predictors of giving.

Let's work through some examples with several of these pieces of information.

Whether a home phone is listed

Take a look at Chart 1. It shows the percentage of alums at a large university who fall at three different levels of lifetime giving: $0; $1-74; and $75 or more. (By the way, if you're surprised at how poor this giving picture looks, take a look at what it is for your own school. You may be in for a bit of a shock.)



Now take a look at Chart 2. It shows the difference in the percentages at each of these three levels of giving for alums who have a home phone listed in the database versus those who don't. Clearly, alums who have a home phone listed are better givers than those who don't. Of those with a home phone listed, 48% have given something to the university (however little). Of those *without* a home phone listed, only 28% have given anything.

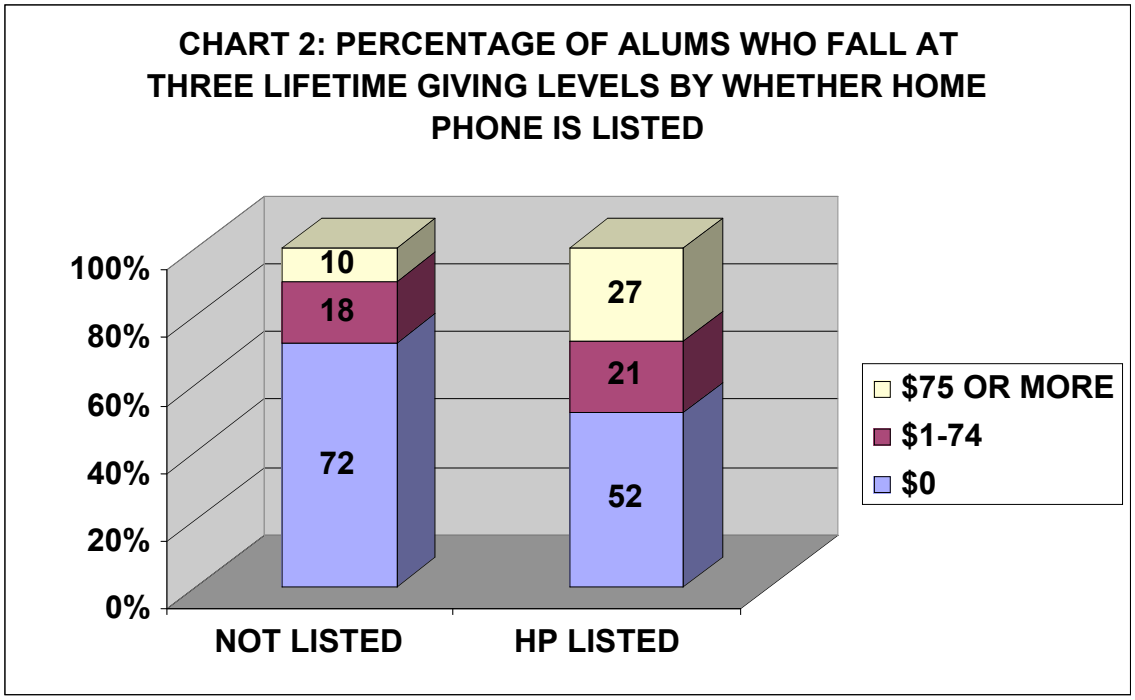
Before moving on to another example, let's deal with a question I often get when I talk at professional meetings on data mining and predictive modeling:

> "Wait a minute. Aren't people with a home phone listed more likely to be givers simply because we can reach them by phone? And also because they're more likely to give us their phone numbers when they make a gift? Aren't we talking about a spurious correlation here?"

It's a good question and I'm always glad when it gets asked. Here's how I respond:

Frankly, we don't know how those numbers got in there. If you were to go to your IT folks and ask them to identify (for each record) how the telephone number got into the database, they'd throw up their hands in despair. The more outspoken ones would say: "We don't know. Plus it would be a huge waste of our time to try and figure it out!" The truth is, all we know is that it's in there or it isn't.

What's important here is that the presence or absence of the phone number is *not* what I would call a "proxy" for giving. For example, a proxy variable would be "MEMBER OF THE $1,000 CLUB."



Alums who are members of this club are givers (and good ones at that). Using their membership as a predictor of giving wouldn't make sense; we already know they're givers. But "home phone listed" is not an automatic measure of giving. There are plenty of alums who have a home phone listed who haven't given a penny. All we know is that alums with a home phone listed are more likely to be givers than alums without a home phone listed. So (and this gets to the nub of my argument) if we're going after new donors (those who haven't yet given us a penny), which alums are our better bets? Those with a home listed or those without a home phone listed?

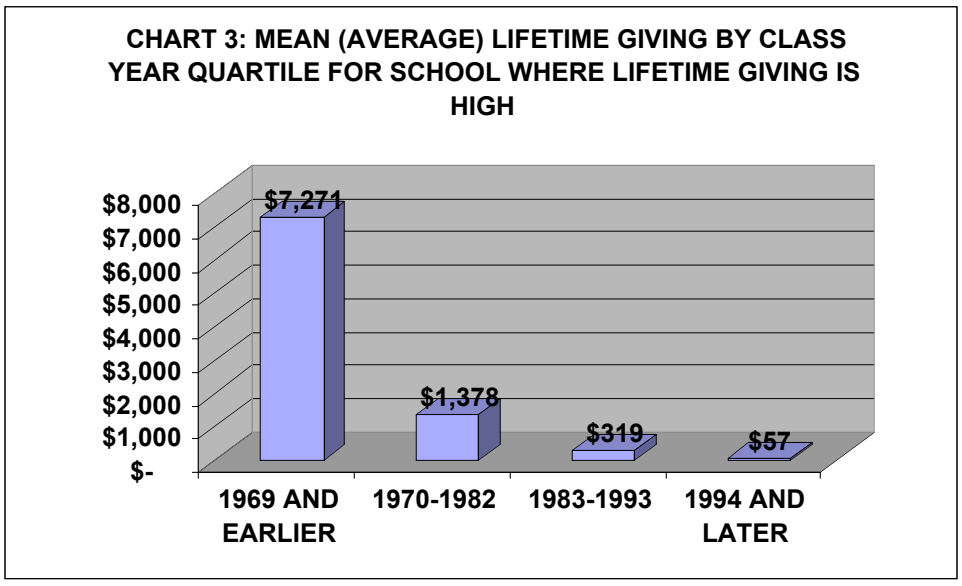
Whether a record is either in the oldest or youngest 25% of alums

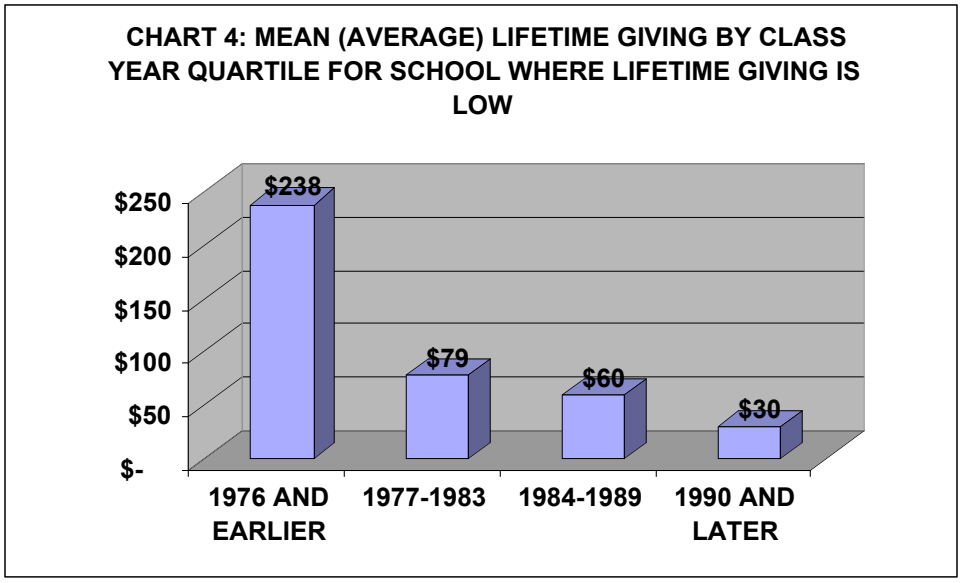
I don't think it's any secret that the longer alums have been out of school the more likely they are to have made a gift to the school and (hopefully) a large one. However, I don't think university advancement folks have a good sense of how disproportionate the giving is between older and younger alums. You can read more about this phenomenon in another paper I've written

(http://www.case.org/files/Bookstore/PDF/Wylie%20White%20Papers/Deep_Pockets_Where_the_Alumni_Money_Is.pdf). But for now let's take a look at this trend in two very different schools, one where the level of lifetime giving is high and one where it is quite low.

As you look at Charts 3 and 4 you'll see that the alums in Chart 3 are giving much more than the alums in Chart 4. This is especially the case when you compare the oldest quartiles (the oldest 25%) at the two schools. The "oldsters" in Chart 3 have given an average of over $7,000 lifetime where as their counterparts in Chart 4 have only given about $240 lifetime.

However, when you compare the lifetime giving levels between the oldest and youngest quartiles, there is quite a difference. Admittedly, the difference between this oldest and youngest group is much greater in Chart 3 than in Chart 4. But even in Chart 4 the oldest alums have given almost eight times as much (on average) as the youngest alums.





If I could show you data from all the schools I've looked at over the years, you'd see that the difference between the oldest and youngest quartiles in Chart 3 is much more typical than the same differences in Chart 4. And it's probably much more typical of your own institution.

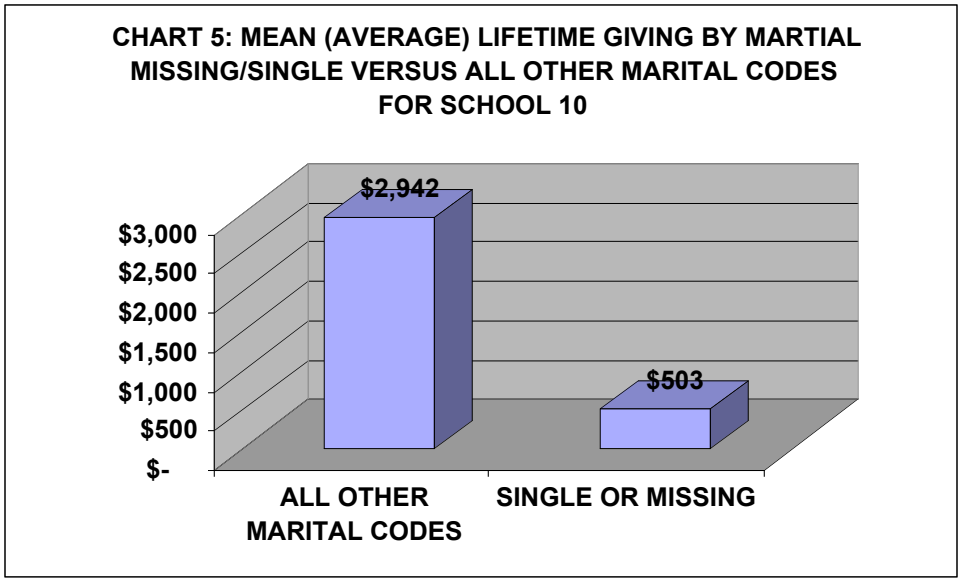
Whether a record is listed as "missing" or "single" in the marital status field

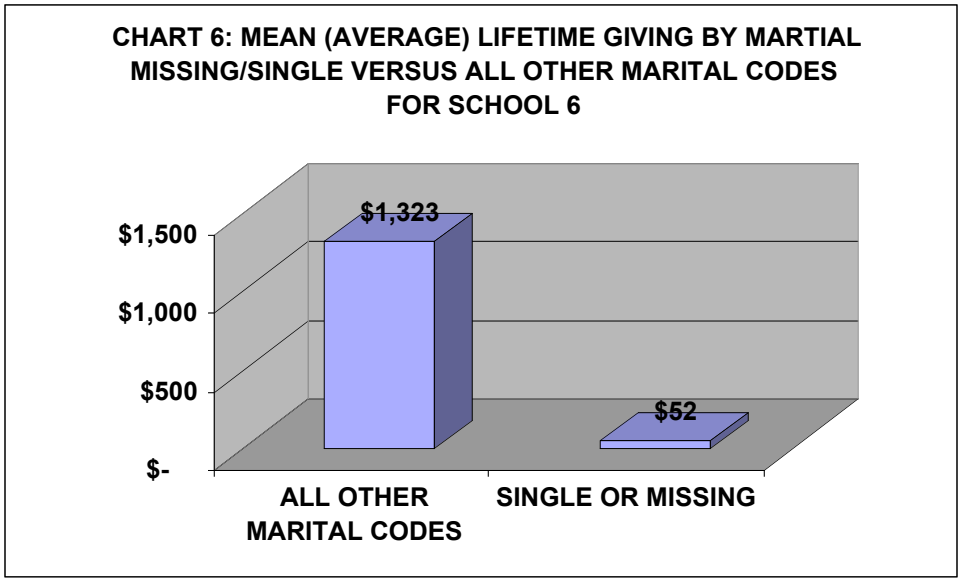
I do a lot of presentations at conferences about data mining and predictive modeling in advancement. At these things people will ask me how I discovered the basic predictors I'm laying out in this paper. No matter how many times I get asked, I hesitate before answering because the truthful answer is: "I'm not exactly sure."

And I'm not. That's the thing about data mining. It's an exploratory process. Sometimes you have a clear idea of what might work as a predictor; sometimes you don't. Sometimes you just stumble onto something. And I think that's what happened with the marital status field. I began to notice that almost all the four-year higher education institutions I worked with had a marital status field.

Then I noticed that for a lot of the records there was simply no code at all in this field. (So I would assign them a code of "missing.") I also noticed that a lot of the records had been assigned a code of "single." When I compared the lifetime giving rate of these "missing" and "single" records with those that had any other code listed, I saw a big difference. The other codes, regardless of what they were ("married," "divorced," "widowed," etc.), gave substantially more than the singles and the missings.

To show how pronounced this difference can be, look at Charts 5 and 6 that show the mean (average) lifetime giving for these two sets of marital codes for two of the ten schools I used for this paper.





How These Pieces of Info Can Be Combined Into a Simple Score

We'll get into the details of this in a moment, but first let's confront a reality you may be considering before I even raise it: getting help from your IT folks. Unfortunately, none of the advancement/development/fundraising software I'm aware of is easy to use when it comes to data analysis. In fact, I think the scale goes from "sort of okay" down to "terrible." This makes the job of your overworked IT people (the ones who haven't already been lured away by higher paying jobs in the private sector) all the more difficult.

So keep all that in mind when you ask for their help on a project like this. Be persistent because they'll tend to "backburner" tasks like this one in favor of all the "urgent" stuff that needs to be done in your shop. But be nice and patient and understanding, too.

Probably the best way to start this project is to ask your IT person to pull a random sample of 10,000 "solicitable" alum records from your database that can be put into an Excel file with the following fields:

- UNIQUE ID NUMBER (the number you use to keep track of all records in your database)

- HOME PHONE LISTED (1 if a home telephone # is listed for the record, 0 if no number is listed)

- BUSINESS PHONE LISTED (1 if a business telephone # is listed for the record, 0 if no number is listed)

- EMAIL LISTED (1 if any e-mail address is listed for the record, 0 if no email address of any kind is listed)

- SINGLE OR MISSING (1 if the record is listed as "single" or has no code at all listed in your marital status field, 0 if the record has any other code listed in your marital status field)

- PREFERRED YEAR OF GRADUATION (this should be a Y2K compliant year for whichever graduating class you associate the record with)

- LIFETIME GIVING (the total number of dollars the record has given to your institution since he or she has been in your database

Once this file has been created, use Excel (or a stats software package if you have one and can use it) to identify the oldest and youngest 25% of alums by preferred year of graduation. Create two more fields in the Excel file: OLDEST 25% and YOUNGEST 25%. In each field assign a 1 to all records who belong to that quartile and a 0 to all the other records.

Okay, here's the basic formula you can use to combine these fields into a score for each record in your Excel file:

SCORE = HOME PHONE LISTED(0/1) + BUSINESS PHONE LISTED(0/1) + EMAIL(home and/or business) LISTED(0/1) + OLDEST GRAD CLASS QUARTILE(0/1) - YOUNGEST GRAD CLASS QUARTILE(0/1) – MARITAL CODE MISSING(0/1) - SINGLE(0/1) + 3 (The constant of 3 is a way to avoid negative and zero scores that can be confusing.)

You're right. It does look a little daunting. But it really isn't. Here's a little data from a school that you can use to see if you understand the formula and how it works. Notice

that I've provided the score for the first two records in this data set but have left the score for the last three records blank.

Let's work through how the score was computed for these first two records. Then you can compute the score for the last three to make sure you understand.

Id	Hp Listed	Bp Listed	Email Listed	1969 And Earlier	1994 And Later	Marital Missing or Unknown	Single	Score
1018	0	0	0	0	1	0	1	1
1019	1	0	1	0	1	0	1	3
1020	0	0	0	0	1	0	0	
1021	1	0	0	0	0	0	0	
1022	1	0	0	1	0	0	0	

We'll start with the first record, Id # 1018. Notice that this alum graduated in the youngest 25% (1994 and later) and is listed as single in the marital status field. So our formula says that alum should get a -1 and another -1 making for a -2. But then we have to add a 3 on to that making for a score of 1.

Now let's do record #1019. This alum has both a home phone listed and an e-mail listed making for a +2, but the alum also is in the youngest class quartile and is listed as single making for a -2. So the +2 and the -2 cancel to 0 and then we add 3 to get, yep, a score of 3 for that person.

Now see if you can come up with a score for record #'s 1020, 1021, and 1022. Then compare your answers with the data below:

Id	Hp Listed	Bp Listed	Email Listed	1969 And Earlier	1994 And Later	Marital Missing or Unknown	Single	Score
1018	0	0	0	0	1	0	1	1
1019	1	0	1	0	1	0	1	3
1020	0	0	0	0	1	0	0	2
1021	1	0	0	0	0	0	0	4
1022	1	0	0	1	0	0	0	5

Once you understand how the score works, you can:

1. Add these 0/1 variables together using the formula to create a score for each of the 10,000 records in your file.
2. Compute the mean (average) lifetime dollars received from alums at each score level.

If you need a little assistance to do these two steps, some math or stats student or professor in your school should be able to help you.

How This Score Works For Ten Different Higher Ed Institutions

Earlier I mentioned that over the years I've foraged through a lot of alumni databases for good predictors of giving. In this section I've laid out how the score you've just been looking at works for 10 schools that are very representative of the 100+ institutions I've studied. In a moment I'd like you to browse through the data to see how remarkably well the score works for these schools -- schools that differ greatly in terms of geographic location, size, money raised, whether they are public or private, and on and on. But first let's go over the data for School 1 to make sure you understand what I've done in these tables. Going from left to right, here's a description of what's in each column of the table:

- **Score**. This column simply lists the scores I computed for the school. Here the scores go from 1 to 7.

- **Count**. The next column displays the exact number of records for each score level in the random sample of alumni records the school gave me.

- **Sum Total Giving**. This column shows the sum of lifetime giving for records at each score level. For example, for this school there are 1,069 alums at score level 1 who have given a combined lifetime total of $3,291 to the school. At score level 7 there are 1,483 alums who have given a combined lifetime total of $23,686,400 to the school. (No, you didn't misread anything. This is a good example of how powerful this simple little score can be.)

- **Mean Total Giving**. This column shows the mean (the arithmetic average) of lifetime giving for records at each score level. I arrived at that number by dividing the sum total giving by the count. For example, for this school there are 1,069 alums at score level 1 who have given a combined lifetime total of $3,291 to the school. Their mean level of lifetime giving is three dollars. At score level 7 there are 1,483 alums who have given a combined lifetime total of $23,686,400 to the school. Their mean level of lifetime giving is $15,972.

- **Median Total Giving**. This column shows the median of lifetime giving for records at each score level. A couple of things about the median. One, in every single instance throughout these ten tables the median is less than the mean, usually a *lot* less. That's because the median (the amount above and below which half the alums at a given score level fall in terms of lifetime giving) is not affected by large donors who can pull the mean way up. Notice, for example, that at score level 7 the mean is $15,972 but the median is $261. That's because there is one alum at this core level who has given almost ten million dollars to the school. The second thing about the median is that, for some score levels, there is a blank amount listed. That simply means that over half of the alums at that score level have given nothing to the school.

- **Maximum Lifetime Amount**. This column shows the largest lifetime amount an alum at each score level has given the school.

SCHOOL 1

SCORE = 'HP LISTED' + 'BP LISTED' + 'EMAIL LSTED' + '1969 AND EARLIER' - '1994 AND LATER' - 'SINGLE' - 'MARITAL MISSING OR UNKNOWN' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	1069	$ 3,291	$ 3	$ -	$ 425
2	5974	$ 2,907,250	$ 487	$ -	$ 2,724,200
3	10068	$ 1,941,510	$ 193	$ -	$ 345,570
4	12290	$ 7,657,680	$ 623	$ -	$ 1,061,890
5	13116	$ 24,780,600	$ 1,889	$ 27	$ 903,859
6	7501	$ 56,208,800	$ 7,494	$ 113	$ 14,526,200
7	1483	$ 23,686,400	$ 15,972	$ 261	$ 9,605,500

Now go ahead and browse through the tables for the remaining nine schools. (You'll see the exact formula I used for each school. The score is pretty much the same for all the schools. The only differences you'll notice are the years I used to define the oldest and youngest grad year quartiles and the way I defined the missing marital codes.)

SCHOOL 2

SCORE = 'HP LISTED' + 'BP LISTED' + 'HOME EMAIL LISTED' + 'WORK EMAIL LISTED' + '1962 AND EARLIER' - 'SINGLE' - 'MARITAL STATUS UNKNOWN' - '1991 OR LATER' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	646	$ 20,931	$ 32	$ -	$ 5,770
2	1684	$ 140,937	$ 84	$ -	$ 13,700
3	2428	$ 855,777	$ 352	$ 20	$ 151,065
4	3061	$ 8,960,600	$ 2,927	$ 160	$ 1,958,140
5	3872	$ 22,911,000	$ 5,917	$ 570	$ 1,572,620
6	2296	$ 23,738,800	$ 10,339	$ 1,015	$ 2,020,940
7	520	$ 5,978,130	$ 11,496	$ 1,350	$ 1,284,210
8	34	$ 902,304	$ 26,538	$ 3,733	$ 481,550

SCHOOL 3

SCORE = 'HOME PHONE LISTED' + 'BUS PH LISTED' + 'HOME EMAIL LISTED' + 'BUS EMAIL LISTED' + '1976 AND EARLIER' - '1990 AND LATER' - 'MARITAL STATUS MISSING' - 'MARITAL STATUS UNKNOWN' - 'SINGLE' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	2053	$ 21,314	$ 10	$ -	$ 760
2	5981	$ 154,049	$ 26	$ -	$ 21,500
3	5701	$ 243,130	$ 43	$ -	$ 4,238
4	3555	$ 367,552	$ 103	$ -	$ 27,258
5	1302	$ 250,948	$ 193	$ -	$ 18,250
6	416	$ 318,775	$ 766	$ 20	$ 199,240
7	152	$ 505,127	$ 3,323	$ 43	$ 365,525
8	12	$ 46,665	$ 3,889	$ 163	$ 42,100

SCHOOL 4

SCORE = 'HAS EMAIL' + 'HOME PHONE LISTED' + 'BUS PHONE LISTED' + '1974 AND EARLIER' - '1993 AND LATER' - 'MARITAL CODE MISSING' - 'SINGLE' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	114	$ 706	$ 6	$ -	$ 235
2	1071	$ 6,466	$ 6	$ -	$ 1,000
3	2682	$ 39,191	$ 15	$ -	$ 2,035
4	2982	$ 91,911	$ 31	$ -	$ 4,670
5	2118	$ 188,079	$ 89	$ -	$ 20,600
6	653	$ 1,408,500	$ 2,157	$ -	$ 1,211,510
7	77	$ 67,004	$ 870	$ 50	$ 41,210

SCHOOL 5

SCORE = 'HP LISTED' + 'BP LISTED' + 'EMAIL LISTED' - 'LISTED AS MARITAL MISSING' - 'LISTED AS SINGLE' + '1916-1966' - '1993 AND LATER' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	447	$ 2,588	$ 6	$ -	$ 352
2	1451	$ 51,989	$ 36	$ -	$ 5,215
3	2247	$ 241,718	$ 108	$ 3	$ 16,340
4	2472	$ 1,387,380	$ 561	$ 92	$ 79,992
5	2757	$ 6,820,120	$ 2,474	$ 322	$ 737,724
6	2169	$ 10,745,900	$ 4,954	$ 610	$ 4,250,050
7	273	$ 1,895,070	$ 6,942	$ 1,155	$ 341,868

SCHOOL 6

SCORE = 'HP LISTED' + 'EMP PH LISTED' + 'EMAIL LISTED' - 'MARITAL MISSING' - 'SINGLE' + '1980 AND EARLIER' - '1998 AND LATER' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	600	$ 3,405	$ 6	$ -	$ 425
2	1957	$ 19,265	$ 10	$ -	$ 3,000
3	2108	$ 103,468	$ 49	$ -	$ 25,973
4	1994	$ 200,829	$ 101	$ -	$ 16,505
5	1931	$ 645,547	$ 334	$ 12	$ 31,975
6	1112	$ 4,680,200	$ 4,209	$ 70	$ 2,180,360
7	298	$ 1,461,470	$ 4,904	$ 295	$ 465,415

SCHOOL 7

SCORE = 'HP LISTED' + 'BP LISTED' + 'EMAIL LISTED' + '1968 OR EARLIER' - '1994 OR LATER' - 'MISSING MARITAL' - 'SINGLE' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	664	$ 3,577	$ 5	$ -	$ 1,102
2	2125	$ 95,016	$ 45	$ -	$ 32,500
3	1975	$ 1,796,250	$ 909	$ -	$ 633,651
4	917	$ 1,424,000	$ 1,553	$ 10	$ 765,804
5	582	$ 1,982,870	$ 3,407	$ 100	$ 951,381
6	167	$ 901,767	$ 5,400	$ 250	$ 135,485
7	15	$ 118,126	$ 7,875	$ 100	$ 95,969

SCHOOL 8

SCORE = 'HP LISTED' + 'BP LISTED' + 'EMAIL LISTED' +'1923-1967' - '1994-2002' - 'SINGLE' - 'MARITAL MISSING' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	144	$ 370	$ 3	$ -	$ 255
2	912	$ 5,686	$ 6	$ -	$ 339
3	1704	$ 97,005	$ 57	$ -	$ 7,210
4	2580	$ 1,509,950	$ 585	$ 25	$ 519,932
5	3198	$ 7,639,950	$ 2,389	$ 115	$ 3,382,280
6	1222	$ 2,103,510	$ 1,721	$ 241	$ 563,206
7	156	$ 256,614	$ 1,645	$ 468	$ 31,901

SCHOOL 9

SCORE = 'HP LISTED' + 'BP LISTED' + 'EMAIL LISTED' - 'SINGLE' - 'MARITAL UNKNOWN' + '1974 AND EARLIER'- '1995 AND LATER' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	522	$ 537	$ 1	$ -	$ 125
2	1766	$ 11,392	$ 6	$ -	$ 3,535
3	1973	$ 33,522	$ 17	$ -	$ 3,500
4	2489	$ 117,030	$ 47	$ -	$ 3,465
5	2357	$ 358,932	$ 152	$ 5	$ 35,096
6	851	$ 325,813	$ 383	$ 35	$ 64,540
7	42	$ 35,559	$ 847	$ 193	$ 11,830

SCHOOL 10

SCORE = 'HP LISTED' + 'BP LISTED' + 'EMAIL LISTED' + '1924-1964' - '1990-2004' - 'SINGLE' - 'MARITAL MISSING' + 3

SCORE	COUNT	SUM TOTAL GIVING	MEAN TOTAL GIVING	MEDIAN TOTAL GIVING	MAXIMUM LIFETIME AMOUNT
1	81	$ 1,655	$ 20	$ -	$ 270
2	256	$ 5,201	$ 20	$ -	$ 500
3	325	$ 44,825	$ 138	$ -	$ 6,388
4	422	$ 597,375	$ 1,416	$ 45	$ 318,025
5	612	$ 2,066,430	$ 3,377	$ 250	$ 357,419
6	544	$ 2,302,410	$ 4,232	$ 475	$ 524,125
7	65	$ 344,355	$ 5,298	$ 760	$ 106,730

Over the last several months I've shown these charts to lots of people either individually or in group presentations. The reactions I get seem to fall into two camps:

1. "Whoa!! That's really something. This is good stuff!"

2. "Yeah, I guess I'm not that surprised. When you add all that stuff together it makes sense that the people with the higher scores would be giving more."

If you're in the first group, great! I've made the sale and you don't need more convincing. If you're in the second group, the last thing you need from me is more exhortation. Maybe the thing to do is just think about it awhile and see what happens. Maybe you'll change your mind; maybe you won't. Another thing you can do is ask someone else (who hasn't been exposed to any of the stuff I said or written about data mining) to take a look at what I've laid out here. Their "take" on it may be thought provoking and helpful.

But for now I'll assume you're convinced this little score has some potential. I'll even assume you're willing to try it out at your own institution to see how it works and whether (limited as it is) it might help you quickly save money and make more on annual fund appeals.

If you are willing to try it out, here's what I'd recommend:

1. Pick a sample of alums whom you plan to appeal to by phone, snail mail, or email over the next several months.

2. With the help of your IT folks, compute scores for each alum you'll be appealing to and get those scores into your database.

3. Go ahead and appeal to these alums as if the scores didn't exist. That is, don't make any decisions about segmenting these alums based on the scores. Just do what you would normally do as if you'd never read this paper.

4. Once the appeal has ended, do some analysis by score level of what you received. For example, for each score level, compute the percentage of alums who gave anything at all to the appeal by check or credit card. Compute the percentage of alums who made a pledge. Compute the mean and median dollars given by check or credit card. Compute the median and median dollars pledged.

If you do all that, you'll learn a huge amount even if the results you get aren't as stellar as I expect they'll be. My expectations are much less important than your willingness to take on the role of applied scientist here. To test ideas, theories, and just plain hunches that may help you do a better job of achieving your institution's advancement mission. That's what we need more of in this business. A lot more of.