For office use only	Team Control Number	For office use only
T1 ________________	**42642**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**C**	F4 ________________

2016

MCM/ICM

Summary Sheet

(Your team's summary should be included as the first page of your electronic submission.)

Type a summary of your results on this page. Do not include the name of your school, advisor, or team members on this page.

In this essay, we intend to develop a strategy for the Greatgrant Foundation to invest 100,000,000 dollars in some colleges. In order to assess the candidate school's potential, we take three main aspects into consideration and build up **the Principal Component Analysis Model (PCA)** to conform the weighting coefficient of all the indicators. The three main initial elements considered are:

- the student resources (mainly be measured by the academic achievements of the students)
- the future development of the undergraduates (lying on the mean earnings of the graduates)
- the financial situations and policies of the school (showed by the standard deviation of the average net price for different family incomes).

After the calculations, we give a score to each candidate school to measure the potential and give out a prioritized ranking list. For the reason that, if we donate to a school, the return-on-investment (ROI) is related to the score of the school, we select the top 50 institutions on the list and decide to invest on them.

Considering the fact that these schools are highly different, it is inappropriate to invest the same amount money in them. Therefore, we build up **the Cluster Analysis Model** to divide them into several classes. The classification is due to the behavior in the three aspects of the school. And to find out the distinct characteristics of each class will help us draw up different investment plans for them. For example, the majors in the schools of the last class are mainly about medical care. As a result, we plan to invest a bit more money in these institutions than the average to encourage more students to choose these schools and work in medical industry after graduation. Furthermore, the enrollment of the undergraduates is another determined element. As to maximize the return-on-investment (ROI), we should take a full consideration of it, when we calculate the amount of the investment of each school.

When it comes to the time duration, we suggest that the foundation make a new ranking list every year and adjust the investment strategy according to the list. It can also stimulate candidate schools and their undergraduates to perform better.

We test the **stability** of our model and find that the coefficients are all stable to a great extent. There is no big difference between the original result and the altered one. Thus we claim that our **Principle Component Analysis Model** is considered to be stable.

The Investment Strategy of the Goodgrant Private Funding: Principle Component Analysis and Cluster Analysis Approach

Contents

1 Introduction **3**

1.1 Problem background 3

1.2 Our work 3

2 Nomenclatures **4**

3 Assumptions **5**

4 Data Processing **5**

4.1 Collect data 5

4.2 Preprocess data 6

5 Two Models for the Investment Strategy **8**

5.1 Principle Component Analysis (PCA) 8

5.2 Cluster analysis 11

6 The Optimal Investment Strategy **14**

6.1 Adjustment and the allocation of the funds 14

6.2 Return On Investment (ROI) 16

6.3 Time duration 17

7 Other Solutions and Future Work 17

7.1 Other solutions and discussions 17

7.2 Future work 18

8 Error Analysis and Stability Analysis 19

8.1 The error analysis 19

8.2 The stability analysis 19

9 Strengths and Weaknesses 20

9.1 Strengths 20

9.2 Weaknesses 20

Reference 21

10 A Letter to Mr. A. Chiang 22

1 Introduction

1.1 Problem background

The Goodgrant Foundation intends to donate a total of \$100,000,000 to an appropriate group of schools per year for five years to help improve educational performance of these undergraduates attending colleges and universities in the United States.

In order to produce a strong positive effect on student performance, the Goodgrant Foundation needs an optimal investment strategy with a 1 to N optimized and prioritized candidate list of schools, which is determined based on each candidate school's demonstrated potential for effective use of private funding, and an estimated return on investment (ROI).

1.2 Our work

To build up a model to attach the optimal investment strategy, we face mainly five problems:

- Choose meaningful and defendable data which should be considered in this model.
- Build up models to assess the potential for these candidate schools and the return on investment (ROI) of the charitable organization.
- Provide an optimized and prioritized candidate list of schools to recommend and adjust the model we've constructed.
- Do the error analysis and stability analysis of some important parameters.
- Make an optimal investment strategy including some details.

So as to assess the potential of these schools, we take three main aspects into consideration [1].

The first one is about the academic achievements of these undergraduates admitted, because we tend to invest in the schools with better student resources. In fact, students with better academic achievements usually have higher talents and work harder in the university. More scholarships can stimulate them to perform better.

The second one is about the future development of these students, mainly about how much the graduates earn in the future of these schools. To a certain extent, the predicted future salaries of the graduates indicates their professionalism, which is closely related

[1] Actually, the rationality why we consider these 3 aspects will be discussed quantitatively, according to the principle analysis.

to the professional education level of the schools. Furthermore, the salaries also reflect the social contribution of the graduates. For all the reasons above, we are inclined to invest in the schools whose graduates have preferable future development.

The third one is about the financial situation of these schools. As a matter of fact, schools with complete financial policies can use the investment more reasonable. That is to say, if we donate our money to these schools with fair financial policies, the donation can yield greater returns on investment.

Furthermore, It is a fact that some undergraduates who come from poor families find it difficult to afford the pay in the institutions of higher learning. As a result, if these people had a chance to get more scholarship, they will work harder and put more effort on academic learning instead of other things such as doing part-time jobs. Different schools have different policies about the tuition fees of students from different family social classes. If we tend to invest more money in the schools whose financial policies are in favor of poor students, it is more likely that our investment be given to these students, and thus our return on the investment (ROI) will be higher [1].

2 Nomenclatures

NPI1	The average net price for $0-$30,000 family income
NPI2	The average net price for $30,001-$48,000 family income
NPI3	The average net price for $48,001-$75,000 family income
NPI4	The average net price for $75,001-$110,000 family income
NPI5	The average net price for $110,000+ family income
STDEVP	The standard deviation of the average net price for different family income
SATVRMID	The midpoint of SAT scores at the institution (critical reading)
SATMTMID	The midpoint of SAT scores at the institution (math)
SATWRMID	The midpoint of SAT scores at the institution (writing)
ACTENMID	The midpoint of the ACT English score
ACTMTMID	The midpoint of the ACT math score
ACTCMMID	The midpoint of the ACT cumulative score
SAT_AVG	The average SAT equivalent score of students admitted
md_earn_p10	The mean earnings of students working and not enrolled 10 years after entry
gt_25k_p6	The share of students earning over $25,000/year (threshold earnings) 6 years after entry

Table 1: Nomenclatures

[1] The claim will be confirmed in the 5th part: Models of our essay.

3 Assumptions

- The SAT scores and the ACT scores can indicate the talent and the academic abilities of the undergraduates.
- The mean earnings of graduates can show the professional education level of the schools as well as their social contributions.
- The future development of these freshmen can be predicted by the mean salaries of the graduates investigated in that year.

4 Data Processing

4.1 Collect data

First, we collect two kinds of data to be calculated in our models in order to assess the potential of these candidate schools.

One kind of these data is about the achievement scores of the undergraduates, because the achievement scores always indicate the talent of these students and how hardworking they are. We can also see the ranking of these schools from it roughly. To avoid leaving out any important elements, we choose as much related factors as possible, such as *The average SAT equivalent score of students admitted (SAT_AVG)*, *The midpoint of SAT scores at the institution (critical reading)*, *The midpoint of SAT scores at the institution (math)*, *The midpoint of SAT scores at the institution (writing)*, *The midpoint of the ACT cumulative score (ACTCMMID)*, *The midpoint of the ACT math score* and *The midpoint of the ACT English score* [1].

Another kind of these data is about the future development of the undergraduates of these candidate schools. In the paper, we choose *The mean earnings of students working and not enrolled 10 years after entry* (mn_earn_p10) and *The share of students earning over \$25,000/year (threshold earnings) 6 years after entry* (gt_25k_p6) [2] to estimate the future development of these undergraduates. These data help us to know the potential of the undergraduates.

[1] We are not going to deal with *The midpoint of the ACT writing score*, because this part is not necessary when applying for universities. Actually, there are 7505 missing values among the whole sample of 7805 schools, thus we will not consider it.

[2] We are not going to deal with *150% completion rate for less-than-four-year institutions, pooled in two-year rolling averages and suppressed for small n size* and *200% completion rate for less-than-four-year institutions, pooled in two-year rolling averages and suppressed for small n size*, because these 2 variables are considered to be related to the net price of institution, which is what we don't want to take into consideration in this part.

Besides, it is a fact that some undergraduates who come from poor families find it difficult to afford the pay in the institutions of higher learning. As a result, if these people had a chance to get more scholarship, they will work harder and they can put more effort on academic learning instead of doing part-time jobs. That is to say, for the purpose of helping improve educational performance and making the return on investment (ROI) highest, we should take the financial situation of these universities and colleges into account as an important parameter.

We choose *The average net price for $0-$30,000 family income* $(NPI1)$, *The average net price for $30,001-$48,000 family income* $(NPI2)$, *The average net price for $48,001-$75,000 family income* $(NPI3)$, *The average net price for $75,001-$110,000 family income* $(NPI4)$, and *The average net price for $110,000+ family income* $(NPI5)$. Through the preprocessing of these data, we can find whether these schools have some policies to help these poor families and the level of their financial aid. All these results will help us to make the optimal investment strategy.

In order to reduce the risk of our investment, *Whether the schools are on Heightened Cash Monitoring 2 by the Department of Education* is another factor that should be considered, which indicates the financial sustainability of these schools. However, we have inspected whether these candidate schools are on Heightened Cash Monitoring 2 (HCM2) by the Department of Education, and we find that all these schools with an average SAT equivalent score of students admitted exceeding 1000 are not on Heightened Cash Monitoring 2 (HCM2), which means these schools are not troubled with any financial problems.

Moreover, taking the types of these schools into consideration, *The enrollment of undergraduate degree-seeking students* also helps determine the investment amount per school.

4.2 Preprocess data

4.2.1 Initial screening according to the average SAT equivalent score

To make sure that the academic performance of all the schools that have been selected to invest is above the average level, we only consider the schools whose average SAT equivalent score of students admitted is higher than 1000 (including 1000) [1]. We find that only 930 schools of the potential candidate schools meet the requirements.

4.2.2 Standard deviation of the average net price for different family income

In this paper, we need to know how these schools' tuition policies help these undergraduates from poor families, so that we can invest ones without appropriate tuition

[1]According to the given data(a sample of 7804 colleges, with 2520 data), the mean of SAT_AVG is 1056.6894, while the median is 932.5, indicating that the standard is reasonable.

policies to help these poor people to perform better in their universities or colleges.

We calculate the standard deviation of the average net price for different family income of these schools to do some deviation analysis. We get the data of the average net price for families whose income are $0-$30,000, $30,001-$48,000, $48,001-$75,000, $75,001-$110,000 and $110,000+ from the U.S. National Center on Education Statistics (www.nces.ed.gov/ipeds). Then we have [1]

$$STDEVP = \left(\sum_{1 \le k \le 5} NPIk^2 - \frac{1}{5} \left(\sum_{1 \le k \le 5} NPIk \right)^2 \right)^{1/2}.$$

4.2.3 The measurement of the similarity of the variable quantities

Considering the fact that some variables may have a high relativity, which will influence the result of the model, we measure the reliability of these variables. According to the reliability analysis, the Cronbach's Alpha $\alpha = .856$, illustrating that the internal equity level is high.

In addition, we calculate the correlation coefficient of these variables. The consequence is showed in the following chart (Table 2, page 8). For all the following correlation coefficients, $p < .001$.

	$SATV-RMID$	$SATM-TMID$	$SATW-RMID$	$ACTM-TMID$	$ACTE-NMID$	$ACTC-MMID$	SAT_-AVG
$SATVR-MID$	1.000	.912	.973	.856	.900	.916	.946
$SATMT-MID$	.912	1.000	.939	.921	.860	.917	.951
$SATWR-MID$	.973	.939	1.000	.884	.915	.929	.955
$ACTMT-MID$	.856	.921	.884	1.000	.896	.937	.946
$ACTEN-MID$	.900	.860	.915	.896	1.000	.946	.943
$ACTCM-MID$	.916	.917	.929	.937	.946	1.000	.982
SAT_AVG	.946	.951	.955	.946	.943	.982	1.000

Table 2: Correlation Matrix of SAT and ACT Scores

Therefore, We can point out the truth that the total scores can indicate the single-subject ones, so in the following models, we will only take *the average SAT equivalent*

[1] We do not adopt the one weighed by the number of students from different families. What we want to measure is the strength of financial assistance that the school has given to students from families of lower income level instead of the amount on average (related with the ratio of students). Thus the value should not be influenced by the ratio of students.

score of students admitted (SAT_AVG) and *the midpoint of the ACT cumulative score (ACTCMMID)* [2].

5 Two Models for the Investment Strategy

5.1 Principle Component Analysis (PCA)

After the analysis above, we can find that there are three main aspects need to be considered.They are:

- the student resources: depending on *The average SAT equivalent score of students admitted (SAT_AVG)* and *The midpoint of the ACT cumulative score (ACTCMMID).*
- the future development prediction of the undergraduates: lying on *The mean earnings of students 10 years after entry* (md_earn_p10) and *The shares of students earning over $25,000/year (threshold earnings) 6 years after entry* (gt_25k_p6).
- the financial situations and policies of the school: showed by *The standard deviation of the average net prices for different family income (STDEVP).*

When it comes to obtaining the weight of these three aspects, subjective judgment is inexact. So we choose the Principal Component Analysis (PCA) as the way to conform the weighting coefficient of all the indicators in the evaluation system. The significance of KMO and Bartlett's test of sphericity is $p < .001$, showing that factor analysis is considered to be proper.

5.1.1 The screening of the components

By calculation, we obtain all the principal components as the chart shows. Considering the correlation matrix $\mathbf{R}$ of the given variables, the equation

$$\det(\lambda \mathbf{I} - \mathbf{R}) = 0$$

leads to the conclusion that there are initially totally 5 components, whose eigenvalues are 3.485, 1.011, 0.328, 0.159 and 0.018. And their weighting coefficients of the total sum of square are 69.701%, 20.213%, 6.557%, 3.179% and 0.351% respectively. The total weight of the first three components is 96.470%, more than 90%, so we choose the first three components as the principal components (Table 3, page 9) [1].

[2]It is common to select these two variables for further analysis, since they can measure the students' comprehensive academic ability.

[1]The result convinces us that the qualitative analysis in introduction is reliable.

Total Variance Explained						
	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.485	69.701	69.701	2.016	40.329	40.329
2	1.011	20.213	89.914	1.791	35.816	76.145
3	.328	6.557	96.470	1.016	20.325	96.470
4	.159	3.179	99.649			
5	.018	.351	100.000			

Table 3: The Solution of PCA

Rotated Component Matrix			
Component	1	2	3
$ACTCMMID$	.915	.242	.309
SAT_AVG	.908	.250	.322
$STDEVP$	.465	.202	.860
gt_25k_p6	.136	.957	.069
md_earn_p10	.348	.844	.269

Table 4: Rotated Matrix of PCA

Furthermore, we get the rotated solution shown above with the varimax orthogonal rotation method, and the weighing coefficients are 40.329%, 35.816%, 20.325%. Now we can get the weighing coefficient of the three main component, and the rotated matrix is shown as the chart shows (Table 4, page 9).

According to the coefficients of the rotated matrix, denote the three components as $Score$, $Income$ and $Fund$. Notice that the components are all standardized by the formula

$$z_X = \frac{X - \bar{X}}{s_X},$$

Now we can define the new variable $Level$ with

$$Level = 0.40329Score + 0.35816Income + 0.20325Fund.$$

The ranking of the candidate schools can thus be computed according to the variable $Level$ (considered to be the level of the candidate schools), we claim that this variable can identify our return on investment (ROI).

5.1.2 Result

And the Principal Component Analysis (PCA) provides us with an ordered list of the candidate schools. We decide to invest in the top 50 schools. They are showed in the following table. These schools are optimized and prioritized on the basis of the schools potential for effective use of private funding. And the estimated return on investment (ROI) has a positive correlation with the demonstrated potential (Table 5, page 10).

The sort of the candidate schools according to the PCA			
1	Massachusetts Institute of Technology	26	Lehigh University
2	Harvard University	27	Stevens Institute of Technology
3	Stanford University	28	Rice University
4	Harvey Mudd College	29	Claremont McKenna College
5	Duke University	30	Brown University
6	University of Pennsylvania	31	University of Southern California
7	California Institute of Technology	32	Villanova University
8	Georgetown University	33	Babson College
9	Yale University	34	Case Western Reserve University
10	Columbia University in the City of New York	35	Rose-Hulman Institute of Technology
11	University of Notre Dame	36	Northeastern University
12	Princeton University	37	Georgia Institute of Technology-Main Campus
13	University of Chicago	38	Bucknell University
14	Dartmouth University	39	Bowdoin College
15	Tufts University	40	MCPHS University
16	Cornell University	41	Middlebury College
17	Johns Hopkins University	42	Amherst College
18	Washington and Lee University	43	Emory University
19	Washington University in Saint Louis	44	Wellesley College
20	Northwestern University	45	Colgate University
21	Carnegie Mellon University	46	Williams College
22	Boston College	47	Bentley University
23	Vanderbilt University	48	Colorado School of Mines
24	Albany College of Pharmacy and Health Sciences	49	Lafayette College
25	Rensselaer Polytechnic Institute	50	Davidson College

Table 5: The Sort of the Candidate Schools According to PCA

5.2 Cluster analysis

As the selected schools maybe highly different to some extent, we should divide them into several categories and make diverse investment plan for each category. To insure the rationality of the classifying and to maximize the return on the investment (ROI), we build up a model and use the cluster analysis to classify the selected 50 schools.

5.2.1 The distance formula

In this model, we use the Euclid distance and the group average method to process the data. The Euclid distance is defined by the following formula:

$$d(x,y) = \left(\sum_{1 \le k \le p} (x_k - y_k)^2 \right)^{1/2},$$

The group average distance is defined by the formula (where we denote $|G_1| = n_1, |G_2| = n_2$):

$$D(G_1, G_2) = \frac{\sum_{x_i \in G_1} \left(\sum_{x_j \in G_2} d(x_i, x_j) \right)}{n_1 n_2}.$$

The cases with a less distance are clustered together as a class.

5.2.2 The dendrogram and the result

The result of the cluster analysis can be showed by the following dendrogram (Figure 1, page 11; Table 6, page 12).

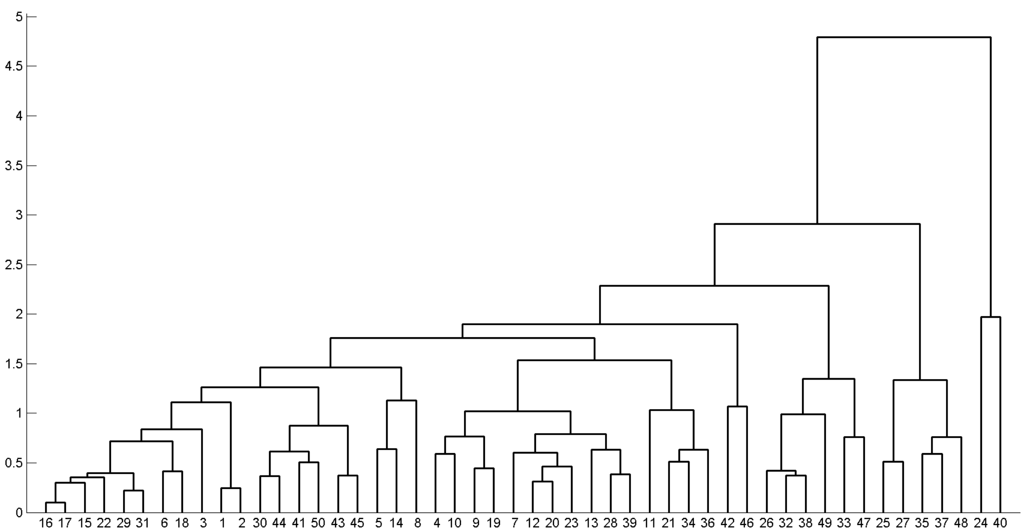

Figure 1: The Dendrogram

The sort of the candidate schools according to the PCA			
1	Massachusetts Institute of Technology	1	Brown University
1	Harvard University	1	University of Southern California
1	Stanford University	1	Northeastern University
1	Harvey Mudd College	1	Northeastern University
1	Duke University	1	Bowdoin College
1	University of Pennsylvania	1	Middlebury College
1	California Institute of Technology	1	Amherst College
1	Georgetown University	1	Emory University
1	Yale University	1	Wellesley College
1	Columbia University in the City of New York	1	Colgate University
1	University of Notre Dame	1	Williams College
1	Princeton University	1	Davidson College
1	University of Chicago	2	Lehigh University
1	Dartmouth University	2	Villanova University
1	Tufts University	2	Babson College
1	Cornell University	2	Bucknell University
1	Johns Hopkins University	2	Bentley University
1	Washington and Lee University	2	Lafayette College
1	Washington University in Saint Louis	3	Rensselaer Polytechnic Institute
1	Northwestern University	3	Stevens Institute of Technology
1	Carnegie Mellon University	3	Rose-Hulman Institute of Technology
1	Boston College	3	Georgia Institute of Technology-Main Campus
1	Vanderbilt University	3	Colorado School of Mines
1	Rice University	4	MCPHS University
1	Claremont McKenna College	4	Albany College of Pharmacy and Health Sciences

Table 6: The Sort of the Candidate Schools According to Clustering

5.2.3 Analysis

According to the cluster analysis, we have divided the 50 selected schools into four categories. It can be seen from the following pie chart that 37 institutions belong to the first class, which account for 74%. And there are 6 institutions in the second class, accounting for 12%, 5 institutions in the third class, accounting for 10%, and 2 institutions in the fourth class, accounting for 4% (figure 2, page 13).

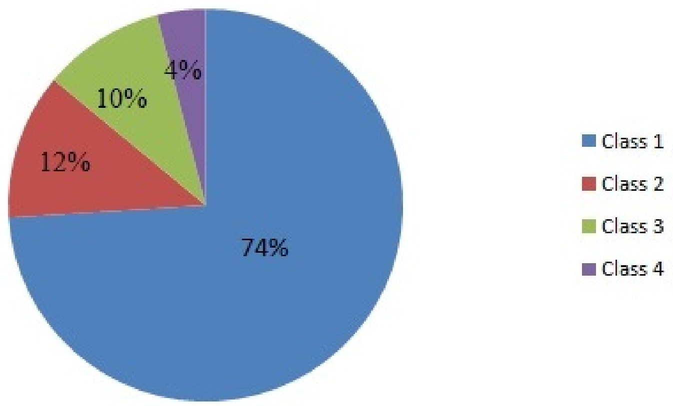

Figure 2: The Ratio of the 4 classes of Schools

Our analysis is based on three main aspects of these schools. They are:

- Aspect One: **the student resources** (Measurement: the SAT scores and the ACT scores of the students admitted)
- Aspect Two: **the future development prediction of the undergraduates** (Measurement: the mean earning of the graduates)
- Aspect Three: **the financial situations and policies of the school** (Measurement: the standard deviation of the average net prices for different family income).

These four classes perform differently in the following three aspects. And to find out the difference among them will help us to draw up different investment plans for them.

Schools in class one play a balanced role in these three aspects, and their students resources are better generally.

Schools in class two perform commonly in the student resources. However, they are always excellent in the last two aspects, which means that the graduates from these schools have superiority in the job market. Furthermore, we can notice that nearly each school in class two has dominant majors. For example, the major of business and management in Babson College and Bentiey University are first-class in the U.S..

In class three, schools perform well in the first two aspects, while they all have a dramatically low score in the third aspect. It is a classical characteristic of the universities of science and engineering. And in these 5 institutions, the degrees in engineering are more than 50%.

The majors in the two schools in the last class are mainly about health professions and related programs. As a result, they perform worse in the first and the third aspects comparing to the others. However, the mean salary of graduates is dramatically higher than other universities. These two schools have cultivated a great number of professional people that are skillful in the pharmaceutical field.

We can also find that there are only two public universities among these 50 schools. They are Georgia Institute of Technology-Main Campus and Colorado School of Mines.

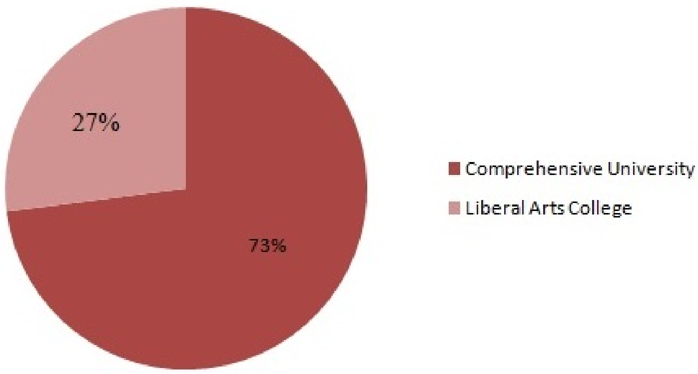

Figure 3: The Ratio of Comprehensive Schools and Liberal Arts Colleges in Class 1

Meanwhile, there are 27 comprehensive universities and 10 liberal arts colleges in the first class, which is demonstrated in the pie graph above (Figure 3, page 14).

6 The Optimal Investment Strategy

6.1 Adjustment and the allocation of the funds

Through inspecting the enrollment of the undergraduates, we can find that the number of the undergraduates of the selected schools varies from 803 to 18087. That is to say, if we want to get an optimal investment strategy, the amount of the investment per school should be determined by not only the potential for effective use of private funding, but also the number of the enrolled students.

The Goodgrant Foundation intends to donate a total of US100 million per year, and there are 50 selected schools in total. In average, each school will get US2 million. However, considering the different characteristics of these schools, we adjust the amount of the investment each school gets according to which class it belongs to and the number of the enrolled students of the school.

We can notice that nearly all the schools are well-known private schools in the first class. And besides the allowance from the government, many alumni and corporations will donate to them, so that they have much steadier financial income compared with the others. Meanwhile, the number of the enrolled students in comprehensive universities in class one is 7329 in average, while the number of the enrolled students in liberal arts colleges in class one is 1905 in average. So, we plan to invest in \$2 million to each comprehensive university in the first class, and \$1 million to each liberal arts college in the first class (table 7, page 15).

The Funds Allocation of Class One				
Kind of the institution	Enrollment	Amount of investment	Number of the institutions	Total amount of investment in class 1
Comprehensive University	7329 (AVG)	2 million dollars	27	64 million dollars
Liberal Arts College	1905 (AVG)	1 million dollars	10	

Table 7: The Funds Allocation of Class One

For the six universities in the second class, we donate a total \$14 million to them, and we do some adjustment of the amount per school according to the enrollment of the undergraduates. The funds allocation of the second class is showed as the following table (Table 8, page 15).

The Funds Allocation of Class Two			
Name of the institution	Enrollment	Amount of investment	Total amount of investment in class 2
Lehigh University	4904	2.4 million dollars	14 million dollars
Villanova University	6856	2.5 million dollars	
Babson College	2106	2.2 million dollars	
Bucknell University	3498	2.3 million dollars	
Bentley University	4172	2.4 million dollars	
Lafayette College	2452	2.2 million dollars	

Table 8: The Funds Allocation of Class Two

These years, the computer science industry and engineering industry are developing so rapidly that professionals in those fields are in great need. Therefore, we plan to invest a bit more in the schools in the third class to encourage students who are good at science and engineering to choose the school in class three. The detailed allocation plan is indicated in the following chart (Table 9, page 16).

The Funds Allocation of Class Three			
Name of the institution	Enrollment	Amount of investment	Total amount of investment in class 3
Rensselaer Polytechnic Institute	5379	2.5 million dollars	15 million dollars
Stevens Institute of Technology	2662	2.5 million dollars	
Rose-Hulman Institute of Technology	2165	2.5 million dollars	
Georgia Institute of Technology - Main Campus	13975	5 million dollars	
Colorado School of Mines	4197	2.5 million dollars	

Table 9: The Funds Allocation of Class Three

The majors in the two schools in the last class are mainly about health professions and related programs. Pharmaceuticals industry plays a important role in improving human beings' health. Therefore, we decide to invest \$2.5 million in Albany College of Pharmacy and \$4.5 million in MCPHS University, hoping to encourage more students to work in medical industry to make contributions to the society (Table 10, page 16).

The Funds Allocation of Class Four			
Name of the institution	Enrollment	Amount of investment	Total amount of investment in class 3
Albany College of Pharmacy	1075	2.5 million dollars	7 million dollars
MCPHS University	3808	4.5 million dollars	

Table 10: The Funds Allocation of Class Four

6.2 Return On Investment (ROI)

In general, the return on investment can be divided into two parts, one social and the other financial. As a charitable organization, we mainly concern the social return on investment. In other words, the social contribution of the college students.

Therefore, the return-on-investment (ROI) is assessed through the three main parameters, the students resources, the future development prediction of the undergraduates and the financial situations and policies of the school. It is a fact that the better the students resources and the future development prediction of the undergraduates, the bigger the return-on-investment (ROI). Also a school with excellent financial policies always can take advantage of the donation better.

In order to maximize the return-on-investment, we study the characteristics of these five classes of institutions and make different detailed plans for them. For example, we invest more in the institutions who are good at medical majors and majors about science and engineering.

6.3 Time duration

Considering the fact that there may be some changes in the coming five years, it is better to make out an optimized and prioritized candidate list of schools to invest every year. And we can adjust the schools, the investment amount of each school tightly according to the list. For example, we can donate to those schools who improve a lot and stop donating to the ones who fall behind rapidly. It can also stimulate candidate schools and their undergraduates to perform better.

7 Other Solutions and Future Work

7.1 Other solutions and discussions

7.1.1.On principle component analysis

When deciding the coefficients of the final scores in principle component analysis, we consider the possibility of defining $Level$ as

$$Level = 0.40329Score + 0.35816Income - 0.20325Fund.$$

The explanation is that it is possible the less financial assistance the schools gave to the students, the more potential it has got for us to invest in. Yet the result doesn't seem reasonable. According to this method, we found many schools of high education qualities are not included in the final list, which is contradictory to many researches on college education.

What's more important is that if we consider the students' earning level after entry, do principle component analysis upon the variable mn_earn_p10 and gt_25k_p6, and denote the principle component as $Earning$, it has a positive relation with the variable $STDEVP$, with the coefficient $r = .463$, $p < .001$. Therefore we know $STDEVP$ or the related variable $Fund$ should be considered to have a positive effect on the potential of the students' performance and our investment. So the model was abandoned.

7.1.2.On cluster analysis

At first, we can not decide how many classes should the selected schools be divided in. With the help of the cluster analysis, we try to divide them into three groups, four groups and five groups. And we notice that it is hard to summarize the similarities and differences in the classes and among the different classes when we set up three or five groups (the 3-class model is coarse, and the 4-class model involves some isolated schools).

However, when the schools are divided into four groups, schools of each group have reasonable and distinct common features. Therefore, we finally divide these selected schools into four classes, and make different donation plans for each of them.

7.1.3.The attempt to fit the three factors into a formula

We also try to fit a formula for the three factor we consider so that we can calculate the exact return on investment. We do principle component analysis upon the variable SAT_AVG and $ACTCMMID$, mn_earn_p10 and gt_25k_p6, and then denote each component as $Grade$, $Earning$. The linear regression with $Earning$ as a dependent variable does not perform well, with the adjusted R square $R^2_{adj} = .29$. What's more, the polynomial model

$$Earning = \lambda(Grade - G_0)(STDEVP - S_0)$$

doesn't do well either, with adjusted R square $R^2_{adj} = .64$. Considering the poor predicting ability, therefore we abandoned this kind of model.

7.2 Future work

As has been mentioned, computer science and engineering are playing more and more important roles, which means that they have great potential to develop. Thus, The Goodgrant Foundation should put emphasis on the schools of the third class, and improve the amount of donation to them according to the rapid development of the related industry.

Yet, in our strategy, we only select 50 schools to donate, and it's a pity that there are only 5 schools from the third class, many schools whose related majors are in the first class aren't contained in our list.

Besides, the predicting ability of our model still needs to be examined by more data. To some extent, our model is still a little inaccurate.

8 Error Analysis and Stability Analysis

8.1 The error analysis

There may be some errors in the establishment of our investment strategy.

- Due to the fact that some schools are in lack of the data of the single-subject scores of ACT, some inevitable errors appear when we estimate the academic achievements of them, which may also influences the final ranking.
- We havent studied the detailed financial income situation of these schools, and just decide the amount of the investment of each school by the class it belongs to and the number of the students of that school, which causes some errors, too. In fact, the amount of the government financial allowance is another important factor.

8.2 The stability analysis

Consider the formula

$$Level = 0.40329Score + 0.35816Income + 0.20325Fund,$$

we would like to test the stability of the three coefficients. We consider an error of 2.5% of each coefficient, and select the top fifty schools once more to compare it with the original result.

A +2.5% error for the coefficient of $Score$ leads to the result that 13 schools' rankings alter 1 rank, 1 school's alters 2 ranks, and 1 school's alters 3, while 1 school is out of top 50. The paired t test shows that $t = -.30, p = .976$. A -2.5% error for the coefficient of $Score$ leads to the result that 8 schools' rankings alter 1 rank, 3 schools' alter 2 ranks, while 1 school is out of top 50. The paired t test shows that $t = .30, p = .976$.

A +2.5% error for the coefficient of $Income$ leads to the result that 14 schools' rankings alter 1 rank, 1 school's alters 2 ranks, while 1 school is out of top 50. The paired t test shows that $t = .106, p = .916$. A -2.5% error for the coefficient of $Income$ leads to the result that 14 schools' rankings alter 1 rank, 3 schools' alter 2 ranks, 3 schools' alter 3, and 1 school's alter 6, while 1 school is out of top 50. The paired t test shows that $t = -.106, p = .916$.

A +2.5% error for the coefficient of $Fund$ leads to the result that 11 schools' rankings alter 1 rank, 1 school's alters 2 rank, while 1 school is out of top 50. The paired t test shows that $t = -.131, p = .896$. A -2.5% error for the coefficient of $Fund$ leads to the result that only 10 schools' rankings alter 1 rank. The paired t test shows that $t = .131, p = .896$.

In conclusion, the three coefficients are all stable to a great extent, and there is no significant difference between the original result and the altered one. Thus we claim that our principle component analysis model is considered to be stable.

9 Strengths and Weaknesses

9.1 Strengths

- Having done the error analysis and the stability analysis, our model is stable to a great extent, and thus is reliable.
- We include all the important elements to assess the students resources of the candidate schools. The total scores and the single-subject scores of both ACT and SAT are discussed in the model.
- To predict the future development of the undergraduates, we take the average salary of graduates into account. To some extent, the average salary of graduates can show the social contribution of these graduates and whether the professional education of the school is competitive in the employee market.
- In order to maximize the return-on-investment (ROI), we assess the financial policies of the candidate schools, and take the standard deviation of the average net prices for different family income into account.
- The use of Principal Component Analysis (PCA) helps us to confirm the weighting coefficient of all the indicators objectively and select the schools to invest reasonably.
- By taking advantage of the cluster analysis, we divide selected schools into 4 classes according to their characteristics. And we make different plans for these classes so as to improve the return-on-investment (ROI).
- In order to make full use of the donation, we determine the investment amount per school due to the enrollment of undergraduate degree-seeking students of the school and the characteristic of it.

9.2 Weaknesses

- When estimating the student resources, we don't take the moral characters of the students into consideration.

- When it comes to the future development of the undergraduates, we take the average salary of the graduates. However, the job market is changing quickly, and the current situation may be different from the future one.
- When we estimate the return-on-investment (ROI), we have not take the edge effect into account.

References

[1] www.nces.ed.gov/ipeds,

[2] https://collegescorecard.ed.gov,

[3] https://en.wikipedia.org/wiki/Charitable_organization#United_States,

[4] https://en.wikipedia.org/wiki/Return_on_investment,

[5] https://en.wikipedia.org/wiki/ACT_,

[6] http://colleges.usnews.rankingsandreviews.com/best-colleges.

10 A Letter to Mr. A. Chiang

Dear Mr. Alpha Chiang:

We team has been asked to develop a model to determine an optimal investment strategy for the Goodgrant Foundation. It's our honor to participate in your project to help improve educational performance of undergraduates in the United States.

To make full use of the funds and maximize the return-on-investment (ROI), we build up a model and estimate the candidate school's potential for effective use of private funding with the help of Principal Component Analysis (PCA). PCA is a modeling approach which can help people to conform the weighting coefficient of all the indicators in the evaluation system. The three main initial elements considered in the estimation are:

- The student resources: depending on the average SAT equivalent score of students admitted and the midpoint of the ACT cumulative score.
- The future development of the undergraduates: lying on the mean earnings of students 10 years after entry and the shares of students earning over $25,000/year 6 years after entry.
- The financial situations and policies of the school: showed by the standard deviation of the average net prices for different family incomes.

After the calculation, we give a score to each candidate school to measure the potential and give out a ranking. A school's score also indicates the amount of return if we invest in this school, so we determine to donate to the top 50 institutions on the list.

As the selected schools are highly different, it is inappropriate to invest the same amount of money in them. Therefore, we divide the selected schools into four classes with a modeling approach called Cluster Analysis. And each class has distinct characteristics. According to these characteristics and the enrollment of the undergraduates, we can give out a detailed plan for the funds investing in all the four classes of schools.

- Class One: Nearly all the schools are well-known private schools. Considering the fact that the number of the undergraduates in comprehensive universities is far more than that in liberal arts colleges, we plan to invest US 2 million to each comprehensive university and US 1 million to each liberal arts college.
- Class Two: For the six universities in the second class, we donate a total US 14 million to them, some adjustments are done due to the enrollment of the undergraduates.
- Class Three: In these 5 institutions, the majors about engineering are all accounting for more than 50%. To encourage more students to choose these schools, we plan to invest a bit more in the schools in the third class than the average.
- Class Four: The majors in the two schools in the last class are mainly about medical care. Therefore, we decide to invest US 2.5 million and US 4.5 million respectively, hoping to encourage more students to work in the medical industry.

Considering the fact that there may be some changes in the coming five years, we suggest that the foundation make the investment strategy every year, which can also simulate these schools and their undergraduates to perform better.

To estimate the return-on-investment (ROI), we tend to invest in those school who has the very potential to improve educational performance. And for different schools, we find out that only when we invest an appropriate amount of money according to the enrollment and characteristics of the school, we can maximize the total ROI.

Hope our strategy will be helpful for you. And the chart in the end is the optimal investment strategy, containing the schools selected and the amount of funds per school.

Yours sincerely, Team # 42642

million dollars	institution name	million dollars	institution name
2	Massachusetts Institute of Technology	2	Brown University
2	Harvard University	2	University of Southern California
2	Stanford University	2	Case Western Reserve University
1	Harvey Mudd College	2	Northeastern University
2	Duke University	1	Bowdoin College
2	University of Pennsylvania	1	Middlebury College
2	California Institute of Technology	1	Amherst College
2	Georgetown University	2	Emory University
2	Yale University	1	Wellesley College
2	Rice University	1	Colgate University
2	University of Notre Dame	1	Williams College
2	Princeton University	1	Davidson College
2	University of Chicago	2.4	Lehigh University
2	Dartmouth College	2.5	Villanova University
2	Tufts University	2.2	Babson College
2	Cornell University	2.3	Bucknell University
2	Johns Hopkins University	2.4	Bentley University
1	Washington and Lee University	2.2	Lafayette College
2	Washington University in St Louis	2.5	Rensselaer Polytechnic Institute
2	Northwestern University	2.5	Stevens Institute of Technology
2	Carnegie Mellon University	2.5	Rose-Hulman Institute of Technology
2	Boston College	5	Georgia Institute of Technology-Main Campus
2	Vanderbilt University	2.5	Colorado School of Mines
2	Columbia University in the City of New York	2.5	Albany College of Pharmacy and Health Sciences
1	Claremont McKenna College	4.5	MCPHS University