1 Summary

The point of interest is that is it possible to predict a charity's total contribution amount based on its financial state, organization type and accountability. My model suggests that total contribution is effected by fundraising expenses, financial score, organization type and leader compensation. It was determined that there is a positive effect between fundraising expenses and total contribution and there is a 7% increase in total contribution for each point increase in total contributions. Lastly, the effect of leader compensation is heavily determined by the org type they are a part of and in the case of religion, leader compensation has no effect on total contribution.

2 Introduction

Charities all over the world exist to obtain funding from the populace and put that funding towards each charity's mission. Using charity fundraising information as well as other financial metric and details about those charities in the united state from two years ago, I want to determine if it is possible to predict fundraising given fundraising expenses, leader compensation, and what type of organization the charity.

3 Data

The data is scraped from https://www.charitynavigator.org/ by Rashedul Haque and published to data world. Furthermore this was done in 2019 so the data is fairly recent and contains pre-pandemic numbers so health organizations might not have the boost they could have had during the height of the pandemic. This scraped data includes the following information for each charity: accountability_score, administrative_expenses, charity_name, charity_url, city, cn_advisory, compensation_leader_compensation, compensation_leader_expense_percent, compensation_leader_title, excess_or_deficit_for_year, financial_score, fundraising_expenses, net_assets, organization_type, other_revenue, overall_score, payments_to_affiliates, program_expenses, state, and total_contributions. Just taking a quick glance at these columns we can quickly gather that there is a lot of financial information on each charity and given that these pertain to the year 2019 there will be a very high likelihood that all these different financial metrics are correlated.

First starting with looking into our response variable of interest which is total contribution. In the figure below we see that across all the charities that funding is heavily skewed with very few charities having billions in funding. To resolve this non-normal trend I have logged total contribution and it looks more normal after that with a peak at 14.5. This similar distribution existed with a lot of the other financial information such as leader compensation, payments to affiliates, expenses, revenue and so on. So all those metrics were taken as a log.

Furthermore, my initial guess of the relationship between other financial predictors are extremely large which was noted by doing ggpairs between all the columns. Taking a look at some of the key relationships between total contribution vs leader compensation, financial score, and fundraising expenses yielded the following plots:

The plots above show that for the log of leader compensation there is a steep positive relationship with log of total contribution besides a few scattered points. This suggests that it might be significant predictor. For financial score, we see that it has a cone shape starting at around a score of 60 and

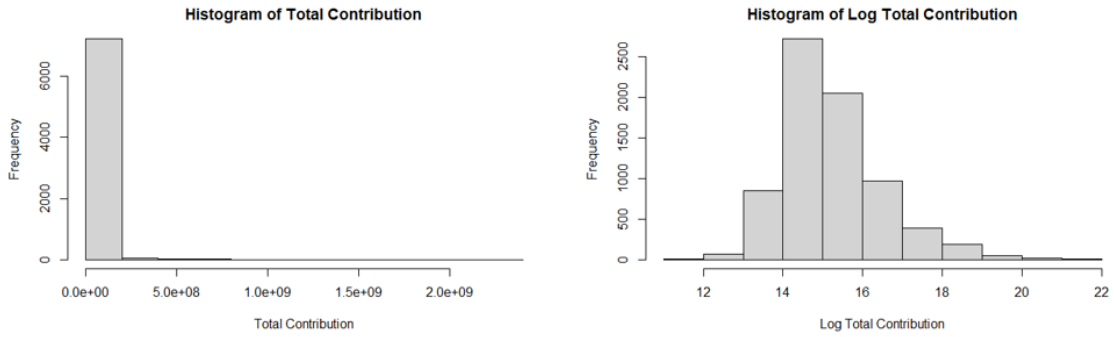

Figure 1: Total Contribution before and after Log

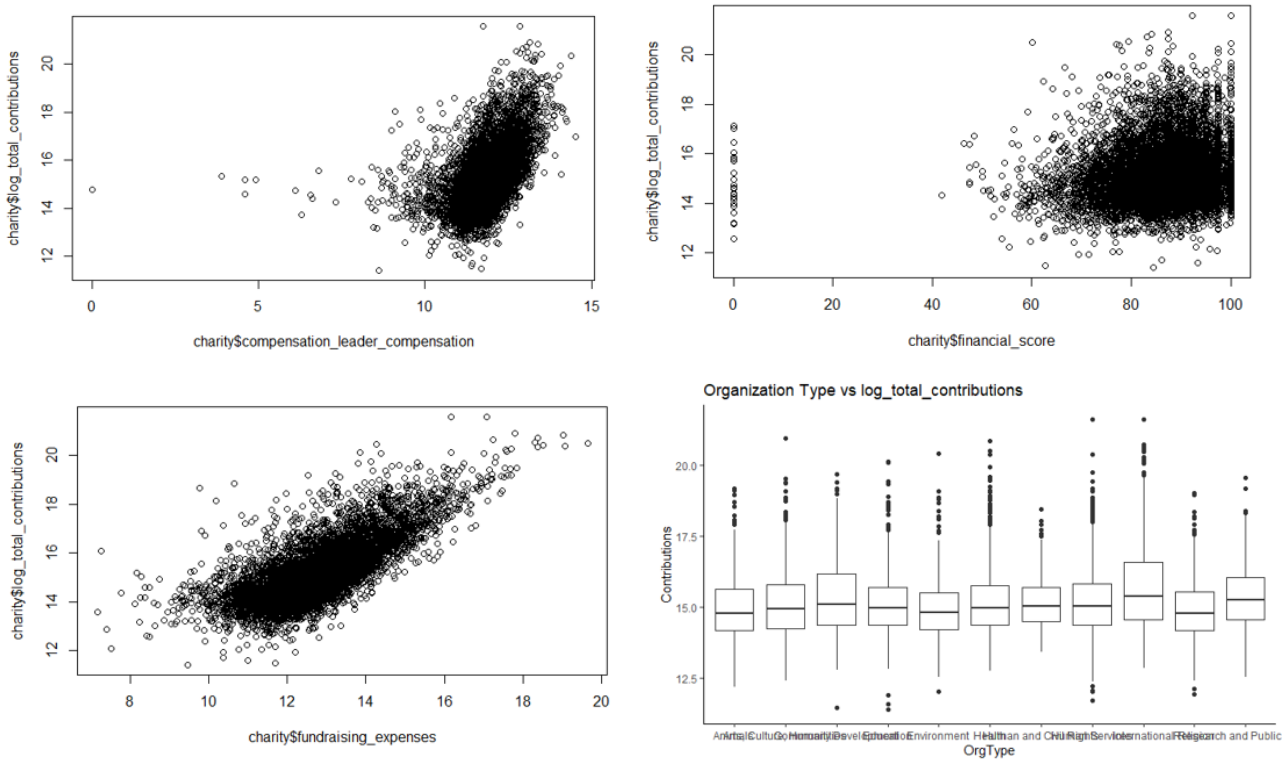

Figure 2: Log total contribution vs log leader compensation, financial score, log fundraising expenses, and by charity type

this cone expands more upward than downward. This suggests a slight positive relationship between financial score and log total contribution. Lastly, for log fundraising expenses we see a distinct positive relationship with log total contribution.

When looking at a categorical variable of org type (only the primary category was used) we see that overall there seems to be a relatively even total contribution for medians but given that this is a on a log scale there are certain organizations in each grouping that pull in a lot more funding than the rest.

I looked into interactions but the only note able one was orgtype and leadership contributions but even that looked very slight especially between all the categories.

4 Model

The model development process began with developing a simple linear model between total compensation and fundraising expenses and as you can expect that fundraising was significant and had a coefficient of 0.72 and a standard error of 0.0076 and a p-value of very close to 0. Furthermore this satisfied all of the normality assumptions. From this we can state that there is a relationship between the two but how does this compare to all the other predictors?

To answer that I used both AIC and BIC on log total contribution vs. accountability_score, compensation_leader_compensation, organization_type:compensation_leader_compensation, fundraising_expenses:organization_type, fundraising_expenses:accountability_score, compensation_leader_compensation:accountability_score,fundraising_expenses, organization_type, and financial_score.

As you can see from the suggested full model before AIC and BIC all the heavily correlated financial variables were dropped as they did not add anything significant to the overall model and just muddied the effects of those listed in the equation above.

Running AIC resulted in retaining pretty much the full model but if we punished for the excessive predictors via BIC it resulted in the following relationship where X is the matrix of fundraising_expenses, financial_score, organization_type, compensation_leader_compensation, and organization_type:compensation_leader_compensation.

$$log(total\_contributions)_i = \beta * X + \epsilon_i \tag{1}$$

The BIC model was deemed the best as it properly punished including too many predictors and removing those that were not significant.

The plots for the linearity assumption are in the appendix but they show that for leader compensation it is fair spread out without any clear pattern although it does scatter as an oval around the 0 line which is fine. As far as the linearity assumption goes for fundraising expenses we can see that there is some what of a patter in terms of a clustering in a triangle pattern. I have tried making this an exponent and a few other transformations but they did not add enough value to take away from the interpret-ability of the model. Lastly for financial score we can see that there is no pattern in the distribution. Through those plots we can say that the linearity assumption is satisfied.

Looking at the residuals vs leverage plot (appendix) we see that only two points have any leverage larger than .2 but their cooks distance is within .5 therefore it they are not substantial enough to warrant removal. Upon further investigation those two points gained a lot in total contribution with very little spent in terms of expenses.

The normalQQ plot(Appendix) follows the 45 degree dotted line till the positive 2nd quantile where it veers off drastically that mean that the distribution of the residuals is relatively normal except for a small portion of the points which is not significant enough to break the normality assumption. The residual vs fitted plot(Appendix) shows that although there is a slight trend noted by the red line the points that make up the sloped portion are very sparse meaning they are not enough to break the independence and equal variance assumptions.

The model produced the following coefficient values:

From this summary of the BIC model we can see that the significant predictors are fundraising expense, financial score, organization type, leader compensation, and organization type:leader compensation. In the case of log fundraising expenses, the total contribution tends to increase by 69% of fundraising expenses as both sides are base 10 logged. In the case of financial score, for every 1 point in crease in financial score there is a 7% increase in total contribution. Furthermore based on our model we can see that there is an interaction between organization type and leader compensation. So

	log_total_contributions		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.36	-1.56 – 0.84	0.553
fundraising_expenses	0.69	0.67 – 0.70	**<0.001**
financial_score	0.03	0.03 – 0.03	**<0.001**
organization_type [Arts, Culture, Humanities]	1.01	-0.39 – 2.41	0.157
organization_type [Community Development]	-1.00	-2.48 – 0.49	0.188
organization_type [Education]	1.20	-0.45 – 2.85	0.154
organization_type [Environment]	0.60	-1.37 – 2.56	0.552
organization_type [Health]	0.81	-0.68 – 2.30	0.286
organization_type [Human and Civil Rights]	3.32	1.70 – 4.93	**<0.001**
organization_type [Human Services]	2.85	1.52 – 4.18	**<0.001**
organization_type [International]	2.58	1.11 – 4.04	**0.001**
organization_type [Religion]	4.45	2.86 – 6.05	**<0.001**
organization_type [Research and Public Policy]	2.20	0.38 – 4.02	**0.018**
compensation_leader_compensation	0.35	0.25 – 0.45	**<0.001**
compensation_leader_compensation	0.35	0.25 – 0.45	**<0.001**
organization_type [Arts, Culture, Humanities] * compensation_leader_compensation	-0.09	-0.21 – 0.03	0.132
organization_type [Community Development] * compensation_leader_compensation	0.12	-0.00 – 0.25	0.058
organization_type [Education] * compensation_leader_compensation	-0.09	-0.23 – 0.05	0.226
organization_type [Environment] * compensation_leader_compensation	-0.04	-0.21 – 0.13	0.630
organization_type [Health] * compensation_leader_compensation	-0.06	-0.18 – 0.07	0.381
organization_type [Human and Civil Rights] * compensation_leader_compensation	-0.26	-0.40 – -0.13	**<0.001**
organization_type [Human Services] * compensation_leader_compensation	-0.21	-0.33 – -0.10	**<0.001**
organization_type [International] * compensation_leader_compensation	-0.16	-0.28 – -0.03	**0.012**
organization_type [Religion] * compensation_leader_compensation	-0.35	-0.49 – -0.21	**<0.001**
organization_type [Research and Public Policy] * compensation_leader_compensation	-0.16	-0.31 – -0.01	**0.041**
Observations	7290		
R^2 / R^2 adjusted	0.650 / 0.649		

Figure 3: Summary of BIC Model

if we want to understand the effect of leader compensation on total compensation we can take religious org as an example and we can see that for a single point increase (on a log scale) for the leader compensation given he is part of a religious org is 0 as (.35 - .35) which suggests that for a religious group their total contributions is not effected by leadership contribution where as for environmental organization the coefficient for log leadership compensation is .31 (.35 - .04) which means that for a single point increase on a log scale for leader compensation the total contribution increases by .31 on a log scale.

Looking at VIF scores only fundraising expenses and financial scores had low VIF values whereas all other predictors of religion and leader compensation and their interactions had high VIF scores. I determined that it is a point to note but not make or break as there is an interaction between those predictors to compensate.

5 Conclusion

Based on our model we can conclude that there is a relationship between total contributions and fundraising expense, financial score, org type and leader compensation. Based on our model, a single point increase on a log base 10 scale for fundraising expenses results in a .69 increase on a log scale for total contributions. Furthermore, for each point increase in financial score results in a 7% increase in total contribution. Also from our model we can gauge that certain org types draw more funding than others such as Religion drawing in far more contributions compared to community development. The effect or leader compensation on total contribution is heavily tied to the org type they are a part of. In the case of religion in has no bearing but for other org type such as health there is a substantial effect.

One limitation that I noticed within this is that there is a lot of financial metrics for each Charity which were heavily correlated which means that only 1 or 2 of those could be used. As a further buildup on top of this, the VIF values are really high for all values besides fundraising expenses and financial scores.

6 Appendix

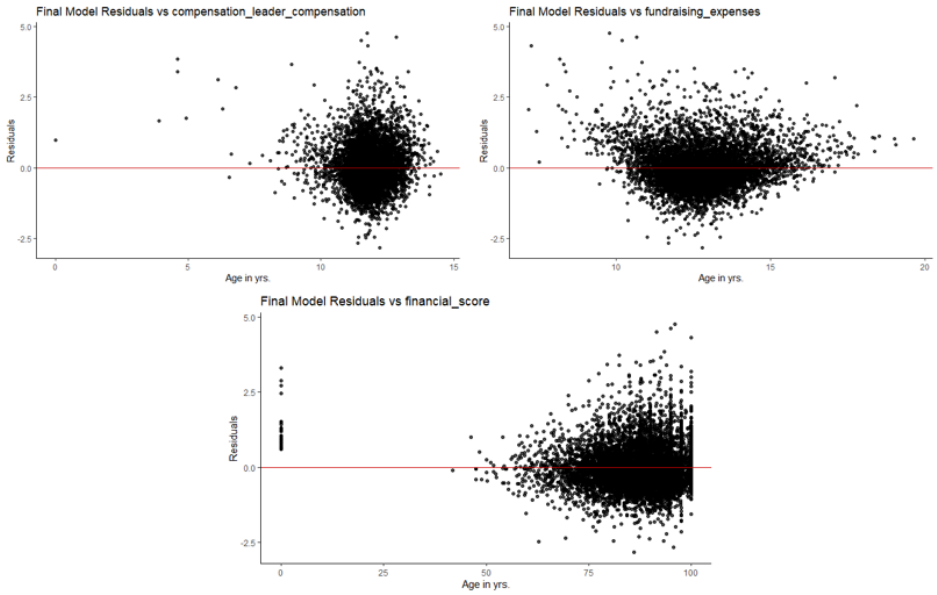

Figure 4: Plots for linearity assumption

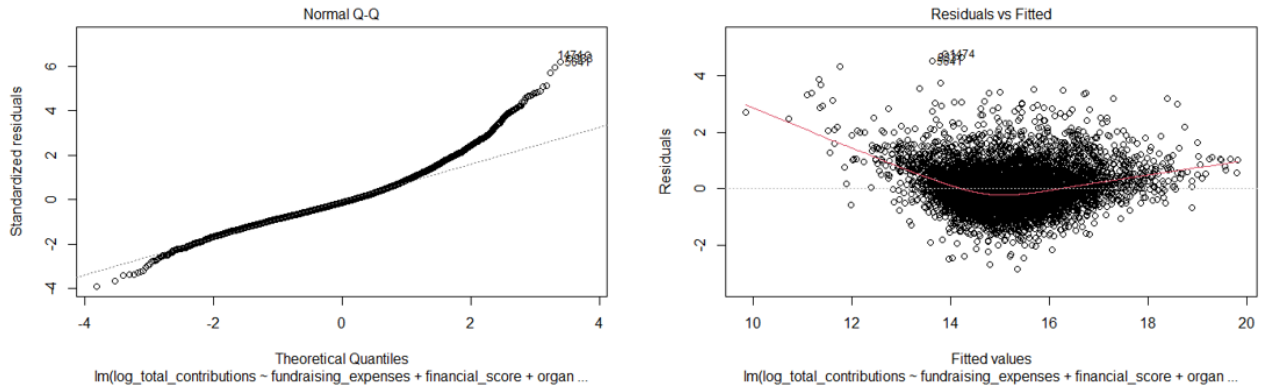

Figure 5: Normality, Independence, and Equal Variance Assumption Plots

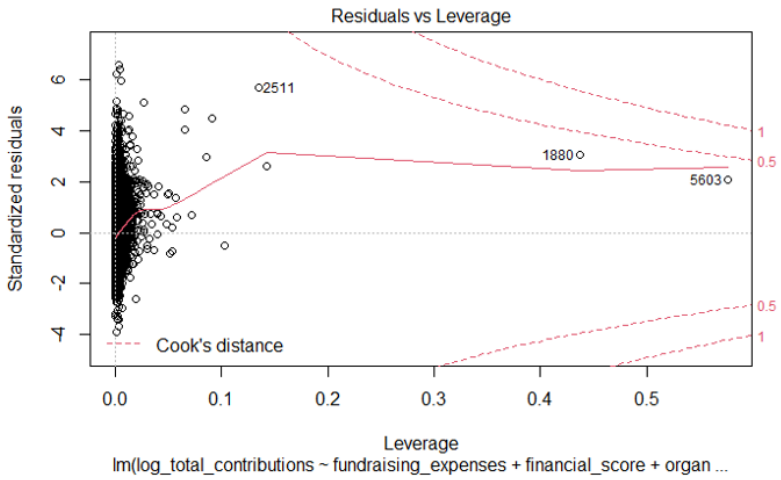

Figure 6: Outliers and Leverage