**Abstract**

The primary goal of this study is to maximize profit while streamlining out marketing campaigns to make them more cost effective. To accomplish this, we used various statistical modeling methods to predict the amount that will be donated and classify if a person is likely to be a donator. We found that the Poisson model best predicts the dollar amount expected from a potential donor. From this we concluded that to maximize donations we should focus our direct marketing campaign on the follow factors: families that live in region 4 or 5, have children, have higher income brackets, and are previous donor.

**Introduction**

Our current mail in donor strategy is currently not cost effective and will cost us more in the long term. The goal of this study is to make our direct marketing campaigns more effective by maximizing profit from our most likely donors. We plan to answer this question by identifying who are most likely to be donors and then using those characteristics to predict how much someone is going to donate. By answering these two questions, it will allow us to focus our efforts on a specific set of clienteles that will help maximize our efforts.

**Methodology**

**Data Exploration**

During our data exploration we looked for the factors that we believe would impact the amount a donor will contribute. These factors included the regions a specific person lived in, the number of children a potential donor has, various levels of income, and their donation frequency. First, we wanted to see if there was a pre-existing relationship between any of the values in the data set by using a correlation matrix. In Figure 1 we see that the positive correlations are focused on income specifically the median and average family income. There is a negative correlation between these same income factors and the percentage of people categorized as low income in a neighborhood. Second, we looked for a possible relationship between gender and their likelihood of being a donor. We utilized the bar chart in Figure 2 to visualize this relationship. Here we see that gender was not a large indicator of whether someone was a donor, but we see there are generally fewer female donors than male donors. Finally, we explored any potential relationships between income and donors. We used the boxplot in Figure 3 to explain this relationship. In general, we see there is an almost negligible difference between the median income of donors and non-donors.

|  |  |  |
| --- | --- | --- |
| A blue and red triangle with white text  Description automatically generated  Figure 1. Correlation Plot of Variables in Data Set | A graph with a red and blue square  Description automatically generated with medium confidence  Figure 2. Bar Chart of Donors by Gender | A graph of a family income by donor  Description automatically generated  Figure 3. Median Income by Donors |

**Models**

For the first model we fit a Poisson model (Figure 4) and selected all the factors that did not explicitly mention income. The same factors are fixed values in all of the models.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

Figure 4. Poisson Model

The next model we created was a negative binomial model to handle potential overdispersion in the original Poisson model (Figure 5).

Figure 5. Negative Binomial Model

Finally, we created a quasi-Poisson model (Figure 6) to completely handle the overdispersion in the data and allow us to relax the variance mean ratio.

Figure 6. Quasi Poisson Model

To handle the classification, we used a quasi-binomial model (Figure 7) to classify the log likelihood of a potential person being a donor.

Figure 6. Quasi Binomial Model

From our data exploration we believed that income was overall negligible and only included the variables that did not relate to income. Those variables include regions, number of children, gender, dollar amount gifted, and frequency of gifting. We will also be evaluating our models based on the mean squared error (MSE).

**Results**

When comparing the models we used mean squared error as the metric to compare the effectiveness of maximizing the dollar amount of donations. Based on the mean squared error values in Table 1, we found that the Poisson model performed the best.

|  |  |
| --- | --- |
| Model | MSE |
| Poisson | 1.867523 |
| Negative Binomial | 140.6796 |
| Quasi Poisson | 140.6796 |
| Quasi Binomial (Classification) | 169.5454 |

Table 1. Mean Squared Error for Each Model

**Discussion**

Based on our findings in the results section and only using mean squared error, the Poisson model demonstrates that our organization should focus our direct marketing efforts towards people with the following qualities: Live in region 4 or 5, have children, have a high household income, and are current donors. With this information we can focus our main efforts on current donors and market towards individuals who are not current donors in the remaining categories. A potential drawback on over marketing to current donors is that the donor could become fatigued or grow apathetic from receiving so many letters in the mail about donating.

**Conclusion**

Through this study, our goal was to maximize the dollar amount of donations for an organization by enhancing their direct marketing campaign strategy. We found that we should tailor our marketing to people who have children, people in a higher income bracket, and to those who have a high recent donation gift and high average dollar amount of gifts to date. These findings emphasize us to invest in our current clientele base and then continue to advertise to families with children and in higher income brackets.

**Future Work**

With time permitting creating an additional spatial model (GLMM) that considered the random effects of regions on the dollar amount of donations. In addition to this, we would recommend building an additional model that offset for the income variable with seven categories. We expect these changes to yield a smaller mean squared error (MSE) and improve interpretability of the models. In addition to this we would edit the models to include a full explanation of each factor and differentiate between them more for readability.