# C:\Users\zpdenton\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\DC9B90C4.tmpExecutive Summary

St. Jude’s Children’s Hospital is one of the industry leaders for advancement in healthcare treatments, cures, prevention, and research. With a moto that no child is denied help based on race, religion of family income. The primary means of support for their company is through donations raised the ALSAC, their national fundraising organization. The organizations focus it to their marketing campaign as a vessel to generate these charitable contributions. So how can St. Jude’s maximize the donations they receive? The goal of this report is to analyze and answer this business question by creating a predictive model to classify if a person could be a donor and profile characteristics of these potential donors.

## Analysis

The model that produced the lowest misclassification rate was the logistic stepwise regression in the [optimal binning flow](#model). The regression produced a misclassification rate of [0.407](#regession). Meaning we can miss-classify a recipient by up to 40%. The model found significant variables household income, dollar amount of largest gift to date, number of children, lifetime number of promotions received to date, and number of months from last donation. I think it is important to note that while the model did not say find if the recipient was a homeowner that a majority of donors were.

After testing the model against new data, 47 % of the recipients were predicted to be donors and should receive direct-mailing for donation when in reality 50% of recipients donated. With this set of data, the model had 94% (predicted/actual) accuracy in predicting.

## Deployment

In conclusion, Is this a good model? We have established that, at a minimum, we can predict with 60% accuracy if someone will be a donor. In addition, we have more knowledge on what metrics are important to classifying if someone is a donor or not. I would conclude this model could be better but is an acceptable start and has proved the company with valuable insight.

For St. Jude’s to maximize donations I recommend that ALSAC target recipients similar to the ones that were classified as donors (EM\_CLASIFACATION=1) \*. Furthermore, look at the most frequent values for where they donated for the five significant variables: household income, dollar amount of largest gift to date, number of children, lifetime number of promotions, and number of months from last donation. St. Jude’s should [target recipients](#Recomendations) with an income level of 4, have 1 child, have not donated before, have had 24 or more promotions and haven’t donated in 28-32 months.

*(\* The linked Excel Workbook attached)*

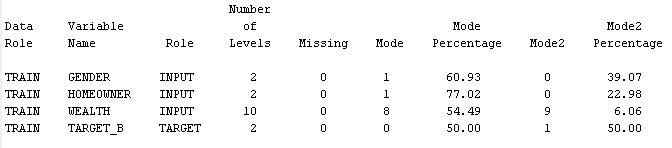
# Appendix

## Data Understanding

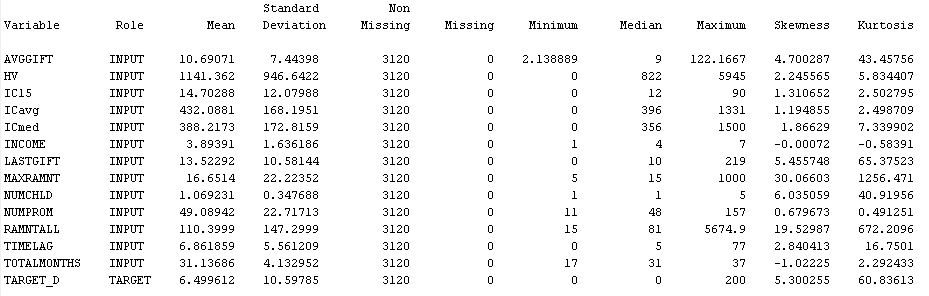
St. Jude’s marketing team has provided fundraising data seeking insights and answers to several strategic business questions:

* Understanding the current direct-mailing campaign data.
* Identify how many should receive the direct-mailing.
* Create a model to predict if the mail recipient will be a donor or non-doner.
* Determine who will be more likely to donate during the direct-mail fundraiser.
* What decision should the ALSAC make to maximize the expected net profit?

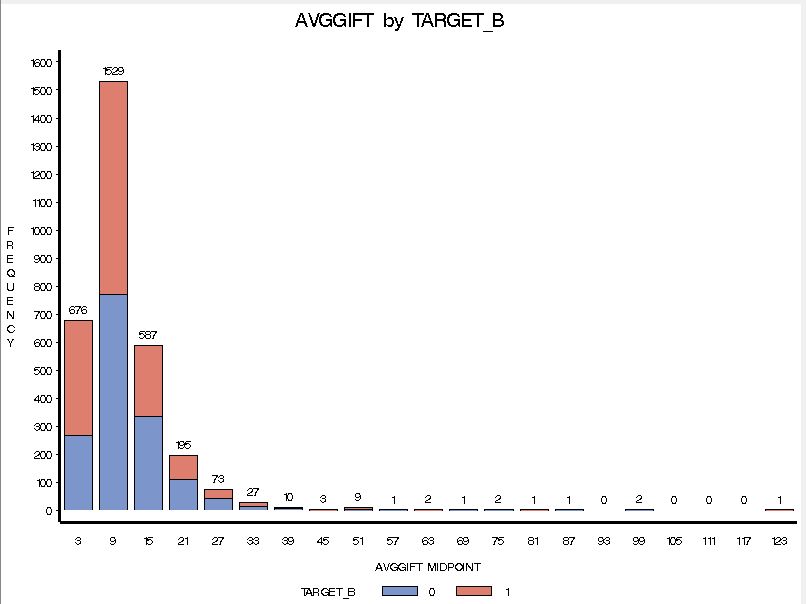
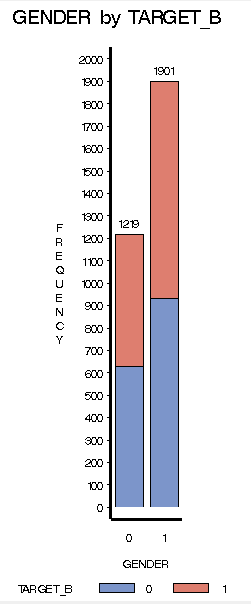
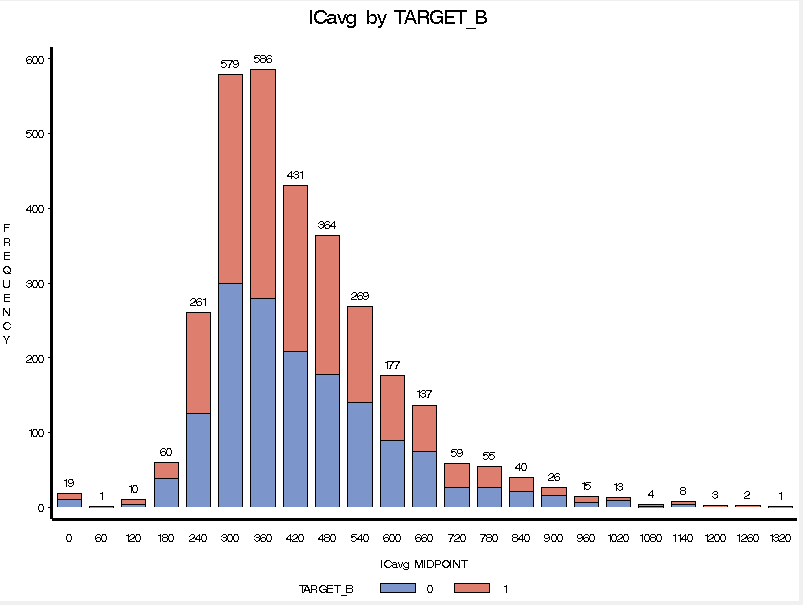
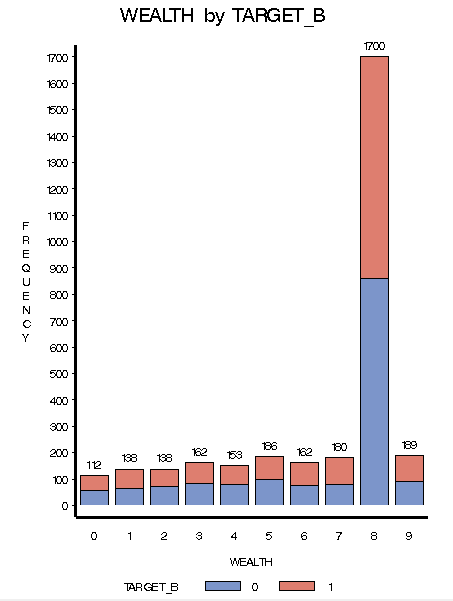
Phase one in this analysis is understanding the data to understand the variables and identify issues to address in phase two of the project life cycle. The figures below show summary statistic for the fundraising data before and transformations. First notable things variables that need their data type converted to nominal or binary types. There are no missing values which add to the integrity of the data. Many variables have large distributions (based of skewness and kurtosis) and large standard deviations to support this.



*Class Variable Summary Statistics*



*Interval Variable Summary Statistics*



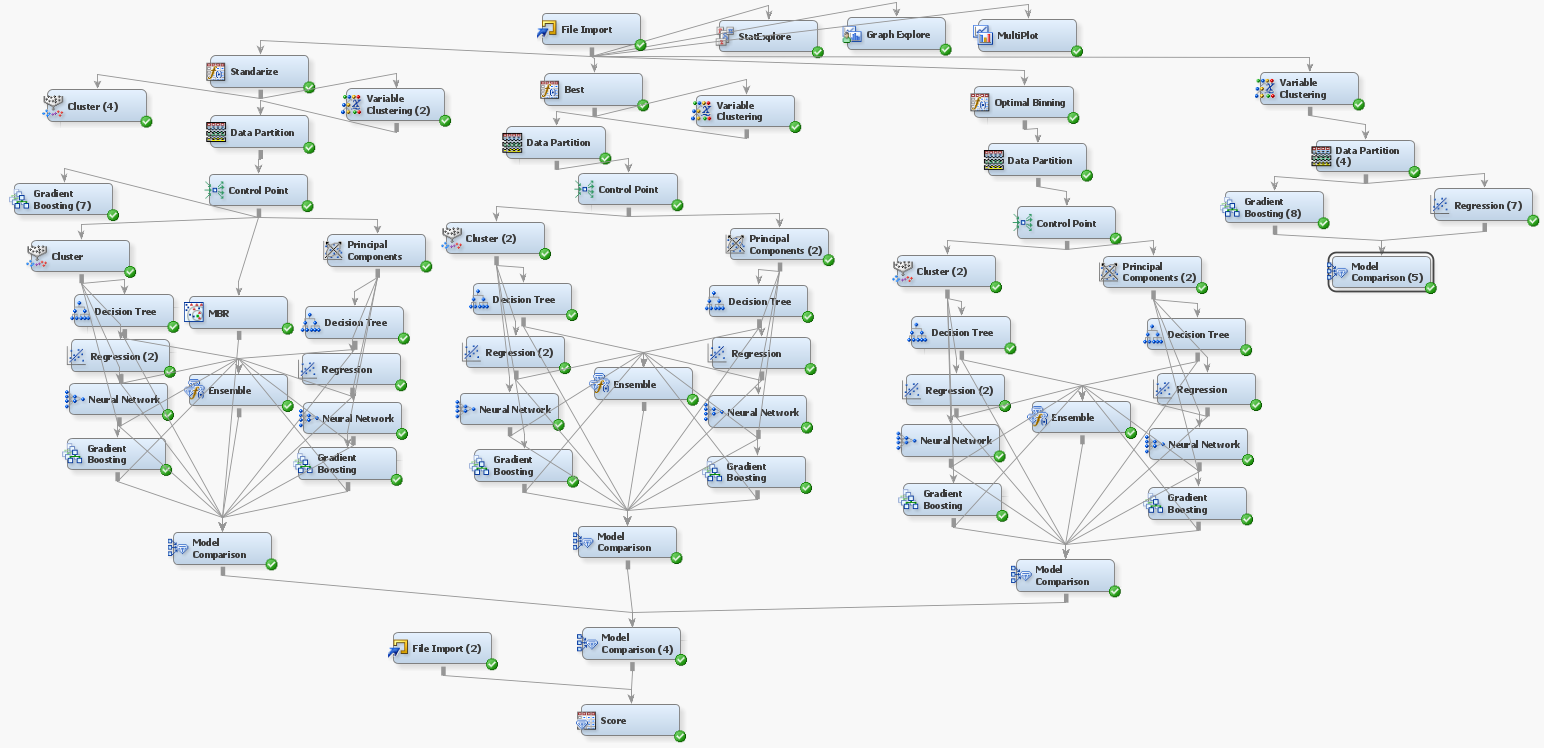
Looking at the class variables, females are recipients 61% of the time, homeowners are recipients 77%, and non-donors are recipients 50%. Recipients in the wealth level of 8 were most common to donate. The average gift was $10.7. Average family income in potential neighborhood was about $432,000. Currently females, that are homeowners with a home value of $816,000-846,000, and have a wealth level of 8, income level of 4, who have one child, and have not donated before. are currently the most targeted group in the marketing campaign.

## Data Preparation

There were several ways I changed the data for the modeling phase. I created three different ways to transform the data each with their own model flows. One was standardizing all interval inputs, two was best fit for interval inputs, and three was optimal binning. These three were done in order to make the data more normally distributed.

## Modeling

There were three model flows created for each of the transformations created. The model flow can be seen below in figure 3.



**Standardization**

* Variable clustering
* Data partition 60/40
* Control Point
* Knn (k=10)
* Cluster
  + DT (2,5)
  + Logistic stepwise regression
  + Default neural network
  + Gradient boost
* Principle components
  + DT (2,5)
  + Logistic stepwise regression
  + Default neural network
  + Gradient boost
* Ensemble of cluster flow and Principle components

**Best**

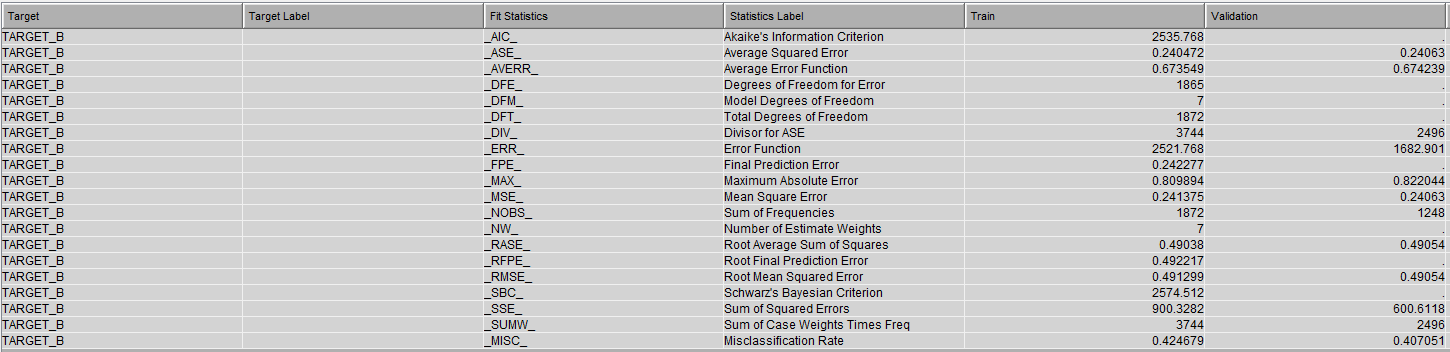
* Variable clustering
* Data partition 60/40
* Control Point
* Cluster
  + DT (2,5)
  + Logistic stepwise regression
  + Default neural network
  + Gradient boost
* Principle components
  + DT (2,5)
  + Logistic stepwise regression
  + Default neural network
  + Gradient boost
* Ensemble of cluster flow and Principle components

**Optimal Binning**

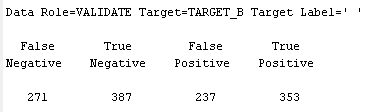
* Variable clustering
* Data partition 60/40
* Control Point
* Cluster
  + DT (2,5)
  + Logistic stepwise regression
  + Default neural network
  + Gradient boost
* Principle components
  + DT (2,5)
  + Logistic stepwise regression
  + Default neural network
  + Gradient boost
* Ensemble of cluster flow and Principle components

## Model Evaluation

After the modeling was complete all models were ran into model comparison nodes. With a class target of a donor or non-donor, we look at the misclassification metric to judge our models. The model that produced the lowest misclassification rate was the logistic stepwise regression in the optimal binning flow. The regression produced a misclassification rate of 0.407. The model selected significant variables household income (INCOME), dollar amount of largest gift to date (OPT\_MAXRAMNT), number of children (OPT\_NUMCHLD), lifetime number of promotions received to date (OPT\_NUMPROM), and number of months from last donation (OPT\_TOTALMONTHS)



*Output Statistics for Logistic Regression*



*Type 1 and Type 2 Errors*

## Deployment

Is this a good model? We have established that we can miss-classify a recipient by up to 40%. I would conclude this error rate could be better but is acceptable. The model does tell us what metrics are important to classifying if someone is a donor or not. St. Jude now knows that household income, dollar amount of largest gift to date, number of children, lifetime number of promotions received to date, and number of months from last donation are import to weather someone will donate or not.

After testing the model against new data and 47 % (590) of the recipients were predicted to be donors and should receive direct-mailing for donation when in reality 50% (624) of recipients donated. With this data the model had 94% (590 predicted/624actual) accuracy in predicting. I recommend that St. Jude target recipients similar to the ones that were classified as donors (EM\_CLASIFACATION=1). Furthermore, look at the most frequent values for where they donated for the five significant variables. In conclusion, St. Jude can now mis predict up to 40% accuracy if someone will be a potential donor. Based on the graphs below, St. should target recipients with an income level of 4, have 1 child, have not donated before, have had 24 or more promotions and haven’t donated in 28-32 months.

## Significant Variables

