# What Gives?

Vanessa Ma | Shyamsunder Sriram

BUSN 37304

# Context
## Optimizing Philanthropy

Paralyzed Veterans of America

- **Nature:** not-for-profit organization
- **Operations:** provides programs and services for US veterans with spinal cord injuries or disease
- **How they make money:** direct mail fundraising with an in-house database of over 13 million donors

**Goal: Increase Philanthropic Returns for Direct Mailing Campaigns**

- **Maximize profit:** increase the funds received from upcoming philanthropic campaigns, through 2 possible ways:
  - Increase recurring donor pool
  - Increase number of big ticket donors
- **Lower costs:** reduce number of 'failed' mailings, i.e. mailings to people who will not respond
- **Possible particular group of interest:** reining back lapsed donors
  - Defined as: donors who last gave 12-14 months ago
  - Believe that the longer a donor goes without giving, the more likely he / she will fall off the map

# Objective
## Can we increase returns by selective mail?

**1** Predicting **WHO** Will Donate

**2** Predicting **HOW MUCH** They Will Give

*The presentation is structured according to the best practices shown in session 1 of this class, providing a ‘recipe’ for good algorithm-making:

# Key Recommendation and Model Results

## Yes! Go Forth and Target

**$14,468**

**RAW PROFIT**

**$2,139**

**ABOVE BASELINE***

*The presentation is structured according to the best practices shown in session 1 of this class, providing a ‘recipe’ for good algorithm-making:

# Contents*



*The presentation is structured according to the best practices shown in session 1 of this class, providing a ‘recipe’ for good algorithm-making:

# Theory & Model

# Theory

## Two Utilities for Two Separate Decisions

### Decision 1: To Donate Or Not?

***Similar*** to choice to buy or not to buy

- Functionally not different from committing to a "method" of spending money
- Binary decision, classification question
- Hypothesis: driven by connection to the organization (i.e. census, interest variables)

Utility theory applies:

- $\boldsymbol{U_1(X_i^1) = \alpha^1 + \beta^1 X_i^1 + \epsilon}$
- $X_i^1$: for individual i, a characteristic set (1) that helps them understand the utility gain from this decision (e.g. neighbourhood, connections)
- $\alpha^1$ : baseline utility value.
- **committing to donating gives them more utility than not donating**

### Decision 2: If I Donate, How Much?

***Dissimilar*** to purchasing!

- The donation is not fixed – there is no price, customer freely decides
- Continuous variable, regression question
- Hypothesis: largely driven by ability to give (i.e. income, wealth)

Utility theory applies:

- $\boldsymbol{U_2(X_i^2) = \alpha^2 + \beta^2 D(X_i^2) + \epsilon}$
- $X_i^2$: for same individual i, different characteristic set
- $D(\ldots)$: function that outputs giving amount based on individual characteristics
- $\alpha^2$ : baseline utility value, likely 0.
- **donating to this organization gives them more utility than giving to any other**

# Theory

## Modelling the Two Utilities Separately

### Logistic Regression

- **To Donate or Not** is a binary variable
  - **classification question**
- Linear models would not make sense
- Chosen for simplicity

**Output: probability of donating**

$$\boldsymbol{P(W_i = 1) = \frac{1}{1 + e^{-(\alpha + \beta X_i)}}}$$

- $P(\ldots)$: probability distribution of logit
- $W_i$ : individual i's choice to donate (1) or not to donate (0)
- $\alpha + \beta X_i$ : individual's utility regarding the choice to donate or not to donate

### Generalized Linear Regression

- **How Much to Donate** is a continuous variable
- Conditional probability question, given that a person is classified with response = donate
- Linear regression model would work well
- Chosen for simplicity and lower requirements for computing power

**Output: expected donation value**

$$\boldsymbol{D(X_i) = mX_i + C}$$

- $D(\ldots)$ : function for donation value given individual i's parameters
- $m$ : weights vector for predictive parameters

# Data

Exploratory Data Analysis

# Dataset Overview

Results a mailing that was dropped in June 1997 to a total of 3.5 million PVA donors

- included a gift "premium" of personalized name & address labels plus an assortment of 10 note cards and envelopes
- donors were acquired through similar mailings in the past.

**Types of Variables**

## Demographics

- First record, frequency of giving
- Age, location, gender, income

**Mail Order Responses**

- Number of times donor has responded to other types of mailing offers
- E.g. Craft hobbies, gardening, magazines

**Sources of Overlay Data**

- Solicitation limits, connection to PVA-related organizations e.g. Military, Government
- Third party data origin

**Interests**

- Donor interests collected from third-party sources (binary variables)
- E.g. stereo, pets, collectables

**Census**

- Characteristics of donor neighbourhood from 1990 Census
- E.g. race, education level, poverty level

**Promotions**

- Giving history for last 24 mailings
- Lifetime giving statistics: total amount, frequency, first gift, average gift etc

# Propensity Scores

## UPSHOT: MAILINGS ARE NOT RANDOMIZED EXPERIMENTS

**Propensity Scoring** to verify randomized nature of the data

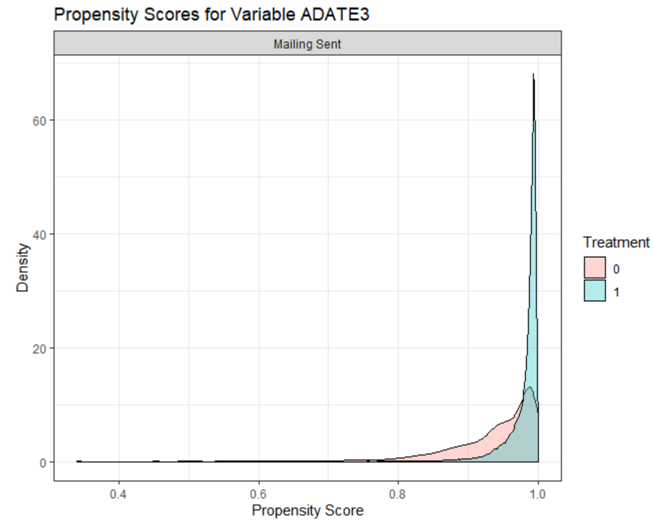
- Only conducted on 5 of 24 available mailings
    - ADATE3, ADATE7 examples shown on right
- Key assumption made: replicable mailing policy, such that results from one mailing can be cross-applied to all mailings)

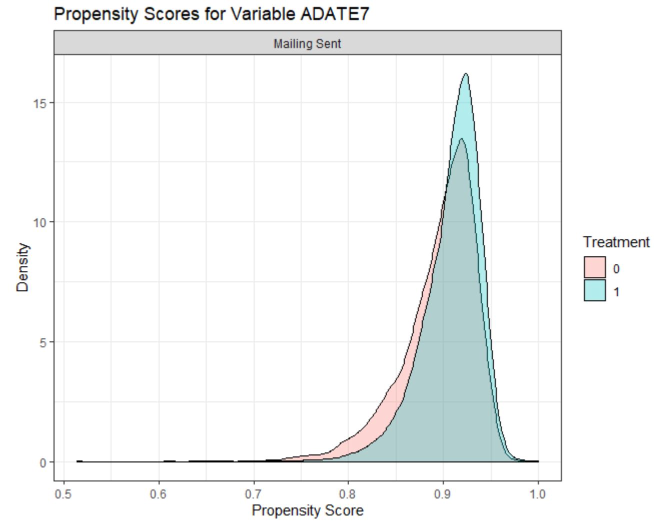
### Findings

- skew heavily towards 1 (Should be at 0.5)
- indicating non-randomization / possible selection bias

### Conclusion

- Current policy is a "blast mailing" policy
- Mail to nearly all without discrimination



# Covariate (Im)Balance

## UPSHOT: THERE IS NO CONTROL GROUP

**The “accidental” non-mailed minorities – can they be treated as a control?**

- Due to sheer amount of covariates (>500), only 10 were analysed

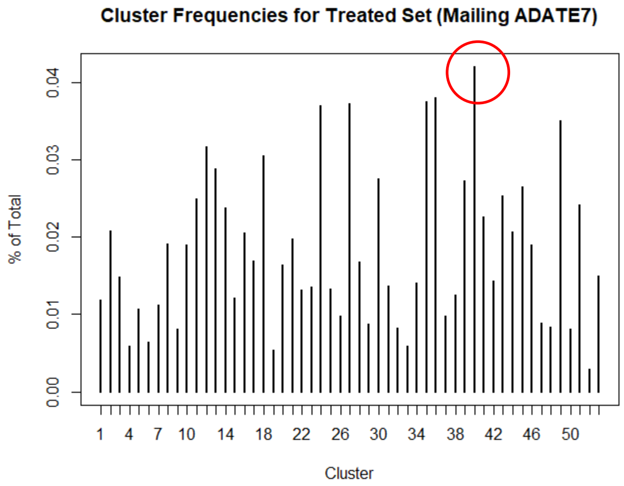
**Cluster as example (right):**

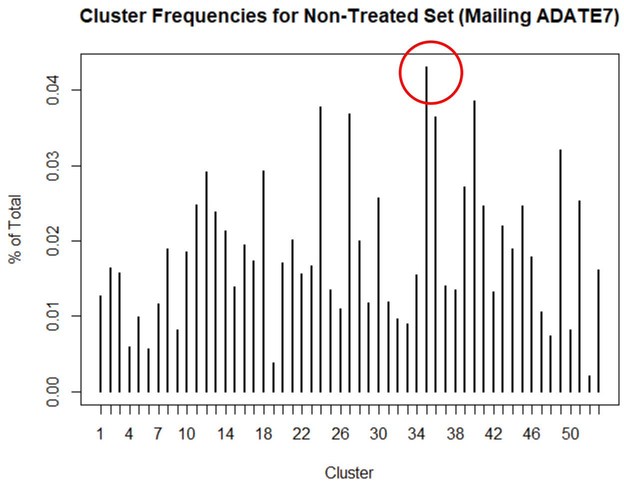
- Serves as needed **proxy** for weighted average of a few key variables:
- E.g. socioeconomic status, gender, geographical location, age
- Coded for and provided by PVA

**Findings**

- Differing % make-up of Treated and Non-Treated Set – <u>at the least, different modes</u>
- Similarly non-comparable results found for other tested variables
  - Income, Homeowners, Wealth (Socioeconomic)
  - Domain, PVA State, Age (Demographics)
  - Veterans, Fisher, Collect1 (Logged Interests)

**Conclusion: No**



# Defining a Target Measure

| LIFT | = | TARGET PROFITS | - | BASELINE |
|---|---|---|---|---|

## Target Profits

*Typically*

- For every customer: Spend* – Cost
- Choose to target N^ people out of population N, depending on targeting policy
- If Lift > 0, then the policy is good!

*For us*

- No meaningful difference, directly applicable for N^ donors targeted (N^ <= N)
- Cost = $0.68, flat rate for sending a promotion
- Spend* = Donation = $??
  - Value determined by donor
  - $0 if no response

**Profits = SUM(Donations)  - Cost * N^**

## Baseline

*Typically*

- For each customer: Spend (without marketing)
- **Assumes:**
- there are other methods to acquire the same product / service / item
- Control group exists upon which to make these estimates

*For us, **Key Assumptions Violated:***

- **no other way of donating except for campaign**
- **no control group exists**

Thus, baseline is status quo:

- Profits if everyone (N) was mailed

**Baseline = SUM(Donations) – Cost * N**

* Spend* refers to amount spent post marketing campaign. Presumably, spend* > spend.

# Final Profit Model

**Model Goal**

- Let $X$ be a random variable that represents whether a prospective donor donates.
  - $X = 1$ if the person is a donor
  - $X = 0$ if the person is NOT a donor.
- Let $Y$ be a random variable that represents the value of the donation given.
- Let $Z$ be the cost of mailing.
- We want to figure out the predicted donations for each donor of the joint distribution $(X, Y)$.

**Model Approach**

- Calculate $f_1(X)$ from our **logistic regression** model.
- Calculate $f_2(Y \mid X = 1)$ from our **linear regression** model.
  - In other words $f_2(Y \mid X = 1)$ translates to: given a set of people willing to donate, how much do you predict that they will donate.
  - Thus $f_2(Y \mid X = 0) = 0$ since people that do not donate give a total donation of zero.
- Subtract cost of mailing from predicted donations. We can do this because we inherently assume that people we do not mail to do not donate.

$$\textbf{Predicted Profits} = \sum_{x \in X, y \in Y} f(x, y) = f(y \mid x)\, f(x) - z$$

# Methods

# Feature Engineering

- **Problem:** The dataset is not clean - many values are missing, and most variables are non-binary, posing as difficulties in running predictive models.
- **Solution:** Cleaned data using helper functions, and dropped some variables out of >500 original variables

| REMOVED | IMPUTE TO MEAN | IMPUTE TO 0 | AS INDICATOR |
|---|---|---|---|
| **70 variables** | **68 variables** | **16 variables** | **4 variables** |
| Usual justification<br>• Insufficient data: over 50% missing<br>• Inconsistent data<br>• Redundant data | Impute missing values to mean of dataset. | Impute missing values to zero. | • Variable only has entries in a fraction of the data<br>• more important that the variable exists rather than the value |

**Data Splits:**

**Given**

- KDD provided 50% training data and 50% testing data.

**Modification**

- Split KDD training data into 80-20 training:val
- **Final: Train 40% Validation 10% Test 50%**

# “Quick Stepwise” Procedure Outline

**Main Problems – Exact methods**

- Running this dataset (90000 x 500) using CPU would take a long time for random forest using R. Logistic regressions do not converge.
- Cumbersome to run stepwise methods for both logistic and linear regressions for all predictors

*Need a <u>reasonable approximation</u> to the stepwise procedure in order to remove unnecessary predictors in our model.*

**Pseudocode for ”Quick Stepwise”**

**---**

**numtrials =** 5

**oldmodel =** regression on full model (without intercepts)

for trial in range(**numtrials**){

  **predictorset** Remove predictors that have p-values < 0.08

  **newmodel =** regression using predictors on **predictorset** (without intercept)

  **oldmodel = newmodel**

return **newmodel** }

**Why Quick Stepwise works**

**---**

- Many useless predictors are non-significant
- To be have p-values less than 0.08, 5 times in separate regression requires to predictor to be very significant.
- Reducing the p-value threshold below 0.08 or increasing number above 5 makes it too conservative dropping certain significant predictors that improve our model.
- Since our goal is prediction, we don’t mind having a few substandard predictors with a model with mostly significant predictors.

# Implementation

# Predicting Responses
## Logistic Regression Implementation

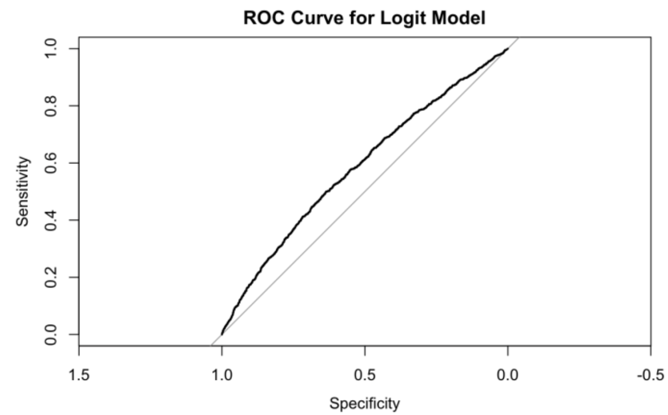
### Final Model Coefficients

```
Coefficients:
              Estimate
HOMEOWNRH   -9.423e-03
HOMEOWNRU   -1.161e-02
NUMCHLD     -3.407e-03
INCOME       2.743e-03
MBCRAFT      6.014e-03
ETH10        1.589e-03
HHD8         1.418e-03
HUR1        -2.963e-04
HUR2         2.934e-04
RHP2        -1.150e-03
TPE3         8.180e-03
TPE4        -8.611e-03
TPE5        -8.939e-03
TPE6        -9.312e-03
TPE12        5.251e-04
EC1          2.656e-04
VC4         -2.717e-04
HC21         5.026e-04
CARDPM12     1.362e-03
RAMNT_3      3.415e-03
RAMNT_8      5.479e-04
RAMNT_9      4.172e-04
RAMNT_12     2.265e-04
RAMNT_13     3.943e-04
RAMNT_14     6.511e-04
RAMNT_19     2.257e-04
RAMNTALL    -4.830e-05
NGIFTALL     9.365e-04
CARDGIFT     1.282e-03
MAXRAMNT     1.012e-04
LASTGIFT    -5.895e-04
---
```

After implementing ”Quick Stepwise” we got a result of **31 predictors** that incorporated data from the following:

- % of Homeowners
- Number of children
- Income
- Statistics on methods of transportation
- Median years of school completed by adults age 25+,
- % Veterans serving in last 23 years.
- % Houses with telephones
- Number of card promotions received in last 12 months
- Past promotion giving history

## AUC_ROC = 0.585



# Predicting Responses
## Logistic Regression Results

- We calculated the predicted logit values on the validation dataset and organized them by decile.
- We organized the logit values by decile and captured the amount of donors captured in our validation dataset.
- The % Donors captured generally decreases as we lower our threshold. This indicates stability in our model.

| Conversions by decile | % Donors Captured in Validation Set |
|---|---|
| Top 10% | 17.33% |
| Top 20% | 12.62% |
| Top 30% | 12.00% |
| Top 40% | 10.36% |
| Top 50% | 8.51% |
| Top 60% | 9.23% |
| Top 70% | 8.51% |
| Top 80% | 7.49% |
| Top 90% | 6.97% |
| All of the data | 6.97% |

# Predicting Donations
# Linear Regression

## Final Model Coefficients

```
Coefficients:
            Estimate
ETH14       0.235313
ETHC1       0.070619
EIC14       0.132659
AC2         0.101813
RAMNT_7     0.098910
RAMNT_8     0.057342
RAMNT_10   -0.073629
RAMNT_11    0.033000
RAMNT_12    0.006177
RAMNT_14    0.051580
RAMNT_17    0.095553
RAMNTALL    0.012203
NGIFTALL   -0.121073
MINRAMNT    0.180471
LASTGIFT    0.701332
---
```

After implementing ”Quick Stepwise” for our linear model we got **15** predictors that incorporated data from the following:

- Statistics on method of transportation
- Percent employed in educational services
- Communities with white people of age < 15
- Past giving history
- Percent adults age 60-64

## Optimal Linear Regression Fit

**Precautions taken**

- Removed predictors that were highly correlated and had high collinearity (if there was any)
- Limited correlation between residuals and predictors (checking residual plots)
- Removed outlier points. By checking studentized residuals.

**$R^2$ = 0.8166**

**Adjusted $R^2$ = 0.8159**

**Key Insight**

Note that in both regression models, there is a heavy reliance on <u>past giving history</u>.

# Profit Methodology

| Deciles | Logit Thresholds |
|---|---|
| Top 10% | 0.072 |
| Top 20% | 0.064 |
| Top 30% | 0.058 |
| Top 40% | 0.054 |
| Top 50% | 0.049 |
| Top 60% | 0.045 |
| Top 70% | 0.041 |
| Top 80% | 0.036 |
| Top 90% | 0.030 |
| All of the data | 0 |

**Logit Thresholds for Targeting**

- These logit probabilities indicate the probability that a given person will donate
- Our targeting scheme will target consumers who have a greater probability of buying above a certain threshold probability
- People in the top 10% are over twice as likely to buy than people in the bottom 10%!

## Profit Formula Used

| Variable | Description |
|---|---|
| LP | Logit Probabilities |
| D | Predicted Donations |
| T | Targeting indicator<br>(if LP > threshold then 1 else 0) |
| C | Cost of mailing promotion |

**Profit = T * (LP * D - C)**

# Profit Results on Validation Set

### Profit Predictions on Validation Set

| Logit Score Thresholds | Incremental Profits by Decile | Profits Including Decile and Up |
|---|---|---|
| Top 10% | $649.30 | $649.30 |
| Top 20% | $478.17 | $1127.26 |
| Top 30% | $360.06 | $1487.20 |
| Top 40% | $365.79 | $1852.96 |
| Top 50% | $336.84 | $2189.65 |
| Top 60% | $298.00 | $2487,77 |
| Top 70% | $170.86 | $2658.28 |
| Top 80% | $90.89 | $2749.35 |
| Top 90% | - $8.55 | $2740.78 |
| All of the data | - $308.19 | $2365.94 |

### Validation Results

- Our model answers the question by evaluating people who are NOT our donors, by taking the top 80% of logit scores
- Through this approach we get excellent results in our validation set
- This exercise has helped us choose our threshold for the testing dataset

| | |
|---|---|
| Actual Profits | $2664.74 |
| **Predicted Profits** | **$2749.35** |

# Profit Results on Testing Set

- While we underestimated the baseline profits, our net profits were very accurate. This is very encouraging because we replicated the same results form our validation set.

| | KDD Testing Set* Actual Results (provided by organization) | Our Predicted Results |
|---|---|---|
| Net Profits | $14,712 | **$14,468** |
| Baseline Profits if entire population is targeted | $10,560 | **$12,329** |
| Increase in Net Donation | $4,152 | **$2,139** |
| % Increase | 39.32% | **$17.35%** |

*Achieved similar results as organizing committee despite utilizing much simpler models than winners*

:

* https://www.kdnuggets.com/meetings-past/kdd98/gain-kddcup98-release.html (KDD Actual figures taken from the winner's site.) :

# Monitoring & Adaptation

*Since we did not work directly with the client, these are recommendations only*

# Monitoring

## Model Evaluation

***Evaluate the model yearly***

- As new mailings are sent out every year
- Crucial, as giving history (RAMNT_??) variables tend to be predictive of future giving behaviour
- New donors can be sourced and added to the dataset with every mailing

***Possible Adjustments***

- Probability thresholds from logistic regression for deciles
- Ensure that all deciles that yield donations are profitable.

## Tracking Model Performance

***Key indicators***

- **AUC_ROC:** measuring logistic regression model fit
- **Capture rate by decile.** The more skewed it is towards higher deciles, the better the model.
- **Adjusted R^2** Of those who gave, were the amounts comparable with predictions?
- **Linear regression assumptions.** If it does not, we have to incorporate a different machine learning framework. Linear models are powerful only if they obey the assumptions.

# Adaptation

## New Data Elements

***New Mailing Types***

- Rise of digital / non-print advertising campaigns in philanthropy
- “Premiums” sent with the mailing campaigns may change
- Response variables remain unchanged

***New Data Variables***

- Online footprint variables
- Old interest variables no longer predictive

***Incorporating / Selecting New Predictors***

- Model automatically does this whenever re-run through the step-wise procedure
- Provided data is coded correctly

## New Campaign Types

***May need to consider a new baseline if***

- Donors are able to give passively without need for targeted campaigns
- Why? Because the “blanket policy” baseline will now overestimate profits from targeting

***How to calculate a new baseline***

- Run a randomized experiment to create a control group
- Duration: a year, or the usual time window for responses to a campaign
- Establish baseline donation amount that does not require cost of targeting and promotion

# Limitations

## Dataset

- **Many variables either had dirty or insufficient data**
  - Data was not consistently coded, and we did not have a way of inferring what certain entries meant
  - Many variables had over >50% blanks, could not possibly impute the rest
- **Some of these variables may have been predictive if coded accurately**
  - Gender
  - Wealth
- **Some indicator variables are predictive but not defensible**
  - E.g. CLUSTER – each cluster has unique socioeconomic, interest and demographic characteristics. Some cluster numbers were predictive, but the assignment of clusters was not known to us

## Model

- **Limited cost information affecting profitability calculus**
  - PVA has different types of mailings, and they presumably have an effect on responses as well as cost
  - Differentiated cost information was not given, and we were forced to treat all mailing types as the same
- **May cause donor pool shrinkage**
  - **The catch:** All who are not targeted in the upcoming campaign will definitely not donate.
  - **Feedback loop:** if you don’t donate this time (whether not targeted, or no response), model unlikely to target you again
  - **Mitigation:** must constantly add new donors from new sources

- *

# Alternative Approaches

## Classification

**We could have tried multiple methods and pick the one with the best accuracy:**

- Random Forest / Ensemble Methods / Decision Tree
- Support Vector Machines

## Clustering

- **K-means to discover heterogenous relationships.** Since there was limited heterogeneity, we could have done K-means and trained separate regression models for each cluster, giving us a more nuanced result.

## Regression

- **Forward Stepwise Method in choosing predictors.**
- **Selective Inference** (to choose predictors judiciously). Using this framework, we can use the same predictors across multiple mailings because of the strength of significance.
- **Regression Trees** (in case linear model assumptions do not hold.)
  - In our model we faced a mild heteroskedasticity problem. In other mailings we cannot guarantee the validity of the linear model. Regression trees can help us get out of this problem.

# The Bottom Line

1. **Modelled philanthropy as a joint distribution of 2 utility functions**
   - Give / Not Give
   - How Much

2. **Gain by saving on cost:**

   by not mailing those who would not have given (omitting bottom 20%)

**$14,468**

***Raw Profit***

**$2,139**

***Above Baseline***