



EMPOWERING GENEROSITY: THE STRATEGIC EDGE OF DATA IN CAMPAIGNS

DONORS

TABLE OF CONTENT



1. A detailed overview of our donor base: characteristics.etc
2. Effectiveness of the model? : predictions vs random chance
 - Lift, Cumulative Gains
3. Benefits for Stakeholders
4. Re-activation campaign
 - Cost savings or increased donations by using the model



DonorID Cleaning

- Removing 'D' from donorID values
- Converting donorID type to integer

SELECTION CAMPAIGNS



Data Type Conversion for Columns

- Converting 'date' column type to datetime
- Converting 'letterSent' column type to integer
- Converting 'CostUnit' column type to float

Feature Engineering

- Deriving the total cost of each campaign
- Extracting Year, Month and Day into separate columns
- Adding Season and Quarter columns

CAMPAIGN



Cleaning Region and Province Columns

- Replacing Region and Province missing values with the most frequent values when Language is equal to English

Cleaning Gender Column

- Filling null values with “gender not mentioned”

DONOR



Cleaning ZipCode Column

- Replacing ZipCode zero values with the most frequent ZipCode when Region is equal to 'Wallonia' and province is equal to 'Walloon Brabant'
- Replacing ZipCode null values with the most frequent ZipCode when Region is equal to 'Flanders' and province is equal to 'Flemish Brabant'

Cleaning DateBirth Column

- Converting "dateOfBirth" column type to dateTime

DONOR



Feature Engineering

- Applying One-Hot Encoding on region, province, gender and language columns
- Extracting 'yearOfBirth', 'MonthOfBirth' and 'dayOfBirth' from 'dateOfBrith' column
- Adding 'Age' and 'AgeGroup' columns based on the current date

DONOR



- **Removing Duplicates**
- **Cleaning Gift Table Columns**
- **Feature Engineering**
- **Grouping By donorID to have unique donor Ids**

GIFT



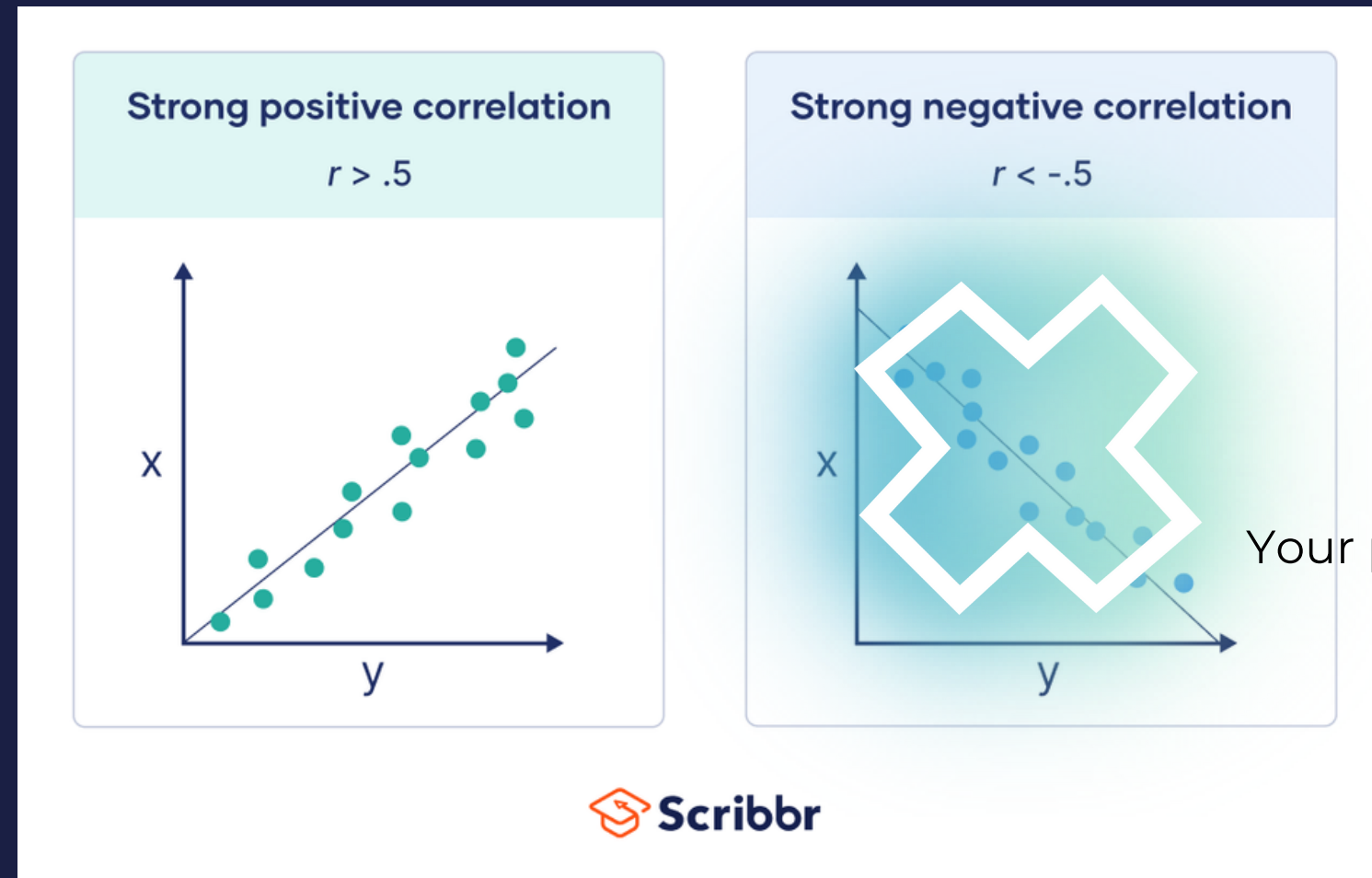
Grouping By donorID to have unique donor Ids

- Calculating total, average, minimum and maximum donation amount per donor
- Extracting first and last dates of donations
- Calculating number of campaigns where a donor donated

GIFT



Feature selection



['R_is_Flanders',
'R_is_Wallonia',
'P_is_East_Flanders',
'P_is_Hainaut',
'P_is_Limburg',
'P_is_Walloon_Brabant',
'is_Gender_Not_Mentioned',
'is_NL_speaker',
'is_FR_speaker',
'year_of_birth',
'age_2018',
'age_2019',
'age_2020',
'age_group_2018',
'age_group_2020',
'total_donation_amount',
'nb_of_donations',
'average_donation_amount',
'maximum_donation_amount',
'nb_of_campaigns']

PROVINCE

REGION

AGE

GENDER AND
LANGUAGE

DONATION DETAILS

PEARSON'S
CORRELATION
COEFFICIENT



12. Run all the models for the test_basetable

+ Code + Markdown

```
# Initiate the models
```

```
tree      = DecisionTreeClassifier(class_weight='balanced')
logistic   = LogisticRegression(solver = "lbfgs", max_iter = 800, class_weight='balanced')
randomForest = RandomForestClassifier(n_estimators = 100, class_weight='balanced')
boostedTree = GradientBoostingClassifier()
svm        = SVC(gamma = "scale", probability = True, class_weight='balanced')
neuralNet   = MLPClassifier()
neighbors   = KNeighborsClassifier()
```

[87]

```
models = {"tree"      :tree,
          "logistic"   :logistic,
          "randomForest":randomForest,
          "boostedTree":boostedTree,
          "svm"        :svm,
          "neuralNet"  :neuralNet,
          "neighbors"  :neighbors
}
```

[88]

Explanation

1. Decision Tree (tree):
 - characteristics.
2. Logistic Regression (logistic):
 - probabilities of donors taking specific actions, such as donating.
3. Random Forest (randomForest):
 - working together, like a team
4. Boosted Tree (boostedTree):
 - learns from its mistakes
5. Neural Network (neuralNet):
 - recognizing complex patterns in donor behavior.
6. Support Vector Machine (SVM):
 - set a boundary
7. Neural Network (neuralNet):
 - recognizing different shades in a picture.
8. K-Nearest Neighbors (neighbors):
 - like neighbors on the same street.

MODELS



Effectiveness of the model?

AUC

CONFUSION MATRIX

CUMULATIVE GAINS
CHART

ROC

LIFT CHART

CUMULATIVE GAINS
CHART



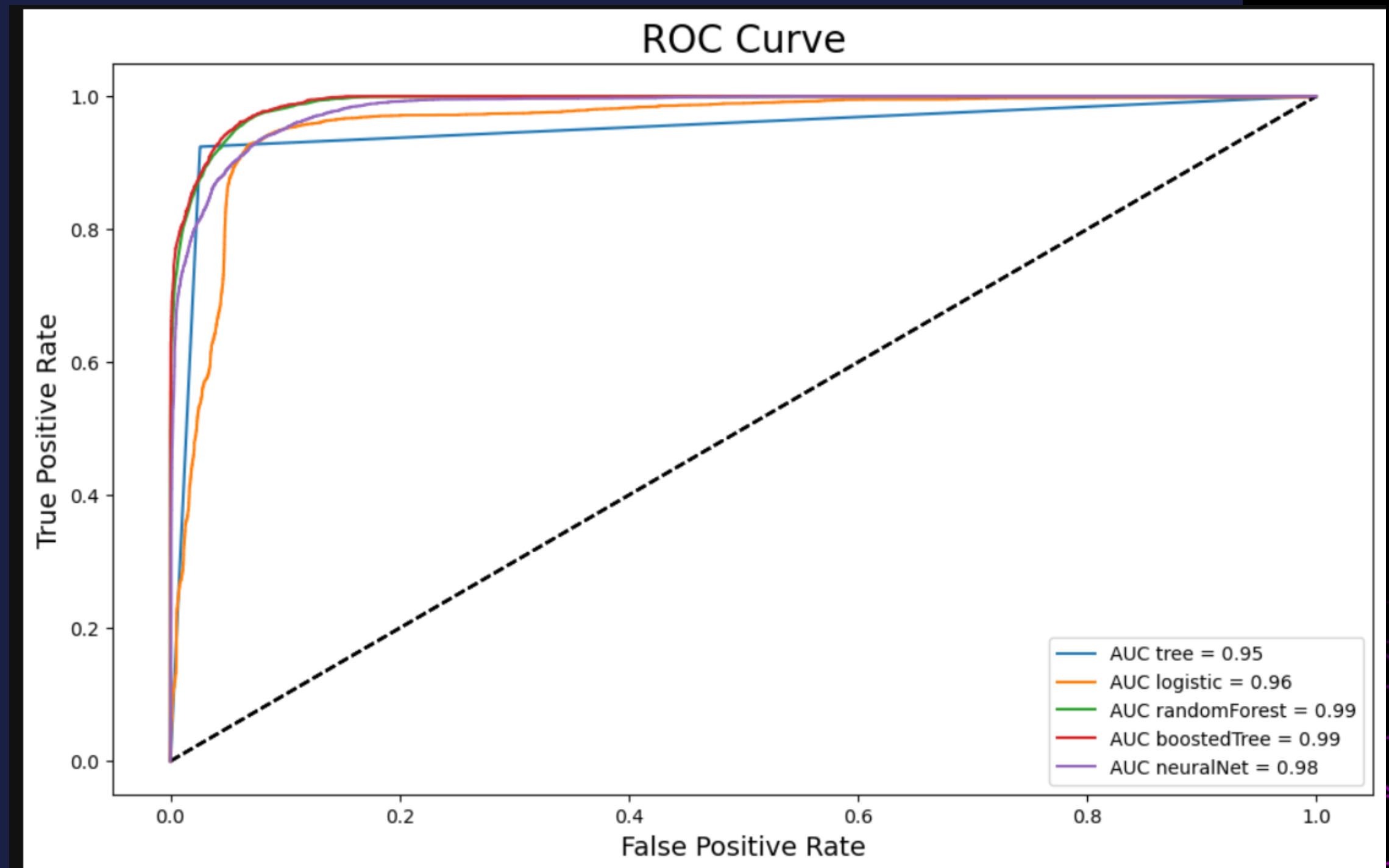
AUC and Accuracy for each model

	logistic	randomFo rest	boostedTree	boostedTree	neuralNet
Accuracy	0.962956	0.912654	0.953402	0.953402	0.925327
AUC	0.948706	0.958740	0.991316	0.991316	0.976460

DATAFRAME(PERFORMANCES)

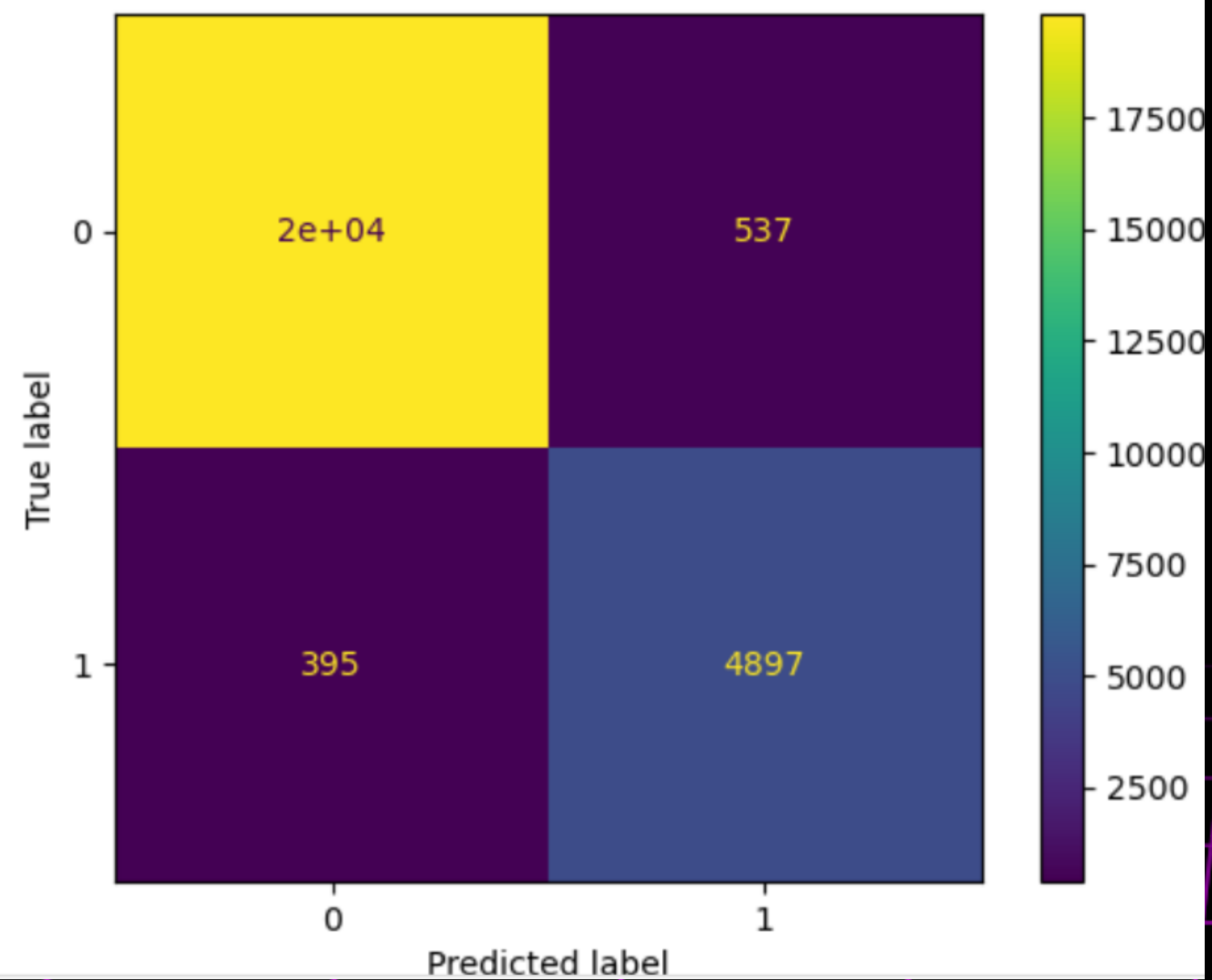
RECEIVER OPERATING CHARACTERISTIC CURVE

AUC Score for tree 0.95
AUC Score for logistic 0.96
AUC Score for randomForest 0.99
AUC Score for boostedTree 0.99
AUC Score for neuralNet 0.99



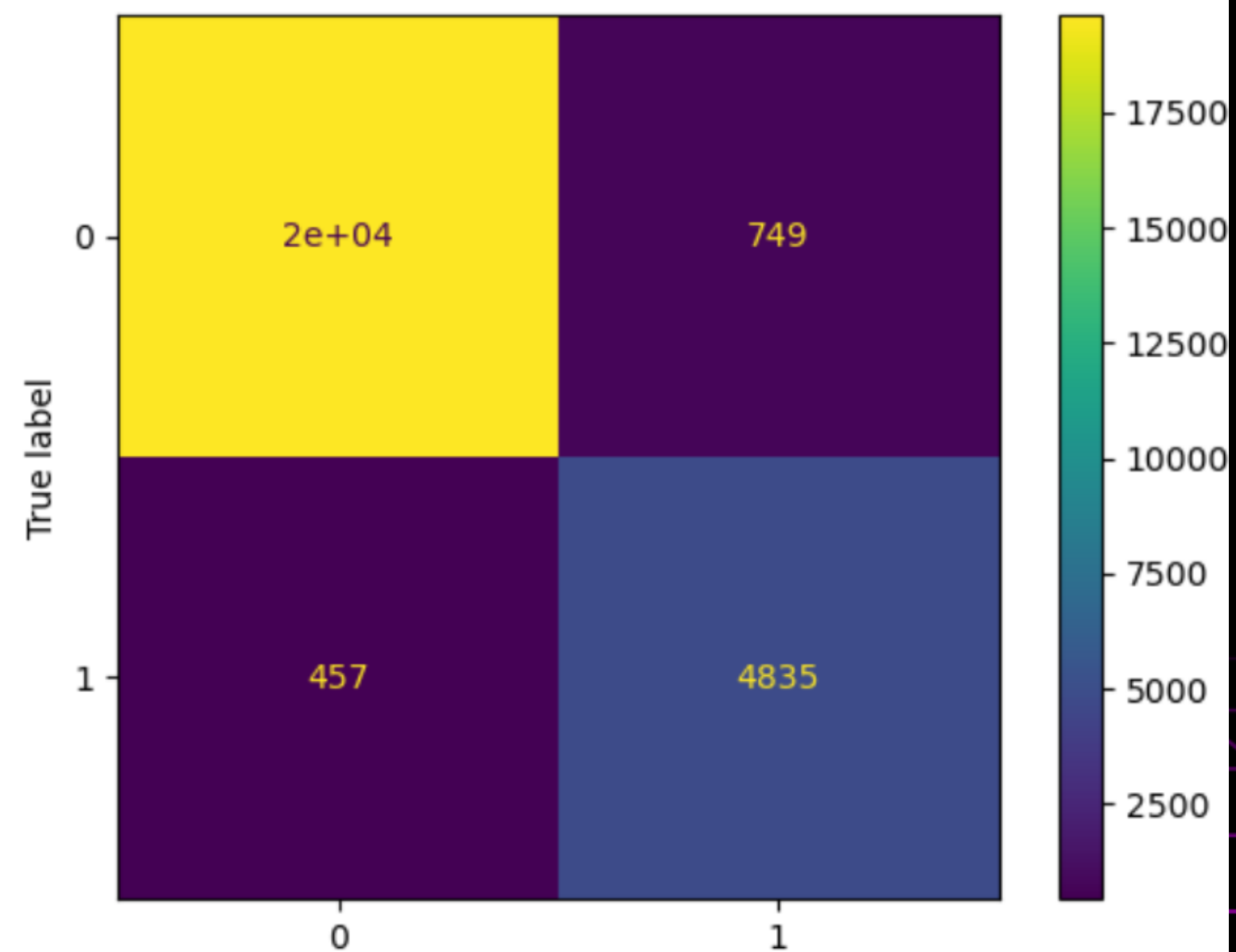
99.97992773986351
96.36576330668747

Confusion Matrix - tree



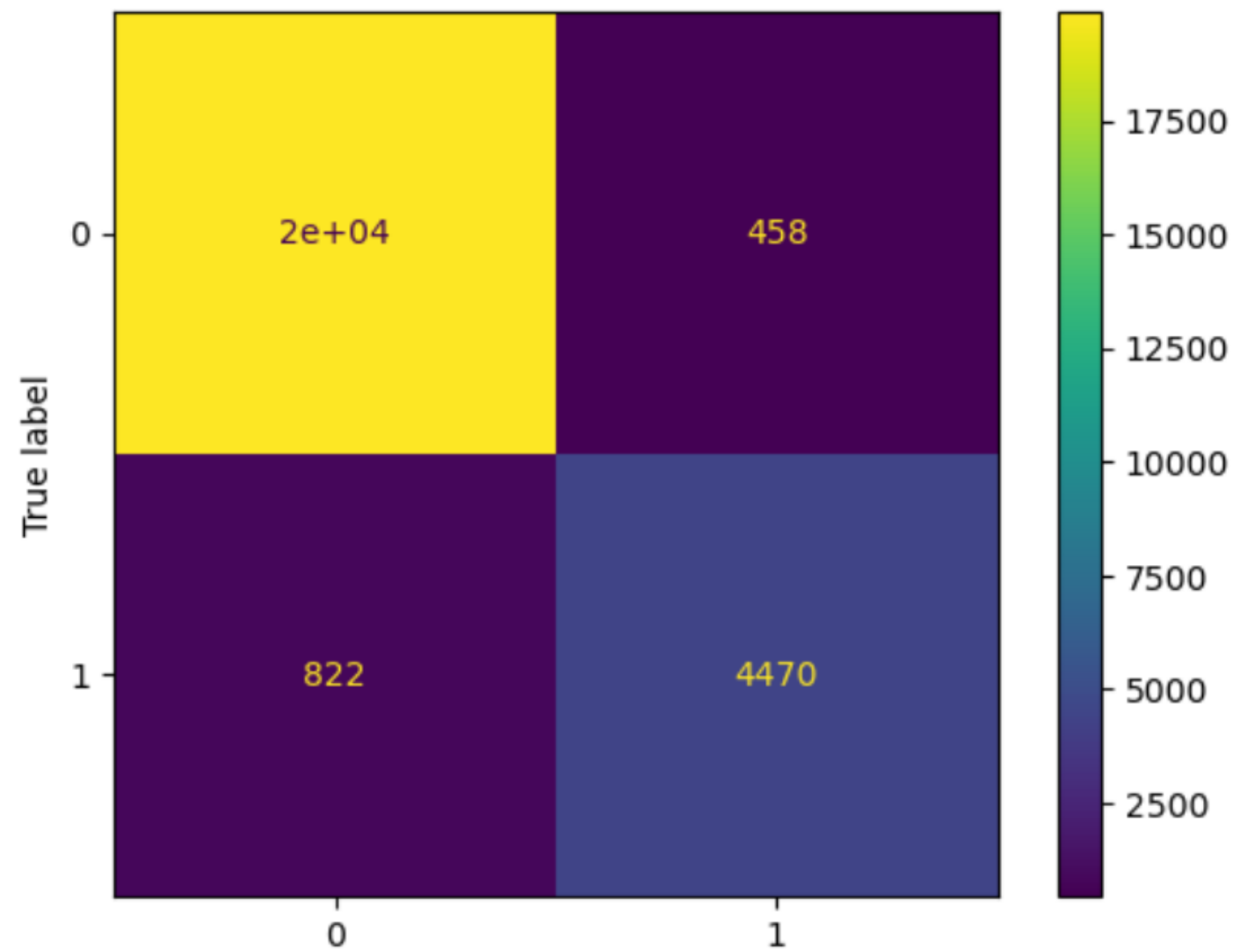
99.97992773986351
95.29732891401834

Confusion Matrix - randomForest



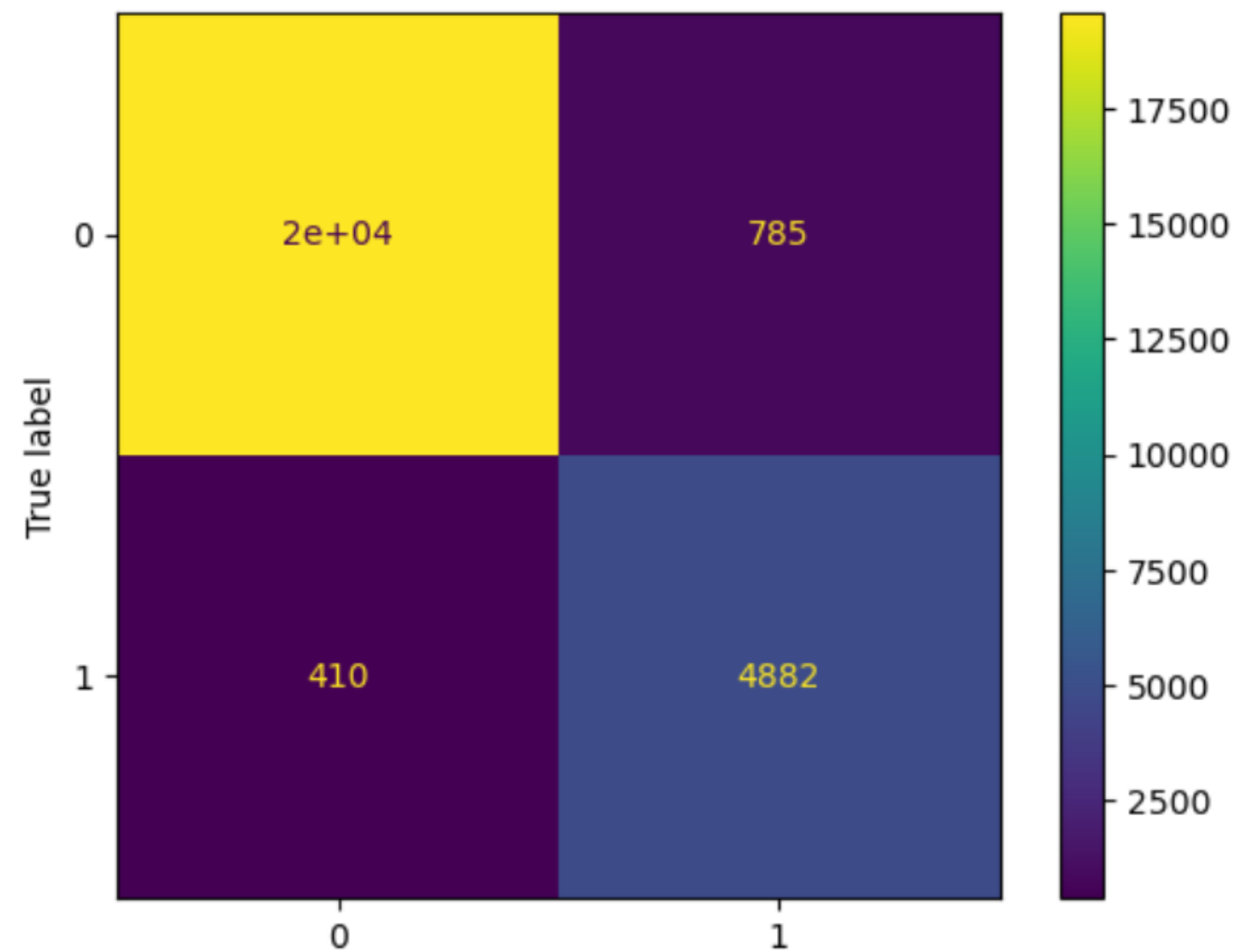
94.44858633939324
95.00877364008579

Confusion Matrix - neuralNet

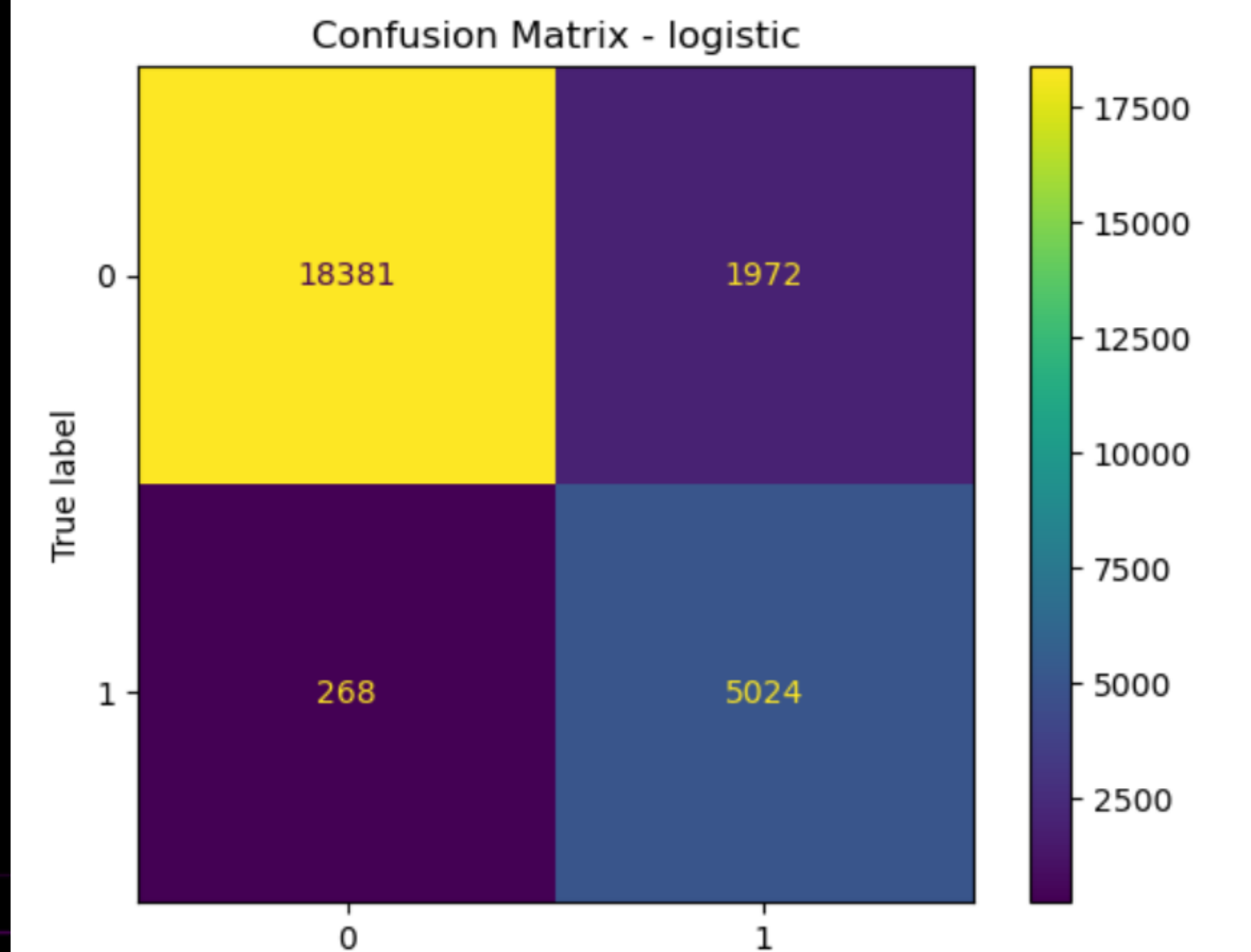


95.58697023570568
95.34022226554883

Confusion Matrix - boostedTree

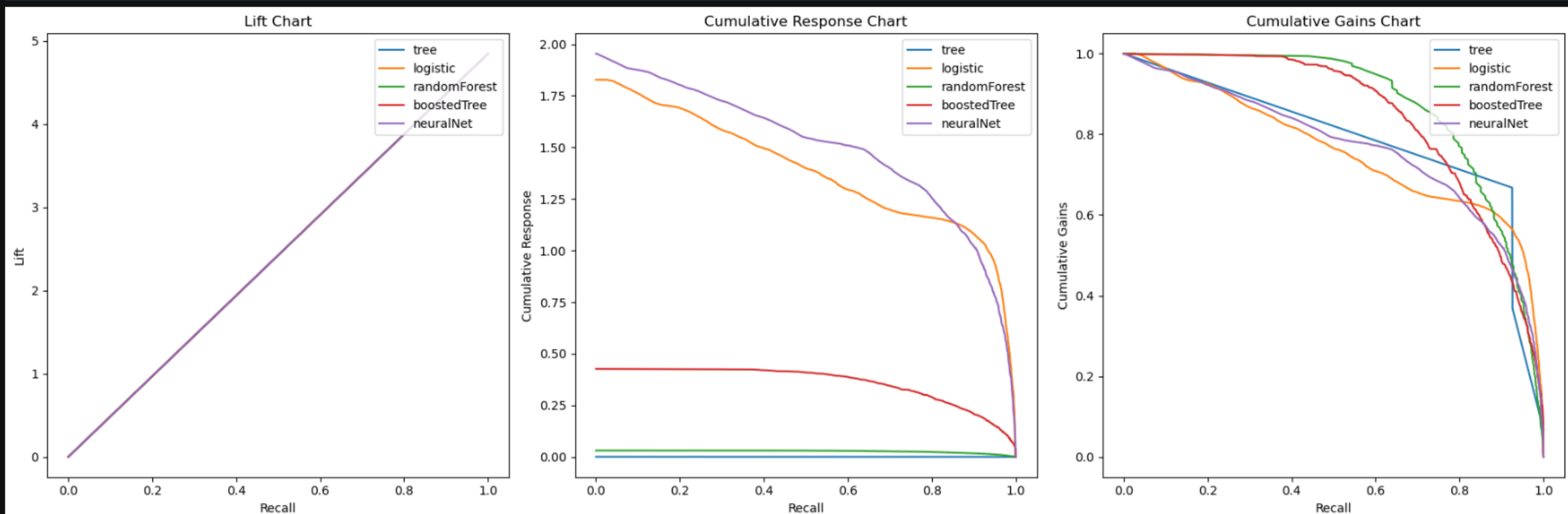


92.59620347536847
91.26535387015012



Better sensitivity

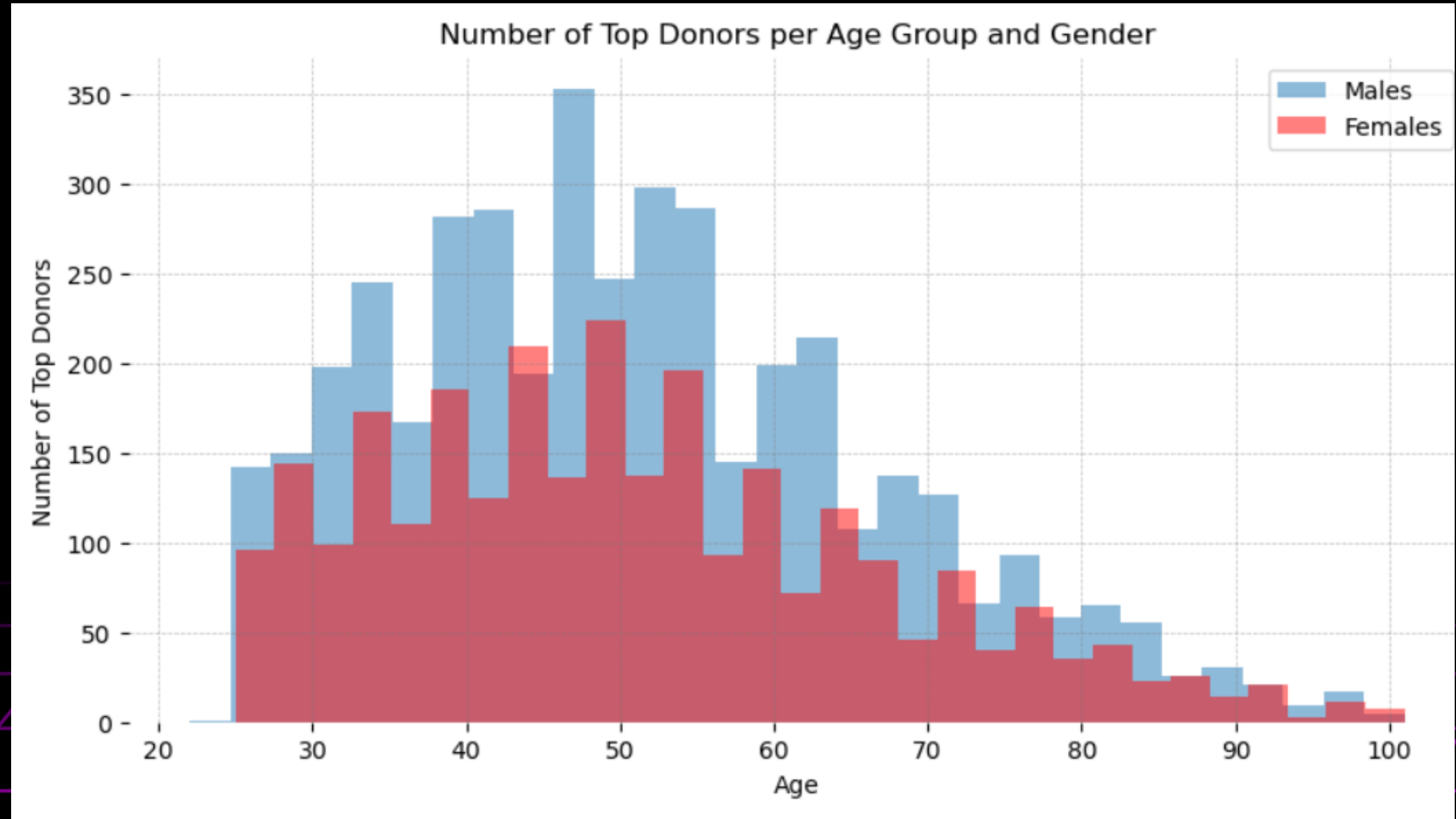
Operative and Resources Optimization

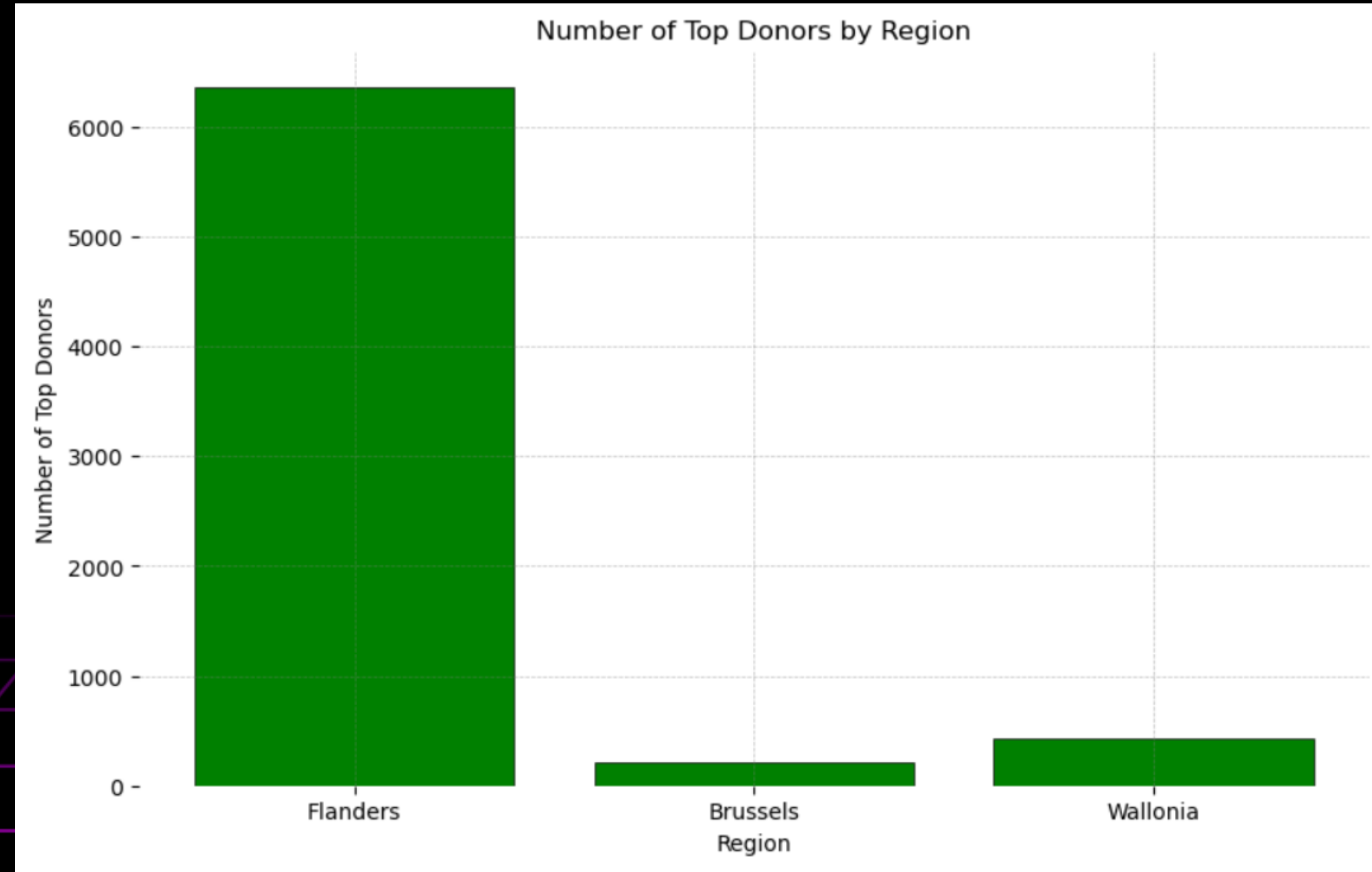


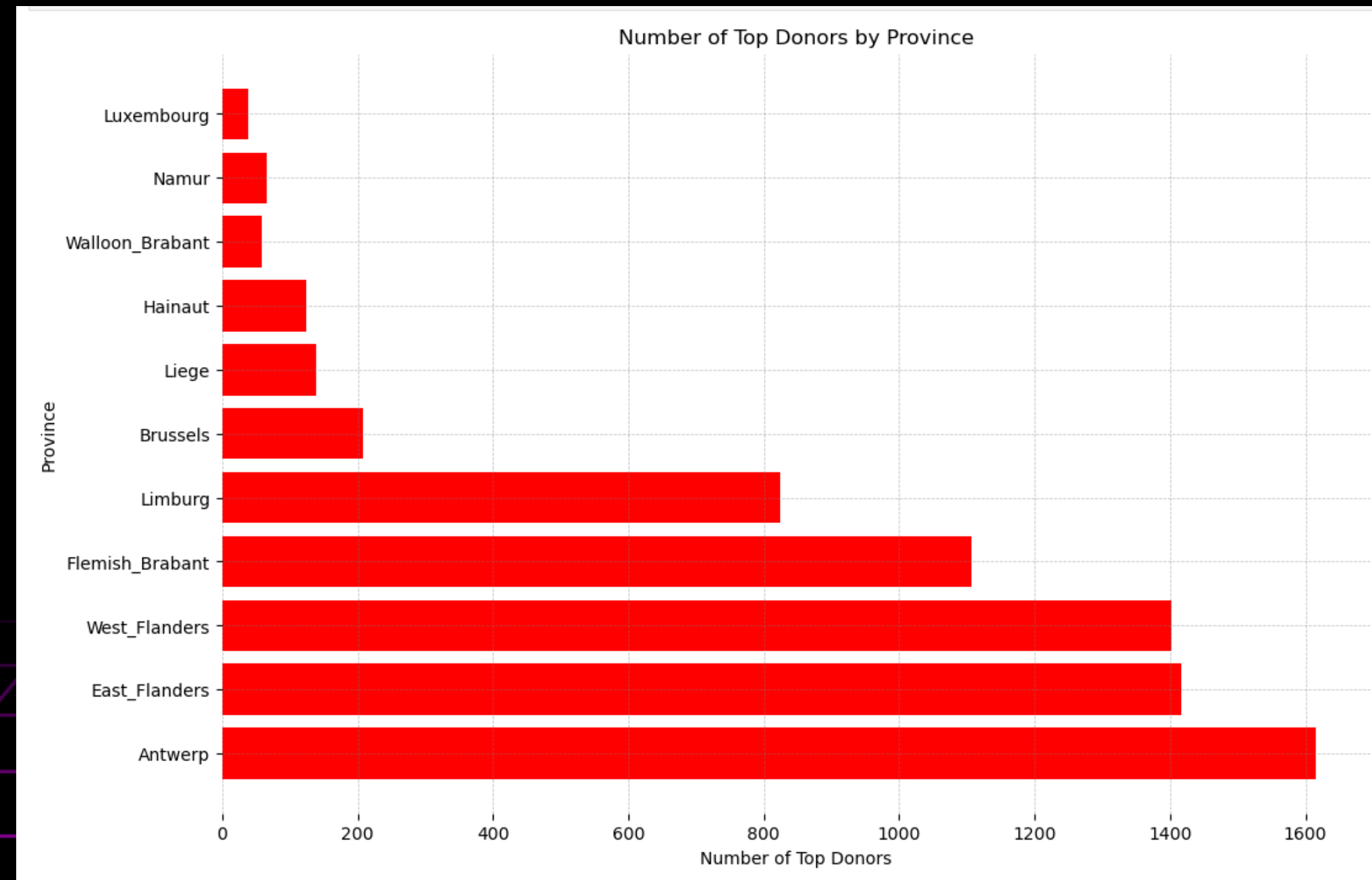
LIFT CHART

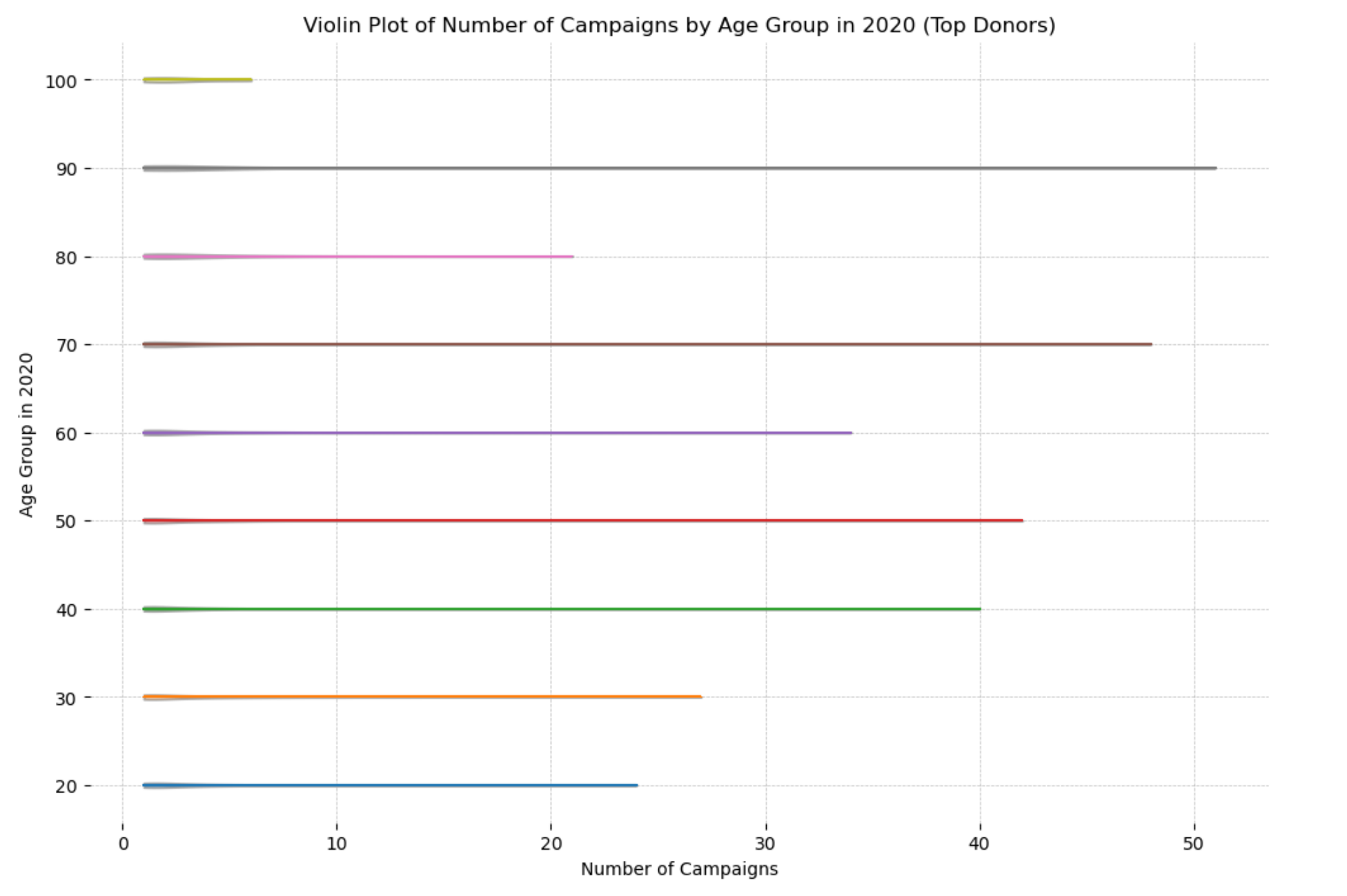
CUMULATIVE
RESPONSE CHART

CUMULATIVE GAINS
CHART











Donor Profile

- The majority of donors are between the ages of 30 and 60, and they participate in over 30 initiatives.
- Male are more likely to donate than female.
- The vast majority of contributors come from Flanders.
- Regions of Antwerp, East Flanders, and West Flanders (above 1300).

PROFILING



Donation Profile

- 27.28% of donors are most likely to donate.
- 5596 euros, which will be the total cost of the campaign (€0.80 per letter).
- To be expected is a total donation of 209,880 euros.

PROFILING



THANK YOU