Introduction and basetable structure

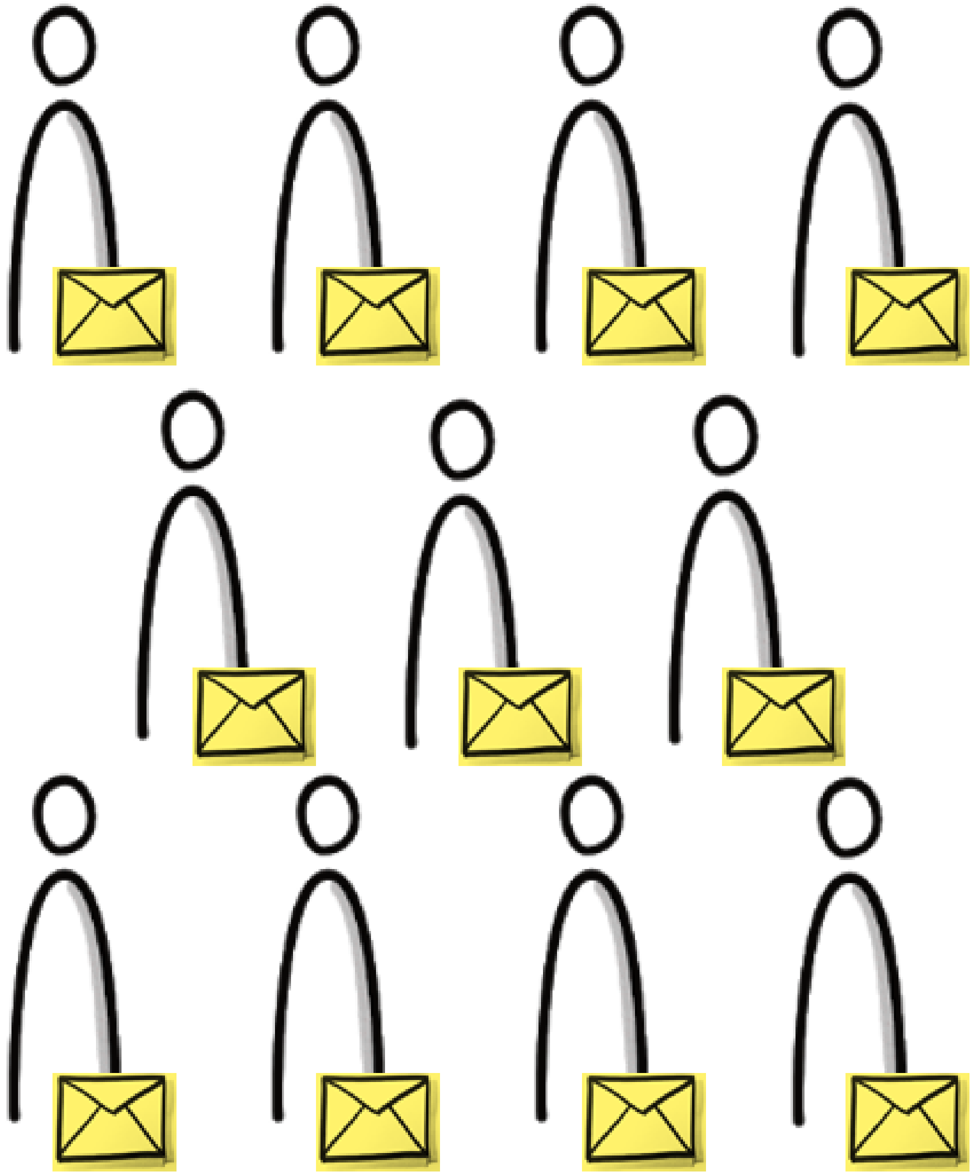
INTRODUCTION TO PREDICTIVE ANALYTICS IN PYTHON

Nele Verbiest, Ph.D

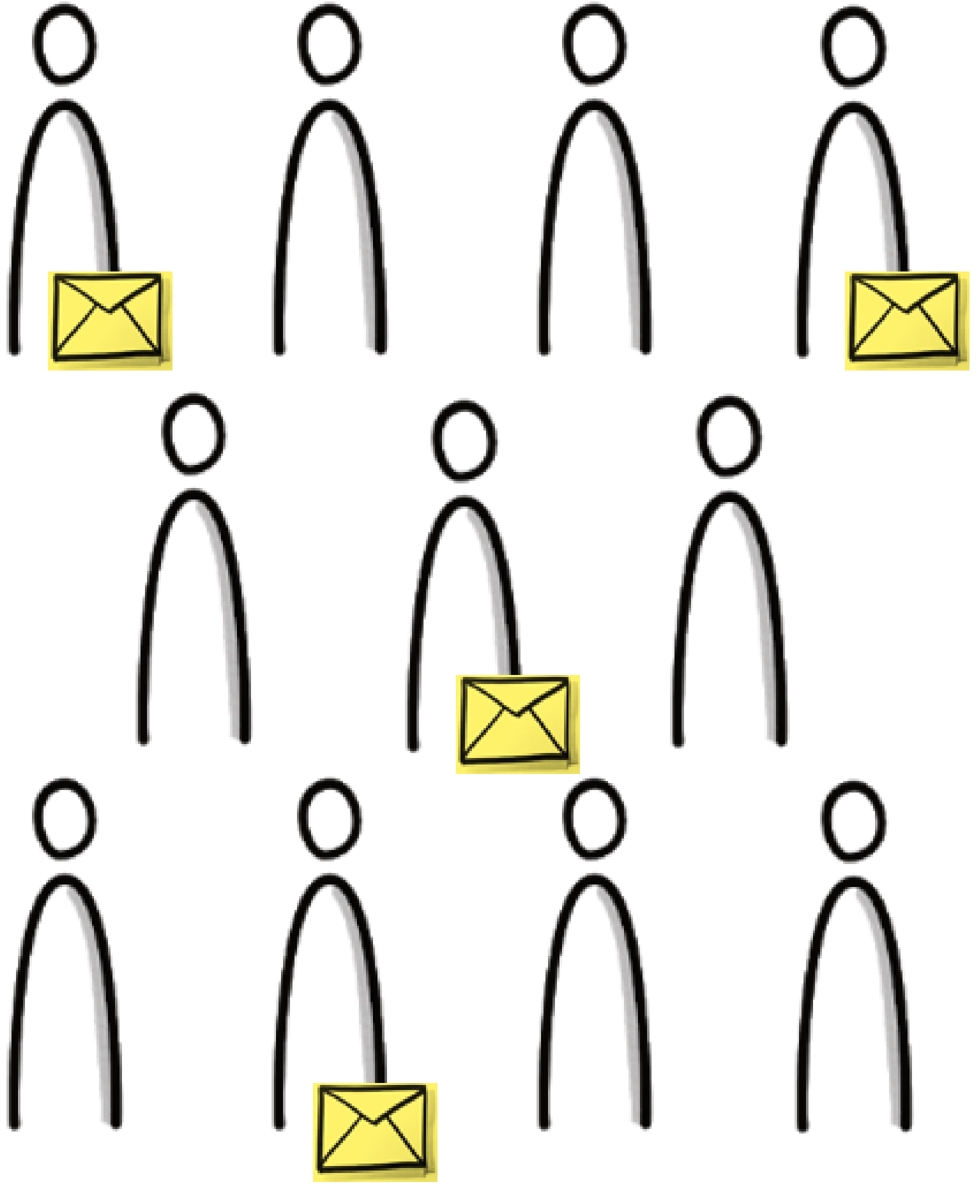
Data Scientist @PythonPredictions

Predictive analytics in fundraising

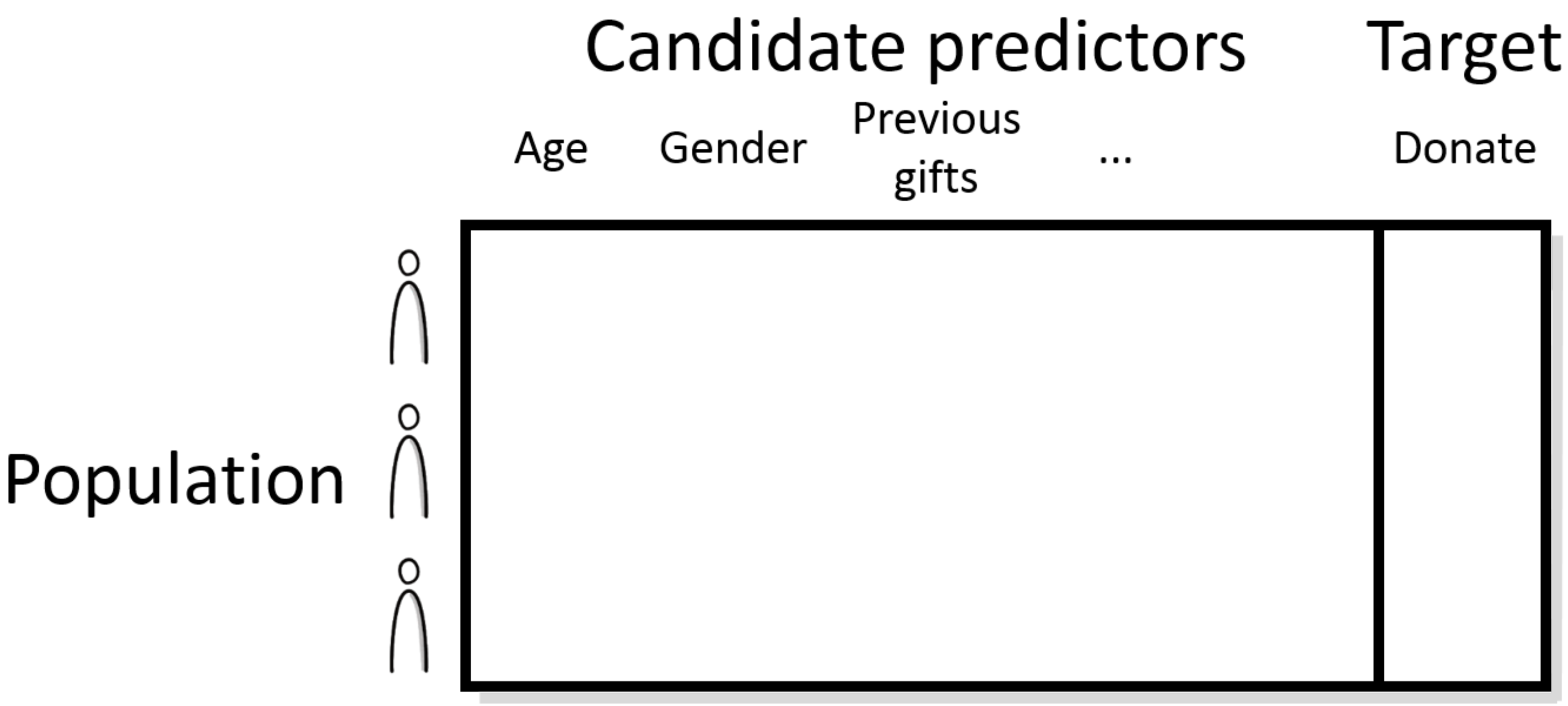
Address all donors



Address donors most likely to donate



The analytical basetable



Candidate predictors

Target

Age

Gender

Previous gifts

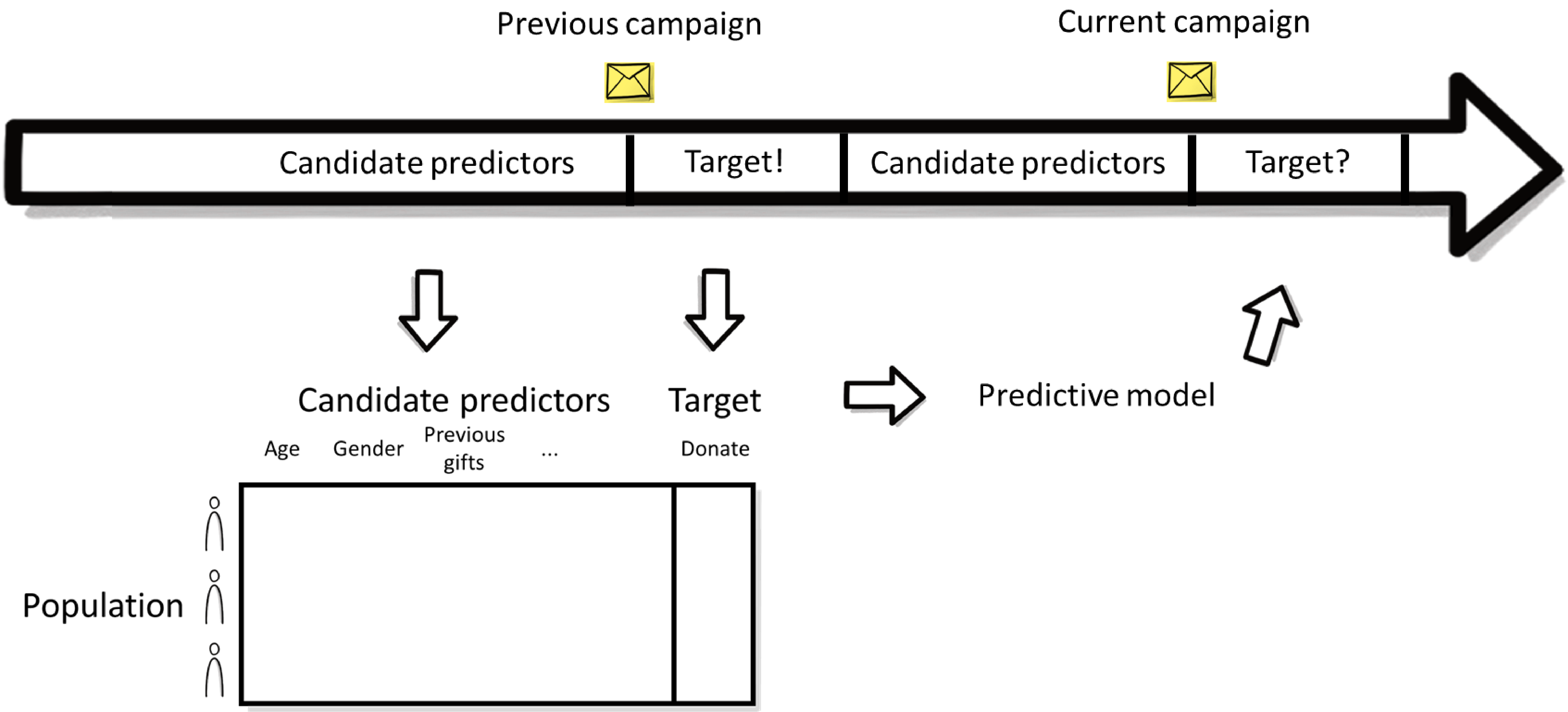
...

Donate

Population

```
import pandas as pd
basetable = pd.DataFrame("import_basetable.csv")
population_size = len(basetable)
targets = sum(basetable["Target"])
```

The timeline



Let's practice!

INTRODUCTION TO PREDICTIVE ANALYTICS IN PYTHON

Logistic regression

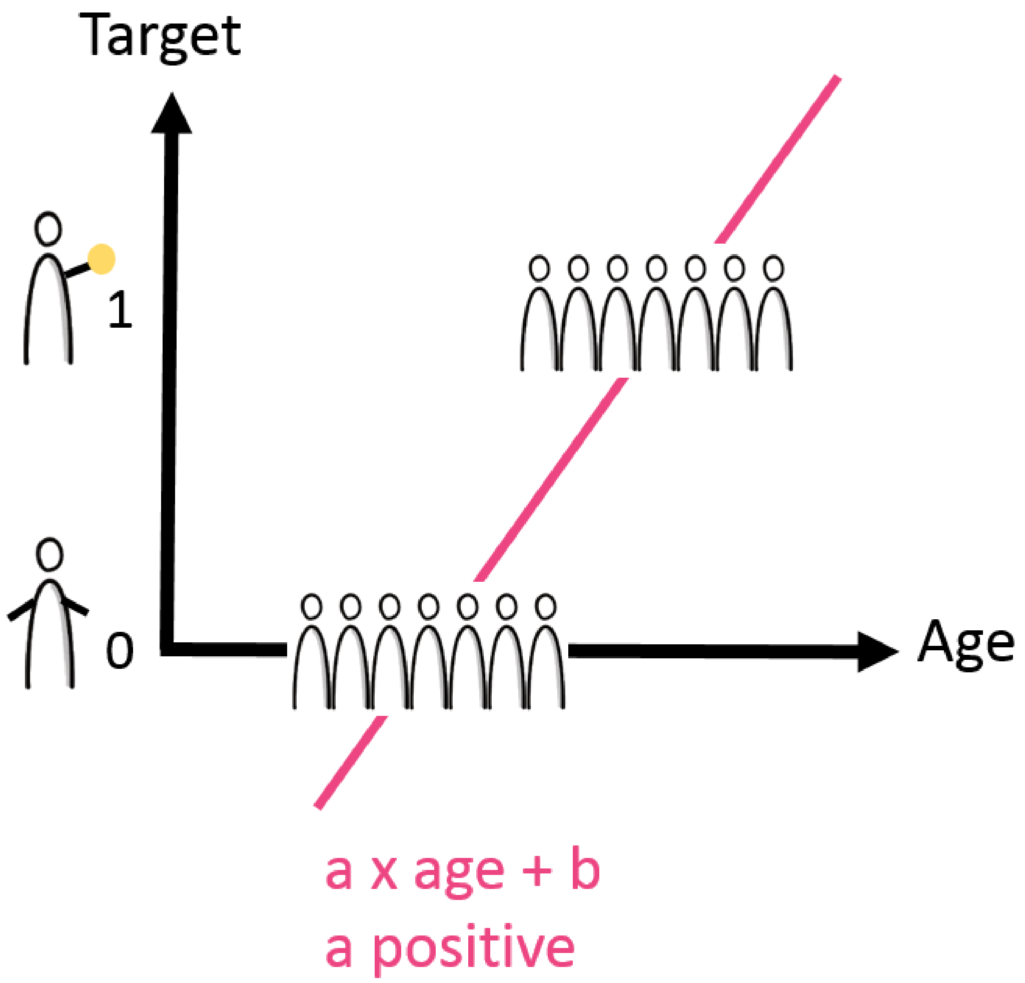
INTRODUCTION TO PREDICTIVE ANALYTICS IN PYTHON

Nele Verbiest, Ph.D

Data Scientist @PythonPredictions

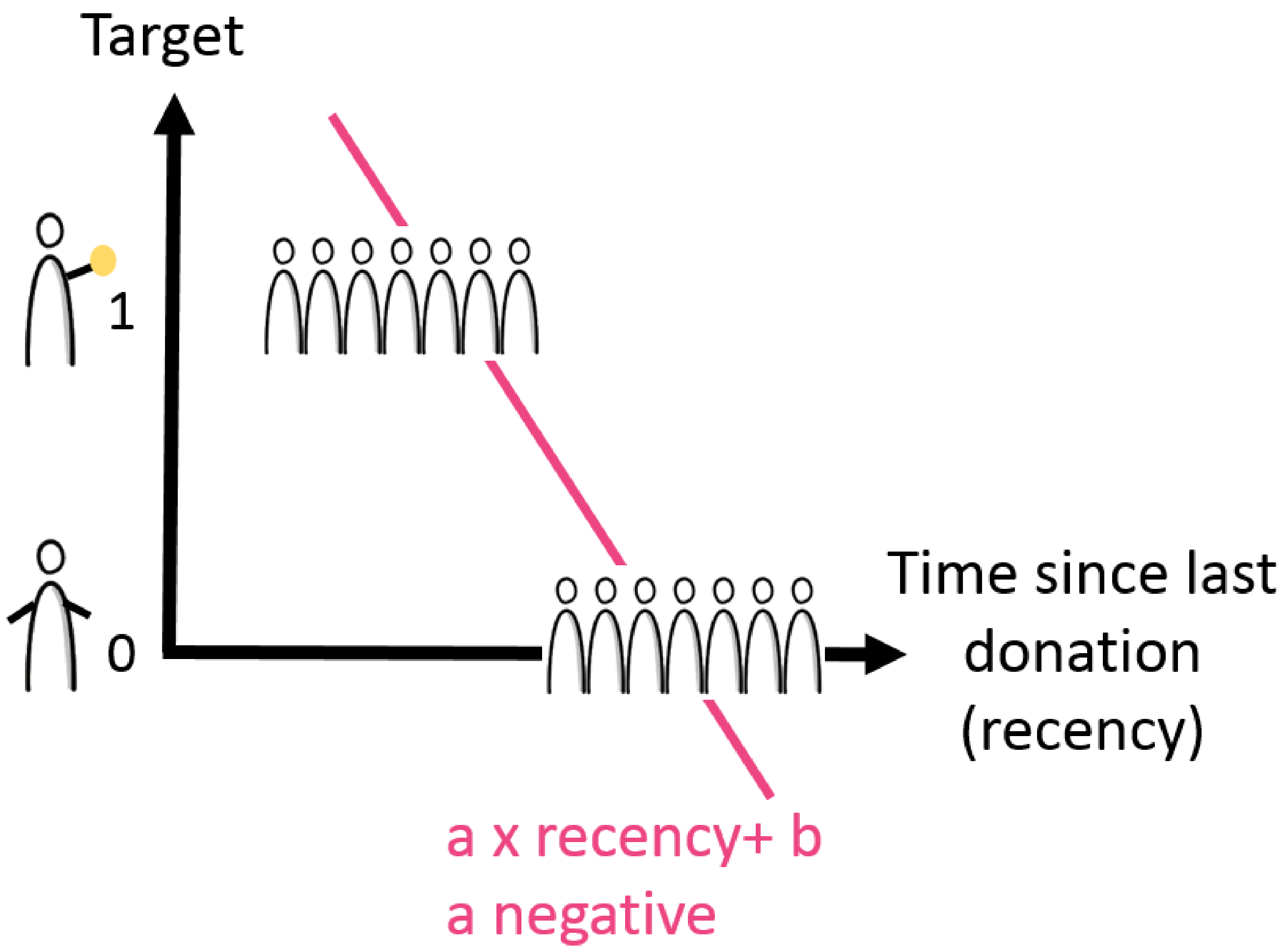
Logistic regression: intuition

Older people are
more likely to donate



a x age + b
a positive

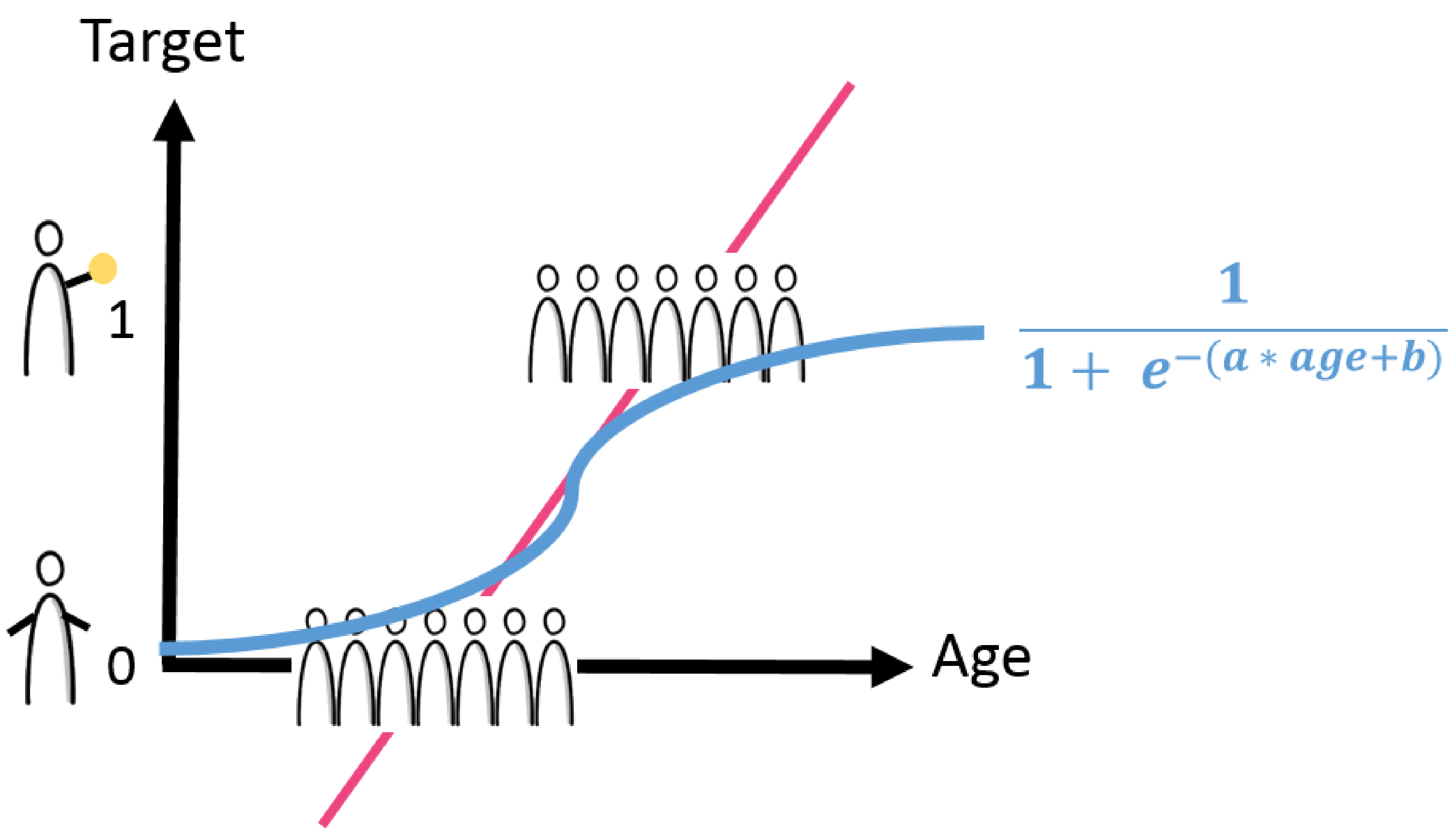
People who donated
long time ago are less
likely to donate



a x recency+ b
a negative

Logistic regression: the logit function

- Output of $a * age + b$ is a real number
- We want to predict a 0 or a 1
- Logit function transforms $a * age + b$ to a probability



Logistic regression in Python

```
from sklearn import linear_model
logreg = linear_model.LogisticRegression()
X = basetable[["age"]]
y = basetable[["target"]]
logreg.fit(X,y)
print(logreg.coef_)
```

```
[[ 0.02449202]]
```

```
print(logreg.intercept_)
```

```
[-4.3299131]
```

Multivariate logistic regression

Univariate: $ax %20 b$

Multivariate: $a_1x_1 %20 a_2x_2 %20 ... %20 a_nx_n %20 b$

```
X = basetable[["age","max_gift","income_low"]]
y = basetable[["target"]]
logreg.fit(X,y)
print(logreg.coef_)
```

```
[[ 0.0243308   0.03906065 -0.76793773]]
```

```
print(logreg.intercept_)
```

```
[-8.80643545]
```

Let's practice!

Using the logistic regression model

INTRODUCTION TO PREDICTIVE ANALYTICS IN PYTHON

Nele Verbiest, Ph.D

Data Scientist @PythonPredictions

The logistic regression function

```
0.545 * gender_F
+ 0.021 * age
-0.001 * time_since_last_gift
-3.39
```

- Female (gender_F=1)
- age 72
- 120 days since last gift

```
0.545 * 1
+ 0.021 * 72
-0.001 * 120
-3.39
= -1.45
```

$$\frac{1}{1+e^{-(-1.45)}} = 0.19$$

Making predictions in Python

- Female (gender_F=1)
- Age 72
- 120 days since last gift

```
logreg.predict_proba([1, 72, 120])
```

```
array([[ 0.8204144,  0.1795856]])
```

Making predictions in Python

```
new_data = current_data[["gender_F","age","time_since_last_gift"]]


predictions = logreg.predict_proba(new_data)
```

Let's practice!

INTRODUCTION TO PREDICTIVE ANALYTICS IN PYTHON