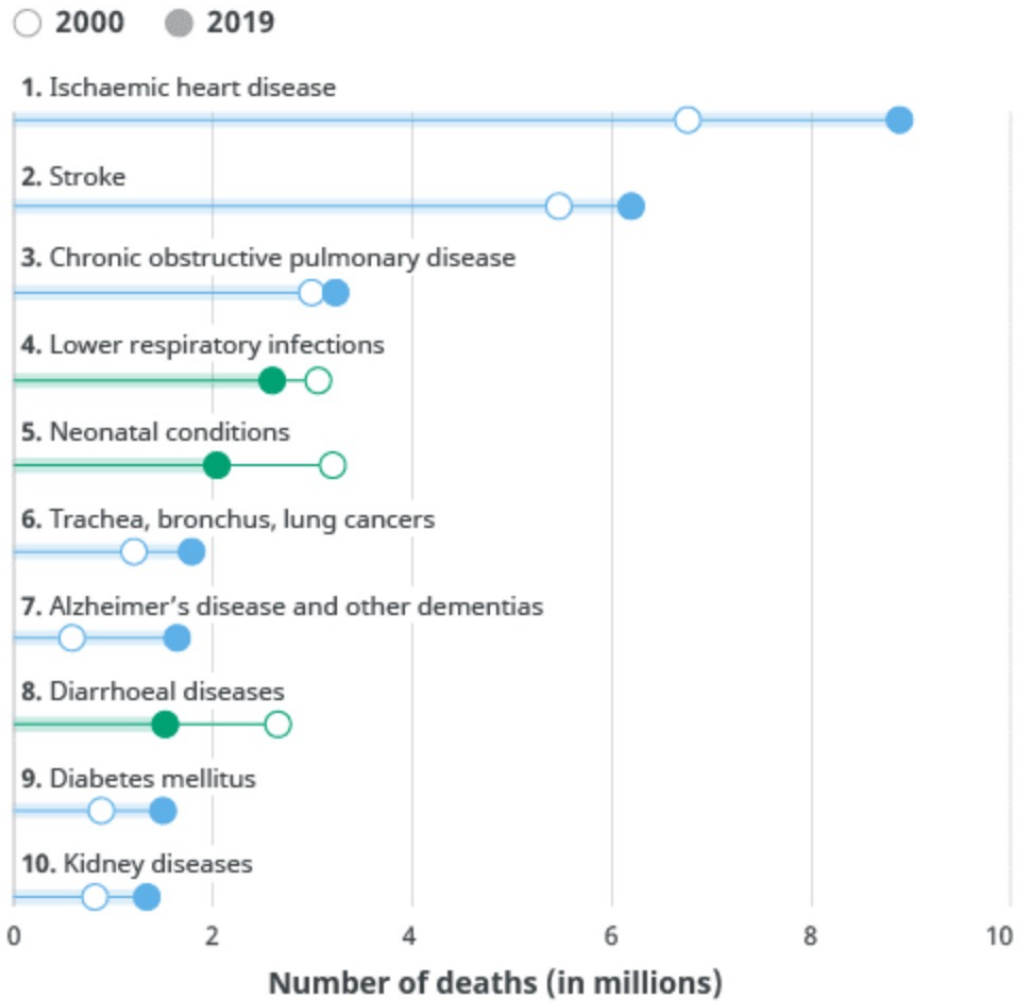
Jonathan Suárez

Jonhy para los amigos

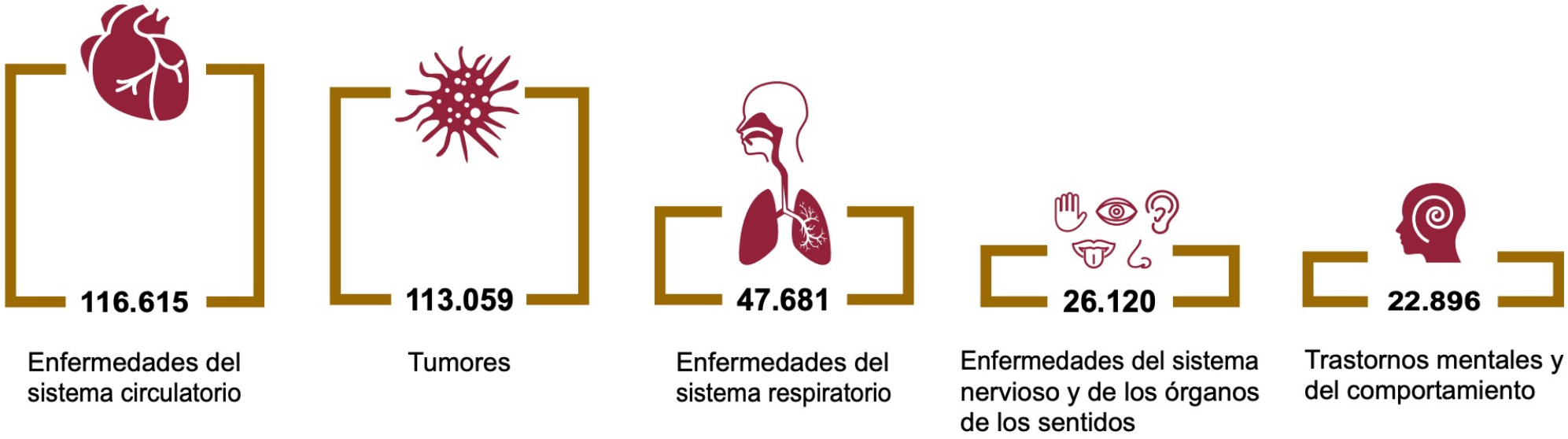
How to become a superhuman

Leading causes of death (worldwide)



https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Leading causes of death (Spain)



https://www.ine.es/infografias/infografia_causas_muerte.pdf

Types of cardiovascular disease

Congestive Heart Disease

coronary heart disease

Angina

Heart attack

Stroke

Causes of cardiovascular disease

High blood cholesterol or triglyceride levels

High blood pressure

Stress

Obesity

Lack of exercise

Thyroid disorders

Longstanding alcohol abuse

Age

smoking

Causes of cardiovascular disease

High blood cholesterol or triglyceride levels

High blood pressure

Stress

Obesity

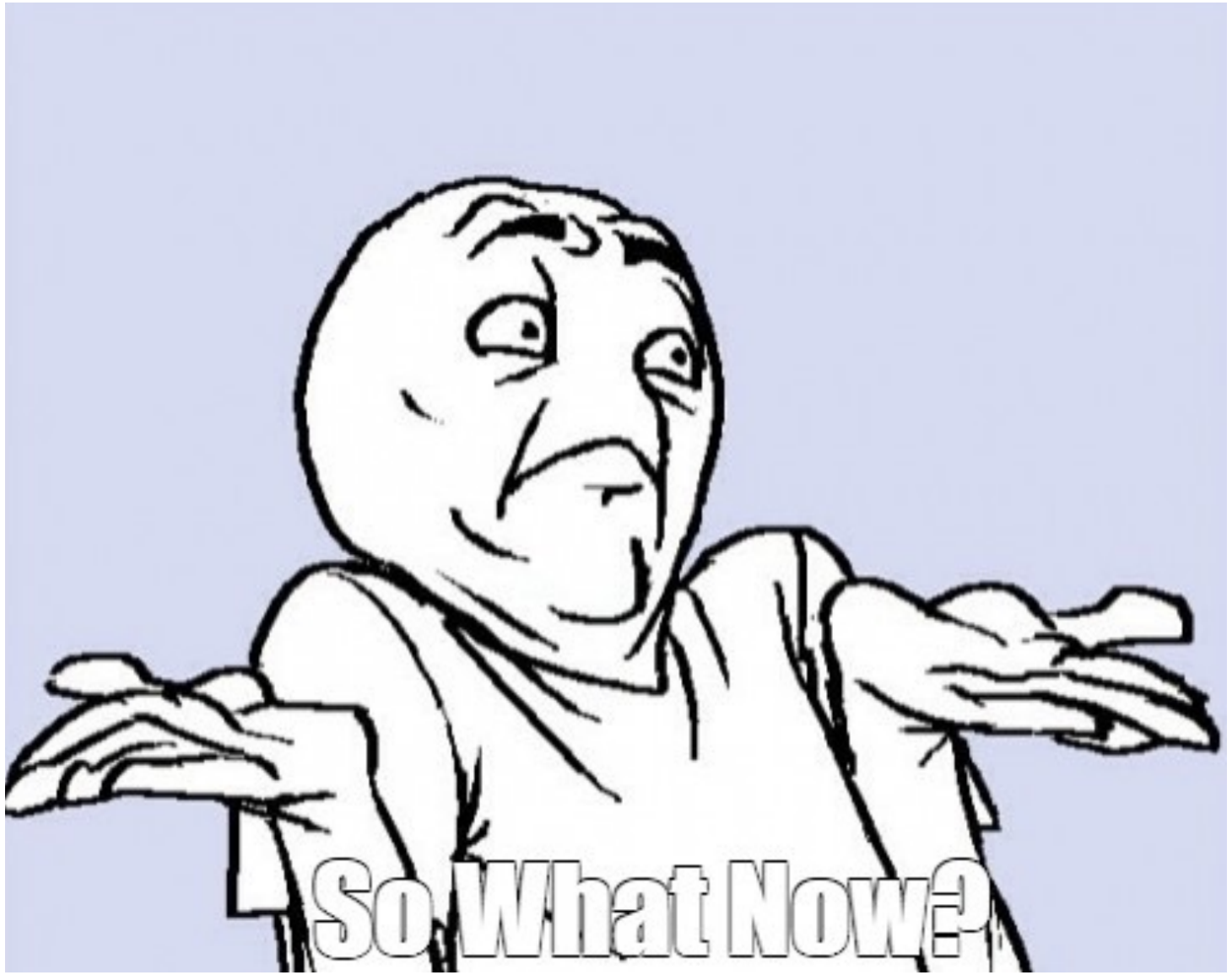
Lack of exercise

Thyroid disorders

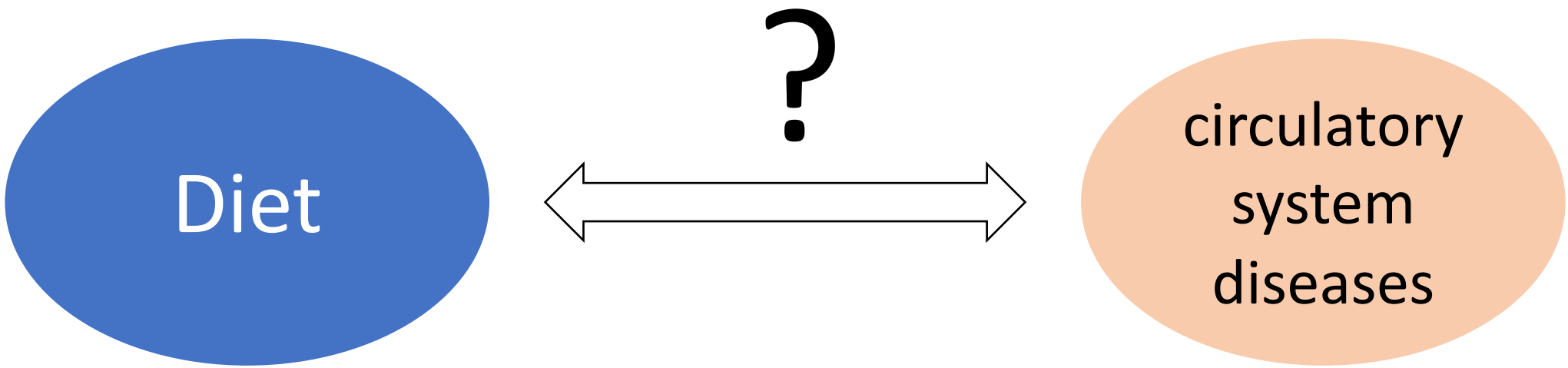
Longstanding alcohol abuse

Age

smoking


So What Now?

Project goal



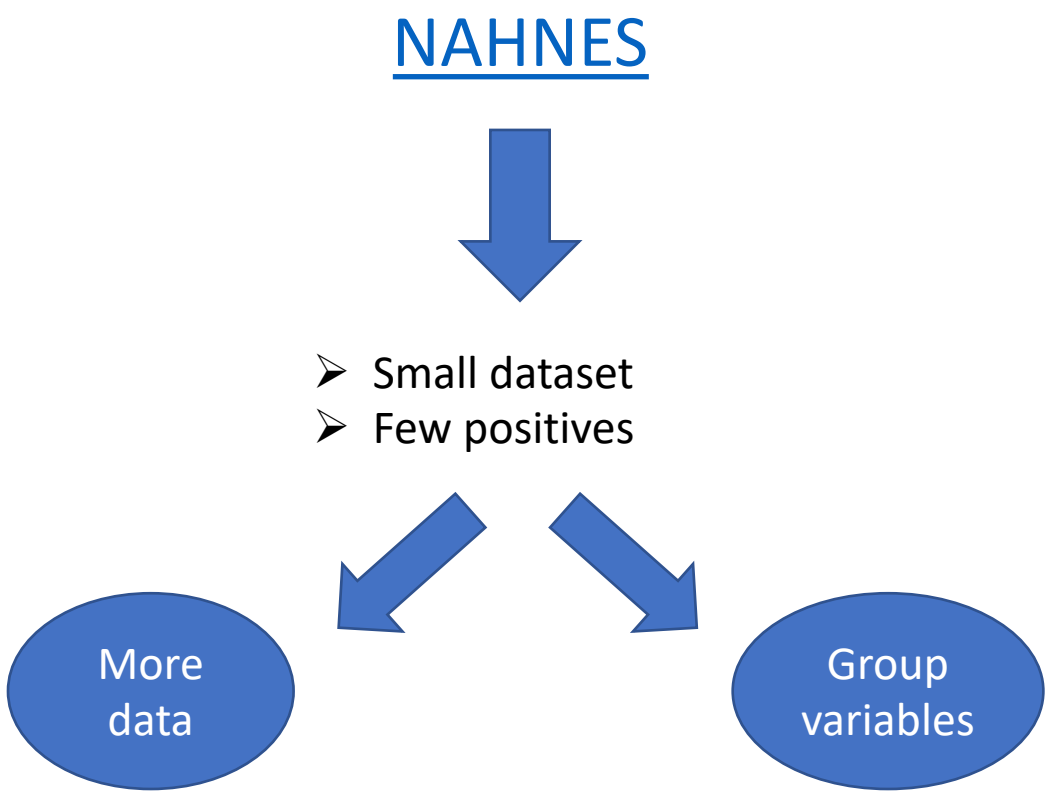
Data – Source & Wrangling

NAHNES

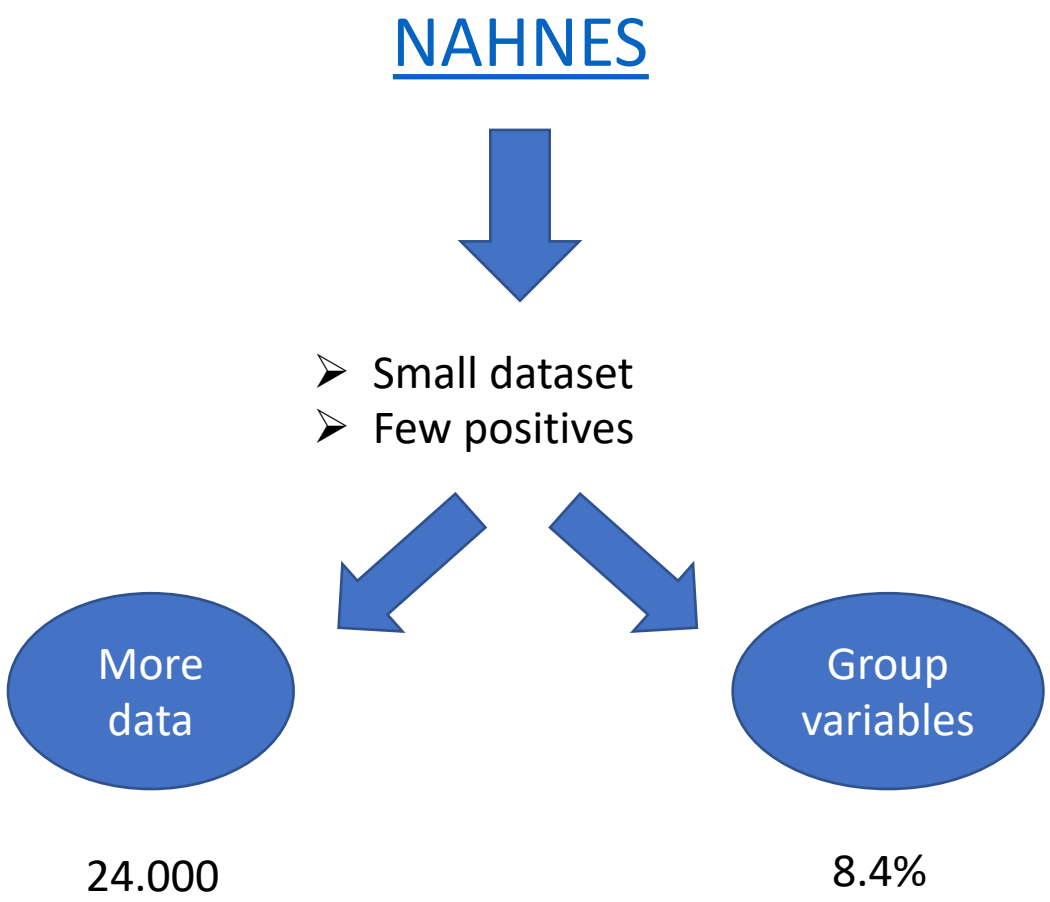
- Small dataset
- Few positives

8.000 -> 5% positives

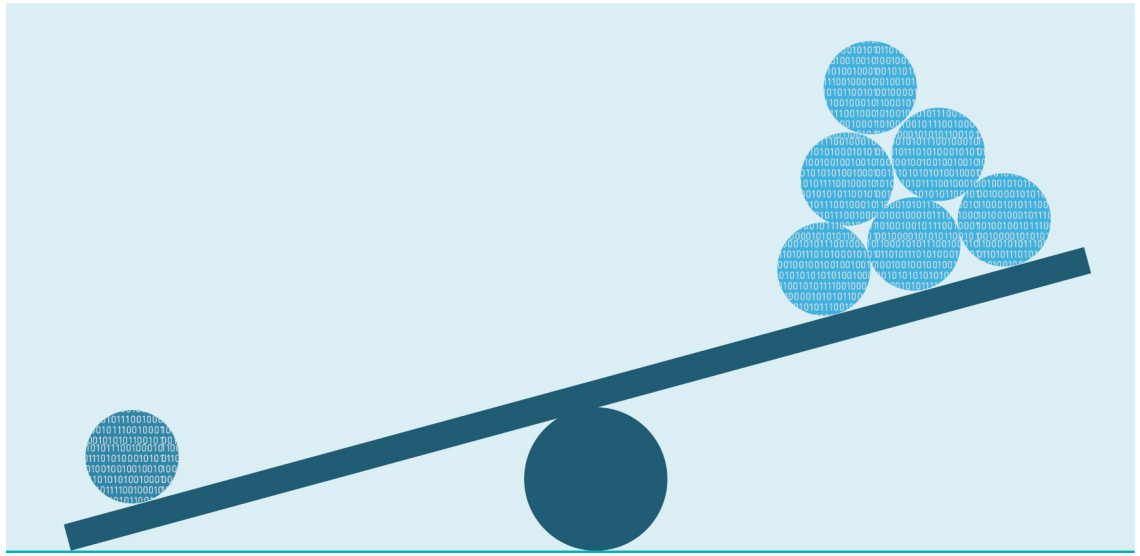
Data – Source & Wrangling



Data – Source & Wrangling

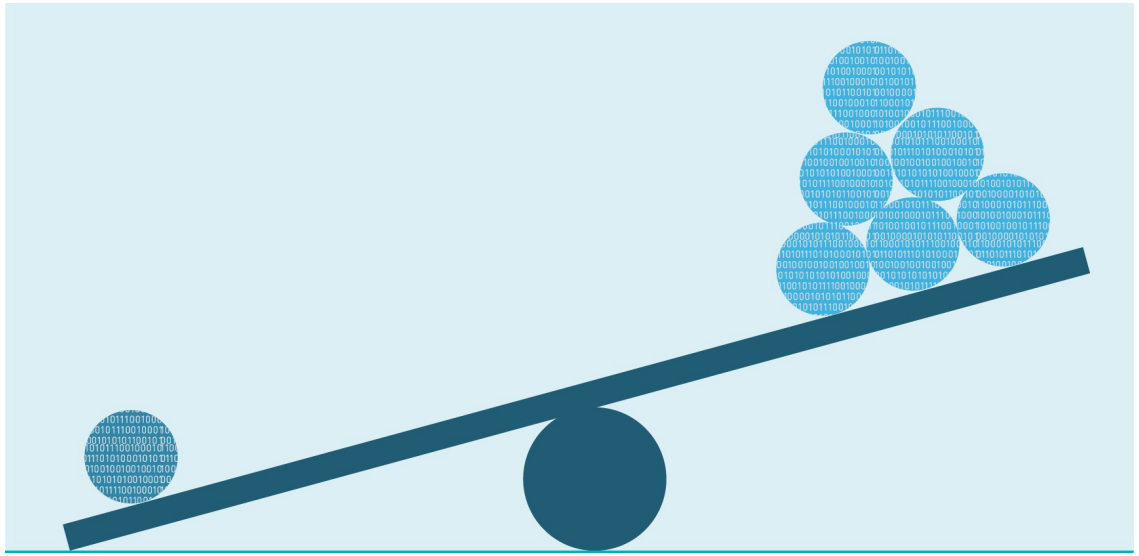


Still quite some imbalance…



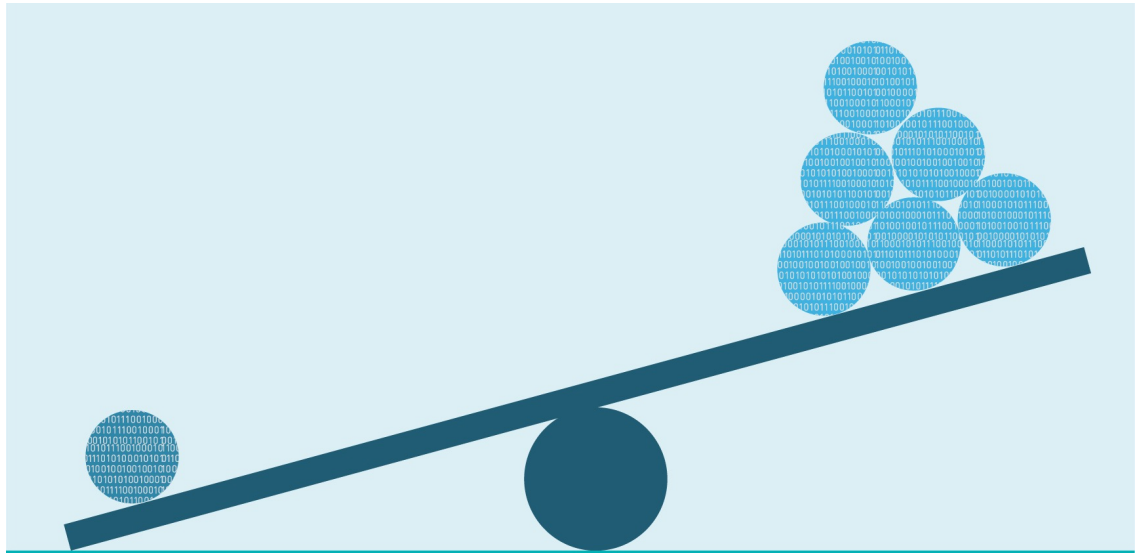
```
Congestive heart failure:
 0.0    16311
1.0       574
Name: MCQ160B, dtype: int64
--------------------------------------------------
Cononary heart disease:
 0.0    16169
1.0       716
Name: MCQ160C, dtype: int64
--------------------------------------------------
Angina pectoris:
 0.0    16470
1.0       415
Name: MCQ160D, dtype: int64
--------------------------------------------------
Heart attack:
 0.0    16172
1.0       713
Name: MCQ160E, dtype: int64
--------------------------------------------------
Stroke:
 0.0    16225
1.0       660
Name: MCQ160F, dtype: int64
--------------------------------------------------
Any cardiovascular disease:
 0    27371
1     1842
Name: MCQ160H, dtype: int64
```

Possible solutions

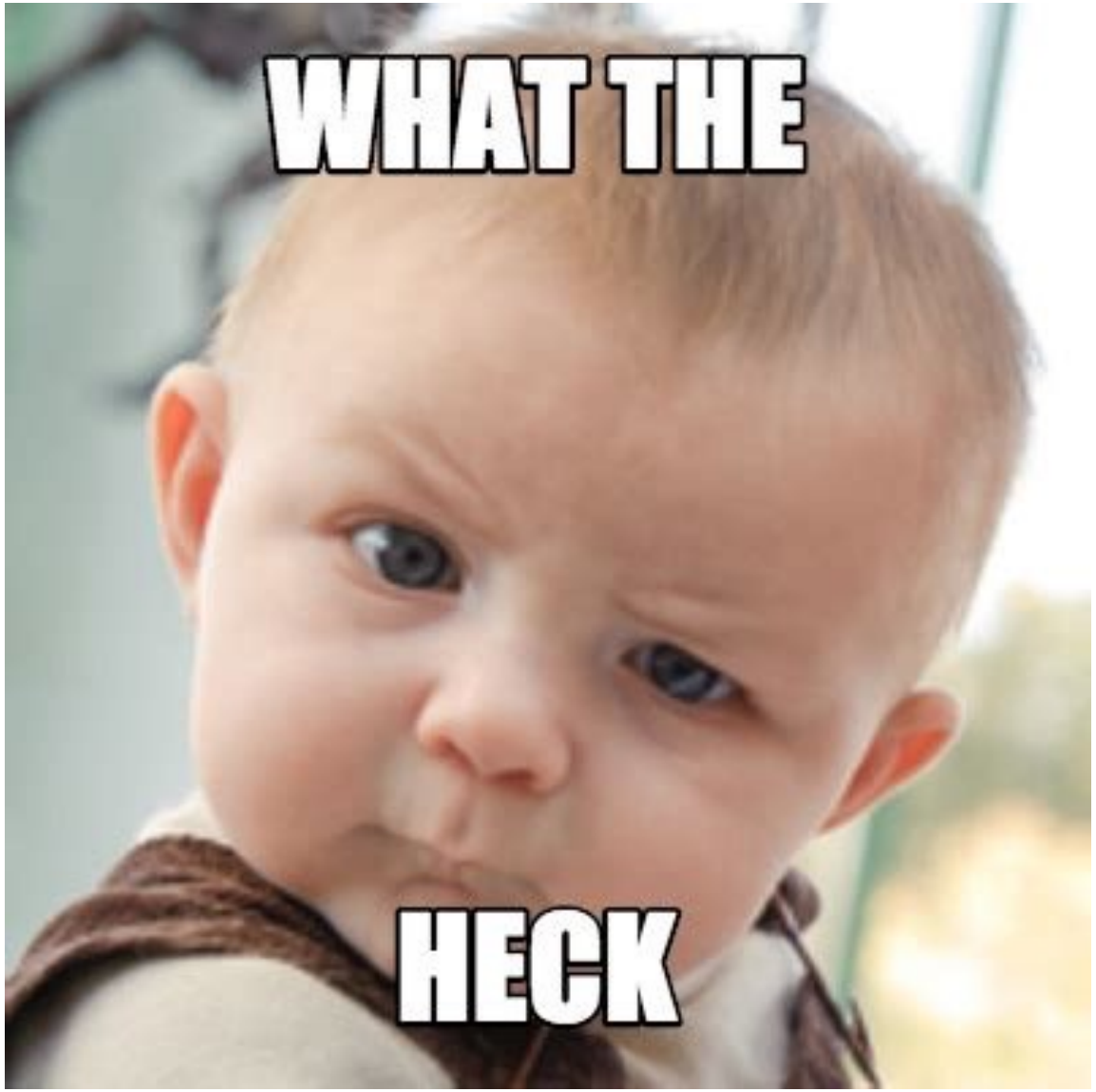


- Reduce negative sample size
- Oversample positive values
- Adjust the loss function of the models

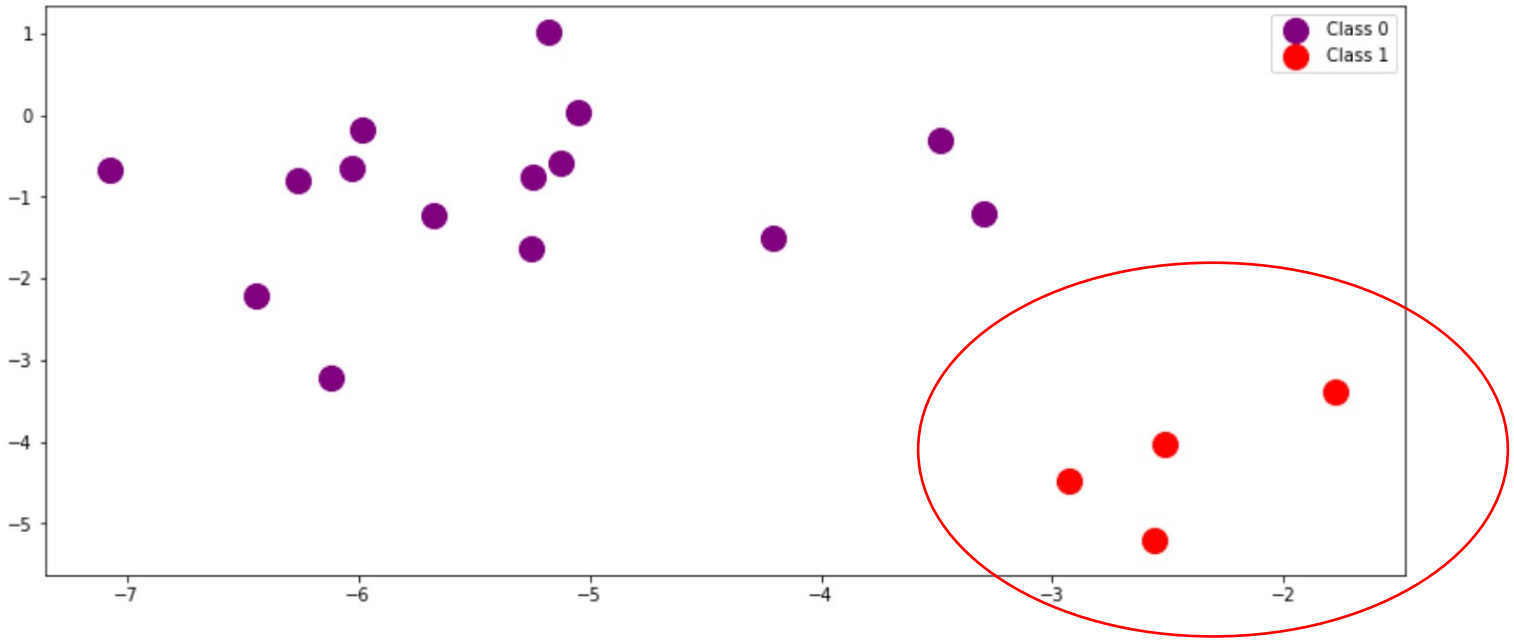
Solution



- Reduce negative sample size
- Oversample positive values
- Adjust the loss function of the models

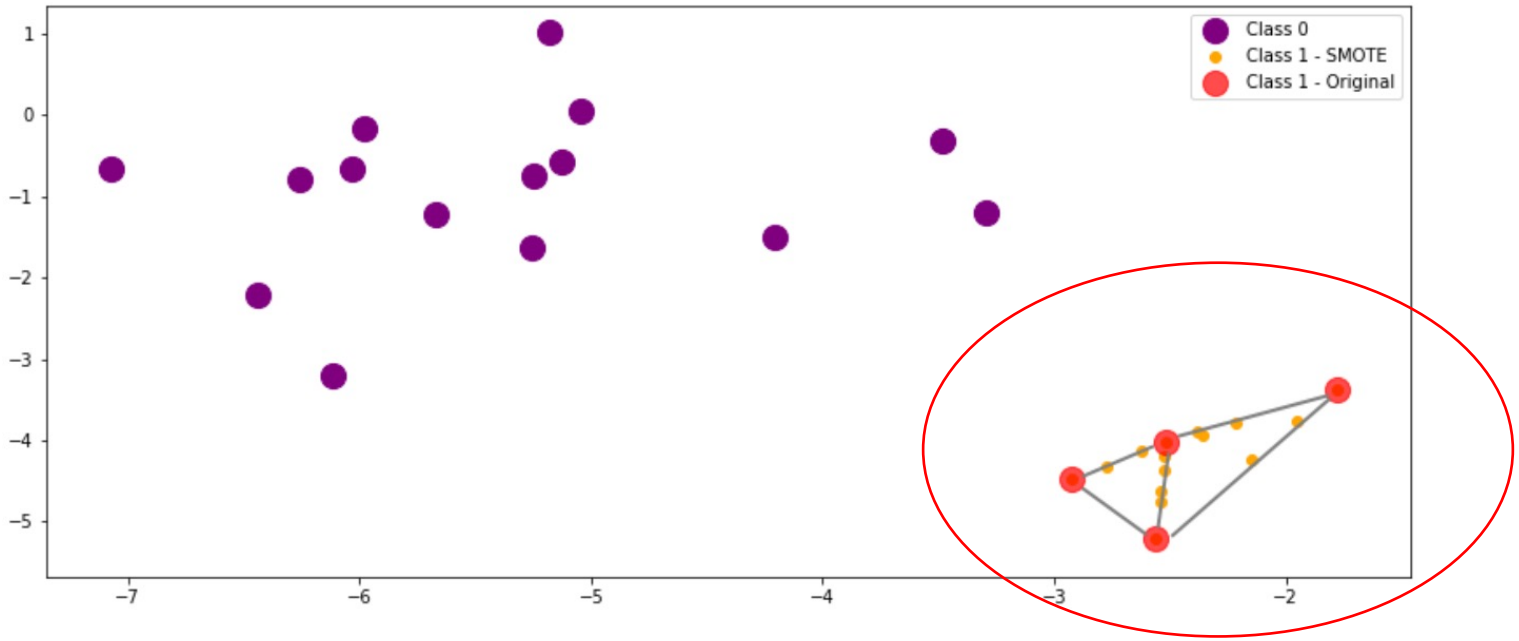

WHAT THE
HECK

Balancing data with SMOTE



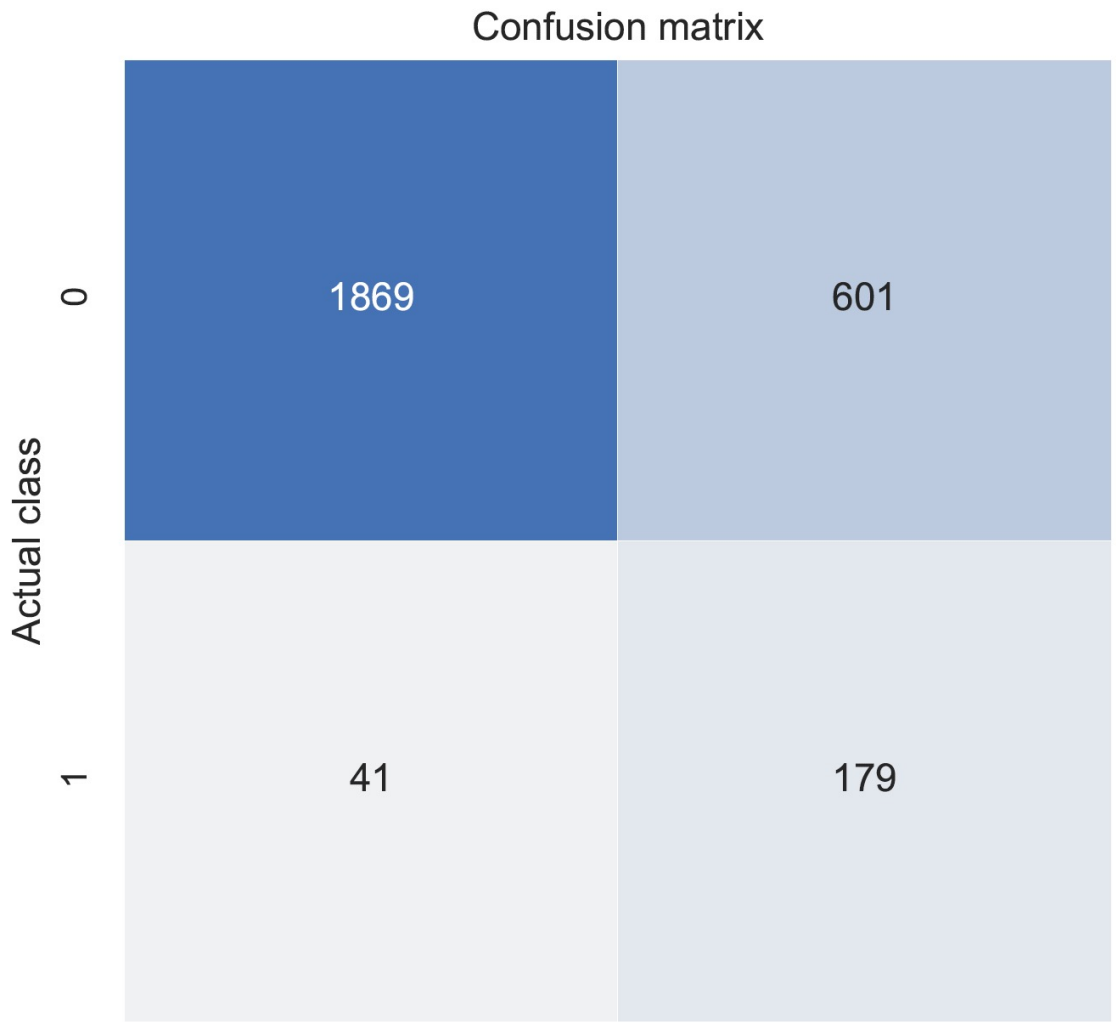
https://medium.com/analytics-vidhya/balance-your-data-using-smote-98e4d79fcddb

Balancing data with SMOTE

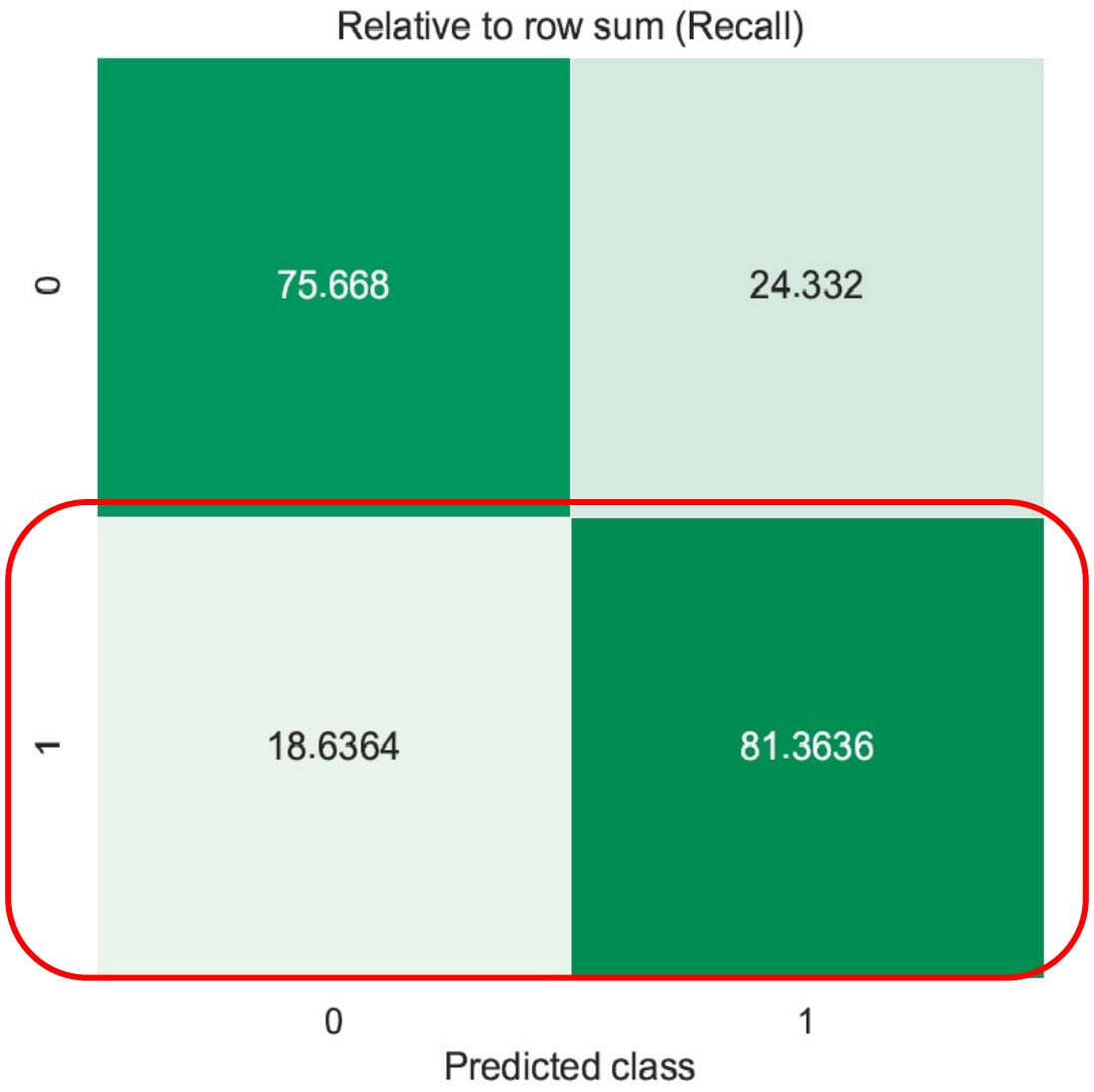


https://medium.com/analytics-vidhya/balance-your-data-using-smote-98e4d79fcddb

Predictions



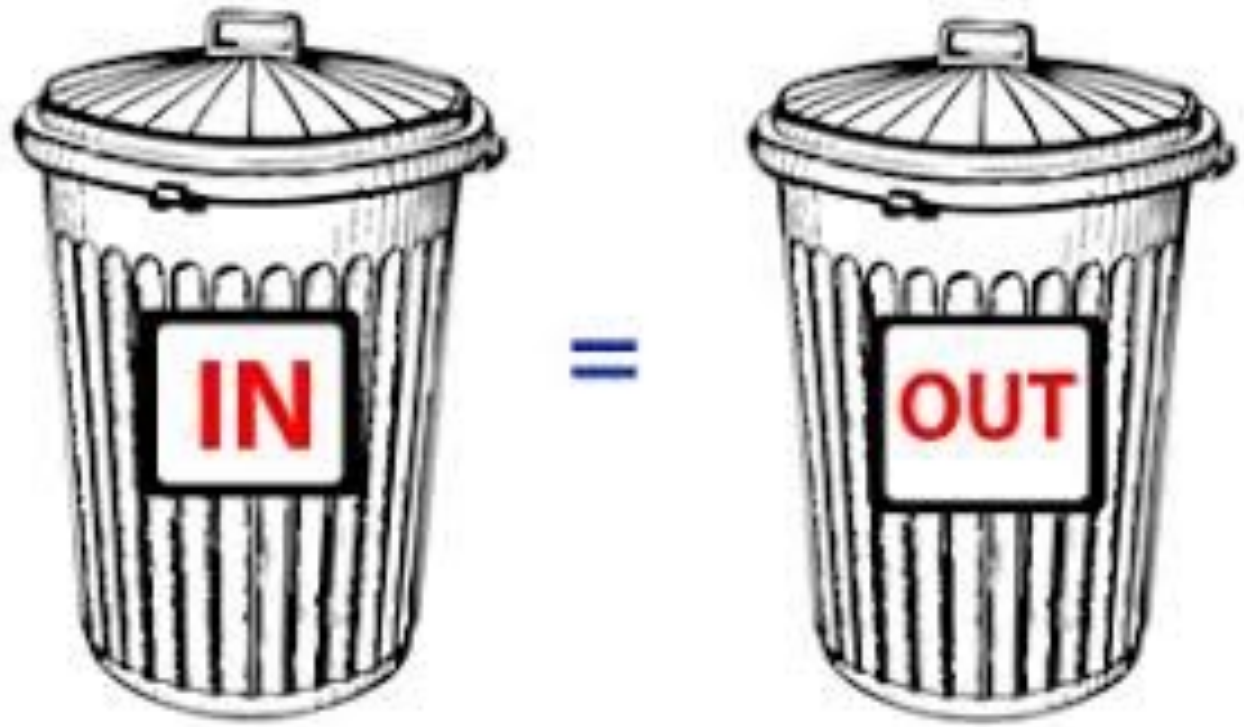
Recall


Relative to row sum (Recall)
0
1
75.668
24.332
18.6364
81.3636
0
1
Predicted class

Precision

Relative to col sum (Precision)
97.8534
77.0513
2.1466
22.9487
0
1

What
Next?


IN
=
OUT