
UCI HEART DISEASE DATASET MODELING



GOAL

1. Build a model that will predict if a patient has cardiovascular disease.

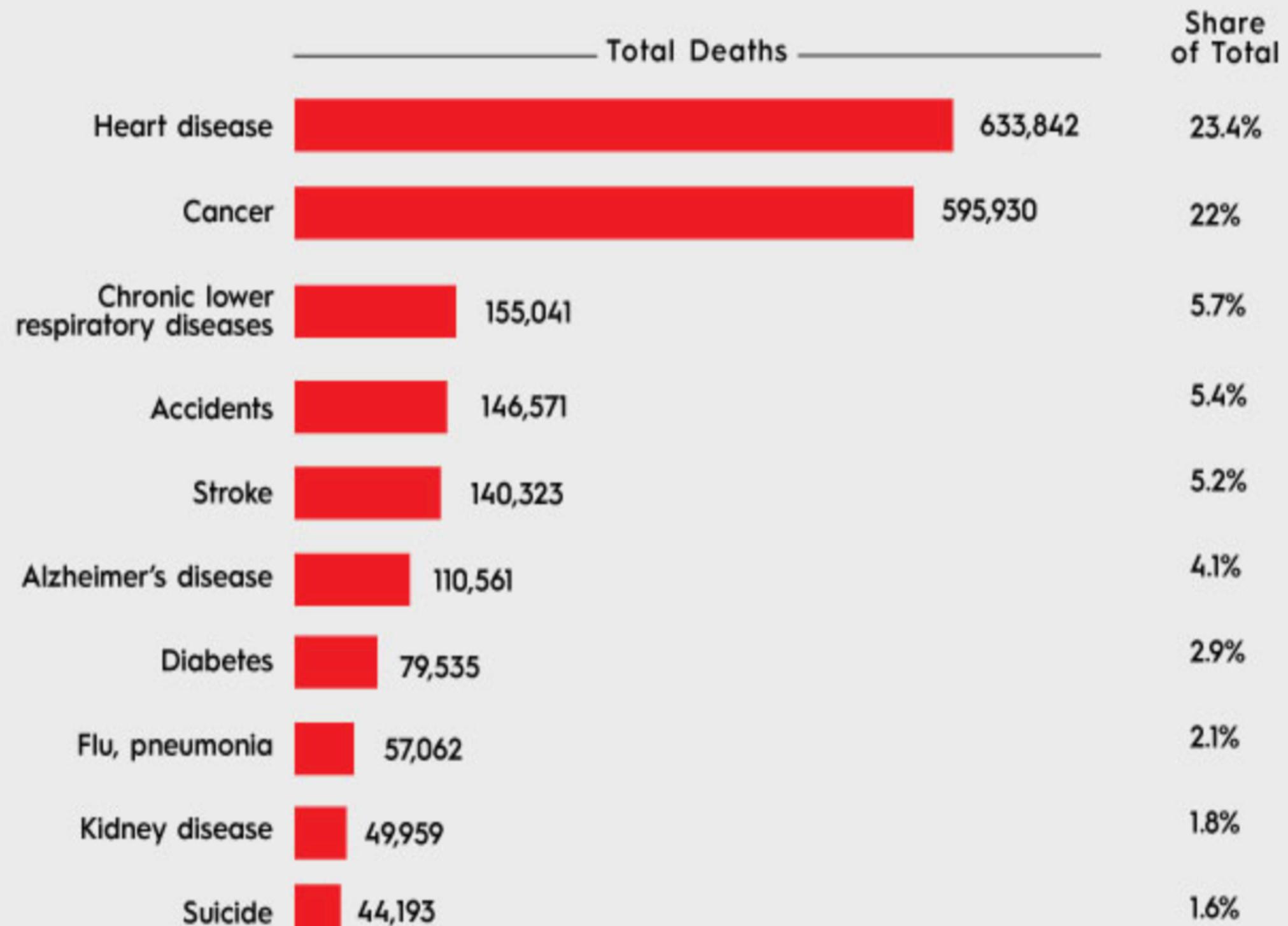
2. Make recommendations that will improve the diagnostic precision and improve life expectancy in people with heart disease.



Leading Causes of Death

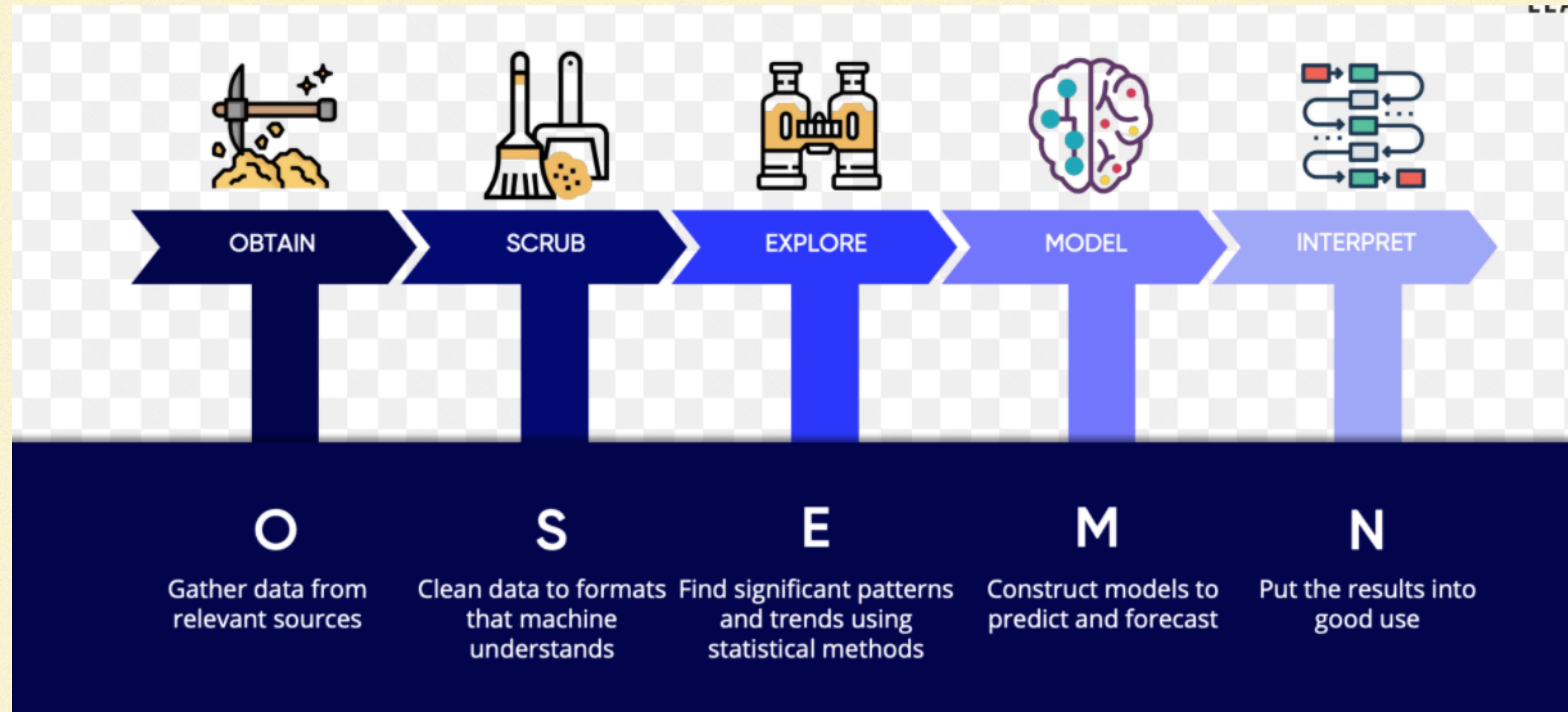
By AMERICAN HEART ASSOCIATION NEWS

Heart disease continues to kill more Americans than any other cause, followed by stroke at No. 5, according to 2015 federal data.



Source: Centers for Disease Control and Prevention

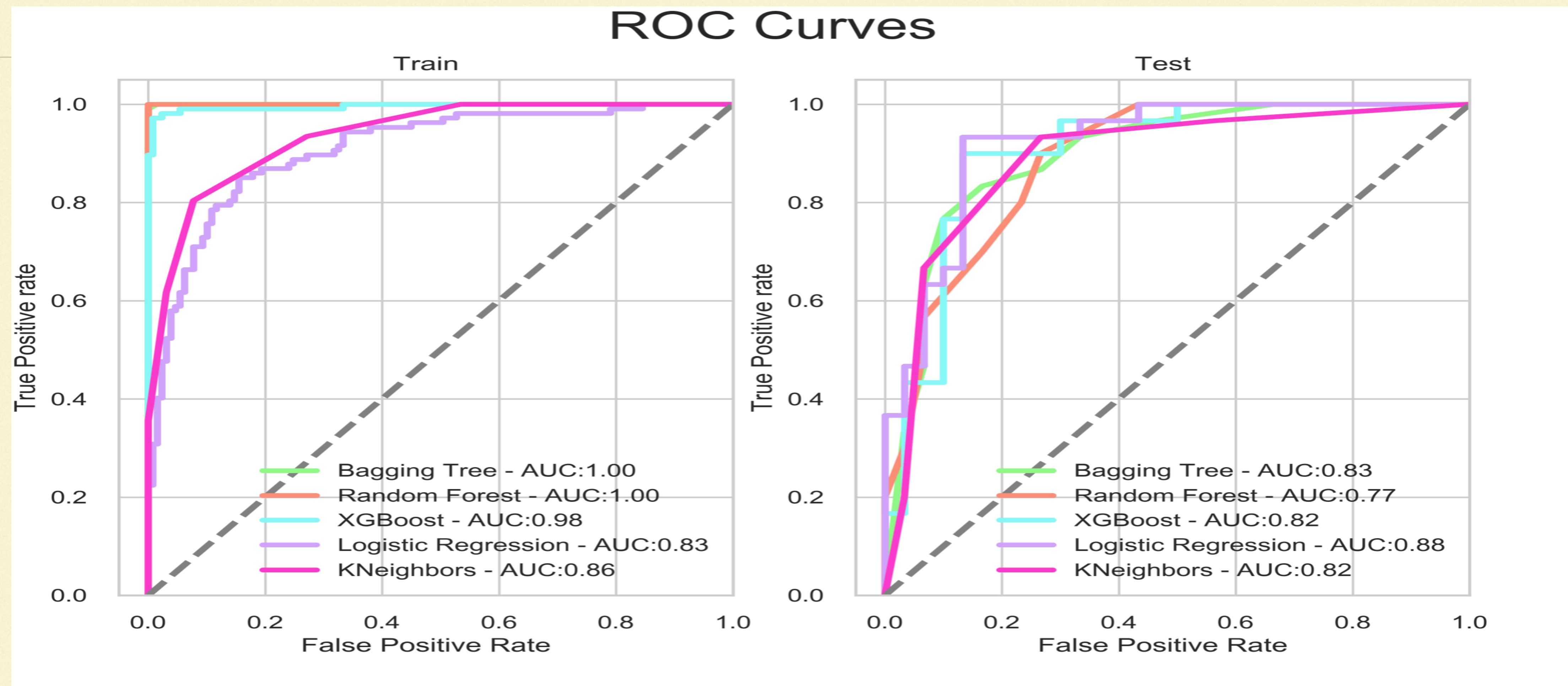
Published Dec. 8, 2016



METHODOLOGY USED

MODELS USED

ROC Curves



MODELS USED

Classifier	Runtime	Train Precision	Test Precision	Train Recall	Test Recall	Train F1	Test F1	Train Accuracy	Test Accuracy	Train ROC AUC	Test ROC AUC
Bagging Tree	0.01	1.00	0.83	0.99	0.83	1.00	0.83	1.00	0.83	1.00	0.83
Random Forest	0.01	1.00	0.81	1.00	0.70	1.00	0.75	1.00	0.77	1.00	0.77
XGBoost	0.04	0.99	0.85	0.96	0.77	0.98	0.81	0.98	0.82	0.98	0.82
Logistic Regression	0.00	0.81	0.85	0.82	0.93	0.82	0.89	0.83	0.88	0.83	0.88
KNeighbors	0.00	0.90	0.83	0.80	0.80	0.85	0.81	0.87	0.82	0.86	0.82

- Logistic Regression Model performed the best with a run time of 0.001 seconds , 0.93 – Recall Score and 0.88 – Accuracy Score.

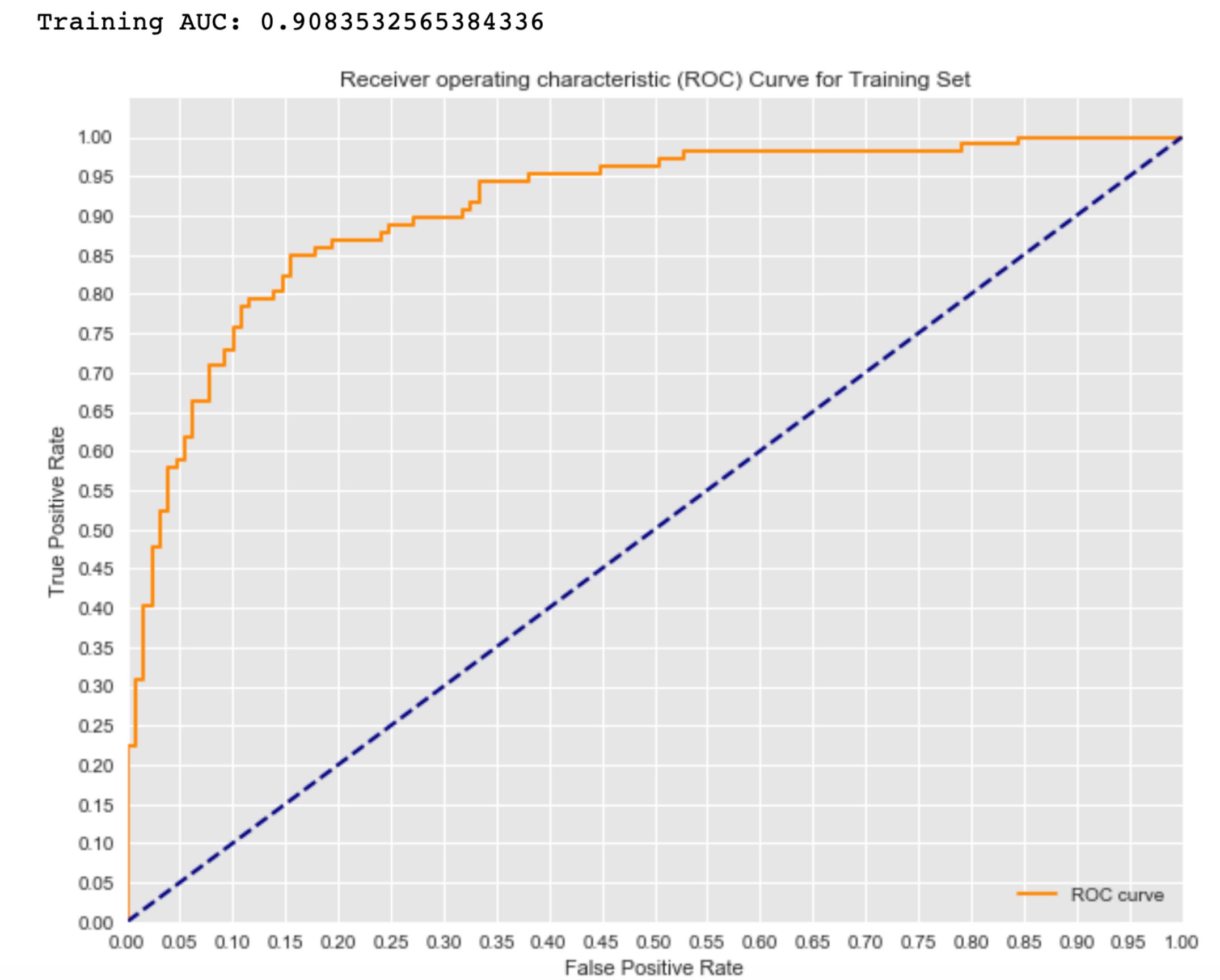
FINAL MODEL

Training Precision: 0.8148148148148148
Testing Precision: 0.8484848484848485

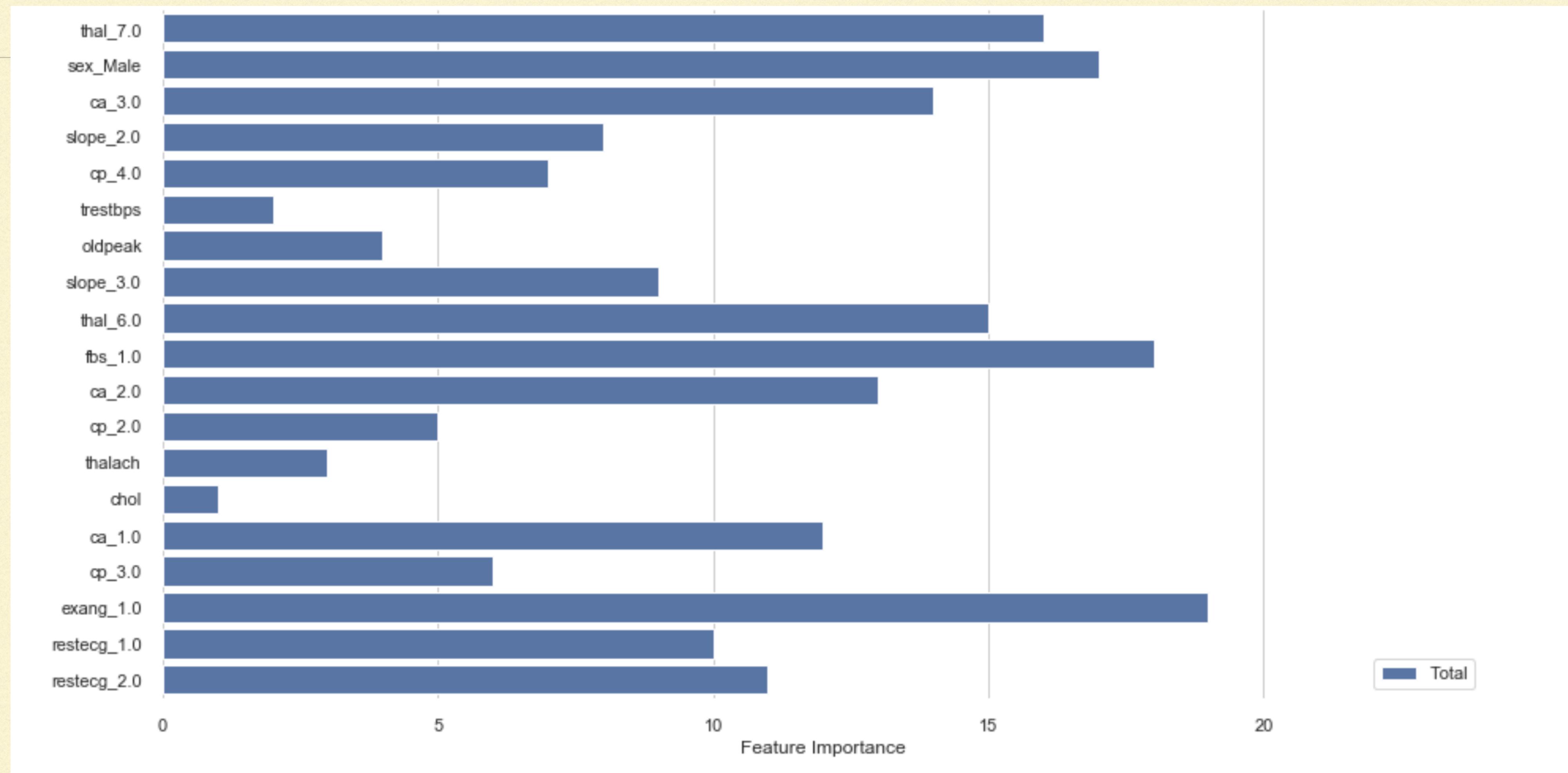
Training Recall: 0.822429906542056
Testing Recall: 0.9333333333333333

Training Accuracy: 0.8347457627118644
Testing Accuracy: 0.8833333333333333

Training F1-Score: 0.8186046511627906
Testing F1-Score: 0.8888888888888889

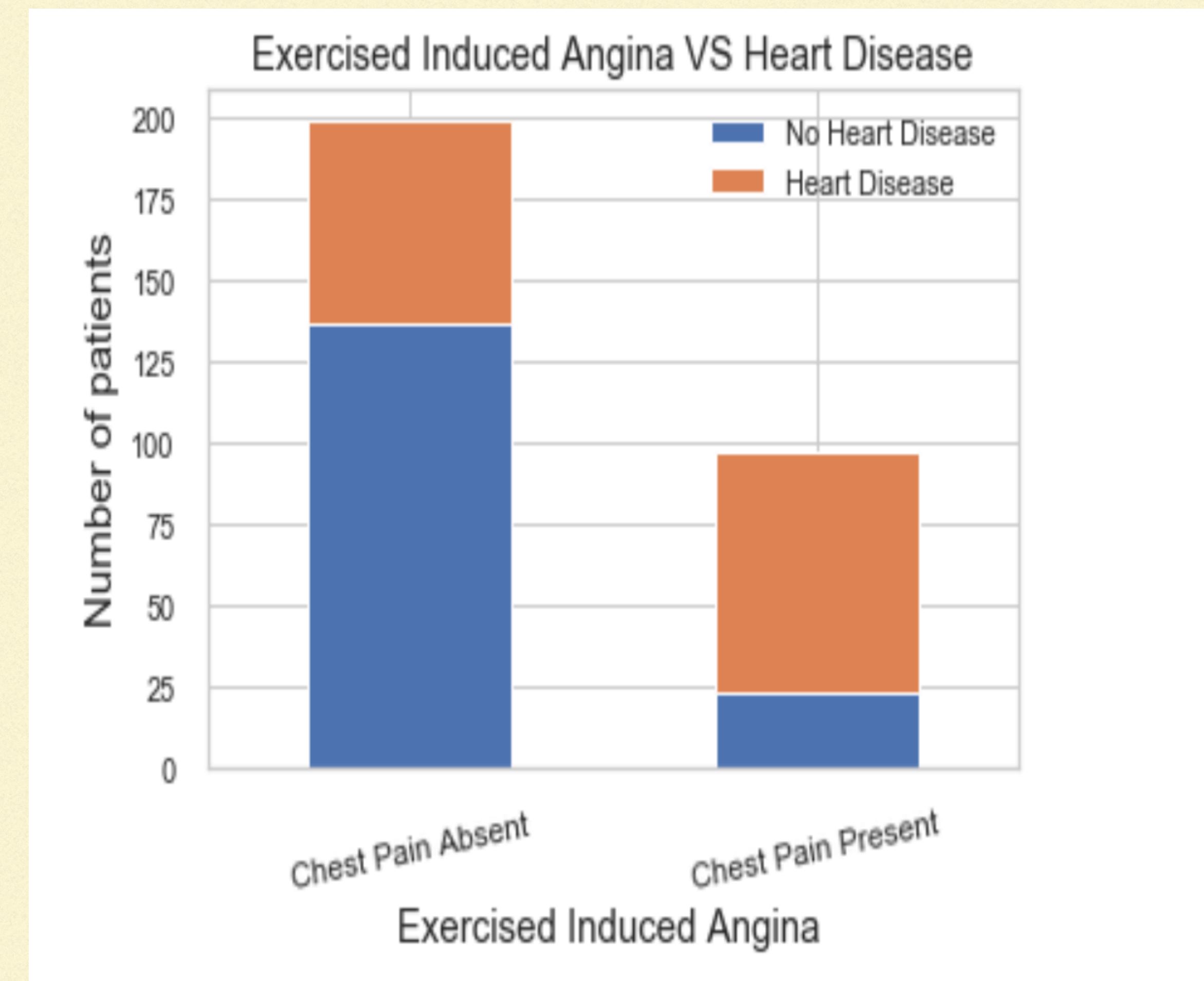


FEATURE IMPORTANCE



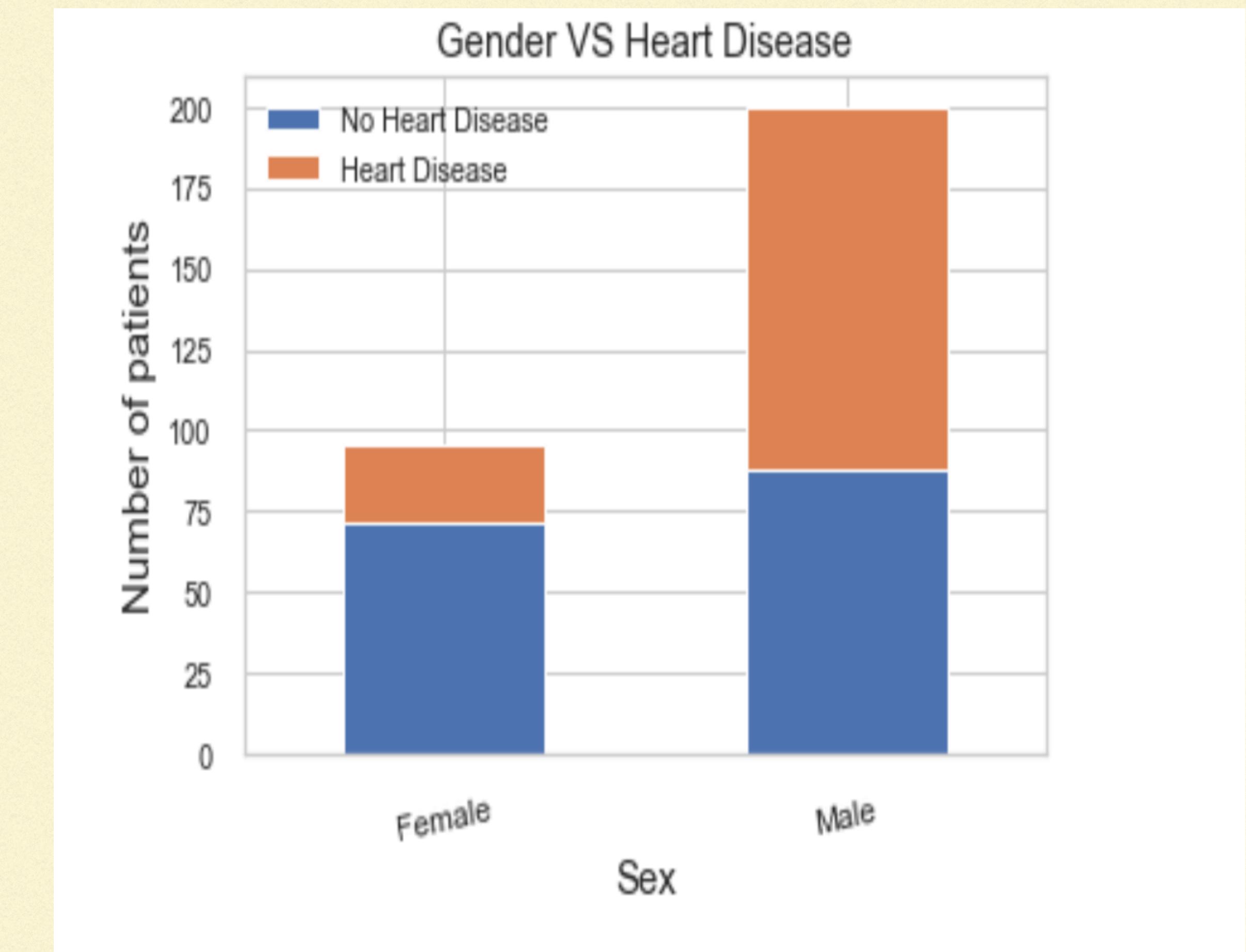
EXERCISED INDUCED ANGINA

A patient with exercise induced angina is more likely to have cardiovascular disease than a patient who doesn't experience it.



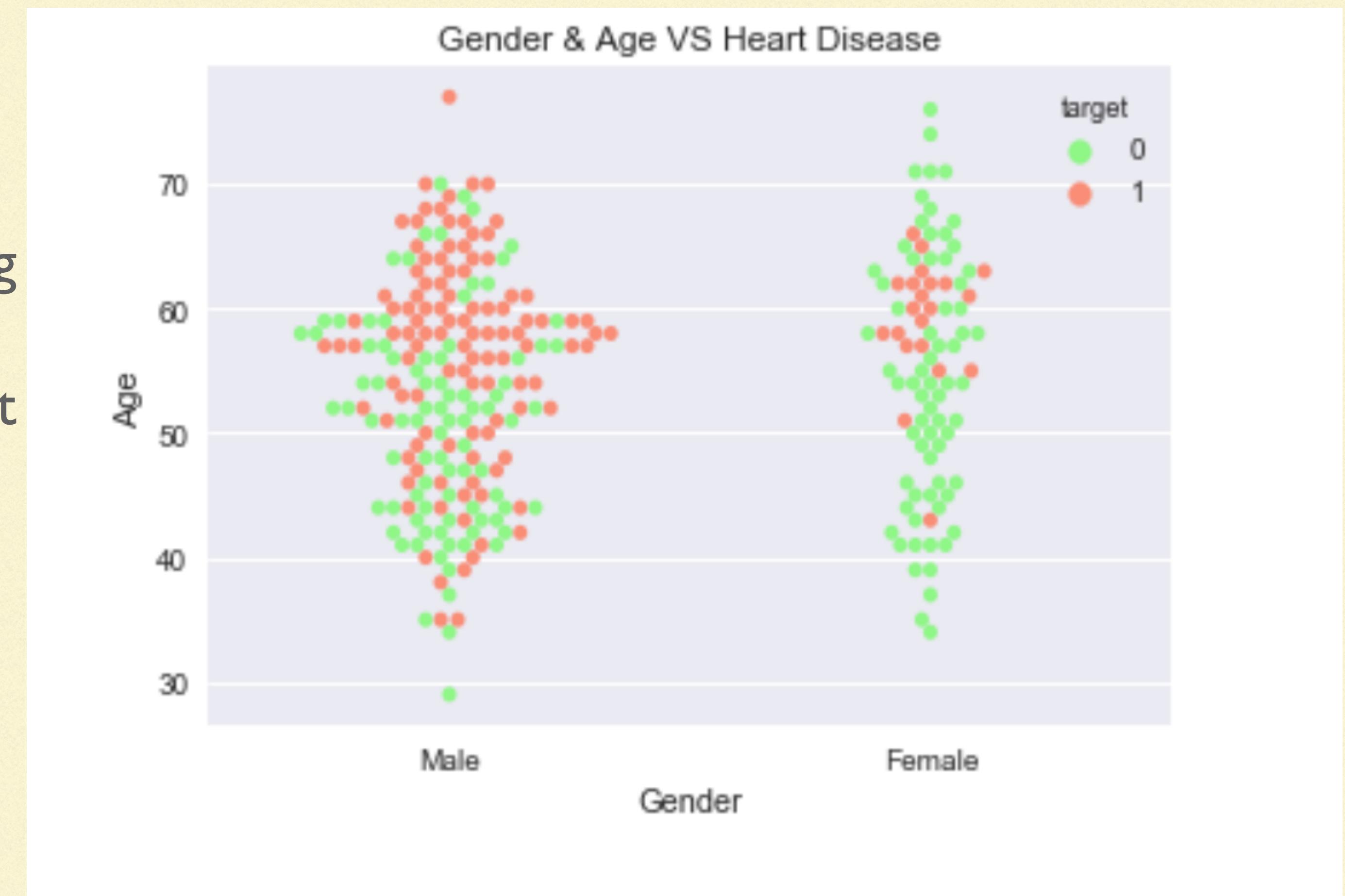
HOW GENDER INFLUENCES HEART DISEASE

The incidence of cardiovascular disease
is higher in man.



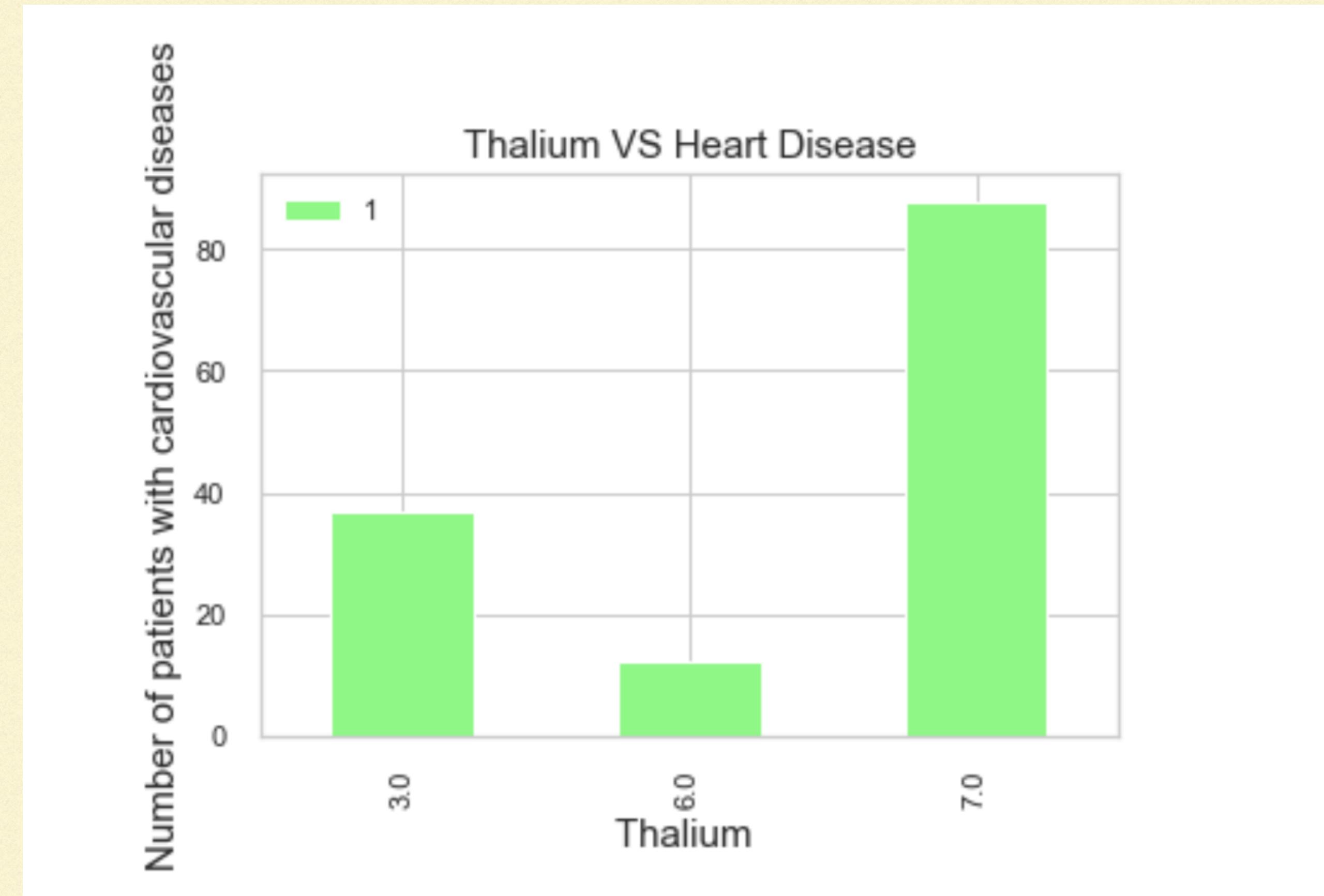
HOW GENDER AND AGE INFLUENCE HEART DISEASE

The previous statement is true for young patients. In the over 50 years old, age segment, gender doesn't play a significant role in predicting heart disease.



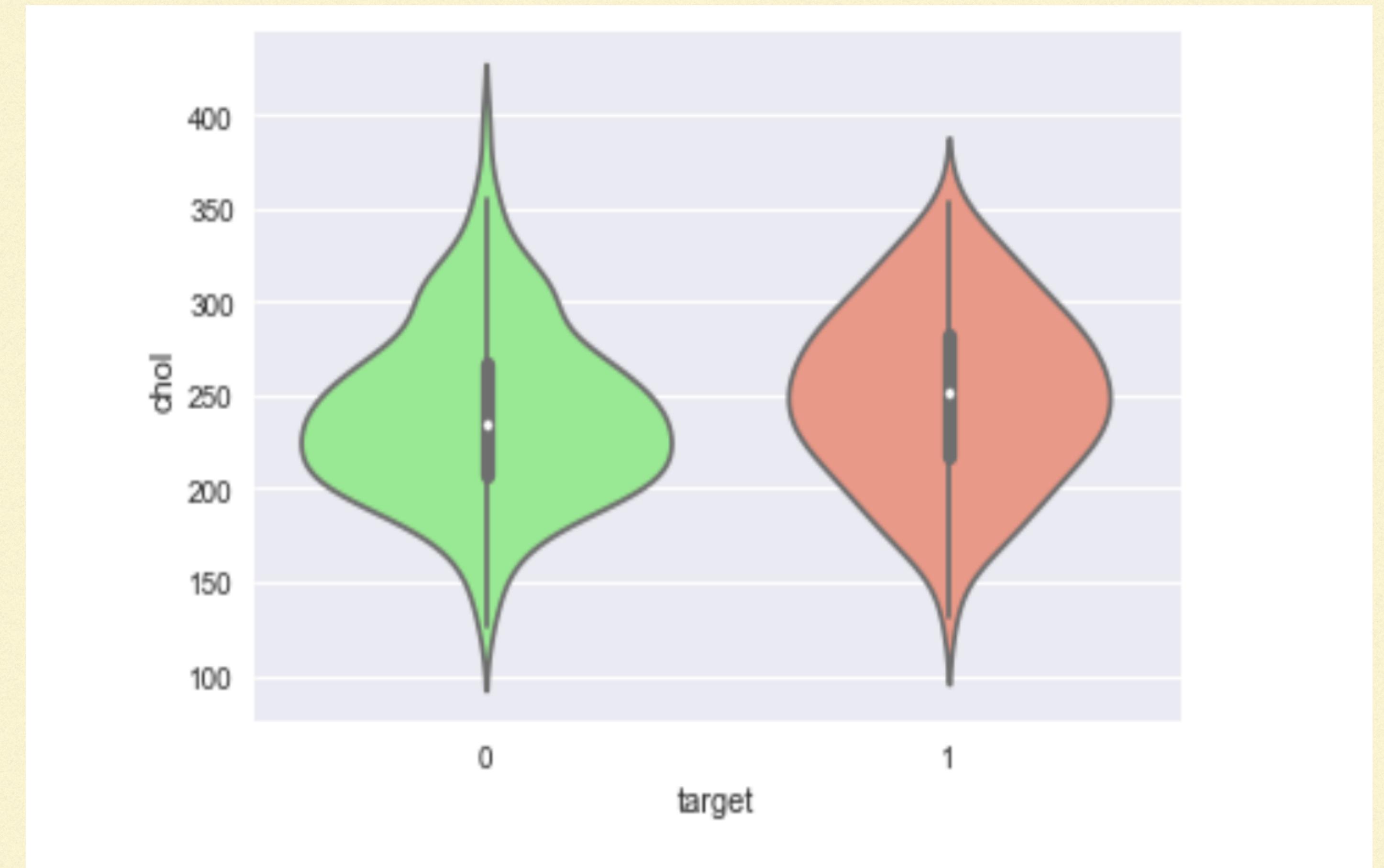
THALIUM STRESS TEST

Reversible cold spots appearing only during exercise, marked on the image with 7.0 usually indicate some blockage of the coronary arteries.



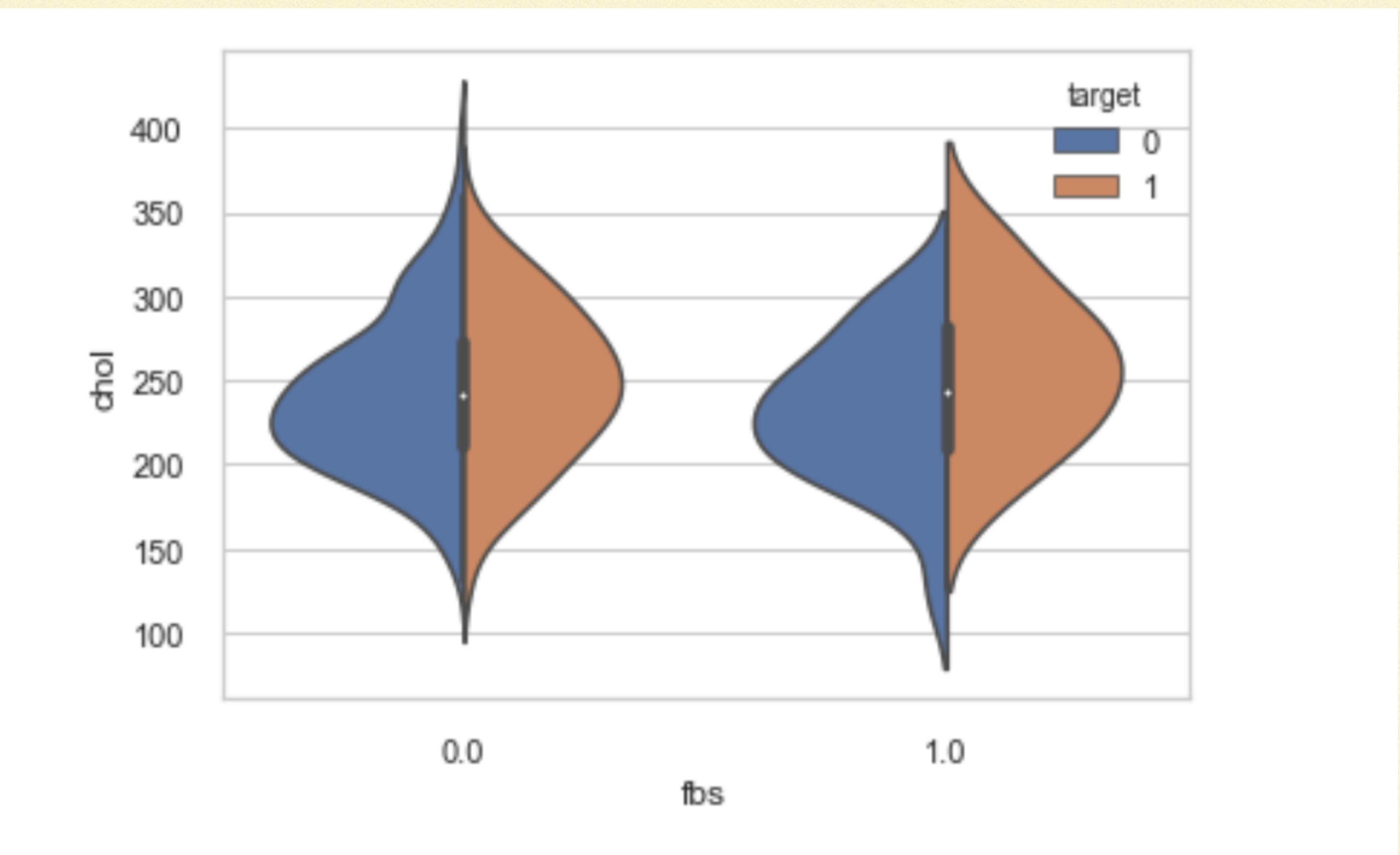
CHOLESTEROL

Higher cholesterol seem to increase the probability of having heart disease but not with much.



CHOLESTEROL

If you take cholesterol and blood sugar into consideration the difference is even less visible.



FINDINGS

Doctor:

A good model will improve the diagnostic precision and the time a doctor spends on it.

Patient:

An early fast diagnostic will fasten the recovery and save money.

FUTURE WORK

-Improve the model:

1. Add a BMI column.
2. Find a better dataset for heart disease. This one is 32 years old. Many things have changed since then.

-Improve the XGBoost model with the hyperparameters found with SearchGrid.

THANK YOU
