



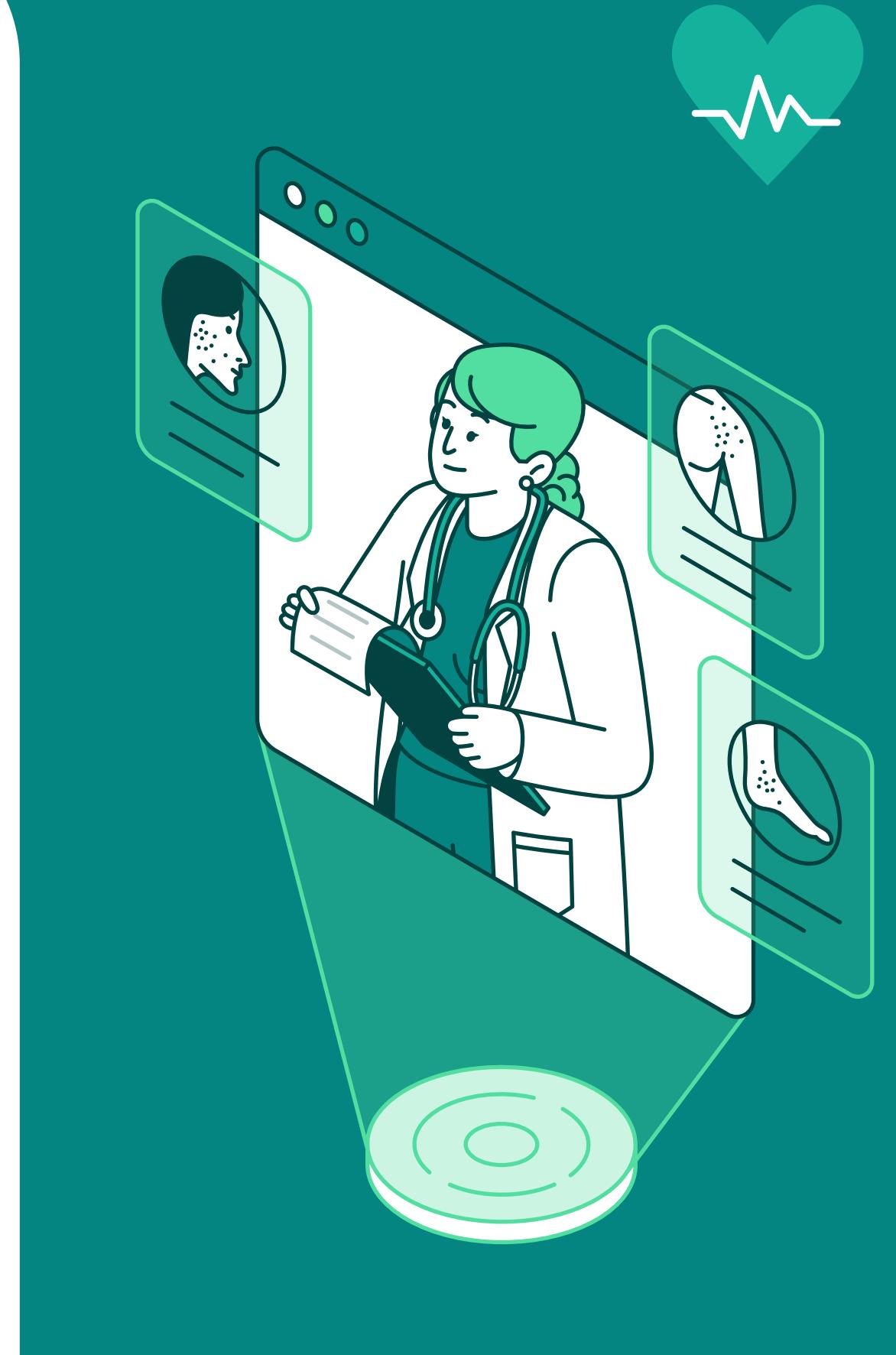
Developing a heart disease prediction dashboard

by Harish Muhammad



Outlines

- Background/context
- Problem Understanding
- Data Understanding
- Data Analysis
- Modeling
- Hyperparameter tuning
- Confusion matrix
- Learning curve
- LIME
- Recommendations
- Demonstration



PROBLEM UNDERSTANDING



Heart disease:

The leading mortality cause worldwide

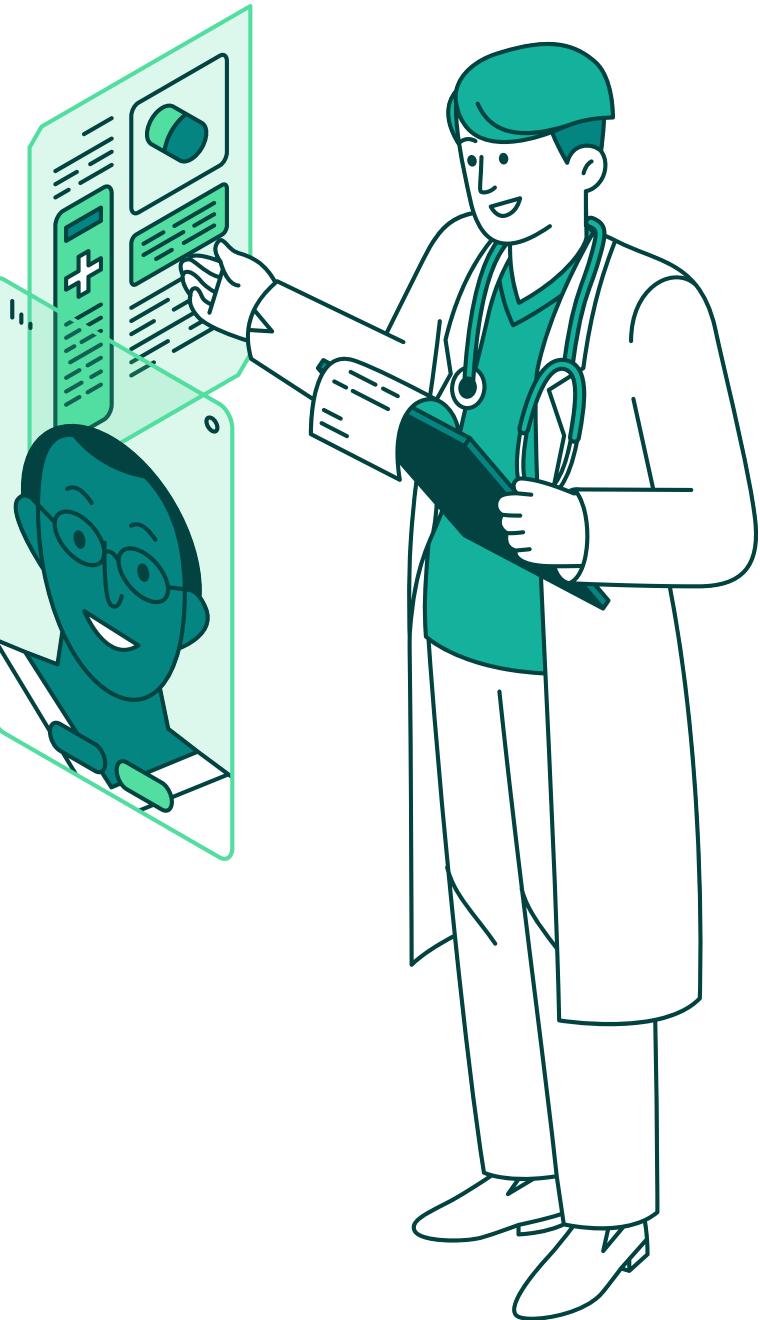


WHO:

17.9 M global deaths annually
caused by heart disease & stroke

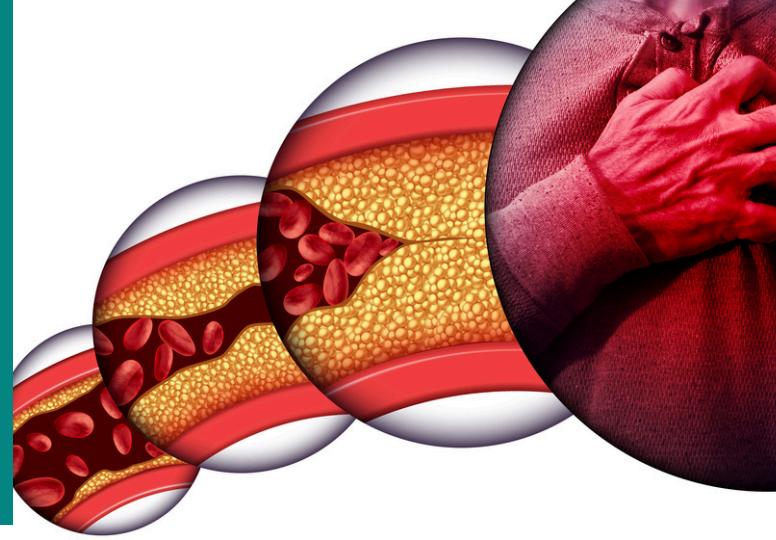


PROBLEM UNDERSTANDING



Assisting medical professionals with fast & accurate
ML prediction for the diagnosis of heart diseases
will be essential to improve health monitoring

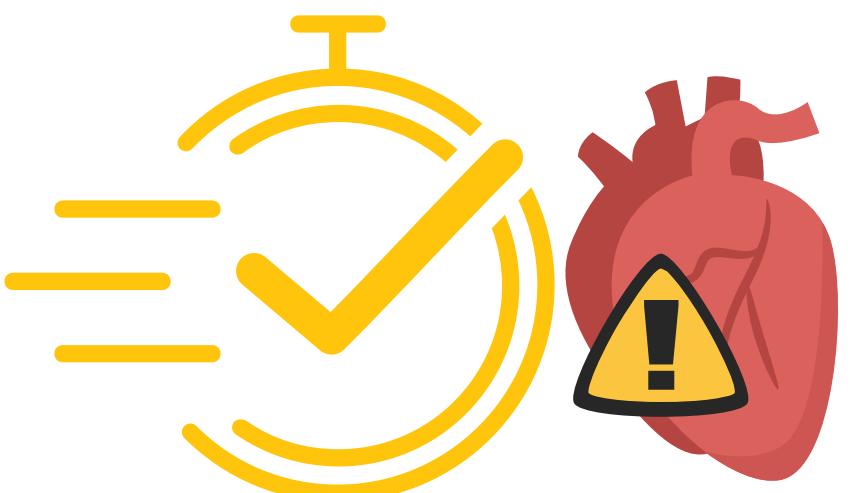
Why heart disease prediction App?



Easy & practical



Cost effective



Fast

Data understanding

Dataset Source



Heart Disease

Donated on 6/30/1988

4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach

Dataset Characteristics

Multivariate

Subject Area

Health and Medicine

Associated Tasks

Classification

Feature Type

Categorical, Integer, Real

Instances

303

Features

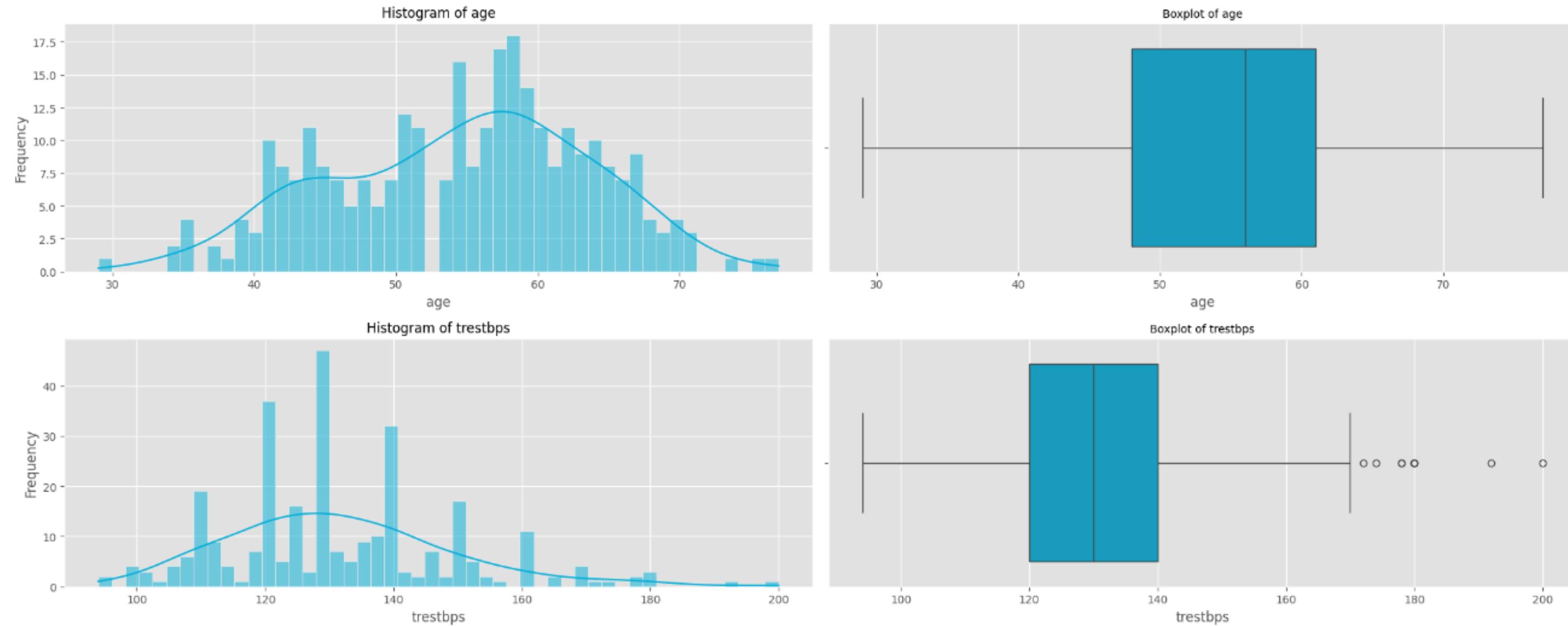
13

13 features

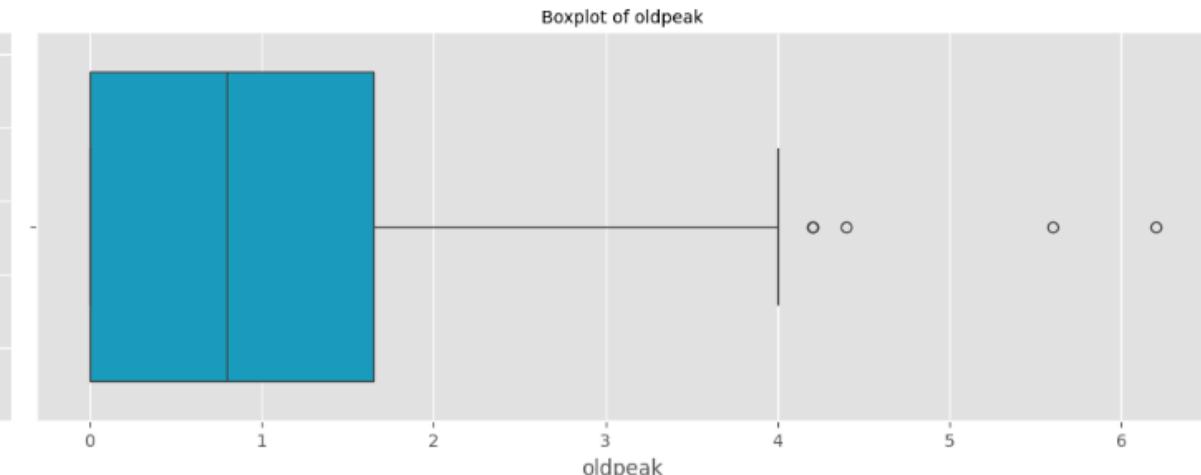
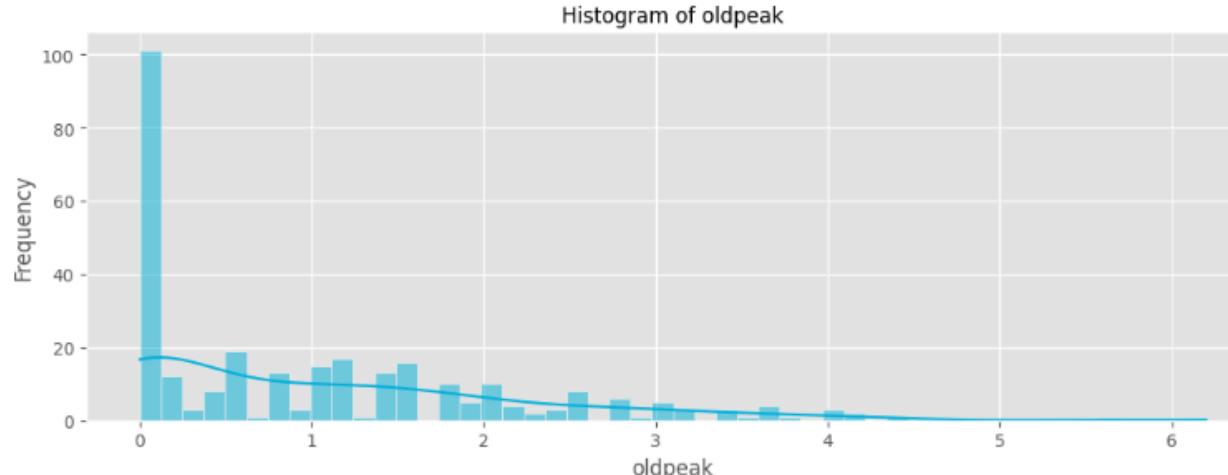
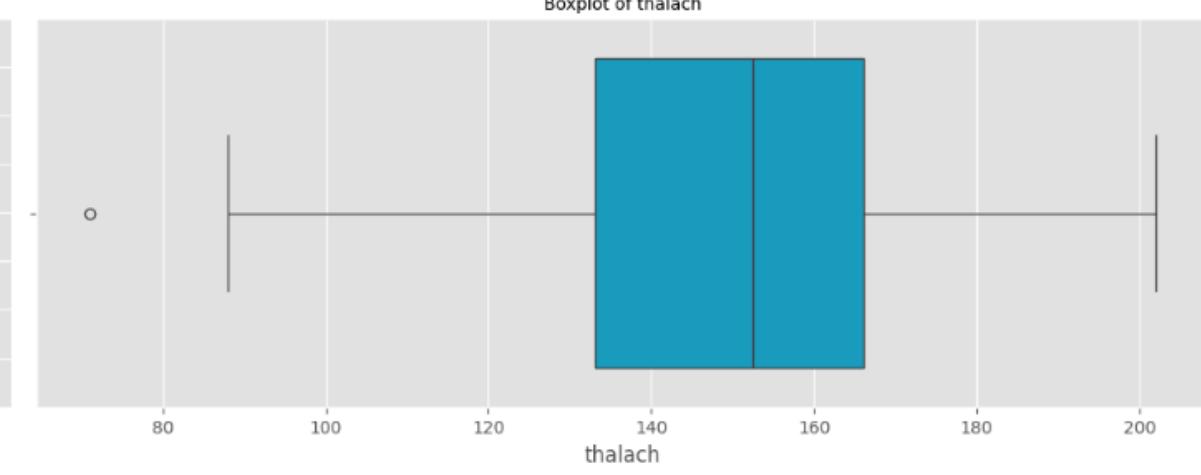
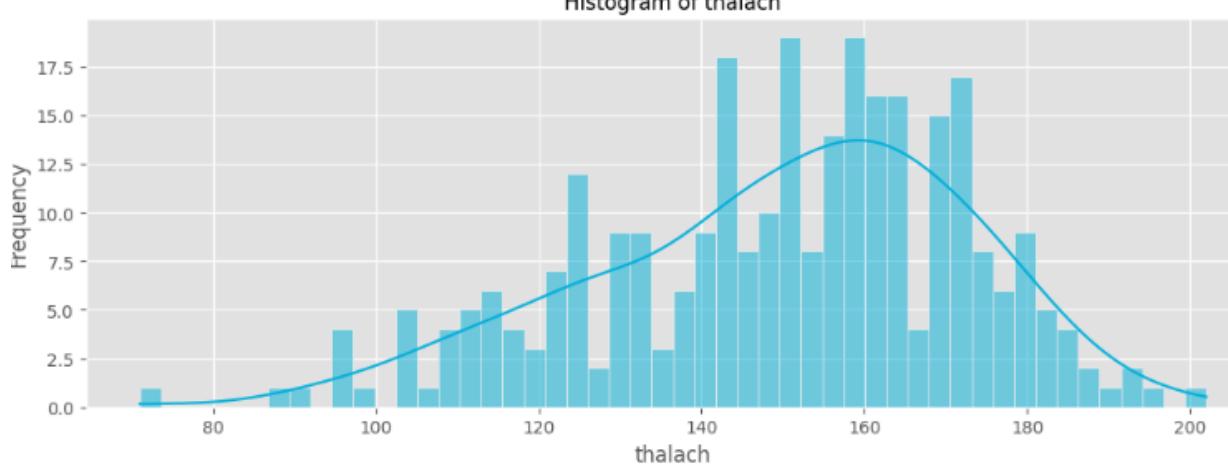
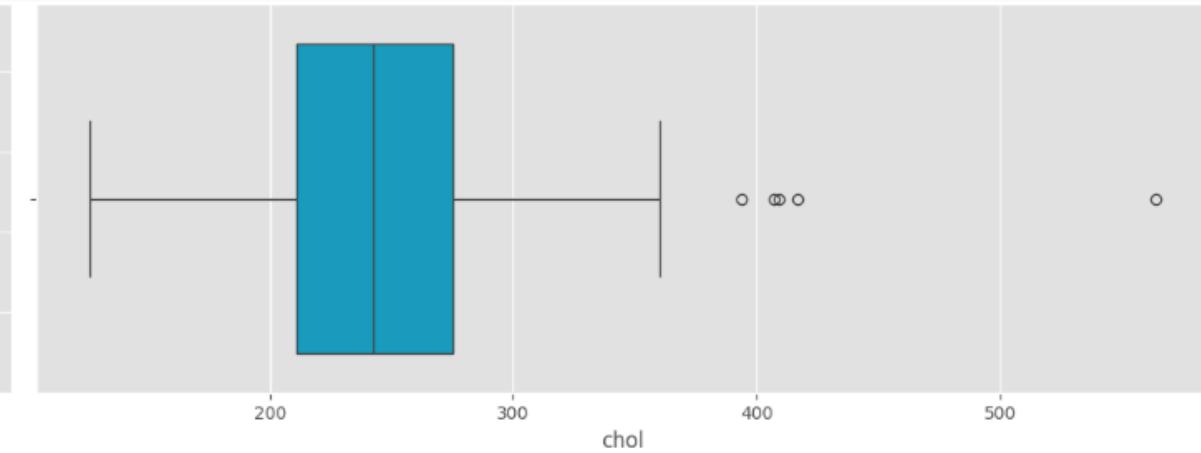
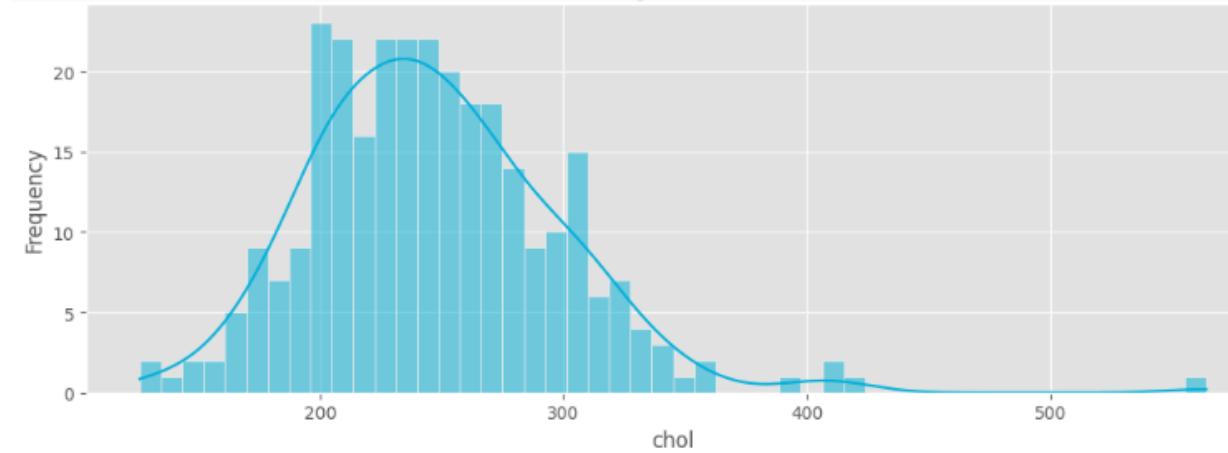
Target ←

Feature	Description
age	The age of the patient measured in years.
sex	The gender of the patient with a value of 1 for male and 0 for female.
cp	The type of chest pain perceived by the patient with 4 possible category values: (1) indicates typical angina chest pain (2) indicates atypical angina chest pain (3) indicates non-anginal pain (4) indicates asymptomatic chest pain.
trestbps	Resting blood pressure or The patient's blood pressure at rest, measured in mmHg (millimeters of mercury).
chol	The serum cholesterol level in the patient's blood, measured in mg/dl (milligrams per deciliter).
fbs	The patient's fasting blood sugar level with a value of 1 if blood sugar level > 120 mg/dl and a value of 0 otherwise.
restecg	The resting electrocardiographic results of patients with 3 possible category values: (0) indicates normal results (1) indicates ST-T wave abnormality (2) indicates left ventricular hypertrophy.
thalach	The maximum heart rate achieved by the patient during exercise testing, measured in bpm (beats per minute).
exang	Exercise induced angina, This variable represents whether the patient experienced exercise-induced angina (chest pain) (0) Patients did not experiencing exercise-induced angina (1) Patients experienced exercise-induced angina
oldpeak	The amount of ST segment depression during physical activity compared to rest.
slope	The slope of the ST segment on the electrocardiogram (EKG) during maximal exercise with 3 category values.
ca	The number of major blood vessels (0-3) visible on fluoroscopy examination.
thal	The result of the thallium scan test with 3 possible category values: (1) indicates normal condition. (2) indicates fixed defect in thalassemia. (3) indicates reversible defect in thalassemia.
Target	The target result of the prediction test (0) Predicted with No heart disease (1) Predicted with Heart disease

NUMERICAL - FEATURES

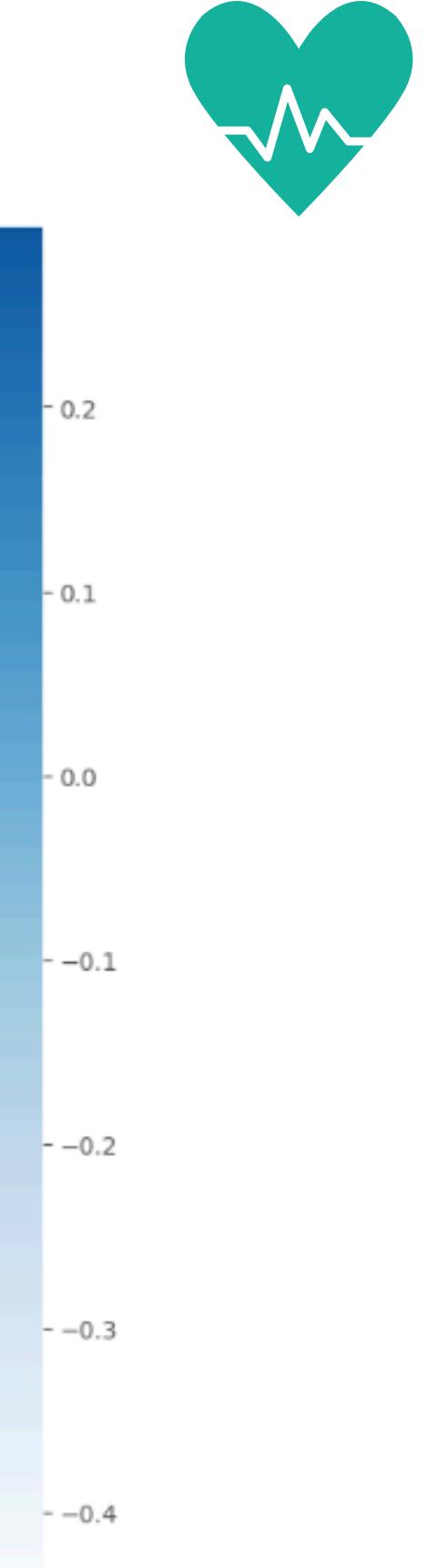


NUMERICAL - FEATURES

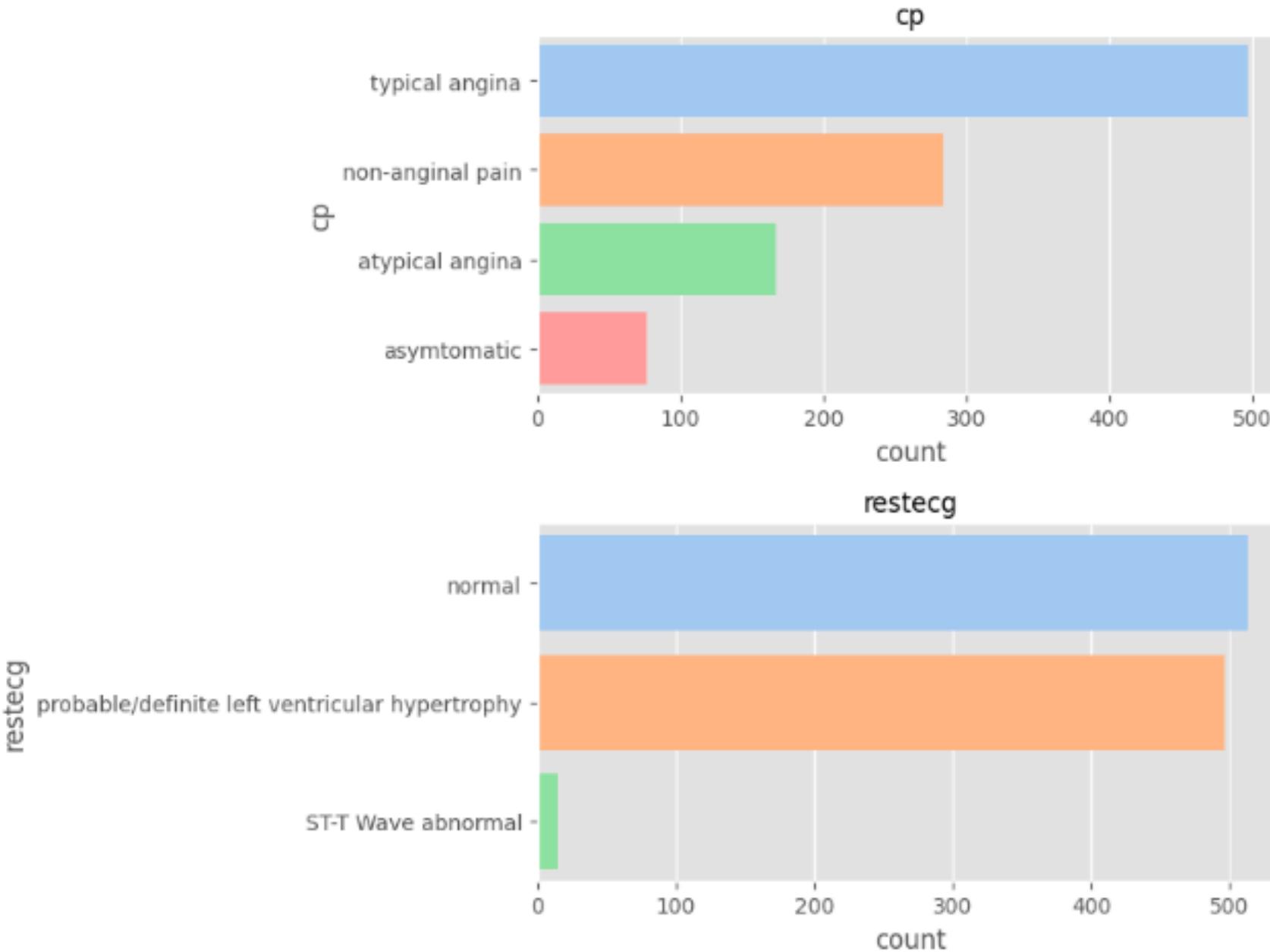
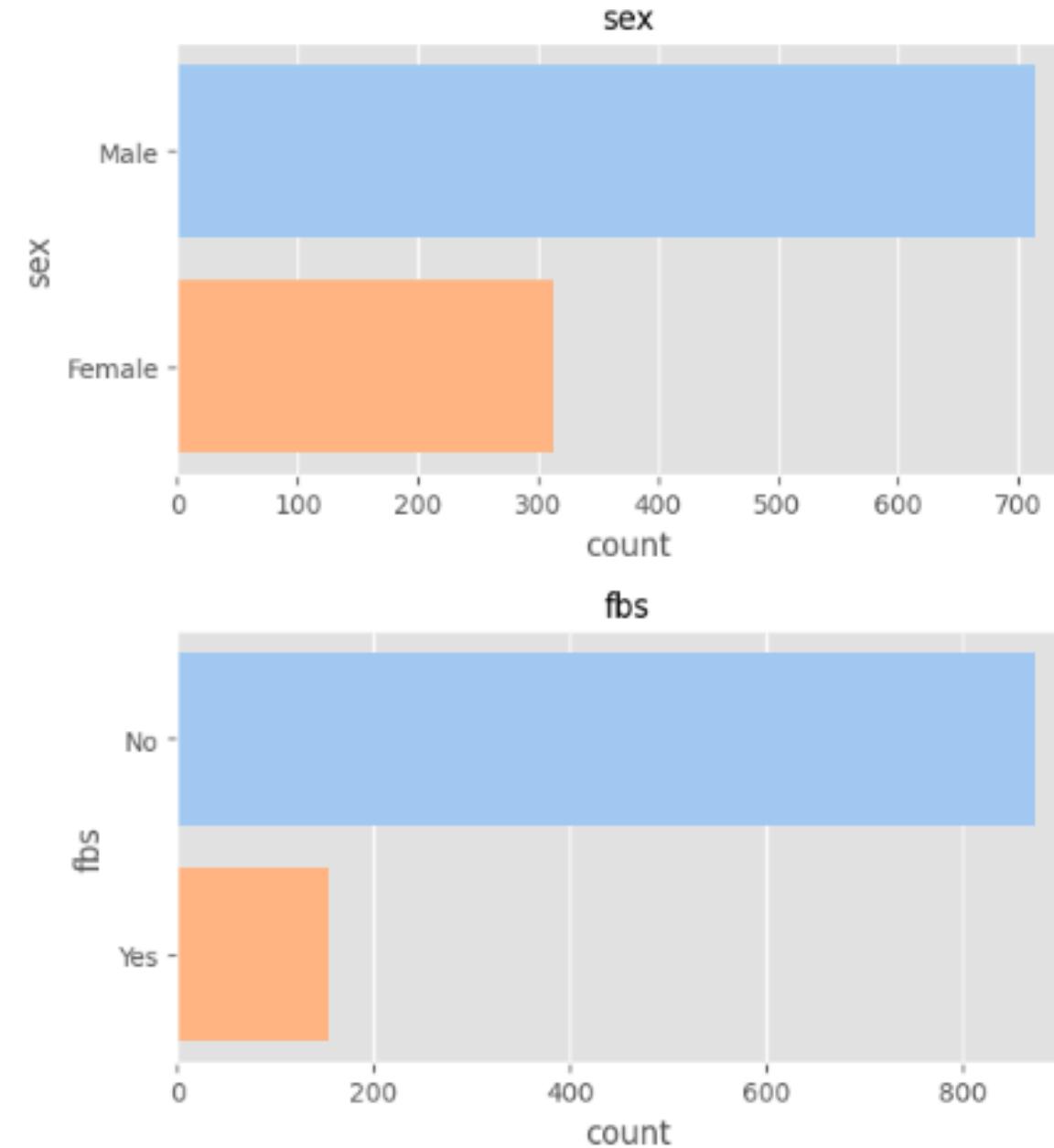


NUMERICAL - EDA

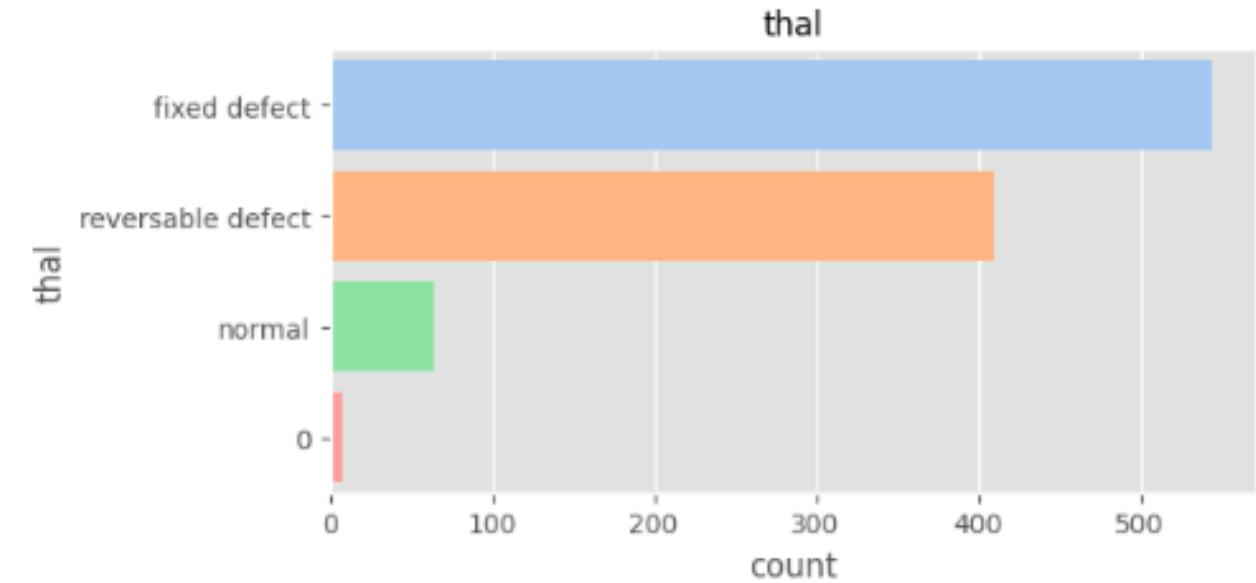
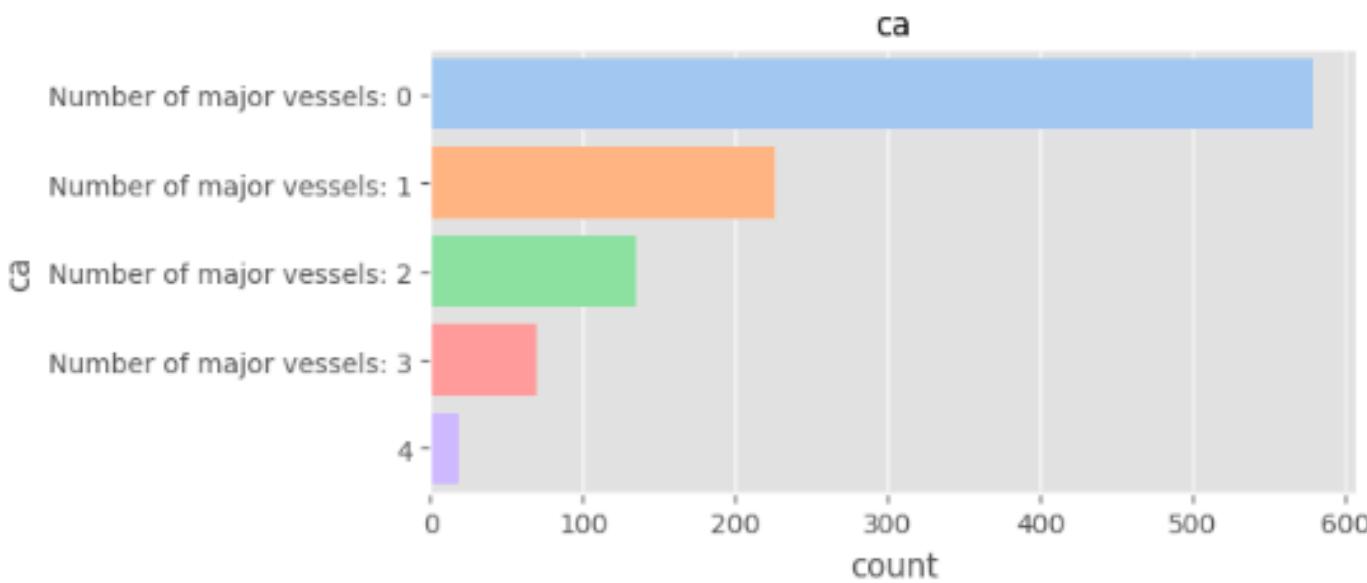
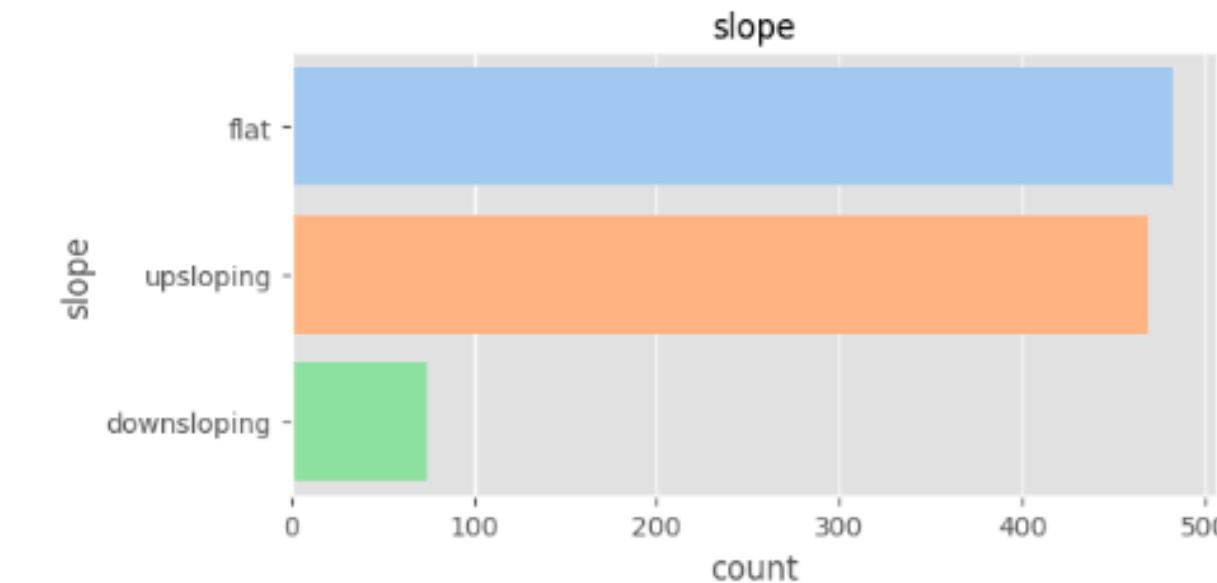
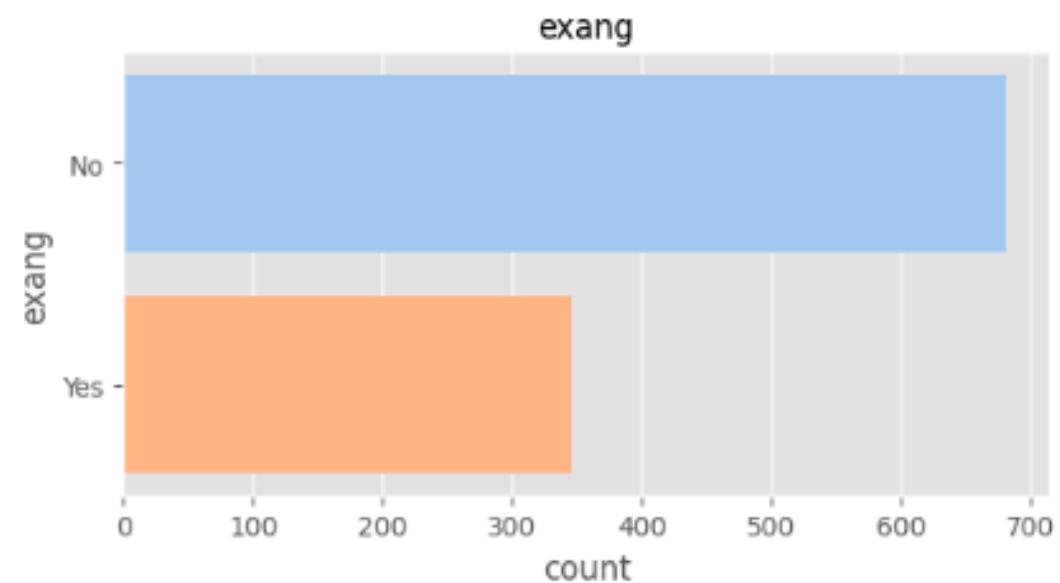
Spearman - correlation heatmap



CATEGORICAL FEATURES



CATEGORICAL FEATURES

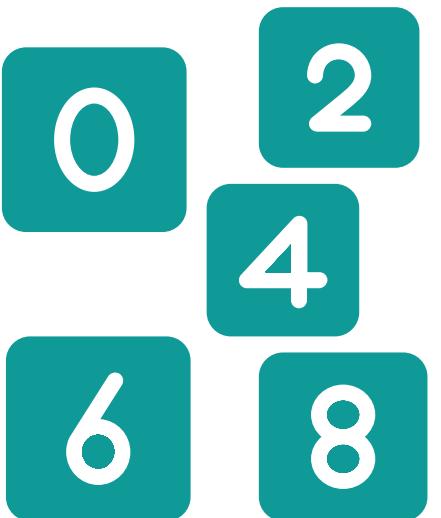


Exploratory Data Analysis (EDA)

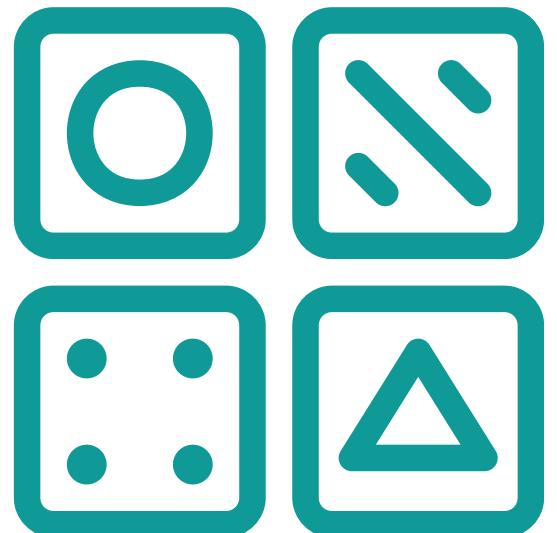
EXPLORE



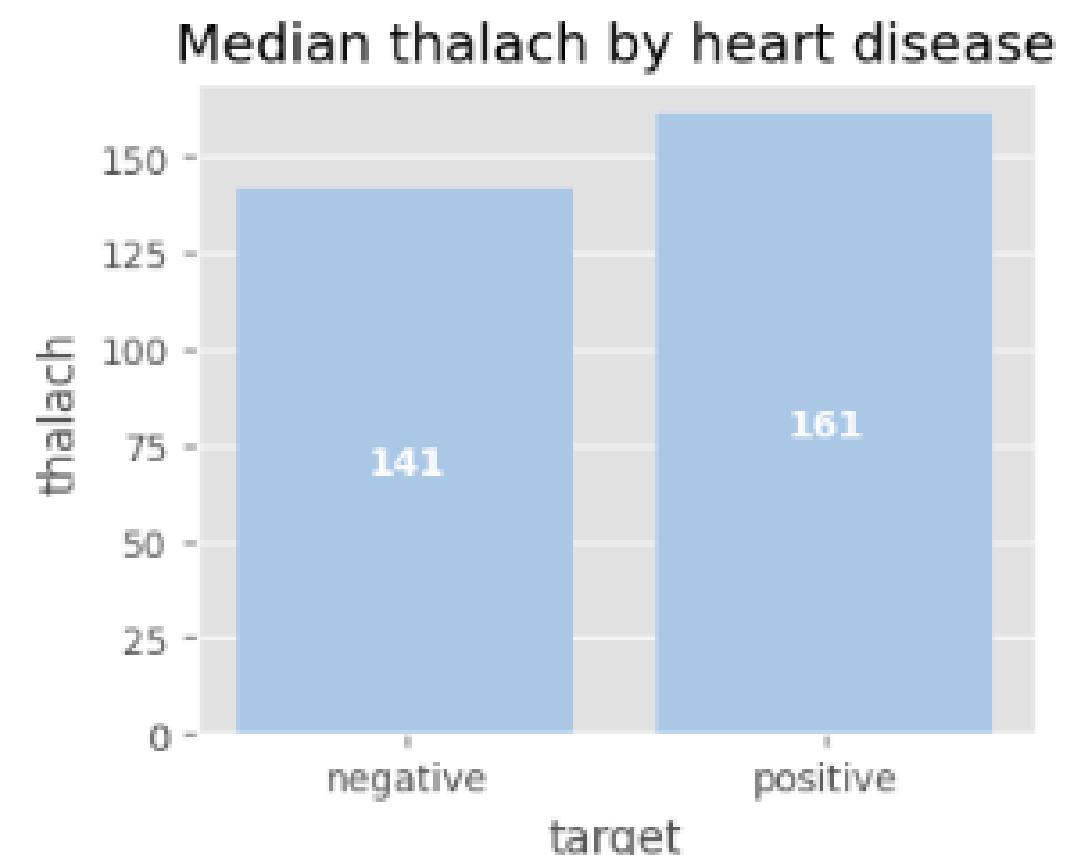
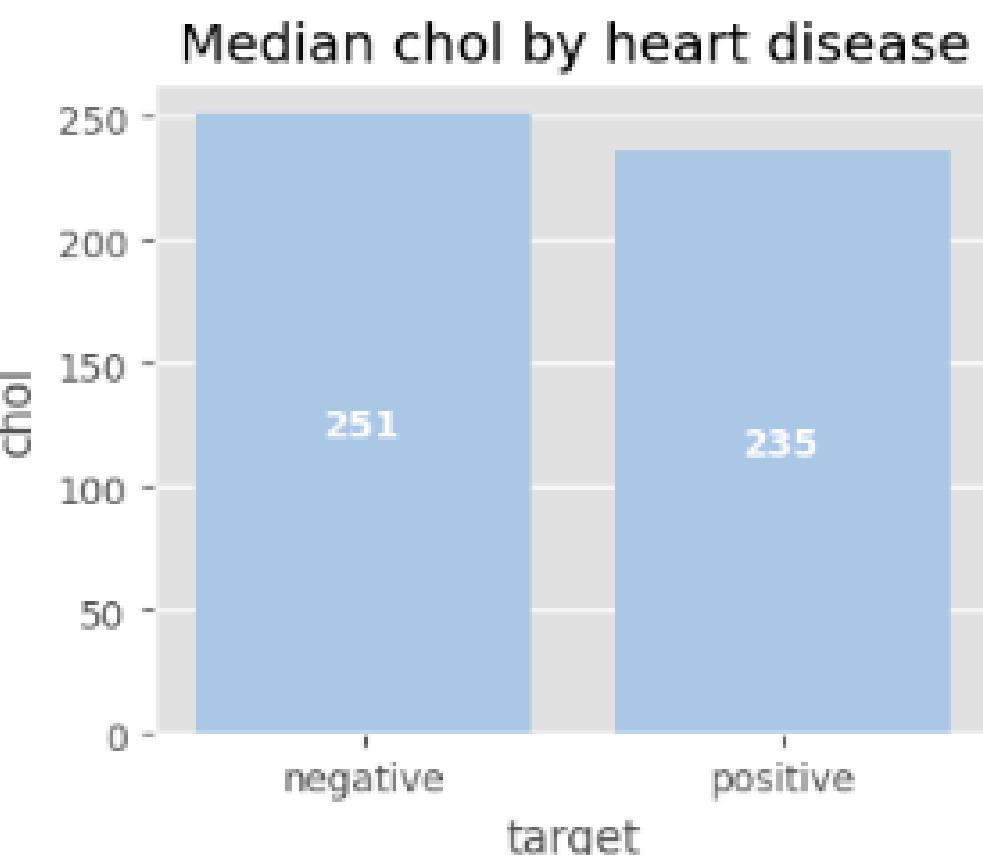
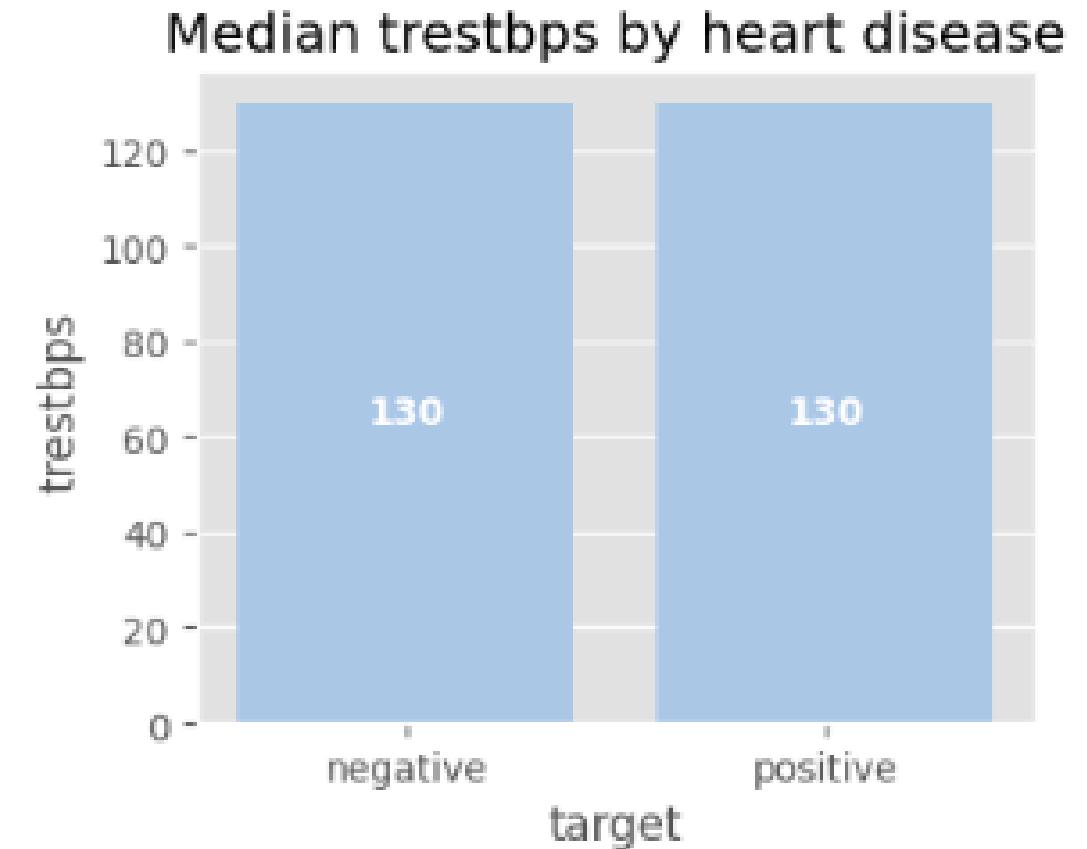
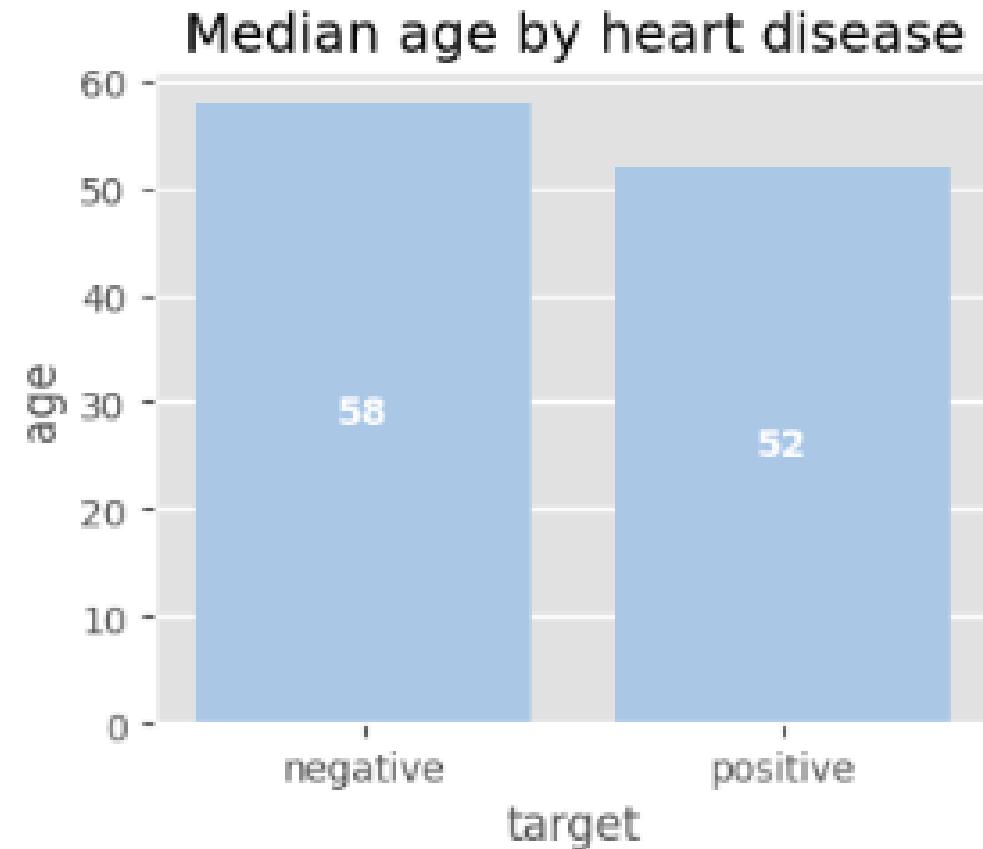
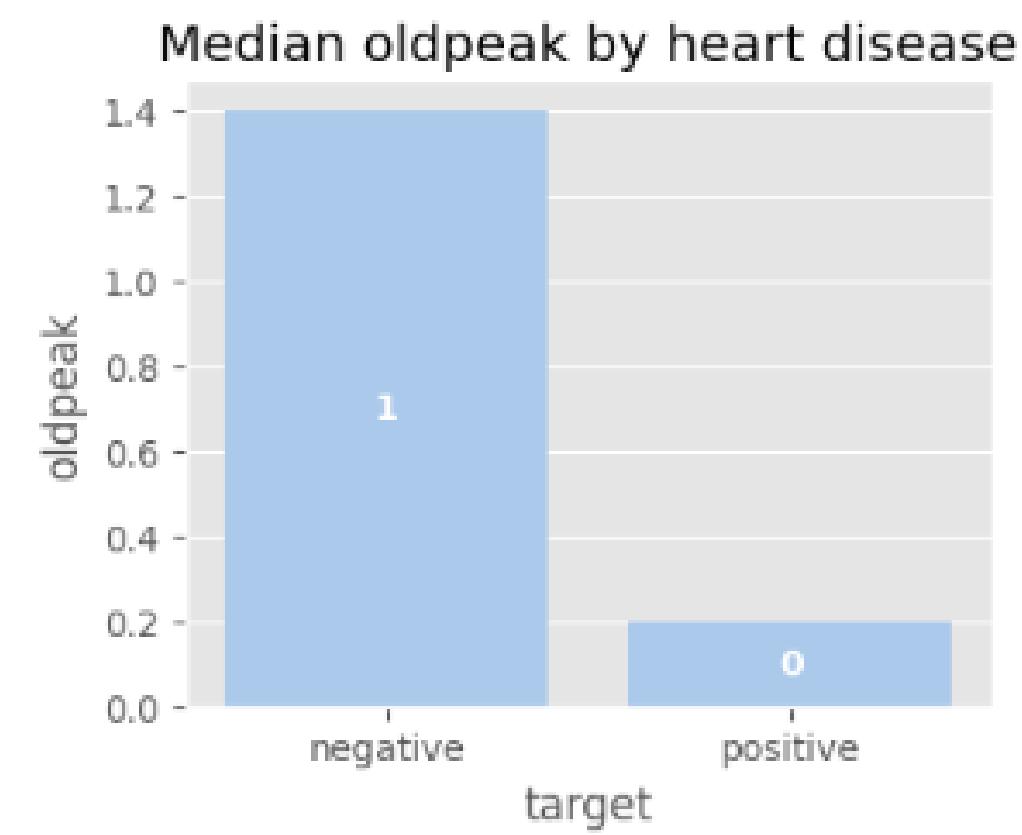
Numerical vs Target



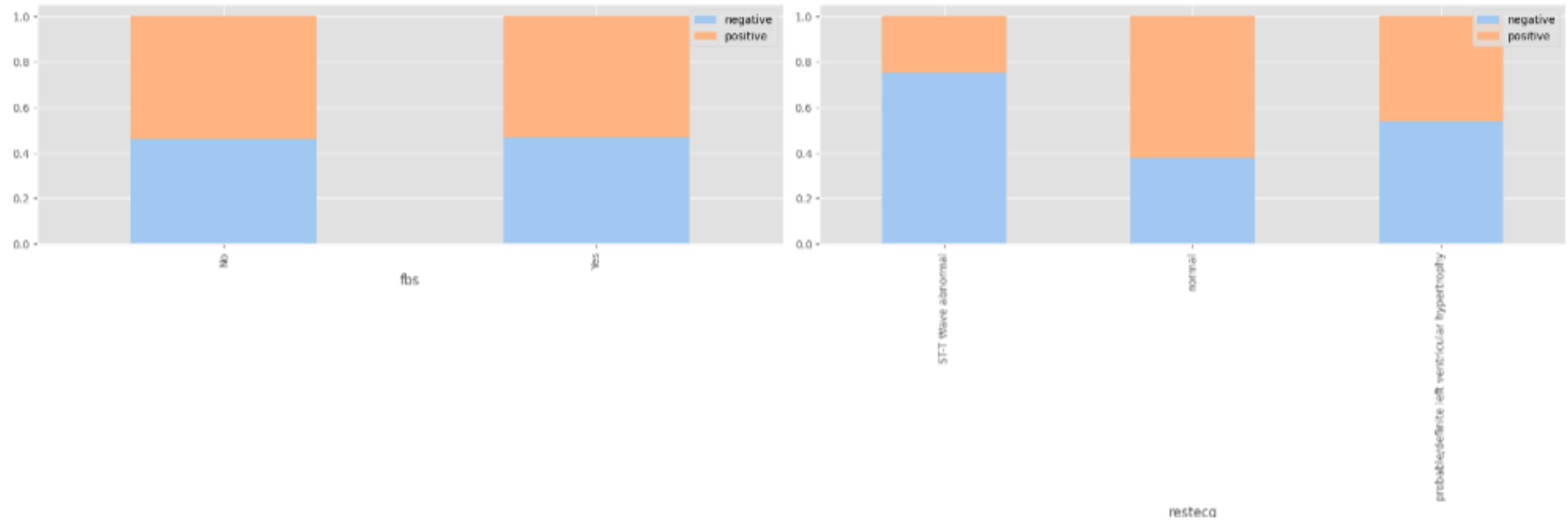
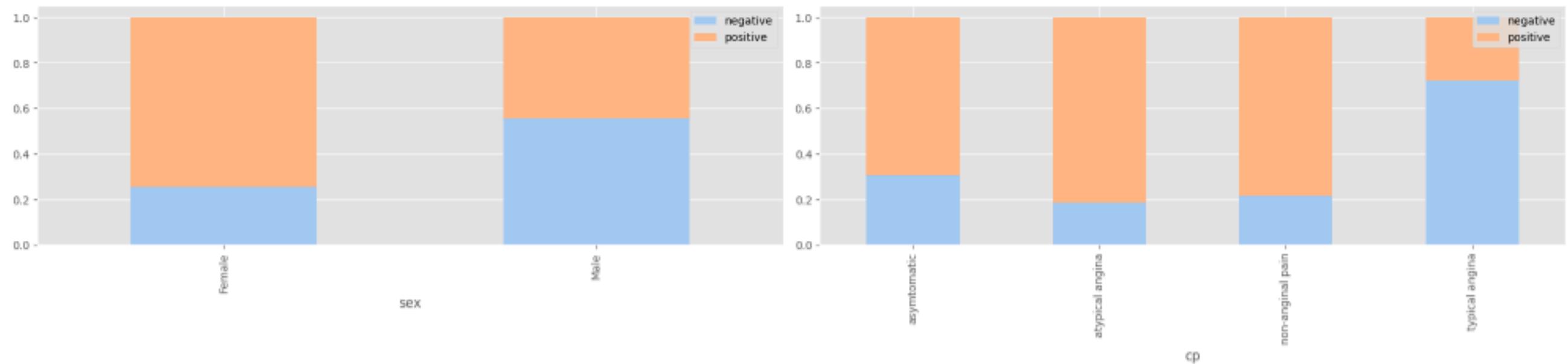
Categorical vs Target



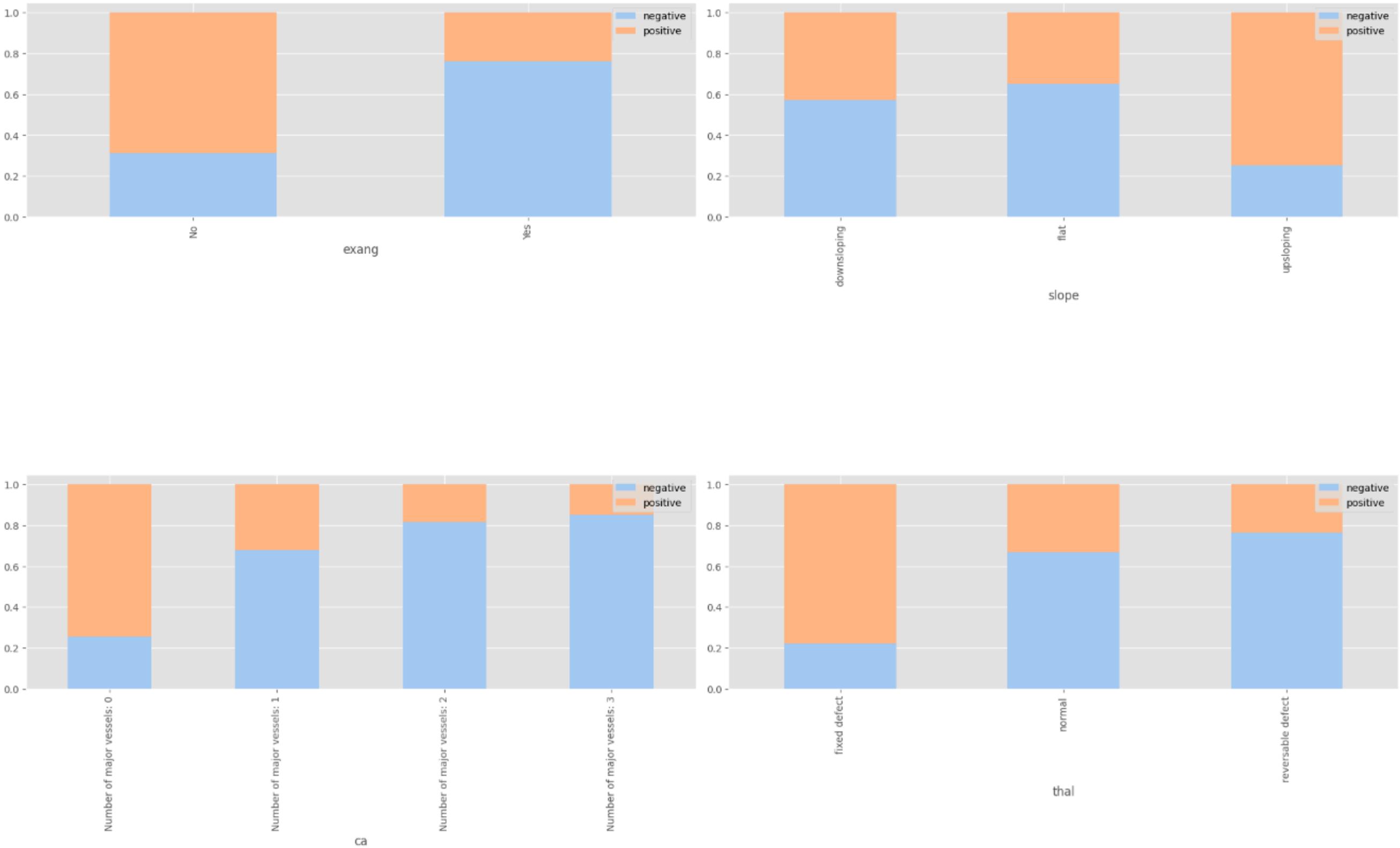
NUMERICAL VS TARGET



CATEGORICAL VS TARGET

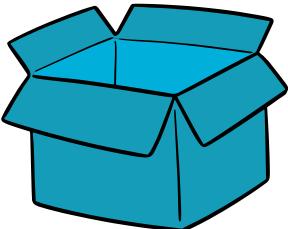


CATEGORICAL VS TARGET



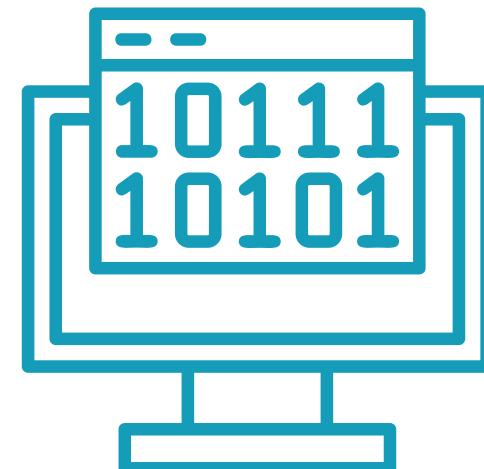
Preprocessing

Handling Outliers
Winsorization



Encoding

OneHot Encoder



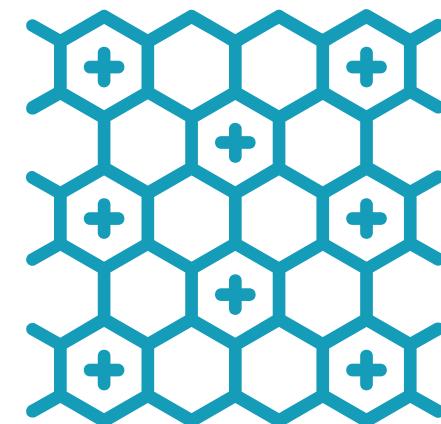
Scaling

MinMaxScaler



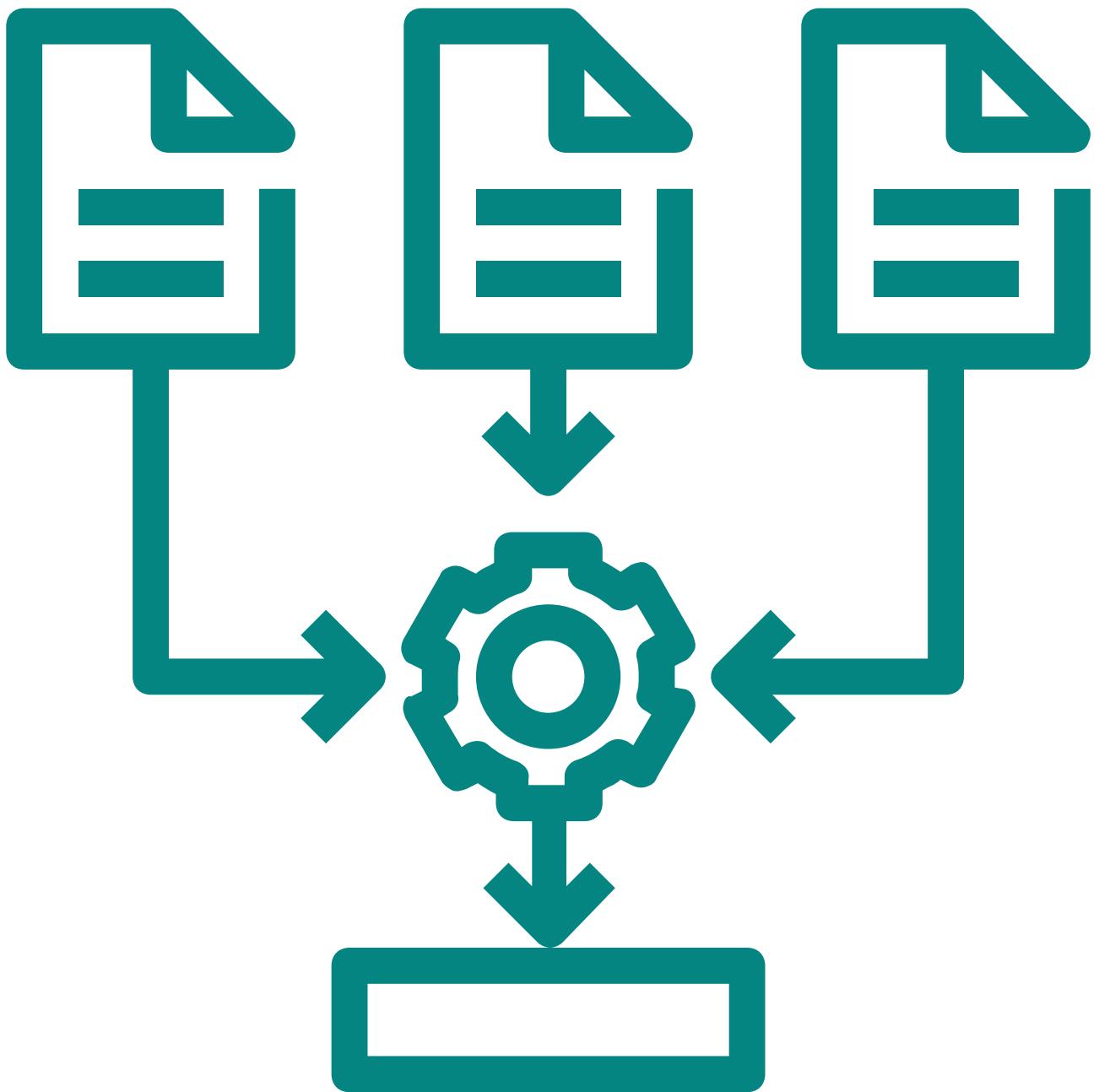
Resampling

SMOTE & SMOTE NC



Benchmark Model

- 01** Logistic Regression
- 02** KNN Classifier
- 03** Decision Tree Classifier
- 04** Random Forest Classifier
- 05** Adaptive Booster Classifier
- 06** Gradient Booster Classifier
- 07** Categorical Booster Classifier
- 08** XGBoost Classifier
- 09** LGBM Classifier



Model Performance

model	mean recall train	mean recall test
KNN	0.851385	0.84375
Logistic Regression	0.851692	0.81250
Random Forest	0.867385	0.81250
LightGBM	0.820308	0.75000
AdaBoost	0.813231	0.71875
GradienBoost	0.828308	0.71875
CatBoost	0.875385	0.71875
XGBoost	0.812308	0.71875
Decision Tree	0.726769	0.68750



KNN



Logistic regression



Random Forest

Hyperparameter Tuning

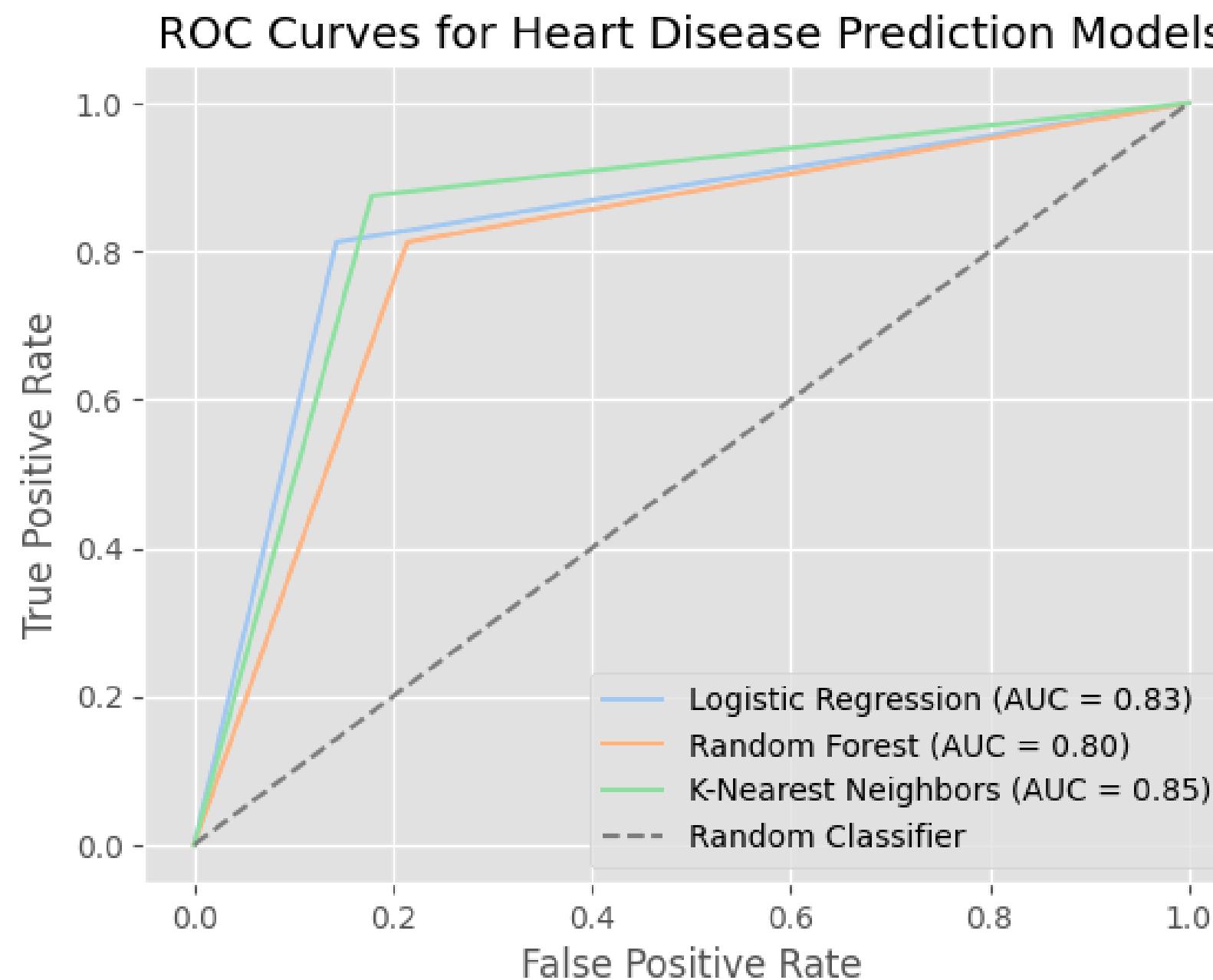
Model Performance Before & After Tuning

Model	Conditions	Train score	Test score
KNN Classifier	Before Tuning	0.851	0.844
KNN Classifier	After Tuning	0.891	0.875
Random Forest Classifier	Before Tuning	0.867	0.813
Random Forest Classifier	After Tuning	0.891	0.813
Logistic Regression Classifier	Before Tuning	0.851	0.813
Logistic Regression Classifier	After Tuning	0.851	0.813



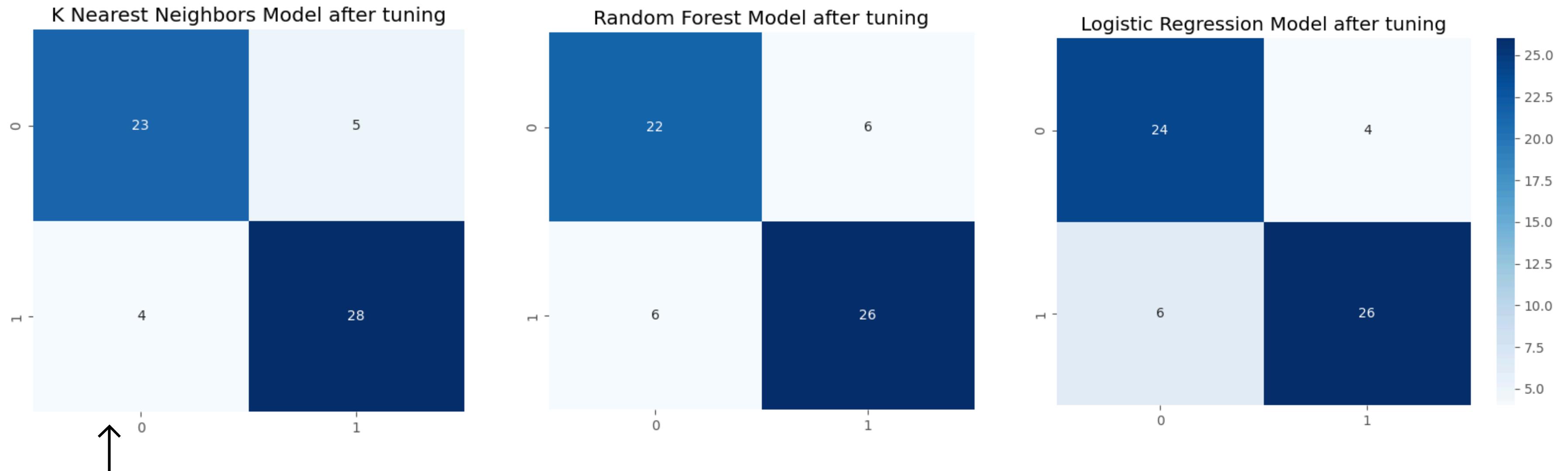
Tuned KNN is the best model

ROC curves



Tuned KNN is the best model

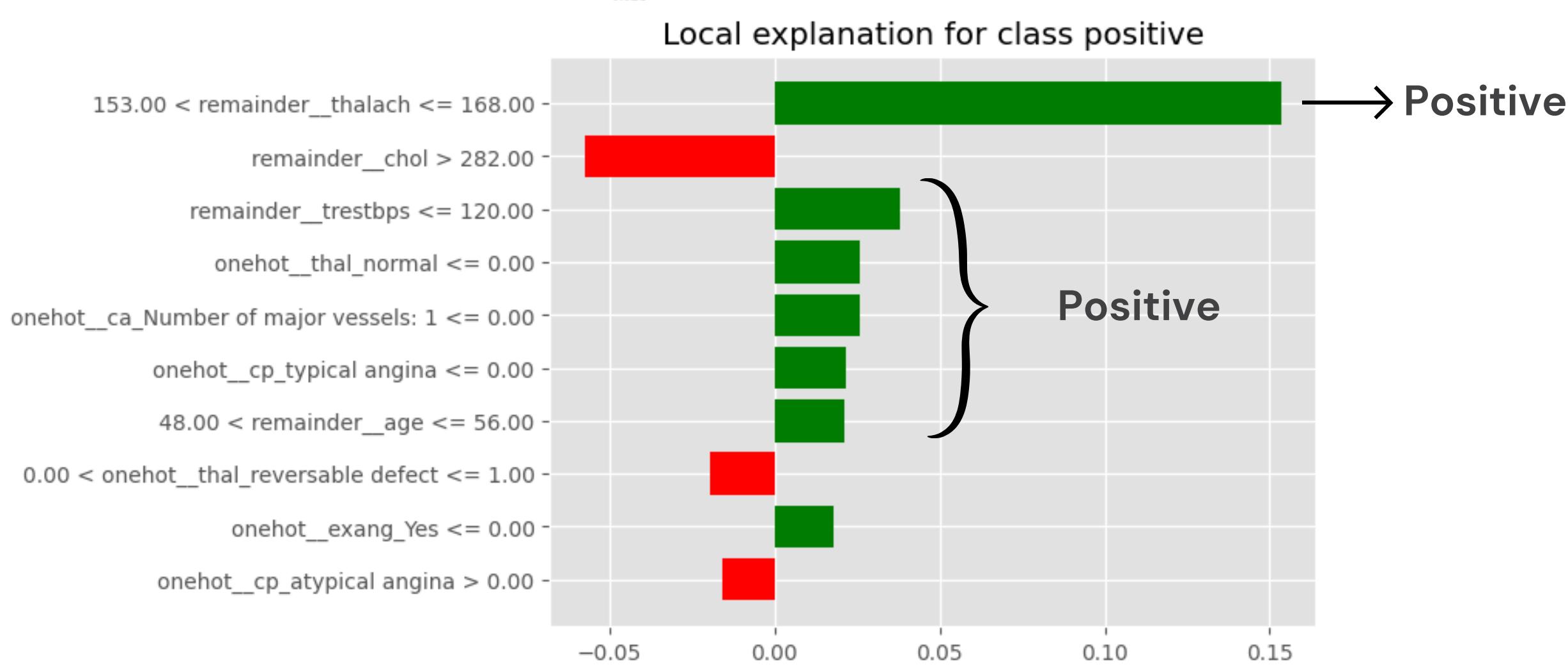
Confusion matrix



The Lowest False Negative

Explainable Model LIME (Local Interpretable Model-agnostic)

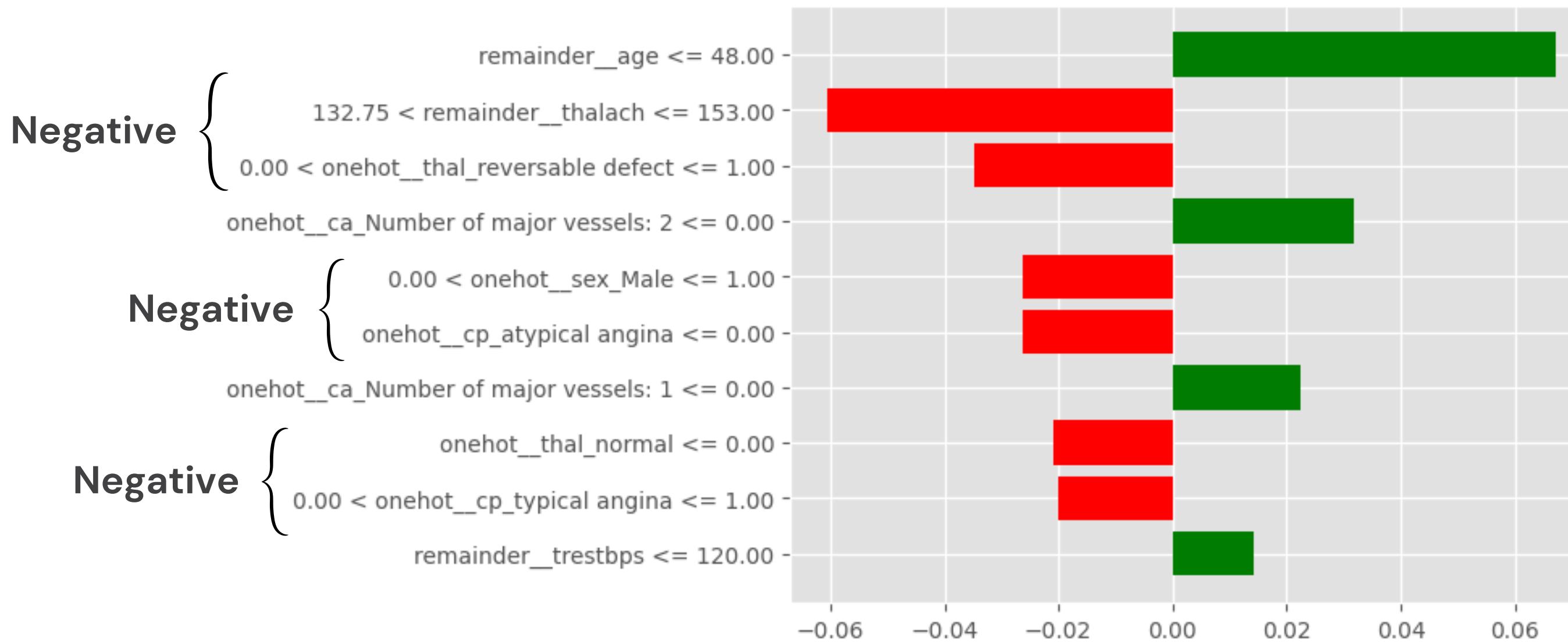
Patient [10] – Positive Heart Diseases



Explainable Model

LIME (Local Interpretable Model-agnostic)

Patient [11] – Negative Heart Diseases



CONCLUSION

Tuned KNN is the best model for predicting heart disease in this scenario, due to its high recall of 0.891, F1-score 0.86 and best ROC AUC score of 0.85.

APP DEMO IN STREAMLIT



Upload your input CSV file

Drag and drop file here
Limit 200MB per file • CSV

Browse files

Manual Input

Chest pain type: 2

Type of Chest pain: Atypical angina

Maximum heart rate achieved: 80

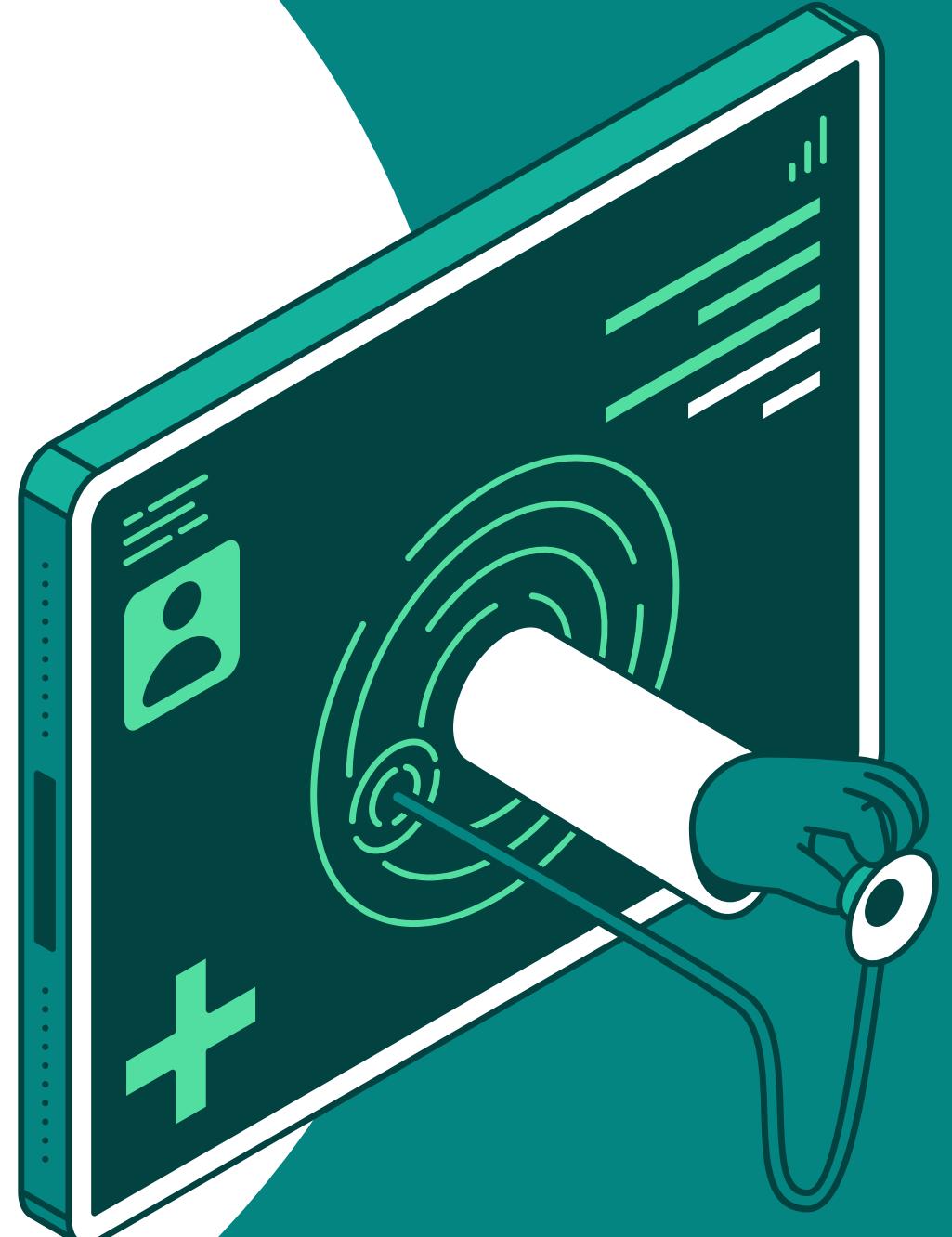
Slope of the peak exercise ST segment: 1

ST depression induced: 1.00

Heart Disease Predictor App

- ☞ This app predicts symptoms of heart disease

The dataset for this prediction was obtained from the [Heart Disease dataset](#) by UC Irvine ML repository .



THANK YOU

My Contact

 <https://github.com/harishmuh>

 www.linkedin.com/in/harish-muhammad-7b600b102/

 harishmuh@gmail.com

