Predicting Heart Disease

Project Overview & Goals

Data Analysis and Visualization

We analyze and visualize heart disease data to discover important patterns and insights for model development.

Predictive Model Development

Our goal is to build a reliable model that predicts heart disease presence or absence in individuals.

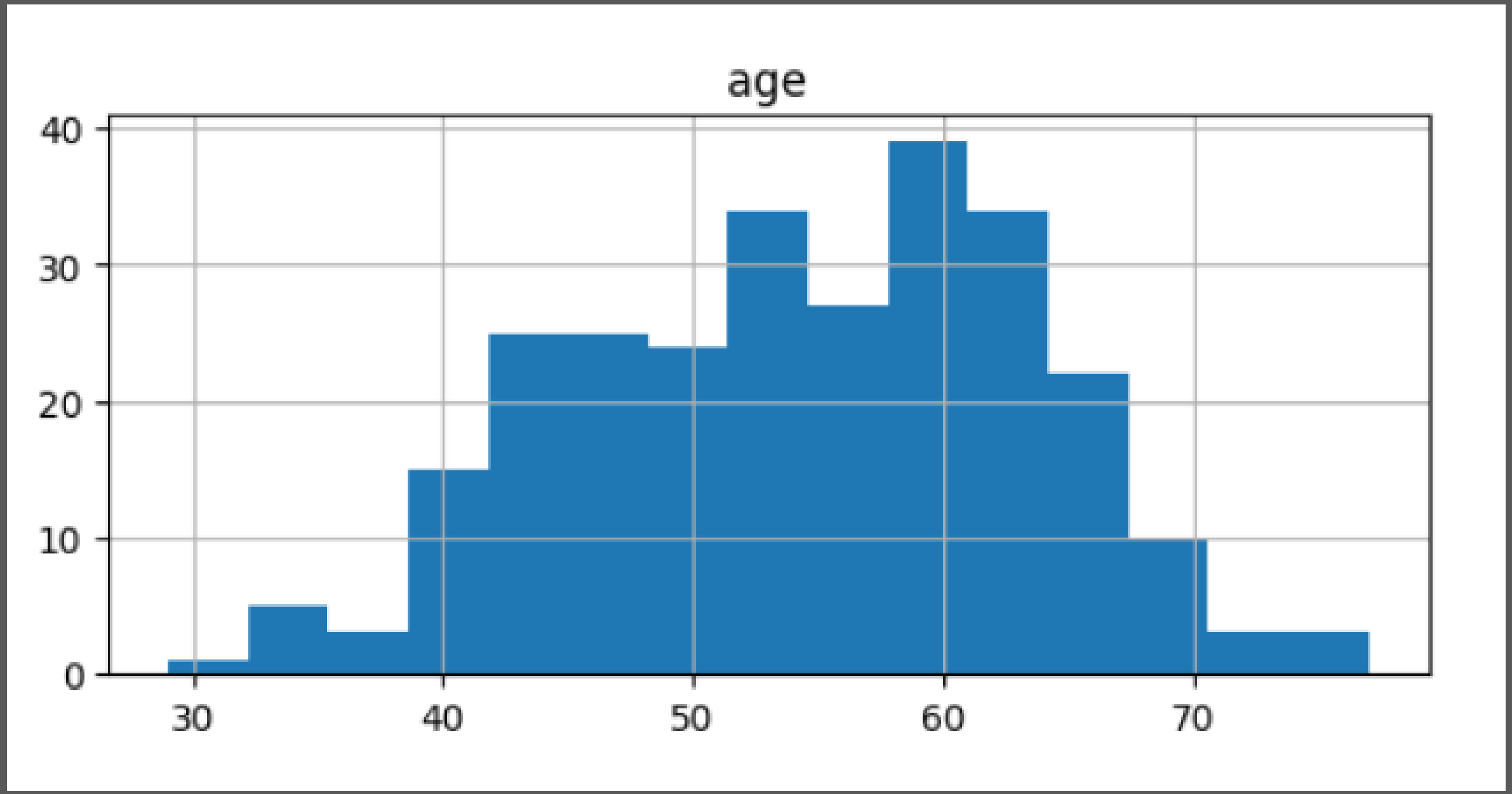
Feature Extraction and Classification

We select key features from the data and use them to classify individuals with high accuracy.

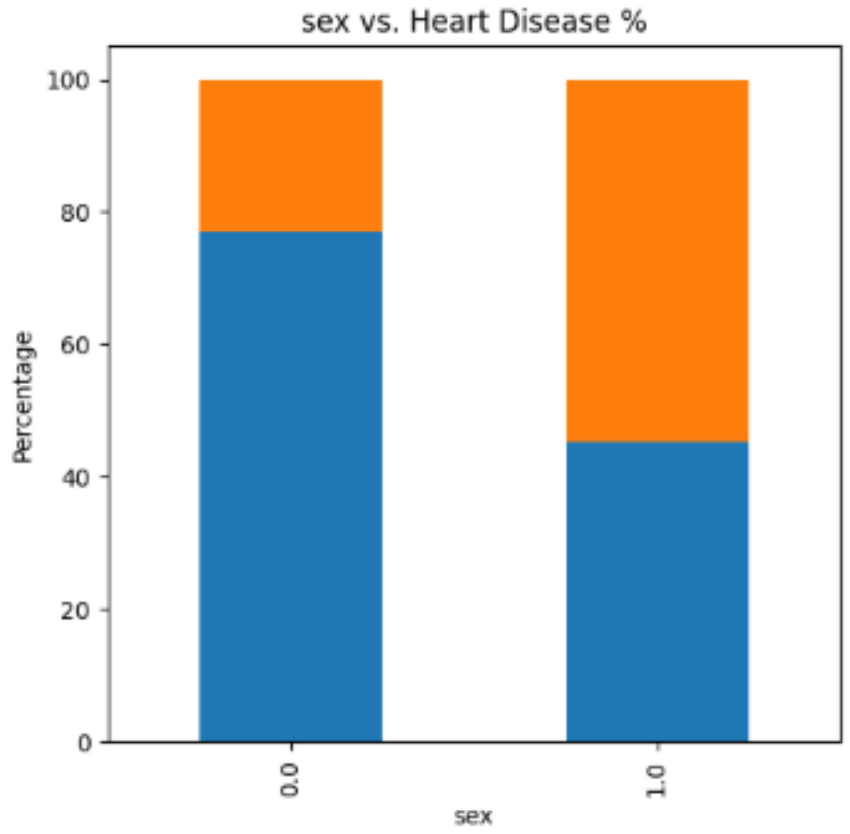
```
df.head(10)
```

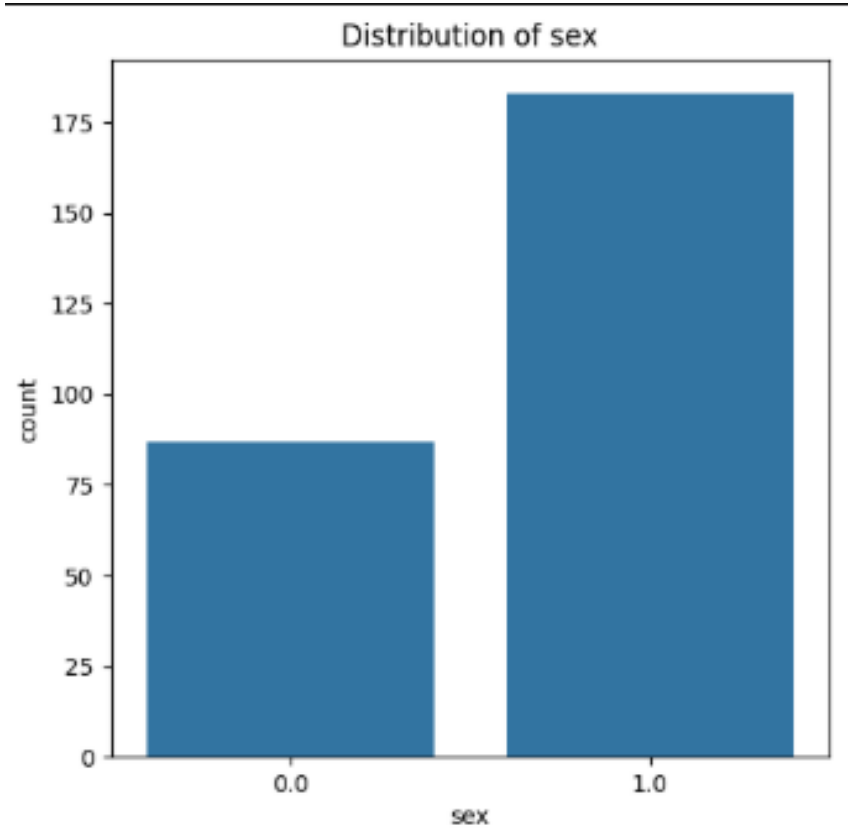
	age	sex	chest-pain	rest-bp	serum-chol	fasting-blood-sugar	electrocardiographic	max-heart-rate	angina	oldpeak	slope	major-vessels	thal	heart-dise
0	70.0	1.0	4.0	130.0	322.0	0.0	2.0	109.0	0.0	2.4	2.0	3.0	3.0	
1	67.0	0.0	3.0	115.0	564.0	0.0	2.0	160.0	0.0	1.6	2.0	0.0	7.0	
2	57.0	1.0	2.0	124.0	261.0	0.0	0.0	141.0	0.0	0.3	1.0	0.0	7.0	
3	64.0	1.0	4.0	128.0	263.0	0.0	0.0	105.0	1.0	0.2	2.0	1.0	7.0	
4	74.0	0.0	2.0	120.0	269.0	0.0	2.0	121.0	1.0	0.2	1.0	1.0	3.0	
5	65.0	1.0	4.0	120.0	177.0	0.0	0.0	140.0	0.0	0.4	1.0	0.0	7.0	
6	56.0	1.0	3.0	130.0	256.0	1.0	2.0	142.0	1.0	0.6	2.0	1.0	6.0	
7	59.0	1.0	4.0	110.0	239.0	0.0	2.0	142.0	1.0	1.2	2.0	1.0	7.0	
8	60.0	1.0	4.0	140.0	293.0	0.0	2.0	170.0	0.0	1.2	2.0	2.0	7.0	
9	63.0	0.0	4.0	150.0	407.0	0.0	2.0	154.0	0.0	4.0	2.0	3.0	7.0	

Data Understanding and Visual

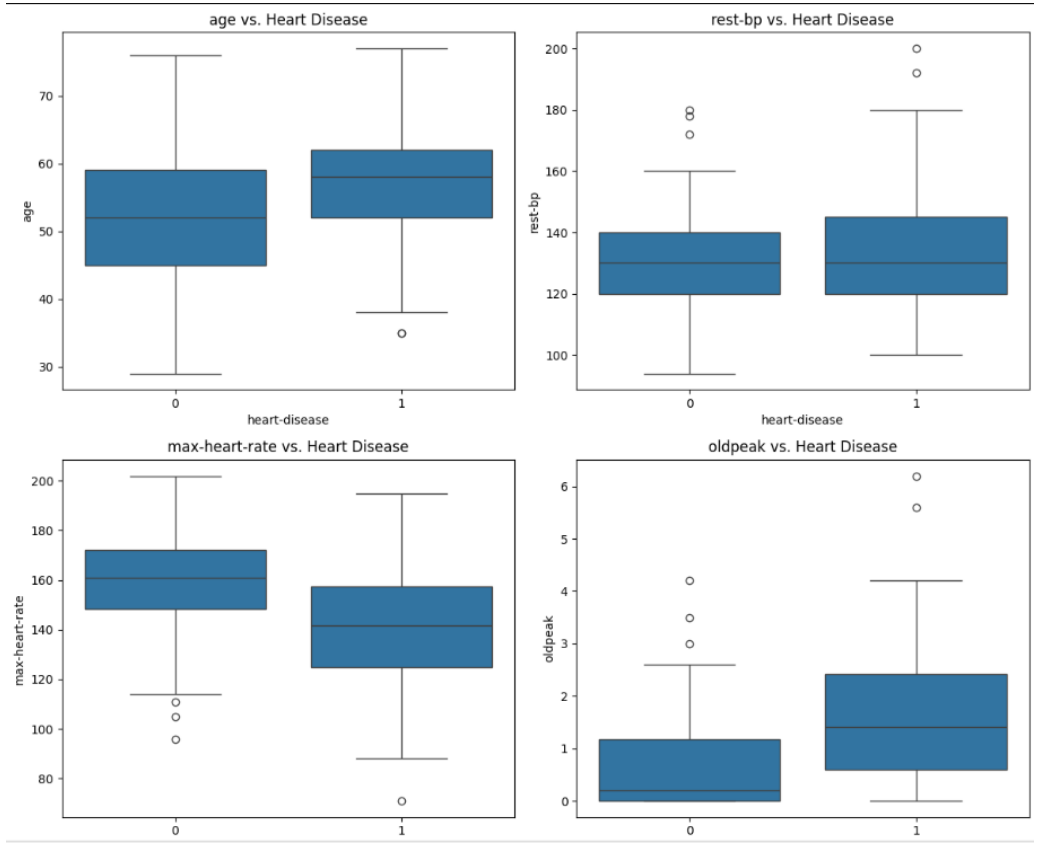

age
40
30
20
10
0
30
40
50
60
70

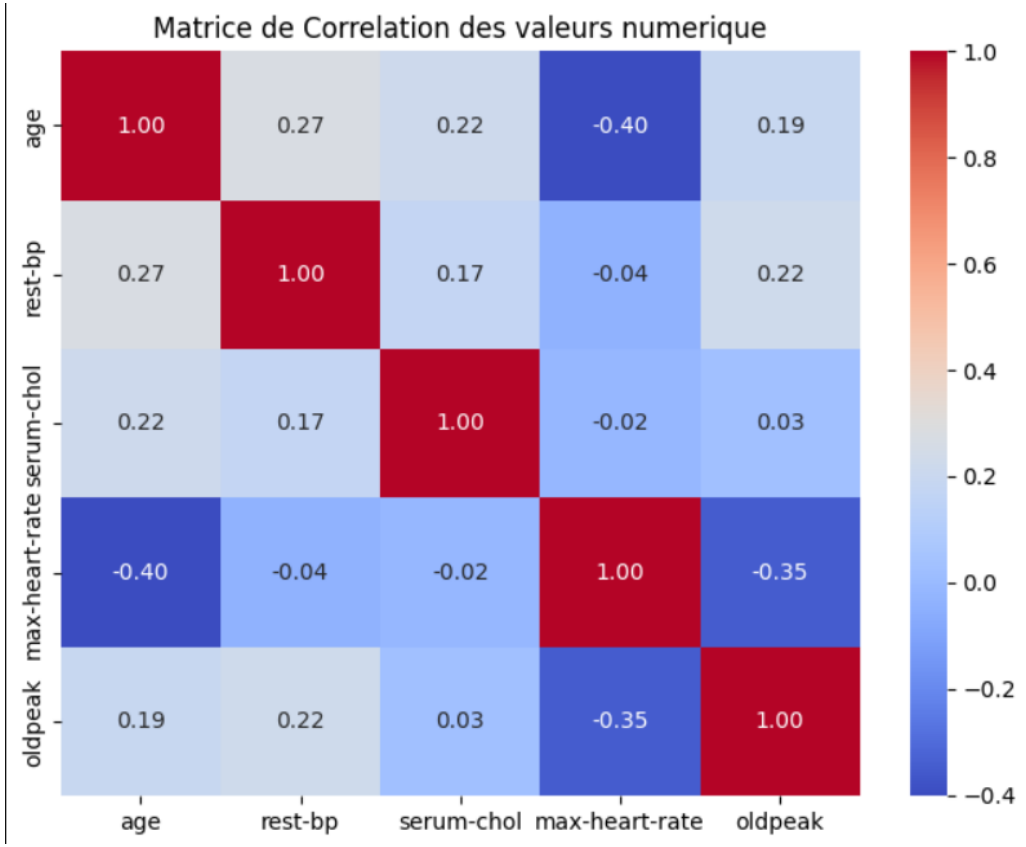
Distribution of Sex





Others Visualizations





Data Science Methodology and Modelisation

Data Collection and Preprocessing

Projects start with gathering and cleaning data to ensure quality and relevance for further analysis.

Logistic Regression for Classification

Logistic Regression is popular for binary classification, providing clear interpretation and probability estimates.

Random Forest for Robust Predictions

Random Forest combines multiple decision trees for accuracy, reducing overfitting and increasing predictive power.

Modele Initail : Performance de la Regression Logistique

Matrice de Confusion:

```
[[39  6]
 [ 2 34]]
```

Rapport de classification:

	precision	recall	f1-score	support
0	0.95	0.87	0.91	45
1	0.85	0.94	0.89	36
accuracy			0.90	81
macro avg	0.90	0.91	0.90	81
weighted avg	0.91	0.90	0.90	81

ROC AUC Score: 0.9364

Matrice de Confusion:

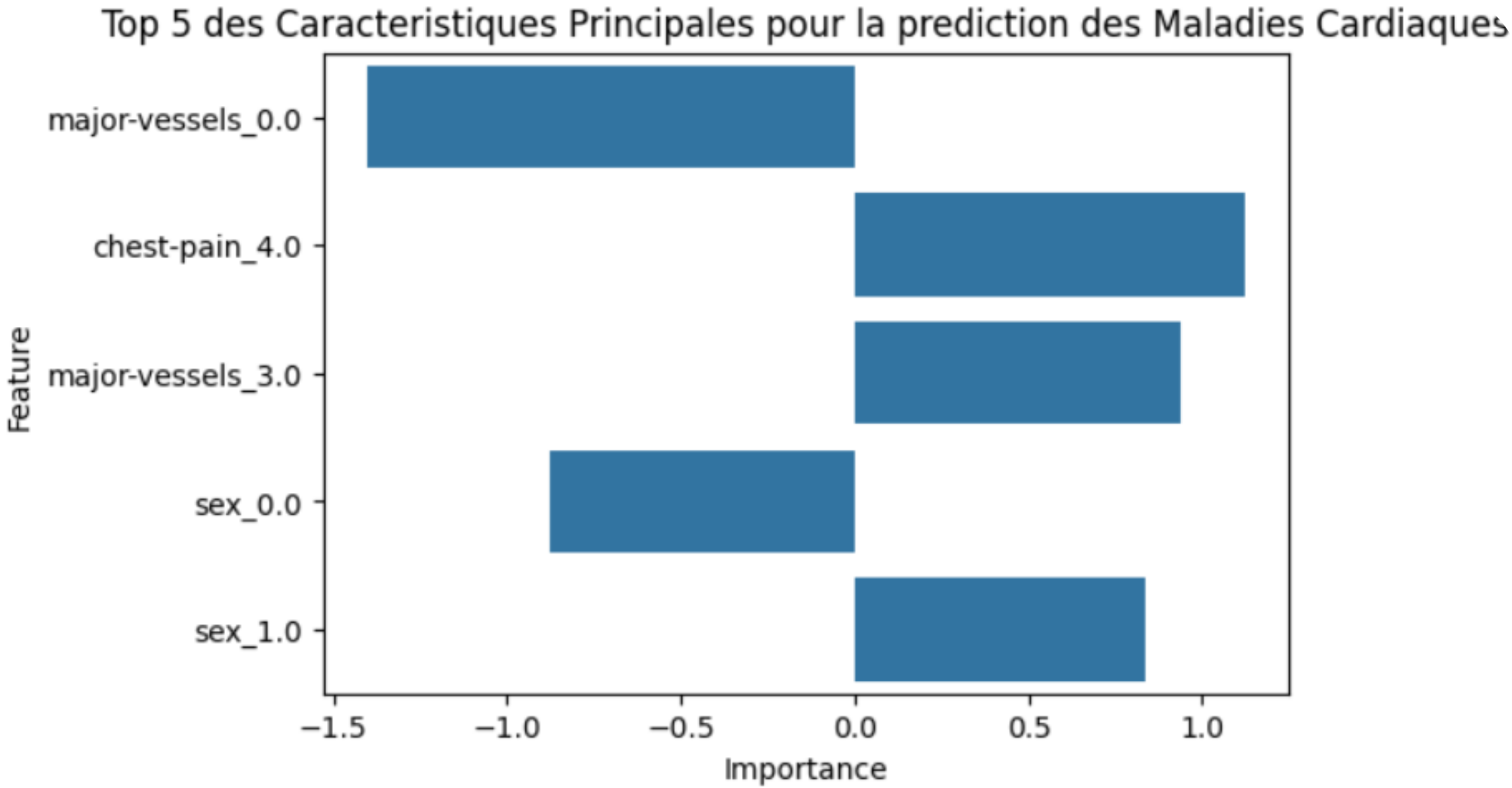
```
[[38  7]
 [ 5 31]]
```

Rapport de Classification:

	precision	recall	f1-score	support
0	0.88	0.84	0.86	45
1	0.82	0.86	0.84	36
accuracy			0.85	81
macro avg	0.85	0.85	0.85	81
weighted avg	0.85	0.85	0.85	81

ROC AUC Score: 0.9299

Results


Top 5 des Caracteristiques Principales pour la prediction des Maladies Cardiaques
major-vessels_0.0
chest-pain_4.0
major-vessels_3.0
sex_0.0
sex_1.0
Feature
−1.5
−1.0
−0.5
0.0
0.5
1.0
Importance

Key Cardiovascular Recommendations

Awareness on Nutrition and Lifestyle

Promoting awareness of nutrition and healthy lifestyles helps reduce the risk of heart disease by encouraging better daily habits.

Community Screening Initiatives

Accessible and affordable community screenings allow for early detection of cardiovascular risks using simple indicators like blood pressure and cholesterol.

National Cardiovascular Database

A national database enhances public health policy by improving data collection and enabling more targeted interventions for cardiovascular disease.

Thank you