

# CARDIOVASCULAR DISEASE DATASET ANALYSIS

7CCSONAR RESEARCH METHODOLOGIES: ANALYSIS AND REPORTING

ASSESSMENT PRESENTATION 2023-24

ALEJANDRO OTERO ORTIZ DE COSCA (K21090108)

# AGENDA



Introduction



Research Objectives



Analysis Results



Conclusions



# INTRODUCTION: THE DATASET

- Cardiovascular Disease dataset.
- Available in Kaggle.
- 1,000 patient records.
- 12 features – common in other studies.
- Patient classification.

# INTRODUCTION: DOMAIN CONTEXT (FEATURES)

## Demographic features

- Gender.
- Age.

## Blood analysis

- Serum cholesterol.
- Fasting blood sugar.

## Fluoroscopy

- Number of major vessels.

## Physical test

- Resting BP.
- Resting ECG.
- Maximum HR.
- Oldpeak and Slope.
- Exercise-induced angina.

## Self-reported

- Chest pain type





# RESEARCH OBJECTIVES

---

Prediction model.

---

*Understand dataset demographics.*

---

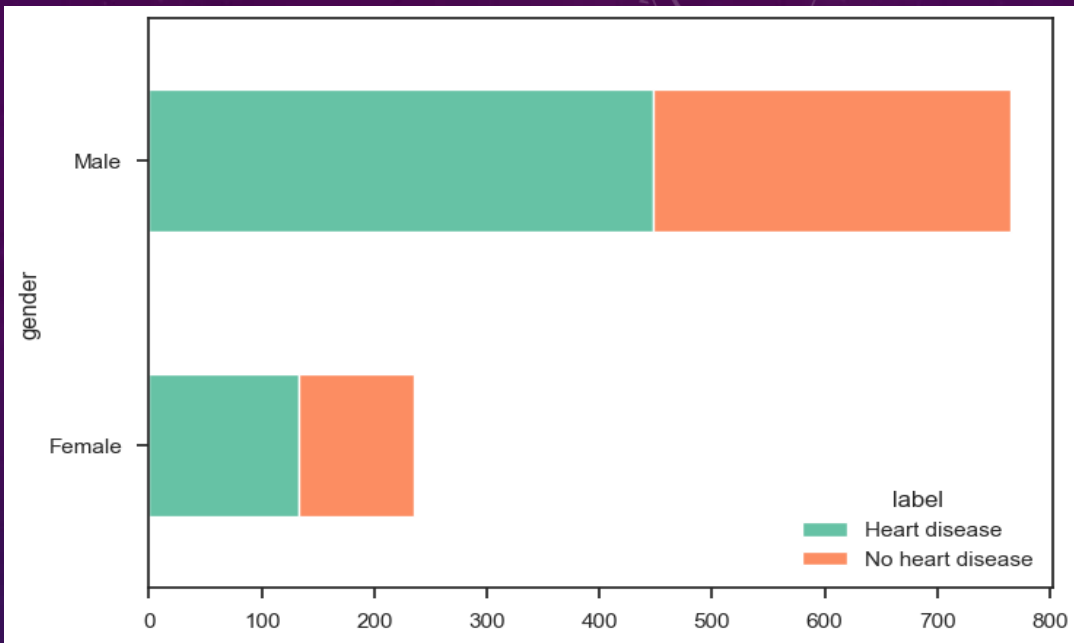
*Find indicators in categorical data.*

---

*Find indicators in numerical data.*



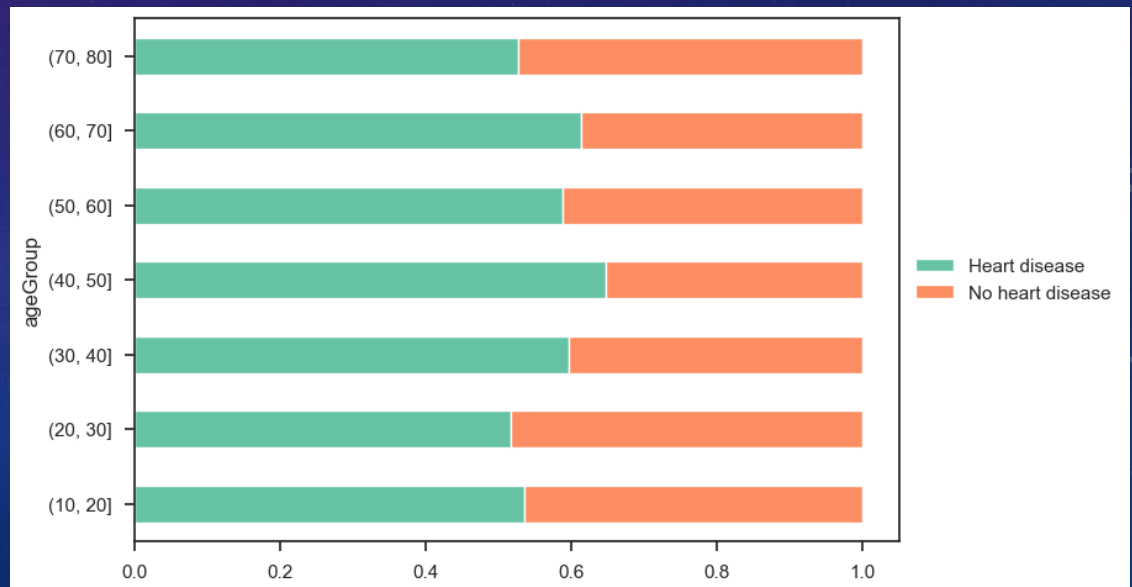
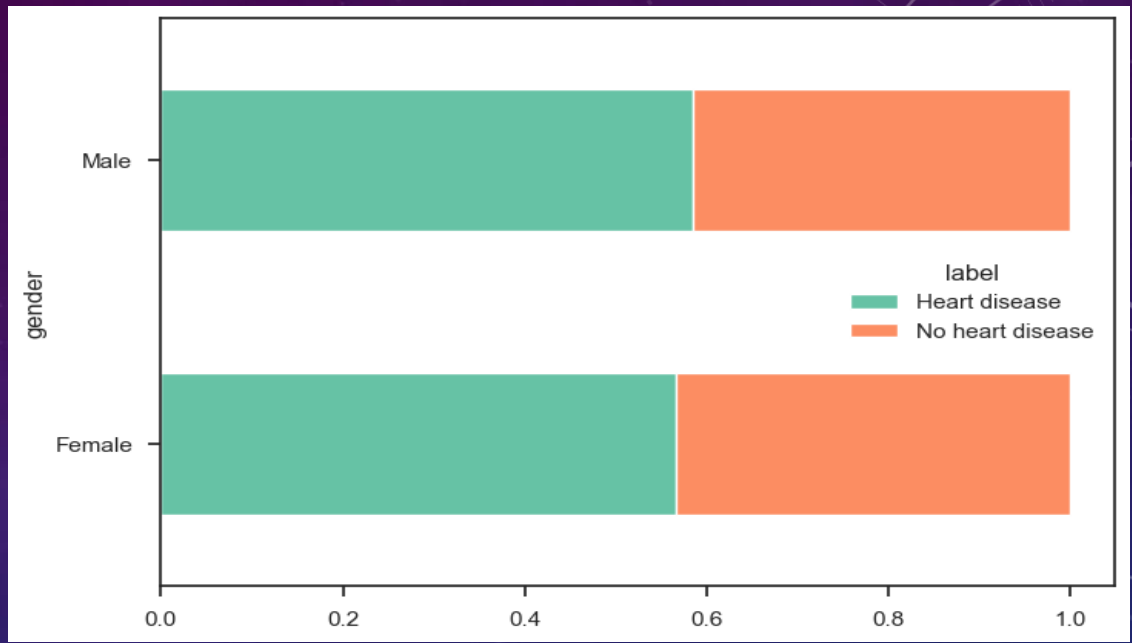
# ANALYSIS RESULTS DEMOGRAPHICS



Top Left: Number of patients of each gender with and without heart disease.

Top Right: Percentage of patients of each gender with and without heart disease.

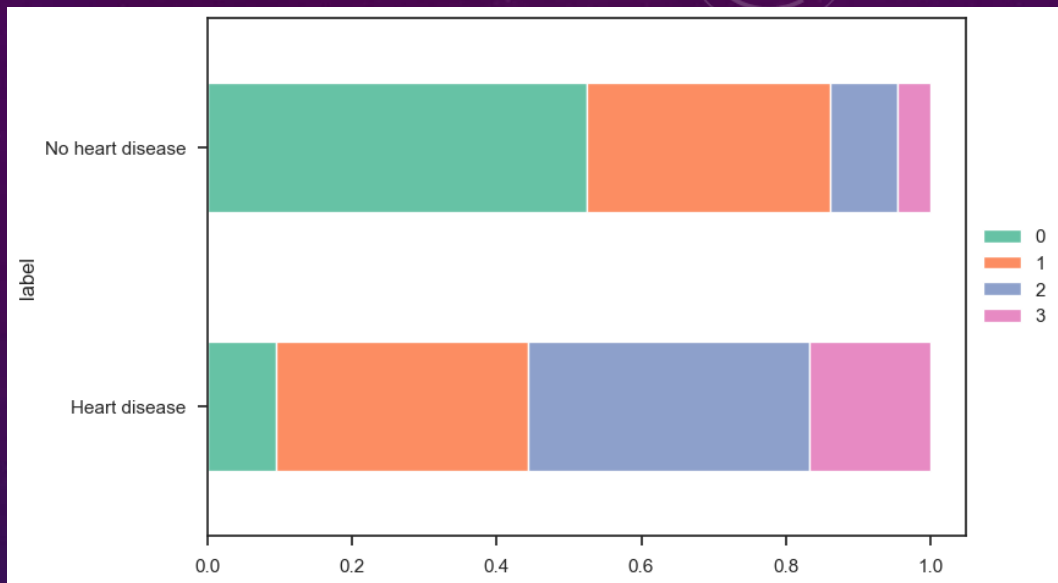
Bottom Right: Percentage of patients of each age group with and without heart disease.





# ANALYSIS RESULTS CATEGORICAL INDICATORS

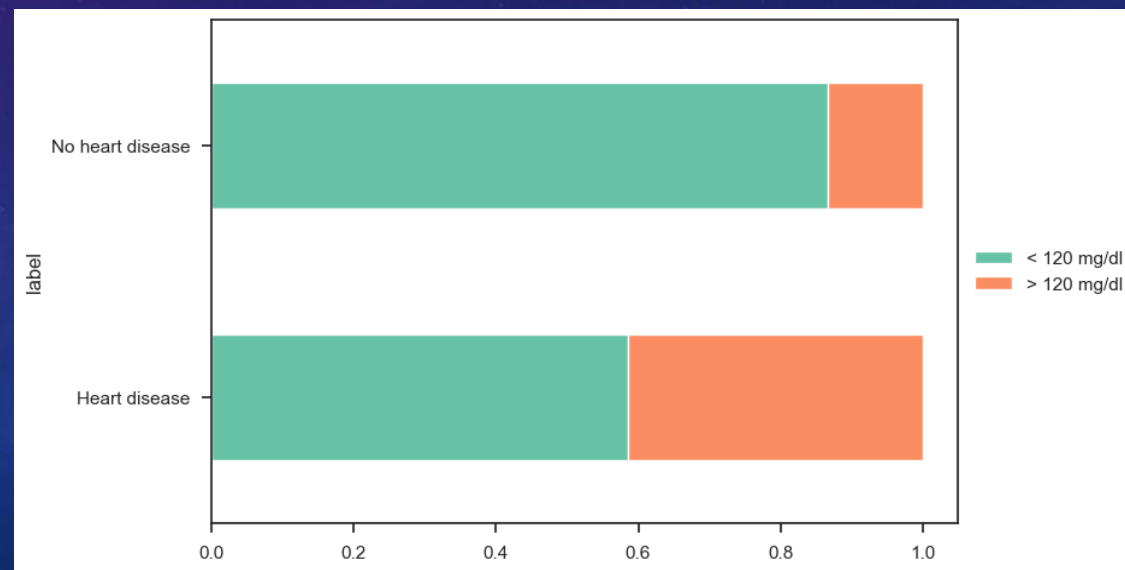
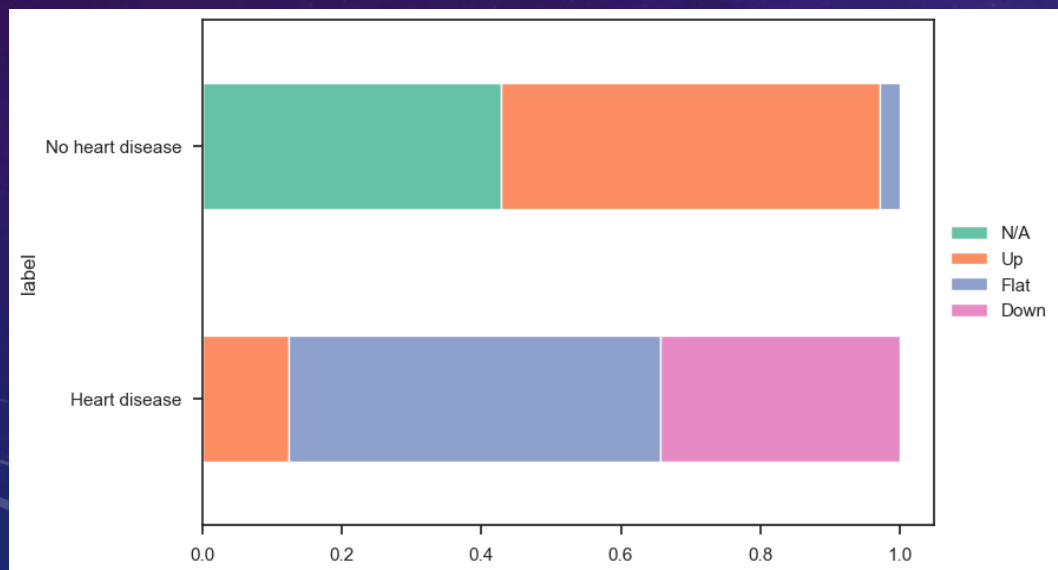


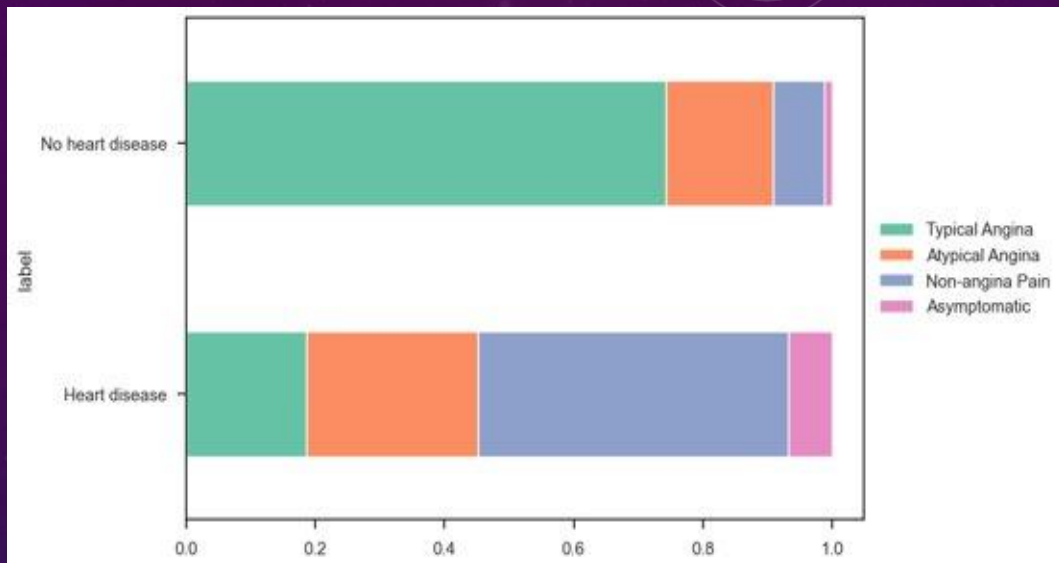


Top Left: Number of visible major vessels in fluoroscopy.

Bottom Left: Shape of the peak slope during exercise.

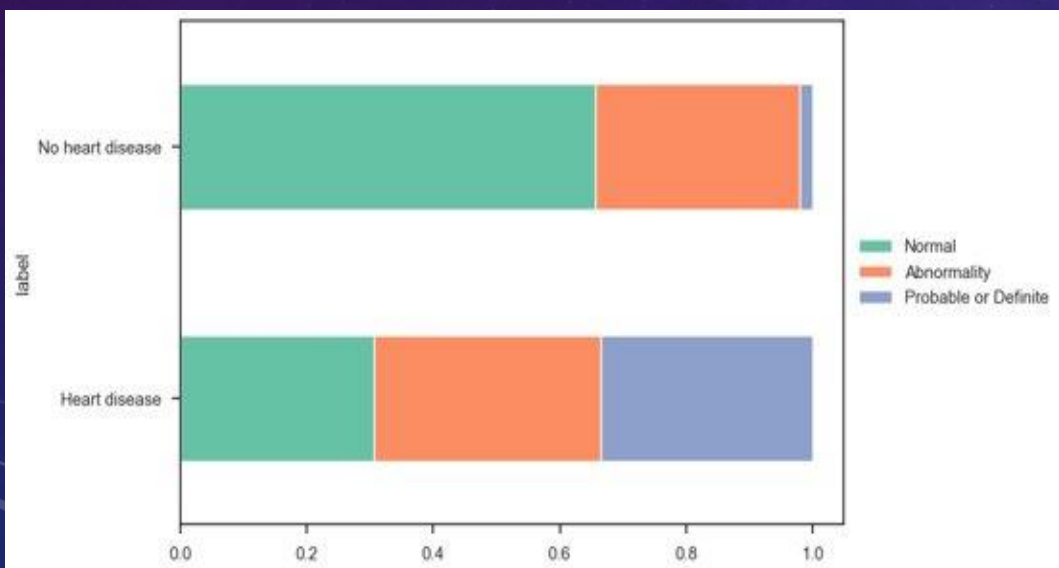
Bottom Right: Fasting blood sugar.





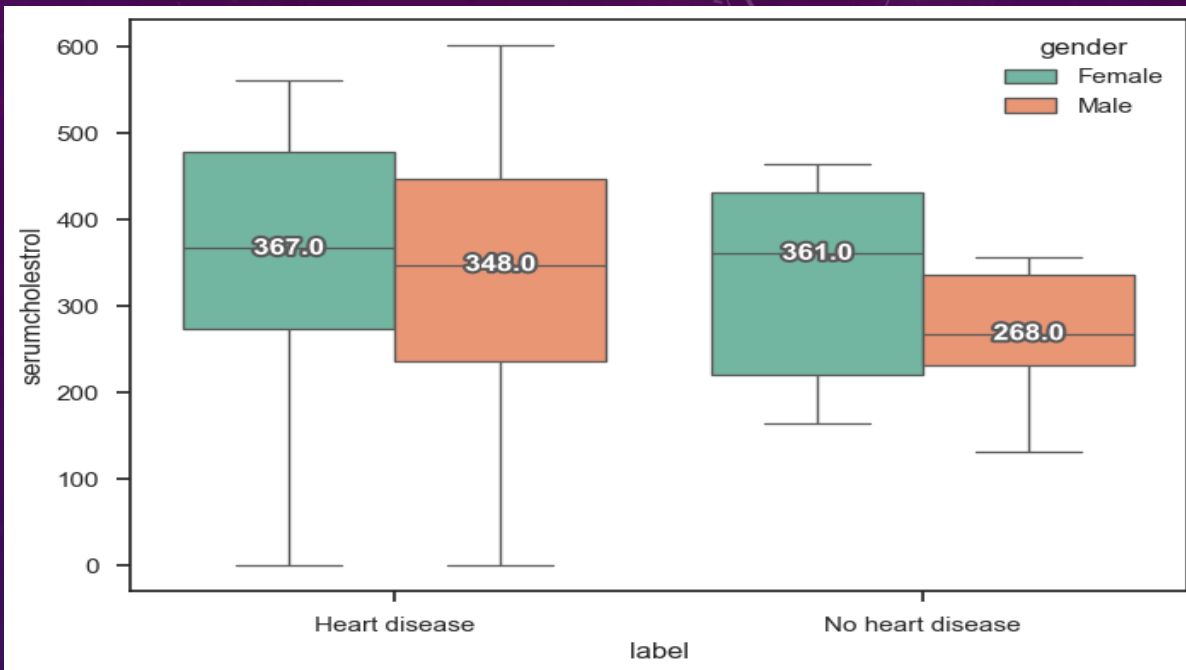
Top: Reported chest pain type.

Bottom: Results of ECG at rest.



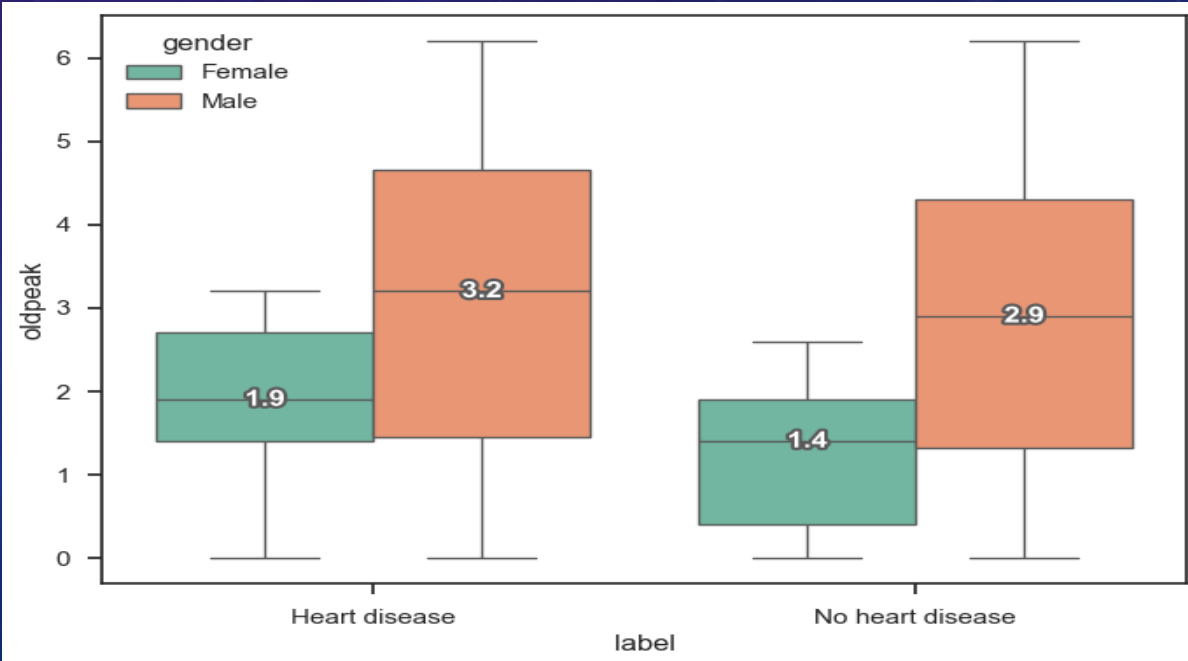
The background of the image is a dark, semi-transparent overlay of financial data. It features a candlestick chart with blue and orange bars, overlaid with various technical indicators including moving averages and oscillators. On the left side, there are circular gauges or dials with numerical scales, some of which are highlighted with white lines. The overall aesthetic is technical and data-driven.

# ANALYSIS RESULTS NUMERICAL INDICATORS



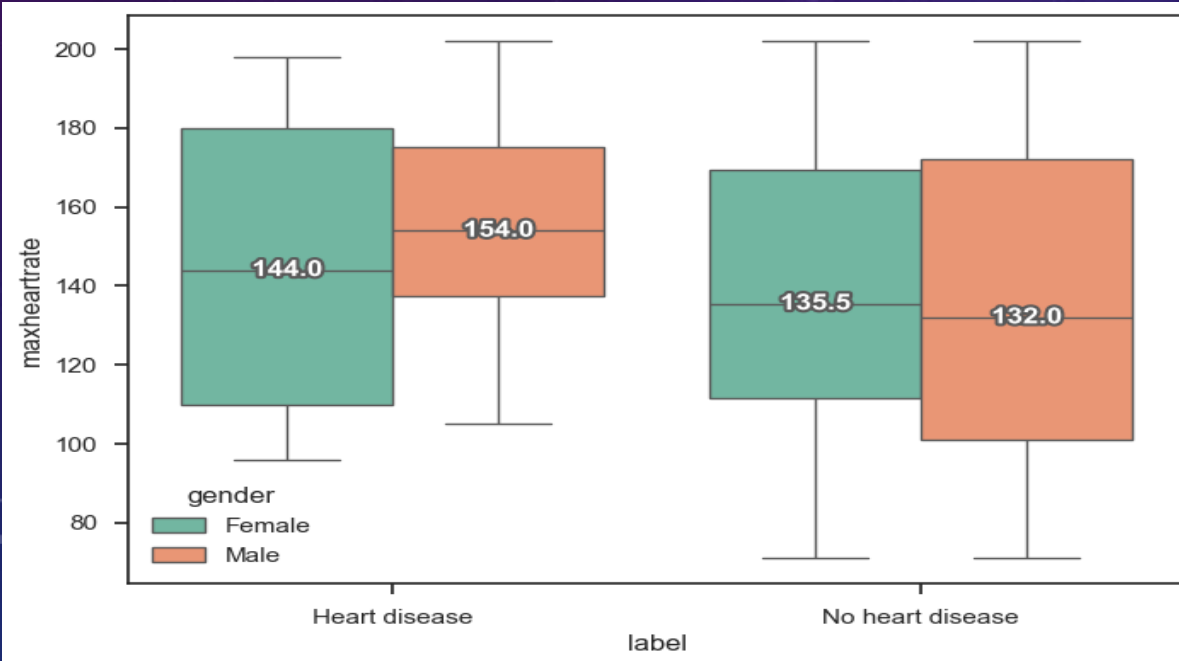
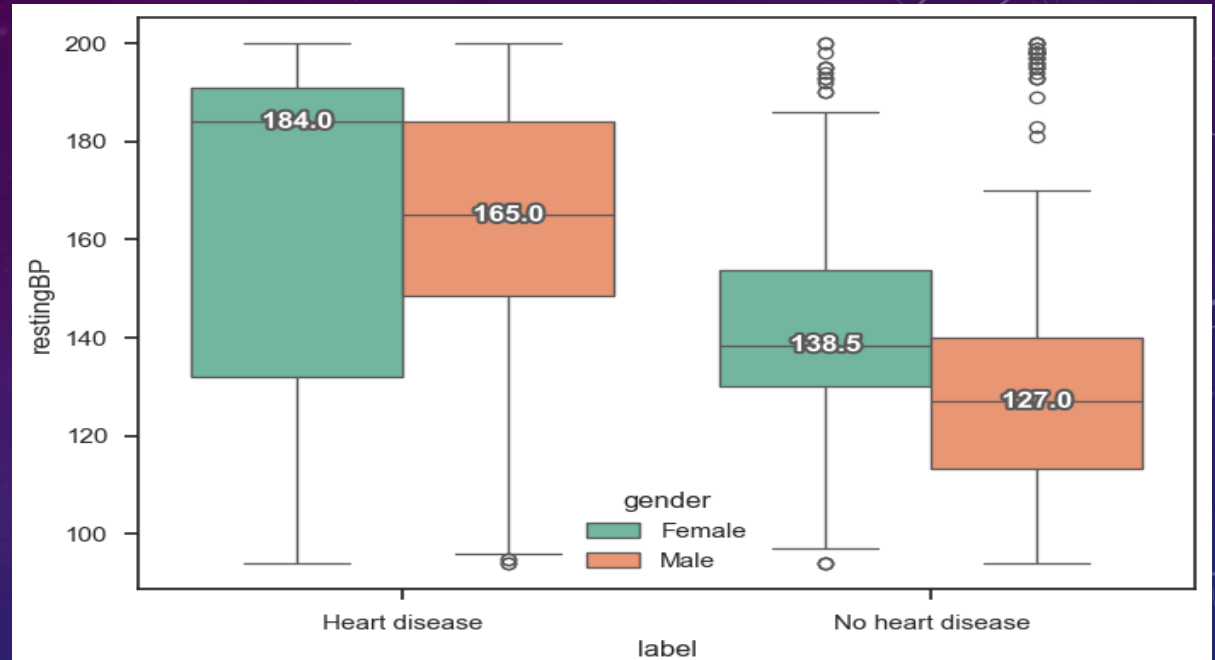
**Serum cholesterol** detected in blood sample (mg/dl).  
In range 126-564 mg/dl.

**Oldpeak**: measurement of the ST depression (in ECG) induced by exercise compared to at rest.  
In range 0-6.2.

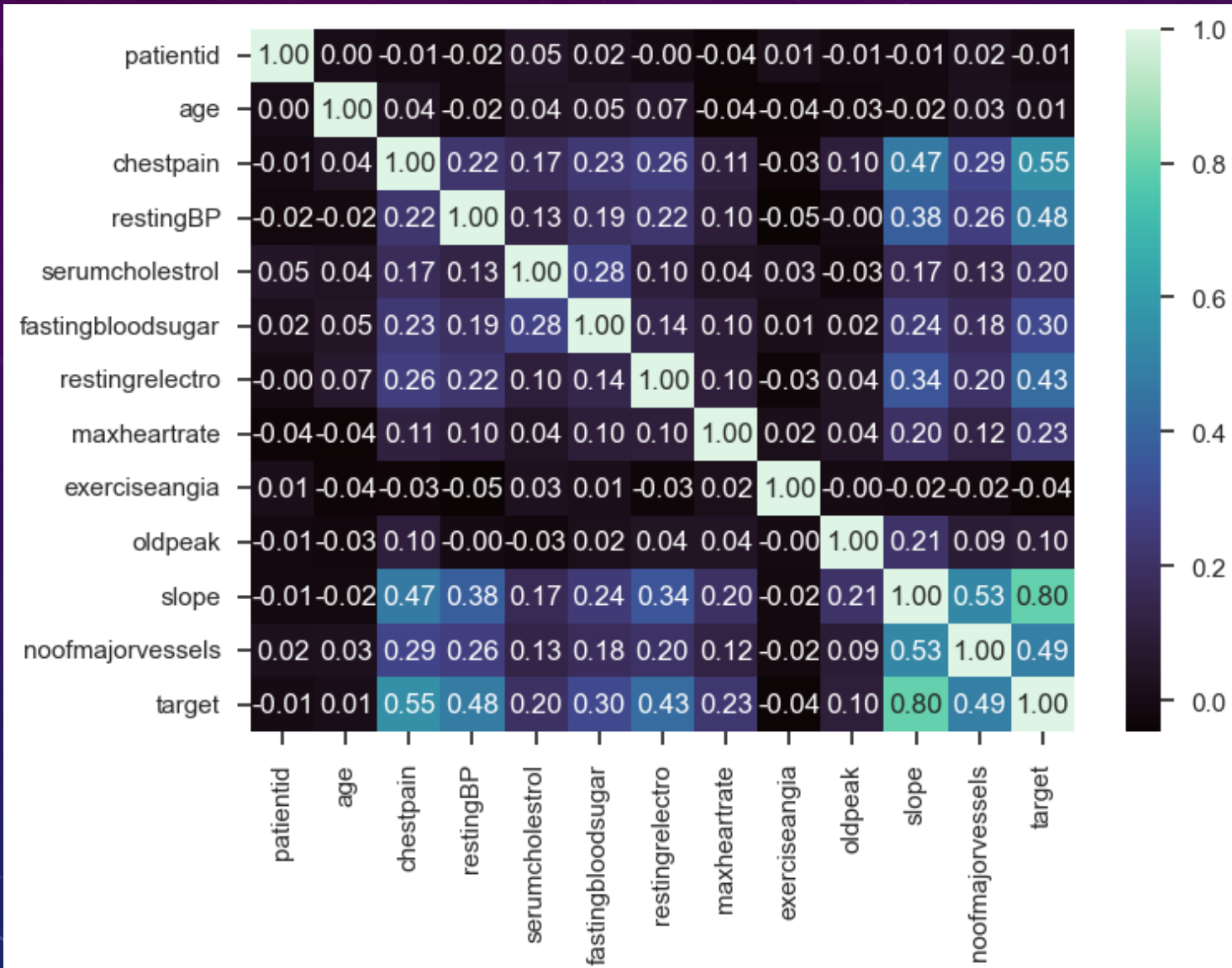




**Resting blood pressure (mm/HG)**  
In range 94-200 mm/HG



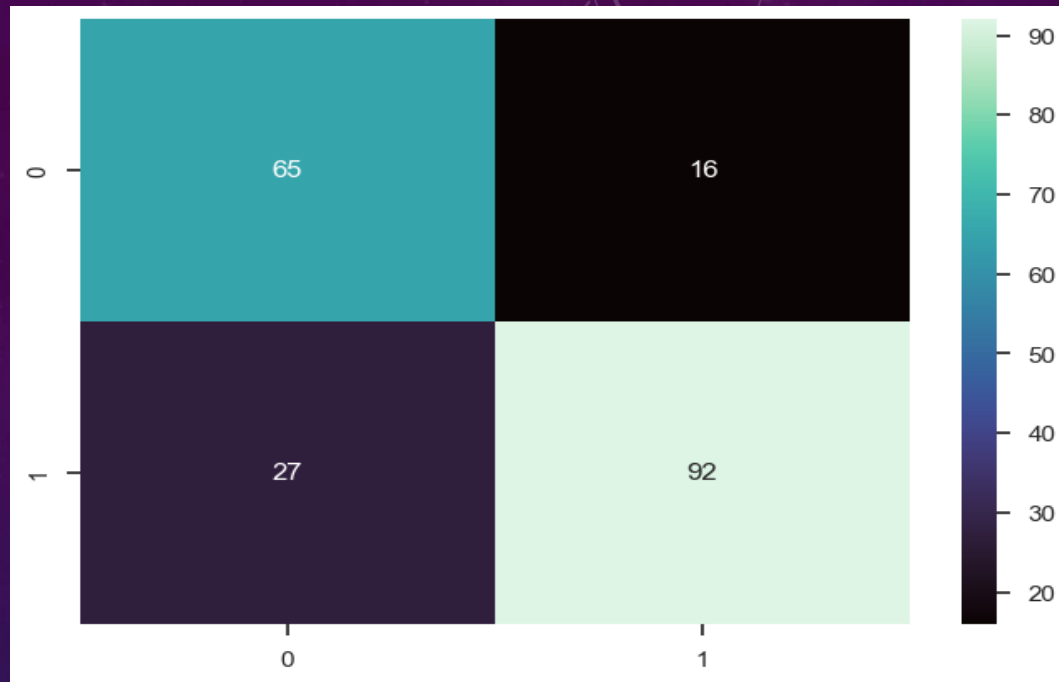
**Maximum heart rate achieved during exercise (bpm).**  
In range 71-202 bpm.



Highest calculated correlation:  
Target and Oldpeak  
**Correlation Matrix**

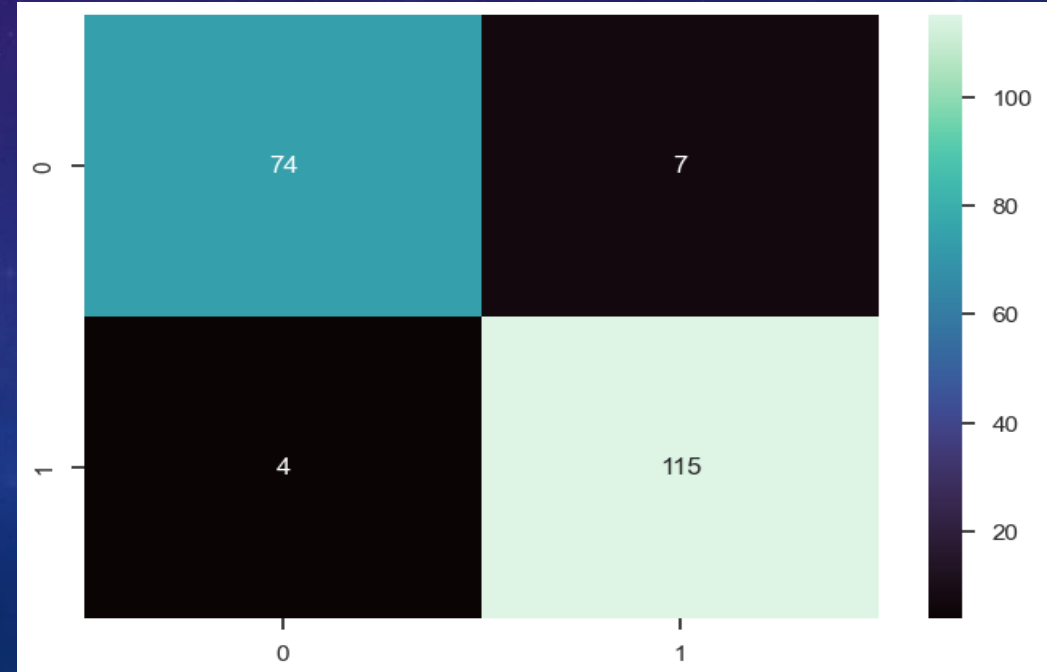
The background of the image is a dark, blue-toned scene featuring a robotic arm, possibly a prosthetic or industrial model, with its joints and segments clearly visible. Overlaid on this are various technical and data-related graphics. On the left side, there are several concentric circles and arcs, some of which are dotted, resembling a radar or targeting system. A large, semi-circular scale with numerical markings (40, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260) is also present. The bottom of the image shows a dense, blurred pattern of light and dark spots, suggesting a data visualization or a complex surface. The overall aesthetic is high-tech and futuristic.

# ANALYSIS RESULTS PREDICTION MODELS

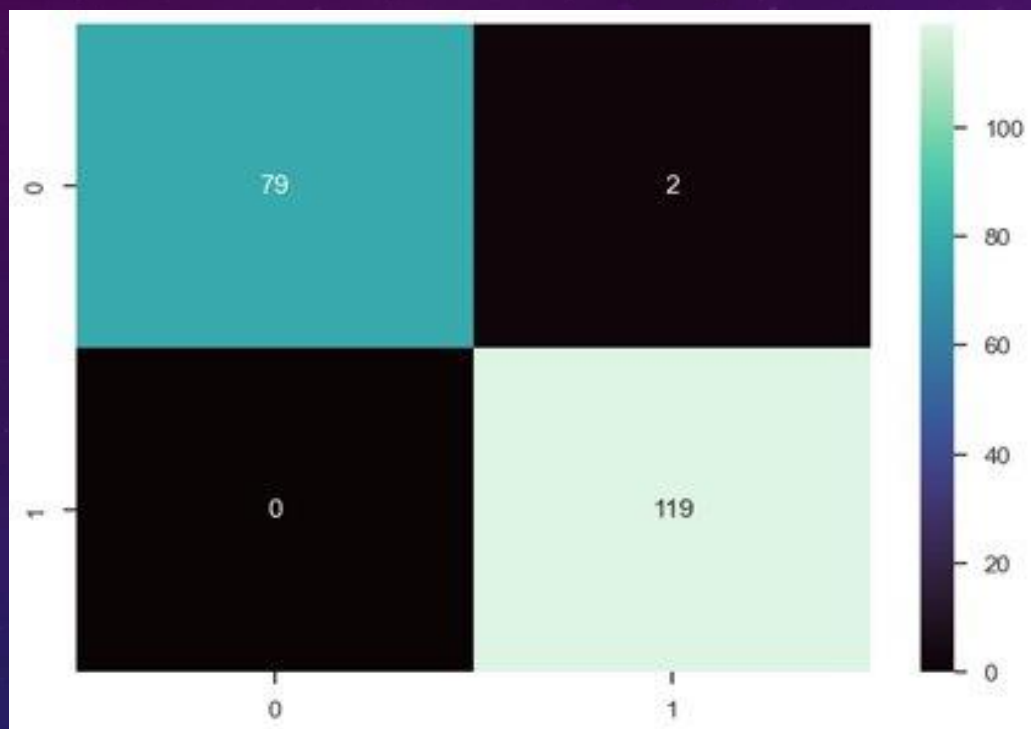


**K-Nearest Neighbours.**  
Avg. Accuracy < 78%.

**Decision Trees.**  
Avg. Accuracy ~ 97%.







**Random Forests.**  
Avg. Accuracy 97-99%

	precision	recall	f1-score	support
0	1.00	0.98	0.99	81
1	0.98	1.00	0.99	119
accuracy			0.99	200
macro avg	0.99	0.99	0.99	200
weighted avg	0.99	0.99	0.99	200

# CONCLUSIONS

- Main objective – Generate a prediction model
  - Both Random Forest and Decision Trees classifiers could be used.
  - These algorithms could be used in automated assisted diagnosis.
- Gender bias
  - There is a need for more female patients.
  - Numerical data highlights the differences.
- Categorical features collected show clear indication of correlation.
- Numerical features collected show indications of correlation.



# THANK YOU!

7CCSONAR RESEARCH METHODOLOGIES: ANALYSIS AND REPORTING

ASSESSMENT PRESENTATION 2023-24

ALEJANDRO OTERO ORTIZ DE COSCA (K21090108)



**KING'S**  
*College*  
**LONDON**