SC1015
MINI PROJECT

Gary Quah & Lau Jing Jie

Table of contents

01 Motivation

02 Exploratory Data Analysis

- Data Insights

03 Machine Learning

- Data Insights

04 Project Outcome

01
Motivation

Motivation

According to Singapore Heart Foundation, “23 people die from cardiovascular disease everyday. Cardiovascular disease accounted for 31.4% of all deaths in 2022, amounting to almost 1 out of 3 deaths in Singapore due to heart disease”

Problem Statement

Based on data provided, are we able to effectively predict if a person has heart disease based on the symptoms exhibited by the person

Dataset Used

kaggle

UCI Heart Disease Data

By MD. REDWAN KARIM SONY

Acknowledgements

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation:Robert Detrano, M.D., Ph.D.

01
Presentation Of Data

Data Cleaning - Renaming Variables

	id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
0	1	63	Male	Cleveland	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	2.3	downsloping	0.0	fixed defect	0
1	2	67	Male	Cleveland	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	1.5	flat	3.0	normal	2
2	3	67	Male	Cleveland	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	2.6	flat	2.0	reversable defect	1
3	4	37	Male	Cleveland	non-anginal	130.0	250.0	False	normal	187.0	False	3.5	downsloping	0.0	normal	0
4	5	41	Female	Cleveland	atypical angina	130.0	204.0	False	lv hypertrophy	172.0	False	1.4	upsloping	0.0	normal	0

	age	sex	chest_pain	rest_blood_pressure	cholesterol_level	diabetic	resting_ecg	max_heart_rate	exercise_angina	oldpeak	slope	number_major_vessels	thal	number
0	63	Male	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	2.3	downsloping	0.0	fixed defect	0
1	67	Male	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	1.5	flat	3.0	normal	2
2	67	Male	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	2.6	flat	2.0	reversable defect	1
3	37	Male	non-anginal	130.0	250.0	False	normal	187.0	False	3.5	downsloping	0.0	normal	0
4	41	Female	atypical angina	130.0	204.0	False	lv hypertrophy	172.0	False	1.4	upsloping	0.0	normal	0

920 rows × 14 columns

Data Cleaning - Empty (NaN) Values

```
dataframe.isnull().sum()
```

age	0
sex	0
chest_pain	0
rest_blood_pressure	59
cholesterol_level	30
diabetic	90
resting_ecg	2
max_heart_rate	55
exercise_angina	55
oldpeak	62
slope	309
number_major_vessels	611
thal	486
number	0

1. Checking for number of NaN values in each column

2. Remove Columns where NaN are too many
   a. Number_Major_Vessels (611 / 920 - **66%**)
   b. Thal (486/920 - **53%**)
   c. Slope (309/920 - **34%**)

3. Removed Oldpeak as its correlated with Slope
   - Avoid possible contamination of data

Data Cleaning - Empty (NaN) Values

age	0
sex	0
chest_pain	0
rest_blood_pressure	59
cholesterol_level	30
diabetic	90
resting_ecg	2
max_heart_rate	55
exercise_angina	55
number	0

1. Checking for number of NaN values in each column again

2. Remove rows where NaN are too many : Dataset is from unique patients where NaN values can't be filled from other rows in the dataset
   a. Rest_blood_pressure (59/920 - 6.4%)
   b. Cholesterol_level (30/920 -3.3%)
   c. Diabetic (90/920 - 9.8%)
   d. Resting_ecg (2/920 - 2.2%)
   e. Max_heart_rate (55/920 - 5.98%)
   f. Exercise_angina (55/920 - 5.98%)

Data Cleaning - Empty (NaN) Values

age	0
sex	0
chest_pain	0
rest_blood_pressure	0
cholesterol_level	0
diabetic	0
resting_ecg	0
max_heart_rate	0
exercise_angina	0
number	0

- Final result : All NaN Values are weeded out from the dataset

Data Cleaning - Empty (NaN) Values

age	0
sex	0
chest_pain	0
rest_blood_pressure	0
cholesterol_level	0
diabetic	0
resting_ecg	0
max_heart_rate	0
exercise_angina	0
number	0

- Final result : All NaN Values are weeded out from the dataset

- 176 rows removed
- 4 Columns removed

920 rows × 14 columns → (744, 10)

Data Cleaning - Invalid Values

There are rows where Rest_blood_pressure, cholesterol_level are 0, as seen from min

- Drop rows that contain the value '0' in these 2 columns

```
dataframe.describe()
```

	age	rest_blood_pressure	cholesterol_level	max_heart_rate	number
count	744.000000	744.000000	744.000000	744.000000	744.000000
mean	53.127688	132.762097	219.822581	138.821237	0.924731
std	9.398811	18.610367	93.735536	25.843072	1.129433
min	28.000000	0.000000	0.000000	60.000000	0.000000
25%	46.000000	120.000000	197.000000	120.000000	0.000000
50%	54.000000	130.000000	231.000000	140.000000	1.000000
75%	60.000000	140.000000	270.250000	160.000000	1.000000
max	77.000000	200.000000	603.000000	202.000000	4.000000

Data Cleaning - Invalid Values

Dataframe is now cleaned.

	age	rest_blood_pressure	cholesterol_level	max_heart_rate	number
count	664.000000	664.000000	664.000000	664.000000	664.000000
mean	52.631024	132.759036	246.307229	141.278614	0.813253
std	9.442100	17.816792	57.561657	25.046787	1.079665
min	28.000000	92.000000	85.000000	69.000000	0.000000
25%	46.000000	120.000000	210.000000	123.000000	0.000000
50%	54.000000	130.000000	239.500000	143.000000	0.000000
75%	59.000000	140.000000	275.000000	160.000000	1.000000
max	77.000000	200.000000	603.000000	202.000000	4.000000

664 rows × 10 columns

Data Cleaning - Invalid Values

Dataframe is now cleaned.

	age	rest_blood_pressure	cholesterol_level	max_heart_rate	number
count	664.000000	664.000000	664.000000	664.000000	664.000000
mean	52.631024	132.759036	246.307229	141.278614	0.813253
std	9.442100	17.816792	57.561657	25.046787	1.079665
min	28.000000	92.000000	85.000000	69.000000	0.000000
25%	46.000000	120.000000	210.000000	123.000000	0.000000
50%	54.000000	130.000000	239.500000	143.000000	0.000000
75%	59.000000	140.000000	275.000000	160.000000	1.000000
max	77.000000	200.000000	603.000000	202.000000	4.000000

`(744, 10)` → 664 rows × 10 columns

Data Cleaning - Invalid Values

Dataframe is now cleaned.

	age	sex	chest_pain	rest_blood_pressure	cholesterol_level	diabetic	resting_ecg	max_heart_rate	exercise_angina	number
0	63	Male	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	0
1	67	Male	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	2
2	67	Male	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	1
3	37	Male	non-anginal	130.0	250.0	False	normal	187.0	False	0
4	41	Female	atypical angina	130.0	204.0	False	lv hypertrophy	172.0	False	0

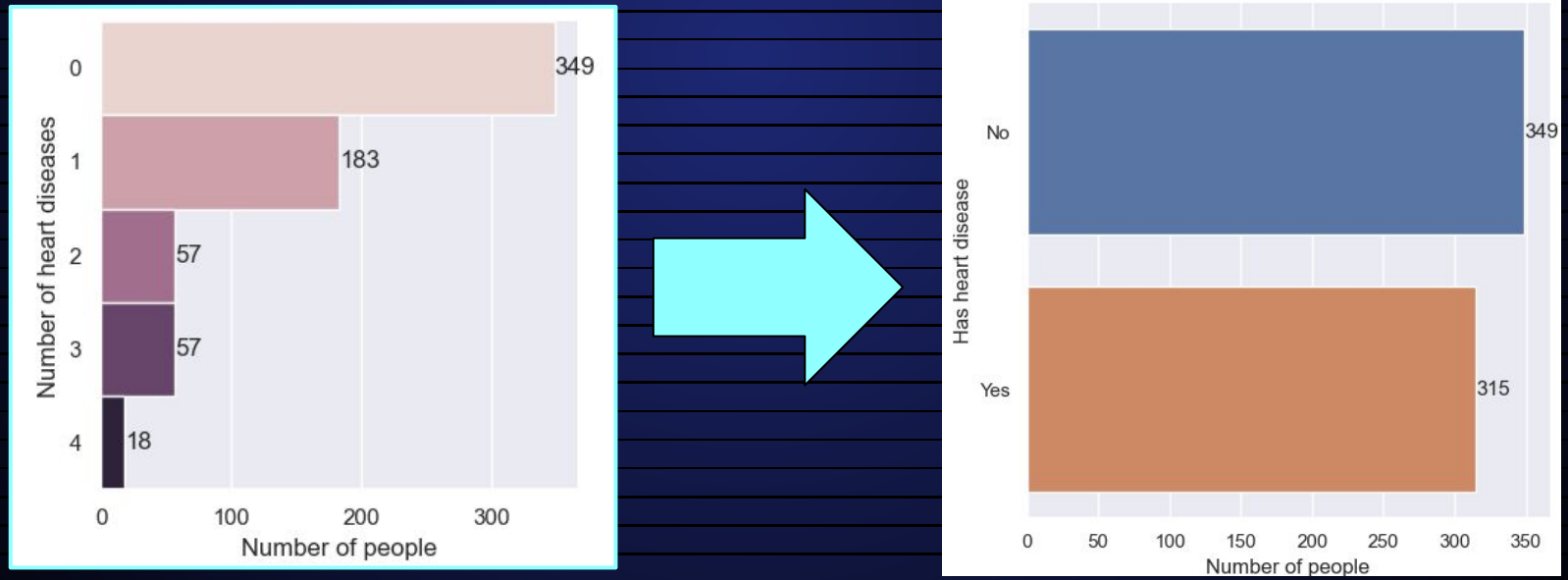
664 rows × 10 columns

02
Exploratory Data Analysis

Letting data speak for itself

EDA - Predictor : “number”

- “Number” represents the number of heart diseases a person has in the dataset
- Replace all >= 1 values with 1
  - ➢ Binary prediction, 1 or 0 : Has heart disease vs no heart disease



EDA - Response Variables

Numerical

1. age
2. rest_blood_pressure
3. cholesterol_level
4. max_heart_rate

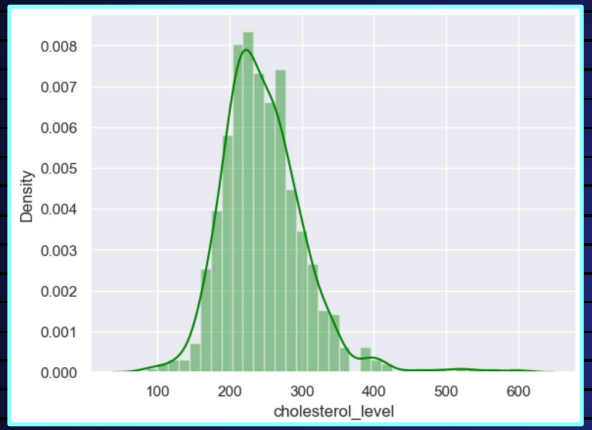
	age	rest_blood_pressure	cholesterol_level	max_heart_rate
count	664.000000	664.000000	664.000000	664.000000
mean	52.631024	132.759036	246.307229	141.278614
std	9.442100	17.816792	57.561657	25.046787
min	28.000000	92.000000	85.000000	69.000000
25%	46.000000	120.000000	210.000000	123.000000
50%	54.000000	130.000000	239.500000	143.000000
75%	59.000000	140.000000	275.000000	160.000000
max	77.000000	200.000000	603.000000	202.000000

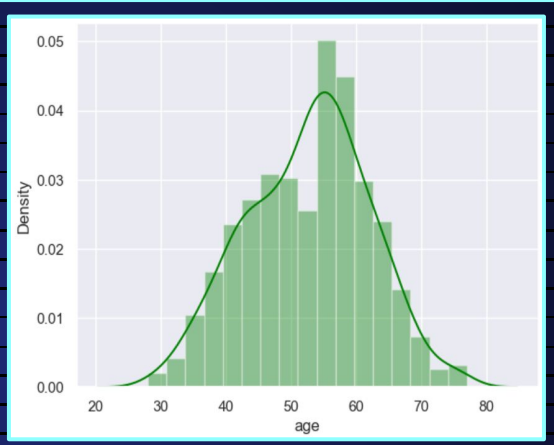
Categorical

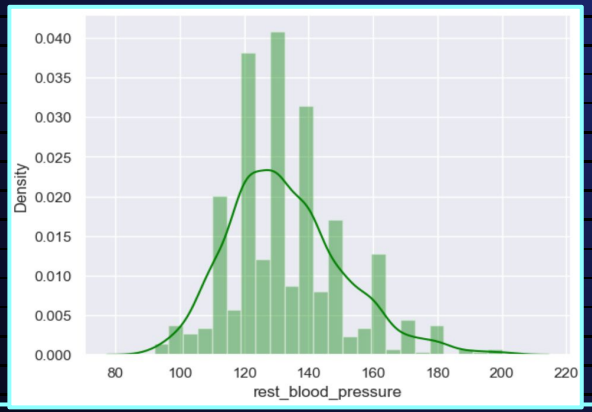
1. chest_pain
2. exercise_angina
3. sex
4. diabetic
5. resting_ecg

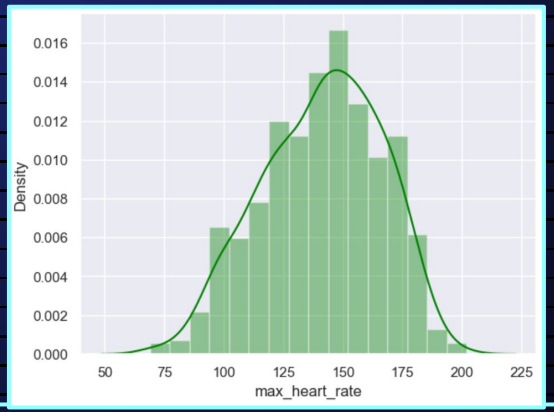
	chest_pain	exercise_angina	sex	diabetic	resting_ecg
count	664	664	664	664	664
unique	4	2	2	2	3
top	asymptomatic	False	Male	False	normal
freq	334	414	493	563	400

EDA - Numericals





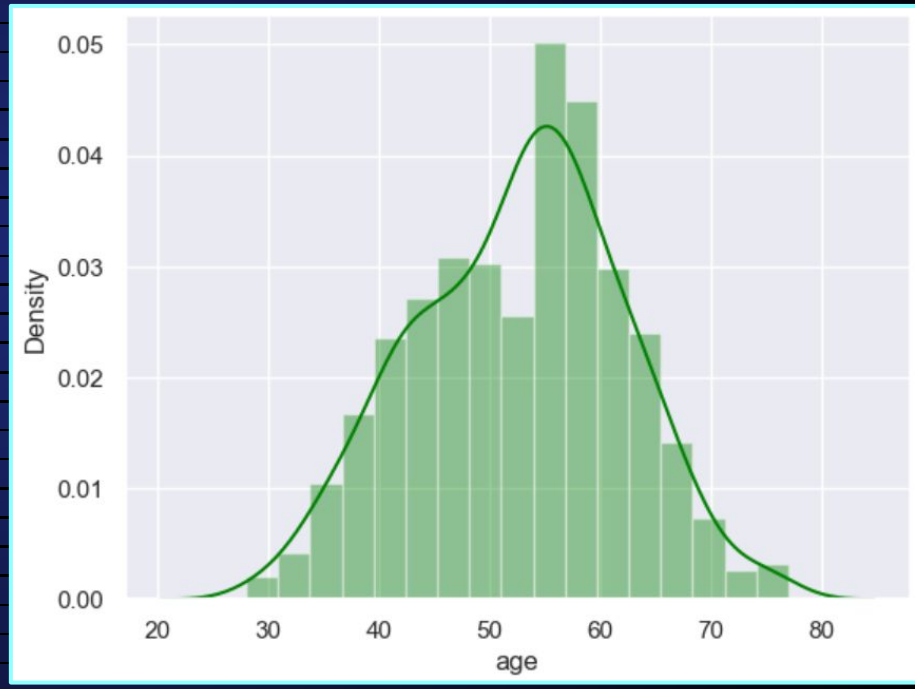




Age

The age range in this data set is between 28 - 77 years old

With an average of 52 years old and standard deviation of 9



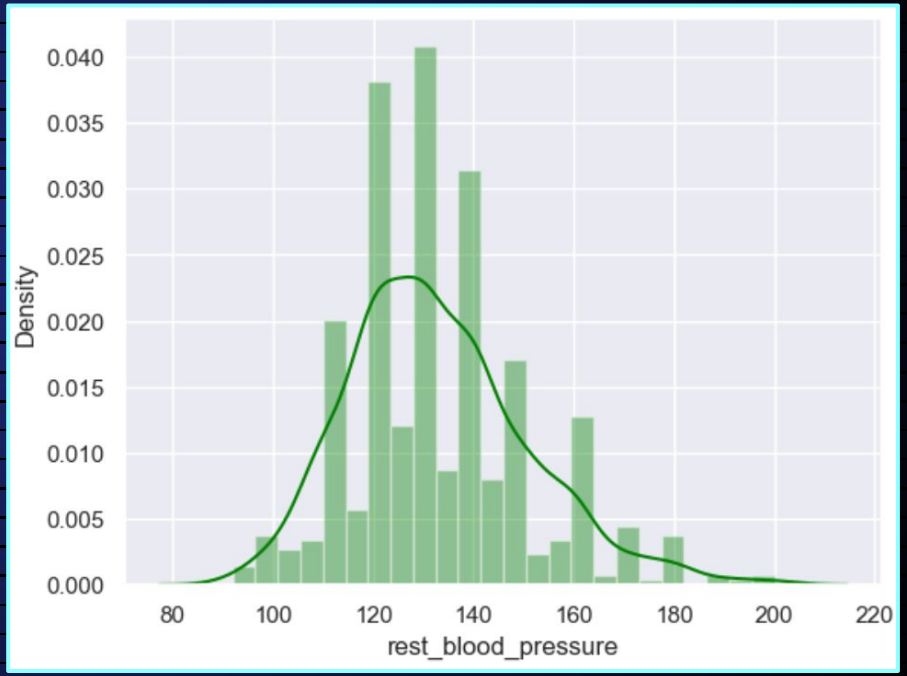
Blood Pressure

Healthy level of blood pressure is below 120. Patients with blood pressure higher than 120 is diagnosed with hypertension

The data shows bp range of 92 to 200

With an average of 132 and standard deviation of 17

We can conclude that most patient in this dataset have hypertension



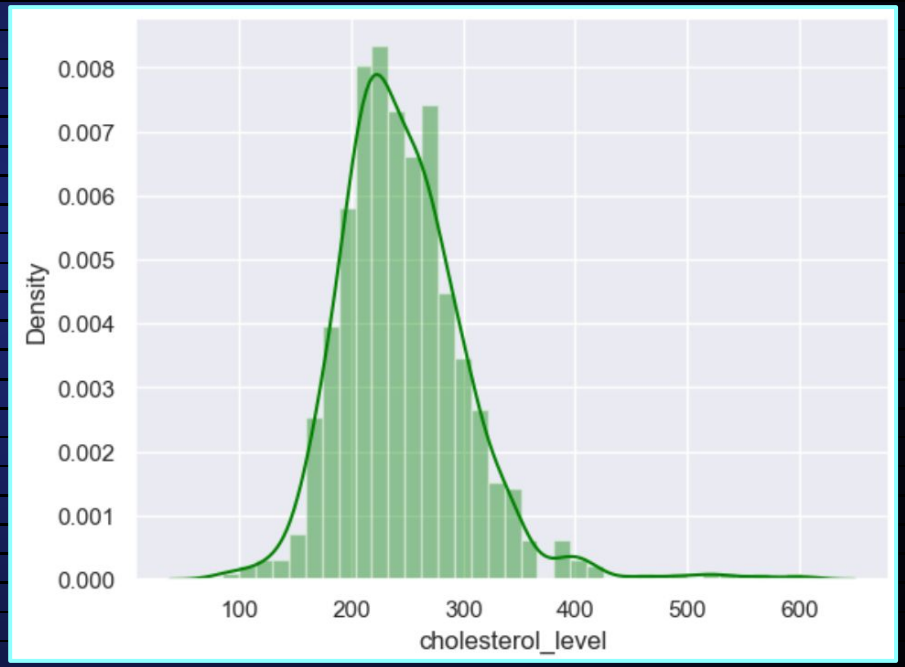
Cholesterol

Healthy level of cholesterol is below 200
Patients with cholesterol higher than 200 is considered high cholesterol level

The data show a range of 85 to 603

With an average of 246 and standard deviation of 57

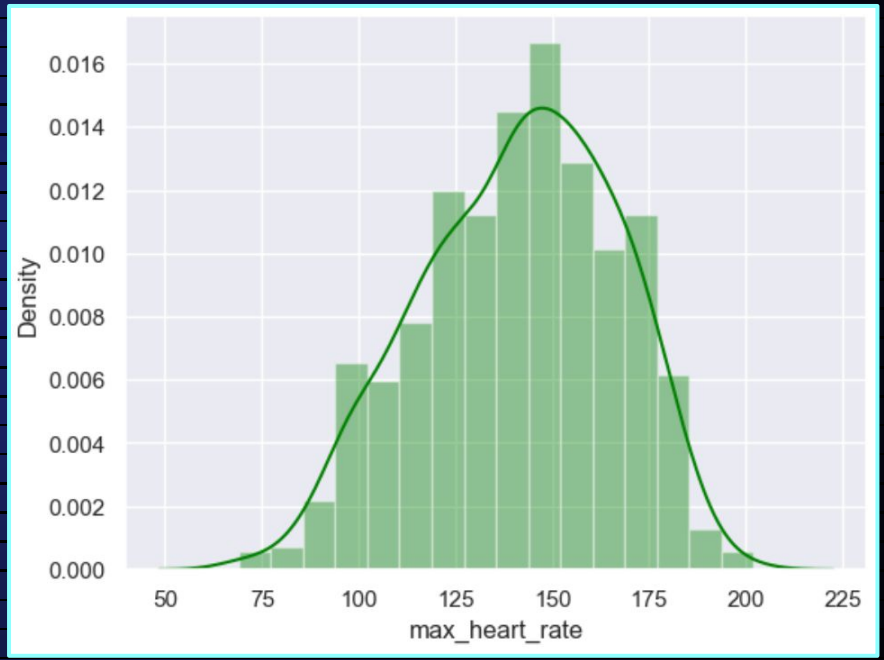
We can conclude that most patient in this dataset have high cholesterol level



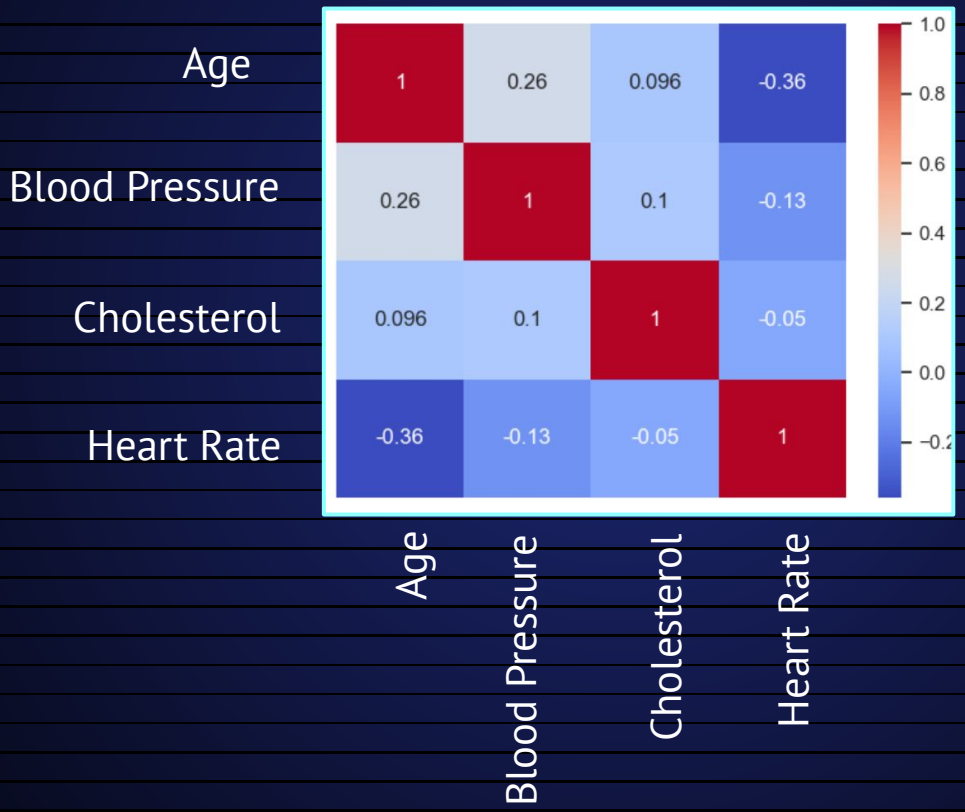
Heart Rate

The data show a heart rate range of 69 to 202

With an average of 141 and standard deviation of 25

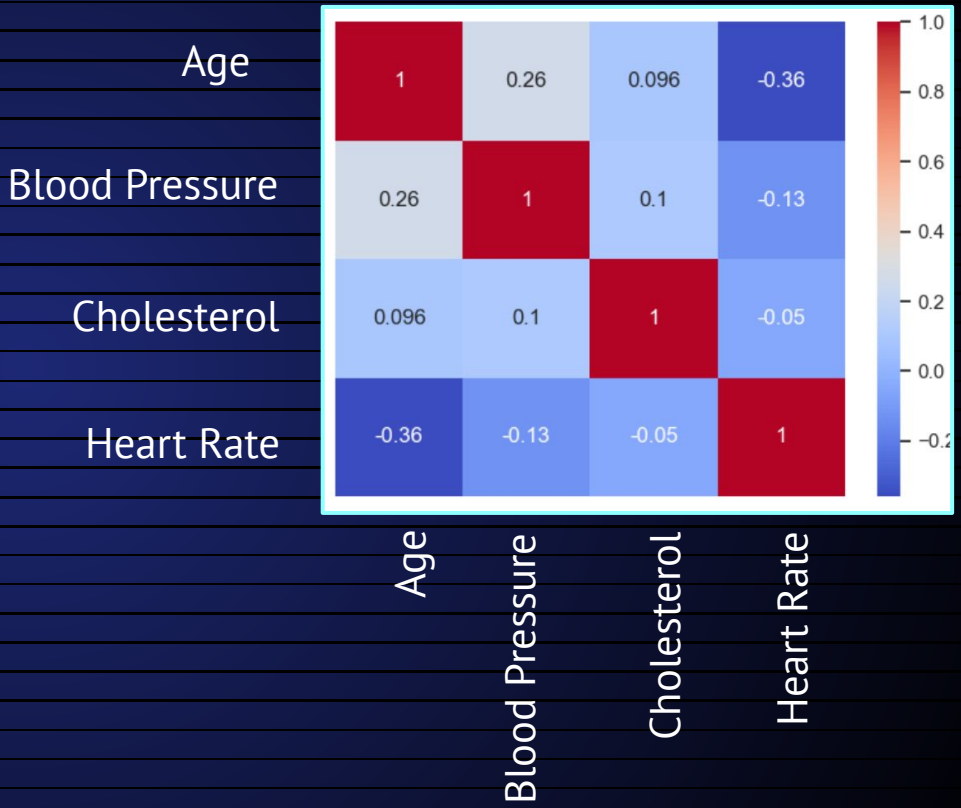


Insights - Numerical



Base on the correlation matrix of the numerical data, we can see that there is no strong indication of any relation.

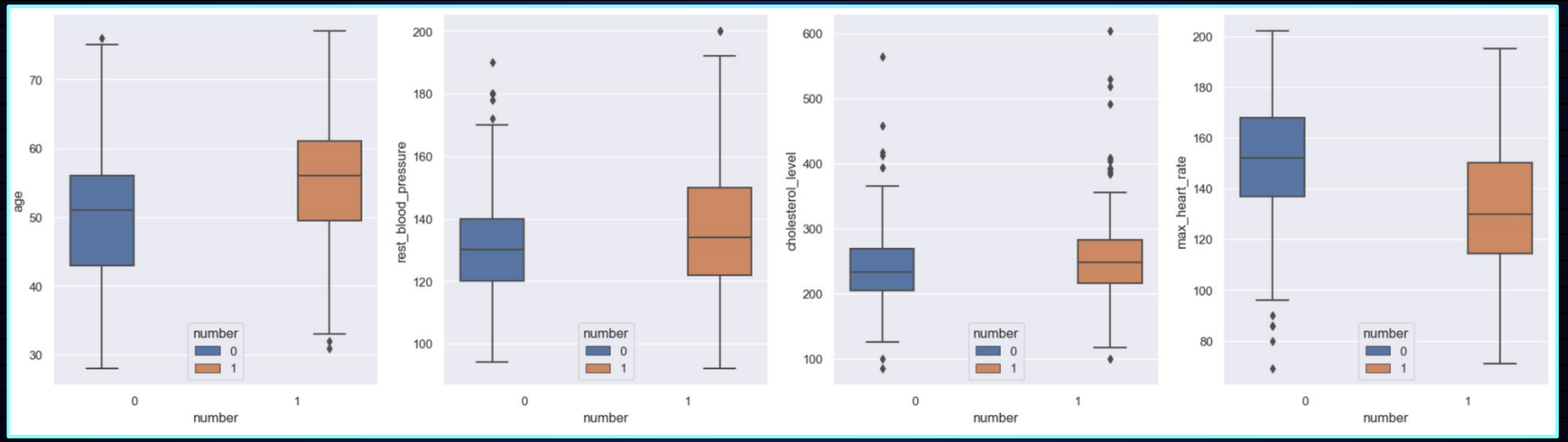
However there is weak negative correlation between age and heart rate with value of -0.36 and a weak positive correlation between age and blood pressure



Insights - Numerical

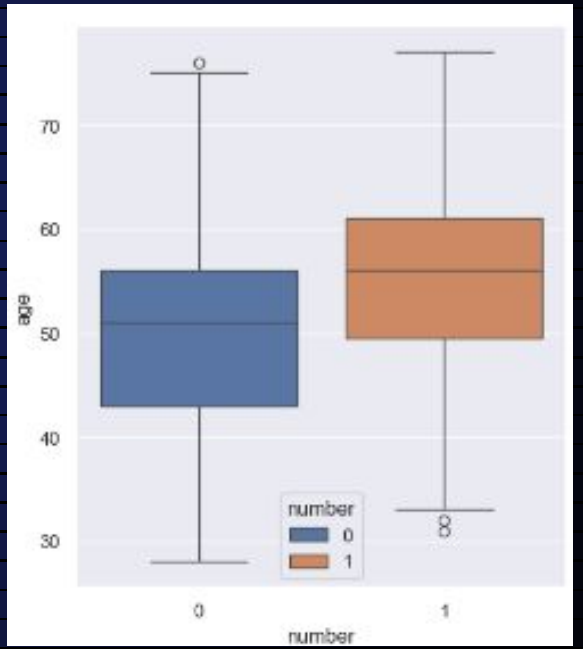
1. Most patients have high blood pressure
2. Most patient have high cholesterol levels
3. Correlation between age and blood pressure and max heart rate

Insights - Numerical Multivariate Data Analysis



Insights - Numerical

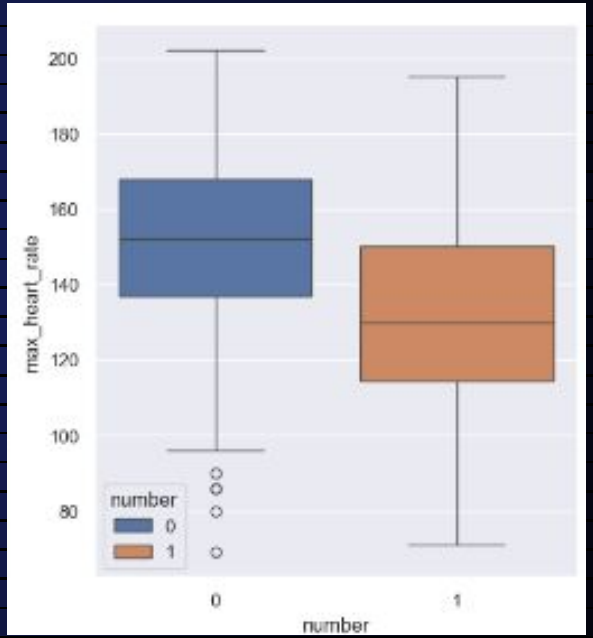
- Age could have a relationship with the presence of heart disease
- "With age, the function of the heart is influenced mainly by the decrease in elasticity and the ability to respond to changes in pressure (compliance) of the arterial system" (Stern, et al).
  - ➢ This inevitably causes more stress to act on the heart as a person ages. This thus makes a person who is older to be more prone to heart diseases as the heart weakens over time.



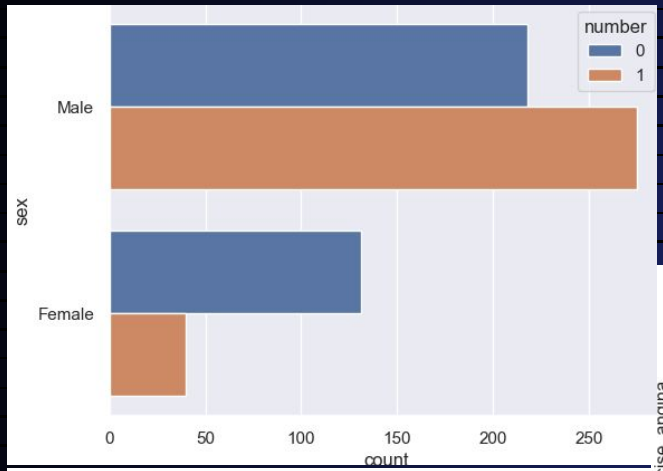
Insights - Numerical

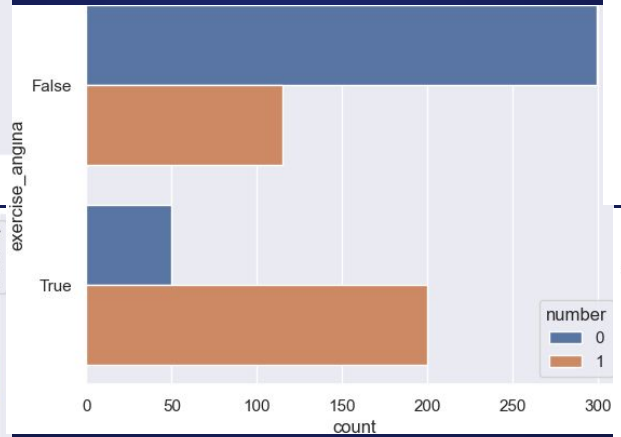
- max_heart_rate could have a relationship with the presence of heart disease

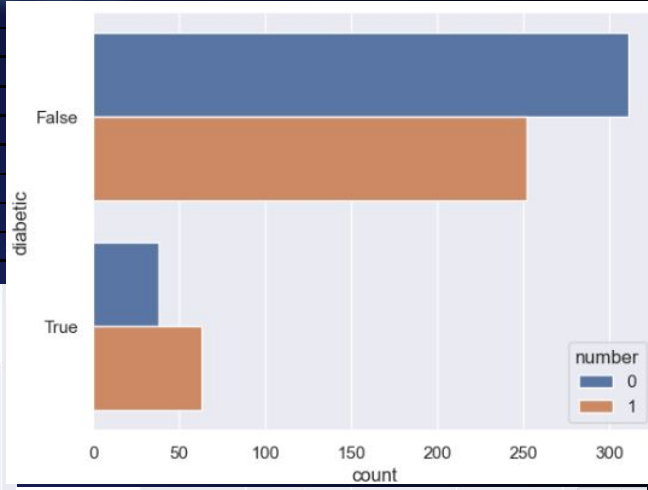
- "The rate at which your heart is beating when it is working its hardest to meet your body's oxygen needs is your maximum heart rate" (LeWine, H. E).
  - ➢ A higher maximum heart rate signifies a healthier heart compared to a lower maximum heart rate
    - ○ Able to circulate oxygen better

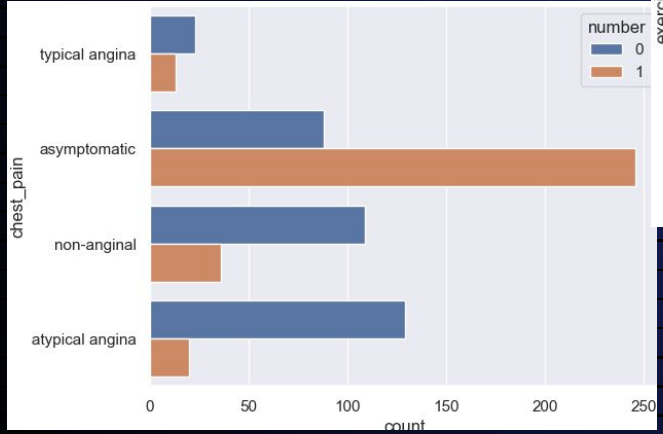


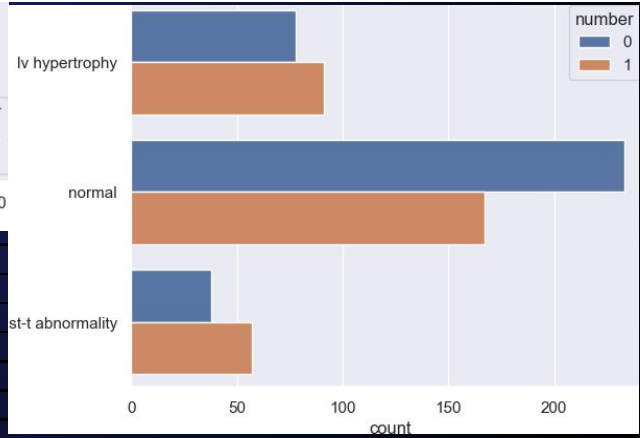
EDA - Categorical







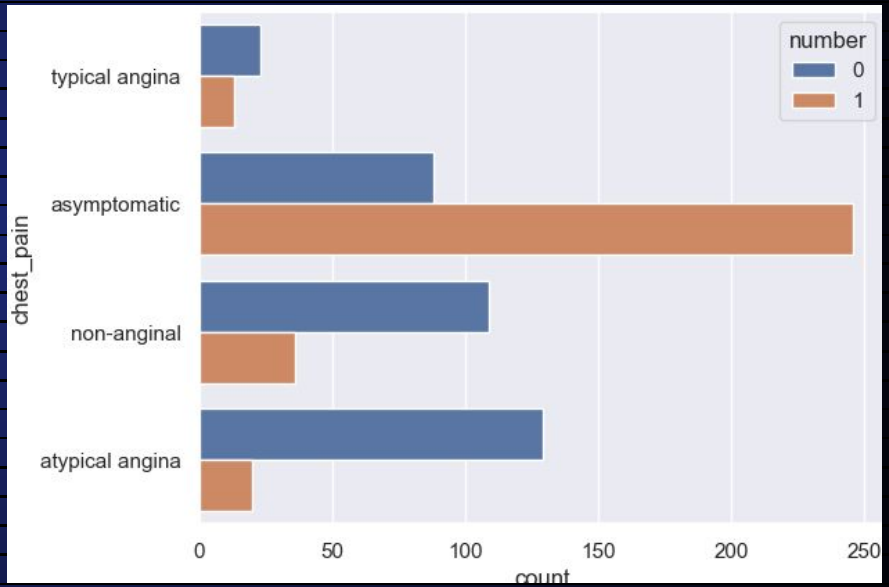




Categorical EDA - chest_pain

From the data...

1. Typical angina (chest discomfort)
   - ➢ No obvious relationship with heart disease
2. Asymptomatic (chest pain / discomfort without symptoms)
   - ➢ People who experience asymptomatic chest pain have a high likelihood to have heart disease
3. Non-anginal (chest pain / discomfort without typical characteristics of angina)
   - ➢ People who have non-anginal chest pain are less likely to have heart disease
4. Atypical angina (chest pain that does not fit typical patterns)
   - ➢ People who have atypical angina are less likely to have chest pain



Categorical EDA - exercise_angina

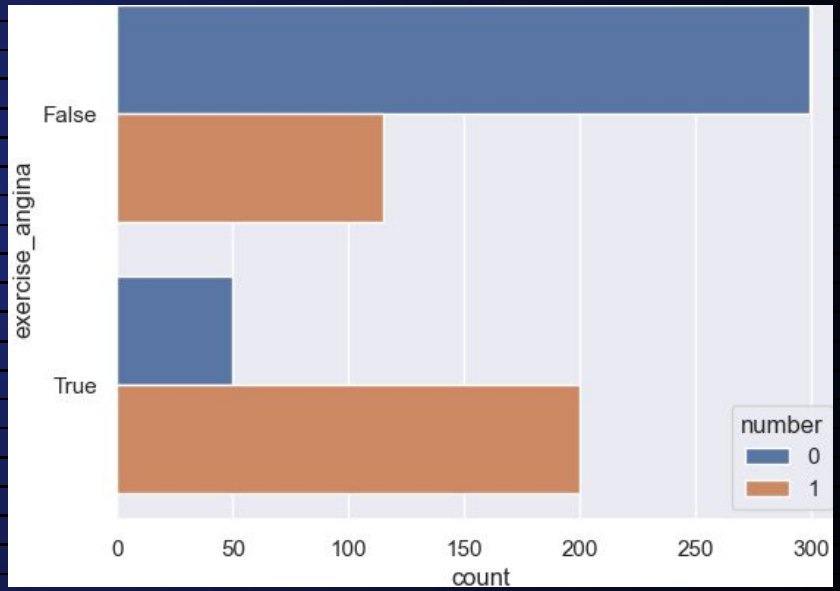
Exercise angina refers to chest pain when exercising

- From the data we can see that people who have exercise angina are more likely to have heart disease than people who don't have exercise angina.

Having exercise angina

=

likelihood of heart disease



Categorical EDA - Sex

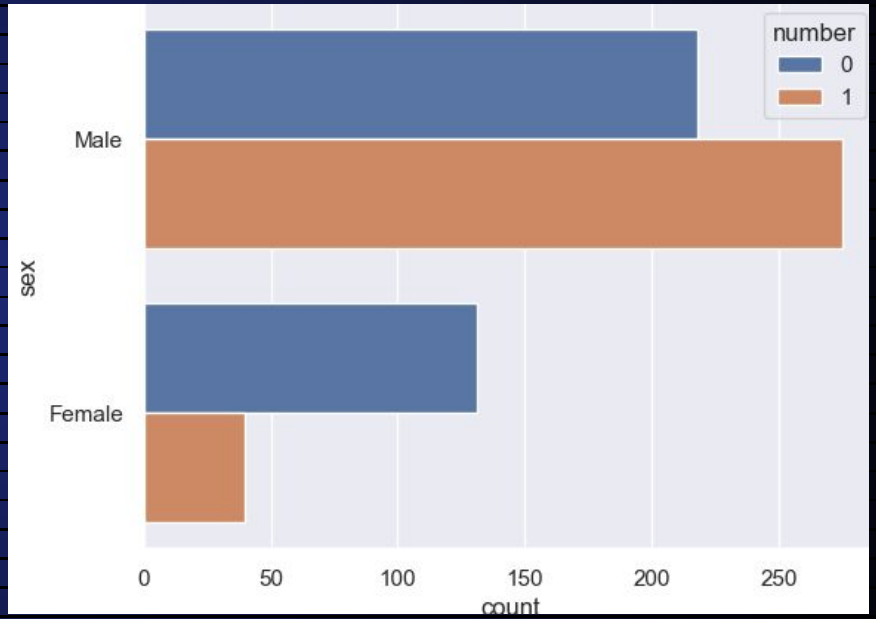
From the data, we can see that males possibly have a higher likelihood of having heart disease compared to females

- 55.7% of Males have heart disease
- 23.4% of Females have heart disease

Male

=

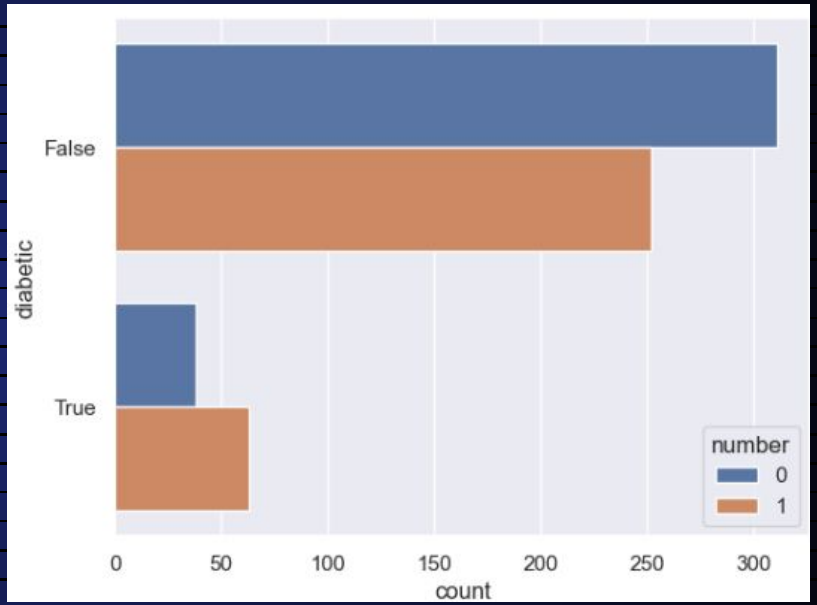
higher odds of having heart disease




```
Number of Males : 493 Males with heart disease :  275  Males without heart disease : 218
Number of Females : 171 Females  with heart disease :  40  Females without heart disease : 131
Percentage of males with heart disease (within males):  55.78093306288032
Percentage of females with heart disease (within females) :  23.391812865497073
```

Categorical EDA - diabetic

From the data, the presence of diabetes has possibility of having a relation between the presence of heart disease, where if one is diabetic, they would have a higher chance of having heart disease

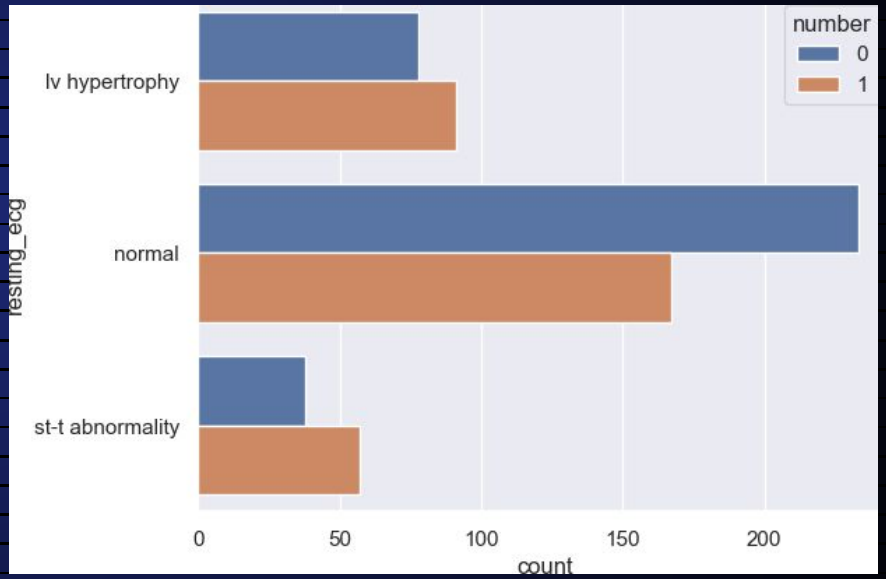


Categorical EDA - resting_ecg

ECG refers to electrocardiogram.

Resting_ecg basically measures the heart's electrical activity at rest.

From the data, there is no obvious relationship between the presence of heart disease and whether one has heart disease.



03
Machine Learning

ML - Binary Classification *(Linear Regression)*

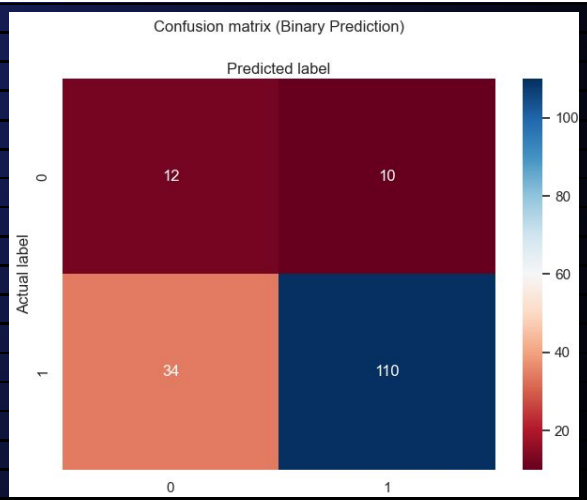
We are able to predict the presence of heart disease with a 73% accuracy, with a TPR of 76%, FPR of 45% and FNR of 24%

From the data, we are able to predict the presence of heart disease with a 73% accuracy, detecting heart disease 76% of the time and failing to detect heart disease 24% of the time.

We falsely identify that heart disease is present 45% of the time, however this is not a big issue since we want to mitigate the presence of heart disease and are interested in correctly detected / failing to detect heart disease.

	precision	recall	f1-score	support
Without heart disease	0.26	0.55	0.35	22
with heart disease	0.92	0.76	0.83	144
accuracy			0.73	166
macro avg	0.59	0.65	0.59	166
weighted avg	0.83	0.73	0.77	166

Confusion Matrix Binary Classification |TP 110 |FP 10 |FN 34 |TN 12
Rates : |TPR 0.76 |FPR 0.45 |FNR 0.24



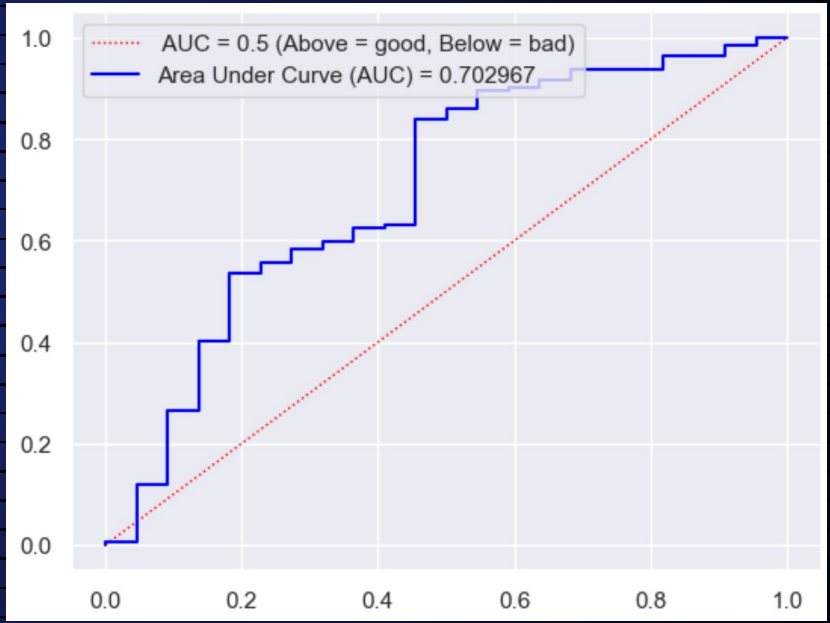
ML - Binary Classification *(Receiver Operating Characteristics)*

Area Under Curve (AUC) of 0.70

- Above 0.5 = better than guessing at random
- Below 0.5 = worse than guessing at random

We are able to predict better than guessing at random

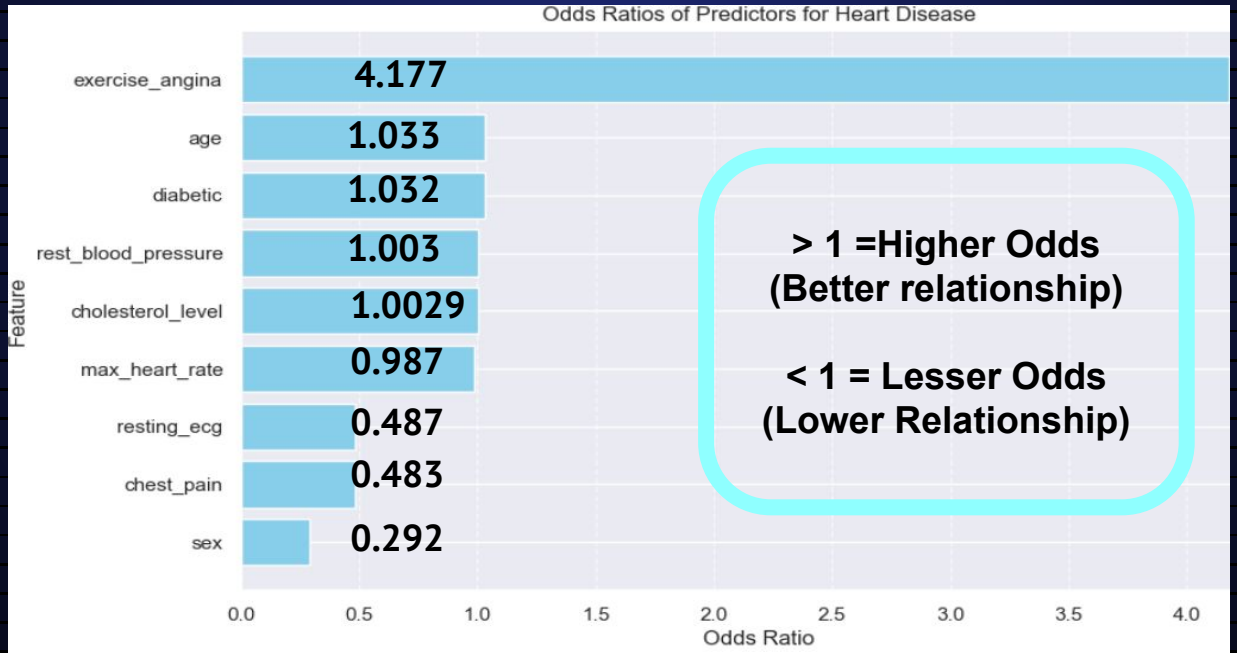
- If a person is speculated to have heart disease, it is good to treat early.



ML - Binary Classification *(Variable Importance)*

Exercise_angina, age, diabetic, rest_blood_pressure, cholesterol_level has a positive indication of the relation between the presence of heart disease.

Exercise angina has the highest odds in predicting heart disease while sex has the least odds.



ML - Random Forest

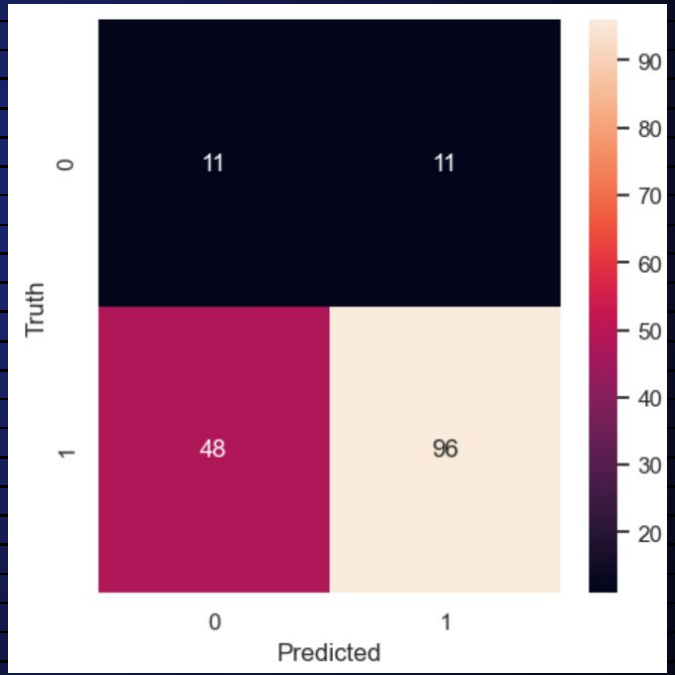
Prediction accuracy of 64%

True Positive Rate of 67%

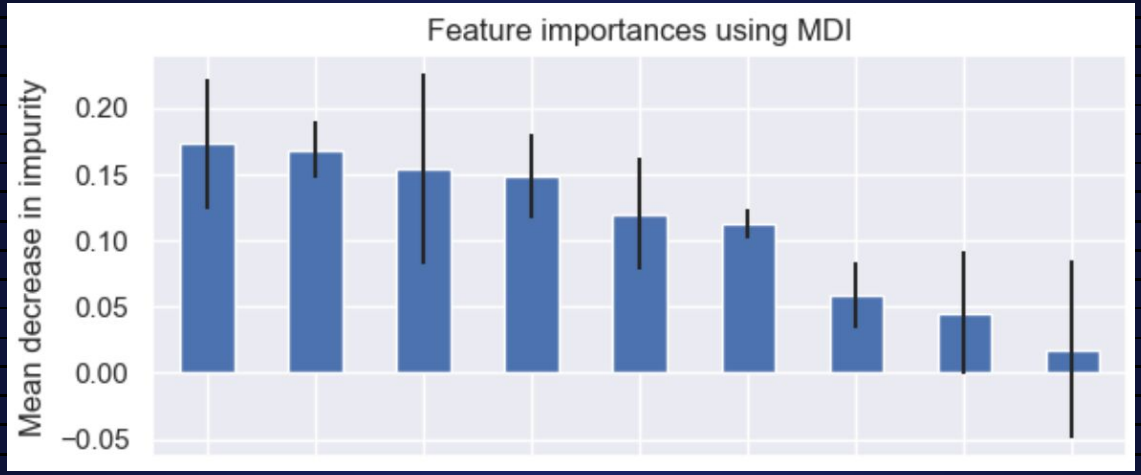
False Positive Rate of 50%

False Negative Rate of 33%

```
Confusion Matrix (Random Forest) |TP 96 |FP 11  |FN 48 |TN 11
Rates : |TPR 0.67 |TNR 0.5 |FPR 0.5 |FNR 0.33
```



ML - Random Forest



ML - Hyperparameter Tuning

Default Parameters:
n_estimators = 100
max_features = sqrt
max_depth = None
max_samples = None
min_samples_split = 2
min_samples_leaf = 1
bootstrap = True

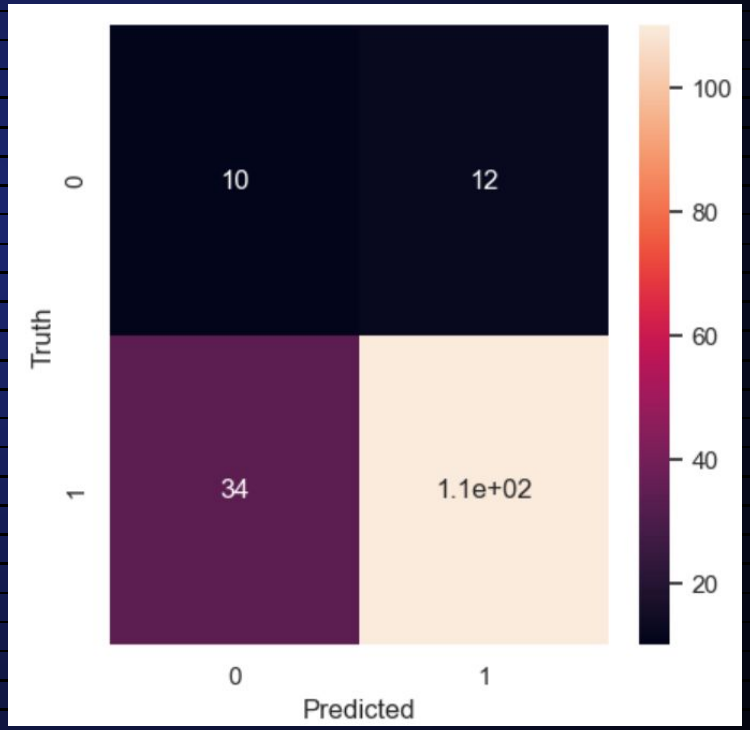
```
# num of decision tree in random forest (default 100)
n_estimators = [20,40,60,80,100,120]

# num of features at every split (default sqrt)
max_features = ['auto', 'sqrt']

# max level of tree (default None)
max_depth = [2,4, 8, None]

# number of samples (default None)
max_samples = [0.5, 0.75, 1.0, None]

# min num of sample to split a node (default 2)
min_samples_split = [2, 5]

# min num of sample at leaf node (default 1)
min_samples_leaf = [1, 2]

# if bootstrap sample used (default True)
bootstrap = [True, False]
```

ML - Randomized Search

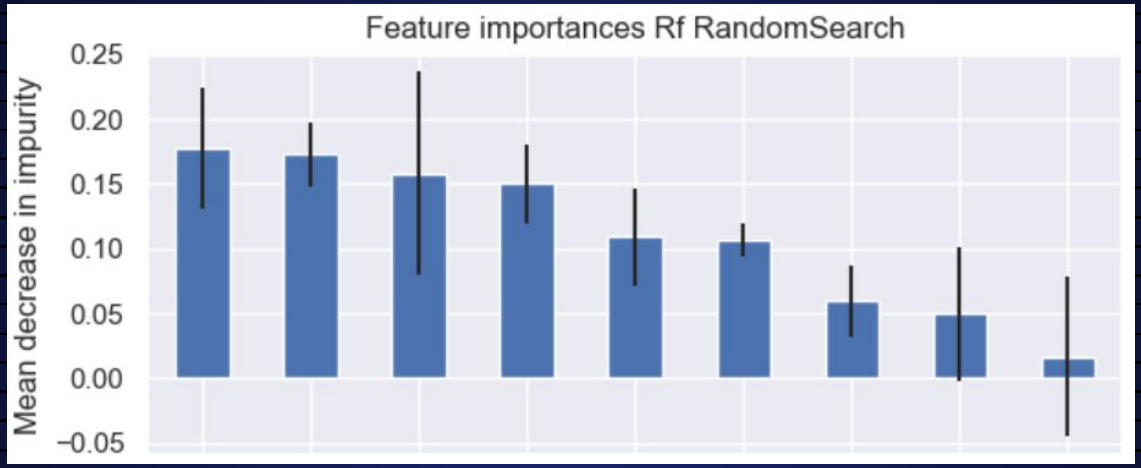
Prediction accuracy: 64% -> 72%
True Positive Rate: 67% -> 76%
False Positive Rate:50% -> 55%
False Negative Rate: 33% -> 24%

n_estimators = 100
max_features = sqrt
max_depth = 4
max_samples = none
min_samples_split = 5
min_samples_leaf = 1
bootstrap = True

```
Confusion Matrix (RF RandomSearch)|TP 110 |FP 12  |FN 34 |TN 10
Rates :  |TPR 0.76 |TNR 0.45 |FPR 0.55 |FNR 0.24
```



ML - Randomized Search

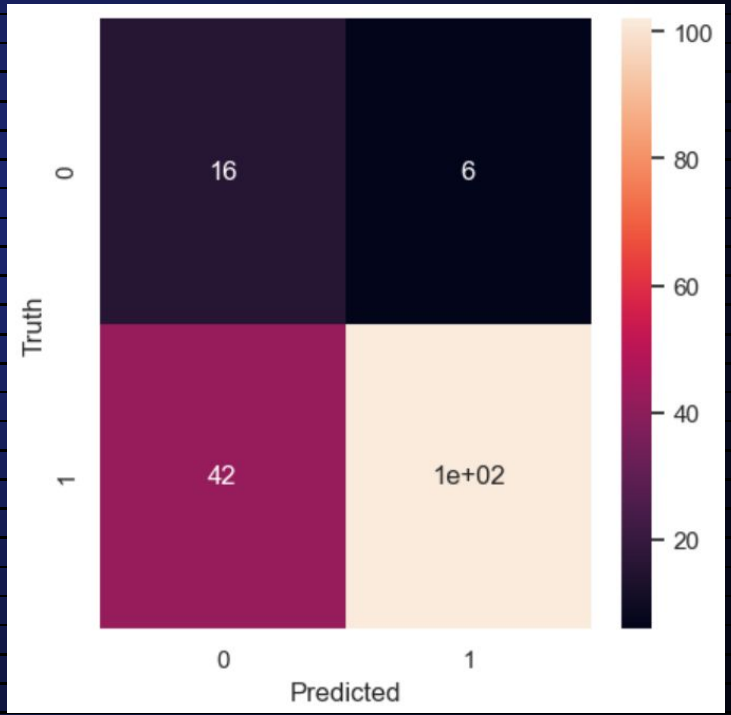


ML - Grid Search

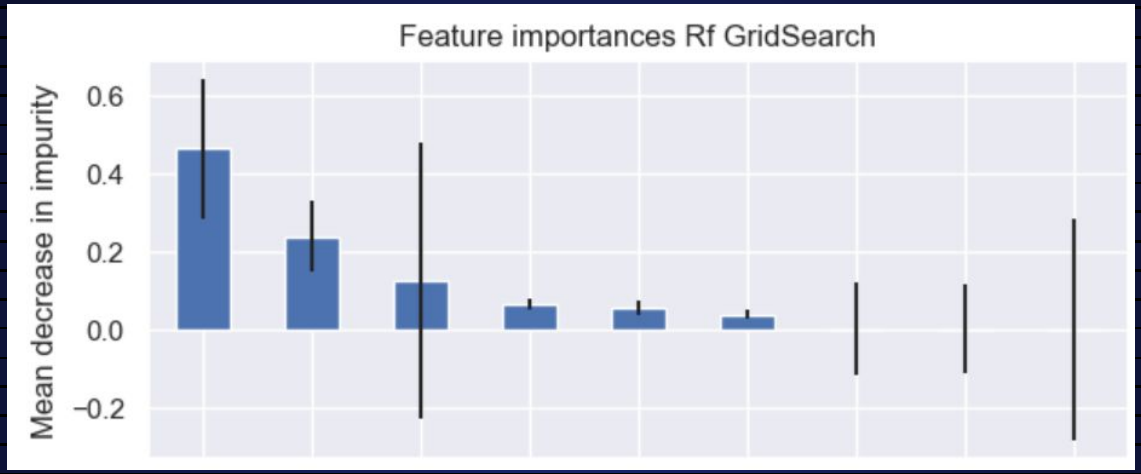
Prediction accuracy: 64% -> 71%
True Positive Rate: 67% -> 71%
False Positive Rate: 50% -> 27%
False Negative Rate: 33% -> 29%

n_estimators = 20
max_features = sqrt
max_depth = 2
max_samples = 1
min_samples_split = 2
min_samples_leaf = 1
bootstrap = True

```
Confusion Matrix (RF Gridsearch) |TP 102 |FP 6  |FN 42 |TN 16
Rates :  |TPR 0.71 |TNR 0.73 |FPR 0.27 |FNR 0.29
```



ML - Grid Search



ML - Comparing

Random Forest

Prediction accuracy: 64%

True Positive Rate: 67%
False Negative Rate: 33%
False Positive Rate: 50%

Variable Importance:

1. Chest Pain
2. Age
3. Cholesterol
4. Heart Rate
5. Exercise Angina
6. Blood Pressure

Randomized Search

Prediction accuracy: 72%

True Positive Rate: 75%
False Negative Rate: 24%
False Positive Rate: 55%

Variable Importance:

1. Age
2. Chest Pain
3. Heart Rate
4. Cholesterol
5. Blood Pressure
6. Exercise Angina

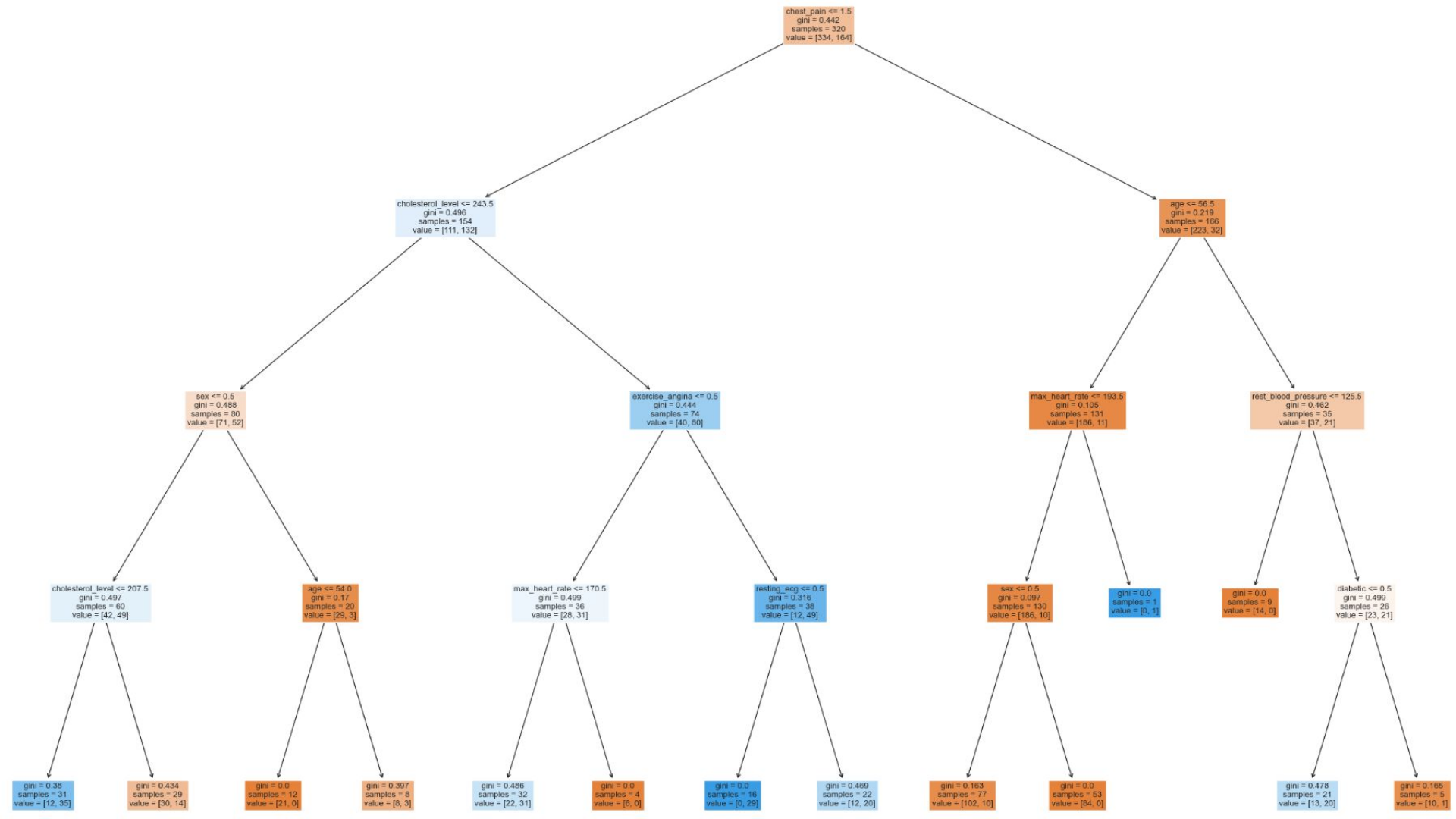
Grid Search

Prediction accuracy of 71%

True Positive Rate of 71%
False Negative Rate: 29%
False Positive Rate of 27%

Variable Importance:

1. Chest Pain
2. Exercise Angina
3. Age

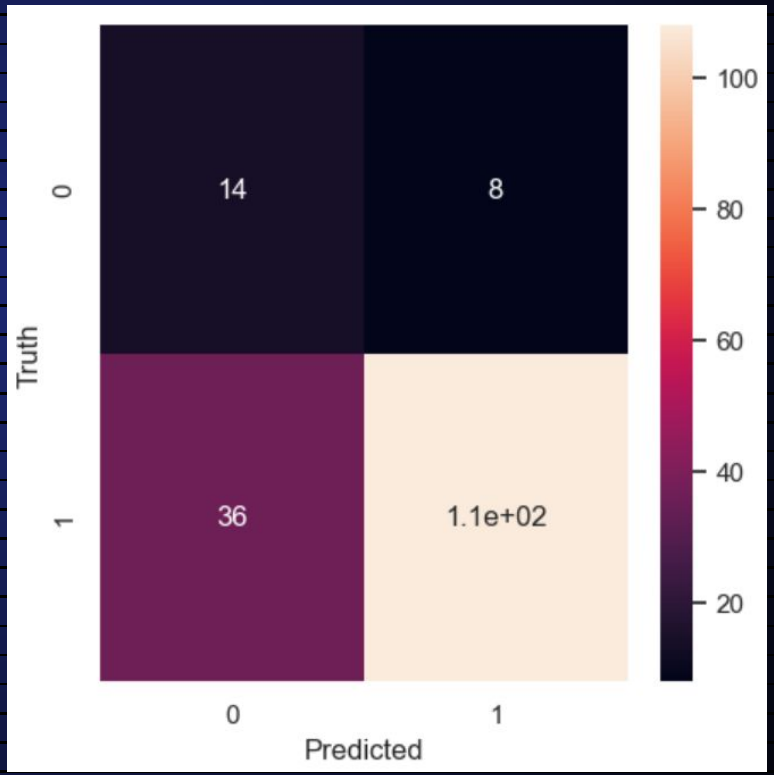

chest_pain <= 1.5
gini = 0.442
samples = 320
value = [334, 164]
cholesterol_level <= 243.5
gini = 0.496
samples = 154
value = [111, 132]
age <= 56.5
gini = 0.219
samples = 166
value = [223, 32]
sex <= 0.5
gini = 0.488
samples = 80
value = [71, 52]
exercise_angina <= 0.5
gini = 0.444
samples = 74
value = [40, 80]
max_heart_rate <= 193.5
gini = 0.105
samples = 131
value = [186, 11]
rest_blood_pressure <= 125.5
gini = 0.462
samples = 35
value = [37, 21]
cholesterol_level <= 207.5
gini = 0.497
samples = 60
value = [42, 49]
age <= 54.0
gini = 0.17
samples = 20
value = [29, 3]
max_heart_rate <= 170.5
gini = 0.499
samples = 36
value = [28, 31]
resting_ecg <= 0.5
gini = 0.316
samples = 38
value = [12, 49]
sex <= 0.5
gini = 0.097
samples = 130
value = [186, 10]
gini = 0.0
samples = 1
value = [0, 1]
gini = 0.0
samples = 9
value = [14, 0]
diabetic <= 0.5
gini = 0.499
samples = 26
value = [23, 21]
gini = 0.35
samples = 31
value = [12, 35]
gini = 0.434
samples = 29
value = [30, 14]
gini = 0.0
samples = 12
value = [21, 0]
gini = 0.397
samples = 8
value = [8, 3]
gini = 0.486
samples = 32
value = [22, 31]
gini = 0.0
samples = 4
value = [6, 0]
gini = 0.0
samples = 16
value = [0, 29]
gini = 0.469
samples = 22
value = [12, 20]
gini = 0.163
samples = 77
value = [102, 10]
gini = 0.0
samples = 53
value = [84, 0]
gini = 0.478
samples = 21
value = [13, 20]
gini = 0.165
samples = 5
value = [10, 1]

ML - Support Vector Machine

Prediction accuracy of 73%

True Positive Rate of 75%
False Negative Rate of 25%
False Positive Rate of 36%

```
Confusion Matrix (RF Gridsearch) |TP 108 |FP 8  |FN 36 |TN 14
Rates :  |TPR 0.75 |TNR 0.64 |FPR 0.36 |FNR 0.25
```



ML - Summary

Random Forest

Prediction accuracy: 64%

True Positive Rate: 67%
False Negative Rate: 33%
False Positive Rate: 50%

Randomized Search

Prediction accuracy: 72%

True Positive Rate: 75%
False Negative Rate: 24%
False Positive Rate: 55%

Grid Search

Prediction accuracy of 71%

True Positive Rate of 71%
False Negative Rate: 29%
False Positive Rate of 27%

Binary Classification

Prediction accuracy: 73%

True Positive Rate: 76%
False Negative Rate: 24%
False Positive Rate: 45%

SVM

Prediction accuracy: 73%

True Positive Rate: 75%
False Negative Rate: 25%
False Positive Rate: 36%

ML - Data Insights

Random Forest

Prediction accuracy: 64%

True Positive Rate: 67%
False Negative Rate: 33%
False Positive Rate: 50%

Randomized Search

Prediction accuracy: 72%

True Positive Rate: 75%
False Negative Rate: 24%
False Positive Rate: 55%

Grid Search

Prediction accuracy of 71%

True Positive Rate of 71%
False Negative Rate: 29%
False Positive Rate of 27%

Binary Classification

Prediction accuracy: 73%

True Positive Rate: 76%
False Negative Rate: 24%
False Positive Rate: 45%

SVM

Prediction accuracy: 73%

True Positive Rate: 75%
False Negative Rate: 25%
False Positive Rate: 36%

ML - Insights

Random Forest

Variable Importance:

1. Chest Pain
2. Age
3. Cholesterol
4. Heart Rate
5. Exercise Angina
6. Blood Pressure

Grid Search

Variable Importance:

1. Chest Pain
2. Exercise Angina
3. Age

Randomized Search

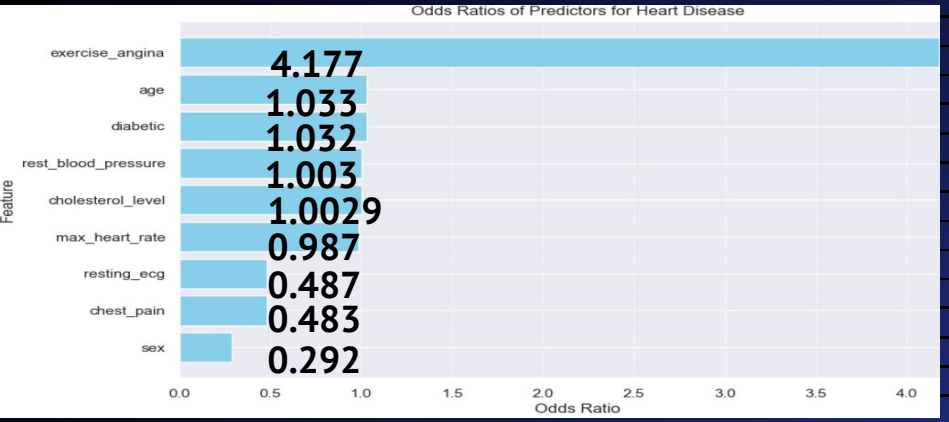
Variable Importance:

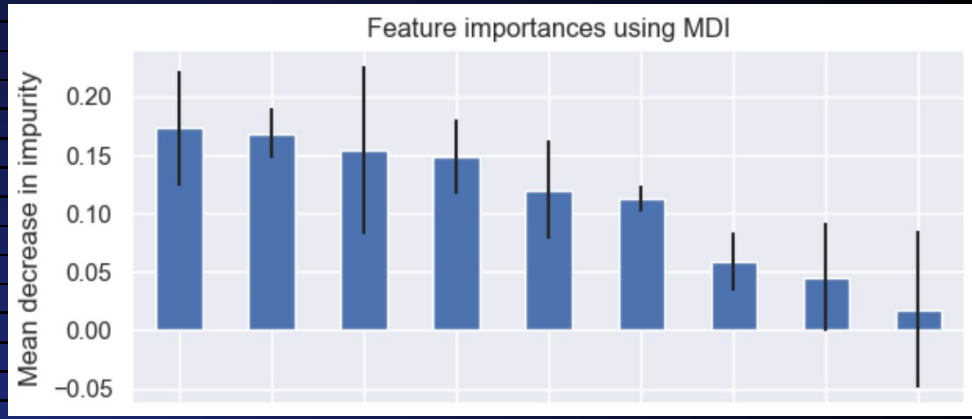
1. Age
2. Chest Pain
3. Heart Rate
4. Cholesterol
5. Blood Pressure
6. Exercise Angina

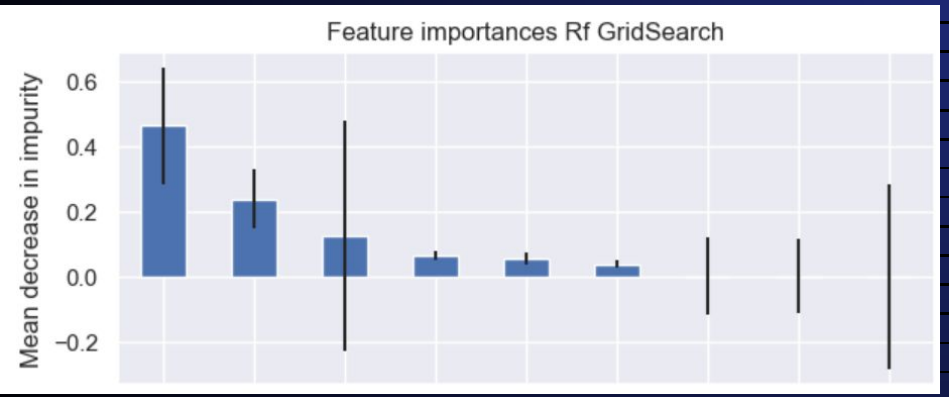
Binary Classification

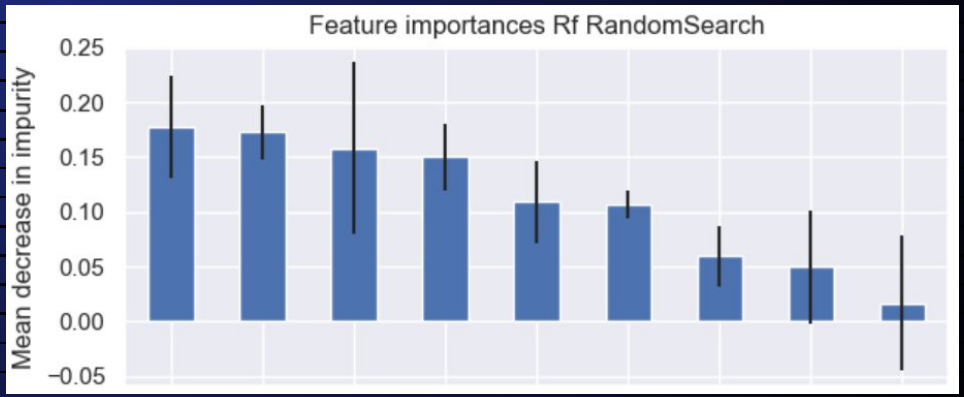
Variable Importance:

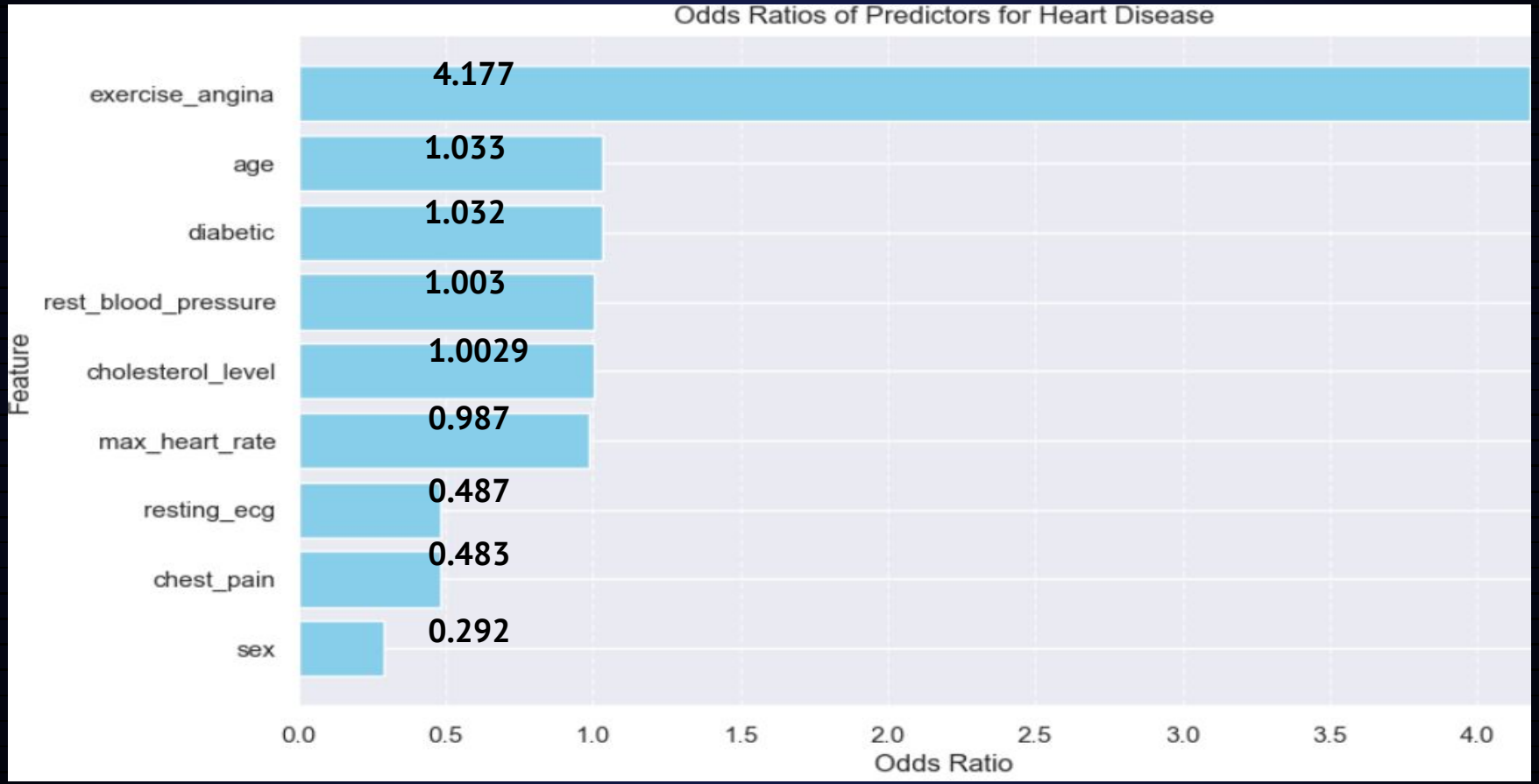
1. Exercise Angina
2. Age
3. Diabetic
4. Blood Pressure
5. Cholesterol
6. Heart Rate


Odds Ratios of Predictors for Heart Disease
exercise_angina
age
diabetic
rest_blood_pressure
cholesterol_level
max_heart_rate
resting_ecg
chest_pain
sex
4.177
1.033
1.032
1.003
1.0029
0.987
0.487
0.483
0.292
Feature
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Odds Ratio


Feature importances using MDI
Mean decrease in impurity
0.20
0.15
0.10
0.05
0.00
−0.05


Feature importances Rf GridSearch
Mean decrease in impurity
0.6
0.4
0.2
0.0
−0.2

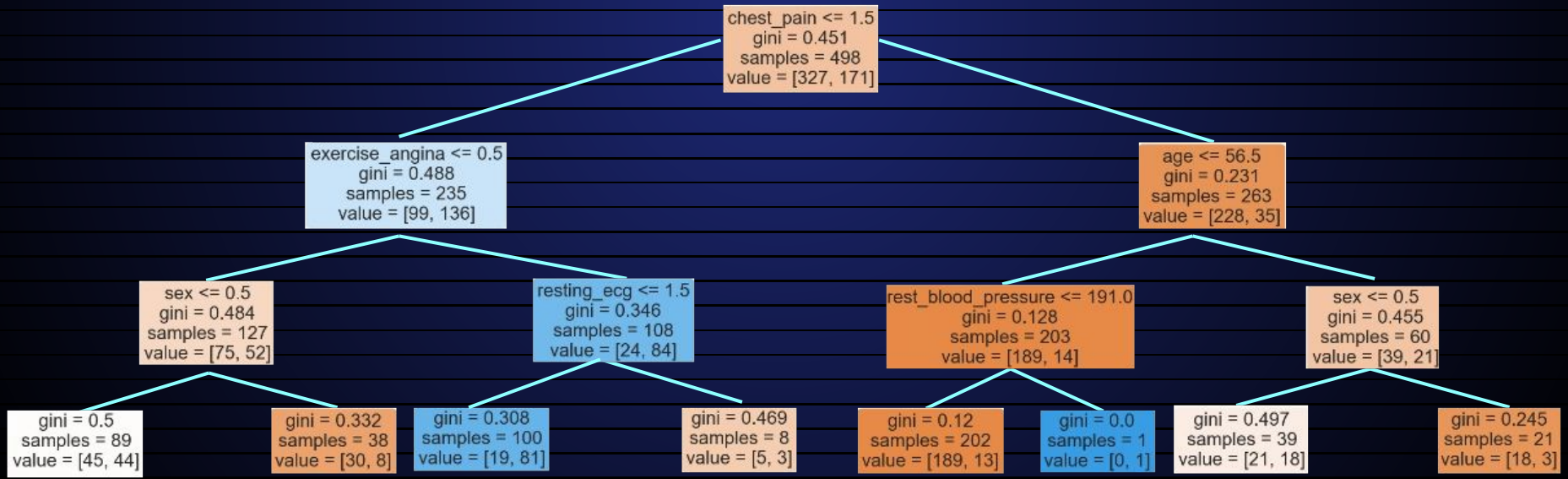

Feature importances Rf RandomSearch
Mean decrease in impurity
0.25
0.20
0.15
0.10
0.05
0.00
−0.05


Odds Ratios of Predictors for Heart Disease
exercise_angina
age
diabetic
rest_blood_pressure
cholesterol_level
max_heart_rate
resting_ecg
chest_pain
sex
4.177
1.033
1.032
1.003
1.0029
0.987
0.487
0.483
0.292
Feature
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Odds Ratio

ML - Insights

No Heart Disease
Chest pain: non-anginal or atypical angina
Age: Less than 56
Blood Pressure: Less than 191

Have Heart Disease
Chest Pain: Typical angina or asymptomatic
Exercise Angina: True
Resting ECG: LV Hypertrophy or Normal



Project Outcome - Recap

Problem Statement

Based on data provided, are we able to effectively predict if a person has heart disease based on the symptoms exhibited by the person

Project Outcome

- **73% Prediction Accuracy**
- **76% True Positive Rate (Successful Identifications)**
- **24% False Negative Rate (Failed identifications)**

- We could potentially save 16 out of the 23 people that die from cardiovascular diseases everyday through early identification of heart disease.

  According to Singapore Heart Foundation, “23 people die from cardiovascular disease everyday. Cardiovascular disease accounted for 31.4% of all deaths in 2022, amounting to almost 1 out of 3 deaths in Singapore due to heart disease”

Addressing the problem

With a 73% accuracy, successfully identifying 76% (TPR) of the cases with heart disease and failing to identify the presence of heart disease 24% (FNR) of the time within people who have heart disease, we could potentially save 16 out of the 23 people that die from cardiovascular diseases everyday through early identification of heart disease.

According to Singapore Heart Foundation, “23 people die from cardiovascular disease everyday. Cardiovascular disease accounted for 31.4% of all deaths in 2022, amounting to almost 1 out of 3 deaths in Singapore due to heart disease”

Future Outlook

- Create an web application
  - Key in symptoms as per the table to check the odds they have heart disease
    - People may be more motivated to do health screenings
      - 60% of Singaporeans in 2022 dont carry out health screenings

https://doctoranywhere.com/preventive-health-survey-2022/

References

1. Stern, et al. (2003, October 7). How aging affects your heart. Aging and Diseases of the Heart. https://www.ahajournals.org/doi/full/10.1161/01.cir.0000086898.96021.b9
2. LeWine, H. E. (2023, June 13). What your heart rate is telling you. Harvard Health. https://www.health.harvard.edu/heart-health/what-your-heart-rate-is-telling-you#:~:text=Your%20maximum%20heart%20rate%20plays,of%20heart%20attack%20and%20death.
3. Singapore Heart Foundation. (2022). Heart disease statistics. https://www.myheart.org.sg/health/heart-disease-statistics/