Heart Disease Correlation Study

Analyzing Risk Factors and Predictors

DA - 01

Heart Disease Dataset Attribute Description

S.No.	Attribute	Code given	Unit	Data type
1	age	Age	in years	Numeric
2	sex	Sex	1, 0	Binary
3	chest pain type	chest pain type	1,2,3,4	Nominal
4	resting blood pressure	resting bp s	in mm Hg	Numeric
5	serum cholesterol	cholesterol	in mg/dl	Numeric
6	fasting blood sugar	fasting blood sugar	1,0 > 120 mg/dl	Binary
7	resting electrocardiogram results	resting ecg	0,1,2	Nominal
8	maximum heart rate achieved	max heart rate	71–202	Numeric
9	exercise induced angina	exercise angina	0,1	Binary
10	oldpeak =ST	oldpeak	depression	Numeric
11	the slope of the peak exercise ST segment	ST slope	0,1,2	Nominal
12	class	target	0,1	Binary

Description of Nominal Attributes

Attribute	Description
Sex	1 = male, 0= female;
Chest Pain Type	-- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
Fasting Blood sugar	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
Resting electrocardiogram results	-- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
Exercise induced angina	1 = yes; 0 = no
the slope of the peak exercise ST segment	-- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
class	1 = heart disease, 0 = Normal

Source

Identifying the independent and dependent attributes

Independent Attributes

- age
- sex
- chest pain type
- resting blood pressure
- serum cholesterol
- fasting blood sugar
- resting electrocardiogram results
- maximum heart rate achieved
- exercise-induced angina
- oldpeak
- the slope of the peak exercise ST segment

Dependent Attribute (Target):

- Class

Characterizing the independent attributes

A. Nominal: cramer's V correlation

1. chest pain type
2. resting electrocardiogram results
3. the slope of the peak exercise ST segment

B. Categorical (Binary): cramer's V correlation

1. sex
2. fasting blood sugar
3. exercise-induced angina

C. Continuous: Point-biserial correlation

1. age
2. resting blood pressure
3. serum cholesterol
4. maximum heart rate achieved
5. oldpeak (ST)

Correlation coefficients in descending order:

Three most highly correlated attributes {

Attributes	Correlation Coefficients
ST slope	0.570
chest pain type	0.530
exercise angina	0.481
oldpeak	0.398
sex	0.311
age	0.262
fasting blood sugar	0.217
resting ecg	0.124
resting bp s	0.121
cholesterol	-0.198
max heart rate	-0.413

The three most highly correlated independent attributes

1. The slope of the peak exercise ST segment
2. Chest Pain Type
3. Exercise induced angina
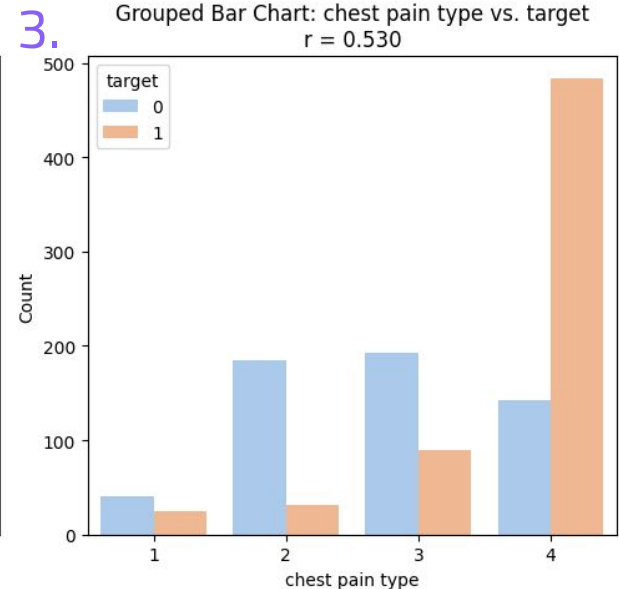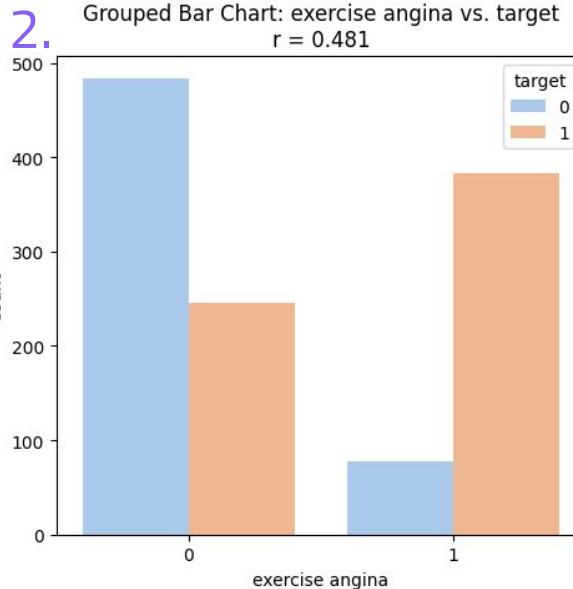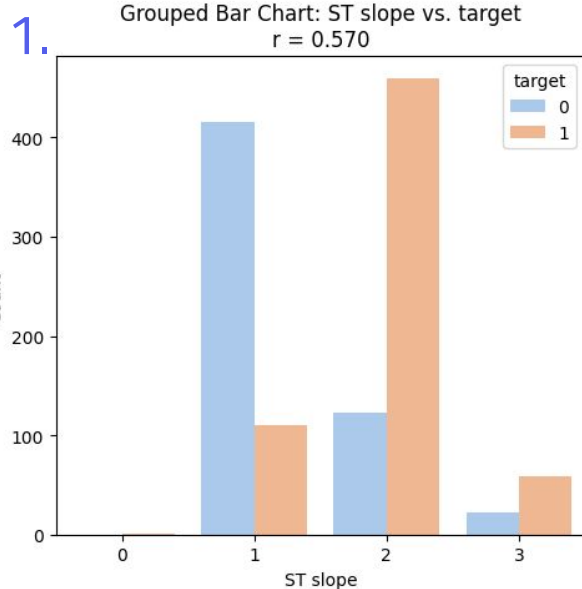
Other Categorical Attributes

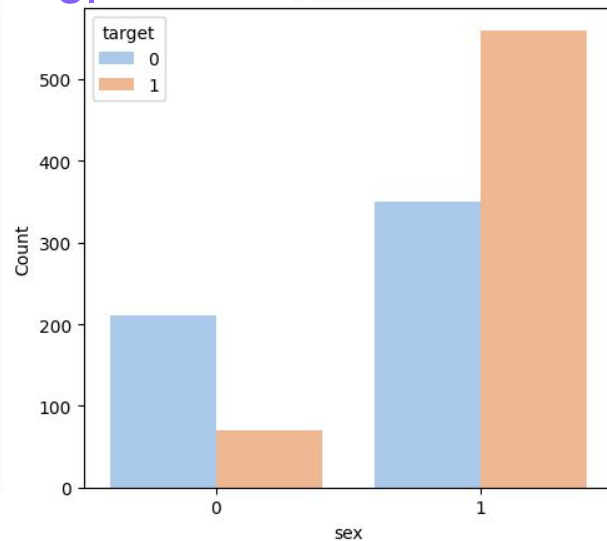- There is no evident correlation between these attributes and Heart Disease

4. Grouped Bar Chart: fasting blood sugar vs. target
r = 0.217
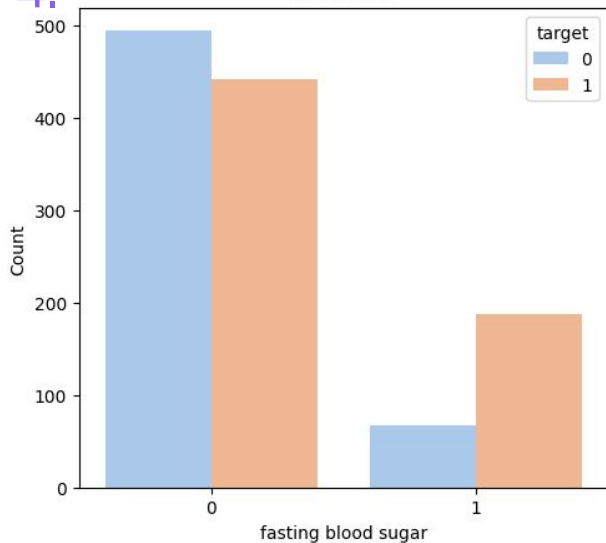
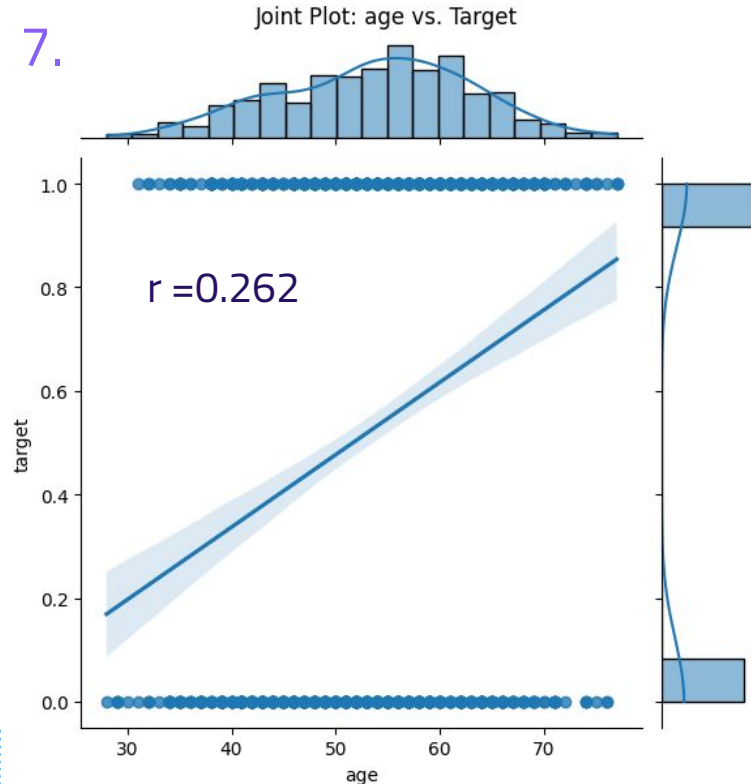**5.** Grouped Bar Chart: resting ecg vs. target
r = 0.124
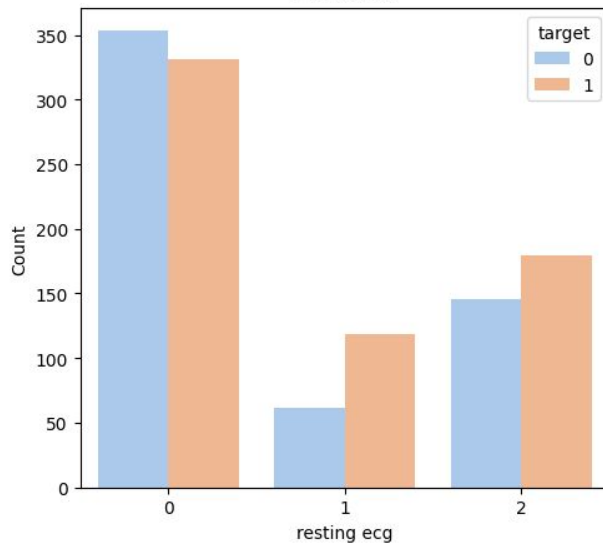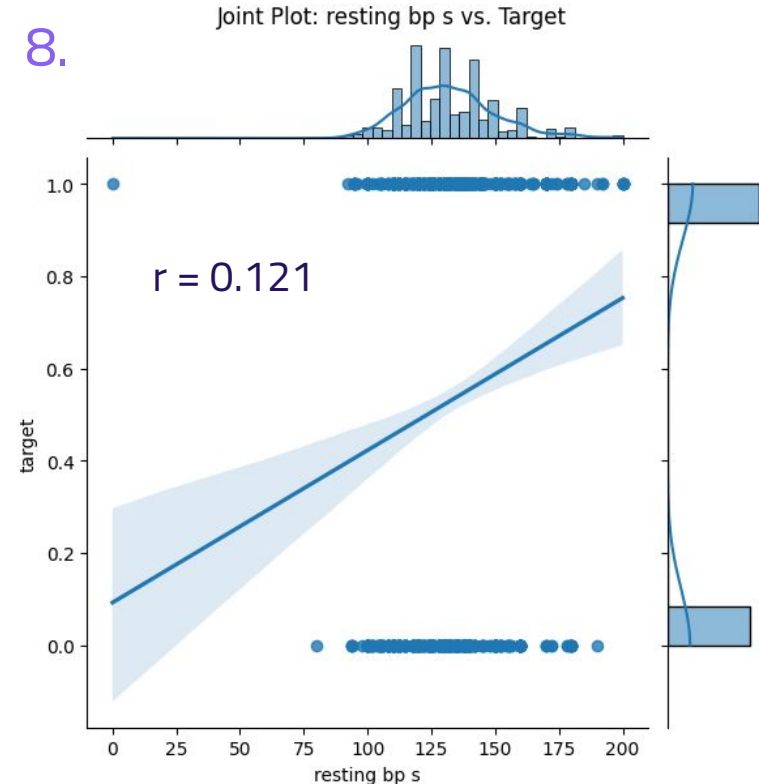
6. Grouped Bar Chart: sex vs. target
r = 0.311

Other continuous attributes

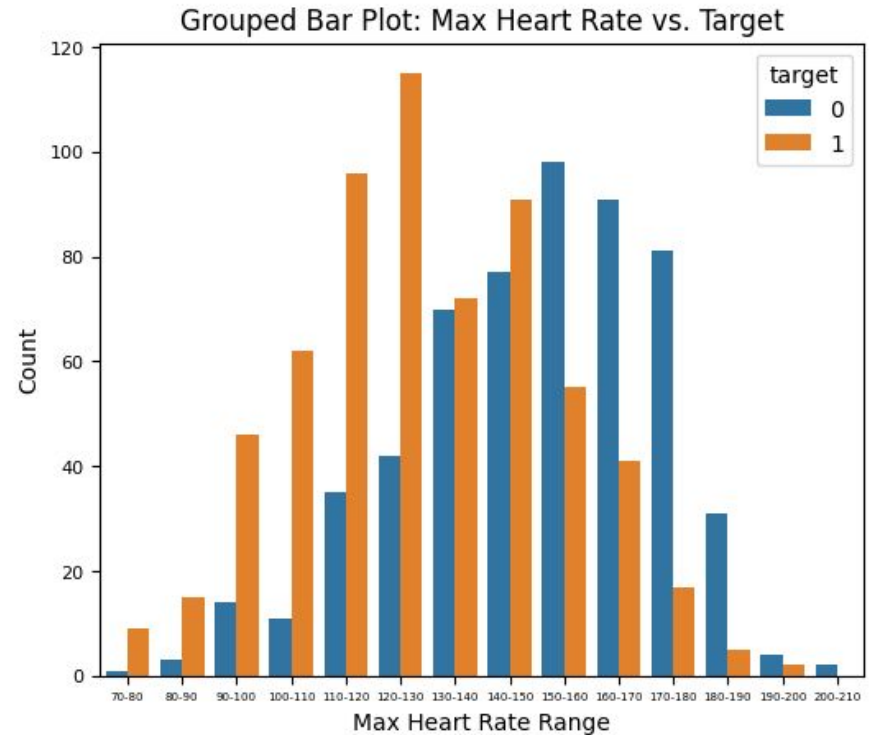- There is no evident correlation between these attributes and Heart Disease
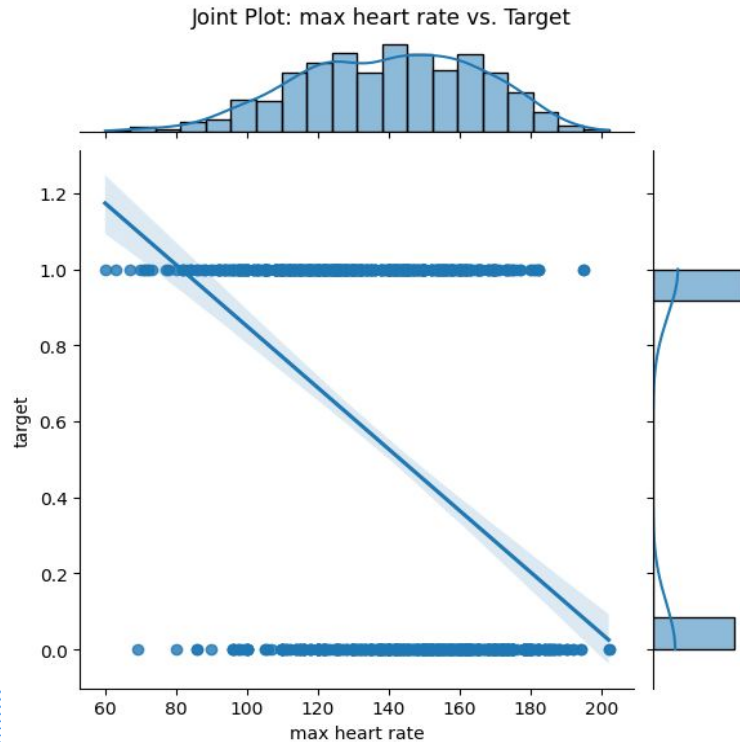
7.

Joint Plot: age vs. Target

r = 0.262

8.

Joint Plot: resting bp s vs. Target

r = 0.121

9. Maximum Heart rate（r = - 0.413）

- There is clear and constant negative correlation between Max Heart Rate and Heart Disease

Joint Plot: max heart rate vs. Target


Grouped Bar Plot: Max Heart Rate vs. Target

10. Old Peak (r = 0.398)

- Despite the abundance of data points with "old peak = 0"and "target = 0".

- A robust correlation is evident for values greater than 0, spanning the range from 0 to 6.

Joint Plot: oldpeak vs. Target


Grouped Bar Plot: oldpeak vs. target
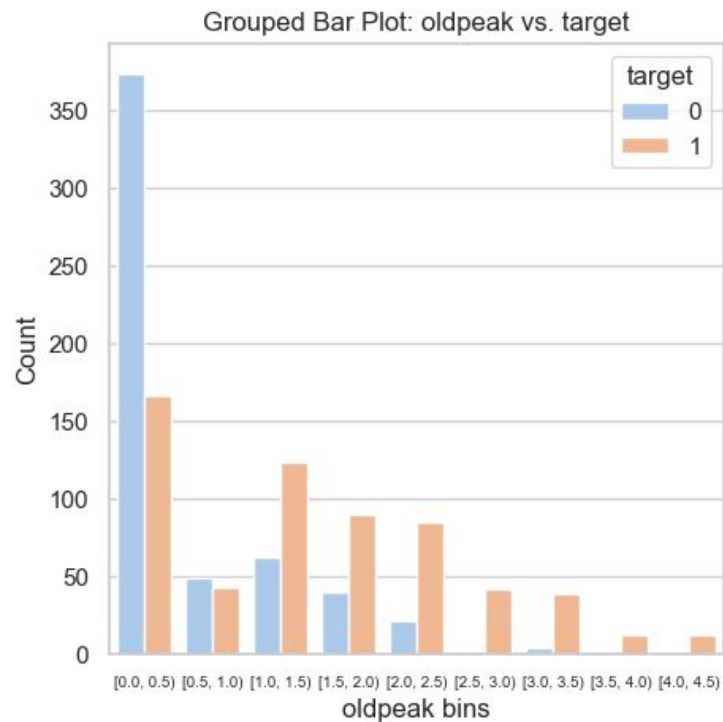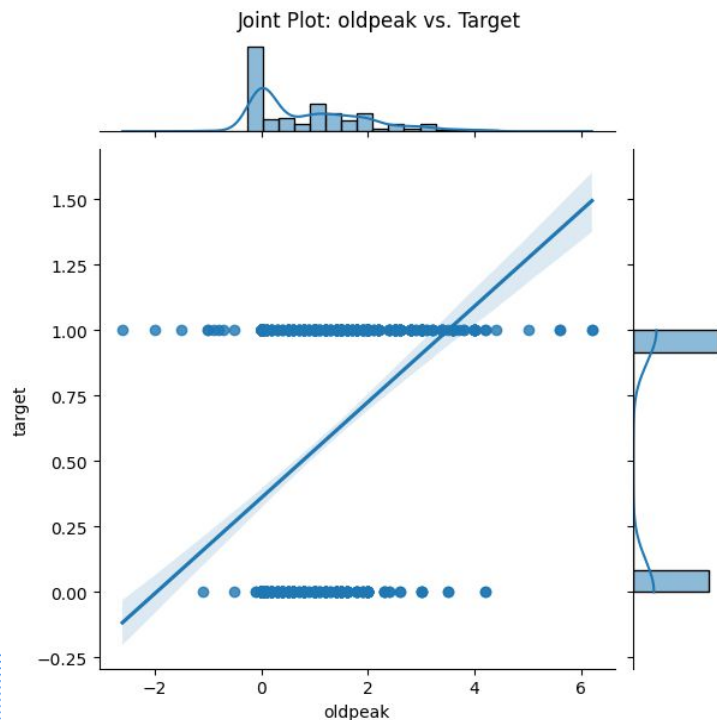
11. Cholesterol

- Contrary to the conventional understanding that associates higher cholesterol levels with an increased risk of heart disease, the dataset reveals an anomaly with a significant number of instances where "Cholesterol = 0" aligns with "target = 1."
- This anomaly led to an apparent negative correlation between cholesterol levels and the occurrence of heart disease.
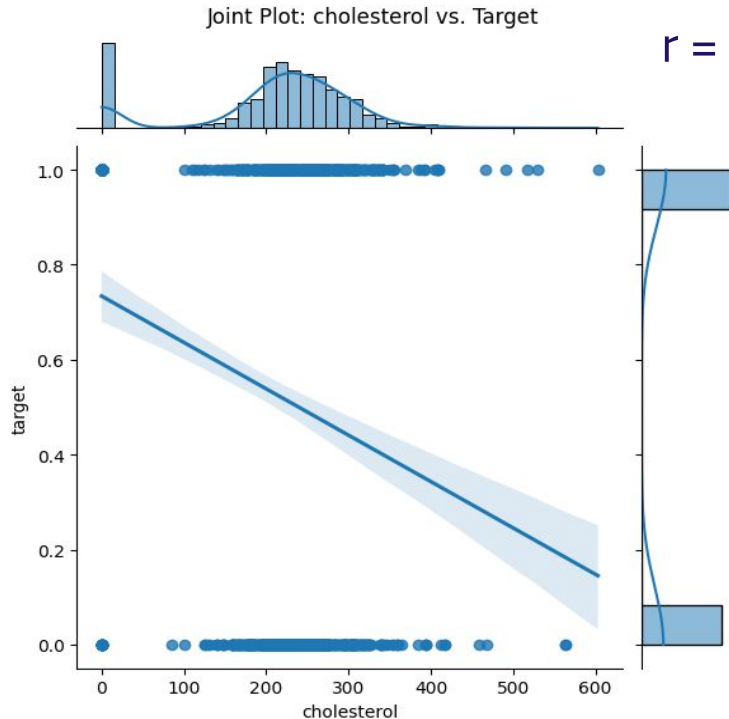
r = - 0.198
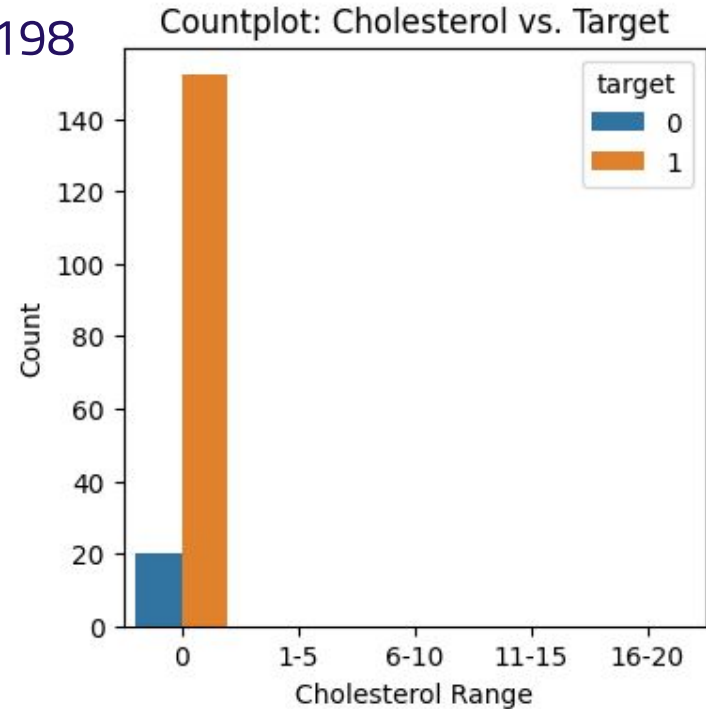
Conclusion :

- Noteworthy correlations emerged from the dataset, with the top three being associated with:
ST slope (0.570)
Chest pain type (0.530)
Exercise-induced angina (0.481)

- The dataset exhibited anomalies, notably with instances of "Cholesterol = 0" coinciding with "target = 1," challenging the expected correlation between higher cholesterol levels and heart disease.

- Despite anomalies, strong correlations were observed in specific ranges, such as for old peak values beyond 0, suggesting nuanced relationships between certain variables and the likelihood of heart disease.

- Further analysis and feature engineering, considering the anomalies, are recommended to enhance the accuracy and interpretability of predictive models.

- The project underscores the importance of thorough data exploration and understanding, as anomalies can significantly impact the interpretation of correlations and relationships within the dataset.

Thirupathi Yuvaraju

21BT10032

Project Code :
https://github.com/Yuvi37/Heart-Disease-Correlation-Study

Thanks !