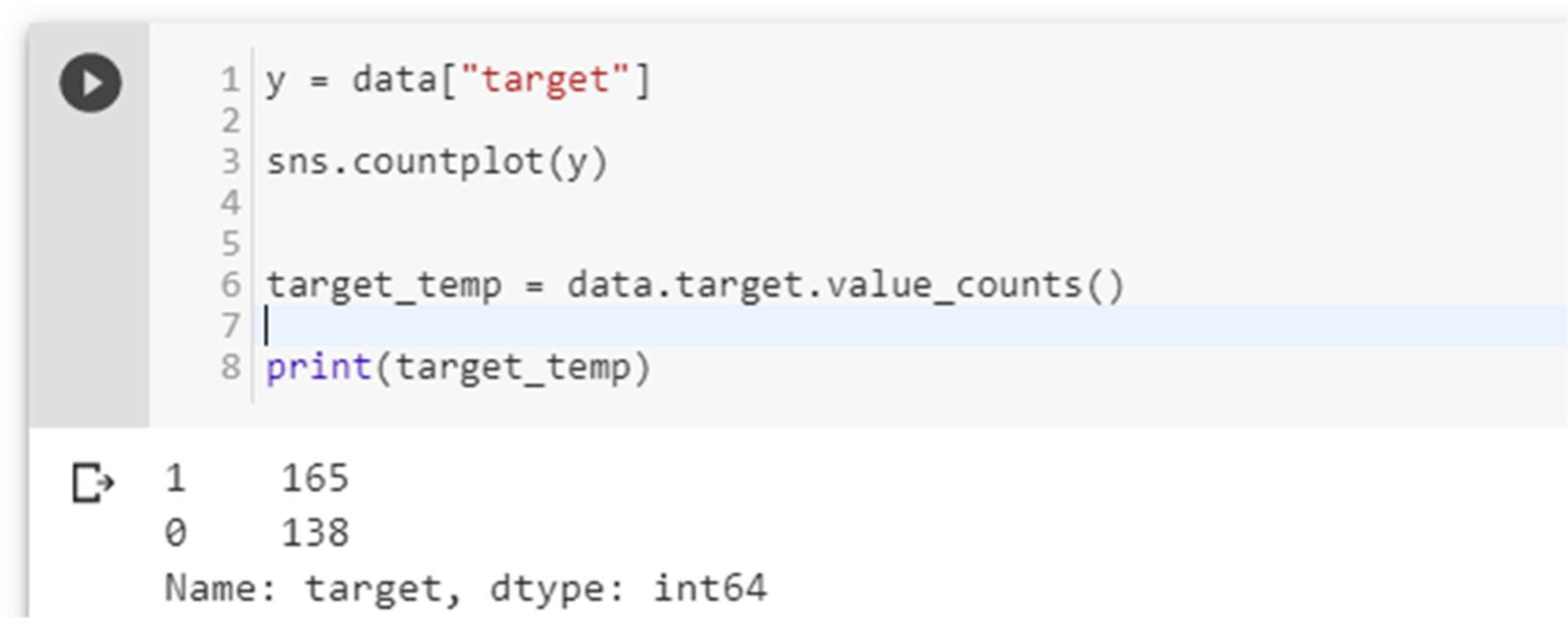
# **Exploratory Data Analysis (EDA)**

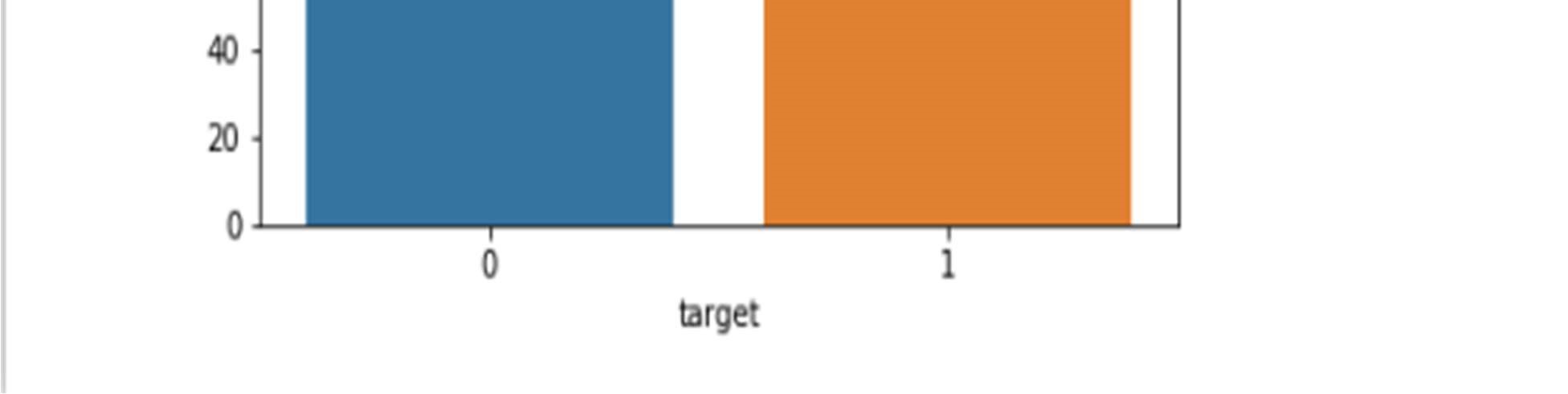
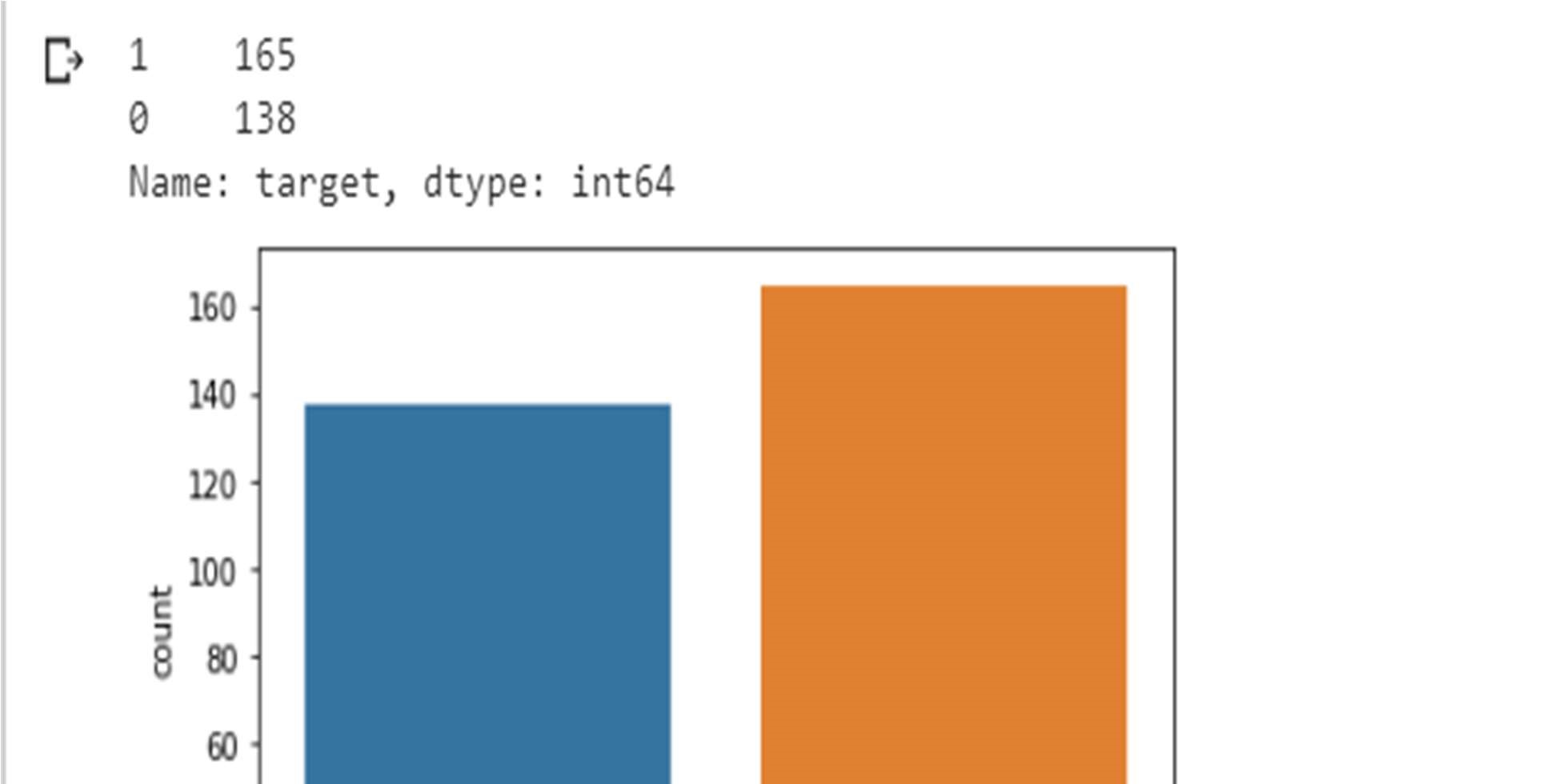


(1 is who have Heart Disease and 0 is who don’t have Heart Disease)

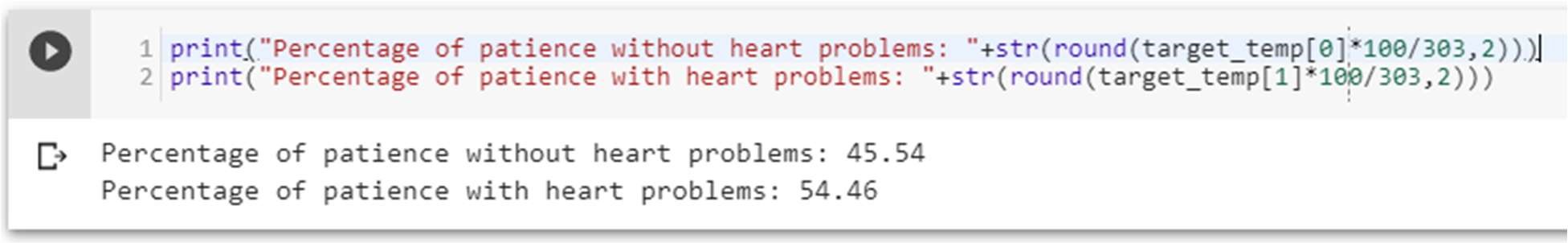
No. of Heart Disease patients is 165. No. of patients who don’t have a heart disease is 138. [Which is a good balance of target data.]

Figure 13

Disease vs Non-Disease Analysis



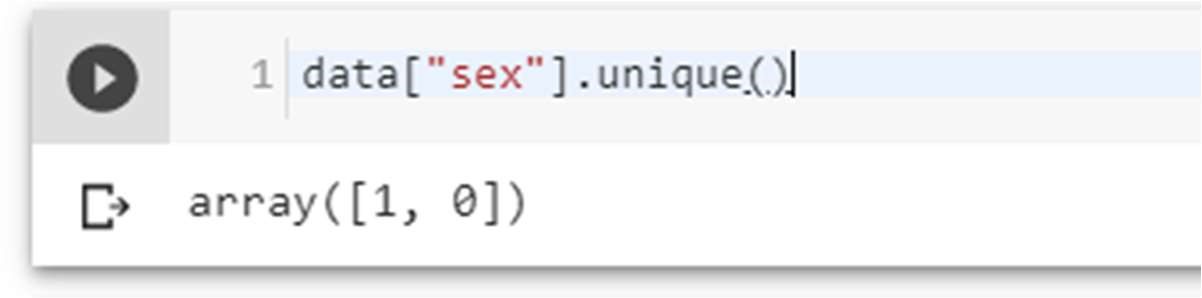
3.5.1 Percentage of patient with or without heart problems in the given dataset



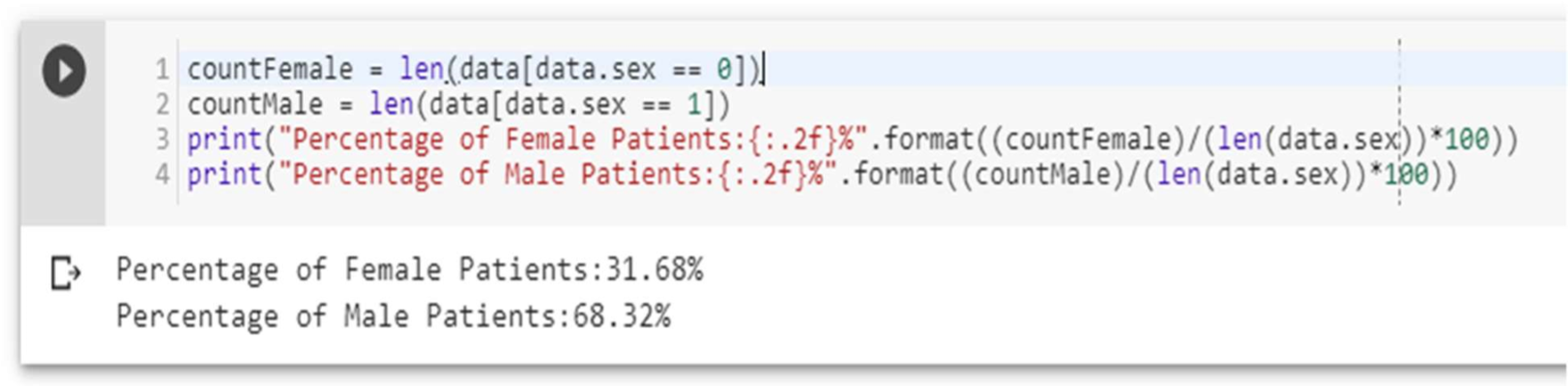
which is a good data distribution

3.5.2 Uniqueness of sex column –

Two sex types: 1 is male and 0 is female



3.5.3 Check the percentage and plot the graph



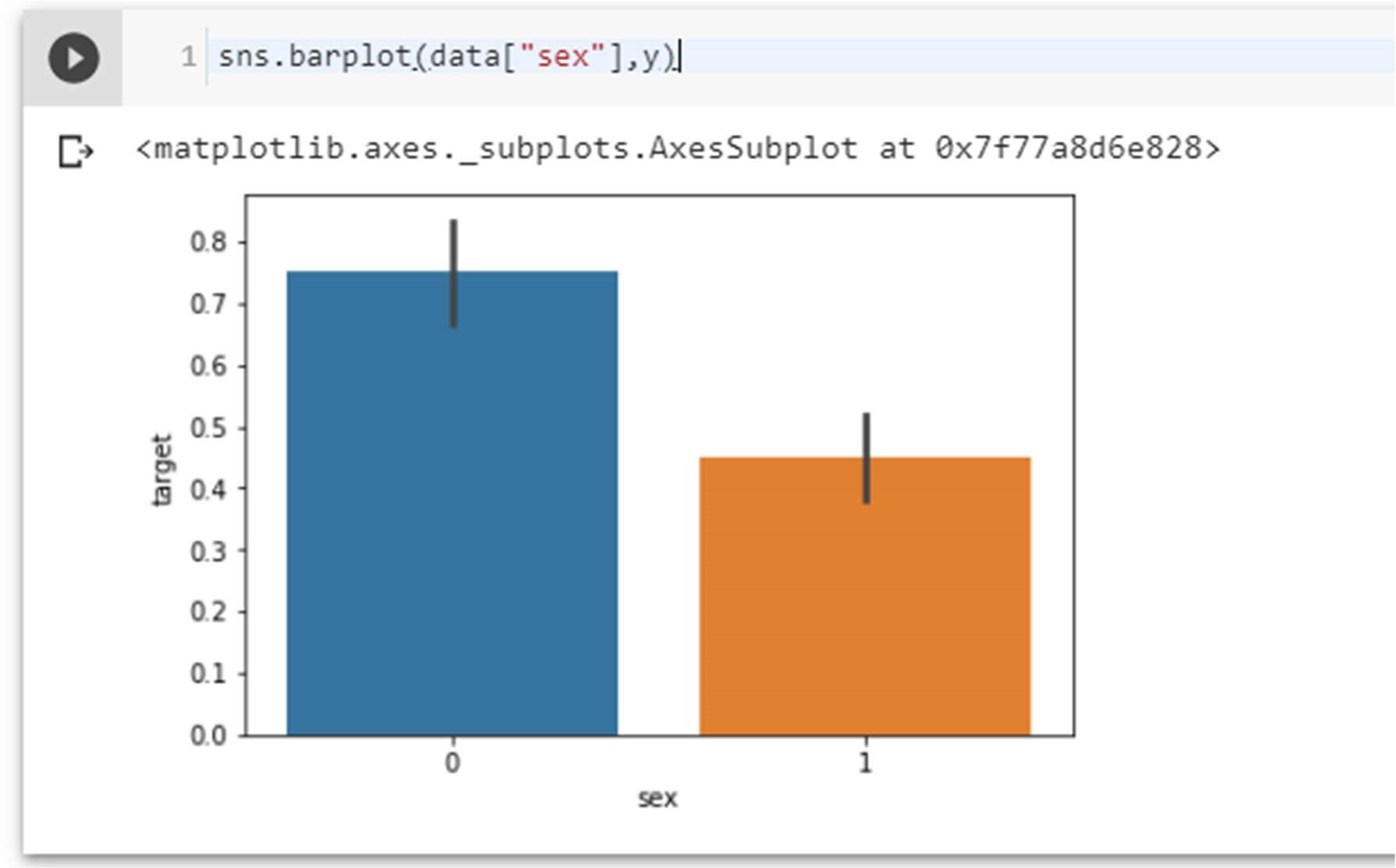


Figure 13 Sex analysis

3.5.4 Heart Disease Frequency for Ages



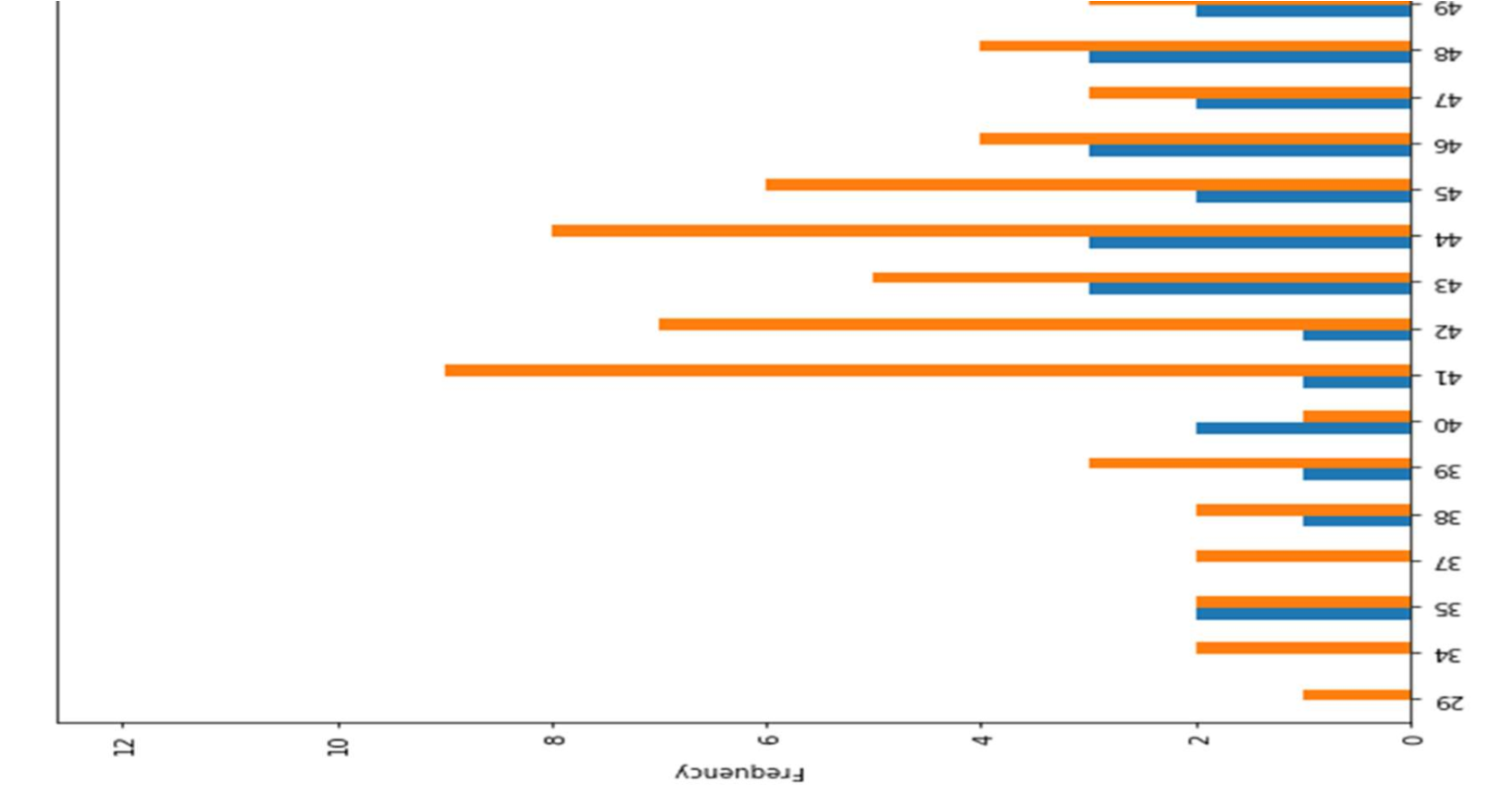
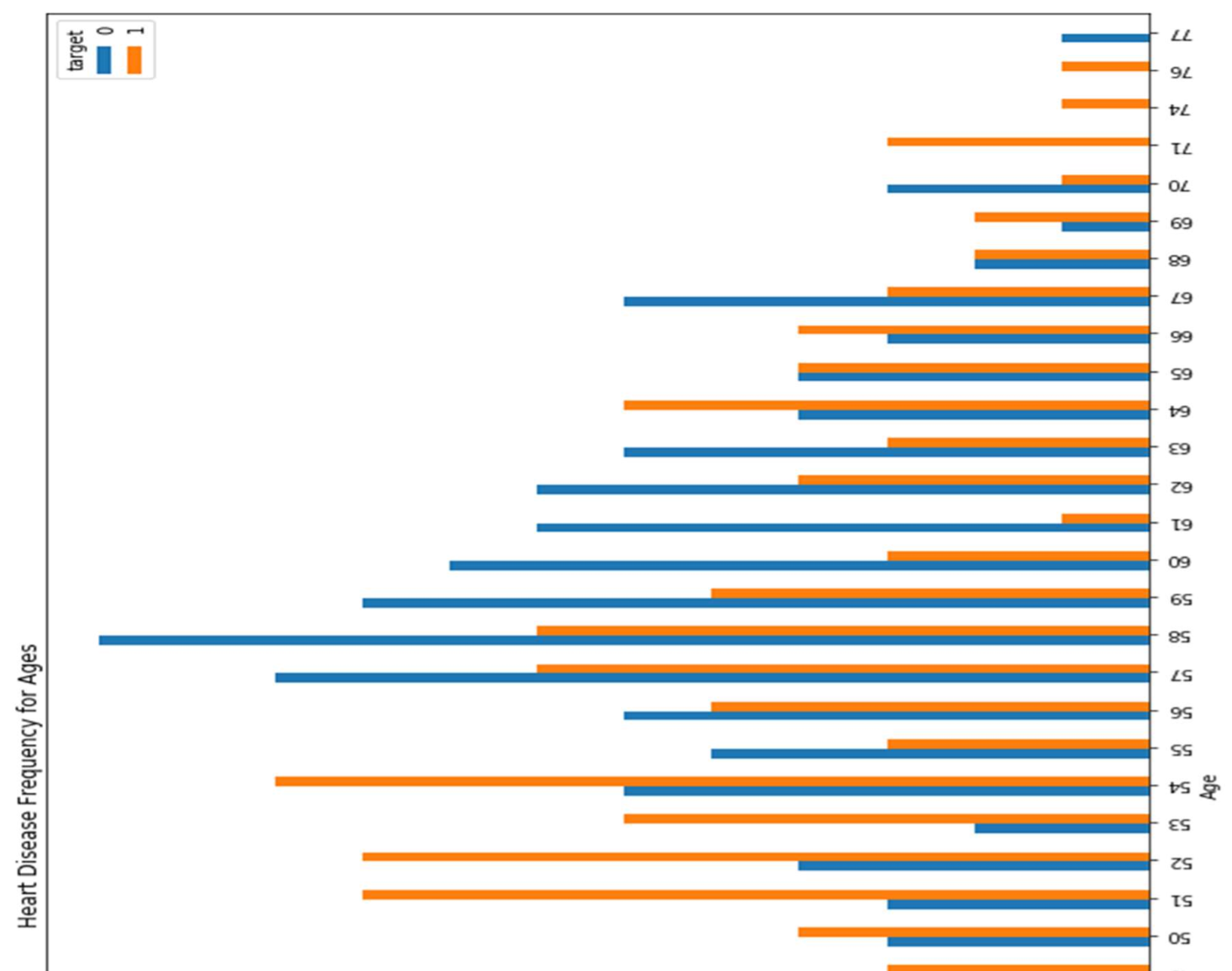


Figure 14 Heart Disease Frequency for ages

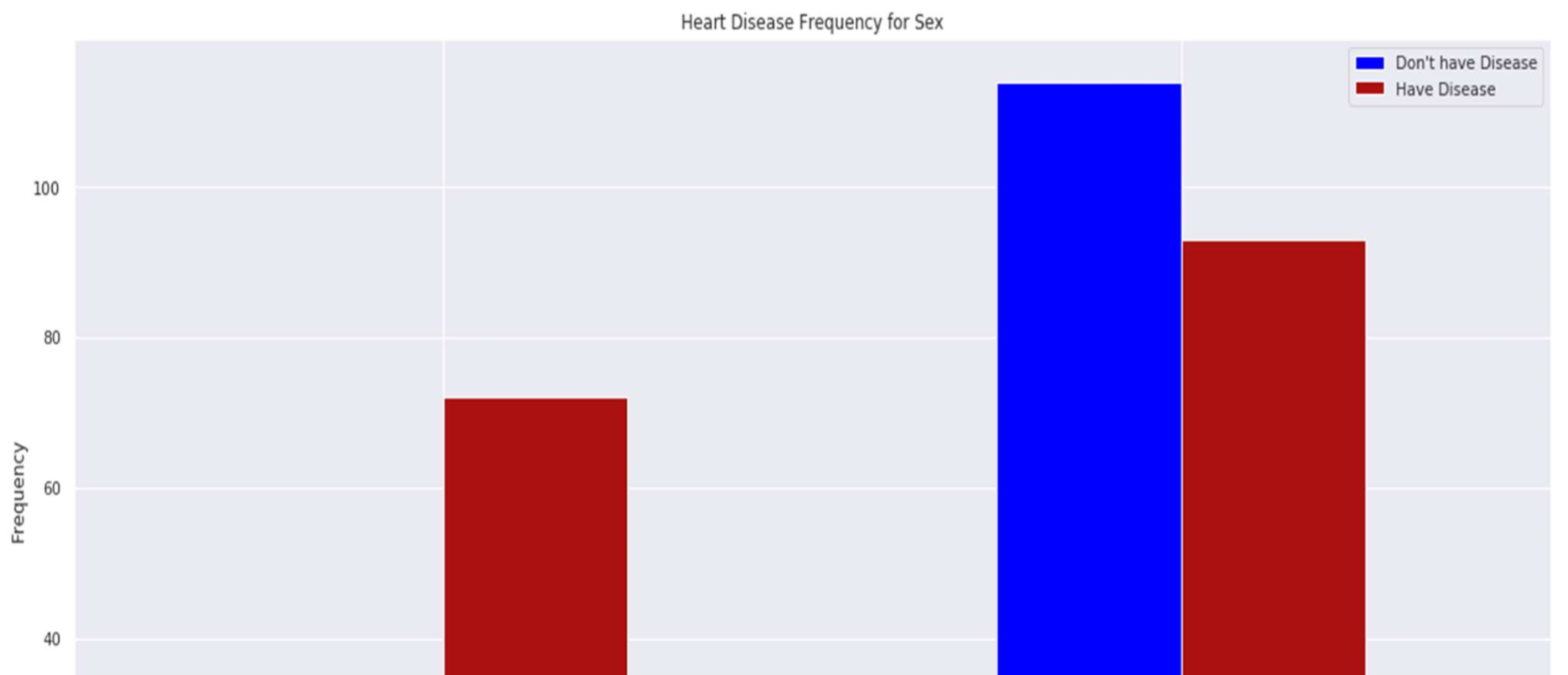
3.5.5 Heart Disease Frequency for sex



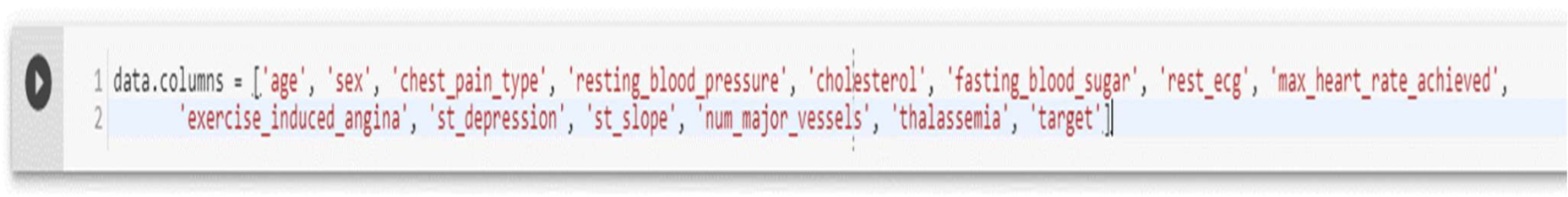
Where 1 is “male”, 0 is “female”.

Figure 15

Heart Disease Frequency for sex

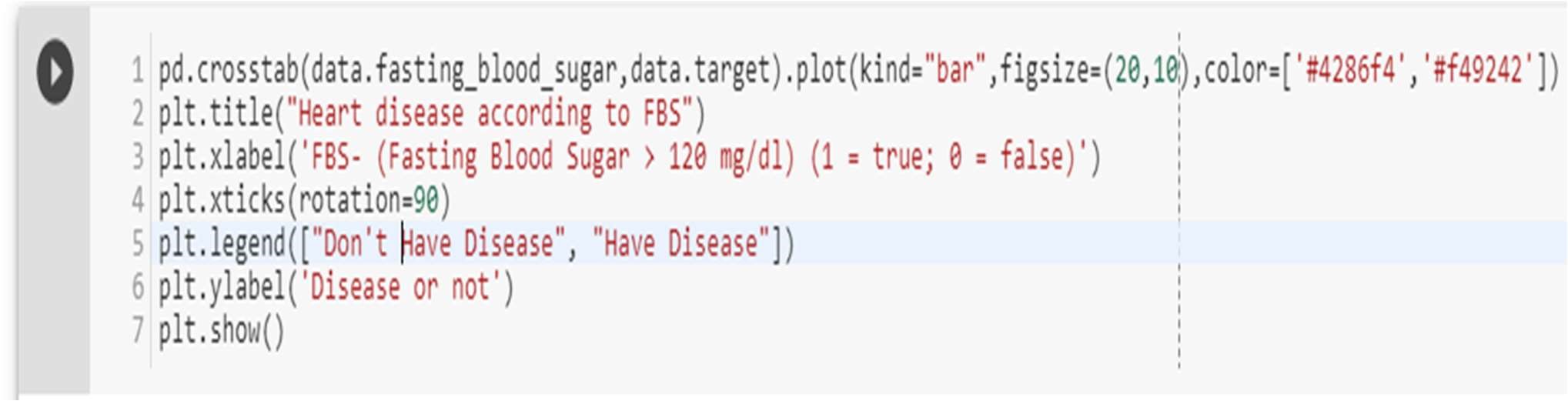


3.5.6 Making the data column names easily recognizable



3.5.7 Checking out Male/Female Heart disease according to Fasting Blood Sugar

Fasting blood sugar test: A blood sample will be taken after an overnight fast. A fasting blood sugar level less than 100 mg/dL (5.6 mmol/L) is normal. A fasting blood sugar level from 100 to 125 mg/dL (5.6 to 6.9 mmol/L) is considered prediabetes. If it's 126 mg/dL (7 mmol/L) or higher on two separate tests, you have diabetes[125].



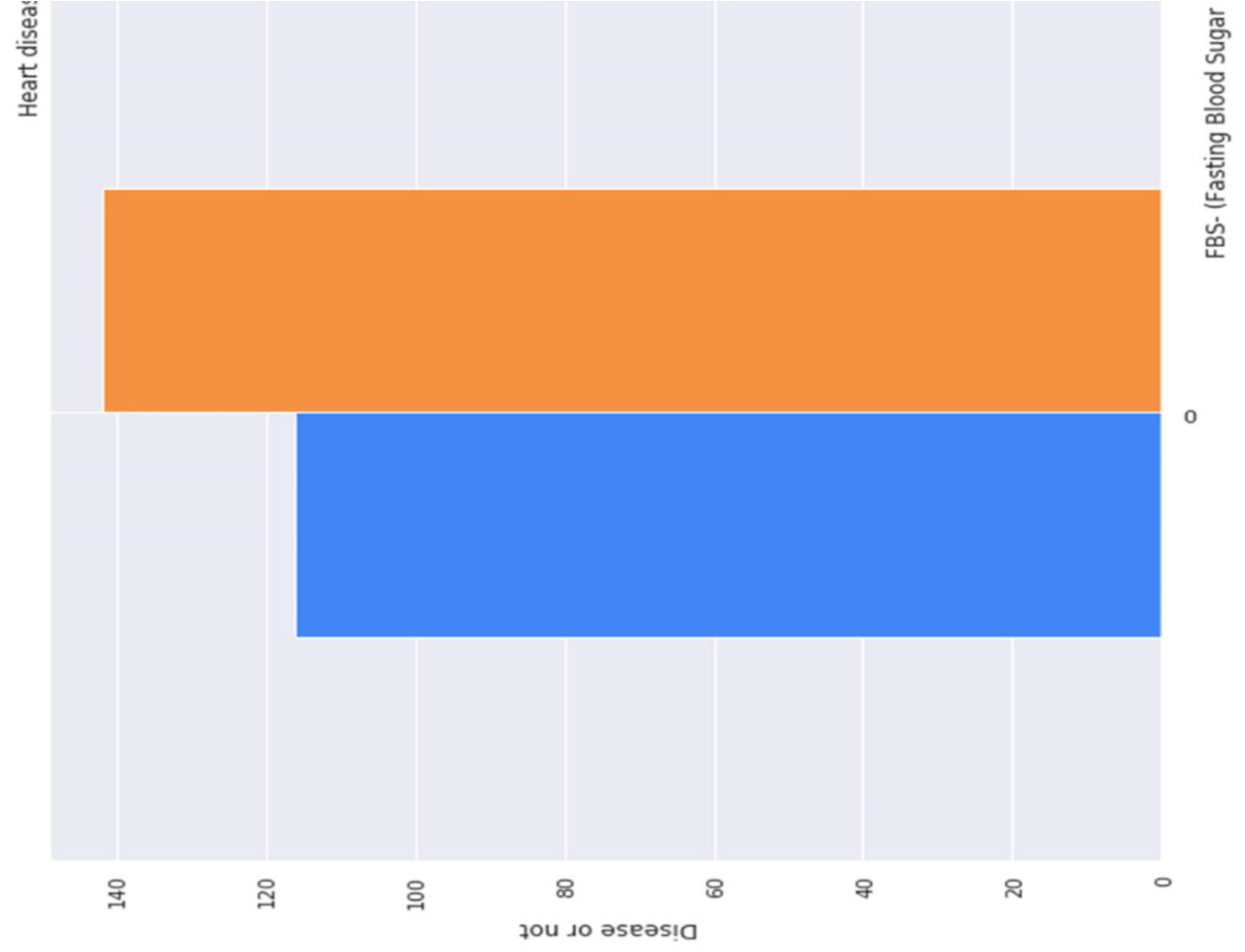
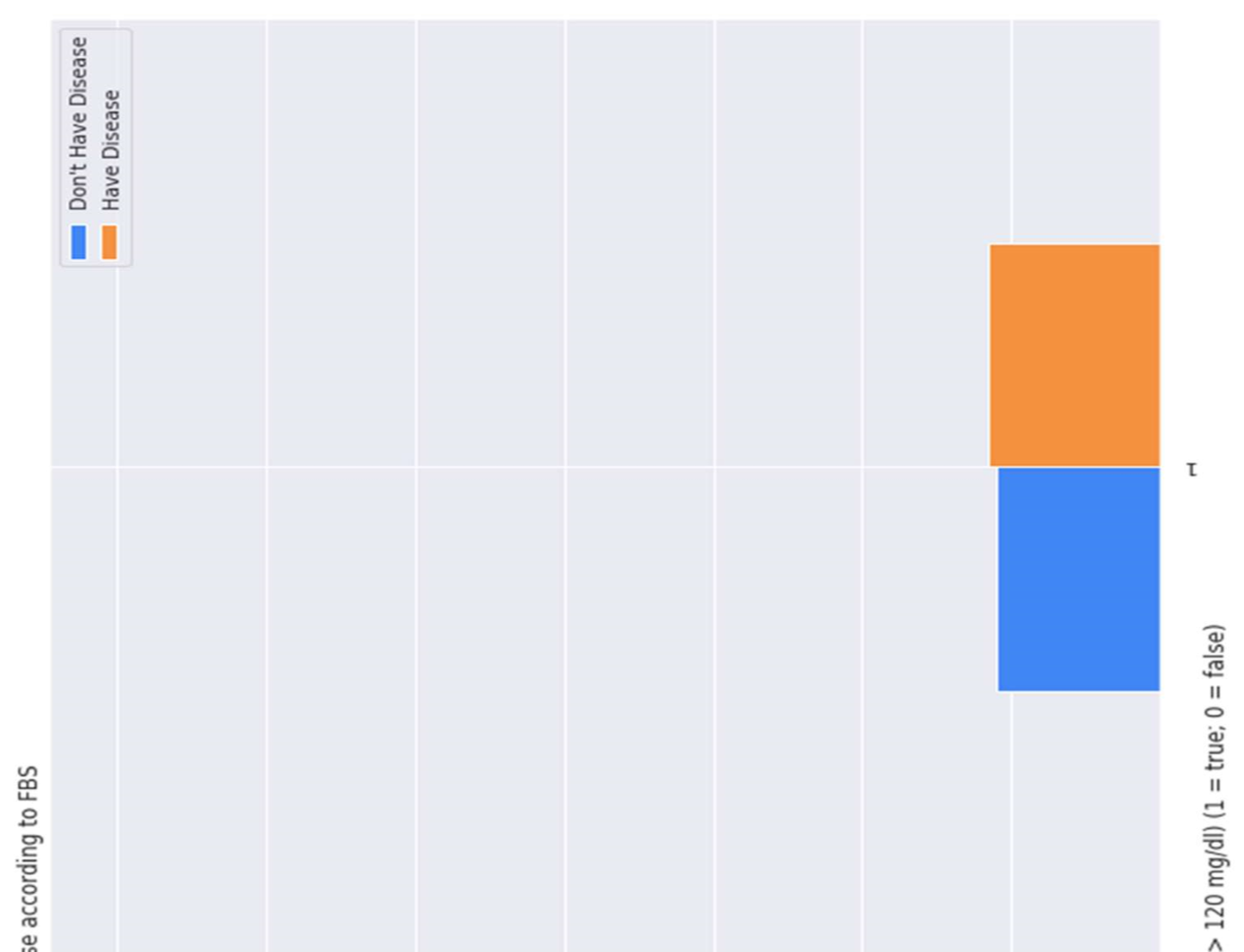


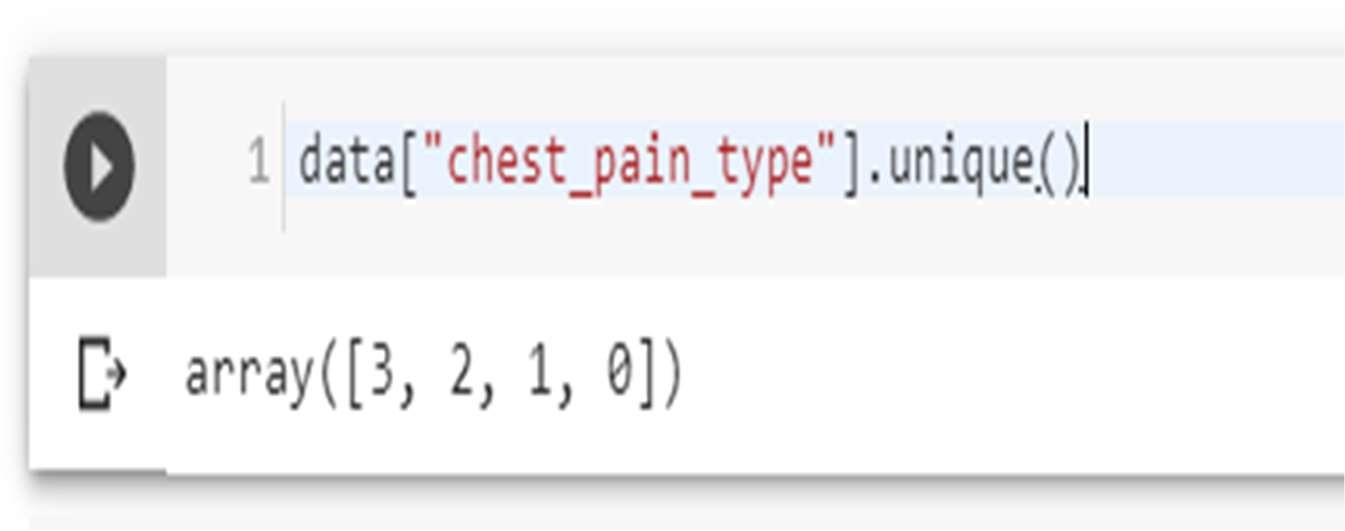
Figure 16 Heart Disease according to FBS

3.5.8 Analyzing the chest pain

There are four types of Angina(chest pain)[126].

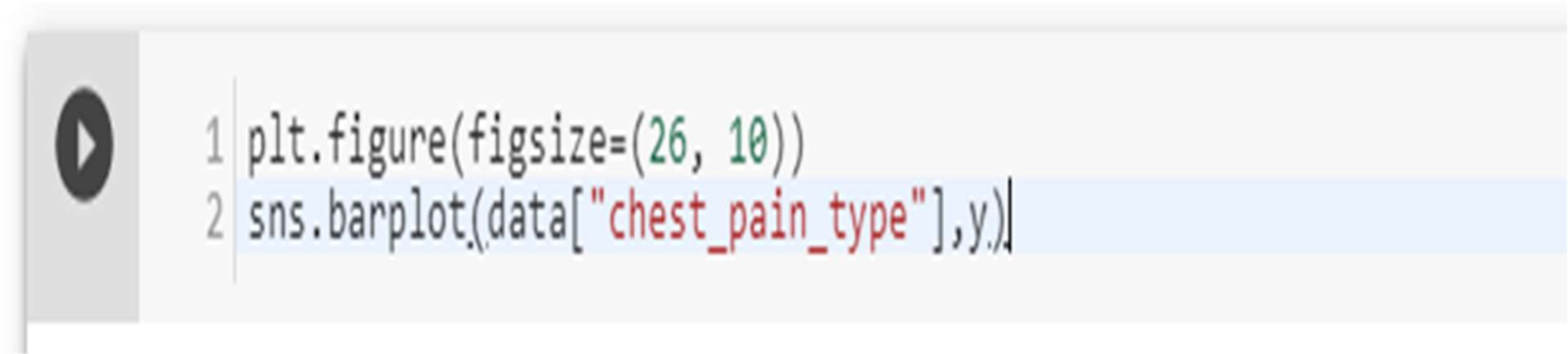
[Value 1: typical angina[122], Value 2: atypical angina[127], Value 3: non-anginal pain[123], Value 4: asymptomatic[128] ].

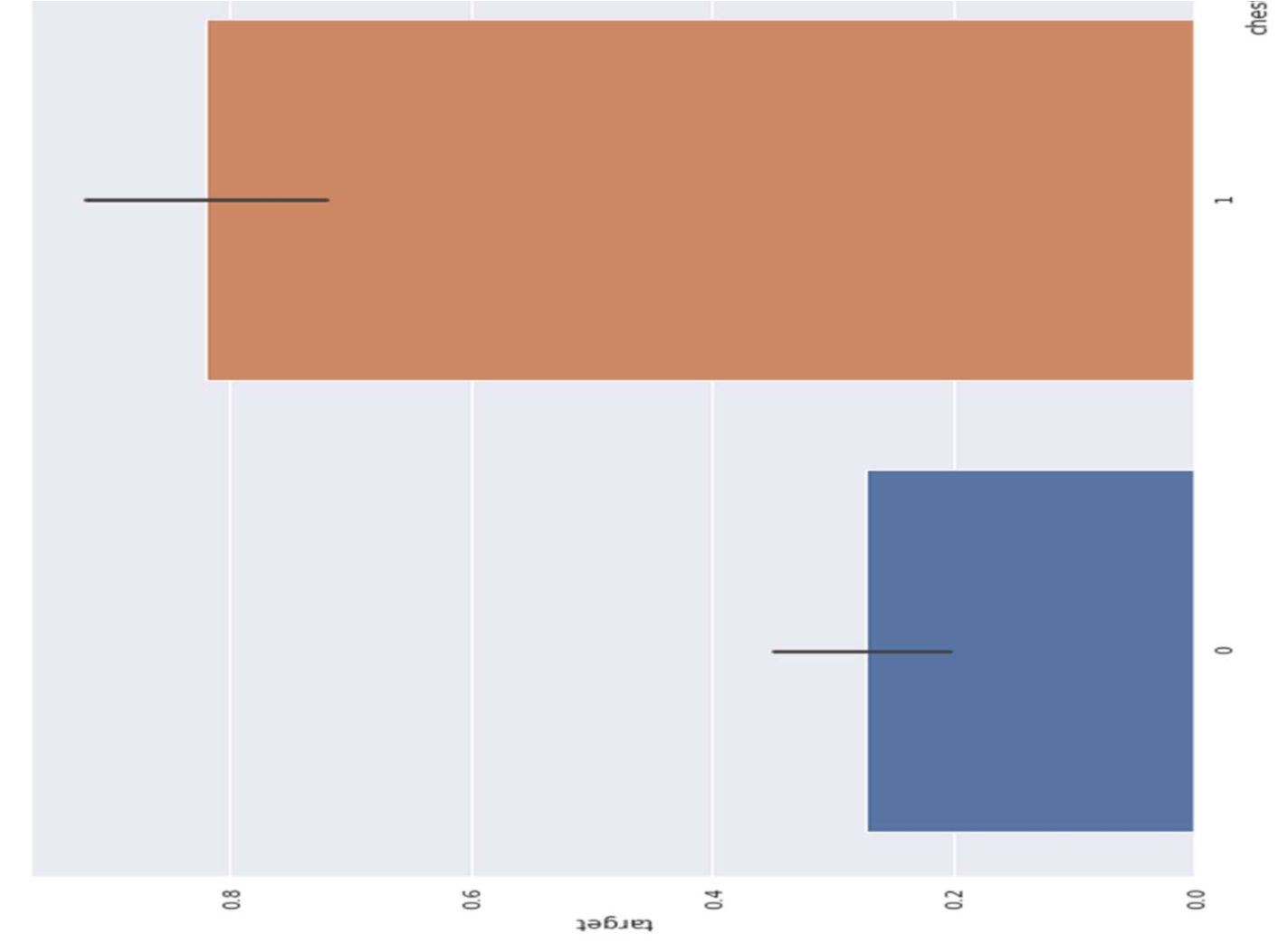
Let’s check how many types of chest pain (Angina) is present in our dataset



4 types are present, 0, 1, 2, 3

Let’s plot chest pain types against target

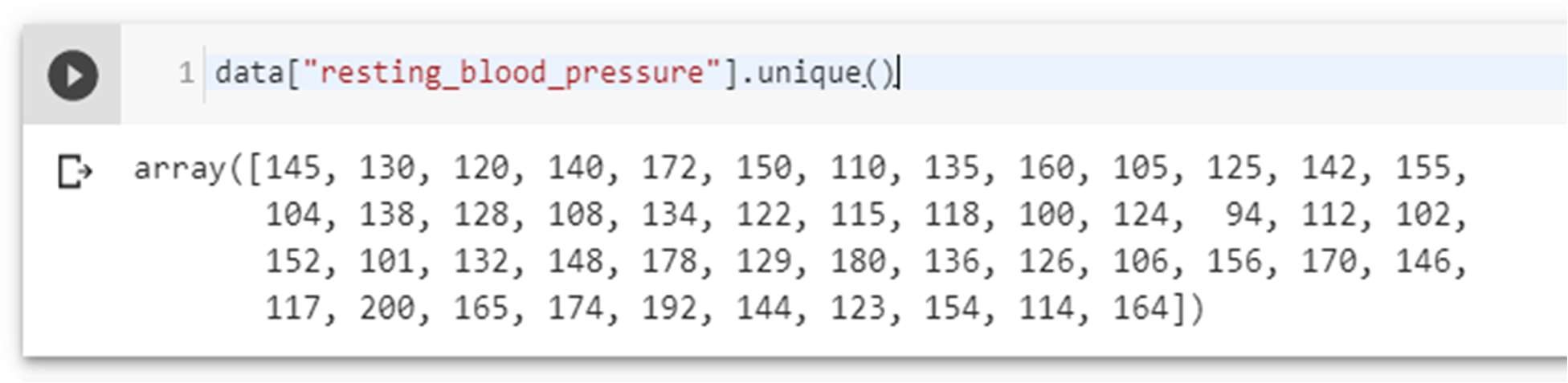




3.5.9 Analyzing the resting blood pressure

Resting blood pressure in mm Hg on admission to the hospital [129][124].

Let’s check out the unique resting blood pressures in our dataset



Let’s plot the resting blood pressure of our dataset against target



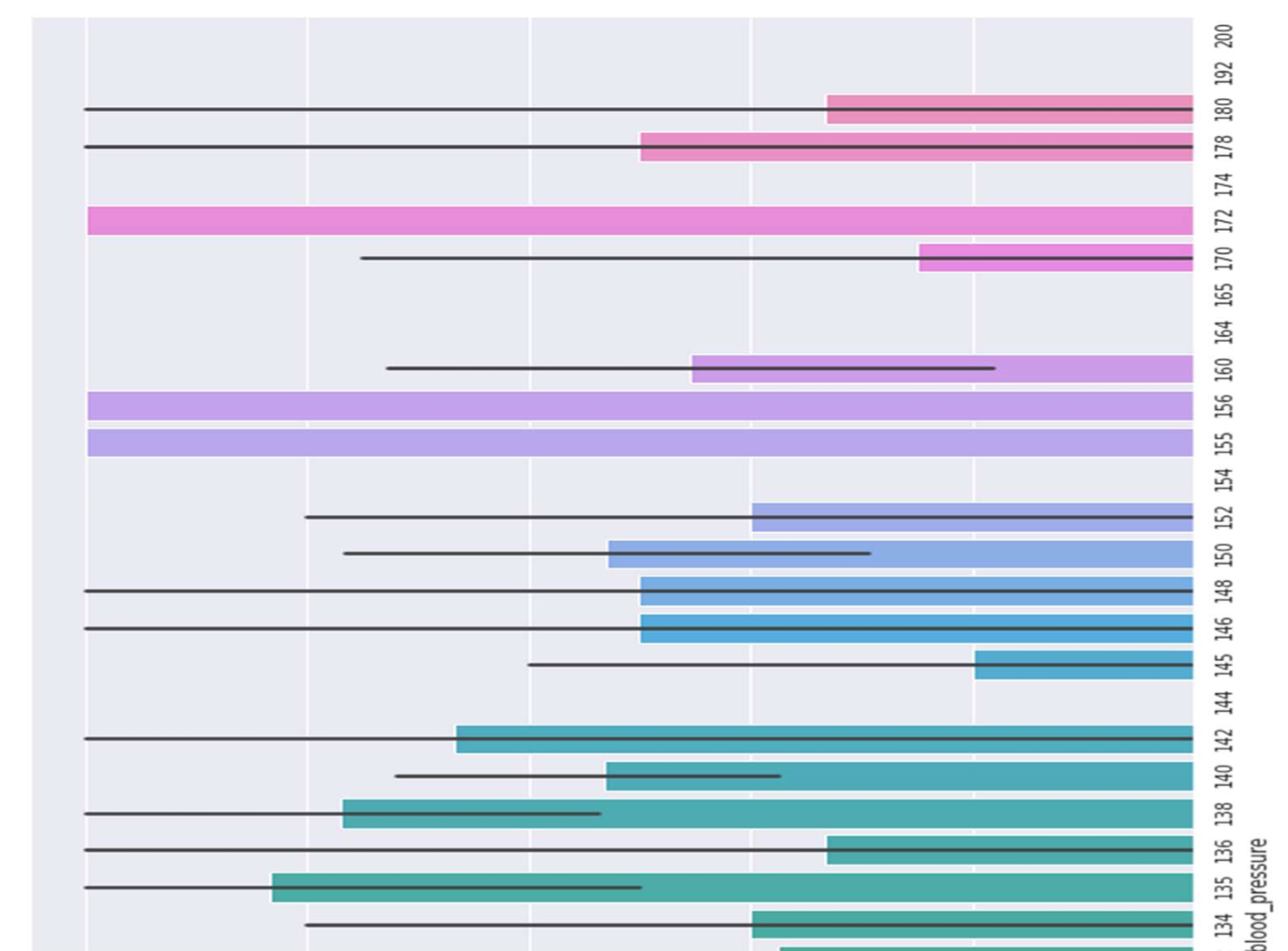
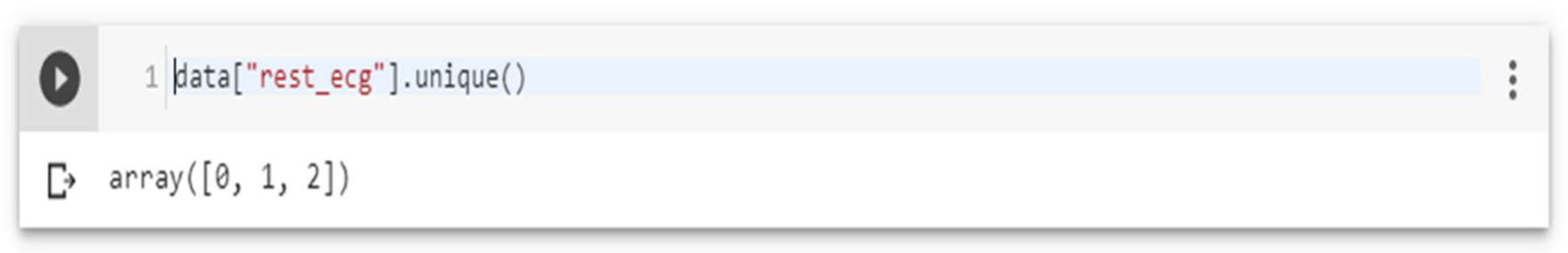


Figure 18 Resting Blood Pressure Analysis

3.5.10 Analyzing the resting electrocardiographic measurement

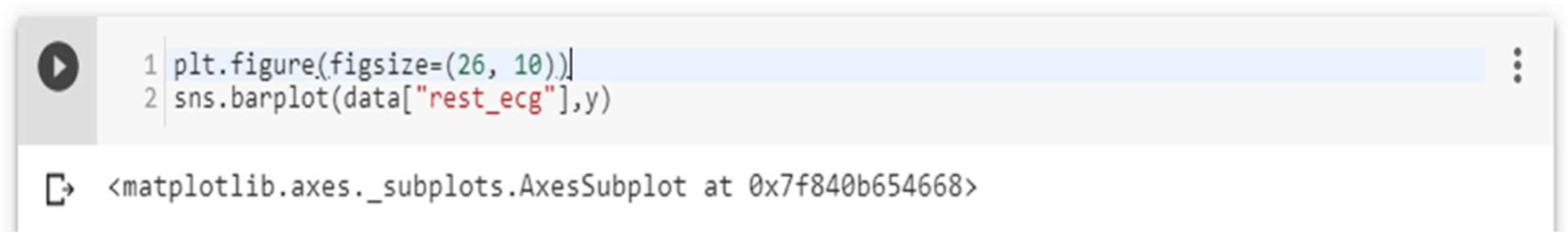
0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria[130][131][132].

Let’s check uniqueness of our dataset



There are 3 types of resting ECG values are present: 0, 1, 2.

Let’s then plot the ECG values against target column



People with resting ECG value: 1 and 0 are much likely to have a heart disease than with the value 2 of resting ECG.

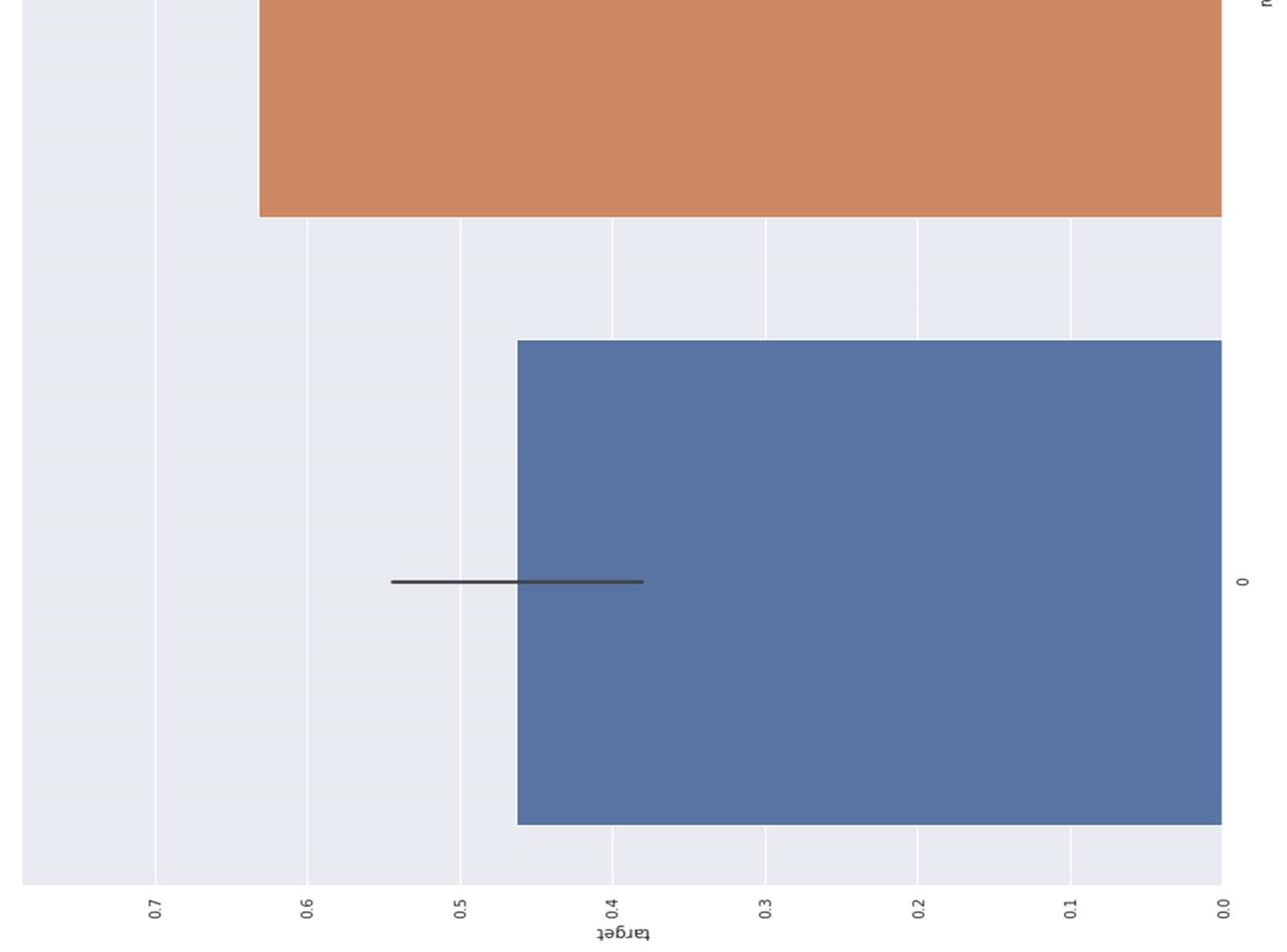
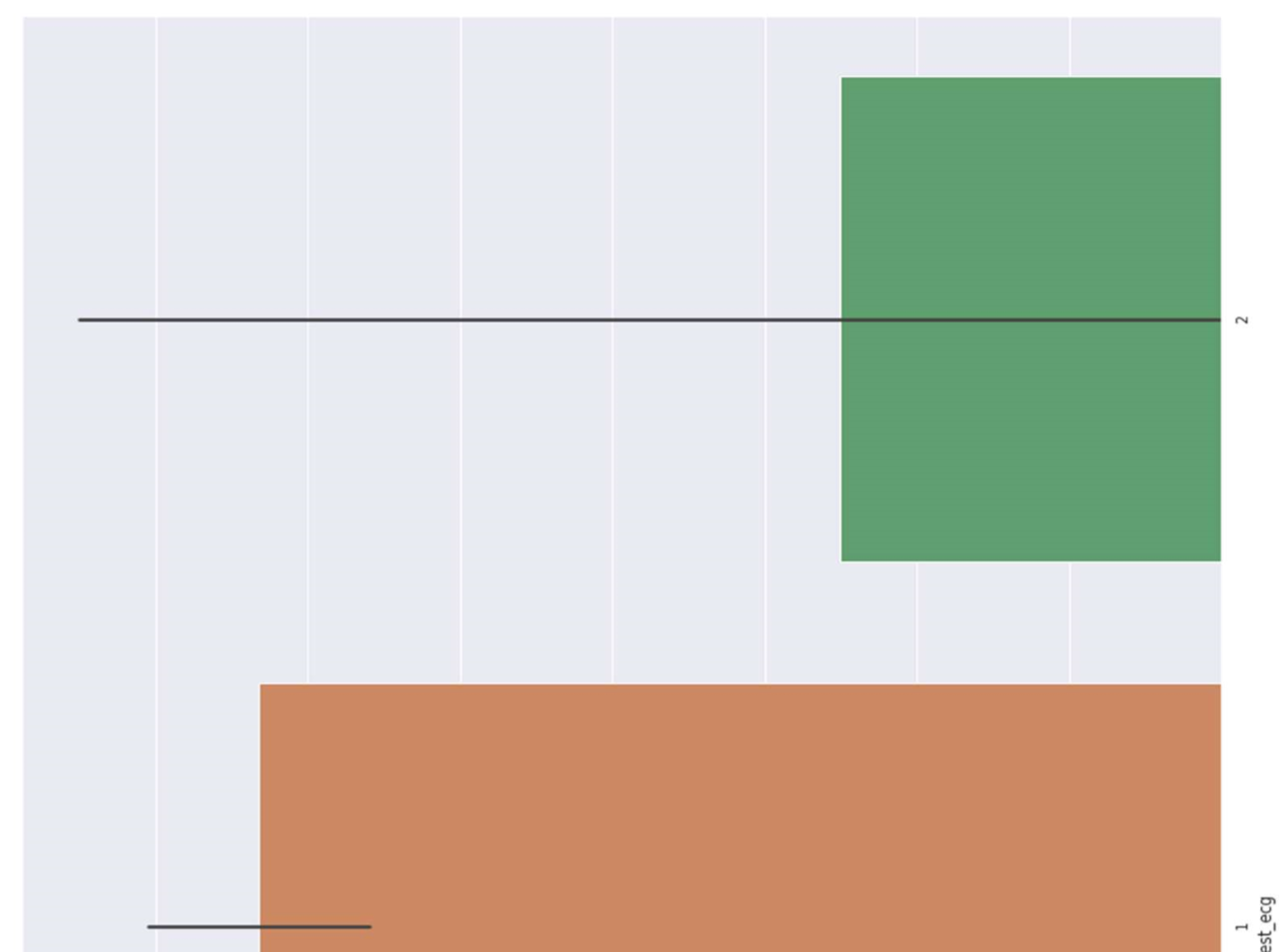


Figure 19 ECG analysis

3.5.11 Analyzing Exercise Induced angina 1 means yes, and 0 means no.

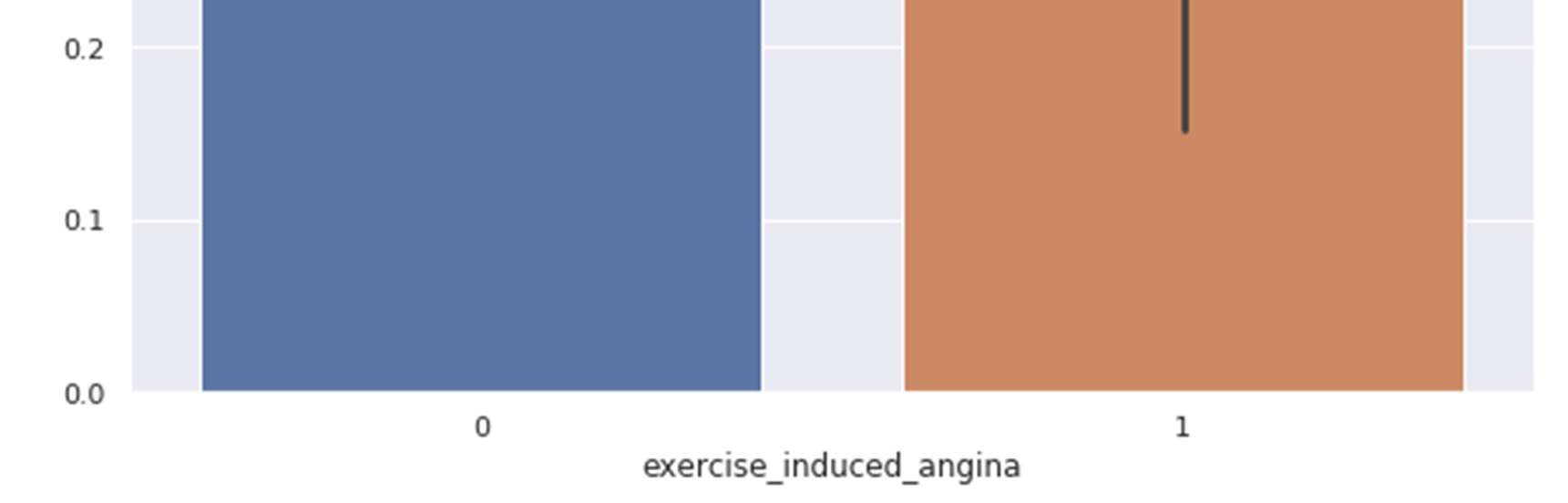
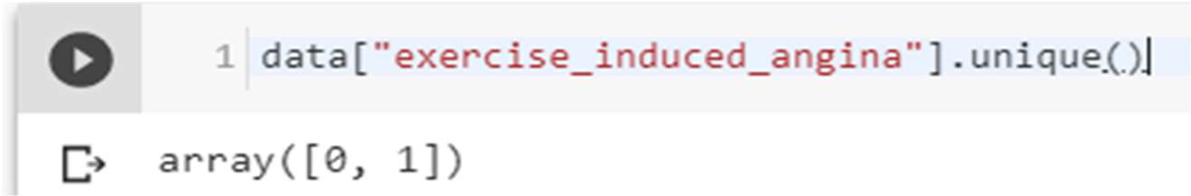
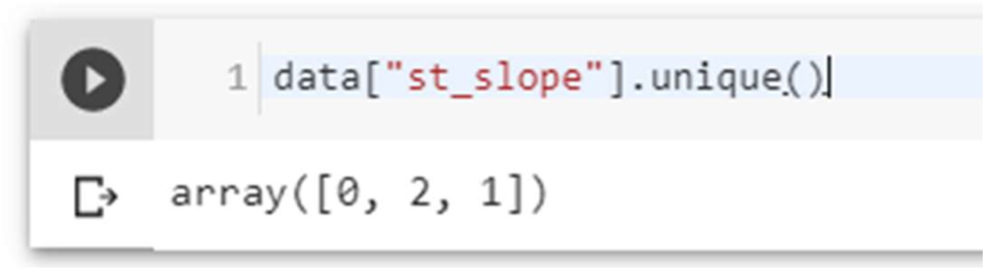


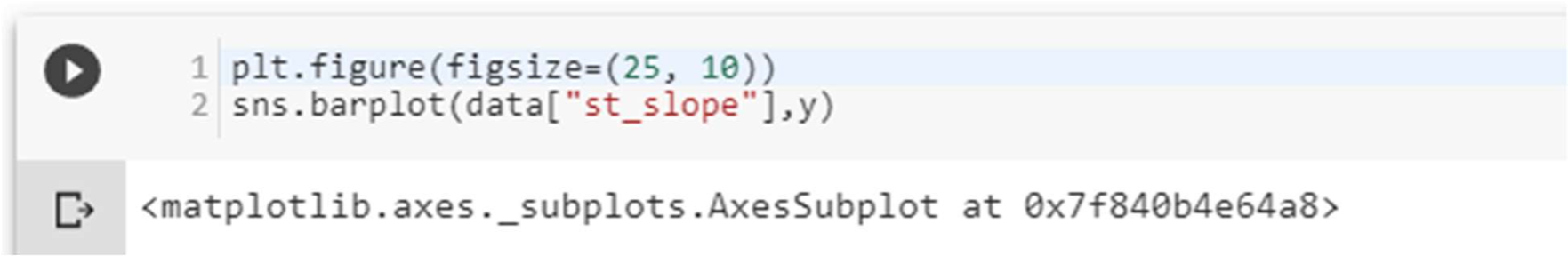
Figure 20 Exercised Induced Angina Analysis

3.5.12 Slope of the peak exercise ST segment

The treadmill electrocardiogram (ECG) stress test is widely used to screen for obstructive coronary artery disease (CAD). The presence of ST segment changes, either depression or elevation, on the ECG during the treadmill test often suggests presence of CAD and warrants further management. We herein present three cases, with evidence of ischemia on the treadmill ECG stress test. In addition, we discuss the use of the treadmill ECG stress test, including its indications, contraindications, reasons for termination and interpretation of the ST-segment changes, heart rate, as well as blood pressure responses to exercise[133].



Value 1: upsloping, Value 2: flat, Value 3: down sloping.



Slope '2' causes heart pain much more than Slope '0' and '1'.

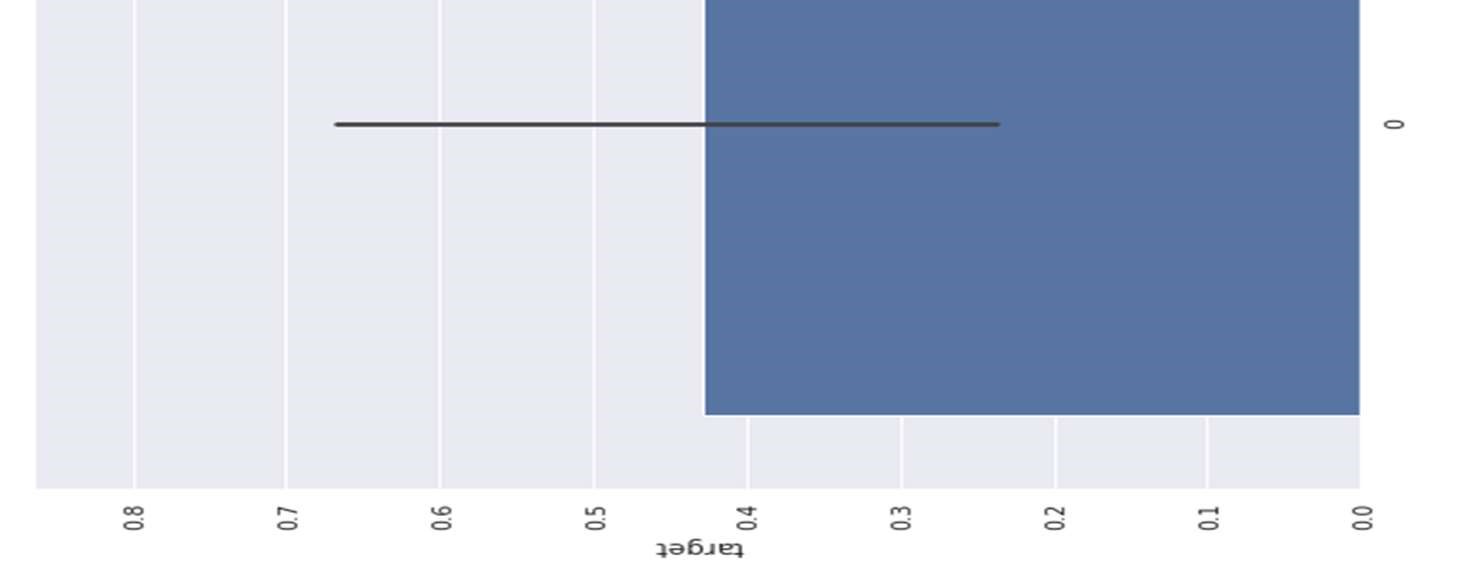
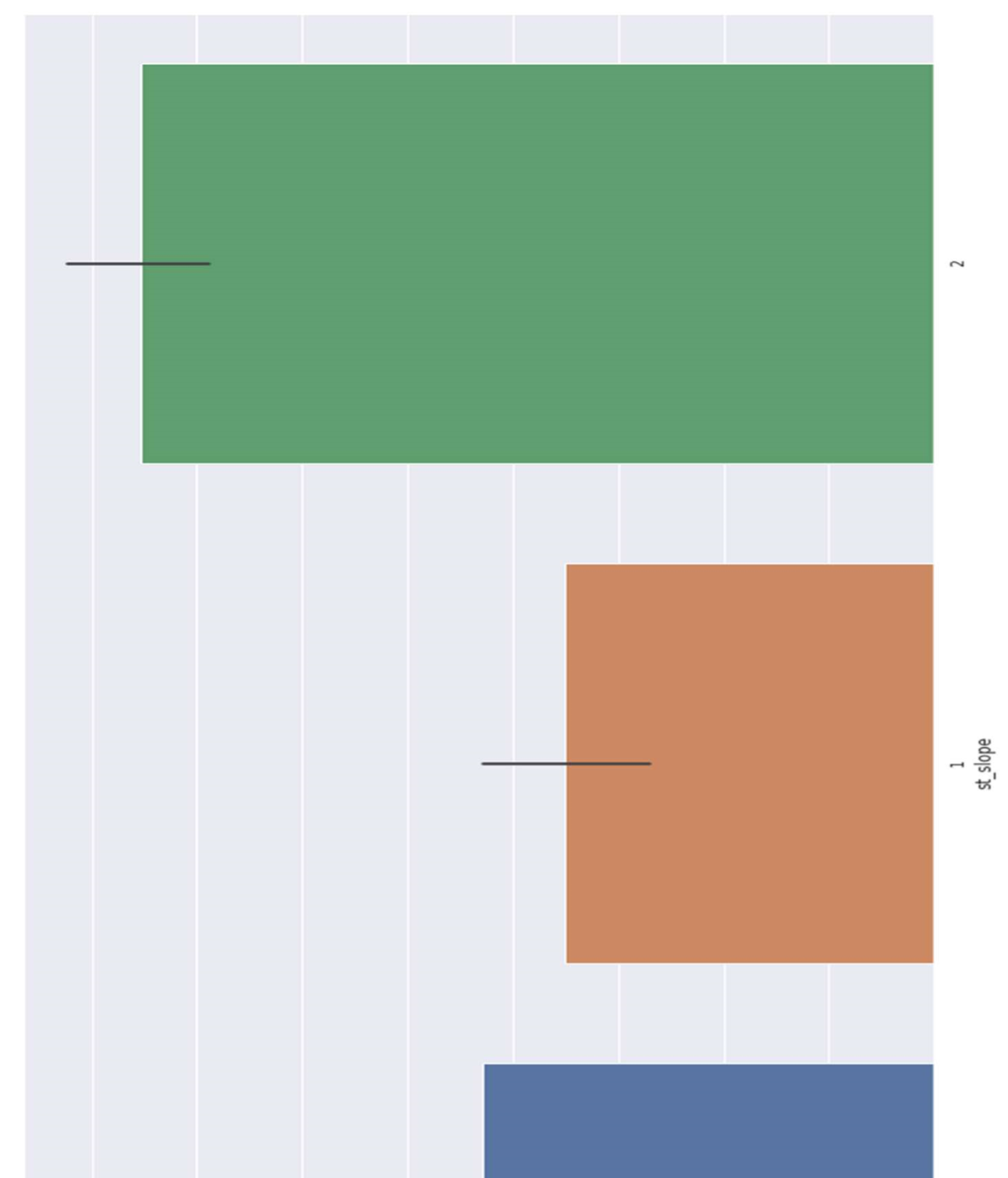


Figure 21 ST Slope Analysis

3.5.13

Analyzing no. of major vessels colored by fluoroscopy

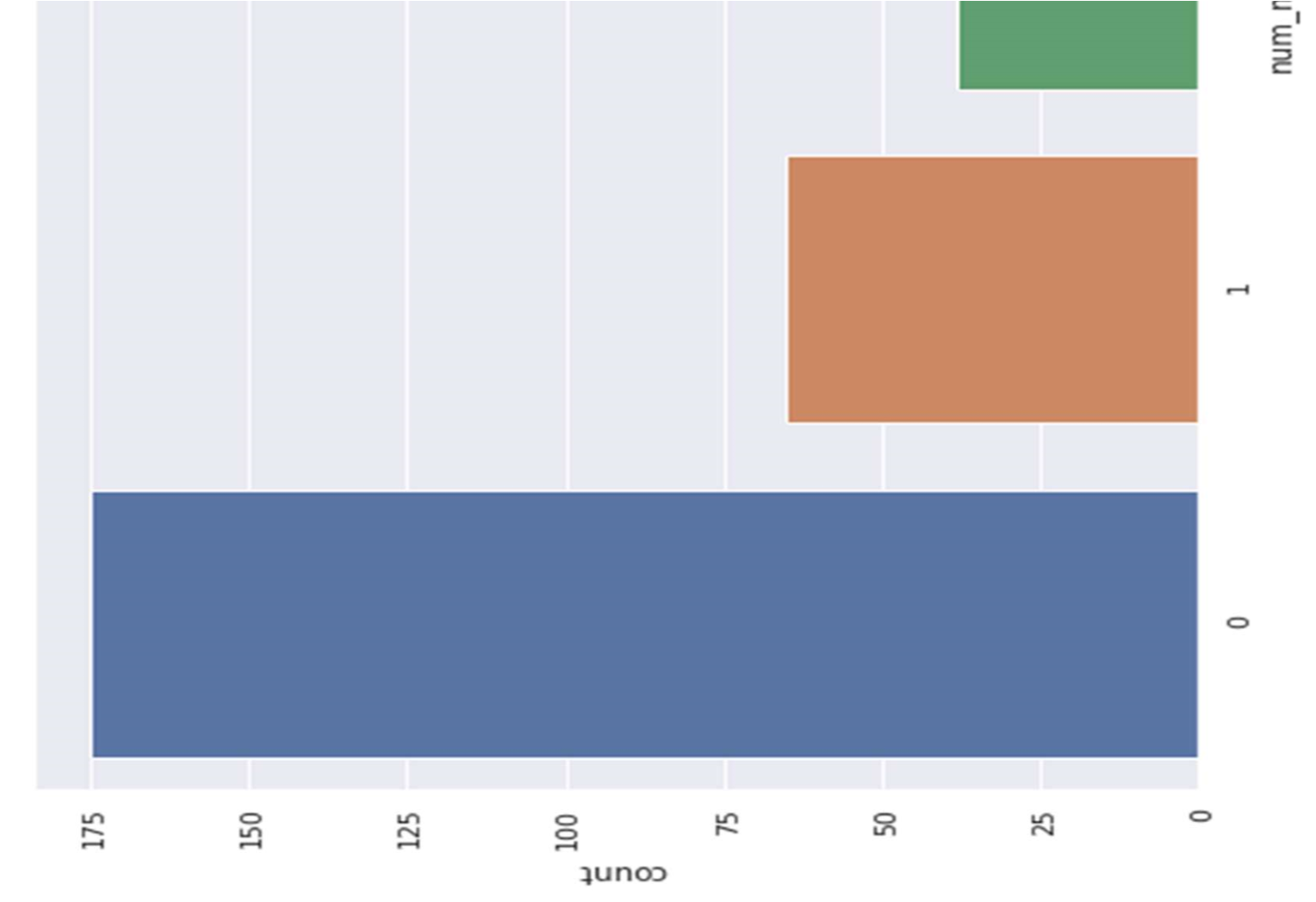
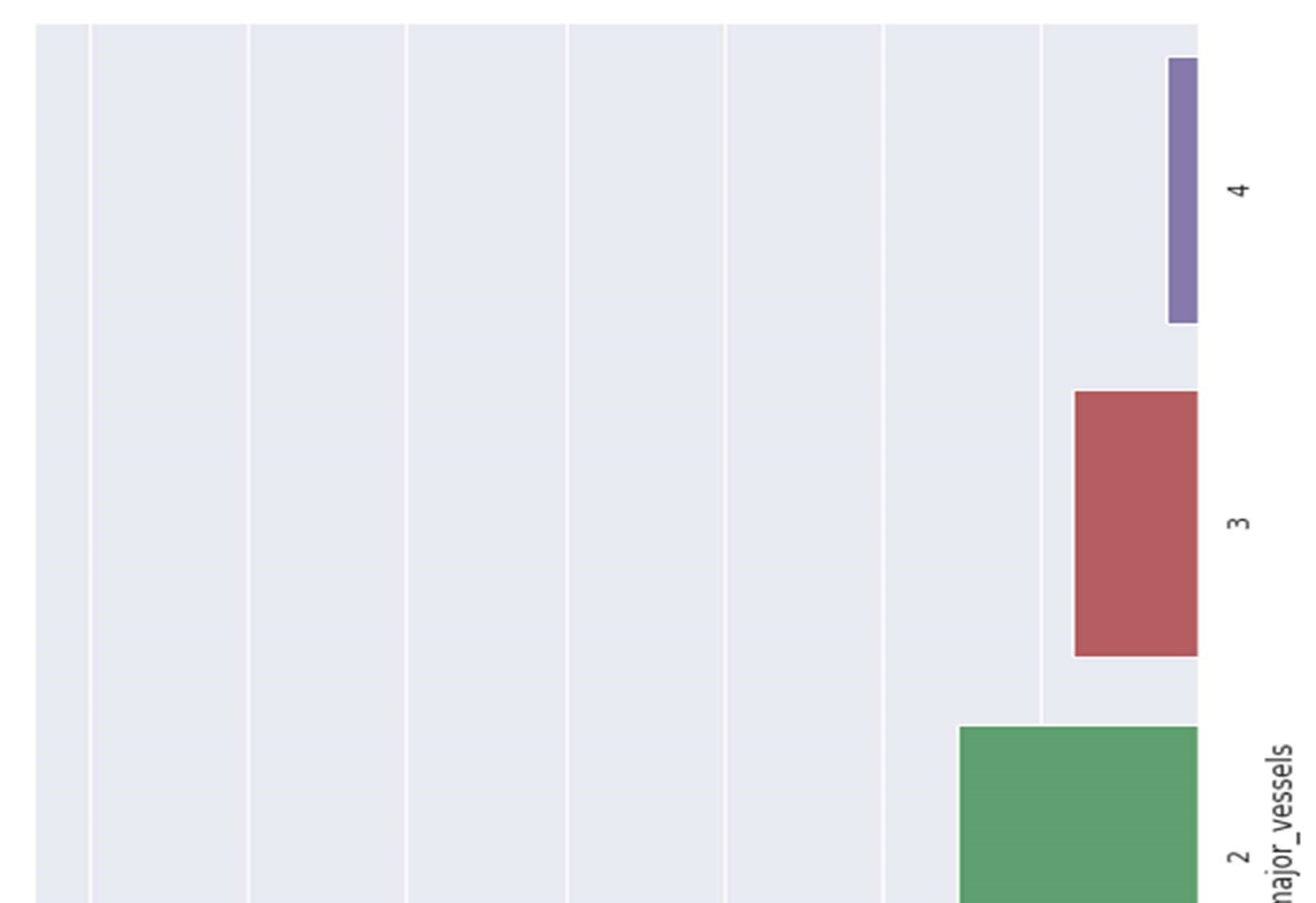


Figure 22 Analyzing no. of major vessels colored by fluoroscopy

3.5.13.1

Comparing with target:

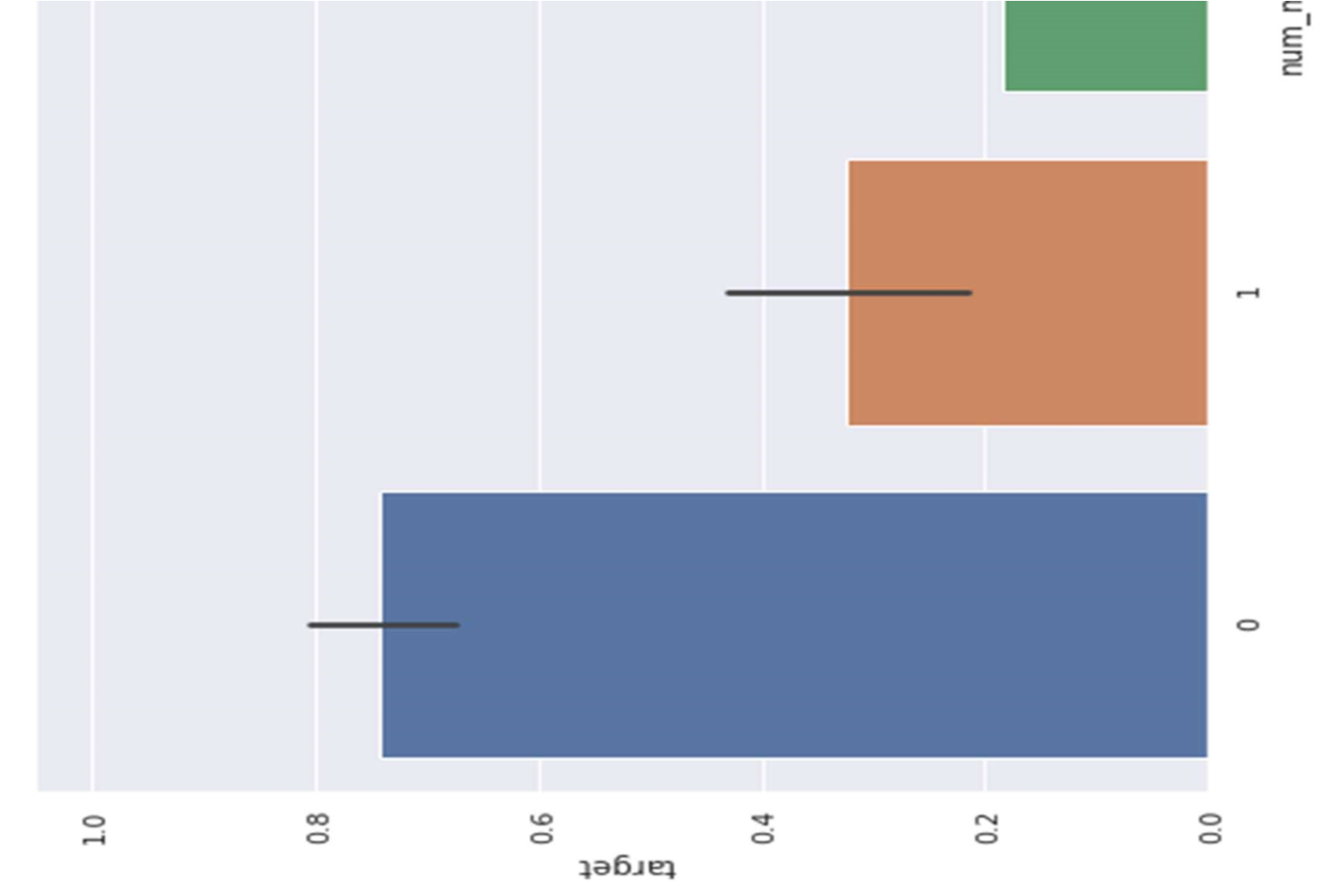
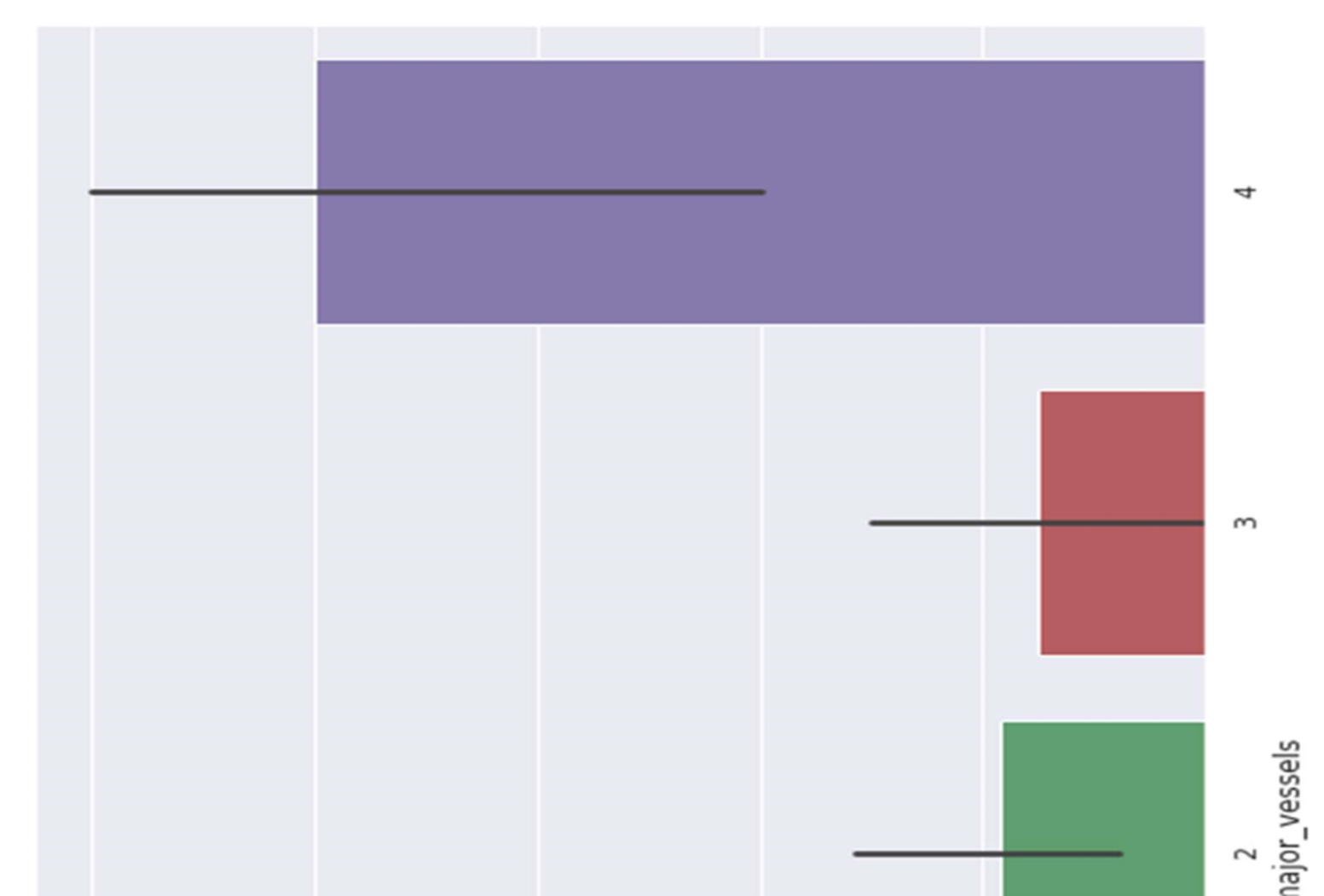


Figure 23 Comparing with targets

3.5.14 Analyzing thalassemia

Four alpha-globin and two beta-globin protein chains make up hemoglobin. The two main types of thalassemia are alpha and beta.

Alpha thalassemia - In alpha thalassemia, the hemoglobin does not produce enough alpha protein. To make alpha-globin protein chains we need four genes, two on each chromosome 16. We get two from each parent. If one or more of these genes is missing, alpha thalassemia will result. The severity of thalassemia depends on how many genes are faulty, or mutated.

* One faulty gene: The patient has no symptoms. A healthy person who has a child with symptoms of thalassemia is a carrier. This type is known as alpha thalassemia minima.
* Two faulty genes: The patient has mild anemia. It is known as alpha thalassemia minor.
* Three faulty genes: The patient has hemoglobin H disease, a type of chronic anemia.

They will need regular blood transfusions throughout their life.

* Four faulty genes: Alpha thalassemia major is the most severe form of alpha thalassemia. It is known to cause hydrops fetalis, a serious condition in which fluid accumulates in parts of the fetus' body. A fetus with four mutated genes cannot produce normal hemoglobin and is unlikely to survive, even with blood transfusions.

Alpha thalassemia is common in southern China, Southeast Asia, India, the Middle East, and Africa.

Beta Thalassemia - We need two globin genes to make beta-globin chains, one from each parent. If one or both genes are faulty, beta thalassemia will occur.

Severity depends on how many genes are mutated.

* One faulty gene: This is called beta thalassemia minor.
* Two faulty genes: There may be moderate or severe symptoms. This is known as thalassemia major. It used to be called Colley's anemia.

Beta thalassemia is more common among people of Mediterranean ancestry. Prevalence is higher in North Africa, West Asia, and the Maldives Islands. So, we’ll mainly work with

Alpha Thalassemia, And value 0 is one faulty gene, Value 1 is Two faulty genes, Value 2 is Three faulty genes, Value 3 is Four faulty genes.

Thalassemia distribution

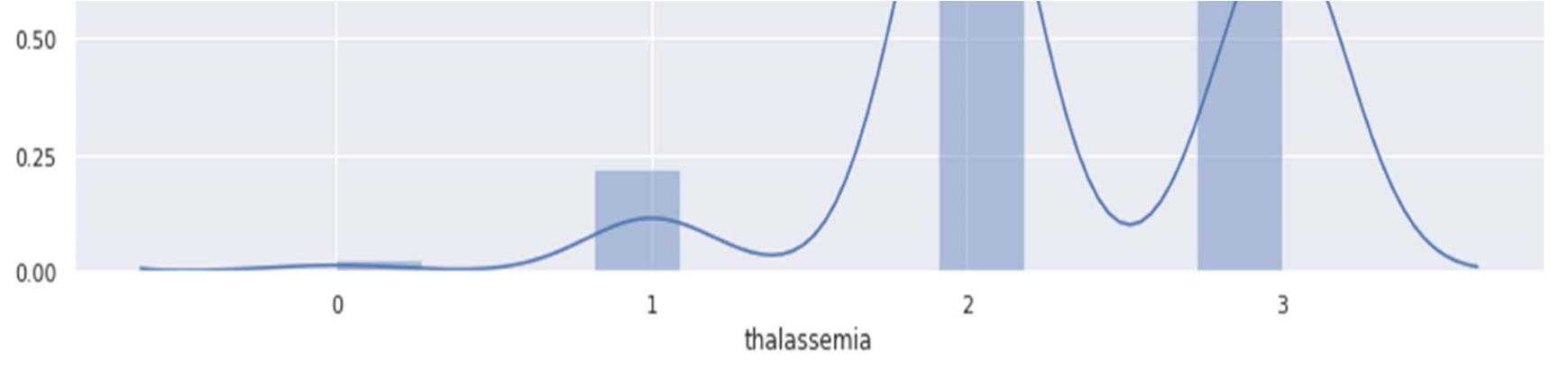
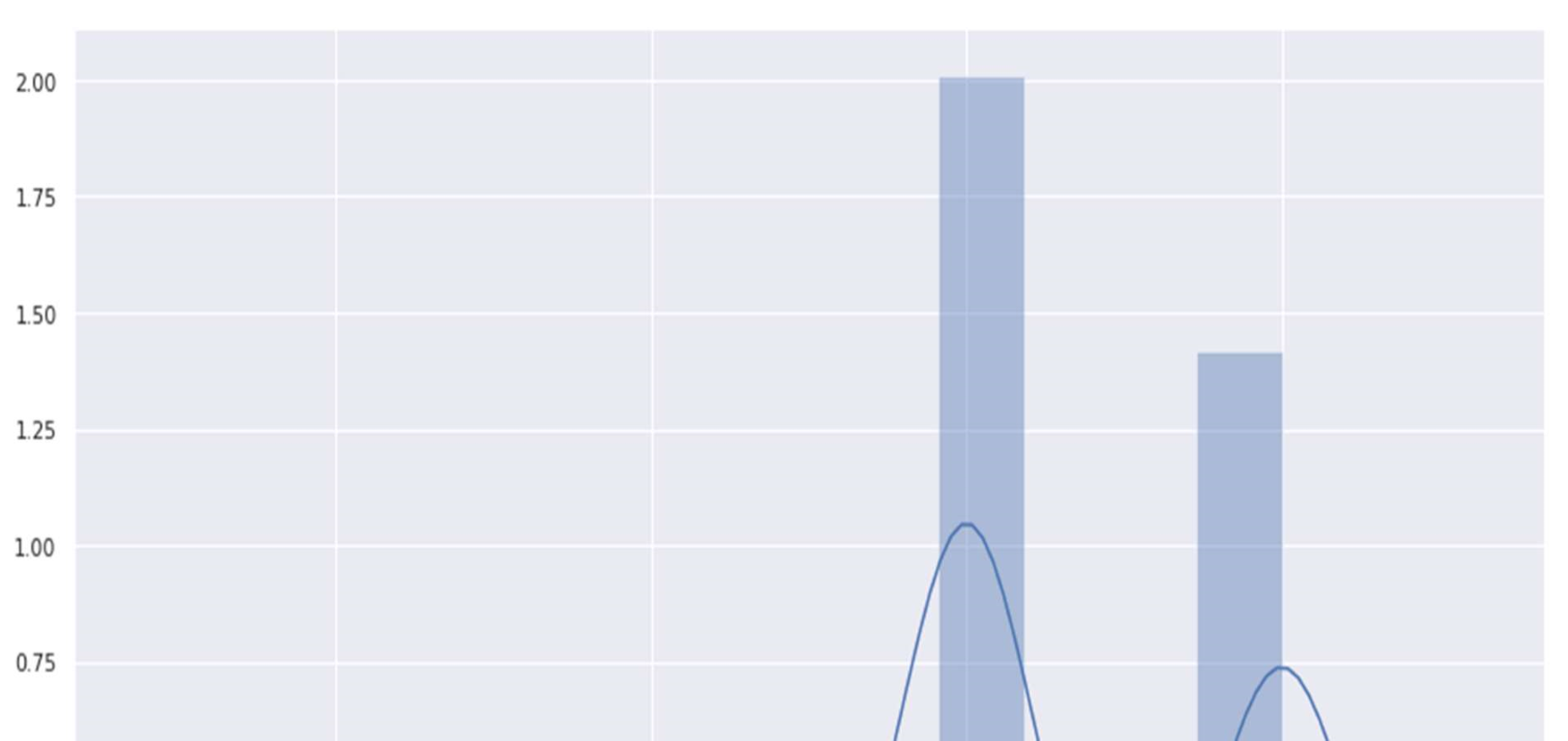


Figure 24 Thalassemia distribution

Against target:

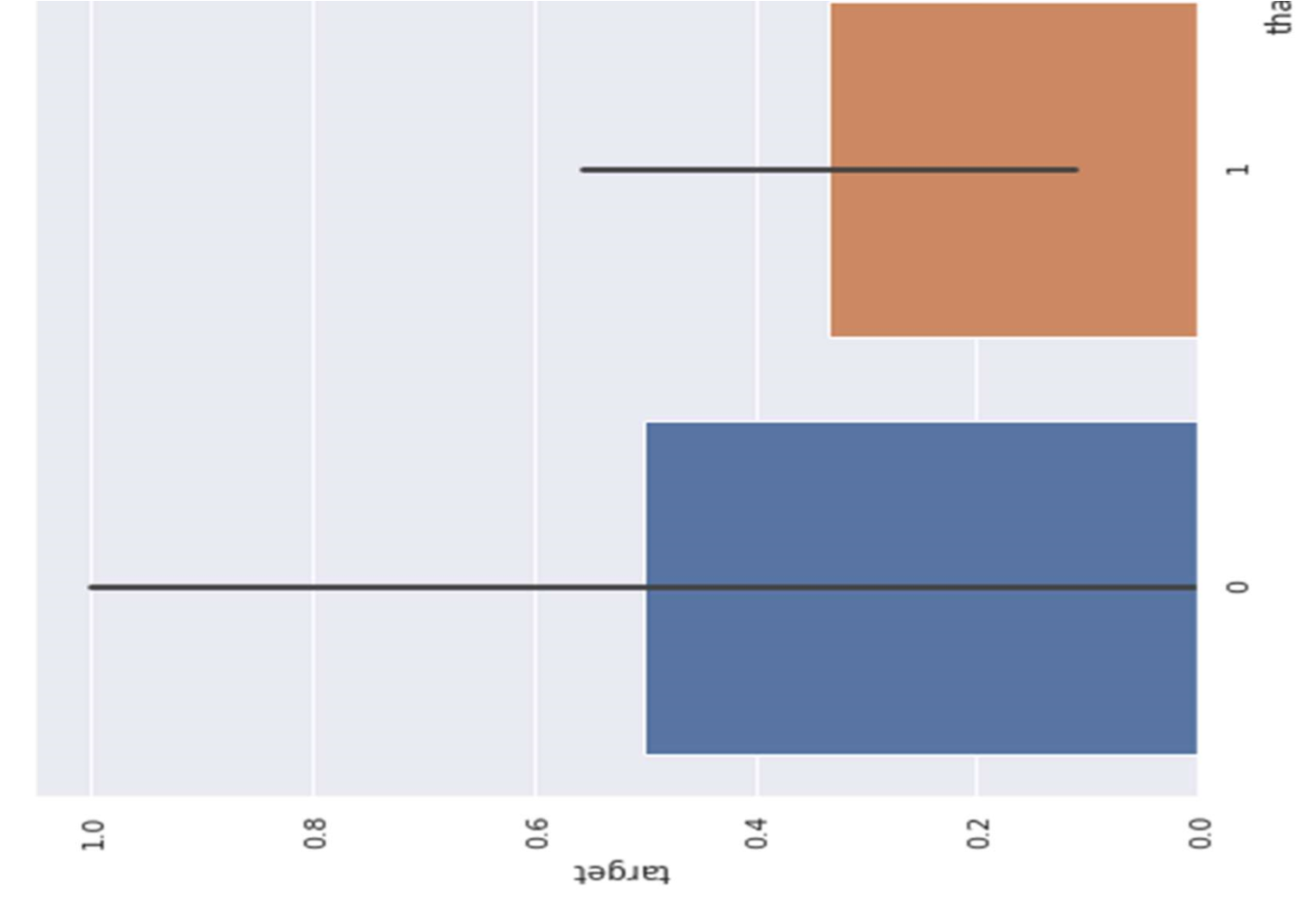
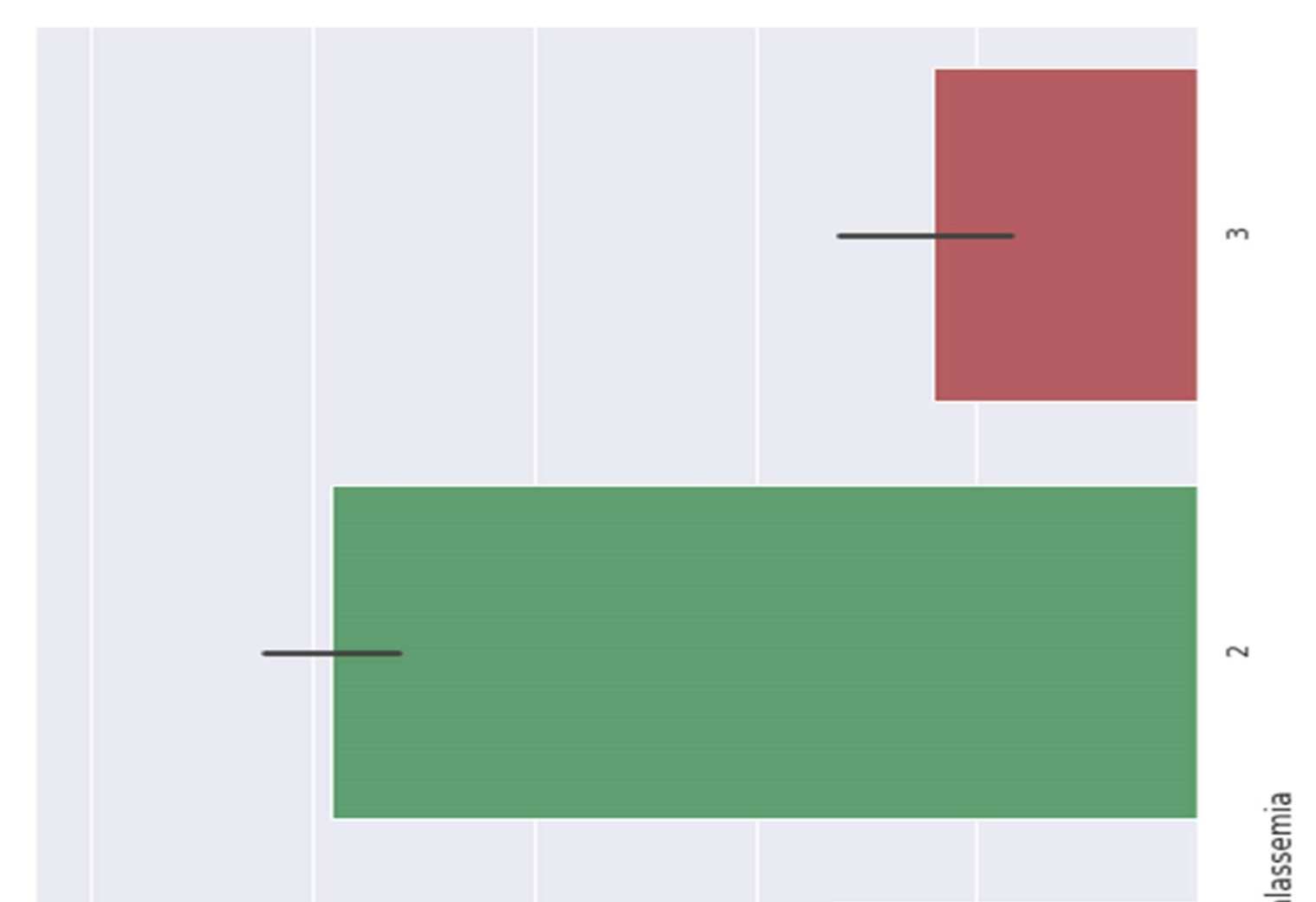


Figure 25 Thalassemia against target

3.5.15

Thalassemia vs cholesterol

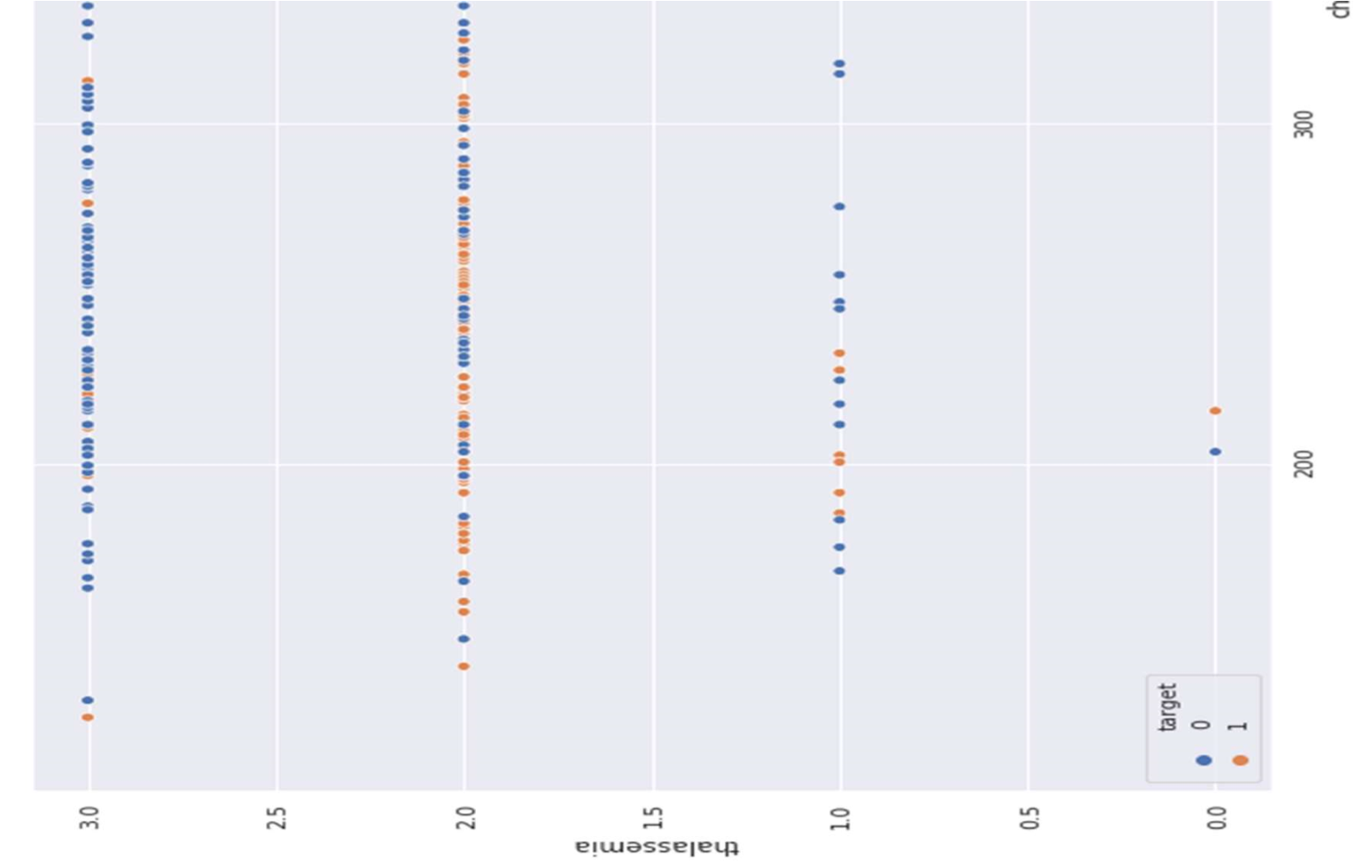
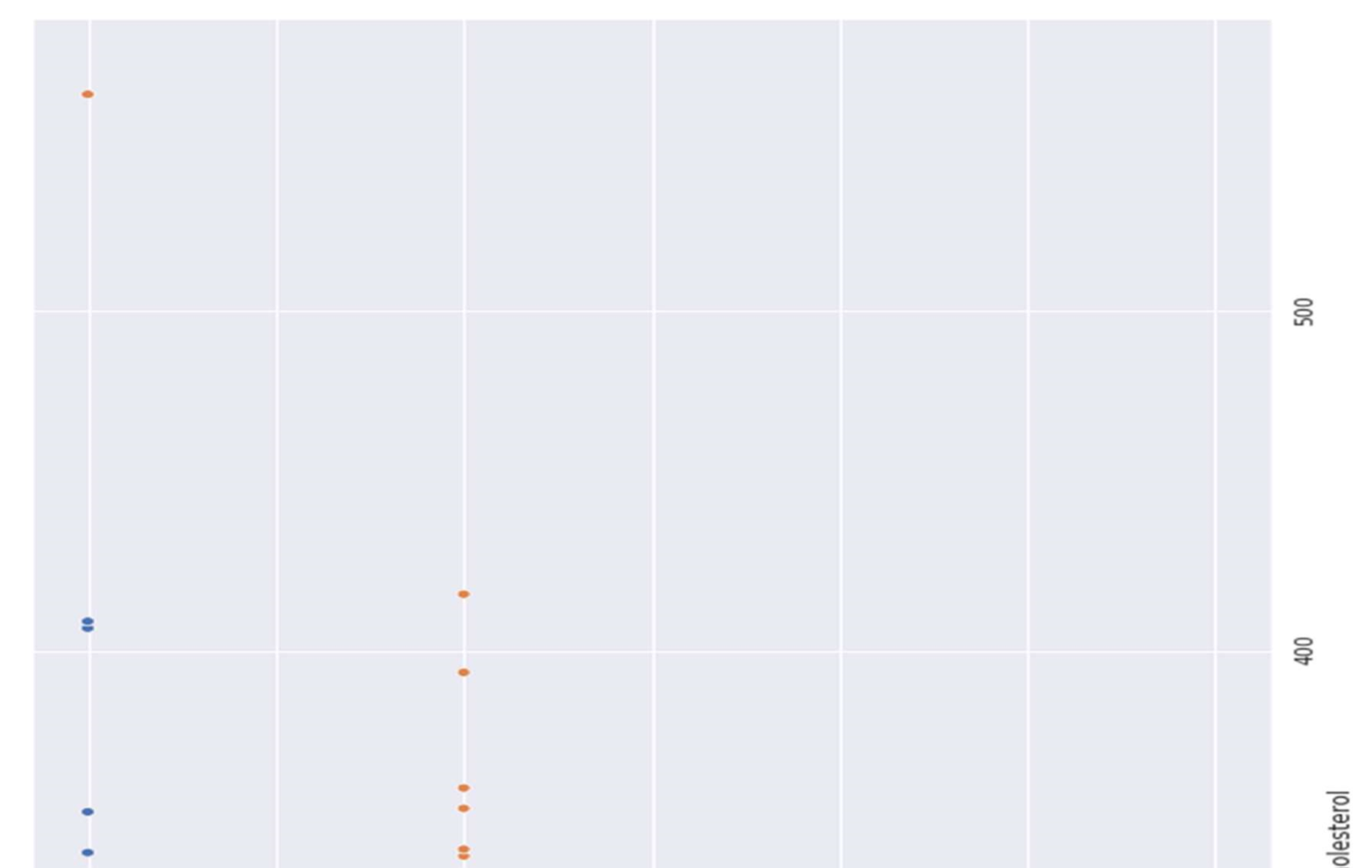


Figure 26 Thalassemia vs cholesterol

Thalassemia vs resting blood pressure

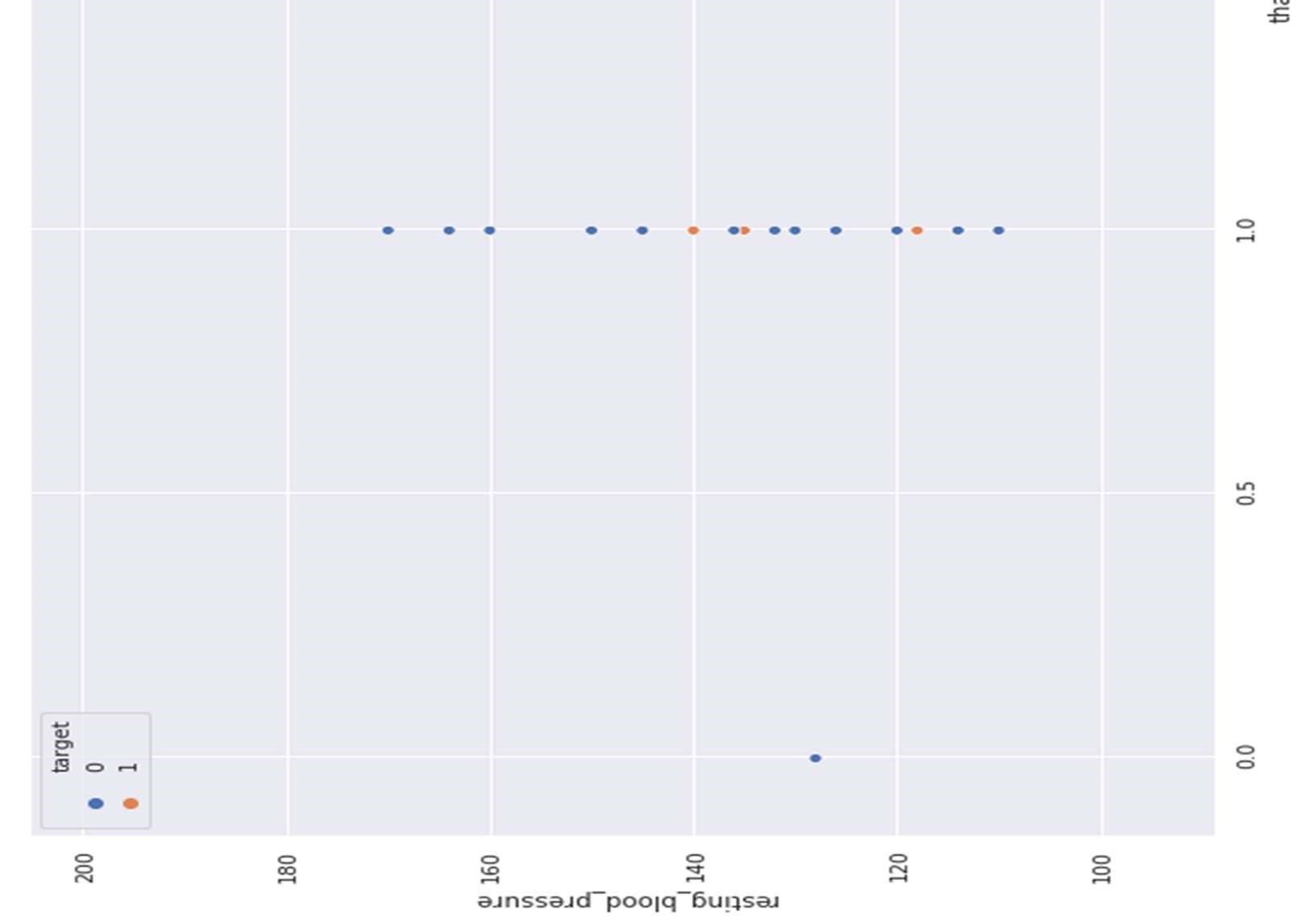
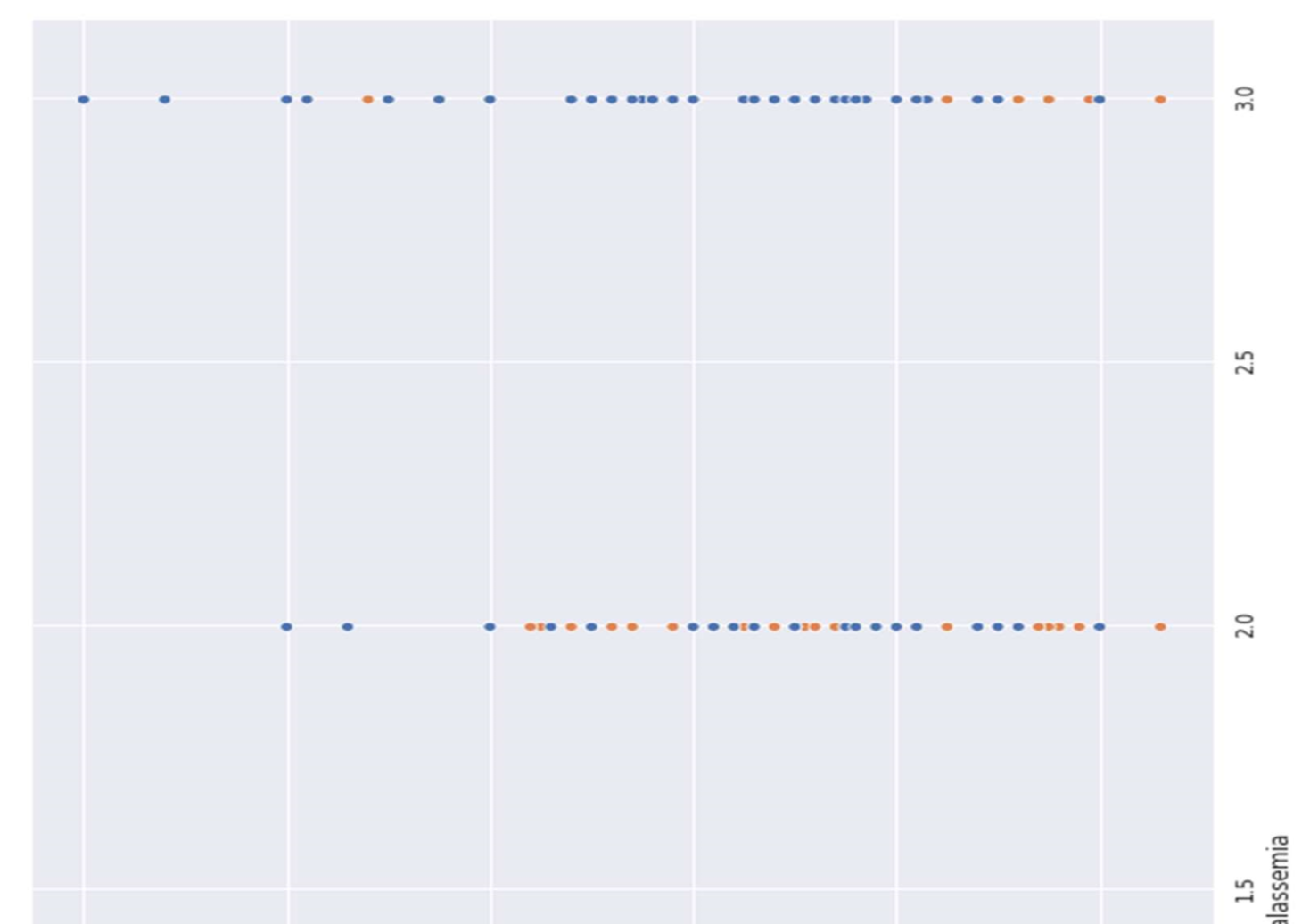


Figure 27 Thalassemia vs resting blood pressure