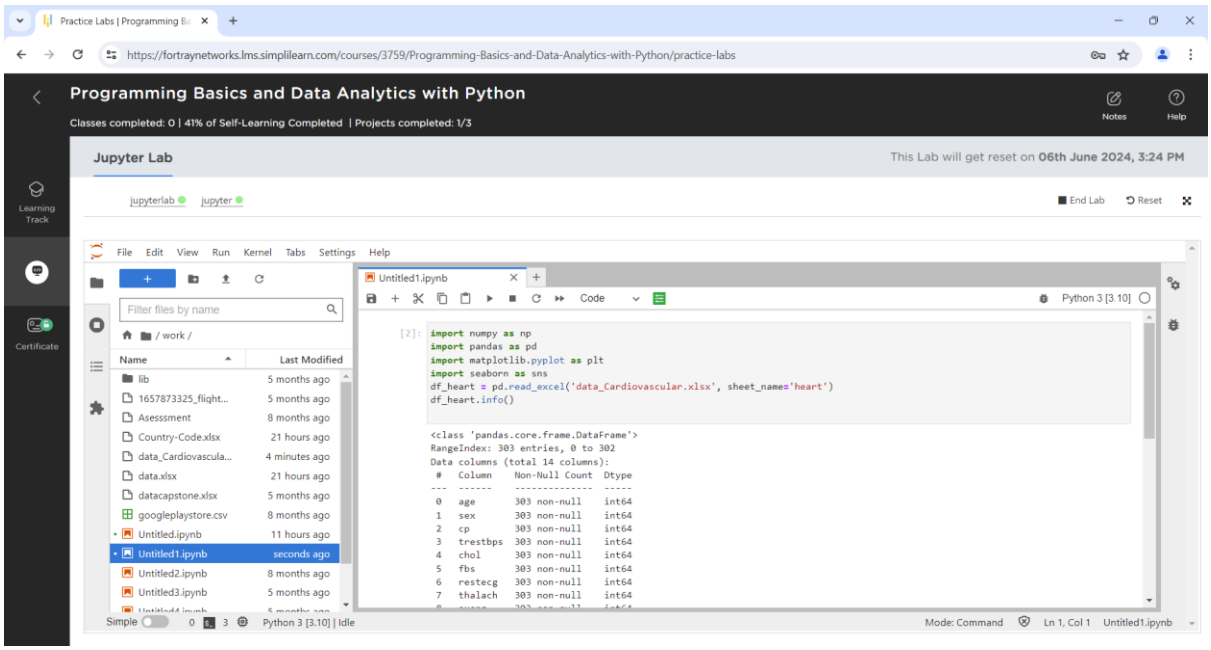
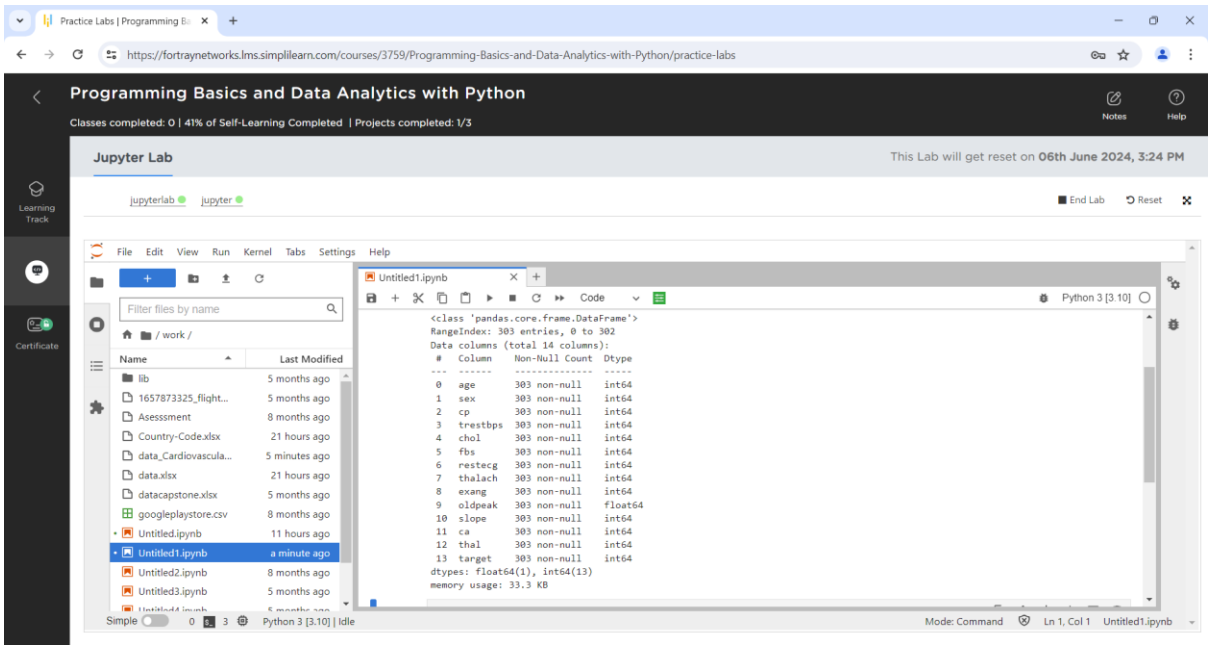
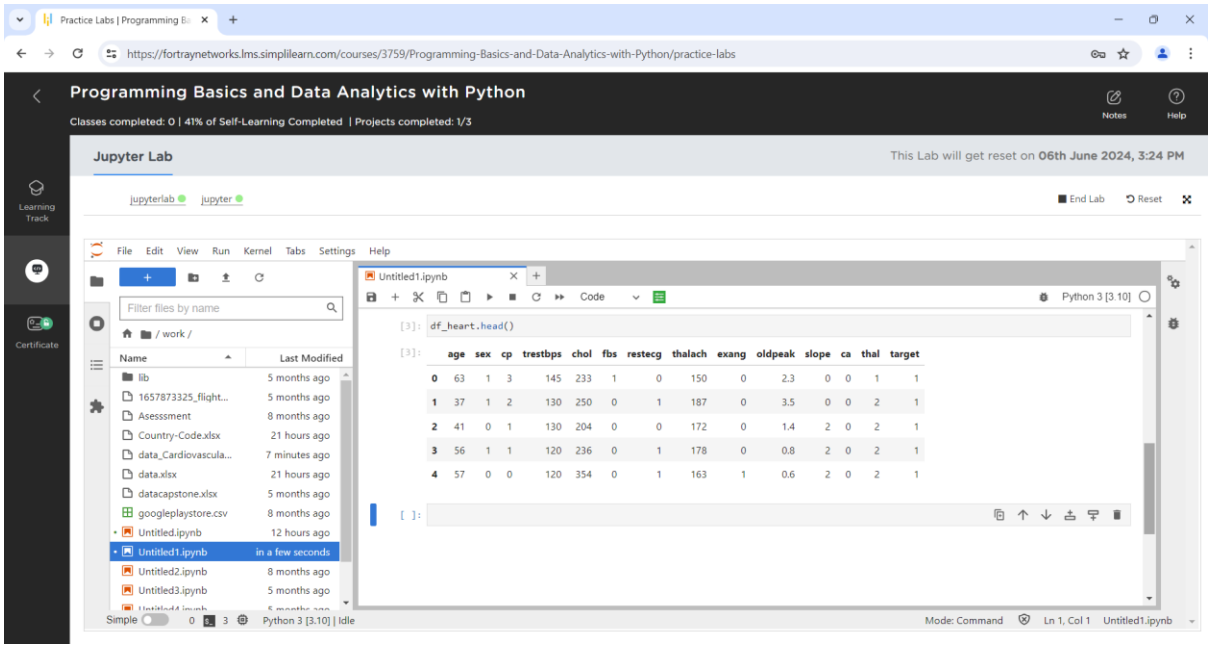
Preliminary analysis:

Perform preliminary data inspection and report the findings as the structure of the data, missing values, duplicates etc.

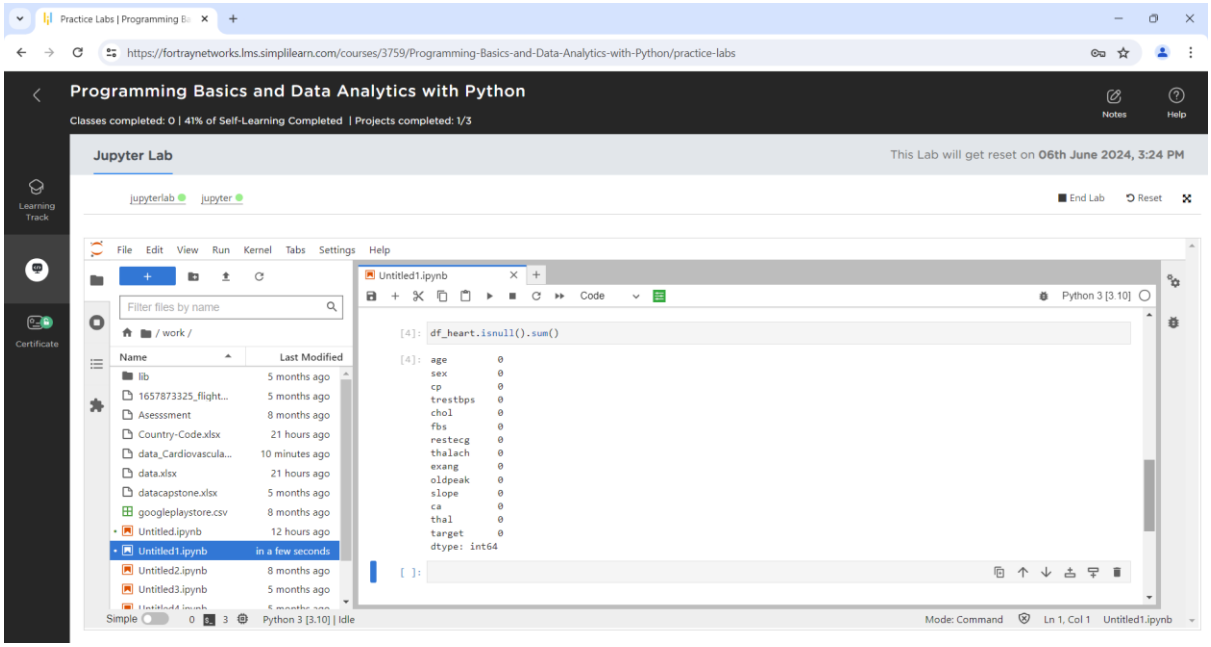
Based on the findings from the previous question remove duplicates (if any) , treat missing values using appropriate strategy.



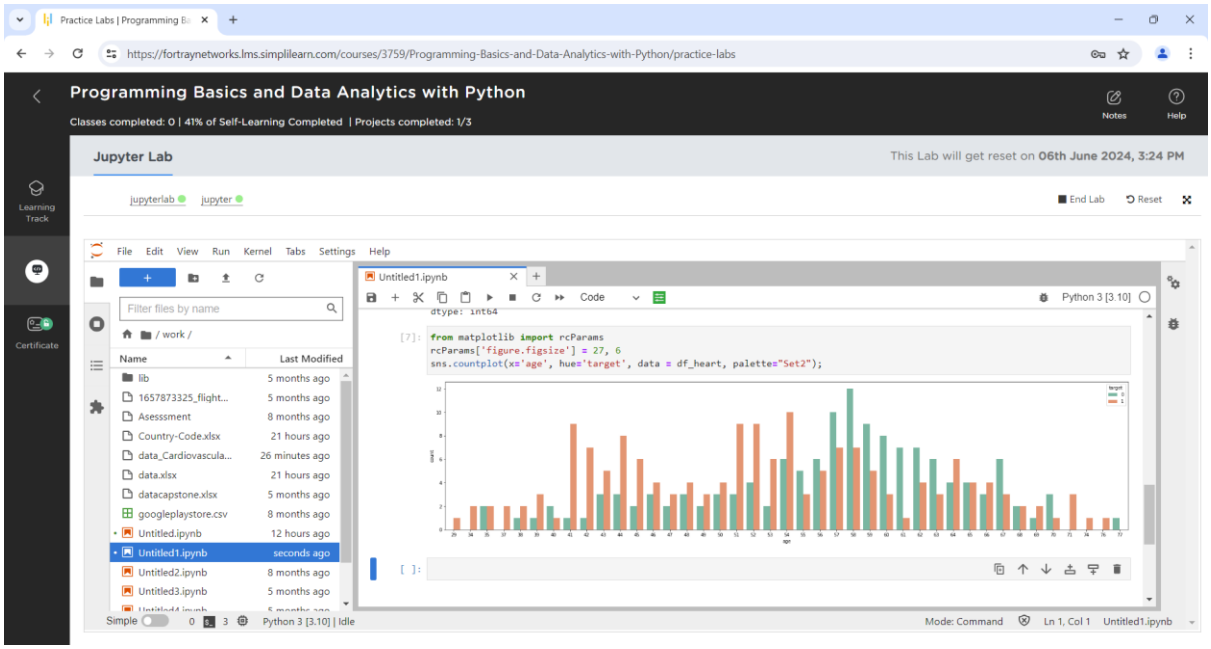



No null values detected:

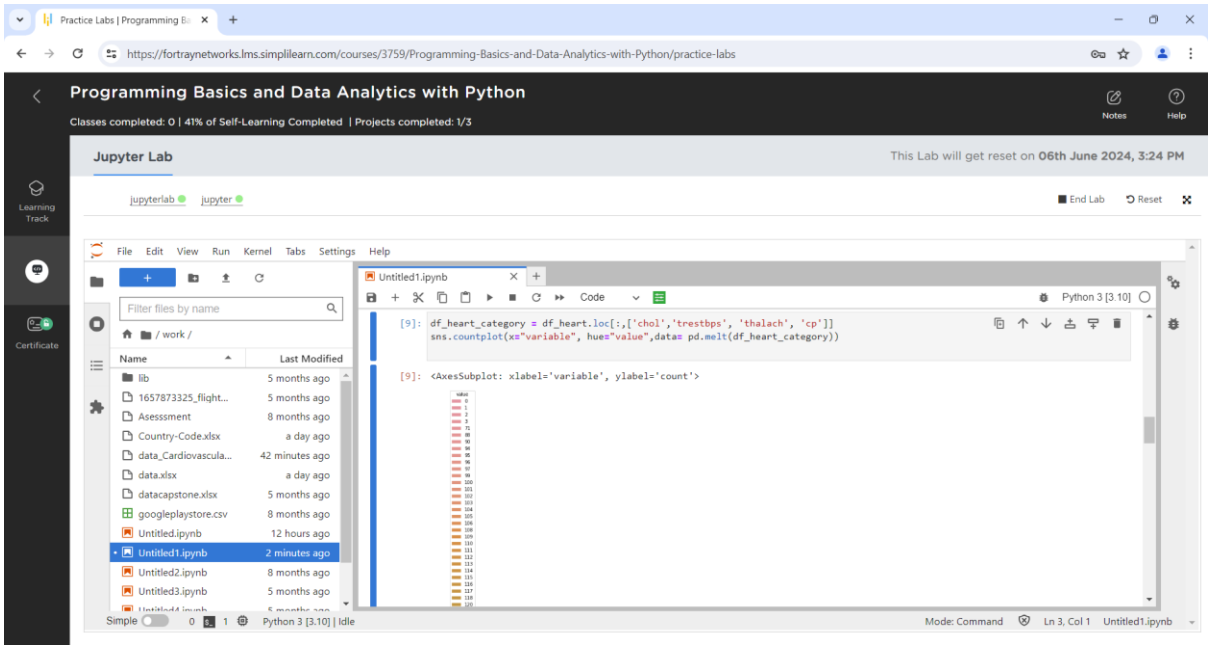


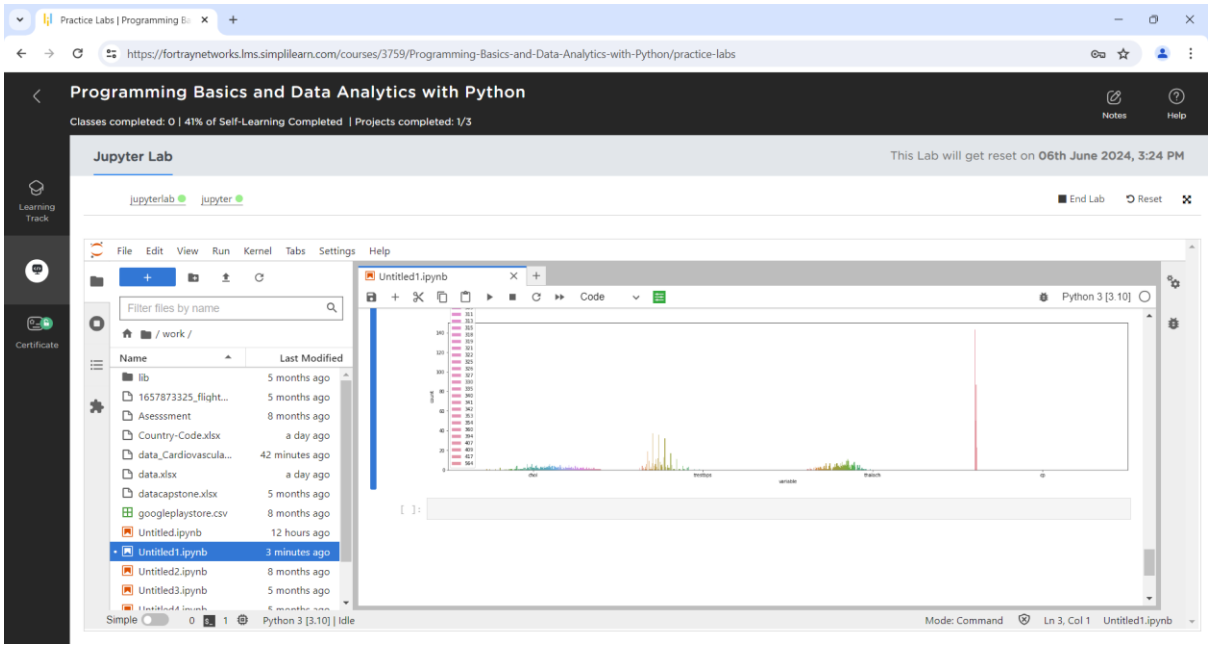
Get a preliminary statistical summary of the data. Explore the measures of central tendencies and the spread of the data overall.

Identify the data variables which might be categorical in nature. Describe and explore these variables using appropriate tools e.g. count plot

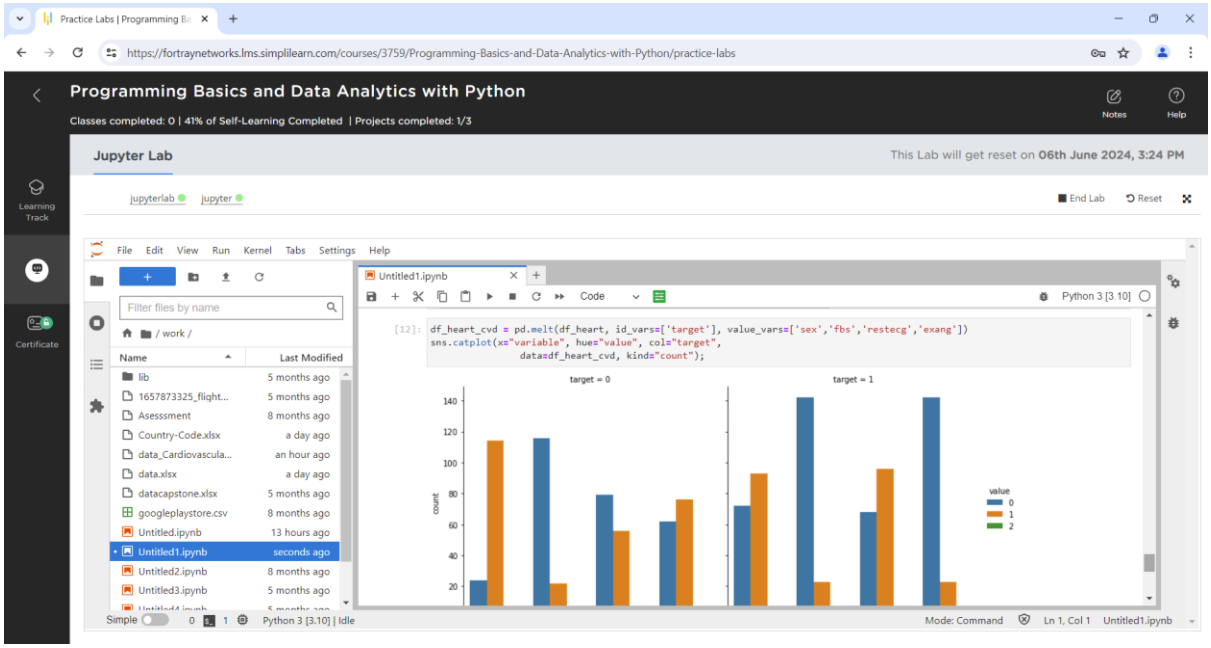


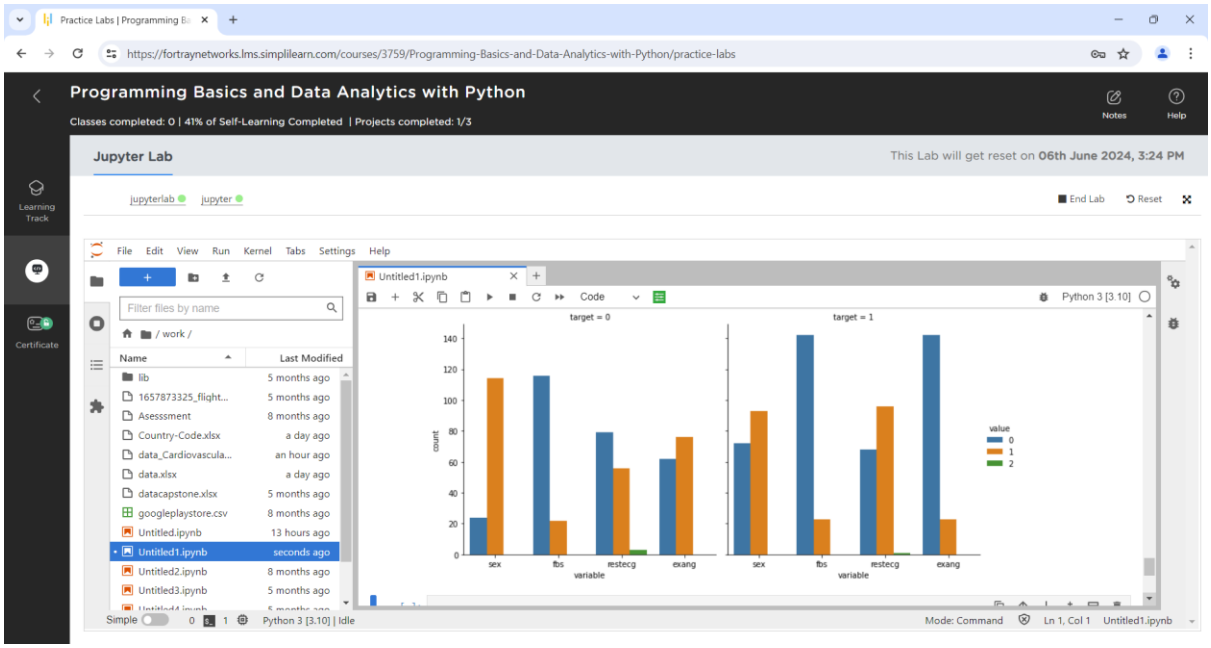

## Let’s look at categorical variables and their distribution.

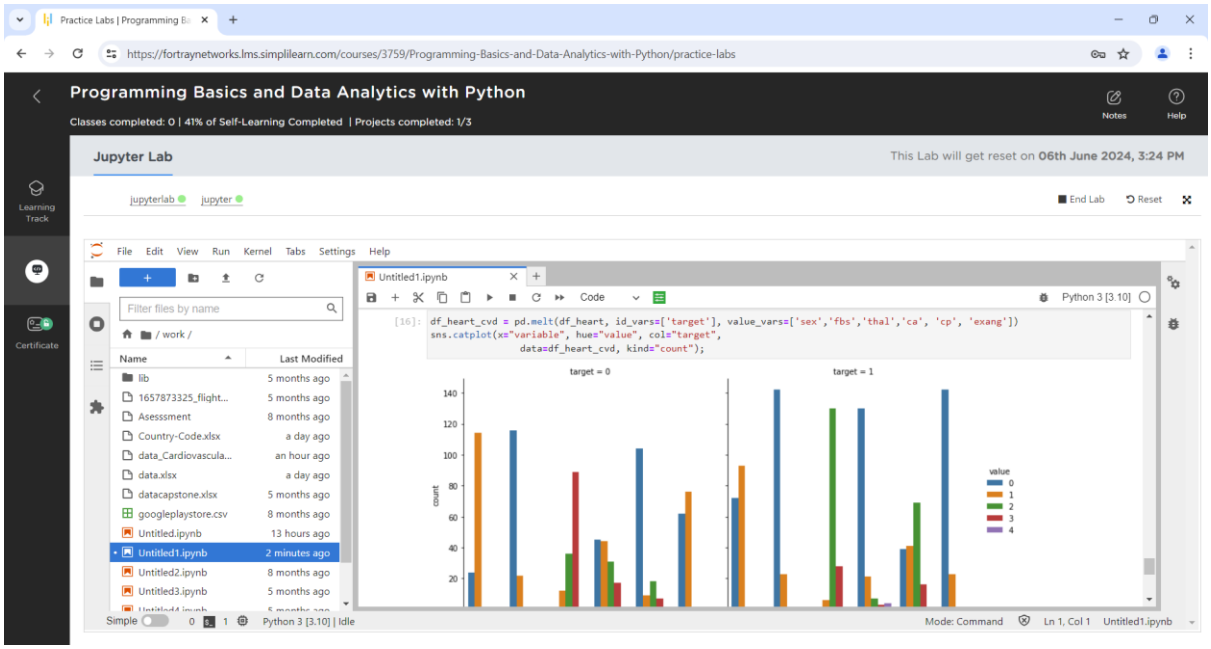


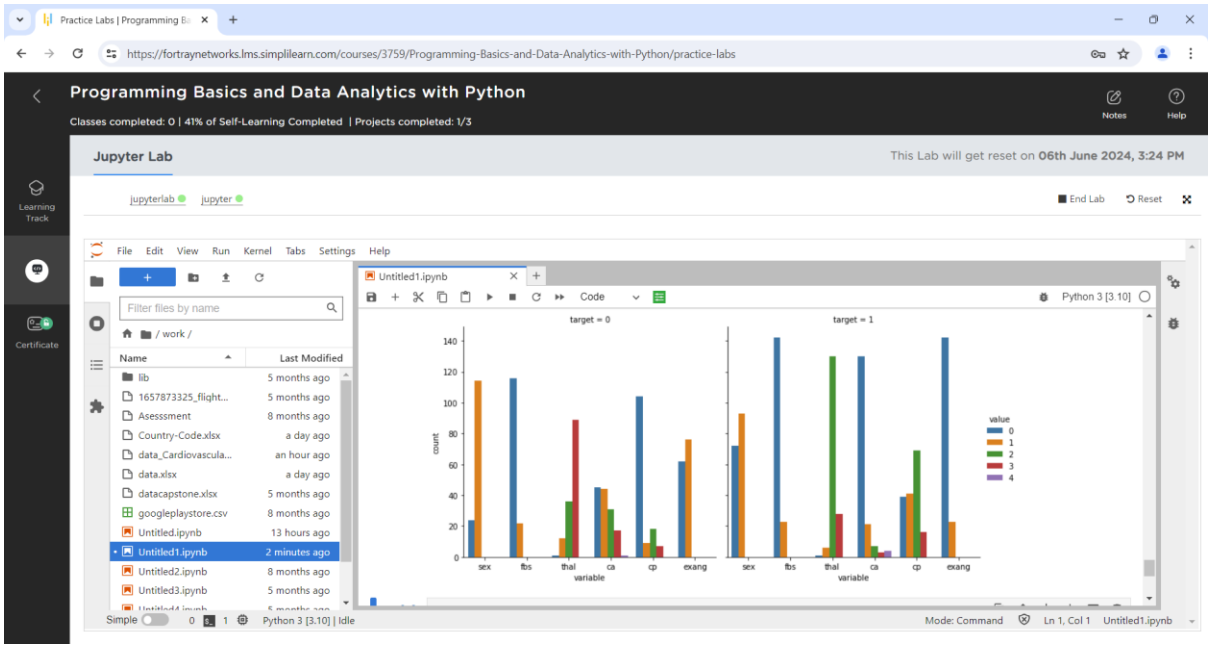



Cholesterol, blood pressure and chest pain are correlated.



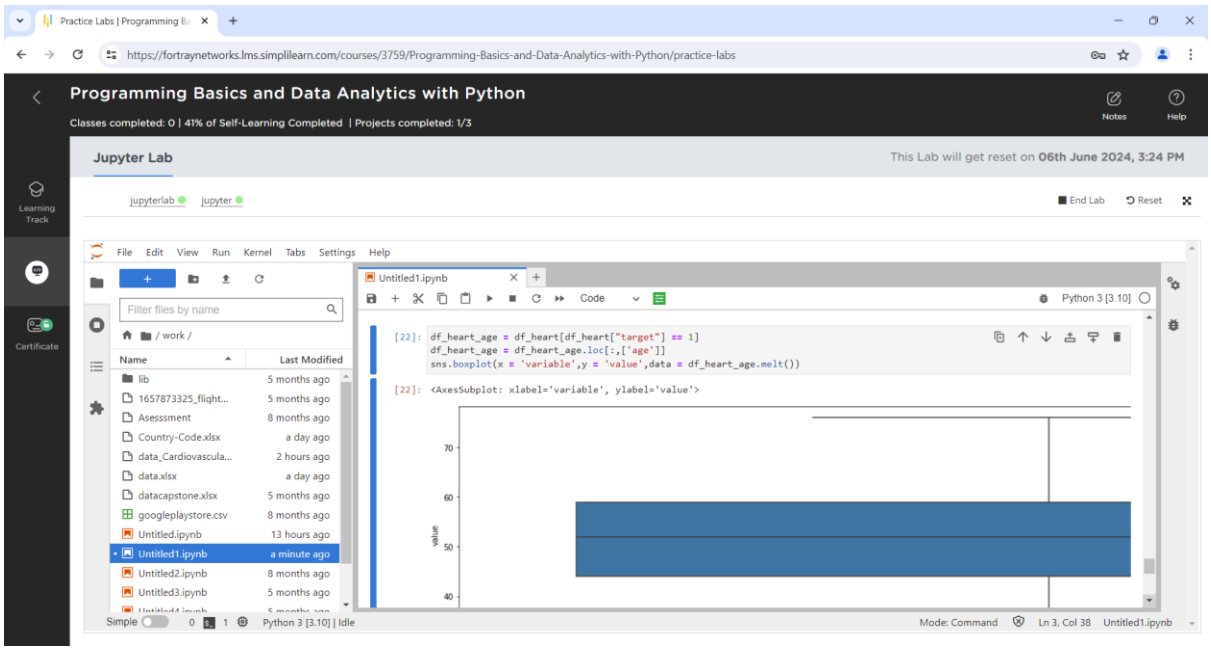

Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Jupyter Lab
This Lab will get reset on 06th June 2024, 3:24 PM
Learning Track
Certificate
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ work /
Name
Last Modified
lib 5 months ago
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx a day ago
data_Cardiovascula... an hour ago
data.xlsx a day ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb 13 hours ago
Untitled1.ipynb seconds ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled1.ipynb
Code
Python 3 [3.10]
target = 0
target = 1
count
variable
value
sex
fbs
restecg
exang
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled1.ipynb

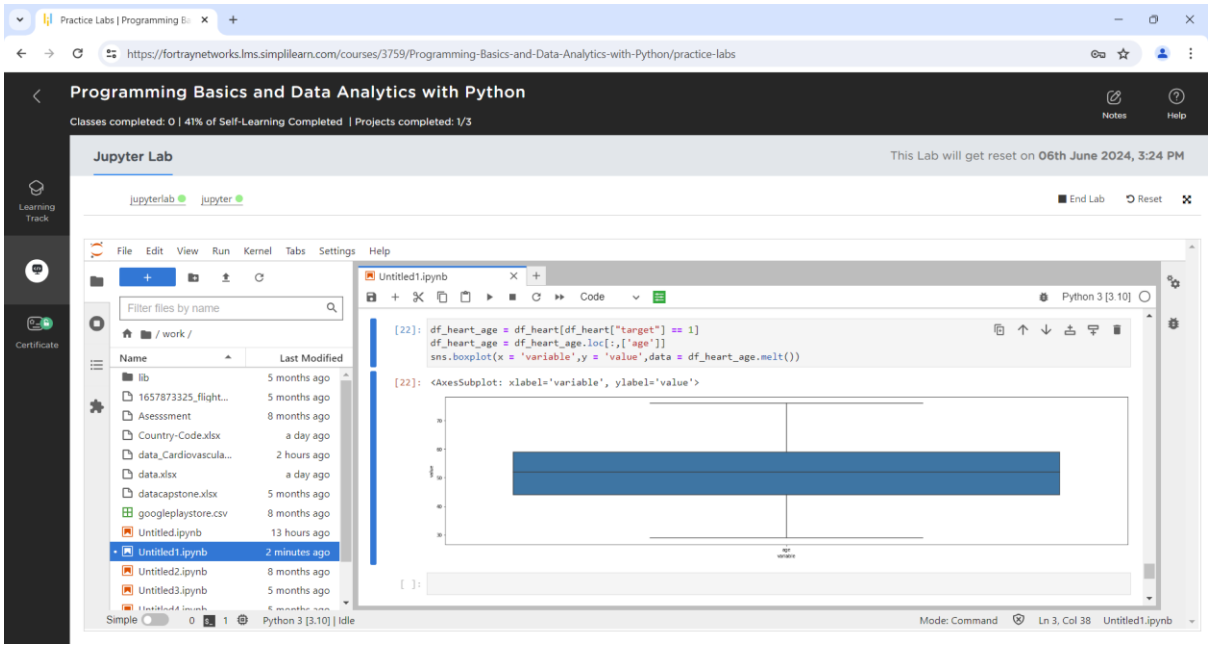

Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Jupyter Lab
This Lab will get reset on 06th June 2024, 3:24 PM
Learning Track
Certificate
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ work /
Name
Last Modified
lib 5 months ago
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx a day ago
data_Cardiovascula... an hour ago
data.xlsx a day ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb 13 hours ago
Untitled1.ipynb 2 minutes ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled1.ipynb
Code
Python 3 [3.10]
[16]:
df_heart_cvd = pd.melt(df_heart, id_vars=['target'], value_vars=['sex','fbs','thal','ca', 'cp', 'exang'])
sns.catplot(x="variable", hue="value", col="target",
data=df_heart_cvd, kind="count");
target = 0
target = 1
count
variable
value
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled1.ipynb




Patients with cardio vascular disease are less active and have high blood sugar level and cholesterol.

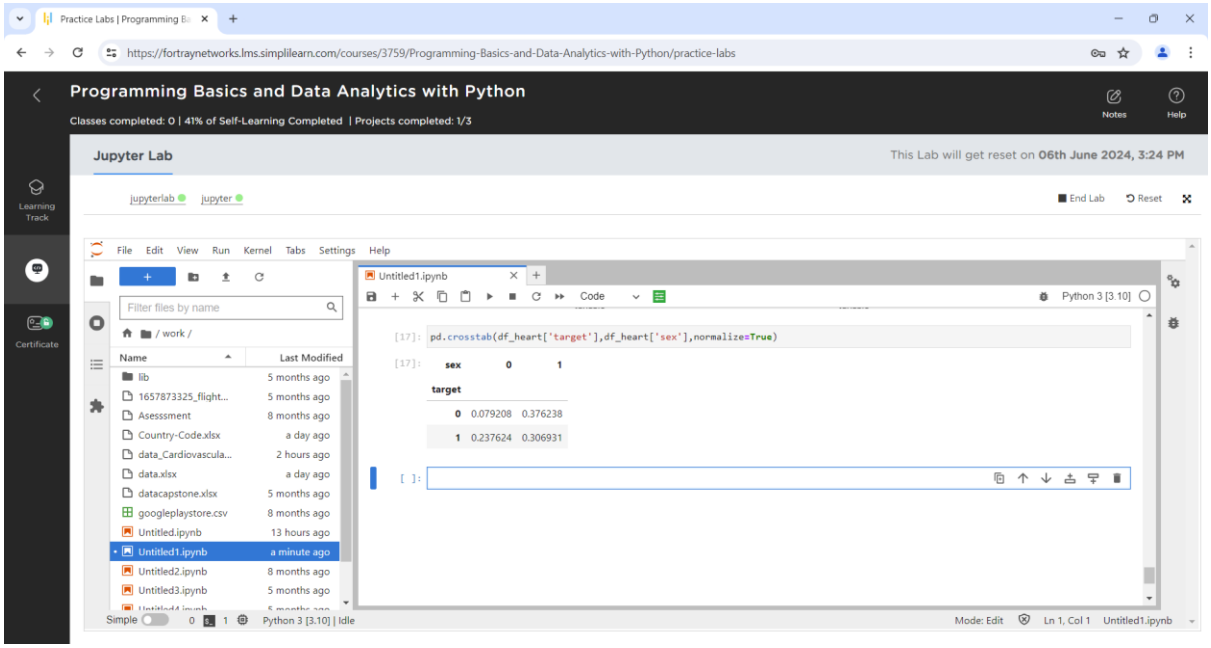
Study the occurrence of CVD across Age.






Cardio vascular disease happens in the age range between 45 and 60.

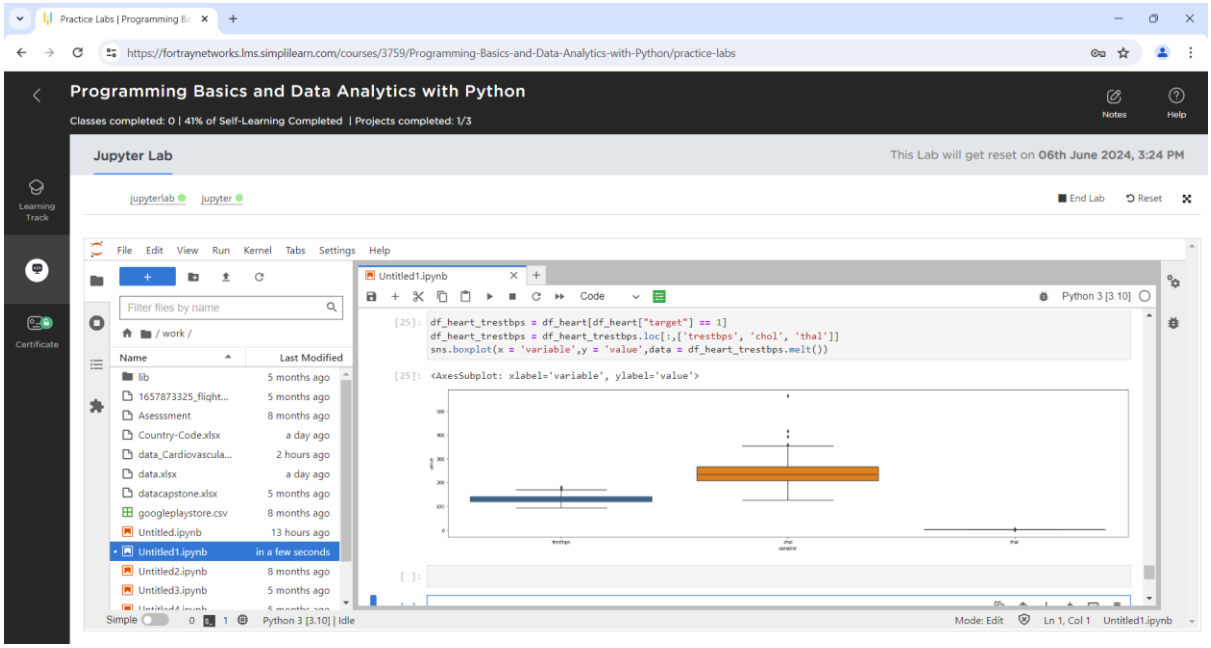
Let us check cardio vascular disease ratio by gender:



Can we detect heart attack based on anomalies in Resting Blood Pressure of the patient?

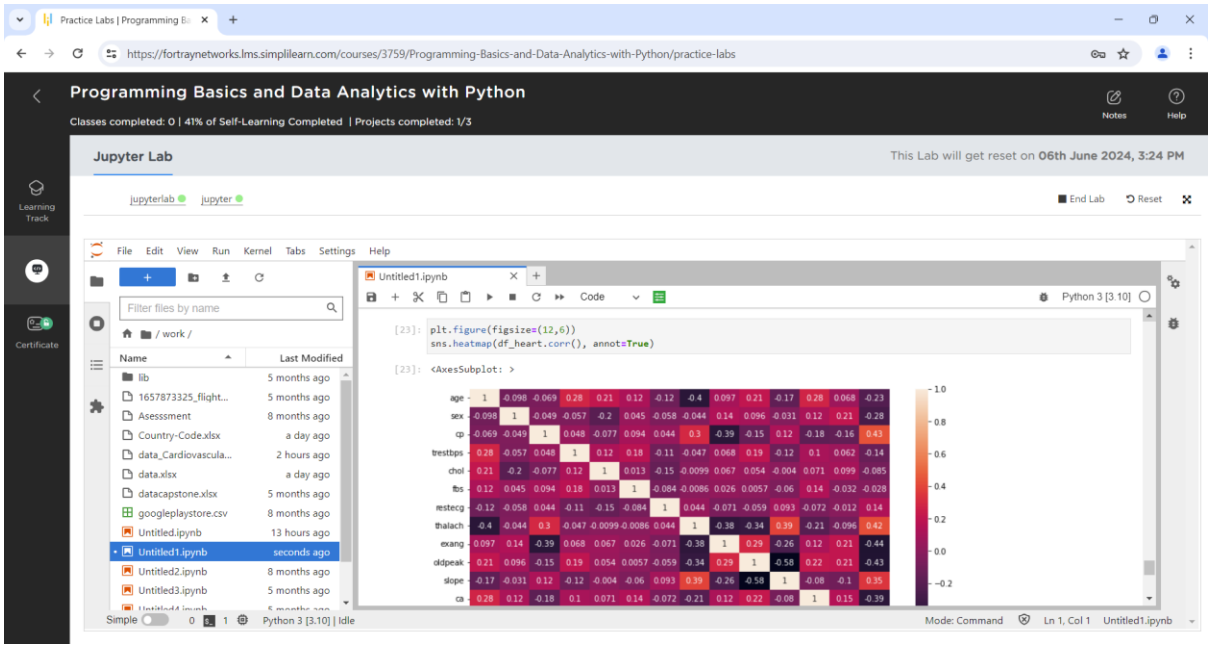
Describe the relationship between Cholesterol levels and our target variable.

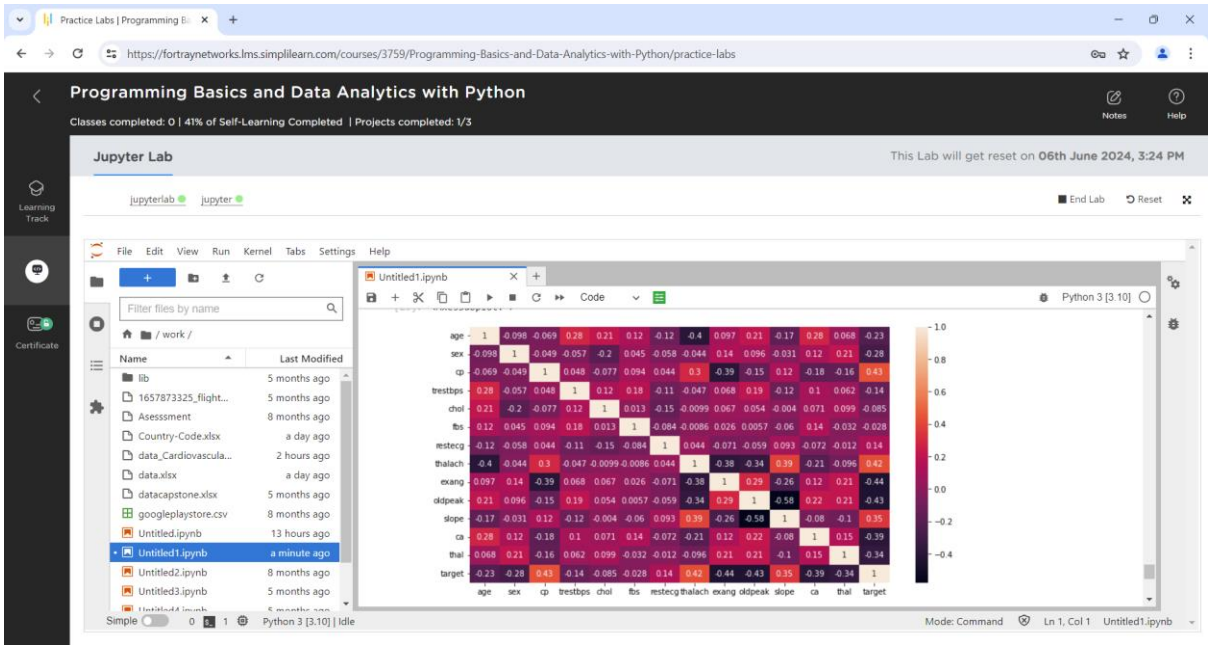
Is thalassemia a major cause of CVD?



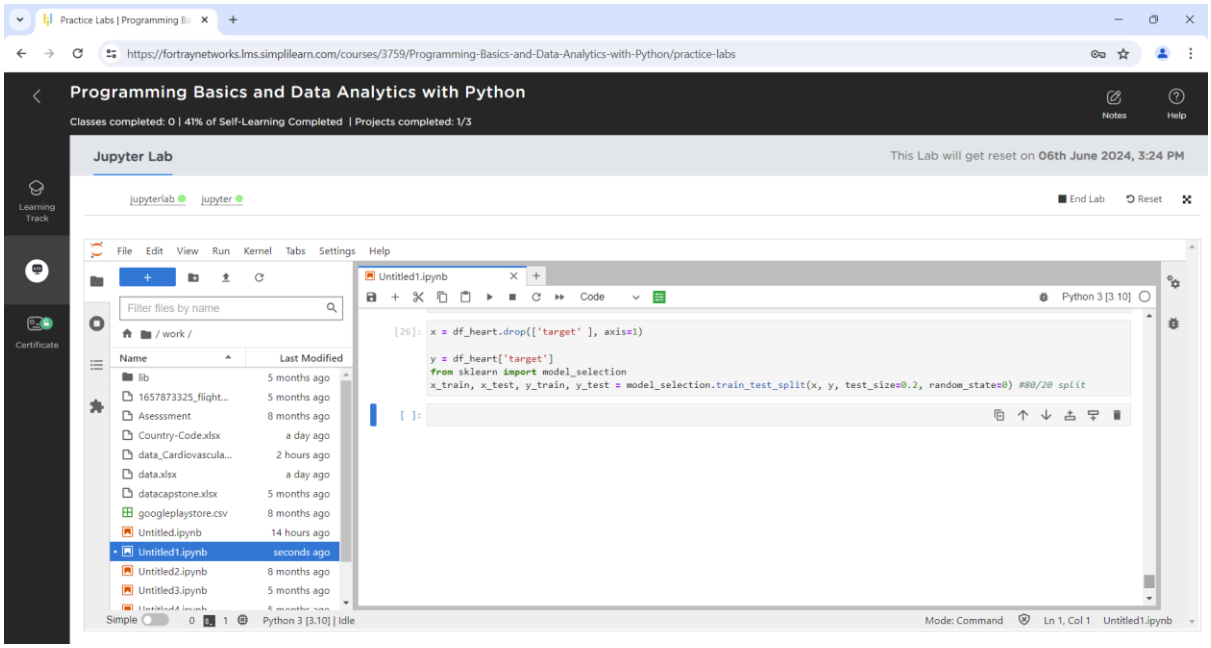

It can be concluded that high blood pressure, cholesterol and thalassemia are major causes of cardiovascular diseases.

Correlation matrix:

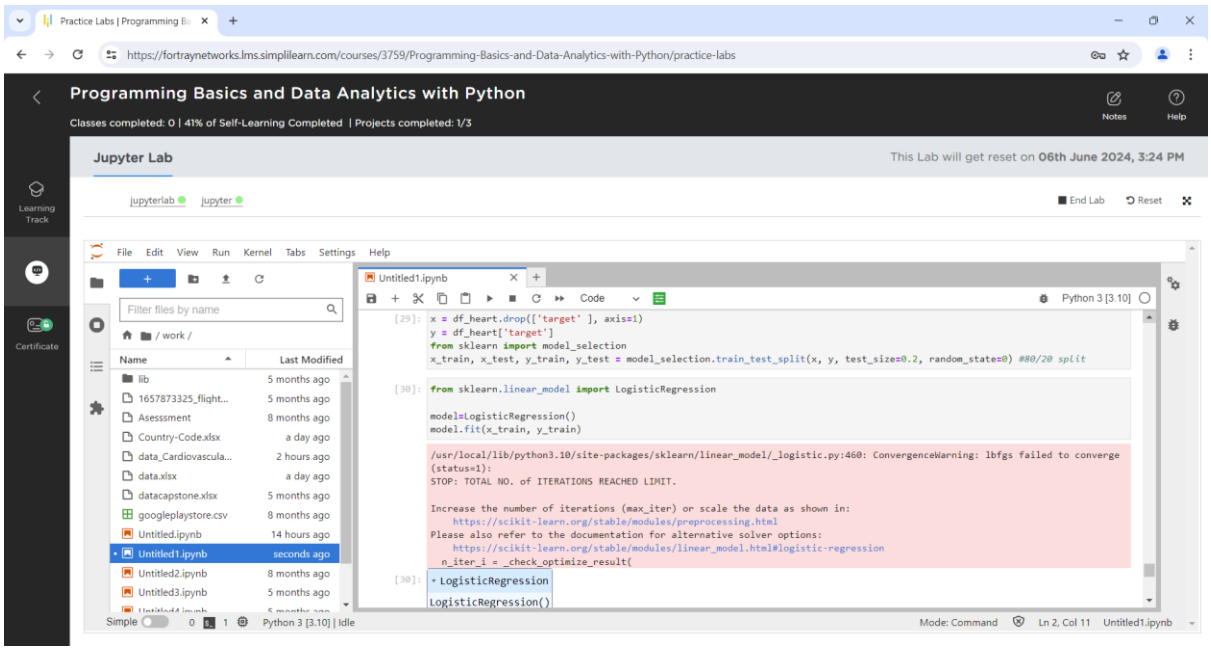


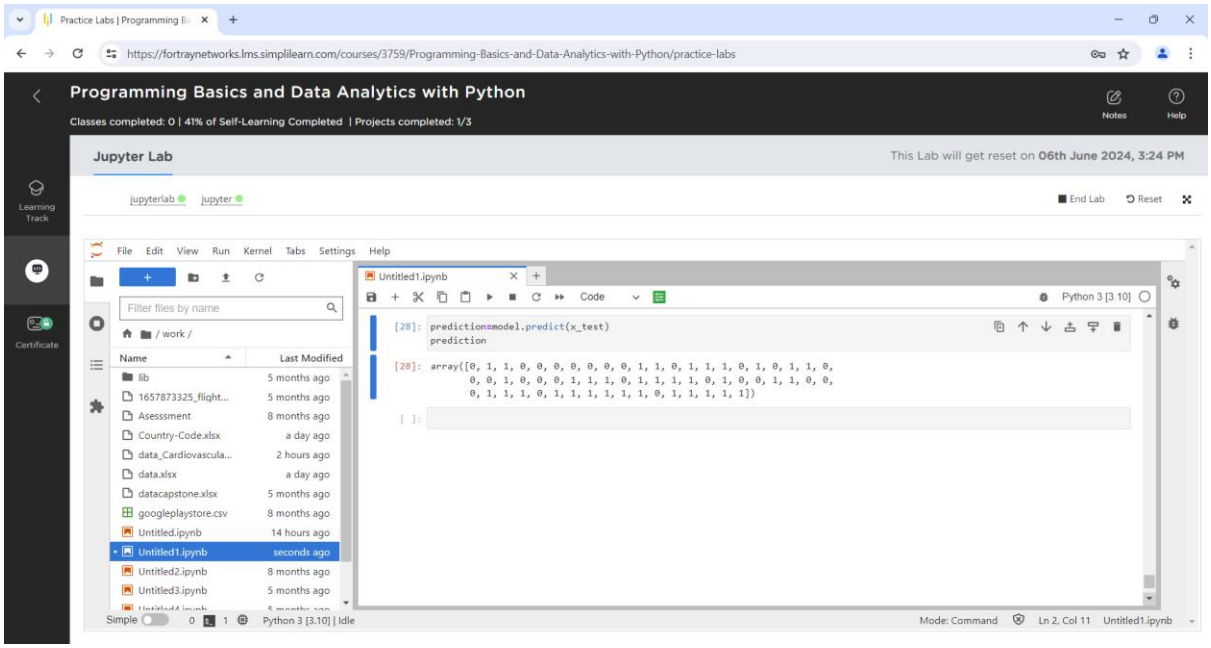



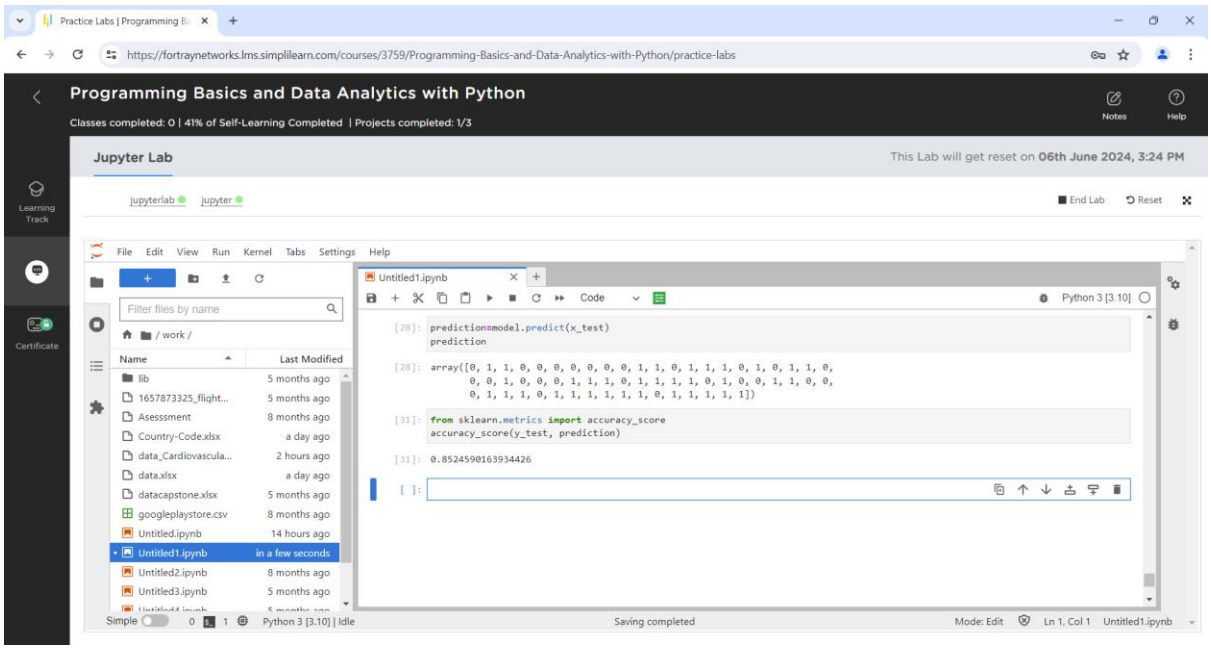

Splitting the dataset into test and train:



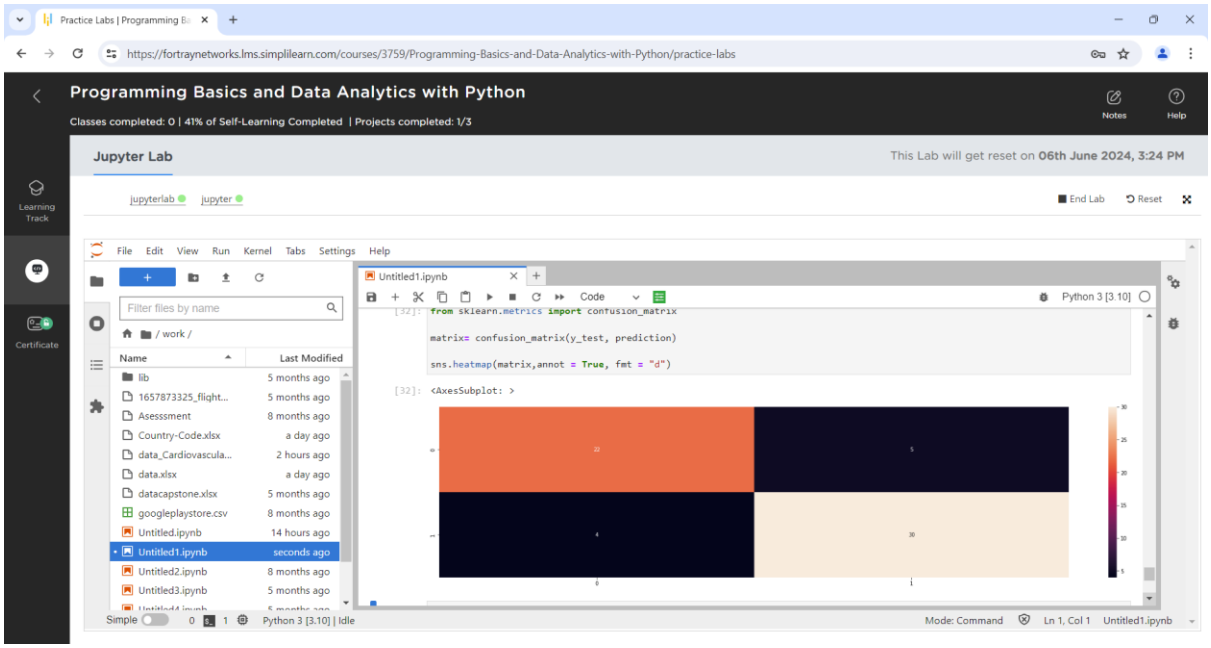

# Logistics Regression:






Confusion matrix:


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 3:24 PM
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ work /
Name
Last Modified
lib 5 months ago
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx a day ago
data_Cardiovascula... 2 hours ago
data.xlsx a day ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb 14 hours ago
Untitled1.ipynb seconds ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled1.ipynb
Code
Python 3 [3.10]
[32]: from sklearn.metrics import confusion_matrix
matrix= confusion_matrix(y_test, prediction)
sns.heatmap(matrix,annot = True, fmt = "d")
[32]: <AxesSubplot: >
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled1.ipynb