

## 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.

The screenshot shows the Jupyter Lab interface with a file browser on the left and a code editor on the right. The file browser lists files in the /work/ directory, including 'lib', '1657873325\_flight...', 'Assessment', 'Country-Code.xlsx', 'data\_Cardiovascula...', 'data.xlsx', 'datacapstone.xlsx', 'googleplaystore.csv', 'Untitled1.ipynb', 'Untitled2.ipynb', 'Untitled3.ipynb', and 'Untitled4.ipynb'. The code editor shows the following code:

```
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df_heart = pd.read_excel('data_Cardiovascula.xlsx', sheet_name='heart')
df_heart.info()
```

The output of the code is displayed below the code cell:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype  
---  -
 0   age           303 non-null    int64  
 1   sex           303 non-null    int64  
 2   cp            303 non-null    int64  
 3   trestbps      303 non-null    int64  
 4   chol          303 non-null    int64  
 5   fbs           303 non-null    int64  
 6   restecg       303 non-null    int64  
 7   thalach       303 non-null    int64  
 8   exang         303 non-null    int64  
 9   oldpeak       303 non-null    float64 
10   slope         303 non-null    int64  
11   ca            303 non-null    int64  
12   thal          303 non-null    int64  
13   target        303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

The screenshot shows the Jupyter Lab interface with the same file browser on the left. The code editor shows the following code:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype  
---  -
 0   age           303 non-null    int64  
 1   sex           303 non-null    int64  
 2   cp            303 non-null    int64  
 3   trestbps      303 non-null    int64  
 4   chol          303 non-null    int64  
 5   fbs           303 non-null    int64  
 6   restecg       303 non-null    int64  
 7   thalach       303 non-null    int64  
 8   exang         303 non-null    int64  
 9   oldpeak       303 non-null    float64 
10   slope         303 non-null    int64  
11   ca            303 non-null    int64  
12   thal          303 non-null    int64  
13   target        303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

Practice Labs | Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Programming Basics and Data Analytics with Python

This Lab will get reset on 06th June 2024, 3:24 PM

Jupyter Lab

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name Last Modified

| Name                  | Last Modified    |
|-----------------------|------------------|
| lib                   | 5 months ago     |
| 1657873325_flight...  | 5 months ago     |
| Assessment            | 8 months ago     |
| Country-Code.xlsx     | 21 hours ago     |
| data_Cardiovascula... | 7 minutes ago    |
| data.xlsx             | 21 hours ago     |
| datacapstone.xlsx     | 5 months ago     |
| googleplaystore.csv   | 8 months ago     |
| Untitled1.ipynb       | 12 hours ago     |
| Untitled1.ipynb       | in a few seconds |
| Untitled2.ipynb       | 8 months ago     |
| Untitled3.ipynb       | 5 months ago     |
| Untitled4.ipynb       | 6 months ago     |

Untitled1.ipynb

```
[3]: df_heart.head()
```

|   | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|---|-----|-----|----|----------|------|-----|---------|---------|-------|---------|-------|----|------|--------|
| 0 | 63  | 1   | 3  | 145      | 233  | 1   | 0       | 150     | 0     | 2.3     | 0     | 0  | 1    | 1      |
| 1 | 37  | 1   | 2  | 130      | 250  | 0   | 1       | 187     | 0     | 3.5     | 0     | 0  | 2    | 1      |
| 2 | 41  | 0   | 1  | 130      | 204  | 0   | 0       | 172     | 0     | 1.4     | 2     | 0  | 2    | 1      |
| 3 | 56  | 1   | 1  | 120      | 236  | 0   | 1       | 178     | 0     | 0.8     | 2     | 0  | 2    | 1      |
| 4 | 57  | 0   | 0  | 120      | 354  | 0   | 1       | 163     | 1     | 0.6     | 2     | 0  | 2    | 1      |

Mode: Command Ln 1, Col 1 Untitled1.ipynb

No null values detected:

Practice Labs | Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Programming Basics and Data Analytics with Python

This Lab will get reset on 06th June 2024, 3:24 PM

Jupyter Lab

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name Last Modified

| Name                  | Last Modified    |
|-----------------------|------------------|
| lib                   | 5 months ago     |
| 1657873325_flight...  | 5 months ago     |
| Assessment            | 8 months ago     |
| Country-Code.xlsx     | 21 hours ago     |
| data_Cardiovascula... | 10 minutes ago   |
| data.xlsx             | 21 hours ago     |
| datacapstone.xlsx     | 5 months ago     |
| googleplaystore.csv   | 8 months ago     |
| Untitled1.ipynb       | 12 hours ago     |
| Untitled1.ipynb       | in a few seconds |
| Untitled2.ipynb       | 8 months ago     |
| Untitled3.ipynb       | 5 months ago     |
| Untitled4.ipynb       | 6 months ago     |

Untitled1.ipynb

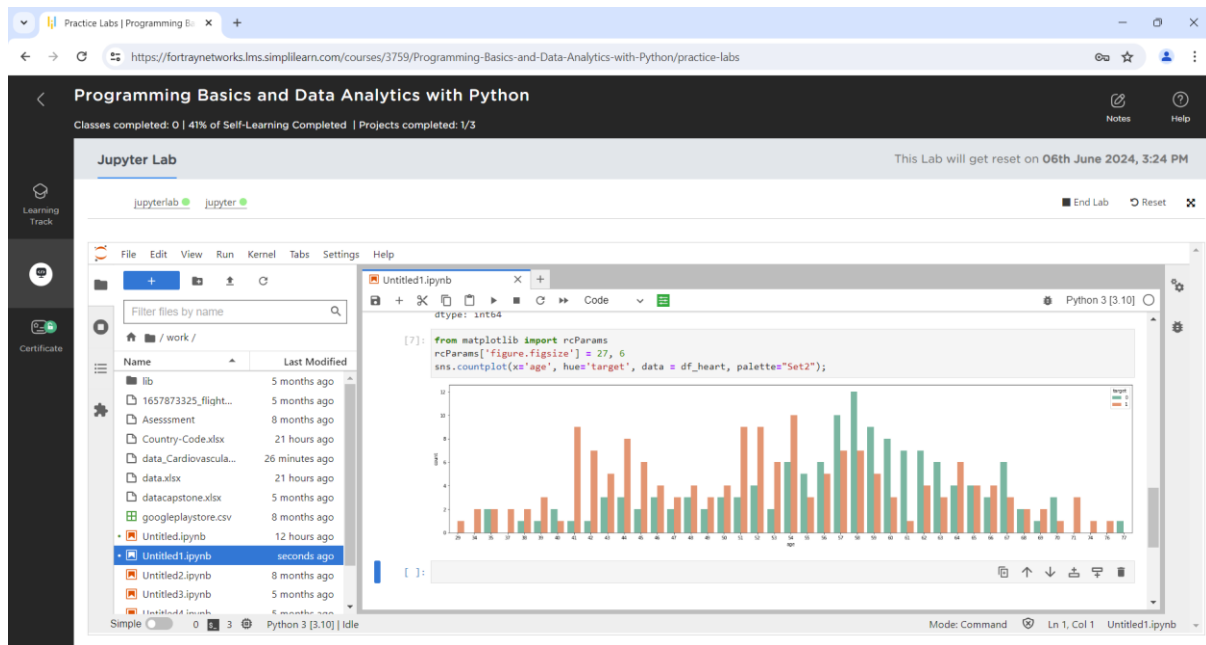
```
[4]: df_heart.isnull().sum()
```

|          | age   | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
|----------|-------|-----|----|----------|------|-----|---------|---------|-------|---------|-------|----|------|--------|
| age      | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| sex      | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| cp       | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| trestbps | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| chol     | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| fbs      | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| restecg  | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| thalach  | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| exang    | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| oldpeak  | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| slope    | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| ca       | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| thal     | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| target   | 0     |     |    |          |      |     |         |         |       |         |       |    |      |        |
| dtype:   | int64 |     |    |          |      |     |         |         |       |         |       |    |      |        |

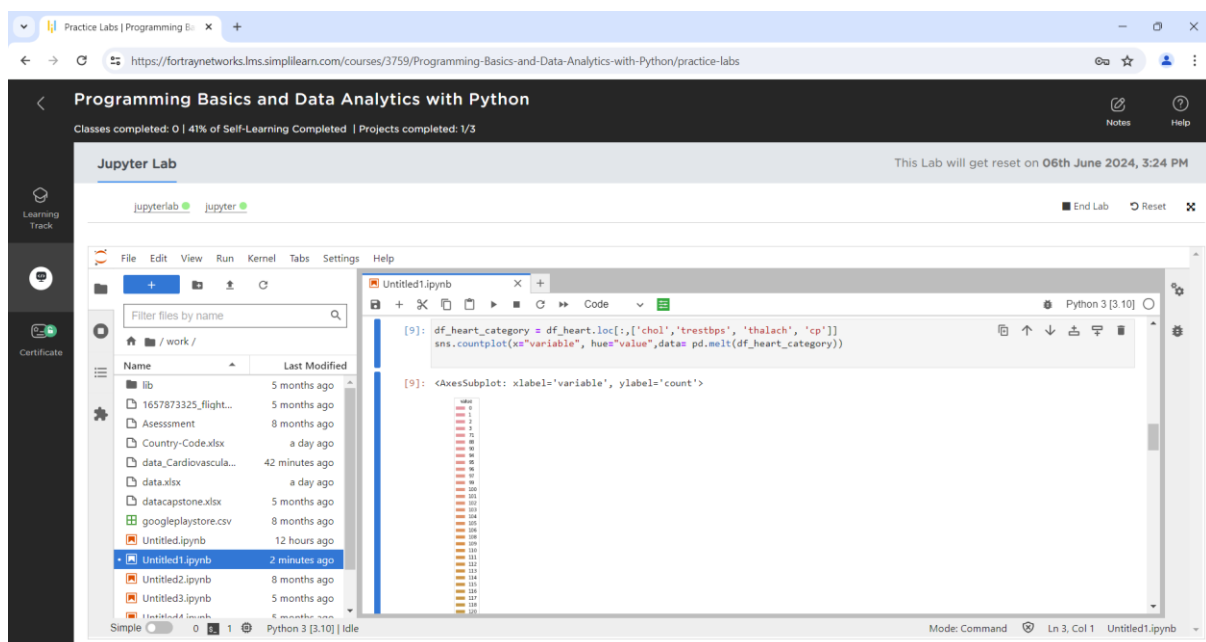
Mode: Command Ln 1, Col 1 Untitled1.ipynb

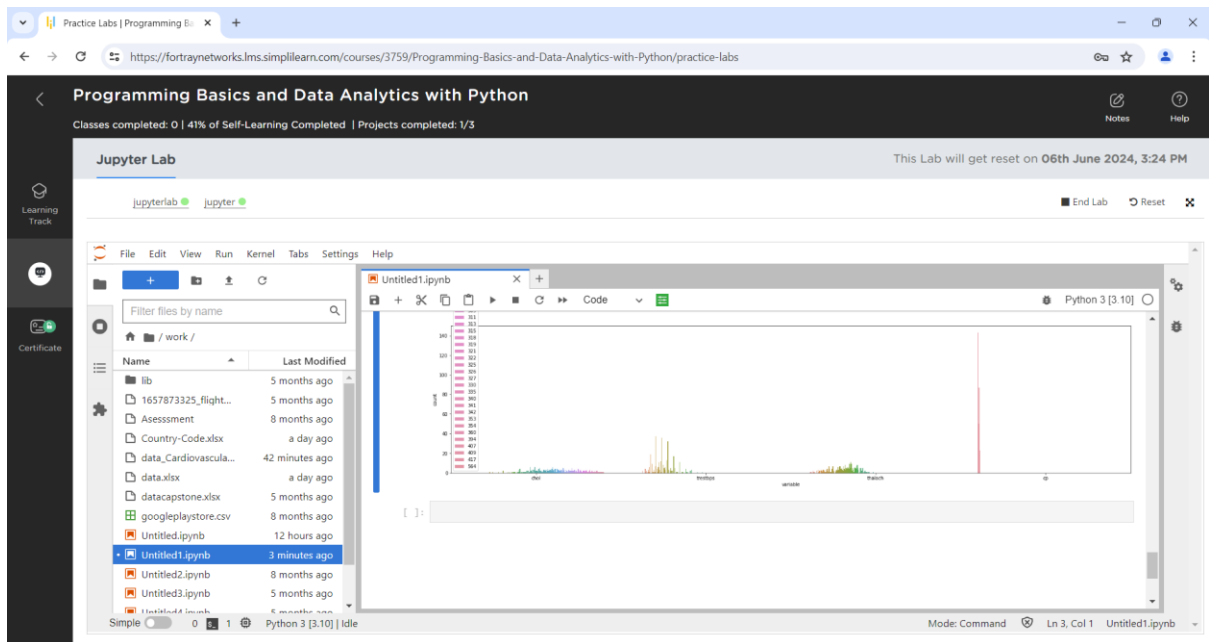
Get a preliminary statistical summary of the data. Explore the measures of central tendencies and the spread of the data overall.

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

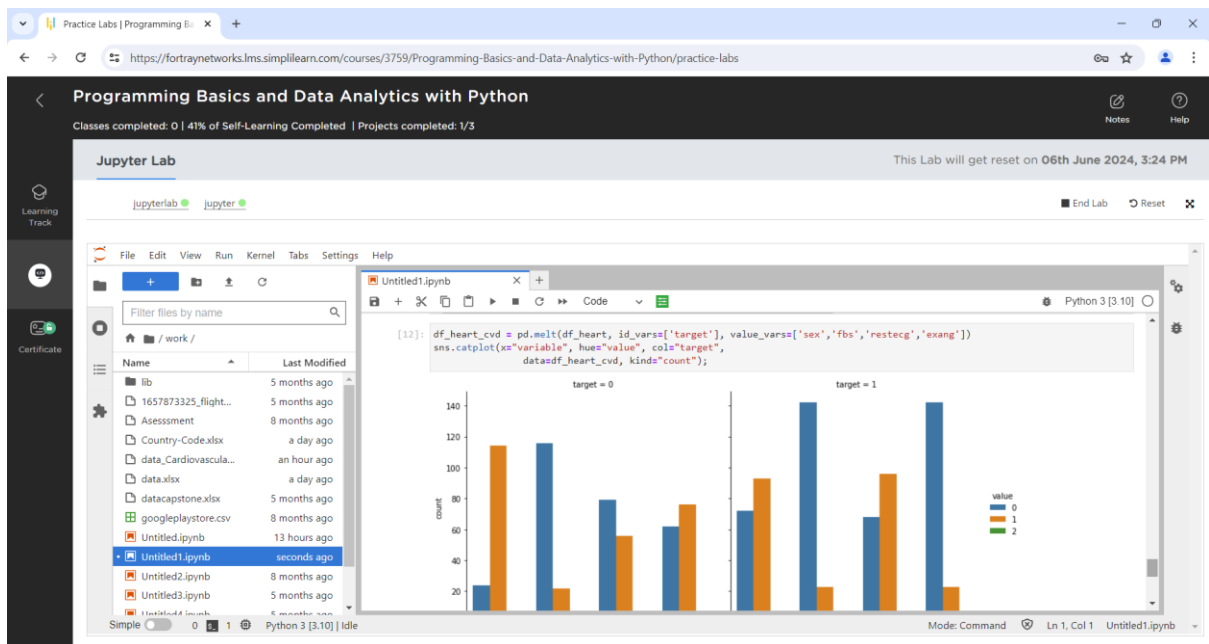


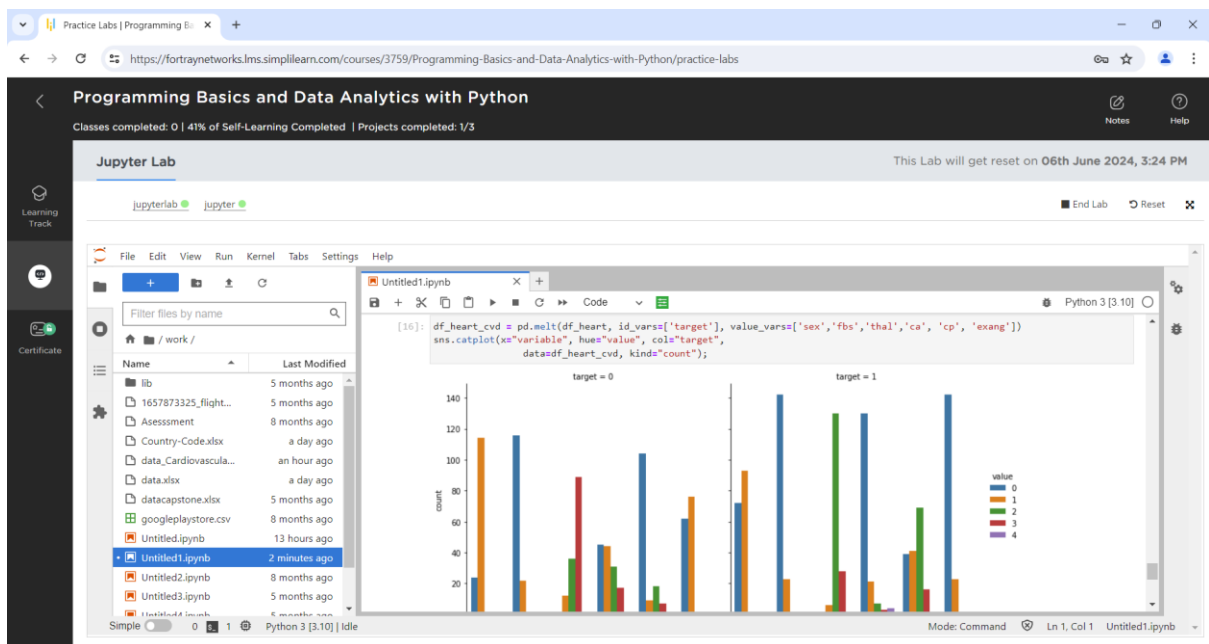
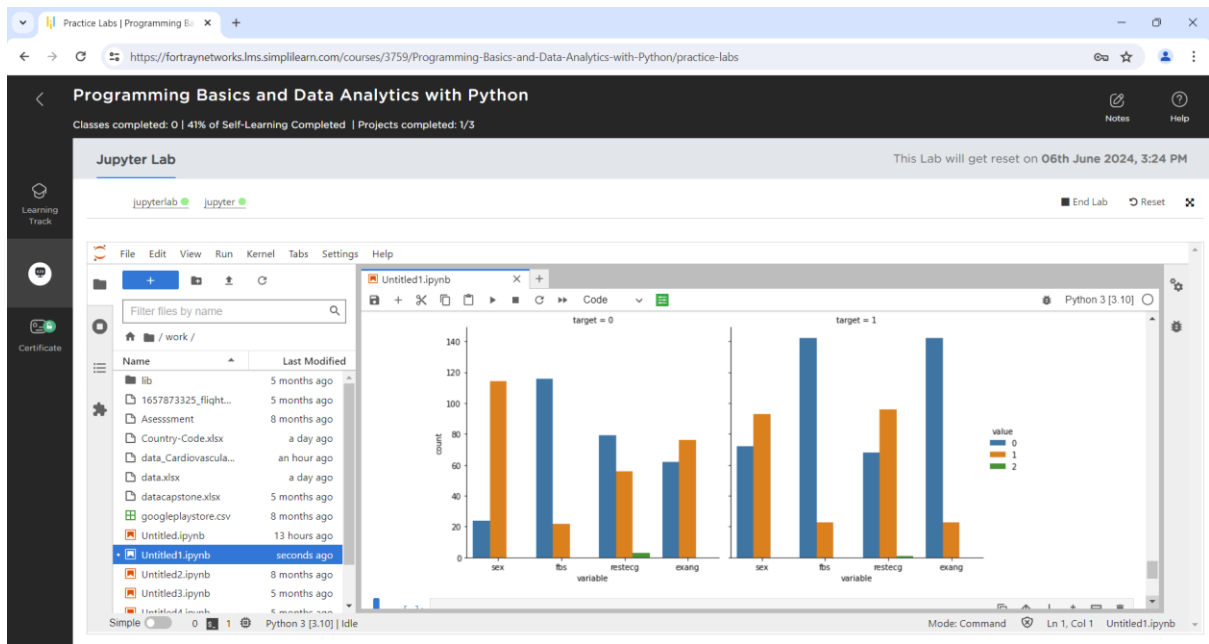
Let's look at categorical variables and their distribution.

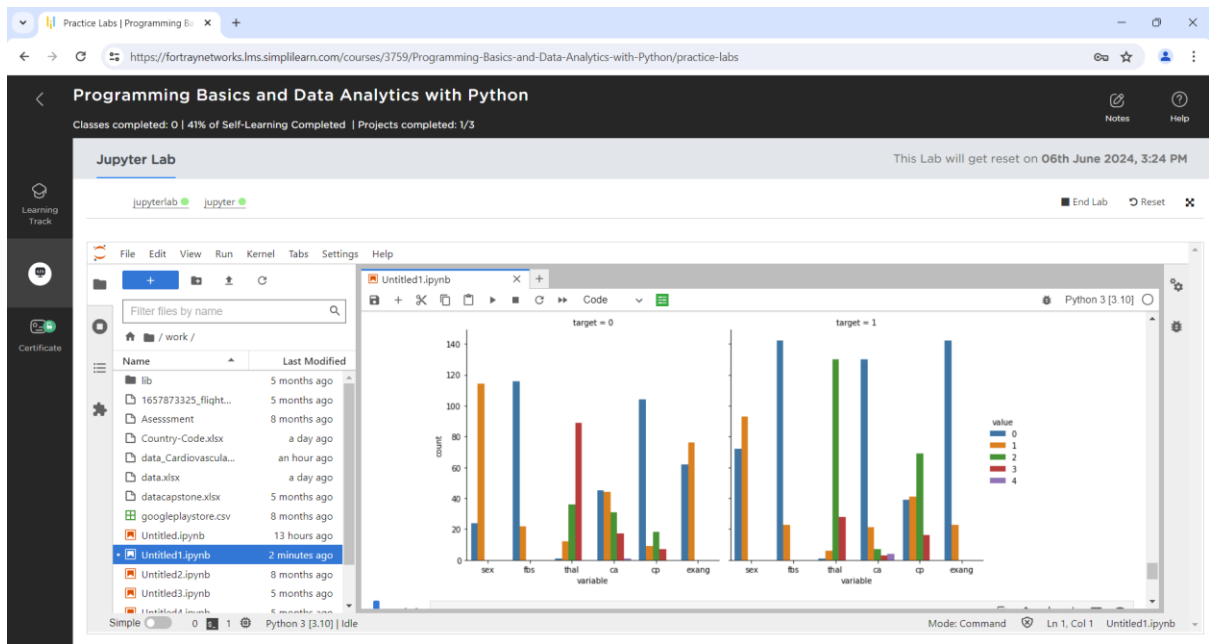




Cholesterol, blood pressure and chest pain are correlated.

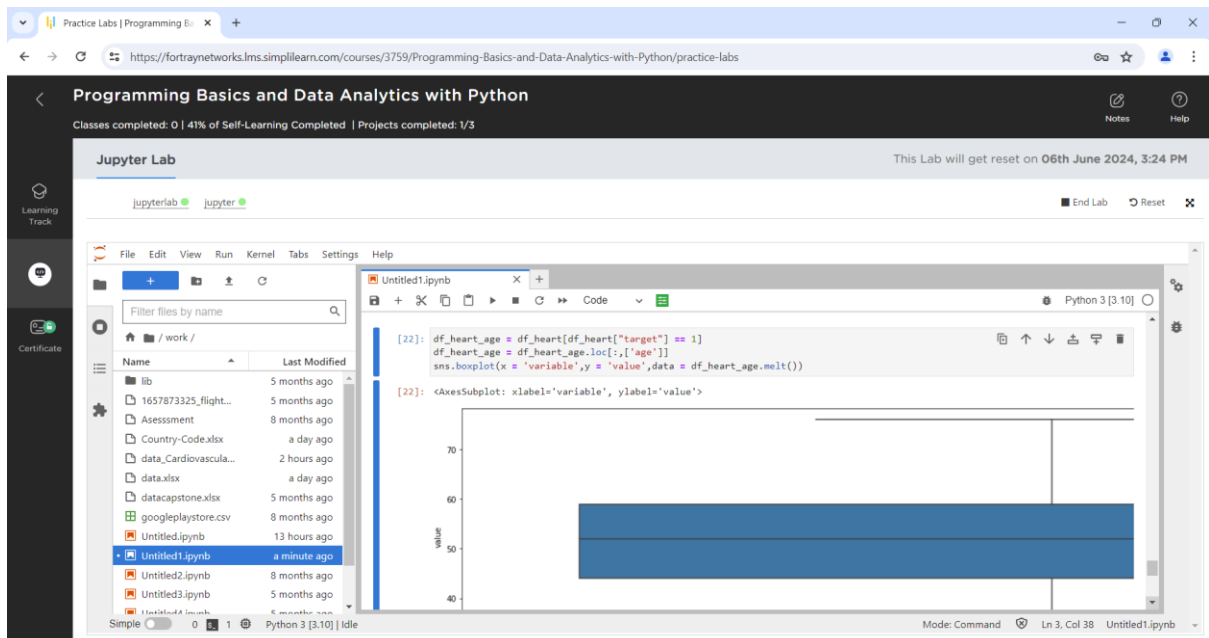


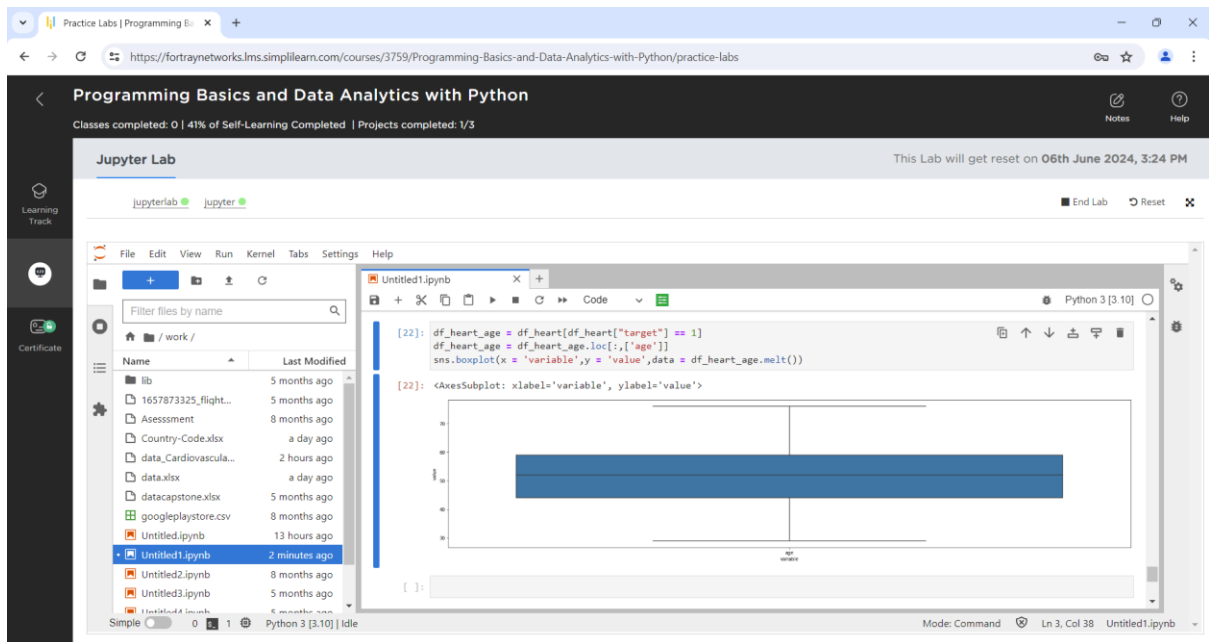




Patients with cardio vascular disease are less active and have high blood sugar level and cholesterol.

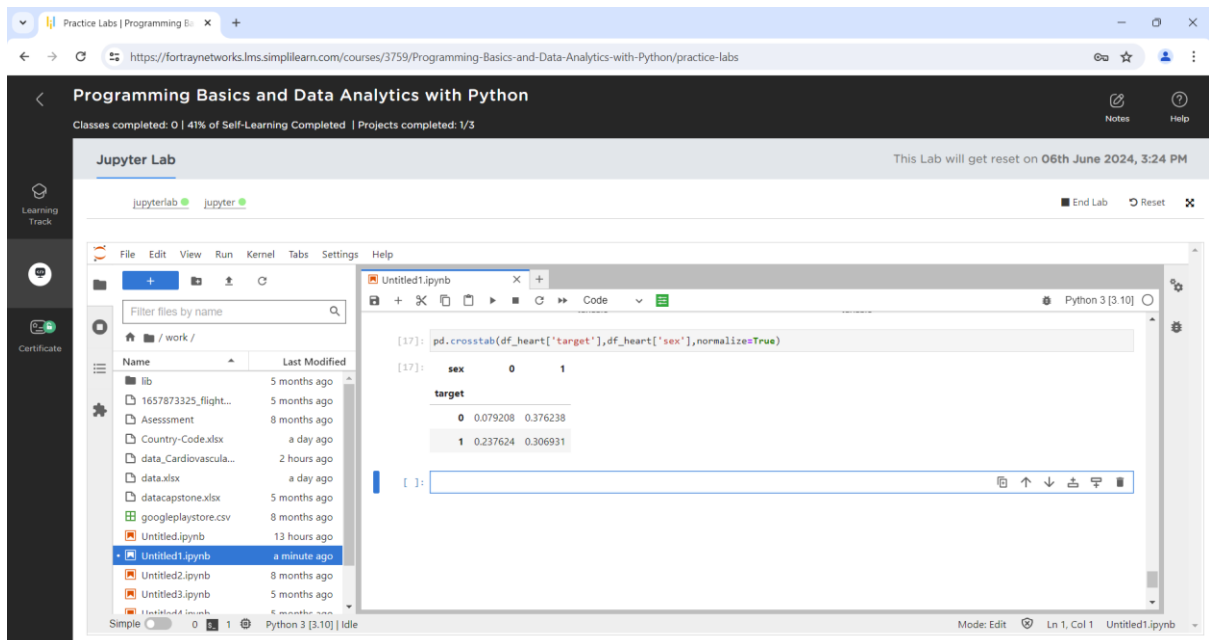
Study the occurrence of CVD across Age.





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

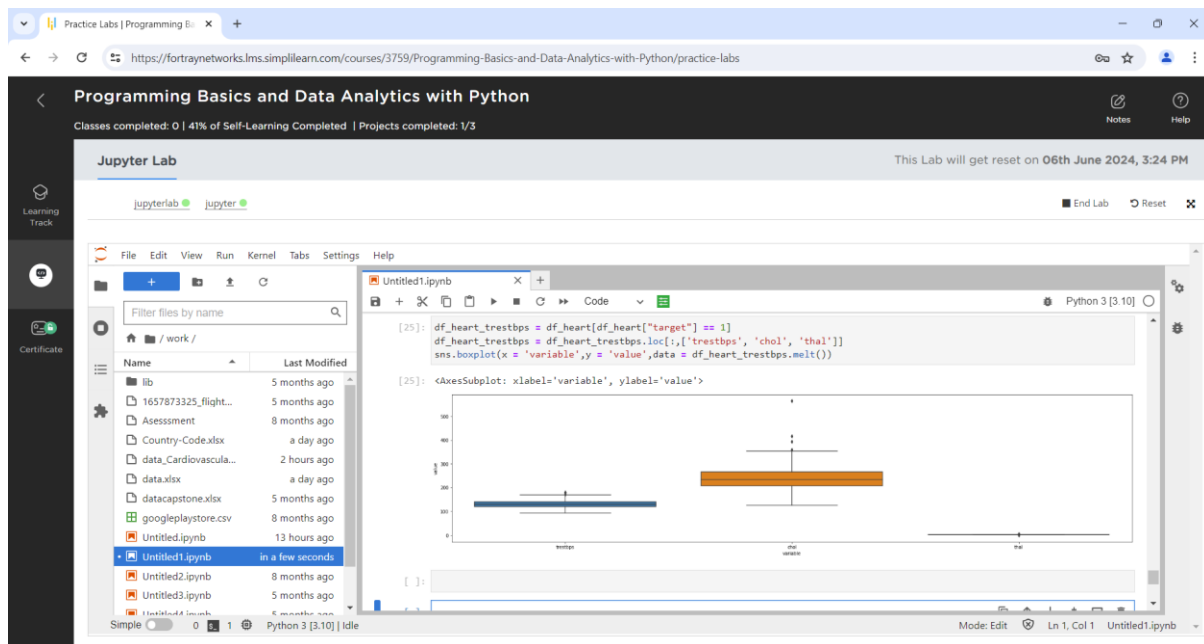
Let us check cardio vascular disease ratio by gender:



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.

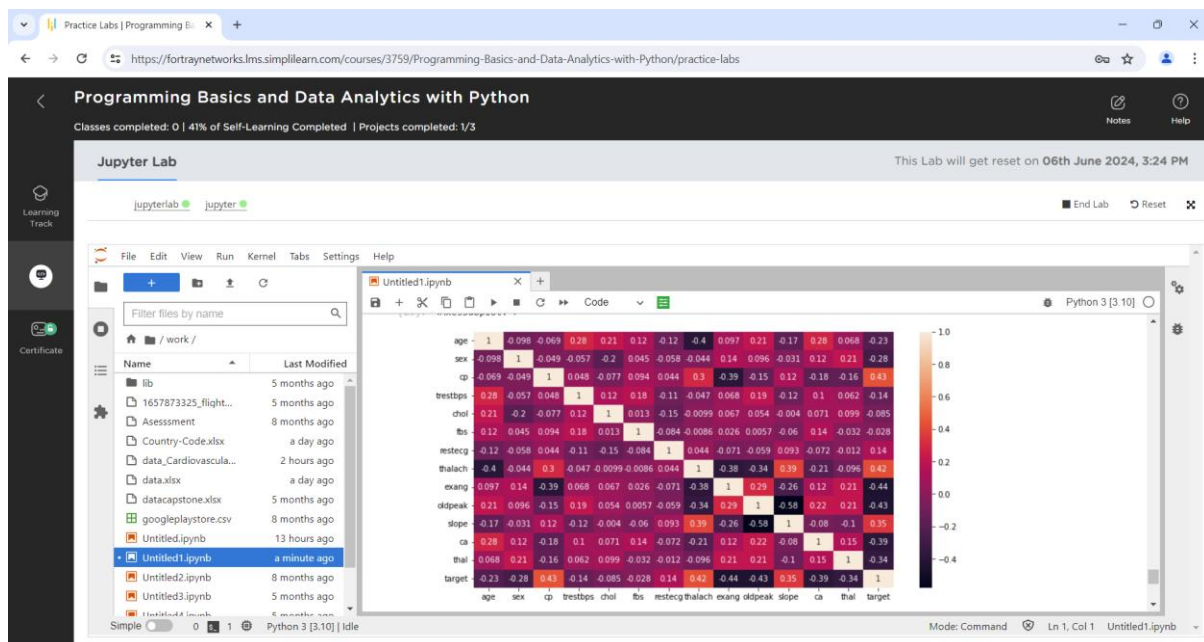
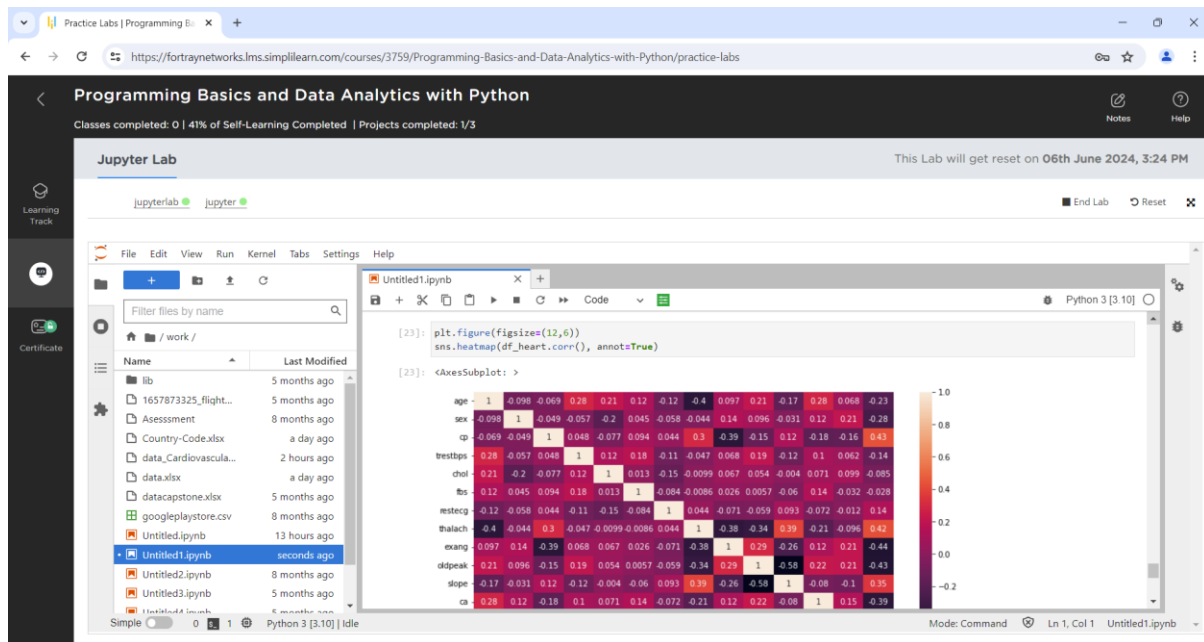
Is thalassemia a major cause of CVD?



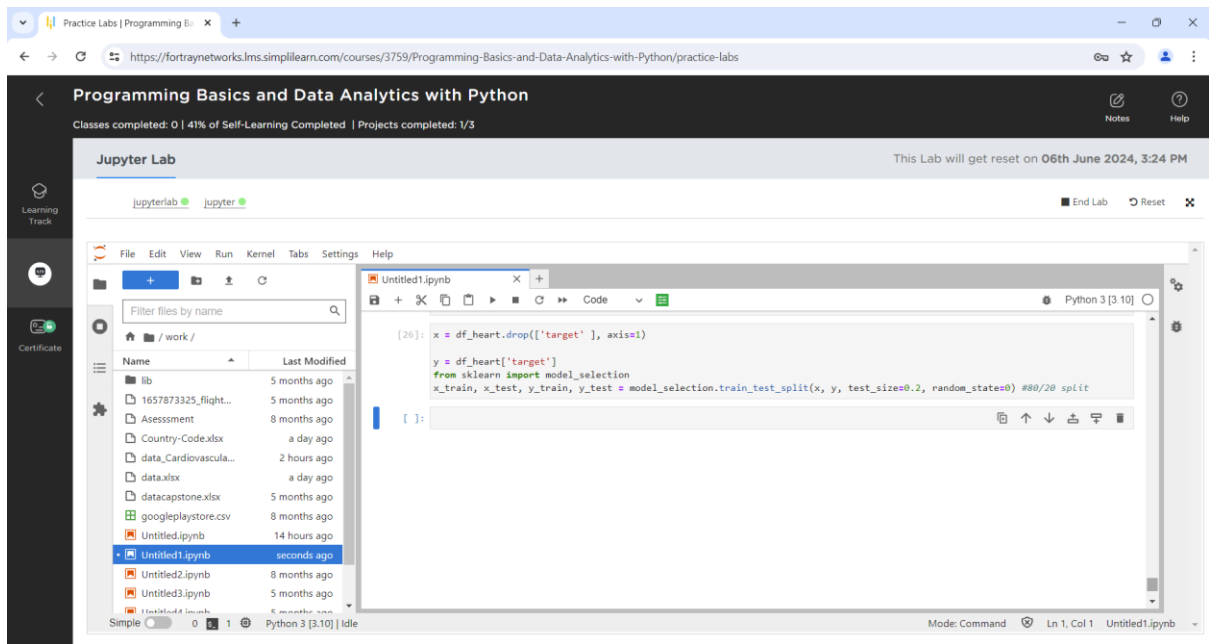
It can be concluded that high blood pressure, cholesterol and thalassemia are major causes of cardiovascular diseases.



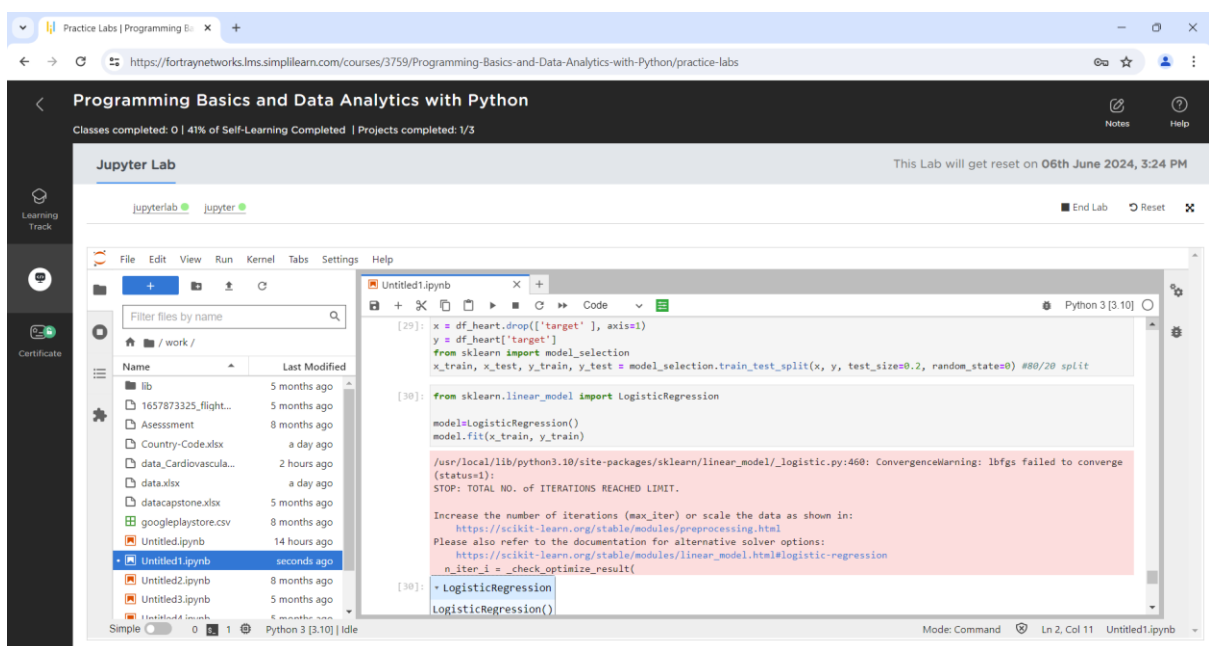
# Correlation matrix:

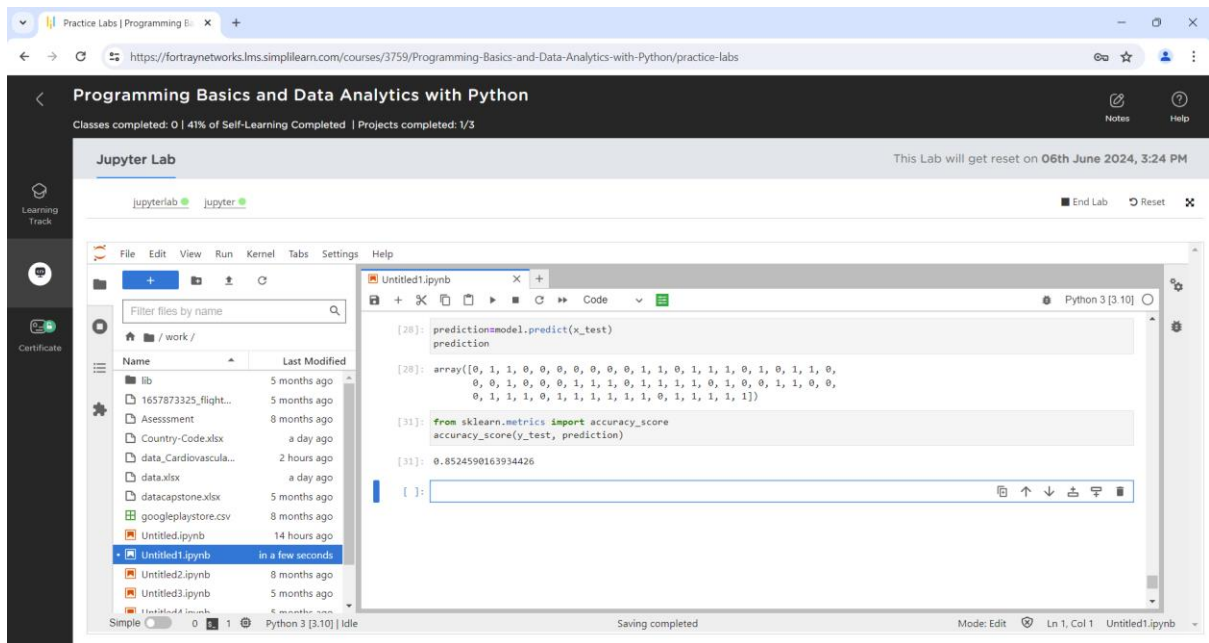
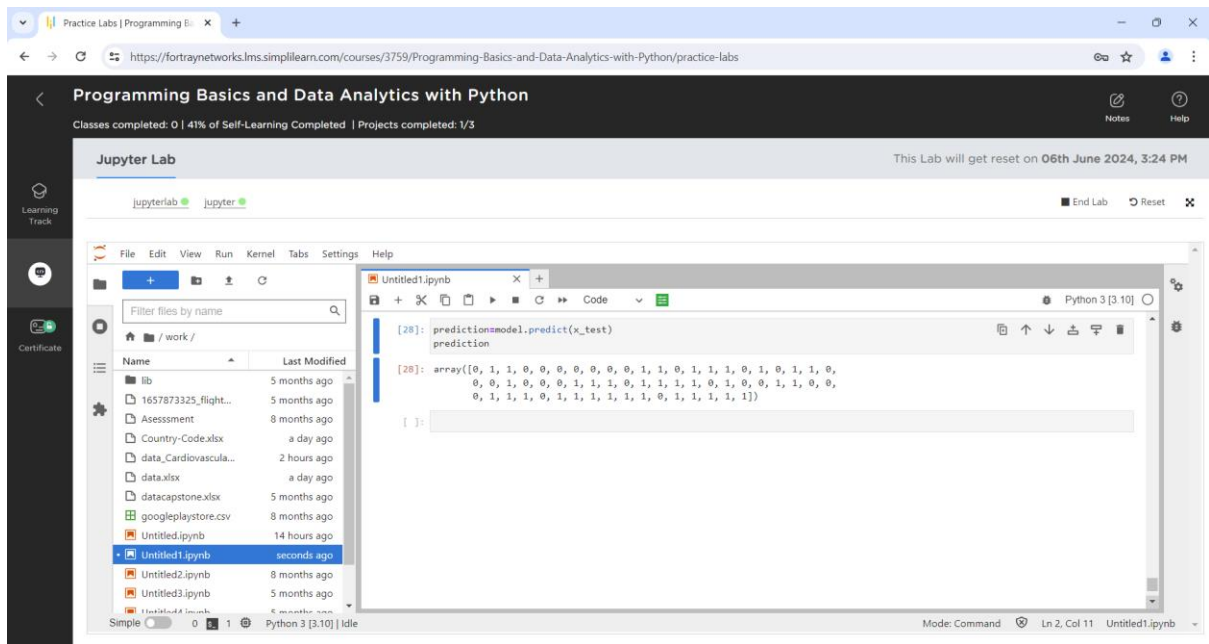


# Splitting the dataset into test and train:



## Logistics Regression:





Confusion matrix:

Practice Labs | Programming Basics and Data Analytics with Python

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

jupyterlab

jupyter

End Lab

Reset

File Edit View Run Kernel Tabs Settings Help

Filter files by name

lib

1657873325\_flight...

Assessment

Country-Code.xlsx

data\_Cardiovascula...

data.xlsx

datacapstone.xlsx

googleplaystore.csv

Untitled1.ipynb

Untitled1.ipynb

Untitled2.ipynb

Untitled3.ipynb

Untitled4.ipynb

Last Modified

5 months ago

5 months ago

8 months ago

a day ago

2 hours ago

a day ago

5 months ago

8 months ago

14 hours ago

seconds ago

8 months ago

5 months ago

6 months ago

Untitled1.ipynb

```
[32]: from sklearn.metrics import confusion_matrix

matrix= confusion_matrix(y_test, prediction)

sns.heatmap(matrix,annot = True, fet = "d")

[32]: <AxesSubplot: >
```

Simple

0

1

Python 3 [3.10] Idle

Mode: Command

Ln 1, Col 1

Untitled1.ipynb