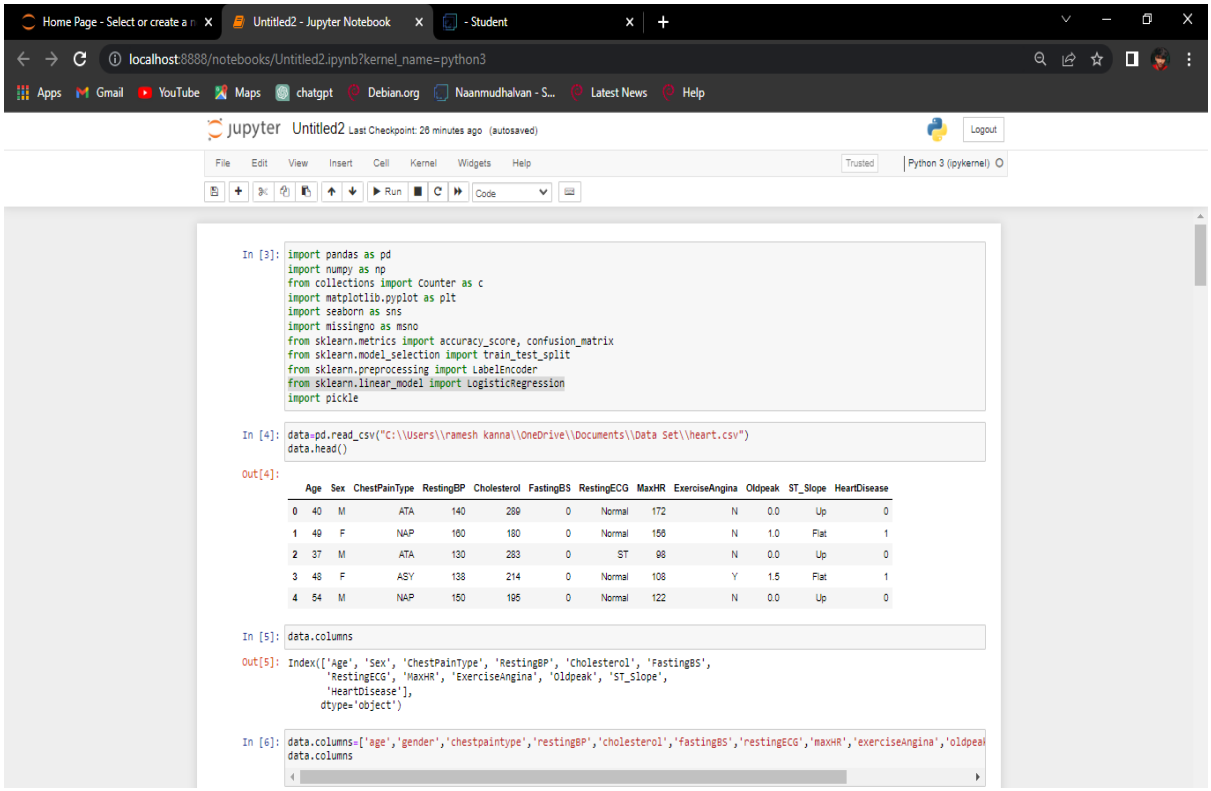
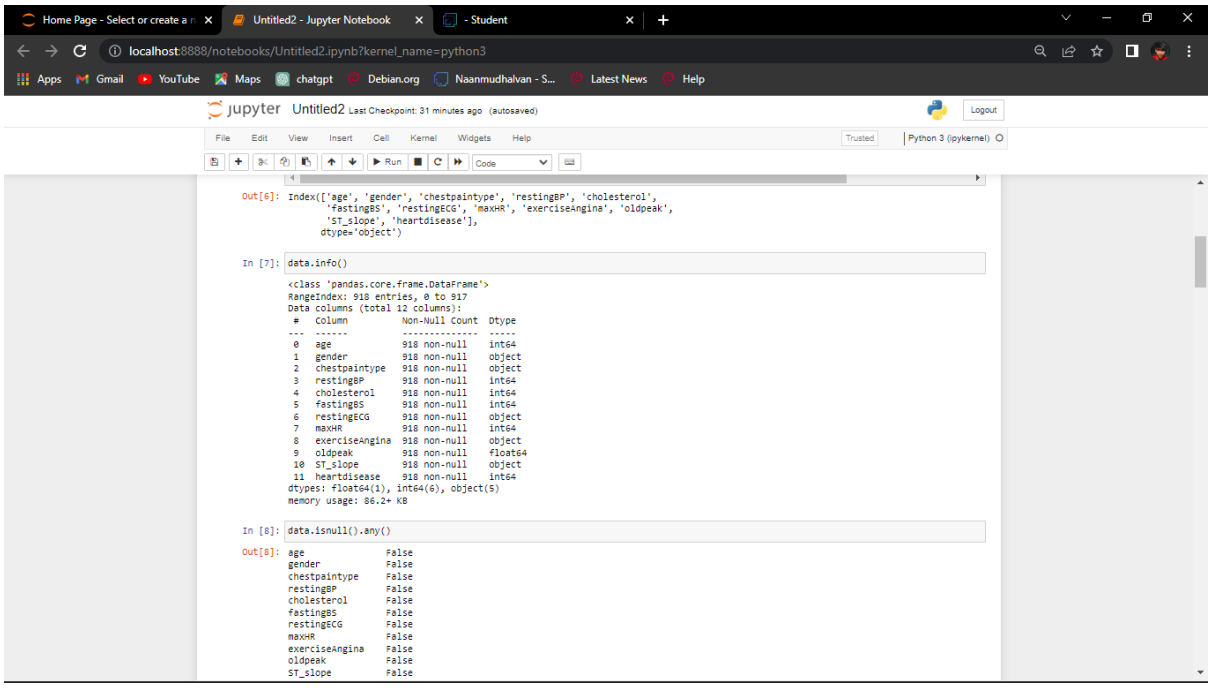
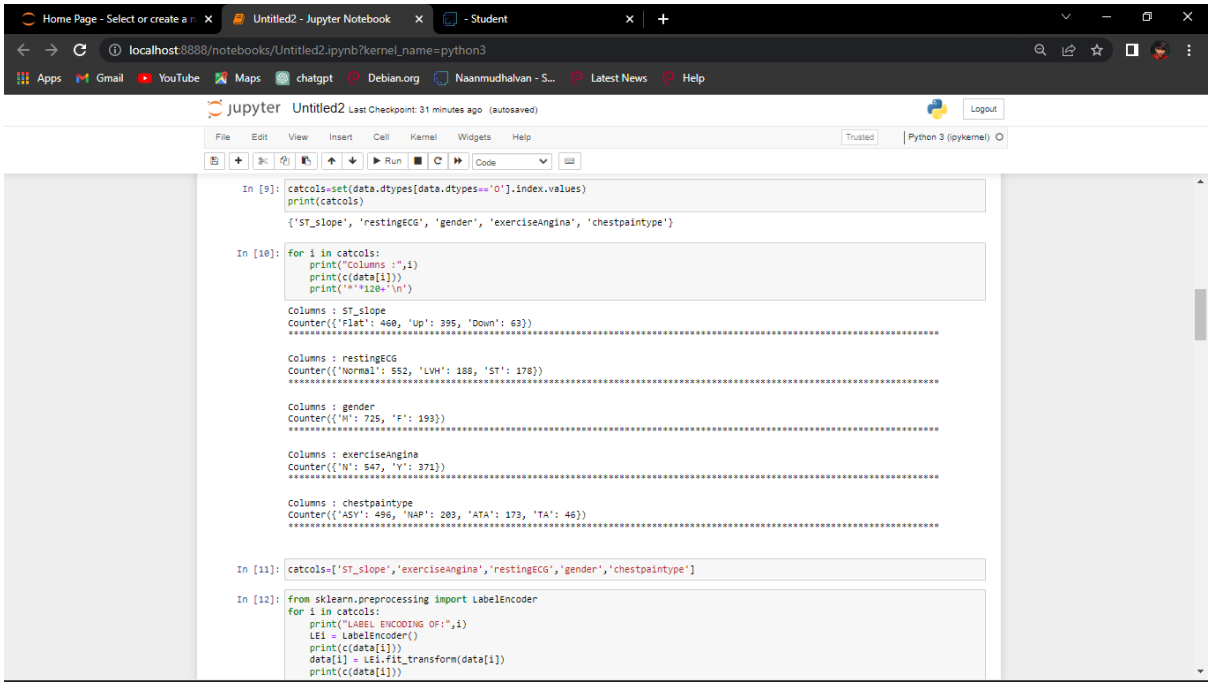
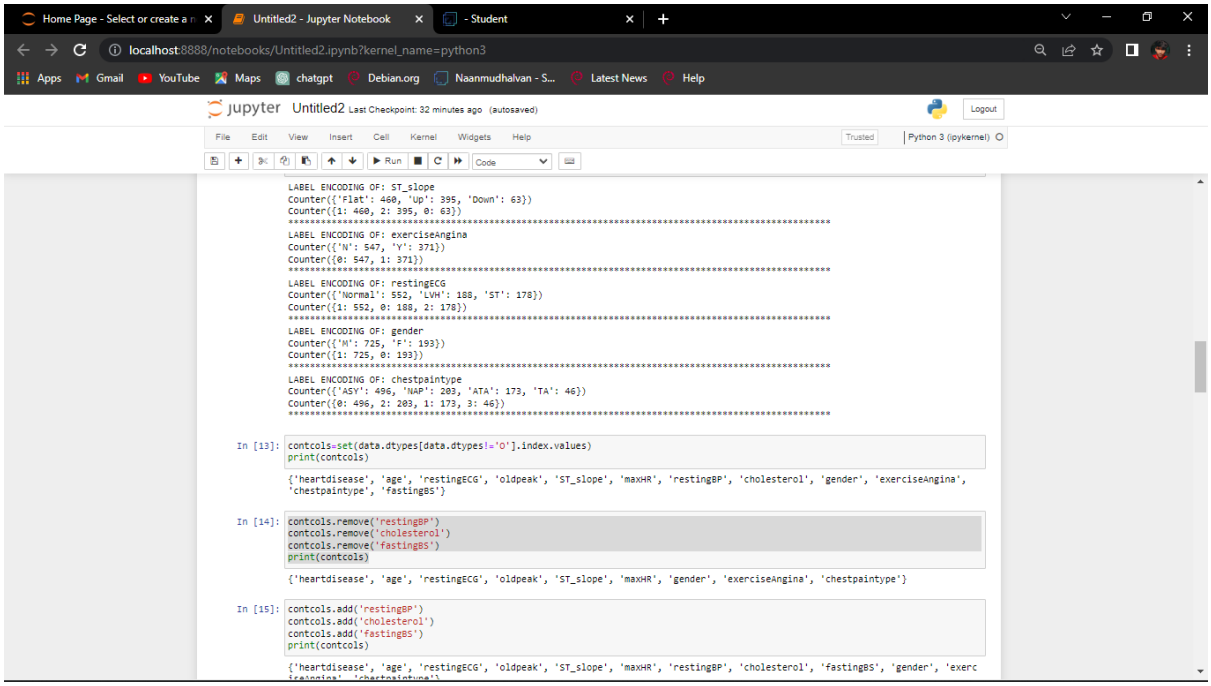

Home Page - Select or create a n
Untitled2 - Jupyter Notebook
- Student
localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3
Apps Gmail YouTube Maps chatgpt Debian.org Naanmudhalvan - S... Latest News Help
jupyter Untitled2 Last Checkpoint: 26 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [3]:
import pandas as pd
import numpy as np
from collections import Counter as c
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
import pickle
In [4]:
data=pd.read_csv("C:\\Users\\ramesh kanna\\OneDrive\\Documents\\Data Set\\heart.csv")
data.head()
Out[4]:
Age Sex ChestPainType RestingBP Cholesterol FastingBS RestingECG MaxHR ExerciseAngina Oldpeak ST_Slope HeartDisease
0 40 M ATA 140 289 0 Normal 172 N 0.0 Up 0
1 49 F NAP 160 180 0 Normal 156 N 1.0 Flat 1
2 37 M ATA 130 283 0 ST 98 N 0.0 Up 0
3 48 F ASY 138 214 0 Normal 108 Y 1.5 Flat 1
4 54 M NAP 150 195 0 Normal 122 N 0.0 Up 0
In [5]:
data.columns
Out[5]:
Index(['Age', 'Sex', 'ChestPainType', 'RestingBP', 'Cholesterol', 'FastingBS',
'RestingECG', 'MaxHR', 'ExerciseAngina', 'Oldpeak', 'ST_Slope',
'HeartDisease'],
dtype='object')
In [6]:
data.columns=['age','gender','chestpaintype','restingBP','cholesterol','fastingBS','restingECG','maxHR','exerciseAngina','oldpeak
data.columns


Home Page - Select or create a n
Untitled2 - Jupyter Notebook
- Student
localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3
Apps Gmail YouTube Maps chatgpt Debian.org Naanmudhalvan - S... Latest News Help
jupyter Untitled2 Last Checkpoint: 31 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Out[6]:
Index(['age', 'gender', 'chestpaintype', 'restingBP', 'cholesterol',
'fastingBS', 'restingECG', 'maxHR', 'exerciseAngina', 'oldpeak',
'ST_slope', 'heartdisease'],
dtype='object')
In [7]:
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 age 918 non-null int64
1 gender 918 non-null object
2 chestpaintype 918 non-null object
3 restingBP 918 non-null int64
4 cholesterol 918 non-null int64
5 fastingBS 918 non-null int64
6 restingECG 918 non-null object
7 maxHR 918 non-null int64
8 exerciseAngina 918 non-null object
9 oldpeak 918 non-null float64
10 ST_slope 918 non-null object
11 heartdisease 918 non-null int64
dtypes: float64(1), int64(6), object(5)
memory usage: 86.2+ KB
In [8]:
data.isnull().any()
Out[8]:
age False
gender False
chestpaintype False
restingBP False
cholesterol False
fastingBS False
restingECG False
maxHR False
exerciseAngina False
oldpeak False
ST_slope False


Home Page - Select or create a n
Untitled2 - Jupyter Notebook
- Student
localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3
Apps
Gmail
YouTube
Maps
chatgpt
Debian.org
Naanmudhalvan - S...
Latest News
Help
jupyter
Untitled2
Last Checkpoint: 31 minutes ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
In [9]:
catcols=set(data.dtypes[data.dtypes=='O'].index.values)
print(catcols)
{'ST_slope', 'restingECG', 'gender', 'exerciseAngina', 'chestpaintype'}
In [10]:
for i in catcols:
print("Columns :",i)
print(c(data[i]))
print('*'*120+'\n')
Columns : ST_slope
Counter({'Flat': 460, 'Up': 395, 'Down': 63})
Columns : restingECG
Counter({'Normal': 552, 'LVH': 188, 'ST': 178})
Columns : gender
Counter({'M': 725, 'F': 193})
Columns : exerciseAngina
Counter({'N': 547, 'Y': 371})
Columns : chestpaintype
Counter({'ASY': 496, 'NAP': 203, 'ATA': 173, 'TA': 46})
In [11]:
catcols=['ST_slope','exerciseAngina','restingECG','gender','chestpaintype']
In [12]:
from sklearn.preprocessing import LabelEncoder
for i in catcols:
print("LABEL ENCODING OF:",i)
LEi = LabelEncoder()
print(c(data[i]))
data[i] = LEi.fit_transform(data[i])
print(c(data[i]))


Home Page - Select or create a n
Untitled2 - Jupyter Notebook
- Student
localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3
Apps
Gmail
YouTube
Maps
chatgpt
Debian.org
Naanmudhalvan - S...
Latest News
Help
jupyter
Untitled2
Last Checkpoint: 32 minutes ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
LABEL ENCODING OF: ST_slope
Counter({'Flat': 460, 'Up': 395, 'Down': 63})
Counter({1: 460, 2: 395, 0: 63})
LABEL ENCODING OF: exerciseAngina
Counter({'N': 547, 'Y': 371})
Counter({0: 547, 1: 371})
LABEL ENCODING OF: restingECG
Counter({'Normal': 552, 'LVH': 188, 'ST': 178})
Counter({1: 552, 0: 188, 2: 178})
LABEL ENCODING OF: gender
Counter({'M': 725, 'F': 193})
Counter({1: 725, 0: 193})
LABEL ENCODING OF: chestpaintype
Counter({'ASY': 496, 'NAP': 203, 'ATA': 173, 'TA': 46})
Counter({0: 496, 2: 203, 1: 173, 3: 46})
In [13]:
contcols=set(data.dtypes[data.dtypes!='O'].index.values)
print(contcols)
{'heartdisease', 'age', 'restingECG', 'oldpeak', 'ST_slope', 'maxHR', 'restingBP', 'cholesterol', 'gender', 'exerciseAngina',
'chestpaintype', 'fastingBS'}
In [14]:
contcols.remove('restingBP')
contcols.remove('cholesterol')
contcols.remove('fastingBS')
print(contcols)
{'heartdisease', 'age', 'restingECG', 'oldpeak', 'ST_slope', 'maxHR', 'gender', 'exerciseAngina', 'chestpaintype'}
In [15]:
contcols.add('restingBP')
contcols.add('cholesterol')
contcols.add('fastingBS')
print(contcols)
{'heartdisease', 'age', 'restingECG', 'oldpeak', 'ST_slope', 'maxHR', 'restingBP', 'cholesterol', 'fastingBS', 'gender', 'exerc

Home Page - Select or create a n | Untitled2 - Jupyter Notebook | - Student

localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3

Apps | Gmail | YouTube | Maps | chatgpt | Debian.org | Naanmudhalvan - S... | Latest News | Help

jupyter Untitled2 Last Checkpoint: 32 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [15]:
```
contcols.add('restingBP')
contcols.add('cholesterol')
contcols.add('fastingBS')
print(contcols)
```

```
{'heartdisease', 'age', 'restingECG', 'oldpeak', 'ST_slope', 'maxHR', 'restingBP', 'cholesterol', 'fastingBS', 'gender', 'exerciseAngina', 'chestpaintype'}
```

In [16]:
```
data['exerciseAngina'] = data.exerciseAngina.replace('\to','N')
c(data['exerciseAngina'])
```

Out[16]:
```
Counter({0: 547, 1: 371})
```

In [17]:
```
data.describe()
```

Out[17]:

	age	gender	chestpaintype	restingBP	cholesterol	fastingBS	restingECG	maxHR	exerciseAngina	oldpeak	ST_slope	heartdisease
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000
mean	53.510893	0.789760	0.781046	132.396514	198.799564	0.233115	0.699107	136.809368	0.404139	0.887364	1.361656	0.553377
std	9.432617	0.407701	0.956519	18.514154	109.384145	0.423046	0.631671	25.460334	0.490992	1.066570	0.607056	0.497414
min	28.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	60.000000	0.000000	-2.600000	0.000000	0.000000
25%	47.000000	1.000000	0.000000	120.000000	173.250000	0.000000	1.000000	120.000000	0.000000	0.000000	1.000000	0.000000
50%	54.000000	1.000000	0.000000	130.000000	223.000000	0.000000	1.000000	138.000000	0.000000	0.600000	1.000000	1.000000
75%	60.000000	1.000000	2.000000	140.000000	267.000000	0.000000	1.000000	156.000000	1.000000	1.500000	2.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	603.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	1.000000

In [18]:
```
sns.distplot(data.age)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[18]:
```
<AxesSubplot:xlabel='age', ylabel='Density'>
```

Home Page - Select or create a n | Untitled2 - Jupyter Notebook | - Student

localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3

Apps | Gmail | YouTube | Maps | chatgpt | Debian.org | Naanmudhalvan - S... | Latest News | Help

jupyter Untitled2 Last Checkpoint: 33 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

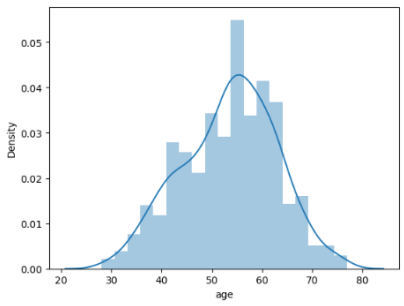
Trusted | Python 3 (ipykernel)

Run | Code

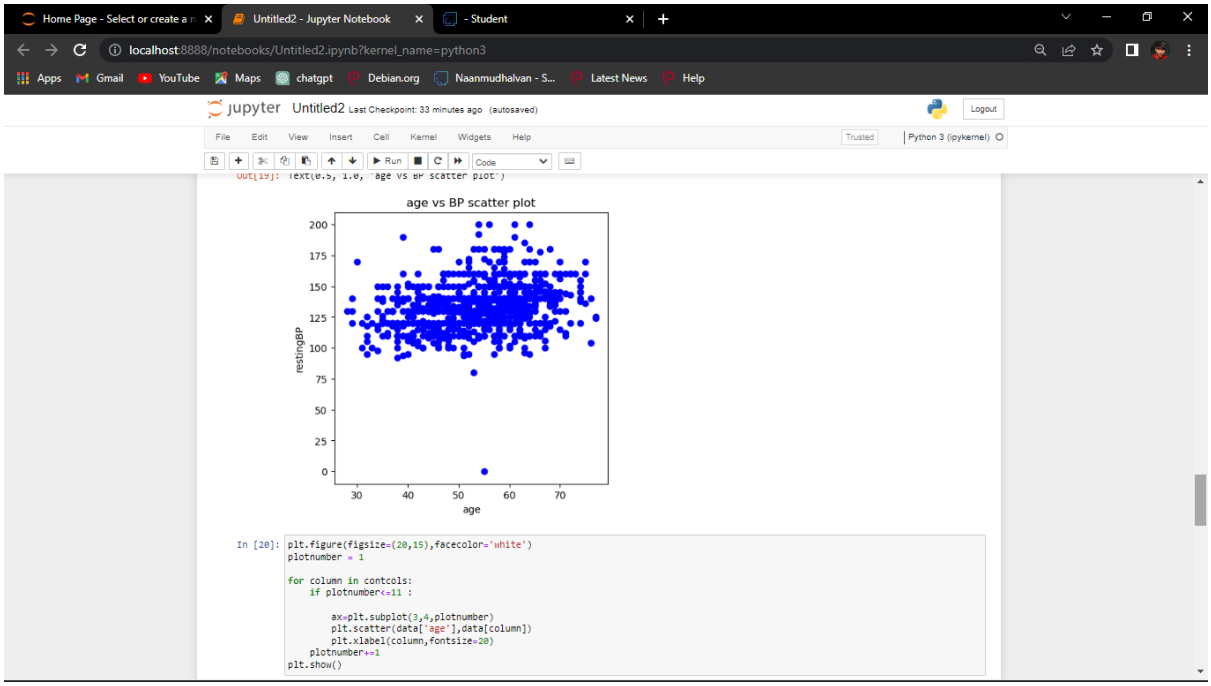
In [18]:
```
sns.distplot(data.age)
```

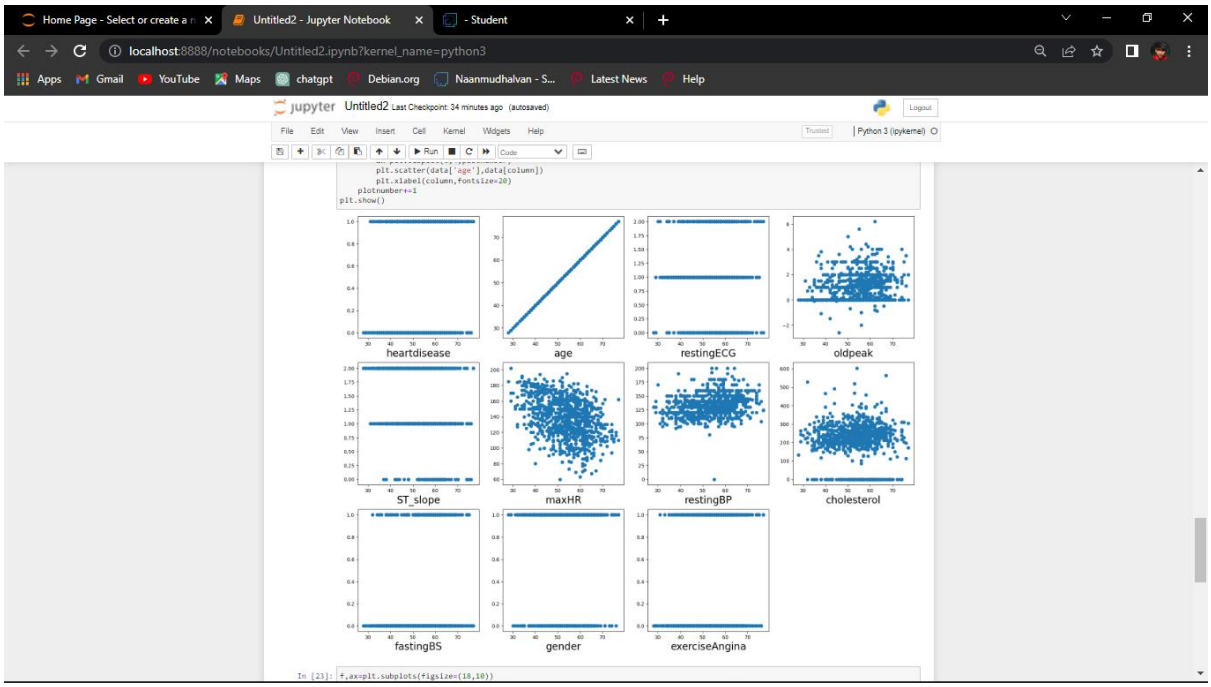
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[18]:
```
<AxesSubplot:xlabel='age', ylabel='Density'>
```



In [19]:
```
import matplotlib.pyplot as plt
fig=plt.figure(figsize=(5,5))
plt.scatter(data['age'],data['restingBP'],color='blue')
plt.xlabel('age')
plt.ylabel('restingBP')
plt.title("age vs BP scatter plot")
```


Home Page - Select or create a n
Untitled2 - Jupyter Notebook
- Student
localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3
Apps
Gmail
YouTube
Maps
chatgpt
Debian.org
Naanmudhalvan - S...
Latest News
Help
jupyter
Untitled2
Last Checkpoint: 33 minutes ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
Out[19]: Text(0.5, 1.0, 'age vs BP scatter plot')
age vs BP scatter plot
restingBP
age
In [20]:
plt.figure(figsize=(20,15),facecolor='white')
plotnumber = 1
for column in contcols:
if plotnumber<=11 :
ax=plt.subplot(3,4,plotnumber)
plt.scatter(data['age'],data[column])
plt.xlabel(column,fontsize=20)
plotnumber+=1
plt.show()


Home Page - Select or create a n
Untitled2 - Jupyter Notebook
- Student
localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3
Apps
Gmail
YouTube
Maps
chatgpt
Debian.org
Naanmudhalvan - S...
Latest News
Help
jupyter
Untitled2
Last Checkpoint: 34 minutes ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
plt.scatter(data['age'],data[column])
plt.xlabel(column,fontsize=20)
plotnumber+=1
plt.show()
heartdisease
age
restingECG
oldpeak
ST_slope
maxHR
restingBP
cholesterol
fastingBS
gender
exerciseAngina
In [21]:
f,ax=plt.subplots(figsize=(18,10))

Home Page - Select or create a n | Untitled2 - Jupyter Notebook | - Student

localhost:8888/notebooks/Untitled2.ipynb?kernel_name=python3

Apps Gmail YouTube Maps chatgpt Debian.org Naanmudhalvan - S... Latest News Help

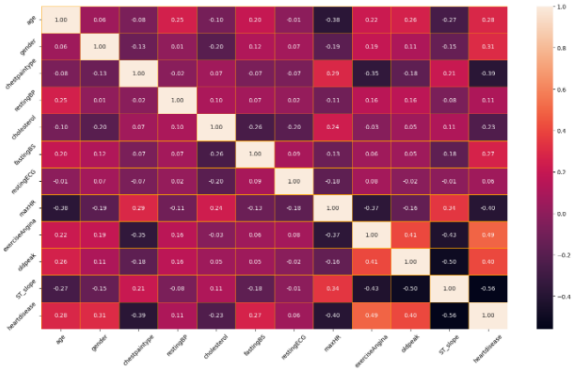
jupyter Untitled2 Last Checkpoint: 34 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

fastingBS gender exerciseAngina

In [23]:
```
f,ax=plt.subplots(figsize=(18,18))
sns.heatmap(data.corr(),annot=True,fmt=".2f",ax=ax,linewidths=0.5,linecolor="orange")
plt.xticks(rotation=45)
plt.yticks(rotation=45)
plt.show()
```



In [29]:
```
from sklearn.linear_model import LogisticRegression
lgr = LogisticRegression()
lgr.fit(X_train,Y_train)
```