

```
File Edit Selection View Go Run Terminal Help
Practical6.ipynb - ML Lab - Visual Studio Code

Practical6.ipynb x
Practical6.ipynb > plt.figure(figsize=(10,6))
+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline ... Python 3.10.0 64-bit

[27] ✓ 2.3s
from sklearn.preprocessing import StandardScaler

[28] ✓ 0.1s
scaler = StandardScaler()

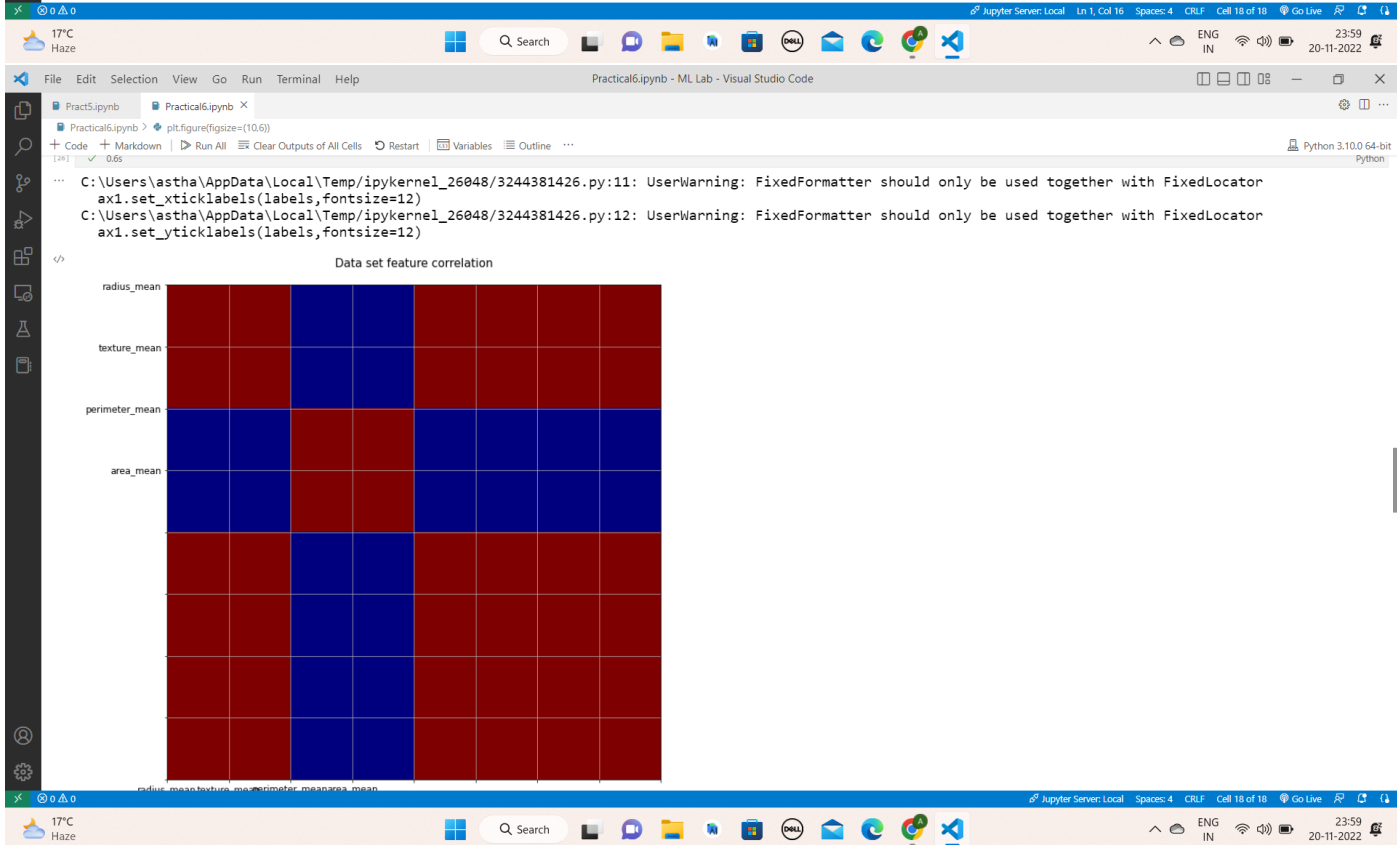
[29] ✓ 0.1s
x = df.drop('radius_mean',axis=1)
y = df['radius_mean']

[30] ✓ 0.1s
x = scaler.fit_transform(x)

[31] ✓ 0.9s
dfnew = pd.DataFrame(data=x,columns=df.columns[1:])

[32] ✓ 0.1s
dfnew.describe()
...
```

	texture_mean	perimeter_mean	area_mean
count	5.690000e+02	5.690000e+02	5.690000e+02
mean	6.868164e-17	-1.248757e-16	-2.185325e-16
std	1.000880e+00	1.000880e+00	1.000880e+00
min	-2.229249e+00	-1.984504e+00	-1.454443e+00
25%	-7.259631e-01	-6.919555e-01	-6.671955e-01
50%	-1.046362e-01	-2.359800e-01	-2.951869e-01
75%	5.841756e-01	4.996769e-01	3.635073e-01



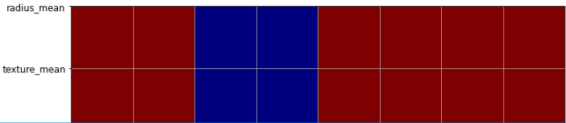
Practical6.ipynb - ML Lab - Visual Studio Code

```
def correlation_matrix(df):
    from matplotlib import pyplot as plt
    from matplotlib import cm as cm
    fig = plt.figure(figsize=(16,12))
    ax1 = fig.add_subplot(111)
    cmap = cm.get_cmap('jet',30)
    cax = ax1.imshow(df.corr(),interpolation="nearest",cmap = cmap)
    ax1.grid(True)
    plt.title('Data set feature correlation\n',fontsize=15)
    labels=df.columns
    ax1.set_xticklabels(labels,fontsize=12)
    ax1.set_yticklabels(labels,fontsize=12)
    #fig.colorbar(cax,ticks=[0.1*i for i in range(-11,11)])
    plt.show()

correlation_matrix(df)
```

C:\Users\astha\AppData\Local\Temp\ipykernel\_26048\3244381426.py:11: UserWarning: FixedFormatter should only be used together with FixedLocator  
ax1.set\_xticklabels(labels,fontsize=12)  
C:\Users\astha\AppData\Local\Temp\ipykernel\_26048\3244381426.py:12: UserWarning: FixedFormatter should only be used together with FixedLocator  
ax1.set\_yticklabels(labels,fontsize=12)

Data set feature correlation

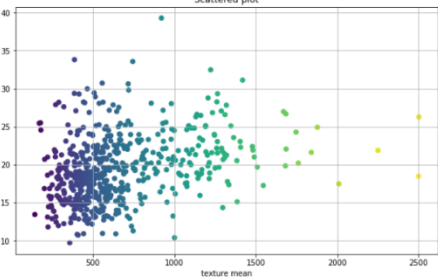


17°C  
Haze

Practical6.ipynb - ML Lab - Visual Studio Code

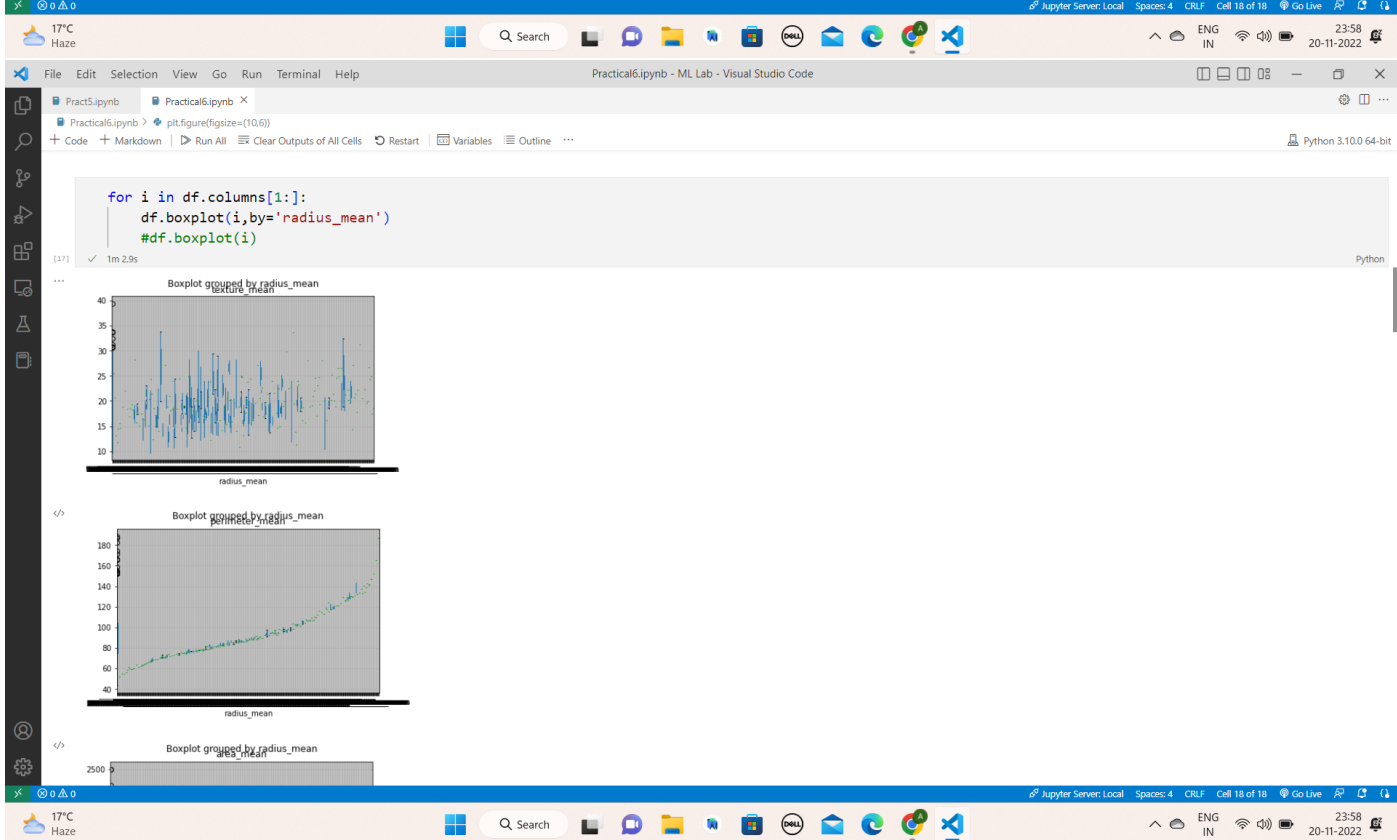
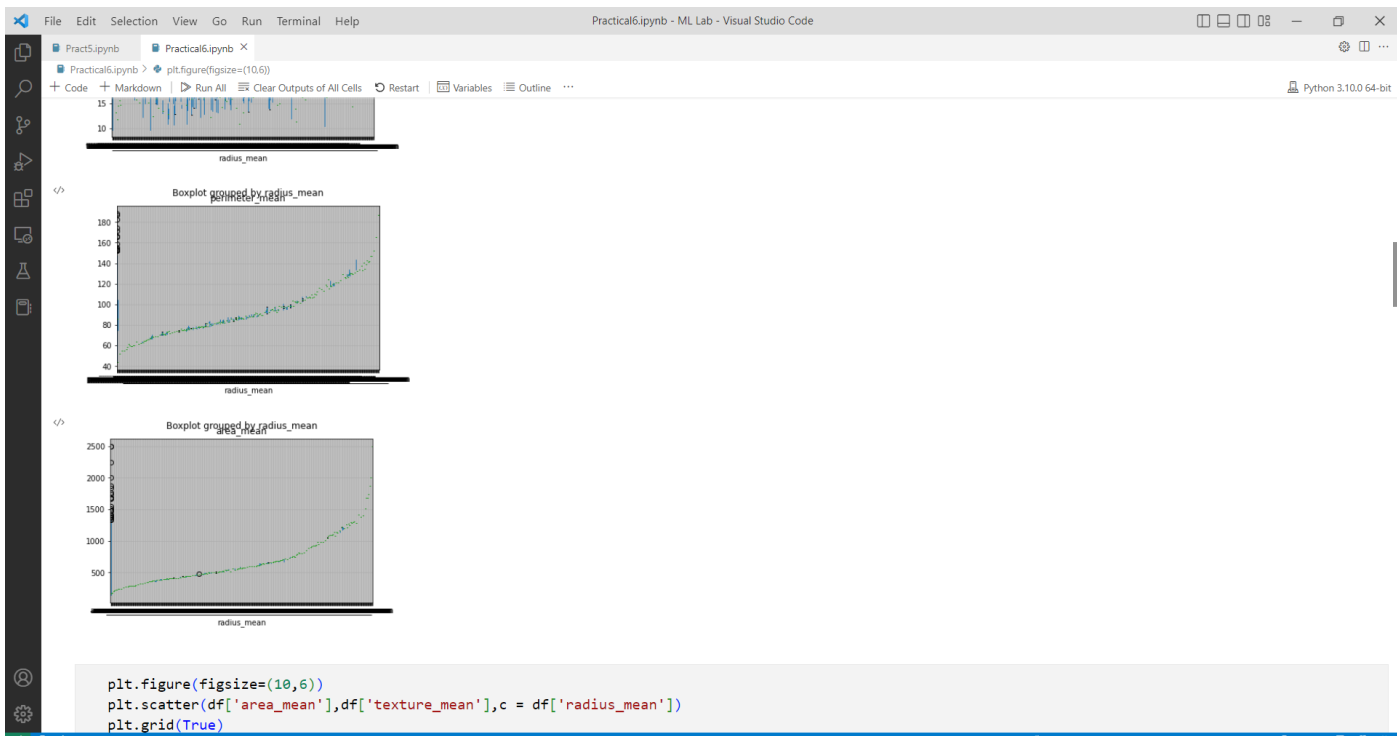
```
plt.figure(figsize=(10,6))
plt.scatter(df['area_mean'],df['texture_mean'],c = df['radius_mean'])
plt.grid(True)
plt.title("Scattered plot")
plt.ylabel("area mean")
plt.xlabel("texture mean")
plt.show()
```

Scattered plot



```
def correlation_matrix(df):
    from matplotlib import pyplot as plt
    from matplotlib import cm as cm
    fig = plt.figure(figsize=(16,12))
```

17°C  
Haze



File Edit Selection View Go Run Terminal Help

Practical6.ipynb - ML Lab - Visual Studio Code

Python 3.10.0 64-bit

Practical6.ipynb

Practical6.ipynb

Practical6.ipynb > plt.figure(figsize=(10,6))

+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline

df.iloc[:,1:].describe()

Python

8 rows x 30 columns

	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean
count	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
mean	14.127292	19.289649	91.969033	654.889104	0.096360	0.104341	0.088799	0.048919	0.181162	0.062798
std	3.524049	4.301036	24.298981	351.914129	0.014064	0.052813	0.079720	0.038803	0.027414	0.007060
min	6.981000	9.710000	43.790000	143.500000	0.052630	0.019380	0.000000	0.000000	0.106000	0.049960
25%	11.700000	16.170000	75.170000	420.300000	0.086370	0.064920	0.029560	0.020310	0.161900	0.057700
50%	13.370000	18.840000	86.240000	551.100000	0.095870	0.092630	0.061540	0.033500	0.179200	0.061540
75%	15.780000	21.800000	104.100000	782.700000	0.105300	0.130400	0.130700	0.074000	0.195700	0.066120
max	28.110000	39.280000	188.500000	2501.000000	0.163400	0.345400	0.426800	0.201200	0.304000	0.097440

df = df[['radius\_mean', 'texture\_mean', 'perimeter\_mean', 'area\_mean']]

df.head()

Python

5 rows x 4 columns

	radius_mean	texture_mean	perimeter_mean	area_mean
0	17.99	10.38	122.80	1001.0
1	20.57	17.77	132.90	1326.0
2	19.69	21.25	130.00	1203.0
3	11.42	20.38	77.58	386.1
4	20.29	14.34	135.10	1297.0

17°C  
Haze

Search

Jupyter Server: Local Spaces: 4 CRLF Cell: 18 of 18 Go Live

23:58  
20-11-2022

File Edit Selection View Go Run Terminal Help

Practical6.ipynb - ML Lab - Visual Studio Code

Python 3.10.0 64-bit

Practical6.ipynb

Practical6.ipynb

Practical6.ipynb > plt.figure(figsize=(10,6))

+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

Python

0.9s

df = pd.read\_csv("breast-cancer.csv")

Python

0.1s

df.head()

Python

5 rows x 32 columns

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...	radius_worst	texture
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...	25.38	
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...	24.99	
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	...	23.57	
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...	14.91	
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...	22.54	

df.iloc[:,1:].describe()

Python

0.2s

radius\_mean texture\_mean perimeter\_mean area\_mean smoothness\_mean compactness\_mean concavity\_mean concave points\_mean symmetry\_mean fractal\_dimension\_mean

17°C  
Haze

Search

Jupyter Server: Local Spaces: 4 CRLF Cell: 18 of 18 Go Live

23:58  
20-11-2022

