

HandsON

Load any data set

```
import numpy as np
import pandas as pd

# Generate sample data
np.random.seed(0)#Configuration for random number
data = pd.DataFrame(np.random.randn(100, 2), columns=['X', 'Y'])#Create a
#two features data, here naming X and Y of size 100 samples
```

```
[ ] data.head(5)# display the geneted random data
```



	X	Y
0	1.764052	0.400157
1	0.978738	2.240893

Finding covariance



```
# Calculate covariance matrix  
#cov_matrix = np.cov(data['X'], data['Y'])#numpy  
cov_matrix=data.corr()#pandas  
print(cov_matrix)
```



	X	Y
X	1.000000	-0.024005
Y	-0.024005	1.000000

Finding correlation

```
# Calculate correlation matrix  
#cor_matrix = np.corrcoef(data['X'], data['Y'])#numpy  
cor_matrix=data.corr()  
print(cor_matrix)
```

```
↔  
      X      Y  
X  1.000000 -0.024005  
Y -0.024005  1.000000
```

```
[ ]
```

Handling outlier

```
▶ import matplotlib.pyplot as plt
import seaborn as sns
def create_box_plot(data):
    plt.figure(figsize=(8, 6))
    sns.boxplot(data=data)
    plt.title('Box Plot')
    plt.xlabel('Data')
    plt.ylabel('Values')
    plt.show()
# Example data
data = {
    'Group A': [10, 12, 15, 18, 20, 22], #add -1 and see outlier
    'Group B': [11, 14, 17, 20, 23, 26],
    'Group C': [8, 10, 12, 14, 16, 18]
}
data1=pd.DataFrame(data)
create_box_plot(data1)
```

Handling outlier

