

CS-23334 FUNDAMENTALS OF DATA SCIENCE

ABENANTHAN P 240701005

Experiment 4

Date: 21.08.2025

4. Experiment to detect outliers in a given data set.

Aim:

To Understand the procedure to identify the outliers in a given dataset

Description:

Identify outliers in dataset using pandas library in python

Algorithm:

Step 1: Data Collection and Preprocessing

Step 2: Statistical Methods for Outlier Detection

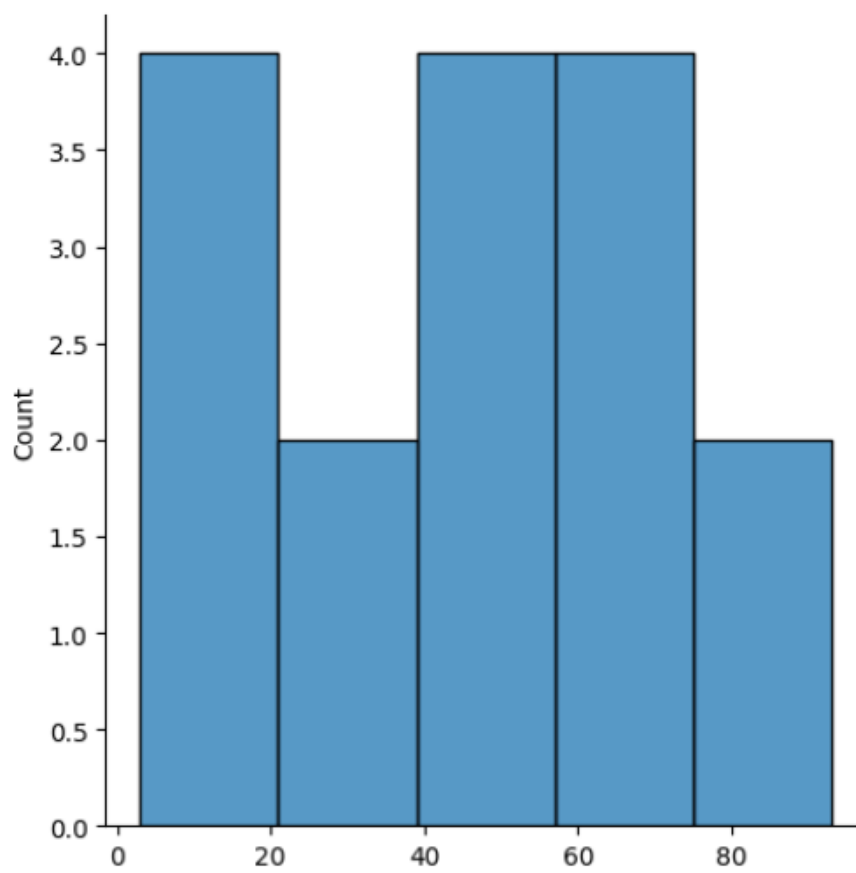
Step 3: Visualization of Outliers

Step 4: Interpretation and Handling of Outliers

Code With Output:

```
import numpy as np
array=np.random.randint(1,100,16)
print(array)
array.mean()
[51 37  3 13 50  3 71 62 46 23 39 12 59 93 73 85]
np.float64(45.0)
np.percentile(array,25)
np.float64(20.5)
np.percentile(array,50)
np.float64(48.0)
np.percentile(array,75)
np.float64(64.25)
np.percentile(array,100)
np.float64(93.0)
def outDetection(array):
    sorted(array)
    Q1,Q3=np.percentile(array,[25,75])
    IQR=Q3-Q1
    lr=Q1-(1.5*IQR)
    ur=Q3+(1.5*IQR)
    return lr,ur
lr,ur=outDetection(array)
lr,ur
(np.float64(-45.125), np.float64(129.875))
```

```
import seaborn as sns
%matplotlib inline
sns.displot(array)
<seaborn.axisgrid.FacetGrid at 0x78ecde4606b0>
```



```
sns.distplot(array)
```

```
/tmp/ipython-input-1133588802.py:1: UserWarning:
```

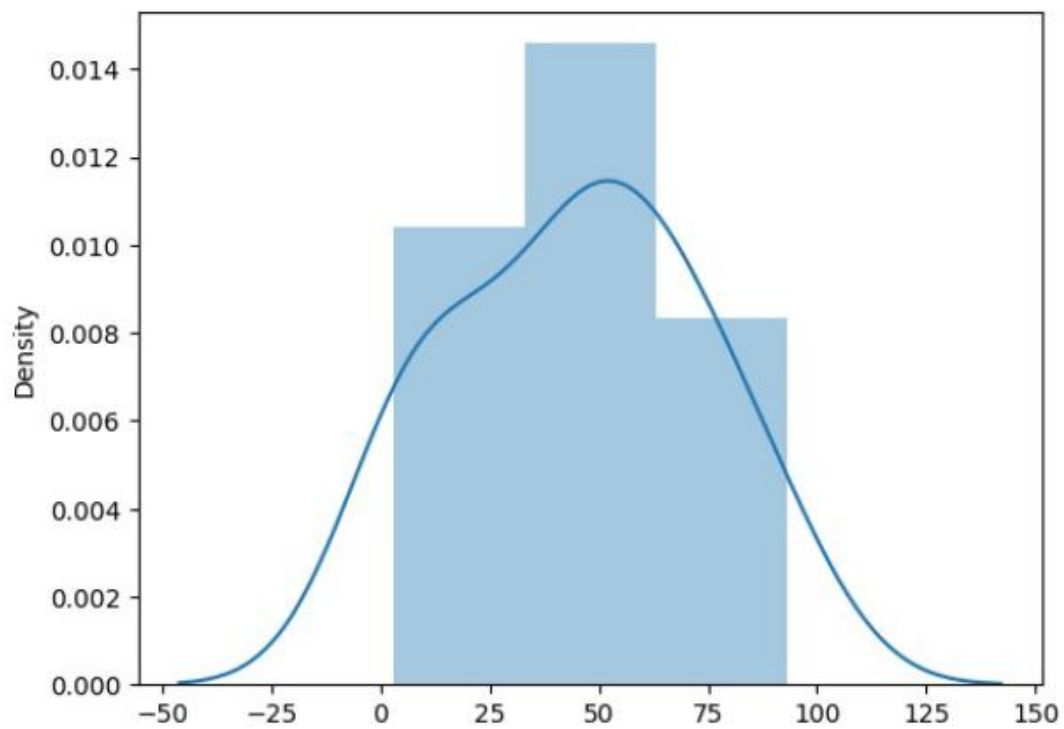
```
`distplot` is a deprecated function and will be removed in seaborn  
v0.14.0.
```

```
Please adapt your code to use either `displot` (a figure-level  
function with  
similar flexibility) or `histplot` (an axes-level function for  
histograms).
```

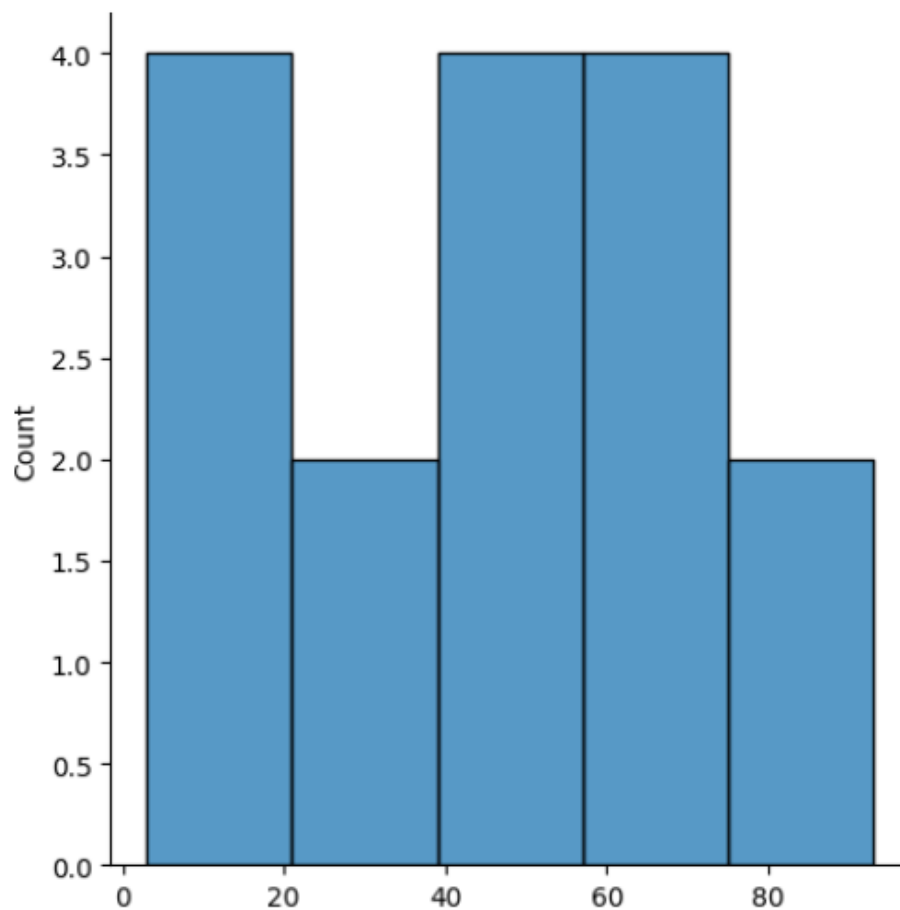
```
For a guide to updating your code to use the new functions, please see  
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
sns.distplot(array)
```

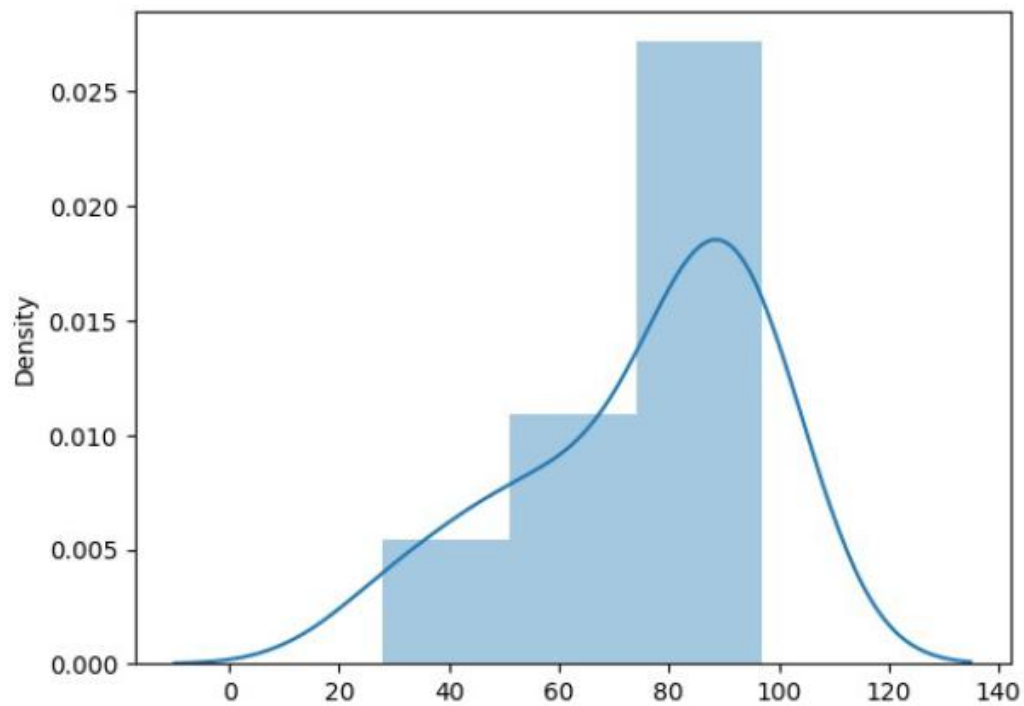
```
<Axes: ylabel='Density'>
```



```
new_array=array[(array>lr) & (array<ur)]
new_array
array([51, 37,  3, 13, 50,  3, 71, 62, 46, 23, 39, 12, 59, 93, 73,
      85])
sns.displot(new_array)
<seaborn.axisgrid.FacetGrid at 0x78ecde22e4e0>
```



```
lr1,url=outDetection(new_array)
lr1,url
(np.float64(-45.125), np.float64(129.875))
sns.distplot(final_array)
/tmp/ipython-input-209491988.py:1: UserWarning:
`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.
Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).
For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
sns.distplot(final_array)
<Axes: ylabel='Density'>
```

**Result:**

Thus the experiment to detect outliers in dataset is completed using python.