CS23334-FUNDAMENTALS OF DATA SCIENCE

DEVA DHARSHINI P(240701107)

4.) OUTLIER DETECTION FOR GIVEN SETS

Aim:

To detect and identify outliers in a given dataset using statistical methods such as the Interquartile Range (IQR) or Z-score technique.

Code:

Aim:

```
import numpy as np
array=np.random.randint(1,100,16)
array
```

```
array([45, 21, 96, 15, 36, 63, 79, 97, 11,  8, 92,  2, 89, 87, 84, 72])
```

```
array.mean()
```

```
56.0625
```

```
np.percentile(array,25)
```

```
19.5
```

```
np.percentile(array,50)
```

```
67.5
```

```
np.percentile(array,75)
```

```
87.5
```

```
np.percentile(array,100)
```

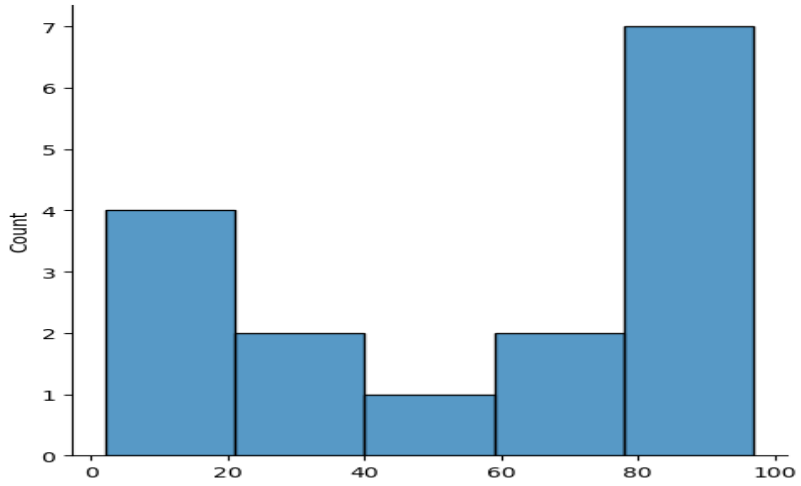
```
97.0
```

```
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
```

```
(-82.5, 189.5)
```

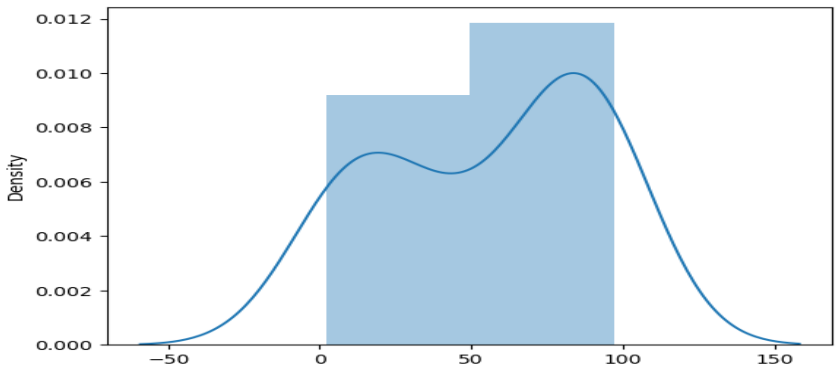
```
import seaborn as sns
%matplotlib inline
sns.displot(array)
```

```
<seaborn.axisgrid.FacetGrid at 0x1c5af76f340>
```




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

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

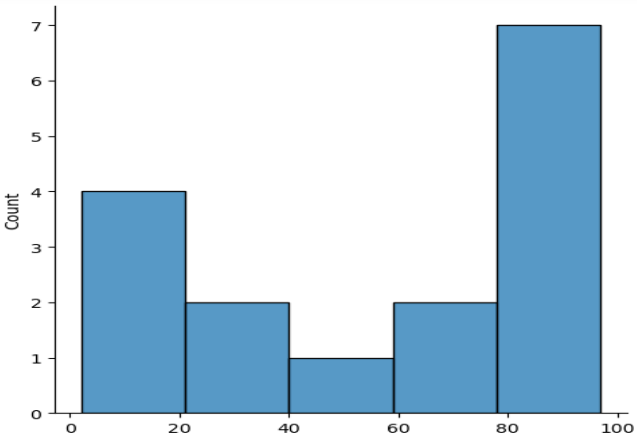



```
new_array=array[(array>lr) & (array<ur)]
new_array
```

```
array([45, 21, 96, 15, 36, 63, 79, 97, 11,  8, 92,  2, 89, 87, 84, 72])
```

```
sns.displot(new_array)
```

```
<seaborn.axisgrid.FacetGrid at 0x1c5af76f2e0>
```




```
lr1,ur1=outDetection(new_array)
lr1,ur1
```

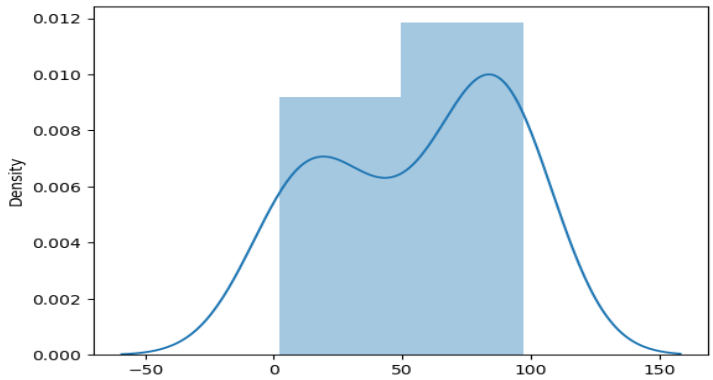
```
(-82.5, 189.5)
```

```
final_array=new_array[(new_array>lr1) & (new_array<ur1)]
final_array
```

```
array([45, 21, 96, 15, 36, 63, 79, 97, 11,  8, 92,  2, 89, 87, 84, 72])
```

```
sns.distplot(final_array)
```

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



Result:

The outliers in the dataset were successfully detected and highlighted using the chosen outlier detection method.