# Get Outlier using Z_Score and IQR

February 20, 2025

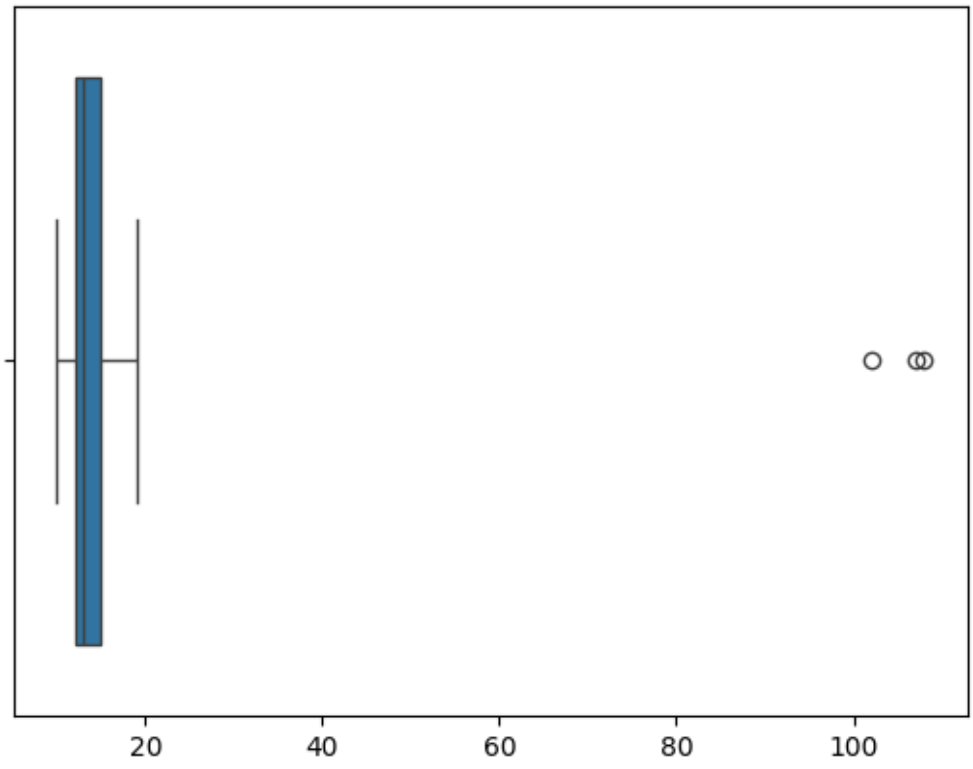
## 1 Outlier Detection and Dropping them using Z_Score and IQR

[ ]:
```
##Outliers
```

[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

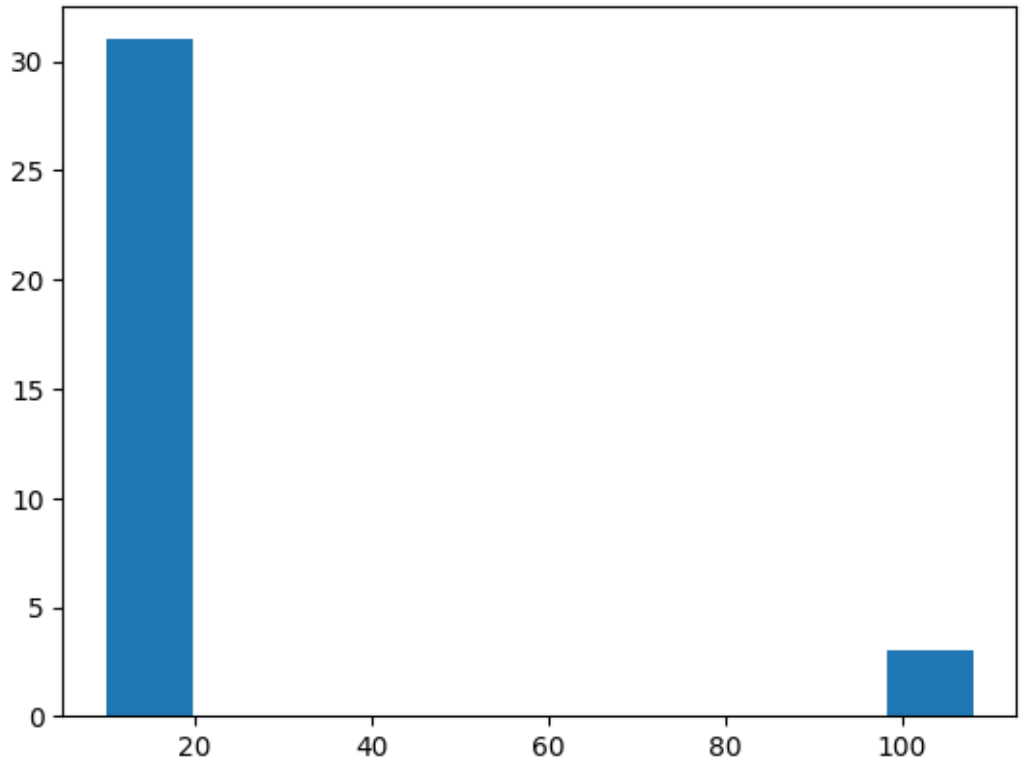
[ ]:
```
##Define a Dataset
dataset=[11,10,12,14,12,15,14,13,15,102,12,14,17,19,107,10,13,12,14,12,108,12,11,14,13,15,10,1
sns.boxplot(x=dataset)
```

[ ]: <Axes: >



[ ]:
```
plt.hist(dataset)
```

[ ]:
```
(array([31.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  3.]),
 array([ 10. ,  19.8,  29.6,  39.4,  49.2,  59. ,  68.8,  78.6,  88.4,
         98.2, 108. ]),
 <BarContainer object of 10 artists>)
```




[ ]:
```
##Z-Score Implementation
outliers=[]

def detect_outliers(data):
  threshold=3
  mean=np.mean(data)
  SD=np.std(data)
  for i in data:
    z_score=(i-mean)/SD
    if np.abs(z_score)>threshold:
      outliers.append(i)

  return outliers
```

[ ]:
```
detect_outliers(dataset)
```

[ ]:
```
[102, 107, 108]
```

##Inter-Quartile Range (IQR) 1. Sort the data 2. Calculate Q1 and Q2 3. IQR=Q3-Q1 4. LF=Q1-1.5(IQR), UF=Q3+1.5(IQR)

[ ]:
```
dataset=sorted(dataset)
dataset
```

[ ]:
```
[10,
 10,
 10,
 10,
 10,
 11,
 11,
 12,
 12,
 12,
 12,
 12,
 12,
 12,
 13,
 13,
 13,
 13,
 14,
 14,
 14,
 14,
 14,
 14,
 15,
 15,
 15,
 15,
 15,
 17,
 19,
 102,
 107,
 108]
```

[ ]:
```
Q1,Q3=np.percentile(dataset,[25,75])
print("25 percentile:",Q1,"75 percentile:",Q3)
```

```
25 percentile: 12.0 75 percentile: 15.0
```

[ ]:
```
##Lowerfence and HigherFence
IQR=Q3-Q1
print("IQR:",IQR)

LF=Q1-1.5*IQR
HF=Q3+1.5*IQR

print("Lower-Fence:",LF,"Higher-fence",HF)
```

```
IQR: 3.0
Lower-Fence: 7.5 Higher-fence 19.5
```

[ ]:
```
dataset
```

[ ]:
```
[10,
 10,
 10,
 10,
 10,
 11,
 11,
 12,
 12,
 12,
 12,
 12,
 12,
 12,
 12,
 13,
 13,
 13,
 13,
 14,
 14,
 14,
 14,
 14,
 14,
 14,
 15,
 15,
 15,
 15,
 15,
 15,
 17,
 19,
```

```
 102,
 107,
 108]
```

[ ]:
```
def removeOutliers(data):
  new_list=[]
  for i in data:
    if i<HF and i>LF:
      new_list.append(i)

  return new_list

new_set=removeOutliers(dataset)
```

[ ]:
```
new_set
```

[ ]:
```
[10,
 10,
 10,
 10,
 10,
 11,
 11,
 12,
 12,
 12,
 12,
 12,
 12,
 12,
 12,
 13,
 13,
 13,
 13,
 13,
 14,
 14,
 14,
 14,
 14,
 14,
 14,
 15,
 15,
 15,
 15,
 15,
 15,
 17,
 19]
```