```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df=pd.read_csv("IOT.csv")
df.describe()
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"temp\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 34492.906619653186,\n
\"min\": 2.0,\n        \"max\": 97601.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
35.05494820749787,\n          35.0,\n          97601.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 97606 entries, 0 to 97605
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   id          97606 non-null  object
 1   room_id     97606 non-null  object
 2   noted_date  97603 non-null  object
 3   temp        97601 non-null  float64
 4   out/in      97601 non-null  object
dtypes: float64(1), object(4)
memory usage: 3.7+ MB
```

```
df=df.rename(columns={'out/in':'status'})
df.columns
```

```
Index(['id', 'room_id', 'noted_date', 'temp', 'status'],
dtype='object')
```

```
df.shape
```

```
(97606, 5)
```

```
df.isna().sum()
```

```
id            0
room_id       0
noted_date    3
temp          5
status        5
dtype: int64
```

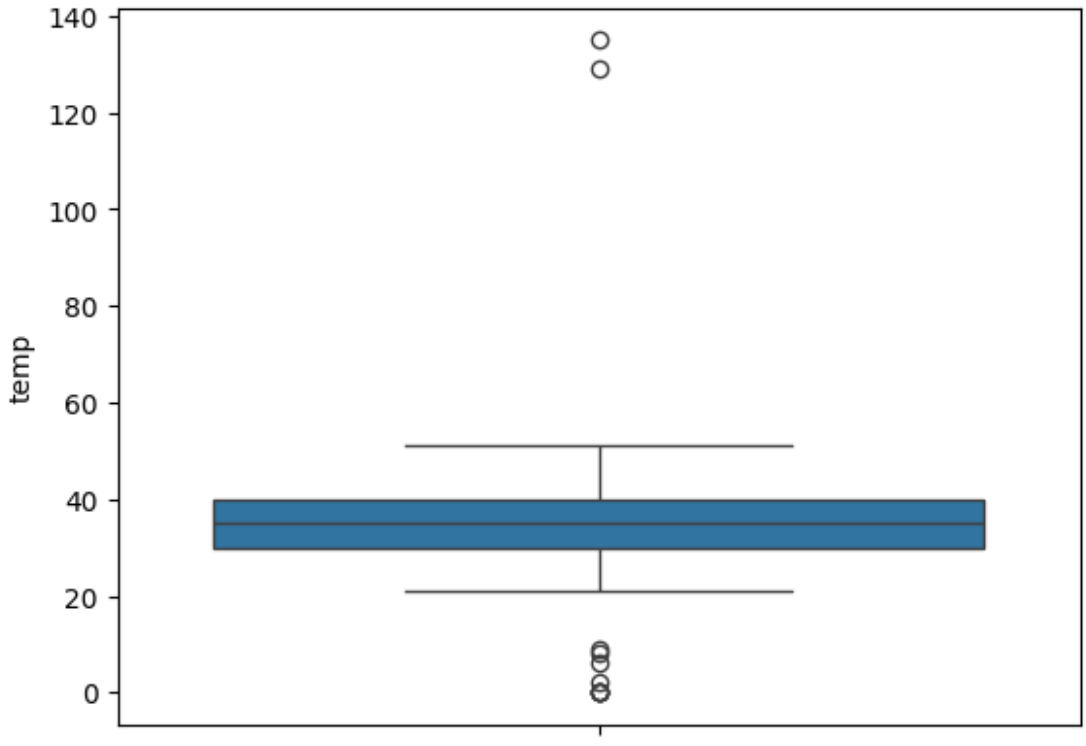
```
df.status.fillna(method='ffill',inplace=True)
df.noted_date.fillna(method='bfill',inplace=True)
df.temp.fillna(0,inplace=True)
df.status.isna().sum()
df.noted_date.isna().sum()
df.temp.isna().sum()
```

```
<ipython-input-23-de89c0662822>:1: FutureWarning: Series.fillna with
'method' is deprecated and will raise in a future version. Use
obj.ffill() or obj.bfill() instead.
  df.status.fillna(method='ffill',inplace=True)
<ipython-input-23-de89c0662822>:2: FutureWarning: Series.fillna with
'method' is deprecated and will raise in a future version. Use
obj.ffill() or obj.bfill() instead.
  df.noted_date.fillna(method='bfill',inplace=True)
```

```
0
```

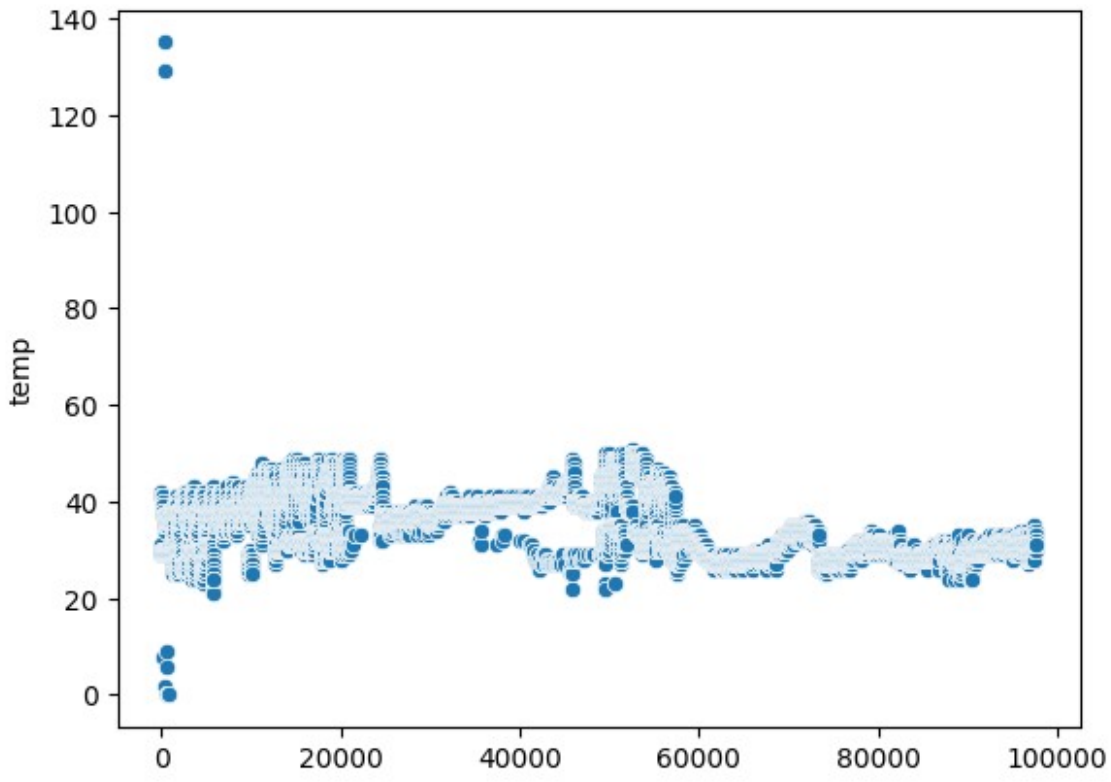
```
sns.boxplot(df.temp)
```

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




```
sns.scatterplot(df.temp)
```

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




```
q1=np.percentile(df.temp,25)
q3=np.percentile(df.temp,75)
iqr= q3-q1
lower_bound= q1-1.5*iqr
upper_bound=q3+1.5*iqr
outliers=df[(df.temp>upper_bound) | (df.temp<lower_bound)]
print('qi:',q1)
print('q3:',q3)
print('iqr:',iqr)
print('lower_bound:',lower_bound)
print('upper_bound:',upper_bound)
print('outliers:',outliers)
sns.boxplot(data='outliers')
```

```
qi: 30.0
q3: 40.0
iqr: 10.0
lower_bound: 15.0
upper_bound: 55.0
outliers:                                          id     room_id
noted_date   temp  \
237  __export__.temp_log_195412_c0870399  Room Admin  08/12/2018 05:23
8.0
346  __export__.temp_log_195071_7362ce83  Room Admin  08/12/2018 03:53
2.0
```

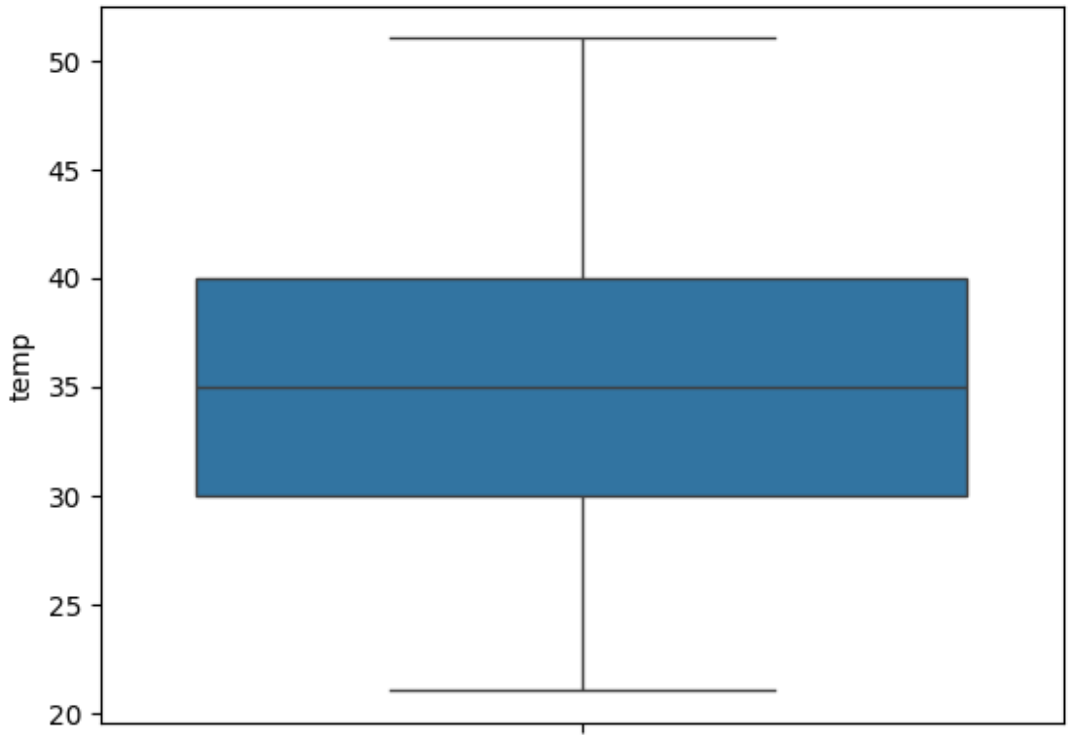
```
390  __export__.temp_log_194939_76c28cf6  Room Admin  08/12/2018 03:07
129.0
397  __export__.temp_log_194928_dc455534  Room Admin  08/12/2018 03:04
135.0
512  __export__.temp_log_194663_3e0f66ee  Room Admin  08/12/2018 01:37
6.0
668  __export__.temp_log_194287_eddd802a  Room Admin  07/12/2018 22:59
0.0
700  __export__.temp_log_194199_f4c417cf  Room Admin  07/12/2018 22:24
9.0
713  __export__.temp_log_194173_f1417c58  Room Admin  07/12/2018 22:12
0.0
791  __export__.temp_log_193962_4aba8cff  Room Admin  07/12/2018 20:56
0.0
818  __export__.temp_log_193887_a130bb6a  Room Admin  07/12/2018 20:32
0.0
869  __export__.temp_log_193798_d00e48f1  Room Admin  07/12/2018 20:02
0.0

    status
237    Out
346     In
390     In
397    Out
512     In
668     In
700    Out
713    Out
791    Out
818     In
869    Out

<Axes: >
```

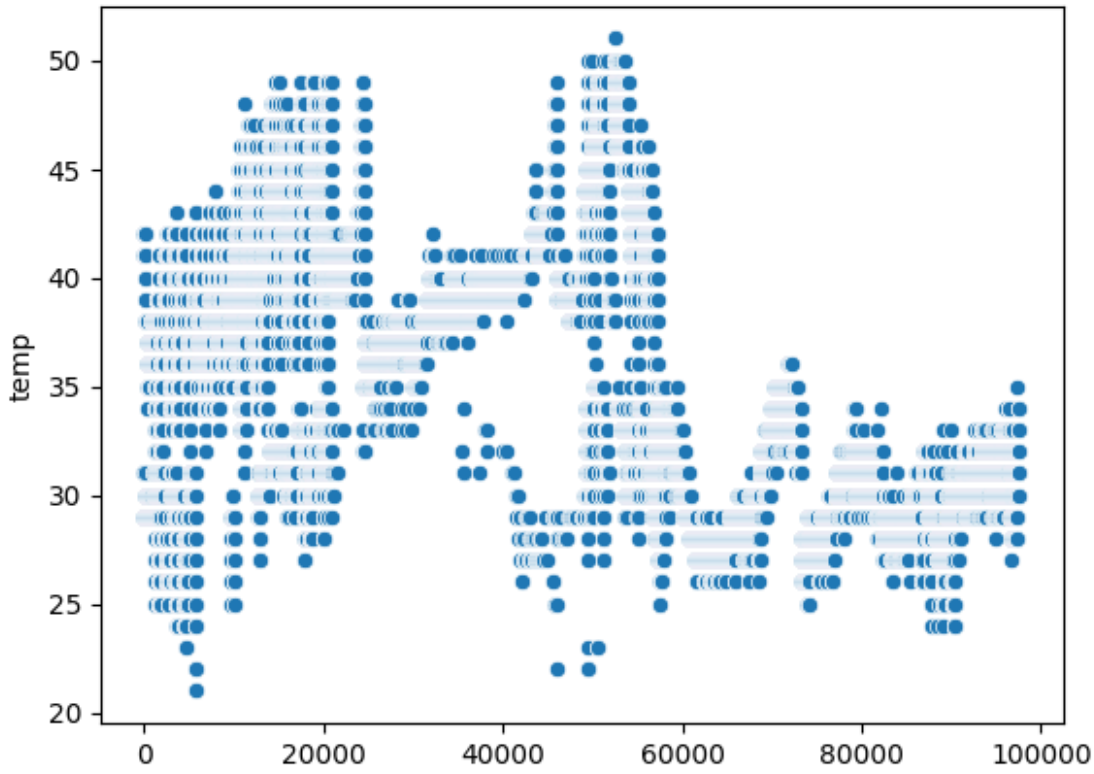

```
df=df[(df.temp>=lower_bound)&(df.temp<=upper_bound) ]
sns.boxplot(data=df['temp'])
```

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




```
sns.scatterplot(data=df['temp'])
```

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




```
df['temp'].value_counts()
```

```
temp
39.0    10203
28.0     8831
29.0     7918
40.0     7798
31.0     7236
30.0     6613
37.0     5721
32.0     5408
27.0     4631
41.0     4354
36.0     3963
38.0     3866
42.0     3447
33.0     3437
34.0     2613
43.0     2004
44.0     1774
35.0     1581
45.0     1508
46.0     1201
47.0     1044
48.0      971
```

```
26.0      699
49.0      401
25.0      224
24.0       66
50.0       55
22.0       19
23.0        5
21.0        2
51.0        2
Name: count, dtype: int64
```