

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

```
data=pd.read_csv("/content/sample_data/new_insurance_data (1) (1).csv")
```

```
data.head()
```

	age	sex	bmi	children	smoker	Claim_Amount	past_consultations	num_of_steps	Hospital_expenditure	Number_of_past_hospitalizat:
0	18.0	male	23.21	0.0	no	29087.54313	17.0	715428.0	4720920.992	
1	18.0	male	30.14	0.0	no	39053.67437	7.0	699157.0	4329831.676	
2	18.0	male	33.33	0.0	no	39023.62759	19.0	702341.0	6884860.774	
3	18.0	male	33.66	0.0	no	28185.39332	11.0	700250.0	4274773.550	
4	18.0	male	34.10	0.0	no	14697.85941	16.0	711584.0	3787293.921	

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   age                                   1329 non-null   float64
1   sex                                   1338 non-null   object
2   bmi                                   1335 non-null   float64
3   children                             1333 non-null   float64
4   smoker                               1338 non-null   object
5   Claim_Amount                         1324 non-null   float64
6   past_consultations                   1332 non-null   float64
7   num_of_steps                         1335 non-null   float64
8   Hospital_expenditure                 1334 non-null   float64
9   Number_of_past_hospitalizations      1336 non-null   float64
10  Anual_Salary                         1332 non-null   float64
11  region                               1338 non-null   object
12  charges                              1338 non-null   float64
dtypes: float64(10), object(3)
memory usage: 136.0+ KB
```

```
data.isna().sum()
```

	0
age	9
sex	0
bmi	3
children	5
smoker	0
Claim_Amount	14
past_consultations	6
num_of_steps	3
Hospital_expenditure	4
Number_of_past_hospitalizations	2
Anual_Salary	6
region	0
charges	0

```
data=data.dropna()
```



McAfee | WebAdvisor

Your download's being scanned.  
We'll let you know if there's an issue.

```
data.isna().sum()
```

```

age      0
sex      0
bmi      0
children 0
smoker   0
Claim_Amount 0
past_consultations 0
num_of_steps 0
Hospital_expenditure 0
NUmber_of_past_hospitalizations 0
Anual_Salary 0
region   0
charges  0

```

```
data.duplicated().sum()
```

```
0
```

```
data.describe()
```

```

age      bmi      children  Claim_Amount  past_consultations  num_of_steps  Hospital_expenditure  NUmber_of_past_hospit:
count  1287.000000  1287.000000  1287.000000  1287.000000  1287.000000  1.287000e+03  1.287000e+03  1
mean    39.419580   30.655746   1.094794  33498.578933   15.247086  9.113778e+05  1.579929e+07
std     14.051987    6.086323   1.202782  15611.720348    7.457366  9.083364e+04  2.653821e+07
min     18.000000   15.960000   0.000000   2912.590584    1.000000  6.991570e+05  2.945253e+04
25%     27.000000   26.220000   0.000000  20905.974820    9.000000  8.489270e+05  4.114357e+06
50%     40.000000   30.360000   1.000000  33968.774160   15.000000  9.152840e+05  7.567064e+06
75%     51.000000   34.600000   2.000000  45152.414195   20.000000  9.726140e+05  1.093014e+07
max     64.000000   53.130000   5.000000  77277.988480   40.000000  1.107872e+06  2.616317e+08

```

Start coding or [generate](#) with AI.

```

ftr=[]                ## no need of this
for i in data.columns:
    if data[i].dtypes=='float64' or data[i].dtypes=='int64':
        ftr.append(i)

```

```
ftr
```

```

['age',
 'bmi',
 'children',
 'Claim_Amount',
 'past_consultations',
 'num_of_steps',
 'Hospital_expenditure',
 'NUmber_of_past_hospitalizations',
 'Anual_Salary',
 'charges']

```

```

data1=pd.DataFrame(ftr)
data1

```



McAfee | WebAdvisor



Your download's being scanned.  
We'll let you know if there's an issue.



	0
0	age
1	bmi
2	children
3	Claim_Amount
4	past_consultations
5	num_of_steps
6	Hospital_expenditure
7	NUmber_of_past_hospitalizations
8	Anual_Salary
9	charges



data.shape

 (1287, 13)

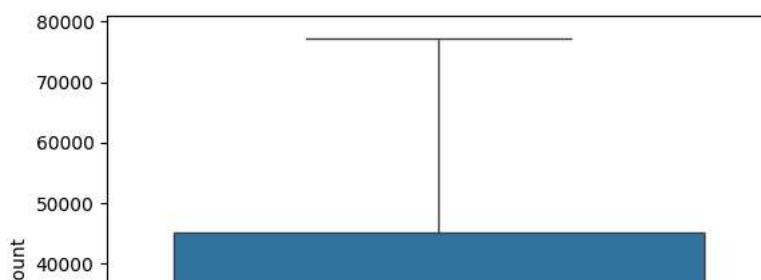
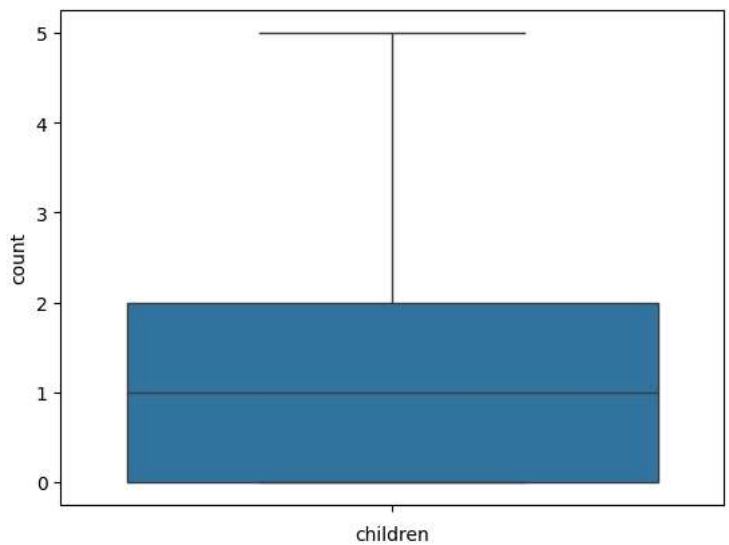
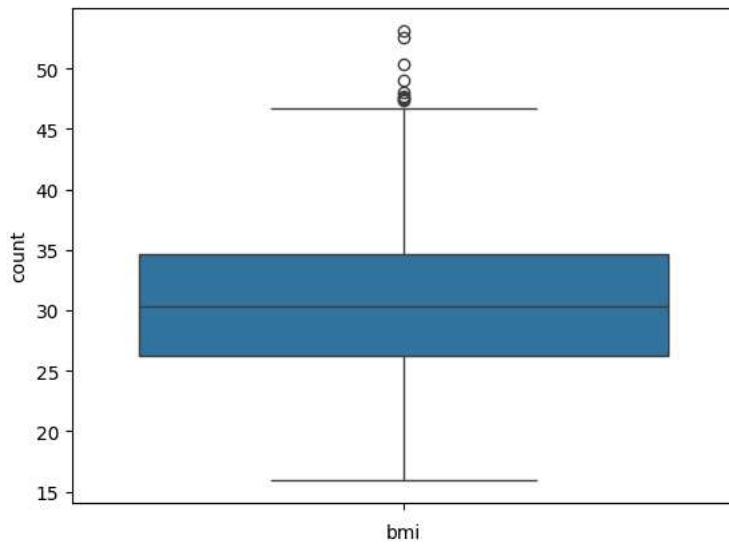
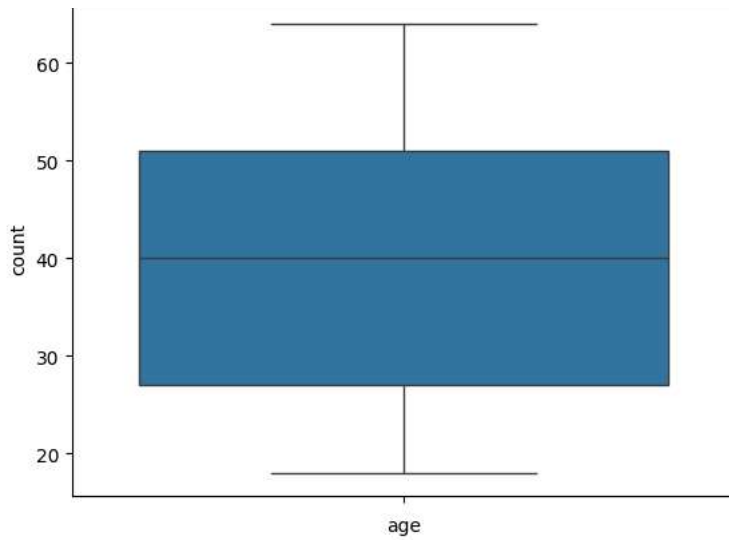
```
for i in data.columns:
    if data[i].dtypes=='float64' or data[i].dtypes=='int64':    ## no need of outlier in this case
        sns.boxplot(data[i])
        plt.xlabel(i)
        plt.ylabel('count')
        plt.show()
```



**McAfee** | WebAdvisor



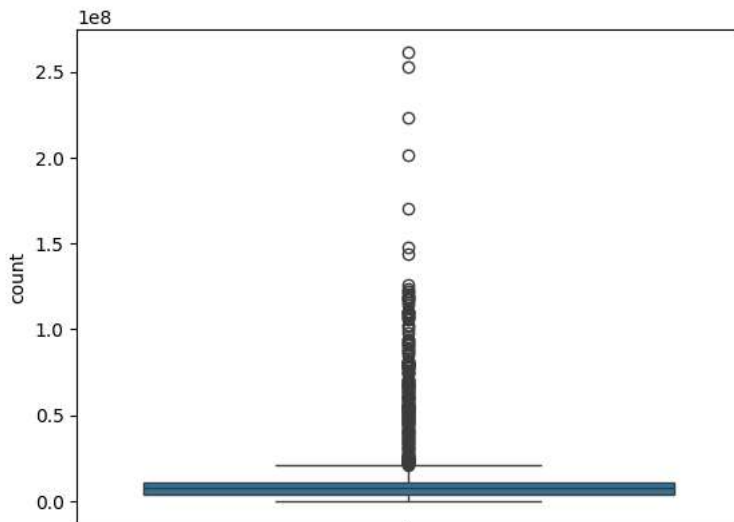
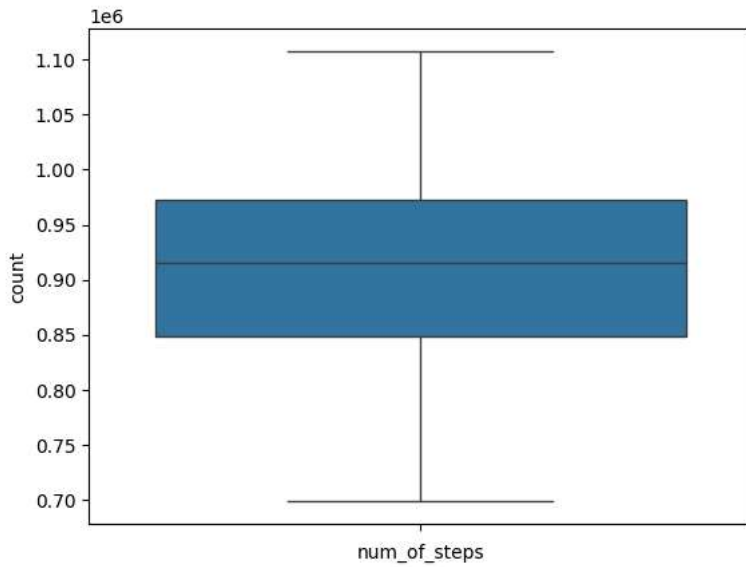
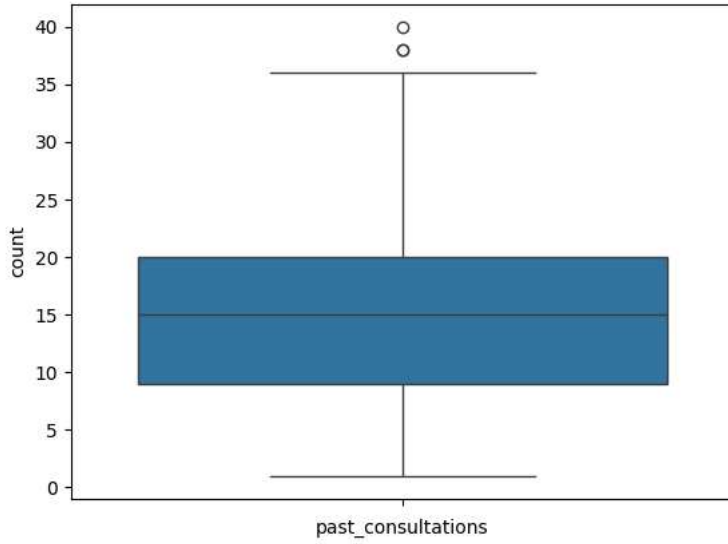
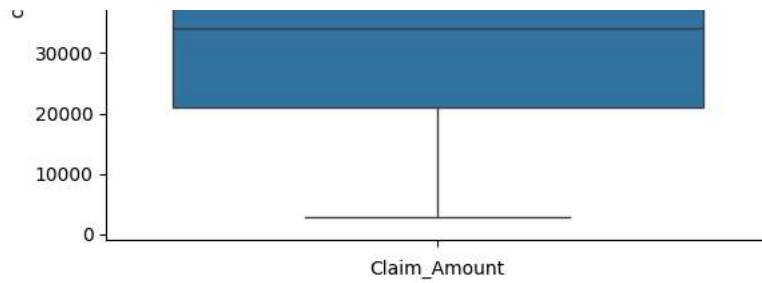
Your download's being scanned.  
We'll let you know if there's an issue.



McAfee | WebAdvisor

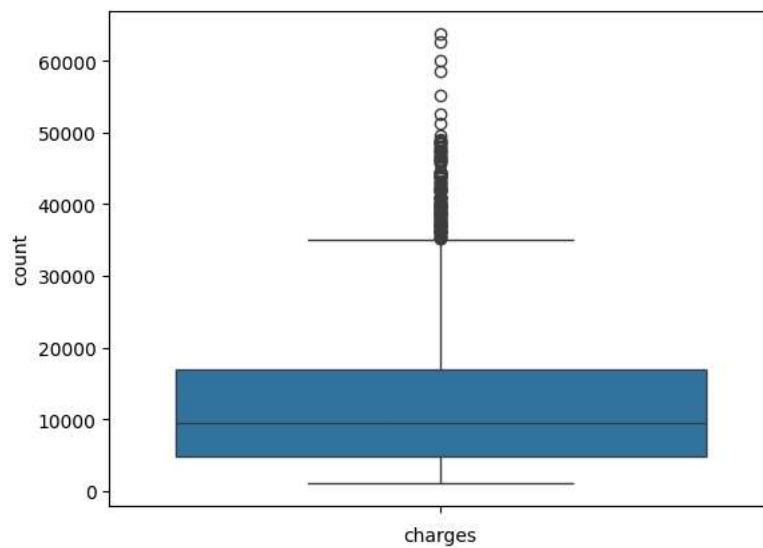
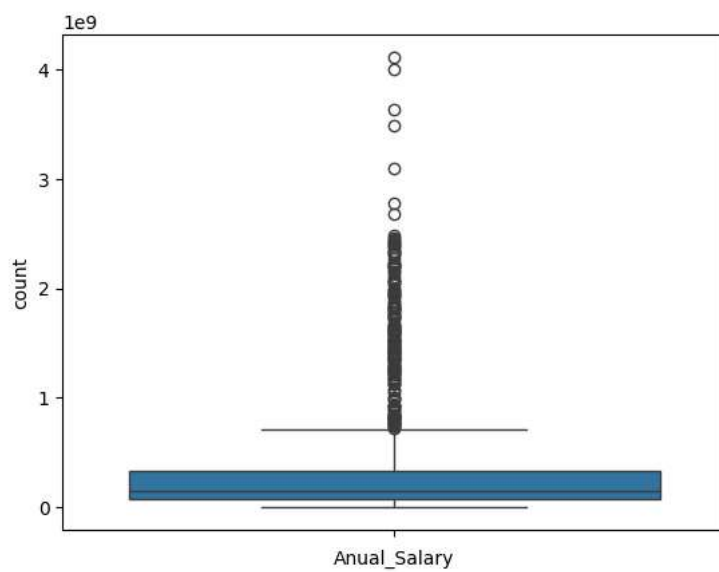
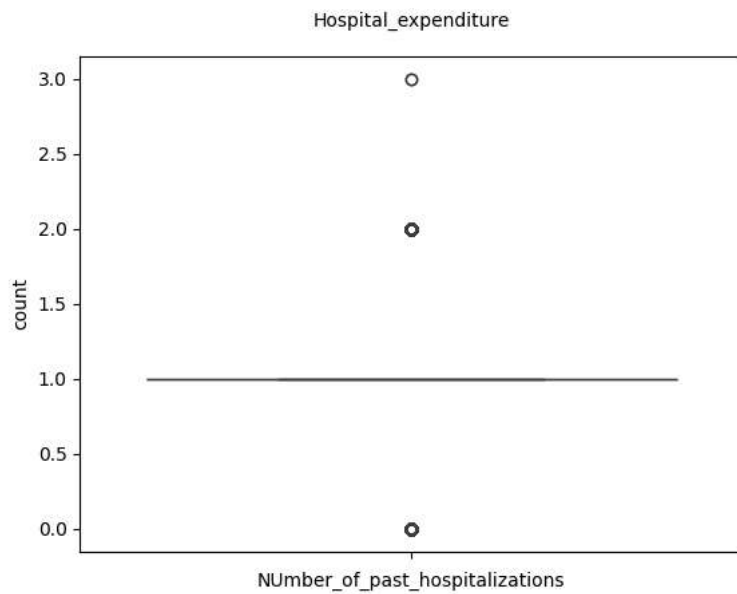


Your download's being scanned.  
We'll let you know if there's an issue.



 McAfee | WebAdvisor ×

 Your download's being scanned.  
We'll let you know if there's an issue.



McAfee | WebAdvisor



Your download's being scanned.  
We'll let you know if there's an issue.