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

```
Start coding or generate with AI.
```

```
Start coding or generate with AI.
```

```
Start coding or generate with AI.
```

```
df=pd.read_csv("insurance.csv")
df
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

```
df.head()
```

```
df['age'].mean()
```

```
df.isnull().sum()
```

	0
age	0
sex	0
bmi	0
children	0
smoker	0
region	0
charges	0

dtype: int64

```
df.nunique()
```

	0
age	47
sex	2
bmi	548
children	6
smoker	2
region	4
charges	1337

dtype: int64

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

age	count
18	69
19	68
46	29
52	29
50	29
47	29
48	29
51	29
45	29
20	29
24	28
27	28
28	28
25	28
23	28
49	28
54	28
53	28
22	28
21	28
26	28
31	27
41	27
44	27
43	27
42	27
29	27
30	27
40	27
32	26

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

57	26
34	26

```
55    26
count
56    26
sex
male      676
female    662
37    25
dtype: int64
69    25
```

```
Start coding or generate with AI.
```

38 25
62 23
60 23
63 23
61 23
64 22

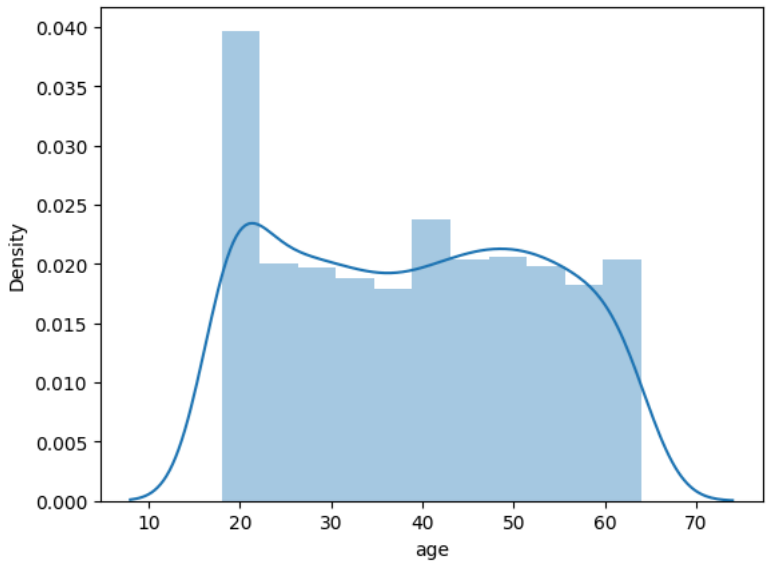
1 univariant

2 bivarient

3 multivarient

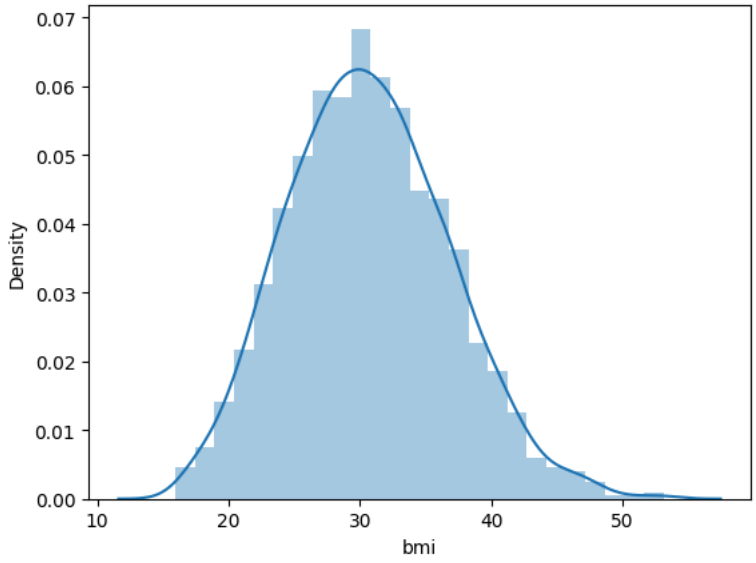
```
sns.distplot(df['age'])
```

```
/tmp/ipython-input-3234920688.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(df['age'])
<Axes: xlabel='age', ylabel='Density'>
```



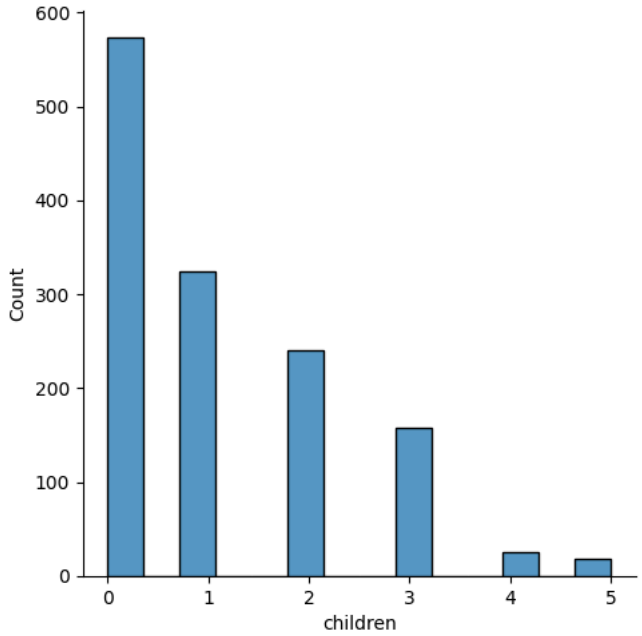

```
sns.distplot(df['bmi'])
```

```
/tmp/ipython-input-4168411822.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(df['bmi'])
<Axes: xlabel='bmi', ylabel='Density'>
```



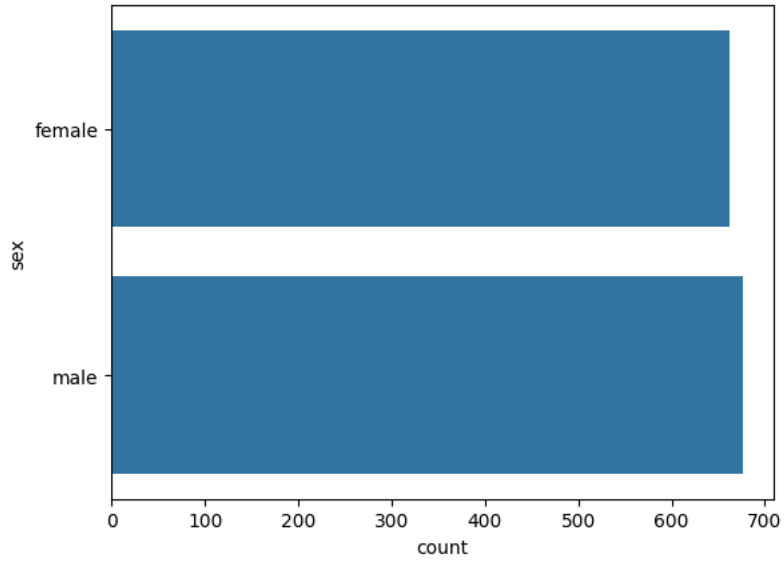

```
sns.displot(df['children'])
```

```
<seaborn.axisgrid.FacetGrid at 0x7f13d0731a60>
```



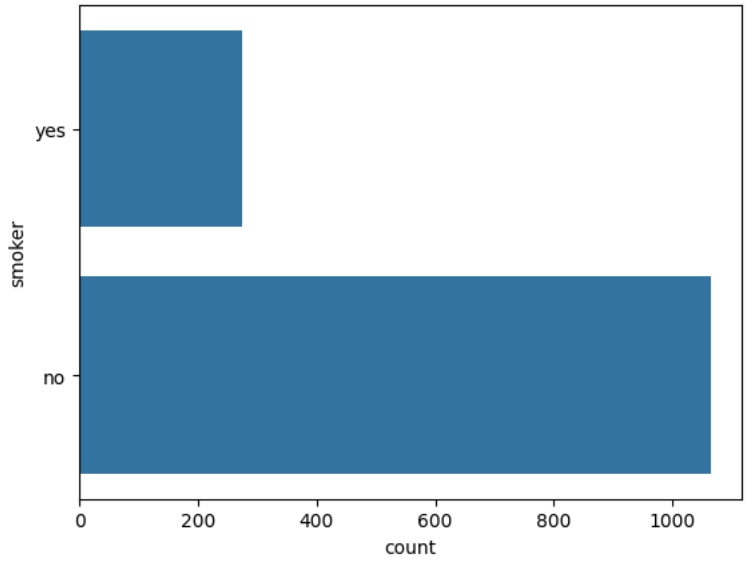

```
sns.countplot(df['sex'])
```

```
<Axes: xlabel='count', ylabel='sex'>
```



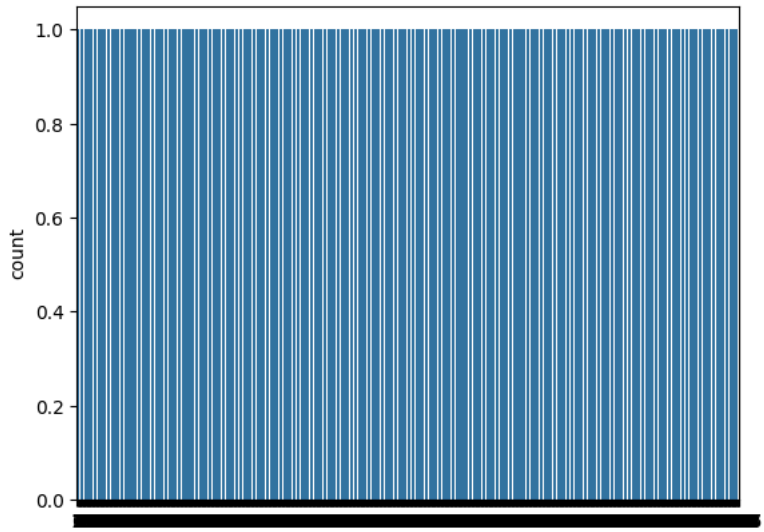

```
sns.countplot(df['smoker'])
```

```
<Axes: xlabel='count', ylabel='smoker'>
```



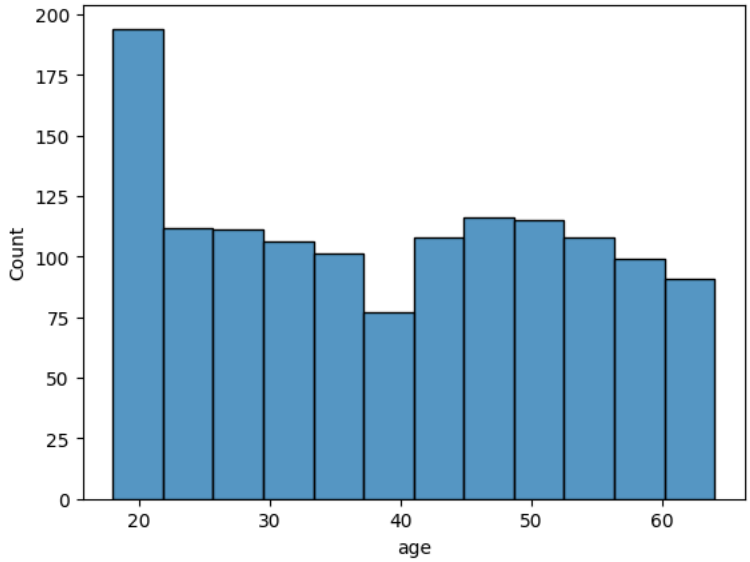

```
sns.countplot(df['children'])
```

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



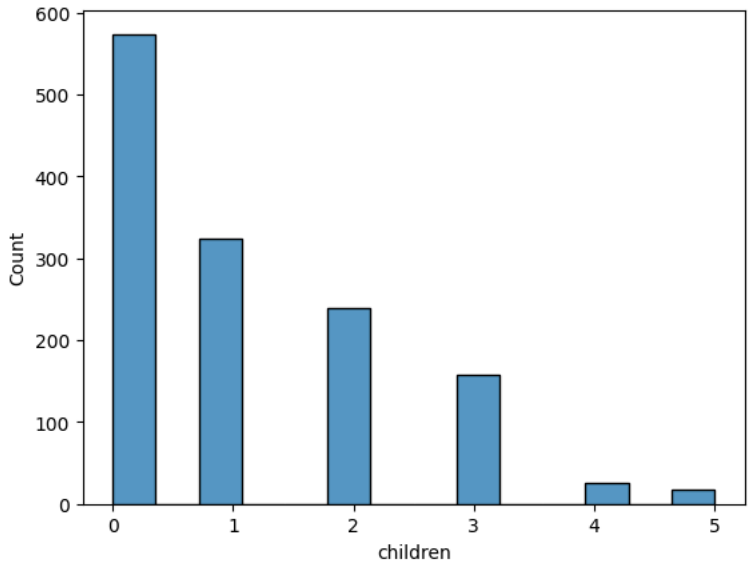

```
sns.histplot(df['age'])
```

```
<Axes: xlabel='age', ylabel='Count'>
```



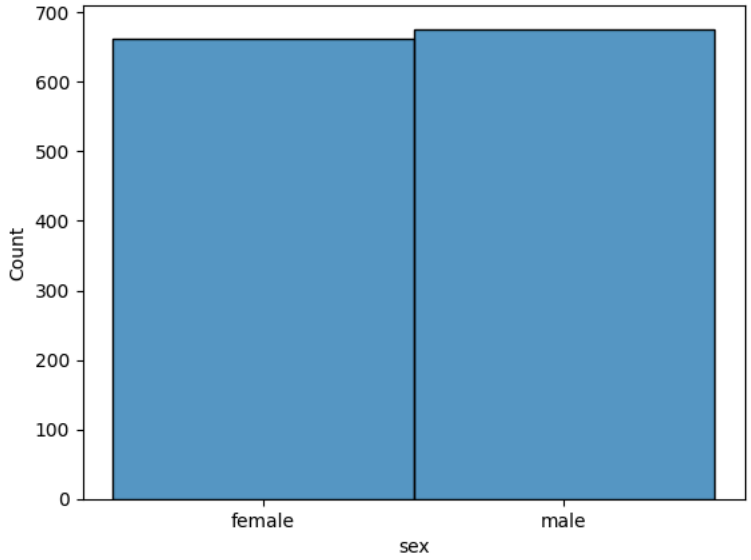

```
sns.histplot(df['children'])
```

```
<Axes: xlabel='children', ylabel='Count'>
```



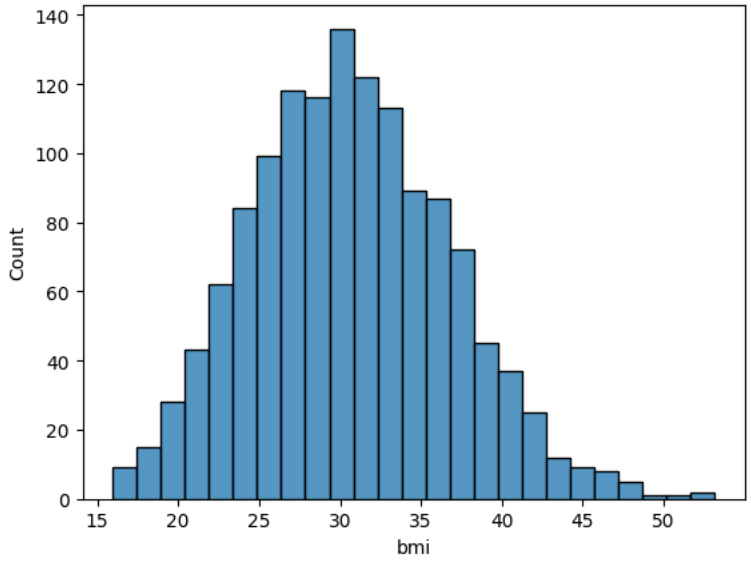

```
sns.histplot(df['sex'])
```

```
<Axes: xlabel='sex', ylabel='Count'>
```



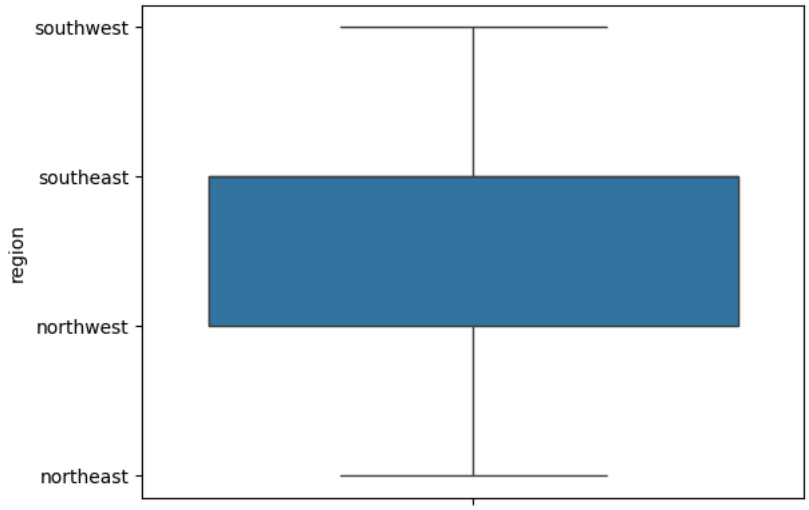

```
sns.histplot(df['bmi'])
```

```
<Axes: xlabel='bmi', ylabel='Count'>
```



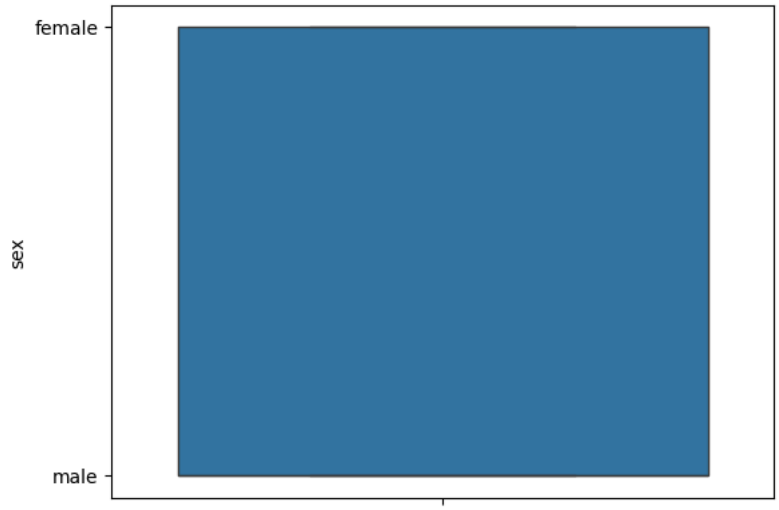

```
sns.boxplot(df['region'])
```

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



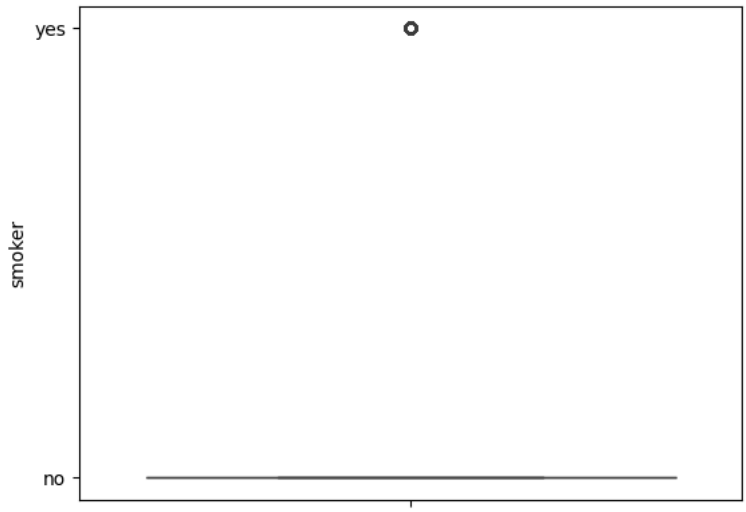

```
sns.boxplot(df['sex'])
```

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



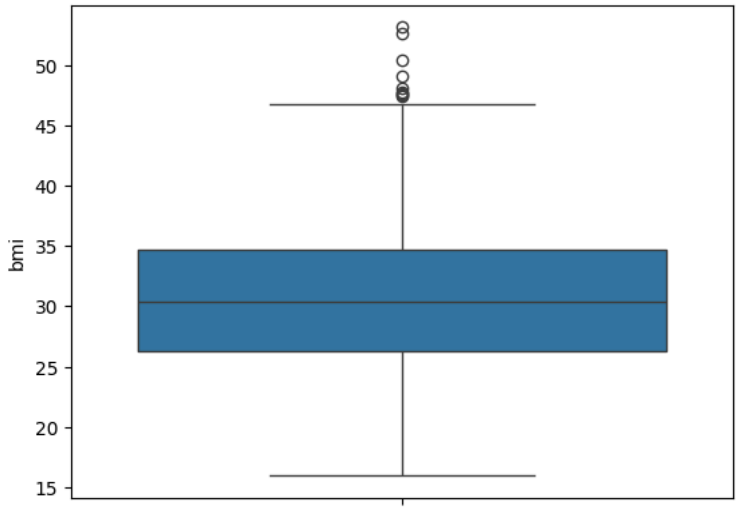
```
sns.boxplot(df['smoker'])
```

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



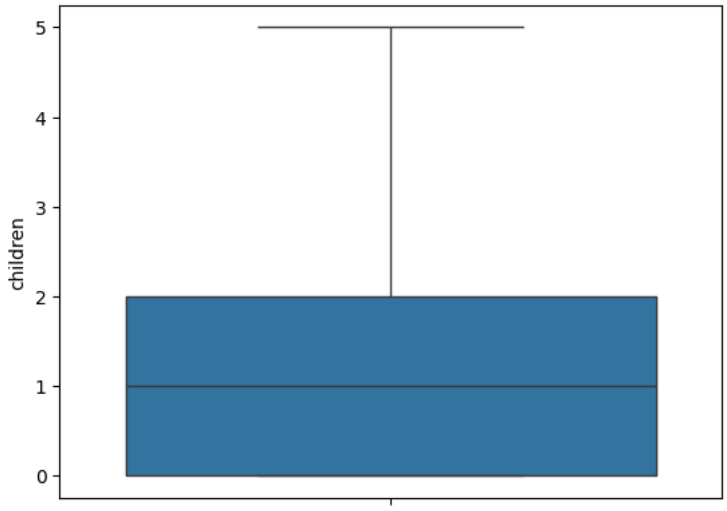

```
sns.boxplot(df['bmi'])
```

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



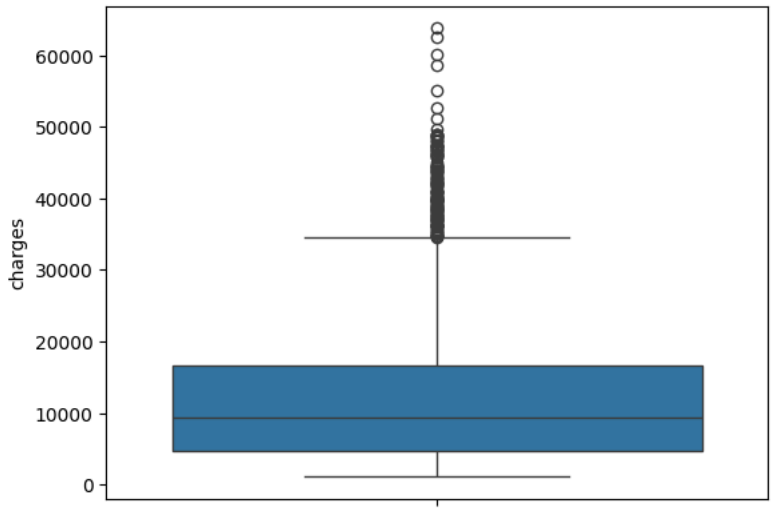

```
sns.boxplot(df['children'])
```

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




```
sns.boxplot(df['charges'])
```

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



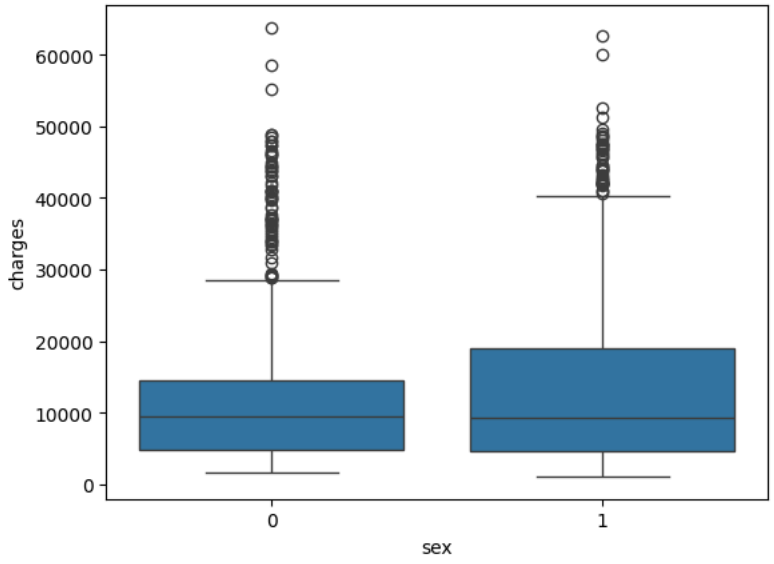

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
a=['sex','age','smoker']
for i in a:
 df[i]=le.fit_transform(df[i])
```

```
df.head()
```

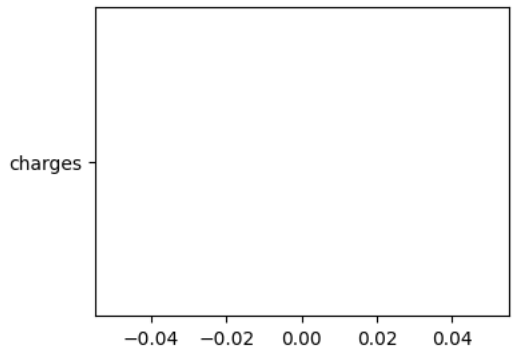
	age	sex	bmi	children	smoker	region	charges
0	1	0	27.900	0	1	southwest	16884.92400
1	0	1	33.770	1	0	southeast	1725.55230
2	10	1	33.000	3	0	southeast	4449.46200
3	15	1	22.705	0	0	northwest	21984.47061
4	14	1	28.880	0	0	northwest	3866.85520

```
sns.boxplot(x="sex",y="charges",data=df)
```

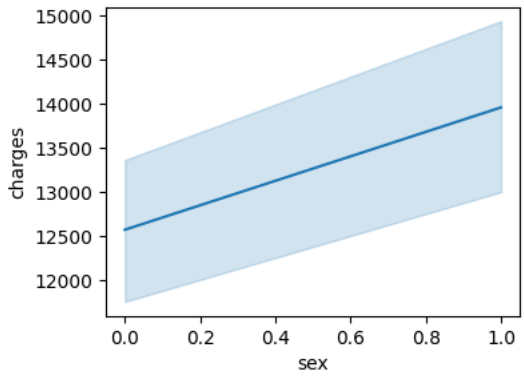
```
<Axes: xlabel='sex', ylabel='charges'>
```



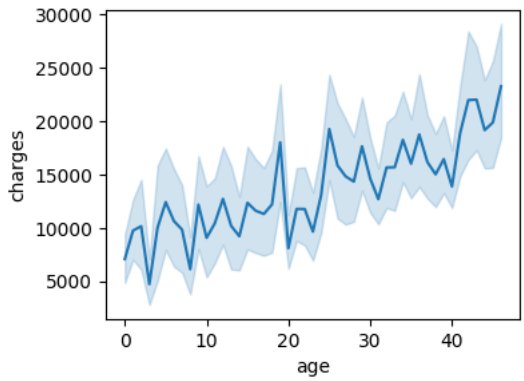
```
plt.figure(figsize=(4,3))
sns.lineplot("charges")
plt.show()
```




```
plt.figure(figsize=(4,3))
sns.lineplot(x="sex",y="charges",data=df)
plt.show()
```




```
plt.figure(figsize=(4,3))
sns.lineplot(x="age",y="charges",data=df)
plt.show()
```

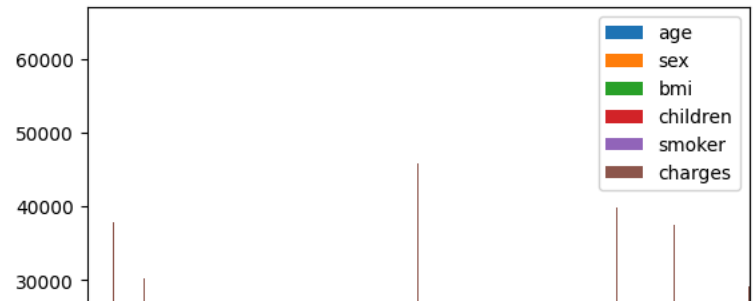



```
df.corr()
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
```

```
df.plot(kind='bar')
```

```
<Axes: >
```




```
 3 frames
n _interleave(self, dtype, na_value)
```