

```
import pandas as pd
import seaborn as sns
```

```
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()
```

	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

```
df.tail()
```

	age	sex	bmi	children	smoker	region	charges
1333	50	male	30.97	3	no	northwest	10600.5483
1334	18	female	31.92	0	no	northeast	2205.9808
1335	18	female	36.85	0	no	southeast	1629.8335
1336	21	female	25.80	0	no	southwest	2007.9450
1337	61	female	29.07	0	yes	northwest	29141.3603

```
df["age"].mean()  
  
np.float64(39.20702541106129)
```

```
df["age"].mode()  
  
age  
0    18  
  
dtype: int64
```

```
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()
```


count	
age	
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

df["sex"].value_counts()

22	28
21	count
26	28
sex	
male	676
female	662
41	27
44	27
dtype: int64	
43	27
42	27
29	27
30	27
40	27

1 univariate

2 bivariate

3 multivariate

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

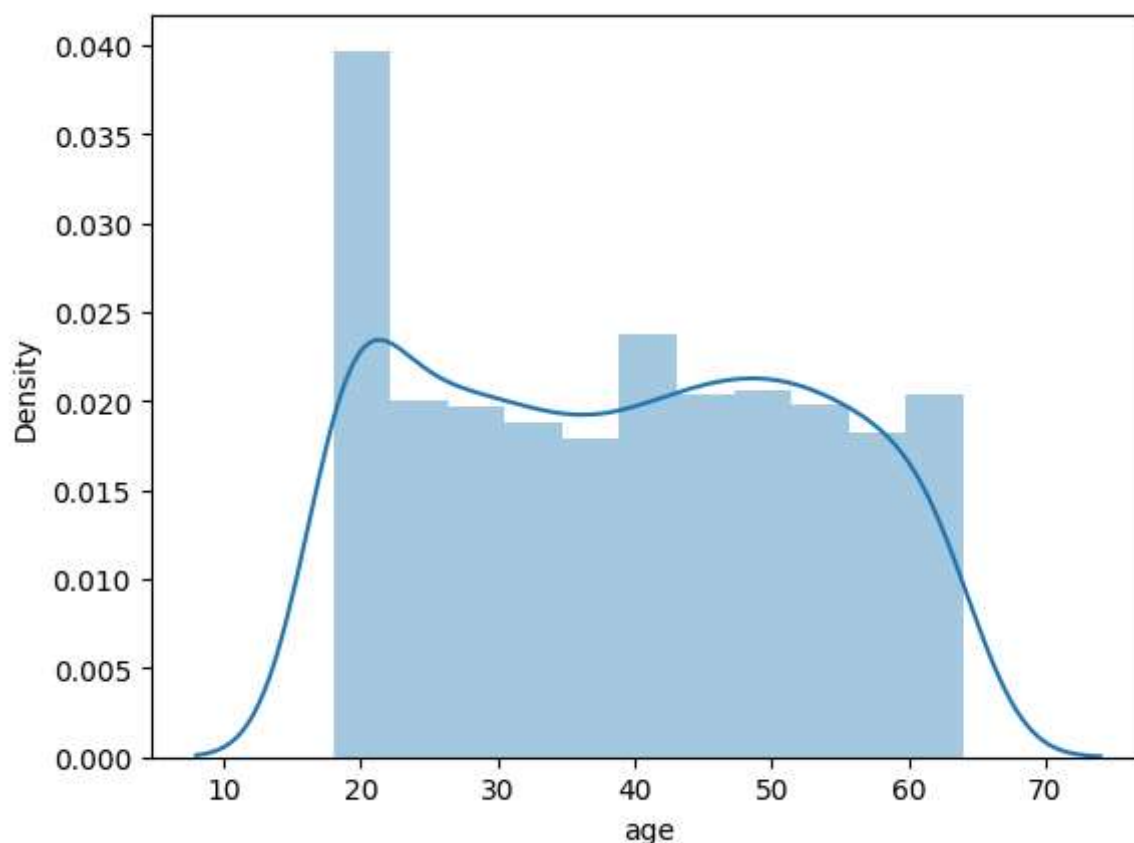
```
/tmp/ipython-input-316555093.py:1: UserWarning:
```

```
`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["sex"])
```

```
/tmp/ipython-input-4018499019.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["sex"])
```

ValueError Traceback (most recent call last)

```
/tmp/ipython-input-4018499019.py in <cell line: 0>()
```

```
----> 1 sns.distplot(df["sex"])
```

1 frames

```
/usr/local/lib/python3.12/dist-packages/pandas/core/series.py in __array__(self, dtype, copy)
```

```

1029         """
1030         values = self._values
-> 1031         arr = np.asarray(values, dtype=dtype)
1032         if using_copy_on_write() and astype_is_view(values.dtype,
arr.dtype):
1033             arr = arr.view()
```

ValueError: could not convert string to float: 'female'



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

```
/tmp/ipython-input-2859878543.py:1: UserWarning:
```

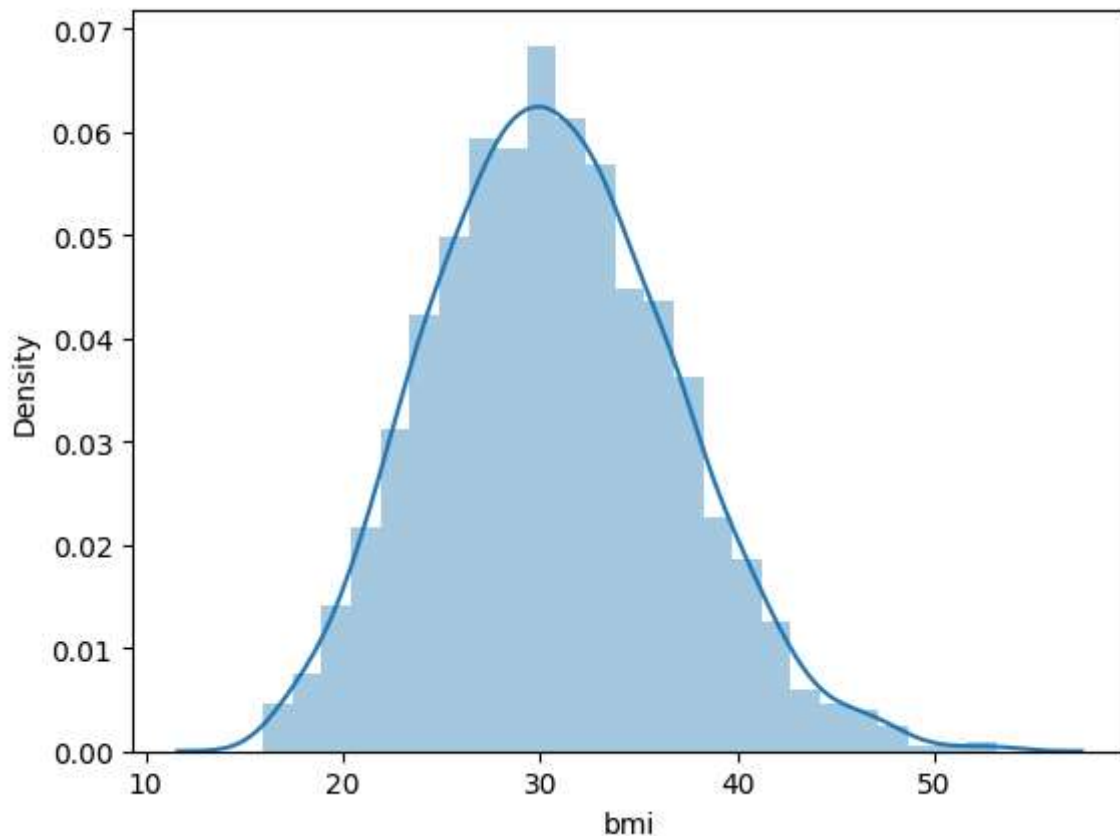
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

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.distplot(df["children"])
```



```
/tmp/ipython-input-2551773931.py:1: UserWarning:
```

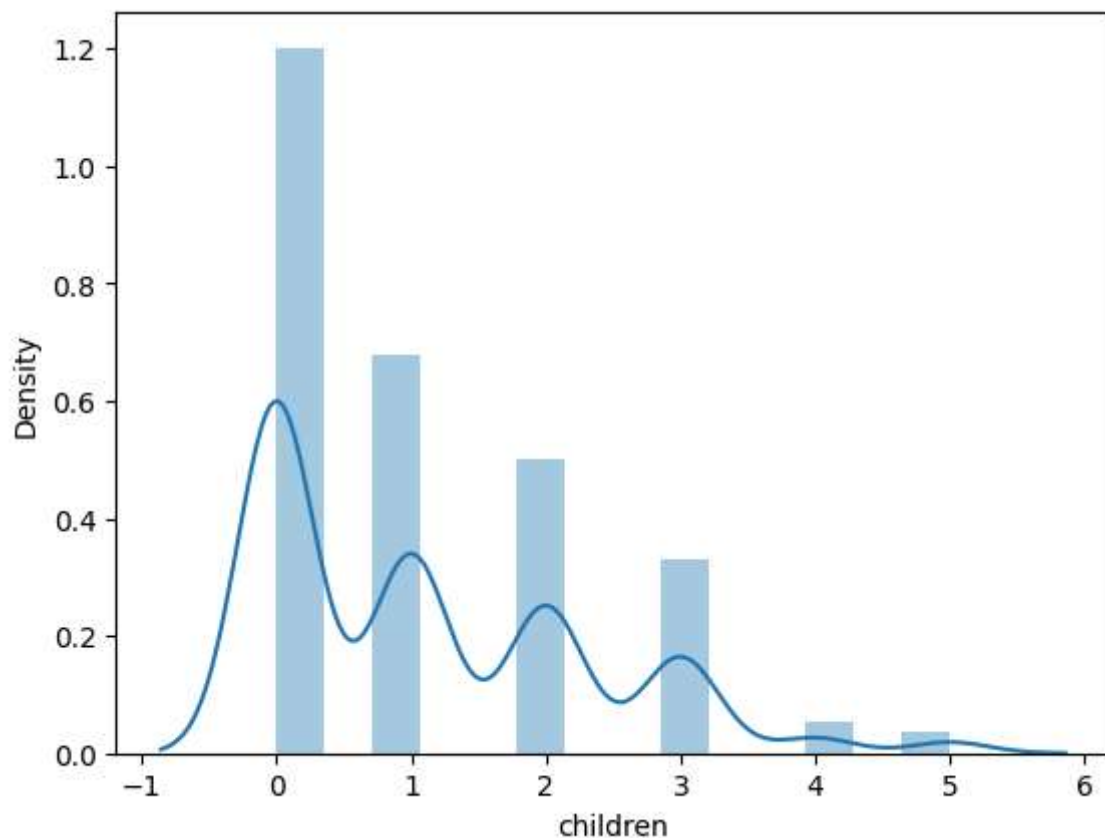
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

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["children"])  
<Axes: xlabel='children', ylabel='Density'>
```



```
sns.distplot(df["smoker"])
```

```
/tmp/ipython-input-1410226135.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["smoker"])
```

ValueError Traceback (most recent call last)

```
/tmp/ipython-input-1410226135.py in <cell line: 0>()
```

```
----> 1 sns.distplot(df["smoker"])
```

1 frames

```
/usr/local/lib/python3.12/dist-packages/pandas/core/series.py in __array__(self, dtype, copy)
```

```

1029         """
1030         values = self._values
-> 1031         arr = np.asarray(values, dtype=dtype)
1032         if using_copy_on_write() and astype_is_view(values.dtype,
arr.dtype):
1033             arr = arr.view()
```

ValueError: could not convert string to float: 'yes'



```
sns.distplot(df["charges"])
```

```
/tmp/ipython-input-3454039009.py:1: UserWarning:
```

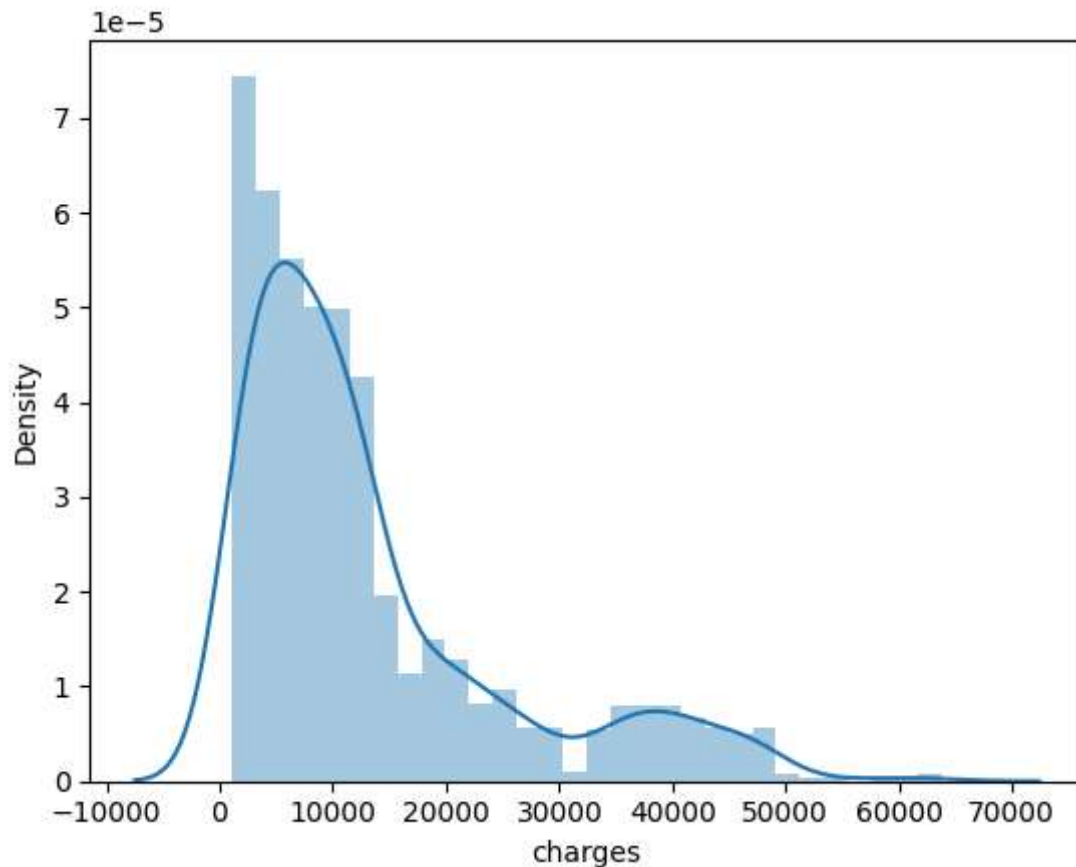
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

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

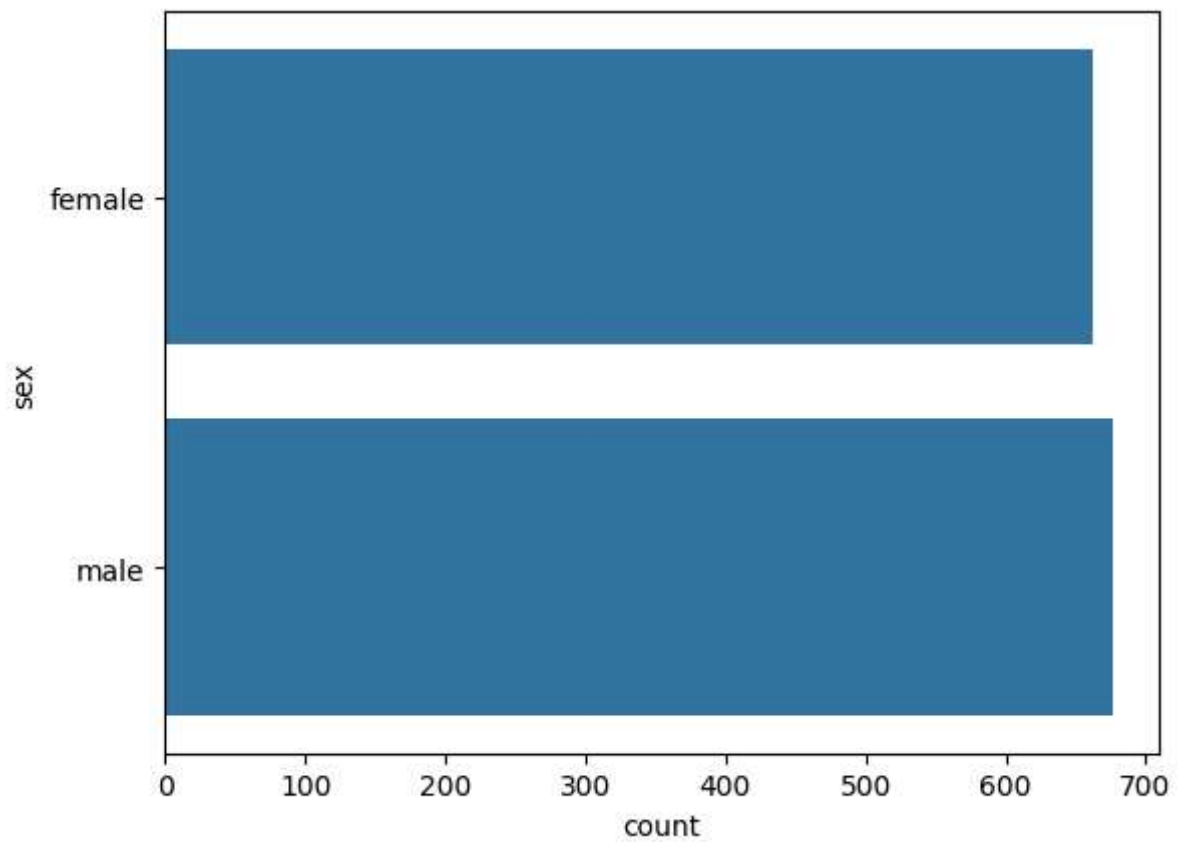
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df["charges"])  
<Axes: xlabel='charges', ylabel='Density'>
```



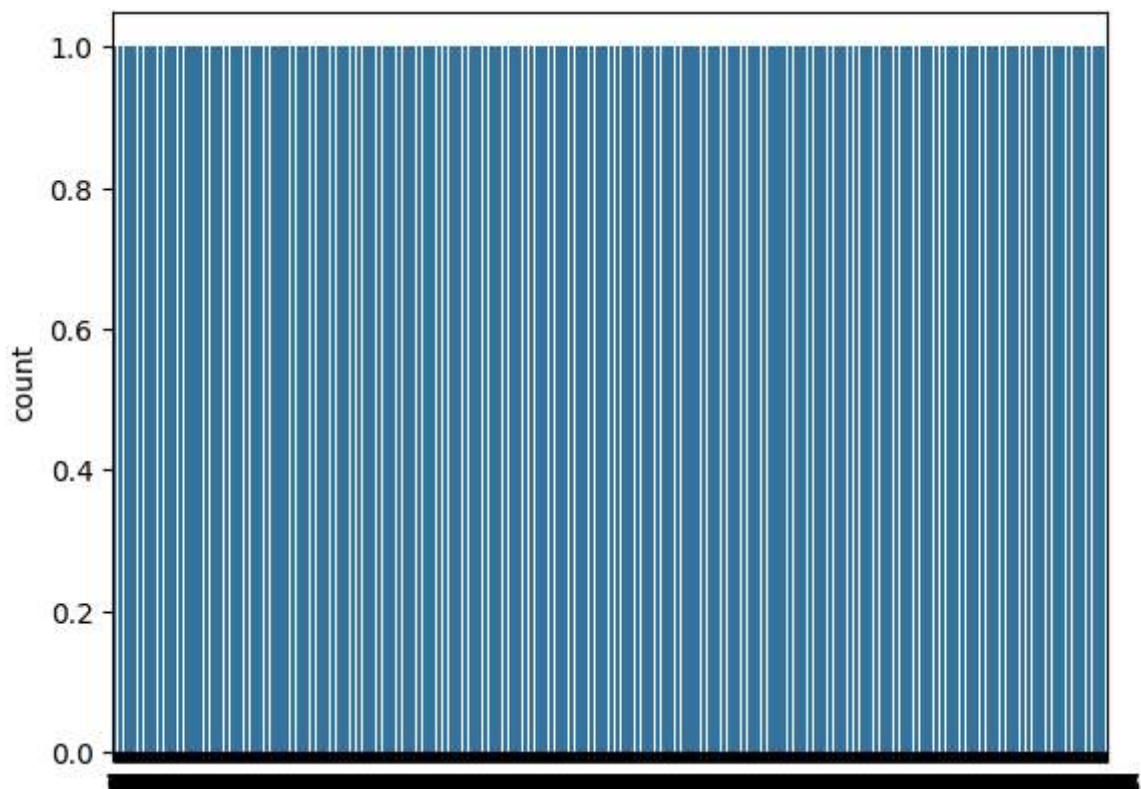
```
sns.countplot(df["sex"])
```

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



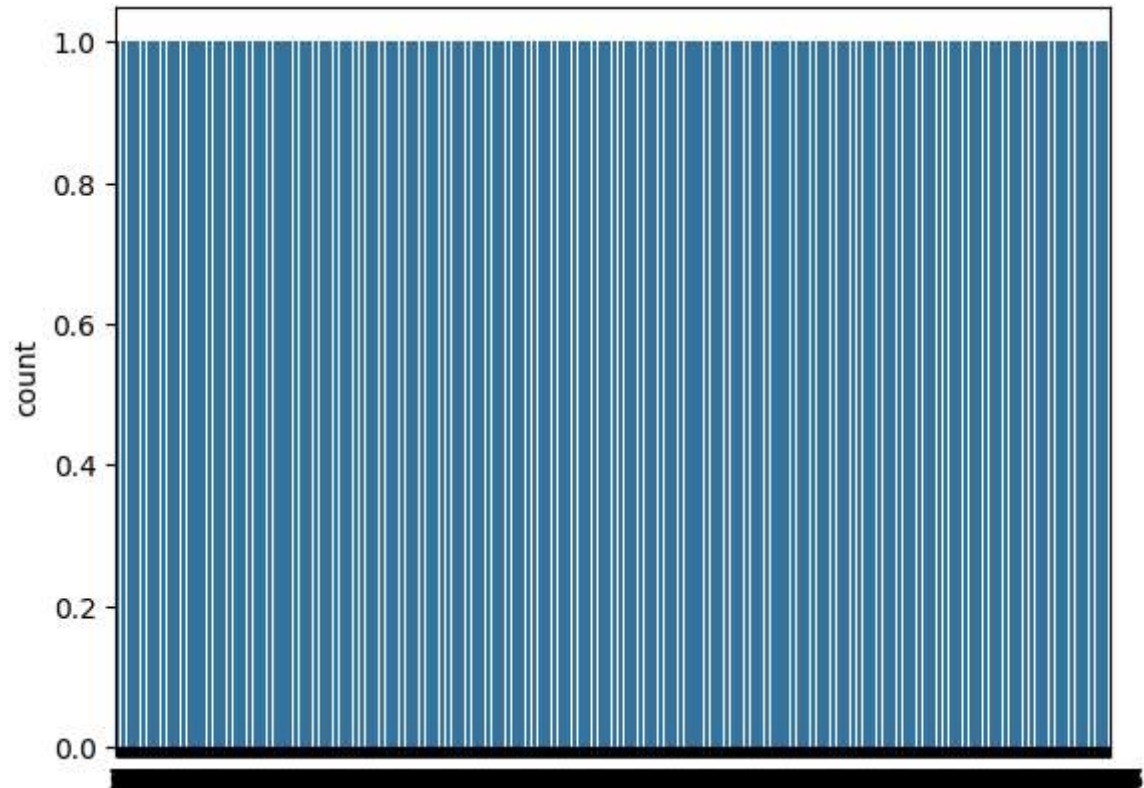
```
sns.countplot(df["age"])
```

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



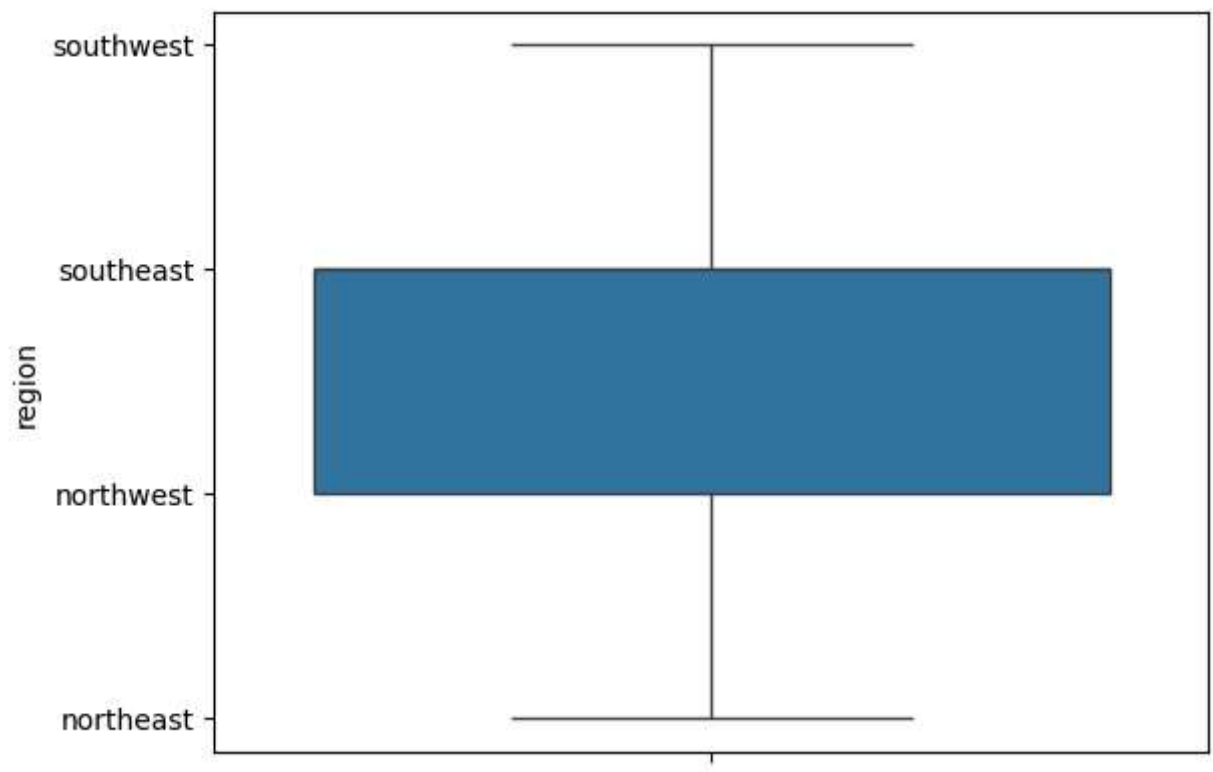
```
sns.countplot(df["bmi"])
```

<Axes: ylabel='count'>



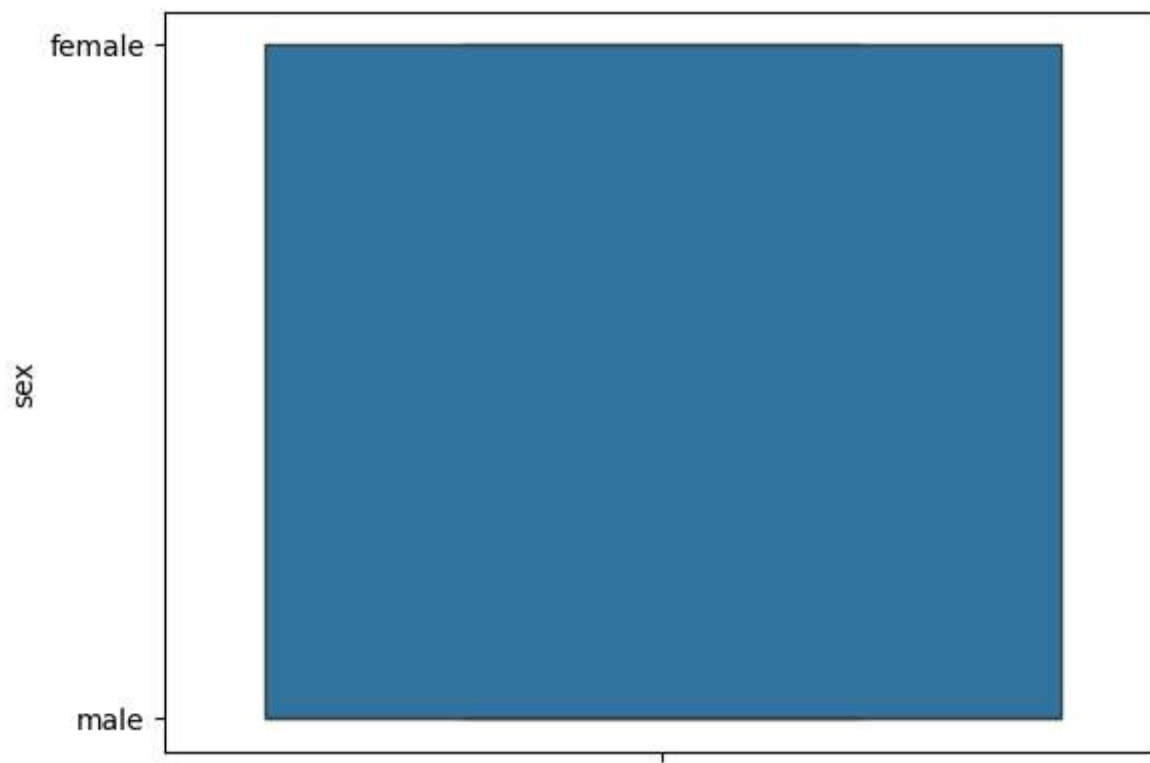
```
sns.boxplot(df["region"])
```

<Axes: ylabel='region'>



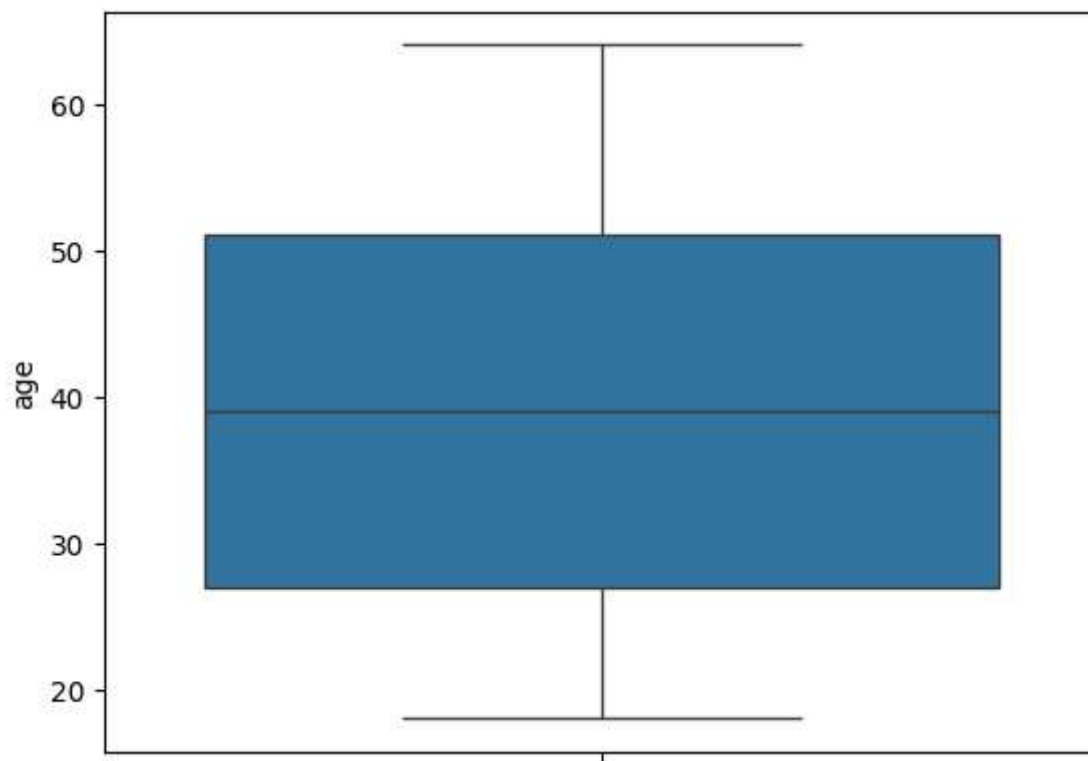
```
sns.boxplot(df["sex"])
```

<Axes: ylabel='sex'>



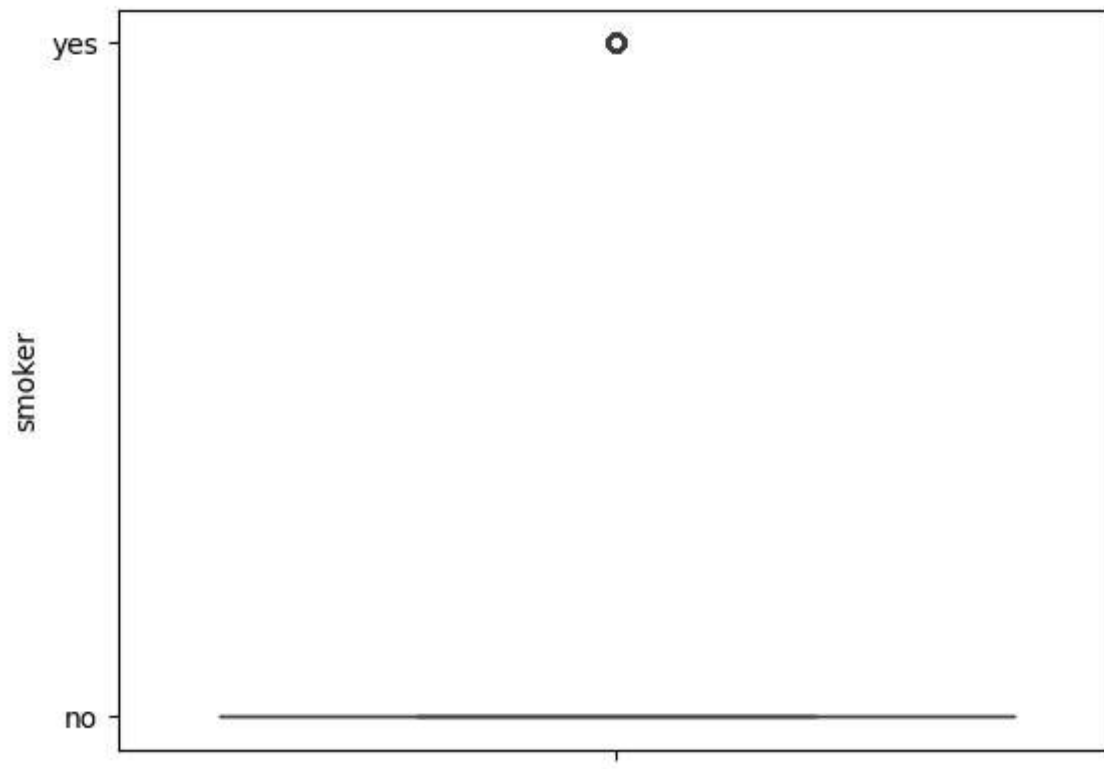
```
sns.boxplot(df["age"])
```

<Axes: ylabel='age'>



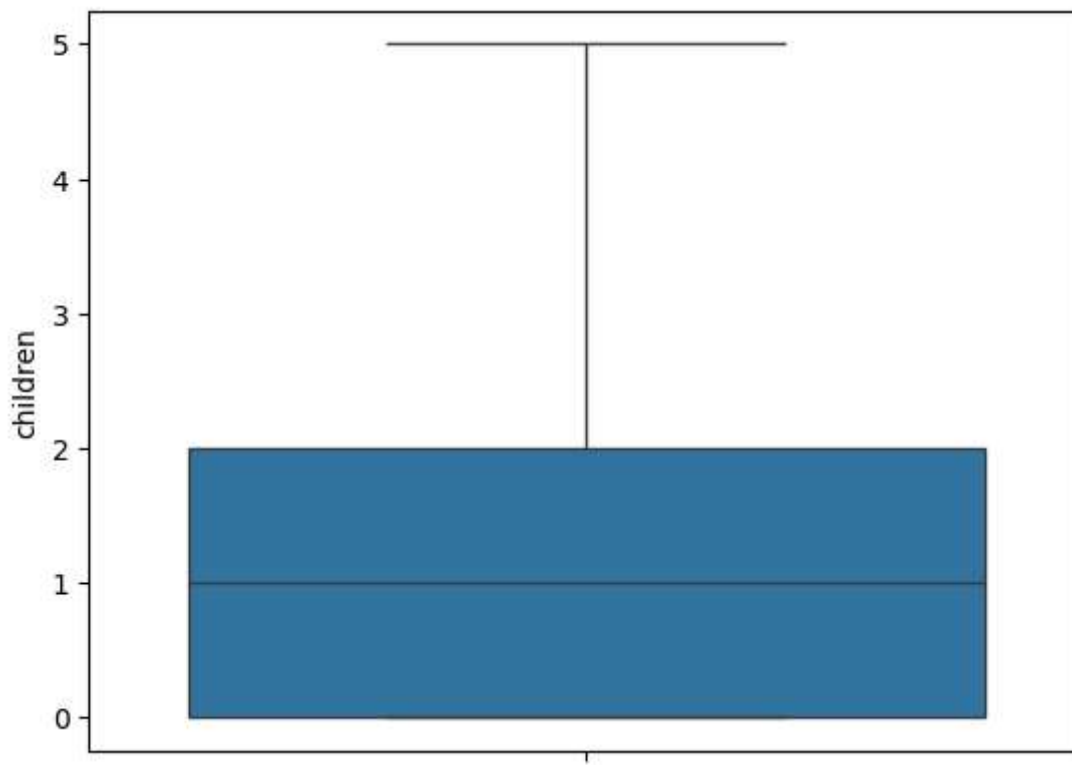
```
sns.boxplot(df["smoker"])
```

<Axes: ylabel='smoker'>



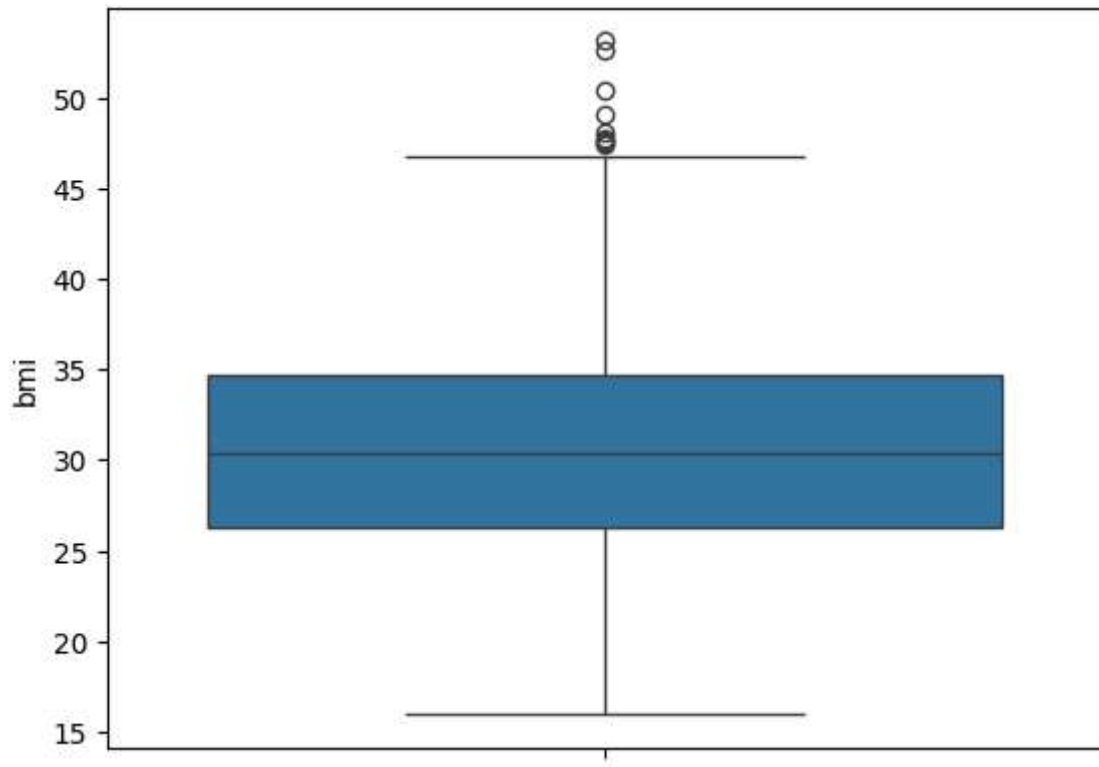
```
sns.boxplot(df["children"])
```

<Axes: ylabel='children'>



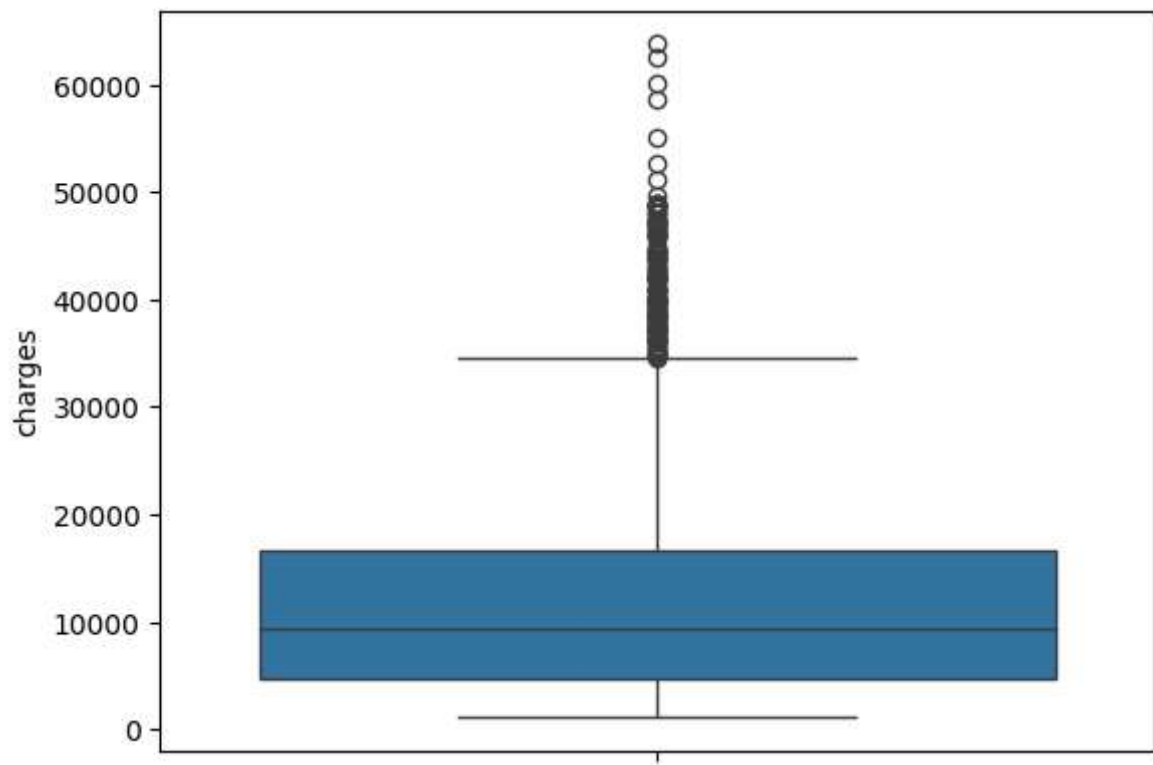
```
sns.boxplot(df["bmi"])
```

<Axes: ylabel='bmi'>



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



<Axes: ylabel='charges'>




```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df["sex"]=le.fit_transform(df["sex"])
df["age"]=le.fit_transform(df["age"])
df["region"]=le.fit_transform(df["region"])
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
a=["sex","smoker","region"]
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	yes	3	16884.92400	
1	0	1	33.770	1	no	2	1725.55230	
2	10	1	33.000	3	no	2	4449.46200	
3	15	1	22.705	0	no	1	21984.47061	
4	14	1	28.880	0	no	1	3866.85520	

Next steps:

[Generate code with df](#)
[New interactive sheet](#)

list orderd hota square[] se banti hai aur ye new table hoti hai

✓ bivariate analysis

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

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

