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

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

```
print("Dataset loaded successfully!")
```

```
Dataset loaded successfully!
```

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

```
np.float64(39.20702541106129)
```

```
df.isnull()
```

	age	sex	bmi	children	smoker	region	charges
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
1333	False	False	False	False	False	False	False
1334	False	False	False	False	False	False	False
1335	False	False	False	False	False	False	False
1336	False	False	False	False	False	False	False
1337	False	False	False	False	False	False	False

1338 rows × 7 columns

```
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
52	29
50	29
48	29
51	29
45	29
24	28
27	28
28	28
26	28
23	28
49	28
54	28
57	26
34	26
56	26
35	25
58	25
37	25
59	25
39	25
36	25
38	25
62	23
60	23
63	23

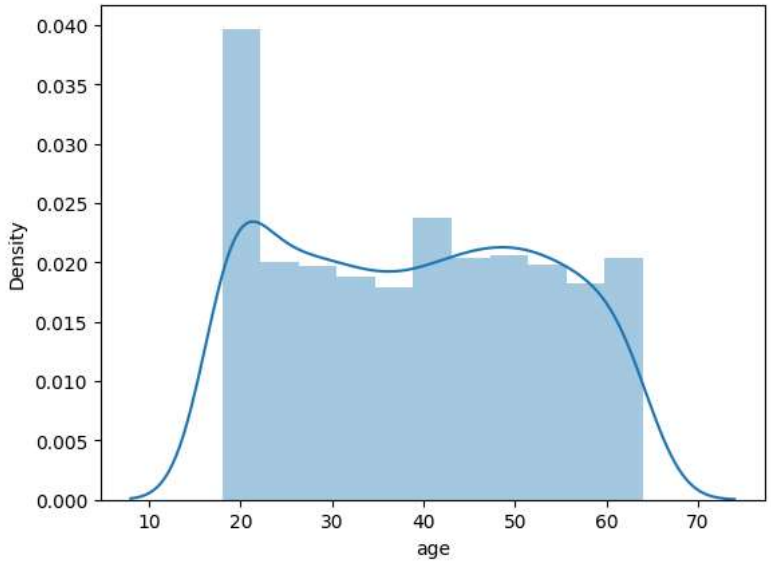
```
df['sex'].value_counts()
```

sex	count
male	676
female	662

dtype: int64

```
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'>
```

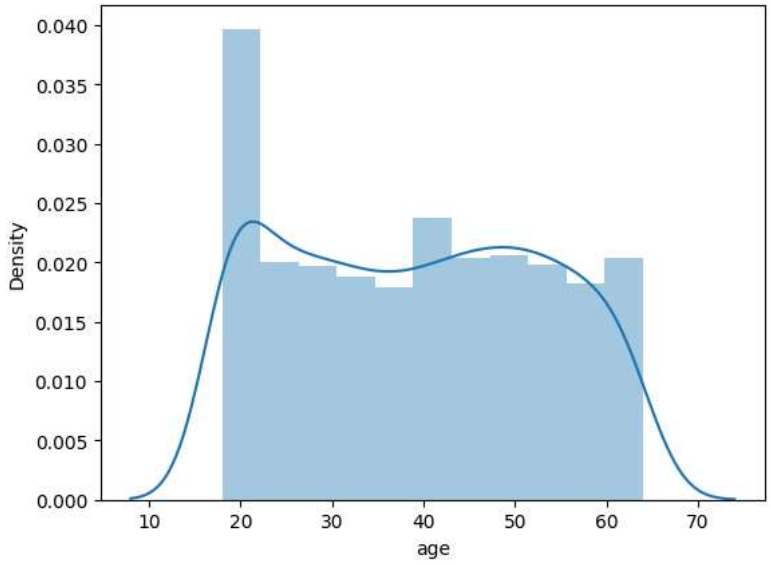


Double-click (or enter) to edit

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

```
61      23
64      22
dtype: int64
```

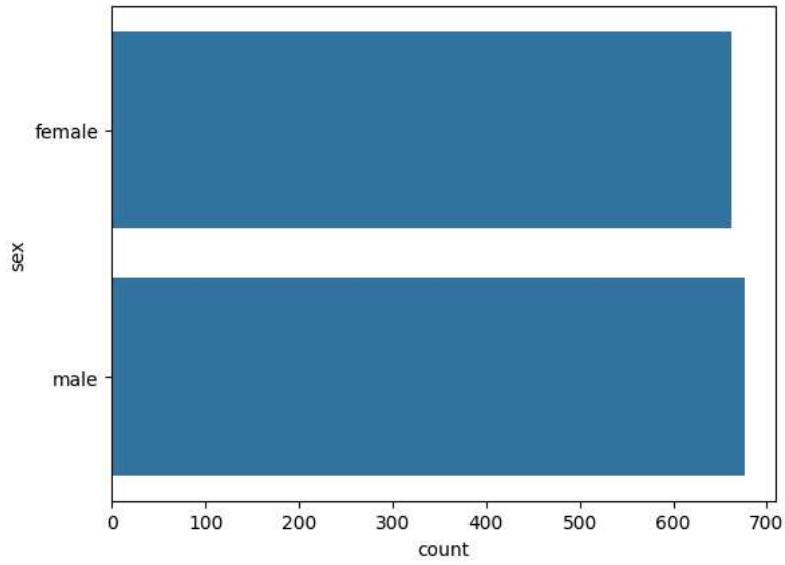
```
/tmp/ipython-input-3234920688.py:1: UserWarning:

`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'>
```



Start coding or generate with AI.

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

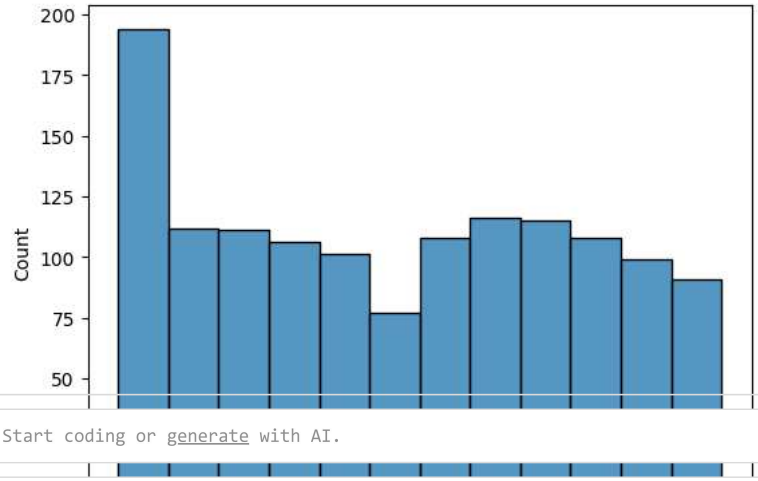
```
<Axes: xlabel='count', ylabel='sex'>
```



Start coding or generate with AI.

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

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




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

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

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