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

```
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.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1338 non-null   int64
 1   sex       1338 non-null   object
 2   bmi       1338 non-null   float64
 3   children  1338 non-null   int64
 4   smoker    1338 non-null   object
 5   region    1338 non-null   object
 6   charges   1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

```
df.describe()
```

| | age | bmi | children | charges |
|---|---|---|---|---|
| **count** | 1338.000000 | 1338.000000 | 1338.000000 | 1338.000000 |
| **mean** | 39.207025 | 30.663397 | 1.094918 | 13270.422265 |
| **std** | 14.049960 | 6.098187 | 1.205493 | 12110.011237 |
| **min** | 18.000000 | 15.960000 | 0.000000 | 1121.873900 |
| **25%** | 27.000000 | 26.296250 | 0.000000 | 4740.287150 |
| **50%** | 39.000000 | 30.400000 | 1.000000 | 9382.033000 |
| **75%** | 51.000000 | 34.693750 | 2.000000 | 16639.912515 |
| **max** | 64.000000 | 53.130000 | 5.000000 | 63770.428010 |

```
df.shape
```

```
(1338, 7)
```

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

| | 0 |
|---|---|
| **age** | 0 |
| **sex** | 0 |
| **bmi** | 0 |
| **children** | 0 |
| **smoker** | 0 |
| **region** | 0 |
| **charges** | 0 |

**dtype:** int64

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

```
np.float64(1.0949177877429)
```

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

| smoker | count |
|---|---|
| **no** | 1064 |
| **yes** | 274 |

**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 |
| 52 | 29 |
| 50 | 29 |
| 48 | 29 |
| 51 | 29 |
| 45 | 29 |
| 24 | 28 |
| 27 | 28 |
| 28 | 28 |
| 26 | 28 |
| 23 | 28 |
| 49 | 28 |
| 54 | 28 |

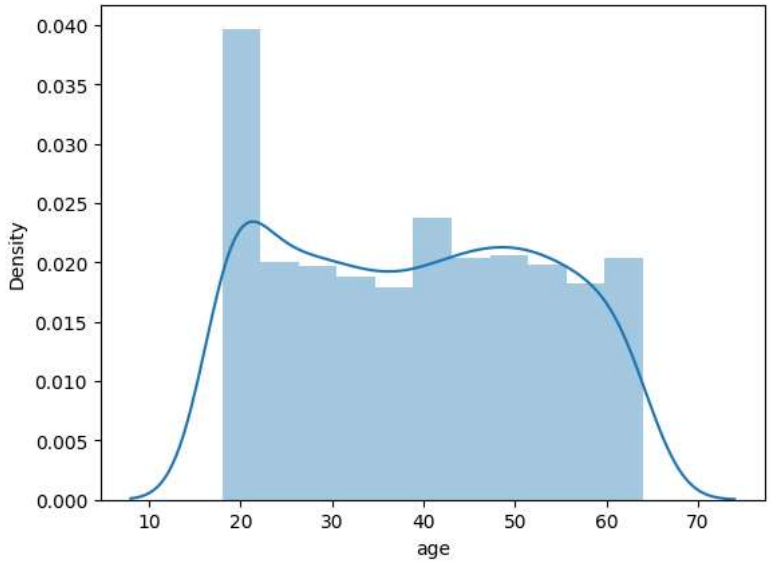
```
df['sex'].value_counts()
```

| 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'>
```

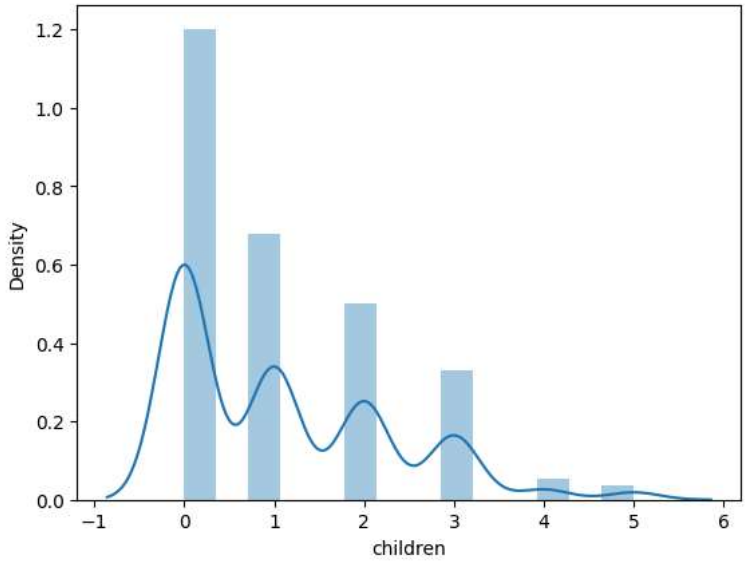



```
sns.distplot(df['children'])
```

| | |
|---|---|
| 55 | 26 |
| 56 | 26 |
| 35 | 25 |
| 58 | 25 |
| 37 | 25 |
| 59 | 25 |
| 39 | 25 |
| 36 | 25 |
| 38 | 25 |
| 62 | 23 |
| 60 | 23 |
| 63 | 23 |

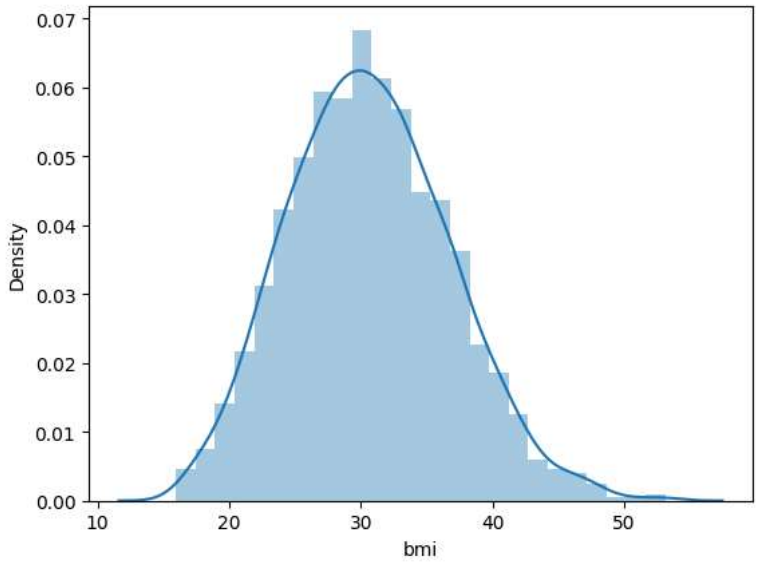
```
61    23
64    22
dtype: int64
```

```
/tmp/ipython-input-2914109155.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'>
```



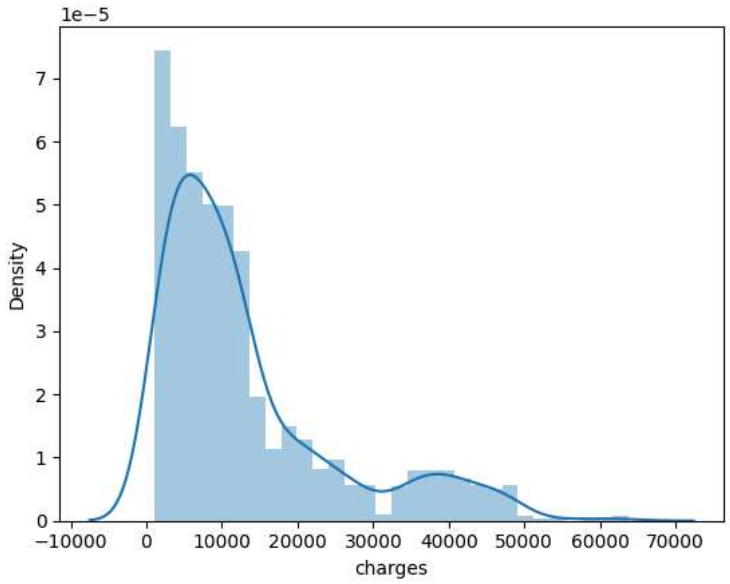

```
sns.distplot(df['bmi'])
```

```
/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'>
```



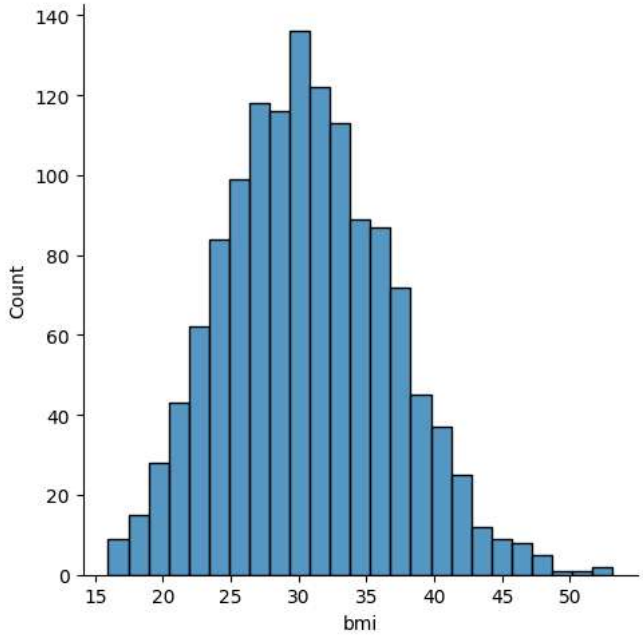

```
sns.distplot(df['charges'])
```

```
/tmp/ipython-input-1319113370.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'>
```



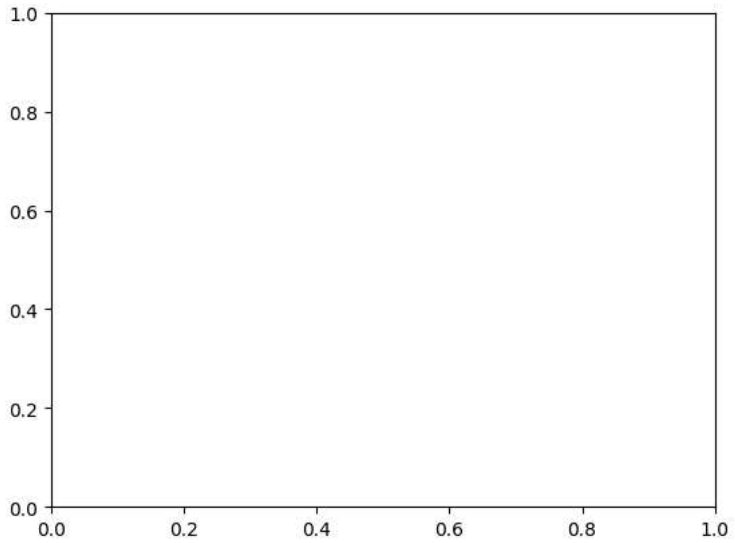

```
sns.displot(df['bmi'])
```

```
<seaborn.axisgrid.FacetGrid at 0x7ab1b2517050>
```



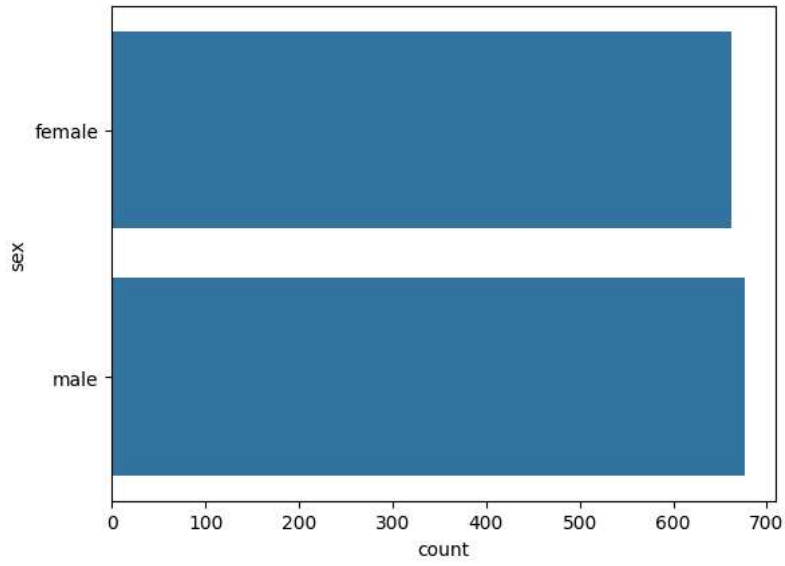

```
sns.distplot(df['smoker'])
```

```
/tmp/ipython-input-3810471177.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-3810471177.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'
```



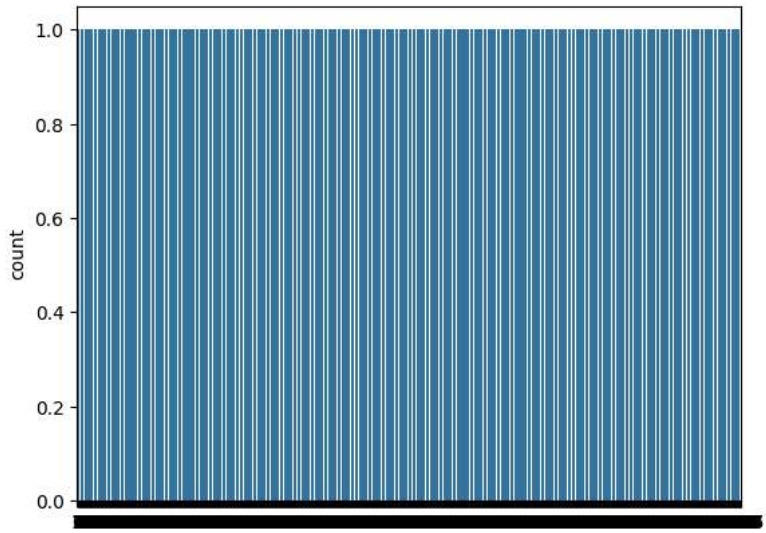

```
sns.countplot(df['sex'])
```

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



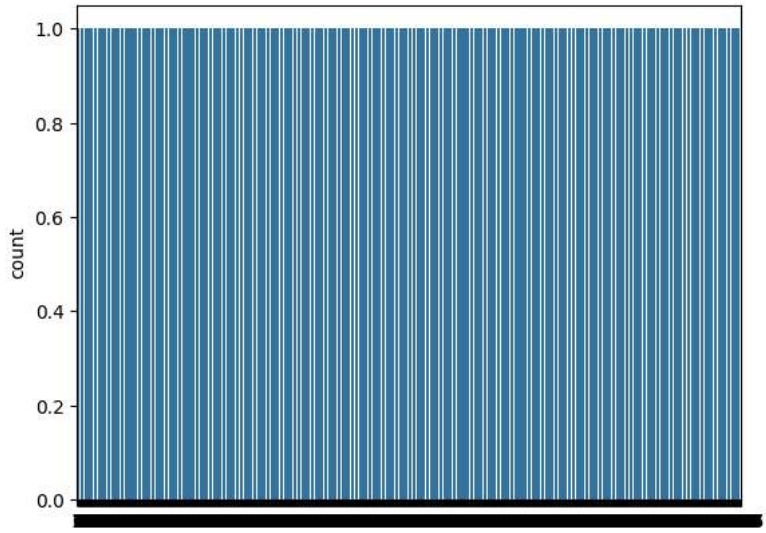

```
sns.countplot(df['children'])
```

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



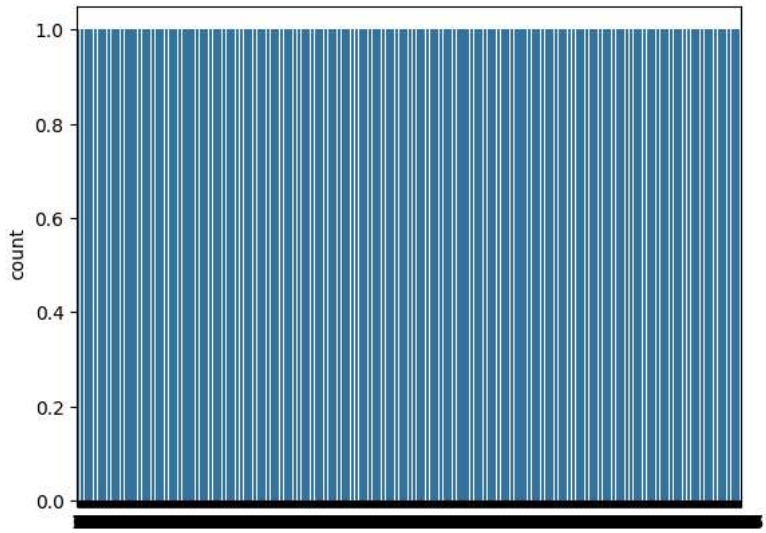

```
sns.countplot(df['age'])
```

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



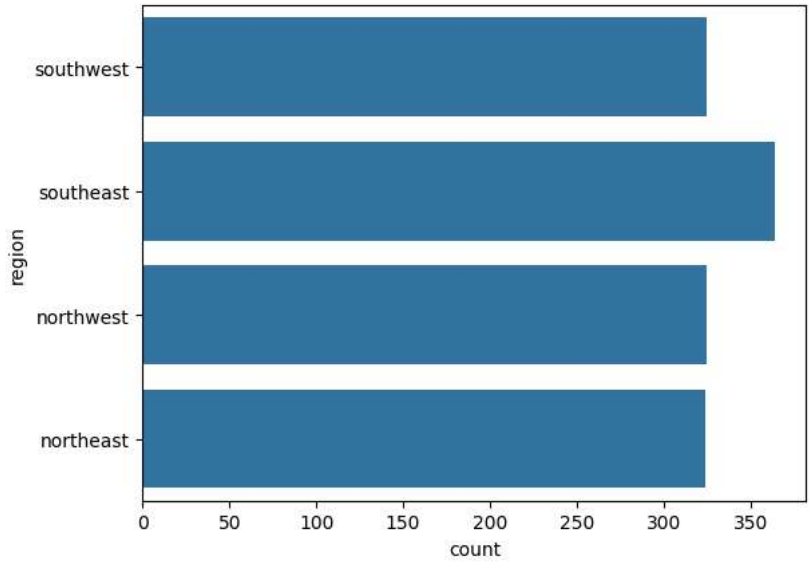

```
sns.countplot(df['bmi'])
```

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



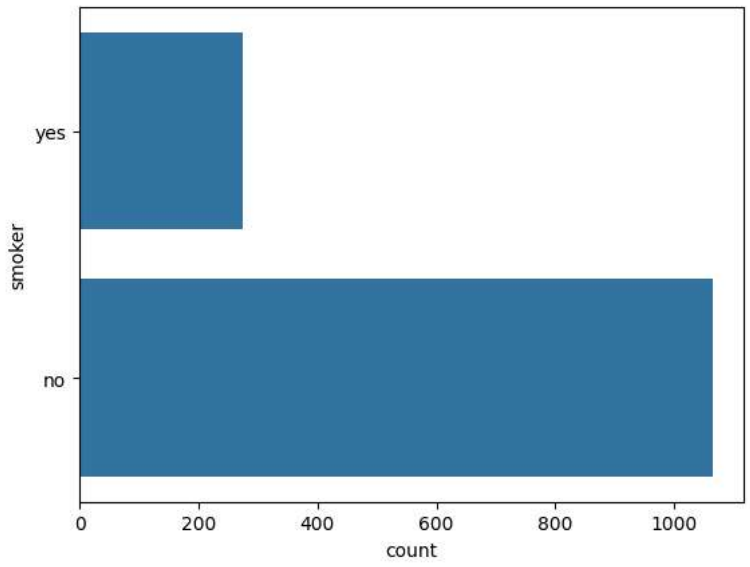
```
sns.countplot(df['region'])
```

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



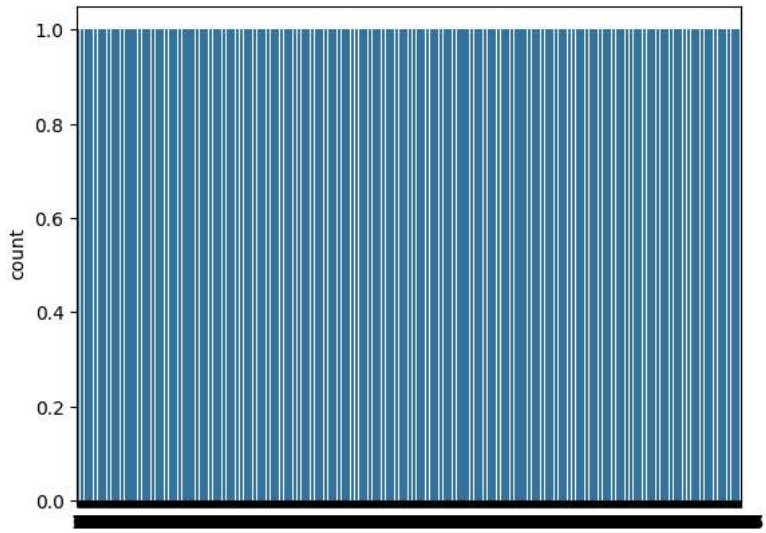

```
sns.countplot(df['smoker'])
```

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



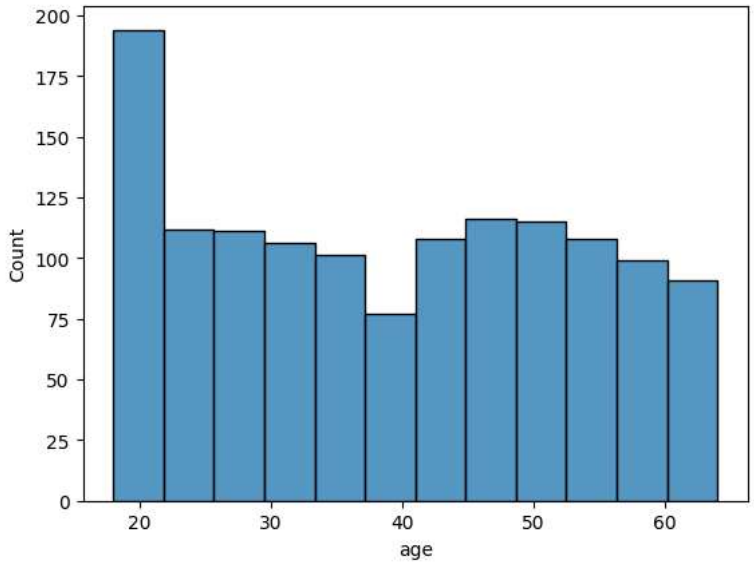

```
sns.countplot(df['charges'])
```

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



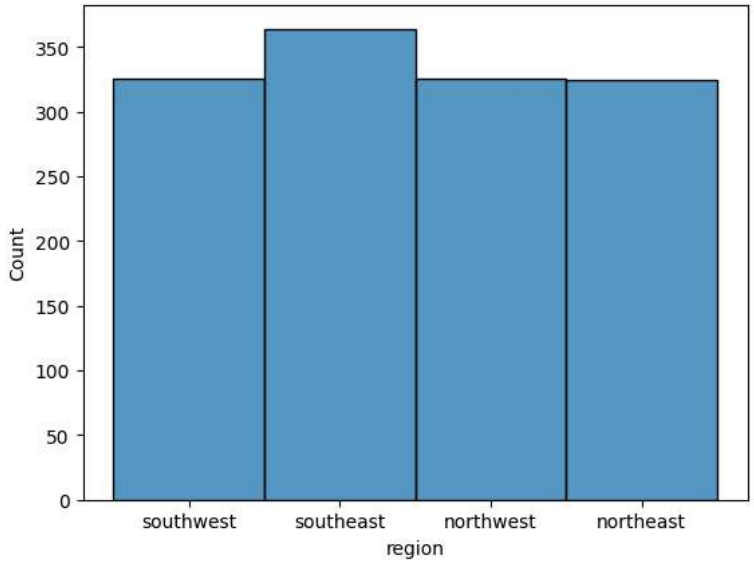

```
sns.histplot(df['age'])
```

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



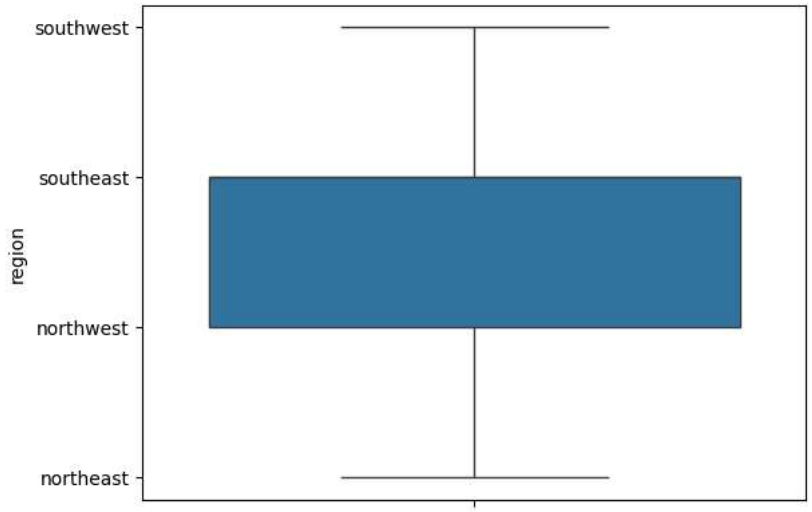

```
sns.histplot(df['region'])
```

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



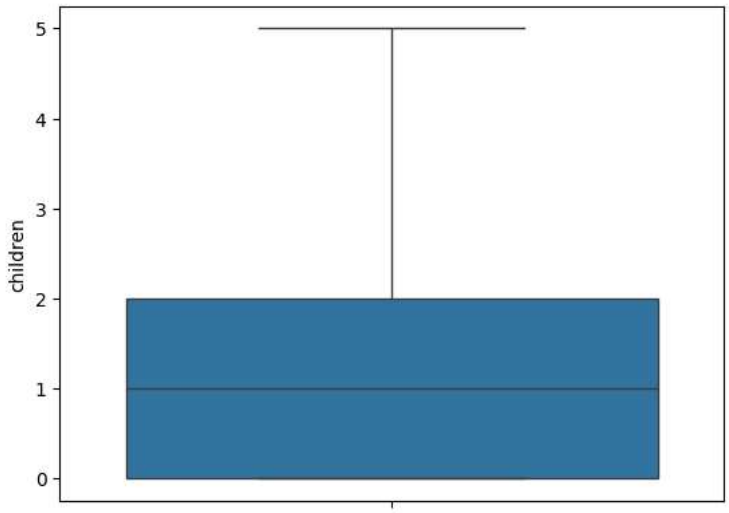

```
sns.boxplot(df['region'])
```

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



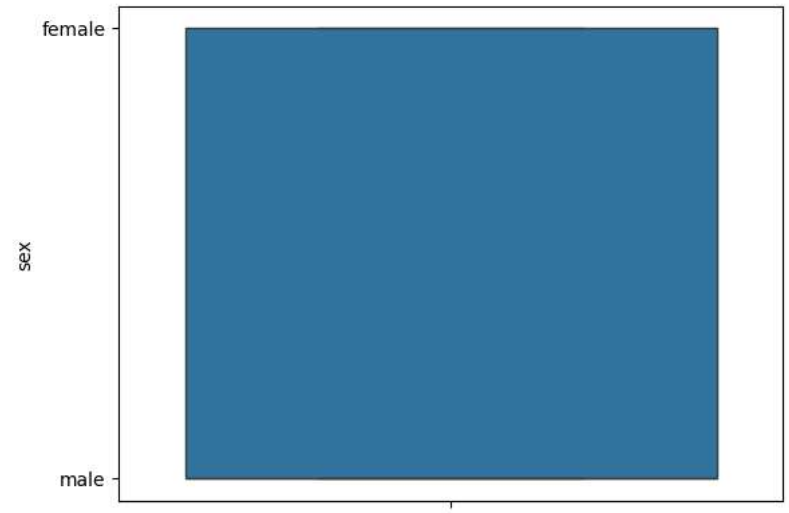

```
sns.boxplot(df['children'])
```

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



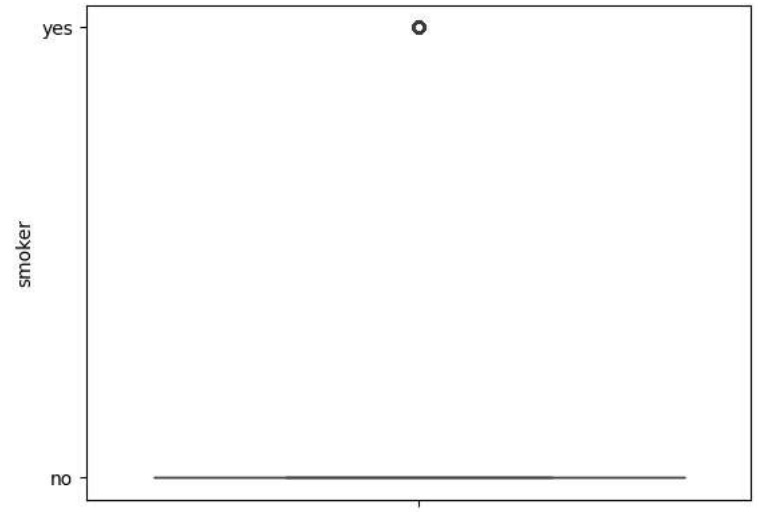
```
sns.boxplot(df['sex'])
```

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



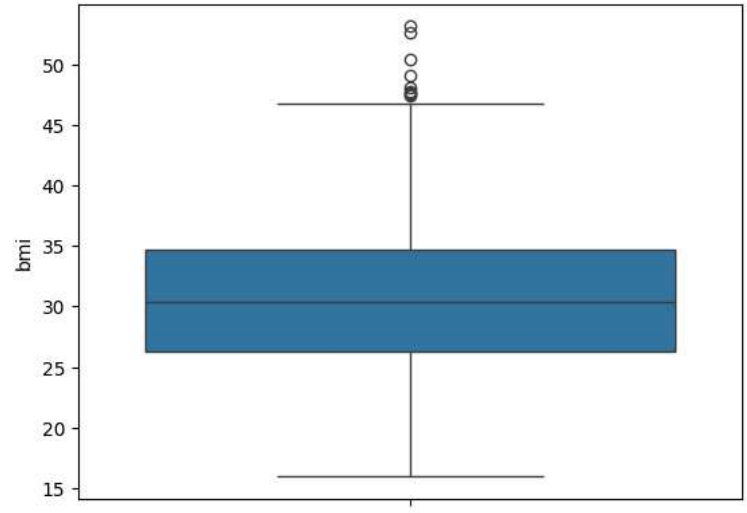

```
sns.boxplot(df['smoker'])
```

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




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

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



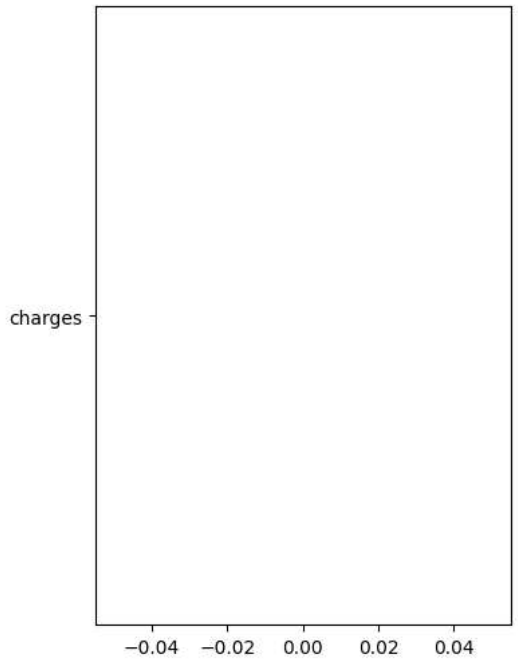
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['sex']=le.fit_transform(df['sex'])
df['smoker']=le.fit_transform(df['smoker'])
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-1182244122.py in <cell line: 0>()
      1 from sklearn.preprocessing import LabelEncoder
      2 le=LabelEncoder()
----> 3 df['sex']=le.fit_transform(df['sex'])
      4 df['smoker']=le.fit_transform(df['smoker'])

NameError: name 'df' is not defined
```

```
df.head()
```

```
from sklearn.preprocessing import labelEncoder
le=labelEncoder()
xcxxxbn
```

```
plt.figure(figsize=(4,6))
sns.lineplot("charges")
plt.show()
```




```
plt.figure(figsize=(4,6))
sns.lineplot(x="sex",y="charges",data as df)
plt.show()
```

```
  File "/tmp/ipython-input-2027948714.py", line 2
    sns.lineplot(x="sex",y="charges",data as df)
                                        ^
SyntaxError: positional argument follows keyword argument
```

```
plt.figure(figsize=(4,6))
sns.lineplot(x="sex",y="charges",data=df)
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-3253847595.py in <cell line: 0>()
      1 plt.figure(figsize=(4,6))
----> 2 sns.lineplot(x="sex",y="charges",data=df)
      3 plt.show()
```