

```
import numpy as np
import pandas as pd
```

```
df=pd.read_csv('/content/insurance - insurance.csv')
df.head(2)
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.90	0	yes	southwest	16884.9240
1	18	male	33.77	1	no	southeast	1725.5523

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

```
x=df.drop(columns=['charges'])
y=df['charges']
```

```
x.head(2)
```

	age	sex	bmi	children	smoker	region
0	19	female	27.90	0	yes	southwest
1	18	male	33.77	1	no	southeast

Next steps: [Generate code with x](#) [New interactive sheet](#)

```
from sklearn.preprocessing import LabelEncoder
```

```
lb=LabelEncoder()
```

```
x['sex']=lb.fit_transform(x['sex'])
x['smoker']=lb.fit_transform(x['smoker'])
x['region']=lb.fit_transform(x['region'])
```

```
x.head(2)
```

	age	sex	bmi	children	smoker	region
0	19	0	27.90	0	1	3
1	18	1	33.77	1	0	2

Next steps: [Generate code with x](#) [New interactive sheet](#)

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

```
from sklearn.tree import DecisionTreeRegressor
```

```
df=DecisionTreeRegressor()
```

```
df.fit(x_train,y_train)
```

▼ DecisionTreeRegressor ⓘ ?
DecisionTreeRegressor()

```
y_pred=df.predict(x_test)
```

```
from sklearn.metrics import r2_score
```

```
r2_score(y_test,y_pred)
```

```
0.763153840958151
```

```
df=pd.read_csv('/content/covid_toy - covid_toy.csv')
df.head(2)
```

	age	gender	fever	cough	city	has_covid
0	60	Male	103.0	Mild	Kolkata	No
1	27	Male	100.0	Mild	Delhi	Yes

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

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

	0
age	0
gender	0
fever	10
cough	0
city	0
has_covid	0

dtype: int64

```
df['fever']=df['fever'].fillna(df['fever'].mean())
```

```
from sklearn.preprocessing import LabelEncoder
```

```
lb=LabelEncoder()
```

```
df['gender']=lb.fit_transform(df['gender'])
df['city']=lb.fit_transform(df['city'])
df['cough']=lb.fit_transform(df['cough'])
df['has_covid']=lb.fit_transform(df['has_covid'])
```

```
df.head(2)
```

	age	gender	fever	cough	city	has_covid
0	60	1	103.0	0	2	0
1	27	1	100.0	0	1	1

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

```
x=df.drop(columns=['has_covid'])
y=df['has_covid']
```

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
dt=DecisionTreeClassifier()
```

```
dt.fit(x_train,y_train)
```

▼ DecisionTreeClassifier ⓘ ?
DecisionTreeClassifier()

```
y_pred=dt.predict(x_test)
```

```
from sklearn.metrics import accuracy_score
```

```
accuracy_score(y_test,y_pred)
```

0.5