


```
import pandas as pd
```

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

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

| | 0 |
|-----------|---|
| age | 0 |
| gender | 0 |
| fever | 0 |
| cough | 0 |
| city | 0 |
| has_covid | 0 |

dtype: int64

```
df=df.drop(columns=['gender'])
```

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

| | age | fever | cough | has_covid |
|---|-----|-------|-------|-----------|
| 0 | 60 | 103.0 | Mild | No |
| 1 | 27 | 100.0 | Mild | Yes |



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=42)
```

```
print(x_train.shape)  
print(x_test.size)  
print(y_train.shape)  
print(x_test.shape)  
print(y_test.shape)
```

```
(80, 5)  
100  
(80,)  
(20, 5)  
(20,)
```