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

```
df=pd.read_csv('/content/covid_toy - covid_toy (1).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.drop(columns=['age','fever'])
```

	gender	cough	city	has_covid
0	Male	Mild	Kolkata	No
1	Male	Mild	Delhi	Yes
2	Male	Mild	Delhi	No
3	Female	Mild	Kolkata	No
4	Female	Mild	Mumbai	No
...	...	...	...	...
95	Female	Mild	Bangalore	No
96	Female	Strong	Kolkata	Yes
97	Female	Mild	Bangalore	No
98	Female	Strong	Mumbai	No
99	Female	Strong	Kolkata	Yes

100 rows × 4 columns

```
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=2)
```

```
a=pd.get_dummies(df[['gender','city','cough','has_covid']],drop_first=True)
```

```
a=a.astype(int)
```