```
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

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

	0
age	0
sex	0
bmi	0
children	0
smoker	0
region	0
charges	0

dtype: int64

```
df['sex']=df['sex'].map({'female':1,'male':0})
```

```
df['smoker']=df['smoker'].map({'yes':1,'no':0})
```

```
df['region']=df['region'].map({"yes":1,"no":0})
```

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

	age	sex	bmi	children	smoker	region	charges
0	19	1	27.90	0	1	NaN	16884.9240
1	18	0	33.77	1	0	NaN	1725.5523

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

charges	count
1639.56310	2
2221.56445	1
19798.05455	1
13063.88300	1
13555.00490	1
...	...
1149.39590	1
37079.37200	1
4738.26820	1
2897.32350	1
4762.32900	1

1337 rows × 1 columns

dtype: int64

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

	age	sex	bmi	children	smoker	region
0	19	1	27.900	0	1	NaN
1	18	0	33.770	1	0	NaN
2	28	0	33.000	3	0	NaN
3	33	0	22.705	0	0	NaN
4	32	0	28.880	0	0	NaN
...	...	...	...	...	...	...
1333	50	0	30.970	3	0	NaN
1334	18	1	31.920	0	0	NaN
1335	18	1	36.850	0	0	NaN
1336	21	1	25.800	0	0	NaN
1337	61	1	29.070	0	1	NaN

1338 rows × 6 columns

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

```
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=3)
```

APPLY NORMALIZATION

```
from sklearn.preprocessing import MinMaxScaler
```

```
MM = MinMaxScaler()
```

```
np.round(x_train.describe(),2)
```

	age	sex	children	smoker	region	charges
count	1070.00	1070.0	1070.0	1070.0	0.0	1070.00
mean	39.24	0.5	1.1	0.2	NaN	13222.99
std	13.95	0.5	1.2	0.4	NaN	12066.25
min	18.00	0.0	0.0	0.0	NaN	1131.51
25%	27.00	0.0	0.0	0.0	NaN	4686.74
50%	39.00	0.5	1.0	0.0	NaN	9369.62
75%	51.00	1.0	2.0	0.0	NaN	16791.38
max	64.00	1.0	5.0	1.0	NaN	63770.43

```
mn = MM.fit_transform(x_train)
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/_array_api.py:776: RuntimeWarning: All-NaN slice encountered
  return xp.asarray(numpy.nanmin(X, axis=axis))
/usr/local/lib/python3.12/dist-packages/sklearn/utils/_array_api.py:793: RuntimeWarning: All-NaN slice encountered
  return xp.asarray(numpy.nanmax(X, axis=axis))
```

```
new=pd.DataFrame(mn,columns=x_train.columns)
```

```
np.round(new.describe(),2)
```

	age	sex	children	smoker	region	charges
count	1070.00	1070.0	1070.00	1070.0	0.0	1070.00
min	0.00	0.0	0.00	0.0	NaN	0.00
25%	0.20	0.0	0.00	0.0	NaN	0.06
50%	0.46	0.5	0.20	0.0	NaN	0.13
75%	0.72	1.0	0.40	0.0	NaN	0.25
max	1.00	1.0	1.00	1.0	NaN	1.00

Start coding or generate with AI.