

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

ins = pd.read_csv("insurance.csv")
ins.head()
```

```
Out[1]:
```

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
1	18	male	33.8	1	no	southeast	1725.55
2	28	male	33.0	3	no	southeast	4449.46
3	33	male	22.7	0	no	northwest	21984.47
4	32	male	28.9	0	no	northwest	3866.86

```
In [2]: ins.shape
```

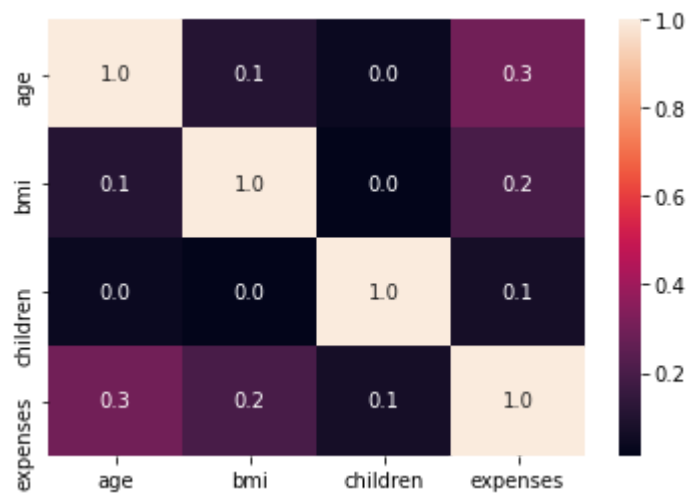
```
Out[2]: (1338, 7)
```

```
In [3]: ins.isnull().sum()
```

```
Out[3]: age      0
sex      0
bmi      0
children  0
smoker   0
region   0
expenses  0
dtype: int64
```

```
In [4]: sns.heatmap(ins.corr(), annot=True, fmt="0.1f")
```

```
Out[4]: <AxesSubplot:>
```



In [5]: `ins.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    expenses    1338 non-null   float64  
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

In [6]: `ins.head()`

Out[6]:

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
1	18	male	33.8	1	no	southeast	1725.55
2	28	male	33.0	3	no	southeast	4449.46
3	33	male	22.7	0	no	northwest	21984.47
4	32	male	28.9	0	no	northwest	3866.86

In [7]: `# Label Encoding`

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
ins["sex"] = le.fit_transform(ins["sex"])
ins["region"] = le.fit_transform(ins["region"])
ins["smoker"] = le.fit_transform(ins["smoker"])
```

In [8]: `ins.head()`

Out[8]:

	age	sex	bmi	children	smoker	region	expenses
0	19	0	27.9	0	1	3	16884.92
1	18	1	33.8	1	0	2	1725.55
2	28	1	33.0	3	0	2	4449.46
3	33	1	22.7	0	0	1	21984.47
4	32	1	28.9	0	0	1	3866.86

In [9]: *# Seperate inputs & output*

```
X = ins.iloc[:, :-1]
Y = ins.iloc[:, -1]
```

In [11]: *# train test split*

```
from sklearn.model_selection import train_test_split

x1,x2,y1,y2 = train_test_split(X,Y, test_size=0.2, random_state=42)
```

In [12]: **from** sklearn.linear_model **import** LinearRegression

```
model = LinearRegression()
model.fit(x1,y1)
```

Out[12]:

```
▼ LinearRegression
LinearRegression()
```

In [13]: `yp = model.predict(x2)`

In [15]: `yp[:2]`

Out[15]: `array([8931.64285422, 7108.65837503])`

In [16]: `y2[:2]`

Out[16]: `764 9095.07
887 5272.18
Name: expenses, dtype: float64`

In [17]: `model.score(x2,y2)`

Out[17]: `0.7833214205203848`

In [18]: `x2[:3]`

Out[18]:

	age	sex	bmi	children	smoker	region
764	45	0	25.2	2	0	0
887	36	0	30.0	0	0	1
890	64	0	26.9	0	1	1

```
In [19]: n=[64,0,26.9,0,1,1]
n = np.array(n).reshape(1,-1)
model.predict(n)
```

C:\Users\Indoskill\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
warnings.warn(

```
Out[19]: array([36911.24588592])
```

```
In [ ]:
```