In [35]:
```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings("ignore")
```

In [36]:
```
data=pd.read_csv("C:/Users/NARESH SANA/Downloads/insurance.csv")
```

In [37]:
```
data
```

Out[37]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

In [38]:
```
data.head()
```

Out[38]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

In [39]:
```
data.tail()
```

Out[39]:

	age	sex	bmi	children	smoker	region	charges
1333	50	male	30.97	3	no	northwest	10600.5483
1334	18	female	31.92	0	no	northeast	2205.9808
1335	18	female	36.85	0	no	southeast	1629.8335
1336	21	female	25.80	0	no	southwest	2007.9450
1337	61	female	29.07	0	yes	northwest	29141.3603

In [40]:
```
data.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   charges   1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

In [41]:
```
data1= data.drop(['region'],axis=1)
data1
```

Out[41]:

	age	sex	bmi	children	smoker	charges
0	19	female	27.900	0	yes	16884.92400
1	18	male	33.770	1	no	1725.55230
2	28	male	33.000	3	no	4449.46200
3	33	male	22.705	0	no	21984.47061
4	32	male	28.880	0	no	3866.85520
...	...	...	...	...	...	...
1333	50	male	30.970	3	no	10600.54830
1334	18	female	31.920	0	no	2205.98080
1335	18	female	36.850	0	no	1629.83350
1336	21	female	25.800	0	no	2007.94500
1337	61	female	29.070	0	yes	29141.36030

1338 rows × 6 columns

In [42]:
```
data.isnull().sum()
```

Out[42]:
```
age         0
sex         0
bmi         0
children    0
smoker      0
region      0
charges     0
dtype: int64
```

In [43]:
```
x = data1.iloc[:,:-1]
```

In [44]:
```
x
```

Out[44]:

	age	sex	bmi	children	smoker
0	19	female	27.900	0	yes
1	18	male	33.770	1	no
2	28	male	33.000	3	no
3	33	male	22.705	0	no
4	32	male	28.880	0	no
...	...	...	...	...	...
1333	50	male	30.970	3	no
1334	18	female	31.920	0	no
1335	18	female	36.850	0	no
1336	21	female	25.800	0	no
1337	61	female	29.070	0	yes

1338 rows × 5 columns

In [45]:
```
y=data1.iloc[:,-1]
```

In [46]:
```
y
```

Out[46]:
```
0       16884.92400
1        1725.55230
2        4449.46200
3       21984.47061
4        3866.85520
           ...
1333    10600.54830
1334     2205.98080
1335     1629.83350
1336     2007.94500
1337    29141.36030
Name: charges, Length: 1338, dtype: float64
```

In [47]:
```
X=pd.get_dummies(x,dtype=int)
X
```

Out[47]:

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

1338 rows × 7 columns

In [48]:
```
cor_mat = X.corr()
cor_mat
```

Out[48]:

	age	bmi	children	sex_female	sex_male	smoker_no	smoker_yes
age	1.000000	0.109272	0.042469	0.020856	-0.020856	0.025019	-0.025019
bmi	0.109272	1.000000	0.012759	-0.046371	0.046371	-0.003750	0.003750
children	0.042469	0.012759	1.000000	-0.017163	0.017163	-0.007673	0.007673
sex_female	0.020856	-0.046371	-0.017163	1.000000	-1.000000	0.076185	-0.076185
sex_male	-0.020856	0.046371	0.017163	-1.000000	1.000000	-0.076185	0.076185
smoker_no	0.025019	-0.003750	-0.007673	0.076185	-0.076185	1.000000	-1.000000
smoker_yes	-0.025019	0.003750	0.007673	-0.076185	0.076185	-1.000000	1.000000

In [1]:
```
#import seaborn as sns
#sns.heatmap(cor_mat,vmax=1,vmin=-1,annot=True,linewidth='bwr')
```

In [49]:
```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=20,random_stat
```

In [50]:
```
X_test.head()
```

Out[50]:

	age	bmi	children	sex_female	sex_male	smoker_no	smoker_yes
671	29	31.160	0	1	0	1	0
1159	32	41.100	0	1	0	1	0
1157	23	23.180	2	1	0	1	0
738	23	31.730	3	0	1	0	1
56	58	31.825	2	1	0	1	0

In [51]:
```
X_train.count()
```

Out[51]:
```
age           1318
bmi           1318
children      1318
sex_female    1318
sex_male      1318
smoker_no     1318
smoker_yes    1318
dtype: int64
```

In [52]:
```
y_train.count()
```

Out[52]:
```
1318
```

In [53]:
```
y_test.count()
```

Out[53]:
```
20
```

In [54]:
```
from sklearn.linear_model import LinearRegression
reg=LinearRegression()
reg.fit(X_train,y_train)
```

Out[54]:
```
LinearRegression()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [55]:
```
y_pred=reg.predict(X_test)
y_pred
```

Out[55]:
```
array([ 5512.69374566,  9523.56489192,  2266.79687819, 29117.10492733,
       14112.59093009,  2599.36887173,  8760.22404016,  9249.53722935,
       13219.63576253,  5144.14417181, 10388.1849229 ,   845.91668316,
        9101.07611279, 12953.783919  ,  2921.96317579, 11942.78929432,
       39528.59811155,  6136.13854066,  7356.81107715, 13044.78639652])
```

In [56]:
```
from sklearn.metrics import r2_score
r2_score(y_test,y_pred)
```

Out[56]:
```
0.8564740033910105
```

In [ ]:
```

```