

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

```
In [22]: df=pd.read_csv('C://Users//ALWAYSRAMESH//Downloads//insurance.csv')
```

```
In [23]: df.head(50)
```

Out[23]:

	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
5	31	female	25.740	0	no	southeast	3756.62160
6	46	female	33.440	1	no	southeast	8240.58960
7	37	female	27.740	3	no	northwest	7281.50560
8	37	male	29.830	2	no	northeast	6406.41070
9	60	female	25.840	0	no	northwest	28923.13692
10	25	male	26.220	0	no	northeast	2721.32080
11	62	female	26.290	0	yes	southeast	27808.72510
12	23	male	34.400	0	no	southwest	1826.84300
13	56	female	39.820	0	no	southeast	11090.71780
14	27	male	42.130	0	yes	southeast	39611.75770
15	19	male	24.600	1	no	southwest	1837.23700
16	52	female	30.780	1	no	northeast	10797.33620
17	23	male	23.845	0	no	northeast	2395.17155
18	56	male	40.300	0	no	southwest	10602.38500
19	30	male	35.300	0	yes	southwest	36837.46700
20	60	female	36.005	0	no	northeast	13228.84695
21	30	female	32.400	1	no	southwest	4149.73600
22	18	male	34.100	0	no	southeast	1137.01100
23	34	female	31.920	1	yes	northeast	37701.87680
24	37	male	28.025	2	no	northwest	6203.90175
25	59	female	27.720	3	no	southeast	14001.13380
26	63	female	23.085	0	no	northeast	14451.83515
27	55	female	32.775	2	no	northwest	12268.63225
28	23	male	17.385	1	no	northwest	2775.19215
29	31	male	36.300	2	yes	southwest	38711.00000

	age	sex	bmi	children	smoker	region	charges
30	22	male	35.600	0	yes	southwest	35585.57600
31	18	female	26.315	0	no	northeast	2198.18985
32	19	female	28.600	5	no	southwest	4687.79700
33	63	male	28.310	0	no	northwest	13770.09790
34	28	male	36.400	1	yes	southwest	51194.55914
35	19	male	20.425	0	no	northwest	1625.43375
36	62	female	32.965	3	no	northwest	15612.19335
37	26	male	20.800	0	no	southwest	2302.30000
38	35	male	36.670	1	yes	northeast	39774.27630
39	60	male	39.900	0	yes	southwest	48173.36100
40	24	female	26.600	0	no	northeast	3046.06200
41	31	female	36.630	2	no	southeast	4949.75870
42	41	male	21.780	1	no	southeast	6272.47720
43	37	female	30.800	2	no	southeast	6313.75900
44	38	male	37.050	1	no	northeast	6079.67150
45	55	male	37.300	0	no	southwest	20630.28351
46	18	female	38.665	2	no	northeast	3393.35635
47	28	female	34.770	0	no	northwest	3556.92230
48	60	female	24.530	0	no	southeast	12629.89670
49	36	male	35.200	1	yes	southeast	38709.17600

In [45]: `df.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   int32  
 2   bmi          1338 non-null   float64 
 3   children    1338 non-null   int64  
 4   smoker       1338 non-null   int32  
 5   region       1338 non-null   int32  
 6   charges      1338 non-null   float64 
dtypes: float64(2), int32(3), int64(2)
memory usage: 57.6 KB
```

```
In [17]: from sklearn.preprocessing import LabelEncoder
```

```
In [49]: le=LabelEncoder()
df['sex']=le.fit_transform(df['sex'])
df['smoker']=le.fit_transform(df['smoker'])
df['region']=le.fit_transform(df['region'])
```

```
In [27]: df.head(10)
```

Out[27]:

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884.92400
1	18	1	33.770	1	0	2	1725.55230
2	28	1	33.000	3	0	2	4449.46200
3	33	1	22.705	0	0	1	21984.47061
4	32	1	28.880	0	0	1	3866.85520
5	31	0	25.740	0	0	2	3756.62160
6	46	0	33.440	1	0	2	8240.58960
7	37	0	27.740	3	0	1	7281.50560
8	37	1	29.830	2	0	0	6406.41070
9	60	0	25.840	0	0	1	28923.13692

```
In [50]: x=df.drop('charges',axis=1)
y=df['charges']
```

```
In [31]: x
```

Out[31]:

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

1338 rows × 6 columns

In [32]:

y

Out[32]:

```
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 [51]:

```
from sklearn.model_selection import train_test_split
```

In [52]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
```

In [53]:

```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
model.fit(x_train,y_train)
```

Out[53]:

LinearRegression

LinearRegression()

In [54]:

```
y_pred=model.predict(x_test)
```

```
In [55]: y_pred
```

```
Out[55]: array([11051.54909755,  9821.28110689,  37867.57220923,  16125.70579228,
   6920.27132517,  3879.38549816,  1448.91928088,  14390.17797974,
   9022.95151353,  7458.83371884,  4584.60125463,  10309.9886336 ,
   8693.03891958,  4085.25393494,  27551.60737718,  11151.0640722 ,
  11243.0536825 ,  5962.9521121 ,  8181.9015666 ,  26750.7993431 ,
  33448.59842228, 14350.03320383, 11672.89478465, 32235.7832204 ,
  4326.07702625,  9096.53607025,  1045.25196369,  10177.76672094,
  4042.60346751,  10384.28681219,  9035.98983755,  40123.71002379,
 15507.20819083, 13678.264976 ,  24481.97362849,  5059.21988589,
 12889.80609711, 30333.92625689, 33301.25091403,  3431.35787088,
  3941.81614597,  4203.90901434,  30219.19050725,  39245.56885373,
 27762.83744249, 4994.74188765, 11042.48621304,  7760.15047885,
  3569.09734756, 10613.61535955,  5544.95921408,  3397.80923785,
 32701.67144343, 38285.57836702, 16290.50463759,  6965.99677468,
  5895.27536963,  9364.94083823,  9395.1780384 ,  11722.13868077,
  1611.87873326, 38750.4981005 ,  15296.11225478, 11708.42958487,
 14076.39653066, 13904.28564489,  25798.46519738, 31953.12169371,
  1168.25915489, 10184.5995492 ,  12273.00414884, 11867.15734569,
 24808.10644113, 15908.53043993, 11198.67421883, 12631.50869281,
  6433.71238434,  9915.55343233, 29953.19794316,  38768.07351788,
 12011.54405755, 37253.64166612,  4056.21325429,  9255.50826428,
 34537.73817683, 28976.62623495,  8444.02316285,  4738.69241453,
 11959.22562859, 30006.0695852 ,  10041.58386562, 11243.48874027,
  8183.6075869 ,  9280.51490529,  8255.40224617,  7239.23538241,
 35731.00350944, 32878.29978853,  7591.7717691 ,  14921.91368481,
  4184.53547122,  8690.01064385,  6619.75457992, 31535.59819898,
 32775.00677547, 1887.67848916,  8933.68024017,  6520.27249906,
 14475.77105663, 36880.82790297,  10252.51955517, 10775.16399139,
 10192.95246113, 26581.47470665, 39936.28907748,  8453.03671416,
  143.08142864,  8874.82383918,  15117.85425873,  9557.08594807,
 35275.59070316, 7270.62037452,  16826.50981439,  9572.8088055 ,
  8159.95902395, 2952.65859719, 32706.51413703, 31283.9896012 ,
 39216.89699401, 5362.49911669,  9675.40479836,  3778.85297694,
  7946.39718647, 8585.02883773, 31341.17050506, 29551.7714624 ,
 29853.91861524, 9151.88904567, 32625.66390263,  3229.01239018,
  3529.93652932, 11054.17156002, 13442.38216447, 12761.80223436,
  5363.70249634, 15875.56674406, 15252.72853146, 2382.17016287,
 -120.56014234, 10834.07802124,  7372.12214193, 31759.88622234,
 12314.86913452, 2548.30390645,  6284.28252705,  8170.0107525 ,
  4285.24015268, 2331.14818812, 11414.21888159, 12551.18010753,
 7208.95663304, 16615.95420641, 11792.56220606, 13920.69808423,
 3134.30793579, 7262.13973297, 22758.38813544,  7596.99822972,
 5401.65993492, 5339.75438707,  6641.09944767,  5142.27041 ,
  9983.03913716, 5526.89132472,  5628.18992827,  6975.95618531,
  3673.17907317, 5521.32735633, 37913.25218948, 1337.01243212,
 12636.06438156, 8935.78276524, 13661.56267036,  5572.770716 ,
  5181.38538205, 36214.23931831,  4207.49996636, 1896.75580314,
 15163.16594007, 12674.02182014, 34823.20434979, 5093.20670396,
  5580.90282376, 31320.99694717,  5982.46375195, 1940.59597738,
  8389.18364163, 10016.84576515,  8238.45168712,  5687.97489766,
 13133.993244 , 38538.79843345, 13749.62605459, 28607.07797491,
  6685.39503417, 35610.2777963 ,  3716.13611211, 12131.97274228,
  9356.80352592, 6339.94803517, 11268.82158683, 14519.98276598,
  5175.63113265, 4233.99203814,  7768.5658748 , 1150.93881984,
 7861.51707835, 4401.34365822, 13351.6815701 , 4312.6173173 ,
 10007.10646576, 7274.06437597, 9167.81641515, 2307.24223958,
```

```
13115.48905979, 16739.15126554, 15287.81532153, 10516.20091624,
5706.41694356, 2453.74757235, 2105.41152789, 13376.27751148,
14339.33209654, 5021.89418573, 4090.605953 , 9366.33184625,
9969.15202337, 27902.13744062, 7572.78791802, 10543.34735666,
6174.31283586, 29537.11622517, 11027.94086667, 7457.00641596,
10225.18823673, 12176.62074648, 2939.88516441, 10785.5019008 ,
1443.73866643, 6998.45741682, 28534.01786505, 38460.70771961,
6225.11287288, 8374.92523275, 2386.63715545, 305.52277009,
10407.09161901, 4362.55365462, 4960.62107139, 2662.94916464,
7249.10943879, 33132.60876475, 37964.64601751, 14795.4925882 ,
8191.23299164, 16000.85775643, 33002.87197448, 9520.82908522,
33323.6218328 , 3516.92369698, 30497.30098573, 8044.20793611,
14191.55955407, 4083.10066112, 32268.03153529, 8315.02063095,
11483.73451116, 9510.21374527, 4115.91551624, 12660.02379304,
11717.00935707, 8354.17363399, 13269.77614371, 2699.20978031,
10542.22785417, 5391.53803822, 11313.67089224, 31398.88972707,
10029.81495966, 1118.21159472, 462.52270438, 39674.38015203,
9676.2207617 , 7081.42950834, 13999.68415448, 13421.82776838,
26976.22819473, 7100.7475948 , 6788.44279001, 12033.45246932,
2786.36526883, 3902.46962572, 24859.29396183, 26090.84463032,
13371.81813565, 3164.31052787, 5117.58945791, 9325.3139586 ,
12404.934023 , 23317.59720695, 30625.80462982, 10046.65063794,
23804.49406862, 2804.50033014, 11550.33636023, 7549.77464787,
8233.25736365, 312.94242569, 7720.18671702, 35401.11202663,
6152.78016128, 6237.66443531, 178.89028134, 10901.70768044,
6750.59209883, 9992.98880624, 38764.63808826, 27466.58793525,
11489.14198588, 35484.16547957, 15039.21254017, 6810.42408411,
11015.353549 , 6839.24282704, 36512.17018034, 6045.00741789,
11075.16473053, 741.96608741, 24186.22593789, 1881.15932715,
34245.135946 , 11379.34404185, 1787.47141685, 32018.68078654,
6644.72832713, 5200.52497017, 37952.77965645, 2171.21921984,
9595.04338665, 2650.53107838, 12806.89011534, 1008.0849294 ,
11113.81326429, 7092.09772183, 36343.11043514, 7160.77113975,
30438.75625726, 29472.1875026 , 7100.76722883, 10790.23315907,
1879.90465675, 2200.37960027, 3793.92758918, 12706.65566117,
37039.77367021, 9716.69560709, 345.17579449, 11387.82123255,
5036.80429065, 9779.98465515, 5576.16518582, 7260.90814745,
4311.38104385, 28280.88564325, 4363.51594298, -921.56882455,
33139.88333244, 12910.28546987, 35733.85862246, 9841.66334273,
7663.70159627, -94.4834943 , 2560.96139226, 11527.50523772,
5669.76181684, 3293.09776754, 12120.72252044, 7743.17735694,
6920.9128887 , 5459.93977214, 3049.37781558, 31786.79912085,
3518.32063003, 8595.19234733, 4796.86509127, 13024.5060374 ,
14691.59701387, 7163.95913996, 26454.5855911 , 14277.02407734,
17354.84539043, 11476.70520069])
```

In [56]: `from sklearn.metrics import r2_score, mean_absolute_error,mean_squared_error`

In [57]: `MAE=mean_absolute_error(y_test,y_pred)`

In [58]: `MAE`

Out[58]: `4013.6929857811997`

In [59]: `MSE=mean_squared_error(y_test,y_pred)`

```
In [60]: MSE
```

```
Out[60]: 33311355.281859968
```

```
In [61]: RMSE=np.sqrt(MSE)
```

```
In [62]: RMSE
```

```
Out[62]: 5771.599022962351
```

```
In [64]: Y_pred2=model.predict([[19,0,27.900,0,1,3]])
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does  
not have valid feature names, but LinearRegression was fitted with feature names  
warnings.warn(
```

```
In [66]: Y_pred2
```

```
Out[66]: array([24798.33579241])
```

```
In [ ]:
```