Performing all model to the given dataset to choose the best fit. \P

Linear Regression

Data Collection

Out[2]:

		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

Data Cleaning

[1338 rows x 7 columns]>

In [3]:	1	df	.he	ad()								
Out[3]:		age		sex	bmi	children	smoker	region	charges			
	0	19	fen	nale	27.900	0	yes	southwest	16884.92400	_		
	1	18	n	nale	33.770	1	no	southeast	1725.55230			
	2	28	n	nale	33.000	3	no	southeast	4449.46200			
	3	33	n	nale .	22,705	0	no	northwest	21984.47061			
		32						northwest				
	4	32	II	iaie	28.880	0	no	normwest	3866.85520			
In [4]:	1	df	.ta	il()								
Out[4]:		а	ıge	se	x bn	ni childre	en smoke	er regio	on charge:	s		
	133	3	50	mal	le 30.9	7	3 r	o northwe	st 10600.548	3		
	133	4	18	femal	le 31.9	2	0 r	o northea	st 2205.980	8		
	133	5	18	femal	le 36.8	5	0 r	o southea	st 1629.833	5		
	133	6	21	femal	le 25.8	0	0 r	o southwe	st 2007.9450	0		
	133	37	61	femal	le 29.0	7	0 ye	s northwe	st 29141.360	3		
In [5]:	1	df	.de	scrib	oe .							
Out[5]:	<bo< th=""><td>und</td><td>met</td><td>hod</td><td>NDFrai</td><td>me.descr</td><td>ribe of</td><td>age</td><td>e sex</td><td>bmi</td><td>children</td><td>smoker</td></bo<>	und	met	hod	NDFrai	me.descr	ribe of	age	e sex	bmi	children	smoker
	reg	ion			arges			J				
	0		19			27.900		0 yes	southwest		.92400	
	1		18 28			33.770 33.000		1 no 3 no	southeast southeast		.55230 .46200	
	2 3		33			22.705		3 no 0 no	northwest		.47061	
	4		32			28.880		0 no	northwest		.85520	
					•••				• • •			
	133		50			30.970		3 no	northwest		.54830	
	133		18			31.920		0 no	northeast		.98080	
	133	5	18	fem	nale :	36.850		0 no	southeast	1629	.83350	
	133	6	21	fem	nale :	25.800		0 no	southwest	2007	.94500	
	133	7	61	fem	nale :	29.070		0 yes	northwest	29141	.36030	

```
In [6]:
          1 df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 1338 entries, 0 to 1337
        Data columns (total 7 columns):
             Column
                       Non-Null Count Dtype
        ---
             ----
                                       ----
                       1338 non-null
         0
             age
                                       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
                       1338 non-null
                                       object
             region
         6
             charges
                       1338 non-null
                                       float64
        dtypes: float64(2), int64(2), object(3)
        memory usage: 73.3+ KB
In [7]:
          1 df.shape
Out[7]: (1338, 7)
        Data Preprocessing
In [8]:
            df.isna().any()
Out[8]: age
                    False
        sex
                    False
        bmi
                    False
        children
                    False
        smoker
                    False
        region
                    False
        charges
                    False
        dtype: bool
In [9]:
            df.isnull().sum()
```

Out[9]: age

sex bmi

children

smoker

region

charges dtype: int64

0 0

0

0

0

0

Hence, the given dataset has no null values and missing.

```
In [11]:
           1 df['age'].value_counts()
Out[11]: age
         18
                69
         19
                68
                29
         50
         51
                29
                29
         47
         46
                29
         45
                29
                29
         20
               29
         48
                29
         52
                28
         22
         49
                28
                28
         54
                28
         53
                28
         21
                28
         26
                28
         24
                28
         25
                28
         28
                28
         27
                28
         23
                27
         43
                27
          29
                27
          30
                27
         41
         42
                27
                27
         44
                27
         31
         40
                27
                26
         32
         33
                26
                26
         56
                26
         34
                26
         55
                26
         57
                25
         37
         59
                25
         58
                25
                25
         36
                25
         38
         35
                25
                25
         39
         61
                23
                23
         60
         63
                23
                23
         62
         64
                22
         Name: count, dtype: int64
```

t[13]:		age	sex	bmi	children	smoker	region	charges
	0	19	1	27.900	0	yes	southwest	16884.92400
	1	18	0	33.770	1	no	southeast	1725.55230
	2	28	0	33.000	3	no	southeast	4449.46200
	3	33	0	22.705	0	no	northwest	21984.47061
	4	32	0	28.880	0	no	northwest	3866.85520
	1333	50	0	30.970	3	no	northwest	10600.54830
	1334	18	1	31.920	0	no	northeast	2205.98080
	1335	18	1	36.850	0	no	southeast	1629.83350
	1336	21	1	25.800	0	no	southwest	2007.94500
	1337	61	1	29.070	0	yes	northwest	29141.36030

Out[14]:

	age	sex	bmi	children	smoker	region	charges
0	19	1	27.900	0	1	southwest	16884.92400
1	18	0	33.770	1	0	southeast	1725.55230
2	28	0	33.000	3	0	southeast	4449.46200
3	33	0	22.705	0	0	northwest	21984.47061
4	32	0	28.880	0	0	northwest	3866.85520
1333	50	0	30.970	3	0	northwest	10600.54830
1334	18	1	31.920	0	0	northeast	2205.98080
1335	18	1	36.850	0	0	southeast	1629.83350
1336	21	1	25.800	0	0	southwest	2007.94500
1337	61	1	29.070	0	1	northwest	29141.36030

1338 rows × 7 columns

In [15]:

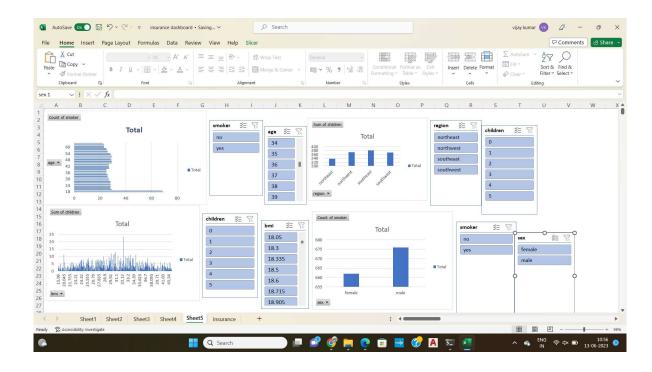
- 1 convert={"region":{"southwest":1,"southeast":2,"northeast":3,"northwest":4
- 2 df=df.replace(convert)
- 3 df

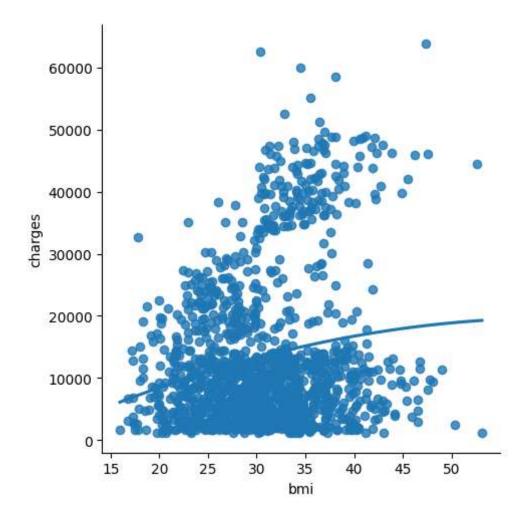
Out[15]:

	age	sex	bmi	children	smoker	region	charges
0	19	1	27.900	0	1	1	16884.92400
1	18	0	33.770	1	0	2	1725.55230
2	28	0	33.000	3	0	2	4449.46200
3	33	0	22.705	0	0	4	21984.47061
4	32	0	28.880	0	0	4	3866.85520
1333	50	0	30.970	3	0	4	10600.54830
1334	18	1	31.920	0	0	3	2205.98080
1335	18	1	36.850	0	0	2	1629.83350
1336	21	1	25.800	0	0	1	2007.94500
1337	61	1	29.070	0	1	4	29141.36030

Data Visualisation

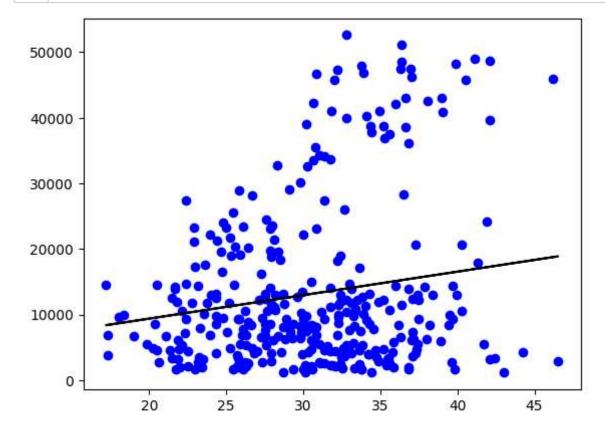
Dataset Dashboard in Excel sheet





0.050136213258239914

```
In [20]: 1  y_pred=lr.predict(x_test)
2  plt.scatter(x_test,y_test,color='b')
3  plt.plot(x_test,y_pred,color='k')
4  plt.show()
```



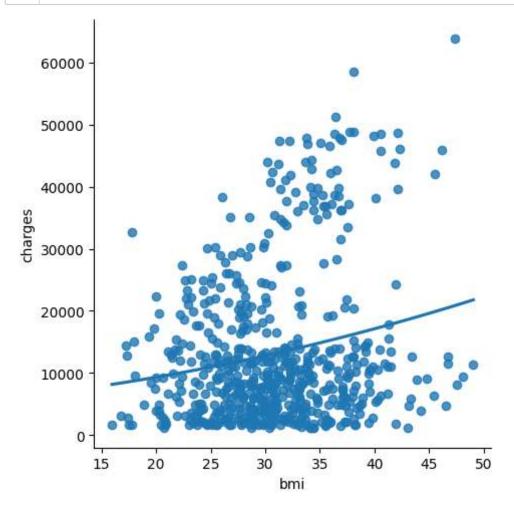
Working with the sub data of the given dataset

In [21]:

1 df600=df[:600] 2 df600

Out[21]:

	age	sex	bmi	children	smoker	region	charges
0	19	1	27.900	0	1	1	16884.92400
1	18	0	33.770	1	0	2	1725.55230
2	28	0	33.000	3	0	2	4449.46200
3	33	0	22.705	0	0	4	21984.47061
4	32	0	28.880	0	0	4	3866.85520
595	46	1	33.725	1	0	3	8823.98575
596	42	1	29.480	2	0	2	7640.30920
597	34	1	33.250	1	0	3	5594.84550
598	43	0	32.600	2	0	1	7441.50100
599	52	1	37.525	2	0	4	33471.97189

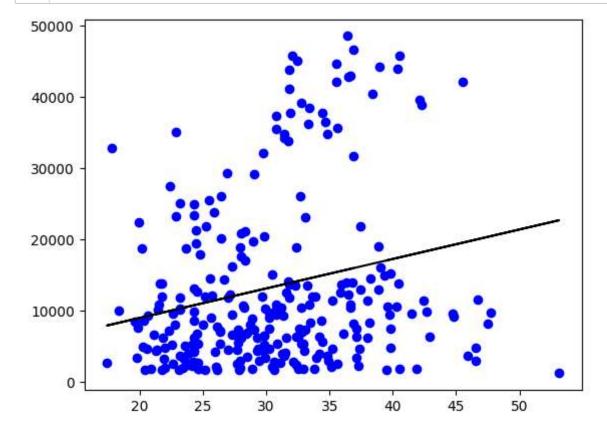


```
In [23]: 1 x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
2 lr=LinearRegression()
```

```
In [24]: 1 lr.fit(x_train,y_train)
2 print(lr.score(x_test,y_test))
```

0.030715187812894063

```
In [25]: 1  y_pred=lr.predict(x_test)
2  plt.scatter(x_test,y_test,color='b')
3  plt.plot(x_test,y_pred,color='k')
4  plt.show()
```



Logistic Regression

```
In [26]: 1 import numpy as np
2 import pandas as pd
3 from sklearn.linear_model import LogisticRegression
4 from sklearn.preprocessing import StandardScaler
```

```
df=pd.read_csv(r"C:\Users\Sushma sree\Downloads\insurance.csv")
In [27]:
                df
             2
Out[27]:
                  age
                                 bmi children smoker
                                                           region
                                                                       charges
                          sex
               0
                       female 27.900
                                             0
                                                                   16884.92400
                   19
                                                        southwest
                                                    yes
               1
                   18
                         male 33.770
                                             1
                                                         southeast
                                                                    1725.55230
                                                     no
               2
                   28
                         male 33.000
                                             3
                                                                    4449.46200
                                                    no
                                                         southeast
                         male 22.705
                                             0
               3
                   33
                                                         northwest
                                                                   21984.47061
                                                    no
                               28.880
               4
                   32
                                             0
                                                         northwest
                                                                    3866.85520
                         male
                                                     no
               ...
            1333
                   50
                         male 30.970
                                             3
                                                                   10600.54830
                                                         northwest
                                                    no
            1334
                       female 31.920
                                             0
                                                                    2205.98080
                   18
                                                     no
                                                         northeast
                               36.850
            1335
                   18
                       female
                                             0
                                                         southeast
                                                                    1629.83350
                                                    no
            1336
                                             0
                   21
                       female 25.800
                                                        southwest
                                                                    2007.94500
            1337
                       female 29.070
                                                    yes
                                                         northwest
                                                                   29141.36030
           1338 rows × 7 columns
In [28]:
                print('This Dataset has %d rows and %d columns'%(df.shape))
           This Dataset has 1338 rows and 7 columns
In [29]:
                df.head()
Out[29]:
               age
                       sex
                              bmi children smoker
                                                        region
                                                                    charges
                    female 27.900
                                          0
            0
                19
                                                     southwest
                                                                16884.92400
                                                yes
            1
                18
                      male 33.770
                                          1
                                                     southeast
                                                                 1725.55230
                                                 no
            2
                28
                      male 33.000
                                          3
                                                                 4449 46200
                                                     southeast
                33
                      male 22.705
                                                      northwest
                                                                21984.47061
                                                 no
                32
                      male 28.880
                                          0
                                                                 3866.85520
                                                     northwest
                                                 no
In [30]:
                df.tail()
Out[30]:
                                bmi children
                                               smoker
                                                          region
                                                                     charges
                  age
                          sex
                                            3
                                                                  10600.5483
            1333
                   50
                         male
                               30.97
                                                        northwest
            1334
                       female 31.92
                                            0
                                                                   2205.9808
                   18
                                                   no
                                                        northeast
            1335
                   18
                       female 36.85
                                            0
                                                        southeast
                                                                   1629.8335
                                                   no
            1336
                       female 25.80
                                            0
                                                        southwest
                                                                   2007.9450
                                                   no
                                            0
            1337
                   61
                       female 29.07
                                                        northwest 29141.3603
                                                  yes
```

```
In [31]:
               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
                                             object
           2
                bmi
                           1338 non-null
                                             float64
           3
                children 1338 non-null
                                             int64
           4
                smoker
                           1338 non-null
                                             object
           5
                           1338 non-null
                region
                                             object
           6
                charges
                           1338 non-null
                                             float64
          dtypes: float64(2), int64(2), object(3)
          memory usage: 73.3+ KB
In [32]:
               df.isnull().sum()
Out[32]: age
                       0
                        0
          sex
          bmi
                        0
          children
                       0
          smoker
                        0
                        0
          region
          charges
                        0
          dtype: int64
               convert={"smoker":{"yes":1, "no":0}}
In [33]:
            2
               df=df.replace(convert)
            3
               df
Out[33]:
                              bmi children smoker
                                                      region
                 age
                        sex
                                                                 charges
              0
                  19
                     female 27.900
                                         0
                                                 1 southwest 16884.92400
              1
                  18
                       male 33.770
                                         1
                                                 0
                                                    southeast
                                                              1725.55230
              2
                  28
                                         3
                       male 33.000
                                                    southeast
                                                              4449.46200
              3
                  33
                       male 22.705
                                         0
                                                 0
                                                    northwest 21984.47061
              4
                  32
                       male 28.880
                                         0
                                                    northwest
                                                              3866.85520
             •••
                                         ...
                                         3
           1333
                  50
                       male 30.970
                                                 0
                                                    northwest 10600.54830
           1334
                  18 female 31.920
                                         0
                                                    northeast
                                                              2205.98080
           1335
                  18 female 36.850
                                         0
                                                 0
                                                    southeast
                                                              1629.83350
                     female 25.800
           1336
                  21
                                         0
                                                    southwest
                                                              2007.94500
           1337
                     female 29.070
                                         0
                                                    northwest 29141.36030
```

```
convert={"sex":{"female":1,"male":0}}
In [34]:
                df=df.replace(convert)
             2
             3
                df
Out[34]:
                              bmi children smoker
                  age sex
                                                        region
                                                                   charges
               0
                   19
                         1 27.900
                                          0
                                                     southwest 16884.92400
               1
                   18
                         0 33.770
                                          1
                                                                 1725.55230
                                                      southeast
               2
                   28
                         0 33.000
                                          3
                                                      southeast
                                                                4449.46200
               3
                   33
                         0 22.705
                                          0
                                                      northwest
                                                               21984.47061
                                                     northwest
               4
                   32
                         0 28.880
                                          0
                                                                 3866.85520
              ...
                   ...
                        ...
                                         ...
                                                            ...
            1333
                   50
                         0 30.970
                                          3
                                                      northwest
                                                              10600.54830
            1334
                   18
                         1 31.920
                                          0
                                                  0
                                                      northeast
                                                                 2205.98080
            1335
                   18
                         1 36.850
                                          0
                                                      southeast
                                                                 1629.83350
            1336
                   21
                         1 25.800
                                          0
                                                                 2007.94500
                                                     southwest
            1337
                   61
                         1 29.070
                                          0
                                                      northwest 29141.36030
           1338 rows × 7 columns
In [35]:
                convert={"region":{"southeast":1, "southwest":2, "northeast":3, "northwest":4
                df=df.replace(convert)
             2
             3
                df
Out[35]:
                  age
                      sex
                              bmi children smoker region
                                                                 charges
               0
                   19
                         1 27.900
                                          0
                                                  1
                                                          2 16884.92400
                                          1
                                                  0
               1
                   18
                         0 33.770
                                                              1725.55230
                   28
                         0 33.000
                                          3
                                                  0
               2
                                                          1
                                                              4449.46200
               3
                   33
                         0 22.705
                                          0
                                                  0
                                                             21984.47061
               4
                   32
                         0 28.880
                                          0
                                                  0
                                                              3866.85520
                                         ...
                   ...
                         0 30.970
                                                  0
                                                          4 10600.54830
            1333
                   50
                                          3
                                                              2205.98080
            1334
                   18
                         1 31.920
                                          0
                                                  0
                                                          3
            1335
                   18
                         1 36.850
                                          0
                                                  0
                                                          1
                                                              1629.83350
            1336
                   21
                         1 25.800
                                          0
                                                  0
                                                              2007.94500
            1337
                   61
                         1 29.070
                                          0
                                                  1
                                                          4 29141.36030
           1338 rows × 7 columns
             1 features_matrix=df.iloc[:,0:4]
In [36]:
```

```
In [37]:
             target_vector=df.iloc[:,-3]
In [38]:
              print('The Feature Matrix has %d Rows and %d columns(s)'%(features matrix
             print('The Target Matrix has %d Rows and %d columns(s)'%(np.array(target)
         The Feature Matrix has 1338 Rows and 4 columns(s)
         The Target Matrix has 1338 Rows and 1 columns(s)
In [41]:
             features_matrix_standardized=StandardScaler().fit_transform(features_matri
In [42]:
             lgr=LogisticRegression(max iter=10000)
In [43]:
              Logistic Regression Model=lgr.fit(features matrix standardized, target vec
In [44]:
             observation=[[1,0,0.99539,-0.0588]]
In [45]:
              predictions=Logistic Regression Model.predict(observation)
             print('The model predicted the observation to belong to class %s'%(predicted)
         The model predicted the observation to belong to class [0]
In [46]:
              print('The algoritham was trained to predict one of the two classes:%s'%()
         The algoritham was trained to predict one of the two classes:[0 1]
In [47]:
             print("The model says the probability of the observation we passed belong:
             print("The Model says the probability of the observation we passed belong:
         The model says the probability of the observation we passed belonging to clas
         s[0] Is 0.8057075871331396
         The Model says the probability of the observation we passed belonging to clas
         s['1'] Is 0.8057075871331396
In [48]:
             x=np.array(df['age']).reshape(-1,1)
             y=np.array(df['smoker']).reshape(-1,1)
In [49]:
             x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.05)
             lr=LogisticRegression()
```

0.8208955223880597

C:\Users\Sushma sree\AppData\Local\Programs\Python\Python310\lib\site-package s\sklearn\utils\validation.py:1143: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_s amples,), for example using ravel().

```
y = column_or_1d(y, warn=True)
```

Decision Tree

Out[52]:

	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	

```
In [53]:
                 df.head()
Out[53]:
                               bmi children smoker
                age
                                                                      charges
                        sex
                                                          region
            0
                 19
                     female
                            27.900
                                           0
                                                                  16884.92400
                                                  yes
                                                       southwest
            1
                 18
                            33.770
                                           1
                                                                   1725.55230
                      male
                                                       southeast
                                                   no
            2
                 28
                      male
                           33.000
                                           3
                                                        southeast
                                                                   4449.46200
                                                   no
            3
                 33
                            22.705
                                           0
                                                                  21984.47061
                      male
                                                        northwest
                                                   no
                 32
                      male 28.880
                                           0
                                                                   3866.85520
                                                        northwest
In [54]:
                 df.tail()
Out[54]:
                                 bmi children
                                                smoker
                                                            region
                                                                       charges
                   age
                           sex
             1333
                    50
                          male
                                30.97
                                             3
                                                          northwest
                                                                    10600.5483
                                                     no
            1334
                    18
                        female 31.92
                                             0
                                                          northeast
                                                                     2205.9808
                                                     no
            1335
                    18
                        female
                                36.85
                                             0
                                                          southeast
                                                                     1629.8335
                                                     no
            1336
                    21
                        female 25.80
                                             0
                                                                     2007.9450
                                                         southwest
                                                     no
                                             0
            1337
                    61
                        female 29.07
                                                         northwest 29141.3603
                                                    yes
In [55]:
                 convert={'smoker':{'yes':1,'no':0}}
              2
                 df=df.replace(convert)
Out[55]:
                                  bmi children smoker
                                                                         charges
                   age
                           sex
                                                             region
                0
                               27.900
                                              0
                    19
                        female
                                                       1
                                                          southwest
                                                                     16884.92400
                1
                    18
                          male 33.770
                                               1
                                                           southeast
                                                                      1725.55230
                2
                    28
                                33.000
                                               3
                                                                      4449.46200
                          male
                                                           southeast
                3
                    33
                          male
                                22.705
                                              0
                                                           northwest
                                                                     21984.47061
                                              0
                4
                    32
                          male 28.880
                                                       0
                                                                      3866.85520
                                                           northwest
                            ...
               •••
                    • • •
                                    ...
                                              •••
                                30.970
                                                                     10600.54830
            1333
                    50
                          male
                                              3
                                                       0
                                                           northwest
            1334
                    18
                        female 31.920
                                              0
                                                           northeast
                                                                      2205.98080
            1335
                    18
                        female 36.850
                                              0
                                                           southeast
                                                                      1629.83350
                        female 25.800
            1336
                                              0
                                                          southwest
                                                                      2007.94500
                        female 29.070
                                                           northwest 29141.36030
            1337
                                              0
```

```
convert={"sex":{"female":1,"male":0}}
In [56]:
               df=df.replace(convert)
            2
            3
               df
Out[56]:
                             bmi children smoker
                 age sex
                                                     region
                                                                charges
              0
                        1 27.900
                                        0
                                                1 southwest 16884.92400
                  19
              1
                        0 33.770
                                        1
                  18
                                                   southeast
                                                             1725.55230
              2
                  28
                        0 33.000
                                        3
                                                0 southeast
                                                             4449.46200
              3
                  33
                        0 22.705
                                                   northwest 21984.47061
              4
                  32
                        0 28.880
                                        0
                                                  northwest
                                                             3866.85520
              ...
                  ...
                       ...
                                       ...
           1333
                  50
                        0 30.970
                                        3
                                                0 northwest 10600.54830
           1334
                  18
                        1 31.920
                                        0
                                                   northeast
                                                             2205.98080
           1335
                  18
                        1 36.850
                                        0
                                                0 southeast
                                                             1629.83350
           1336
                  21
                        1 25.800
                                        0
                                                             2007.94500
                                                0 southwest
           1337
                  61
                        1 29.070
                                        0
                                                   northwest 29141.36030
          1338 rows × 7 columns
            1 x=["bmi","children"]
2 y=["Yes","No"]
In [57]:
            3 all_inputs=df[x]
            4 all_classes=df["sex"]
In [58]:
                (x_train,x_test,y_train,y_test)=train_test_split(all_inputs,all_classes,te
In [59]:
               clf=DecisionTreeClassifier(random_state=0)
In [60]:
               clf.fit(x_train,y_train)
Out[60]:
          DecisionTreeClassifier(random_state=0)
          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 [61]:
               score=clf.score(x_test,y_test)
               print(score)
```

0.4146341463414634

Random Forest

Name: count, dtype: int64

```
In [62]:
                import pandas as pd
               import numpy as np
               import matplotlib.pyplot as plt ,seaborn as sns
               df=pd.read csv(r"C:\Users\Sushma sree\Downloads\insurance.csv")
In [63]:
            1
Out[63]:
                                bmi children smoker
                                                        region
                                                                   charges
                 age
                         sex
                      female 27.900
                                                 yes southwest 16884.92400
              0
                  19
                                           0
              1
                  18
                       male 33.770
                                           1
                                                  no
                                                      southeast
                                                                 1725.55230
                  28
                       male 33.000
                                           3
              2
                                                      southeast
                                                                 4449.46200
                                                  no
              3
                  33
                                           0
                       male 22.705
                                                      northwest
                                                               21984.47061
              4
                  32
                       male 28.880
                                                      northwest
                                                                 3866.85520
                                                  no
              ...
           1333
                  50
                       male 30.970
                                           3
                                                      northwest 10600.54830
                                                  no
                  18 female 31.920
                                           0
                                                      northeast
                                                                 2205.98080
           1334
                                                  no
           1335
                      female 36.850
                                           0
                                                      southeast
                                                                 1629.83350
                  18
                                                  no
           1336
                      female 25.800
                                                      southwest
                                                                 2007.94500
                                                  no
           1337
                  61 female 29.070
                                           0
                                                      northwest 29141.36030
                                                 yes
           1338 rows × 7 columns
In [64]:
            1 df['region'].value_counts()
Out[64]: region
          southeast
                         364
          southwest
                         325
          northwest
                         325
          northeast
                          324
```

```
In [65]:
               df['bmi'].value_counts()
Out[65]: bmi
          32.300
                     13
          28.310
                      9
          30.495
                      8
          30.875
                      8
          31.350
                      8
                     . .
          46.200
                      1
          23.800
                      1
          44.770
                      1
          32.120
                      1
          30.970
                      1
          Name: count, Length: 548, dtype: int64
In [66]:
              m={"sex":{"female":1,"male":0}}
            2
               df=df.replace(m)
              print(df)
                      sex
                                    children smoker
                 age
                               bmi
                                                          region
                                                                       charges
          0
                  19
                        1
                           27.900
                                            0
                                                       southwest
                                                                   16884.92400
                                                  yes
                           33.770
          1
                  18
                                            1
                                                       southeast
                                                                    1725.55230
                        0
                                                   no
          2
                  28
                        0
                           33.000
                                            3
                                                   no
                                                       southeast
                                                                    4449.46200
                           22.705
          3
                  33
                        0
                                            0
                                                       northwest
                                                                   21984.47061
                                                   no
          4
                  32
                           28.880
                                            0
                                                                    3866.85520
                                                   no
                                                       northwest
                               . . .
                                                  . . .
                  50
                           30.970
                                                                   10600.54830
          1333
                        0
                                            3
                                                  no
                                                       northwest
          1334
                  18
                        1
                           31.920
                                            0
                                                       northeast
                                                                    2205.98080
                                                   no
          1335
                  18
                           36.850
                                            0
                                                       southeast
                                                                    1629.83350
                        1
                                                   no
          1336
                  21
                        1
                           25.800
                                            0
                                                       southwest
                                                                    2007.94500
                                                   no
          1337
                  61
                        1
                           29.070
                                            0
                                                       northwest
                                                                   29141.36030
                                                  yes
          [1338 rows x 7 columns]
In [67]:
               n={"smoker":{"yes":1,"no":0}}
              df=df.replace(n)
            2
            3
               print(df)
                               bmi
                                    children
                                               smoker
                                                           region
                      sex
                 age
                                                                         charges
                  19
                           27.900
                                                        southwest
                                                                   16884.92400
          0
                        1
                                            0
          1
                  18
                        0
                           33.770
                                            1
                                                        southeast
                                                                     1725.55230
          2
                                            3
                  28
                           33.000
                                                        southeast
                                                                     4449.46200
          3
                  33
                        0
                           22.705
                                            0
                                                     0
                                                        northwest
                                                                    21984.47061
          4
                                            0
                  32
                           28.880
                                                     0
                                                        northwest
                                                                     3866.85520
                        0
                                          . . .
          . . .
          1333
                  50
                        0
                           30.970
                                            3
                                                     0
                                                        northwest 10600.54830
          1334
                        1
                           31.920
                                            0
                                                        northeast
                                                                     2205.98080
                  18
          1335
                  18
                        1
                           36.850
                                            0
                                                        southeast
                                                                     1629.83350
          1336
                  21
                           25.800
                                            0
                                                        southwest
                                                                     2007.94500
                        1
          1337
                  61
                        1
                           29.070
                                            0
                                                        northwest
                                                                    29141.36030
          [1338 rows x 7 columns]
```

Out[68]: RandomForestClassifier()

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 [69]:
           1 rf=RandomForestClassifier()
           2 params={'max depth':[2,3,5,20],
             'min samples leaf':[5,10,20,50,100,200],
           4 'n_estimators':[10,25,30,50,100,200]}
In [70]:
           1 x=df[['bmi']]
           2 y=df['smoker']
In [71]:
           1 from sklearn.model selection import GridSearchCV
           2 | grid_search=GridSearchCV(estimator=rf,param_grid=params,cv=2,scoring="acc
           3 grid_search.fit(x_train,y_train)
Out[71]: GridSearchCV(cv=2, estimator=RandomForestClassifier(),
                      param_grid={'max_depth': [2, 3, 5, 20],
                                   'min samples leaf': [5, 10, 20, 50, 100, 200],
                                   'n_estimators': [10, 25, 30, 50, 100, 200]},
                      scoring='accuracy')
```

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 [72]: 1 grid_search.best_score_
Out[72]: 0.5165746447526108

In [73]: 1 rf_best=grid_search.best_estimator_
2 print(rf_best)

RandomForestClassifier(max_depth=3, min_samples_leaf=200, n_estimators=200)

In [74]: 1 rf_best.feature_importances_
Out[74]: array([0.67662889, 0.32337111])

In [75]: 1 rf=RandomForestClassifier(random_state=0)
```

```
In [76]: 1 rf.fit(x_train,y_train)
```

Out[76]: RandomForestClassifier(random_state=0)

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 [77]: 1 score=rf.score(x_test,y_test)
2 print(score)
```

0.4878048780487805

Conclusion:

After performing all models the to the given dataset. Among all models "Logistic Regression" has 80% accuracy

```
In [ ]: 1
```