{

"cells":[

{

"cell\_type": "code","execution\_count": 100,"id":"c476cd32",

"metadata":{},

"outputs": [],"source":[

"#importallnessesarylibraries**\n**","importpandasaspd**\n**",

"importnumpyasnp**\n**",

"importmatplotlib.pyplotasplt**\n**","importseabornassns"

]

},

{

"cell\_type": "code","execution\_count": 101,"id":"d9574953",

"metadata": {},"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>sex</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>smoker</th>**\n**",

" <th>region</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**"," <tr>**\n**",

" <th>0</th>**\n**",

" <td>19</td>**\n**",

" <td>female</td>**\n**",

" <td>27.900</td>**\n**",

" <td>0</td>**\n**",

" <td>yes</td>**\n**",

" <td>southwest</td>**\n**",

" <td>16884.92400</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1</th>**\n**",

" <td>18</td>**\n**",

" <td>male</td>**\n**",

" <td>33.770</td>**\n**",

" <td>1</td>**\n**",

" <td>no</td>**\n**",

" <td>southeast</td>**\n**"," <td>1725.55230</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>2</th>**\n**",

" <td>28</td>**\n**",

" <td>male</td>**\n**",

" <td>33.000</td>**\n**",

" <td>3</td>**\n**",

" <td>no</td>**\n**",

" <td>southeast</td>**\n**"," <td>4449.46200</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>3</th>**\n**",

" <td>33</td>**\n**",

" <td>male</td>**\n**",

" <td>22.705</td>**\n**",

" <td>0</td>**\n**",

" <td>no</td>**\n**",

" <td>northwest</td>**\n**",

" <td>21984.47061</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>4</th>**\n**",

" <td>32</td>**\n**",

" <td>male</td>**\n**",

" <td>28.880</td>**\n**",

" <td>0</td>**\n**",

" <td>no</td>**\n**",

" <td>northwest</td>**\n**"," <td>3866.85520</td>**\n**",

" </tr>**\n**","</tbody>**\n**","</table>**\n**","</div>"

],

"text/plain":[

" age sex bmichildrensmoker region charges**\n**","0 19female27.900 0 yessouthwest16884.92400**\n**",

"1 18 male33.770 1 nosoutheast 1725.55230**\n**",

"2 28 male33.000 3 nosoutheast 4449.46200**\n**",

"3 33 male22.705 0 nonorthwest21984.47061**\n**",

"4 32 male28.880 0 nonorthwest 3866.85520"

]

},

"execution\_count": 101,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["#loaddata**\n**",

"data=pd.read\_csv(r**\"**C:**\\**Users**\\sagamvishnuvardhanReddy\\**Downloads**\\**interen\_bridgre\_project\_daataset**\\**insurance.csv**\"**)**\n**","data.head()"

]

},

{

"cell\_type": "code","execution\_count": 102,"id":"ddf49a46",

"metadata": {},"outputs":[

{

"name": "stdout","output\_type": "stream","text": [

"<class'pandas.core.frame.DataFrame'>**\n**","RangeIndex: 1338 entries, 0 to 1337**\n**","Datacolumns(total7columns):**\n**",

"# Column Non-NullCountDtype**\n**"," **\n**",

"0 age 1338 non-null int64 **\n**","1 sex 1338 non-null object **\n**","2 bmi 1338 non-null float64**\n**"," 3 children 1338 non-null int64 **\n**","4 smoker 1338 non-null object **\n**","5 region 1338 non-null object **\n**"," 6charges1338 non-nullfloat64**\n**","dtypes:float64(2),int64(2),object(3)**\n**","memoryusage:73.3+KB**\n**"

]

}

],

"source":[

"data.info()"

]

},

{

"cell\_type": "code","execution\_count": 103,"id":"9db9ab59",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"(1338,7)"

]

},

"execution\_count": 103,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["data.shape"

]

},

{

"cell\_type": "code","execution\_count": 104,"id":"34b3abcc",

"metadata": {},"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**"," <tr>**\n**",

" <th>count</th>**\n**",

" <td>1338.000000</td>**\n**"," <td>1338.000000</td>**\n**"," <td>1338.000000</td>**\n**","<td>1338.000000</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>mean</th>**\n**",

" <td>39.207025</td>**\n**",

" <td>30.663397</td>**\n**",

" <td>1.094918</td>**\n**",

" <td>13270.422265</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>std</th>**\n**",

" <td>14.049960</td>**\n**",

" <td>6.098187</td>**\n**",

" <td>1.205493</td>**\n**",

" <td>12110.011237</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>min</th>**\n**",

" <td>18.000000</td>**\n**",

" <td>15.960000</td>**\n**",

" <td>0.000000</td>**\n**",

" <td>1121.873900</td>**\n**",

|  |  |
| --- | --- |
| " | </tr>**\n**", |
| " | <tr>**\n**", |
| " | <th>25%</th>**\n**", |
| " | <td>27.000000</td>**\n**", |
| " | <td>26.296250</td>**\n**", |
| " | <td>0.000000</td>**\n**", |
| " | <td>4740.287150</td>**\n**", |
| " | </tr>**\n**", |
| " | <tr>**\n**", |
| " | <th>50%</th>**\n**", |
| " | <td>39.000000</td>**\n**", |
| " | <td>30.400000</td>**\n**", |
| " | <td>1.000000</td>**\n**", |
| " | <td>9382.033000</td>**\n**", |
| " | </tr>**\n**", |
| " | <tr>**\n**", |
| " | <th>75%</th>**\n**", |
| " | <td>51.000000</td>**\n**", |
| " | <td>34.693750</td>**\n**", |
| " | <td>2.000000</td>**\n**", |
| " | <td>16639.912515</td>**\n**", |
| " | </tr>**\n**", |
| " | <tr>**\n**", |
| " | <th>max</th>**\n**", |
| " | <td>64.000000</td>**\n**", |
| " | <td>53.130000</td>**\n**", |
| " | <td>5.000000</td>**\n**", |
| " | <td>63770.428010</td>**\n**", |
| " | </tr>**\n**", |
| " | </tbody>**\n**", |

]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| "</table>**\n**","</div>"  ],  "text/plain": [" | age | bmi | children | charges**\n**", |
| "count1338.000000 | | 1338.000000 | 1338.000000 | 1338.000000**\n**", |
| "mean 39.207025 | | 30.663397 | 1.094918 | 13270.422265**\n**", |
| "std 14.049960 | | 6.098187 | 1.205493 | 12110.011237**\n**", |
| "min 18.000000 | | 15.960000 | 0.000000 | 1121.873900**\n**", |
| "25% 27.000000 | | 26.296250 | 0.000000 | 4740.287150**\n**", |
| "50% 39.000000 | | 30.400000 | 1.000000 | 9382.033000**\n**", |
| "75% 51.000000 | | 34.693750 | 2.000000 | 16639.912515**\n**", |
| "max 64.000000 | | 53.130000 | 5.000000 | 63770.428010" |

},

"execution\_count": 104,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["data.describe()"

]

},

{

"cell\_type": "code","execution\_count": 105,"id":"4aa40681",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"age 0**\n**",

"sex 0**\n**",

"bmi 0**\n**",

"children 0**\n**",

"smoker 0**\n**",

"region 0**\n**",

"charges 0**\n**","dtype:int64"

]

},

"execution\_count": 105,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["data.isnull().sum()"

]

},

{

"cell\_type": "code","execution\_count": 106,"id":"703f3ba3",

"metadata": {},"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>sex</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>smoker</th>**\n**",

" <th>region</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**"," <tr>**\n**",

" <th>581</th>**\n**",

" <td>19</td>**\n**",

" <td>male</td>**\n**",

" <td>30.59</td>**\n**",

" <td>0</td>**\n**",

" <td>no</td>**\n**",

" <td>northwest</td>**\n**",

" <td>1639.5631</td>**\n**",

" </tr>**\n**","</tbody>**\n**","</table>**\n**","</div>"

],

"text/plain":[

" age sex bmichildrensmoker region charges**\n**","581 19male30.59 0 nonorthwest1639.5631"

]

},

"execution\_count": 106,"metadata":{},

"output\_type":"execute\_result"

}

],

"source":[

"#remove duplicates**\n**","data[data.duplicated()]"

]

},

{

"cell\_type": "code","execution\_count": 107,"id":"f35097ba",

"metadata":{},

"outputs": [],"source":[

"data=data.drop\_duplicates()"

]

},

{

"cell\_type": "code","execution\_count": 108,"id":"95bc5094",

"metadata": {},"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>sex</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>smoker</th>**\n**",

" <th>region</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**","</tbody>**\n**","</table>**\n**","</div>"

],

"text/plain": ["EmptyDataFrame**\n**",

"Columns:[age,sex,bmi,children,smoker,region,charges]**\n**","Index:[]"

]

},

"execution\_count": 108,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["data[data.duplicated()]"

]

},

{

"cell\_type": "code","execution\_count": 109,"id":"23eb8b6d",

"metadata":{},

"outputs": [],"source":[

"#encodingcatagoricalvariables**\n**",

"fromsklearn.preprocessingimportOrdinalEncoder**\n**","le=OrdinalEncoder()"

]

},

{

"cell\_type": "code","execution\_count": 110,"id":"8ed938e3",

"metadata":{},

"outputs": [],"source":[

"**\n**","data[['sex']]=le.fit\_transform(data[['sex']])**\n**",

"data[['smoker']]=le.fit\_transform(data[['smoker']])**\n**",

"data[['region']]=le.fit\_transform(data[['region']])**\n**"

]

},

{

"cell\_type": "code","execution\_count": 111,"id":"224825f1",

"metadata": {"scrolled":true

},

"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>sex</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>smoker</th>**\n**",

" <th>region</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**"," <tr>**\n**",

" <th>0</th>**\n**",

" <td>19</td>**\n**",

" <td>0.0</td>**\n**",

" <td>27.900</td>**\n**",

" <td>0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>3.0</td>**\n**",

" <td>16884.92400</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1</th>**\n**",

" <td>18</td>**\n**",

" <td>1.0</td>**\n**",

" <td>33.770</td>**\n**",

" <td>1</td>**\n**",

" <td>0.0</td>**\n**",

" <td>2.0</td>**\n**",

" <td>1725.55230</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>2</th>**\n**",

" <td>28</td>**\n**",

" <td>1.0</td>**\n**",

" <td>33.000</td>**\n**",

" <td>3</td>**\n**",

" <td>0.0</td>**\n**",

" <td>2.0</td>**\n**",

" <td>4449.46200</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>3</th>**\n**",

" <td>33</td>**\n**",

" <td>1.0</td>**\n**",

" <td>22.705</td>**\n**",

" <td>0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>21984.47061</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>4</th>**\n**",

" <td>32</td>**\n**",

" <td>1.0</td>**\n**",

" <td>28.880</td>**\n**",

" <td>0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>3866.85520</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>...</th>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1333</th>**\n**",

" <td>50</td>**\n**",

" <td>1.0</td>**\n**",

" <td>30.970</td>**\n**",

" <td>3</td>**\n**",

" <td>0.0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>10600.54830</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1334</th>**\n**",

" <td>18</td>**\n**",

" <td>0.0</td>**\n**",

" <td>31.920</td>**\n**",

" <td>0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>2205.98080</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1335</th>**\n**",

" <td>18</td>**\n**",

" <td>0.0</td>**\n**",

" <td>36.850</td>**\n**",

" <td>0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>2.0</td>**\n**",

" <td>1629.83350</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1336</th>**\n**",

" <td>21</td>**\n**",

" <td>0.0</td>**\n**",

" <td>25.800</td>**\n**",

" <td>0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>3.0</td>**\n**",

" <td>2007.94500</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1337</th>**\n**",

" <td>61</td>**\n**",

" <td>0.0</td>**\n**",

" <td>29.070</td>**\n**",

" <td>0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>29141.36030</td>**\n**",

" </tr>**\n**","</tbody>**\n**","</table>**\n**",

"<p>1337rowsÃ7columns</p>**\n**","</div>"

],

"text/plain":[

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| " age | sex | bmi | children | smoker | region | charges**\n**", |
| "0 19 | 0.0 | 27.900 | 0 | 1.0 | 3.0 | 16884.92400**\n**", |
| "1 18 | 1.0 | 33.770 | 1 | 0.0 | 2.0 | 1725.55230**\n**", |
| "2 28 | 1.0 | 33.000 | 3 | 0.0 | 2.0 | 4449.46200**\n**", |
| "3 33 | 1.0 | 22.705 | 0 | 0.0 | 1.0 | 21984.47061**\n**", |
| "4 32 | 1.0 | 28.880 | 0 | 0.0 | 1.0 | 3866.85520**\n**", |
| "... ... | ... | ... | ... | ... | ... | ...**\n**", |
| "1333 50 | 1.0 | 30.970 | 3 | 0.0 | 1.0 | 10600.54830**\n**", |
| "1334 18 | 0.0 | 31.920 | 0 | 0.0 | 0.0 | 2205.98080**\n**", |
| "1335 18 | 0.0 | 36.850 | 0 | 0.0 | 2.0 | 1629.83350**\n**", |
| "1336 21 | 0.0 | 25.800 | 0 | 0.0 | 3.0 | 2007.94500**\n**", |
| "1337 61 | 0.0 | 29.070 | 0 | 1.0 | 1.0 | 29141.36030**\n**", |
| "**\n**", |  |  |  |  |  |  |

"[1337rowsx7columns]"

]

},

"execution\_count": 111,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["data"

]

},

{

"cell\_type": "code","execution\_count": 112,"id":"e50c7c94",

"metadata": {},"outputs":[

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAjkAAAGdCAYAAADwjmIIAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["#scaling**\n**",

"#%matplotlib inline**\n**","sns.barplot(data=data)**\n**",

"plt.xticks(range(len(data.columns)),data.columns)**\n**","plt.show()"

]

},

{

"cell\_type": "code","execution\_count": 113,"id":"e304e016",

"metadata":{},

"outputs": [],"source":[

"#scaling**\n**",

"from sklearn.preprocessing import MinMaxScaler**\n**","scaler=MinMaxScaler()"

]

},

{

"cell\_type": "code","execution\_count": 114,"id":"be9aed67",

"metadata":{},

"outputs": [],"source":[

"data\_scaled=scaler.fit\_transform(data)**\n**"

]

},

{

"cell\_type": "code","execution\_count": 115,"id":"7b323bdc",

"metadata":{},

"outputs": [],"source":[

"data=pd.DataFrame(data\_scaled,columns=data.columns)"

]

},

{

"cell\_type": "code","execution\_count": 116,"id":"31b06761",

"metadata": {},"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>sex</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>smoker</th>**\n**",

" <th>region</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**"," <tr>**\n**",

" <th>0</th>**\n**",

" <td>0.021739</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.321227</td>**\n**",

" <td>0.0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>1.000000</td>**\n**",

" <td>0.251611</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1</th>**\n**",

" <td>0.000000</td>**\n**",

" <td>1.0</td>**\n**",

" <td>0.479150</td>**\n**",

" <td>0.2</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.666667</td>**\n**",

" <td>0.009636</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>2</th>**\n**",

" <td>0.217391</td>**\n**",

" <td>1.0</td>**\n**",

" <td>0.458434</td>**\n**",

" <td>0.6</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.666667</td>**\n**",

" <td>0.053115</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>3</th>**\n**",

" <td>0.326087</td>**\n**",

" <td>1.0</td>**\n**",

" <td>0.181464</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.333333</td>**\n**",

" <td>0.333010</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>4</th>**\n**",

" <td>0.304348</td>**\n**",

" <td>1.0</td>**\n**",

" <td>0.347592</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.333333</td>**\n**",

" <td>0.043816</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>...</th>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" <td>...</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1332</th>**\n**",

" <td>0.695652</td>**\n**",

" <td>1.0</td>**\n**",

" <td>0.403820</td>**\n**",

" <td>0.6</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.333333</td>**\n**",

" <td>0.151299</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1333</th>**\n**",

" <td>0.000000</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.429379</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.000000</td>**\n**",

" <td>0.017305</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1334</th>**\n**",

" <td>0.000000</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.562012</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.666667</td>**\n**",

" <td>0.008108</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1335</th>**\n**",

" <td>0.065217</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.264730</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.0</td>**\n**",

" <td>1.000000</td>**\n**",

" <td>0.014144</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>1336</th>**\n**",

" <td>0.934783</td>**\n**",

" <td>0.0</td>**\n**",

" <td>0.352704</td>**\n**",

" <td>0.0</td>**\n**",

" <td>1.0</td>**\n**",

" <td>0.333333</td>**\n**",

" <td>0.447249</td>**\n**",

" </tr>**\n**","</tbody>**\n**","</table>**\n**",

"<p>1337rowsÃ7columns</p>**\n**","</div>"

],

"text/plain":[

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| " | age | sex | bmi | children | smoker | region | charges**\n**", |
| "0 | 0.021739 | 0.0 | 0.321227 | 0.0 | 1.0 | 1.000000 | 0.251611**\n**", |
| "1 | 0.000000 | 1.0 | 0.479150 | 0.2 | 0.0 | 0.666667 | 0.009636**\n**", |
| "2 | 0.217391 | 1.0 | 0.458434 | 0.6 | 0.0 | 0.666667 | 0.053115**\n**", |
| "3 | 0.326087 | 1.0 | 0.181464 | 0.0 | 0.0 | 0.333333 | 0.333010**\n**", |
| "4 | 0.304348 | 1.0 | 0.347592 | 0.0 | 0.0 | 0.333333 | 0.043816**\n**", |
| "... | ... | ... | ... | ... | ... | ... | ...**\n**", |
| "1332 | 0.695652 | 1.0 | 0.403820 | 0.6 | 0.0 | 0.333333 | 0.151299**\n**", |
| "1333 | 0.000000 | 0.0 | 0.429379 | 0.0 | 0.0 | 0.000000 | 0.017305**\n**", |
| "1334 | 0.000000 | 0.0 | 0.562012 | 0.0 | 0.0 | 0.666667 | 0.008108**\n**", |
| "1335 | 0.065217 | 0.0 | 0.264730 | 0.0 | 0.0 | 1.000000 | 0.014144**\n**", |
| "1336 | 0.934783 | 0.0 | 0.352704 | 0.0 | 1.0 | 0.333333 | 0.447249**\n**", |
| "**\n**", |  |  |  |  |  |  |  |

"[1337rowsx7columns]"

]

},

"execution\_count": 116,"metadata":{},

"output\_type":"execute\_result"

}

],

"source":[

"**\n**","data"

]

},

{

"cell\_type": "code","execution\_count": 117,"id":"c2d3c04a",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 117,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source":[

"#oulier detection**\n**","sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 118,"id":"f668eefa",

"metadata":{},

"outputs": [],"source":[

"Q1=data['bmi'].quantile(0.25)**\n**",

"Q3 = data['bmi'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['bmi']>=lower\_bound)&(data['bmi']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 119,"id":"da46b1fc",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 119,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 120,"id":"57d87711",

"metadata":{},

"outputs": [],"source":[

"Q1 = data['charges'].quantile(0.25)**\n**","Q3 = data['charges'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['charges']>=lower\_bound)&(data['charges']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 121,"id":"e61d3341",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 121,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 122,"id":"a3420b81",

"metadata":{},

"outputs": [],"source":[

"Q1 = data['charges'].quantile(0.25)**\n**","Q3 = data['charges'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['charges']>=lower\_bound)&(data['charges']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 123,"id":"e7bf9a6e",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 123,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count":124,

"id":"7489c7a9",

"metadata":{},

"outputs": [],"source":[

"Q1=data['bmi'].quantile(0.25)**\n**",

"Q3 = data['bmi'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['bmi']>=lower\_bound)&(data['bmi']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 125,"id":"21473f0d",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 125,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 126,"id":"e28d1b60",

"metadata":{},

"outputs": [],"source":[

"Q1 = data['charges'].quantile(0.25)**\n**","Q3 = data['charges'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['charges']>=lower\_bound)&(data['charges']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 127,"id":"1cf57a43",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 127,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 128,"id":"9c56d2ad",

"metadata":{},

"outputs": [],"source":[

"Q1 = data['smoker'].quantile(0.25)**\n**","Q3 = data['smoker'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['smoker']>=lower\_bound)&(data['smoker']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 129,"id":"7c9ca6fe",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 129,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 130,"id":"5d9e7947",

"metadata":{},

"outputs": [],"source":[

"Q1 = data['charges'].quantile(0.25)**\n**","Q3 = data['charges'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['charges']>=lower\_bound)&(data['charges']<=upper\_bound)]"

]

},

{

"cell\_type": "code","execution\_count": 131,"id":"ed4f5139",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 131,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.boxplot(data=data)"

]

},

{

"cell\_type": "code","execution\_count": 132,"id":"65c5a5da",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<AxesSubplot:>"

]

},

"execution\_count": 132,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source":[

"Q1 = data['charges'].quantile(0.25)**\n**","Q3 = data['charges'].quantile(0.75)**\n**","IQR=Q3-Q1**\n**",

"**\n**",

"**\n**",

"lower\_bound=Q1-1.5\*IQR**\n**","upper\_bound=Q3+1.5\*IQR**\n**","**\n**",

"**\n**",

"data=data[(data['charges']>=lower\_bound)&(data['charges']<=upper\_bound)]**\n**","sns.boxplot(data=data)"

]

},

{

"cell\_type":"code",

"execution\_count":133,

"id":"41779711",

"metadata": {},"outputs":[

{

"data":{

"text/html": ["<div>**\n**",

"<stylescoped>**\n**",

" .dataframetbodytrth:only-of-type{**\n**"," vertical-align:middle;**\n**",

" }**\n**",

"**\n**",

" .dataframe tbody tr th {**\n**"," vertical-align: top;**\n**","}**\n**",

"**\n**",

" .dataframe thead th {**\n**"," text-align:right;**\n**"," }**\n**",

"</style>**\n**",

"<table border=**\"**1**\"** class=**\"**dataframe**\"**>**\n**","<thead>**\n**",

" <tr style=**\"**text-align: right;**\"**>**\n**"," <th></th>**\n**",

" <th>age</th>**\n**",

" <th>sex</th>**\n**",

" <th>bmi</th>**\n**",

" <th>children</th>**\n**",

" <th>smoker</th>**\n**",

" <th>region</th>**\n**",

" <th>charges</th>**\n**",

" </tr>**\n**","</thead>**\n**"," <tbody>**\n**"," <tr>**\n**",

" <th>age</th>**\n**",

" <td>1.000000</td>**\n**",

" <td>-0.016574</td>**\n**",

" <td>0.122776</td>**\n**",

" <td>0.030059</td>**\n**",

" <td>NaN</td>**\n**",

" <td>0.002888</td>**\n**",

" <td>0.798984</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>sex</th>**\n**",

" <td>-0.016574</td>**\n**",

" <td>1.000000</td>**\n**",

" <td>0.010545</td>**\n**",

" <td>0.008622</td>**\n**",

" <td>NaN</td>**\n**",

" <td>-0.017247</td>**\n**",

" <td>-0.061555</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>bmi</th>**\n**",

" <td>0.122776</td>**\n**",

" <td>0.010545</td>**\n**",

" <td>1.000000</td>**\n**",

" <td>0.000046</td>**\n**",

" <td>NaN</td>**\n**",

" <td>0.141389</td>**\n**",

" <td>0.134148</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>children</th>**\n**",

" <td>0.030059</td>**\n**",

" <td>0.008622</td>**\n**",

" <td>0.000046</td>**\n**",

" <td>1.000000</td>**\n**",

" <td>NaN</td>**\n**",

" <td>0.023565</td>**\n**",

" <td>0.157323</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>smoker</th>**\n**",

" <td>NaN</td>**\n**",

" <td>NaN</td>**\n**",

" <td>NaN</td>**\n**",

" <td>NaN</td>**\n**",

" <td>NaN</td>**\n**",

" <td>NaN</td>**\n**",

" <td>NaN</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>region</th>**\n**",

" <td>0.002888</td>**\n**",

" <td>-0.017247</td>**\n**",

" <td>0.141389</td>**\n**",

" <td>0.023565</td>**\n**",

" <td>NaN</td>**\n**",

" <td>1.000000</td>**\n**",

" <td>-0.062612</td>**\n**",

" </tr>**\n**",

" <tr>**\n**",

" <th>charges</th>**\n**",

" <td>0.798984</td>**\n**",

" <td>-0.061555</td>**\n**",

" <td>0.134148</td>**\n**",

" <td>0.157323</td>**\n**",

" <td>NaN</td>**\n**",

" <td>-0.062612</td>**\n**",

" <td>1.000000</td>**\n**",

" </tr>**\n**","</tbody>**\n**","</table>**\n**","</div>"

],

"text/plain":[

" age sex bmichildrensmoker regioncharges**\n**","age 1.000000-0.0165740.1227760.030059 NaN0.0028880.798984**\n**",

"sex -0.0165741.0000000.0105450.008622 NaN-0.017247-0.061555**\n**",

"bmi 0.1227760.0105451.0000000.000046 NaN0.1413890.134148**\n**",

"children0.0300590.0086220.0000461.000000 NaN0.0235650.157323**\n**",

"smoker NaN NaN NaN NaN NaN NaN NaN**\n**","region 0.002888-0.0172470.1413890.023565 NaN1.000000-0.062612**\n**",

"charges 0.798984-0.0615550.1341480.157323 NaN-0.0626121.000000"

]

},

"execution\_count": 133,"metadata":{},

"output\_type":"execute\_result"

}

],

"source":[

"data.corr()"

]

},

{

"cell\_type": "code","execution\_count": 134,"id":"b23f6bde",

"metadata":{},

"outputs": [],"source":[

"skewness=data.skew()"

]

},

{

"cell\_type": "code","execution\_count": 135,"id":"e2bc4a95",

"metadata": {},"outputs":[

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAjcAAAGwCAYAAABVdURTAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["plt.bar(skewness.index,skewness)**\n**","plt.xlabel('variables')**\n**","plt.ylabel('skewness')**\n**","plt.show()"

]

},

{

"cell\_type": "code","execution\_count": 136,"id":"2f891d1b",

"metadata":{},

"outputs": [],"source":[

"data['children']=np.sqrt(data['children'])**\n**"

]

},

{

"cell\_type": "code","execution\_count": 137,"id":"8a654e7c",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"age 0.068391**\n**",

"sex 0.064258**\n**",

"bmi0.100436**\n**","children0.171790**\n**","smoker0.000000**\n**","region-0.026116**\n**","charges0.546046**\n**","dtype:float64"

]

},

"execution\_count": 137,"metadata":{},

"output\_type":"execute\_result"

}

],

"source":[

"data.skew()"

]

},

{

"cell\_type": "code","execution\_count": 138,"id":"c393593c",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<BarContainerobjectof7artists>"

]

},

"execution\_count": 138,"metadata":{},

"output\_type":"execute\_result"

},

{

"data":{

"image/png":"iVBORw0KGgoAAAANSUhEUgAAAiMAAAGdCAYAAADAAnMpAAAAOXRFWHRTb2Z0d2FyZQBNYXRwbG90bGliIHZlcnNpb24zLjYuMCwgaHR0cHM6Ly9tYXRwbG90bGliLm9yZy89olMNAAAACXBIWXMAAA9hAAAPYQGoP6dp"text/plain":[

"<Figuresize640x480with1Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["plt.bar(data.skew().index,data.skew())"

]

},

{

"cell\_type": "code","execution\_count": 139,"id":"7f45e5dc",

"metadata": {},"outputs":[

{

"data":{

"text/plain": ["<AxesSubplot:>"

]

},

"execution\_count": 139,"metadata":{},

"output\_type":"execute\_result"

},

{

"data": {"image/png":"text/plain":[

"<Figuresize640x480with2Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source":["sns.heatmap(data,annot=True,cmap='Greens')"

]

},

{

"cell\_type": "code","execution\_count": 140,"id":"a4d6d802",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"<seaborn.axisgrid.PairGridat0x18ea7410460>"

]

},

"execution\_count": 140,"metadata":{},

"output\_type":"execute\_result"

},

{

"data": {"image/png":

"text/plain":[

"<Figuresize1750x1750with56Axes>"

]

},

"metadata":{},"output\_type":"display\_data"

}

],

"source": ["sns.pairplot(data)"

]

},

{

"cell\_type": "code","execution\_count": 141,"id":"3918a903",

"metadata":{},

"outputs": [],"source":[

"#modelbuilding**\n**",

"from sklearn.linear\_model import LinearRegression**\n**","model=LinearRegression()"

]

},

{

"cell\_type": "code","execution\_count": 142,"id":"b2b9005d",

"metadata":{},

"outputs": [],"source":[

"fromsklearn.decompositionimportPCA**\n**","pca=PCA(0.95)**\n**",

"data\_pca=pca.fit\_transform(data.drop('charges',axis='columns'))"

]

},

{

"cell\_type": "code","execution\_count": 147,"id":"d2e26715",

"metadata":{},

"outputs": [],"source":[

"from sklearn.model\_selection import train\_test\_split**\n**","train\_x,test\_x,train\_y,test\_y=train\_test\_split(data\_pca,data.charges,test\_size=0.2,random\_state=10)"

]

},

{

"cell\_type": "code","execution\_count": 148,"id":"fb651e4d",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"LinearRegression()"

]

},

"execution\_count": 148,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["model.fit(train\_x,train\_y)"

]

},

{

"cell\_type": "code","execution\_count": 149,"id":"74ecf910",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"0.6810590806361383"

]

},

"execution\_count": 149,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["model.score(test\_x,test\_y)"

]

},

{

"cell\_type": "code","execution\_count": 151,"id":"66d7ef8c",

"metadata":{},

"outputs": [],"source":[

"test\_pred=model.predict(test\_x)**\n**","train\_pre=model.predict(train\_x)"

]

},

{

"cell\_type": "code","execution\_count": 152,"id":"78e23a28",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"0.653781806666839"

]

},

"execution\_count": 152,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["model.score(train\_x,train\_y)"

]

},

{

"cell\_type": "code","execution\_count": 153,"id":"a7cb336f",

"metadata":{},

"outputs": [],"source":[

"fromsklearn.metricsimportr2\_score"

]

},

{

"cell\_type": "code","execution\_count": 154,"id":"d7a57157",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"0.6810590806361383"

]

},

"execution\_count": 154,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["#test\_score**\n**",

"r2\_score(test\_y,test\_pred)"

]

},

{

"cell\_type": "code","execution\_count": 155,"id":"a3a0bc5a",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"0.653781806666839"

]

},

"execution\_count": 155,"metadata":{},

"output\_type":"execute\_result"

}

],

"source": ["#train\_score**\n**",

"r2\_score(train\_y,train\_pre)"

]

},

{

"cell\_type": "code","execution\_count": 157,"id":"d4b1b7f3",

"metadata":{},

"outputs": [],"source":[

"fromsklearn.metricsimportmean\_absolute\_error,mean\_squared\_error"

]

},

{

"cell\_type": "code","execution\_count": 158,"id":"8fa496a3",

"metadata": {},"outputs":[

{

"data": {"text/plain":[

"0.018635568259422683"

]

},

"execution\_count":158,

"metadata":{},

"output\_type":"execute\_result"

}

],

"source":[

"#mae**\n**","mean\_absolute\_error(test\_y,test\_pred)"

]

},

{

"cell\_type": "code","execution\_count": 159,"id":"961a2c71",

"metadata":{},"outputs":[

{

"data": {"text/plain":[

"0.0015773874262789821"

]

},

"execution\_count": 159,"metadata":{},

"output\_type":"execute\_result"

}

],

"source":[

"#mse**\n**","mean\_squared\_error(test\_y,test\_pred)"

]

},

{

"cell\_type": "code","execution\_count": null,"id":"1e6330ca",

"metadata":{},

"outputs":[],

"source":[]

}

],

"metadata": {"kernelspec":{

"display\_name":"Python3(ipykernel)","language":"python",

"name":"python3"

},

"language\_info": {"codemirror\_mode":{

"name":"ipython","version":3

},

"file\_extension": ".py","mimetype": "text/x-python","name": "python","nbconvert\_exporter": "python","pygments\_lexer": "ipython3","version":"3.9.13"

}

},

"nbformat":4,

"nbformat\_minor":5

}