untitled10-1

June 13, 2024

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

df = pd.read_csv('insurance.csv')
```

[2]:
```
df.head()
```

[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

[3]:
```
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   region    1338 non-null   object 
 6   charges   1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

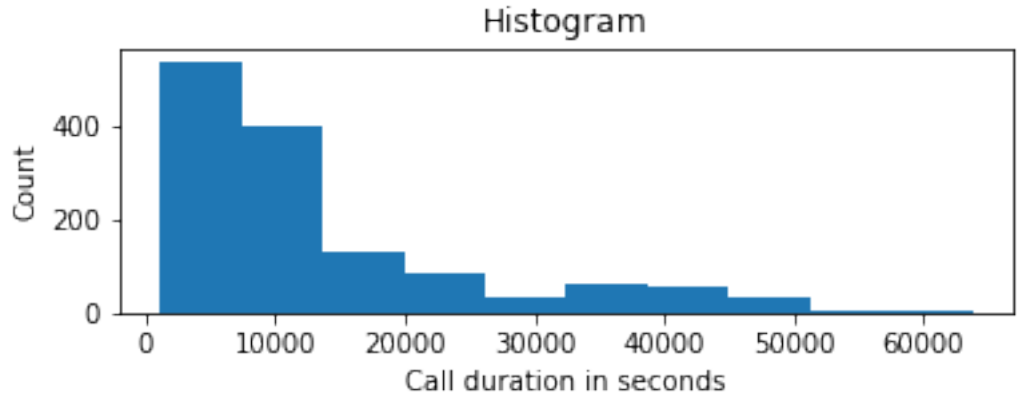
[5]:
```
df.isna().sum()
```

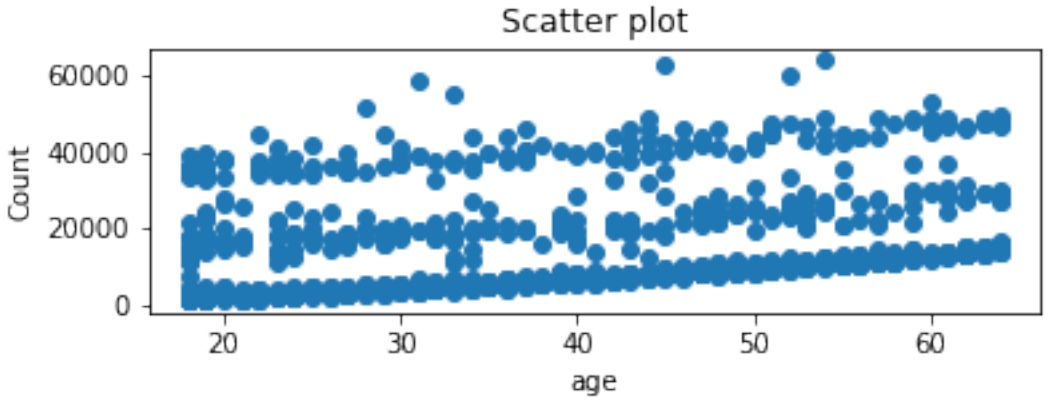
[5]:
```
age         0
sex         0
bmi         0
```

```
children    0
smoker      0
region      0
charges     0
dtype: int64
```

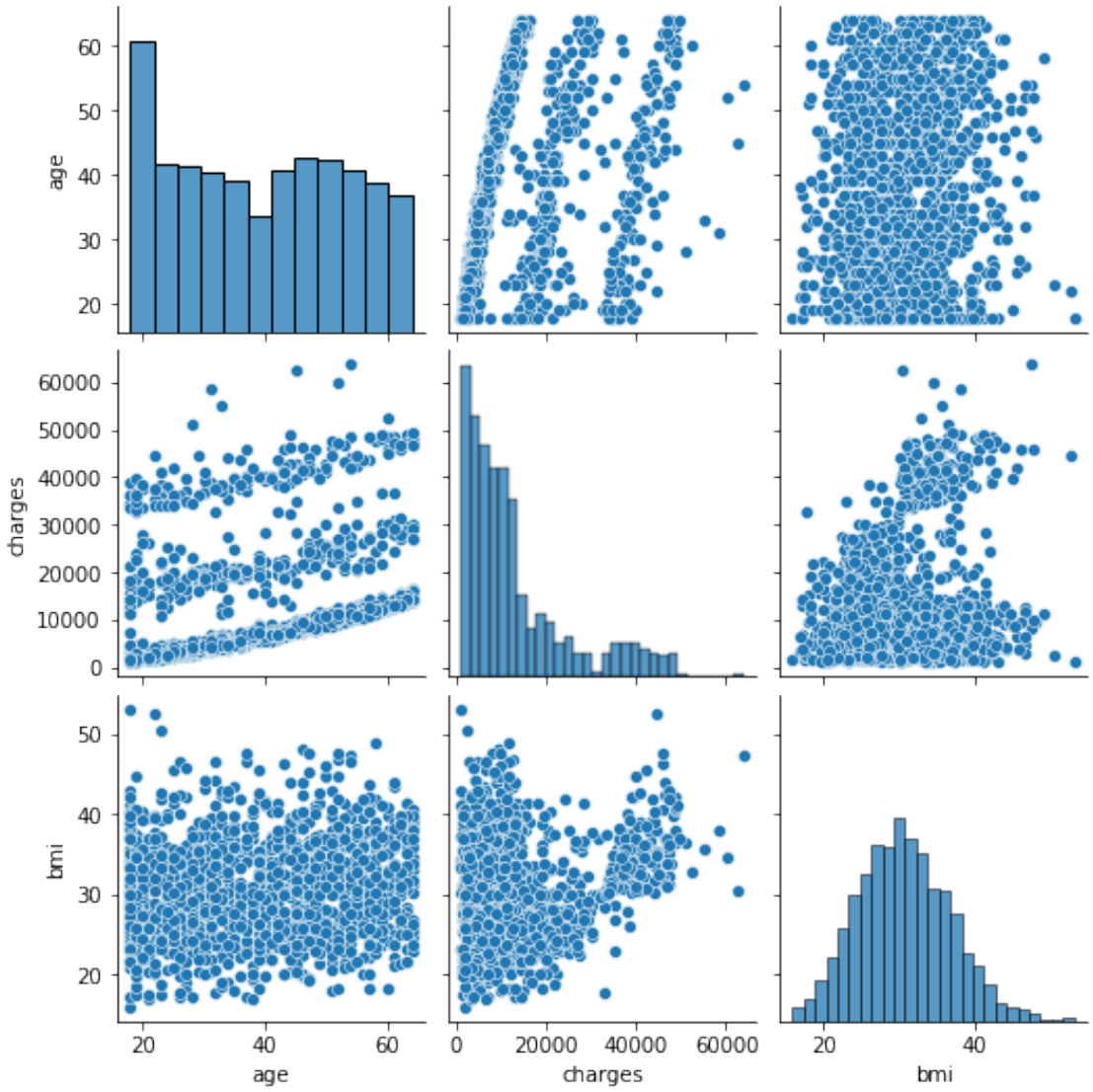
[9]:
```
plt.figure()
plt.subplot(2, 1, 1)
plt.hist(df['charges'])
plt.title('Histogram')
plt.xlabel('Call duration in seconds')
plt.ylabel('Count')
plt.show()

plt.subplot(2, 1, 2)
plt.scatter(x='age', y='charges', data= df)
plt.title('Scatter plot')
plt.xlabel('age')
plt.ylabel('Count')
plt.show()
```





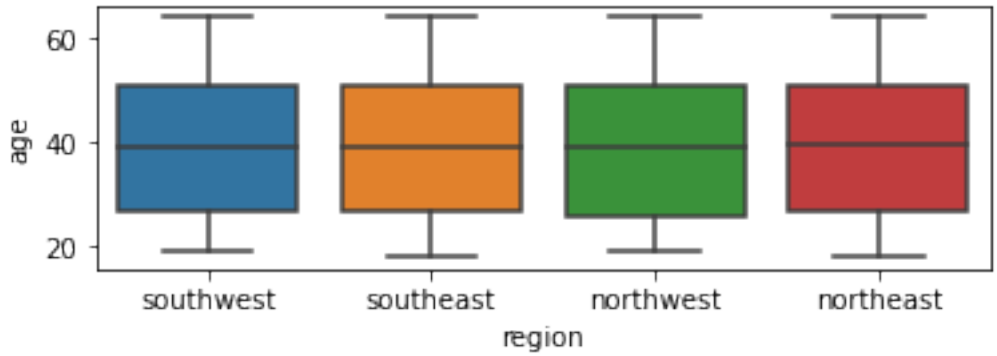
[13]:
```
cols =['age','region','charges','bmi']
sns.pairplot(df[cols])
plt.show()
```

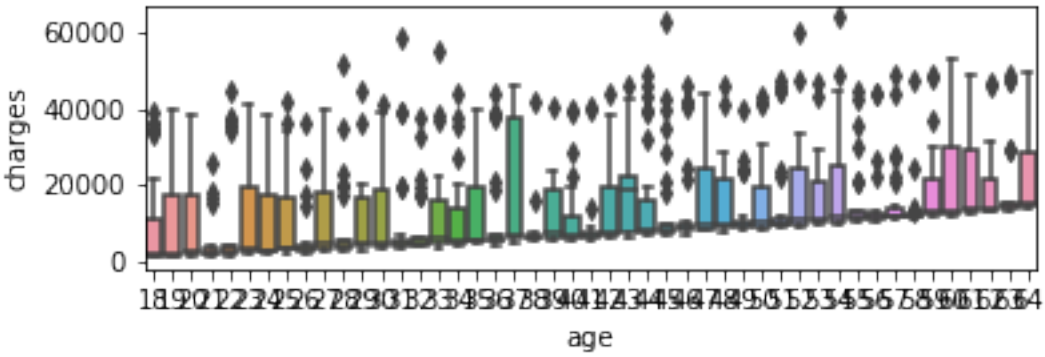


[15]:
```
plt.subplot(2,1,1)
sns.boxplot(x= 'region', y='age', data=df)
plt.show()

plt.subplot(2,1,2)
sns.boxplot(x ='age', y= 'charges', data=df)
```

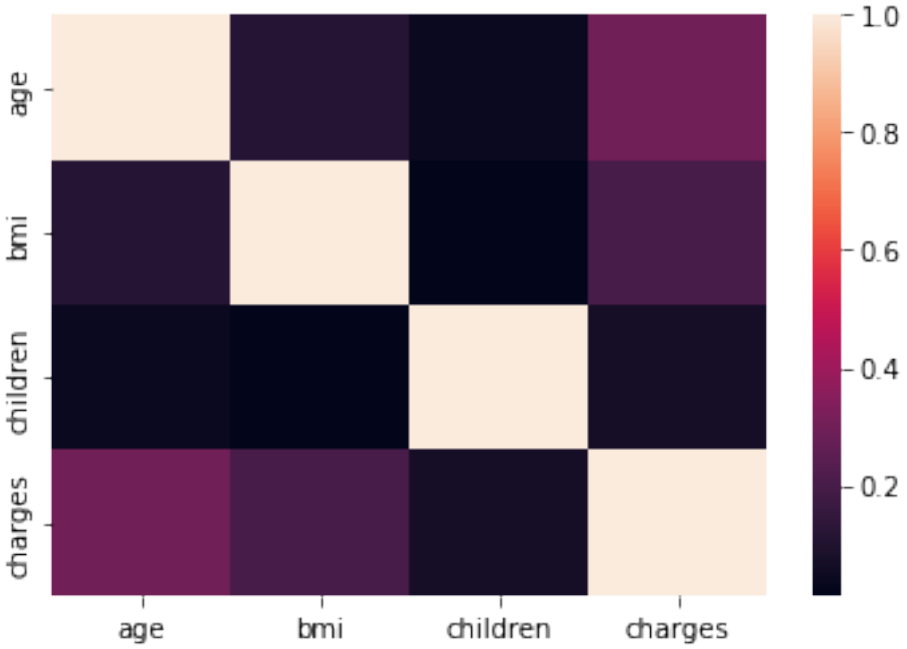
```
plt.show()
```





[16]:
```
sns.heatmap(df.corr())
plt.show()
```

```
/tmp/ipykernel_190/2975651719.py:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df.corr())
```


age
bmi
children
charges
age
bmi
children
charges
1.0
0.8
0.6
0.4
0.2