

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

```
In [2]: data = pd.read_csv('https://raw.githubusercontent.com/mwaskom/seaborn-data/master/tips.csv')
```

```
In [3]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
---  --
 0   total_bill  244 non-null    float64
 1   tip         244 non-null    float64
 2   sex         244 non-null    object  
 3   smoker      244 non-null    object  
 4   day         244 non-null    object  
 5   time        244 non-null    object  
 6   size        244 non-null    int64   
dtypes: float64(2), int64(1), object(4)
memory usage: 13.5+ KB
```

```
In [4]: data.head()
```

```
Out[4]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
In [5]: data.tail()
```

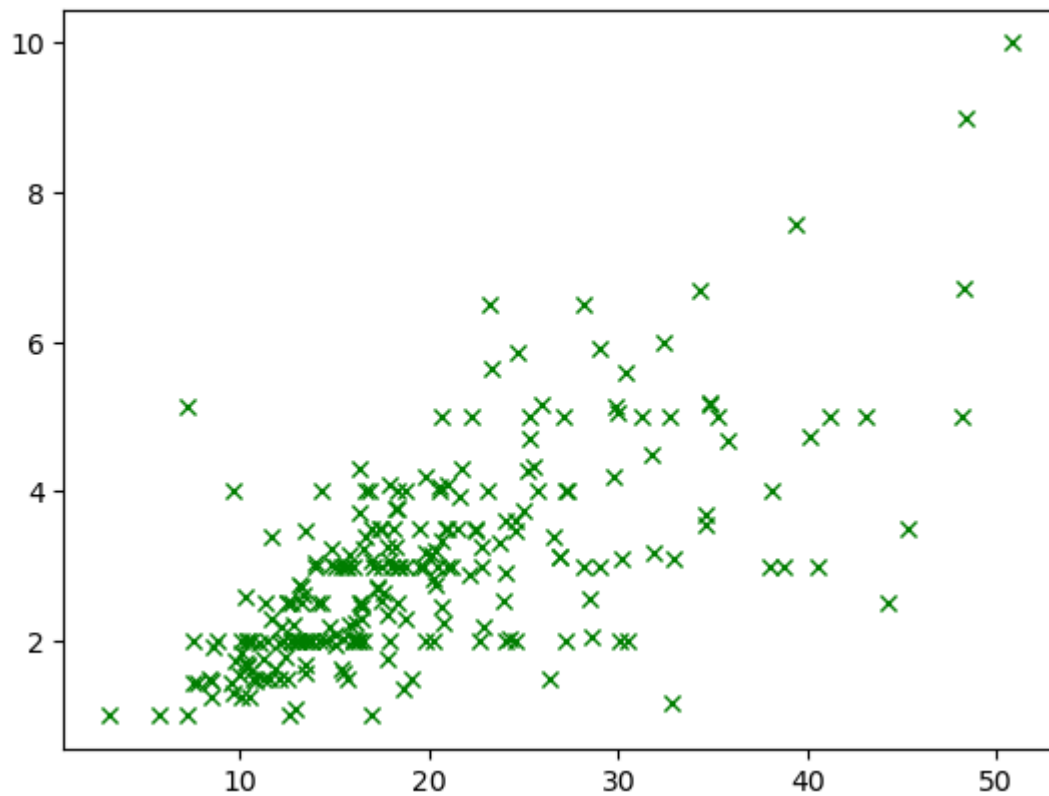
```
Out[5]:
```

	total_bill	tip	sex	smoker	day	time	size
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

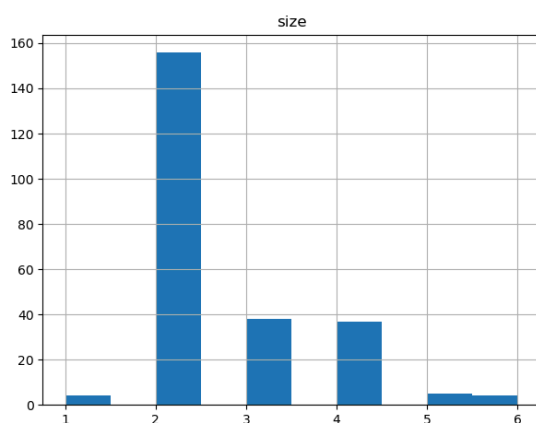
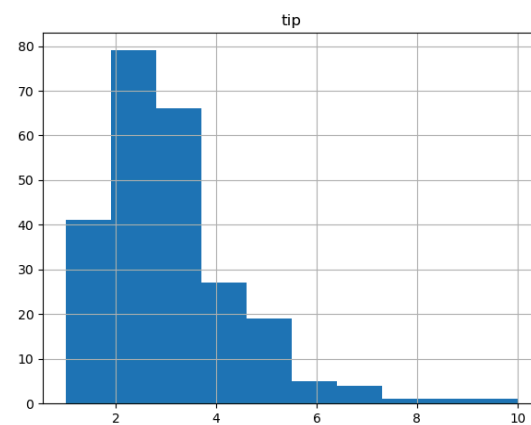
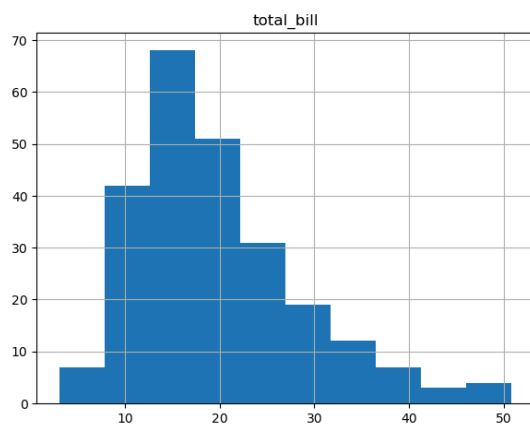
```
In [6]: data.isnull().sum()
```

```
Out[6]: total_bill    0  
        tip          0  
        sex          0  
        smoker       0  
        day          0  
        time         0  
        size         0  
        dtype: int64
```

```
In [7]: plt.plot(data.total_bill, data.tip, "gx")  
        plt.show()
```



```
In [9]: data.hist(bins=10, figsize=(16,12))  
        plt.show()
```



```
In [14]: y= data['total_bill']
        x= data['tip']
```

```
In [15]: x=np.array(x)
        x.reshape(244,1)
        x=pd.DataFrame(x)
        y=np.array(y)
        y.reshape(244,1)
        y=pd.DataFrame(y)
```

```
In [16]: from sklearn.model_selection import train_test_split
        xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size= 0.33, random_state=42)
        from sklearn.linear_model import LinearRegression
```

```
In [17]: linreg=LinearRegression()
```

```
In [18]: linreg.fit(xtrain, ytrain)
```

```
Out[18]: ▼ LinearRegression
        LinearRegression()
```

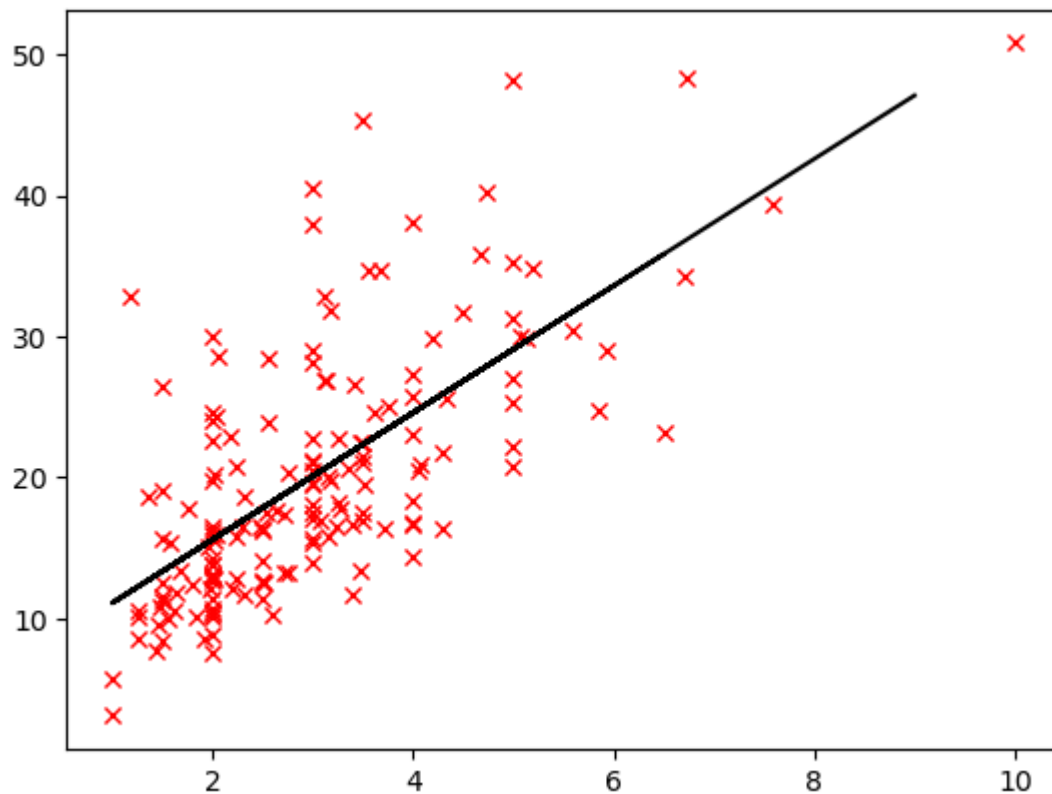
```
In [19]: b0 = linreg.intercept_
        b1 = linreg.coef_
```

```
In [20]: print('b0= ',b0)
        print('b1= ',b1)

b0= [6.57259258]
b1= [[4.5016331]]
```

```
In [21]: ypred = linreg.predict(xtest)
```

```
In [24]: plt.plot(xtrain, ytrain, 'rx')  
plt.plot(xtest, ypred, "black")  
plt.show()
```



```
In [27]: from sklearn import metrics  
# print result of MAE  
print(metrics.mean_absolute_error(ytest, ypred))  
  
#print result of MSE  
print(metrics.mean_squared_error(ytest, ypred))  
  
#print result of RMSE  
print(np.sqrt(metrics.mean_squared_error(ytest, ypred)))
```

```
4.59551637169004  
43.98946964623296  
6.632455777932708
```