TASK 4

SALES PREDICTION USING PYTHON

Importing libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

IMPORT DATASET

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

```
df.head(3)
```

	TV	Radio	Newspaper	Sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	12.0

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.shape
```

```
(200, 4)
```

```
df.describe()
```

	TV	Radio	Newspaper	Sales
count	200.000000	200.000000	200.000000	200.000000
mean	147.042500	23.264000	30.554000	15.130500
std	85.854236	14.846809	21.778621	5.283892
min	0.700000	0.000000	0.300000	1.600000
25%	74.375000	9.975000	12.750000	11.000000
50%	149.750000	22.900000	25.750000	16.000000
75%	218.825000	36.525000	45.100000	19.050000
max	296.400000	49.600000	114.000000	27.000000

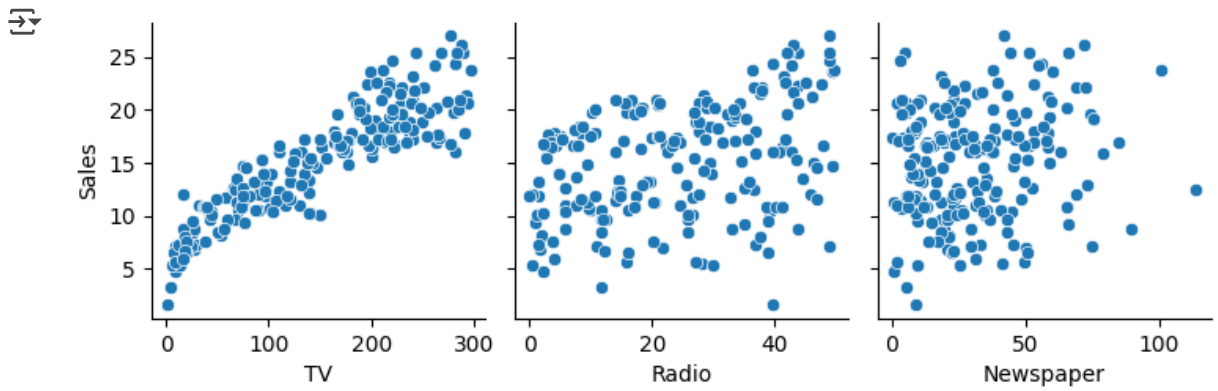
Basic observation

Avg expense spend is highest on TV.

Avg expense spend is lowest on Radio

Max sale is 27 & min is 1.6

```
sns.pairplot(df,x_vars=['TV','Radio','Newspaper'],y_vars='Sales',kind='scatter')
plt.show()
```



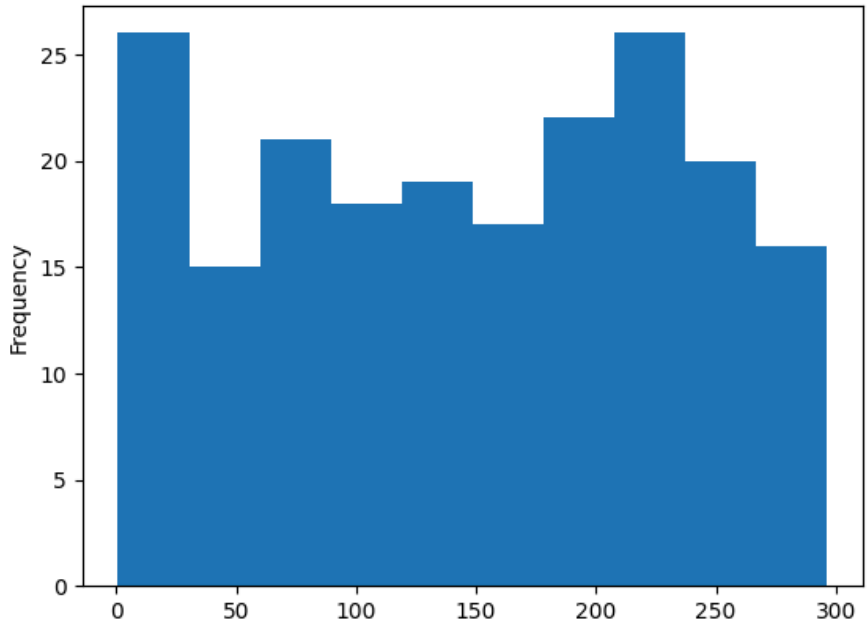
Pair Plot Observation

When advertising cost increases in TV ads the sales will increses as well.while the for newsaper and Radio it is unpredictable

if we cannot determine the correlation using a scatter plot. we can use the seaborn heatmap to visualize the data

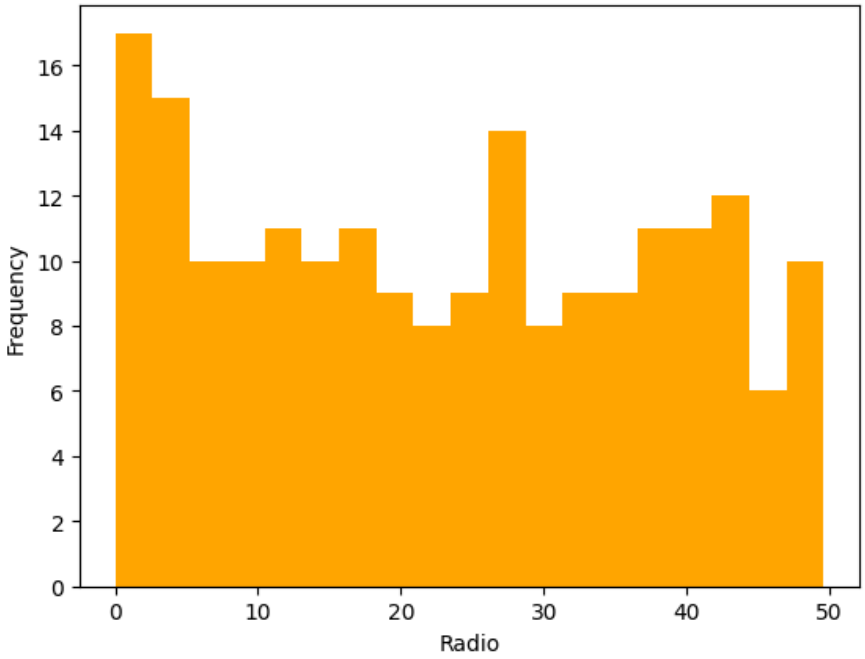
```
df['TV'].plot.hist(bins=10)
```

```
<Axes: ylabel='Frequency'>
```



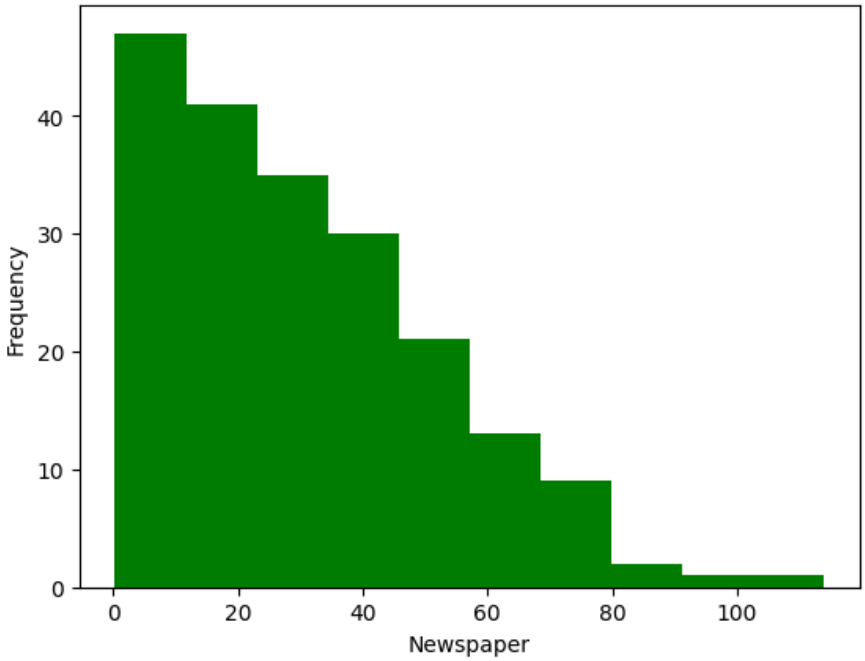

```
df["Radio"].plot.hist(bins=19,color="orange",xlabel="Radio")
```

```
<Axes: xlabel='Radio', ylabel='Frequency'>
```



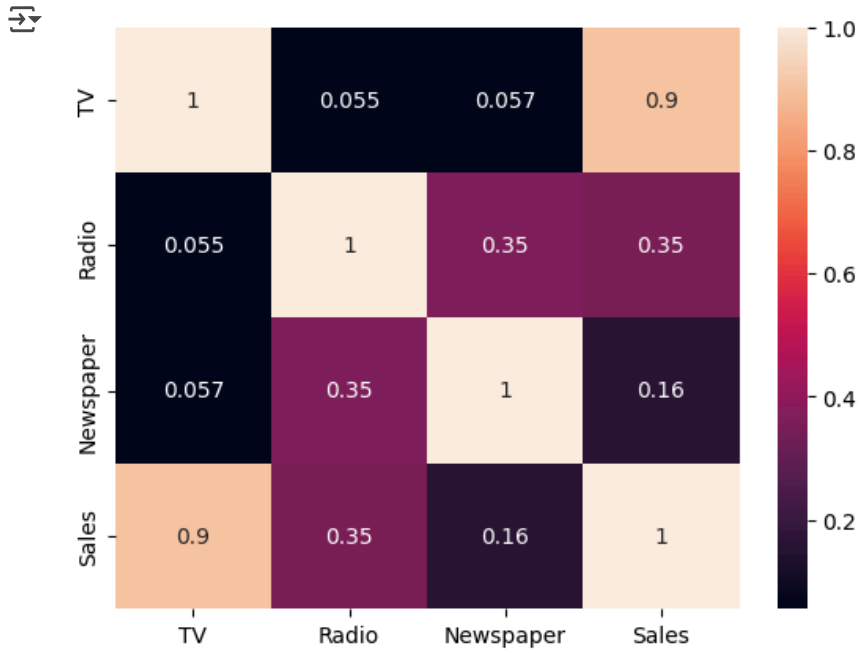

```
df["Newspaper"].plot.hist(bins=10,color="Green",xlabel="Newspaper")
```

```
<Axes: xlabel='Newspaper', ylabel='Frequency'>
```



The Majority Sales is the result of Now advertising cost in Newspaper

```
sns.heatmap(df.corr(),annot=True)
plt.show()
```



Sales is highly correlated with the TV

Lets train our model using linear regression as it is correlated with only

To generate independent variable represent X and Y represents the largest variable in a simple linear regression model

```
x=df['TV']
y=df['Sales']
```

```
from sklearn.model_selection import train_test_split
xTrain,xTest,yTrain,yTest=train_test_split(df[['TV']],df[['Sales']],test_size=0.3,random_state=0)
```

```
print(xTrain)
```

```
        TV
131  265.2
96   197.6
181  218.5
19   147.3
153  171.3
..     ...
67   139.3
192   17.2
117   76.4
47   239.9
172   19.6

[140 rows x 1 columns]
```

```
print(xTest)
```

```
107   90.4
98   289.7
177  170.2
182   56.2
5      8.7
```

```
61   261.3
125   87.2
180  156.6
154  187.8
80    76.4
7    120.2
33   265.6
130    0.7
37    74.7
74   213.4
183  287.6
145  140.3
45   175.1
159  131.7
60    53.5
123  123.1
179  165.6
185  205.0
122  224.0
44    25.1
16    67.8
55   198.9
150  280.7
111  241.7
22    13.2
189   18.7
129   59.6
4    180.8
83    68.4
106   25.0
134   36.9
66    31.5
26   142.9
113  209.6
168  215.4
63   102.7
8      8.6
75    16.9
118  125.7
143  104.6
71   109.8
124  229.5
184  253.8
97   184.9
149   44.7
24    62.3
30   292.9
160  172.5
40   202.5
56     7.3
```

```
print(yTrain)
```

```
      Sales
131   17.7
96    16.7
181   17.2
19    14.6
153   16.0
..     ...
67    13.4
192    5.9
117    9.4
47    23.2
172    7.6

[140 rows x 1 columns]
```

```
print(yTest)
```

```
107   12.0
98    25.4
177   16.7
182    8.7
5      7.2
146   18.2
12     9.2
152   16.6
61    24.2
125   10.6
180   15.5
154   20.6
80    11.8
7     13.2
33    17.4
130    1.6
37    14.7
74    17.0
183   26.2
145   10.3
45    16.1
159   12.9
60     8.1
123   15.2
179   17.6
185   22.6
122   16.6
44     8.5
16    12.5
55    23.7
150   16.1
111   21.8
22     5.6
189    6.7
129    9.7
4     17.9
83    13.6
106    7.2
134   10.8
66    11.0
26    15.0
113   20.9
168   17.1
63    14.0
8      4.8
75     8.7
118   15.9
143   10.4
71    12.4
124   19.7
184   17.6
97    20.5
149   10.1
24     9.7
30    21.4
160   16.4
40    16.6
56     5.5
```

```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
s=model.fit(xTrain,yTrain)
s
```

```
▾ LinearRegression
LinearRegression()
```

```
res=model.predict(xTest)
print(res)
```

```
[12.09159447]
[22.99968079]
[16.45920756]
[10.21976029]
[ 7.6199906 ]
[20.28497391]
[ 8.4464437 ]
[17.95886418]
[21.44529217]
[11.91645209]
[15.71485245]
[17.42249065]
[11.32534656]
[13.72260788]
[21.68063975]
[ 7.18213465]
[11.23230217]
[18.82362968]
[22.88474361]
[14.82272095]
[16.72739433]
[14.35202581]
[10.07198391]
[13.88133066]
[16.20744039]
[18.36388094]
[19.40378881]
[ 8.51759529]
[10.85465142]
[18.03001578]
[22.50709285]
[20.3725451 ]
[ 7.86628457]
[ 8.16731053]
[10.40584907]
[17.03936669]
[10.88749061]
[ 8.51212209]
[ 9.16343282]
[ 8.86788005]
[14.96502414]
[18.61564811]
[18.93309367]
[12.76479799]
[ 7.6145174 ]
[ 8.06879294]
[14.02363385]
[12.86878878]
[13.15339515]
[19.70481478]
[21.03480222]
[17.26376787]
[ 9.59034237]
[10.55362545]
[23.17482317]
[16.58509115]
[18.22705095]
[ 7.54336581]]
```

```
model.coef_
```

```
array([[0.05473199]])
```

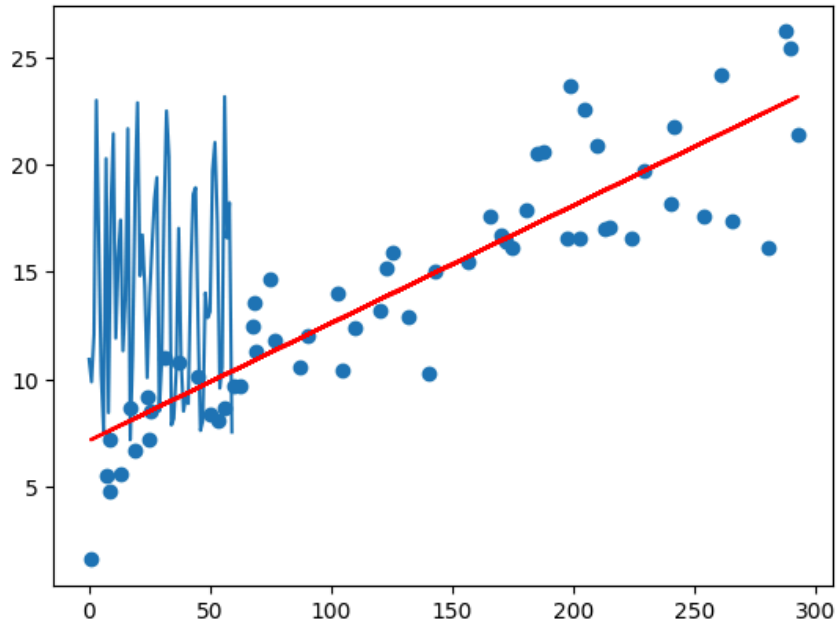
```
model.intercept_
```

```
array([7.14382225])
```

visulization the linear plot

```
plt.plot(res)
plt.scatter(xTest,yTest)
plt.plot(xTest,7.143822+0.054731*xTest,'r')
```

```
[<matplotlib.lines.Line2D at 0x7814e76d1790>]
```




```
yTrain_pred=model.predict(xTrain)
yTrain_pred
```

```
[13.49273351],
[11.25419497],
[ 9.22363801],
[16.27311878],
[ 8.61063968],
[13.73902748],
[21.53286336],
[19.97847475],
[20.26855431],
[22.79717242],
[19.31621762],
[ 9.305736  ],
[19.62271679],
[18.6813265 ],
[11.16115058],
[11.97665729],
[18.04096217],
[13.20265394],
[21.75179134],
[20.13719753],
[18.07927457],
[17.42796385],
[14.76798896],
[ 8.08521254],
[11.32534656],
[20.27402751],
[ 8.21656933]])
```

```
res=(yTrain-yTrain_pred)
print(res)
```

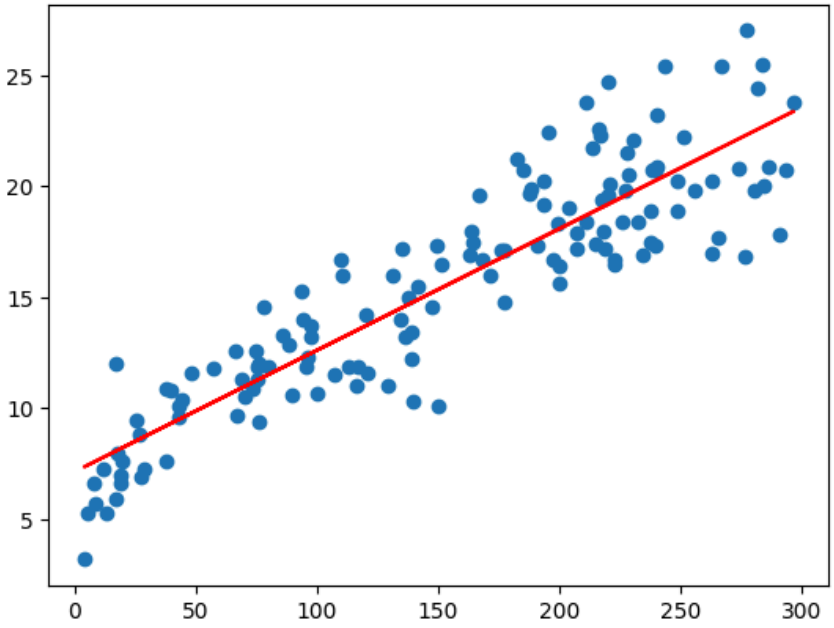
```
         Sales
131 -3.958747
96  -1.258864
181 -1.902763
19  -0.605845
153 -0.519413
..        ...
67  -1.367989
192 -2.185213
117 -1.925347
47   2.925972
172 -0.616569

[140 rows x 1 columns]
```

visualization the regression line

```
plt.scatter(xTrain,yTrain)
plt.plot(xTrain,7.143822+0.054731*xTrain,'r')
```

```
[<matplotlib.lines.Line2D at 0x7814e773a2d0>]
```



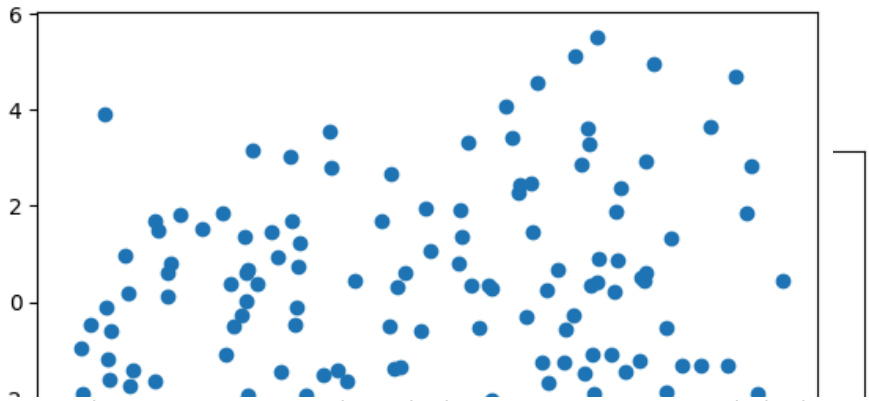
ploting the histogram using the residual values

```
fig=plt.figure()
sns.distplot(res,bins=15)
plt.xlabel('res')
plt.show()
```

```
<ipython-input-30-64ebfee4856b>:2: UserWarning:

plt.scatter(xTrain,res)
plt.show()
```

```
similar flexibility) or `histplot` (an axes-level function for histograms).
```



Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.