Date - 17/10/2023

Team ID - 3893

Project Title - Future Sales Prediction

Importing Dependencies

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import statsmodels.api as sm
```

Loading Dataset

```
dataset = pd.read_csv("C:\\Users\\Student\\Downloads\\
future_sales_prediction.csv")
```

Data Exploration

```
dataset
```

	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
3	151.5	41.3	58.5	16.5
4	180.8	10.8	58.4	17.9
..	...	...	...	...
195	38.2	3.7	13.8	7.6
196	94.2	4.9	8.1	14.0
197	177.0	9.3	6.4	14.8
198	283.6	42.0	66.2	25.5
199	232.1	8.6	8.7	18.4

```
[200 rows x 4 columns]
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   TV         200 non-null    float64
 1   Radio      200 non-null    float64
 2   Newspaper  200 non-null    float64
 3   Sales      200 non-null    float64
dtypes: float64(4)
memory usage: 6.3 KB
```

```
dataset.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
```

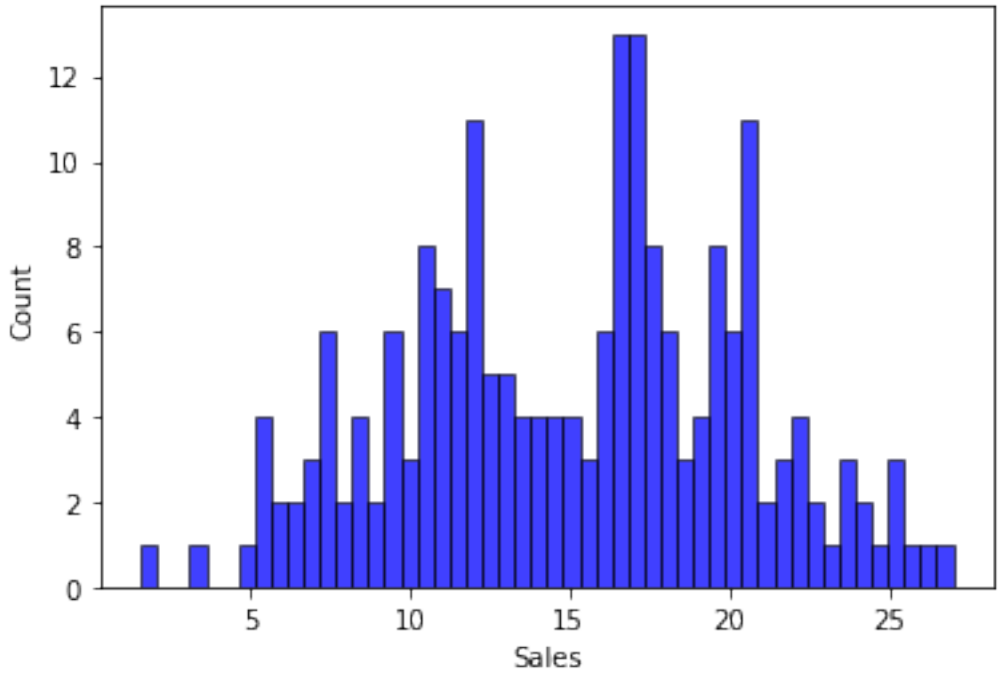
```
dataset.columns
```

```
Index(['TV', 'Radio', 'Newspaper', 'Sales'], dtype='object')
```

Pre-Processing and Visualisation of Data

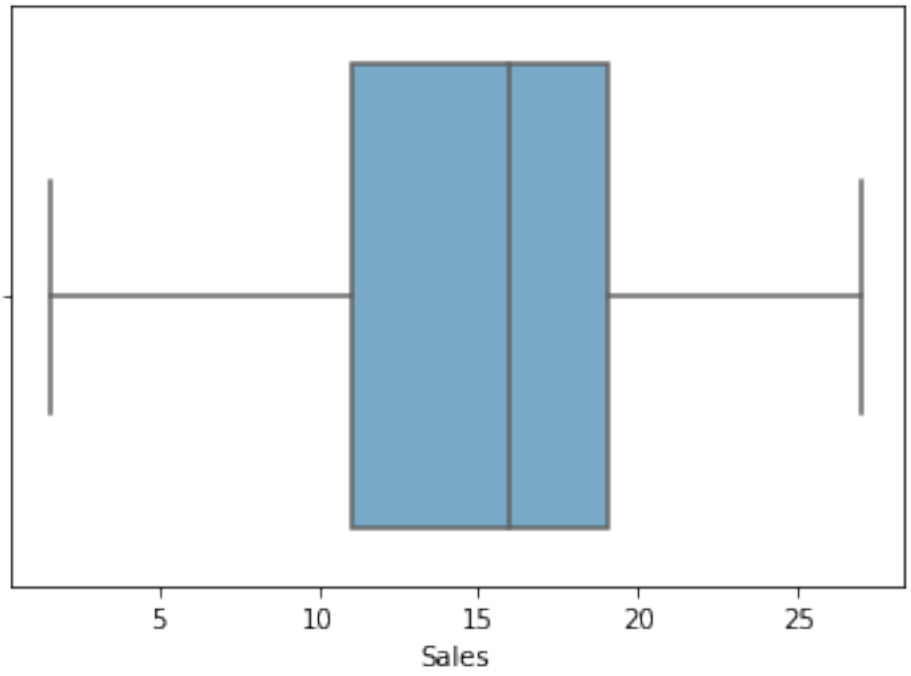
```
sns.histplot(dataset,x='Sales',bins=50,color='b')
```

```
<AxesSubplot:xlabel='Sales', ylabel='Count'>
```



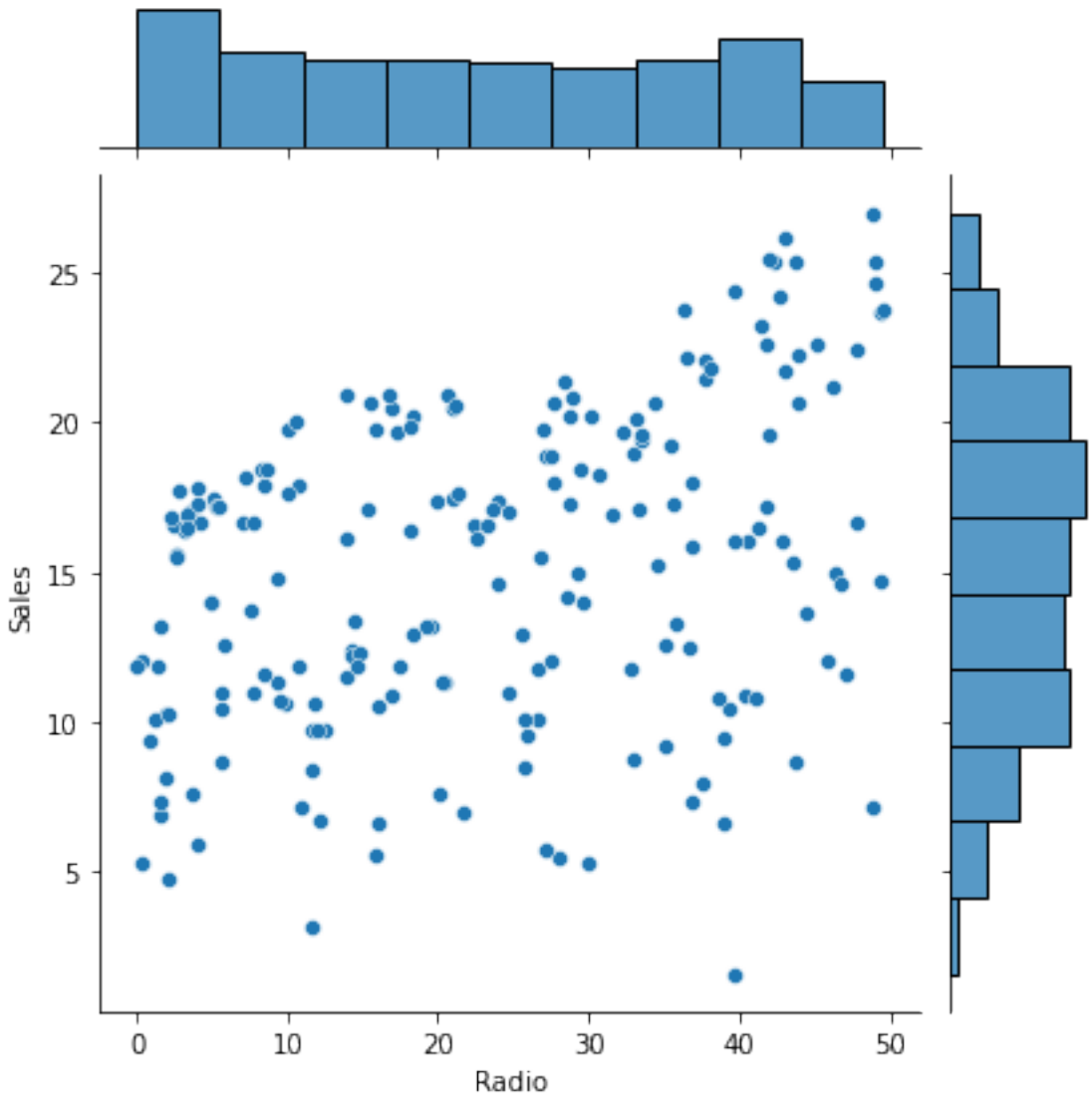

```
sns.boxplot(dataset,x='Sales',palette='Blues')
```

```
<AxesSubplot:xlabel='Sales'>
```



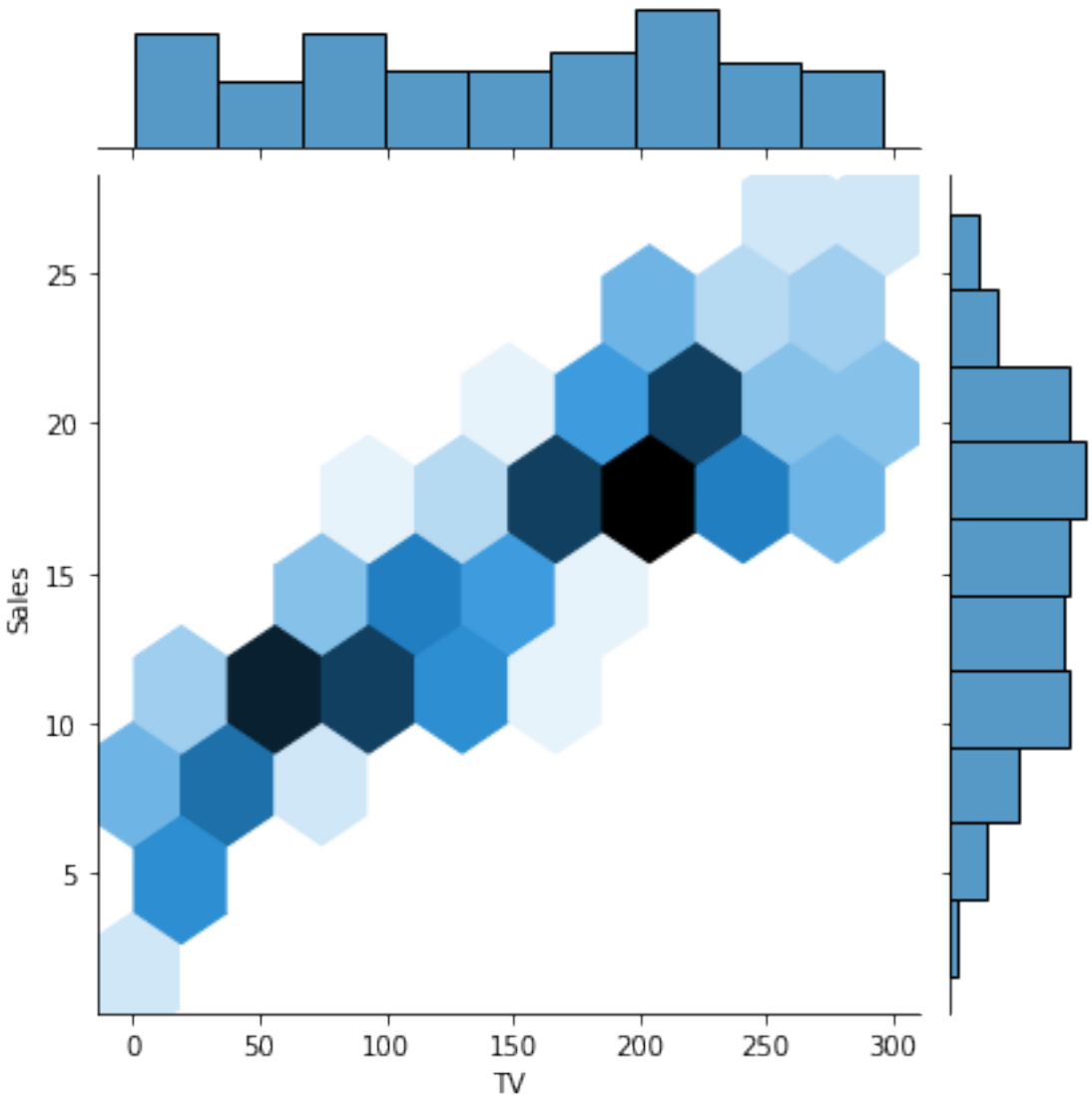

```
sns.jointplot(dataset,x='Radio',y='Sales',kind='scatter')
```

```
<seaborn.axisgrid.JointGrid at 0x4066830>
```




```
sns.jointplot(dataset,x='TV',y='Sales',kind='hex')
```

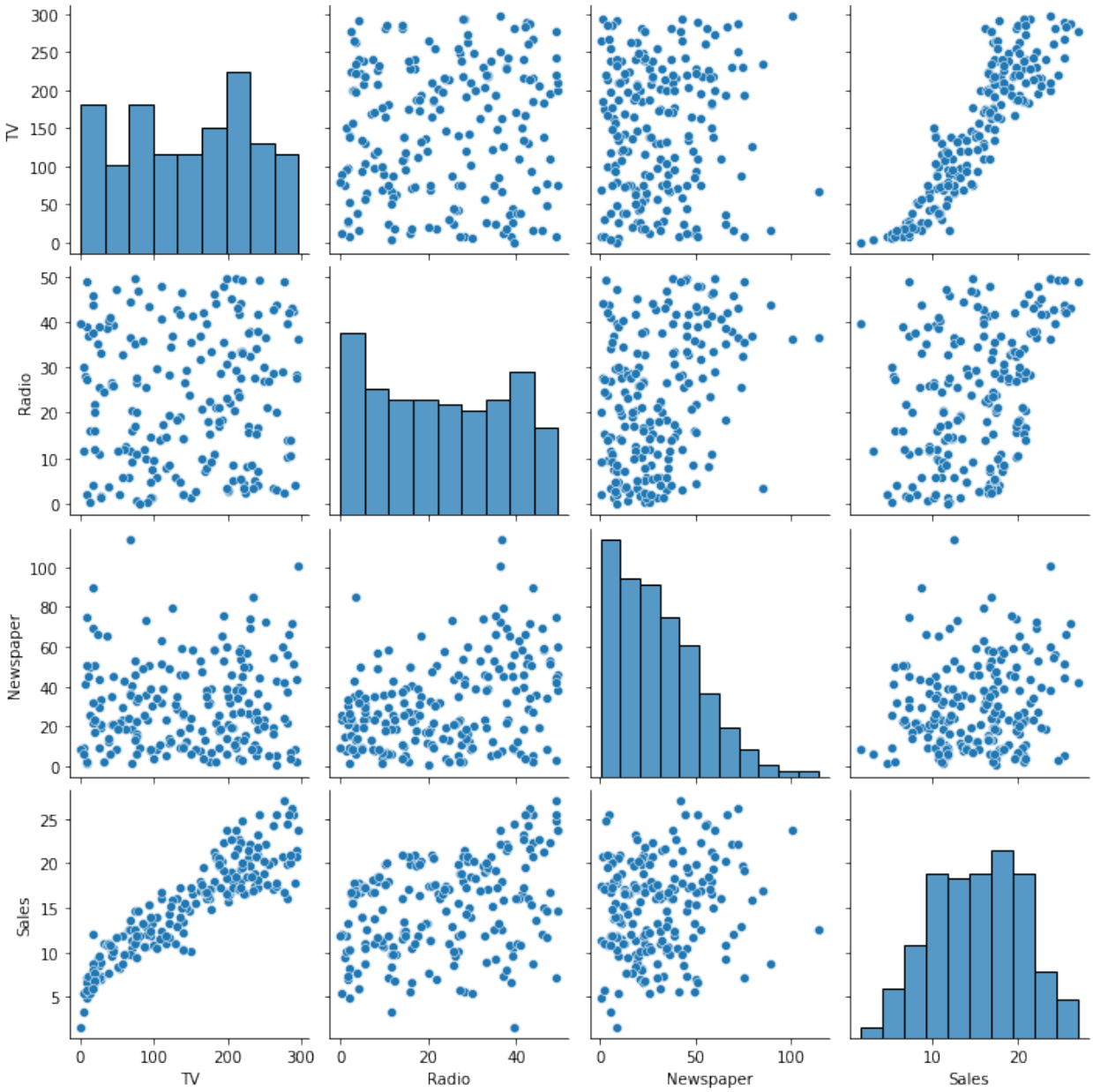
```
<seaborn.axisgrid.JointGrid at 0xd534070>
```




```
plt.figure(figsize=(12,8))
sns.pairplot(dataset)
```

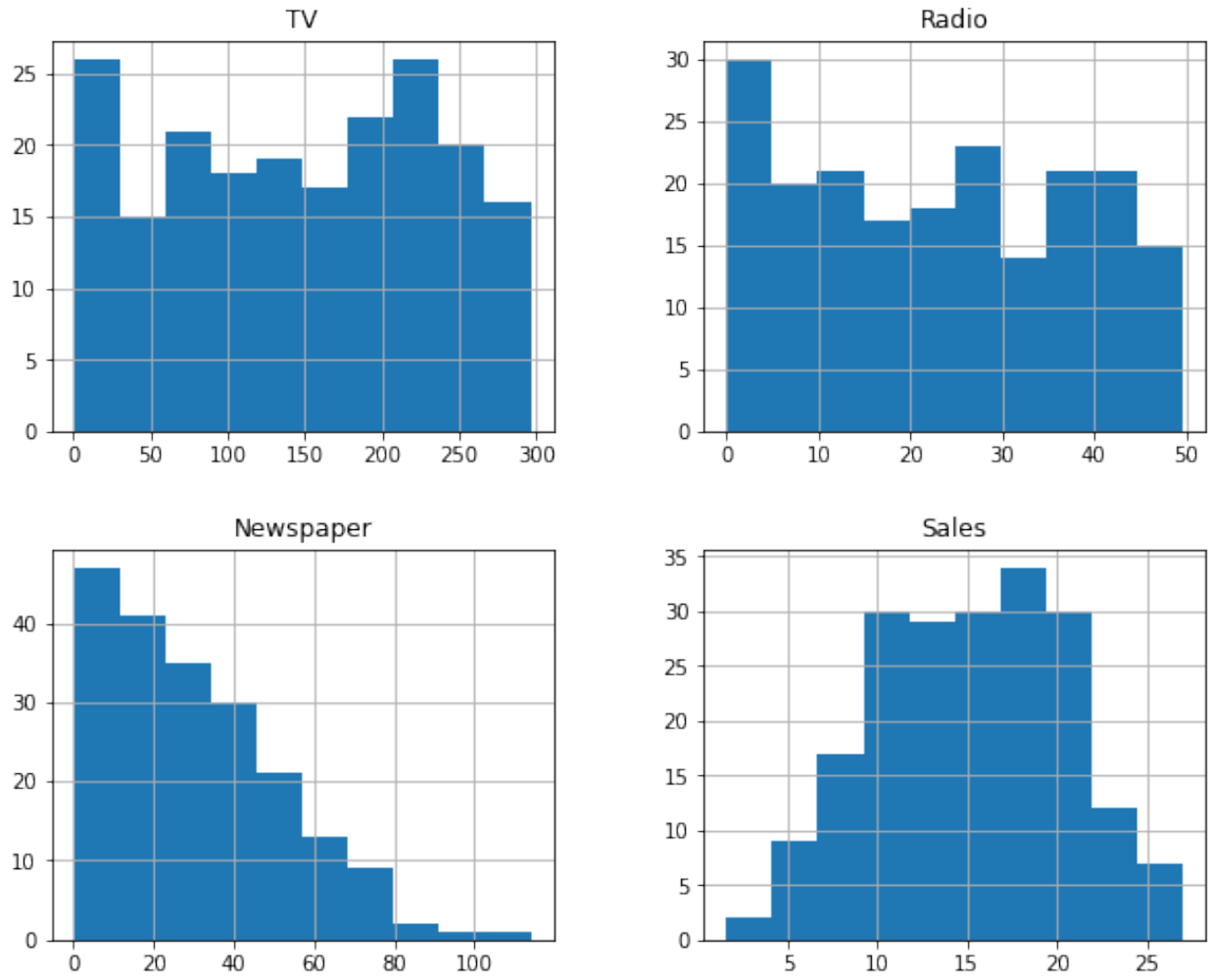
```
<seaborn.axisgrid.PairGrid at 0x1294f8f0>
```

```
<Figure size 864x576 with 0 Axes>
```




```
dataset.hist(figsize=(10,8))
```

```
array([[<AxesSubplot:title={'center':'TV'}>,
        <AxesSubplot:title={'center':'Radio'}>],
       [<AxesSubplot:title={'center':'Newspaper'}>,
        <AxesSubplot:title={'center':'Sales'}>]], dtype=object)
```



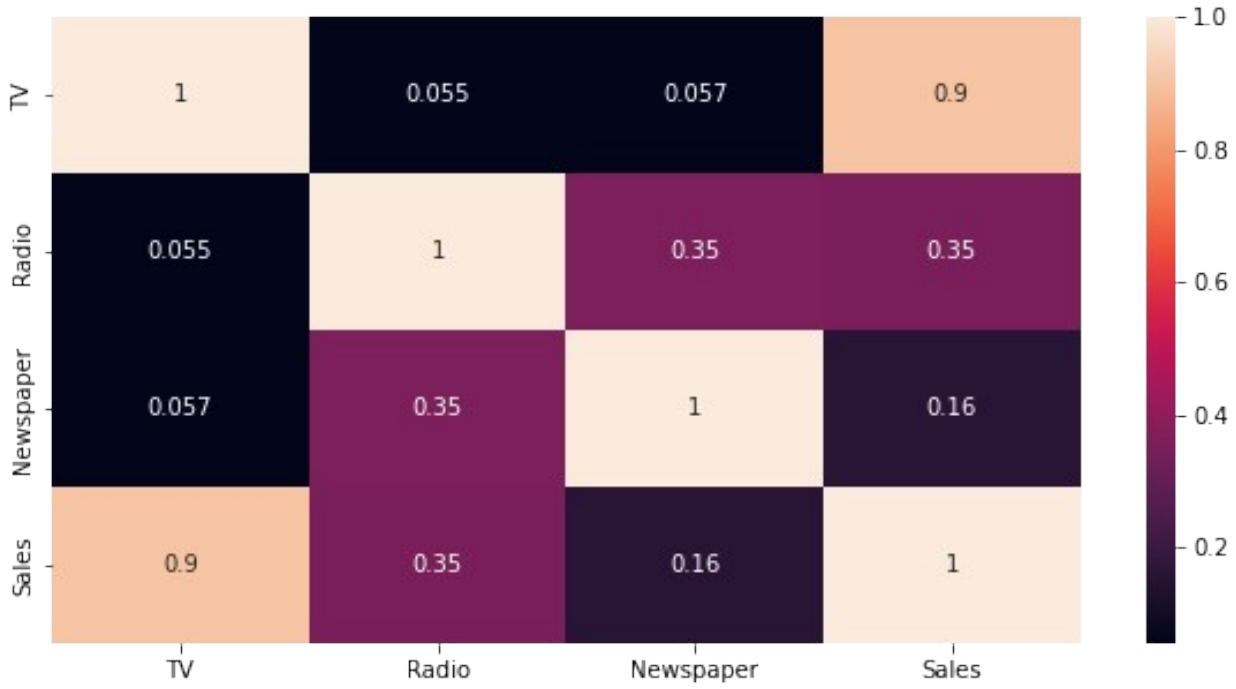
Visualisation Correlation

```
dataset.corr()
```

	TV	Radio	Newspaper	Sales
TV	1.000000	0.054809	0.056648	0.901208
Radio	0.054809	1.000000	0.354104	0.349631
Newspaper	0.056648	0.354104	1.000000	0.157960
Sales	0.901208	0.349631	0.157960	1.000000

```
plt.figure(figsize=(10,5))
sns.heatmap(dataset.corr(),annot=True)
```

```
<AxesSubplot:>
```


TV
Radio
Newspaper
Sales
1
0.055
0.057
0.9
0.055
1
0.35
0.35
0.057
0.35
1
0.16
0.9
0.35
0.16
1
TV
Radio
Newspaper
Sales
1.0
0.8
0.6
0.4
0.2