

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

Reading csv File

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

df

	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]

Head Method

```
df.head(2)
```

	TV	Radio	Newspaper	Sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4

Tail Method

```
df.tail(2)
```

	TV	Radio	Newspaper	Sales
198	283.6	42.0	66.2	25.5
199	232.1	8.6	8.7	18.4

Info Method

```
df.info
```

```
<bound method DataFrame.info of          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]>
```

IsNull Method

```
df.isnull().sum()
```

```
TV          0
Radio        0
Newspaper    0
Sales        0
dtype: int64
```

Describe Method

```
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

Total No of Rows and Columns

```
print ("Total no of Rows ==>", df.shape[1])  
print ("Total no of Columns ==>", df.shape[0])
```

```
Total no of Rows ==> 4  
Total no of Columns ==> 200
```

Second Highest in Column

```
second_highest = df.apply(lambda x: x.nlargest(2).iloc[-1])
```

```
print(second_highest)
```

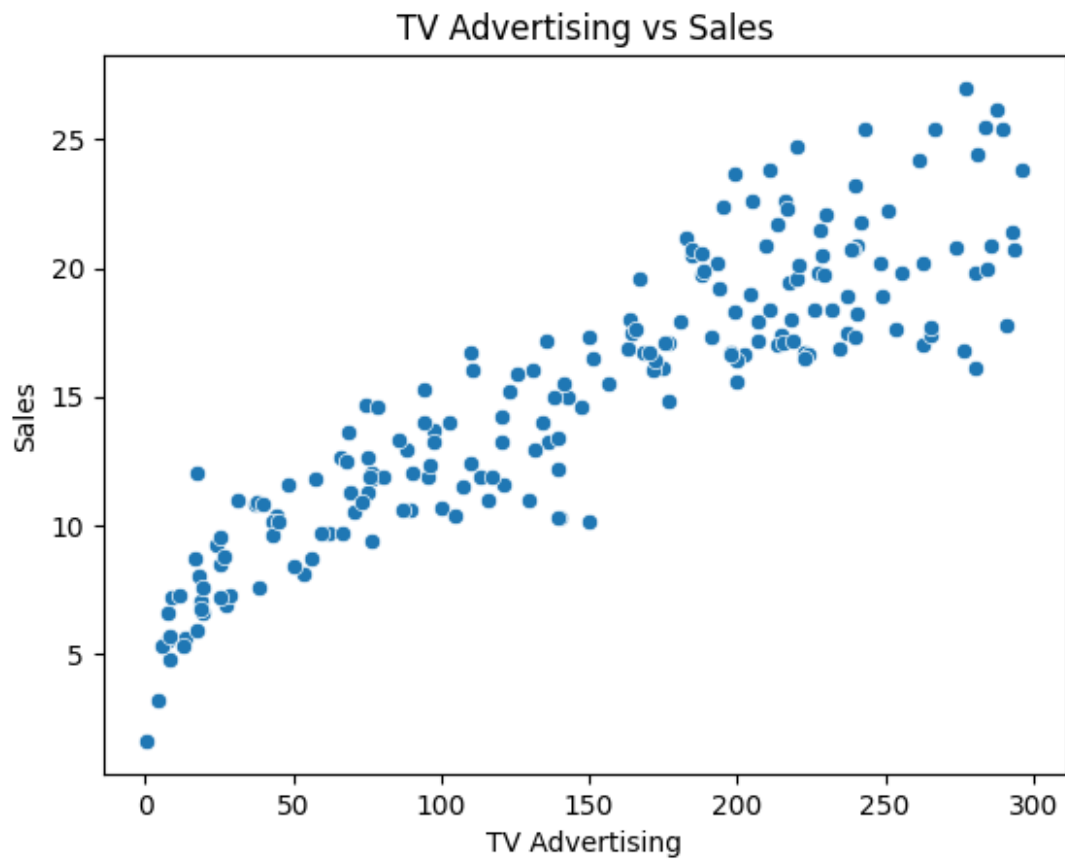
```
TV          293.6  
Radio        49.4  
Newspaper   100.9  
Sales        26.2  
dtype: float64
```

```
len(df.drop_duplicates())
```

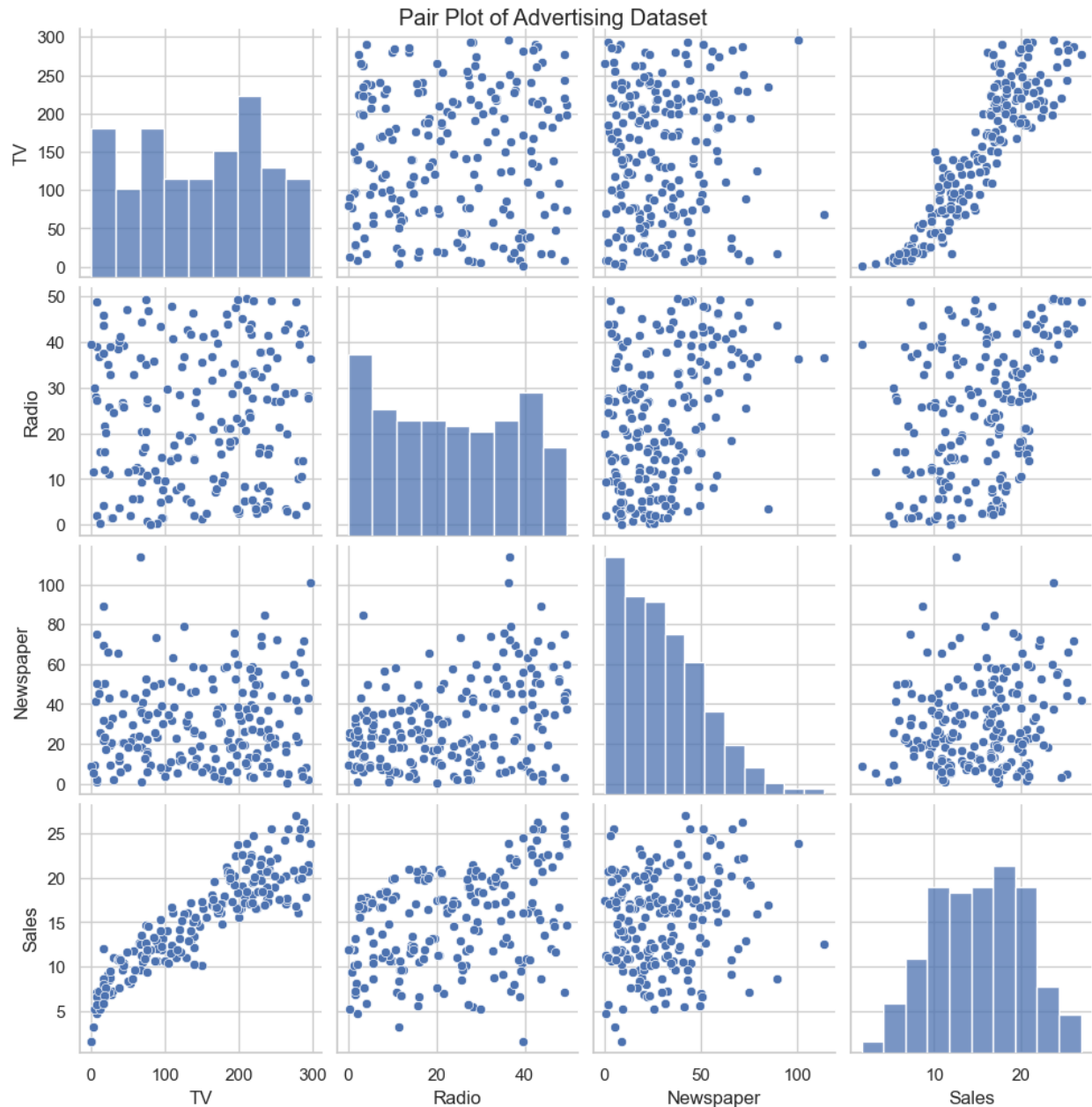
```
200
```

Data Visualization

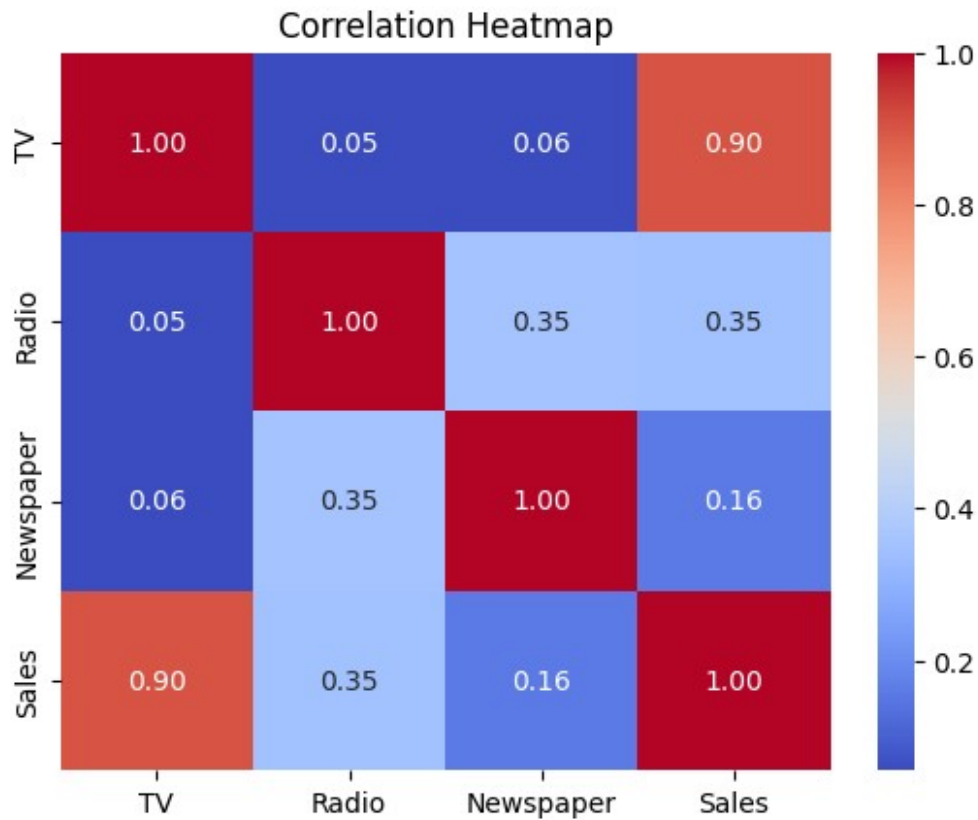
```
sb.scatterplot(data=df, x='TV', y='Sales')  
plt.title('TV Advertising vs Sales')  
plt.xlabel('TV Advertising')  
plt.ylabel('Sales')  
plt.show()
```



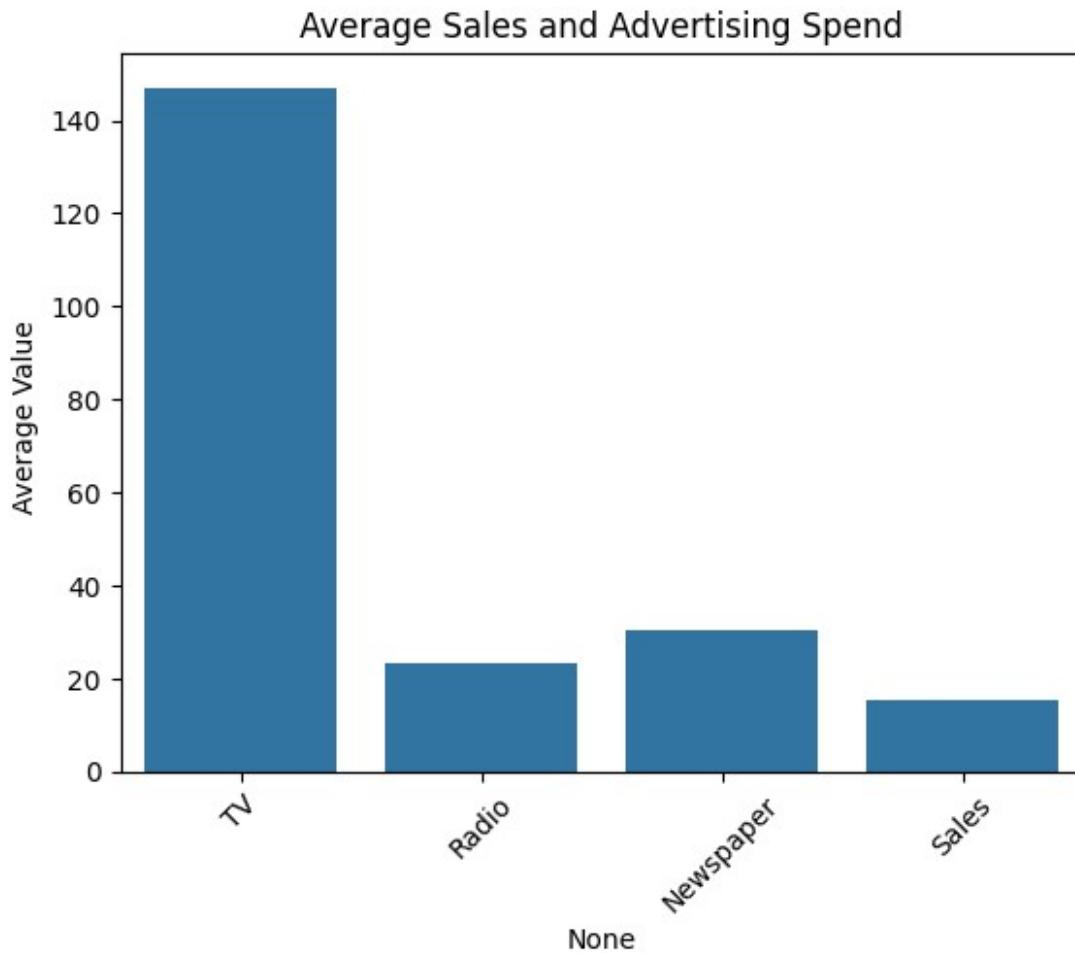
```
sb.pairplot(df)
plt.suptitle('Pair Plot of Advertising Dataset', y=1.00)
plt.show()
```



```
correlation = df.corr()
sb.heatmap(correlation, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```



```
average_sales = df[['TV', 'Radio', 'Newspaper', 'Sales']].mean()  
sb.barplot(x=average_sales.index, y=average_sales.values)  
plt.title('Average Sales and Advertising Spend')  
plt.ylabel('Average Value ')  
plt.xticks(rotation=45)  
plt.show()
```



```
sb.set(style='whitegrid')  
  
# Example of a customized scatter plot  
sb.scatterplot(data=df, x='Radio', y='Sales', hue='TV', size='TV',  
               sizes=(20, 200))  
plt.title('Radio Advertising vs Sales (Colored by TV Spend)')  
plt.xlabel('Radio Advertising Spend ' )  
plt.ylabel('Sales')  
plt.show()
```

