ASSIGNMENT - 3

Finding the relationship between advertising media ans sales

```
In [1]:    # importing packages pandas and numpy
           import pandas as pd
           import numpy as np
```

```
In [2]:    # loading the dataset
           advertising = pd.read_csv("Advertising.csv")
```

Advertising dataset provides us the information about the sales of a certain product in 200 different market along with its budget for each of those markets in 3 different media.

```
In [3]:    advertising.head()
```

Out[3]:

	Unnamed: 0	TV	radio	newspaper	sales
0	1	230.1	37.8	69.2	22.1
1	2	44.5	39.3	45.1	10.4
2	3	17.2	45.9	69.3	9.3
3	4	151.5	41.3	58.5	18.5
4	5	180.8	10.8	58.4	12.9

```
In [4]:    # removing column "unnamed: 0" since it is of no use
           advertising = advertising.drop(['Unnamed: 0'], axis=1)
```

```
In [5]:    advertising.head()
```

Out[5]:

	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	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

Advertising dataframe contains 4 columns and 200 entries with no missing values.

```
In [6]:    advertising.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: 6.4 KB
```

Our goal is to determine the relation between advertising and sales in order to indirectly increase the sales by adjusting the budget for advertisement.

```
In [7]:    # importing packages seaborn and matplotlib
           import seaborn as sns
           import matplotlib.pyplot as plt
```

```
In [8]:    # visualizing the dataset
           # plotting 3 different media with the sales in order to understand the relation between them.
```
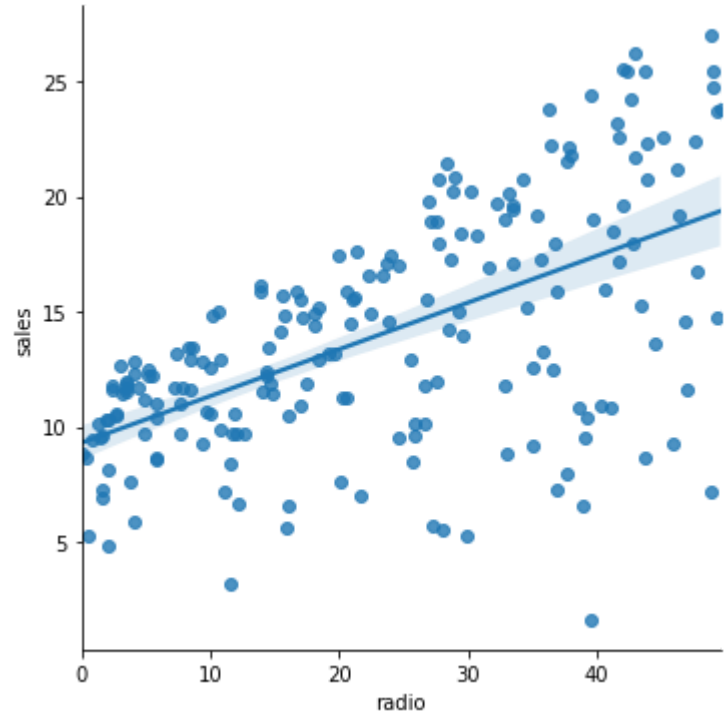
```
In [9]:    # TV vs sales
           sns.lmplot(x="TV", y="sales", data = advertising)
```

Out[9]: <seaborn.axisgrid.FacetGrid at 0x7f7d2f385d30>
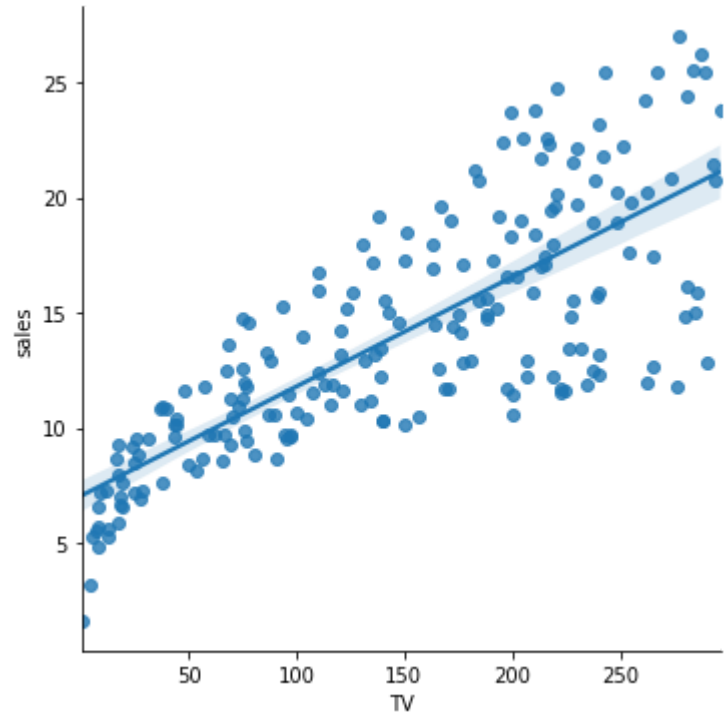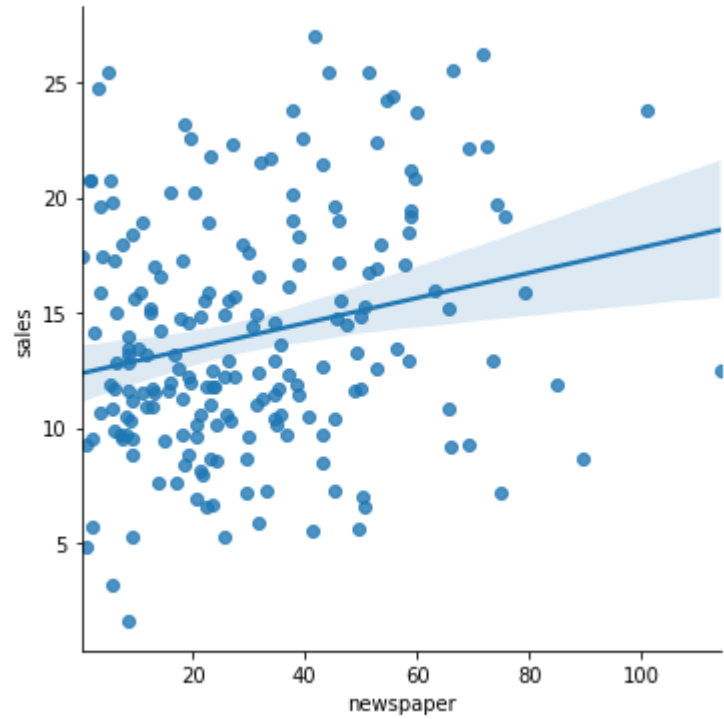


```
In [10]:   # radio vs sales
           sns.lmplot(x="radio", y="sales", data = advertising)
```

Out[10]: <seaborn.axisgrid.FacetGrid at 0x7f7d2f360250>


```
In [11]:   # newspaper vs sales
           sns.lmplot(x="newspaper", y="sales", data = advertising)
```

Out[11]: <seaborn.axisgrid.FacetGrid at 0x7f7d2d2144f0>

The relationship between the features and the predictor have to be linear. Hence, visually inspecting their scatter plots in order to check linearity.