import the libraries

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from sklearn.metrics import accuracy_score
from sklearn.metrics import mean_squared_error
import seaborn as sns
```

load the dataset

In [2]:
```
data = pd.read_csv("/home/tamanna/Downloads/advertising.csv")
```

In [3]:
```
data
```

Out[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
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 × 4 columns

data preprocessing

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

Out[4]:

	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

In [5]:
```
data.shape
```

Out[5]:
```
(200, 4)
```

In [6]:
```
data.columns
```

Out[6]:
```
Index(['TV', 'Radio', 'Newspaper', 'Sales'], dtype='object')
```

In [7]:
```
data.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
```

In [8]:
```
data.describe()
```

Out[8]:

	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

In [9]:
```
data.isnull().sum()
```

Out[9]:
```
TV           0
Radio        0
Newspaper    0
Sales        0
dtype: int64
```

In [10]:
```
x = data[['TV', 'Radio', 'Newspaper']]
```

In [11]:
```
x
```

Out[11]:

	TV	Radio	Newspaper
0	230.1	37.8	69.2
1	44.5	39.3	45.1
2	17.2	45.9	69.3
3	151.5	41.3	58.5
4	180.8	10.8	58.4
...	...	...	...
195	38.2	3.7	13.8
196	94.2	4.9	8.1
197	177.0	9.3	6.4
198	283.6	42.0	66.2
199	232.1	8.6	8.7

200 rows × 3 columns

In [12]:
```
y = data["Sales"]
```

In [13]:
```
y
```

Out[13]:
```
0      22.1
1      10.4
2      12.0
3      16.5
4      17.9
       ...
195     7.6
196    14.0
197    14.8
198    25.5
199    18.4
Name: Sales, Length: 200, dtype: float64
```

splitting the data into training and testing dataset

In [14]:
```
x_train, x_test, y_train, y_test = train_test_split(x,y, random_state = 2, t
```

In [15]:
```
print(x.shape, x_train.shape, x_test.shape)
```

```
(200, 3) (180, 3) (20, 3)
```

In [16]:
```
model = LinearRegression()
```

In [17]:
```
model.fit(x_train, y_train)
```

Out[17]:
```
▾ LinearRegression

LinearRegression()
```

In [18]:
```
x_train_prediction = model.predict(x_train)
```

calculating mse

In [19]:
```
mean_square_error = mean_squared_error(y_train, x_train_prediction)
```

In [20]:
```
mean_square_error
```

Out[20]:
```
2.630161965000496
```

calculating rmse

In [21]:
```
root_mean_square_error = np.sqrt(mean_square_error)
```

In [22]:
```
root_mean_square_error
```

Out[22]:
```
1.6217774092027846
```