

## import libararies

```
In [2]:  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import plotly.express as px  
import seaborn as sns  
from sklearn.model_selection import train_test_split  
from sklearn.linear_model import LinearRegression  
from sklearn.metrics import accuracy_score  
from sklearn.metrics import mean_absolute_error,  
mean_squared_error,r2_score  
from sklearn.feature_selection import SelectKBest, f_regression  
from sklearn.feature_selection import VarianceThreshold
```

## Load The Dataset

```
In [2]:  
df = pd.read_csv("Advertising.csv")  
  
In [3]:  
df.head()
```

	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

```
In [4]:  
df.shape  
  
Out[4]:  
(200, 4)
```

## relation between feature and response

```
In [5]:  
fig = px.scatter_3d(df, x='TV', y='radio', z='newspaper', color='sales',  
size='sales')  
fig.show()
```

## Multiple linear regression -estimating cooefficients

```
In [6]:  
x = df.iloc[:,0:3]  
y = df.iloc[:, -1]  
  
In [7]:
```

```

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,
random_state=42)                                     In [8]:  

x_train.shape                                         Out[8]:  

(160, 3)                                              In [9]:  

x_test.shape                                         Out[9]:  

(40, 3)                                               In [10]:  

lr = LinearRegression()                            In [11]:  

lr.fit(x_train,y_train)                           Out[11]:  

LinearRegression  

LinearRegression()  

.....
```

---

```

y_pred = lr.predict(x_test)                         In [12]:  

lr.coef_ ## (b1,b2,b3)                           In [13]:  

array([0.04472952, 0.18919505, 0.00276111])    Out[13]:  

lr.intercept_ ## (b0)                             In [14]:  

2.979067338122631                               Out[14]:
```

## Model Evaluation Using Train Test Split

```

print("MAE", mean_absolute_error(y_test, y_pred))
print("MSE", mean_squared_error(y_test, y_pred))
print("R2", r2_score(y_test, y_pred))
MAE 1.4607567168117597
MSE 3.1740973539761015
R2 0.8994380241009121
```

In [15]:

## feature selection

```

sel = VarianceThreshold(threshold=0.05)           In [23]:  

sel.fit(x_train)                                 In [24]:  

.....
```

---

```

VarianceThreshold                                         Out[24]:  

VarianceThreshold(threshold=0.05)
```

In [25]:

```

sum(sel.get_support())
```

3

```
columns = x_train.columns[sel.get_support()]
columns
Index(['TV', 'radio', 'newspaper'], dtype='object')
X_train = sel.transform(x_train)
X_test = sel.transform(x_test)

X_train = pd.DataFrame(x_train, columns=columns)
X_test = pd.DataFrame(x_test, columns=columns)

print(X_train.shape)
print(X_test.shape)
(160, 3)
(40, 3)
```

Out[25]:

In [27]:

In [28]:

Out[28]:

In [29]:

In [30]:

In []: