LINEAR REGRESSION PROJECT-Copy1

In [2]:

```
!pip install opendatasets
```

```
Requirement already satisfied: opendatasets in c:\users\om\anaconda3\
anaconda\lib\site-packages (0.1.20)
Requirement already satisfied: kaggle in c:\users\om\anaconda3\
anaconda\lib\site-packages (from opendatasets) (1.5.12)
Requirement already satisfied: click in c:\users\om\anaconda3\
anaconda\lib\site-packages (from opendatasets) (8.0.3)
Requirement already satisfied: tqdm in c:\users\om\anaconda3\anaconda\
lib\site-packages (from opendatasets) (4.62.3)
Requirement already satisfied: colorama in c:\users\om\anaconda3\
anaconda\lib\site-packages (from click->opendatasets) (0.4.4)
Requirement already satisfied: certifi in c:\users\om\anaconda3\
anaconda\lib\site-packages (from kaggle->opendatasets) (2021.10.8)
Requirement already satisfied: six>=1.10 in c:\users\om\anaconda3\
anaconda\lib\site-packages (from kaggle->opendatasets) (1.16.0)
Requirement already satisfied: python-dateutil in c:\users\om\
anaconda3\anaconda\lib\site-packages (from kaggle->opendatasets)
(2.8.2)
Requirement already satisfied: requests in c:\users\om\anaconda3\
anaconda\lib\site-packages (from kaggle->opendatasets) (2.26.0)
Requirement already satisfied: urllib3 in c:\users\om\anaconda3\
anaconda\lib\site-packages (from kaggle->opendatasets) (1.26.7)
Requirement already satisfied: python-slugify in c:\users\om\
anaconda3\anaconda\lib\site-packages (from kaggle->opendatasets)
(5.0.2)
Requirement already satisfied: text-unidecode>=1.3 in c:\users\om\
anaconda3\anaconda\lib\site-packages (from python-slugify->kaggle-
>opendatasets) (1.3)
Requirement already satisfied: idna<4,>=2.5 in c:\users\om\anaconda3\
anaconda\lib\site-packages (from requests->kaggle->opendatasets) (3.2)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\
om\anaconda3\anaconda\lib\site-packages (from requests->kaggle-
>opendatasets) (2.0.4)
```

In [3]:

```
import opendatasets as od
```

In [4]:

```
dataset = 'https://www.kaggle.com/pramodkhade/salary-lin-regression'
```

In [5]:

```
od.download(dataset)
```

```
Skipping, found downloaded files in ".\salary-lin-regression" (use
force=True to force download)
```

In [8]:

```
import os
```

In [6]:

```
data_dir = '.\salary-lin-regression'
```

In [9]:

```
os.listdir(data_dir)
```

Out[9]:

```
['.ipynb_checkpoints',
 'salary linear  regression.ipynb',
 'Salary_Data.csv',
 'Untitled.ipynb']
```

In [11]:

```
import pandas as pd
```

In [12]:

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

In [13]:

```
df
```

Out[13]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0

	YearsExperience	Salary
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

In [16]:

```
import matplotlib.pyplot as plt
%matplotlib inline
```

In [24]:

```
plt.scatter(df['YearsExperience'],df['Salary'])
```

Out[24]:

```
<matplotlib.collections.PathCollection at 0x2d471281ee0>
```

In [26]:

```
plt.scatter(df['YearsExperience'],df['Salary'])
plt.xlim([0,12])
plt.ylim([0,13000])
```

Out[26]:

```
(0.0, 13000.0)
```

In [36]:

```
 import pandas as pd
X = df[['YearsExperience']]
y = df['Salary']
```

In [37]:

```
from sklearn.linear_model import LinearRegression
```

In [38]:

```
model = LinearRegression()
model.fit(X,y)
```

Out[38]:

```
LinearRegression()
```

In [40]:

```
model.intercept_
```

Out[40]:

```
25792.20019866871
```

In [42]:

```
model.coef_
```

Out[42]:

```
array([9449.96232146])
```

In [43]:

```
model.predict([[0],[10]])
```

Out[43]:

```
array([ 25792.20019867, 120291.82341322])
```

In [44]:

```
plt.scatter(df['YearsExperience'],df['Salary'])
plt.xlim([0,12])
plt.ylim([0,13000])
plt.plot[[0,10],model.predict([[0],[10]])]
```

```
---------------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
~\AppData\Local\Temp/ipykernel_9512/778412229.py in <module>
      2 plt.xlim([0,12])
      3 plt.ylim([0,13000])
----> 4 plt.plot[[0,10],model.predict([[0],[10]])]

TypeError: 'function' object is not subscriptable
```

In [45]:

```
model.score(X,y)
```

Out[45]:

```
0.9569566641435086
```

In [46]:

```
from sklearn.model_selection import train_test_split
```

In [51]:

```
 X_train, X_test, y_train, y_test= train_test_split(X,y,test_size=
0.3)
```

In [52]:

```
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

Out[52]:

```
((21, 1), (9, 1), (21,), (9,))
```

In [53]:

```
model_1 = LinearRegression()
model_1.fit(X_train,y_train)
```

Out[53]:

```
LinearRegression()
```

In [55]:

```
model_1.score(X_test,y_test)
```

Out[55]:

```
0.9509157849555185
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: