

```
In [41]: #importing the libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [42]: #importing the dataset and splitting into dependent and independent variable
df = pd.read_csv('C:/Users/prash/OneDrive/Desktop/Machine+Learning+A-Z+(Codes+and+Datasets)/Machine Learning A-Z (Code
x = df.iloc[:, :-1].values
y = df.iloc[:, -1].values
```

```
In [43]: df.head(4)
```

```
Out[43]:
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99

```
In [44]: #encoding the categorical data
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
ct = ColumnTransformer(transformers = [('encoder', OneHotEncoder(), [3])], remainder = 'passthrough')
x = np.array(ct.fit_transform(x))
```

```
In [45]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0)
```

```
In [46]: from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x_train, y_train)
```

```
Out[46]: LinearRegression()
```

```
In [47]: y_pred = regressor.predict(x_test)
np.set_printoptions(precision = 2)
print(np.concatenate((y_pred.reshape(len(y_pred),1),y_test.reshape(len(y_test),1)),1))

[[103015.2  103282.38]
 [132582.28 144259.4 ]
 [132447.74 146121.95]
 [ 71976.1   77798.83]
 [178537.48 191050.39]
 [116161.24 105008.31]
 [ 67851.69  81229.06]
 [ 98791.73  97483.56]
 [113969.44 110352.25]
 [167921.07 166187.94]]
```

```
In [ ]:
```