

```
In [1]: import numpy as np
import pandas as pd
```

```
In [2]: dataset=pd.read_csv(r"D:\ML_Course\Works_on_python\Multi_Linear_Regression\50_Startups.csv")
```

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

Out[3]:

| | 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 |
| 4 | 142107.34 | 91391.77 | 366168.42 | Florida | 166187.94 |

```
In [4]: dataset["State"].unique()
```

Out[4]: array(['New York', 'California', 'Florida'], dtype=object)

```
In [5]: dataset.isnull().any()
```

Out[5]: R&D Spend False
Administration False
Marketing Spend False
State False
Profit False
dtype: bool

```
In [6]: from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
dataset["State"]=le.fit_transform(dataset["State"])
```

```
In [7]: dataset.head()
```

```
Out[7]:
```

| | R&D Spend | Administration | Marketing Spend | State | Profit |
|---|-----------|----------------|-----------------|-------|-----------|
| 0 | 165349.20 | 136897.80 | 471784.10 | 2 | 192261.83 |
| 1 | 162597.70 | 151377.59 | 443898.53 | 0 | 191792.06 |
| 2 | 153441.51 | 101145.55 | 407934.54 | 1 | 191050.39 |
| 3 | 144372.41 | 118671.85 | 383199.62 | 2 | 182901.99 |
| 4 | 142107.34 | 91391.77 | 366168.42 | 1 | 166187.94 |

```
In [8]: x=dataset.iloc[:,0:4].values  
y=dataset.iloc[:,4:5].values
```

```
In [9]: x.shape
```

```
Out[9]: (50, 4)
```

```
In [10]: y.shape
```

```
Out[10]: (50, 1)
```

```
In [11]: from sklearn.preprocessing import OneHotEncoder  
one=OneHotEncoder()  
z=one.fit_transform(x[:,3:4]).toarray()  
x=np.delete(x,3,axis=1)  
x=np.concatenate((z,x),axis=1)
```

```
In [12]: x.shape
```

```
Out[12]: (50, 6)
```

```
In [13]: 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 [14]: x_test.shape
```

```
Out[14]: (10, 6)
```

```
In [15]: x_train.shape
```

```
Out[15]: (40, 6)
```

```
In [16]: y_train.shape
```

```
Out[16]: (40, 1)
```

```
In [17]: y_test.shape
```

```
Out[17]: (10, 1)
```

```
In [18]: from sklearn.linear_model import LinearRegression  
mlr=LinearRegression()  
mlr.fit(x_train,y_train)
```

```
Out[18]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
In [19]: import pickle #used for ML  
pickle.dump(mlr,open('multilinear.pkl','wb'))
```

```
In [ ]:
```