

▼ CHANDAN KUMAR

ID: GO_STP_13267

```
1 import numpy as np
2 import pandas as pd
```

```
1 startup = pd.read_csv("/content/50_Startups.csv")
2 startup.head()
```

	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

```
1 startup.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   R&D Spend              50 non-null    float64
1   Administration         50 non-null    float64
2   Marketing Spend        50 non-null    float64
3   State                  50 non-null    object  
4   Profit                 50 non-null    float64
dtypes: float64(4), object(1)
memory usage: 2.1+ KB
```

```
1 startup.describe()
```

	R&D Spend	Administration	Marketing Spend	Profit
count	50.000000	50.000000	50.000000	50.000000

```
1 startup.corr()
```

	R&D Spend	Administration	Marketing Spend	Profit
R&D Spend	1.000000	0.241955	0.724248	0.972900
Administration	0.241955	1.000000	-0.032154	0.200717
Marketing Spend	0.724248	-0.032154	1.000000	0.747766
Profit	0.972900	0.200717	0.747766	1.000000

```
1 startup.isnull().sum()
```

```
R&D Spend      0
Administration  0
Marketing Spend  0
State           0
Profit          0
dtype: int64
```

```
1 startup["State"].unique()
```

```
array(['New York', 'California', 'Florida'], dtype=object)
```

```
1 startup["State"] = startup["State"].replace(["New York"], 1)
2 startup["State"] = startup["State"].replace(["California"], 2)
3 startup["State"] = startup["State"].replace(["Florida"], 3)
```

```
1 features = startup[["R&D Spend", "Administration", "Marketing Spend", "State"]]
2 features.head()
3 features.shape
```

```
(50, 4)
```

```
1 target = startup[["Profit"]]
2 target.head()
3 target.shape
```

```
(50, 1)
```

```
1 from sklearn.model_selection import train_test_split
2 xtrain, xtest, ytrain, ytest = train_test_split(features, target, test_size = 0.2, rand
3 print(xtrain.shape)
4 print(xtest.shape)
5 print(ytrain.shape)
6 print(ytest.shape)
```

```
(40, 4)
(10, 4)
(40, 1)
(10, 1)
```

```
1 from sklearn.linear_model import LinearRegression
2 model = LinearRegression()
```

```
1 model.fit(xtrain, ytrain)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
1 ypred = model.predict(xtest)
2 ypred
```

```
array([[103083.22418255],
       [132499.19098469],
       [132504.92013514],
       [ 72044.22079157],
       [178578.88804813],
       [116196.18897916],
       [ 67901.09868183],
       [ 98843.64614762],
       [114025.15059486],
       [167965.96780569]])
```

```
1 from sklearn.metrics import r2_score
2 accuracy = r2_score(ytest, ypred)
3 accuracy*100
```

```
93.48088470484865
```

