

```
In [11]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

```
In [12]: df = pd.read_csv('./Datasets/exp_1_B.csv')
df.head()
```

```
Out[12]:
```

	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 [13]: from sklearn.preprocessing import StandardScaler,OneHotEncoder
from sklearn.linear_model import LinearRegression
from sklearn.compose import ColumnTransformer
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
```

```
In [14]: rnd_mean = df['R&D Spend'].mean()
marketing_mean = df['Marketing Spend'].mean()
df['R&D Spend'] = df['R&D Spend'].replace(0,rnd_mean)
df['Marketing Spend'] = df['Marketing Spend'].replace(0,marketing_mean)
```

```
In [15]: x = df.drop(['Profit'],axis=1)
y = df['Profit']
```

```
In [16]: encoding = ('encoding',OneHotEncoder(sparse_output=False,drop='first'),[3])
scaling = ('scaling',StandardScaler(),[0,1,2])
step1 = ColumnTransformer(transformers=[scaling,encoding],remainder='passthru')
step2 = LinearRegression()
```

```
In [17]: pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])
```

```
In [18]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_st
pipe.fit(x_train,y_train)
y_pred = pipe.predict(x_test)
train_score = pipe.score(x_train,y_train)
test_score = pipe.score(x_test,y_test)
print(f"Train Score: {round(train_score*100,2)}")
print(f"Test Score: {round(test_score*100,2)}")
```

Train Score: 80.06

Test Score: 85.31

```
In [19]: plt.scatter(df['R&D Spend'],df['Profit'],color="red")
plt.scatter(df['Administration'],df['Profit'],color="green")
plt.scatter(df['Marketing Spend'],df['Profit'],color="blue")
plt.show()
```

