

```

from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import
r2_score, mean_absolute_error, mean_squared_error
from sklearn.model_selection import train_test_split

df=pd.read_csv(r"C:\my pythonfiles\50_Startups.csv")
df.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

```
df.isnull().sum()
```

```

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

```

```

df.drop_duplicates(inplace=True)
df.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

```

state_encode
0      2
1      0
2      1
3      2
4      1

```

## Data preprocessing

```
state_e=LabelEncoder()  
df['state_encode']=state_e.fit_transform(df['State'])  
df.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

	state_encode
0	2
1	0
2	1
3	2
4	1

```
X=df[['R&D Spend','Administration','Marketing Spend','state_encode']]  
Y=df[['Profit']]
```

```
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=42)
```

```
profit_model=LinearRegression()  
profit_model.fit(X_train,Y_train)
```

```
LinearRegression()
```

```
rd=float(input("Enter the gross domestic spending on research and  
development: "))  
admn=float(input("Enter your administrative expenses: "))  
exp=float(input("Enter your organization's total expenditure: "))  
st=input("Enter your state name: ")
```

```
Enter the gross domestic spending on research and development:  
34365689  
Enter your administrative expenses: 3445668  
Enter your organization's total expenditure: 1234677  
Enter your state name: New York
```

```
state_enc=state_e.transform([st])[0]  
print(state_enc)
```

2

```
result=profit_model.predict([[rd,admn,exp,state_enc]])  
print(result)
```

```
[[27481604.71519688]]
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py:439:  
UserWarning: X does not have valid feature names, but LinearRegression  
was fitted with feature names  
warnings.warn(  

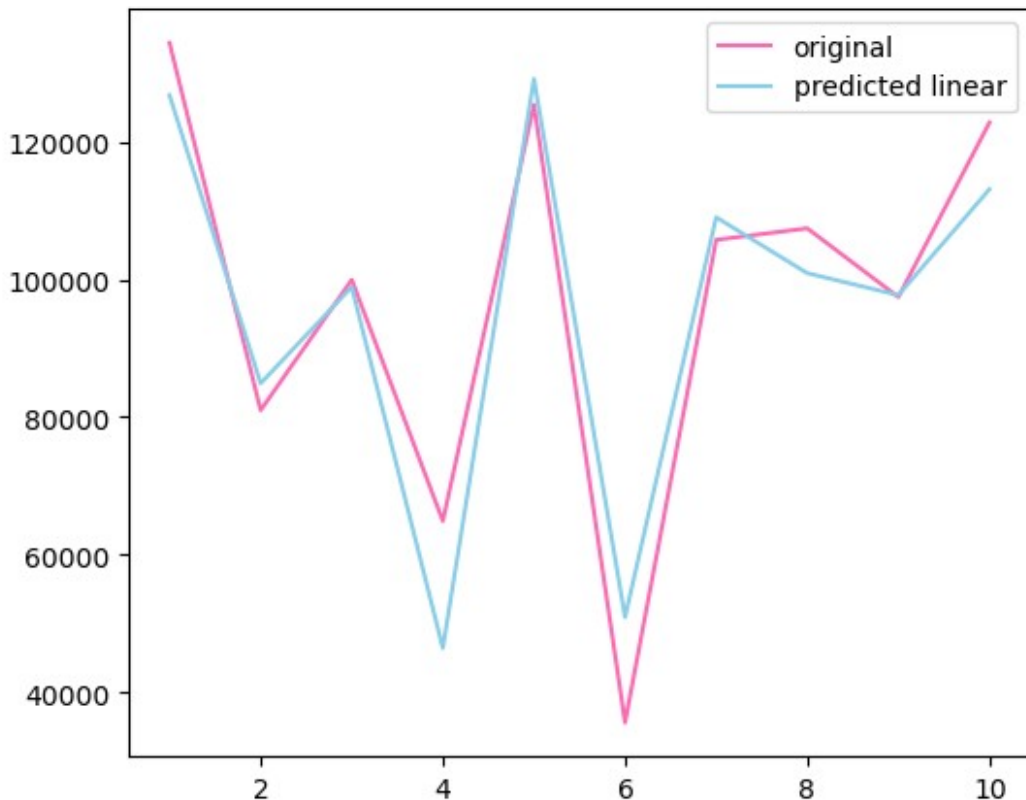
```

```
model_predictions=profit_model.predict(X_test)
```

```
len(Y_test)
```

10

```
import matplotlib.pyplot as plt  
plt.figure(figsize=(6,5))  
plt.plot(np.arange(1,11),Y_test,label="original",color='Hotpink')  
plt.plot(np.arange(1,11),model_predictions,label="predicted  
linear",color='Skyblue')  
plt.legend()  
plt.show()
```



```
r2score=r2_score(Y_test,model_predictions)
print(r2score)
if(r2score>0.5):
    print("Model is good fit")
else:
    print("Model is not good fit")

0.9000614254946402
Model is good fit

mse=mean_squared_error(Y_test,model_predictions)
print(mse)

80929465.49097784

mae=mean_absolute_error(Y_test,model_predictions)
print(mae)

6979.17574672139
```