Task- Predicting a Startups Profit/Success Rate using Multiple Linear Regression in Python

Here 50 startups dataset containing 5 columns like “R&D Spend”, “Administration”, “Marketing Spend”, “State”, “Profit”.

In this dataset first 3 columns provides you spending on Research , Administration and Marketing respectively. State indicates startup based on that state. Profit indicates how much profits earned by a startup.

Clearly, we can understand that it is a multiple linear regression problem, as the independent variables are more than one.

Prepare a prediction model for profit of 50_Startups data in Python

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd


dataset = pd.read_csv('/content/50_Startups.csv')


X = dataset.iloc[:, :-1]
y = dataset.iloc[:, 4]


#Convert the column into categorical columns
states=pd.get_dummies(X['State'],drop_first=True)


# Drop the state coulmn
X=X.drop('State',axis=1)


# concat the dummy variables
X=pd.concat([X,states],axis=1)
```

```
# Splitting the dataset into the Training set and Test set
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)


# Fitting Multiple Linear Regression to the Training set
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

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

```
# Predicting the Test set results
y_pred = regressor.predict(X_test)
from sklearn.metrics import r2_score
score=r2_score(y_test,y_pred)
score
```

```
0.9347068473282423
```