```
#importing the required files
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#read the given dataset and check whether the dataset contains any missing value or not
data = pd.read_csv("50_Startups.csv")
print(data.head())
print(data.describe())

#By using seaborn function look at the correlation between the features
sns.heatmap(data.corr(), annot=True)
plt.show()

#since the task is to predict the profit regression algorithm.so we are using linear regression algorithm to train the profit prediction model
x = data[["R&D Spend", "Administration", "Marketing Spend"]]
y = data["Profit"]
x = x.to_numpy()
y = y.to_numpy()
y = y.reshape(-1, 1)
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2, random_state=42)

#now train the regression model on the given data and look at the predicted profit
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(xtrain, ytrain)
ypred = model.predict(xtest)
data = pd.DataFrame(data={"Predicted Profit": ypred.flatten()})
```

```
print(data.head())
```