In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
dataset = pd.read_csv(r'C:\Users\hp\OneDrive\Documents\Desktop\Investment.
```

In [3]:
```
x = dataset.iloc[:, :-1]

y = dataset.iloc[:,4]
```

In [4]:
```
x = pd.get_dummies(x)
```

In [5]:
```
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,
```

In [6]:
```
from sklearn.linear_model import LinearRegression
```

In [24]:
```
regressor = LinearRegression()
```

In [25]:
```
regressor.fit(x_train, y_train)
```

Out[25]:
```
LinearRegression()
```

In [9]:
```
y_pred = regressor.predict(x_test)
```

In [10]:
```
slope = regressor.coef_
slope
```

Out[10]:
```
array([ 7.73467193e-01,  3.28845975e-02,  3.66100259e-02,  8.66383692e+0
1,
       -8.72645791e+02,  7.86007422e+02])
```

In [11]:
```
cons = regressor.intercept_
cons
```

Out[11]: 42467.52924855311

In [12]:
```
bias = regressor.score(x_train, y_train)
bias
```

Out[12]: 0.9501847627493607

In [13]:
```
variance = regressor.score(x_test, y_test)
variance
```

Out[13]: 0.9347068473282424

We built the model so far

In [14]:
```
import statsmodels.formula.api as sm
```

In [15]:
```
x = np.append(arr = np.ones((50,1)).astype(int), values = x, axis = 1)
```

In [16]:
```
import statsmodels.api as sm

x_opt = x[:,[0,1,2,3,4,5]]
```

OrdinaryLeast sequares

In [18]:
```
regressor_OLS = sm.OLS(endog=y, exog=x_opt).fit()

regressor_OLS.summary()
```

Out[18]:

OLS Regression Results

Dep. Variable:	Profit	**R-squared:**	0.951
Model:	OLS	**Adj. R-squared:**	0.945
Method:	Least Squares	**F-statistic:**	169.9
Date:	Wed, 16 Aug 2023	**Prob (F-statistic):**	1.34e-27
Time:	12:35:27	**Log-Likelihood:**	-525.38
No. Observations:	50	**AIC:**	1063.
Df Residuals:	44	**BIC:**	1074.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.008e+04	6952.587	7.204	0.000	3.61e+04	6.41e+04
x1	0.8060	0.046	17.369	0.000	0.712	0.900
x2	-0.0270	0.052	-0.517	0.608	-0.132	0.078
x3	0.0270	0.017	1.574	0.123	-0.008	0.062
x4	41.8870	3256.039	0.013	0.990	-6520.229	6604.003
x5	240.6758	3338.857	0.072	0.943	-6488.349	6969.701

Omnibus:	14.782	**Durbin-Watson:**	1.283
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	21.266
Skew:	-0.948	**Prob(JB):**	2.41e-05
Kurtosis:	5.572	**Cond. No.**	1.47e+06

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.47e+06. This might indicate that there are strong multicollinearity or other numerical problems.

In [19]:
```
import statsmodels.api as sm

x_opt = x[:,[0,1,2,3]]
```

OrdinaryleastSquares

In [20]:
```
regressor_OLS = sm.OLS(endog=y, exog=x_opt).fit()

regressor_OLS.summary()
```

Out[20]:

OLS Regression Results

Dep. Variable:	Profit	**R-squared:**	0.951
Model:	OLS	**Adj. R-squared:**	0.948
Method:	Least Squares	**F-statistic:**	296.0
Date:	Wed, 16 Aug 2023	**Prob (F-statistic):**	4.53e-30
Time:	12:36:00	**Log-Likelihood:**	-525.39
No. Observations:	50	**AIC:**	1059.
Df Residuals:	46	**BIC:**	1066.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.012e+04	6572.353	7.626	0.000	3.69e+04	6.34e+04
x1	0.8057	0.045	17.846	0.000	0.715	0.897
x2	-0.0268	0.051	-0.526	0.602	-0.130	0.076
x3	0.0272	0.016	1.655	0.105	-0.006	0.060

Omnibus:	14.838	**Durbin-Watson:**	1.282
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	21.442
Skew:	-0.949	**Prob(JB):**	2.21e-05
Kurtosis:	5.586	**Cond. No.**	1.40e+06

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.4e+06. This might indicate that there are
strong multicollinearity or other numerical problems.

In [21]:
```
x_opt = x[:, [0,1]]
regressor_OLS = sm.OLS(endog=y, exog=x_opt).fit()
regressor_OLS.summary()
```

Out[21]:

OLS Regression Results

Dep. Variable:	Profit	**R-squared:**	0.947
Model:	OLS	**Adj. R-squared:**	0.945
Method:	Least Squares	**F-statistic:**	849.8
Date:	Wed, 16 Aug 2023	**Prob (F-statistic):**	3.50e-32
Time:	12:36:23	**Log-Likelihood:**	-527.44
No. Observations:	50	**AIC:**	1059.
Df Residuals:	48	**BIC:**	1063.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.903e+04	2537.897	19.320	0.000	4.39e+04	5.41e+04
x1	0.8543	0.029	29.151	0.000	0.795	0.913

Omnibus:	13.727	**Durbin-Watson:**	1.116
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	18.536
Skew:	-0.911	**Prob(JB):**	9.44e-05
Kurtosis:	5.361	**Cond. No.**	1.65e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.65e+05. This might indicate that there are strong multicollinearity or other numerical problems.

In [ ]:
```
```