### # Task-1: Model building using constant value as 1.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
# Import dataset
data=pd.read csv(r"D:\Full Stack Data Science\4 Sep (Multiple Regression)\MLR\Investment.csv")
data
X=data.iloc[:,:-1]
y=data.iloc[:,4]
X=pd.get_dummies(X).astype(int)
# Splitting the data into train and test
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)
# Applying MLR
from sklearn.linear model import LinearRegression
model=LinearRegression()
model.fit(X_train, y_train)
y pred=model.predict(X test)
import statsmodels.formula.api as sm
X=np.append(arr=np.ones((50,1)).astype(int),values=X,axis=1)
```

import statsmodels.api as sm

 $X_{opt}=X[:,[0,1,2,3,4,5]]$ 

# Ordinary Least Squares

 $ols = sm.OLS(endog = y, exog = X\_opt).fit()$ 

ols.summary()

### OLS Regression Results

Dep. Varia	hla:		Dno	fit	R-squa	======== nod:	.========	0.951
	DIE.				100 ST.	0.945		
Model:				OLS	Adj. F			
Method:		Least Squares				istic:		169.9
Date: Time: No. Observations: Df Residuals: Df Model:		Tue, 05 Sep 2023		2023	Prob (	(F-statisti	ic):	1.34e-27
		11:40:24			Log-Li	kelihood:		-525.38
		44		50	AIC:			1063. 1074.
				44	BIC:			
				5				
Covariance	Type:	1	nonrob	ust				
	coef	std	err		t	P> t	[0.025	0.975]
const	5.008e+04	6952	.617	7	7.204	0.000	3.61e+04	6.41e+04
x1	0.8060	0.	.046	17	7.368	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	42.0063	3256	.058	e	0.013	0.990	-6520.148	6604.161
x5	240.7605	3338	.877	e	0.072	0.943	-6488.304	6969.825
Omnibus:	=======	======	1/1	783	Durhin	 n-Watson:		1.283
Prob(Omnib	ue).	0.001			Jarque		21.267	
Skew:	u3).			948	Prob()	i.	2.41e-05	
					199			
Kurtosis:			٥.	572	Cond.	NO.		1.47e+06

import statsmodels.api as sm

# Removing x4 whose p-value is grater than 0.05 (4th column)

 $X_{opt}=X[:,[0,1,2,3,5]]$ 

# Ordinary Least Squares

ols=sm.OLS(endog=y,exog=X\_opt).fit()

ols.summary()

#### OLS Regression Results

Dep. Variable:		Pro	fit R-	squared:	R-squared:			
Model:			OLS Ad	i. R-squared	:	0.945		
Method: Date:		Least Squar	res F-	statistic:		169.9		
		ue, 05 Sep 20		b (F-statis	tic):	1.34e-27		
Time:		11:40	:24 Los	g-Likelihood	:	-525.38		
No. Observations: Df Residuals:			50 AI			1063.		
			44 BI			1074.		
Df Model:			5	raine		V		
Covarianc	e Type:	nonrob	ust					
	coef	std err	7-	t P> t	[0.025	0.975]		
const	5.008e+04	6952.617	7.20	4 0.000	3.61e+04	6.41e+04		
x1	0.8060	0.046	17.368	0.000	0.712	0.900		
x2	-0.0270	0.052	-0.51	7 0.608	-0.132	0.078		
<b>x</b> 3	0.0270	0.017	1.57	4 0.123	-0.008	0.062		
x4	42.0063	3256.058	0.01	0.990	-6520.148	6604.161		
x5	240.7605	3338.877	0.07	0.943	-6488.304	6969.825		
====== Omnibus:		14.	783 Dui	 rbin-Watson:		1.283		
Prob(Omni	bus):	0.001		rque-Bera (J	21.267			
Skew:	•	-0.9		ob(JB):	2.41e-05			
Kurtosis:				nd. No.	1.47e+06			

# Removing x4 whose p-value is grater than 0.05 (5th column)

 $X_{opt}=X[:,[0,1,2,3]]$ 

# Ordinary Least Squares

ols=sm.OLS(endog=y,exog=X\_opt).fit()

ols.summary()

## OLS Regression Results

Dep. Variable:		Pro	fit	R-squa	0.951		
Model:	LUDIC!		DLS	Adj. R	0.948		
Method: Date: Time: No. Observations: Df Residuals: Df Model:		Least Squa		F-stat	296.0 4.53e-30		
		Tue, 05 Sep 2					
		11:41			kelihood:	6	-525.39
		50 46 3		A STATE OF THE STA			1059.
				BIC:			1066.
				7.7.7.7			
Covariand	ce Type:	nonrob	ust				
	coef	std err		t	P> t	[0.025	0.975]
const	5.012e+04	6572.384	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
<b>x</b> 3	0.0272	0.016	1.	.655	0.105	-0.006	0.060
Omnibus:		14.	===== 839	Durbin	-Watson:		1.282
Prob(Omnibus):		0.001		Jarque	:	21.443	
Skew:	esconden di Soldmil	-0.9	949				2.21e-05
Kurtosis:	:	5.	587	Cond.	507		1.40e+06

# Removing x2 whose p-value is grater than 0.05 (2nd column)

 $X_{opt}=X[:,[0,1,3]]$ 

# Ordinary Least Squares

ols=sm.OLS(endog=y,exog=X\_opt).fit()

ols.summary()

# OLS Regression Results

Dep. Vari	able:	Profit OLS Least Squares			R-squared: Adj. R-squared:			0.950 0.948 450.8
Model:								
Method:					tatisti			
Date:	10.	Tue, 05	Sep 202	23 Pro	b (F-st	atistic)	):	2.16e-31
Time:		Arma amore	11:41:5	53 Log	-Likeli	hood:		-525.54
No. Obser	vations:			50 AIC				1057.
Df Residuals:			4	47 BIC	:			1063.
Df Model:			2	2				
Covarianc	e Type:	nonrobust						
=======	coef	std	err	 t	. P	====== > t	[0.025	0.975]
const	4.698e+04	2689	.941	17.464	. 0	.000	4.16e+04	5.24e+04
x1	0.7966	0.	.041	19.265	0	.000	0.713	0.880
x2	0.0299	0.	.016	1.927	0	.060	-0.001	0.061
Omnibus:	========	======	14.67	78 Dur	bin-Wat	====== son :	========	1.257
Prob(Omnibus): Skew:		0.001		a Jar	Jarque-Bera (JB):			21.162
			-0.93	9 Pro	Prob(JB):			2.54e-05
Kurtosis:			5.57	7E Co.	d. No.			5.32e+05

# Removing x2 whose p-value is grater than 0.05 (3rd column)

X\_opt=X[:,[0,1]]

# Ordinary Least Squares

ols=sm.OLS(endog=y,exog=X\_opt).fit()

ols.summary()

#### OLS Regression Results

Dep. Variable: Model: Method: Date: Time: No. Observations: Df Residuals: Df Model: Covariance Type:		Prof	it R-s	squared:		0.947
		0	LS Ad	j. R-squared:	0.945	
		Least Squar	es F-s	statistic:	849.8	
		Tue, 05 Sep 20	23 Pro	ob (F-statisti	ic):	3.50e-32
		11:42:	02 Log	g-Likelihood:		-527.44
			50 AI	c:		1059.
			48 BIG	C:		1063.
			1			
		nonrobu	st			
	coe	f stderr		t P> t	[0.025	0.975]
const	4.903e+04	4 2537.900	19.320	0.000	4.39e+04	5.41e+04
x1	0.854	0.029	29.15	0.000	0.795	0.913
Omnibus:	:=======	 13.7	27 Dui	 rbin-Watson:		1.116
Prob(Omnib	ous):	0.0	01 Jai	rque-Bera (JB)	):	18.538
Skew:		-0.9	11 Pro	ob(JB):		9.43e-05
Kurtosis:		5.3	2215	nd. No.		1.65e+05

## # Task-2: Calculate Intercept value and replace constant as Intercept value.

# Intercept Value

c=model.intercept\_

c

42467.9

import statsmodels.formula.api as sm

X=np.append(arr= np.full((50, 1), 42467.9),values=X,axis=1)

import statsmodels.api as sm

 $X_{opt}=X[:,[0,1,2,3,4,5]]$ 

# Ordinary Least Squares

ols=sm.OLS(endog=y,exog=X\_opt).fit()

ols.summary()

#### 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:	Tue, 05 Sep 2023	Prob (F-statistic):	1.34e-27
Time:	11:55:08	Log-Likelihood:	-525.38
No. Observations:	50	AIC:	1063.
Df Residuals:	44	BIC:	1074.
Df Model:	5		
Covenience Tunes	nannahust		

Covariance Type: nonrobust

coef	std err	t	P> t	[0.025	0.975]
1.1793	0.164	7.204	0.000	0.849	1.509
0.8060	0.046	17.368	0.000	0.712	0.900
-0.0270	0.052	-0.517	0.608	-0.132	0.078
0.0270	0.017	1.574	0.123	-0.008	0.062
42.0063	3256.058	0.013	0.990	-6520.148	6604.161
240.7605	3338.877	0.072	0.943	-6488.304	6969.825
	1.1793 0.8060 -0.0270 0.0270 42.0063	1.1793 0.164 0.8060 0.046 -0.0270 0.052 0.0270 0.017 42.0063 3256.058	1.1793 0.164 7.204 0.8060 0.046 17.368 -0.0270 0.052 -0.517 0.0270 0.017 1.574 42.0063 3256.058 0.013	1.1793 0.164 7.204 0.000 0.8060 0.046 17.368 0.000 -0.0270 0.052 -0.517 0.608 0.0270 0.017 1.574 0.123 42.0063 3256.058 0.013 0.990	1.1793       0.164       7.204       0.000       0.849         0.8060       0.046       17.368       0.000       0.712         -0.0270       0.052       -0.517       0.608       -0.132         0.0270       0.017       1.574       0.123       -0.008         42.0063       3256.058       0.013       0.990       -6520.148

Omnibus:	14.783	Durbin-Watson:	1.283
Prob(Omnibus):	0.001	Jarque-Bera (JB):	21.267
Skew:	-0.948	Prob(JB):	2.41e-05
Kurtosis:	5.572	Cond. No.	8.45e+05

import statsmodels.api as sm

# Removing x4 whose p-value is grater than 0.05 (4th column)

$$X_{opt}=X[:,[0,1,2,3,5]]$$

# Ordinary Least Squares

 $ols = sm.OLS(endog = y, exog = X\_opt).fit()$ 

ols.summary()

```
# Removing x4 whose p-value is grater than 0.05 (5th column)
X opt=X[:,[0,1,2,3]]
# Ordinary Least Squares
ols=sm.OLS(endog=y,exog=X opt).fit()
ols.summary()
# Removing x2 whose p-value is grater than 0.05 (2nd column)
X_{opt}=X[:,[0,1,3]]
# Ordinary Least Squares
ols=sm.OLS(endog=y,exog=X opt).fit()
ols.summary()
# Removing x2 whose p-value is grater than 0.05 (3rd column)
X opt=X[:,[0,1]]
# Ordinary Least Squares
ols=sm.OLS(endog=y,exog=X opt).fit()
ols.summary()
.....
                     OLS Regression Results
______
Dep. Variable:
                       Profit
                              R-squared:
                                                       0.947
                          OLS Adj. R-squared:
Model:
                                                       0.945
                 Least Squares F-statistic:
Method:
                                                       849.8
           Tue, 05 Sep 2023
Date:
                             Prob (F-statistic):
                                                   3.50e-32
                             Log-Likelihood:
Time:
                     11:58:11
                                                     -527.44
No. Observations:
                              AIC:
                           50
                                                       1059.
Df Residuals:
                           48
                              BIC:
                                                       1063.
Df Model:
                           1
Covariance Type: nonrobust
______
            coef std err
                               t
                                     P>|t| [0.025 0.975]
1.1546 0.060 19.320 0.000 1.034
const
                                                      1.275
                    0.029 29.151
           0.8543
                                     0.000 0.795
______
Omnibus:
                      13.727 Durbin-Watson:
                                                      1.116
Omnibus.
Prob(Omnibus):
                              Jarque-Bera (JB):
                                                      18.538
                        0.001
Skew:
                        -0.911
                              Prob(JB):
                                                    9.43e-05
Kurtosis:
                        5.361 Cond. No.
                                                        4.60
```