Investment Dataset (MLR)

 # 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)
```

Investment Dataset (MLR)

```
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:	11:40:24	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	6952.617	7.204	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	-0.517	0.608	-0.132	0.078
x3	0.0270	0.017	1.574	0.123	-0.008	0.062
x4	42.0063	3256.058	0.013	0.990	-6520.148	6604.161
x5	240.7605	3338.877	0.072	0.943	-6488.304	6969.825

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.	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()
```

Investment Dataset (MLR)

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:	11:40:24	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	6952.617	7.204	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	-0.517	0.608	-0.132	0.078
x3	0.0270	0.017	1.574	0.123	-0.008	0.062
x4	42.0063	3256.058	0.013	0.990	-6520.148	6604.161
x5	240.7605	3338.877	0.072	0.943	-6488.304	6969.825

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.	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:	Profit	R-squared:	0.951
Model:	OLS	Adj. R-squared:	0.948
Method:	Least Squares	F-statistic:	296.0
Date:	Tue, 05 Sep 2023	Prob (F-statistic):	4.53e-30
Time:	11:41:39	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	6572.384	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
x3	0.0272	0.016	1.655	0.105	-0.006	0.060

Omnibus:	14.839	Durbin-Watson:	1.282
Prob(Omnibus):	0.001	Jarque-Bera (JB):	21.443
Skew:	-0.949	Prob(JB):	2.21e-05
Kurtosis:	5.587	Cond. No.	1.40e+06

Investment Dataset (MLR)

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. Variable:                 Profit   R-squared:                       0.950
Model:                            OLS   Adj. R-squared:                  0.948
Method:                 Least Squares   F-statistic:                     450.8
Date:                Tue, 05 Sep 2023   Prob (F-statistic):           2.16e-31
Time:                        11:41:53   Log-Likelihood:                -525.54
No. Observations:                  50   AIC:                             1057.
Df Residuals:                      47   BIC:                             1063.
Df Model:                           2
Covariance 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:                       14.678   Durbin-Watson:                   1.257
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               21.162
Skew:                          -0.939   Prob(JB):                     2.54e-05
Kurtosis:                       5.575   Cond. 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:                 Profit   R-squared:                       0.947
Model:                            OLS   Adj. R-squared:                  0.945
Method:                 Least Squares   F-statistic:                     849.8
Date:                Tue, 05 Sep 2023   Prob (F-statistic):           3.50e-32
Time:                        11:42:02   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   2537.900     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):               18.538
Skew:                          -0.911   Prob(JB):                     9.43e-05
Kurtosis:                       5.361   Cond. 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		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.1793	0.164	7.204	0.000	0.849	1.509
x1	0.8060	0.046	17.368	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	0.013	0.990	-6520.148	6604.161
x5	240.7605	3338.877	0.072	0.943	-6488.304	6969.825

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()

Investment Dataset (MLR)

```
# 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
Model:                            OLS   Adj. R-squared:                  0.945
Method:                 Least Squares   F-statistic:                     849.8
Date:                Tue, 05 Sep 2023   Prob (F-statistic):           3.50e-32
Time:                        11:58:11   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          1.1546      0.060     19.320      0.000       1.034       1.275
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):               18.538
Skew:                          -0.911   Prob(JB):                     9.43e-05
Kurtosis:                       5.361   Cond. No.                         4.60
```