```
import pandas as pd

dataset = pd.read_csv("50_Startups.csv")
#dataset = pd.read_csv("50_Startups_withYear.csv")

dataset
```

```
    R&D Spend  Administration  Marketing Spend       State     Profit
0   165349.20       136897.80        471784.10    New York  192261.83
1   162597.70       151377.59        443898.53  California  191792.06
2   153441.51       101145.55        407934.54     Florida  191050.39
3   144372.41       118671.85        383199.62    New York  182901.99
4   142107.34        91391.77        366168.42     Florida  166187.94
5   131876.90        99814.71        362861.36    New York  156991.12
6   134615.46       147198.87        127716.82  California  156122.51
7   130298.13       145530.06        323876.68     Florida  155752.60
8   120542.52       148718.95        311613.29    New York  152211.77
9   123334.88       108679.17        304981.62  California  149759.96
10  101913.08       110594.11        229160.95     Florida  146121.95
11  100671.96        91790.61        249744.55  California  144259.40
12   93863.75       127320.38        249839.44     Florida  141585.52
13   91992.39       135495.07        252664.93  California  134307.35
14  119943.24       156547.42        256512.92     Florida  132602.65
15  114523.61       122616.84        261776.23    New York  129917.04
16   78013.11       121597.55        264346.06  California  126992.93
17   94657.16       145077.58        282574.31    New York  125370.37
18   91749.16       114175.79        294919.57     Florida  124266.90
19   86419.70       153514.11             0.00    New York  122776.86
20   76253.86       113867.30        298664.47  California  118474.03
21   78389.47       153773.43        299737.29    New York  111313.02
22   73994.56       122782.75        303319.26     Florida  110352.25
23   67532.53       105751.03        304768.73     Florida  108733.99
24   77044.01        99281.34        140574.81    New York  108552.04
25   64664.71       139553.16        137962.62  California  107404.34
26   75328.87       144135.98        134050.07     Florida  105733.54
27   72107.60       127864.55        353183.81    New York  105008.31
28   66051.52       182645.56        118148.20     Florida  103282.38
29   65605.48       153032.06        107138.38    New York  101004.64
30   61994.48       115641.28         91131.24     Florida   99937.59
31   61136.38       152701.92         88218.23    New York   97483.56
32   63408.86       129219.61         46085.25  California   97427.84
33   55493.95       103057.49        214634.81     Florida   96778.92
34   46426.07       157693.92        210797.67  California   96712.80
35   46014.02        85047.44        205517.64    New York   96479.51
36   28663.76       127056.21        201126.82     Florida   90708.19
37   44069.95        51283.14        197029.42  California   89949.14
38   20229.59        65947.93        185265.10    New York   81229.06
39   38558.51        82982.09        174999.30  California   81005.76
40   28754.33       118546.05        172795.67  California   78239.91
41   27892.92        84710.77        164470.71     Florida   77798.83
```

```
42   23640.93         96189.63        148001.11  California   71498.49
43   15505.73        127382.30         35534.17    New York   69758.98
44   22177.74        154806.14         28334.72  California   65200.33
45    1000.23        124153.04          1903.93    New York   64926.08
46    1315.46        115816.21        297114.46     Florida   49490.75
47       0.00        135426.92             0.00  California   42559.73
48     542.05         51743.15             0.00    New York   35673.41
49       0.00        116983.80         45173.06  California   14681.40
```

```
datase=pd.get_dummies(dataset,drop_first=True)
```

```
datase
```

```
    R&D Spend  Administration  Marketing Spend     Profit
State_Florida  \
0   165349.20       136897.80        471784.10  192261.83
False
1   162597.70       151377.59        443898.53  191792.06
False
2   153441.51       101145.55        407934.54  191050.39
True
3   144372.41       118671.85        383199.62  182901.99
False
4   142107.34        91391.77        366168.42  166187.94
True
5   131876.90        99814.71        362861.36  156991.12
False
6   134615.46       147198.87        127716.82  156122.51
False
7   130298.13       145530.06        323876.68  155752.60
True
8   120542.52       148718.95        311613.29  152211.77
False
9   123334.88       108679.17        304981.62  149759.96
False
10  101913.08       110594.11        229160.95  146121.95
True
11  100671.96        91790.61        249744.55  144259.40
False
12   93863.75       127320.38        249839.44  141585.52
True
13   91992.39       135495.07        252664.93  134307.35
False
14  119943.24       156547.42        256512.92  132602.65
True
15  114523.61       122616.84        261776.23  129917.04
False
16   78013.11       121597.55        264346.06  126992.93
False
17   94657.16       145077.58        282574.31  125370.37
```

```
False
18   91749.16       114175.79        294919.57  124266.90
True
19   86419.70       153514.11             0.00  122776.86
False
20   76253.86       113867.30        298664.47  118474.03
False
21   78389.47       153773.43        299737.29  111313.02
False
22   73994.56       122782.75        303319.26  110352.25
True
23   67532.53       105751.03        304768.73  108733.99
True
24   77044.01        99281.34        140574.81  108552.04
False
25   64664.71       139553.16        137962.62  107404.34
False
26   75328.87       144135.98        134050.07  105733.54
True
27   72107.60       127864.55        353183.81  105008.31
False
28   66051.52       182645.56        118148.20  103282.38
True
29   65605.48       153032.06        107138.38  101004.64
False
30   61994.48       115641.28         91131.24   99937.59
True
31   61136.38       152701.92         88218.23   97483.56
False
32   63408.86       129219.61         46085.25   97427.84
False
33   55493.95       103057.49        214634.81   96778.92
True
34   46426.07       157693.92        210797.67   96712.80
False
35   46014.02        85047.44        205517.64   96479.51
False
36   28663.76       127056.21        201126.82   90708.19
True
37   44069.95        51283.14        197029.42   89949.14
False
38   20229.59        65947.93        185265.10   81229.06
False
39   38558.51        82982.09        174999.30   81005.76
False
40   28754.33       118546.05        172795.67   78239.91
False
41   27892.92        84710.77        164470.71   77798.83
True
```

```
42   23640.93         96189.63         148001.11   71498.49
False
43   15505.73        127382.30          35534.17   69758.98
False
44   22177.74        154806.14          28334.72   65200.33
False
45    1000.23        124153.04           1903.93   64926.08
False
46    1315.46        115816.21         297114.46   49490.75
True
47       0.00        135426.92              0.00   42559.73
False
48     542.05         51743.15              0.00   35673.41
False
49       0.00        116983.80          45173.06   14681.40
False

    State_New York
0             True
1            False
2            False
3             True
4            False
5             True
6            False
7            False
8             True
9            False
10           False
11           False
12           False
13           False
14           False
15            True
16           False
17            True
18           False
19            True
20           False
21            True
22           False
23           False
24            True
25           False
26           False
27            True
28           False
29            True
30           False
```

```
31             True
32            False
33            False
34            False
35             True
36            False
37            False
38             True
39            False
40            False
41            False
42            False
43             True
44            False
45             True
46            False
47            False
48             True
49            False
```

```
dataset=datase.replace({True: 1,False: 0})
```

```
dataset
```

```
    R&D Spend  Administration  Marketing Spend     Profit
State_Florida  \
0   165349.20       136897.80        471784.10  192261.83
0
1   162597.70       151377.59        443898.53  191792.06
0
2   153441.51       101145.55        407934.54  191050.39
1
3   144372.41       118671.85        383199.62  182901.99
0
4   142107.34        91391.77        366168.42  166187.94
1
5   131876.90        99814.71        362861.36  156991.12
0
6   134615.46       147198.87        127716.82  156122.51
0
7   130298.13       145530.06        323876.68  155752.60
1
8   120542.52       148718.95        311613.29  152211.77
0
9   123334.88       108679.17        304981.62  149759.96
0
10  101913.08       110594.11        229160.95  146121.95
1
11  100671.96        91790.61        249744.55  144259.40
0
```

```
12   93863.75       127320.38        249839.44  141585.52
1
13   91992.39       135495.07        252664.93  134307.35
0
14  119943.24       156547.42        256512.92  132602.65
1
15  114523.61       122616.84        261776.23  129917.04
0
16   78013.11       121597.55        264346.06  126992.93
0
17   94657.16       145077.58        282574.31  125370.37
0
18   91749.16       114175.79        294919.57  124266.90
1
19   86419.70       153514.11             0.00  122776.86
0
20   76253.86       113867.30        298664.47  118474.03
0
21   78389.47       153773.43        299737.29  111313.02
0
22   73994.56       122782.75        303319.26  110352.25
1
23   67532.53       105751.03        304768.73  108733.99
1
24   77044.01        99281.34        140574.81  108552.04
0
25   64664.71       139553.16        137962.62  107404.34
0
26   75328.87       144135.98        134050.07  105733.54
1
27   72107.60       127864.55        353183.81  105008.31
0
28   66051.52       182645.56        118148.20  103282.38
1
29   65605.48       153032.06        107138.38  101004.64
0
30   61994.48       115641.28         91131.24   99937.59
1
31   61136.38       152701.92         88218.23   97483.56
0
32   63408.86       129219.61         46085.25   97427.84
0
33   55493.95       103057.49        214634.81   96778.92
1
34   46426.07       157693.92        210797.67   96712.80
0
35   46014.02        85047.44        205517.64   96479.51
0
36   28663.76       127056.21        201126.82   90708.19
```

```
1
37   44069.95         51283.14        197029.42   89949.14
0
38   20229.59         65947.93        185265.10   81229.06
0
39   38558.51         82982.09        174999.30   81005.76
0
40   28754.33        118546.05        172795.67   78239.91
0
41   27892.92         84710.77        164470.71   77798.83
1
42   23640.93         96189.63        148001.11   71498.49
0
43   15505.73        127382.30         35534.17   69758.98
0
44   22177.74        154806.14         28334.72   65200.33
0
45    1000.23        124153.04          1903.93   64926.08
0
46    1315.46        115816.21        297114.46   49490.75
1
47       0.00        135426.92             0.00   42559.73
0
48     542.05         51743.15             0.00   35673.41
0
49       0.00        116983.80         45173.06   14681.40
0

    State_New York
0                1
1                0
2                0
3                1
4                0
5                1
6                0
7                0
8                1
9                0
10               0
11               0
12               0
13               0
14               0
15               1
16               0
17               1
18               0
19               1
```

```
20               0
21               1
22               0
23               0
24               1
25               0
26               0
27               1
28               0
29               1
30               0
31               1
32               0
33               0
34               0
35               1
36               0
37               0
38               1
39               0
40               0
41               0
42               0
43               1
44               0
45               1
46               0
47               0
48               1
49               0
```

```
dataset.columns
```

```
Index(['R&D Spend', 'Administration', 'Marketing Spend', 'Profit',
       'State_Florida', 'State_New York'],
      dtype='object')
```

```
independent=dataset[['R&D Spend', 'Administration', 'Marketing
Spend','State_Florida', 'State_New York']]
```

```
dependent=dataset[['Profit']]
```

```
#Training code
#Now we need to split the Training and Test data set from parant
dataset. To split, we are using sklearn as class and call function as
model_selection-> train_test_split
from sklearn.model_selection import train_test_split
#input parameters are passed in function train_test_split with
x,y,size for test data(30 percent from x & y) and random state then
assign to variable x_train,x_test,y_train and y_test
```

```
x_train,x_test,y_train,y_test=train_test_split(independent, dependent,
test_size=0.30, random_state=0)

from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
x_train=sc.fit_transform(x_train)
x_test=sc.transform(x_test)

x_train

array([[ 1.17644103,  0.84515251,  0.94354978,  2.        , -
0.76870611],
       [ 0.96420324,  1.27283565,  0.42738817,  2.        , -
0.76870611],
       [-1.47369826,  0.0153175 , -1.52350329, -0.5       ,
1.30088727],
       [-1.48308929, -2.79556363, -1.53809178, -0.5       ,
1.30088727],
       [-0.14952431,  1.13637282, -0.71716495, -0.5       ,
1.30088727],
       [ 0.85312042, -0.04431628,  0.46771725, -0.5       ,
1.30088727],
       [-0.22353674, -0.3151007 , -0.83981652,  2.        , -
0.76870611],
       [-0.19454707,  0.21199679, -1.18497259, -0.5       , -
0.76870611],
       [ 0.10478723, -0.08388412,  0.48740807, -0.5       , -
0.76870611],
       [-1.0096458 , -1.07019473, -0.4040623 , -0.5       , -
0.76870611],
       [ 0.06872897, -0.38396487,  0.75036616, -0.5       , -
0.76870611],
       [-1.17638797,  0.14067421, -1.26581817, -0.5       ,
1.30088727],
       [ 0.97648631,  0.9689421 ,  0.84958395, -0.5       ,
1.30088727],
       [ 0.39131191,  0.45560401,  0.3979037 , -0.5       , -
0.76870611],
       [-0.16880669,  0.6131351 , -0.48098026, -0.5       , -
0.76870611],
       [ 1.2088001 , -0.92947267,  1.24226224, -0.5       ,
1.30088727],
       [ 0.44593006,  0.82758768,  0.62707846, -0.5       ,
1.30088727],
       [-0.90483959, -0.20234037, -0.21407884, -0.5       , -
0.76870611],
       [-1.49419935, -0.26298539, -1.19196207, -0.5       , -
0.76870611],
       [ 1.83846539,  1.07214791,  1.86319367, -0.5       , -
0.76870611],
       [ 0.42966802,  0.13827054,  0.37625394,  2.        , -
0.76870611],
       [-0.59092478, -2.81342077, -0.02839248, -0.5       , -
0.76870611],
       [ 0.08492419, -0.95017758, -0.46096486, -0.5       ,
1.30088727],
       [ 1.26493068,  0.90993408, -0.55948669, -0.5       , -
0.76870611],
       [-0.11002675, -0.69903054,  0.79713886,  2.        , -
0.76870611],
       [-0.90669595,  0.12801572,  0.00300304,  2.        , -
0.76870611],
       [ 0.11250125,  1.16515207,  0.75858643, -0.5       ,
1.30088727],
       [ 0.27709192,  1.15508553, -1.53809178, -0.5       ,
1.30088727],
       [ 1.03371959, -0.585363  ,  0.79877008, -0.5       , -
0.76870611],
       [-0.70388936, -1.58289852, -0.19719396, -0.5       , -
0.76870611],
       [-1.46723718, -0.30831009,  0.73848951,  2.        , -
0.76870611],
       [ 1.46491286, -0.19745694,  1.39810017, -0.5       ,
1.30088727],
       [ 1.89486118,  0.51005662,  2.07686138, -0.5       ,
1.30088727],
       [-1.49419935,  0.4529585 , -1.53809178, -0.5       , -
0.76870611],
       [-1.0396359 ,  1.20524087, -1.32098255, -0.5       , -
0.76870611]])
```

```
#Model Creation by using Support Vector regression algorithm, here
importing SVR algorithm from sklearn.svm
from sklearn.svm import SVR
#regressor=SVR(kernel="linear")#linear
#regressor=SVR(kernel="linear",C=1000, gamma="auto")#linear
#regressor=SVR(kernel="rbf")#non linear
#regressor=SVR(kernel="rbf",C=0.01)#non linear
#regressor=SVR(kernel="rbf",C=10)#non linear
#regressor=SVR(kernel="rbf",C=100)#non linear
regressor=SVR(kernel="rbf",C=1000, gamma=0.1, epsilon=0.1)#non linear
#regressor=SVR(kernel="poly",C=100, gamma="auto", degree=3,
epsilon=0.1, coef0=1)#non linear
#regressor=SVR(kernel="sigmoid")#non linear
regressor.fit(x_train,y_train)
```

```
C:\Users\Maheshwaran\anaconda3\Lib\site-packages\sklearn\utils\
validation.py:1184: DataConversionWarning: A column-vector y was
passed when a 1d array was expected. Please change the shape of y to
```

```
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

```
SVR(C=1000, gamma=0.1)
```

```
#from where the orgin start
regressor.intercept_
```

```
array([108527.21813668])
```

```
#how many support vector / data points taken
regressor.n_support_
```

```
array([35])
```

```
regressor.support_
```

```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15,
16,
       17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
33,
       34])
```

```
y_pred=regressor.predict(x_test)
```

```
#import r2_score from sklean.metrics functions
from sklearn.metrics import r2_score
#compare actual test (y_test) vs predicted value on x_test which is
assign in y_pred
r_score=r2_score(y_test,y_pred)
```

```
r_score
```

```
0.029274514540696406
```