

Train a machine learning model so that it's capable of predicting the profit a company after their expense.

```
In [1]: import pandas as pd
```

```
In [2]: data=pd.read_csv("50_Startups (1).csv")
data
```

|    | 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  |

```
In [3]: #checking null values
data.isnull().sum()
```

|                 |       |
|-----------------|-------|
| R&D Spend       | 0     |
| Administration  | 0     |
| Marketing Spend | 0     |
| State           | 0     |
| Profit          | 0     |
| dtype:          | int64 |

```
In [4]: #sorting data by statewide in ascending order to better understand the data
sorted_data=data.sort_values(by="State",ascending=True)
sorted_data
```

|    | R&D Spend | Administration | Marketing Spend | State      | Profit    |
|----|-----------|----------------|-----------------|------------|-----------|
| 49 | 0.00      | 116983.80      | 45173.06        | California | 14681.40  |
| 42 | 23640.93  | 96189.63       | 148001.11       | California | 71498.49  |
| 40 | 28754.33  | 118546.05      | 172795.67       | California | 78239.91  |
| 34 | 46426.07  | 157693.92      | 210797.67       | California | 96712.80  |
| 16 | 78013.11  | 121597.55      | 264346.06       | California | 126992.93 |
| 44 | 22177.74  | 154806.14      | 28334.72        | California | 65200.33  |
| 25 | 64664.71  | 139553.16      | 137962.62       | California | 107404.34 |
| 13 | 91992.39  | 135495.07      | 252664.93       | California | 134307.35 |
| 11 | 100671.96 | 91790.61       | 249744.55       | California | 144259.40 |
| 20 | 76253.86  | 113867.30      | 298664.47       | California | 118474.03 |
| 39 | 38558.51  | 82982.09       | 174999.30       | California | 81005.76  |
| 6  | 134615.46 | 147198.87      | 127716.82       | California | 156122.51 |
| 37 | 44069.95  | 51283.14       | 197029.42       | California | 89949.14  |
| 32 | 63408.86  | 129219.61      | 46085.25        | California | 97427.84  |
| 47 | 0.00      | 135426.92      | 0.00            | California | 42559.73  |
| 1  | 162597.70 | 151377.59      | 443898.53       | California | 191792.06 |
| 9  | 123334.88 | 108679.17      | 304981.62       | California | 149759.96 |
| 28 | 66051.52  | 182645.56      | 118148.20       | Florida    | 103282.38 |
| 26 | 75328.87  | 144135.98      | 134050.07       | Florida    | 105733.54 |
| 30 | 61994.48  | 115641.28      | 91131.24        | Florida    | 99937.59  |
| 35 | 55493.95  | 103057.49      | 214634.81       | Florida    | 96778.92  |
| 23 | 67532.53  | 105751.03      | 304768.73       | Florida    | 108733.99 |
| 36 | 28663.76  | 127056.21      | 201126.82       | Florida    | 90708.19  |
| 41 | 27892.92  | 84710.77       | 164470.71       | Florida    | 77798.83  |
| 18 | 91749.16  | 114175.79      | 294919.57       | Florida    | 124266.90 |
| 14 | 119943.24 | 156547.42      | 256512.92       | Florida    | 132602.65 |
| 12 | 93863.75  | 127320.38      | 249839.44       | Florida    | 141585.52 |
| 10 | 101913.08 | 110594.11      | 229160.95       | Florida    | 146121.95 |
| 7  | 130298.13 | 145530.06      | 323876.68       | Florida    | 155752.60 |
| 46 | 1315.46   | 115816.21      | 297114.46       | Florida    | 49490.75  |
| 4  | 142107.34 | 91391.77       | 366168.42       | Florida    | 166187.94 |
| 2  | 153441.51 | 101145.55      | 407934.54       | Florida    | 191050.39 |
| 22 | 73994.56  | 122782.75      | 303319.26       | Florida    | 110352.25 |
| 45 | 1000.23   | 124153.04      | 1903.93         | New York   | 64926.08  |
| 38 | 20229.59  | 65947.93       | 185265.10       | New York   | 81229.06  |
| 43 | 15505.73  | 127382.30      | 35534.17        | New York   | 69758.98  |
| 0  | 165349.20 | 136897.80      | 471784.10       | New York   | 192261.83 |
| 31 | 61136.38  | 152701.92      | 88218.23        | New York   | 97483.56  |
| 29 | 65605.48  | 153032.06      | 107138.38       | New York   | 101004.64 |
| 27 | 72107.60  | 127864.55      | 353183.81       | New York   | 105008.31 |
| 48 | 542.05    | 51743.15       | 0.00            | New York   | 35673.41  |
| 21 | 78389.47  | 153773.43      | 299737.29       | New York   | 111313.02 |
| 19 | 86419.70  | 153514.11      | 0.00            | New York   | 122776.86 |
| 17 | 94657.16  | 145077.58      | 282574.31       | New York   | 125370.37 |
| 15 | 114523.61 | 122616.84      | 261776.23       | New York   | 129917.04 |
| 8  | 120542.52 | 148718.95      | 311613.29       | New York   | 152211.77 |
| 5  | 131876.90 | 99814.71       | 362861.36       | New York   | 156991.12 |
| 3  | 144372.41 | 118671.85      | 383199.62       | New York   | 182901.99 |
| 35 | 46014.02  | 85047.44       | 205517.64       | New York   | 96479.51  |
| 24 | 77044.01  | 99281.34       | 140574.81       | New York   | 108552.04 |

```
In [5]: #Using the function of pandas to get the dummies of data and then concatenating sorted data and dummy data
dummies=pd.get_dummies(sorted_data.State)
dummies
```

|    | California | Florida | New York |
|----|------------|---------|----------|
| 49 | 1          | 0       | 0        |
| 42 | 1          | 0       | 0        |
| 40 | 1          | 0       | 0        |
| 34 | 1          | 0       | 0        |
| 16 | 1          | 0       | 0        |
| 44 | 1          | 0       | 0        |
| 25 | 1          | 0       | 0        |
| 13 | 1          | 0       | 0        |
| 11 | 1          | 0       | 0        |
| 20 | 1          | 0       | 0        |
| 39 | 1          | 0       | 0        |
| 6  | 1          | 0       | 0        |
| 37 | 1          | 0       | 0        |
| 32 | 1          | 0       | 0        |
| 47 | 1          | 0       | 0        |
| 1  | 1          | 0       | 0        |
| 9  | 1          | 0       | 0        |
| 28 | 0          | 1       | 0        |
| 26 | 0          | 1       | 0        |
| 30 | 0          | 1       | 0        |
| 33 | 0          | 1       | 0        |
| 23 | 0          | 1       | 0        |
| 36 | 0          | 1       | 0        |
| 41 | 0          | 1       | 0        |
| 18 | 0          | 1       | 0        |
| 14 | 0          | 1       | 0        |
| 12 | 0          | 1       | 0        |
| 10 | 0          | 1       | 0        |
| 7  | 0          | 1       | 0        |
| 46 | 0          | 1       | 0        |
| 4  | 0          | 1       | 0        |
| 2  | 0          | 1       | 0        |
| 22 | 0          | 1       | 0        |
| 45 | 0          | 0       | 1        |
| 38 | 0          | 0       | 1        |
| 43 | 0          | 0       | 1        |
| 0  | 0          | 0       | 1        |
| 31 | 0          | 0       | 1        |
| 29 | 0          | 0       | 1        |
| 27 | 0          | 0       | 1        |
| 48 | 0          | 0       | 1        |
| 21 | 0          | 0       | 1        |
| 19 | 0          | 0       | 1        |
| 17 | 0          | 0       | 1        |
| 15 | 0          | 0       | 1        |
| 8  | 0          | 0       | 1        |
| 5  | 0          | 0       | 1        |
| 3  | 0          | 0       | 1        |
| 35 | 0          | 0       | 1        |
| 24 | 0          | 0       | 1        |

```
In [6]: sorted_data_dummies=pd.concat([sorted_data,dummies],axis="columns")
sorted_data_dummies
```

|    |           |           |           |            |           |   |   |   |
|----|-----------|-----------|-----------|------------|-----------|---|---|---|
| 42 | 23640.93  | 96189.63  | 148001.11 | California | 71498.49  | 1 | 0 | 0 |
| 40 | 28754.33  | 118546.05 | 172795.67 | California | 78239.91  | 1 | 0 | 0 |
| 34 | 46426.07  | 157693.92 | 210797.67 | California | 96712.80  | 1 | 0 | 0 |
| 16 | 78013.11  | 121597.55 | 264346.06 | California | 126992.93 | 1 | 0 | 0 |
| 44 | 22177.74  | 154806.14 | 28334.72  | California | 65200.33  | 1 | 0 | 0 |
| 25 | 64664.71  | 139553.16 | 137962.62 | California | 107404.34 | 1 | 0 | 0 |
| 13 | 91992.39  | 135495.07 | 252664.93 | California | 134307.35 | 1 | 0 | 0 |
| 11 | 100671.96 | 91790.61  | 249744.55 | California | 144259.40 | 1 | 0 | 0 |
| 20 | 76253.86  | 113867.30 | 298664.47 | California | 118474.03 | 1 | 0 | 0 |
| 39 | 38558.51  | 82982.09  | 174999.30 | California | 81005.76  | 1 | 0 | 0 |
| 6  | 134615.46 | 147198.87 | 127716.82 | California | 156122.51 | 1 | 0 | 0 |
| 37 | 44069.95  | 51283.14  | 197029.42 | California | 89949.14  | 1 | 0 | 0 |
| 32 | 63408.86  | 129219.61 | 46085.25  | California | 97427.84  | 1 | 0 | 0 |
| 47 | 0.00      | 135426.92 | 0.00      | California | 42559.73  | 1 | 0 | 0 |
| 1  | 162597.70 | 151377.59 | 443898.53 | California | 191792.06 | 1 | 0 | 0 |
| 9  | 123334.88 | 108679.17 | 304981.62 | California | 149759.96 | 1 | 0 | 0 |
| 28 | 66051.52  | 182645.56 | 118148.20 | Florida    | 103282.38 | 0 | 1 | 0 |
| 26 | 75328.87  | 144135.98 | 134050.07 | Florida    | 105733.54 | 0 | 1 | 0 |
| 30 | 61994.48  | 115641.28 | 91131.24  | Florida    | 99937.59  | 0 | 1 | 0 |
| 35 | 55493.95  | 103057.49 | 214634.81 | Florida    | 96778.92  | 0 | 1 | 0 |
| 23 | 67532.53  | 105751.03 | 304768.73 | Florida    | 108733.99 | 0 | 1 | 0 |
| 36 | 28663.76  | 127056.21 | 201126.82 | Florida    | 90708.19  | 0 | 1 | 0 |
| 41 | 27892.92  | 84710.77  | 164470.71 | Florida    | 77798.83  | 0 | 1 | 0 |
| 18 | 91749.16  | 114175.79 | 294919.57 | Florida    | 124266.90 | 0 | 1 | 0 |
| 14 | 119943.24 | 156547.42 | 256512.92 | Florida    | 132602.65 | 0 | 1 | 0 |
| 12 | 93863.75  | 127320.38 | 249839.44 | Florida    | 141585.52 | 0 | 1 | 0 |
| 10 | 101913.08 | 110594.11 | 229160.95 | Florida    | 146121.95 | 0 | 1 | 0 |
| 7  | 130298.13 | 145530.06 | 323876.68 | Florida    | 155752.60 | 0 | 1 | 0 |
| 46 | 1315.46   | 115816.21 | 297114.46 | Florida    | 49490.75  | 0 | 1 | 0 |
| 4  | 142107.34 | 91391.77  | 366168.42 | Florida    | 166187.94 | 0 | 1 | 0 |
| 2  | 153441.51 | 101145.55 | 407934.54 | Florida    | 191050.39 | 0 | 1 | 0 |
| 22 | 73994.56  | 122782.75 | 303319.26 | Florida    | 110352.25 | 0 | 1 | 0 |
| 45 | 1000.23   | 124153.04 | 1903.93   | New York   | 64926.08  | 0 | 0 | 1 |
| 38 | 20229.59  | 65947.93  | 185265.10 | New York   | 81229.06  | 0 | 0 | 1 |
| 43 | 15505.73  | 127382.30 | 35534.17  | New York   | 69758.98  | 0 | 0 | 1 |
| 0  | 165349.20 | 136897.80 | 471784.10 | New York   | 192261.83 | 0 | 0 | 1 |
| 31 | 61136.38  | 152701.92 | 88218.23  | New York   | 97483.56  | 0 | 0 | 1 |
| 29 | 65605.48  | 153032.06 | 107138.38 | New York   | 101004.64 | 0 | 0 | 1 |
| 27 | 72107.60  | 127864.55 | 353183.81 | New York   | 105008.21 | 0 | 0 | 1 |
| 48 | 542.05    | 51743.15  | 0.00      | New York   | 35673.41  | 0 | 0 | 1 |
| 21 | 78389.47  | 153773.43 | 299737.29 | New York   | 111313.02 | 0 | 0 | 1 |
| 19 | 86419.70  | 153514.11 | 0.00      | New York   | 122776.86 | 0 | 0 | 1 |
| 17 | 94657.16  | 145077.58 | 282574.31 | New York   | 125370.37 | 0 | 0 | 1 |
| 15 | 114523.61 | 122616.84 | 261776.23 | New York   | 129917.04 | 0 | 0 | 1 |
| 8  | 120542.52 | 148718.95 | 311613.29 | New York   | 152211.77 | 0 | 0 | 1 |
| 5  | 131876.90 | 99814.71  | 362861.36 | New York   | 156991.12 | 0 | 0 | 1 |
| 3  | 144372.41 | 118671.85 | 383199.62 | New York   | 182901.99 | 0 | 0 | 1 |
| 35 | 46014.02  | 85047.44  | 205517.64 | New York   | 96479.51  | 0 | 0 | 1 |
| 24 | 77044.01  | 99281.34  | 140574.81 | New York   | 108552.04 | 0 | 0 | 1 |

[7]:

```
#New dropping the two columns state(not needed now already dummy is present) and New york(drping to use the one hot coder function)
main_data=sorted_data.dummies.drop(["'State'",'New York'],axis='columns')
main_data
```