Basic Model Prediction

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

In [3]:
```
dataset = pd.read_csv("C:\\Users\\mdsha\\Downloads\\archive (8)\\Salary_Data.c
```

In [4]:
```
dataset
```

Out[4]:

	YearsExperience	Age	Salary
0	1.1	21.0	39343
1	1.3	21.5	46205
2	1.5	21.7	37731
3	2.0	22.0	43525
4	2.2	22.2	39891
5	2.9	23.0	56642
6	3.0	23.0	60150
7	3.2	23.3	54445
8	3.2	23.3	64445
9	3.7	23.6	57189
10	3.9	23.9	63218
11	4.0	24.0	55794
12	4.0	24.0	56957
13	4.1	24.0	57081
14	4.5	25.0	61111
15	4.9	25.0	67938
16	5.1	26.0	66029
17	5.3	27.0	83088
18	5.9	28.0	81363
19	6.0	29.0	93940
20	6.8	30.0	91738
21	7.1	30.0	98273
22	7.9	31.0	101302
23	8.2	32.0	113812
24	8.7	33.0	109431
25	9.0	34.0	105582
26	9.5	35.0	116969
27	9.6	36.0	112635
28	10.3	37.0	122391
29	10.5	38.0	121872

In [9]:
```
x = dataset[['YearsExperience','Age']]
```

In [10]:
```
x
```

Out[10]:

	YearsExperience	Age
0	1.1	21.0
1	1.3	21.5
2	1.5	21.7
3	2.0	22.0
4	2.2	22.2
5	2.9	23.0
6	3.0	23.0
7	3.2	23.3
8	3.2	23.3
9	3.7	23.6
10	3.9	23.9
11	4.0	24.0
12	4.0	24.0
13	4.1	24.0
14	4.5	25.0
15	4.9	25.0
16	5.1	26.0
17	5.3	27.0
18	5.9	28.0
19	6.0	29.0
20	6.8	30.0
21	7.1	30.0
22	7.9	31.0
23	8.2	32.0
24	8.7	33.0
25	9.0	34.0
26	9.5	35.0
27	9.6	36.0
28	10.3	37.0
29	10.5	38.0

In [11]:
```
x = dataset[['YearsExperience','Age']].values
```

In [12]:
```
x
```

Out[12]:
```
array([[ 1.1, 21. ],
       [ 1.3, 21.5],
       [ 1.5, 21.7],
       [ 2. , 22. ],
       [ 2.2, 22.2],
       [ 2.9, 23. ],
       [ 3. , 23. ],
       [ 3.2, 23.3],
       [ 3.2, 23.3],
       [ 3.7, 23.6],
       [ 3.9, 23.9],
       [ 4. , 24. ],
       [ 4. , 24. ],
       [ 4.1, 24. ],
       [ 4.5, 25. ],
       [ 4.9, 25. ],
       [ 5.1, 26. ],
       [ 5.3, 27. ],
       [ 5.9, 28. ],
       [ 6. , 29. ],
       [ 6.8, 30. ],
       [ 7.1, 30. ],
       [ 7.9, 31. ],
       [ 8.2, 32. ],
       [ 8.7, 33. ],
       [ 9. , 34. ],
       [ 9.5, 35. ],
       [ 9.6, 36. ],
       [10.3, 37. ],
       [10.5, 38. ]])
```

In [13]:
```
y = dataset[['Salary']]
```

In [14]:
```
y
```

Out[14]:

	Salary
0	39343
1	46205
2	37731
3	43525
4	39891
5	56642
6	60150
7	54445
8	64445
9	57189
10	63218
11	55794
12	56957
13	57081
14	61111
15	67938
16	66029
17	83088
18	81363
19	93940
20	91738
21	98273
22	101302
23	113812
24	109431
25	105582
26	116969
27	112635
28	122391
29	121872

In [18]:
```
dataset.isnull().sum
```

Out[18]:
```
<bound method NDFrame._add_numeric_operations.<locals>.sum of     YearsExper
ience    Age  Salary
0             False  False   False
1             False  False   False
2             False  False   False
3             False  False   False
4             False  False   False
5             False  False   False
6             False  False   False
7             False  False   False
8             False  False   False
9             False  False   False
10            False  False   False
11            False  False   False
12            False  False   False
13            False  False   False
14            False  False   False
15            False  False   False
16            False  False   False
17            False  False   False
18            False  False   False
19            False  False   False
20            False  False   False
21            False  False   False
22            False  False   False
23            False  False   False
24            False  False   False
25            False  False   False
26            False  False   False
27            False  False   False
28            False  False   False
29            False  False   False>
```

In [19]:
```
from sklearn.model_selection import train_test_split
```

In [20]:
```
x_train,x_test,y_train,y_test =train_test_split(x,y,test_size=0.2,random_state
```

In [21]:
```
x_train
```

Out[21]:
```
array([[ 9.6, 36. ],
       [ 4. , 24. ],
       [ 5.3, 27. ],
       [ 7.9, 31. ],
       [ 2.9, 23. ],
       [ 5.1, 26. ],
       [ 3.2, 23.3],
       [ 4.5, 25. ],
       [ 8.2, 32. ],
       [ 6.8, 30. ],
       [ 1.3, 21.5],
       [10.5, 38. ],
       [ 3. , 23. ],
       [ 2.2, 22.2],
       [ 5.9, 28. ],
       [ 6. , 29. ],
       [ 3.7, 23.6],
       [ 3.2, 23.3],
       [ 9. , 34. ],
       [ 2. , 22. ],
       [ 1.1, 21. ],
       [ 7.1, 30. ],
       [ 4.9, 25. ],
       [ 4. , 24. ]])
```

In [22]:
```
x_test
```

Out[22]:
```
array([[ 1.5, 21.7],
       [10.3, 37. ],
       [ 4.1, 24. ],
       [ 3.9, 23.9],
       [ 9.5, 35. ],
       [ 8.7, 33. ]])
```

In [23]:
```
y_train
```

Out[23]:

	Salary
27	112635
11	55794
17	83088
22	101302
5	56642
16	66029
8	64445
14	61111
23	113812
20	91738
1	46205
29	121872
6	60150
4	39891
18	81363
19	93940
9	57189
7	54445
25	105582
3	43525
0	39343
21	98273
15	67938
12	56957

In [24]:
```
y_test
```

Out[24]:

	Salary
2	37731
28	122391
13	57081
10	63218
26	116969
24	109431

In [26]:
```
from sklearn.linear_model import LinearRegression
```

In [27]:
```
lr  = LinearRegression()
```

In [28]:
```
lr.fit(x_train,y_train)
```

Out[28]:
```
LinearRegression()
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [29]:
```
lr.predict(x_test)
```

Out[29]:
```
array([[ 43505.10119893],
       [124587.04346091],
       [ 62780.35130369],
       [ 61459.79568568],
       [115931.97767592],
       [107276.91189093]])
```

In [33]:
```
lr.predict([[ 1.1, 21. ]])
```

Out[33]:
```
array([[39809.97642761]])
```

In [35]:
```
lr.predict([[ 1.5, 21.7]])
```

Out[35]:
```
array([[43505.10119893]])
```

In [38]:
```
lr.predict([[5,30]])
```

Out[38]:
```
array([[80422.40182651]])
```

In [ ]:
```

```