In [18]:

```python
import pandas as pd
import numpy as np

#reading the dataset
df=pd.read_csv("salary.csv")
df.head()
```

Out[18]:

	sx	rk	yr	dg	yd	sl
0	male	full	25	doctorate	35	36350
1	male	full	13	doctorate	22	10000
2	male	full	10	doctorate	23	10001
3	female	full	7	doctorate	11	12345
4	male	full	19	masters	13	23456

In [19]:

```python
print(df.info)
```

```
<bound method DataFrame.info of         sx         rk  yr         dg  yd     sl
0       male       full  25  doctorate  35  36350
1       male       full  13  doctorate  22  10000
2       male       full  10  doctorate  23  10001
3     female       full   7  doctorate  11  12345
4       male       full  19    masters  13  23456
5       male       full  16  doctorate  12  34567
6     female       full   0    masters  14  36443
7       male       full  16  doctorate  15  21212
8       male       full  13    masters  16  12121
9       male       full  13    masters  33  12123
10      male       full  12  doctorate  41  12124
11      male  associate  15  doctorate   7  12125
12      male       full   9  doctorate   8  23234
13      male  associate   9    masters   9  23232
14      male       full   9  doctorate  44  23233
15      male       full   7  doctorate   1  23456
16      male       full  13  doctorate  15  34343
17      male  associate  11    masters  16  34344
18      male  associate  10    masters  17  34345
19      male       full   6    masters  28  34346
20      male  assistant  16    masters  34  34347
21      male  associate   8    masters  56  45454
22      male       full   7  doctorate  44  45456
23    female       full   8  doctorate  33  45455
24      male  associate   9  doctorate  23  56565
25      male       full   5  doctorate  24  56567
26      male  associate  11  doctorate  32  56566
27    female       full   5  doctorate  31  56564
28      male  associate   3    masters  21  56654
29      male  associate   3    masters  66  55555
30    female  assistant  10    masters  77  66666
31      male  associate  11    masters  76  77777
32      male  assistant   9    masters  75  88888
33    female  associate   4    masters  11  22222
34    female  associate   6    masters  22  33333
35      male  associate   1  doctorate  33  23500
36    female  assistant   8  doctorate  44  12300
37      male  assistant   4  doctorate  55  23300
38      male  assistant   4  doctorate   2  45000
39      male  assistant   4  doctorate   3  67000
40      male  assistant   3  doctorate   4  38000
41      male  assistant   3    masters   5  43434
```

```
42    male  associate  0  doctorate   6  54545
43  female  assistant  3  doctorate   8  54544
44    male  assistant  2  doctorate   9  32323
45    male  assistant  2  doctorate  10  32322
46  female  assistant  2  doctorate  21  54549
47  female  assistant  2  doctorate  12  11223
48    male  assistant  1  doctorate  32  11224
49  female  assistant  1  doctorate  23  11225
50  female  assistant  1  doctorate  25  11229
51  female  assistant  0  doctorate  43  35000>
```

In [26]:

```
df.shape
```

Out[26]:

(52, 6)

In [44]:

```
dummies = pd.get_dummies(df)
dummies
```

Out[44]:

	yr	dg	yd	sl	sx_female	sx_male	rk_assistant	rk_associate	rk_full
0	25	0	35	36350	0	1	0	0	1
1	13	0	22	10000	0	1	0	0	1
2	10	0	23	10001	0	1	0	0	1
3	7	0	11	12345	1	0	0	0	1
4	19	1	13	23456	0	1	0	0	1
5	16	0	12	34567	0	1	0	0	1
6	0	1	14	36443	1	0	0	0	1
7	16	0	15	21212	0	1	0	0	1
8	13	1	16	12121	0	1	0	0	1
9	13	1	33	12123	0	1	0	0	1
10	12	0	41	12124	0	1	0	0	1
11	15	0	7	12125	0	1	0	1	0
12	9	0	8	23234	0	1	0	0	1
13	9	1	9	23232	0	1	0	1	0
14	9	0	44	23233	0	1	0	0	1
15	7	0	1	23456	0	1	0	0	1
16	13	0	15	34343	0	1	0	0	1
17	11	1	16	34344	0	1	0	1	0
18	10	1	17	34345	0	1	0	1	0
19	6	1	28	34346	0	1	0	0	1
20	16	1	34	34347	0	1	1	0	0
21	8	1	56	45454	0	1	0	1	0
22	7	0	44	45456	0	1	0	0	1
23	8	0	33	45455	1	0	0	0	1
24	9	0	23	56565	0	1	0	1	0
25	5	0	24	56567	0	1	0	0	1
26	11	0	32	56566	0	1	0	1	0
27	5	0	31	56564	1	0	0	0	1

	ys	dg	yd	sl	sx_female	sx_male	rk_assistant	rk_associate	rk_full
29	3	1	66	55555	0	1	0	1	0
30	10	1	77	66666	1	0	1	0	0
31	11	1	76	77777	0	1	0	1	0
32	9	1	75	88888	0	1	1	0	0
33	4	1	11	22222	1	0	0	1	0
34	6	1	22	33333	1	0	0	1	0
35	1	0	33	23500	0	1	0	1	0
36	8	0	44	12300	1	0	1	0	0
37	4	0	55	23300	0	1	1	0	0
38	4	0	2	45000	0	1	1	0	0
39	4	0	3	67000	0	1	1	0	0
40	3	0	4	38000	0	1	1	0	0
41	3	1	5	43434	0	1	1	0	0
42	0	0	6	54545	0	1	0	1	0
43	3	0	8	54544	1	0	1	0	0
44	2	0	9	32323	0	1	1	0	0
45	2	0	10	32322	0	1	1	0	0
46	2	0	21	54549	1	0	1	0	0
47	2	0	12	11223	1	0	1	0	0
48	1	0	32	11224	0	1	1	0	0
49	1	0	23	11225	1	0	1	0	0
50	1	0	25	11229	1	0	1	0	0
51	0	0	43	35000	1	0	1	0	0

In [45]:

```python
merged = pd.concat([df,dummies],axis='columns')
merged
```

Out[45]:

	sx	rk	yr	dg	yd	sl	yr	dg	yd	sl	sx_female	sx_male	rk_assistant	rk_associate	rk_full
0	male	full	25	0	35	36350	25	0	35	36350	0	1	0	0	1
1	male	full	13	0	22	10000	13	0	22	10000	0	1	0	0	1
2	male	full	10	0	23	10001	10	0	23	10001	0	1	0	0	1
3	female	full	7	0	11	12345	7	0	11	12345	1	0	0	0	1
4	male	full	19	1	13	23456	19	1	13	23456	0	1	0	0	1
5	male	full	16	0	12	34567	16	0	12	34567	0	1	0	0	1
6	female	full	0	1	14	36443	0	1	14	36443	1	0	0	0	1
7	male	full	16	0	15	21212	16	0	15	21212	0	1	0	0	1
8	male	full	13	1	16	12121	13	1	16	12121	0	1	0	0	1
9	male	full	13	1	33	12123	13	1	33	12123	0	1	0	0	1
10	male	full	12	0	41	12124	12	0	41	12124	0	1	0	0	1
11	male	associate	15	0	7	12125	15	0	7	12125	0	1	0	1	0
12	male	full	9	0	8	23234	9	0	8	23234	0	1	0	0	1
13	male	associate	9	1	9	23232	9	1	9	23232	0	1	0	1	0
14	male	full	9	0	44	23233	9	0	44	23233	0	1	0	0	1

	sx	rk	yr	dg	yd	sl	yr	dg	yd	sl	sx_female	sx_male	rk_assistant	rk_associate	rk_full
15	male	full	7	0	1	23456	7	0	1	23456	0	1	0	0	1
16	male	full	13	0	15	34343	13	0	15	34343	0	1	0	0	1
17	male	associate	11	1	16	34344	11	1	16	34344	0	1	0	1	0
18	male	associate	10	1	17	34345	10	1	17	34345	0	1	0	1	0
19	male	full	6	1	28	34346	6	1	28	34346	0	1	0	0	1
20	male	assistant	16	1	34	34347	16	1	34	34347	0	1	1	0	0
21	male	associate	8	1	56	45454	8	1	56	45454	0	1	0	1	0
22	male	full	7	0	44	45456	7	0	44	45456	0	1	0	0	1
23	female	full	8	0	33	45455	8	0	33	45455	1	0	0	0	1
24	male	associate	9	0	23	56565	9	0	23	56565	0	1	0	1	0
25	male	full	5	0	24	56567	5	0	24	56567	0	1	0	0	1
26	male	associate	11	0	32	56566	11	0	32	56566	0	1	0	1	0
27	female	full	5	0	31	56564	5	0	31	56564	1	0	0	0	1
28	male	associate	3	1	21	56654	3	1	21	56654	0	1	0	1	0
29	male	associate	3	1	66	55555	3	1	66	55555	0	1	0	1	0
30	female	assistant	10	1	77	66666	10	1	77	66666	1	0	1	0	0
31	male	associate	11	1	76	77777	11	1	76	77777	0	1	0	1	0
32	male	assistant	9	1	75	88888	9	1	75	88888	0	1	1	0	0
33	female	associate	4	1	11	22222	4	1	11	22222	1	0	0	1	0
34	female	associate	6	1	22	33333	6	1	22	33333	1	0	0	1	0
35	male	associate	1	0	33	23500	1	0	33	23500	0	1	0	1	0
36	female	assistant	8	0	44	12300	8	0	44	12300	1	0	1	0	0
37	male	assistant	4	0	55	23300	4	0	55	23300	0	1	1	0	0
38	male	assistant	4	0	2	45000	4	0	2	45000	0	1	1	0	0
39	male	assistant	4	0	3	67000	4	0	3	67000	0	1	1	0	0
40	male	assistant	3	0	4	38000	3	0	4	38000	0	1	1	0	0
41	male	assistant	3	1	5	43434	3	1	5	43434	0	1	1	0	0
42	male	associate	0	0	6	54545	0	0	6	54545	0	1	0	1	0
43	female	assistant	3	0	8	54544	3	0	8	54544	1	0	1	0	0
44	male	assistant	2	0	9	32323	2	0	9	32323	0	1	1	0	0
45	male	assistant	2	0	10	32322	2	0	10	32322	0	1	1	0	0
46	female	assistant	2	0	21	54549	2	0	21	54549	1	0	1	0	0
47	female	assistant	2	0	12	11223	2	0	12	11223	1	0	1	0	0
48	male	assistant	1	0	32	11224	1	0	32	11224	0	1	1	0	0
49	female	assistant	1	0	23	11225	1	0	23	11225	1	0	1	0	0
50	female	assistant	1	0	25	11229	1	0	25	11229	1	0	1	0	0
51	female	assistant	0	0	43	35000	0	0	43	35000	1	0	1	0	0

In [46]:

```python
final = merged1.drop(['sx','rk'],axis='columns')
final
```

Out[46]:

	yr	dg	yd	sl	female	male
0	25	0	35	36350	0	1
1	13	0	22	10000	0	1

	yr	dg	yd	sl	female	male
2	10	0	23	10001	0	1
3	7	0	11	12345	1	0
4	19	1	13	23456	0	1
5	16	0	12	34567	0	1
6	0	1	14	36443	1	0
7	16	0	15	21212	0	1
8	13	1	16	12121	0	1
9	13	1	33	12123	0	1
10	12	0	41	12124	0	1
11	15	0	7	12125	0	1
12	9	0	8	23234	0	1
13	9	1	9	23232	0	1
14	9	0	44	23233	0	1
15	7	0	1	23456	0	1
16	13	0	15	34343	0	1
17	11	1	16	34344	0	1
18	10	1	17	34345	0	1
19	6	1	28	34346	0	1
20	16	1	34	34347	0	1
21	8	1	56	45454	0	1
22	7	0	44	45456	0	1
23	8	0	33	45455	1	0
24	9	0	23	56565	0	1
25	5	0	24	56567	0	1
26	11	0	32	56566	0	1
27	5	0	31	56564	1	0
28	3	1	21	56654	0	1
29	3	1	66	55555	0	1
30	10	1	77	66666	1	0
31	11	1	76	77777	0	1
32	9	1	75	88888	0	1
33	4	1	11	22222	1	0
34	6	1	22	33333	1	0
35	1	0	33	23500	0	1
36	8	0	44	12300	1	0
37	4	0	55	23300	0	1
38	4	0	2	45000	0	1
39	4	0	3	67000	0	1
40	3	0	4	38000	0	1
41	3	1	5	43434	0	1
42	0	0	6	54545	0	1
43	3	0	8	54544	1	0
44	2	0	9	32323	0	1
45	2	0	10	32322	0	1
46	2	0	21	54549	1	0

	yr	dg	yd	sl	female	male
47	2	0	12	11223	1	0
48	1	0	32	11224	0	1
49	1	0	23	11225	1	0
50	1	0	25	11229	1	0
51	0	0	43	35000	1	0

In [47]:

```python
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

In [49]:

```python
X = final.drop('sl',axis='columns')
X
```

Out[49]:

	yr	dg	yd	female	male
0	25	0	35	0	1
1	13	0	22	0	1
2	10	0	23	0	1
3	7	0	11	1	0
4	19	1	13	0	1
5	16	0	12	0	1
6	0	1	14	1	0
7	16	0	15	0	1
8	13	1	16	0	1
9	13	1	33	0	1
10	12	0	41	0	1
11	15	0	7	0	1
12	9	0	8	0	1
13	9	1	9	0	1
14	9	0	44	0	1
15	7	0	1	0	1
16	13	0	15	0	1
17	11	1	16	0	1
18	10	1	17	0	1
19	6	1	28	0	1
20	16	1	34	0	1
21	8	1	56	0	1
22	7	0	44	0	1
23	8	0	33	1	0
24	9	0	23	0	1
25	5	0	24	0	1
26	11	0	32	0	1
27	5	0	31	1	0
28	3	1	21	0	1
29	3	1	66	0	1

	yr	dg	yd	female	male
30	10	1	77	1	0
31	11	1	76	0	1
32	9	1	75	0	1
33	4	1	11	1	0
34	6	1	22	1	0
35	1	0	33	0	1
36	8	0	44	1	0
37	4	0	55	0	1
38	4	0	2	0	1
39	4	0	3	0	1
40	3	0	4	0	1
41	3	1	5	0	1
42	0	0	6	0	1
43	3	0	8	1	0
44	2	0	9	0	1
45	2	0	10	0	1
46	2	0	21	1	0
47	2	0	12	1	0
48	1	0	32	0	1
49	1	0	23	1	0
50	1	0	25	1	0
51	0	0	43	1	0

In [50]:

```
y = final.sl
y
```

Out[50]:

```
0      36350
1      10000
2      10001
3      12345
4      23456
5      34567
6      36443
7      21212
8      12121
9      12123
10     12124
11     12125
12     23234
13     23232
14     23233
15     23456
16     34343
17     34344
18     34345
19     34346
20     34347
21     45454
22     45456
23     45455
24     56565
25     56567
26     56566
27     56564
28     56654
29     55555
```

```
29      33333
30      66666
31      77777
32      88888
33      22222
34      33333
35      23500
36      12300
37      23300
38      45000
39      67000
40      38000
41      43434
42      54545
43      54544
44      32323
45      32322
46      54549
47      11223
48      11224
49      11225
50      11229
51      35000
Name: sl, dtype: int64
```

In [51]:

```
model.fit(X,y)
```

Out[51]:

```
LinearRegression()
```

In [52]:

```
model.predict([[32,0,28,1,0]])
```

Out[52]:

```
array([8400.69947492])
```

In [53]:

```
model.score(X,y)
```

Out[53]:

```
0.19893917655583604
```

In [55]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

In [62]:

```
dfle = df
dfle.sx = le.fit_transform(dfle.sx)
dfle
```

Out[62]:

	sx	rk	yr	dg	yd	sl
0	1	full	25	0	35	36350
1	1	full	13	0	22	10000
2	1	full	10	0	23	10001
3	0	full	7	0	11	12345
4	1	full	19	1	13	23456
5	1	full	16	0	12	34567

	sx	rk	yr	dg	yd	sl
6						3644
7	1	full	16	0	15	21212
8	1	full	13	1	16	12121
9	1	full	13	1	33	12123
10	1	full	12	0	41	12124
11	1	associate	15	0	7	12125
12	1	full	9	0	8	23234
13	1	associate	9	1	9	23232
14	1	full	9	0	44	23233
15	1	full	7	0	1	23456
16	1	full	13	0	15	34343
17	1	associate	11	1	16	34344
18	1	associate	10	1	17	34345
19	1	full	6	1	28	34346
20	1	assistant	16	1	34	34347
21	1	associate	8	1	56	45454
22	1	full	7	0	44	45456
23	0	full	8	0	33	45455
24	1	associate	9	0	23	56565
25	1	full	5	0	24	56567
26	1	associate	11	0	32	56566
27	0	full	5	0	31	56564
28	1	associate	3	1	21	56654
29	1	associate	3	1	66	55555
30	0	assistant	10	1	77	66666
31	1	associate	11	1	76	77777
32	1	assistant	9	1	75	88888
33	0	associate	4	1	11	22222
34	0	associate	6	1	22	33333
35	1	associate	1	0	33	23500
36	0	assistant	8	0	44	12300
37	1	assistant	4	0	55	23300
38	1	assistant	4	0	2	45000
39	1	assistant	4	0	3	67000
40	1	assistant	3	0	4	38000
41	1	assistant	3	1	5	43434
42	1	associate	0	0	6	54545
43	0	assistant	3	0	8	54544
44	1	assistant	2	0	9	32323
45	1	assistant	2	0	10	32322
46	0	assistant	2	0	21	54549
47	0	assistant	2	0	12	11223
48	1	assistant	1	0	32	11224
49	0	assistant	1	0	23	11225
50	0	assistant	1	0	25	11229

In [63]:

```
dfle = df
dfle.dg = le.fit_transform(dfle.dg)
dfle
```

Out[63]:

	sx	rk	yr	dg	yd	sl
0	1	full	25	0	35	36350
1	1	full	13	0	22	10000
2	1	full	10	0	23	10001
3	0	full	7	0	11	12345
4	1	full	19	1	13	23456
5	1	full	16	0	12	34567
6	0	full	0	1	14	36443
7	1	full	16	0	15	21212
8	1	full	13	1	16	12121
9	1	full	13	1	33	12123
10	1	full	12	0	41	12124
11	1	associate	15	0	7	12125
12	1	full	9	0	8	23234
13	1	associate	9	1	9	23232
14	1	full	9	0	44	23233
15	1	full	7	0	1	23456
16	1	full	13	0	15	34343
17	1	associate	11	1	16	34344
18	1	associate	10	1	17	34345
19	1	full	6	1	28	34346
20	1	assistant	16	1	34	34347
21	1	associate	8	1	56	45454
22	1	full	7	0	44	45456
23	0	full	8	0	33	45455
24	1	associate	9	0	23	56565
25	1	full	5	0	24	56567
26	1	associate	11	0	32	56566
27	0	full	5	0	31	56564
28	1	associate	3	1	21	56654
29	1	associate	3	1	66	55555
30	0	assistant	10	1	77	66666
31	1	associate	11	1	76	77777
32	1	assistant	9	1	75	88888
33	0	associate	4	1	11	22222
34	0	associate	6	1	22	33333
35	1	associate	1	0	33	23500
36	0	assistant	8	0	44	12300

	sx	rk	yr	dg	yd	sl
37	1	assistant	4	0	55	23300
38	1	assistant	4	0	2	45000
39	1	assistant	4	0	3	67000
40	1	assistant	3	0	4	38000
41	1	assistant	3	1	5	43434
42	1	associate	0	0	6	54545
43	0	assistant	3	0	8	54544
44	1	assistant	2	0	9	32323
45	1	assistant	2	0	10	32322
46	0	assistant	2	0	21	54549
47	0	assistant	2	0	12	11223
48	1	assistant	1	0	32	11224
49	0	assistant	1	0	23	11225
50	0	assistant	1	0	25	11229
51	0	assistant	0	0	43	35000

In [65]:

```
dfle = df
dfle.rk = le.fit_transform(dfle.rk)
dfle
```

Out[65]:

	sx	rk	yr	dg	yd	sl
0	1	2	25	0	35	36350
1	1	2	13	0	22	10000
2	1	2	10	0	23	10001
3	0	2	7	0	11	12345
4	1	2	19	1	13	23456
5	1	2	16	0	12	34567
6	0	2	0	1	14	36443
7	1	2	16	0	15	21212
8	1	2	13	1	16	12121
9	1	2	13	1	33	12123
10	1	2	12	0	41	12124
11	1	1	15	0	7	12125
12	1	2	9	0	8	23234
13	1	1	9	1	9	23232
14	1	2	9	0	44	23233
15	1	2	7	0	1	23456
16	1	2	13	0	15	34343
17	1	1	11	1	16	34344
18	1	1	10	1	17	34345
19	1	2	6	1	28	34346
20	1	0	16	1	34	34347
21	1	1	8	1	56	45454
22	1	2	7	0	44	45456

	sx	rk	yr	dg	yd	sl
23	0	2	8	0	33	45455
24	1	1	9	0	23	56565
25	1	2	5	0	24	56567
26	1	1	11	0	32	56566
27	0	2	5	0	31	56564
28	1	1	3	1	21	56654
29	1	1	3	1	66	55555
30	0	0	10	1	77	66666
31	1	1	11	1	76	77777
32	1	0	9	1	75	88888
33	0	1	4	1	11	22222
34	0	1	6	1	22	33333
35	1	1	1	0	33	23500
36	0	0	8	0	44	12300
37	1	0	4	0	55	23300
38	1	0	4	0	2	45000
39	1	0	4	0	3	67000
40	1	0	3	0	4	38000
41	1	0	3	1	5	43434
42	1	1	0	0	6	54545
43	0	0	3	0	8	54544
44	1	0	2	0	9	32323
45	1	0	2	0	10	32322
46	0	0	2	0	21	54549
47	0	0	2	0	12	11223
48	1	0	1	0	32	11224
49	0	0	1	0	23	11225
50	0	0	1	0	25	11229
51	0	0	0	0	43	35000

In [67]:

```python
X = dfle[['sx','rk','dg','yd']].values
X
```

Out[67]:

```
array([[ 1,  2,  0, 35],
       [ 1,  2,  0, 22],
       [ 1,  2,  0, 23],
       [ 0,  2,  0, 11],
       [ 1,  2,  1, 13],
       [ 1,  2,  0, 12],
       [ 0,  2,  1, 14],
       [ 1,  2,  0, 15],
       [ 1,  2,  1, 16],
       [ 1,  2,  1, 33],
       [ 1,  2,  0, 41],
       [ 1,  1,  0,  7],
       [ 1,  2,  0,  8],
       [ 1,  1,  1,  9],
       [ 1,  2,  0, 44],
       [ 1,  2,  0,  1],
       [ 1,  2,  0, 15],
       [ 1,  1,  1, 16],
       [ 1,  1,  1, 17]
```

```
       [ 1,   1,   1,   17]],
       [ 1,   2,   1,   28],
       [ 1,   0,   1,   34],
       [ 1,   1,   1,   56],
       [ 1,   2,   0,   44],
       [ 0,   2,   0,   33],
       [ 1,   1,   0,   23],
       [ 1,   2,   0,   24],
       [ 1,   1,   0,   32],
       [ 0,   2,   0,   31],
       [ 1,   1,   1,   21],
       [ 1,   1,   1,   66],
       [ 0,   0,   1,   77],
       [ 1,   1,   1,   76],
       [ 1,   0,   1,   75],
       [ 0,   1,   1,   11],
       [ 0,   1,   1,   22],
       [ 1,   1,   0,   33],
       [ 0,   0,   0,   44],
       [ 1,   0,   0,   55],
       [ 1,   0,   0,    2],
       [ 1,   0,   0,    3],
       [ 1,   0,   0,    4],
       [ 1,   0,   1,    5],
       [ 1,   1,   0,    6],
       [ 0,   0,   0,    8],
       [ 1,   0,   0,    9],
       [ 1,   0,   0,   10],
       [ 0,   0,   0,   21],
       [ 0,   0,   0,   12],
       [ 1,   0,   0,   32],
       [ 0,   0,   0,   23],
       [ 0,   0,   0,   25],
       [ 0,   0,   0,   43]], dtype=int64)
```

In [68]:

```
y = dfle.sl
y
```

Out[68]:

```
0      36350
1      10000
2      10001
3      12345
4      23456
5      34567
6      36443
7      21212
8      12121
9      12123
10     12124
11     12125
12     23234
13     23232
14     23233
15     23456
16     34343
17     34344
18     34345
19     34346
20     34347
21     45454
22     45456
23     45455
24     56565
25     56567
26     56566
27     56564
28     56654
29     55555
30     66666
```

```
31    77777
32    88888
33    22222
34    33333
35    23500
36    12300
37    23300
38    45000
39    67000
40    38000
41    43434
42    54545
43    54544
44    32323
45    32322
46    54549
47    11223
48    11224
49    11225
50    11229
51    35000
Name: sl, dtype: int64
```

In [71]:

```python
from sklearn.preprocessing import OneHotEncoder
ohe = OneHotEncoder()
```

In [73]:

```python
X = ohe.fit_transform(X).toarray()
X
```

Out[73]:

```
array([[0., 1., 0., ..., 0., 0., 0.],
       [0., 1., 0., ..., 0., 0., 0.],
       [0., 1., 0., ..., 0., 0., 0.],
       ...,
       [1., 0., 1., ..., 0., 0., 0.],
       [1., 0., 1., ..., 0., 0., 0.],
       [1., 0., 1., ..., 0., 0., 0.]])
```

In [74]:

```python
model.fit(X,y)
```

Out[74]:

```
LinearRegression()
```

In [76]:

```python
model.score(X,y)
```

Out[76]:

```
0.8555070591731344
```

In [ ]: