

```
In [1]: import pandas as pd
import numpy as np
from sklearn import linear_model
```

```
In [2]: df = pd.read_csv("homeprices.csv")
df
```

```
Out[2]:
```

	Unnamed: 0	area	bedrooms	age	price
0	NaN	2600	3.0	20	550000
1	NaN	3000	4.0	15	565000
2	NaN	3200	NaN	18	610000
3	NaN	3600	3.0	30	595000
4	NaN	4000	5.0	8	760000
5	NaN	4100	6.0	8	810000

```
In [4]: import math

d = math.floor(df.bedrooms.median())
d
```

```
Out[4]: 4
```

```
In [12]: df.bedrooms.fillna(d, inplace = True)
df
```

```
Out[12]:
```

	Unnamed: 0	area	bedrooms	age	price
0	NaN	2600	3.0	20	550000
1	NaN	3000	4.0	15	565000
2	NaN	3200	4.0	18	610000
3	NaN	3600	3.0	30	595000
4	NaN	4000	5.0	8	760000
5	NaN	4100	6.0	8	810000

```
In [9]: df
reg = linear_model.LinearRegression()
```

```
In [13]: reg.fit(df[['area', 'bedrooms', 'age']], df.price)
```

```
Out[13]: LinearRegression()
```

```
In [15]: reg.coef_
```

```
Out[15]: array([ 112.06244194, 23388.88007794, -3231.71790863])
```

```
In [16]: reg.intercept_
```

```
Out[16]: 221323.00186540408
```

```
In [17]: reg.predict([[3000,3,40]])
```

```
Out[17]: array([498408.25158031])
```

```
In [18]: reg.predict([[2500,4,5]])
```

```
Out[18]: array([578876.03748933])
```

```
In [19]: df = pd.read_csv("hiring.csv")
```

```
In [28]: df
```

```
Out[28]:
```

	Unnamed: 0	experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
0	NaN	0	8.0	9	50000
1	NaN	0	8.0	6	45000
2	NaN	five	6.0	7	60000
3	NaN	two	10.0	10	65000
4	NaN	seven	9.0	6	70000
5	NaN	three	7.0	10	62000
6	NaN	ten	NaN	7	72000
7	NaN	eleven	7.0	8	80000

```
In [26]: df.experience = df.experience.fillna("zero")
df
```

```
Out[26]:
```

	Unnamed: 0	experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
0	NaN	0	8.0	9	50000
1	NaN	0	8.0	6	45000
2	NaN	five	6.0	7	60000
3	NaN	two	10.0	10	65000
4	NaN	seven	9.0	6	70000
5	NaN	three	7.0	10	62000
6	NaN	ten	NaN	7	72000
7	NaN	eleven	7.0	8	80000

```
In [27]: df.experience = df.experience.apply(w2n.word_to_num)
df
```

```
-----
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_11272\3776223095.py in <module>
----> 1 df.experience = df.experience.apply(w2n.word_to_num)
      2 df

NameError: name 'w2n' is not defined
```

```
In [57]: d = pd.read_csv("hiring.csv")
d
```

Out[57]:

	Unnamed: 0	experience	test_score	interview_score	salary
0	NaN	0	8.0	9	50000
1	NaN	0	8.0	6	45000
2	NaN	5	6.0	7	60000
3	NaN	2	10.0	10	65000
4	NaN	7	9.0	6	70000
5	NaN	3	7.0	10	62000
6	NaN	10	NaN	7	72000
7	NaN	11	7.0	8	80000

```
In [51]: import math
p = math.floor(d.test_score.mean())
p
```

Out[51]: 7

```
In [58]: d.test_score.fillna(p, inplace = True)
d
```

Out[58]:

	Unnamed: 0	experience	test_score	interview_score	salary
0	NaN	0	8.0	9	50000
1	NaN	0	8.0	6	45000
2	NaN	5	6.0	7	60000
3	NaN	2	10.0	10	65000
4	NaN	7	9.0	6	70000
5	NaN	3	7.0	10	62000
6	NaN	10	7.0	7	72000
7	NaN	11	7.0	8	80000

```
In [59]: reg = linear_model.LinearRegression()
```

```
In [60]: reg.fit(d[['experience', 'test_score', 'interview_score']], d.salary)
```

Out[60]: LinearRegression()

```
In [61]: reg.coef_
```

Out[61]: array([2922.26901502, 2221.30909959, 2147.48256637])

```
In [62]: reg.intercept_
```

Out[62]: 14992.65144669314

```
In [63]: reg.predict([[2,9,6]])
```

Out[63]: array([53713.86677124])

```
In [64]: reg.predict([[12,10,10]])
```

Out[64]: array([93747.79628651])

```
In [ ]:
```