

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

```
In [2]: df = pd.read_csv('bedroom.csv')
df
```

Out[2]:

	area	bedrooms	age	price
0	2600.0	3.0	20.0	550000.0
1	3000.0	4.0	15.0	565000.0
2	3200.0	NaN	18.0	610000.0
3	3600.0	3.0	30.0	595000.0
4	4000.0	5.0	8.0	760000.0
5	4100.0	6.0	8.0	810000.0
6	NaN	NaN	NaN	NaN

```
In [4]: newdf=df.drop('price',axis='columns')
newdf
```

Out[4]:

	area	bedrooms	age
0	2600.0	3.0	20.0
1	3000.0	4.0	15.0
2	3200.0	NaN	18.0
3	3600.0	3.0	30.0
4	4000.0	5.0	8.0
5	4100.0	6.0	8.0
6	NaN	NaN	NaN

```
In [6]: price=df.price
price
```

Out[6]:

0	550000.0
1	565000.0
2	610000.0
3	595000.0
4	760000.0
5	810000.0
6	NaN

Name: price, dtype: float64

```
In [13]: price.median()
```

```
Out[13]: 602500.0
```

```
In [14]: price.fillna(price.median())
```

```
Out[14]: 0    550000.0  
         1    565000.0  
         2    610000.0  
         3    595000.0  
         4    760000.0  
         5    810000.0  
         6    602500.0  
         Name: price, dtype: float64
```

```
In [17]: newdf.bedrooms.median()
```

```
Out[17]: 4.0
```

```
In [21]: newdf.bedrooms = df.bedrooms.fillna(newdf.bedrooms.median())  
         newdf
```

```
Out[21]:
```

	area	bedrooms	age
0	2600.0	3.0	20.0
1	3000.0	4.0	15.0
2	3200.0	4.0	18.0
3	3600.0	3.0	30.0
4	4000.0	5.0	8.0
5	4100.0	6.0	8.0
6	NaN	4.0	NaN

```
In [10]: newdf.age.median()
```

```
Out[10]: 16.5
```

```
In [22]: newdf.age = df.age.fillna(newdf.age.median())
newdf
```

Out[22]:

	area	bedrooms	age
0	2600.0	3.0	20.0
1	3000.0	4.0	15.0
2	3200.0	4.0	18.0
3	3600.0	3.0	30.0
4	4000.0	5.0	8.0
5	4100.0	6.0	8.0
6	NaN	4.0	16.5

```
In [24]: newdf.area.median()
```

Out[24]: 3400.0

```
In [25]: newdf.area = df.area.fillna(newdf.area.median())
newdf
```

Out[25]:

	area	bedrooms	age
0	2600.0	3.0	20.0
1	3000.0	4.0	15.0
2	3200.0	4.0	18.0
3	3600.0	3.0	30.0
4	4000.0	5.0	8.0
5	4100.0	6.0	8.0
6	3400.0	4.0	16.5

```
In [28]: price.median()
```

Out[28]: 602500.0

```
In [35]: price=price.fillna(price.median())
price
```

Out[35]:

0	550000.0
1	565000.0
2	610000.0
3	595000.0
4	760000.0
5	810000.0
6	602500.0

Name: price, dtype: float64

```
In [37]: reg = linear_model.LinearRegression()  
reg.fit(newdf,price)
```

```
Out[37]: 

▼ LinearRegression  
LinearRegression()


```

```
In [38]: reg.predict(newdf)
```

```
Out[38]: array([511357.12356788, 597599.47604163, 611589.11793099, 592257.78565509,  
749906.1499924 , 794354.37961529, 635435.96719673])
```

```
In [39]: reg.score(newdf,price)
```

```
Out[39]: 0.9348241332988168
```

```
In [ ]:
```