```
In [9]:  import pandas as pd
         import numpy as np
         from sklearn import linear_model
```

```
In [21]:  data = pd.read_csv('homeprices2.csv')
          data
```

Out[21]:

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

Getting the median of bedrooms column

```
In [30]:  median_bedroom = int(data.bedrooms.median())
          median_bedroom
```

Out[30]: 4

Filling the null values with median

```
In [31]:  data['bedrooms'].fillna(median_bedroom, inplace=True)
          data
```

Out[31]:

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

Training the model

```
In [49]:  model = linear_model.LinearRegression()
          model.fit(data[['area', 'bedrooms', 'age']], data[['price']])
```

Loading [MathJax]/extensions/Safe.js

Out[49]: ▼ LinearRegression

LinearRegression()

Getting the coeff and intercept

In [53]:
```python
print(model.coef_) # this will return three coeff for area, bedrooms and age
print(model.intercept_)
```

```
[[  112.06244194 23388.88007794 -3231.71790863]]
[221323.0018654]
```

Predicting 2 data

In [51]:
```python
model.predict([[3000, 3, 40], [2500, 4, 5]])
```

```
C:\Users\User\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklea
rn\base.py:465: UserWarning: X does not have valid feature names, but LinearR
egression was fitted with feature names
  warnings.warn(
```

Out[51]:
```
array([[498408.25158031],
       [578876.03748933]])
```

Calculating the best line formula

In [54]:
```python
# y = m1*x1 + m2*x2 + m3*x3 + b
# price = m1*area + m2*bedrooms + m3*age + b
112.06244194 * 3000 + 23388.88007794 * 3 + -3231.71790863 * 40 + 221323.0018
```

Out[54]: 498408.25157402