In [68]:

```python
import pandas
import json
from sklearn import linear_model
```

In [69]:

```python
#Opens json
with open('xaaa.json', encoding="utf8") as f:
    data = json.load(f)

#Save only immobile's informations
im_list = []
for d in data:
    im_list.append(d['_source'])

#Convert to a data frame
data = pandas.DataFrame(im_list)
```

In [70]:

```python
#Displaying dataframe columns
data.columns
```

Out[70]:

```
Index(['anunciante', 'area_comum', 'area_construida', 'area_privativa',
       'area_terreno', 'area_total', 'area_util', 'bairro', 'banheiro',
       'cidade', 'cidade_uf', 'codigo', 'condominio', 'contato_imobiliaria',
       'data_inclusao', 'default', 'descricao', 'edificio', 'endereco',
       'error', 'finalidade', 'garagem', 'geohash', 'geometry', 'iptu',
       'isparticular', 'isportal', 'latitude', 'lista_fotos', 'longitude',
       'nome_corretor', 'numero_corretor', 'opcionais', 'quarto', 'scrapin
g',
       'status', 'suite', 'tipo_imovel', 'tipo_negocio', 'title', 'uf', 'ur
l',
       'valor', 'zoneamento'],
      dtype='object')
```

In [71]:

```python
#Selecting only relevant columns
data = data[["area_util", "banheiro", "garagem", "quarto", "valor"]]
```

In [72]:

```python
data.columns
```

Out[72]:

```
Index(['area_util', 'banheiro', 'garagem', 'quarto', 'valor'], dtype='objec
t')
```

In [77]:

```python
data.head(4)
```

Out[77]:

	area_util	banheiro	garagem	quarto	valor
767	104.0	3.0	2.0	3.0	647000.0
768	149.0	3.0	4.0	3.0	650000.0
769	113.0	4.0	1.0	3.0	650000.0
770	165.0	2.0	3.0	4.0	650000.0

In [73]:

```python
#Removing de null values
data = data.dropna(subset = ['banheiro','area_util','quarto', 'garagem', 'banheiro','valor'
```

In [74]:

```python
#Linear regression

X = data[['area_util','banheiro','garagem','quarto']]
Y = data['valor']

# with sklearn
regr = linear_model.LinearRegression()
regr.fit(X, Y)
```

Out[74]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
        normalize=False)
```

In [75]:

```python
print('Intercept: \n', regr.intercept_)
print('Coefficients: \n', regr.coef_)
```

```
Intercept:
 96167.9061063031
Coefficients:
 [  145.87314981 55000.38650363 62282.0402052   62628.31711994]
```

In [78]:

```python
# prediction with sklearn
area_util = 104
banheiro = 3
garagem = 2
quarto = 3
print ('Predicted Imobille Price: \n', regr.predict([[area_util, banheiro, garagem, quarto]
```

```
Predicted Imobille Price:
 [588788.90496738]
```

In [ ]: