



Programación. Python

Dataframes con pandas

Cristóbal Pareja Flores :

DataFrames

```
import pandas
realestate = pandas.read_csv('realestate.csv')
realestate
```

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.631913	-121.434879
1	51 OMAHA CT	SACRAMENTO	95823	CA	3	1	1167	Residential	Wed May 21 00:00:00 EDT 2008	68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	95815	CA	2	1	796	Residential	Wed May 21 00:00:00 EDT 2008	68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAMENTO	95824	CA	2	1	797	Residential	Wed May 21 00:00:00 EDT 2008	81900	38.519470	-121.435768
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3	1	1122	Condo	Wed May 21 00:00:00 EDT 2008	89921	38.662595	-121.327813
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895	38.681659	-121.351705
7	2561 19TH AVE	SACRAMENTO	95820	CA	3	1	1177	Residential	Wed May 21 00:00:00	91002	38.535092	-121.481367

DataFrames

```
import pandas
realestate = pandas.read_csv('realestate.csv')
realestate
```

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	1177	Residential	Wed May 21 00:00:00	91000	38.535092	-121.481367
1	51 OMAHA CT	SACRAMENTO	95823	CA	3	1						
2	2796 BRANCH ST	SACRAMENTO	95815	CA	2	1						
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1						
4	6001 MCMAHON DR	SACRAMENTO	95824	CA	2	1						
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3	1						
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2						
7	2561 19TH AVE	SACRAMENTO	95820	CA	3	1						

```
type(realestate)
```

```
pandas.core.frame.DataFrame
```

```
realestate.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 985 entries, 0 to 984
```

```
Data columns (total 12 columns):
```

```
street      985 non-null object
```

```
city        985 non-null object
```

```
zip         985 non-null int64
```

```
state       985 non-null object
```

```
beds        985 non-null int64
```

```
baths       985 non-null int64
```

```
sq__ft      985 non-null int64
```

```
type        985 non-null object
```

```
sale_date   985 non-null object
```

```
price       985 non-null int64
```

```
latitude    985 non-null float64
```

```
longitude   985 non-null float64
```

```
dtypes: float64(2), int64(5), object(5)
```

```
memory usage: 92.4+ KB
```

DataFrames

```
import pandas
realestate = pandas.read_csv('realest
realestate
```

```
realestate['price']
```

```
0    59222
1    68212
2    68880
3    69307
4    81900
5    89921
6    90895
7    91002
...
```

```
980    232425
981    234000
982    235000
983    235301
984    235738
```

```
Name: price, dtype: int64
```

					6					te	price	latitude	longitude
					7								
0	3526 HIGH ST	SACRAMENTO	95838	CA	...				00	59222	38.631913	-121.434879	
1	51 OMAHA CT	SACRAMENTO	95823	CA	980					00	68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	95815	CA	981					00	68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	982					00	69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAMENTO	95824	CA	983					00	81900	38.519470	-121.435768
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	984					00	89921	38.662595	-121.327813
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	Name: price, dtype: int64					00	90895	38.681659	-121.351705
7	2561 19TH AVE	SACRAMENTO	95820	CA	3	1	1177	Residential	Wed May 21 00:00:00 EDT 2008	00	91002	38.535092	-121.481367

DataFrames

```
import pandas
realestate = pandas.read_csv('realestate.csv')
realestate
```

	street	city	zip
0	3526 HIGH ST	SACRAMENTO	95838
1	51 OMAHA CT	SACRAMENTO	95823
2	2796 BRANCH ST	SACRAMENTO	95815
3	2805 JANETTE WAY	SACRAMENTO	95815
4	6001 MCMAHON DR	SACRAMENTO	95824
5	5828 PEPPERMILL CT	SACRAMENTO	95841
6	6048 OGDEN NASH WAY	SACRAMENTO	95842
7	2561 19TH AVE	SACRAMENTO	95820

```
realestate['price']
```

```
realestate[['price','zip']]
```

	price	zip
0	59222	95838
1	68212	95823
2	68880	95815
3	69307	95815
...
981	234000	95823
982	235000	95610
983	235301	95758
984	235738	95762

985 rows x 2 columns

	price	latitude	longitude
0	59222	38.631913	-121.434879
1	68212	38.478902	-121.431028
2	68880	38.618305	-121.443839
3	69307	38.616835	-121.439146
64	81900	38.519470	-121.435768
65	89921	38.662595	-121.327813
66	90895	38.681659	-121.351705
67	91002	38.535092	-121.481367

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.631913	-121.434879
1	51 OMAHA CT	SACRAMENTO	selcols = realestate.columns[2:6] selcols							68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	Index(['zip', 'state', 'beds', 'baths'], dtype='object')							68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	realestate[selcols]							69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAMENTO								81900	38.519470	-121.435768
5	5828 PEPPERMILL CT	SACRAMENTO	0	95838	CA	2	1			89921	38.662595	-121.327813
6	6048 OGDEN NASH WAY	SACRAMENTO	1	95823	CA	3	1			90895	38.681659	-121.351705
7	2561 19TH AVE	SACRAMENTO	2	95815	CA	2	1			91002	38.535092	-121.481367
			3	95815	CA	2	1					
							
			981	95823	CA	3	2					
			982	95610	CA	3	2					
			983	95758	CA	4	2					
			984	95762	CA	3	2					

985 rows × 4 columns

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude	
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.631913	-121.434879	
1	51 OMAHA CT	SACRAMENTO	selcols = realestate.columns[2:6] selcols							68212	38.478902	-121.431028	
2	2796 BRANCH ST	SACRAMENTO	Index(['zip', 'state', 'beds', 'baths'], dtype='object')							68880	38.618305	-121.443839	
3	2805 JANETTE WAY	SACRAMENTO	realestate[selcols]							69307	35	-121.439146	
4	6001 MCMAHON DR	SACRAMENTO	zip	state							70	-121.435768	
5	5828 PEPPERMILL CT	SACRAMENTO	0	95838	CA	['zip', 'state', 'beds', 'baths', 'price']					95	-121.327813	
6	6048 OGDEN NASH WAY	SACRAMENTO	1	95823	CA	realestate[mixed]					59	-121.351705	
7	2561 19TH AVE	SACRAMENTO	2	95815	CA						81000	42	-121.481367
			3	95815	CA	zip	state	beds	baths	price			
			0	95838	CA	2	1		59222	
			981	95823	CA	1	95823	CA	3	1		68212	
			982	95610	CA	2	95815	CA	2	1		68880	
			983	95758	CA	3	95815	CA	2	1		69307	
			984	95762	CA	4	95824	CA	2	1		81000	

985 rows × 4 columns

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.631913	-121.434879
1	51 OMAHA CT	SACRAMENTO	95823	CA	3	1	1167	Residential	Wed May 21 00:00:00 EDT 2008	68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	95815	CA	2	1	796	Residential	Wed May 21 00:00:00 EDT 2008	68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	69307	38.616835	-121.439146

4	6001 MCMAHON DR	SACRAMENTO	95824	CA	2	realestate.iloc[3]						
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3	street						
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	city						
7	2561 19TH AVE	SACRAMENTO	95820	CA	3	zip						

```
2805 JANETTE WAY
SACRAMENTO
95815
CA
2
1
852
Residential
Wed May 21 00:00:00 EDT 2008
69307
38.6168
-121.439
Name: 3, dtype: object
```


	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.631913	-121.434879
1	51 OMAHA CT	SACRAMENTO	95823	CA	3	1	1167	Residential	Wed May 21 00:00:00 EDT 2008	68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	95815	CA	2	1	796	Residential	Wed May 21 00:00:00 EDT 2008	68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAMENTO	95824	CA	2							
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3							
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3							
7	2561 19TH AVE	SACRAMENTO	95820	CA	3							

realestate.iloc[3]

street2805 JANETTE WAY
citySACRAMENTO
zip95815
stateCA
beds2
baths1

realestate.iloc[[3,6]]

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	69307	38.616835	-121.439146
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895	38.681659	-121.351705

longitude-121.439

Name: 3, dtype: object

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00	59222	38.631913	-121.434879
1	51 OMAHA CT	SACRAMENTO	95823	CA						68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	95815	CA						68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA						69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAMENTO	95824	CA						81900	38.519470	-121.435768
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA						89921	38.662595	-121.327813
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA						90895	38.681659	-121.351705
7	2561 19TH AVE	SACRAMENTO	95820	CA						91002	38.535092	-121.481367

realestate.iloc[3]['city']

'SACRAMENTO'

realestate.iloc[3][1]

'SACRAMENTO'

realestate[['price','zip']][3:8]

	price	zip
3	69307	95815
4	81900	95824
5	89921	95841
6	90895	95842
7	91002	95820

Operaciones con dataframes

```
realestate.sort_values(by='price', ascending=False)
```

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price	latitude	longitude
864	9401 BARREL RACER CT	WILTON	95693	CA	4	3	4400	Residential	Fri May 16 00:00:00 EDT 2008	884790	38.415298	-121.194858
863	2982 ABERDEEN LN	EL DORADO HILLS	95762	CA	4	3	0	Residential	Fri May 16 00:00:00 EDT 2008	879000	38.706692	-121.058869
...

```
realestate.describe()
```

	zip	beds	baths	sq__ft	price	latitude	longitude
count	985.000000	985.000000	985.000000	985.000000	985.000000	985.000000	985.000000
mean	95750.697462	2.911675	1.776650	1314.916751	234144.263959	38.607732	-121.355982
std	85.176072	1.307932	0.895371	853.048243	138365.839085	0.145433	0.138278
min	95603.000000	0.000000	0.000000	0.000000	1551.000000	38.241514	-121.551704
25%	95660.000000	2.000000	1.000000	952.000000	145000.000000	38.482717	-121.446127
50%	95762.000000	3.000000	2.000000	1304.000000	213750.000000	38.626582	-121.376220
75%	95828.000000	4.000000	2.000000	1718.000000	300000.000000	38.695589	-121.295778
max	95864.000000	8.000000	5.000000	5822.000000	884790.000000	39.020808	-120.597599

Operaciones con dataframes

```
cond = realestate['baths'] == 2  
cond
```

```
0    False  
1    False  
2    False  
3    False  
4    False  
5    False  
6     True  
7    False
```

```
realestate[cond]
```

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price	latitude	longitude
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895	38.681659	-121.351705
8	11150 TRINITY RIVER DR Unit 114	RANCHO CORDOVA	95670	CA	2	2	941	Condo	Wed May 21 00:00:00 EDT 2008	94905	38.621188	-121.270555
9	7325 10TH ST	RIO LINDA	95673	CA	3	2	1146	Residential	Wed May 21 00:00:00 EDT 2008	98937	38.700909	-121.442979
10	645 MORRISON AVE	SACRAMENTO	95838	CA	3	2	909	Residential	Wed May 21 00:00:00 EDT 2008	100309	38.637663	-121.451520
11	4085 FAWN CIR	SACRAMENTO	95823	CA	3	2	1289	Residential	Wed May 21 00:00:00 EDT 2008	106250	38.470746	-121.458918

Operaciones con dataframes

```
two_baths = realestate['baths'] == 2
three_beds = realestate['beds'] == 3
realestate[two_baths & three_beds]
```

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price	latitude	longitude
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895	38.681659	-121.351705
9	7325 10TH ST	RIO LINDA	95673	CA	3	2	1146	Residential	Wed May 21 00:00:00 EDT 2008	98937	38.700909	-121.442979
10	645 MORRISON AVE	SACRAMENTO	95838	CA	3	2	909	Residential	Wed May 21 00:00:00 EDT 2008	100309	38.637663	-121.451520
11	4085 FAWN CIR	SACRAMENTO	95823	CA	3	2	1289	Residential	Wed May 21 00:00:00 EDT 2008	106250	38.470746	-121.458918
19	113 LEEWILL AVE	RIO LINDA	95673	CA	3	2	1356	Residential	Wed May 21 00:00:00 EDT 2008	121630	38.689999	-121.463220

Operaciones con dataframes

```
two_baths = realestate['baths'] == 2
three_beds = realestate['beds'] == 3
realestate[two_baths & three_beds]
```

	street	city	zip	state	beds	baths	sq_
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	11
9	7325 10TH ST	RIO LINDA	95673	CA	3	2	11
10	645 MORRISON AVE	SACRAMENTO	95838	CA	3	2	9
11	4085 FAWN CIR	SACRAMENTO	95823	CA	3	2	12
19	113 LEEWILL AVE	RIO LINDA	95673	CA	3	2	13

```
realestate.groupby(['zip', 'baths']).size()
```

zip	baths	
95603	2	2
	3	3
95608	1	4
	2	15
	3	1
95610	2	5
	3	1
	4	1
95614	2	1
95619	2	1
95621	1	7
	2	20
	3	1
95622	1	1

Operaciones con dataframes

```
two_baths = realestate['baths'] == 2
three_beds = realestate['beds'] == 3
realestate[two_baths & three_beds]
```

```
realestate.groupby(['zip', 'baths']).size().unstack()
```

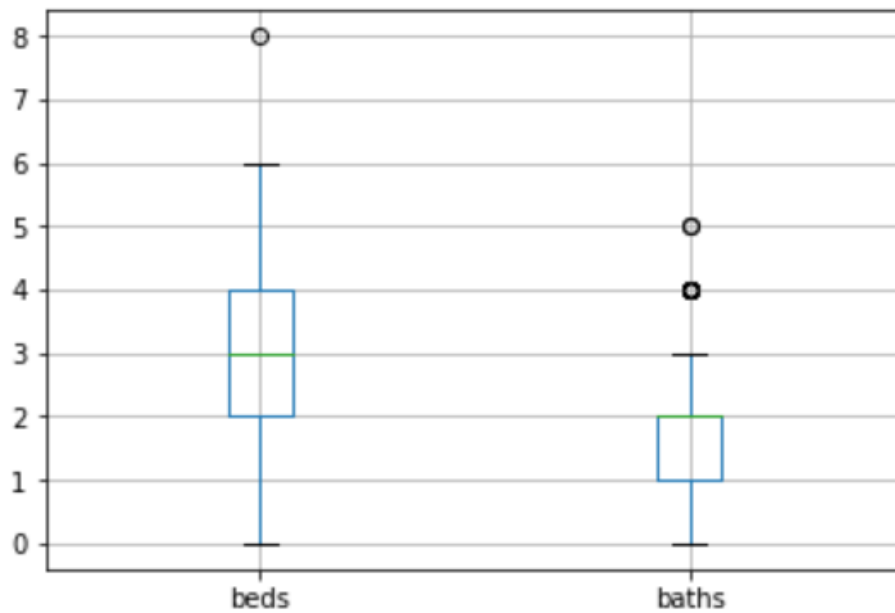
	baths	0	1	2	3	4	5
	zip						
6	95603	NaN	NaN	2.0	3.0	NaN	NaN
9	95608	NaN	4.0	15.0	1.0	NaN	NaN
10	95610	NaN	NaN	5.0	1.0	1.0	NaN
11	95614	NaN	NaN	1.0	NaN	NaN	NaN
19	95619	NaN	NaN	1.0	NaN	NaN	NaN
	95621	NaN	7.0	20.0	1.0	NaN	NaN
	95623	NaN	1.0	1.0	NaN	NaN	NaN
	95624	3.0	1.0	22.0	8.0	NaN	NaN

```
realestate.groupby(['zip', 'baths']).size()
```

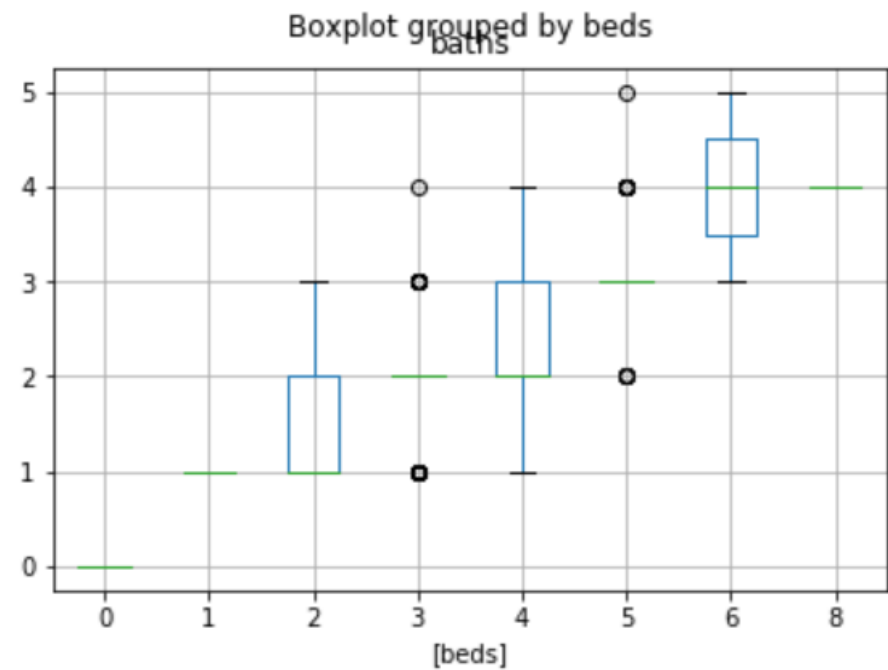
	zip	baths	
	95603	2	2
		3	3
11	95608	1	4
		2	15
11		3	1
	95610	2	5
9		3	1
		4	1
12	95614	2	1
	95619	2	1
13	95621	1	7
		2	20
		3	1
	95623	1	1

Visualización

```
realestate[['beds', 'baths']].boxplot()
```



```
realestate[['beds', 'baths']].boxplot(by="beds")
```



Visualización

Matplotlib

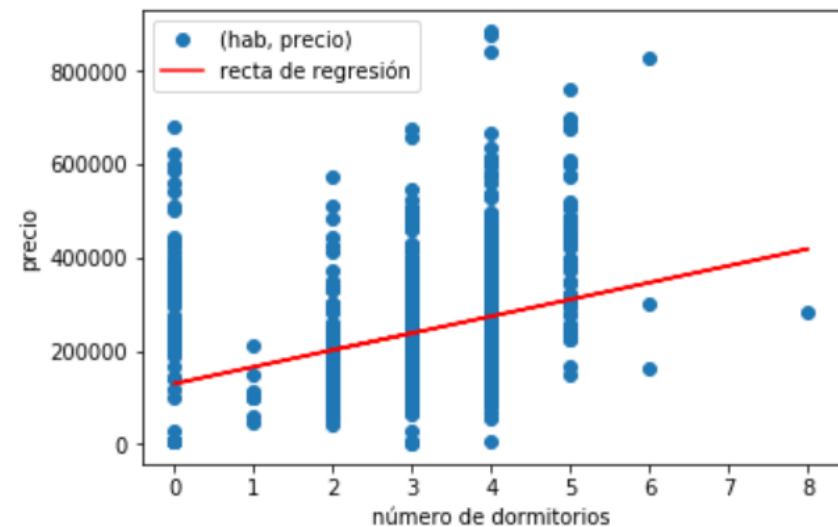
```
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt

xs = tabla_habs_precios['beds']
ys = tabla_habs_precios['price']
slope, intercept, r_value, p_value, std_err = \
    stats.linregress(xs, ys)
recta_regres = lambda x: intercept + slope*x

plt.plot(xs, ys, 'o', label='(hab, precio)')
plt.plot(xs, recta_regres(xs), 'r', \
         label='recta de regresión')
plt.legend(loc = 'upper left')
plt.xlabel('número de dormitorios')
plt.ylabel('precio')
plt.show()
```

```
tabla_habs_precios = realestate[['beds', 'price']]
tabla_habs_precios
```

	beds	price
0	2	59222
1	3	68212
2	2	68880
3	2	69307
4	2	81900



Visualización Seaborn

```
import seaborn
```

```
seaborn.pairplot(realestate)
```

<https://seaborn.pydata.org/>

