ntic master
revolucionamos la comunicación

UNIVERSIDAD COMPLUTENSE DE MADRID

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.cs
realestate
```

	street	city	zip	state	beds	baths	sq	gitude
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1		34879
1	51 OMAHA CT	SACRAMENTO	95823	CA	3	1		31028
2	2796 BRANCH ST	SACRAMENTO	95815	CA	2	1		43839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1		39146
4	6001 MCMAHON DR	SACRAMENTO	95824	CA	2	1		35768
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3	1		27813
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2		51705
7	2561 19TH AVE	SACRAMENTO	95820	CA	3	1	1177	Residential Wed May 21 00:00:00 91002 38.535092 -121.481367

```
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
```

	street	city	zip	state	bed					te	price	latitude	longitude
0	3526 HIGH ST	SACRAMENTO	95838	CA						00 08	59222	38.631913	-121.434879
1	51 OMAHA CT	SACRAMENTO	95823	CA						00 08	68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAMENTO	95815	CA						00 08	68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAMENTO	95815	CA						00 08	69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAMENTO	95824	CA						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('re
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 × 2 columns

te	price	latitude	longitude
00 08	59222	38.631913	-121.434879
00 08	68212	38.478902	-121.431028
00 08	68880	38.618305	-121.443839
00 08	69307	38.616835	-121.439146
64 00 2008	81900	38.519470	-121.435768
00:00:00 T 2008	89921	38.662595	-121.327813
00:00:00 T 2008	90895	38.681659	-121.351705
00:00:00	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	SACRAME								68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAME								68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAME								69307	38.616835	-121.439146
4	6001 MCMAHON DR	SACRAME								81900	38.519470	-121.435768
5	5828 PEPPERMILL CT	SACRAME								89921	38.662595	-121.327813
6	6048 OGDEN NASH WAY	SACRAME								90895	38.681659	-121.351705
7	2561 19TH AVE	SACRAME								91002	38.535092	-121.481367

```
selcols = realestate.columns[2:6]
selcols
```

```
Index(['zip', 'state', 'beds', 'baths'], dtype='object')
```

```
realestate[selcols]
```

	zip	state	beds	baths
0	95838	CA	2	1
1	95823	CA	3	1
2	95815	CA	2	1
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	SACRAME								68212	38.478902	-121.431028
2	2796 BRANCH ST	SACRAME								68880	38.618305	-121.443839
3	2805 JANETTE WAY	SACRAME									5	-121.439146
4	6001 MCMAHON DR	SACRAME									0	-121.435768
5	5828 PEPPERMILL CT	SACRAME									5	-121.327813
6	6048 OGDEN NASH WAY	SACRAME									9	-121.351705
7	2561 19TH AVE	SACRAME									2	-121.481367

```
selcols = realestate.columns[2:6]
selcols
```

```
Index(['zip', 'state', 'beds', 'baths'], dtype='object')
```

```
realestate[selco
```

	zip	state
0	95838	CA
1	95823	CA
2	95815	CA
3	95815	CA
...	...	...
981	95823	CA
982	95610	CA
983	95758	CA
984	95762	CA

985 rows × 4 columns

```
mixed = list(selcols)+['price']
mixed
```

```
['zip', 'state', 'beds', 'baths', 'price']
```

```
realestate[mixed]
```

	zip	state	beds	baths	price
0	95838	CA	2	1	59222
1	95823	CA	3	1	68212
2	95815	CA	2	1	68880
3	95815	CA	2	1	69307
4	95824	CA	2	1	81900

	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						70	-121.435768
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3						95	-121.327813
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3						59	-121.351705
7	2561 19TH AVE	SACRAMENTO	95820	CA	3						92	-121.481367

```
realestate.iloc[3]
```

```
street                    2805 JANETTE WAY
city                            SACRAMENTO
zip                                  95815
state                                   CA
beds                                     2
baths                                    1
sq__ft                                 852
type                           Residential
sale_date    Wed May 21 00:00:00 EDT 2008
price                                69307
latitude                           38.6168
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 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						70	-121.435768
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3						95	-121.327813
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3						59	-121.351705
7	2561 19TH AVE	SACRAMENTO	95820	CA	3						92	-121.481367

```
realestate.iloc[3]
```

```
street                2805 JANETTE WAY
city                        SACRAMENTO
zip                            95815
state                             CA
beds                               2
baths                              1
```

```
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_	de
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	11	05
9	7325 10TH ST	RIO LINDA	95673	CA	3	2	11	79
10	645 MORRISON AVE	SACRAMENTO	95838	CA	3	2	9	20
11	4085 FAWN CIR	SACRAMENTO	95823	CA	3	2	12	18
19	113 LEEWILL AVE	RIO LINDA	95673	CA	3	2	13	20

```
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
```

Operaciones con dataframes

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

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

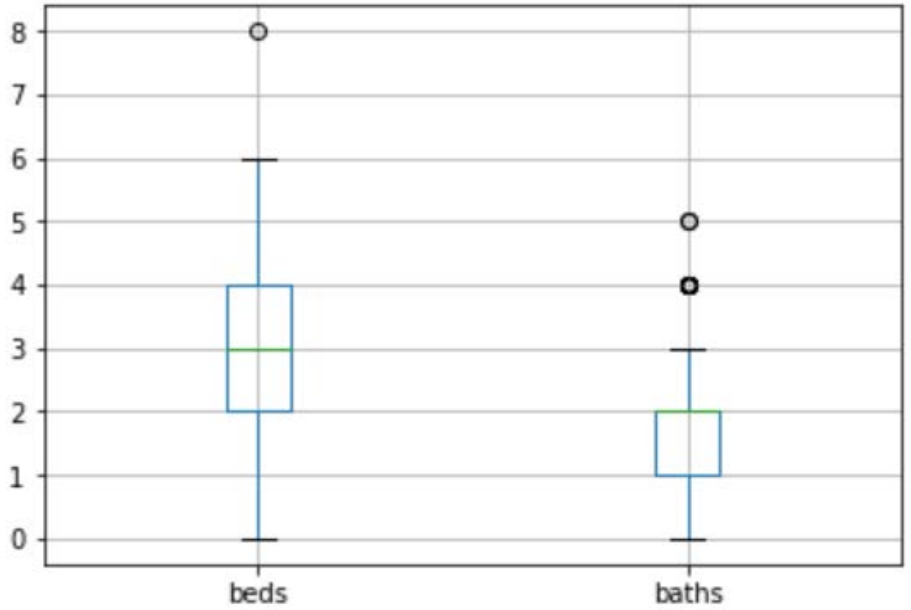
baths	0	1	2	3	4	5
zip						
95603	NaN	NaN	2.0	3.0	NaN	NaN
95608	NaN	4.0	15.0	1.0	NaN	NaN
95610	NaN	NaN	5.0	1.0	1.0	NaN
95614	NaN	NaN	1.0	NaN	NaN	NaN
95619	NaN	NaN	1.0	NaN	NaN	NaN
95621	NaN	7.0	20.0	1.0	NaN	NaN
95623	NaN	1.0	1.0	NaN	NaN	NaN
95624	3.0	1.0	22.0	8.0	NaN	NaN

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

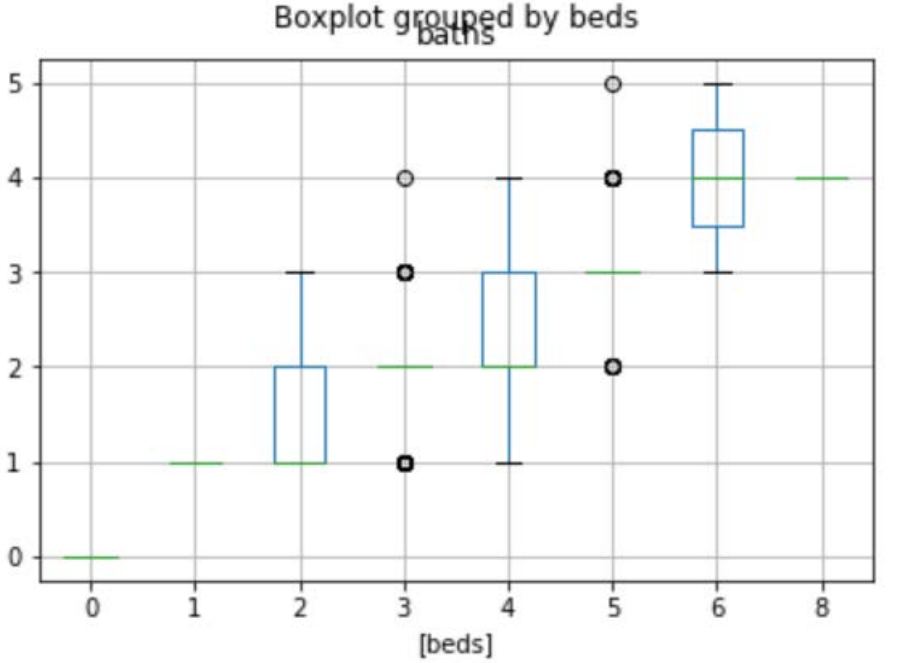
```
zip    baths
95603  2        2
       3        3
95608  1        4
       2       15
       3        1
95610  2        5
       3        1
       4        1
95614  2        1
95619  2        1
95621  1        7
       2       20
       3        1
```

Visualización

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




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



Visualización Matplotlib

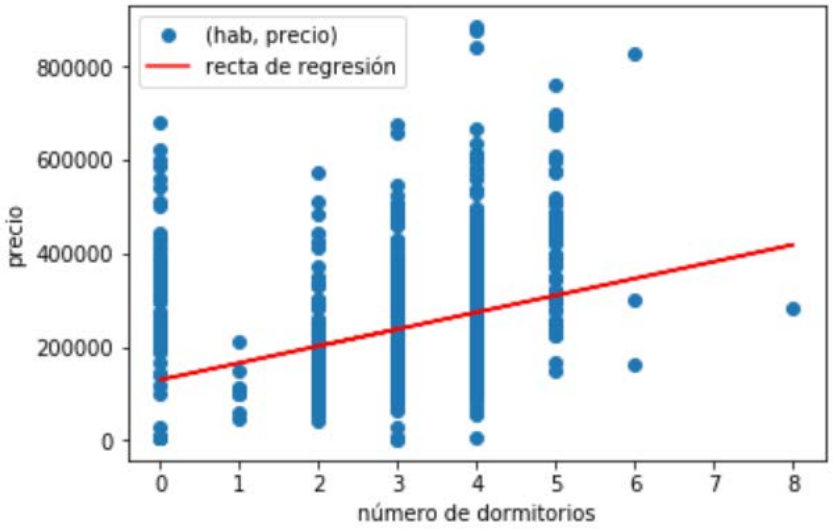
```
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt

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/