

# DataFrames, tablas de datos

Un dataframe es una estructura bidimensional, con filas y columnas, esto es, una tabla de datos. El `DataFrame` es el tipo de datos esencial de la librería `pandas`.

In [1]:

```
import pandas
```

Podemos construir un `DataFrame` de muchas formas. Uno de los más empleados es a partir de un archivo `csv`, que maneja esencialmente estas estructuras de tabla, con columnas que indican campos y filas que indican registros, cuyas celdas son los valores asociados a dichos campos.

In [2]:

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

Out[2]:

	street	city	zip	state	beds	baths	sq_ft	type	sale_date	price	lati
0	3526 HIGH ST	SACRAMENTO	95838	CA	2	1	836	Residential	Wed May 21 00:00:00 EDT 2008	59222	38.63
1	51 OMAHA CT	SACRAMENTO	95823	CA	3	1	1167	Residential	Wed May 21 00:00:00 EDT 2008	68212	38.47
2	2796 BRANCH ST	SACRAMENTO	95815	CA	2	1	796	Residential	Wed May 21 00:00:00 EDT 2008	68880	38.61
3	2805 JANETTE	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	60207	38.61

In [3]:

```
type(realestate)
```

Out[3]:

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

In [4]:



```
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
```

Vamos a comentar a continuación cómo podemos acceder a los diferentes elementos de la tabla

## Acceso a columnas

In [5]:



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

Out[5]:

0	59222
1	68212
2	68880
3	69307
4	81900
5	89921
6	90895
7	91002
8	94905
9	98937
10	100309
11	106250
12	106852
13	107502
14	108750
15	110700
16	113263
17	116250
18	120000
19	121630
20	122000
21	122682
22	123000
23	124100
24	125000
25	126640
26	127281
27	129000
28	131200
29	132000
	...
955	208250
956	208318
957	209347
958	211500
959	212000
960	213000
961	216000
962	216021
963	219000
964	219794
965	220000
966	220000
967	220000
968	220000
969	223000
970	224000
971	224000
972	224252
973	224500
974	225000
975	228000
976	229027
977	229500

```
978    230000
979    230000
980    232425
981    234000
982    235000
983    235301
984    235738
```

```
Name: price, Length: 985, dtype: int64
```

In [6]:



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

Out[6]:

	price	zip
0	59222	95838
1	68212	95823
2	68880	95815
3	69307	95815
4	81900	95824
5	89921	95841
6	90895	95842
7	91002	95820
8	94905	95670
9	98937	95673
10	100309	95838
11	106250	95823
12	106852	95815
13	107502	95822
14	108750	95842
15	110700	95842
16	113263	95673
17	116250	95621
18	120000	95833
19	121630	95673
20	122000	95621
21	122682	95823
22	123000	95660
23	124100	95834
24	125000	95660
25	126640	95843
26	127281	95820
27	129000	95832
28	131200	95827
29	132000	95828
...	...	...
955	208250	95747
956	208318	95660
957	209347	95632

	price	zip
958	211500	95624
959	212000	95843
960	213000	95828
961	216000	95624
962	216021	95828
963	219000	95610
964	219794	95757
965	220000	95758
966	220000	95678
967	220000	95608
968	220000	95632
969	223000	95842
970	224000	95834
971	224000	95821
972	224252	95660
973	224500	95682
974	225000	95662
975	228000	95757
976	229027	95648
977	229500	95823
978	230000	95823
979	230000	95758
980	232425	95829
981	234000	95823
982	235000	95610
983	235301	95758
984	235738	95762

985 rows × 2 columns

In [7]:

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

Out[7]:

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

In [8]:

```
realestate[selcols]
```

Out[8]:

	zip	state	beds	baths
0	95838	CA	2	1
1	95823	CA	3	1
2	95815	CA	2	1
3	95815	CA	2	1
4	95824	CA	2	1
5	95841	CA	3	1
6	95842	CA	3	2
7	95820	CA	3	1
8	95670	CA	2	2
9	95673	CA	3	2
10	95838	CA	3	2
11	95823	CA	3	2
12	95815	CA	1	1
13	95822	CA	3	1
14	95842	CA	2	2
15	95842	CA	2	2
16	95673	CA	2	1
17	95621	CA	2	1
18	95833	CA	2	1
19	95673	CA	3	2
20	95621	CA	3	2
21	95823	CA	4	2
22	95660	CA	4	2
23	95834	CA	3	2
24	95660	CA	3	2
25	95843	CA	3	2
26	95820	CA	3	1
27	95832	CA	3	2
28	95827	CA	4	2
29	95828	CA	3	2
...	...	...	...	...
955	95747	CA	3	2
956	95660	CA	3	1
957	95632	CA	3	2

	zip	state	beds	baths
958	95624	CA	4	2
959	95843	CA	3	2
960	95828	CA	3	2
961	95624	CA	3	2
962	95828	CA	3	1
963	95610	CA	3	2
964	95757	CA	3	2
965	95758	CA	3	2
966	95678	CA	3	2
967	95608	CA	4	2
968	95632	CA	4	2
969	95842	CA	4	2
970	95834	CA	0	0
971	95821	CA	3	1
972	95660	CA	3	1
973	95682	CA	3	2
974	95662	CA	3	1
975	95757	CA	3	2
976	95648	CA	3	2
977	95823	CA	4	2
978	95823	CA	4	2
979	95758	CA	3	2
980	95829	CA	4	3
981	95823	CA	3	2
982	95610	CA	3	2
983	95758	CA	4	2
984	95762	CA	3	2

985 rows × 4 columns

In [9]:

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

Out[9]:

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



In [10]:



realestate[mixed]

Out[10]:

	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
5	95841	CA	3	1	89921
6	95842	CA	3	2	90895
7	95820	CA	3	1	91002
8	95670	CA	2	2	94905
9	95673	CA	3	2	98937
10	95838	CA	3	2	100309
11	95823	CA	3	2	106250
12	95815	CA	1	1	106852
13	95822	CA	3	1	107502
14	95842	CA	2	2	108750
15	95842	CA	2	2	110700
16	95673	CA	2	1	113263
17	95621	CA	2	1	116250
18	95833	CA	2	1	120000
19	95673	CA	3	2	121630
20	95621	CA	3	2	122000
21	95823	CA	4	2	122682
22	95660	CA	4	2	123000
23	95834	CA	3	2	124100
24	95660	CA	3	2	125000
25	95843	CA	3	2	126640
26	95820	CA	3	1	127281
27	95832	CA	3	2	129000
28	95827	CA	4	2	131200
29	95828	CA	3	2	132000
...	...	...	...	...	...
955	95747	CA	3	2	208250
956	95660	CA	3	1	208318
957	95632	CA	3	2	209347

	zip	state	beds	baths	price
<b>958</b>	95624	CA	4	2	211500
<b>959</b>	95843	CA	3	2	212000
<b>960</b>	95828	CA	3	2	213000
<b>961</b>	95624	CA	3	2	216000
<b>962</b>	95828	CA	3	1	216021
<b>963</b>	95610	CA	3	2	219000
<b>964</b>	95757	CA	3	2	219794
<b>965</b>	95758	CA	3	2	220000
<b>966</b>	95678	CA	3	2	220000
<b>967</b>	95608	CA	4	2	220000
<b>968</b>	95632	CA	4	2	220000
<b>969</b>	95842	CA	4	2	223000
<b>970</b>	95834	CA	0	0	224000
<b>971</b>	95821	CA	3	1	224000
<b>972</b>	95660	CA	3	1	224252
<b>973</b>	95682	CA	3	2	224500
<b>974</b>	95662	CA	3	1	225000
<b>975</b>	95757	CA	3	2	228000
<b>976</b>	95648	CA	3	2	229027
<b>977</b>	95823	CA	4	2	229500
<b>978</b>	95823	CA	4	2	230000
<b>979</b>	95758	CA	3	2	230000
<b>980</b>	95829	CA	4	3	232425
<b>981</b>	95823	CA	3	2	234000
<b>982</b>	95610	CA	3	2	235000
<b>983</b>	95758	CA	4	2	235301
<b>984</b>	95762	CA	3	2	235738

985 rows × 5 columns

## Acceso a filas

In [11]:

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

Out[11]:

```
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
```

In [12]:

```
realestate.iloc[[3,6]]
```

Out[12]:

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	69307
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895

In [13]:

```
realestate.iloc[3:6]
```

Out[13]:

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	pric
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	6930
4	6001 MCMAHON DR	SACRAMENTO	95824	CA	2	1	797	Residential	Wed May 21 00:00:00 EDT 2008	8190
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3	1	1122	Condo	Wed May 21 00:00:00 EDT 2008	8992

In [14]:

```
list(range(3,8,2))
```

Out[14]:

```
[3, 5, 7]
```

In [15]:

```
realestate.iloc[range(3,8,2)]
```

Out[15]:

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	pric
3	2805 JANETTE WAY	SACRAMENTO	95815	CA	2	1	852	Residential	Wed May 21 00:00:00 EDT 2008	6930
5	5828 PEPPERMILL CT	SACRAMENTO	95841	CA	3	1	1122	Condo	Wed May 21 00:00:00 EDT 2008	8992
7	2561 19TH AVE	SACRAMENTO	95820	CA	3	1	1177	Residential	Wed May 21 00:00:00 EDT 2008	9100

In [16]:



```
realestate.iloc[3]['city']
```

Out[16]:

```
'SACRAMENTO'
```

In [17]:



```
realestate.iloc[3][1]
```

Out[17]:

```
'SACRAMENTO'
```

In [18]:



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

Out[18]:

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

## Operaciones con Dataframes

El módulo `pandas` nos ofrece una amplísima variedad de funciones para trabajar con los `DataFrame`, a continuación, a modo de ejemplo te mostramos algunas.

In [19]:



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

Out[19]:

	street	city	zip	state	beds	baths	sq_ft	type	sale_date
864	9401 BARREL RACER CT	WILTON	95693	CA	4	3	4400	Residential	Fri May 16 00:00:00 EDT 2008
863	2982 ABERDEEN LN	EL DORADO HILLS	95762	CA	4	3	0	Residential	Fri May 16 00:00:00 EDT 2008
334	3935 EL MONTE DR	LOOMIS	95650	CA	4	4	1624	Residential	Tue May 20 00:00:00 EDT 2008
157	315 JUMEL CT	EL DORADO HILLS	95762	CA	6	5	0	Residential	Wed May 21 00:00:00 EDT 2008
553	6222 CALLE MONTALVO CIR	GRANITE BAY	95746	CA	5	3	3670	Residential	Mon May 19 00:00:00 EDT 2008
333	1409 47TH ST	SACRAMENTO	95819	CA	5	2	2484	Residential	Tue May 20 00:00:00 EDT 2008
552	12901 FURLONG DR	WILTON	95693	CA	5	3	3788	Residential	Mon May 19 00:00:00 EDT 2008
332	4128 HILL ST	FAIR OAKS	95628	CA	5	5	2846	Residential	Tue May 20 00:00:00 EDT 2008
862	2065 IMPRESSIONIST WAY	EL DORADO HILLS	95762	CA	0	0	0	Residential	Fri May 16 00:00:00 EDT 2008
551	2400 COUNTRYSIDE DR	PLACERVILLE	95667	CA	3	2	2025	Residential	Mon May 19 00:00:00 EDT 2008
550	4629 DORCHESTER LN	GRANITE BAY	95746	CA	5	3	2896	Residential	Mon May 19 00:00:00 EDT 2008
549	4734 GIBBONS DR	CARMICHAEL	95608	CA	4	3	3357	Residential	Mon May 19 00:00:00 EDT 2008
156	4004 CRESTA WAY	SACRAMENTO	95864	CA	3	3	2325	Residential	Wed May 21 00:00:00 EDT 2008

	street	city	zip	state	beds	baths	sq_ft	type	sale_date
548	104 CATLIN CT	FOLSOM	95630	CA	4	3	2660	Residential	Mon May 19 00:00:00 EDT 2008
861	200 CRADLE MOUNTAIN CT	EL DORADO HILLS	95762	CA	0	0	0	Residential	Fri May 16 00:00:00 EDT 2008
331	8616 ROCKPORTE CT	ROSEVILLE	95747	CA	4	2	0	Residential	Tue May 20 00:00:00 EDT 2008
330	2981 WRINGER DR	ROSEVILLE	95661	CA	4	3	3838	Residential	Tue May 20 00:00:00 EDT 2008
860	8432 BRIGGS DR	ROSEVILLE	95747	CA	5	3	3579	Residential	Fri May 16 00:00:00 EDT 2008
155	4070 REDONDO DR	EL DORADO HILLS	95762	CA	4	3	0	Residential	Wed May 21 00:00:00 EDT 2008
859	4478 GREENBRAE RD	ROCKLIN	95677	CA	0	0	0	Residential	Fri May 16 00:00:00 EDT 2008
154	6030 PALERMO WAY	EL DORADO HILLS	95762	CA	4	3	0	Residential	Wed May 21 00:00:00 EDT 2008
857	4620 BROMWICH CT	ROCKLIN	95677	CA	4	3	0	Residential	Fri May 16 00:00:00 EDT 2008
858	620 KESWICK CT	GRANITE BAY	95746	CA	4	3	2356	Residential	Fri May 16 00:00:00 EDT 2008
856	1800 AVONDALE DR	ROSEVILLE	95747	CA	5	3	0	Residential	Fri May 16 00:00:00 EDT 2008
855	9880 IZILDA CT	SACRAMENTO	95829	CA	5	4	3863	Residential	Fri May 16 00:00:00 EDT 2008
153	868 HILDEBRAND CIR	FOLSOM	95630	CA	0	0	0	Residential	Wed May 21 00:00:00 EDT 2008
329	5709 RIVER OAK WAY	CARMICHAEL	95608	CA	4	2	2222	Residential	Tue May 20 00:00:00 EDT 2008

	street	city	zip	state	beds	baths	sq_ft	type	sale_date
547	9360 MAGOS RD	WILTON	95693	CA	5	2	3741	Residential	Mon May 19 00:00:00 EDT 2008
853	1165 37TH ST	SACRAMENTO	95816	CA	2	1	1252	Residential	Fri May 16 00:00:00 EDT 2008
854	203 CASCADE FALLS DR	FOLSOM	95630	CA	4	3	3229	Residential	Fri May 16 00:00:00 EDT 2008
...	...	...	...	...	...	...	...	...	..
565	76 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
559	40 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
600	7 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
599	15 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
587	108 CRYSTALWOOD WAY	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
598	19 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
597	23 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
596	27 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
595	31 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
594	35 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
593	39 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008



	street	city	zip	state	beds	baths	sq_ft	type	sale_date
592	43 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
591	47 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
590	51 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
589	55 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
588	100 CRYSTALWOOD WAY	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
586	116 CRYSTALWOOD WAY	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
602	3 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
585	374 1ST ST	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
584	386 1ST ST	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
583	398 1ST ST	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
582	63 CRYSTALWOOD CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
572	434 1ST ST	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
581	75 E ST	LINCOLN	95648	CA	3	2	0	Residential	Mon May 19 00:00:00 EDT 2008
580	59 E ST	LINCOLN	95648	CA	3	2	0	Residential	Mon May 19 00:00:00 EDT 2008

	street	city	zip	state	beds	baths	sq__ft	type	sale_date
579	51 E ST	LINCOLN	95648	CA	4	2	0	Residential	Mon May 19 00:00:00 EDT 2008
601	CRYSTALWOOD 7 CIR	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
573	3 E ST	LINCOLN	95648	CA	0	0	0	Residential	Mon May 19 00:00:00 EDT 2008
866	14151 INDIO DR	SLOUGHHOUSE	95683	CA	3	4	5822	Residential	Fri May 16 00:00:00 EDT 2008
865	3720 VISTA DE MADERA	LINCOLN	95648	CA	3	3	0	Residential	Fri May 16 00:00:00 EDT 2008

985 rows × 12 columns

In [20]:

```
realestate.describe()
```

Out[20]:

	zip	beds	baths	sq__ft	price	latitude	longitu
count	985.000000	985.000000	985.000000	985.000000	985.000000	985.000000	985.0000
mean	95750.697462	2.911675	1.776650	1314.916751	234144.263959	38.607732	-121.3559
std	85.176072	1.307932	0.895371	853.048243	138365.839085	0.145433	0.1382
min	95603.000000	0.000000	0.000000	0.000000	1551.000000	38.241514	-121.5517
25%	95660.000000	2.000000	1.000000	952.000000	145000.000000	38.482717	-121.4467
50%	95762.000000	3.000000	2.000000	1304.000000	213750.000000	38.626582	-121.3762
75%	95828.000000	4.000000	2.000000	1718.000000	300000.000000	38.695589	-121.2957
max	95864.000000	8.000000	5.000000	5822.000000	884790.000000	39.020808	-120.5975

In [21]:



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

Out[21]:

```
0      False
1      False
2      False
3      False
4      False
5      False
6       True
7      False
8       True
9       True
10      True
11      True
12     False
13     False
14       True
15       True
16     False
17     False
18     False
19       True
20       True
21       True
22       True
23       True
24       True
25       True
26     False
27       True
28       True
29       True
...
955     True
956    False
957     True
958     True
959     True
960     True
961     True
962    False
963     True
964     True
965     True
966     True
967     True
968     True
969     True
970    False
971    False
972    False
973     True
974    False
975     True
976     True
```

```

977     True
978     True
979     True
980    False
981     True
982     True
983     True
984     True

```

Name: baths, Length: 985, dtype: bool

In [22]:

```
realestate[cond]
```

Out[22]:

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price	lati
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895	38.68
8	11150 TRINITY RIVER DR Unit 114	RANCHO CORDOVA	95670	CA	2	2	941	Condo	Wed May 21 00:00:00 EDT 2008	94905	38.62
9	7325 10TH ST	RIO LINDA	95673	CA	3	2	1146	Residential	Wed May 21 00:00:00 EDT 2008	98937	38.70
10	645 MORRISON	SACRAMENTO	95838	CA	3	2	909	Residential	Wed May 21 00:00:00 EDT 2008	100309	38.63

In [23]:

```

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

```

Out[23]:

	street	city	zip	state	beds	baths	sq__ft	type	sale_date	price	lati
6	6048 OGDEN NASH WAY	SACRAMENTO	95842	CA	3	2	1104	Residential	Wed May 21 00:00:00 EDT 2008	90895	38.68
9	7325 10TH ST	RIO LINDA	95673	CA	3	2	1146	Residential	Wed May 21 00:00:00 EDT 2008	98937	38.70
10	645 MORRISON AVE	SACRAMENTO	95838	CA	3	2	909	Residential	Wed May 21 00:00:00 EDT 2008	100309	38.63
11	4085 FAWN CIR	SACRAMENTO	95823	CA	3	2	1289	Residential	Wed May 21 00:00:00 EDT 2008	106250	38.47

In [24]:

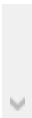


```
realestate[two_baths & three_beds]['zip'].value_counts()
```

Out[24]:

95823	33
95828	19
95843	17
95838	16
95757	14
95758	14
95678	12
95835	12
95621	11
95822	9
95632	9
95624	8
95832	8
95660	8
95834	8
95608	7
95670	7
95833	6
95630	6
95762	6
95648	6
95667	5
95673	5
95826	5
95842	5
95610	5
95827	4
95747	4
95662	4
95626	3
95628	3
95829	3
95682	3
95765	2
95661	2
95831	2
95677	2
95820	1
95819	1
95821	1
95817	1
95815	1
95726	1
95824	1
95825	1
95841	1
95693	1
95614	1
95619	1
95623	1
95633	1
95635	1
95655	1

```
95663      1
95864      1
Name: zip, dtype: int64
```



In [25]:

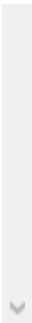


```
realestate.groupby(['zip', 'baths']).size()
```

Out[25]:

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
95623	1	1
	2	1
95624	0	3
	1	1
	2	22
	3	8
95626	2	4
95628	1	3
	2	5
	5	1
95630	0	2
	2	9
	3	6
95631	0	1
95632	0	1
	1	2
	2	15
	..	
95832	1	1
	2	9
	3	2
95833	0	3
	1	4
	2	12
	3	1
95834	0	5
	2	12
	3	5
95835	0	3
	1	2
	2	22
	3	9
	4	1
95838	0	1
	1	11
	2	24
	3	1
95841	1	4
	2	3
95842	1	7

```
      2      14
      3       1
95843 1       1
      2      24
      3       8
95864 1       3
      2       1
      3       1
Length: 185, dtype: int64
```





In [26]:

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

Out[26]:

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
95626	NaN	NaN	4.0	NaN	NaN	NaN
95628	NaN	3.0	5.0	NaN	NaN	1.0
95630	2.0	NaN	9.0	6.0	NaN	NaN
95631	1.0	NaN	NaN	NaN	NaN	NaN
95632	1.0	2.0	15.0	2.0	1.0	NaN
95633	NaN	NaN	1.0	NaN	NaN	NaN
95635	NaN	NaN	1.0	NaN	NaN	NaN
95648	51.0	NaN	15.0	5.0	1.0	NaN
95650	NaN	NaN	1.0	NaN	1.0	NaN
95655	NaN	NaN	1.0	NaN	NaN	NaN
95660	NaN	9.0	12.0	NaN	NaN	NaN
95661	NaN	3.0	3.0	2.0	NaN	NaN
95662	NaN	2.0	8.0	NaN	1.0	NaN
95663	NaN	NaN	1.0	NaN	NaN	NaN
95667	1.0	1.0	6.0	2.0	NaN	NaN
95670	1.0	3.0	13.0	3.0	1.0	NaN
95673	1.0	4.0	8.0	NaN	NaN	NaN
95677	1.0	NaN	3.0	2.0	NaN	NaN
95678	1.0	3.0	13.0	3.0	NaN	NaN
95682	2.0	1.0	7.0	NaN	NaN	NaN
95683	1.0	NaN	1.0	1.0	1.0	NaN
95690	NaN	1.0	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...
95758	2.0	2.0	34.0	6.0	NaN	NaN
95762	3.0	NaN	10.0	8.0	1.0	1.0

baths	0	1	2	3	4	5
zip						
95765	3.0	NaN	6.0	2.0	NaN	NaN
95811	NaN	1.0	1.0	NaN	NaN	NaN
95814	NaN	NaN	NaN	3.0	NaN	NaN
95815	NaN	12.0	6.0	NaN	NaN	NaN
95816	1.0	2.0	NaN	1.0	NaN	NaN
95817	NaN	5.0	2.0	NaN	NaN	NaN
95818	NaN	6.0	1.0	NaN	NaN	NaN
95819	NaN	2.0	2.0	NaN	NaN	NaN
95820	NaN	18.0	5.0	NaN	NaN	NaN
95821	NaN	4.0	2.0	NaN	NaN	NaN
95822	NaN	10.0	14.0	NaN	NaN	NaN
95823	NaN	7.0	50.0	4.0	NaN	NaN
95824	NaN	10.0	2.0	NaN	NaN	NaN
95825	1.0	7.0	5.0	NaN	NaN	NaN
95826	NaN	7.0	9.0	1.0	1.0	NaN
95827	NaN	NaN	7.0	1.0	1.0	NaN
95828	NaN	5.0	34.0	3.0	3.0	NaN
95829	NaN	NaN	5.0	5.0	1.0	NaN
95831	1.0	1.0	6.0	2.0	NaN	NaN
95832	NaN	1.0	9.0	2.0	NaN	NaN
95833	3.0	4.0	12.0	1.0	NaN	NaN
95834	5.0	NaN	12.0	5.0	NaN	NaN
95835	3.0	2.0	22.0	9.0	1.0	NaN
95838	1.0	11.0	24.0	1.0	NaN	NaN
95841	NaN	4.0	3.0	NaN	NaN	NaN
95842	NaN	7.0	14.0	1.0	NaN	NaN
95843	NaN	1.0	24.0	8.0	NaN	NaN
95864	NaN	3.0	1.0	1.0	NaN	NaN

68 rows × 6 columns

# Visualización

In [27]:

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

Out[27]:

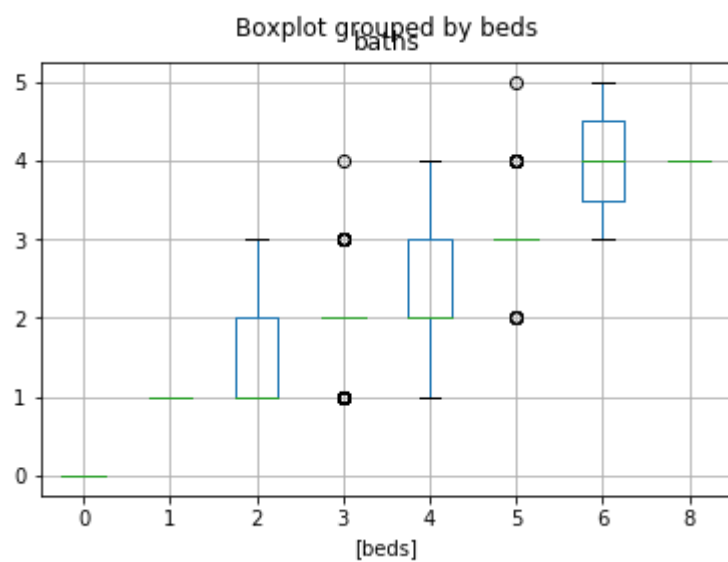
<matplotlib.axes.\_subplots.AxesSubplot at 0x268213fffd0>

In [28]:

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

Out[28]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x26821592f28>



## Matplotlib

In [29]:



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

Out[29]:

	beds	price
0	2	59222
1	3	68212
2	2	68880
3	2	69307
4	2	81900
5	3	89921
6	3	90895
7	3	91002
8	2	94905
9	3	98937
10	3	100309
11	3	106250
12	1	106852
13	3	107502
14	2	108750
15	2	110700
16	2	113263
17	2	116250
18	2	120000
19	3	121630
20	3	122000
21	4	122682
22	4	123000
23	3	124100
24	3	125000
25	3	126640
26	3	127281
27	3	129000
28	4	131200
29	3	132000
...	...	...
955	3	208250
956	3	208318

	<b>beds</b>	<b>price</b>
<b>957</b>	3	209347
<b>958</b>	4	211500
<b>959</b>	3	212000
<b>960</b>	3	213000
<b>961</b>	3	216000
<b>962</b>	3	216021
<b>963</b>	3	219000
<b>964</b>	3	219794
<b>965</b>	3	220000
<b>966</b>	3	220000
<b>967</b>	4	220000
<b>968</b>	4	220000
<b>969</b>	4	223000
<b>970</b>	0	224000
<b>971</b>	3	224000
<b>972</b>	3	224252
<b>973</b>	3	224500
<b>974</b>	3	225000
<b>975</b>	3	228000
<b>976</b>	3	229027
<b>977</b>	4	229500
<b>978</b>	4	230000
<b>979</b>	3	230000
<b>980</b>	4	232425
<b>981</b>	3	234000
<b>982</b>	3	235000
<b>983</b>	4	235301
<b>984</b>	3	235738

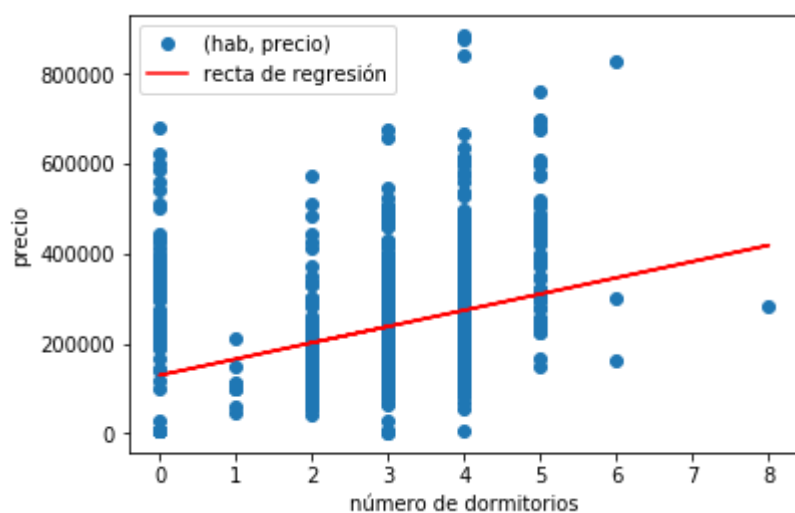
985 rows × 2 columns

In [30]:

```
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()
```



## Seaborn

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

In [31]:

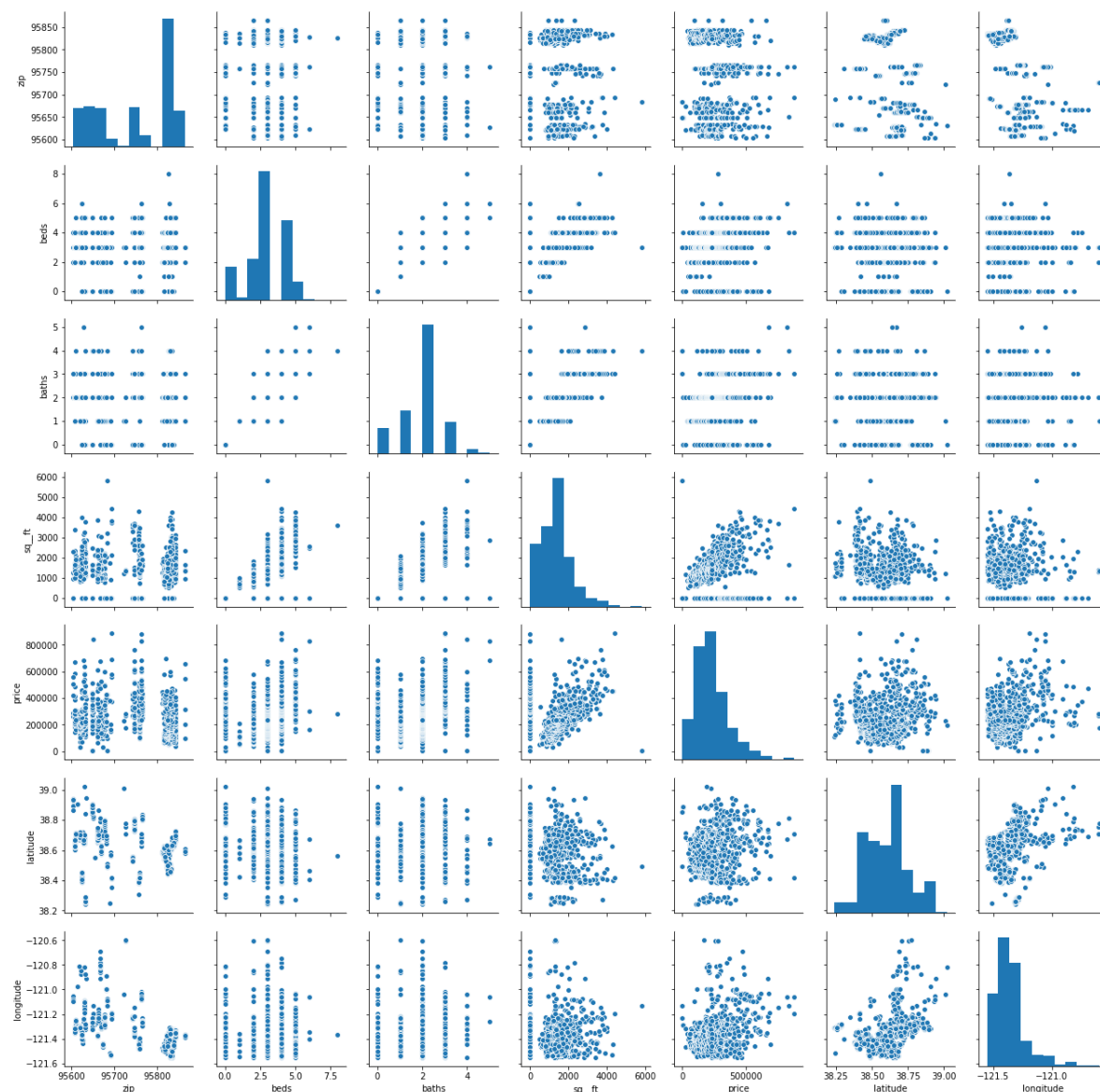
```
import seaborn
```

In [32]:

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

Out[32]:

&lt;seaborn.axisgrid.PairGrid at 0x2682407dcf8&gt;



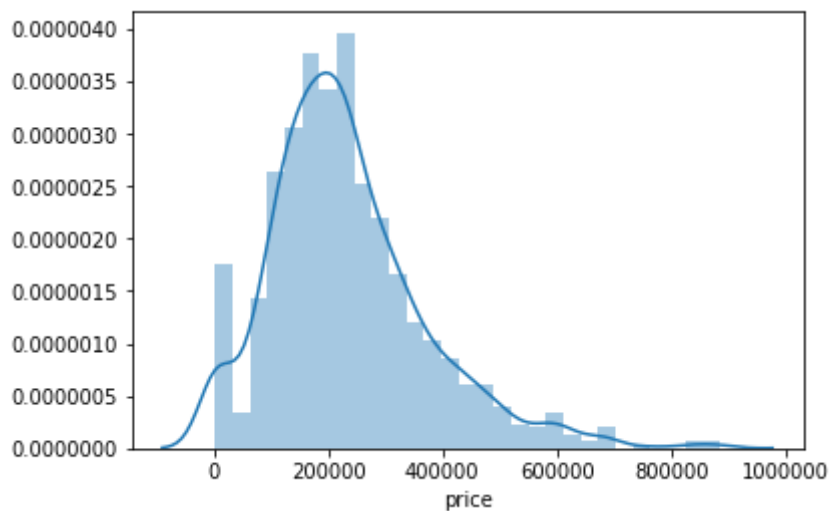
In [33]:

```
price = realestate["price"]  
seaborn.distplot(price)
```



Out[33]:

&lt;matplotlib.axes.\_subplots.AxesSubplot at 0x268254f0a58&gt;



## Referencias

Uno de los objetivos de pandas es dar una funcionalidad de análisis de datos parecida a la proporcionada por R. Podemos encontrar una comparación en la documentación de pandas: [https://pandas.pydata.org/pandas-docs/stable/comparison\\_with\\_r.html](https://pandas.pydata.org/pandas-docs/stable/comparison_with_r.html) ([https://pandas.pydata.org/pandas-docs/stable/comparison\\_with\\_r.html](https://pandas.pydata.org/pandas-docs/stable/comparison_with_r.html))

- <http://pandas.pydata.org/pandas-docs/stable/> (<http://pandas.pydata.org/pandas-docs/stable/>)