



Shahid Beheshti  
University

گزارش سری تمرین 1

واحد درسی داده کاوی دوره الکترونیک

جناب آقای دکتر خردپیشه

رحیم اکبری 99422028

00/01/21

## فهرست مطالب

- (3) - مجموعه داده ها و بررسی ویژگی آنها
- (4) - بررسی تعداد خانه ها و بررسی قیمت ها
- (5) - نمودار توزیع قیمت
- (6) - بررسی بالاترین قیمت منطقه
- (7) - رابطه بین اتاق ها و قیمت ها
- (8) - بررسی تعداد خانه های هر نوع اتاق
- (9) - نمودار میزبانهایی که بیشتر از بقیه خانه هایشان به اجاره رفته

## مجموعه داده های و ویژگی آنها

مجموعه داده ای که در اینجا در نظر گرفته ایم شامل لیستی از خانه های اجاره ای به همراه تعدادی ویژگی می-باشد. در تصویر زیر می-توانیم قسمتی از این مجموعه داده را مشاهده نماییم.

id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_revi
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1
1	2595	Skyliit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1
2	3647	THE VILLAGE OF HARLEM....NEW YORK!	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10

تعدادی از ویژگی های مجموعه داده شامل آیدی خانه، نام خانه، منطقه، قیمت و غیره می-باشد. ابتدا ستون هایی که داده های عددی دارند را جدا کرده و پارامترهای آماری آنها نظیر میانگین، مینیمم، ماکزیمم، انحراف معیار و چارک ها را بررسی می-کنیم.

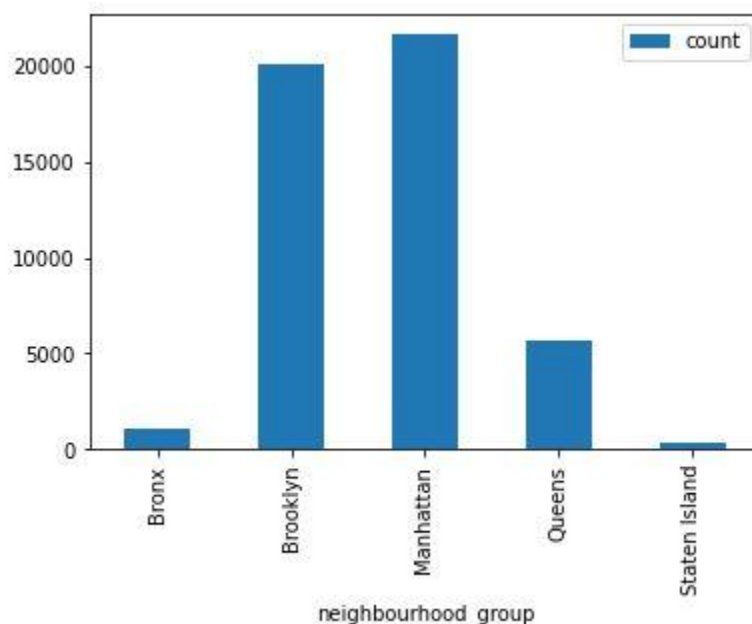
	price	latitude	longitude	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
count	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	38843.000000	48895.000000	48895.000000
mean	152.720687	40.728949	-73.952170	7.029962	23.274466	1.373221	7.143982	112.781327
std	240.154170	0.054530	0.046157	20.510550	44.550582	1.680442	32.952519	131.622289
min	0.000000	40.499790	-74.244420	1.000000	0.000000	0.010000	1.000000	0.000000
25%	69.000000	40.690100	-73.983070	1.000000	1.000000	0.190000	1.000000	0.000000
50%	106.000000	40.723070	-73.955680	3.000000	5.000000	0.720000	1.000000	45.000000
75%	175.000000	40.763115	-73.936275	5.000000	24.000000	2.020000	2.000000	227.000000
max	10000.000000	40.913060	-73.712990	1250.000000	629.000000	58.500000	327.000000	365.000000

## بررسی تعداد خانه ها و بررسی قیمت ها

در مرحله بعد وجود داده‌های از دست‌رفته را بررسی می‌کنیم. پس از آن داده‌های آماری ستون قیمت را چک می‌کنیم. در ستون قیمت می‌بینیم که قیمت صفر نیز وجود دارد. رکوردهایی که قیمت صفر دارند را حذف می‌کنیم.

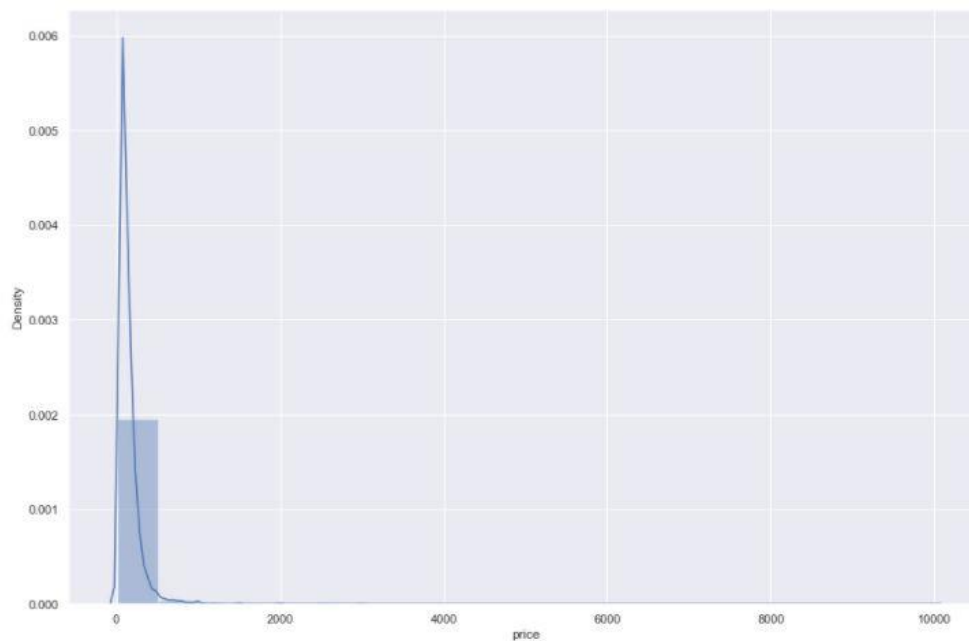
```
count    48895.000000
mean      152.720687
std       240.154170
min        0.000000
25%        69.000000
50%       106.000000
75%       175.000000
max      10000.000000
Name: price, dtype: float64
```

تعداد خانه‌ها را بر حسب هر منطقه می‌شماریم. نمودار زیر نشان می‌دهد که منطقه منهتن و بروکلین بیشترین تعداد خانه را دارند.

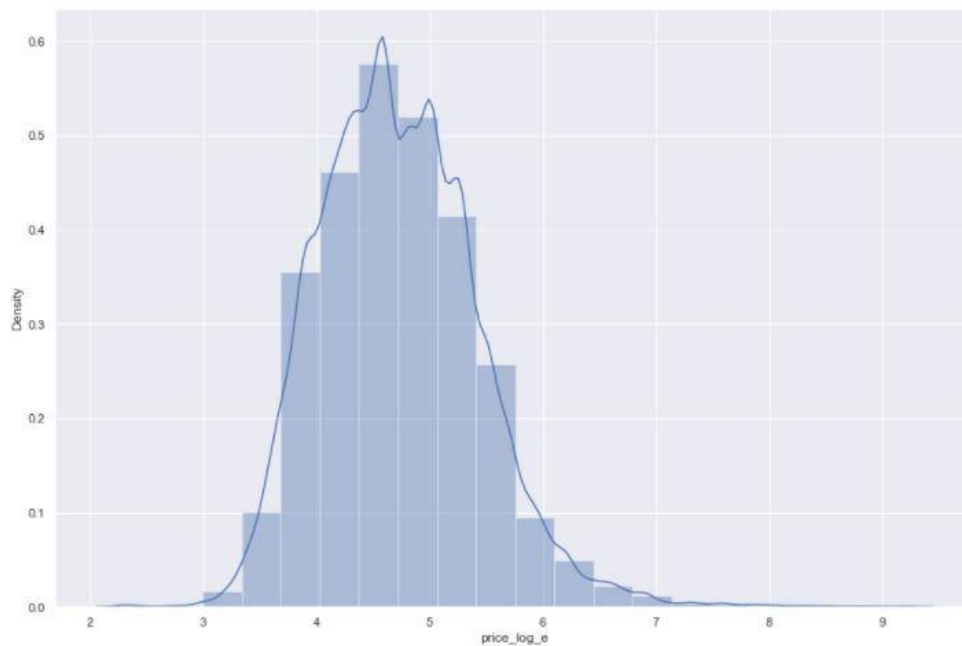


## نمودار توزیع قیمت

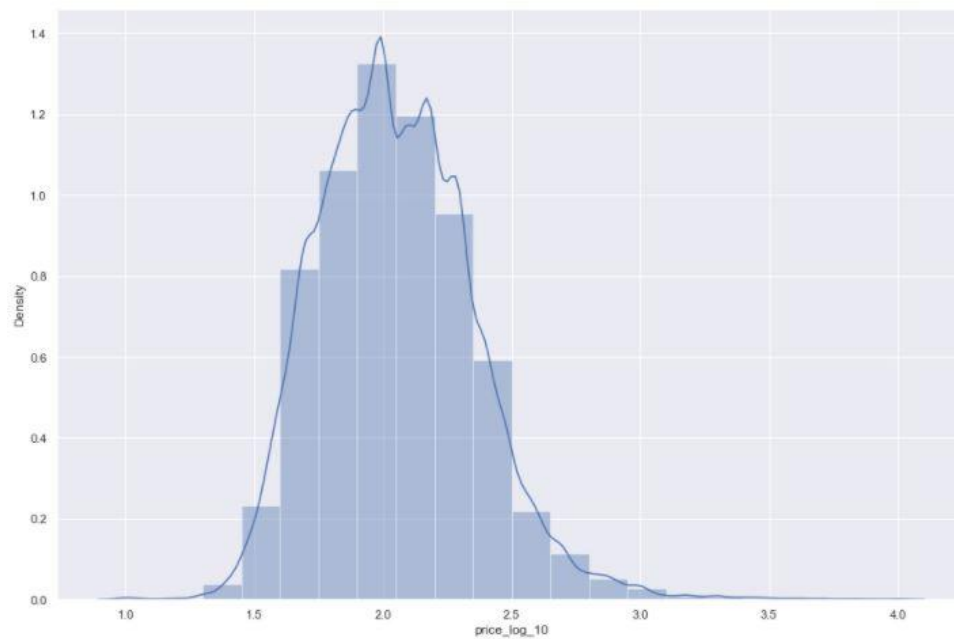
نمودار توزیع قیمت را رسم می‌کنیم. نمودار زیر نشان می‌دهد که توزیع قیمت نرمال نیست.



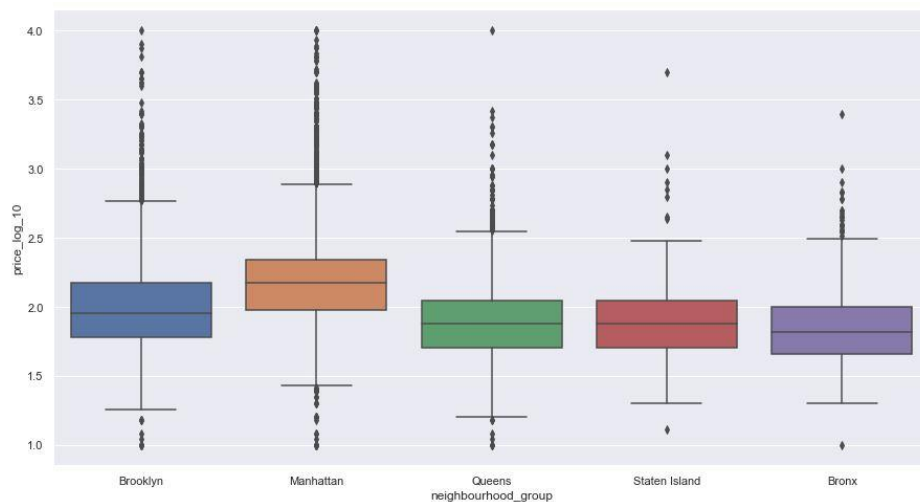
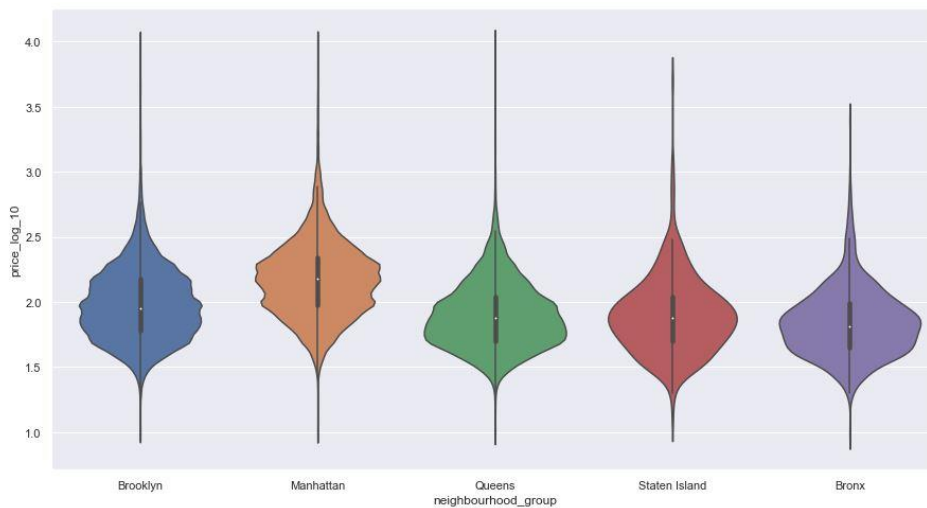
توزیع قیمت را نرمال می‌کنیم و سپس نمودار آن را رسم می‌نماییم.



سپس رابطه بین مناطق مختلف و قیمت را بررسی می‌نماییم.



## برسی بالاترین قیمت منطقه

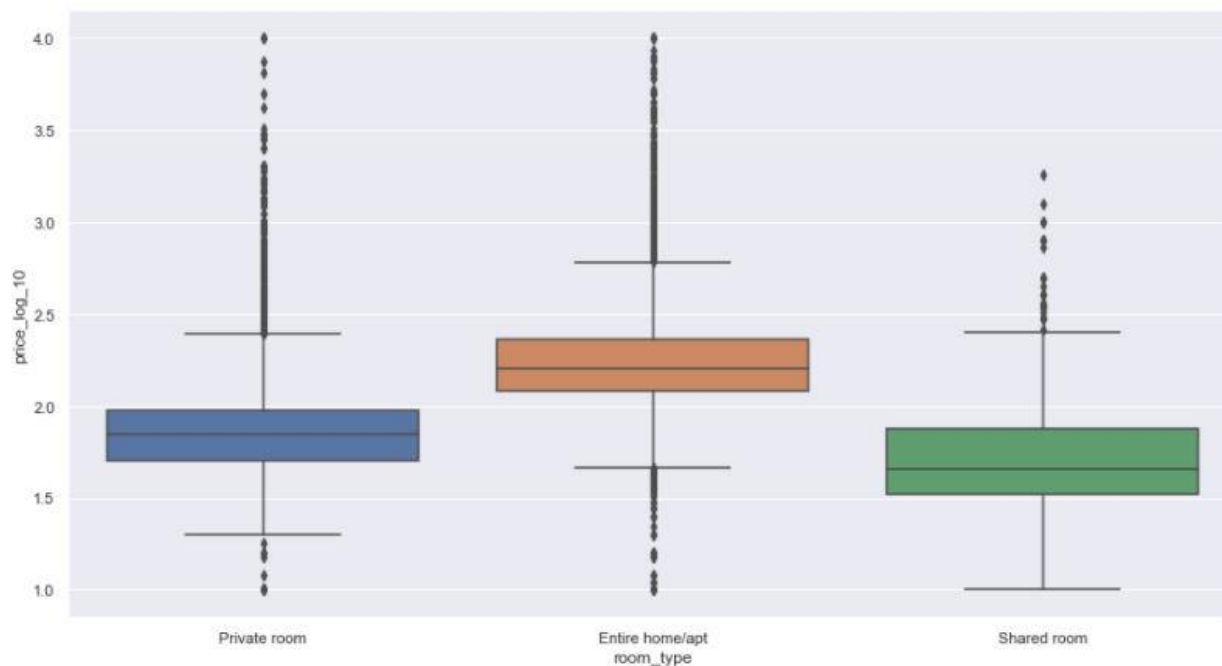
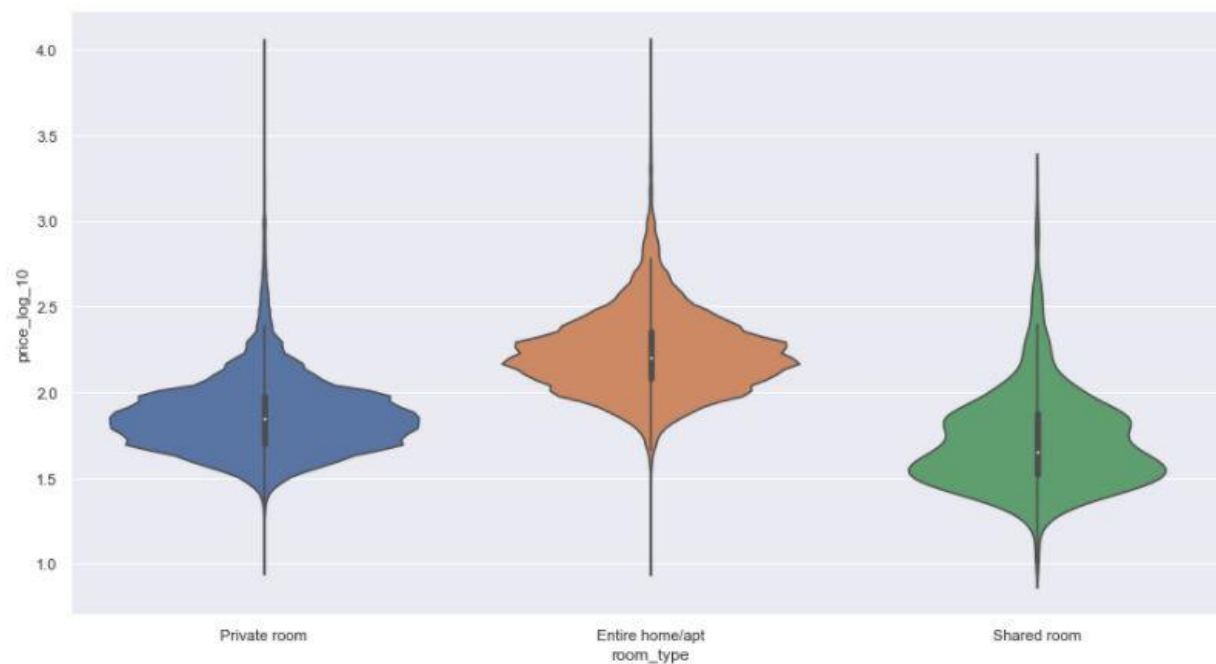


neighbourhood_group	price							
	count	mean	std	min	25%	50%	75%	max
Bronx	1090.0	87.577064	106.725371	10.0	45.0	65.0	99.0	2500.0
Brooklyn	20095.0	124.438915	186.896837	10.0	60.0	90.0	150.0	10000.0
Manhattan	21660.0	196.884903	291.386838	10.0	95.0	150.0	220.0	10000.0
Queens	5666.0	99.517649	167.102155	10.0	50.0	75.0	110.0	10000.0
Staten Island	373.0	114.812332	277.620403	13.0	50.0	75.0	110.0	5000.0

نمودارها و جدول بالا نشان می‌دهد که منطقه منهتن و بروکلین قیمت‌های بالاتری نسبت به سایر مناطق دارند و احتمالاً بخاطر همین مالکان بیشتری را ترغیب به راه‌اندازی خانه‌های اجاره‌ای کرده است.

در ادامه رابطه بین نوع اتاق‌ها و قیمت را بررسی می‌کنیم.

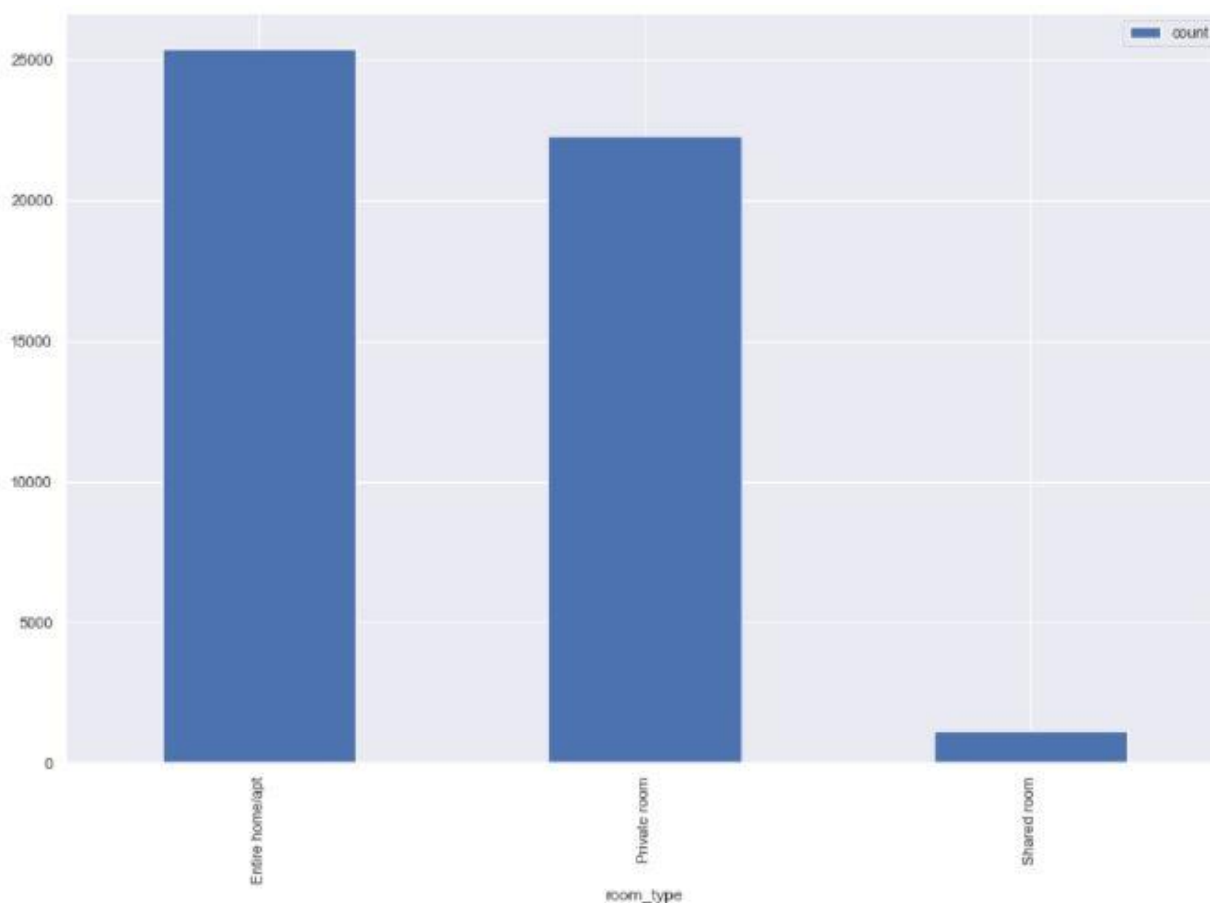
## رابطه بین نوع اتاق ها و قیمت ها



نمودارهای بالا نشان می‌دهد که نوع اتاق entire home قیمت بالاتری نسبت به دو نوع دیگر دارد. حالا تعداد خانه‌های هر نوع را بررسی می‌کنیم.



## برسی تعداد خانه های هر نوع اتاق ها



نمودار بالا نشان می‌دهد که خانه‌هایی که اتاق نوع entire home دارند تعداد بیشتری هستند. پس از آن نوع private room و بعد هم نوع shared room.

## میزبان هایی که بیشتر از بقیه خانه هایشان به اجاره رفته است

در ادامه نیز ده میزبانی که بیشتر از بقیه خانه هایشان اجاره رفته است را در نمودار زیر می بینیم.

