

میهمانان و میزبانان فراوانی از Airbnb برای گسترش امکان سفر استفاده می کنند. امروز ، Airbnb به یکی از خدماتی تبدیل شده که توسط تمام دنیا مورد استفاده قرار می گیرد. تجزیه و تحلیل داده ها در مورد میلیون ها لیست ارائه شده از طریق Airbnb یک عامل حیاتی برای این شرکت است. این میلیون ها لیست داده های زیادی تولید می کنند - داده هایی که می توانند برای امنیت ، تصمیمات تجاری ، درک رفتار و عملکرد مشتریان و ارائه دهندگان (میزبان) ، راهنمایی برای ابتکارات بازاریابی ، اجرای خدمات اضافی نوآورانه و موارد خیلی بیشتر در سیستم تجزیه و تحلیل و استفاده شوند

این مجموعه داده در حدود 49,000 مورد مشاهده با 16 ستون ویژگی در خود دارد و ترکیبی است از مقادیر ترکیبی و عددی.

<i>id</i>	<i>int64</i>
<i>name</i>	<i>object</i>
<i>host_id</i>	<i>int64</i>
<i>host_name</i>	<i>object</i>
<i>neighbourhood_group</i>	<i>object</i>
<i>neighbourhood</i>	<i>object</i>
<i>latitude</i>	<i>float64</i>

<i>longitude</i>	<i>float64</i>
<i>room_type</i>	<i>object</i>
<i>price</i>	<i>int64</i>
<i>minimum_nights</i>	<i>int64</i>
<i>number_of_reviews</i>	<i>int64</i>
<i>last_review</i>	<i>object</i>
<i>reviews_per_month</i>	<i>float64</i>
<i>calculated_host_listings_count</i>	<i>int64</i>
<i>availability_365</i>	<i>int64</i>
	<i>dtype: object</i>

پس از بارگذاری مجموعه داده می توان موارد مختلفی را مشاهده کرد. این 16 ستون مقدار بسیار غنی از اطلاعات را برای کاوش عمیق داده ها فراهم می کند. در حال حاضر برخی از مقادیر از دست رفته و nan را مشاهده میکنیم که نیاز به تمیز کردن دارند. بعداً، ممکن است لازم باشد برای تجزیه و تحلیل و پیش بینی، با نداشت مقادیر خاص به یک و صفر ادامه دهیم.

تعداد مقادیر از دست رفته برای هر ستون و یا ویژگی:

```

      id                                0
      name                             16
      host_id                           0
      host_name                         21
      neighbourhood_group               0
      neighbourhood                     0
      latitude                          0
      longitude                         0
      room_type                         0
      price                             0
      minimum_nights                    0
      number_of_reviews                 0
      last_review                       10052
      reviews_per_month                 10052
      calculated_host_listings_count    0
      availability_365                   0

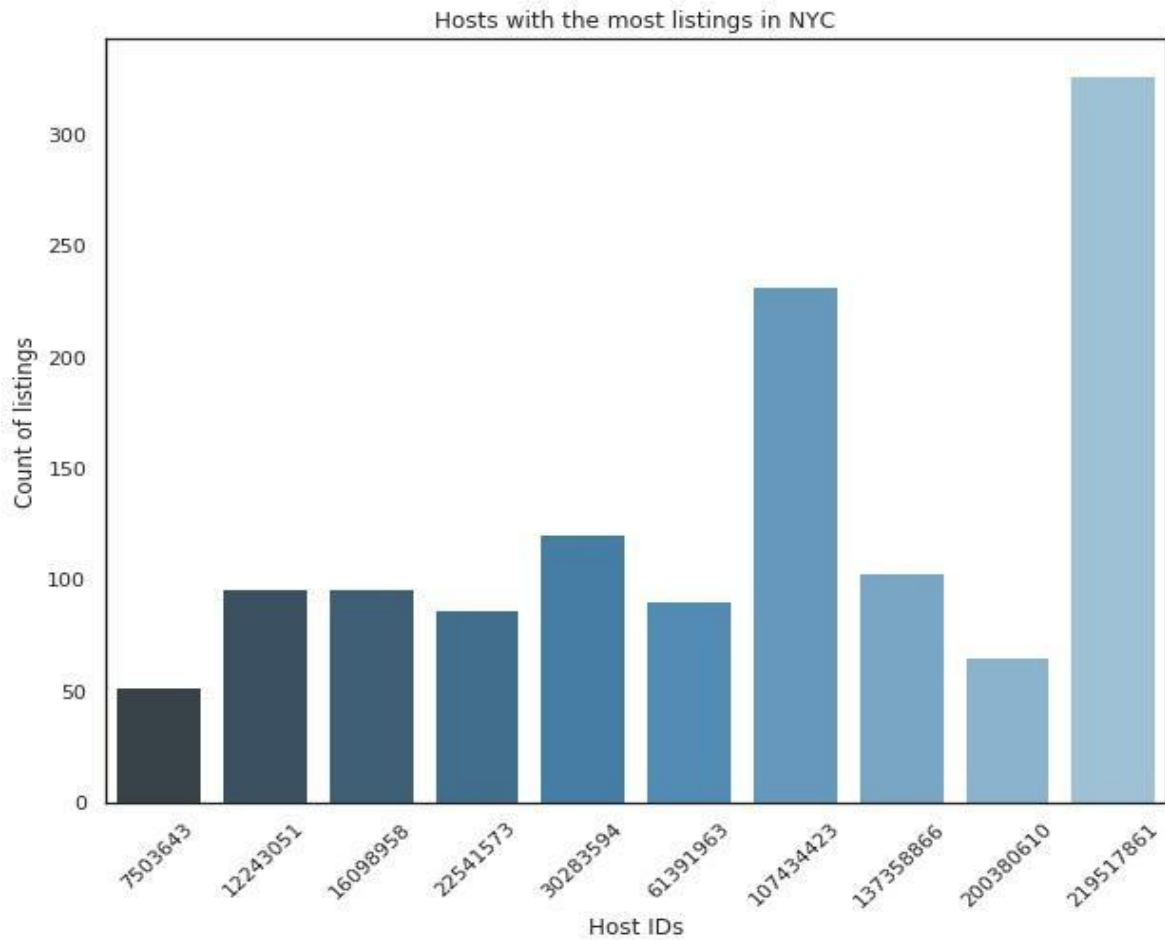
      dtype: int64

```

اطلاعات از دست رفته ای که مشاهده می شوند به راهکار خاص زیادی احتیاج ندارند. با نگاهی به ماهیت مجموعه داده های خود می توانیم موارد دیگری را بیان کنیم: ستون های "نام" و "نام میزبان" برای تجزیه و تحلیل داده های ما بی ربط و ناچیز هستند ، ستون های "last_review" و "review_per_month" به مدیریت بسیار ساده

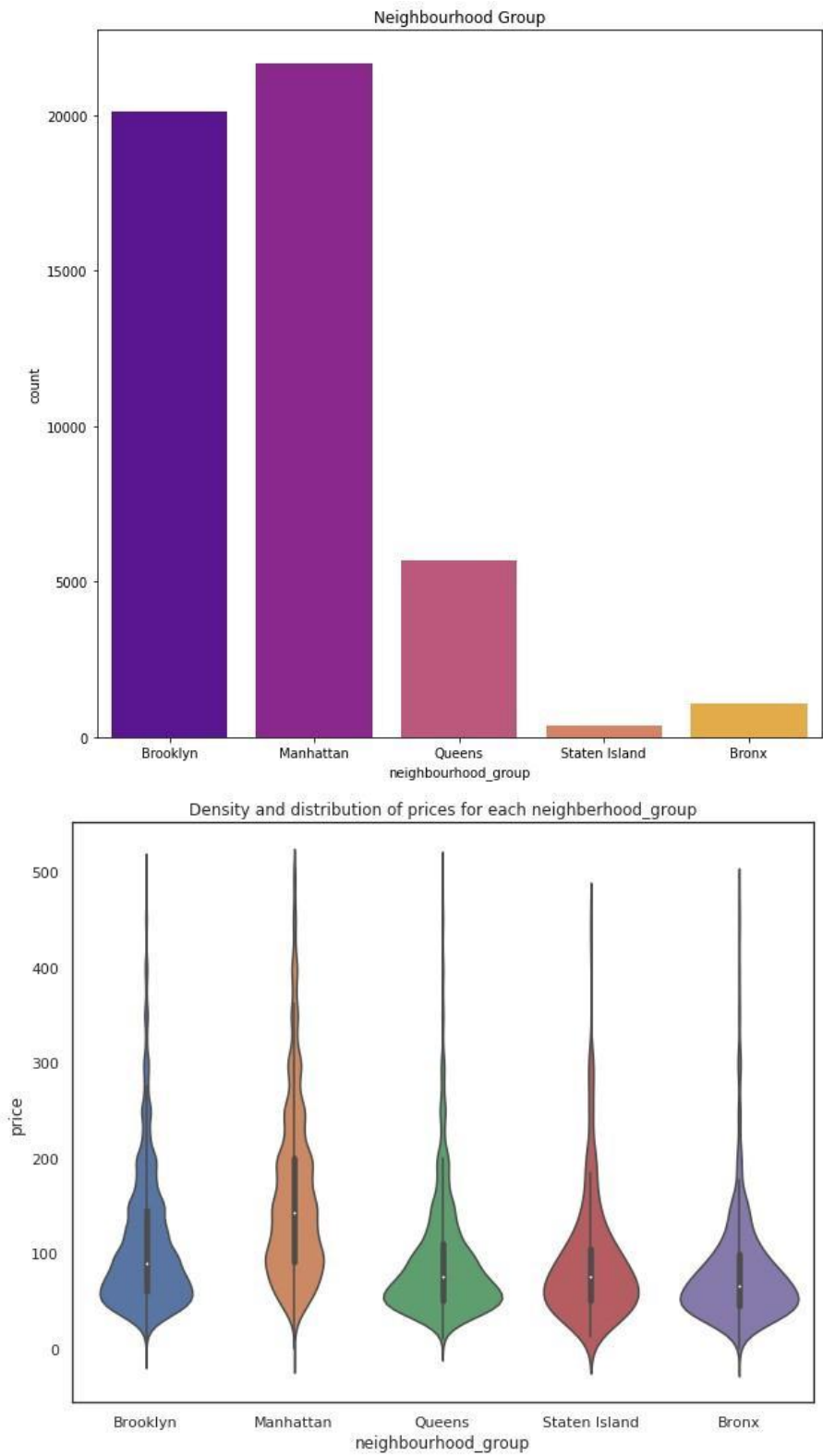
نیاز دارند. برای توضیح بیشتر، "last_review" تاریخ است. برای ستون "review_per_month" می توانیم آن را به سادگی با 0.0 برای مقادیر از دست رفته ضمیمه کنیم. لطفا توجه داشته باشید که ما 'host_name' را به دلیل ناچیز بودن و دلایل اخلاقی نیز رها می کنیم. برای ادامه اکتشاف داده ها و آموزش مدل (که بعداً انجام خواهیم داد) نسبت به افراد خاص بر اساس نام آنها هیچ استدلالی وجود ندارد. این اسامی به انسان واقعی اختصاص داده می شوند.

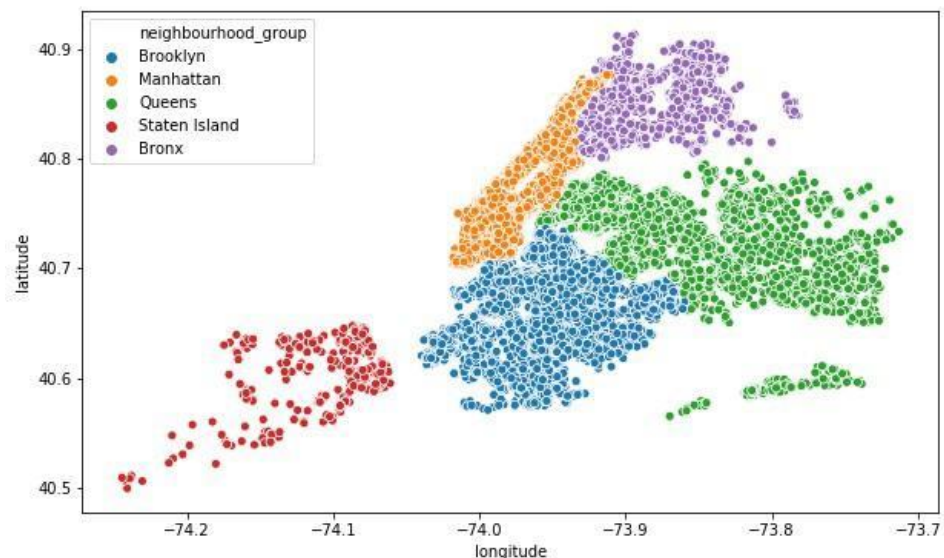
کاوش داده ها با تجزیه و تحلیل آمار و تجسم مقادیر ویژگی ها و همبستگی بین ویژگی های مختلف.



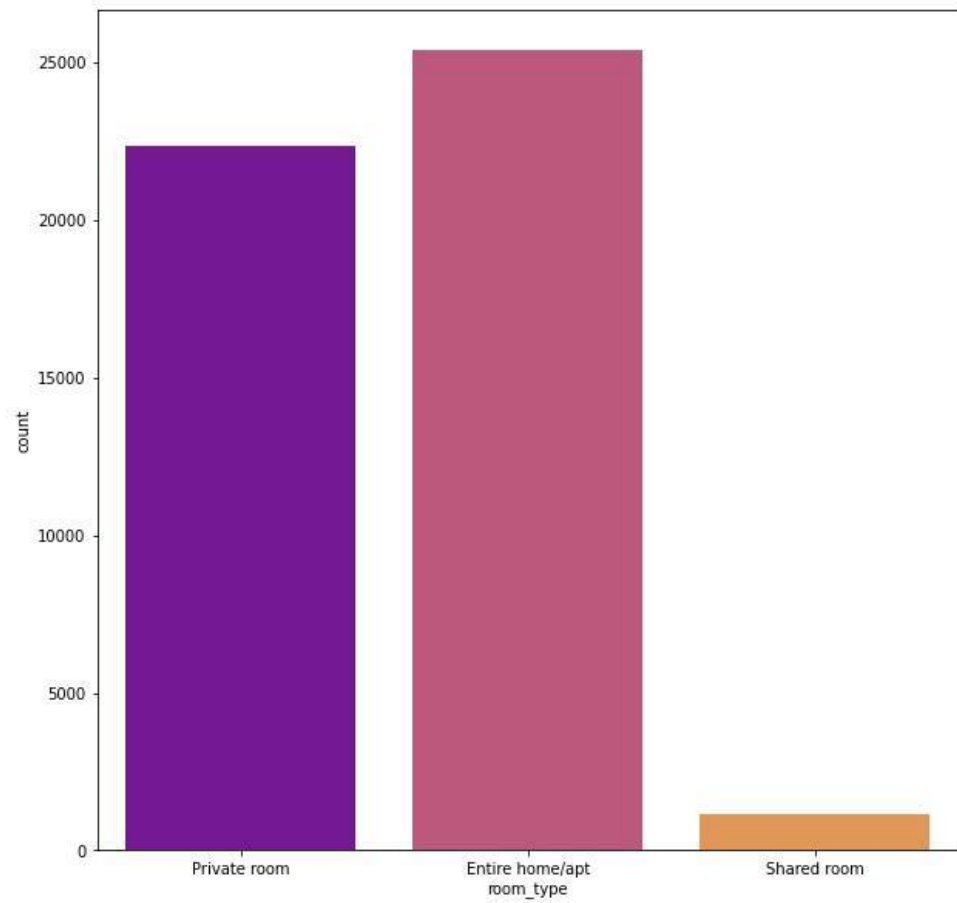
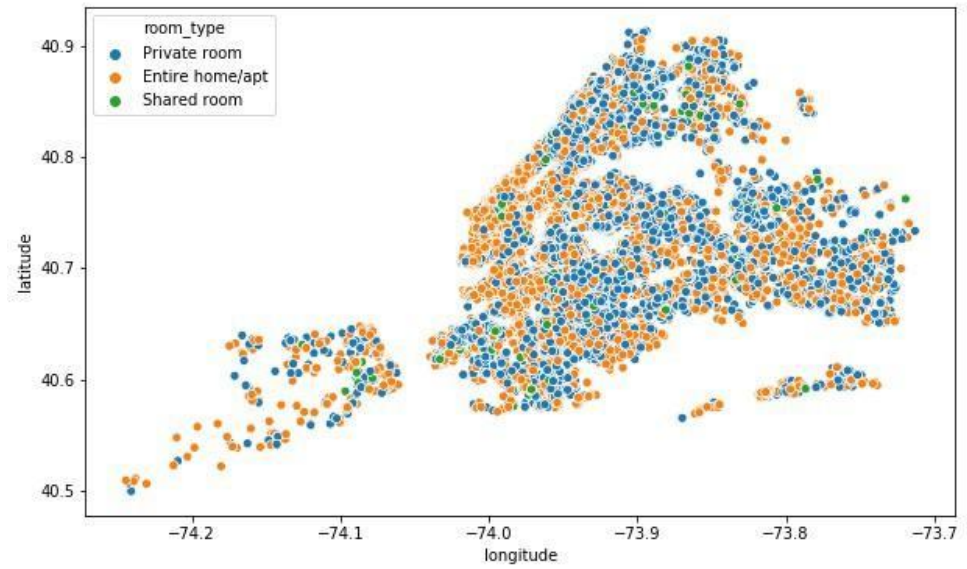
با توجه به نمودار ستونی می بینیم توزیع خوبی بین 10 میزبان برتر با بیشترین لیست وجود دارد. میزبان اول بیش از

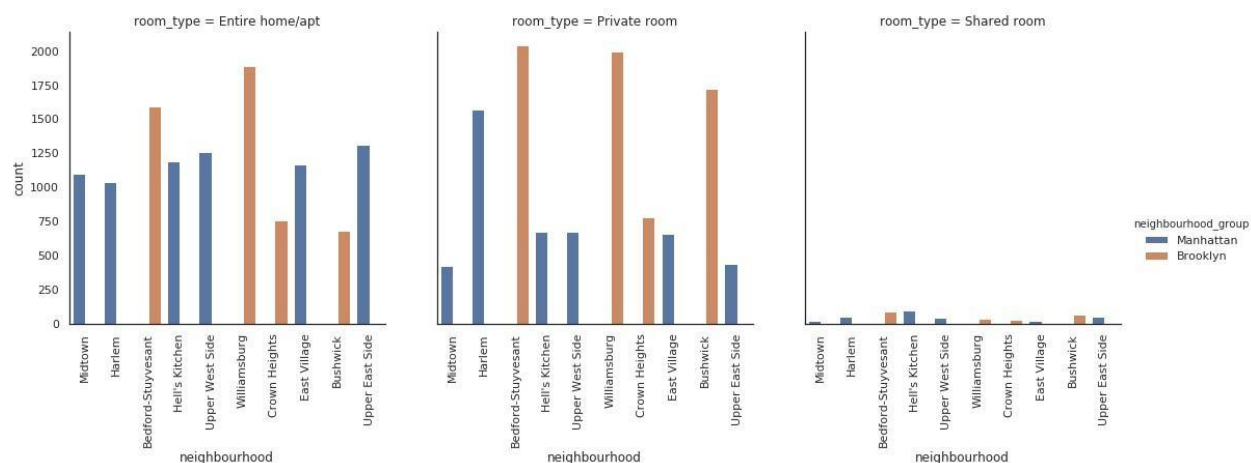
300 لیست دارد.



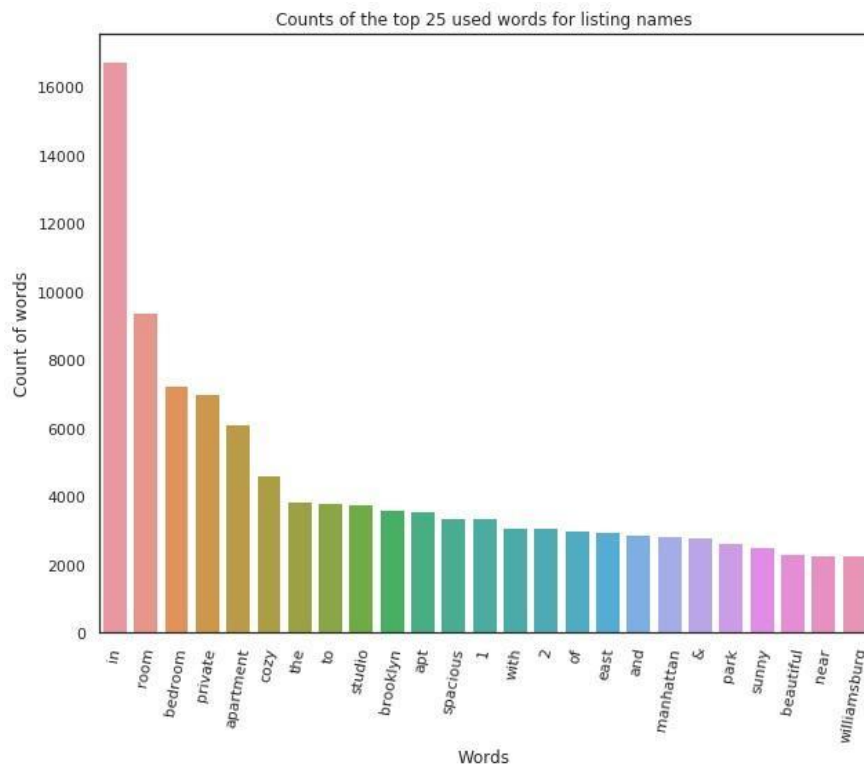


با داشتن یک جدول آماری و یک طرح ویولن ، ما قطعاً می توانیم چند مورد را در مورد توزیع قیمت Airbnb در بخش های NYC مشاهده کنیم. اول ، می توانیم بگوییم که منهن بالاترین طیف قیمت را برای لیست ها با قیمت 150 دلار به عنوان مشاهده متوسط دارد و پس از آن بروکلین با 90 دلار برای هر شب. کوینز و استاتن آیلند توزیع های بسیار مشابهی دارند ، برونکس ارزانترین از همه است. منهن یکی از گرانترین مکانهای زندگی در جهان است ، از طرف دیگر به نظر می رسد که برانکس دارای استانداردهای پایین زندگی است.

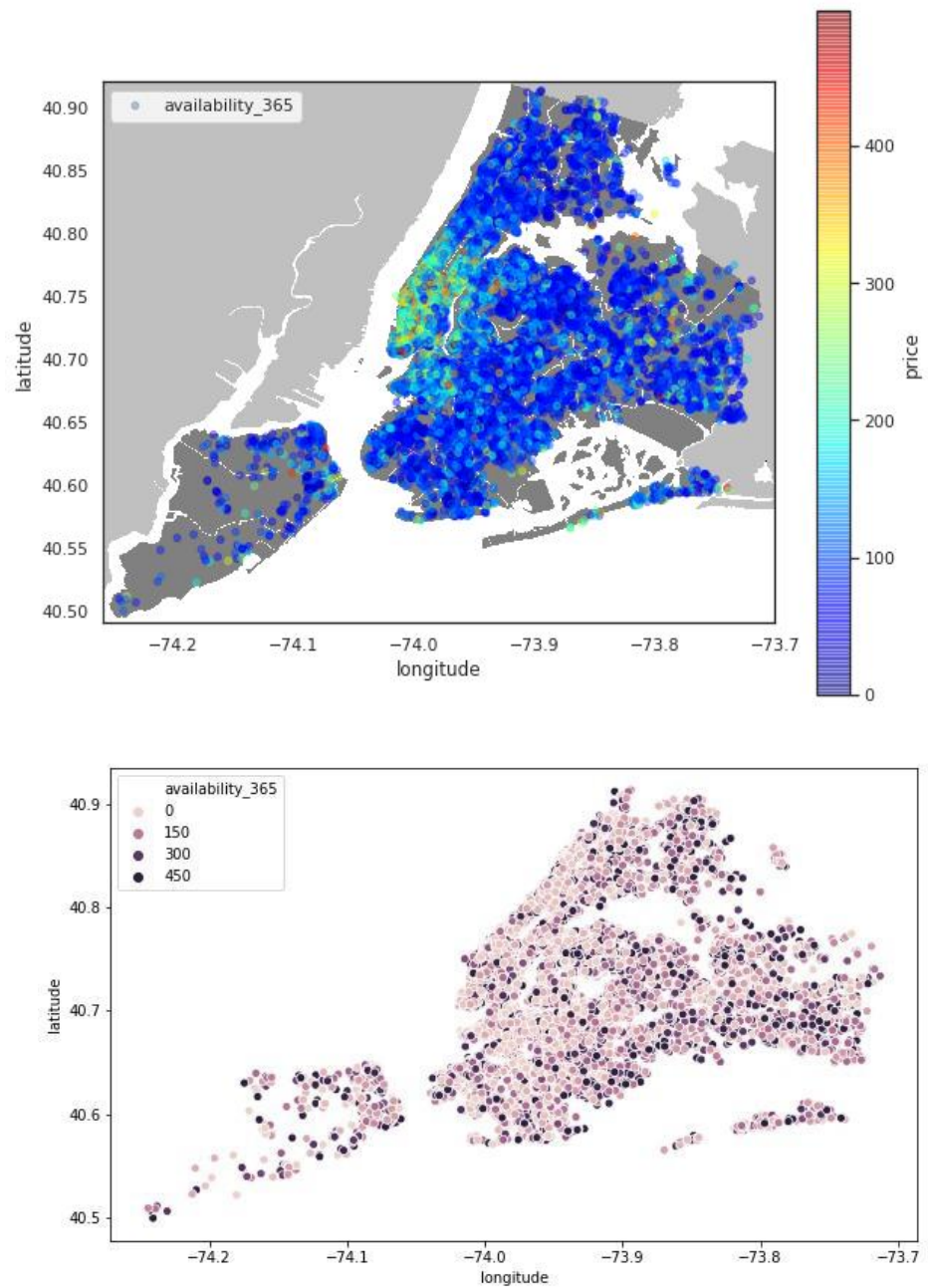




با توجه به نمودار های بالا اول ، می توانیم ببینیم که طرح ما از 3 زیرمجموعه تشکیل شده است - با چنین خروجی ، ما به راحتی می توانیم مقایسه توزیع بین ویژگی ها را ادامه دهیم. محورهای Y و X برای هر زیر قطعه کاملاً یکسان باقی می ماند ، محور Y نمایانگر شماری از مشاهدات محور X است که می خواهیم آنها را بشماریم. با این حال ، 2 عنصر مهم دیگر وجود دارد: ستون و رنگ. پس از تعیین ستون و تعیین رنگ ، قادر به مشاهده و مقایسه محورهای Y و X خود در میان ستون مشخص شده و همچنین کدگذاری رنگی هستیم. مشاهداتی که قطعاً بیشترین تضاد را دارند این است که لیست Airbnb از نوع "اتاق مشترک" به سختی در میان 10 محله پرجمعیت موجود است. سپس ، می توانیم ببینیم که برای این 10 محله فقط 2 بخش ارائه شده است: منهن و بروکلین. تا حدودی انتظار می رفت چون منهن و بروکلین یکی از پر مسافرت ترین مقصد هستند ، بنابراین بیشترین لیست را در دسترس خواهند داشت. همچنین می توانیم مشاهده کنیم که بدفورد-استویوزانت و ویلیامزبورگ محبوب ترین شهرها برای بخش منهن و هارلم برای بروکلین هستند.



ما می‌توانیم با این نمودار ستونی 25 کلمه نامی لیست‌های پرکاربرد را کشف و مرور می‌کنیم. این نشان می‌دهد که میزبان‌ها به سادگی لیست خود را به صورت کوتاه با اصطلاحات بسیار خاص برای جستجوی راحت‌تر توسط یک مسافر بالقوه توصیف می‌کنند. چنین شرایطی عبارتند از "اتاق"، "اتاق خواب"، "خصوصی"، "آپارتمان". این نشان می‌دهد که هیچ عبارتی یا اصطلاحی "محبوب / پرطرفدار" وجود ندارد که برای نام استفاده شود. میزبان از اصطلاحات بسیار ساده‌ای برای توصیف فضا و منطقه‌ای که لیست در آن قرار دارد استفاده می‌کنند. این روش تا حدودی مورد انتظار بود زیرا برخورد با مشتریان چندزبانه می‌تواند مشکل باشد و شما قطعاً می‌خواهید فضای خود را تا حد ممکن به صورت مختصر و قابل فهم توصیف کنید.



در این عکس ها با استفاده از نقاط عرض و طول تمام لیست های NYC را تجسم کردیم. همچنین ، بر اساس قیمت لیست ، برای هر نقطه از نقشه محدوده کدگذاری رنگی اضافه کرده ایم.

```

3 #let's grab 10 most reviewed listings in NYC
4 top_reviewed_listings=airbnb.nlargest(10,'number_of_reviews')
5 top_reviewed_listings

```

neighborhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
Brooklyn	40.66730	-73.76831	Private room	47	1	629	14.58	2	333
Harlem	40.82085	-73.94025	Private room	49	1	607	7.75	3	293
Harlem	40.82124	-73.93838	Private room	49	1	597	7.72	3	342
Harlem	40.82264	-73.94041	Private room	49	1	594	7.57	3	339
Brooklyn	40.66939	-73.76975	Private room	47	1	576	13.40	2	173
Manhattan	40.77006	-73.87683	Private room	46	1	543	11.59	5	163
Manhattan	40.71921	-73.99116	Private room	99	2	540	6.95	1	179
Manhattan	40.76335	-73.87007	Private room	48	1	510	16.22	5	341
Manhattan	40.67926	-73.97711	Entire home/apt	160	1	488	8.14	1	269
Manhattan	40.70283	-73.92131	Private room	60	3	480	6.70	1	0

از این جدول می توان مشاهده کرد که 10 لیست مورد بررسی بیشتر در Airbnb برای NYC دارای قیمت متوسط \$ 65 با بیشتر لیست ها زیر \$ 50 و 9/10 آنها از نوع "اتاق خصوصی" هستند. لیست برتر دارای 629 بررسی (review) است.

نتیجه گیری

به نظر می رسد این مجموعه داده برای سال 2019 یک مجموعه داده بسیار غنی با ستون های متنوع است که به ما امکان می دهد کاوش عمیق داده ها را در هر ستون قابل توجه ارائه دهیم. اول ، ما میزبانانی پیدا کرده ایم که از پلت فرم Airbnb به خوبی بهره می برند و بیشترین لیست را ارائه می دهند. ما دریافتیم که میزبان برتر ما 327 لیست دارد. پس از آن ، ما به تحلیل محله ها و تراکم لیست محله و مناطق محبوب تر از مناطق دیگر پرداختیم. در مرحله بعد ، ما از ستون های عرض و طول جغرافیایی خود برای ایجاد یک نقشه حرارتی جغرافیایی با کد رنگی با قیمت لیست ها استفاده می کنیم. علاوه بر این ، ما به ستون اول با رشته های نام بازگشتیم و مجبور شدیم برای تجزیه هر

عنوان و تجزیه و تحلیل روند موجود در مورد نحوه نامگذاری لیست ها و همچنین تعداد کلمات مورد استفاده توسط میزبان ، کمی بیشتر کدگذاری کنیم. سرانجام ، بیشترین میزبان بازیابی شده را پیدا کردیم و برخی از ویژگی های اضافی را تجزیه و تحلیل کردیم. این تجزیه و تحلیل داده ها برای تصمیم گیری بهتر در زمینه کسب و کار ، کنترل بر روی سیستم ، ابتکارات بازاریابی ، اجرای ویژگی های جدید و موارد دیگر استفاده می شود.