CSCI 4146/6409 - Process of Data Science (Summer 2023)

Assignment 1

Tasneem Hoque
B00841761

Karan Aggarwal
B00912580

1. [1.5] Business Understanding

a) Formulate a business problem that can be solved with the dataset. [0.25]

Airbnb hosts may struggle to determine the most appropriate nightly rates for their listings, especially if they are new to the platform or are unfamiliar with the local market. However, with the U.S. Airbnb Open Data, it is possible to analyze historical booking data and identify patterns and trends in pricing across different geographic locations.

By analyzing the data, businesses could determine the optimal pricing for their listings, taking into account factors such as location, property type, and minimum nights requirement. They could also identify high-demand periods and adjust their pricing strategy accordingly. This could lead to increased bookings and revenue for hosts, while also improving the overall guest experience by offering competitive rates.

b) For the business problem, propose 3 data science solutions and assess their feasibility. Select the final solution and explain your decision. [0.5]

Solution 1: Availability prediction

- A model could be built to predict the availabilbility in the neighbourhood, and an appropriate pricing strategy could be used based on availability of other properties in the area.
- **Required data** - Historical claims of availability based on location.
- **Required business capacity** - Adujust some features of the Airbnb such as minimum nights requirement, or read reviews to understand customer requirements to increase airbnb bookings.

Solution 2: Demand prediction

- A model can be used to predict whether a given listing is likely to have a high or low demand based on a set of relevant features such as location and property type. Once the demand level has been predicted, hosts can adjust their pricing strategy accordingly.
- **Required data** - Data on all previous Airbnb listing with location, property type and their pricing
- **Required business capacity** - Changing the price of the listing based on demands.

Solution 3: Review prediction

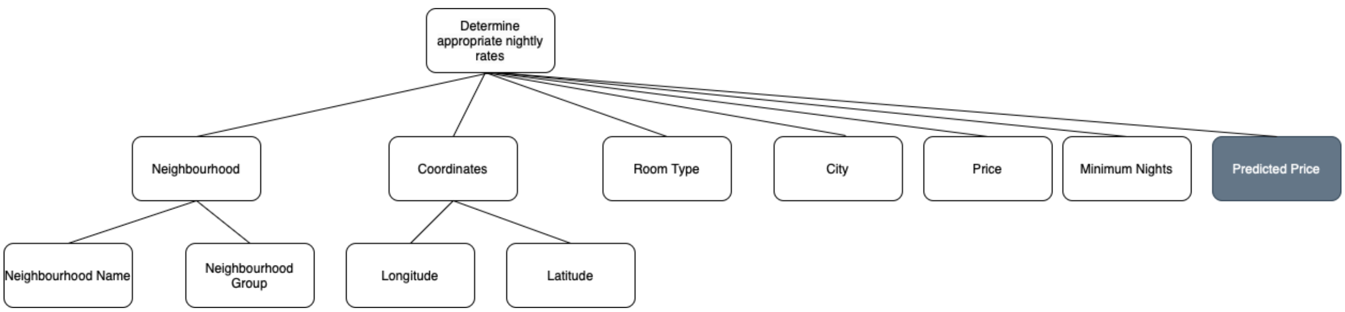
- A model can be used to predict whether a given listing is likely to have a higher price point based on reviews.
- **Required data** - Data on all previous Airbnb listing with total reviews, monthly reviews and pricing.
- **Required business capacity** - Changing the price of the listing based on reviews.

The final solution for this business problem is demand prediction. The location is a major factor to pricing in conjuction to room type, neighbourhood, city and minimum nights. One additional factor that can be taken into account is availability of other properties in the same area. In general, the higher the demand, the more nightly rates the property will be able to collect.

c) For the final solution, identify the prediction subject, its domain concepts, and sub-concepts (if there are any). Draw a hierarchical graph of the concepts. [0.25]

For the final solution we are having the prediction subject as Price.

It's domain concept could be Neighbouhood, Coordinates, Room Type, City, and Minimum Nights. Neighbourhood has sub domain as Neighbourhood Name and Neighbourhood Group. Coordinates has sub domain as Longitude and Latitude.



d) For each domain concept, design descriptive features that best describe a conceptusing data from the dataset. Summarize the resulting ABT in a table with the following columns [0.5]:

i. Feature Name ii. Domain Concept iii. Feature Description iv. Feature Type v. Data Type

1. Room type
   A. Type of room
2. Price
   A. Price equals 0 - Invalid data - Derived Flag
   B. Price > 0 and Price < 10000 - Raw
   C. Price < 100000 and price > 10000 - could be outliers - Derived Flag
   D. Average Price for a room type - Derived Aggregate
   E. Average Price for a neighbourhood - Derived Aggregate
   F. Average Price City wise - Derived Aggregate
3. Neighbourhood Name
   A. Average Price for a neighbourhood - Derived Aggregate
4. Neighbourhood Group
   A. Diversity of Neighbourhood group - Derived Other
5. Coordinates
   A. Coordinates as per the Neighbourhood - Mapping
   B. Coordinates as per the Neighbourhood Group - Mapping

6. City
    A. Average Price City wise - Derived Aggregate
7. Minimum Nights
    A. Average Price to Minimum Nights and City - Derived Aggregate
8. Predicted Price

S.No.	Feature Name	Domain Concept	Feature Description	Feature Type	Data Type
1	Type of Room	Room type	Describes it's a private room, Entire home/apt, or shared room. The costs of the Airbnb varies based on them.	Raw	String
2	Price equals 0 - Invalid data	Price	Since prices are less than $1, it clearly tells that it is an invalid data.	Derived Flag	Number
3	Price > 0 and Price < 10000	Price	Valid raw data.	Derived Flag	Number
4	Price < 100000 and price > 10000	Price	If prices are too high, then there could be a possibility of outliers, since such data are very less.	Derived Flag	Number
5	Average Price for a room type	Price	We can calculate average price of room basis on different factors.	Derived Aggregate	Number
6	Average Price for a neighbourhood	Price	To predict price of an Airbnb we can calculate the average price of the room in a neighbourhood.	Derived Aggregate	Number
7	Average Price City wise	Price	To predict price of an Airbnb we can calculate the average price of the room in a city.	Derived Aggregate	Number
8	Diversity of Neighbourhood group	Neighbourhood Group	To predict price of an Airbnb we can calculate the average price of the room in a neighbourhood group.	Derived Other	String
9	Coordinates as per the Neighbourhood	Coordinates	We can read what all coordinates are lying in a neighbourhood.	Mapping	Number
10	Coordinates as per the city	Coordinates	We can read what all coordinates are lying in a neighbourhood.	Mapping	Number
11	Average Price to Minimum Nights and City	Minimum Nights	As the minimum nights, the price of the room could vary.	Derived Aggregate	Number
12	Predicted Price	Predicted Price	The final feature will be our predicted price.		Number

2. [1] Data Exploration

a) Build the data quality report of the resulting ABT. [0.5]

In [18]:

```
import pandas as pd
import warnings
import matplotlib.pyplot as plt

pd.set_option("display.max_rows", None)
pd.set_option("display.max_columns", None)

data = "AB_US_2023.csv"
details = pd.read_csv(data, low_memory=False)
details.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 232147 entries, 0 to 232146
Data columns (total 18 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   id                              232147 non-null  int64
 1   name                            232131 non-null  object
```

```
 2   host_id                         232147 non-null  int64
 3   host_name                       232134 non-null  object
 4   neighbourhood_group             96500 non-null   object
 5   neighbourhood                   232147 non-null  object
 6   latitude                        232147 non-null  float64
 7   longitude                       232147 non-null  float64
 8   room_type                       232147 non-null  object
 9   price                           232147 non-null  int64
 10  minimum_nights                  232147 non-null  int64
 11  number_of_reviews               232147 non-null  int64
 12  last_review                     183062 non-null  object
 13  reviews_per_month               183062 non-null  float64
 14  calculated_host_listings_count  232147 non-null  int64
 15  availability_365                232147 non-null  int64
 16  number_of_reviews_ltm           232147 non-null  int64
 17  city                            232147 non-null  object
dtypes: float64(3), int64(8), object(7)
memory usage: 31.9+ MB
```

As seen from the table above, we have a total of 232,147 records with 18 columns.
Most of the data in the "neighbourhood group" column is NULL, with only 96,500 non-null values out of the 232,147 records.
The "name" and "host name" columns have a few missing rows.
Additionally, some of the Airbnb listings do not have any reviews provided.

Now, let's have a look on few of the rows to better understand the data.

In [2]:
```
details.loc[0:3]
```

Out[2]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type
0	958	Bright, Modern Garden Unit - 1BR/1BTH	1169	Holly	NaN	Western Addition	37.77028	-122.43317	Entire home/apt
1	5858	Creative Sanctuary	8904	Philip And Tania	NaN	Bernal Heights	37.74474	-122.42089	Entire home/apt
2	8142	Friendly Room Apt. Style - UCSF/USF - San Franc...	21994	Aaron	NaN	Haight Ashbury	37.76555	-122.45213	Private room
3	8339	Historic Alamo Square Victorian	24215	Rosy	NaN	Western Addition	37.77564	-122.43642	Entire home/apt

Based on the data presented above, it can be inferred that the "name" column contains more information than just the name of the Airbnb listing. It appears to provide additional details such as the type of accommodation (e.g. 1BR or sanctuary) and location details.

The remaining columns provide IDs, host names, group and neighborhood information, longitude and latitude coordinates, and other details to help identify and locate each listing. Additionally, the "room type" column should be treated as categorical data and can be split into two categories - "entire home" or "private room", for example.

Next, we will examine the statistics of the data.

In [3]:
```
details_without_id = details.drop('id', axis=1)
details_without_id.describe(include = 'all')
```

Out[3]:

	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitu
count	232131	2.321470e+05	232134	96500	232147	232147.000000	232147.0000
unique	220164	NaN	29368	30	1412	NaN	N.
top	Presidential Suite In A Mansion	NaN	Blueground	City of Los Angeles	Unincorporated Areas	NaN	N.
freq	150	NaN	4305	22204	11882	NaN	N.
mean	NaN	1.582248e+08	NaN	NaN	NaN	36.610585	-98.3014
std	NaN	1.587164e+08	NaN	NaN	NaN	5.126523	19.7069
min	NaN	2.300000e+01	NaN	NaN	NaN	25.957323	-123.0891
25%	NaN	2.299242e+07	NaN	NaN	NaN	33.976225	-118.3151
50%	NaN	1.005783e+08	NaN	NaN	NaN	36.190556	-97.7276
75%	NaN	2.686930e+08	NaN	NaN	NaN	40.717440	-77.0262
max	NaN	5.069384e+08	NaN	NaN	NaN	47.734010	-70.9960

Since a lot other stats are missing, we will be manually checking median, mode, variance, range, kurtosis, skew, covariance and correlation.

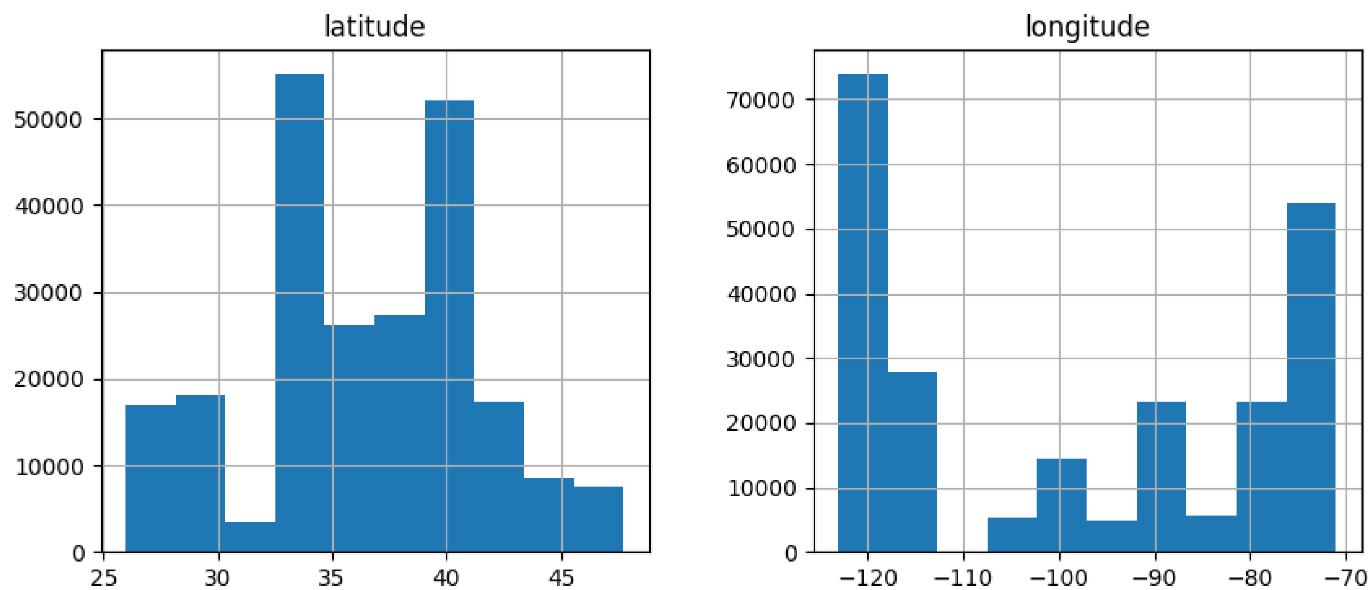
In [4]:
```
    stats = {
        "Median": lambda details: details.median(),
        "Mode": lambda details: details.mode().iloc[0],
        "Variance": lambda details: details.var(),
        "Range": lambda details: details.max() - details.min(),
        "Kurtosis": lambda details: details.kurtosis(),
        "Skew": lambda details: details.skew(),
        "Covariance": lambda details: details.iloc[:,0].cov(details.iloc[:,1]),
        "Correlation": lambda details: details.iloc[:,0].corr(details.iloc[:,1])
    }

    number_details = details.select_dtypes("number").columns
    data = details[number_details]

    report = pd.DataFrame(index=number_details, columns=stats.keys())

    for stat_name, fn in stats.items():
        report[stat_name] = fn(data)

    report
```

Out[4]:

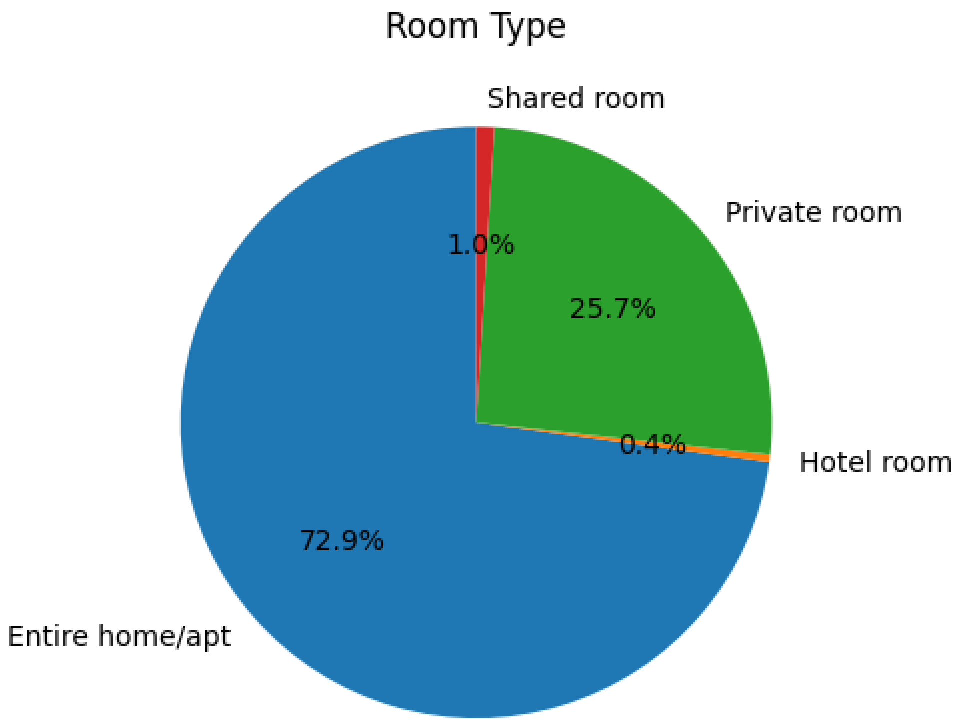
	Median	Mode	Variance	Range	Kurtosis	Skew
id	4.896307e+07	1.398981e+06	1.201305e+35	8.581014e+17	-1.504092	0.639792
host_id	1.005783e+08	1.074344e+08	2.519089e+16	5.069383e+08	-0.723879	0.824883
latitude	3.619056e+01	3.610761e+01	2.628124e+01	2.177669e+01	-0.358114	-0.189426
longitude	-9.772767e+01	-1.151616e+02	3.883631e+02	5.209313e+01	-1.719888	0.055468
price	1.490000e+02	1.500000e+02	1.049899e+06	1.000000e+05	4717.267259	57.065465

minimum_nights	3.000000e+00	1.000000e+00	7.795616e+02	1.249000e+03	417.859111	14.770285
number_of_reviews	9.000000e+00	0.000000e+00	6.504286e+03	3.091000e+03	36.575840	4.350472
reviews_per_month	1.000000e+00	1.000000e+00	3.651202e+00	1.014100e+02	182.251882	6.419711
calculated_host_listings_count	2.000000e+00	1.000000e+00	1.123890e+04	1.002000e+03	43.238413	6.172384
availability_365	1.750000e+02	0.000000e+00	1.814821e+04	3.650000e+02	-1.529300	-0.003551
number_of_reviews_ltm	3.000000e+00	0.000000e+00	4.243581e+02	1.314000e+03	276.796655	8.159665

In [5]:
```
from matplotlib import pyplot
fig, axs = pyplot.subplots(1, 2, figsize=(10, 4))
details.hist(column='latitude', ax=axs[0])
details.hist(column='longitude', ax=axs[1])
pyplot.show()   # show the figure
```

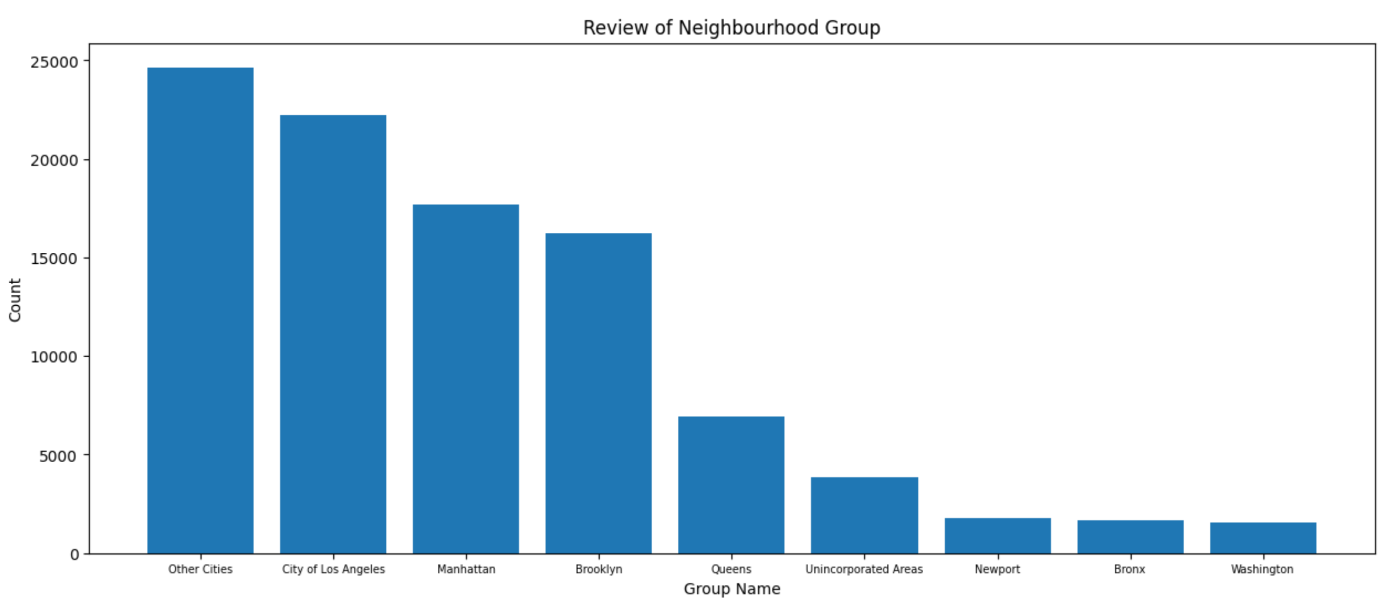


In [6]:
```
room_type_counts = details.groupby('room_type').size()
pyplot.pie(room_type_counts, labels = room_type_counts.index, autopct='%.1f%%', startang
pyplot.title('Room Type')
pyplot.show()
```



The majority of the listings on Airbnb are for entire homes/apartments, while private rooms account for only 25% of the listings. Shared rooms and hotel rooms are relatively infrequently listed on Airbnb.

In [42]:
```
fig, ax = pyplot.subplots(figsize=(15, 6))
neighbourhood_group = details['neighbourhood_group'].value_counts()
newgroup = neighbourhood_group[neighbourhood_group <= 1500]
neighbourhood_group = neighbourhood_group[neighbourhood_group >= 1500]
neighbourhood_group['Other Cities'] = newgroup.sum() + neighbourhood_group['Other Cities
neighbourhood_group = neighbourhood_group.sort_values(ascending=False)
pyplot.bar(neighbourhood_group.index, neighbourhood_group.values)

pyplot.title('Review of Neighbourhood Group')
pyplot.xlabel('Group Name')
pyplot.ylabel('Count')
pyplot.xticks(fontsize=7)
pyplot.show()
```

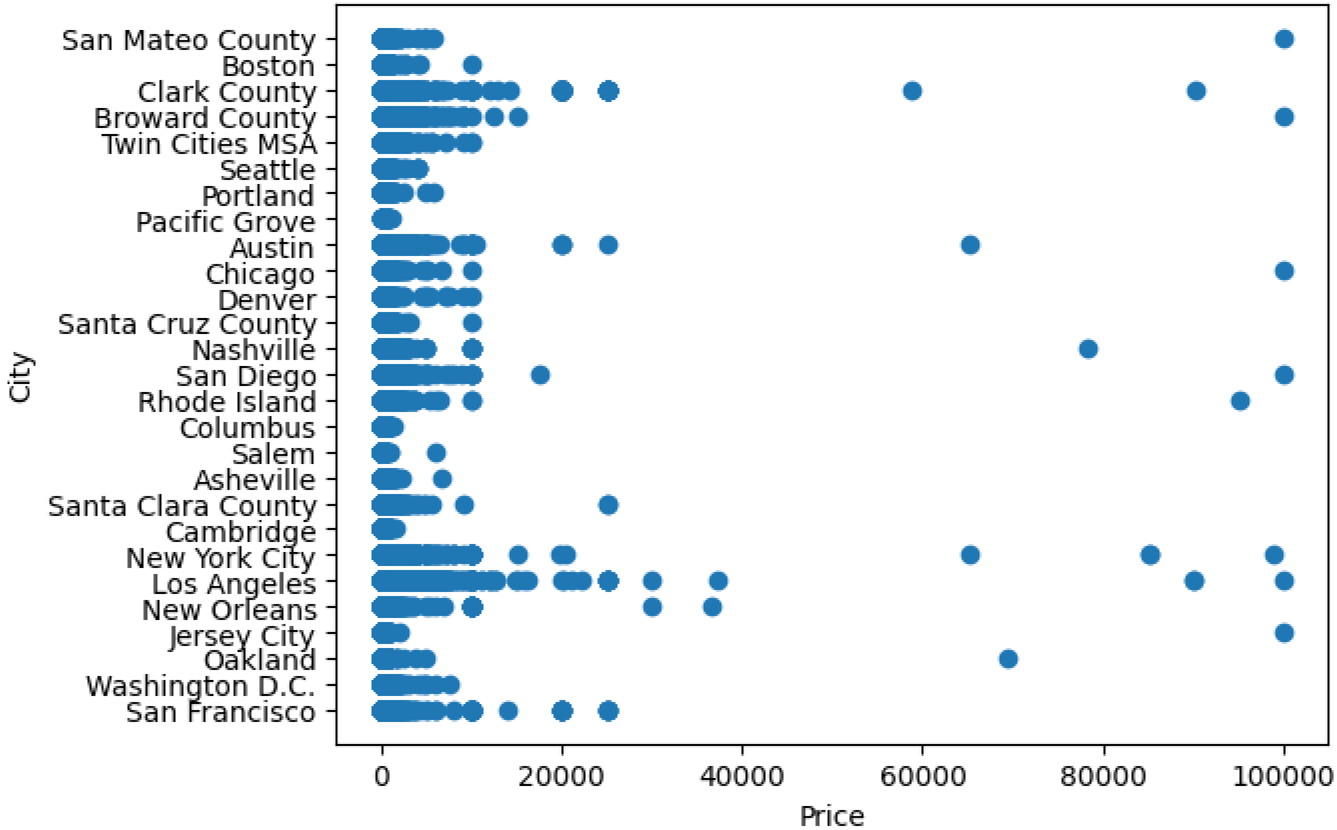


Upon comparing the number of listings for a particular neighborhood group, it has been observed that the

City of Los Angeles has the highest number of listings, with more than 20,000 listings. This suggests that it may be the most popular city among Airbnb owners.

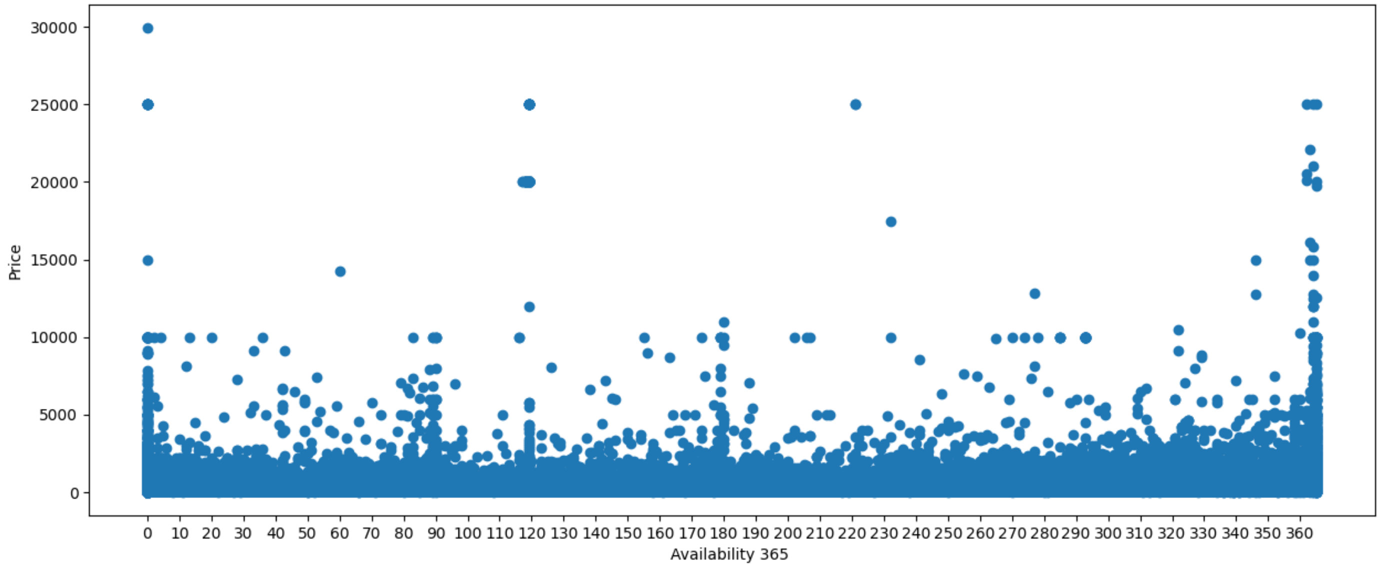
In [8]:
```
pyplot.scatter(details['price'], details['city'])
pyplot.xlabel('Price')
pyplot.ylabel('City')
```

Out[8]:
```
Text(0, 0.5, 'City')
```



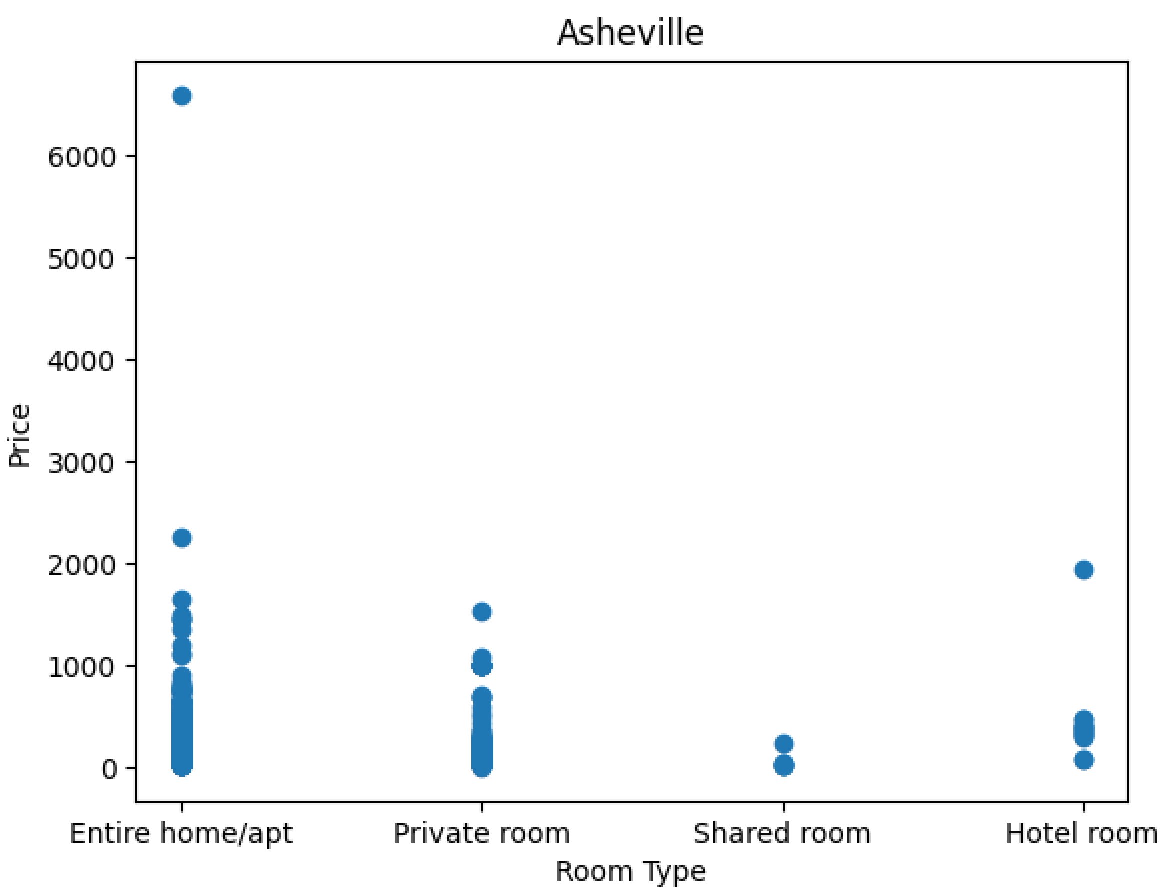
During the comparison of price differences among different cities, it was discovered that Los Angeles is the most expensive city, while Salem is the cheapest. This finding highlights a strong correlation between price and city.

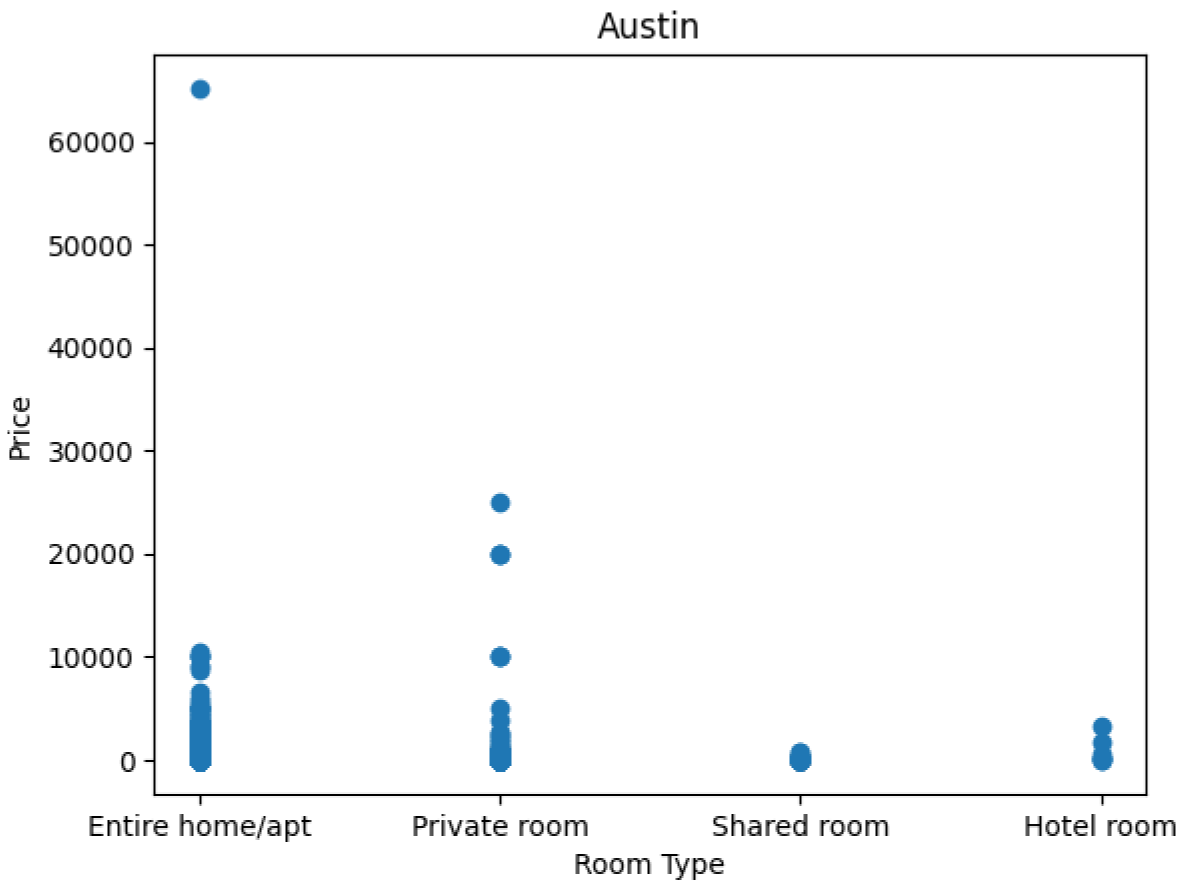
In [9]:
```
fig, ax = pyplot.subplots(figsize=(15, 6))
pyplot.scatter(details['availability_365'][details['price'] < 30000], details['price'][d
pyplot.xlabel('Availability 365 ')
plt.xticks(range(0, 366, 10))
pyplot.ylabel('Price')
pyplot.show()
```

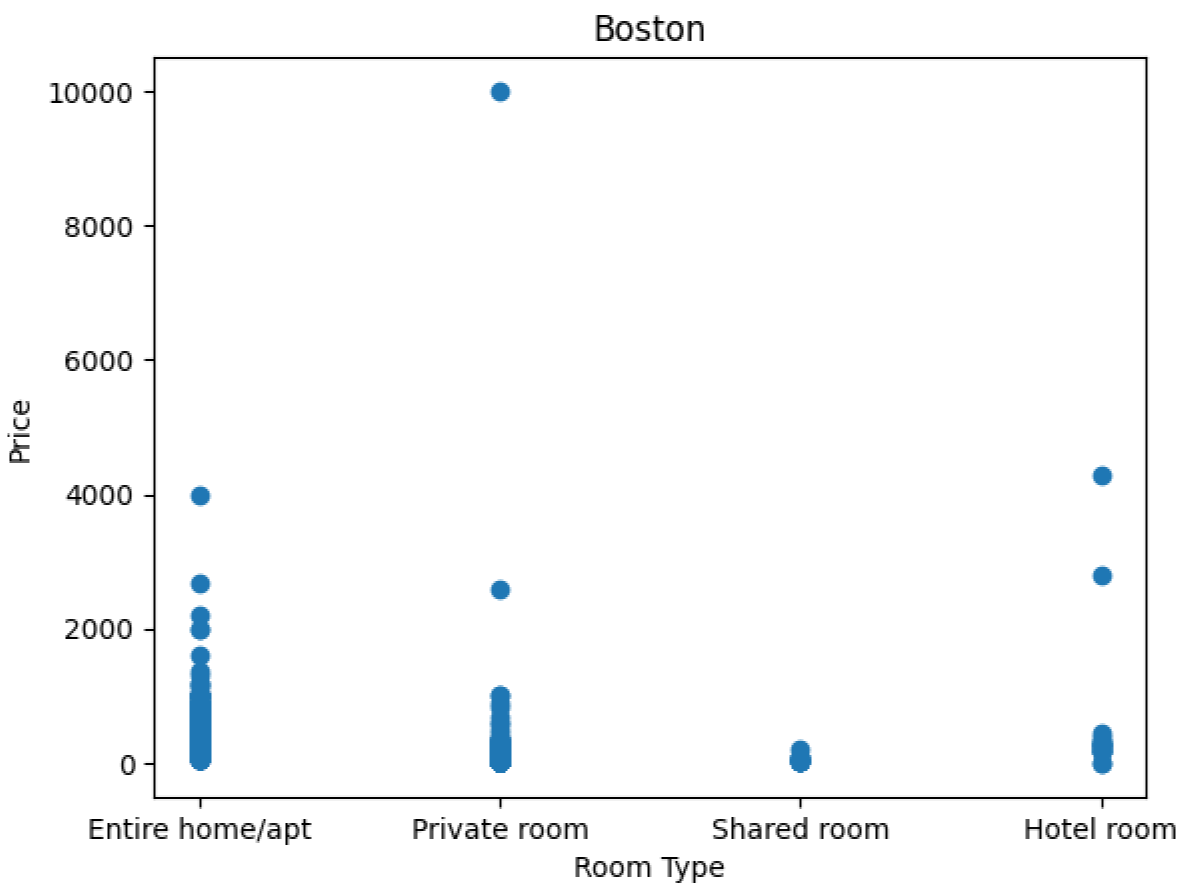


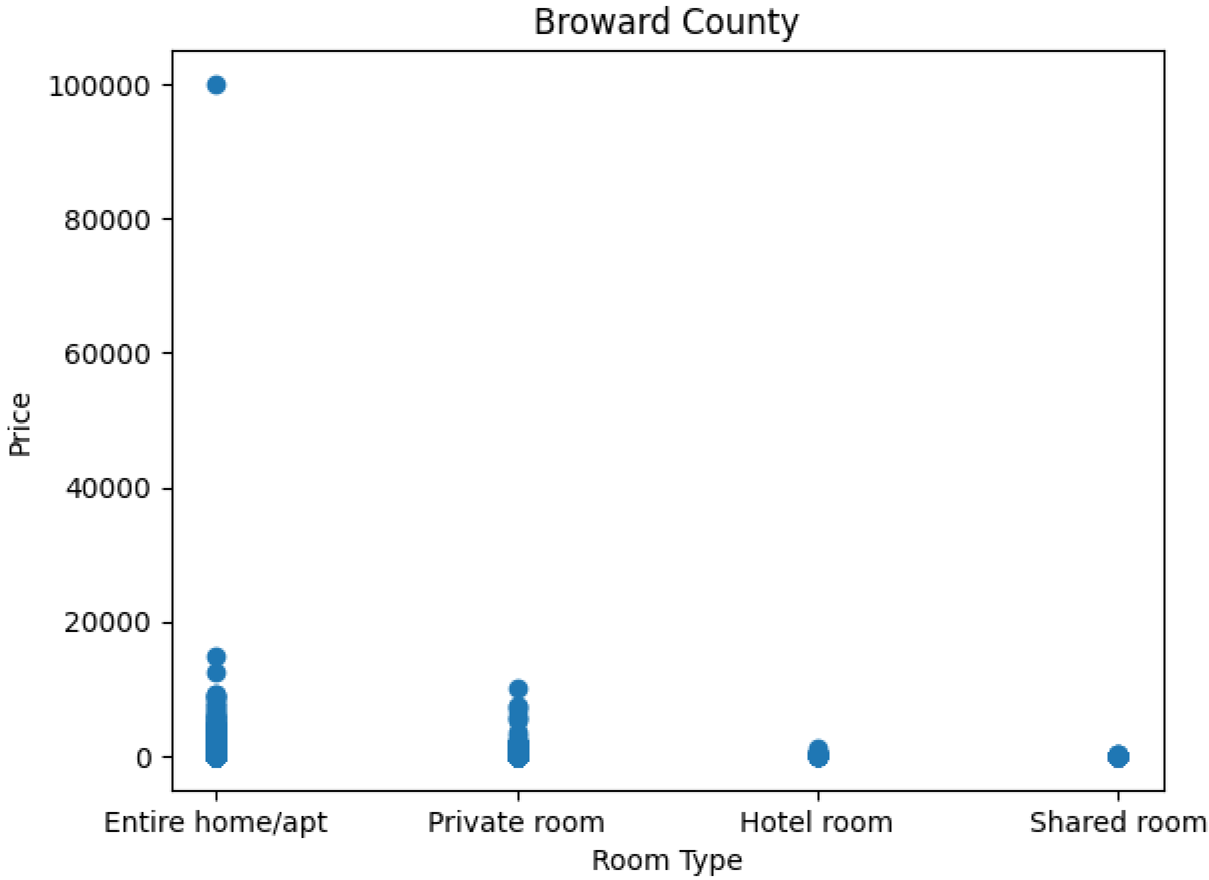
Since there is no direct relation between the availability of the airbnb and the price, the availability column can be dropped.

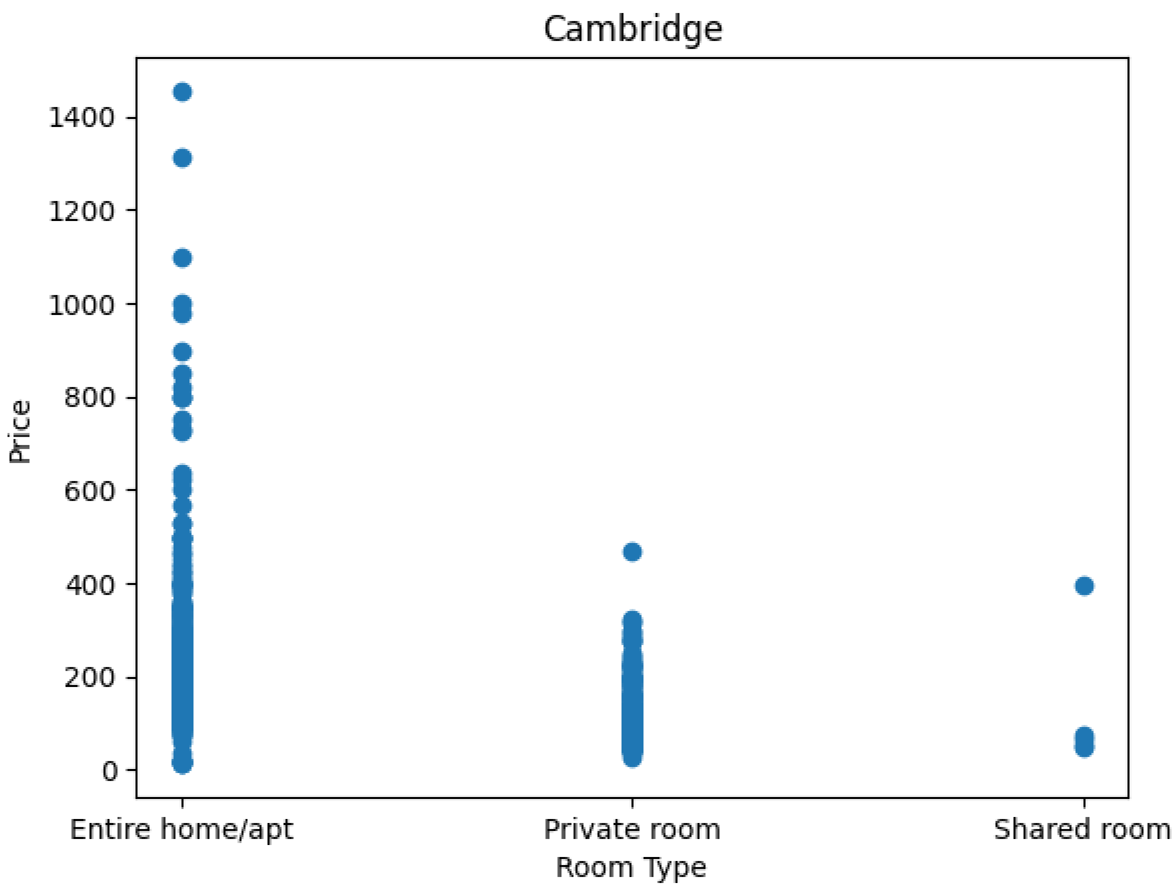
In [10]:
```
city = details.groupby('city')

for city, data in city:
    plt.figure()
    plt.scatter(data['room_type'], data['price'])
    plt.title(city)
    plt.xlabel('Room Type')
    plt.ylabel('Price')
    plt.show()
```

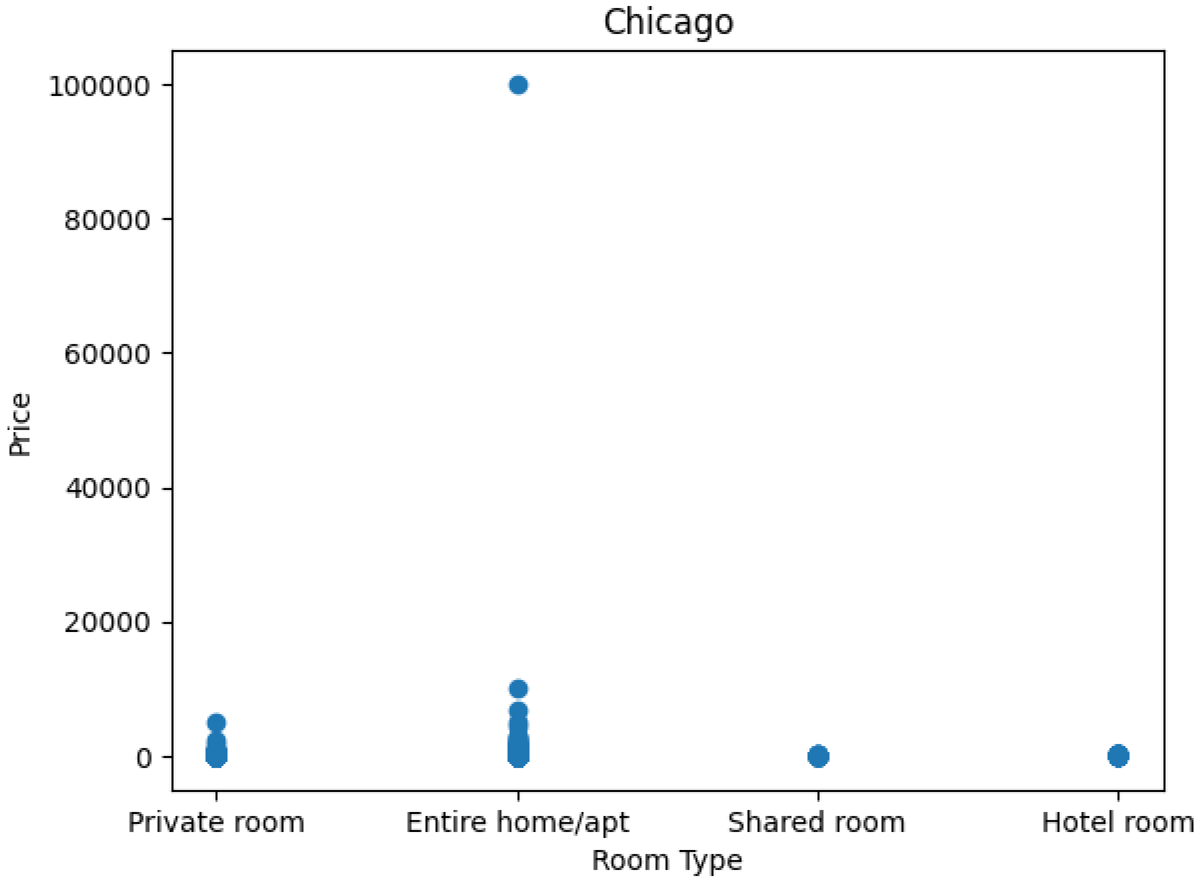


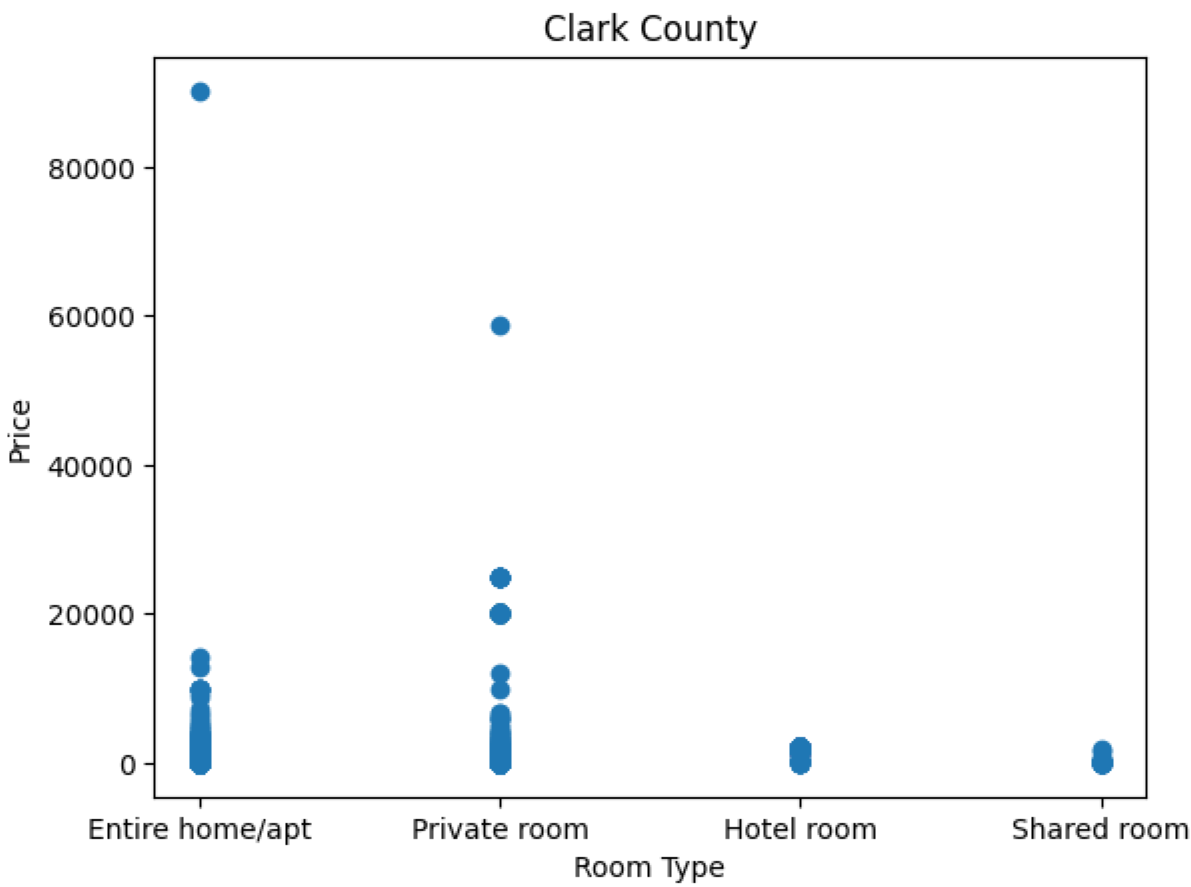

Austin
Price
60000
50000
40000
30000
20000
10000
0
Entire home/apt
Private room
Shared room
Hotel room
Room Type

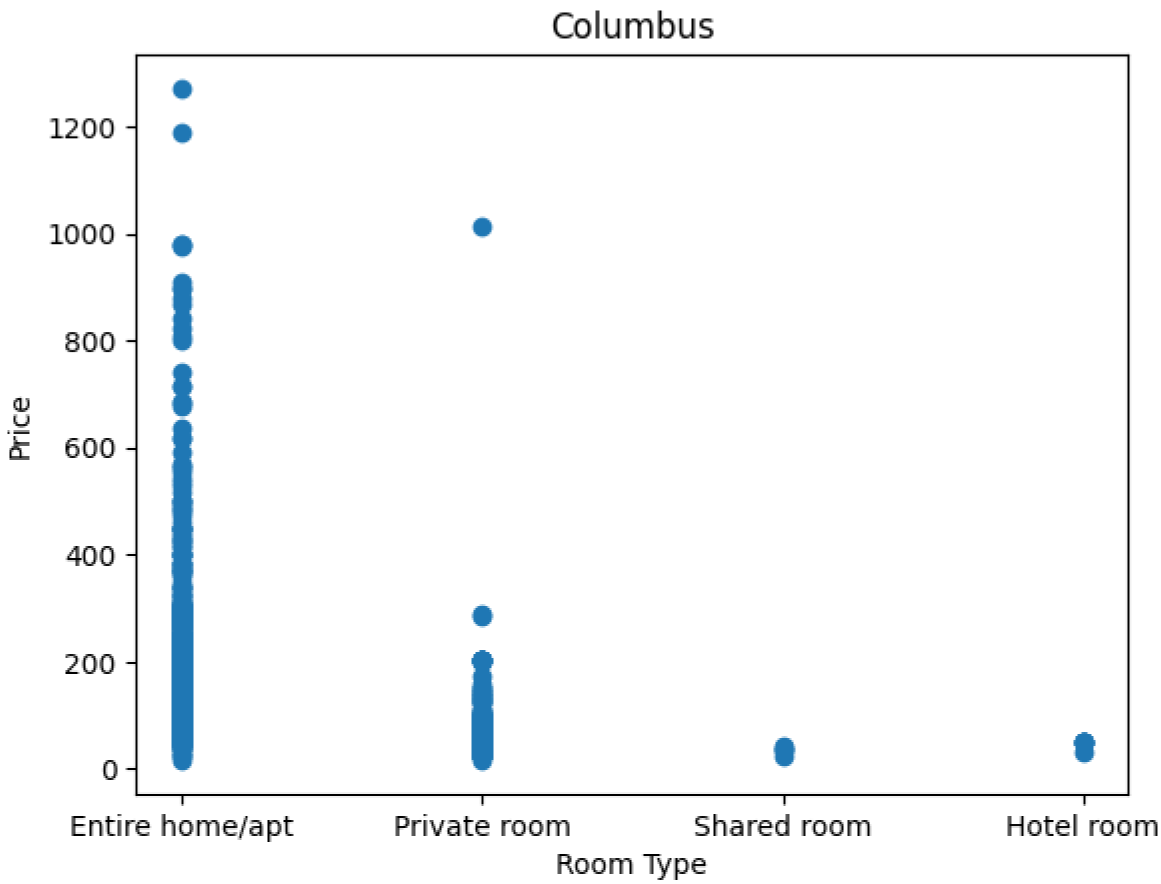

Boston
Price
10000
8000
6000
4000
2000
0
Entire home/apt
Private room
Shared room
Hotel room
Room Type

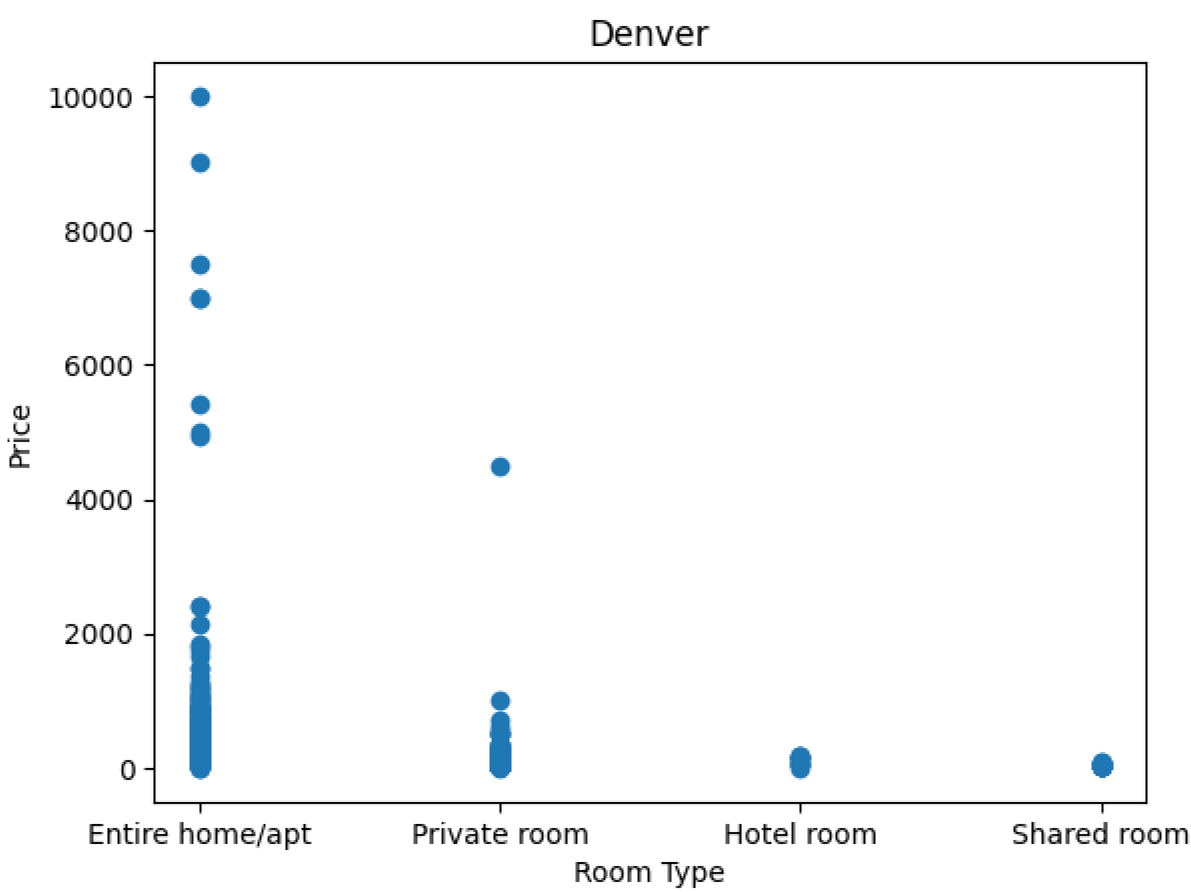

Broward County
Price
100000
80000
60000
40000
20000
0
Entire home/apt
Private room
Hotel room
Shared room
Room Type

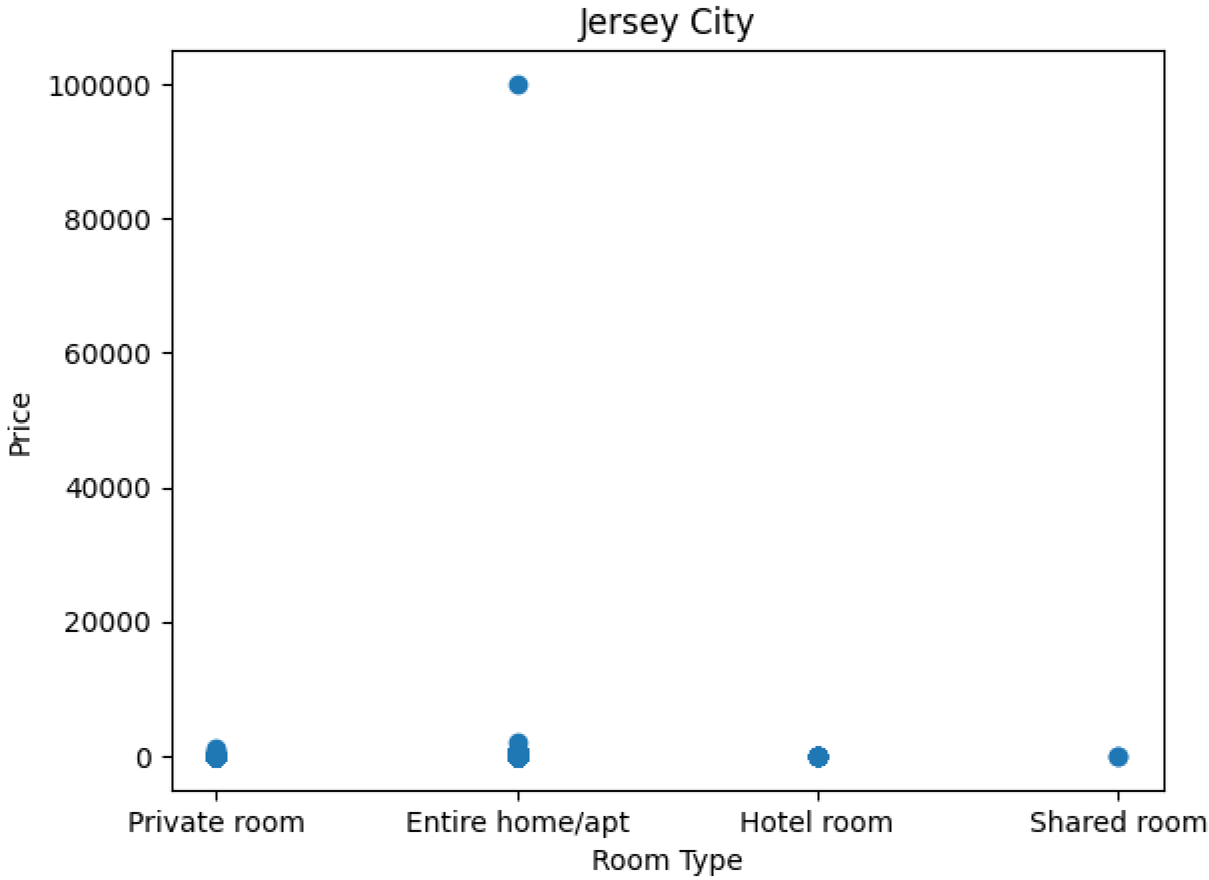

Cambridge
Price
1400
1200
1000
800
600
400
200
0
Entire home/apt
Private room
Shared room
Room Type

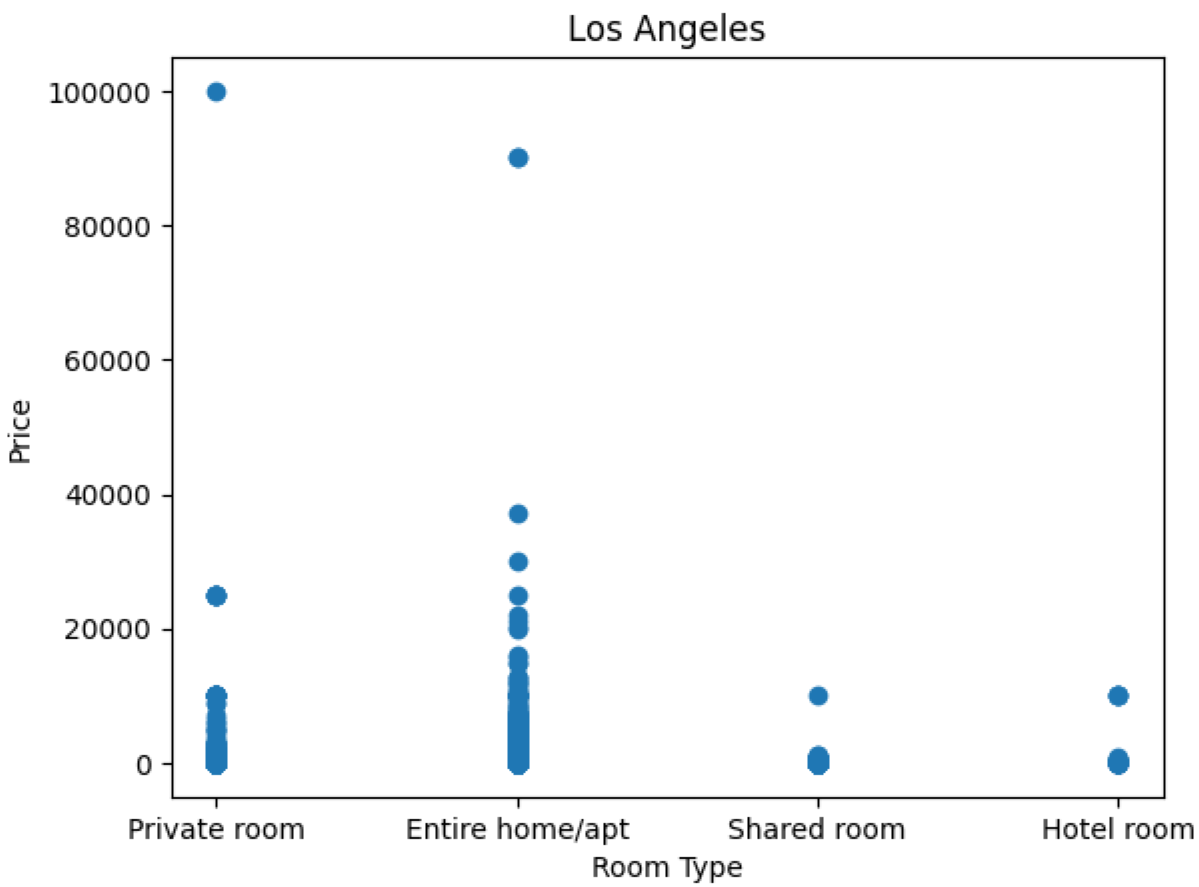

Chicago
100000
80000
60000
40000
20000
0
Price
Private room
Entire home/apt
Shared room
Hotel room
Room Type

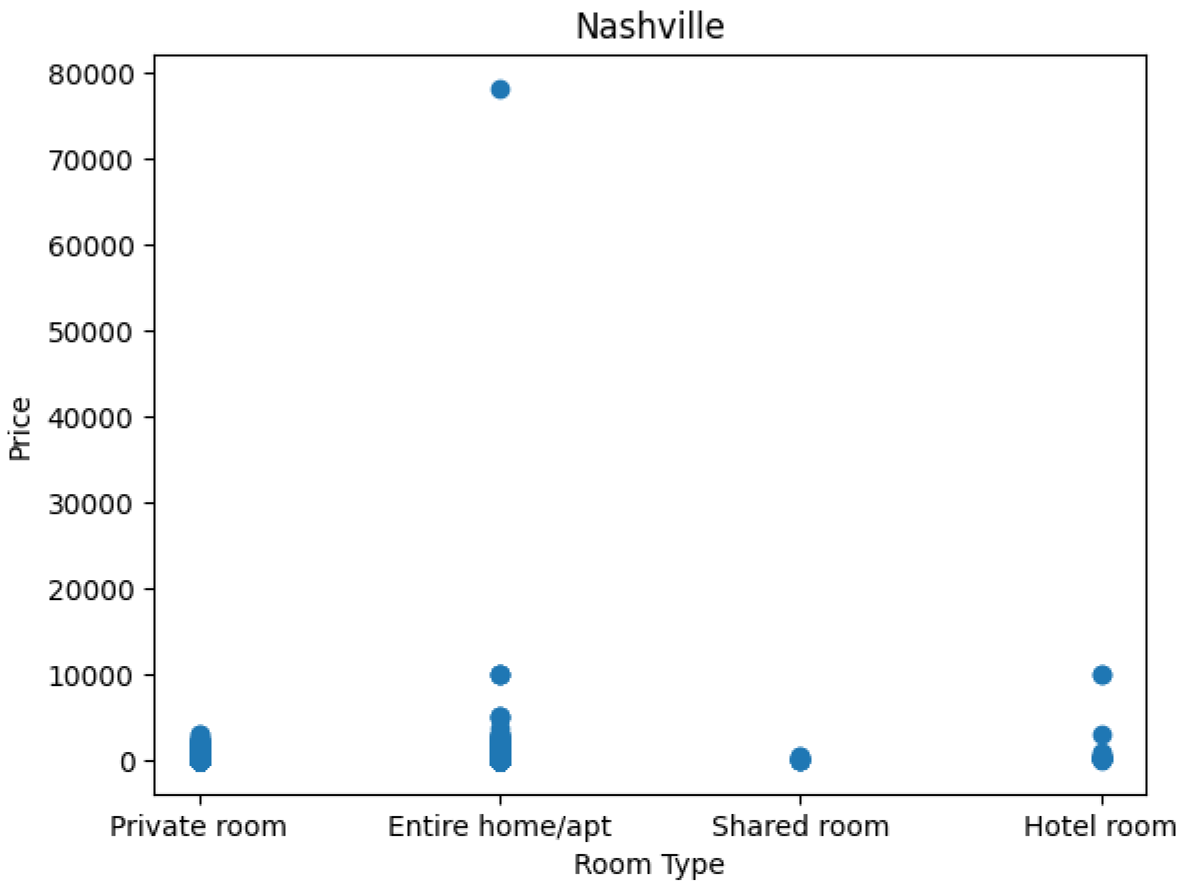

Clark County
80000
60000
40000
20000
0
Price
Entire home/apt
Private room
Hotel room
Shared room
Room Type

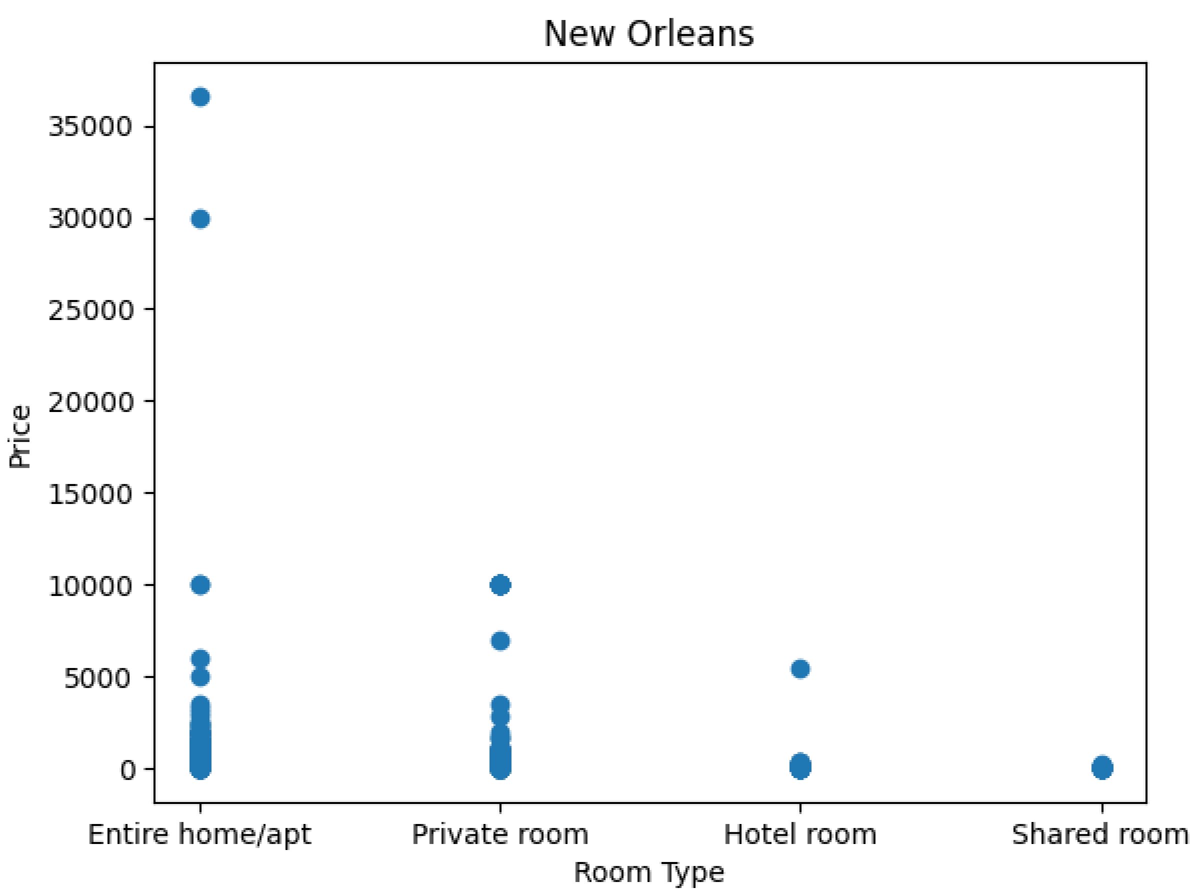

Columbus
Price
0
200
400
600
800
1000
1200
Entire home/apt
Private room
Shared room
Hotel room
Room Type

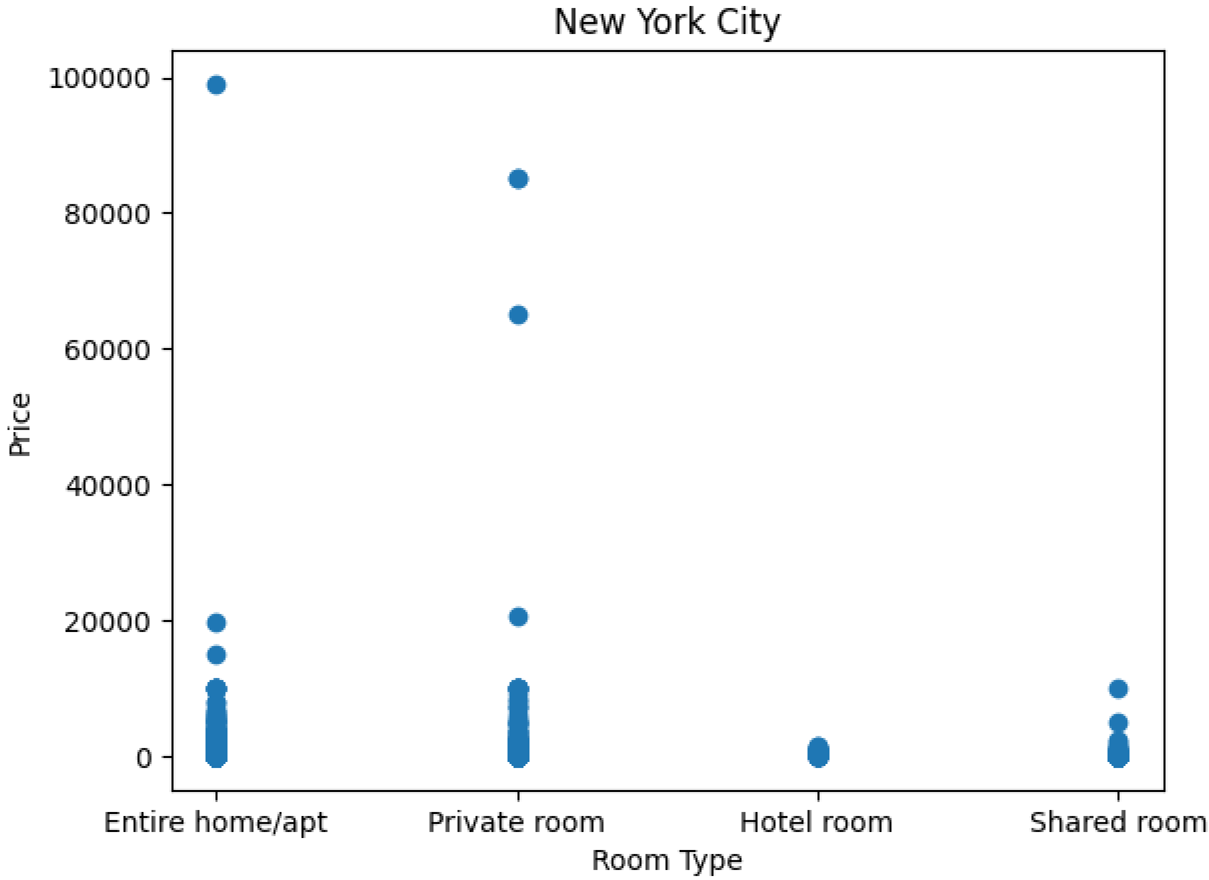

Denver
Price
0
2000
4000
6000
8000
10000
Entire home/apt
Private room
Hotel room
Shared room
Room Type

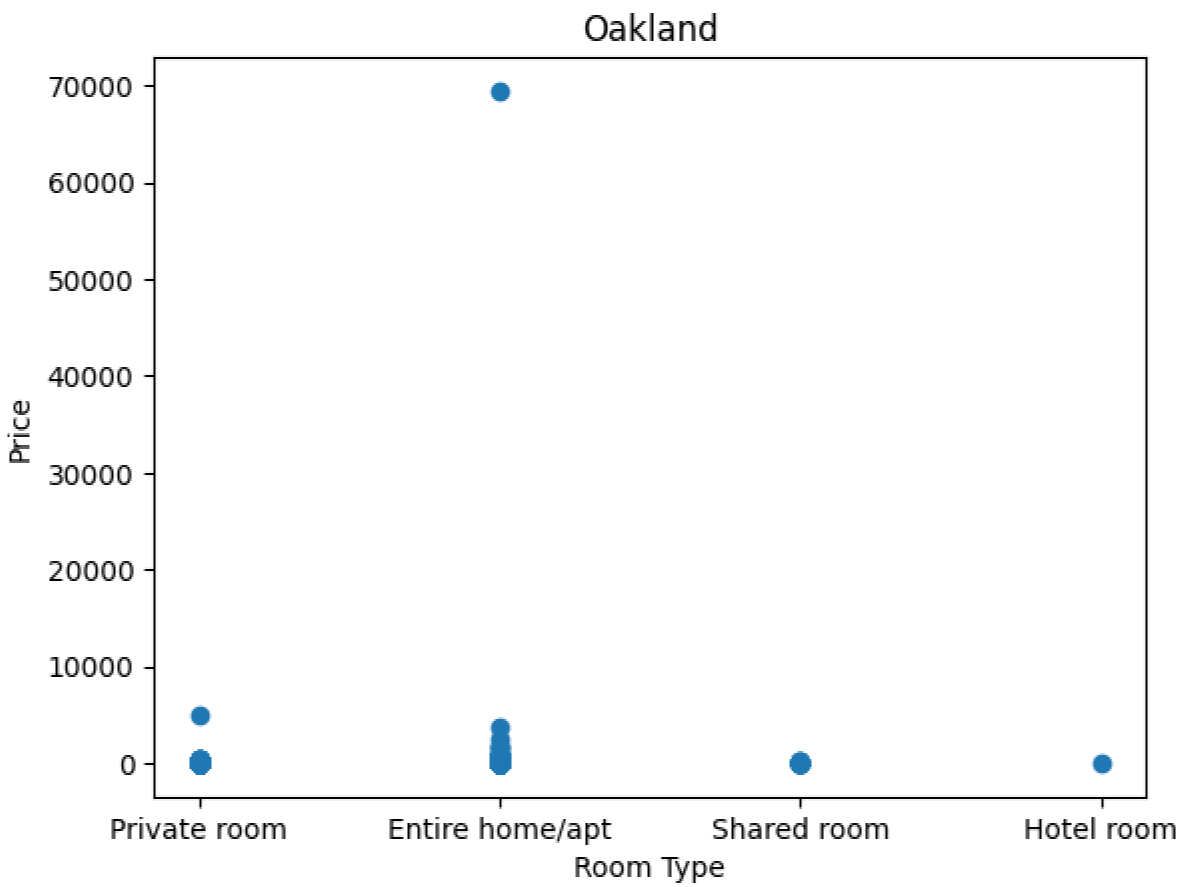

Jersey City
Price
100000
80000
60000
40000
20000
0
Private room
Entire home/apt
Hotel room
Shared room
Room Type

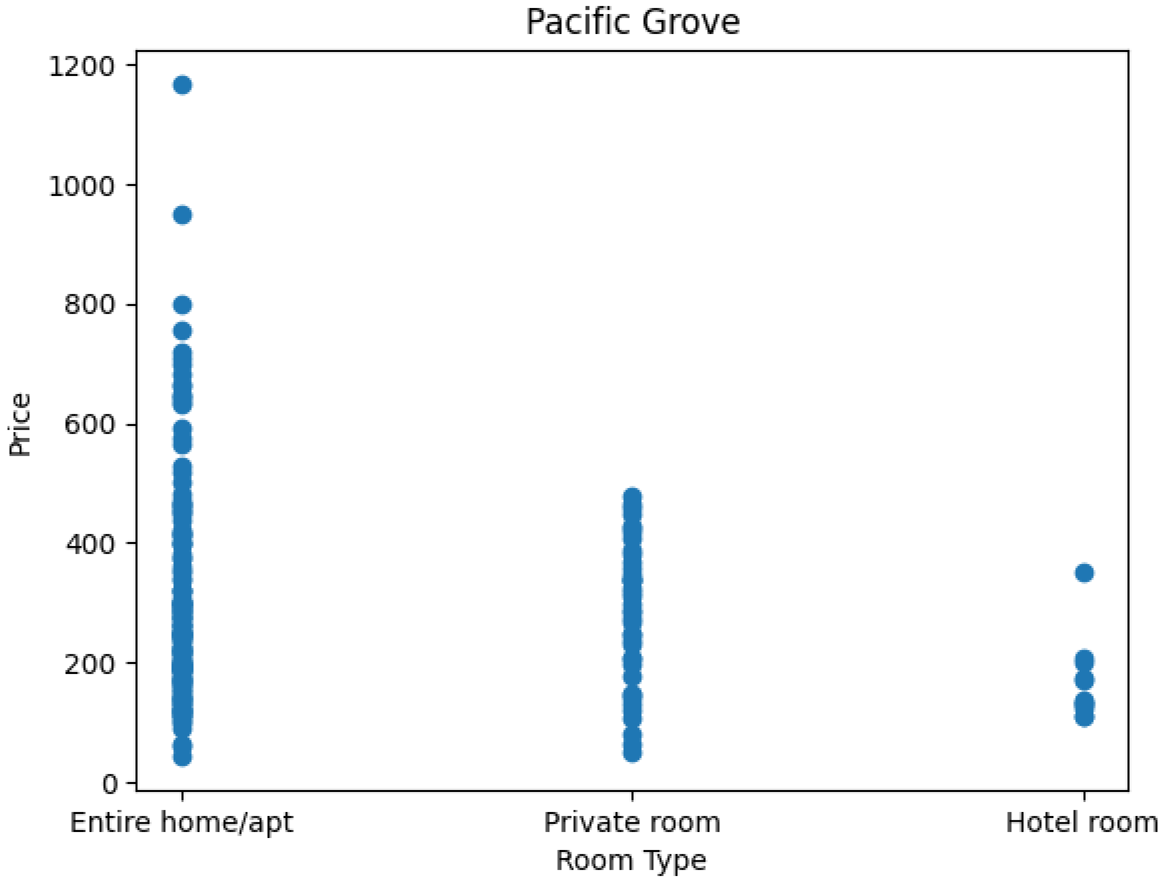

Los Angeles
Price
100000
80000
60000
40000
20000
0
Private room
Entire home/apt
Shared room
Hotel room
Room Type

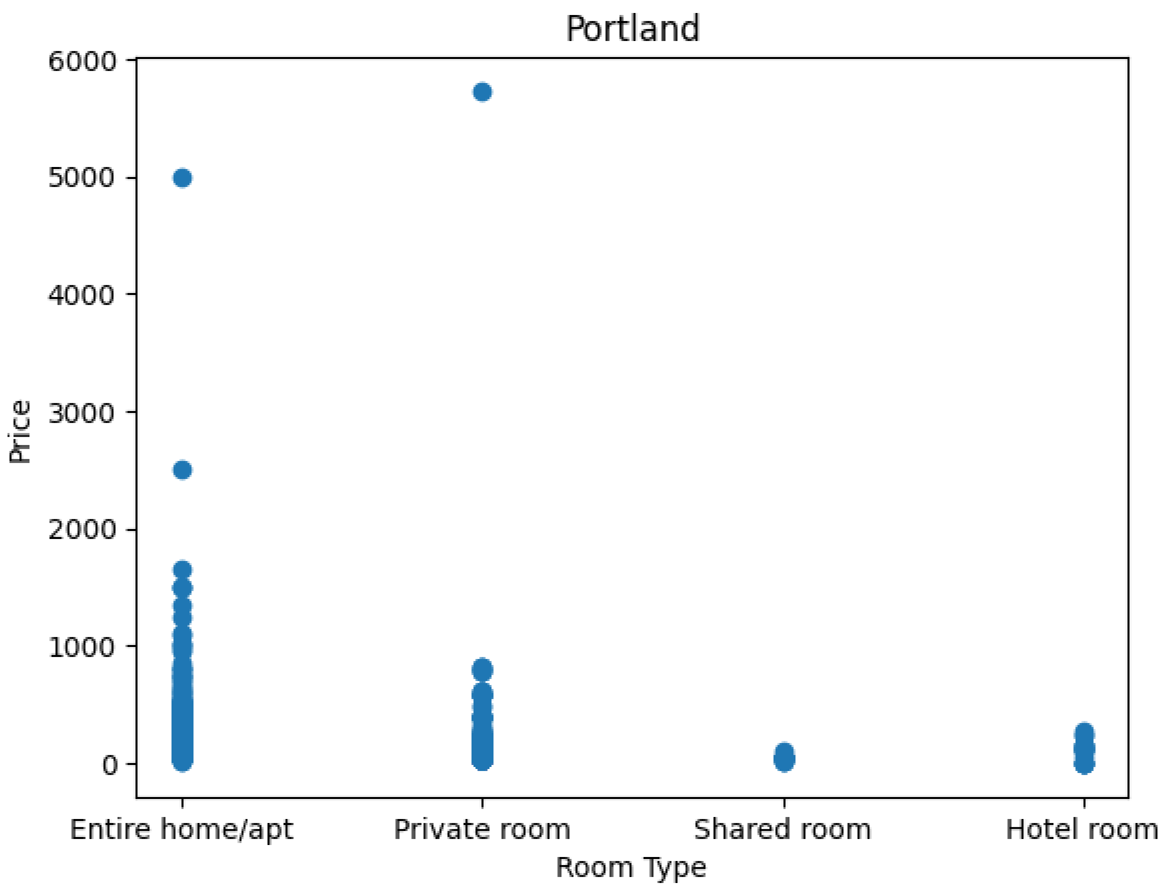

Nashville
80000
70000
60000
50000
40000
30000
20000
10000
0
Price
Private room
Entire home/apt
Shared room
Hotel room
Room Type

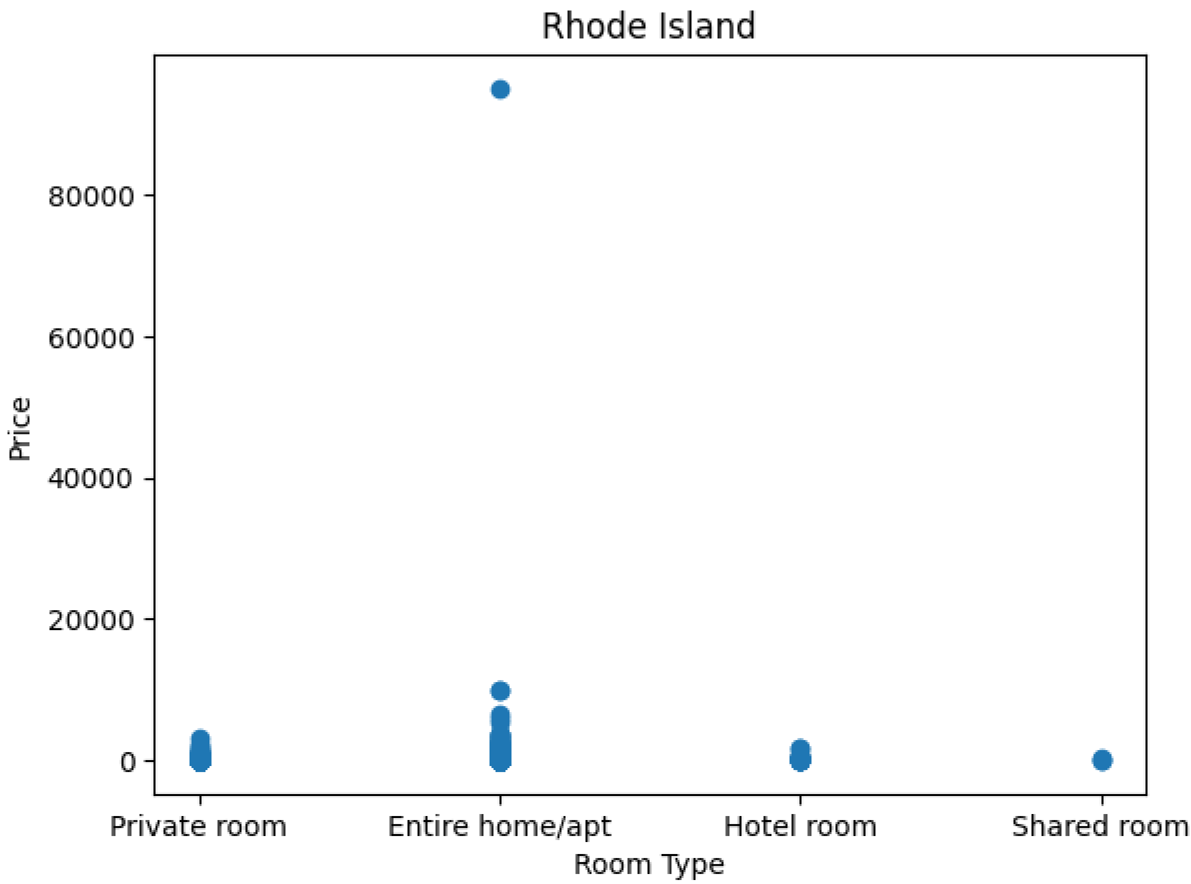

New Orleans
35000
30000
25000
20000
15000
10000
5000
0
Price
Entire home/apt
Private room
Hotel room
Shared room
Room Type

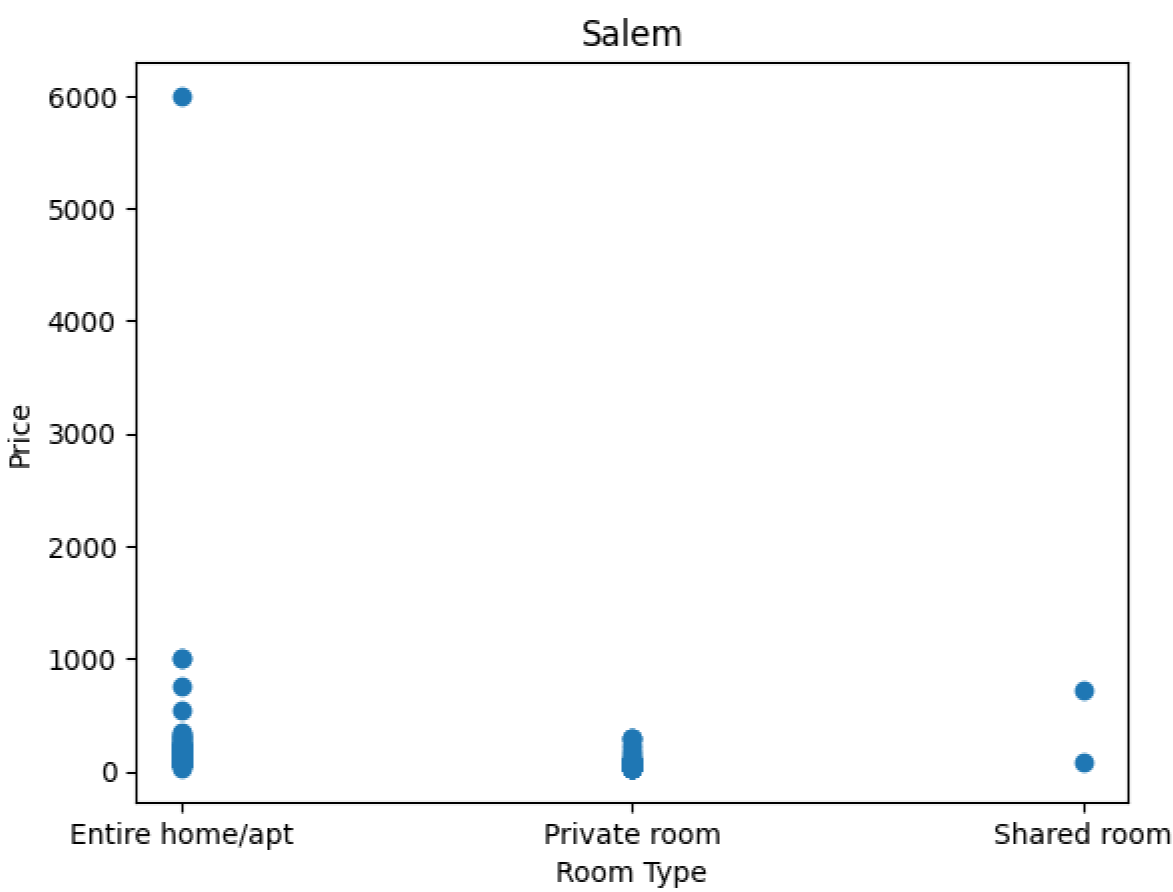

New York City
Price
100000
80000
60000
40000
20000
0
Entire home/apt
Private room
Hotel room
Shared room
Room Type

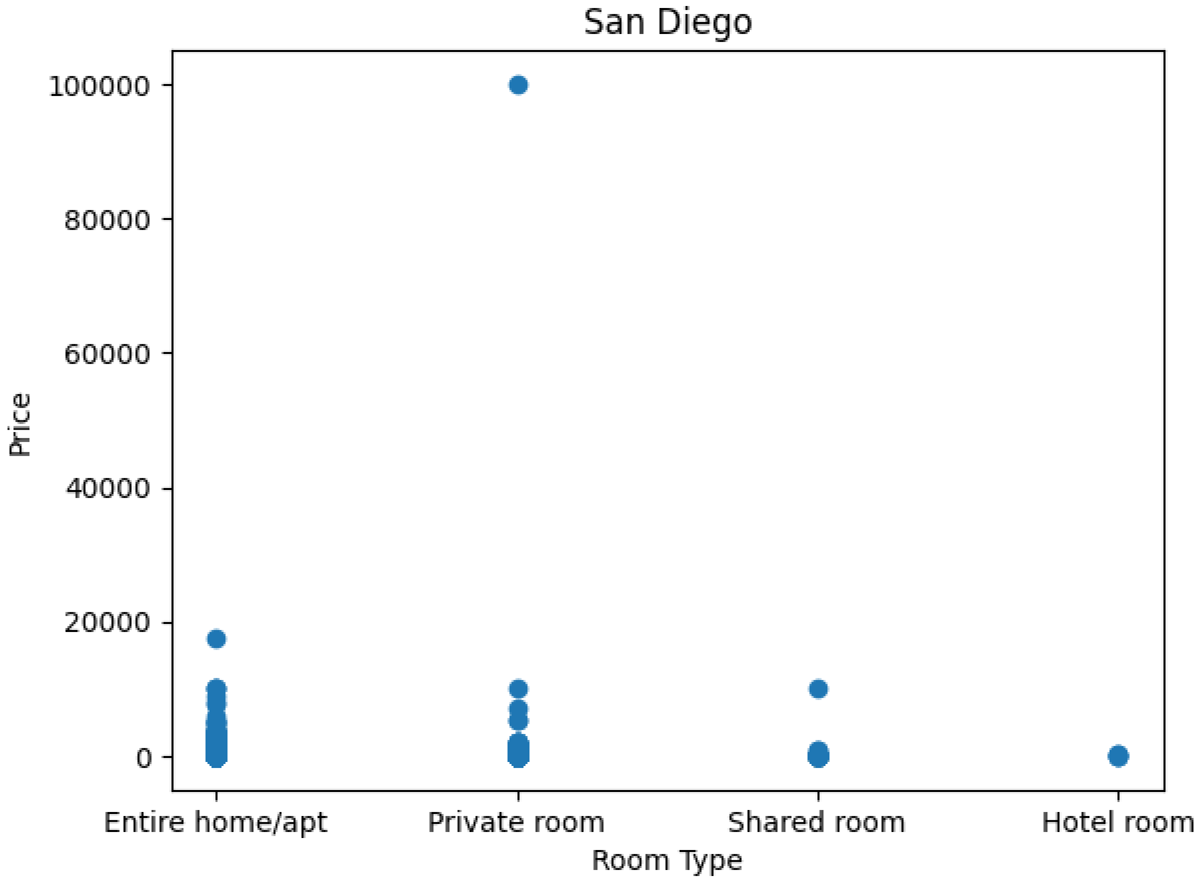

Oakland
Price
70000
60000
50000
40000
30000
20000
10000
0
Private room
Entire home/apt
Shared room
Hotel room
Room Type

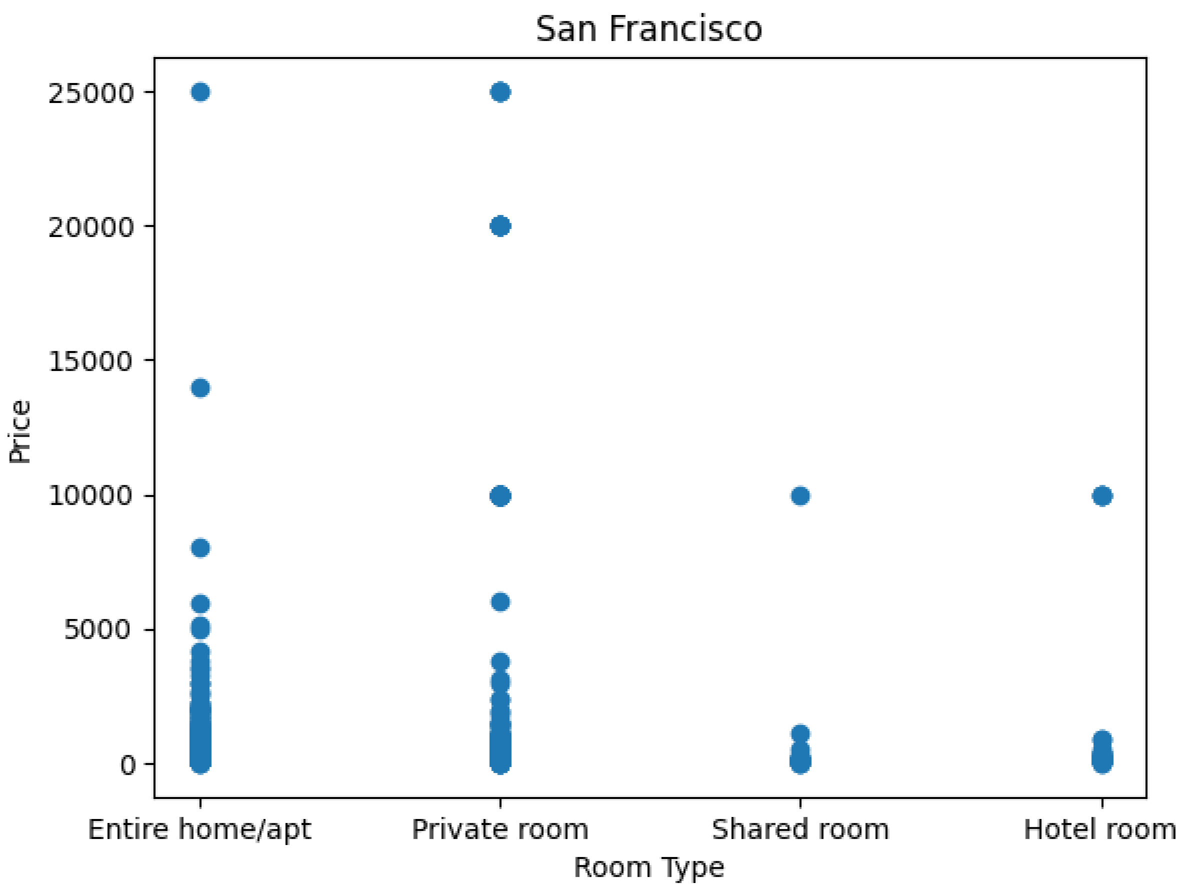

Pacific Grove
Price
Room Type
0
200
400
600
800
1000
1200
Entire home/apt
Private room
Hotel room

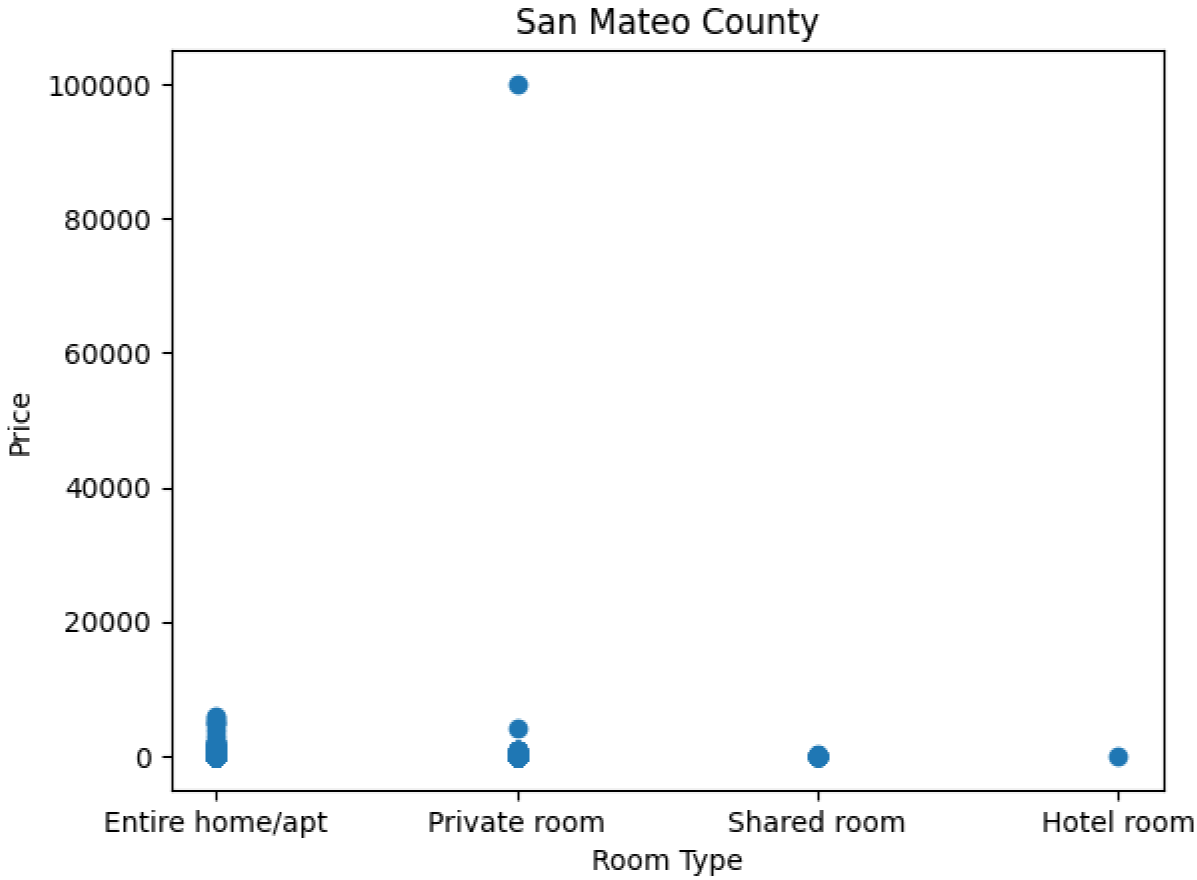

Portland
Price
Room Type
0
1000
2000
3000
4000
5000
6000
Entire home/apt
Private room
Shared room
Hotel room

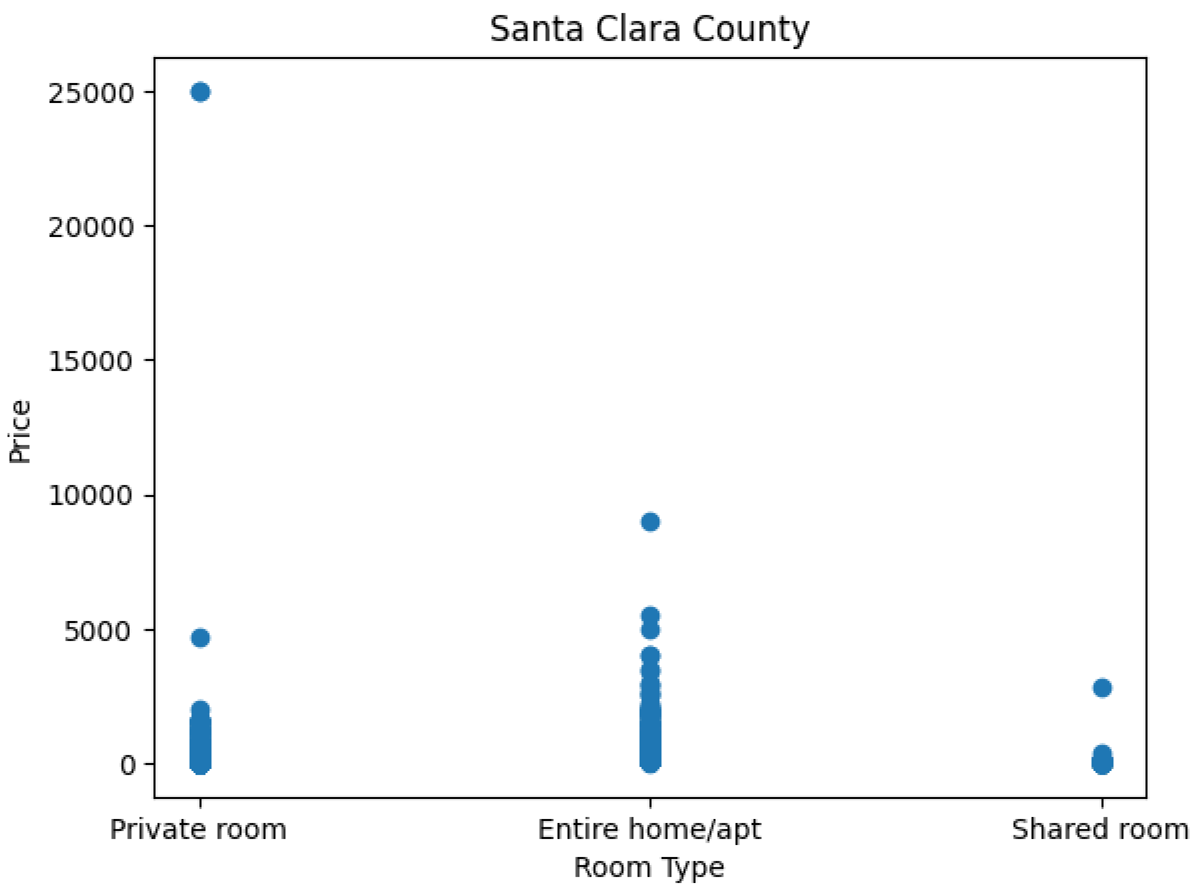

Rhode Island
Price
80000
60000
40000
20000
0
Private room
Entire home/apt
Hotel room
Shared room
Room Type

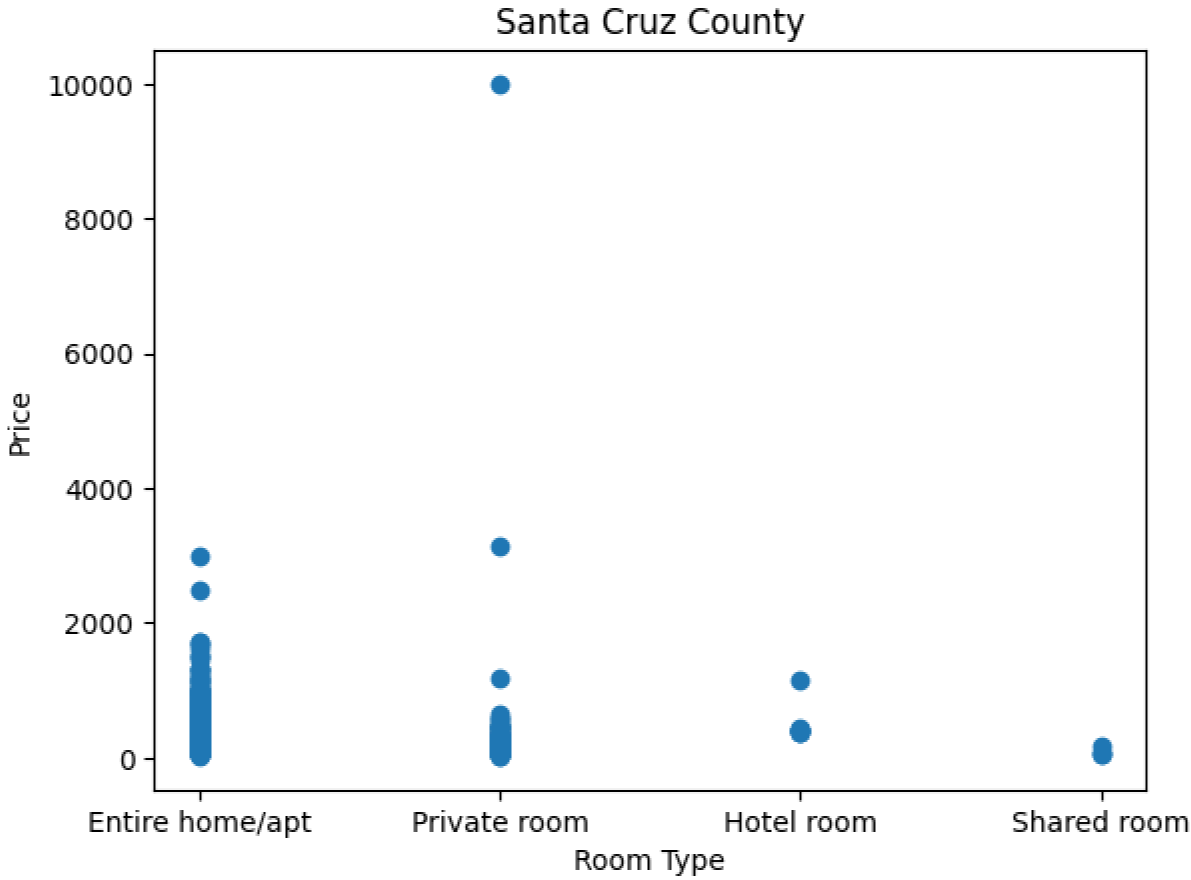

Salem
Price
6000
5000
4000
3000
2000
1000
0
Entire home/apt
Private room
Shared room
Room Type

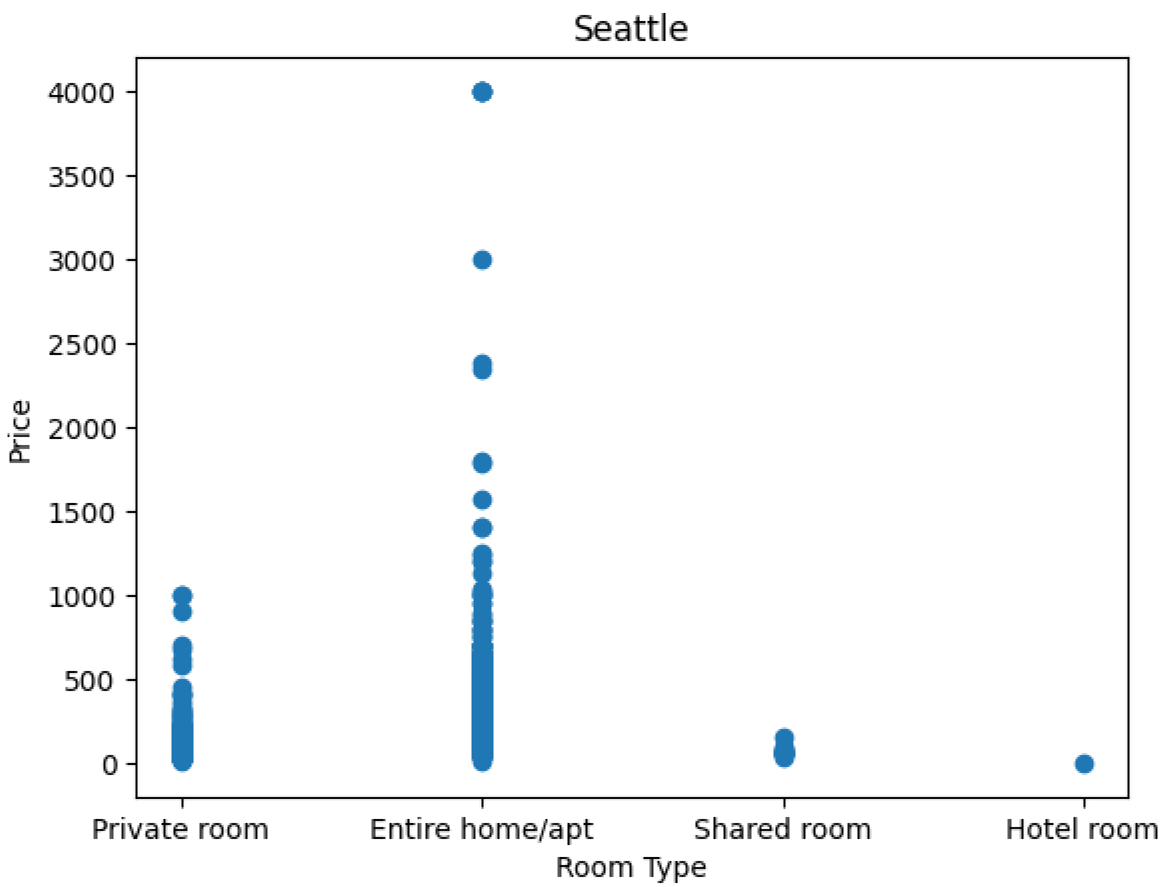

San Diego
Price
100000
80000
60000
40000
20000
0
Entire home/apt
Private room
Shared room
Hotel room
Room Type

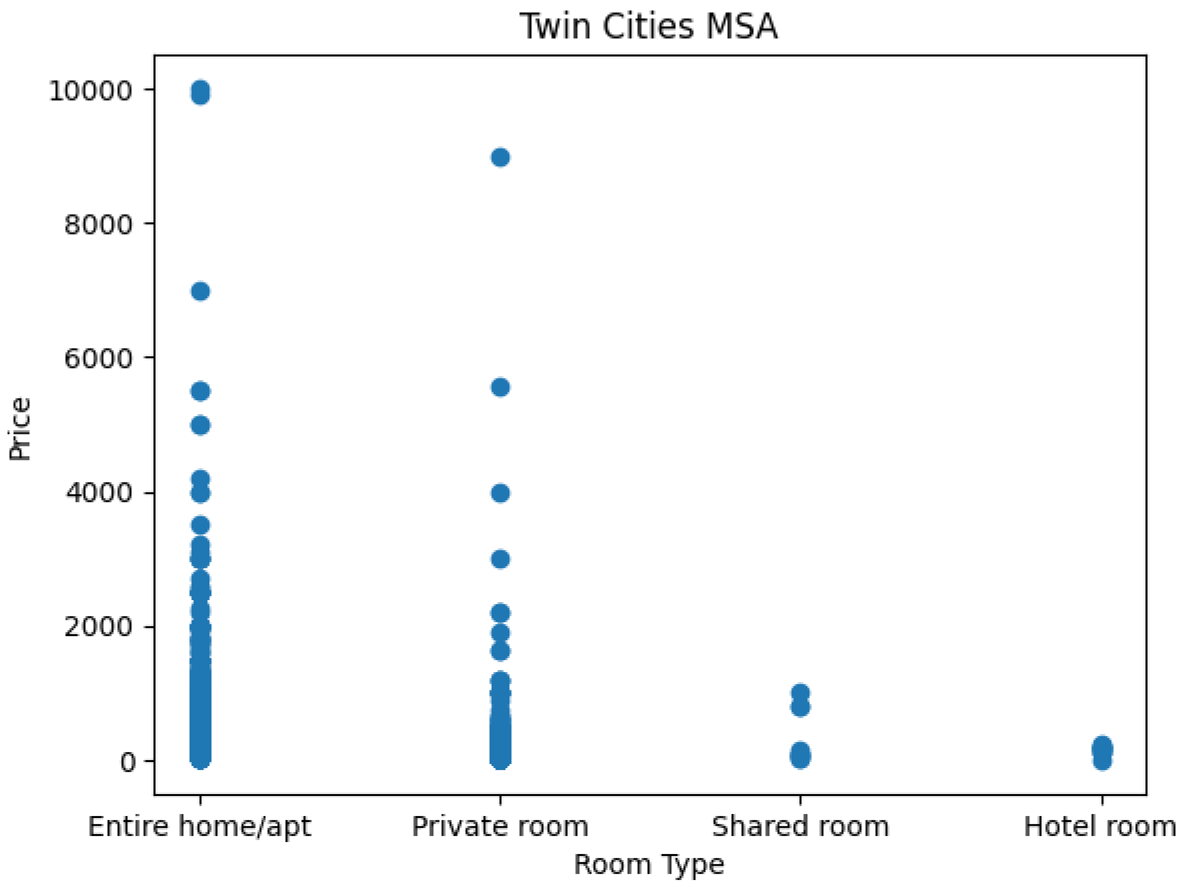

San Francisco
Price
25000
20000
15000
10000
5000
0
Entire home/apt
Private room
Shared room
Hotel room
Room Type

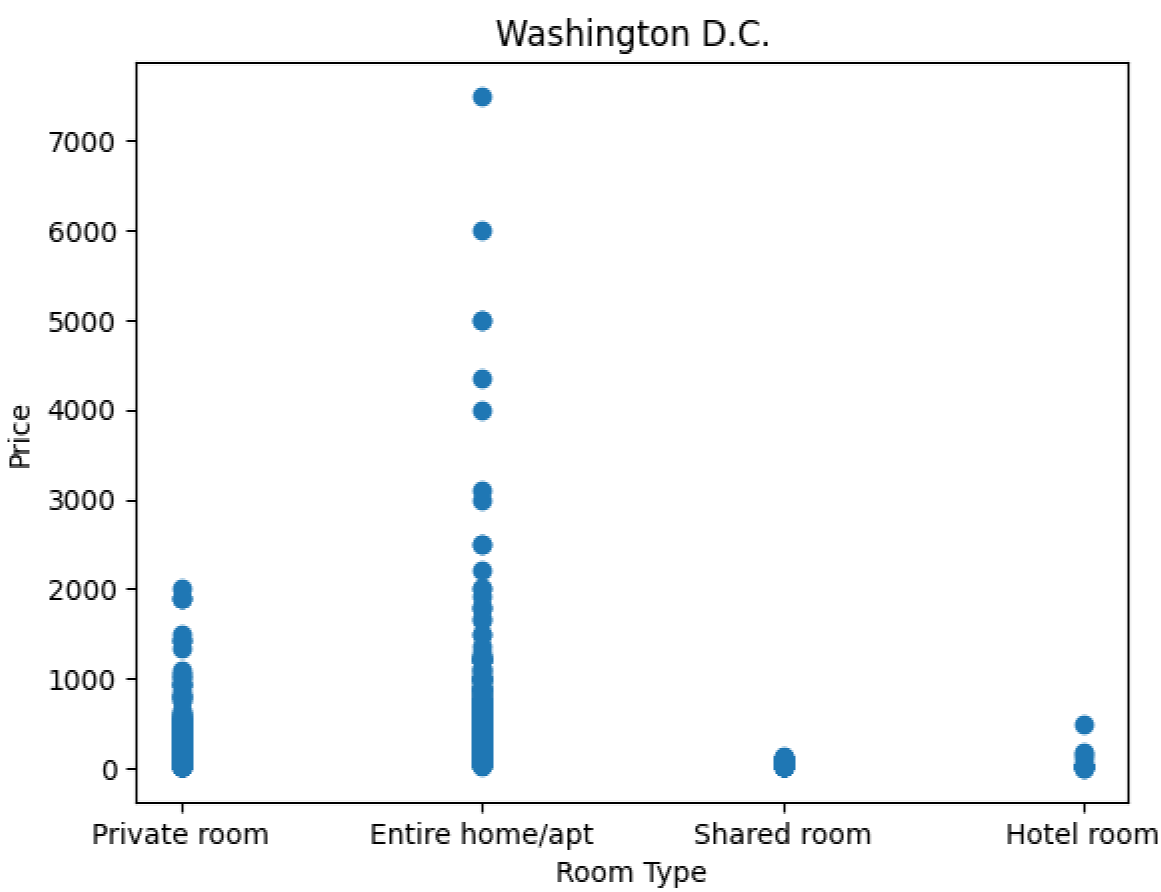

San Mateo County
100000
80000
60000
40000
20000
0
Price
Entire home/apt
Private room
Shared room
Hotel room
Room Type


Santa Clara County
25000
20000
15000
10000
5000
0
Price
Private room
Entire home/apt
Shared room
Room Type


Santa Cruz County
Price
Room Type
10000
8000
6000
4000
2000
0
Entire home/apt
Private room
Hotel room
Shared room

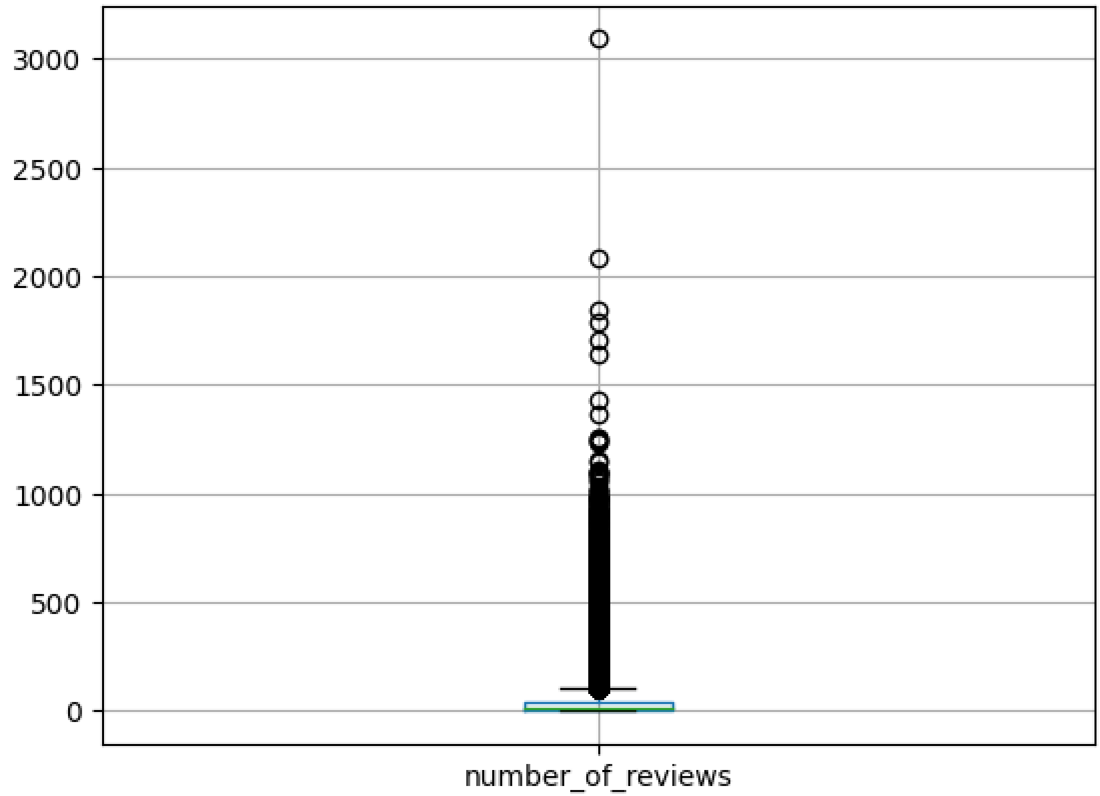

Seattle
Price
Room Type
4000
3500
3000
2500
2000
1500
1000
500
0
Private room
Entire home/apt
Shared room
Hotel room



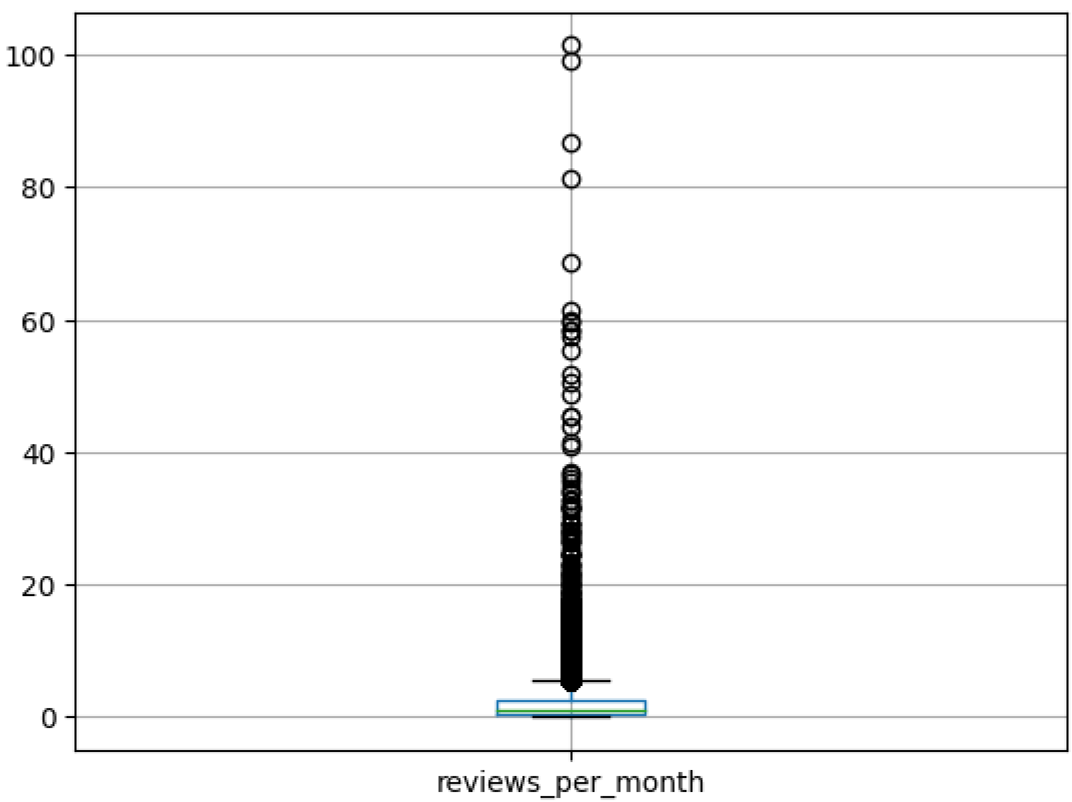


When comparing the prices of different room types across various cities, it was observed that for most cities, entire homes/apartments are the most expensive, followed by private rooms, hotels, and shared rooms, in descending order of price.

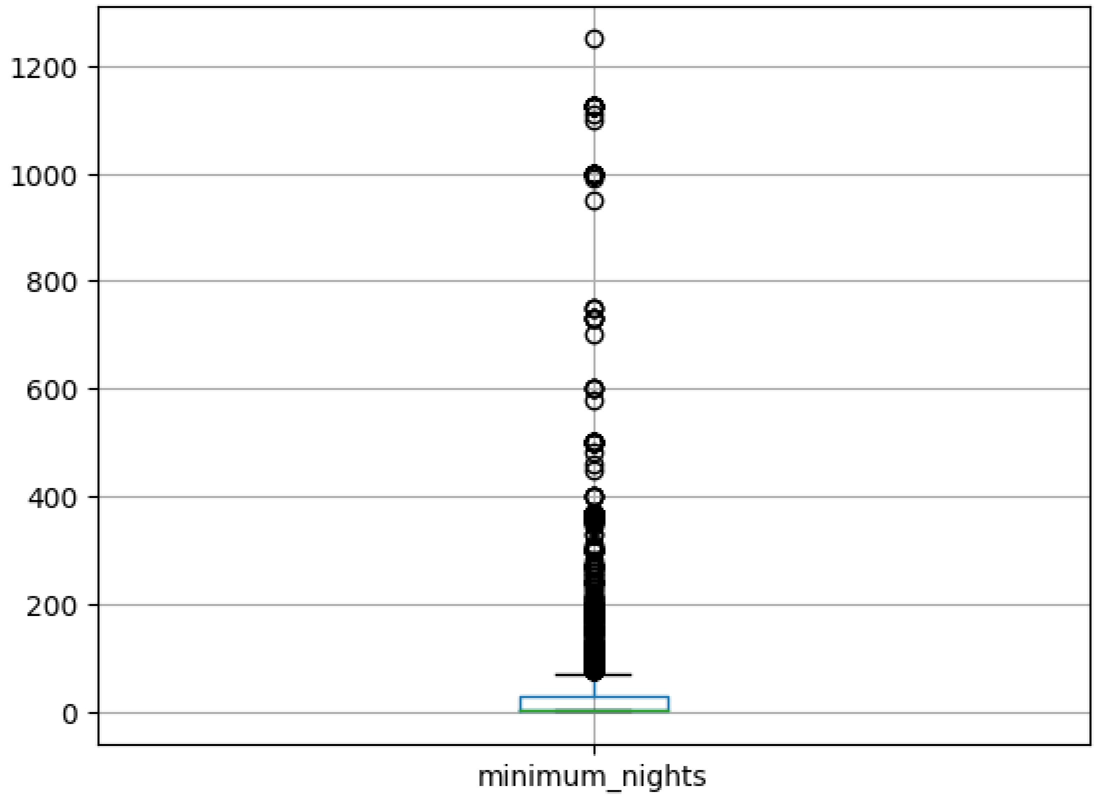
In [11]:
```
details.boxplot(column=['number_of_reviews']);
```



In [12]:
```
details.boxplot(column=['reviews_per_month']);
```



In [13]:
```
details.boxplot(column=['minimum_nights']);
```



We have generated distinct boxplots for minimum nights, reviews per month, and total reviews to compare and test the presence of outliers in our dataset.

b) Identify data quality issues and build the data quality plan. [0.5]

To identify issues with data quality, visualizations can be utilized to detect various anomalies, such as missing values, outliers, and cardinality issues.
The first issue is missing values, which occurs when a considerable amount of data is absent or unknown from a particular column. By visualizing these values, we can detect their quality and identify potential gaps in the data.
The second issue is outliers, which refers to data that has values that are outside the expected range. These anomalies could be caused by junk or irregular data, or by a specific value that is different from the range of values in the data. It is essential to detect and address such anomalies, as they can significantly degrade the overall quality of the data.
The third issue is cardinality, which refers to the diversity of values in a particular dataset. Data with just one specific value may not be useful, while data with an excessive amount of diverse values may make it difficult to predict values accurately. Identifying such irregular data is crucial, as it can degrade the quality of the data and render it less useful for analysis.

In [14]:
```
details.isnull().sum()
```

Out[14]:
```
id                                0
name                             16
host_id                           0
host_name                        13
neighbourhood_group          135647
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
```

```
number_of_reviews                        0
last_review                          49085
reviews_per_month                    49085
calculated_host_listings_count           0
availability_365                         0
number_of_reviews_ltm                    0
city                                     0
dtype: int64
```

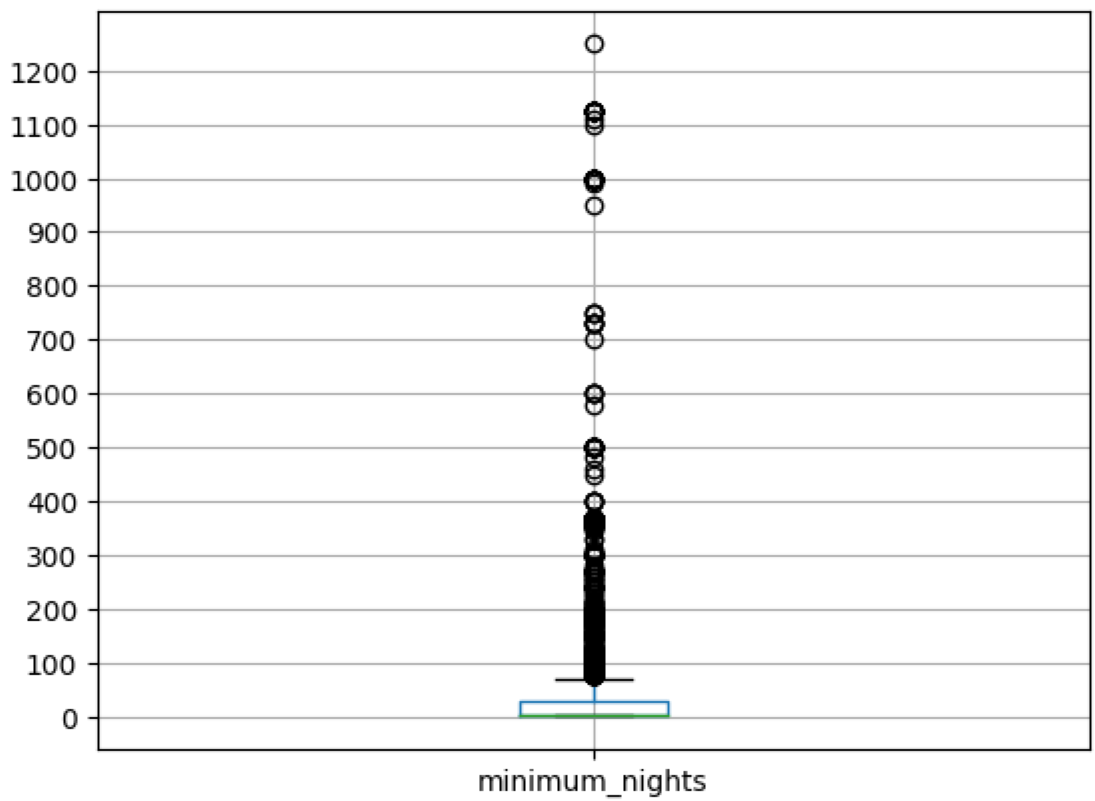
Of all the columns required for our price prediction, most of the data does not contain any null values, with the exception of the neighborhood group column.
Moving forward, we will examine the data for outliers in the room_type, minimum_nights, and price columns.

In [15]:
```
room_type_counts = details.groupby('room_type').size()
print(room_type_counts)
```

```
room_type
Entire home/apt    169142
Hotel room            970
Private room        59759
Shared room          2276
dtype: int64
```

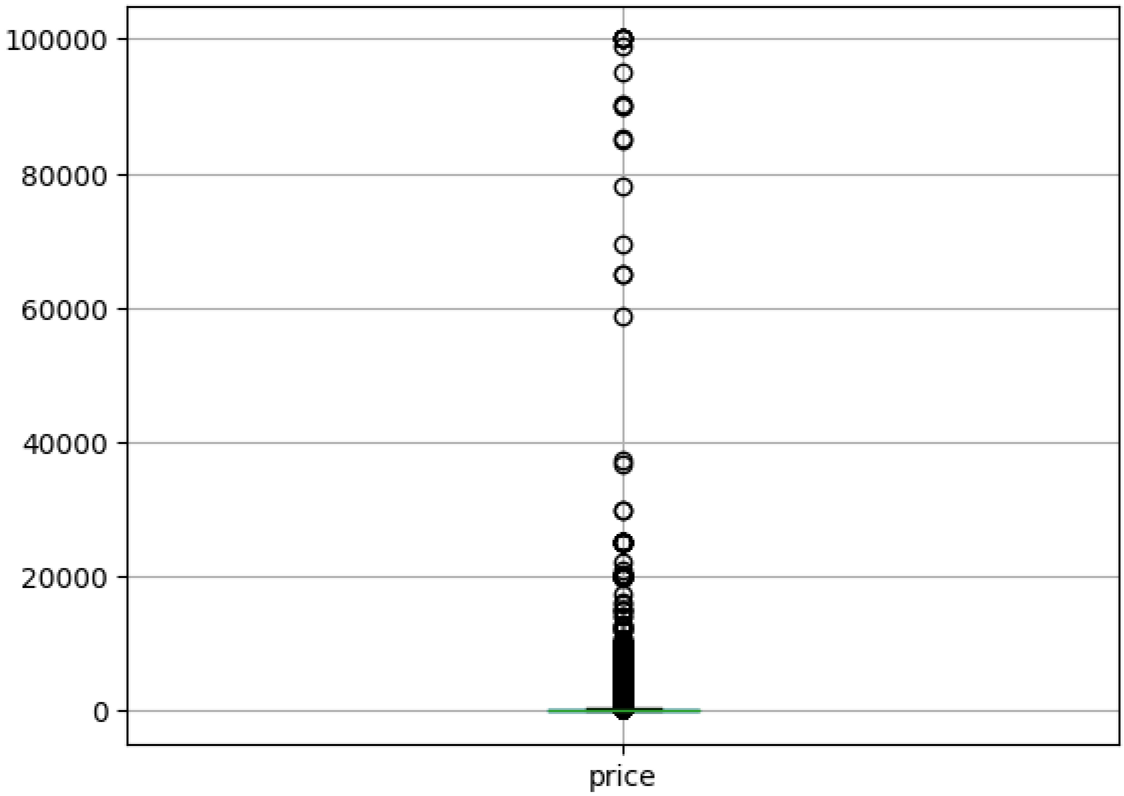
The room_type column does not exhibit any outliers in particular. It Simply contains four distinct types of data: Entire home/apt, Hotel room, Private room, and Shared room.

In [16]:
```
import numpy as np
details.boxplot(column=['minimum_nights']);
pyplot.yticks(np.arange(0, 1300, 100));
```



the Minimum Nights column contains a large number of outlier values.

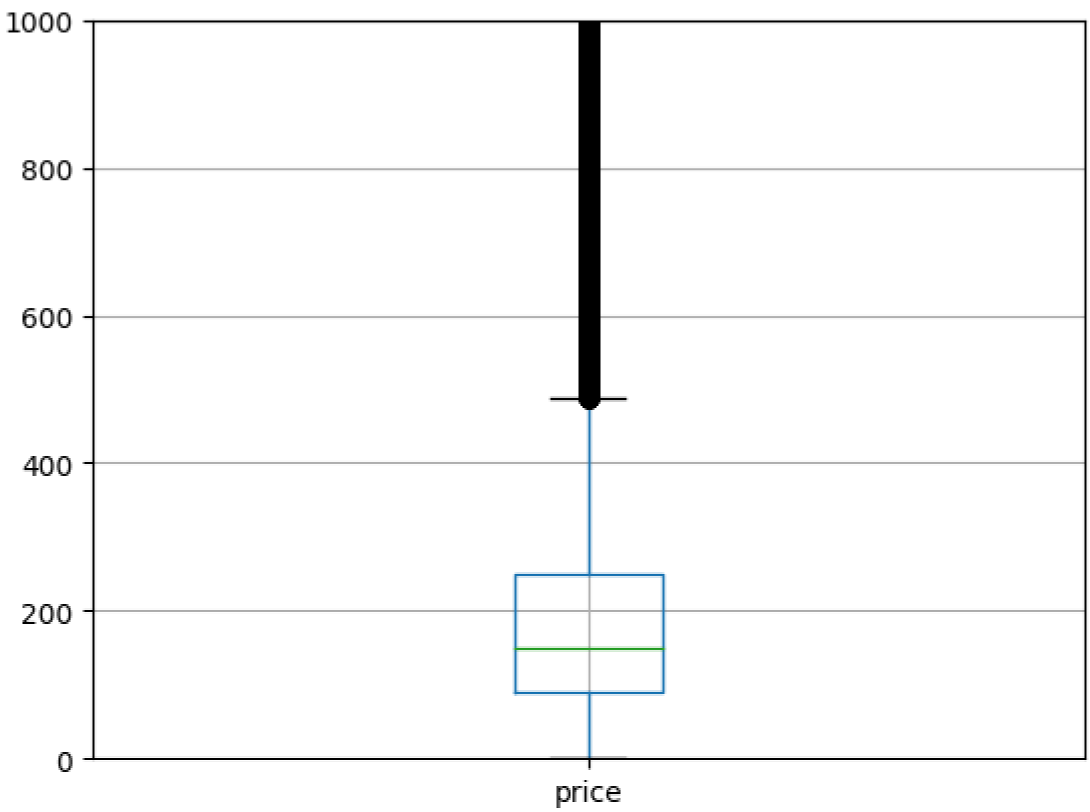
In [17]:
```
details.boxplot(column=['price']);
```



As can be seen, the price column contains a large number of outlier values. Let us further examine this issue.

In [18]:
```
details.boxplot(column=['price']);
pyplot.ylim(0, 1000)
```

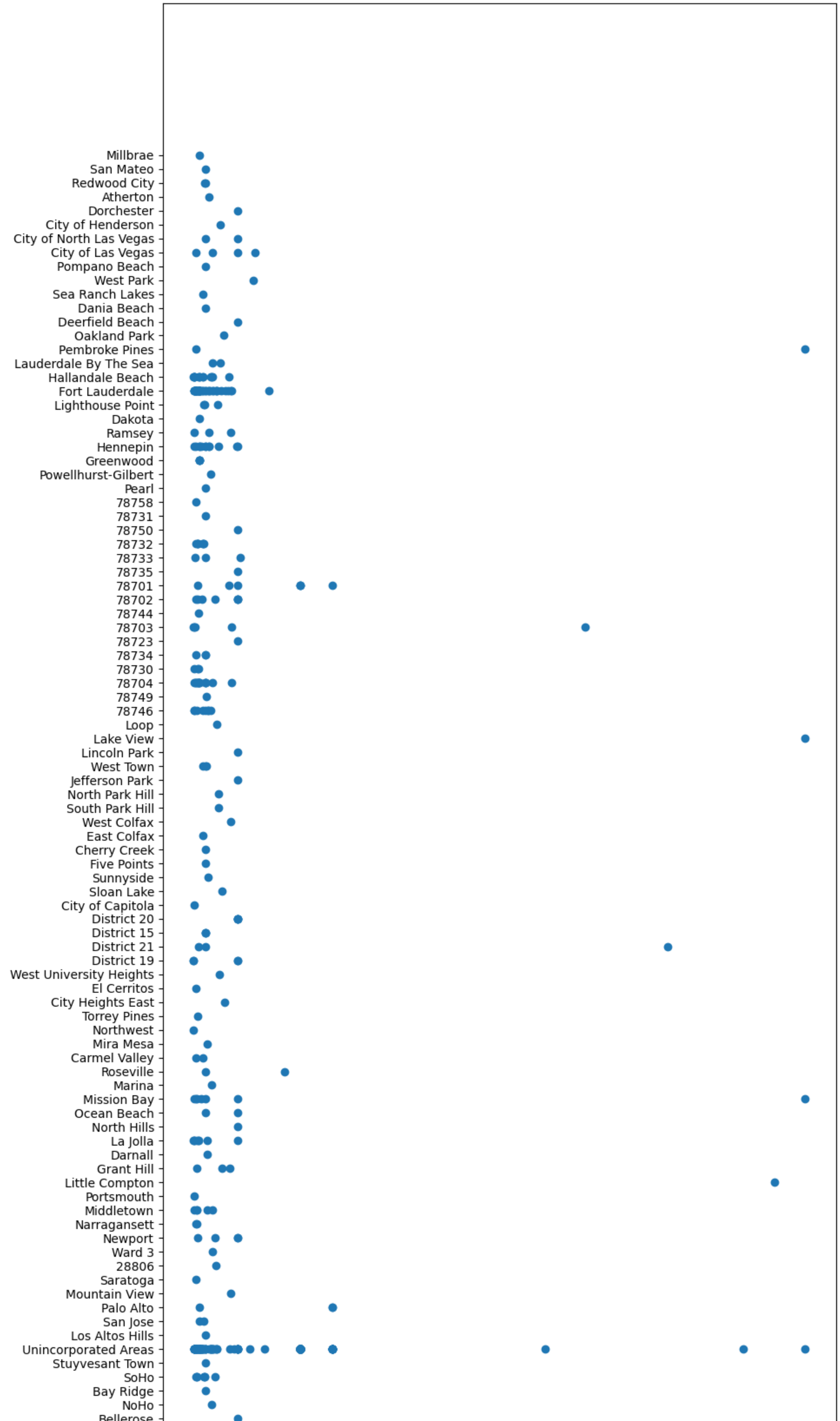
Out[18]:
```
(0.0, 1000.0)
```



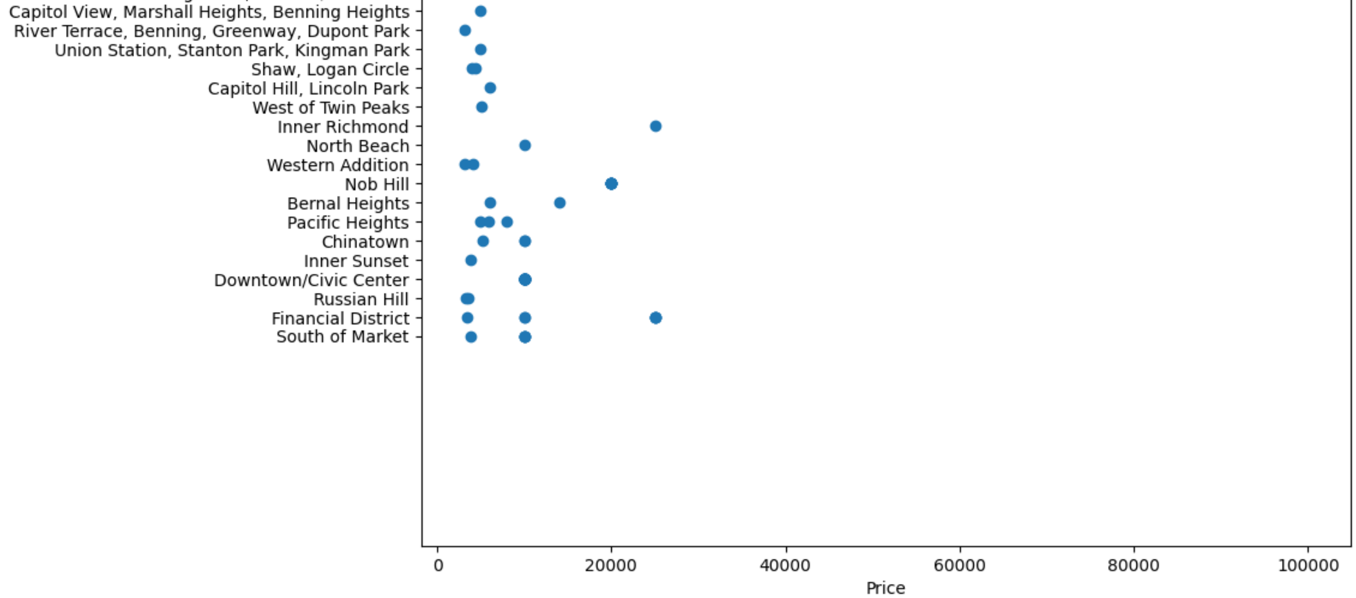
In [19]:
```
fig, ax = pyplot.subplots(figsize=(10, 50))
outliers = details[details['price'] > 3000]
pyplot.scatter(outliers['price'], outliers['neighbourhood'])
```

```
pyplot.xlabel('Price')
pyplot.ylabel('neighbourhood')
```

Out[19]: Text(0, 0.5, 'neighbourhood')




neighbourhood
Nolita
Prospect Heights
Bushwick
Bergen Beach
Greenpoint
Sunset Park
Park Slope
South Ozone Park
Laurelton
Longwood
Flatiron District
Sheepshead Bay
Chelsea
Washington Heights
Richmond Hill
Williamsburg
Midtown
Crown Heights
Murray Hill
Upper East Side
Jackson Heights
Briarwood
Stapleton
West Village
Harlem
Canarsie
West Brighton
Greenwich Village
Tribeca
Lower East Side
Long Island City
Gramercy
Clinton Hill
East Village
Upper West Side
East Harlem
Artesia
Azusa
Hawthorne
Torrance
Fairfax
Reseda
Topanga
Woodland Hills
West Compton
Mar Vista
West Carson
San Gabriel
Pico Rivera
Palos Verdes Estates
Santa Monica
Hancock Park
Venice
Long Beach
Koreatown
Rancho Palos Verdes
Tarzana
Palms
Southeast Antelope Valley
Monterey Park
Pacific Palisades
Los Feliz
Hollywood Hills
Hermosa Beach
Covina
Playa del Rey
Manhattan Beach
North Hollywood
Encino
Burbank
Unincorporated Santa Monica Mountains
Studio City
East Los Angeles
Valley Village
Rosemead
Downtown
Pasadena
Bellflower
Bel-Air
Marina del Rey
Glendale
West Hollywood
Universal City
Hollywood
Valinda
Westlake Village
Beverly Grove
Canoga Park
Sherman Oaks
Beverly Crest
Alhambra
Malibu
Hollywood Hills West
Beverly Hills
Brentwood
Central City
East Carrollton
French Quarter
Central Business District
New Aurora - English Turn
Mid-City
Bywater
Marigny
Mcdonogh
Leonidas
Ward E (councilmember James Solomon)
Bancroft Business-Havenscourt
Lynn-Highland Park
Highland Terrace
Georgetown, Burleith/Hillandale



We visualized the neighbourhood groups with outliers

In [20]:
```
outliers[['price', 'neighbourhood']].sort_values('price')
```

Out[20]:

	price	neighbourhood
109487	3006	Upper West Side
148038	3012	District 19
137989	3018	Northwest
172783	3036	78703
54534	3036	Unincorporated Santa Monica Mountains
201784	3045	Hallandale Beach
24954	3049	Malibu
129266	3049	La Jolla
141506	3057	District 19
6565	3093	Western Addition
9162	3100	River Terrace, Benning, Greenway, Dupont Park
190492	3100	Hennepin
207925	3104	Fort Lauderdale
140318	3105	La Jolla
126153	3107	Portsmouth
62228	3122	West Hollywood
59878	3122	West Hollywood
221106	3125	Unincorporated Areas
30575	3125	Malibu
57214	3126	West Hollywood
60094	3128	West Carson
168761	3135	78730

223133	3136	Unincorporated Areas
27932	3138	Malibu
214880	3142	Unincorporated Areas
214100	3142	Unincorporated Areas
207645	3143	Fort Lauderdale
204633	3143	Fort Lauderdale
221674	3143	Unincorporated Areas
211775	3143	Unincorporated Areas
103123	3143	Nolita
103264	3143	Chelsea
26673	3143	Beverly Hills
149617	3143	City of Capitola
64138	3150	Beverly Hills
207115	3154	Hollywood
134921	3154	Mission Bay
170703	3166	78746
168057	3170	78704
136678	3185	La Jolla
32518	3186	Malibu
178232	3186	78746
107652	3196	Midtown
202261	3200	Hallandale Beach
64327	3200	Beverly Hills
190375	3200	Ramsey
27192	3200	Beverly Hills
60052	3200	Malibu
60049	3200	Malibu
43846	3200	Palms
41236	3200	Hollywood
40652	3200	Burbank
82420	3200	Harlem
127303	3200	Middletown
206532	3200	Fort Lauderdale
164837	3200	78703
198143	3206	Hallandale Beach
198147	3206	Hallandale Beach
201040	3206	Hollywood
223786	3207	Unincorporated Areas

223575	3207	Unincorporated Areas
202113	3208	Hallandale Beach
56276	3230	Malibu
53607	3243	Bel-Air
32655	3250	Bel-Air
1537	3250	Russian Hill
35570	3250	Bel-Air
35136	3250	Beverly Hills
24826	3280	Malibu
176758	3283	78703
218235	3283	Unincorporated Areas
220082	3286	Unincorporated Areas
90690	3300	West Village
61330	3300	Beverly Hills
57621	3300	Beverly Hills
175772	3300	78704
22502	3311	Central City
39898	3314	Hermosa Beach
52223	3325	Manhattan Beach
61548	3325	Malibu
31044	3333	Beverly Hills
34335	3334	Beverly Hills
42708	3339	Hermosa Beach
168342	3350	78733
200369	3354	Fort Lauderdale
38519	3356	Unincorporated Santa Monica Mountains
38989	3357	Encino
206575	3360	Fort Lauderdale
198517	3360	Fort Lauderdale
220437	3371	Unincorporated Areas
198721	3378	Fort Lauderdale
208854	3393	Hollywood
171406	3396	78732
24387	3400	Brentwood
113229	3400	Saratoga
38950	3400	North Hollywood
66149	3400	Beverly Crest

99015	3400	Park Slope
219650	3403	Unincorporated Areas
215255	3413	Unincorporated Areas
130486	3420	Midtown
134275	3433	Mission Bay
207337	3457	Fort Lauderdale
26102	3459	Hollywood Hills West
173145	3460	78758
198263	3462	Fort Lauderdale
21055	3464	French Quarter
65182	3474	West Hollywood
45891	3474	West Hollywood
59086	3478	West Hollywood
58549	3480	Hollywood Hills West
173081	3483	78734
107855	3488	SoHo
110892	3495	Unincorporated Areas
28847	3497	Hollywood Hills West
45415	3500	Tarzana
51203	3500	Hollywood
52331	3500	Hermosa Beach
33336	3500	Malibu
86474	3500	Williamsburg
31583	3500	Beverly Crest
56741	3500	Malibu
64093	3500	Hollywood Hills West
63477	3500	Hollywood Hills West
57845	3500	Malibu
30893	3500	Beverly Crest
91533	3500	Harlem
94712	3500	Midtown
60451	3500	Topanga
29101	3500	Hollywood Hills West
61993	3500	Hollywood Hills West
34767	3500	Beverly Grove
34959	3500	Malibu
29203	3500	Hollywood Hills West
38124	3500	Burbank

124789	3500	Narragansett
27444	3500	Westlake Village
27478	3500	Hollywood Hills West
202123	3500	Pembroke Pines
25593	3500	Hollywood Hills West
168864	3500	78704
212196	3500	City of Las Vegas
190124	3500	Hennepin
24808	3500	Malibu
25734	3500	Beverly Grove
213380	3500	Unincorporated Areas
167359	3500	78702
140216	3500	Carmel Valley
863	3500	Financial District
21381	3500	East Carrollton
139969	3500	El Cerritos
223297	3500	Unincorporated Areas
47197	3510	Pacific Palisades
24608	3514	Hollywood Hills West
62708	3518	Beverly Crest
131230	3528	Mission Bay
126884	3543	Middletown
125893	3544	Narragansett
214667	3545	Unincorporated Areas
34716	3547	Beverly Crest
65477	3550	Malibu
88820	3557	Chelsea
137690	3563	Grant Hill
2372	3571	Russian Hill
36547	3571	Unincorporated Santa Monica Mountains
218304	3571	Unincorporated Areas
173701	3571	78704
198282	3588	Fort Lauderdale
194716	3599	Fort Lauderdale
30894	3600	Beverly Hills
103412	3600	Tribeca
103129	3600	Williamsburg

81942	3600	Upper East Side
24953	3600	Alhambra
126651	3600	Middletown
28708	3600	Hollywood
42561	3600	Pacific Palisades
99774	3600	Greenpoint
24709	3604	Malibu
170782	3605	78746
107853	3610	SoHo
107854	3610	SoHo
30251	3617	Beverly Crest
137619	3621	La Jolla
217125	3637	Unincorporated Areas
172520	3643	78702
139110	3643	Torrey Pines
215794	3657	Unincorporated Areas
170095	3657	78732
169030	3658	78732
173353	3659	78704
206556	3668	Hollywood
177259	3675	78701
49713	3700	Hollywood Hills
58353	3700	Mar Vista
171407	3713	78732
171764	3714	78730
13741	3714	Highland Terrace
123006	3714	Newport
210910	3714	Fort Lauderdale
199702	3715	Fort Lauderdale
87163	3721	Midtown
58359	3737	Beverly Crest
56745	3750	Malibu
68465	3750	East Village
64023	3750	Malibu
61770	3757	Woodland Hills
215638	3758	Unincorporated Areas
58559	3799	Hollywood Hills West
31036	3800	Beverly Hills

34184	3800	Beverly Crest
60197	3800	Malibu
206530	3800	Fort Lauderdale
90280	3800	Flatiron District
34309	3800	Beverly Hills
2653	3800	Inner Sunset
164455	3800	78730
138164	3800	Midtown
2680	3809	South of Market
143383	3820	District 21
216223	3829	Unincorporated Areas
164397	3833	78704
207221	3841	Hollywood
201042	3848	Hallandale Beach
26864	3850	Hollywood Hills West
198766	3857	Fort Lauderdale
128574	3869	La Jolla
56957	3899	Beverly Crest
32078	3900	Malibu
130135	3900	La Jolla
169973	3900	78704
165772	3900	78744
94150	3900	West Village
64309	3900	Hollywood Hills West
215119	3906	Unincorporated Areas
201725	3909	Hollywood
209175	3946	Fort Lauderdale
88375	3950	Washington Heights
214052	3968	Unincorporated Areas
195780	3970	Hallandale Beach
230940	3993	Unincorporated Areas
32892	3995	Bellflower
177584	3995	78704
32557	3995	Beverly Hills
59457	3995	Malibu
33249	3999	Downtown
31001	3999	Beverly Crest

225509	3999	Downtown
191790	3999	Dakota
196296	3999	Hallandale Beach
24457	3999	Beverly Hills
103552	3999	Tribeca
221540	4000	Unincorporated Areas
52981	4000	Santa Monica
223429	4000	Unincorporated Areas
56273	4000	Malibu
172819	4000	78704
187386	4000	Greenwood
49235	4000	Long Beach
54479	4000	Palos Verdes Estates
188293	4000	Greenwood
57111	4000	Beverly Hills
196123	4000	Fort Lauderdale
103639	4000	Midtown
101517	4000	Upper West Side
190874	4000	Hennepin
112020	4000	Palo Alto
190198	4000	Hennepin
64571	4000	Beverly Hills
207631	4000	Fort Lauderdale
59701	4000	Hollywood Hills West
113549	4000	San Jose
189384	4000	Greenwood
210817	4000	Fort Lauderdale
78054	4000	Greenwich Village
194915	4000	Fort Lauderdale
187838	4000	Greenwood
83947	4000	Crown Heights
65516	4000	Malibu
232109	4000	Millbrae
7825	4000	Shaw, Logan Circle
31383	4000	Glendale
26868	4000	Hollywood Hills West
33021	4000	Beverly Hills
41455	4000	Encino

41260	4000	Hollywood Hills
31012	4000	Beverly Hills
223049	4033	Unincorporated Areas
50504	4036	Unincorporated Santa Monica Mountains
210828	4077	Fort Lauderdale
38933	4079	Bel-Air
214619	4091	Unincorporated Areas
215087	4096	Unincorporated Areas
102696	4100	Prospect Heights
103286	4100	Lower East Side
25052	4113	Beverly Hills
223107	4118	Unincorporated Areas
57305	4120	Hollywood Hills West
53011	4138	Unincorporated Santa Monica Mountains
65181	4163	West Hollywood
64136	4163	West Hollywood
61665	4168	Beverly Crest
56990	4169	West Hollywood
34062	4176	Hollywood Hills West
6126	4187	Western Addition
31214	4200	Malibu
32769	4200	Beverly Crest
190191	4200	Hennepin
135939	4213	Mission Bay
57827	4221	Hollywood Hills West
194648	4250	Fort Lauderdale
26038	4250	Hollywood Hills West
32039	4259	Beverly Crest
54925	4281	Unincorporated Santa Monica Mountains
226226	4283	Downtown
196075	4286	Hollywood
194657	4286	Hollywood
63902	4300	Woodland Hills
55602	4302	Monterey Park
218234	4320	Unincorporated Areas
218222	4320	Unincorporated Areas
29582	4341	Beverly Crest

11142	4357	Shaw, Logan Circle
221764	4368	Unincorporated Areas
58788	4375	Malibu
223585	4400	Unincorporated Areas
57722	4400	West Carson
168120	4401	78702
37502	4472	Encino
224890	4480	Unincorporated Areas
28779	4500	West Hollywood
28645	4500	Beverly Hills
216189	4500	Unincorporated Areas
35457	4500	Hollywood Hills West
196970	4500	Fort Lauderdale
132970	4500	Carmel Valley
153000	4500	East Colfax
53871	4500	Hollywood Hills
53350	4500	Hollywood Hills
31047	4500	Beverly Hills
168971	4500	78746
158286	4500	West Town
170044	4500	78732
174616	4500	78732
65137	4500	Beverly Crest
34308	4500	Beverly Hills
25086	4514	Beverly Crest
223455	4535	Unincorporated Areas
34787	4536	Glendale
30423	4537	Hollywood Hills West
26582	4539	Beverly Crest
29364	4553	Beverly Grove
208542	4575	Sea Ranch Lakes
205740	4586	Hallandale Beach
209166	4586	Hallandale Beach
201886	4601	Fort Lauderdale
26141	4620	Hollywood Hills West
65179	4632	West Hollywood
62232	4632	West Hollywood
65962	4638	West Hollywood

107717	4650	SoHo
209311	4650	Lighthouse Point
61331	4700	Beverly Hills
32936	4700	Malibu
49987	4700	Venice
173146	4703	78732
112009	4711	San Jose
52580	4721	Hollywood Hills
29949	4750	Glendale
231752	4750	Redwood City
33488	4750	Pasadena
33224	4750	Pasadena
199501	4757	Hollywood
51293	4768	Bel-Air
221666	4786	Unincorporated Areas
198235	4800	Lighthouse Point
32642	4800	Malibu
64355	4800	Beverly Crest
107715	4813	SoHo
107716	4813	SoHo
61546	4823	Hollywood Hills West
65486	4863	Artesia
26453	4897	Hollywood Hills West
59842	4900	Hollywood Hills West
143127	4900	District 15
3450	4950	Pacific Heights
152672	4950	Cherry Creek
198209	4950	Fort Lauderdale
77340	4964	West Village
35843	4986	Valley Village
43481	4986	Studio City
31148	4995	Malibu
171000	4999	78733
60906	4999	Reseda
43626	4999	Southeast Antelope Valley
73981	5000	Tribeca
106868	5000	Lower East Side

26374	5000	Brentwood
134926	5000	Mission Bay
104370	5000	Upper East Side
32864	5000	Beverly Crest
54246	5000	Venice
142904	5000	District 21
219345	5000	Unincorporated Areas
106695	5000	Bay Ridge
36337	5000	East Los Angeles
7988	5000	Union Station, Stanton Park, Kingman Park
172415	5000	78734
231492	5000	San Mateo
229677	5000	Redwood City
46330	5000	Hollywood
46430	5000	Bel-Air
89653	5000	Sheepshead Bay
18783	5000	Marigny
172084	5000	78731
137008	5000	Roseville
181774	5000	Pearl
85913	5000	East Village
47534	5000	Monterey Park
14495	5000	Lynn-Highland Park
189902	5000	Hennepin
190109	5000	Hennepin
222862	5000	Unincorporated Areas
11562	5000	Capitol View, Marshall Heights, Benning Heights
219461	5000	City of North Las Vegas
140499	5000	Ocean Beach
147296	5000	District 15
62325	5000	Torrance
61362	5000	Fairfax
151552	5000	Five Points
38545	5000	Burbank
164538	5000	78734
213824	5000	Unincorporated Areas
39544	5000	Manhattan Beach
164697	5000	78704

107640	5000	Stuyvesant Town
111035	5000	Los Altos Hills
204417	5000	Dania Beach
25594	5000	Beverly Hills
165230	5000	78704
164389	5000	78704
198744	5000	Fort Lauderdale
205484	5000	Hollywood
30938	5000	Beverly Hills
28776	5000	Universal City
57108	5000	San Gabriel
156536	5000	West Town
45614	5000	Rancho Palos Verdes
55878	5000	Bel-Air
210150	5000	Pompano Beach
130390	5000	East Village
60867	5000	Beverly Crest
147915	5000	District 15
64366	5000	Beverly Crest
166048	5000	78746
164187	5021	78749
57288	5045	Hollywood Hills West
34356	5054	Hollywood Hills West
157163	5060	West Town
67021	5065	East Harlem
28874	5114	Beverly Hills
77461	5143	Stapleton
68863	5143	Gramercy
6671	5150	West of Twin Peaks
62916	5200	Hollywood Hills West
127894	5213	Darnall
137629	5214	La Jolla
29308	5232	Beverly Hills
68514	5250	Chinatown
133696	5260	Mira Mesa
63631	5286	Beverly Hills
125878	5286	Middletown

202357	5300	Fort Lauderdale
50698	5311	Encino
164126	5343	78746
53008	5372	Unincorporated Santa Monica Mountains
169024	5379	78746
19076	5400	Mid-City
151340	5409	Sunnyside
34180	5475	Pasadena
100805	5500	Bergen Beach
113057	5500	Unincorporated Areas
26058	5500	Sherman Oaks
65246	5500	Beverly Hills
189753	5500	Hennepin
54408	5500	Hollywood
229238	5500	Atherton
63863	5500	Hollywood Hills West
50293	5500	Hollywood
33732	5500	Sherman Oaks
26871	5500	Beverly Hills
190346	5500	Hennepin
34304	5500	Beverly Hills
54098	5500	Studio City
60484	5520	Beverly Crest
195913	5540	Fort Lauderdale
190648	5556	Ramsey
87614	5556	Richmond Hill
81327	5562	Jackson Heights
206653	5590	Fort Lauderdale
58920	5607	Beverly Hills
210414	5650	Hollywood
182831	5729	Powellhurst-Gilbert
215148	5732	Unincorporated Areas
173841	5748	78746
217071	5760	Unincorporated Areas
223087	5760	Unincorporated Areas
200453	5772	Hallandale Beach
229275	5800	Unincorporated Areas
61557	5834	Beverly Hills

63998	5857	Malibu
197433	5864	Hallandale Beach
58982	5899	Beverly Hills
24451	5899	Beverly Hills
131554	5920	Marina
18310	5947	Mcdonogh
106096	5950	NoHo
5712	5950	Pacific Heights
58376	5950	Beverly Crest
64320	5952	Beverly Crest
200139	5956	Hollywood
7705	5995	Capitol Hill, Lincoln Park
52658	5999	Monterey Park
46260	5999	Monterey Park
215845	6000	Unincorporated Areas
68385	6000	Upper West Side
54496	6000	Bel-Air
63492	6000	Hollywood Hills West
126868	6000	Middletown
205741	6000	Hallandale Beach
32455	6000	Beverly Hills
25499	6000	Malibu
24755	6000	Malibu
120401	6000	Ward 3
54608	6000	Encino
211030	6000	Fort Lauderdale
202803	6000	Lauderdale By The Sea
54739	6000	Pico Rivera
165753	6000	78704
216222	6000	Unincorporated Areas
4564	6000	Bernal Heights
63293	6000	Malibu
25148	6000	Beverly Hills
74183	6000	Greenwich Village
34415	6031	Hollywood Hills West
214081	6065	Unincorporated Areas
218135	6075	City of Las Vegas

87035	6080	East Harlem
199003	6143	Fort Lauderdale
25464	6200	Canoga Park
31104	6250	Marina del Rey
31109	6250	Marina del Rey
56924	6250	Malibu
215116	6309	Unincorporated Areas
63268	6350	Beverly Hills
61747	6350	Beverly Hills
39838	6429	Covina
31019	6429	Beverly Hills
60705	6500	Hollywood Hills West
123805	6500	Newport
84678	6500	Tribeca
41644	6500	Los Feliz
28644	6500	Beverly Hills
40381	6500	Hollywood Hills
107601	6500	SoHo
68588	6500	Clinton Hill
166433	6500	78702
26396	6500	Malibu
62549	6504	Hawthorne
61947	6570	Hollywood Hills West
222849	6571	Unincorporated Areas
118115	6588	28806
206652	6600	Fort Lauderdale
58615	6667	Hollywood Hills West
163264	6676	Loop
206063	6690	Fort Lauderdale
199427	6714	Fort Lauderdale
213840	6765	Unincorporated Areas
220093	6786	Unincorporated Areas
194600	6850	Lighthouse Point
58110	6875	Malibu
37380	6909	Burbank
29237	6948	Beverly Hills
19560	6975	New Aurora - English Turn
29025	6995	Beverly Hills

155519	7000	North Park Hill
33145	7000	Beverly Hills
189815	7000	Hennepin
155026	7000	South Park Hill
43490	7066	Hollywood
48085	7071	Canoga Park
101155	7075	Bushwick
31000	7079	Beverly Hills
140530	7181	West University Heights
76221	7184	Canarsie
94303	7203	West Brighton
210302	7286	Hollywood
196665	7286	Lauderdale By The Sea
72765	7314	Lower East Side
220466	7321	City of Henderson
43753	7358	Encino
198983	7429	Fort Lauderdale
61778	7500	Hollywood Hills West
202054	7500	Hollywood
11939	7500	Georgetown, Burleith/Hillandale
65808	7500	Malibu
151180	7500	Sloan Lake
62373	7500	Malibu
56204	7538	Downtown
53672	7550	Bel-Air
131063	7586	Grant Hill
52833	7600	Hancock Park
202027	7856	Oakland Park
32291	7900	Malibu
139760	7993	City Heights East
5715	7995	Pacific Heights
109040	8000	Midtown
109008	8000	Midtown
63512	8000	Hollywood Hills West
198541	8071	Fort Lauderdale
94508	8085	South Ozone Park
92449	8090	Laurelton

62467	8321	Malibu
25104	8500	Sherman Oaks
207320	8571	Fort Lauderdale
201883	8667	Hallandale Beach
167823	8691	78701
46297	8699	Monterey Park
217419	8800	Unincorporated Areas
127782	8807	Grant Hill
28083	8899	Alhambra
60251	9000	Woodland Hills
78077	9000	Briarwood
189898	9000	Ramsey
209894	9000	Fort Lauderdale
43482	9000	Monterey Park
112322	9000	Mountain View
199692	9000	Hollywood
154471	9000	West Colfax
197311	9095	Hollywood
172674	9102	78703
203183	9108	Fort Lauderdale
173065	9143	78704
222906	9429	Unincorporated Areas
25196	9500	Beverly Hills
27800	9500	Beverly Crest
190020	9900	Hennepin
32194	9957	Hollywood Hills West
69486	9990	East Village
91862	9994	Longwood
29314	9995	Beverly Crest
33602	9998	Alhambra
169504	9998	78750
3346	9999	Chinatown
62063	9999	Pasadena
2297	9999	Downtown/Civic Center
3334	9999	Financial District
32639	9999	West Hollywood
2295	9999	Downtown/Civic Center
2294	9999	Downtown/Civic Center

3347	9999	Downtown/Civic Center
2300	9999	Downtown/Civic Center
2335	9999	Downtown/Civic Center
2350	9999	Downtown/Civic Center
2301	9999	Downtown/Civic Center
212975	9999	Unincorporated Areas
36403	9999	Studio City
2304	9999	Downtown/Civic Center
2349	9999	Downtown/Civic Center
2308	9999	Downtown/Civic Center
211207	9999	City of Las Vegas
65256	9999	Alhambra
2347	9999	Downtown/Civic Center
69945	9999	East Harlem
36410	9999	Studio City
2352	9999	Downtown/Civic Center
2346	9999	Downtown/Civic Center
3345	9999	Chinatown
108778	9999	Upper East Side
2267	9999	Downtown/Civic Center
30799	9999	Beverly Crest
87877	9999	Upper East Side
2248	9999	Downtown/Civic Center
2239	9999	Downtown/Civic Center
2233	9999	Downtown/Civic Center
2231	9999	Downtown/Civic Center
2230	9999	Downtown/Civic Center
123243	9999	Newport
124625	9999	Newport
155845	9999	Jefferson Park
28786	9999	Beverly Hills
2223	9999	Downtown/Civic Center
149254	9999	Unincorporated Areas
129918	9999	Ocean Beach
2222	9999	Downtown/Civic Center
2221	9999	Downtown/Civic Center
146896	9999	District 20

146895	9999	District 20
146893	9999	District 20
146886	9999	District 20
44890	9999	Studio City
2220	9999	Downtown/Civic Center
2219	9999	Downtown/Civic Center
2217	9999	Downtown/Civic Center
46315	9999	Koreatown
2215	9999	Downtown/Civic Center
2278	9999	Downtown/Civic Center
2279	9999	Downtown/Civic Center
2287	9999	Downtown/Civic Center
215804	9999	Unincorporated Areas
2289	9999	Downtown/Civic Center
20837	9999	Central Business District
20844	9999	Central Business District
20846	9999	Central Business District
20909	9999	Central Business District
2285	9999	Downtown/Civic Center
103260	9999	Bellerose
215811	9999	Unincorporated Areas
20916	9999	Central Business District
20805	9999	Central Business District
41652	9999	Studio City
56283	9999	Santa Monica
20919	9999	Central Business District
2280	9999	Downtown/Civic Center
41653	9999	Studio City
2283	9999	Downtown/Civic Center
20918	9999	Central Business District
21659	9999	Central Business District
2290	9999	Downtown/Civic Center
21175	9999	Central Business District
215863	9999	Unincorporated Areas
215854	9999	Unincorporated Areas
45483	10000	Burbank
226443	10000	Dorchester
39675	10000	Burbank

60901	10000	Malibu
3376	10000	Downtown/Civic Center
3375	10000	Downtown/Civic Center
44993	10000	Hollywood
39279	10000	Hollywood
214006	10000	City of North Las Vegas
3292	10000	South of Market
3291	10000	South of Market
58011	10000	Beverly Hills
3290	10000	South of Market
49917	10000	Long Beach
55040	10000	Palms
39280	10000	Hollywood
42970	10000	Bel-Air
39634	10000	Playa del Rey
56780	10000	Beverly Hills
374	10000	South of Market
29416	10000	Beverly Crest
65670	10000	Beverly Hills
94325	10000	Financial District
164751	10000	78723
172599	10000	78702
129223	10000	North Hills
172667	10000	78702
69507	10000	Long Island City
75132	10000	Long Island City
85652	10000	Midtown
131530	10000	East Village
20905	10000	Central Business District
79354	10000	Upper West Side
30943	10000	Hollywood Hills West
82815	10000	Murray Hill
85646	10000	Midtown
85651	10000	Midtown
143177	10000	District 19
143181	10000	District 19
190447	10000	Hennepin

6232	10000	North Beach
86064	10000	Lower East Side
31010	10000	Beverly Hills
28394	10000	Hollywood Hills West
62257	10000	Beverly Hills
21000	10000	Central Business District
169728	10000	78702
167863	10000	78735
136158	10000	Mission Bay
62908	10000	Azusa
137266	10000	La Jolla
25648	10000	Malibu
167381	10000	78701
202174	10000	Deerfield Beach
34233	10000	Hollywood Hills West
156783	10000	Lincoln Park
31100	10286	Beverly Hills
169025	10450	78733
27329	11000	Beverly Hills
56539	11000	Glendale
65672	12000	Beverly Hills
61833	12000	Beverly Hills
33726	12000	Beverly Hills
217041	12003	Unincorporated Areas
31081	12500	Beverly Hills
209497	12510	West Park
46439	12723	Unincorporated Santa Monica Mountains
61671	12750	Beverly Hills
221362	12857	City of Las Vegas
5740	14000	Bernal Heights
216172	14263	Unincorporated Areas
50712	14949	Unincorporated Santa Monica Mountains
99977	15000	East Harlem
50888	15000	Bel-Air
196715	15000	Fort Lauderdale
29399	15790	Beverly Hills
30954	16121	Hollywood Hills West
132198	17429	Roseville

98099	19750	Upper East Side
218252	20000	Unincorporated Areas
218250	20000	Unincorporated Areas
218262	20000	Unincorporated Areas
218263	20000	Unincorporated Areas
218274	20000	Unincorporated Areas
223050	20000	Unincorporated Areas
218249	20000	Unincorporated Areas
220009	20000	Unincorporated Areas
218244	20000	Unincorporated Areas
218243	20000	Unincorporated Areas
218242	20000	Unincorporated Areas
218240	20000	Unincorporated Areas
223051	20000	Unincorporated Areas
218233	20000	Unincorporated Areas
218232	20000	Unincorporated Areas
218231	20000	Unincorporated Areas
218248	20000	Unincorporated Areas
218276	20000	Unincorporated Areas
223045	20000	Unincorporated Areas
218286	20000	Unincorporated Areas
220006	20000	Unincorporated Areas
219651	20000	Unincorporated Areas
220022	20000	Unincorporated Areas
220039	20000	Unincorporated Areas
220041	20000	Unincorporated Areas
220044	20000	Unincorporated Areas
220045	20000	Unincorporated Areas
220172	20000	Unincorporated Areas
220174	20000	Unincorporated Areas
220176	20000	Unincorporated Areas
223046	20000	Unincorporated Areas
218442	20000	Unincorporated Areas
218421	20000	Unincorporated Areas
223047	20000	Unincorporated Areas
218290	20000	Unincorporated Areas
220007	20000	Unincorporated Areas

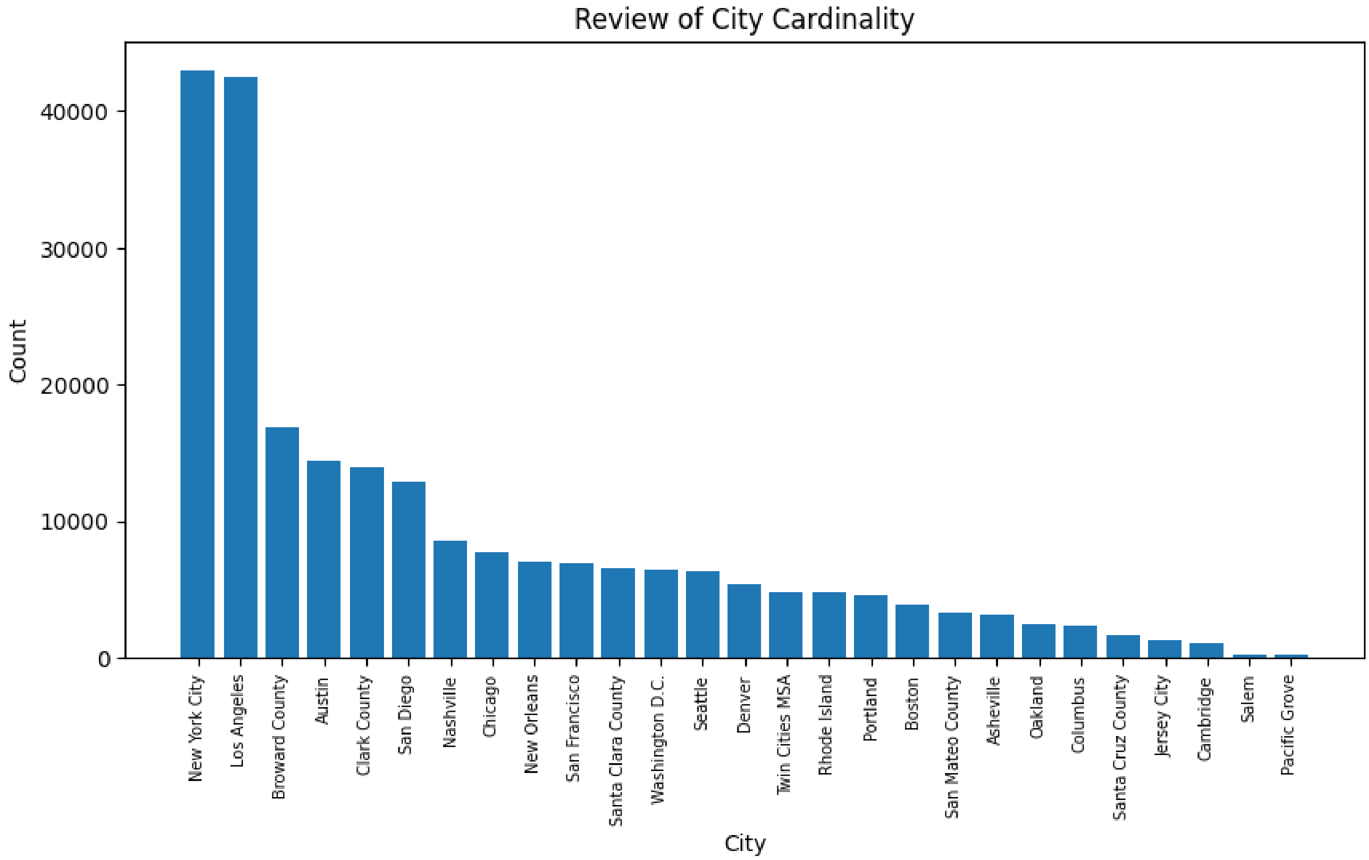
218287	20000	Unincorporated Areas
218277	20000	Unincorporated Areas
218230	20000	Unincorporated Areas
217088	20000	Unincorporated Areas
223060	20000	Unincorporated Areas
218228	20000	Unincorporated Areas
216802	20000	Unincorporated Areas
216801	20000	Unincorporated Areas
223083	20000	Unincorporated Areas
223084	20000	Unincorporated Areas
177090	20000	78701
177094	20000	78701
177095	20000	78701
32718	20000	Malibu
217048	20000	Unincorporated Areas
6828	20000	Nob Hill
6824	20000	Nob Hill
6821	20000	Nob Hill
224460	20000	Unincorporated Areas
5932	20000	Nob Hill
5929	20000	Nob Hill
5927	20000	Nob Hill
5925	20000	Nob Hill
224463	20000	Unincorporated Areas
224489	20000	Unincorporated Areas
6826	20000	Nob Hill
217049	20000	Unincorporated Areas
217040	20000	Unincorporated Areas
217084	20000	Unincorporated Areas
218215	20000	Unincorporated Areas
218214	20000	Unincorporated Areas
218198	20000	Unincorporated Areas
218194	20000	Unincorporated Areas
218191	20000	Unincorporated Areas
218187	20000	Unincorporated Areas
223064	20000	Unincorporated Areas
217195	20000	Unincorporated Areas
217190	20000	Unincorporated Areas

217189	20000	Unincorporated Areas
223080	20000	Unincorporated Areas
223065	20000	Unincorporated Areas
223066	20000	Unincorporated Areas
217179	20000	Unincorporated Areas
223067	20000	Unincorporated Areas
223069	20000	Unincorporated Areas
223071	20000	Unincorporated Areas
217091	20000	Unincorporated Areas
217089	20000	Unincorporated Areas
223073	20000	Unincorporated Areas
222188	20000	Unincorporated Areas
46708	20125	Bel-Air
76520	20500	Harlem
50921	21053	Bel-Air
59741	22092	Hollywood Hills West
28651	25000	Beverly Hills
28670	25000	Beverly Hills
35347	25000	Beverly Hills
113854	25000	Palo Alto
35348	25000	Beverly Hills
3553	25000	Financial District
3613	25000	Financial District
61914	25000	Malibu
3614	25000	Financial District
4439	25000	Financial District
6563	25000	Inner Richmond
172607	25000	78701
216800	25000	Unincorporated Areas
217056	25000	Unincorporated Areas
217086	25000	Unincorporated Areas
217099	25000	Unincorporated Areas
217168	25000	Unincorporated Areas
27588	25000	Beverly Hills
217182	25000	Unincorporated Areas
218221	25000	Unincorporated Areas
113850	25000	Palo Alto

29373	25000	Beverly Hills
218275	25000	Unincorporated Areas
218408	25000	Unincorporated Areas
218279	25000	Unincorporated Areas
217181	25000	Unincorporated Areas
18928	29969	Bywater
52114	30000	Unincorporated Santa Monica Mountains
17898	36606	Leonidas
27892	37234	Valinda
219300	58812	Unincorporated Areas
76129	65115	West Brighton
169452	65155	78703
15197	69429	Bancroft Business-Havenscourt
143364	78200	District 21
101640	85100	East Harlem
99151	85170	Sunset Park
51330	90135	Burbank
60045	90150	West Compton
216614	90180	Unincorporated Areas
126323	95117	Little Compton
101010	99000	Williamsburg
157490	99998	Lake View
16625	99999	Ward E (councilmember James Solomon)
35792	99999	Rosemead
200126	100000	Pembroke Pines
136510	100000	Mission Bay
231699	100000	Unincorporated Areas

From the graph above we should be good to remove outliers from price with values greater than 3000.

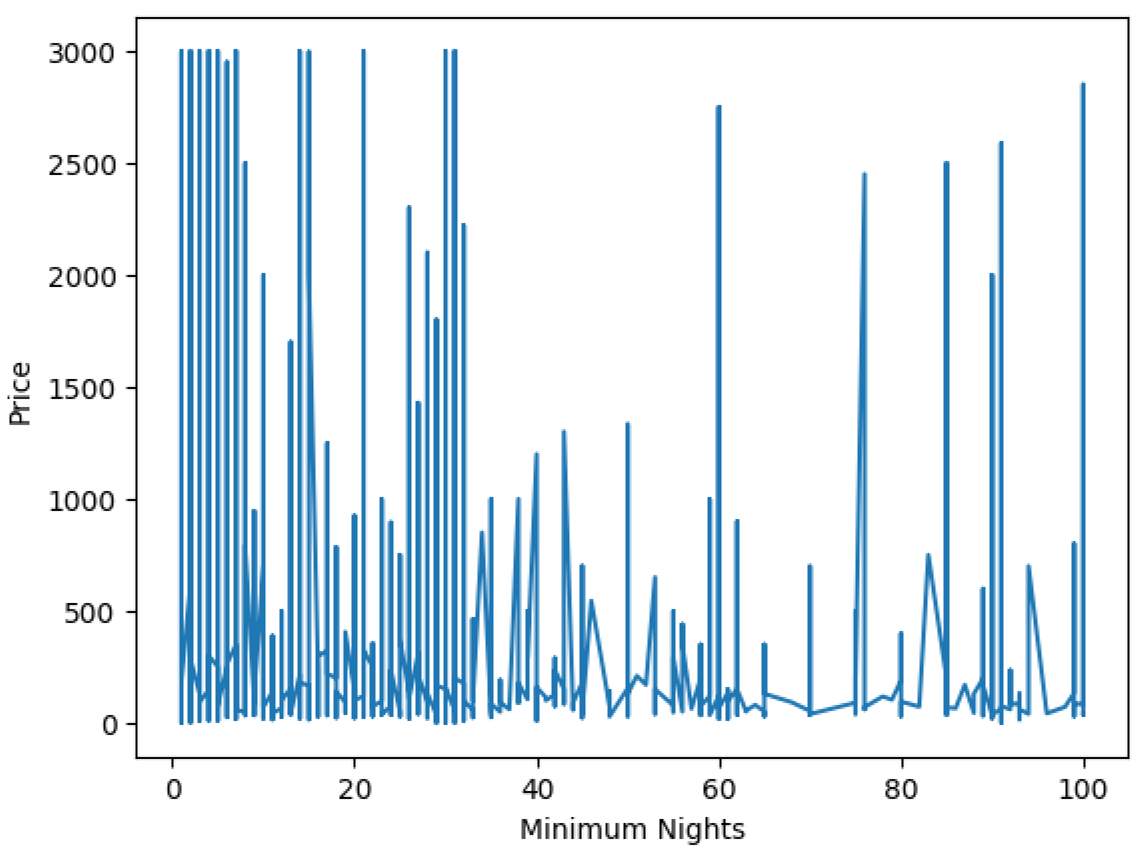
In [21]:
```
city = details['city'].value_counts()
pyplot.figure(figsize=(10,5))

pyplot.bar(city.index, city.values)

pyplot.title('Review of City Cardinality')
pyplot.xlabel('City')
pyplot.ylabel('Count')
pyplot.xticks(fontsize=7, rotation=90)
pyplot.show()

print('Unique values to check cardinality are')
details['city'].nunique()
```




```
Unique values to check cardinality are
```

Out[21]:
```
27
```

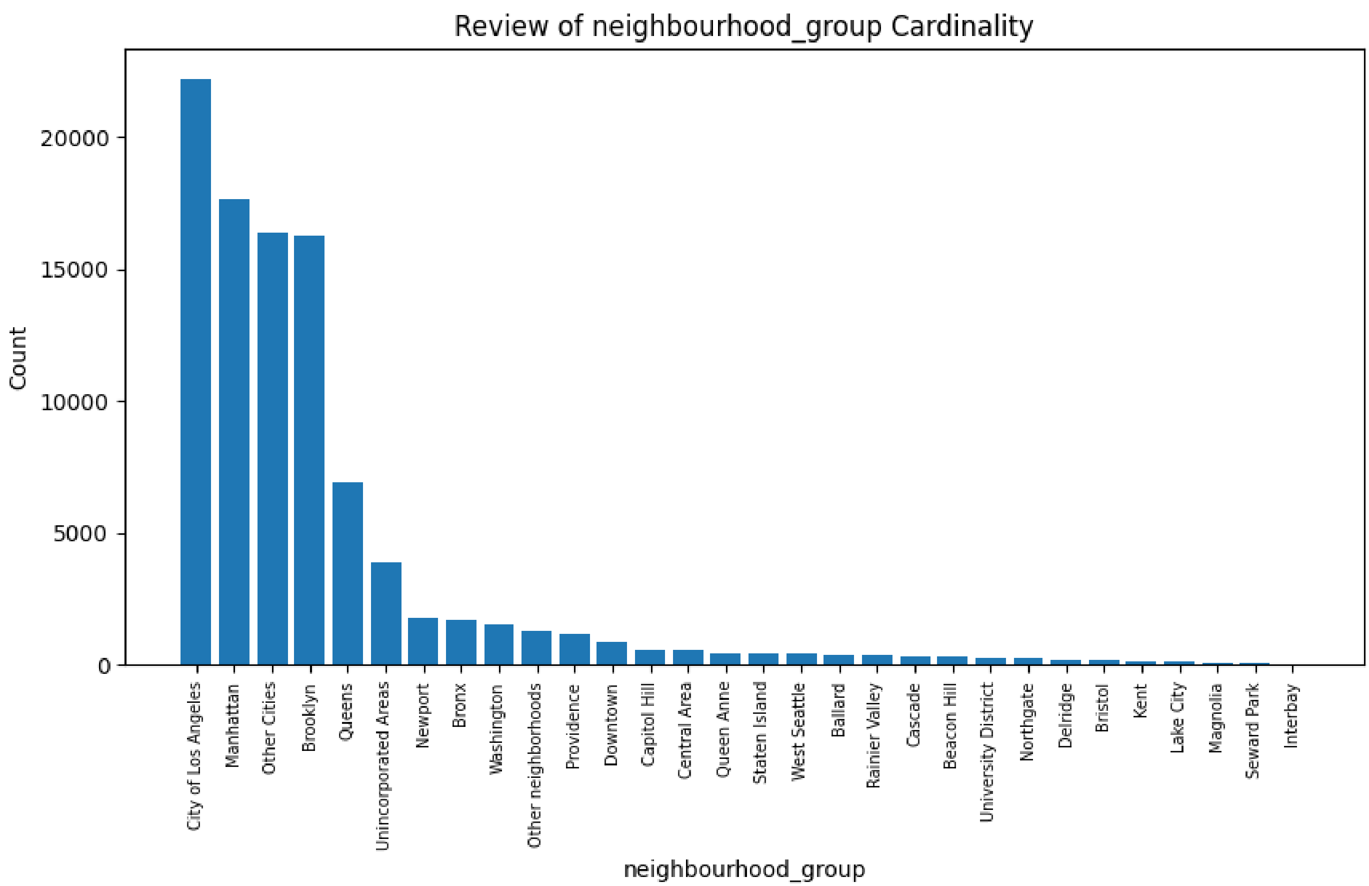
In [22]:
```
# Taking minimum nights <100 and price < 3000 due to cardinality issues inferred from ab
sorted_details = details[(details['minimum_nights'] <= 100) & (details['price'] <= 3000)
pyplot.plot(sorted_details['minimum_nights'], sorted_details['price'])
pyplot.xlabel('Minimum Nights')
pyplot.ylabel('Price')
pyplot.show()
```



Based on the graph, it can be inferred that there is no significant correlation between minimum nights and price. The rate of price fluctuation does not seem to be affected by changes in minimum nights. Moreover, the cardinality of minimum nights is too high, which results in irregular patterns and makes it difficult to analyze the relationship between these variables.

The trend suggests that the price is decreasing as the number of minimum nights increases, except for a few scenarios. This pattern may indicate the presence of outliers. Let us investigate this further.

In [23]:
```
neighbourhood_group = details['neighbourhood_group'].value_counts()
pyplot.figure(figsize=(10,5))

pyplot.bar(neighbourhood_group.index, neighbourhood_group.values)

pyplot.title('Review of neighbourhood_group Cardinality')
pyplot.xlabel('neighbourhood_group')
pyplot.ylabel('Count')
pyplot.xticks(fontsize=7, rotation=90)
pyplot.show()

print('Unique values to check cardinality are')
details['neighbourhood_group'].nunique()
```




```
Unique values to check cardinality are
```

Out[23]:
```
30
```

In [24]:
```
details['neighbourhood'].nunique()
```

Out[24]:
```
1412
```

There are total of 1412 neighbourhood for 232147 data entries. That is not a irregular cardinality.

3. [0.75] Data Preparation

a) Preprocess your data according to the data quality plan.

In [20]:
```
import pandas as pd
```

In [21]:
```
data = pd.read_csv('AB_US_2023.csv')
data.head()
```

```
C:\Users\Tasne\AppData\Local\Temp\ipykernel_28932\3204684119.py:1: DtypeWarning: Columns
(4) have mixed types. Specify dtype option on import or set low_memory=False.
  data = pd.read_csv('AB_US_2023.csv')
```

Out[21]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type
0	958	Bright, Modern Garden Unit - 1BR/1BTH	1169	Holly	NaN	Western Addition	37.77028	-122.43317	Entire home/apt
1	5858	Creative Sanctuary	8904	Philip And Tania	NaN	Bernal Heights	37.74474	-122.42089	Entire home/apt
2	8142	Friendly Room Apt. Style - UCSF/USF - San Franc...	21994	Aaron	NaN	Haight Ashbury	37.76555	-122.45213	Private room
3	8339	Historic Alamo Square Victorian	24215	Rosy	NaN	Western Addition	37.77564	-122.43642	Entire home/apt
4	8739	Mission Sunshine, with Private Bath	7149	Ivan & Wendy	NaN	Mission	37.76030	-122.42197	Private room

In [22]:
```
data.drop(['id', 'name', 'host_id', 'host_name', 'number_of_reviews_ltm', 'reviews_per_m
           'last_review', 'calculated_host_listings_count', 'latitude',
           'availability_365', 'longitude'], axis=1, inplace=True)
```

In [23]:
```
data.isnull().sum() * 100 / len(data)
```

Out[23]:
```
neighbourhood_group    58.431511
neighbourhood           0.000000
room_type               0.000000
price                   0.000000
minimum_nights          0.000000
number_of_reviews       0.000000
city                    0.000000
dtype: float64
```

In [24]:
```
# drop neighbourhood_group because there are too many nan values
data.drop(['neighbourhood_group'], axis=1, inplace=True)
```

In [25]:
```
data.head()
```

Out[25]:

	neighbourhood	room_type	price	minimum_nights	number_of_reviews	city
0	Western Addition	Entire home/apt	202	2	383	San Francisco

1	Bernal Heights	Entire home/apt	235	30	111	San Francisco
2	Haight Ashbury	Private room	56	32	9	San Francisco
3	Western Addition	Entire home/apt	575	9	28	San Francisco
4	Mission	Private room	110	1	770	San Francisco

In [26]:
```
unique = data['city'].unique()
len(unique)
```
Out[26]:
```
27
```
In [27]:
```
unique = data['neighbourhood'].unique()
len(unique)
```
Out[27]:
```
1412
```
In [28]:
```
unique = data['room_type'].unique()
len(unique)
```
Out[28]:
```
4
```
In [29]:
```
# convert string fields to numerical
city_map = dict(enumerate(data['city'].astype('category').cat.categories))
data['city'] = data['city'].astype('category').cat.codes
data['neighbourhood'] = data['neighbourhood'].astype('category').cat.codes
data['room_type'] = data['room_type'].astype('category').cat.codes
```
In [30]:
```
# data.drop(data[data['price']>3000], axis=0, inplace=True)
data.drop(data[data['price'] > 20000].index, inplace = True)
```
In [31]:
```
len(data)
```
Out[31]:
```
232095
```
In [32]:
```
data.sample(n = 5)
```
Out[32]:

	neighbourhood	room_type	price	minimum_nights	number_of_reviews	city
66963	677	2	189	10	121	13
66196	1239	0	120	1	41	10
47234	767	0	675	30	0	10
203404	1215	0	378	2	2	3
81372	710	2	180	30	4	13

4. [1.75] Data Insights

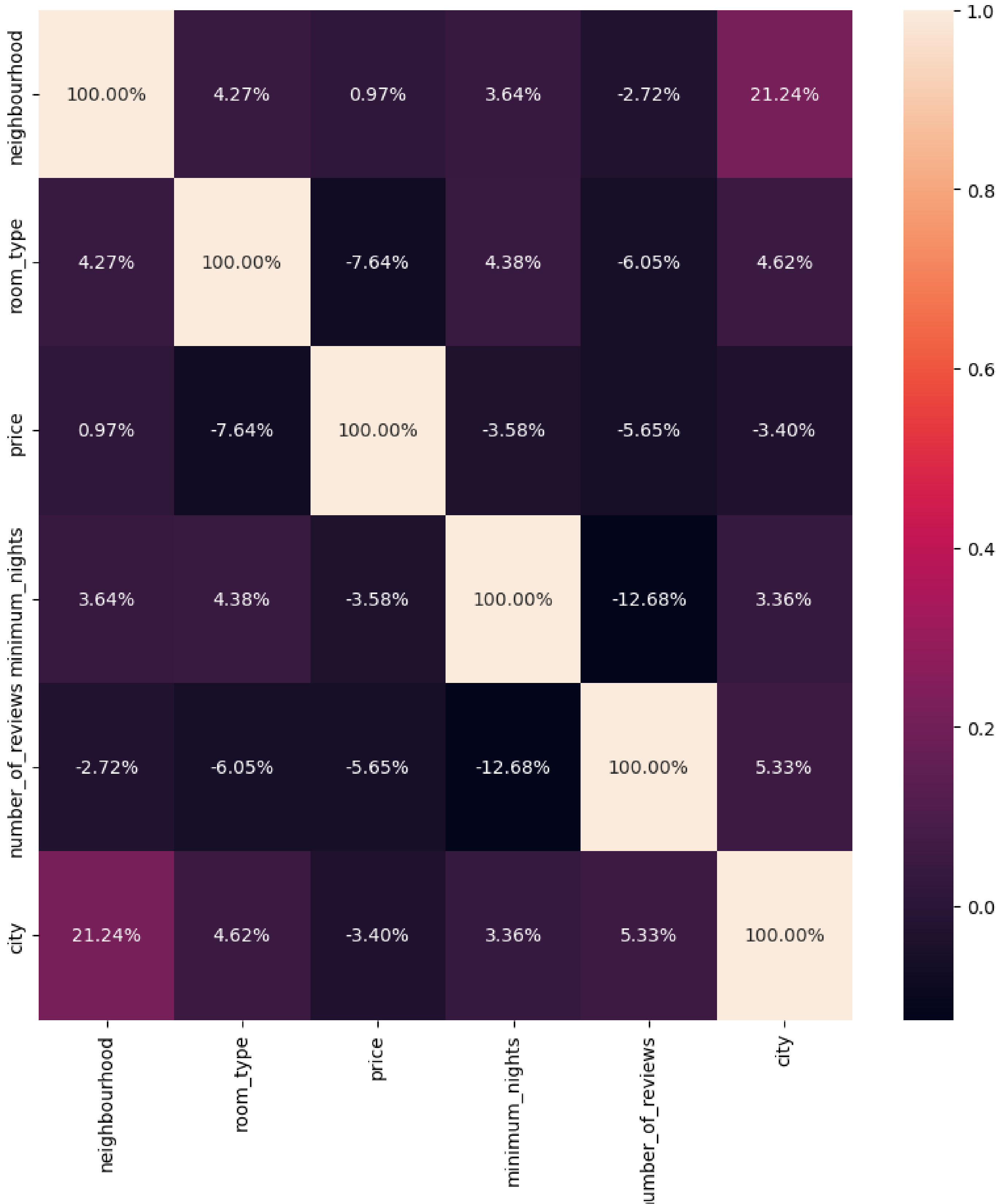
a) Build the correlation heatmap of the features in the ABT. Derive insights from it and relate it to the business problem being addressed. [0.75]

In [33]:
```
import seaborn as sns
```
In [34]:
```
data_corr = data.corr()
fig, ax = plt.subplots(figsize=(10,10))
```

```
sns.heatmap(data_corr, annot=True, fmt='.2%', ax=ax)
```

Out[34]:

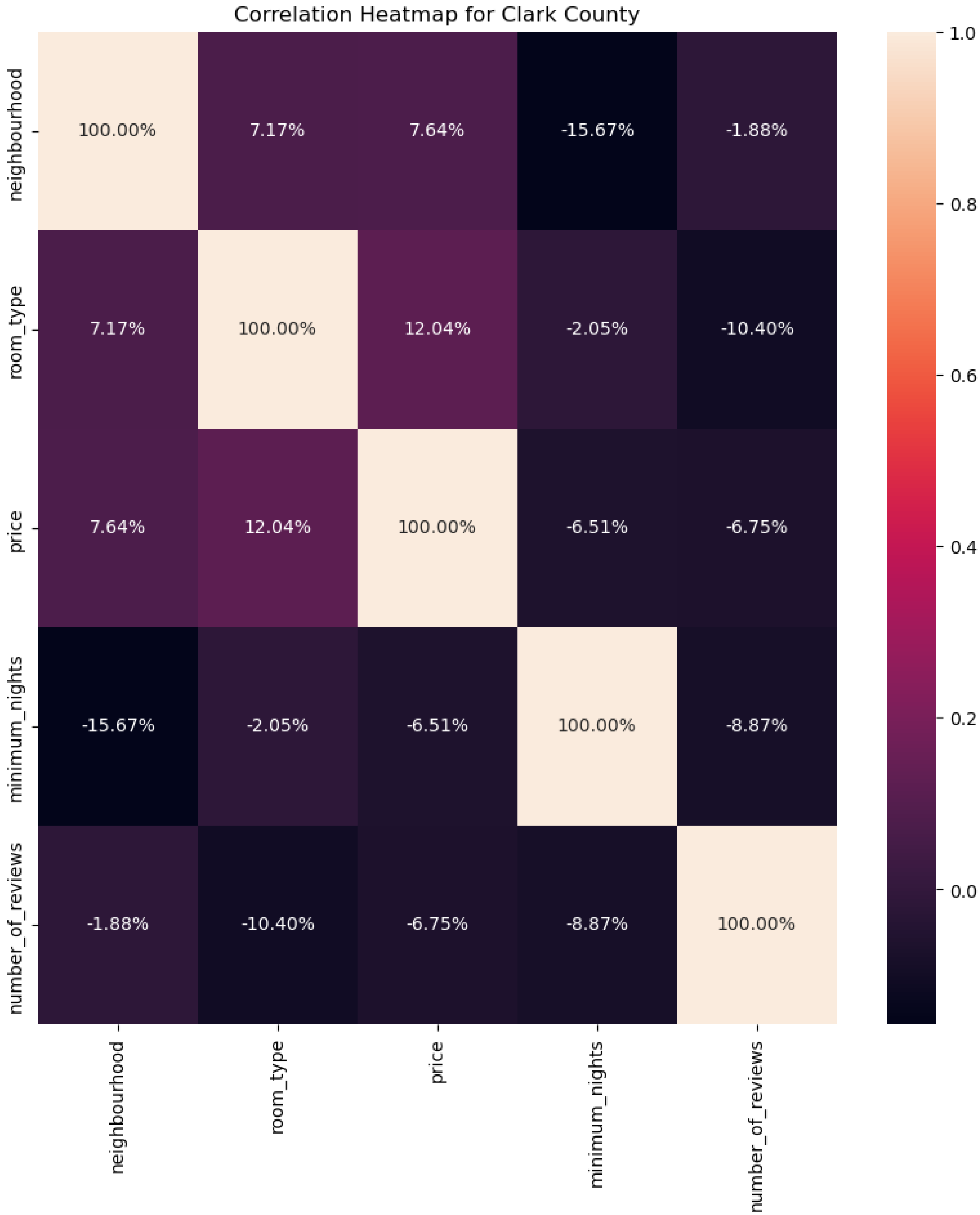
<AxesSubplot:>



The problem we are trying to solve is to predict a reasonable price given city, number of reviews, minimum nights, room type and neighbourhood. The heat map shows what we know theoretically, neighbourhood and city have good correlation to the price. While price have no strong direct correlation to minimum nights, but as seen from the heatmap, it has a strong correlation with neighbourhood and city. Number of reviews only seems to have a good correlation with the city. Needless to say, city and neighbourhood have a really strong correlation. One thing that can be noticed is that room type did not correlate strongly with the price (which was expected), so another heatmap is created below.

In [35]:
```
city = 6
if city in data['city'].unique():
    city_data = data[data['city'] == city]
    city_name = city_map[city]
    city_data.drop(['city'], axis=1, inplace=True)
    city_corr = city_data.corr()
    fig, ax = plt.subplots(figsize=(10,10))
    sns.heatmap(city_corr, annot=True, fmt='.2%', ax=ax)
    plt.title(f"Correlation Heatmap for {city_name}")
    plt.show()
```

```
C:\Users\Tasne\AppData\Local\Temp\ipykernel_28932\736411330.py:5: SettingWithCopyWarnin
g:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
guide/indexing.html#returning-a-view-versus-a-copy
  city_data.drop(['city'], axis=1, inplace=True)
```



In this heatmap, only the airbnb listings in Clark County was taken into account, as evident from the plot, price has a very high correlation to room type. This trend is seen in all the other cities, if a heatmap is plotted for each individual city. This trend is seen because a single room in new york city (busy place) will be more expensive than a whole house in clark county (less busy place). So for our business problem we may have to group together cities before predicting based on room types.

i. What are the descriptive features that highly correlate with the target feature? Propose some hypotheses explaining the correlation. [0.5]

The discriptive features that highly correlate with the target feature are:

1. Neighbourhood: some neighgbourhoods are better than others hence higher price. Also, within a city there are different regions where some are cheaper while others are costlier hence price and neighbourhood are correlated.
2. City: cities with more tourism or densly populated (urban) will tend to have higher prices than with cities with low tourist attraction or less population (rural).
3. Minimum nights: if minimum nights requirement is lower, prices will be higher compared to if the minimum nights requiremnet is higher. Even in the data exploration section above, this trend is evident.

ii. What are the domain concepts that highly correlate with each other? Propose some hypotheses explaining the correlation. [0.25]

Cities and neighbourhood have a high correlation since many neighbourhoods are part of one city.

Room type is highly correlated with city/neighbourhood because some neighbourhood has bigger houses with more rooms, vs some cities have more hotels and since city and neighbourhood correlate to eachother, they are also correlated with room type.

In urban cities, people tend to stay for longer for buisness/medical purposes, where as in tourist areas people come to visit for a day or two therefore minimum nights correlate strongly with city and neighbourhood.

iii. Are there any features that are useless for a predictive model? [0.25]

Room type and price seems not directly related, this is because, a single room in New York city may be the same price as a single room in Nashville. Although the correlation is not high, it will be high once the heatmap is created if grouped by cities.

Number of reviews and price are not correlated, so we can conclude that number of reviews is a useless feature for the predictive model.

References:

1. https://medium.com/ibm-data-science-experience/markdown-for-jupyter-notebooks-cheatsheet-386c05aeebed