

AirBnb Exploratory Analysis

November 16, 2021

1 2. Data Understanding

1.0.1 Import Libraries

```
[2]: import os
import sys
import time
import numpy as np
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
import seaborn as sns
import dataframe_image as dfi
from ast import literal_eval

%matplotlib inline
```

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-data/stylelib/_classic_test.mplstyle:

The text.latex.preview rcparam was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-data/stylelib/_classic_test.mplstyle:

The mathtext.fallback_to_cm rcparam was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-data/stylelib/_classic_test.mplstyle: Support for setting the 'mathtext.fallback_to_cm' rcParam is deprecated since 3.3 and will be removed two minor releases later; use 'mathtext.fallback : 'cm' instead.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-data/stylelib/_classic_test.mplstyle:

The validate_bool_maybe_none function was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-data/stylelib/_classic_test.mplstyle:

The savefig.jpeg_quality rcparam was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-

data/stylelib/_classic_test.mplstyle:

The keymap.all_axes rcparam was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-

data/stylelib/_classic_test.mplstyle:

The animation.avconv_path rcparam was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

In /opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-

data/stylelib/_classic_test.mplstyle:

The animation.avconv_args rcparam was deprecated in Matplotlib 3.3 and will be removed two minor releases later.

```
[3]: # pip install geopandas
```

1.0.2 Load Data

```
[4]: # Loading csv file into dataframe
df_listings = pd.read_csv('listings.csv', low_memory=False)
```

```
[5]: # Number of rows and columns in the dataframe
df_listings.shape
```

```
[5]: (8111, 106)
```

```
[6]: # Displaying all columns
pd.set_option('display.max_columns', None)

# show listings data sample
df_listings.head(10)
```

```
[6]:
```

	id	listing_url	scrape_id	last_scraped	\
0	958	https://www.airbnb.com/rooms/958	2.019100e+13	10/14/19	
1	3850	https://www.airbnb.com/rooms/3850	2.019100e+13	10/14/19	
2	5858	https://www.airbnb.com/rooms/5858	2.019100e+13	10/14/19	
3	7918	https://www.airbnb.com/rooms/7918	2.019100e+13	10/14/19	
4	8142	https://www.airbnb.com/rooms/8142	2.019100e+13	10/14/19	
5	8339	https://www.airbnb.com/rooms/8339	2.019100e+13	10/14/19	
6	8739	https://www.airbnb.com/rooms/8739	2.019100e+13	10/14/19	
7	9225	https://www.airbnb.com/rooms/9225	2.019100e+13	10/14/19	
8	10251	https://www.airbnb.com/rooms/10251	2.019100e+13	10/14/19	
9	10578	https://www.airbnb.com/rooms/10578	2.019100e+13	10/14/19	

	name	\
0	Bright, Modern Garden Unit - 1BR/1B	
1	Charming room for two	
2	Creative Sanctuary	
3	A Friendly Room - UCSF/USF - San Francisco	
4	Friendly Room Apt. Style -UCSF/USF - San Franc...	

5 Historic Alamo Square Victorian
6 Mission Sunshine, with Private Bath
7 Artful Potrero Separate Floor with Garden
8 Victorian Suite in Inner Mission
9 Classic Nob Hill Studio - Roof Deck

summary \

0 New update: the house next door is under const...
1 Your own private room plus access to a shared ...
2 NaN
3 Nice and good public transportation. 7 minute...
4 Nice and good public transportation. 7 minute...
5 Pls email before booking. Interior featured i...
6 Welcome to "The Mission," the sunniest neighbo...
7 A Unique Guest Suite! A Spacious Art Filled ...
8 NaN
9 A cute studio with nice street views and lots ...

space \

0 Newly remodeled, modern, and bright garden uni...
1 This room can fit two people. Nobody else will...
2 We live in a large Victorian house on a quiet ...
3 Settle down, S.F. resident, student, hospital,...
4 Settle down, S.F. resident, student, hospital,...
5 Please send us a quick message before booking ...
6 Your sunny room looks out over a lush garden, ...
7 It is unique, very spacious [800 sq. ft. with...
8 Please read this before you book! Second floor...
9 Features: Elevator, Wood floors, TV, DVD Play...

description experiences_offered \

0	New update: the house next door is under const...	none
1	Your own private room plus access to a shared ...	none
2	We live in a large Victorian house on a quiet ...	none
3	Nice and good public transportation. 7 minute...	none
4	Nice and good public transportation. 7 minute...	none
5	Pls email before booking. Interior featured i...	none
6	Welcome to "The Mission," the sunniest neighbo...	none
7	A Unique Guest Suite! A Spacious Art Filled ...	none
8	Please read this before you book! Second floor...	none
9	A cute studio with nice street views and lots ...	none

neighborhood_overview \

0 *Quiet cul de sac in friendly neighborhood *St...
1 This is a quiet, safe neighborhood on a substa...
2 I love how our neighborhood feels quiet but is...
3 Shopping old town, restaurants, McDonald, Whol...

4 NaN
5 NaN
6 Located between Valencia Street and Dolores Pa...
7 It is a distinct neighborhood, with easy acces...
8 Neighborhood is safe, sunny, lively and fun. ...
9 Very centrally located. Fishermans Wharf, Chi...

notes \

0 Due to the fact that we have children and a do...
1 House Rule footnotes: 1.\tI don't allow check ...
2 All the furniture in the house was handmade so...
3 Wi-Fi signal in common areas. Large eat in k...
4 Wi-Fi signal in common areas. Large eat in k...
5 tax ID on file tax ID on file
6 We live in a dense, urban neighborhood, and ou...
7 NaN
8 San Francisco Office of Short-Term Rentals Bus...
9 NaN

transit \

0 *Public Transportation is 1/2 block away. *Ce...
1 Public transit service to my house is outstand...
2 The train is two blocks away and you can stop ...
3 N Juda Muni and bus stop. Street parking.
4 N Juda Muni, Bus and UCSF Shuttle. small shopp...
5 NaN
6 It's an easy and enjoyable 10 minute stroll to...
7 There are buses within a block to connect to a...
8 SF Muni line 12, bus stop is a block away from...
9 The California 1 bus, which runs from downtown...

access \

0 *Full access to patio and backyard (shared wit...
1 During the hours of 7:00 to 22:00, your access...
2 Our deck, garden, gourmet kitchen and extensiv...
3 NaN
4 NaN
5 Guests have access to everything listed and sh...
6 You'll have your own bedroom and private bathr...
7 Laundry privileges, Spacious Garden with City ...
8 You are welcome to relax and enjoy our backyar...
9 Shared access to the roof deck. The studio is ...

interaction \

0 A family of 4 lives upstairs with their dog. N...
1 I live at the house in my own bedroom, and I w...
2 NaN

3		NaN
4		NaN
5		NaN
6	We are experienced Airbnb hosts, and have had ...	
7	We are available to make suggestions to enhanc...	
8	Feel free to knock on the downstairs door if y...	
9	Usually zero, as the place is set up with door...	

	house_rules	thumbnail_url	\
0	* No Pets - even visiting guests for a short t...	NaN	
1	Airbnb has no default house rules that hosts c...	NaN	
2	Please respect the house, the art work, the fu...	NaN	
3	No party, No smoking, not for any kinds of smo...	NaN	
4	no pet no smoke no party inside the building	NaN	
5	House Manual and House Rules will be provided ...	NaN	
6	PLEASE READ: By booking, you agree to all of t...	NaN	
7	We are respectful of the privacy of our renter...	NaN	
8	Basic rules: No smoking inside the suite or on...	NaN	
9	No smoking, no pets. Sorry.	NaN	

	medium_url	picture_url	\
0	NaN	https://a0.muscache.com/im/pictures/b7c2a199-4...	
1	NaN	https://a0.muscache.com/im/pictures/b6e97d2a-1...	
2	NaN	https://a0.muscache.com/im/pictures/17714/3a7a...	
3	NaN	https://a0.muscache.com/im/pictures/26356/8030...	
4	NaN	https://a0.muscache.com/im/pictures/27832/3b1f...	
5	NaN	https://a0.muscache.com/im/pictures/213fbf05-3...	
6	NaN	https://a0.muscache.com/im/pictures/80811/7a57...	
7	NaN	https://a0.muscache.com/im/pictures/80742/6401...	
8	NaN	https://a0.muscache.com/im/pictures/30649314/e...	
9	NaN	https://a0.muscache.com/im/pictures/2249128/7f...	

	xl_picture_url	host_id	host_url	\
0	NaN	1169	https://www.airbnb.com/users/show/1169	
1	NaN	4921	https://www.airbnb.com/users/show/4921	
2	NaN	8904	https://www.airbnb.com/users/show/8904	
3	NaN	21994	https://www.airbnb.com/users/show/21994	
4	NaN	21994	https://www.airbnb.com/users/show/21994	
5	NaN	24215	https://www.airbnb.com/users/show/24215	
6	NaN	7149	https://www.airbnb.com/users/show/7149	
7	NaN	29674	https://www.airbnb.com/users/show/29674	
8	NaN	35199	https://www.airbnb.com/users/show/35199	
9	NaN	37049	https://www.airbnb.com/users/show/37049	

	host_name	host_since	host_location	\
0	Holly	7/31/08	San Francisco, California, United States	
1	Kevin	12/8/08	San Francisco, California, United States	

2	Philip And Tania	3/2/09	San Francisco, California, United States
3	Aaron	6/17/09	San Francisco, California, United States
4	Aaron	6/17/09	San Francisco, California, United States
5	Rosy	7/2/09	San Francisco, California, United States
6	Ivan & Wendy	1/27/09	San Francisco, California, United States
7	Gae	8/4/09	San Francisco, California, United States
8	Roman & Sarah	8/31/09	San Francisco, California, United States
9	Andrew	9/8/09	San Francisco, California, United States

	host_about	host_response_time	\
0	We are a family with 2 boys born in 2009 and 2...	within a day	
1	I am a retired software developer now raising ...	within an hour	
2	Philip: English transplant to the Bay Area and...	within a day	
3	7 minutes walk to UCSF hospital & school campu...	within an hour	
4	7 minutes walk to UCSF hospital & school campu...	within an hour	
5	I'm an Interior Stylist living in SF. \r\n\r\n...	within a few hours	
6	Ivan is a publisher of a book for 4 year olds...	within an hour	
7	The apartment is on the ground floor of the ho...	within an hour	
8	Coming from SF.	within a few hours	
9	I'm a writer, and encourage visiting writers a...	within a few hours	

	host_response_rate	host_acceptance_rate	host_is_superhost	\
0	100%	NaN	t	
1	100%	NaN	t	
2	80%	NaN	f	
3	86%	NaN	t	
4	86%	NaN	t	
5	100%	NaN	f	
6	100%	NaN	t	
7	100%	NaN	t	
8	100%	NaN	t	
9	100%	NaN	f	

	host_thumbnail_url	\
0	https://a0.muscache.com/im/pictures/efdad96a-3...	
1	https://a0.muscache.com/im/pictures/97d29791-d...	
2	https://a0.muscache.com/im/users/8904/profile_...	
3	https://a0.muscache.com/im/users/21994/profile...	
4	https://a0.muscache.com/im/users/21994/profile...	
5	https://a0.muscache.com/im/pictures/user/6e05b...	
6	https://a0.muscache.com/im/users/7149/profile_...	
7	https://a0.muscache.com/im/users/29674/profile...	
8	https://a0.muscache.com/im/users/35199/profile...	
9	https://a0.muscache.com/im/pictures/user/885cb...	

	host_picture_url	host_neighbourhood	\
0	https://a0.muscache.com/im/pictures/efdad96a-3...	Duboce Triangle	

1	https://a0.muscache.com/im/pictures/97d29791-d...	Inner Sunset
2	https://a0.muscache.com/im/users/8904/profile_...	Bernal Heights
3	https://a0.muscache.com/im/users/21994/profile...	Cole Valley
4	https://a0.muscache.com/im/users/21994/profile...	Cole Valley
5	https://a0.muscache.com/im/pictures/user/6e05b...	Alamo Square
6	https://a0.muscache.com/im/users/7149/profile_...	Mission District
7	https://a0.muscache.com/im/users/29674/profile...	Potrero Hill
8	https://a0.muscache.com/im/users/35199/profile...	Mission District
9	https://a0.muscache.com/im/pictures/user/885cb...	Nob Hill

	host_listings_count	host_total_listings_count	\
0	1.0	1.0	
1	2.0	2.0	
2	2.0	2.0	
3	10.0	10.0	
4	10.0	10.0	
5	2.0	2.0	
6	2.0	2.0	
7	1.0	1.0	
8	1.0	1.0	
9	0.0	0.0	

	host_verifications	host_has_profile_pic	\
0	['email', 'phone', 'facebook', 'reviews', 'kba']	t	
1	['email', 'phone', 'reviews', 'jumio', 'govern...	t	
2	['email', 'phone', 'reviews', 'kba', 'work_ema...	t	
3	['email', 'phone', 'reviews', 'jumio', 'govern...	t	
4	['email', 'phone', 'reviews', 'jumio', 'govern...	t	
5	['email', 'phone', 'reviews', 'kba']	t	
6	['email', 'phone', 'reviews', 'kba', 'work_ema...	t	
7	['email', 'phone', 'facebook', 'reviews', 'kba']	t	
8	['email', 'phone', 'facebook', 'reviews', 'kba']	t	
9	['email', 'phone', 'reviews', 'jumio', 'govern...	t	

	host_identity_verified	street	neighbourhood	\
0	t	San Francisco, CA, United States	Duboce Triangle	
1	t	San Francisco, CA, United States	Inner Sunset	
2	t	San Francisco, CA, United States	Bernal Heights	
3	t	San Francisco, CA, United States	Cole Valley	
4	t	San Francisco, CA, United States	Cole Valley	
5	t	San Francisco, CA, United States	Alamo Square	
6	f	San Francisco, CA, United States	Mission District	
7	t	San Francisco, CA, United States	Potrero Hill	
8	t	San Francisco, CA, United States	Mission District	
9	t	San Francisco, CA, United States	Nob Hill	

neighbourhood_cleansed	neighbourhood_group_cleansed	city	state	\
------------------------	------------------------------	------	-------	---

0	Western Addition	NaN	San Francisco	CA
1	Inner Sunset	NaN	San Francisco	CA
2	Bernal Heights	NaN	San Francisco	CA
3	Haight Ashbury	NaN	San Francisco	CA
4	Haight Ashbury	NaN	San Francisco	CA
5	Western Addition	NaN	San Francisco	CA
6	Mission	NaN	San Francisco	CA
7	Potrero Hill	NaN	San Francisco	CA
8	Mission	NaN	San Francisco	CA
9	Nob Hill	NaN	San Francisco	CA

	zipcode	market	smart_location	country_code	country	\
0	94117	San Francisco	San Francisco, CA	US	United States	
1	94131	San Francisco	San Francisco, CA	US	United States	
2	94110	San Francisco	San Francisco, CA	US	United States	
3	94117	San Francisco	San Francisco, CA	US	United States	
4	94117	San Francisco	San Francisco, CA	US	United States	
5	94117	San Francisco	San Francisco, CA	US	United States	
6	94110	San Francisco	San Francisco, CA	US	United States	
7	94107	San Francisco	San Francisco, CA	US	United States	
8	94110	San Francisco	San Francisco, CA	US	United States	
9	94109	San Francisco	San Francisco, CA	US	United States	

	latitude	longitude	is_location_exact	property_type	room_type	\
0	37.76931	-122.43386	t	Apartment	Entire home/apt	
1	37.75402	-122.45805	t	House	Private room	
2	37.74511	-122.42102	t	Apartment	Entire home/apt	
3	37.76669	-122.45250	t	Apartment	Private room	
4	37.76487	-122.45183	t	Apartment	Private room	
5	37.77525	-122.43637	t	House	Entire home/apt	
6	37.75919	-122.42237	t	Condominium	Private room	
7	37.76259	-122.40543	t	House	Private room	
8	37.75874	-122.41327	t	House	Entire home/apt	
9	37.79143	-122.41544	t	Apartment	Entire home/apt	

	accommodates	bathrooms	bedrooms	beds	bed_type	\
0	3	1.0	1.0	2.0	Real Bed	
1	2	1.0	1.0	1.0	Real Bed	
2	5	1.0	2.0	3.0	Real Bed	
3	2	4.0	1.0	1.0	Real Bed	
4	2	4.0	1.0	1.0	Real Bed	
5	4	1.5	2.0	2.0	Real Bed	
6	3	1.0	1.0	2.0	Real Bed	
7	2	1.0	1.0	1.0	Real Bed	
8	4	1.0	2.0	3.0	Real Bed	
9	2	1.0	0.0	1.0	Real Bed	

	amenities	square_feet	price \
0	{TV,"Cable TV",Internet,Wifi,Kitchen,"Pets liv...	NaN	\$170.00
1	{Internet,Wifi,Kitchen,Breakfast,"Free street ...	NaN	\$99.00
2	{Internet,Wifi,Kitchen,Heating,"Family/kid fri...	NaN	\$235.00
3	{TV,Internet,Wifi,Kitchen,"Free street parking...	NaN	\$65.00
4	{TV,Internet,Wifi,Kitchen,"Free street parking...	NaN	\$65.00
5	{TV,Internet,Wifi,Kitchen,"Free street parking...	NaN	\$875.00
6	{TV,"Cable TV",Internet,Wifi,"Pets live on thi...	NaN	\$139.00
7	{TV,"Cable TV",Internet,Wifi,Kitchen,"Pets liv...	NaN	\$135.00
8	{TV,"Cable TV",Internet,Wifi,Kitchen,"Free str...	NaN	\$225.00
9	{TV,"Cable TV",Internet,Wifi,Kitchen,Elevator,...	NaN	\$120.00

	weekly_price	monthly_price	security_deposit	cleaning_fee	guests_included \
0	\$1,120.00	\$4,200.00	\$100.00	\$100.00	2
1	NaN	NaN	\$0.00	\$10.00	2
2	\$1,600.00	\$5,500.00	NaN	\$100.00	2
3	\$485.00	\$1,685.00	\$200.00	\$50.00	1
4	\$490.00	\$1,685.00	\$200.00	\$50.00	1
5	NaN	NaN	\$0.00	\$225.00	2
6	NaN	\$9,999.00	\$0.00	\$50.00	2
7	\$1,000.00	\$4,000.00	NaN	\$50.00	1
8	\$1,540.00	\$5,400.00	\$500.00	\$100.00	1
9	\$1,150.00	\$2,875.00	\$500.00	\$75.00	2

	extra_people	minimum_nights	maximum_nights	minimum_minimum_nights \
0	\$25.00	1	30	1
1	\$20.00	1	5	1
2	\$0.00	30	60	30
3	\$12.00	32	60	32
4	\$12.00	32	90	32
5	\$150.00	6	1125	6
6	\$60.00	1	14	1
7	\$0.00	1	365	1
8	\$0.00	30	60	30
9	\$0.00	30	180	30

	maximum_minimum_nights	minimum_maximum_nights	maximum_maximum_nights \
0	1	30	30
1	1	5	5
2	30	60	60
3	32	60	60
4	32	90	90
5	6	1125	1125
6	1	14	14
7	1	365	365
8	30	60	60
9	30	180	180

	minimum_nights_avg_ntm	maximum_nights_avg_ntm	calendar_updated	\
0	1.0	30.0	2 months ago	
1	1.0	5.0	3 days ago	
2	30.0	60.0	3 weeks ago	
3	32.0	60.0	3 months ago	
4	32.0	90.0	3 months ago	
5	6.0	1125.0	7 weeks ago	
6	1.0	14.0	2 weeks ago	
7	1.0	365.0	2 days ago	
8	30.0	60.0	4 weeks ago	
9	30.0	180.0	2 months ago	

	has_availability	availability_30	availability_60	availability_90	\
0	t	3	8	16	
1	t	5	32	62	
2	t	0	0	0	
3	t	30	60	90	
4	t	30	60	90	
5	t	30	60	90	
6	t	7	21	47	
7	t	15	41	71	
8	t	13	43	73	
9	t	0	27	57	

	availability_365	calendar_last_scraped	number_of_reviews	\
0	85	10/14/19	217	
1	62	10/14/19	160	
2	0	10/14/19	111	
3	365	10/14/19	18	
4	365	10/14/19	8	
5	90	10/14/19	28	
6	135	10/14/19	704	
7	344	10/14/19	511	
8	348	10/14/19	337	
9	332	10/14/19	18	

	number_of_reviews_ltm	first_review	last_review	review_scores_rating	\
0	52	7/23/09	9/21/19	97.0	
1	36	7/16/09	9/30/19	94.0	
2	0	5/3/09	8/6/17	98.0	
3	1	8/31/09	9/10/19	86.0	
4	0	9/8/14	9/12/18	93.0	
5	1	9/25/09	6/28/19	97.0	
6	87	8/1/09	10/6/19	98.0	
7	82	10/26/09	10/7/19	94.0	
8	33	9/12/09	5/27/19	96.0	

9	0	12/23/11	5/17/15	99.0
---	---	----------	---------	------

	review_scores_accuracy	review_scores_cleanliness	review_scores_checkin	\
0	10.0	10.0	10.0	
1	10.0	10.0	10.0	
2	10.0	10.0	10.0	
3	8.0	8.0	9.0	
4	9.0	9.0	10.0	
5	10.0	10.0	10.0	
6	10.0	10.0	10.0	
7	10.0	9.0	10.0	
8	10.0	10.0	10.0	
9	10.0	10.0	10.0	

	review_scores_communication	review_scores_location	review_scores_value	\
0	10.0	10.0	9.0	
1	10.0	10.0	10.0	
2	10.0	10.0	9.0	
3	9.0	9.0	8.0	
4	10.0	9.0	9.0	
5	10.0	10.0	10.0	
6	10.0	10.0	9.0	
7	10.0	10.0	9.0	
8	10.0	10.0	9.0	
9	10.0	10.0	10.0	

	requires_license	license	jurisdiction_names	instant_bookable	\
0	t	STR-0001256	{"SAN FRANCISCO"}	f	
1	t	STR-0000346	{"SAN FRANCISCO"}	f	
2	t	NaN	{"SAN FRANCISCO"}	f	
3	t	NaN	{"SAN FRANCISCO"}	f	
4	t	NaN	{"SAN FRANCISCO"}	f	
5	t	STR-0000264	{"SAN FRANCISCO"}	f	
6	t	STR-0000028	{"SAN FRANCISCO"}	t	
7	t	STR-0000179	{"SAN FRANCISCO"}	f	
8	t	STR-0001558	{"SAN FRANCISCO"}	f	
9	t	NaN	{"SAN FRANCISCO"}	f	

	is_business_travel_ready	cancellation_policy	\
0	f	moderate	
1	f	strict_14_with_grace_period	
2	f	strict_14_with_grace_period	
3	f	strict_14_with_grace_period	
4	f	strict_14_with_grace_period	
5	f	moderate	
6	f	strict_14_with_grace_period	
7	f	strict_14_with_grace_period	

8	f	moderate
9	f	moderate

	require_guest_profile_picture	require_guest_phone_verification	\
0	f	f	
1	f	f	
2	f	f	
3	f	f	
4	f	f	
5	t	t	
6	f	f	
7	f	f	
8	f	f	
9	f	t	

	calculated_host_listings_count	\
0	1	
1	3	
2	1	
3	9	
4	9	
5	2	
6	2	
7	1	
8	1	
9	1	

	calculated_host_listings_count_entire_homes	\
0	1	
1	0	
2	1	
3	0	
4	0	
5	2	
6	0	
7	0	
8	1	
9	1	

	calculated_host_listings_count_private_rooms	\
0	0	
1	3	
2	0	
3	9	
4	9	
5	0	
6	2	

7	1
8	0
9	0

	calculated_host_listings_count_shared_rooms	reviews_per_month
0	0	1.74
1	0	1.28
2	0	0.87
3	0	0.15
4	0	0.13
5	0	0.23
6	0	5.67
7	0	4.21
8	0	2.74
9	0	0.19

```
[7]: # General statistics of the data
df_listings.describe()
```

```
[7]:
```

	id	scrape_id	thumbnail_url	medium_url	xl_picture_url	\
count	8.111000e+03	8.111000e+03	0.0	0.0	0.0	
mean	2.024561e+07	2.019100e+13	NaN	NaN	NaN	
std	1.228565e+07	0.000000e+00	NaN	NaN	NaN	
min	9.580000e+02	2.019100e+13	NaN	NaN	NaN	
25%	8.905668e+06	2.019100e+13	NaN	NaN	NaN	
50%	2.161057e+07	2.019100e+13	NaN	NaN	NaN	
75%	3.120025e+07	2.019100e+13	NaN	NaN	NaN	
max	3.935418e+07	2.019100e+13	NaN	NaN	NaN	

	host_id	host_acceptance_rate	host_listings_count	\
count	8.111000e+03	0.0	8103.000000	
mean	6.159690e+07	NaN	90.421079	
std	8.017417e+07	NaN	324.526784	
min	2.800000e+01	NaN	0.000000	
25%	4.386571e+06	NaN	1.000000	
50%	2.178983e+07	NaN	2.000000	
75%	1.000763e+08	NaN	12.000000	
max	3.016759e+08	NaN	1735.000000	

	host_total_listings_count	neighbourhood_group_cleansed	latitude	\
count	8103.000000	0.0	8111.000000	
mean	90.421079	NaN	37.766054	
std	324.526784	NaN	0.022937	
min	0.000000	NaN	37.704630	
25%	1.000000	NaN	37.751450	
50%	2.000000	NaN	37.769150	
75%	12.000000	NaN	37.785670	

max	1735.000000	NaN	37.828790
-----	-------------	-----	-----------

	longitude	accommodates	bathrooms	bedrooms	beds \
count	8111.000000	8111.000000	8099.000000	8107.000000	8102.000000
mean	-122.430107	3.160769	1.395975	1.345874	1.744754
std	0.026967	1.896988	0.923213	0.925298	1.158799
min	-122.513060	1.000000	0.000000	0.000000	0.000000
25%	-122.442830	2.000000	1.000000	1.000000	1.000000
50%	-122.424650	2.000000	1.000000	1.000000	1.000000
75%	-122.410615	4.000000	1.500000	2.000000	2.000000
max	-122.368570	16.000000	14.000000	14.000000	14.000000

	square_feet	guests_included	minimum_nights	maximum_nights \
count	124.000000	8111.000000	8.111000e+03	8.111000e+03
mean	1046.177419	1.718407	1.234526e+04	1.284597e+04
std	827.297727	1.334492	1.110357e+06	1.110352e+06
min	0.000000	1.000000	1.000000e+00	1.000000e+00
25%	500.000000	1.000000	2.000000e+00	2.900000e+01
50%	851.000000	1.000000	4.000000e+00	1.800000e+02
75%	1492.500000	2.000000	3.000000e+01	1.125000e+03
max	5500.000000	16.000000	1.000000e+08	1.000000e+08

	minimum_minimum_nights	maximum_minimum_nights	minimum_maximum_nights \
count	8.111000e+03	8.111000e+03	8.111000e+03
mean	1.234509e+04	1.235125e+04	2.775902e+05
std	1.110357e+06	1.110357e+06	2.387044e+07
min	1.000000e+00	1.000000e+00	1.000000e+00
25%	2.000000e+00	2.000000e+00	2.900000e+01
50%	4.000000e+00	5.000000e+00	1.800000e+02
75%	3.000000e+01	3.000000e+01	1.125000e+03
max	1.000000e+08	1.000000e+08	2.147484e+09

	maximum_maximum_nights	minimum_nights_avg_ntm	maximum_nights_avg_ntm \
count	8.111000e+03	8.111000e+03	8.111000e+03
mean	2.775928e+05	1.235001e+04	2.775923e+05
std	2.387044e+07	1.110357e+06	2.387044e+07
min	1.000000e+00	1.000000e+00	1.000000e+00
25%	2.900000e+01	2.000000e+00	2.900000e+01
50%	1.800000e+02	4.000000e+00	1.800000e+02
75%	1.125000e+03	3.000000e+01	1.125000e+03
max	2.147484e+09	1.000000e+08	2.147484e+09

	availability_30	availability_60	availability_90	availability_365 \
count	8111.000000	8111.000000	8111.000000	8111.000000
mean	9.255825	22.885341	39.619529	159.278881
std	10.586687	21.459500	32.125850	134.093744
min	0.000000	0.000000	0.000000	0.000000

25%	0.000000	0.000000	6.000000	28.000000
50%	5.000000	19.000000	38.000000	133.000000
75%	17.000000	42.000000	69.500000	300.000000
max	30.000000	60.000000	90.000000	365.000000

	number_of_reviews	number_of_reviews_ltm	review_scores_rating \
count	8111.000000	8111.000000	6460.000000
mean	45.203181	13.712736	95.421517
std	78.437987	22.523539	7.571031
min	0.000000	0.000000	20.000000
25%	1.000000	0.000000	95.000000
50%	10.000000	2.000000	98.000000
75%	54.000000	18.000000	100.000000
max	757.000000	258.000000	100.000000

	review_scores_accuracy	review_scores_cleanliness \
count	6457.000000	6457.000000
mean	9.752207	9.603686
std	0.732872	0.839768
min	2.000000	2.000000
25%	10.000000	9.000000
50%	10.000000	10.000000
75%	10.000000	10.000000
max	10.000000	10.000000

	review_scores_checkin	review_scores_communication \
count	6456.000000	6458.000000
mean	9.847584	9.817745
std	0.593539	0.659274
min	2.000000	2.000000
25%	10.000000	10.000000
50%	10.000000	10.000000
75%	10.000000	10.000000
max	10.000000	10.000000

	review_scores_location	review_scores_value \
count	6456.000000	6456.000000
mean	9.682776	9.375155
std	0.715906	0.879174
min	2.000000	2.000000
25%	10.000000	9.000000
50%	10.000000	10.000000
75%	10.000000	10.000000
max	10.000000	10.000000

	calculated_host_listings_count \
count	8111.000000

mean	22.329799
std	54.369225
min	1.000000
25%	1.000000
50%	2.000000
75%	10.000000
max	248.000000

calculated_host_listings_count_entire_homes \	
count	8111.000000
mean	18.201085
std	53.333669
min	0.000000
25%	0.000000
50%	1.000000
75%	2.000000
max	248.000000

calculated_host_listings_count_private_rooms \	
count	8111.000000
mean	2.910985
std	6.807334
min	0.000000
25%	0.000000
50%	0.000000
75%	2.000000
max	39.000000

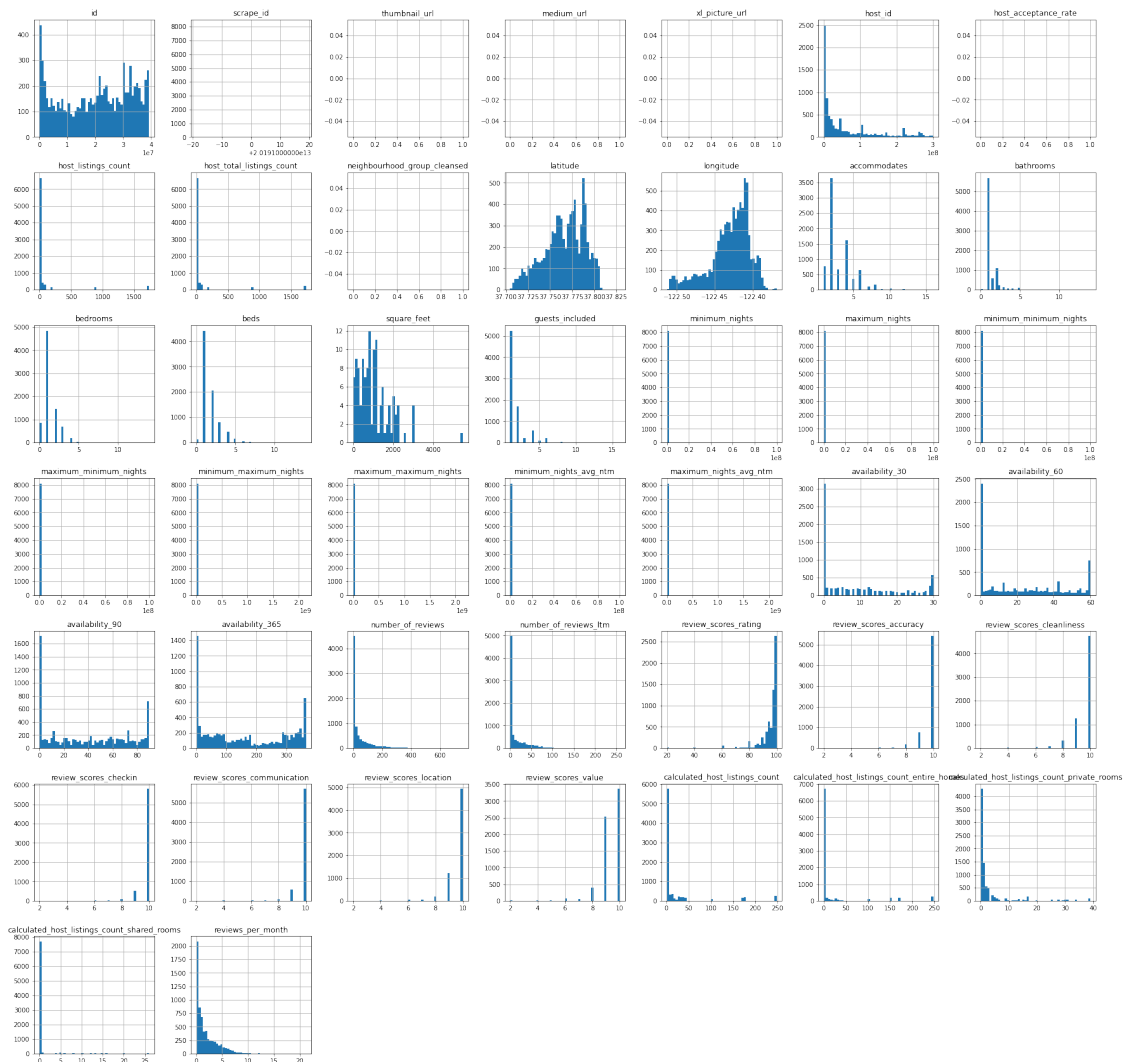
calculated_host_listings_count_shared_rooms		reviews_per_month
count	8111.000000	6506.000000
mean	0.496979	1.873298
std	2.669129	2.096730
min	0.000000	0.010000
25%	0.000000	0.300000
50%	0.000000	1.010000
75%	0.000000	2.840000
max	26.000000	21.190000

```
[8]: # Dataframe columns information
df_listings.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8111 entries, 0 to 8110
Columns: 106 entries, id to reviews_per_month
dtypes: float64(24), int64(20), object(62)
memory usage: 6.6+ MB
```



```
[9]: # Checking the distribution of all variables
df_listings.hist(bins=50, figsize=(30,30))
plt.show()
```



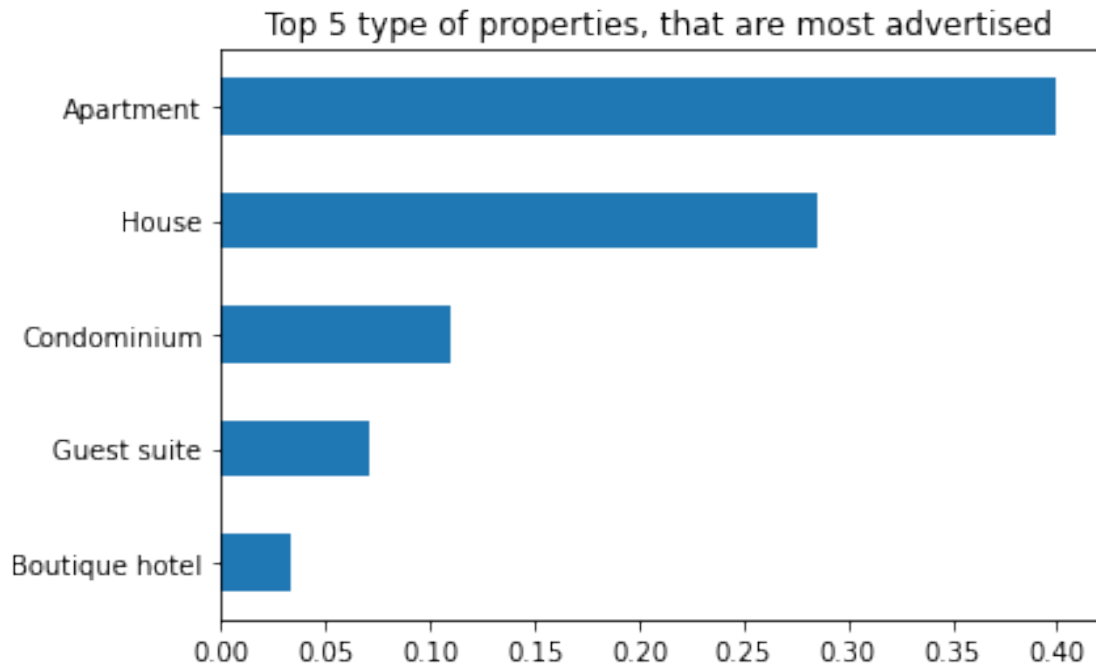
```
[10]: # Finding the columns with more than 75% of values missing
set(df_listings.columns[df_listings.isnull().mean(>0.75])
```

```
[10]: {'host_acceptance_rate',
'medium_url',
'monthly_price',
'neighbourhood_group_cleansed',
'square_feet',
'thumbnail_url',
'weekly_price',
```

```
'xl_picture_url']}
```

```
[11]: # Exploring the data for number of property types
neighbourhood_vals = df_listings['property_type'].value_counts()

((neighbourhood_vals[0:5]/df_listings.shape[0]).sort_values()).
    plot(kind="barh");
plt.title("Top 5 type of properties, that are most advertised");
```

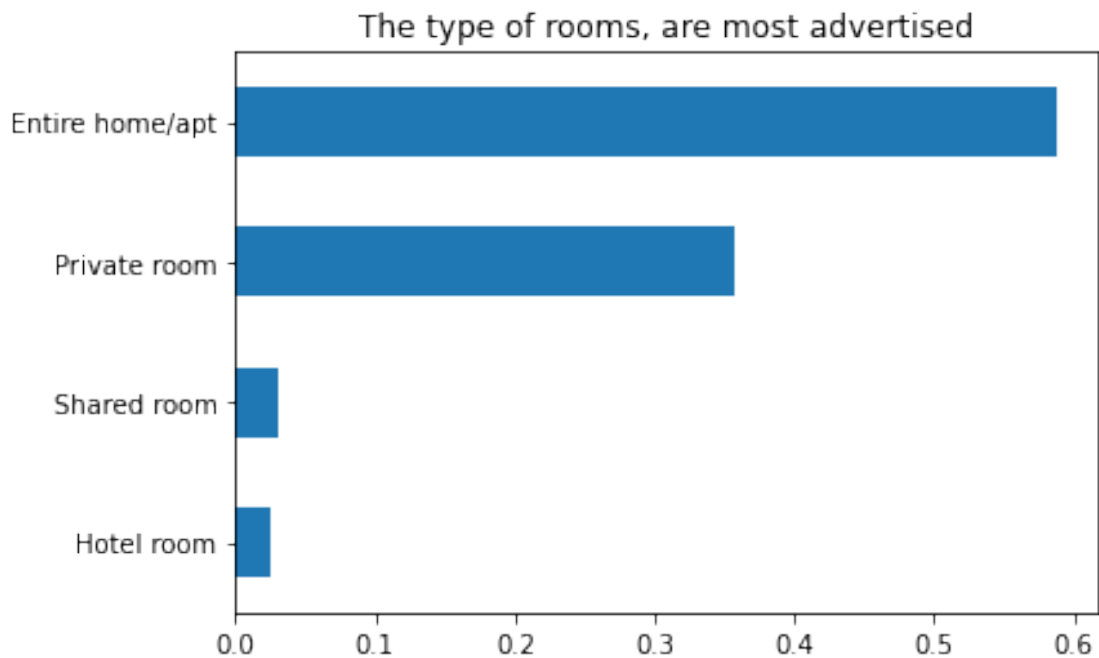


```
[12]: # Checking the property type count above 100
neighbourhood_vals[df_listings['property_type'].value_counts() > 100]
```

```
[12]: Apartment      3243
      House        2314
      Condominium    892
      Guest suite    577
      Boutique hotel  267
      Hotel         157
      Townhouse     142
      Serviced apartment 121
      Name: property_type, dtype: int64
```

```
[13]: # Exploring the data for number of room types
room_vals = df_listings['room_type'].value_counts()
```

```
((room_vals/df_listings.shape[0]).sort_values()).plot(kind="barh");
plt.title("The type of rooms, are most advertised");
```



```
[14]: # Exploring the data of bathrooms_text
df_listings['bathrooms'].value_counts()
```

```
[14]: 1.0    5671
      2.0    1111
      1.5     579
      2.5     234
      3.0     149
      5.0     113
      3.5      63
      4.0      61
      0.0      38
      10.0     19
      0.5      17
      4.5      14
      8.0      14
      6.0       9
      7.0       5
      14.0      1
      6.5       1
      Name: bathrooms, dtype: int64
```

1.1 3. Data Preparation

```
[15]: class format_data:

    """
    Class for applying a new format to the numerical columns so that they can
    ↪ be explored

    data_float function converts object to float
    data_percentage function converts object to percentage

    Parameters:

    Dataframe, "column name"

    Returns:

    Dataframe

    """

    def __init__(self, dataframe):
        self.dataframe = dataframe

    def data_float(self):
        self = self.str.replace('$', '', regex=True).replace(',', '',
        ↪ regex=True).astype(float)

        return self
```

```
[16]: # Function for finding outliers
def find_outliers(df_listings, column_name):

    """
    The function finds outliers and removes them
    from the dataframe.

    Parameters:

    Dataframe, column_name
```

Returns:

Dataframe

"""

Calculate Q1 (25th percentile of the data) for the given feature

```
Q1 = np.percentile(df_listings[column_name], 25)
```

#Calculate Q3 (75th percentile of the data) for the given feature

```
Q3 = np.percentile(df_listings[column_name],75)
```

Use the interquartile range to calculate an outlier step (1.5 times the
↪interquartile range)

```
step = (Q3 - Q1) * 1.5
```

Display the outliers

#print("Data points considered outliers for the feature '{}':".

↪format("price"))

```
filtered_data = df_listings[~((df_listings[column_name] >= Q1 - step) &  
                             (df_listings[column_name] <= Q3 + step))].
```

↪sort_values(by=[column_name])

```
return filtered_data
```

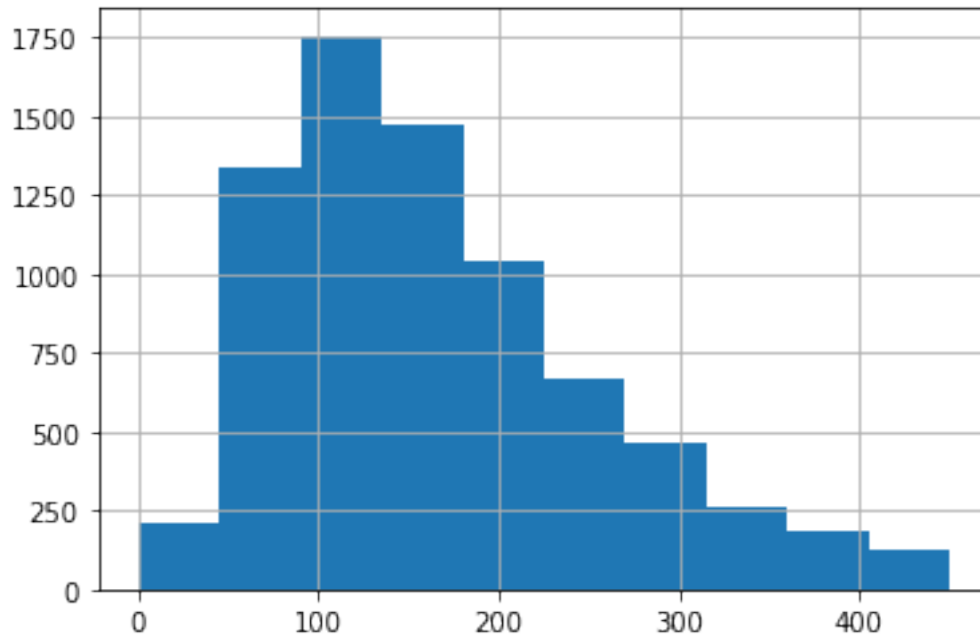
```
[17]: # Applying format to the price column  
float_list = ['price']  
for x in float_list:  
    df_listings[x] = format_data.data_float(df_listings[x])
```

1.1.1 Filter Outliers

```
[18]: # Applying the find_outliers function  
outliers = find_outliers(df_listings, "price").index  
  
df_listings = df_listings.drop(df_listings.index[outliers]).reset_index(drop =  
↪True)
```

```
[19]: # Checking the new distribution of prices  
df_listings["price"].hist()
```

```
[19]: <AxesSubplot:>
```



1.1.2 Business Questions

```
[20]: def chart_plot(data, x_label, y_label):  
  
    """  
    Function creates a plot  
    """  
  
    plt.figure(figsize=(10,5))  
  
    sns.set(style="whitegrid")  
    chart = sns.barplot(x=x_label, y=y_label, data=data)  
  
    chart.set_xticklabels(  
        chart.get_xticklabels(),  
        rotation=60,  
        horizontalalignment='right',  
        fontweight='light',  
        fontsize='medium',  
        rotation_mode='anchor'  
    )
```

```
[21]: def type_grouping(column):
```

```

"""
Create dictionary count

Parameters:

Column name

Returns:

Dictionary

"""

verifications_dict = {}
temp = []

for i in column:
    temp.append(eval(i))

for i in temp:
    try:
        for j in i:
            if j not in verifications_dict:
                verifications_dict[j] = 1
            else:
                verifications_dict[j] += 1
    except:
        continue

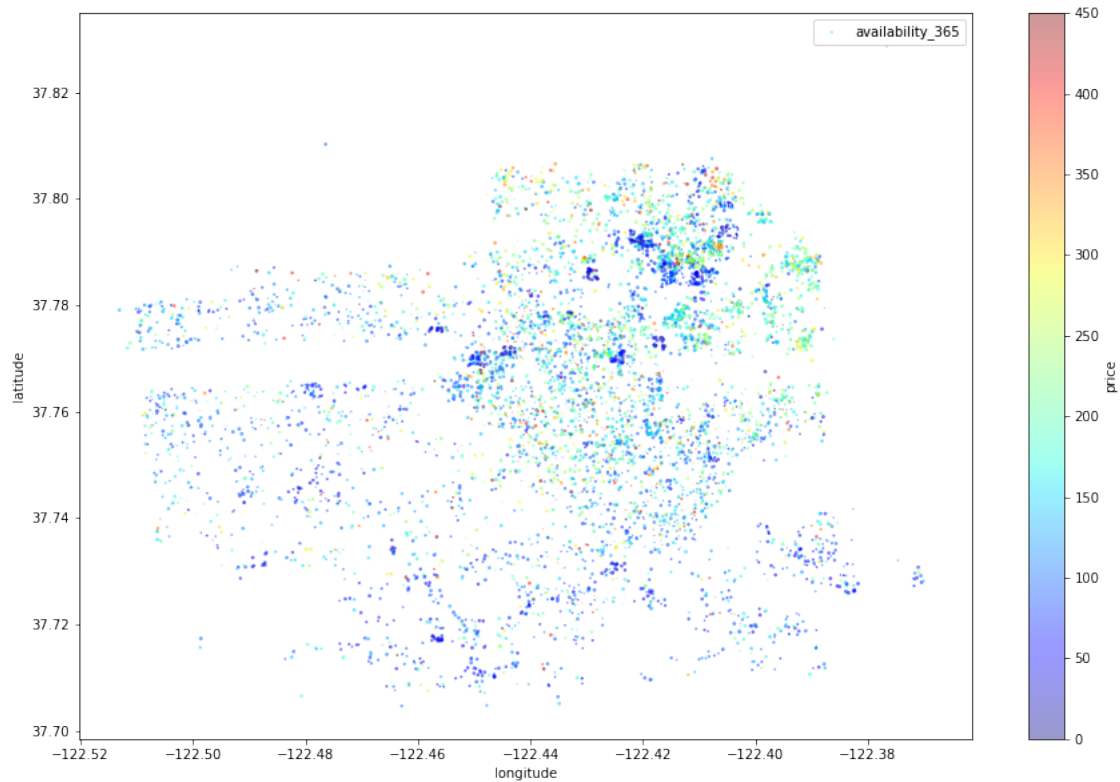
return verifications_dict

```

```

[22]: # Initial map of San Francisco prices by all year availability
df_listings.plot(kind='scatter', x='longitude', y='latitude', alpha=0.4,
→figsize=(15,10),
        s=(df_listings["availability_365"])/100,
→label="availability_365",
        c="price", cmap=plt.get_cmap('jet'),
        colorbar=True,
        sharex=False)
plt.show()

```



[23]: *#Load folium and add map*

```
import folium
folium.Map()
```

[23]: <folium.folium.Map at 0x7f88836c7c50>

```
airbnb = df_listings.iloc[0:8111, :]
map_it = folium.Map(location=[37.77, -122.42], zoom_start = 12)
locations = folium.map.FeatureGroup()
for lat, lng in zip(airbnb.latitude, airbnb.longitude):
    locations.add_child(
        folium.CircleMarker(
            [lat, lng],
            radius=0.2,
            color='black',
            fill=False))
map_it.add_child(locations)
```

[24]: <folium.folium.Map at 0x7f8884268650>

[25]: `airbnb_df = pd.read_csv('listings.csv', parse_dates = ['last_review'])`


```
[26]: nbh_geo_df = gpd.read_file('SanFrancisco.Neighborhoods.json', driver='GeoJSON')
```

```
[27]: airbnb_df = airbnb_df[['id', 'room_type', 'neighbourhood',  
                             'latitude', 'longitude', 'price']]  
nbh_geo_df = nbh_geo_df[['neighborhood', 'geometry']]
```

```
[28]: airbnb_df.head(2)
```

```
[28]:
```

	id	room_type	neighbourhood	latitude	longitude	price
0	958	Entire home/apt	Duboce Triangle	37.76931	-122.43386	\$170.00
1	3850	Private room	Inner Sunset	37.75402	-122.45805	\$99.00

```
[29]: nbh_geo_df.head(2)
```

```
[29]:
```

	neighborhood	geometry
0	Rincon Hill	GEOMETRYCOLLECTION (POLYGON ((-122.39170 37.79...
1	South Beach	GEOMETRYCOLLECTION (POLYGON ((-122.38777 37.78...

```
[30]: from shapely.geometry import Point, shape  
  
locs_geometry = [Point(xy) for xy in zip(airbnb_df.longitude,  
                                           airbnb_df.latitude)]  
  
crs = {'init': 'epsg:4326'}  
# Coordinate Reference Systems, "epsg:4326" is a common projection of WGS84  
# ↪ Latitude/Longitude  
locs_gdf = gpd.GeoDataFrame(airbnb_df, crs=crs, geometry=locs_geometry)
```

```
/opt/anaconda3/lib/python3.7/site-packages/pyproj/crs/crs.py:53: FutureWarning:  
'+init=<authority>:<code>' syntax is deprecated. '<authority>:<code>' is the  
preferred initialization method. When making the change, be mindful of axis  
order changes: https://pyproj4.github.io/pyproj/stable/gotchas.html#axis-order-  
changes-in-proj-6  
    return _prepare_from_string(" ".join(pjargs))
```

```
[31]: import folium  
locs_map = folium.Map(location=[37.77, -122.42],  
                        zoom_start=12, tiles='cartodbpositron')  
  
feature_ea = folium.FeatureGroup(name='Entire home/apt')  
feature_pr = folium.FeatureGroup(name='Private room')  
feature_sr = folium.FeatureGroup(name='Shared room')  
  
for i, v in locs_gdf.iterrows():  
    popup = """  
    Location id : <b>%s</b><br>  
    Room type : <b>%s</b><br>  
    Neighborhood : <b>%s</b><br>  
    Price : <b>%s</b><br>
```

```

""" % (v['id'], v['room_type'], v['neighbourhood'], v['price'])

if v['room_type'] == 'Entire home/apt':
    folium.CircleMarker(location=[v['latitude'], v['longitude']],
                        radius=1,
                        tooltip=popup,
                        color='#FFBA00',
                        fill_color='#FFBA00',
                        fill=True).add_to(feature_ea)
elif v['room_type'] == 'Private room':
    folium.CircleMarker(location=[v['latitude'], v['longitude']],
                        radius=1,
                        tooltip=popup,
                        color='#087FBF',
                        fill_color='#087FBF',
                        fill=True).add_to(feature_pr)
elif v['room_type'] == 'Shared room':
    folium.CircleMarker(location=[v['latitude'], v['longitude']],
                        radius=1,
                        tooltip=popup,
                        color='#FF0700',
                        fill_color='#FF0700',
                        fill=True).add_to(feature_sr)

feature_ea.add_to(locs_map)
feature_pr.add_to(locs_map)
feature_sr.add_to(locs_map)
folium.LayerControl(collapsed=False).add_to(locs_map)

```

[31]: <folium.map.LayerControl at 0x7f8885057990>

[32]: locs_map

[32]: <folium.folium.Map at 0x7f887fa85dd0>

```

[33]: from folium import plugins

cluster_map = folium.Map(location=[37.77, -122.42],
                        zoom_start=13, tiles='cartodbpositron')

marker_cluster = plugins.MarkerCluster().add_to(cluster_map)
for i, v in locs_gdf.iterrows():
    popup = """
    Location id : <b>%s</b><br>
    Room type : <b>%s</b><br>
    Neighbourhood : <b>%s</b><br>
    Price : <b>%s</b><br>
    """

```

```

""" % (v['id'], v['room_type'], v['neighbourhood'], v['price'])

if v['room_type'] == 'Entire home/apt':
    folium.CircleMarker(location=[v['latitude'], v['longitude']],
                        radius=3,
                        tooltip=popup,
                        color='#FFBA00',
                        fill_color='#FFBA00',
                        fill=True).add_to(marker_cluster)
elif v['room_type'] == 'Private room':
    folium.CircleMarker(location=[v['latitude'], v['longitude']],
                        radius=3,
                        tooltip=popup,
                        color='#087FBF',
                        fill_color='#087FBF',
                        fill=True).add_to(marker_cluster)
elif v['room_type'] == 'Shared room':
    folium.CircleMarker(location=[v['latitude'], v['longitude']],
                        radius=3,
                        tooltip=popup,
                        color='#FF0700',
                        fill_color='#FF0700',
                        fill=True).add_to(marker_cluster)

```

```
[34]: cluster_map
```

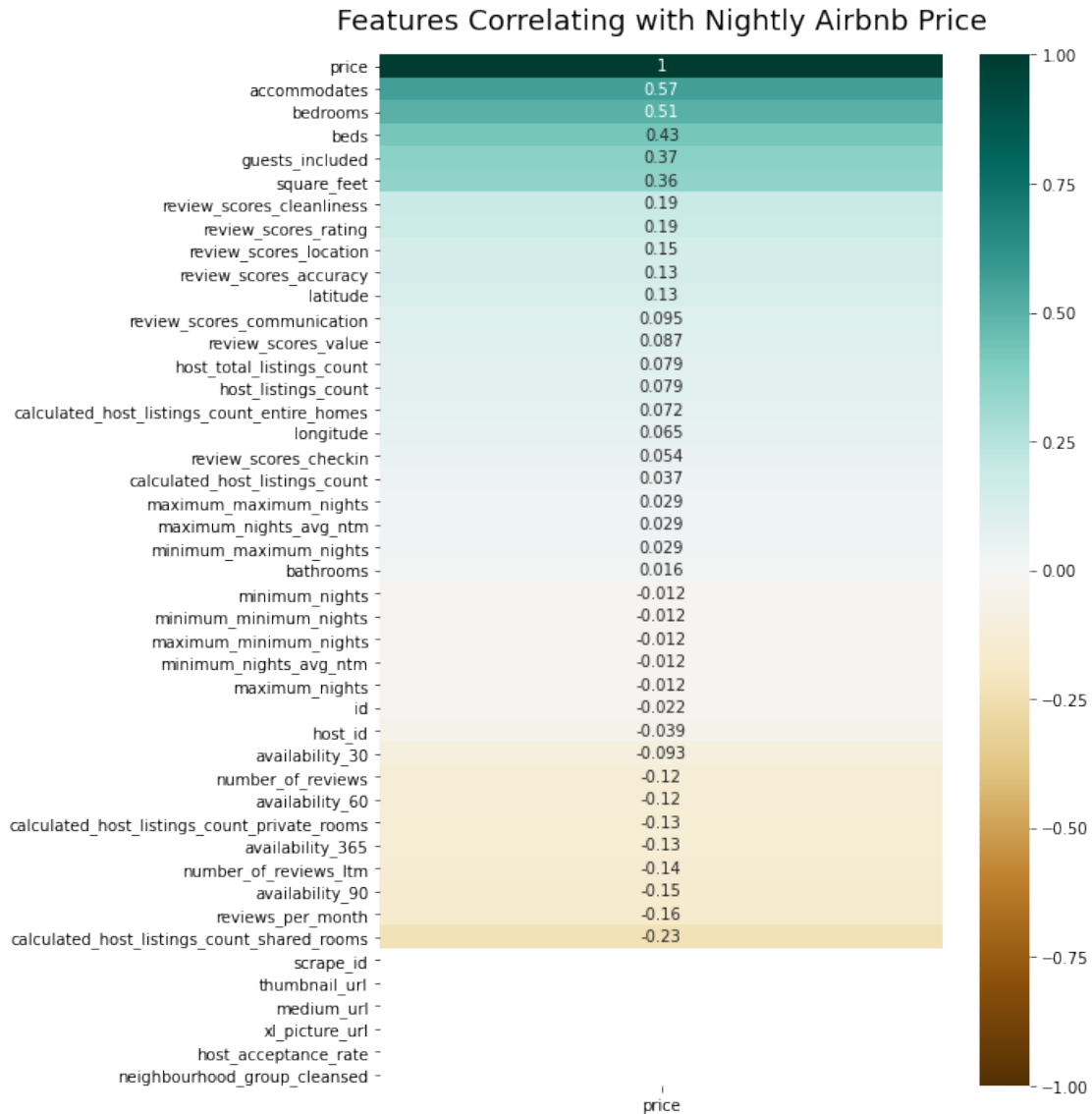
```
[34]: <folium.folium.Map at 0x7f888b27ba50>
```

```

[35]: plt.figure(figsize=(8, 12))
heatmap = sns.heatmap(df_listings.corr()[['price']].sort_values
                      (by='price', ascending=False),
                      vmin=-1, vmax=1, annot=True, cmap='BrBG')

heatmap.set_title('Features Correlating with Nightly Airbnb Price',
                  fontdict={'fontsize':18}, pad=16);

```



Want to also look at `host_acceptance_rate` and `neighbourhood_group_cleansed`

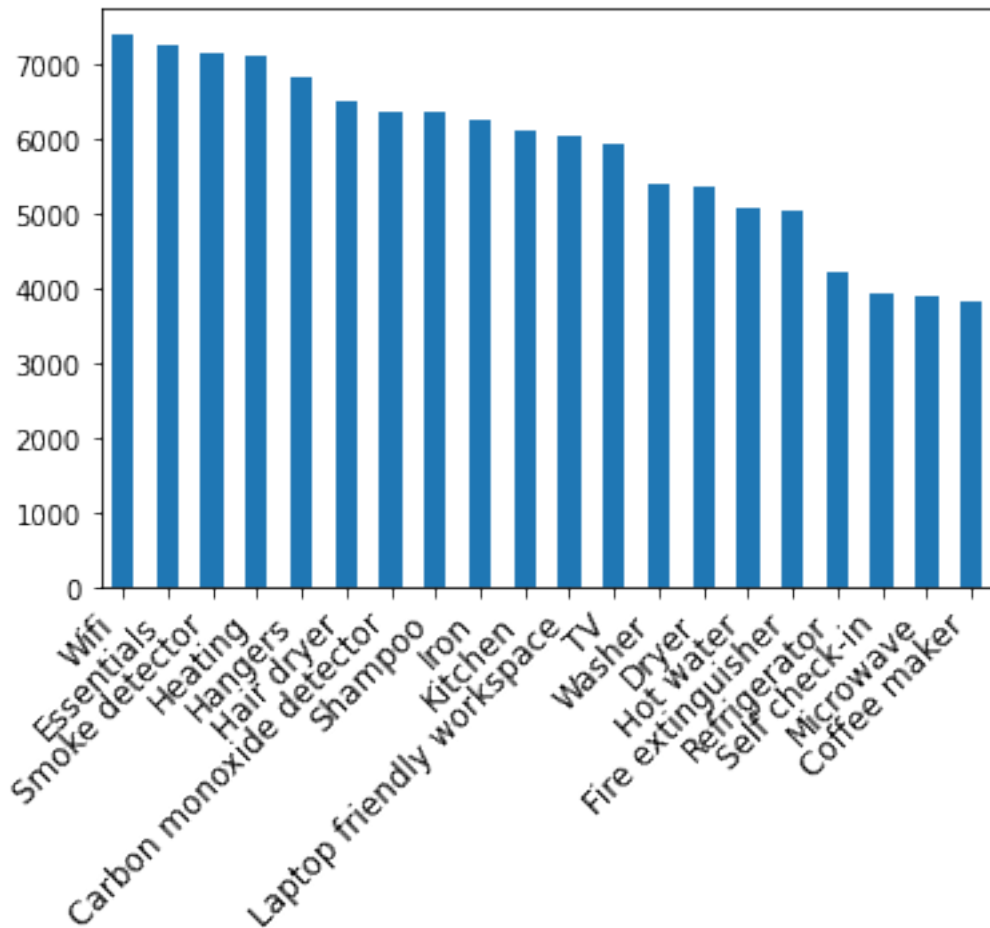
```
[36]: # Cleaning the amenities
df_listings.amenities = df_listings.amenities.str.replace("[{}]", "").str.
      ↪replace("'", "")
df_listings['amenities'].head()
```

/opt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2:

FutureWarning: The default value of regex will change from True to False in a future version.

```
[36]: 0    TV,Cable TV,Internet,Wifi,Kitchen,Pets live on...
      1    Internet,Wifi,Kitchen,Breakfast,Free street pa...
      2    Internet,Wifi,Kitchen,Heating,Family/kid frien...
      3    TV,Internet,Wifi,Kitchen,Free street parking,H...
      4    TV,Internet,Wifi,Kitchen,Free street parking,H...
      Name: amenities, dtype: object
```

```
[37]: pd.Series(np.concatenate(df_listings['amenities'].map(lambda amns: amns.
    ↪split(", "))))\
      .value_counts().head(20)\
      .plot(kind='bar')
ax = plt.gca()
ax.set_xticklabels(ax.get_xticklabels(), rotation=45, ha='right', fontsize=12)
plt.show();
```

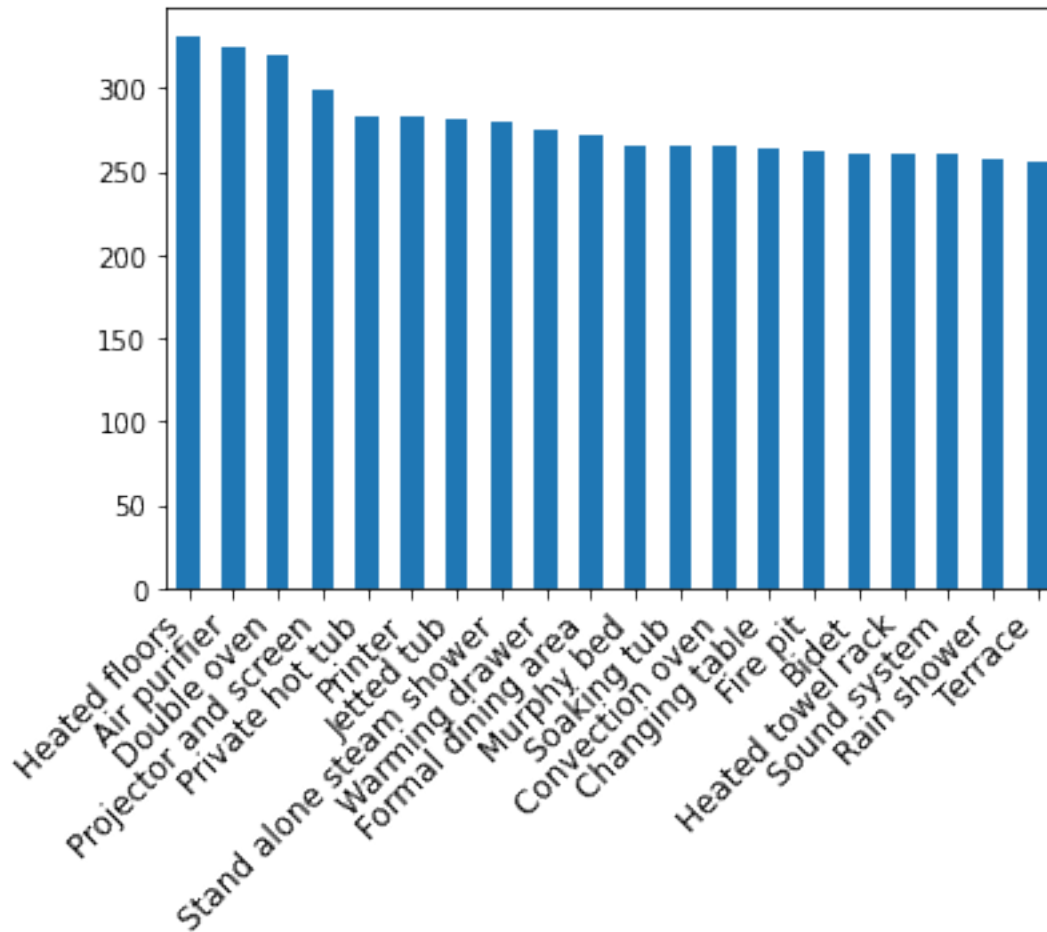


```
[38]: amenities = np.unique(np.concatenate(df_listings['amenities'].map(lambda amns:
    ↪amns.split(", "))))
```

```

amenity_prices = [(amn, df_listings[df_listings['amenities'].map(lambda amns:
    ↪ amn in amns)]['price'].mean()) for amn in amenities if amn != ""]
amenity_srs = pd.Series(data=[a[1] for a in amenity_prices], index=[a[0] for a
    ↪ in amenity_prices])
amenity_srs.sort_values(ascending=False)[:20].plot(kind='bar')
ax = plt.gca()
ax.set_xticklabels(ax.get_xticklabels(), rotation=45, ha='right', fontsize=12)
plt.show();

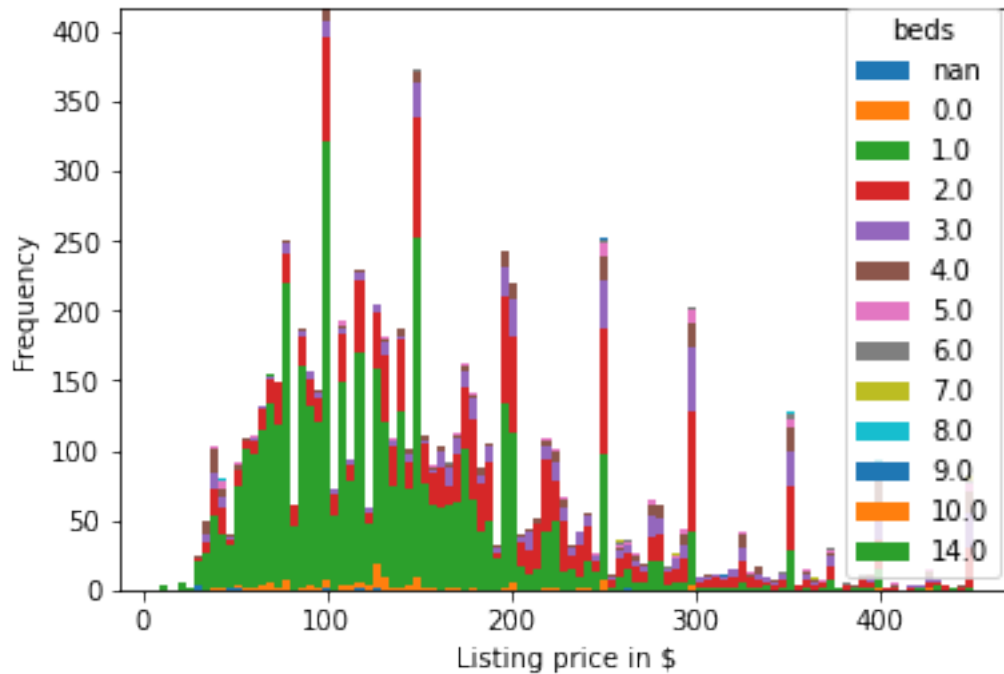
```



```

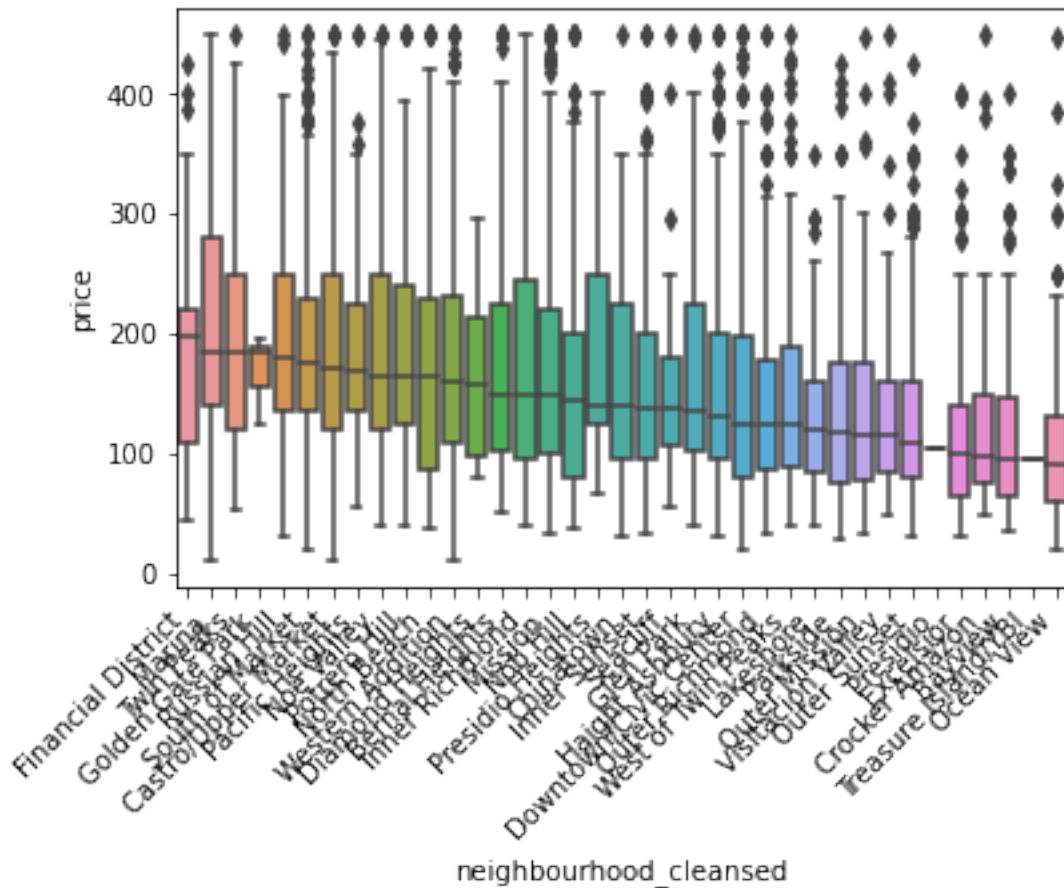
[42]: # bed amount vs price
df_listings.loc[(df_listings.price <= 600) & (df_listings.price > 0)].
    ↪ pivot(columns = 'beds', values = 'price').plot.hist(stacked = True, bins=100)
plt.xlabel('Listing price in $');

```



```
[43]: # neighborhood vs price

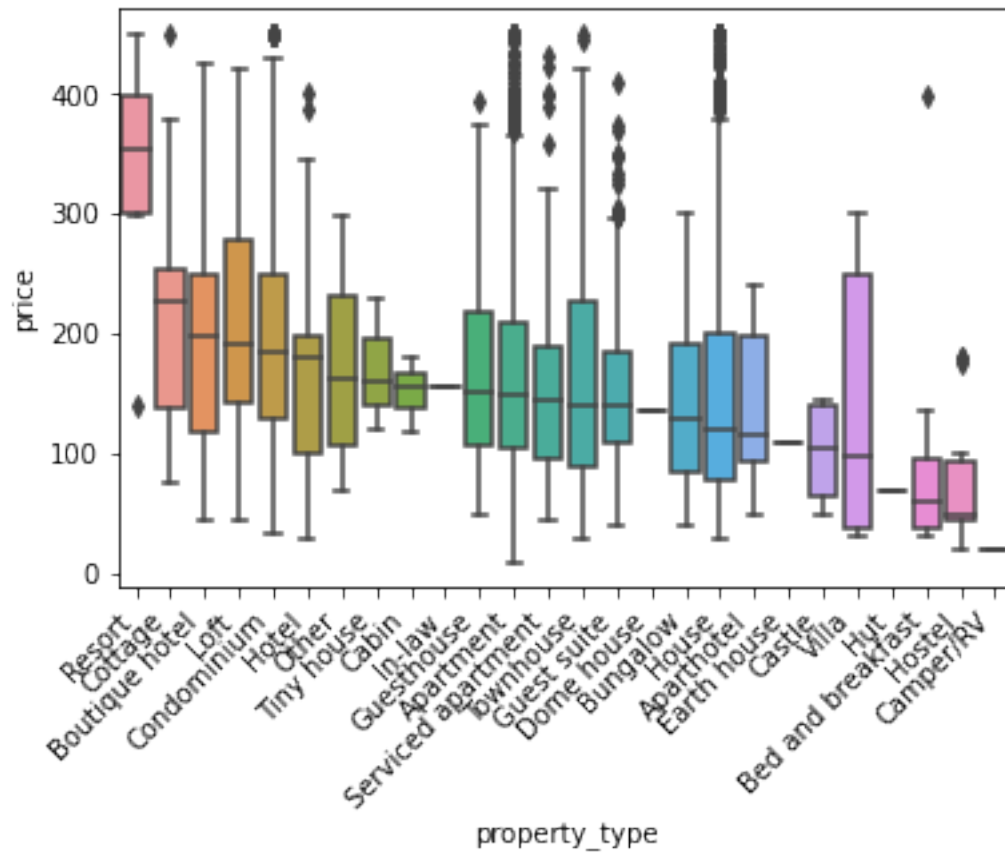
sort_price = df_listings.loc[(df_listings.price <= 600) & (df_listings.price > 0)]\
    .groupby('neighbourhood_cleansed')['price']\
    .median()\
    .sort_values(ascending=False)\
    .index
sns.boxplot(y='price', x='neighbourhood_cleansed', data=df_listings.\
    loc[(df_listings.price <= 600) & (df_listings.price > 0)],\
    order=sort_price)
ax = plt.gca()
ax.set_xticklabels(ax.get_xticklabels(), rotation=45, ha='right')
plt.show();
```



```
[44]: # property type vs price

sort_price = df_listings.loc[(df_listings.price <= 600) & (df_listings.price > 0)]\
    .groupby('property_type')['price']\
    .median()\
    .sort_values(ascending=False)\
    .index

sns.boxplot(y='price', x='property_type', data=df_listings.loc[(df_listings.\
    price <= 600) & (df_listings.price > 0)], order=sort_price)
ax = plt.gca()
ax.set_xticklabels(ax.get_xticklabels(), rotation=45, ha='right')
plt.show();
```

[]: