

Overview

Dataset statistics	
Number of variables	18
Number of observations	47926
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	23.9 MiB
Average record size in memory	522.4 B

Variable types	
Numeric	11
Categorical	6
DateTime	1

Alerts

name has a high cardinality: 46963 distinct values	High cardinality
host_name has a high cardinality: 11425 distinct values	High cardinality
neighbourhood has a high cardinality: 221 distinct values	High cardinality
df_index is highly correlated with id and <u>1 other fields (id, host_id)</u>	High correlation
id is highly correlated with df_index and <u>1 other fields (df_index, host_id)</u>	High correlation
host_id is highly correlated with df_index and <u>1 other fields (df_index, id)</u>	High correlation
number_of_reviews is highly correlated with reviews_per_month	High correlation
reviews_per_month is highly correlated with number_of_reviews	High correlation
df_index is highly correlated with id and <u>1 other fields (id, host_id)</u>	High correlation
id is highly correlated with df_index and <u>1 other fields (df_index, host_id)</u>	High correlation
host_id is highly correlated with df_index and <u>1 other fields (df_index, id)</u>	High correlation
number_of_reviews is highly correlated with reviews_per_month	High correlation
reviews_per_month is highly correlated with number_of_reviews	High correlation
df_index is highly correlated with id	High correlation
id is highly correlated with df_index	High correlation
number_of_reviews is highly correlated with reviews_per_month	High correlation
reviews_per_month is highly correlated with number_of_reviews	High correlation
df_index is highly correlated with id and <u>1 other fields (id, host_id)</u>	High correlation
id is highly correlated with df_index and <u>1 other fields (df_index, host_id)</u>	High correlation
host_id is highly correlated with df_index and <u>1 other fields (df_index, id)</u>	High correlation
neighbourhood_group is highly correlated with latitude and <u>1 other fields (latitude, longitude)</u>	High correlation
latitude is highly correlated with neighbourhood_group and <u>1 other fields (neighbourhood_group, longitude)</u>	High correlation
longitude is highly correlated with neighbourhood_group and <u>1 other fields (neighbourhood_group, latitude)</u>	High correlation
name is uniformly distributed	Uniform
df_index has unique values	Unique
id has unique values	Unique
number_of_reviews has 9510 (19.8%) zeros	Zeros
reviews_per_month has 9510 (19.8%) zeros	Zeros
availability_365 has 17464 (36.4%) zeros	Zeros

Reproduction

Analysis started	2021-10-05 17:38:24.258338
Analysis finished	2021-10-05 17:38:57.753589
Duration	33.5 seconds
Software version	pandas-profiling v3.1.0 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Pandas%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%22

Variables

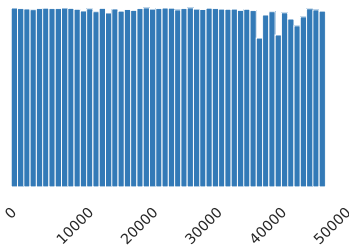
df_index

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: id, host_id)
HIGH CORRELATION (This variable has a high correlation with 2 fields: id, host_id)
HIGH CORRELATION (This variable has a high correlation with 1 fields: id)
HIGH CORRELATION (This variable has a high correlation with 2 fields: id, host_id)
UNIQUE

Distinct	47926
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	24220.06001
Minimum	0
Maximum	48857
Zeros	1
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB

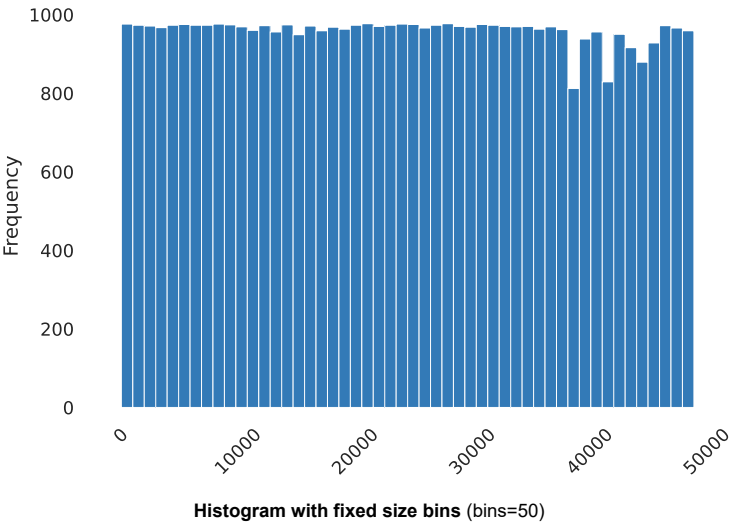


Quantile statistics

Minimum	0
5-th percentile	2406.25
Q1	12051.25
median	24149.5
Q3	36208.75
95-th percentile	46428.75
Maximum	48857
Range	48857
Interquartile range (IQR)	24157.5

Descriptive statistics

Standard deviation	14061.80844
Coefficient of variation (CV)	0.5805852021
Kurtosis	-1.185390867
Mean	24220.06001
Median Absolute Deviation (MAD)	12079
Skewness	0.01948179953
Sum	1160770596
Variance	197734456.5
Monotonicity	Strictly increasing



Value	Count	Frequency (%)
2047	1	< 0.1%
40249	1	< 0.1%
19779	1	< 0.1%
17730	1	< 0.1%
23873	1	< 0.1%
21824	1	< 0.1%
42302	1	< 0.1%
48445	1	< 0.1%
46396	1	< 0.1%
36155	1	< 0.1%
Other values (47916)	47916	> 99.9%

Value	Count	Frequency (%)
0	1	< 0.1%
1	1	< 0.1%
2	1	< 0.1%
3	1	< 0.1%
4	1	< 0.1%
5	1	< 0.1%
6	1	< 0.1%
7	1	< 0.1%
8	1	< 0.1%
9	1	< 0.1%

Value	Count	Frequency (%)
48857	1	< 0.1%
48856	1	< 0.1%
48855	1	< 0.1%
48854	1	< 0.1%
48853	1	< 0.1%
48852	1	< 0.1%
48851	1	< 0.1%
48850	1	< 0.1%
48849	1	< 0.1%
48848	1	< 0.1%

id

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: df.index, host.id)

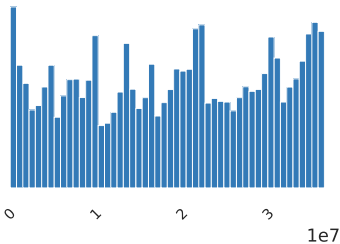
HIGH CORRELATION (This variable has a high correlation with 2 fields: df.index, host.id)

HIGH CORRELATION (This variable has a high correlation with 1 fields: df.index)

HIGH CORRELATION (This variable has a high correlation with 2 fields: df.index, host.id)

UNIQUE

Distinct	47926
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	18861822.12
Minimum	2539
Maximum	36487245
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



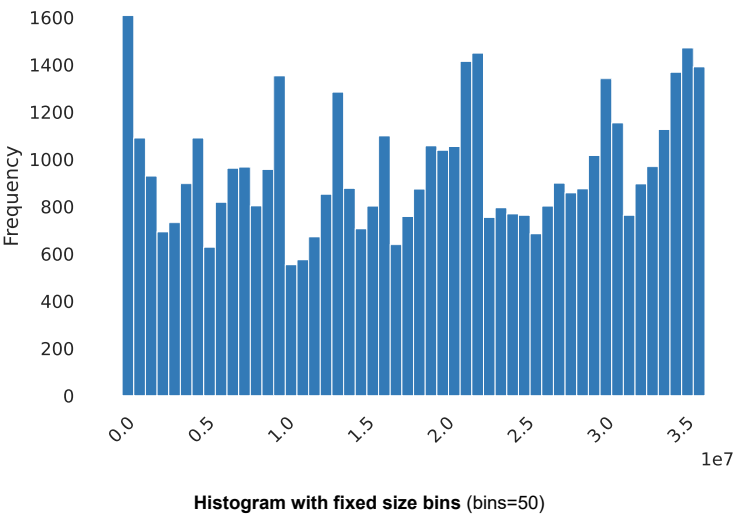
Quantile statistics

Minimum	2539
5-th percentile	1185399
Q1	9386073.25
median	19475094
Q3	28834379.75
95-th percentile	35263722.75
Maximum	36487245
Range	36484706
Interquartile range (IQR)	19448306.5

Descriptive statistics

Standard deviation	10951890.59
Coefficient of variation (CV)	0.5806379956
Kurtosis	-1.218210411
Mean	18861822.12

Median Absolute Deviation (MAD)	9777786.5
Skewness	-0.07293591539
Sum	$9.039716868 \times 10^{11}$
Variance	$1.199439075 \times 10^{14}$
Monotonicity	Strictly increasing



Value	Count	Frequency (%)
11667455	1	< 0.1%
34157056	1	< 0.1%
7851219	1	< 0.1%
24827469	1	< 0.1%
34065052	1	< 0.1%
11222221	1	< 0.1%
15542745	1	< 0.1%
27417822	1	< 0.1%
20055242	1	< 0.1%
11801800	1	< 0.1%
Other values (47916)	47916	> 99.9%

Value	Count	Frequency (%)
2539	1	< 0.1%
2595	1	< 0.1%
3647	1	< 0.1%
3831	1	< 0.1%
5022	1	< 0.1%
5099	1	< 0.1%
5121	1	< 0.1%
5178	1	< 0.1%
5203	1	< 0.1%
5238	1	< 0.1%

Value	Count	Frequency (%)
36487245	1	< 0.1%
36485609	1	< 0.1%
36485431	1	< 0.1%
36485057	1	< 0.1%
36484665	1	< 0.1%
36484363	1	< 0.1%
36484087	1	< 0.1%
36483152	1	< 0.1%
36483010	1	< 0.1%
36482809	1	< 0.1%

name		
Categorical		
		HIGH
CARDINALITY		
UNIFORM		
Distinct	46963	
Distinct (%)	98.0%	
Missing	0	
Missing (%)	0.0%	
Memory size	4.4 MiB	

Length

Max length	179
Median length	36
Mean length	36.69745024
Min length	1

Characters and Unicode

Total characters	182	
Distinct characters	1	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	46326	?
Unique (%)	96.7%	

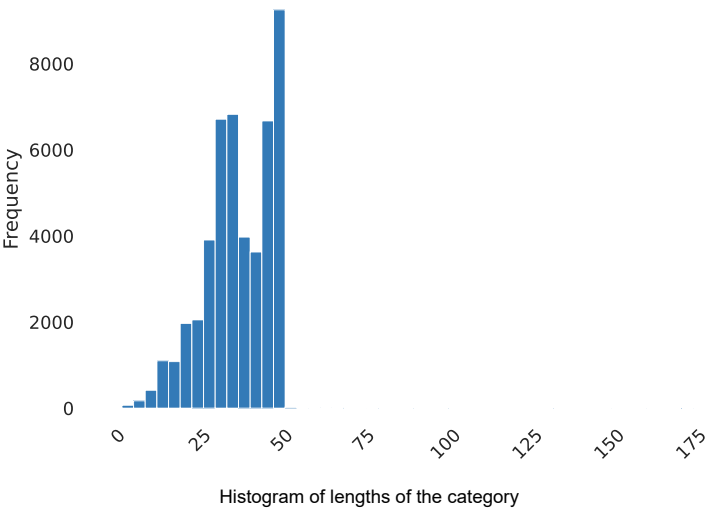
Sample

1st row	Clean & quiet apt home by the park
2nd row	Skylit Midtown Castle
3rd row	THE VILLAGE OF HARLEM....NEW YORK !
4th row	Cozy Entire Floor of Brownstone
5th row	Entire Apt: Spacious Studio/Loft by central park

Common Values

Value	Count	Frequency (%)
Hillside Hotel	18	< 0.1%
Home away from home	17	< 0.1%
New york Multi-unit building	16	< 0.1%
Brooklyn Apartment	12	< 0.1%
Loft Suite @ The Box House Hotel	11	< 0.1%
Private Room	11	< 0.1%
Private room	10	< 0.1%
Artsy Private BR in Fort Greene Cumberland	10	< 0.1%
Private room in Williamsburg	8	< 0.1%
Cozy Brooklyn Apartment	8	< 0.1%
Other values (46953)	47805	99.7%

Length



Value	Count	Frequency (%)
in	16604	5.7%
room	9944	3.4%
bedroom	7531	2.6%
	7279	2.5%
private	7132	2.5%
apartment	6648	2.3%
cozy	4946	1.7%
apt	4533	1.6%
brooklyn	4037	1.4%
studio	3814	1.3%
Other values (12451)	217980	75.0%

Most occurring characters

Value	Count	Frequency (%)
	182	100.0%

Most occurring categories

Value	Count	Frequency (%)
Control	182	100.0%

Most frequent character per category

Control

Value	Count	Frequency (%)
	182	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	182	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
	182	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	182	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	182	100.0%

host_id

Real number ($\mathbb{R}_{\geq 0}$)

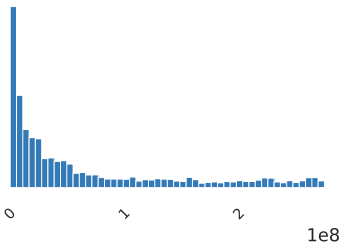
HIGH

CORRELATION (This variable has a high correlation with 2 fields: df.index, id)

HIGH CORRELATION (This variable has a high correlation with 2 fields: df.index, id)

HIGH CORRELATION (This variable has a high correlation with 2 fields: df.index, id)

Distinct	37325
Distinct (%)	77.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	66366719.84
Minimum	2438
Maximum	274321313
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



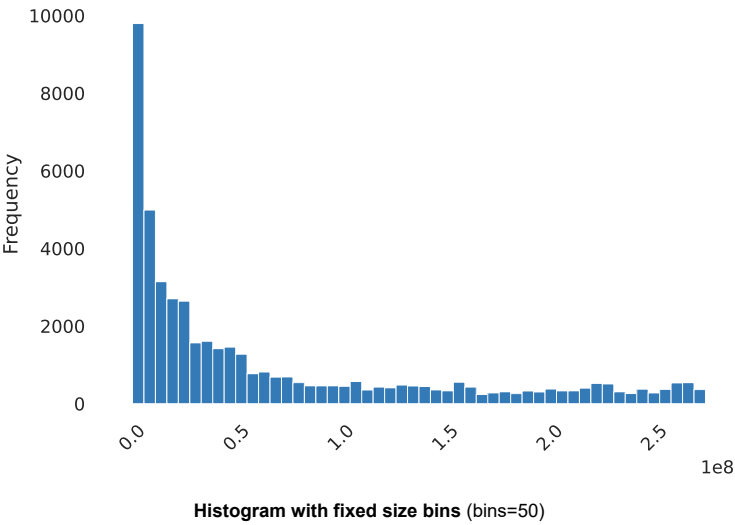
Quantile statistics

Minimum	2438
5-th percentile	797420
Q1	7598992.25
median	29863220.5
Q3	104497453
95-th percentile	242478115.2
Maximum	274321313
Range	274318875
Interquartile range (IQR)	96898460.75

Descriptive statistics

Standard deviation	78126572.85
Coefficient of variation (CV)	1.177195031
Kurtosis	0.2838830002
Mean	66366719.84
Median Absolute Deviation (MAD)	26616062

Skewness	1.246626958
Sum	$3.180691415 \times 10^{12}$
Variance	$6.103761385 \times 10^{15}$
Monotonicity	Not monotonic



Value	Count	Frequency (%)
12243051	96	0.2%
16098958	96	0.2%
61391963	91	0.2%
22541573	87	0.2%
200380610	65	0.1%
7503643	52	0.1%
1475015	52	0.1%
120762452	50	0.1%
2856748	49	0.1%
205031545	49	0.1%
Other values (37315)	47239	98.6%

Value	Count	Frequency (%)
2438	1	< 0.1%
2571	1	< 0.1%
2787	6	< 0.1%
2845	2	< 0.1%
2868	1	< 0.1%
2881	2	< 0.1%
3151	1	< 0.1%
3211	1	< 0.1%
3415	1	< 0.1%
3563	1	< 0.1%

Value	Count	Frequency (%)
274321313	1	< 0.1%
274311461	1	< 0.1%
274307600	1	< 0.1%
274298453	1	< 0.1%
274273284	1	< 0.1%
274225617	1	< 0.1%
274195458	1	< 0.1%
274188386	1	< 0.1%
274103383	1	< 0.1%
274079964	1	< 0.1%

host_name
Categorical

HIGH

CARDINALITY

Distinct	11425
Distinct (%)	23.8%
Missing	0
Missing (%)	0.0%
Memory size	2.9 MiB

Length

Max length	35
Median length	6
Mean length	6.073592622
Min length	1

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	6883	?
Unique (%)	14.4%	

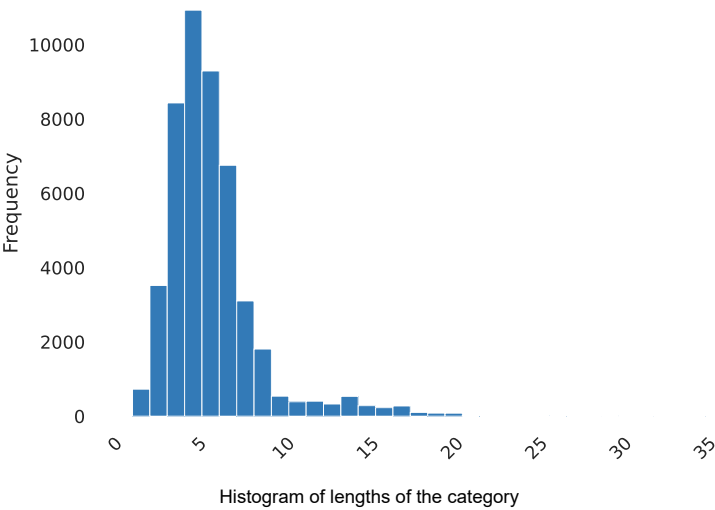
Sample

1st row	John
2nd row	Jennifer
3rd row	Elisabeth
4th row	LisaRoxanne
5th row	Laura

Common Values

Value	Count	Frequency (%)
Michael	417	0.9%
David	402	0.8%
John	293	0.6%
Alex	278	0.6%
Daniel	226	0.5%
Sarah	226	0.5%
Maria	204	0.4%
Jessica	202	0.4%
Mike	193	0.4%
Andrew	190	0.4%
Other values (11415)	45295	94.5%

Length



Value	Count	Frequency (%)
	1120	2.1%
and	622	1.2%
michael	460	0.9%
david	448	0.8%
john	336	0.6%
alex	329	0.6%
laura	292	0.5%
maria	244	0.5%
daniel	242	0.5%
sarah	239	0.4%
Other values (10237)	48970	91.9%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

neighbourhood_group
Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: latitude, longitude)

Distinct	5
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.0 MiB

Length

Max length	13
Median length	8
Mean length	8.173246255
Min length	5

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

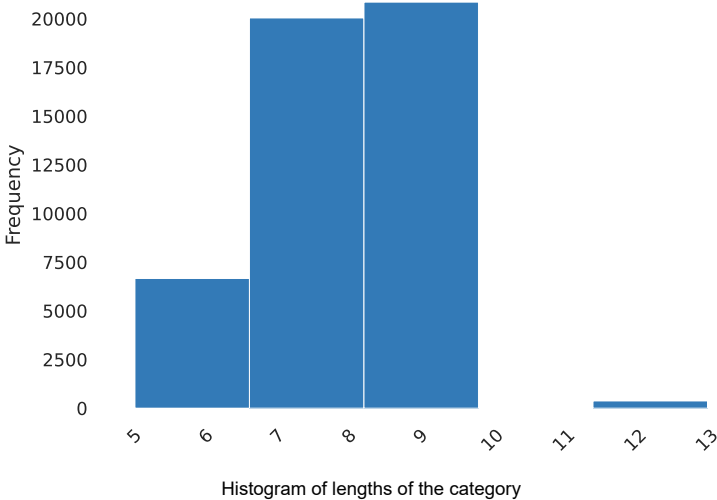
Sample

1st row	Brooklyn
2nd row	Manhattan
3rd row	Manhattan
4th row	Brooklyn
5th row	Manhattan

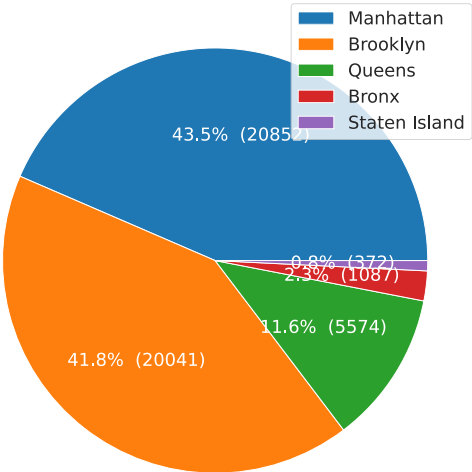
Common Values

Value	Count	Frequency (%)
Manhattan	20852	43.5%
Brooklyn	20041	41.8%
Queens	5574	11.6%
Bronx	1087	2.3%
Staten Island	372	0.8%

Length



Pie chart



Value	Count	Frequency (%)
manhattan	20852	43.2%
brooklyn	20041	41.5%
queens	5574	11.5%
bronx	1087	2.3%
island	372	0.8%
staten	372	0.8%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

neighbourhood
Categorical

HIGH

CARDINALITY

Distinct	221
Distinct (%)	0.5%
Missing	0
Missing (%)	0.0%
Memory size	3.1 MiB

Length

Max length	26
Median length	12
Mean length	11.87564161
Min length	4

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	6	?
Unique (%)	< 0.1%	

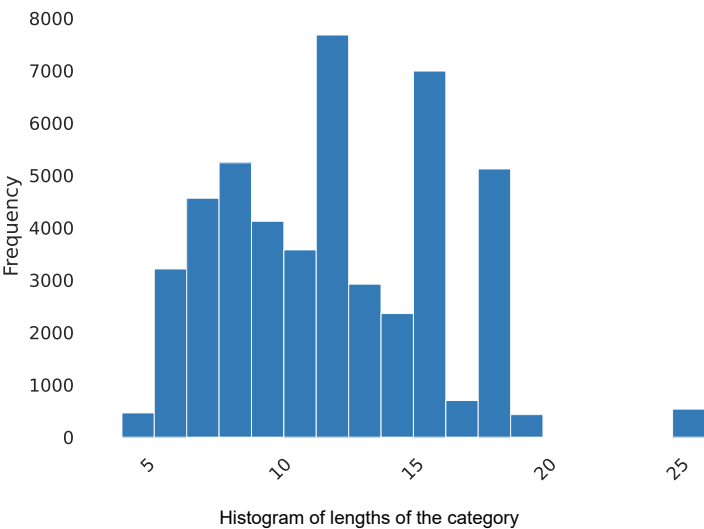
Sample

1st row	Kensington
2nd row	Midtown
3rd row	Harlem
4th row	Clinton Hill
5th row	East Harlem

Common Values

Value	Count	Frequency (%)
Williamsburg	3910	8.2%
Bedford-Stuyvesant	3704	7.7%
Harlem	2635	5.5%
Bushwick	2451	5.1%
Upper West Side	1934	4.0%
Hell's Kitchen	1883	3.9%
East Village	1832	3.8%
Upper East Side	1766	3.7%
Crown Heights	1560	3.3%
Midtown	1489	3.1%
Other values (211)	24762	51.7%

Length



Value	Count	Frequency (%)
east	6519	8.4%
side	4606	6.0%
williamsburg	3910	5.1%
harlem	3743	4.8%
bedford-stuyvesant	3704	4.8%
upper	3700	4.8%
heights	3575	4.6%
village	3117	4.0%
west	2704	3.5%
bushwick	2451	3.2%
Other values (233)	39373	50.9%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

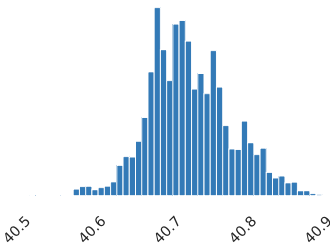
latitude

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: neighbourhood group, longitude)

Distinct	18984
Distinct (%)	39.6%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	40.72879618
Minimum	40.49979
Maximum	40.91306
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



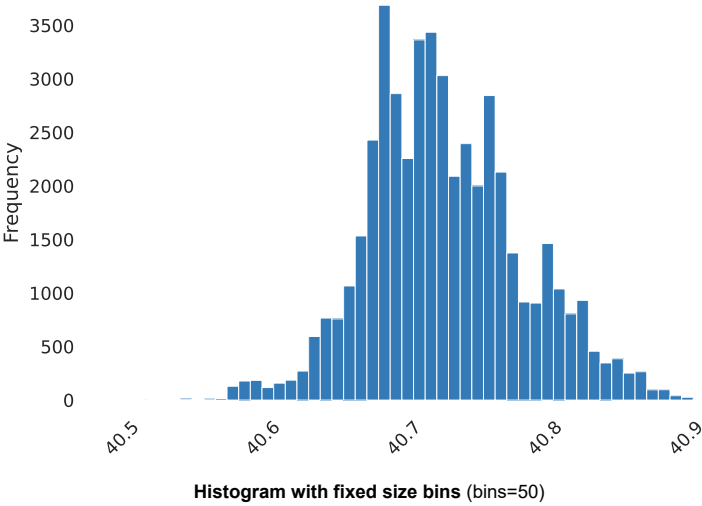
Quantile statistics

Minimum	40.49979
5-th percentile	40.645715
Q1	40.68961
median	40.72278
Q3	40.7633075
95-th percentile	40.826025
Maximum	40.91306
Range	0.41327
Interquartile range (IQR)	0.0736975

Descriptive statistics

Standard deviation	0.05487824725
Coefficient of variation (CV)	0.001347406562
Kurtosis	0.122347916
Mean	40.72879618
Median Absolute Deviation (MAD)	0.03655
Skewness	0.2423492754

Sum	1951968.286
Variance	0.003011622021
Monotonicity	Not monotonic



Value	Count	Frequency (%)
40.71813	18	< 0.1%
40.68444	13	< 0.1%
40.68634	13	< 0.1%
40.69414	13	< 0.1%
40.71171	12	< 0.1%
40.68537	12	< 0.1%
40.76189	12	< 0.1%
40.71353	12	< 0.1%
40.69054	11	< 0.1%
40.69454	11	< 0.1%
Other values (18974)	47799	99.7%

Value	Count	Frequency (%)
40.49979	1	< 0.1%
40.50641	1	< 0.1%
40.50708	1	< 0.1%
40.50868	1	< 0.1%
40.50873	1	< 0.1%
40.50943	1	< 0.1%
40.51133	1	< 0.1%
40.52211	1	< 0.1%
40.52293	1	< 0.1%
40.527	1	< 0.1%

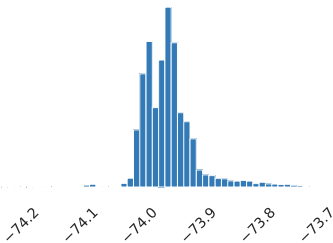
Value	Count	Frequency (%)
40.91306	1	< 0.1%
40.91234	1	< 0.1%
40.91169	1	< 0.1%
40.91167	1	< 0.1%
40.90804	1	< 0.1%
40.90734	1	< 0.1%
40.90527	1	< 0.1%
40.90484	1	< 0.1%
40.90406	1	< 0.1%
40.90391	1	< 0.1%

longitude
Real number (ℝ)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: neighbourhood group, latitude)

Distinct	14578
Distinct (%)	30.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	-73.95160063
Minimum	-74.24442
Maximum	-73.71299
Zeros	0
Zeros (%)	0.0%
Negative	47926
Negative (%)	100.0%
Memory size	374.5 KiB



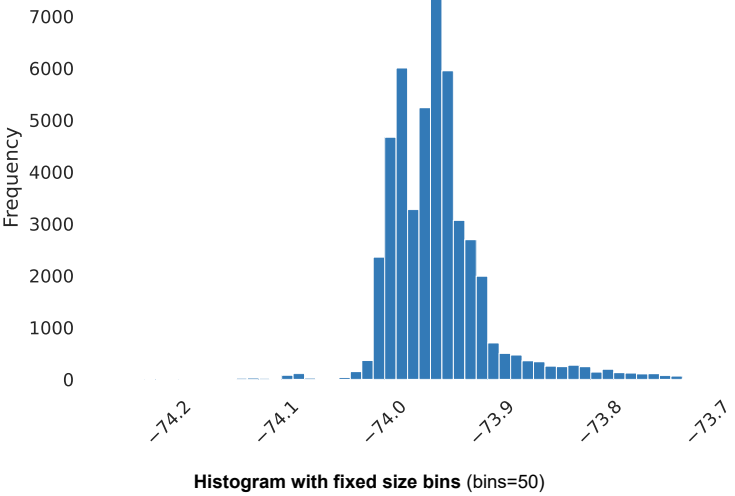
Quantile statistics

Minimum	-74.24442
5-th percentile	-74.0028575
Q1	-73.98242
median	-73.95526
Q3	-73.93583
95-th percentile	-73.8645225
Maximum	-73.71299
Range	0.53143
Interquartile range (IQR)	0.04659

Descriptive statistics

Standard deviation	0.04616488066
Coefficient of variation (CV)	-0.0006242580319
Kurtosis	5.066036523
Mean	-73.95160063
Median Absolute Deviation (MAD)	0.02466
Skewness	1.281648255

Sum	-3544204.412
Variance	0.002131196206
Monotonicity	Not monotonic



Value	Count	Frequency (%)
-73.95677	18	< 0.1%
-73.95427	18	< 0.1%
-73.95405	17	< 0.1%
-73.9506	16	< 0.1%
-73.95136	16	< 0.1%
-73.95332	16	< 0.1%
-73.94791	16	< 0.1%
-73.98439	15	< 0.1%
-73.95725	15	< 0.1%
-73.95669	15	< 0.1%
Other values (14568)	47764	99.7%

Value	Count	Frequency (%)
-74.24442	1	< 0.1%
-74.24285	1	< 0.1%
-74.24084	1	< 0.1%
-74.23986	1	< 0.1%
-74.23914	1	< 0.1%
-74.23803	1	< 0.1%
-74.23059	1	< 0.1%
-74.21238	1	< 0.1%
-74.21017	1	< 0.1%
-74.20941	1	< 0.1%

Value	Count	Frequency (%)
-73.71299	1	< 0.1%
-73.7169	1	< 0.1%
-73.71795	1	< 0.1%
-73.71829	1	< 0.1%
-73.71928	1	< 0.1%
-73.72173	1	< 0.1%
-73.72179	1	< 0.1%
-73.72247	1	< 0.1%
-73.72435	1	< 0.1%
-73.72581	1	< 0.1%

room_type
Categorical

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	3.2 MiB

Length

Max length	15
Median length	15
Mean length	13.5166298
Min length	11

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

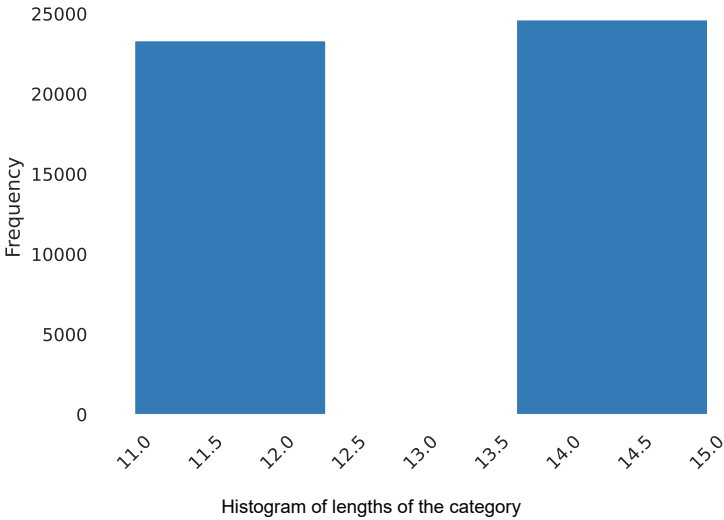
Sample

1st row	Private room
2nd row	Entire home/apt
3rd row	Private room
4th row	Entire home/apt
5th row	Entire home/apt

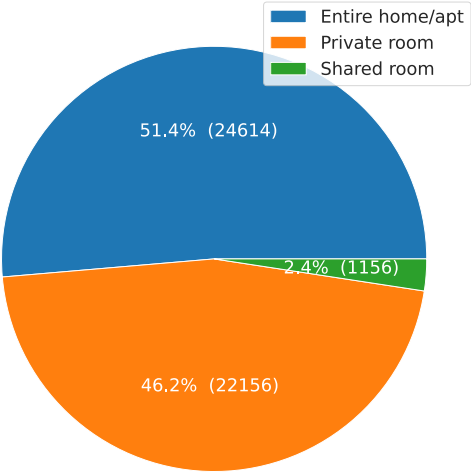
Common Values

Value	Count	Frequency (%)
Entire home/apt	24614	51.4%
Private room	22156	46.2%
Shared room	1156	2.4%

Length



Pie chart



Value	Count	Frequency (%)
home/apt	24614	25.7%
entire	24614	25.7%
room	23312	24.3%
private	22156	23.1%
shared	1156	1.2%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

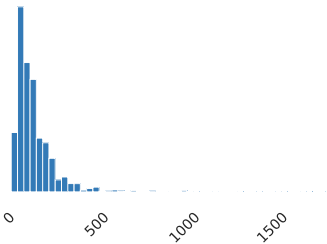
Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

price
Real number ($\mathbb{R}_{\geq 0}$)

Distinct	587
Distinct (%)	1.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	142.7652422
Minimum	10
Maximum	1800
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



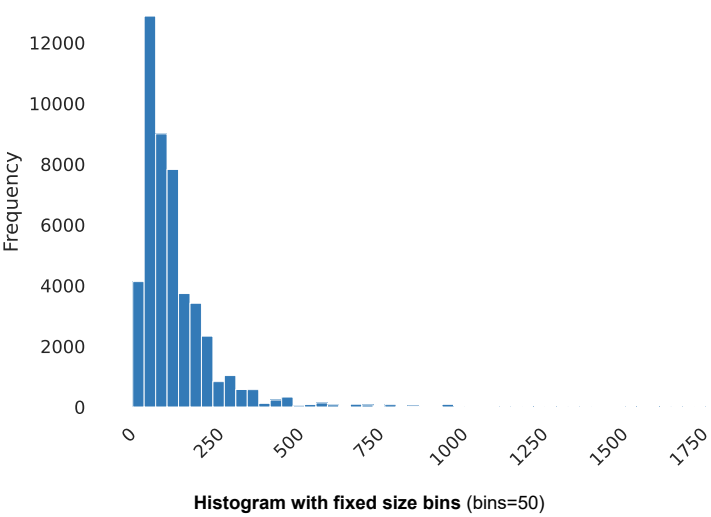
Quantile statistics

Minimum	10
5-th percentile	40
Q1	69
median	105
Q3	175
95-th percentile	350
Maximum	1800
Range	1790
Interquartile range (IQR)	106

Descriptive statistics

Standard deviation	131.4182141
Coefficient of variation (CV)	0.9205196729
Kurtosis	29.28742523
Mean	142.7652422
Median Absolute Deviation (MAD)	45
Skewness	4.224543833
Sum	6842167

Variance	17270.747
Monotonicity	Not monotonic



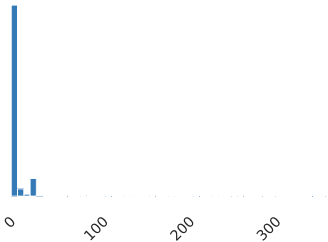
Value	Count	Frequency (%)
150	2043	4.3%
100	2037	4.3%
50	1526	3.2%
60	1455	3.0%
200	1394	2.9%
75	1368	2.9%
80	1272	2.7%
65	1190	2.5%
70	1166	2.4%
120	1127	2.4%
Other values (577)	33348	69.6%

Value	Count	Frequency (%)
10	17	< 0.1%
11	3	< 0.1%
12	4	< 0.1%
13	1	< 0.1%
15	6	< 0.1%
16	6	< 0.1%
18	2	< 0.1%
19	4	< 0.1%
20	33	0.1%
21	6	< 0.1%

Value	Count	Frequency (%)
1800	2	< 0.1%
1799	1	< 0.1%
1795	1	< 0.1%
1763	1	< 0.1%
1750	3	< 0.1%
1749	1	< 0.1%
1731	1	< 0.1%
1700	4	< 0.1%
1680	1	< 0.1%
1600	4	< 0.1%

minimum_nights
Real number ($\mathbb{R}_{\geq 0}$)

Distinct	100
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	6.5220757
Minimum	1
Maximum	365
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



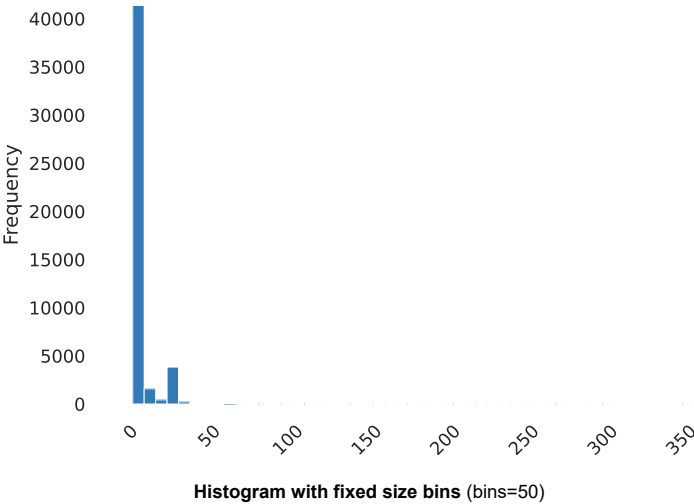
Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	2
Q3	5
95-th percentile	30
Maximum	365
Range	364
Interquartile range (IQR)	4

Descriptive statistics

Standard deviation	16.10130951
Coefficient of variation (CV)	2.468740053
Kurtosis	219.5989898
Mean	6.5220757
Median Absolute Deviation (MAD)	1
Skewness	12.07727139
Sum	312577

Variance	259.2521678
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1	12640	26.4%
2	11490	24.0%
3	7992	16.7%
30	3294	6.9%
4	3293	6.9%
5	3029	6.3%
7	2050	4.3%
6	749	1.6%
14	561	1.2%
10	482	1.0%
Other values (90)	2346	4.9%

Value	Count	Frequency (%)
1	12640	26.4%
2	11490	24.0%
3	7992	16.7%
4	3293	6.9%
5	3029	6.3%
6	749	1.6%
7	2050	4.3%
8	130	0.3%
9	80	0.2%
10	482	1.0%

Value	Count	Frequency (%)
365	27	0.1%
364	1	< 0.1%
360	5	< 0.1%
354	1	< 0.1%
300	6	< 0.1%
299	1	< 0.1%
275	1	< 0.1%
270	2	< 0.1%
265	1	< 0.1%
250	1	< 0.1%

number_of_reviews

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

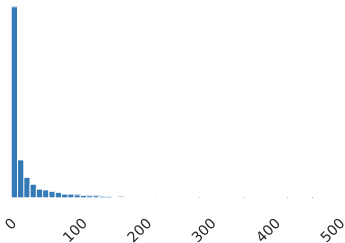
CORRELATION (This variable has a high correlation with 1 fields: reviews per month)

HIGH CORRELATION (This variable has a high correlation with 1 fields: reviews per month)

HIGH CORRELATION (This variable has a high correlation with 1 fields: reviews per month)

ZEROS

Distinct	386
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	23.57901765
Minimum	0
Maximum	488
Zeros	9510
Zeros (%)	19.8%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



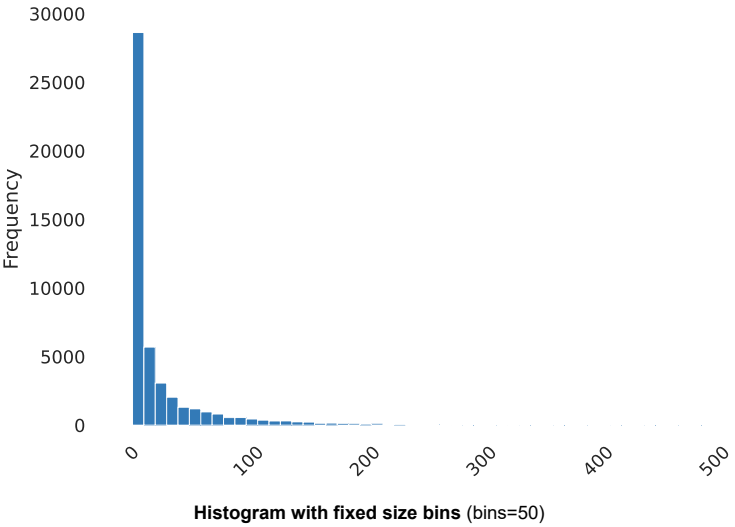
Quantile statistics

Minimum	0
5-th percentile	0
Q1	1
median	5
Q3	24
95-th percentile	115
Maximum	488
Range	488
Interquartile range (IQR)	23

Descriptive statistics

Standard deviation	44.30524514
Coefficient of variation (CV)	1.879011492
Kurtosis	16.28536397
Mean	23.57901765
Median Absolute Deviation (MAD)	5

Skewness	3.481772224
Sum	1130048
Variance	1962.954747
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	9510	19.8%
1	5122	10.7%
2	3408	7.1%
3	2487	5.2%
4	1970	4.1%
5	1592	3.3%
6	1333	2.8%
7	1160	2.4%
8	1103	2.3%
9	943	2.0%
Other values (376)	19298	40.3%

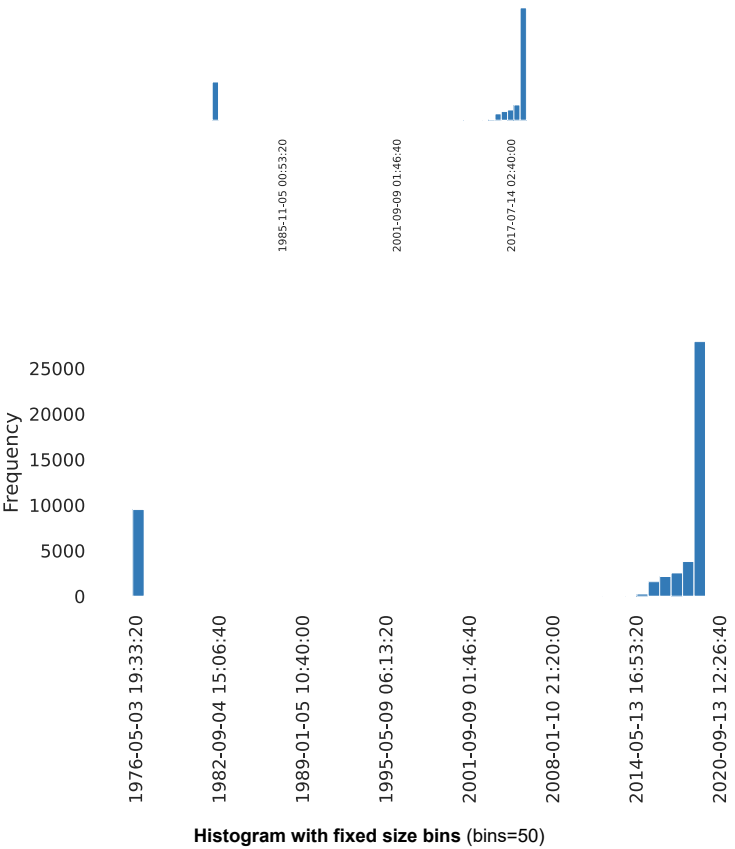
Value	Count	Frequency (%)
0	9510	19.8%
1	5122	10.7%
2	3408	7.1%
3	2487	5.2%
4	1970	4.1%
5	1592	3.3%
6	1333	2.8%
7	1160	2.4%
8	1103	2.3%
9	943	2.0%

Value	Count	Frequency (%)
488	1	< 0.1%
480	1	< 0.1%
474	1	< 0.1%
467	1	< 0.1%
466	1	< 0.1%
459	1	< 0.1%
458	1	< 0.1%
454	1	< 0.1%
451	1	< 0.1%
448	1	< 0.1%

last_review
Date

Distinct	1763
Distinct (%)	3.7%
Missing	0
Missing (%)	0.0%
Memory size	374.5 KiB

Minimum	1976-01-01 01:00:00+00:00
Maximum	2019-07-08 00:00:00+00:00



reviews_per_month

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

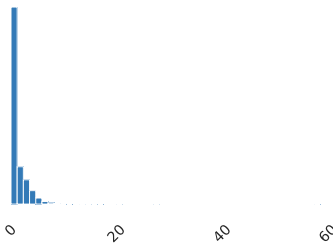
CORRELATION (This variable has a high correlation with 1 fields: number of reviews)

HIGH CORRELATION (This variable has a high correlation with 1 fields: number of reviews)

HIGH CORRELATION (This variable has a high correlation with 1 fields: number of reviews)

ZEROS

Distinct	935
Distinct (%)	2.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.100549806
Minimum	0
Maximum	58.5
Zeros	9510
Zeros (%)	19.8%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



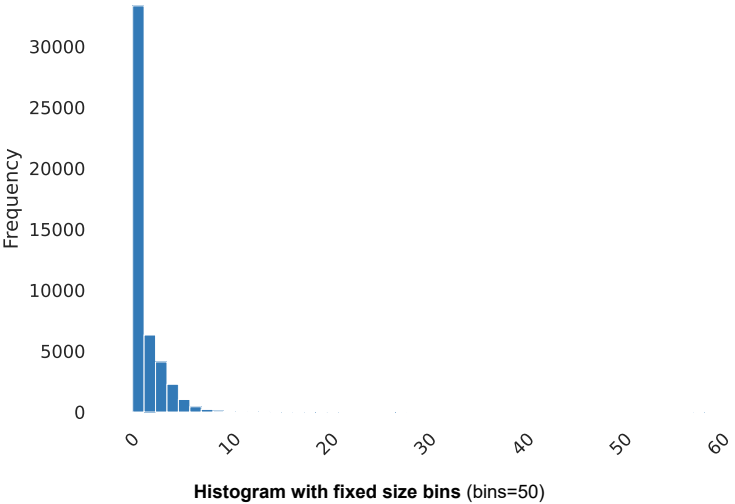
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0.05
median	0.38
Q3	1.61
95-th percentile	4.33
Maximum	58.5
Range	58.5
Interquartile range (IQR)	1.56

Descriptive statistics

Standard deviation	1.599773144
Coefficient of variation (CV)	1.453612672
Kurtosis	43.67746979
Mean	1.100549806
Median Absolute Deviation (MAD)	0.38

Skewness	3.28091007
Sum	52744.95
Variance	2.559274112
Monotonicity	Not monotonic



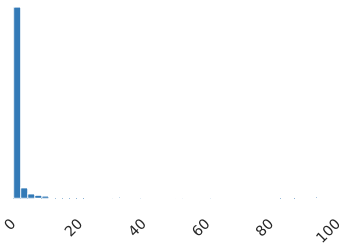
Value	Count	Frequency (%)
0	9510	19.8%
0.02	911	1.9%
0.05	883	1.8%
1	876	1.8%
0.03	797	1.7%
0.16	661	1.4%
0.04	646	1.3%
0.08	593	1.2%
0.09	586	1.2%
0.06	572	1.2%
Other values (925)	31891	66.5%

Value	Count	Frequency (%)
0	9510	19.8%
0.01	42	0.1%
0.02	911	1.9%
0.03	797	1.7%
0.04	646	1.3%
0.05	883	1.8%
0.06	572	1.2%
0.07	458	1.0%
0.08	593	1.2%
0.09	586	1.2%

Value	Count	Frequency (%)
58.5	1	< 0.1%
27.95	1	< 0.1%
20.94	1	< 0.1%
19.75	1	< 0.1%
17.82	1	< 0.1%
16.81	1	< 0.1%
16.03	1	< 0.1%
15.78	1	< 0.1%
15.32	1	< 0.1%
15.23	1	< 0.1%

calculated_host_listings_count
Real number ($\mathbb{R}_{\geq 0}$)

Distinct	43
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.397091349
Minimum	1
Maximum	96
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



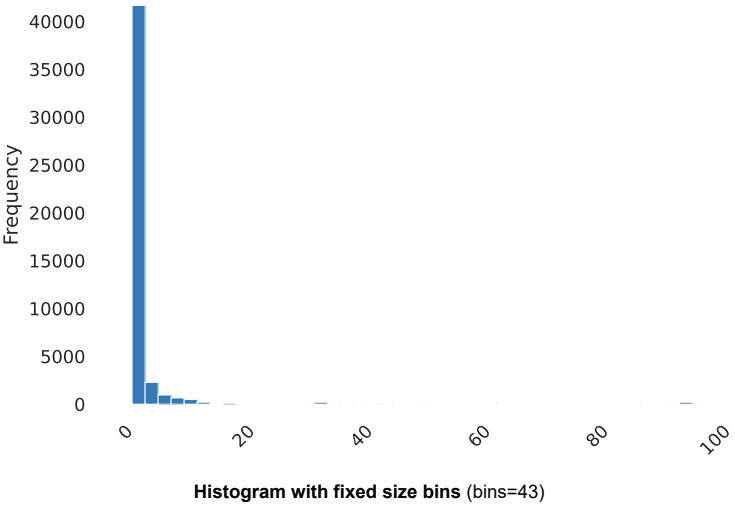
Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	1
Q3	2
95-th percentile	9
Maximum	96
Range	95
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	9.918211126
Coefficient of variation (CV)	2.919618611
Kurtosis	54.33235713
Mean	3.397091349
Median Absolute Deviation (MAD)	0
Skewness	6.988563317
Sum	162809

Variance	98.37091194
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1	32182	67.1%
2	6641	13.9%
3	2843	5.9%
4	1436	3.0%
5	840	1.8%
6	558	1.2%
8	416	0.9%
7	399	0.8%
9	234	0.5%
10	210	0.4%
Other values (33)	2167	4.5%

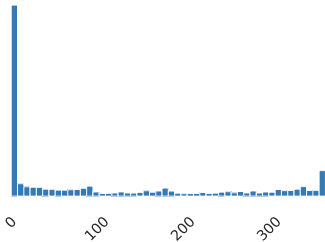
Value	Count	Frequency (%)
1	32182	67.1%
2	6641	13.9%
3	2843	5.9%
4	1436	3.0%
5	840	1.8%
6	558	1.2%
7	399	0.8%
8	416	0.9%
9	234	0.5%
10	210	0.4%

Value	Count	Frequency (%)
96	192	0.4%
91	91	0.2%
87	87	0.2%
65	65	0.1%
52	104	0.2%
50	50	0.1%
49	98	0.2%
47	47	0.1%
43	43	0.1%
39	39	0.1%

availability_365
Real number ($\mathbb{R}_{\geq 0}$)

ZEROS

Distinct	366
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	109.8798564
Minimum	0
Maximum	365
Zeros	17464
Zeros (%)	36.4%
Negative	0
Negative (%)	0.0%
Memory size	374.5 KiB



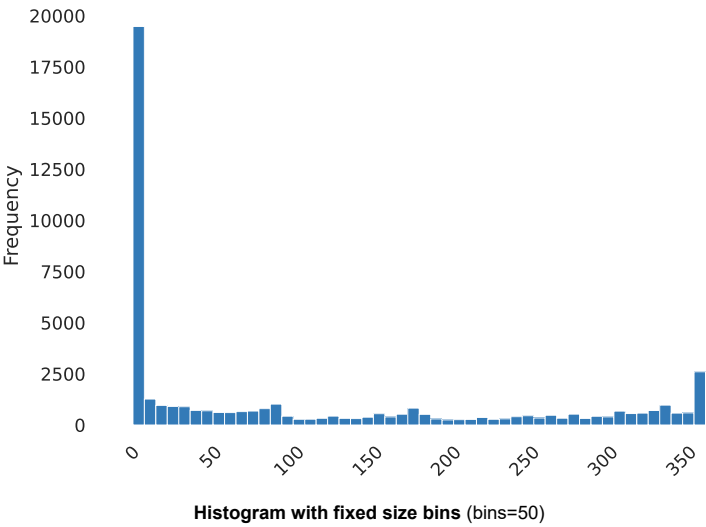
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	41
Q3	217
95-th percentile	359
Maximum	365
Range	365
Interquartile range (IQR)	217

Descriptive statistics

Standard deviation	130.3937594
Coefficient of variation (CV)	1.186693937
Kurtosis	-0.9174765325
Mean	109.8798564
Median Absolute Deviation (MAD)	41
Skewness	0.8067430482
Sum	5266102

Variance	17002.5325
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	17464	36.4%
365	1233	2.6%
364	468	1.0%
1	401	0.8%
89	358	0.7%
5	340	0.7%
3	306	0.6%
179	297	0.6%
90	285	0.6%
2	269	0.6%
Other values (356)	26505	55.3%

Value	Count	Frequency (%)
0	17464	36.4%
1	401	0.8%
2	269	0.6%
3	306	0.6%
4	233	0.5%
5	340	0.7%
6	244	0.5%
7	218	0.5%
8	233	0.5%
9	193	0.4%

Value	Count	Frequency (%)
365	1233	2.6%
364	468	1.0%
363	228	0.5%
362	160	0.3%
361	107	0.2%
360	100	0.2%
359	131	0.3%
358	175	0.4%
357	89	0.2%
356	75	0.2%

month
Categorical

Distinct	12
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	2.9 MiB

Length

Max length	9
Median length	5
Mean length	5.370070525
Min length	3

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

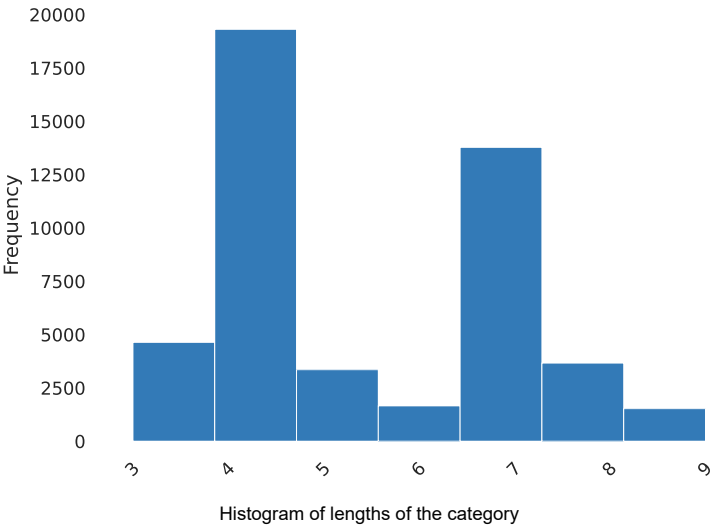
Sample

1st row	October
2nd row	May
3rd row	January
4th row	July
5th row	November

Common Values

Value	Count	Frequency (%)
June	13390	27.9%
January	12243	25.5%
July	5920	12.4%
May	4635	9.7%
April	2086	4.4%
December	1760	3.7%
August	1658	3.5%
October	1534	3.2%
September	1528	3.2%
March	1268	2.6%
Other values (2)	1904	4.0%

Length



Value	Count	Frequency (%)
june	13390	27.9%
january	12243	25.5%
july	5920	12.4%
may	4635	9.7%
april	2086	4.4%
december	1760	3.7%
august	1658	3.5%
october	1534	3.2%
september	1528	3.2%
march	1268	2.6%
Other values (2)	1904	4.0%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

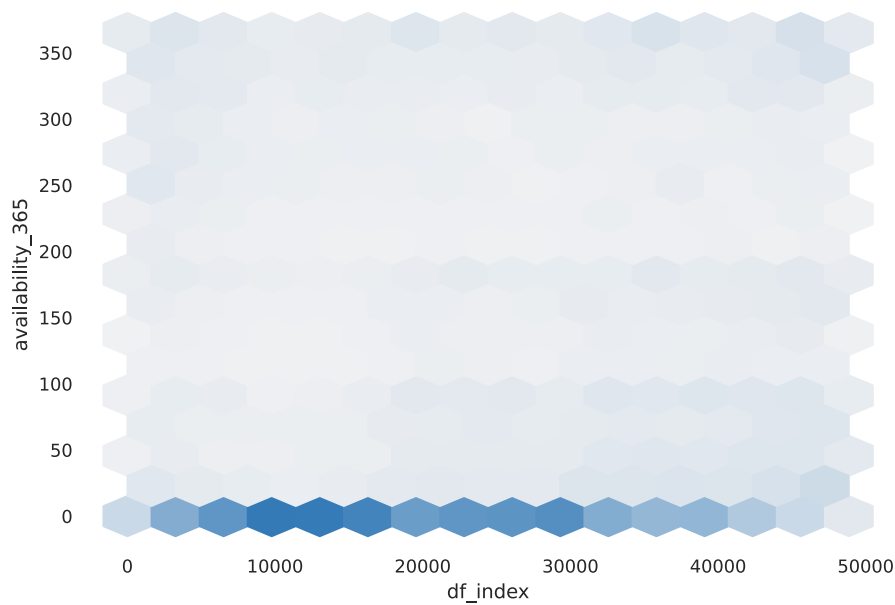
Value	Count	Frequency (%)
No values found.		

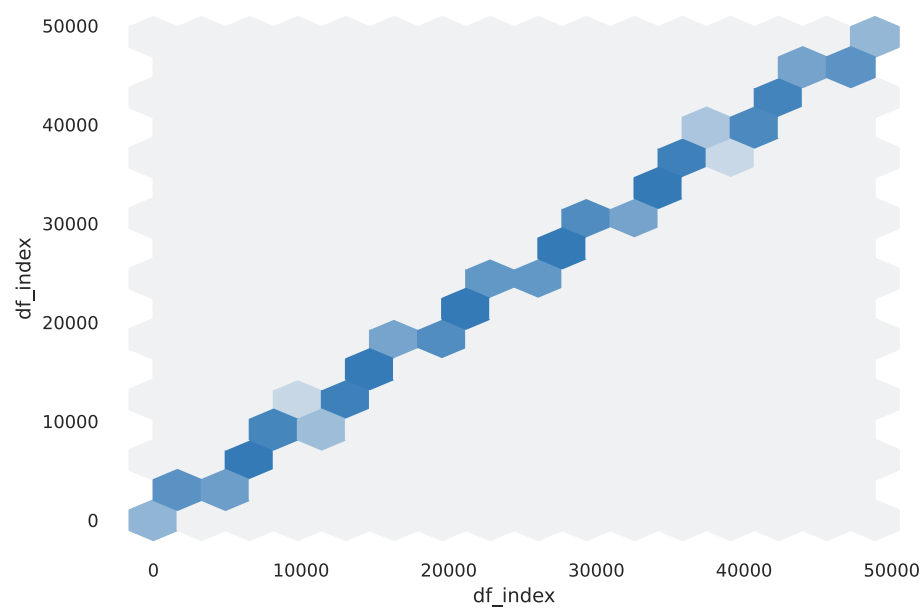
Most frequent character per block

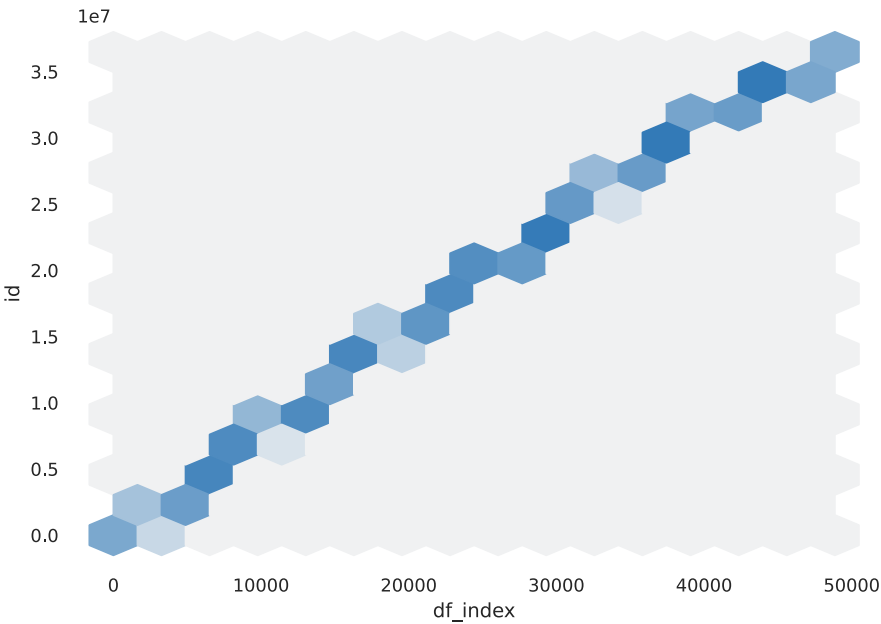
Interactions

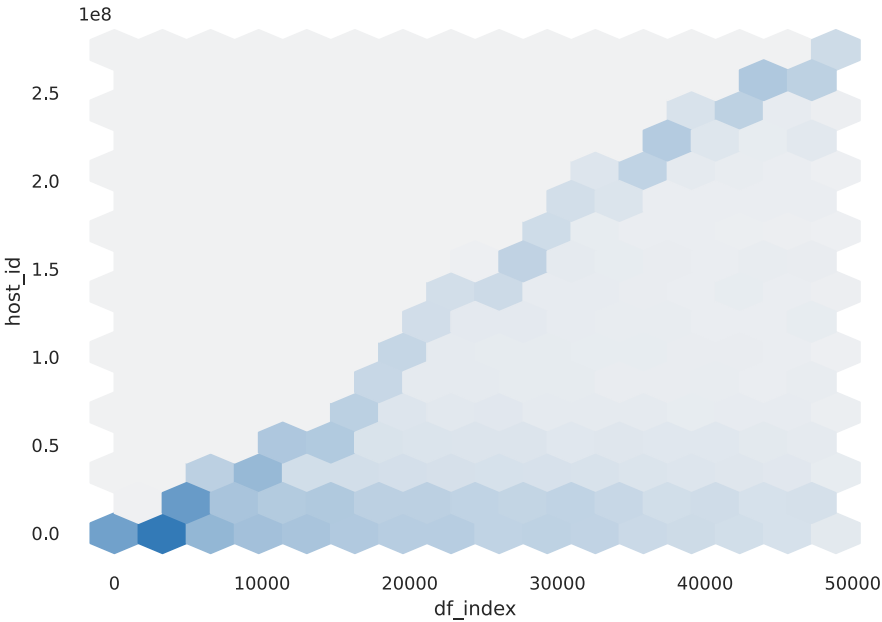
df_index
id
host_id
latitude
longitude
price
minimum_nights
number_of_reviews
reviews_per_month
calculated_host_listing:
availability_365

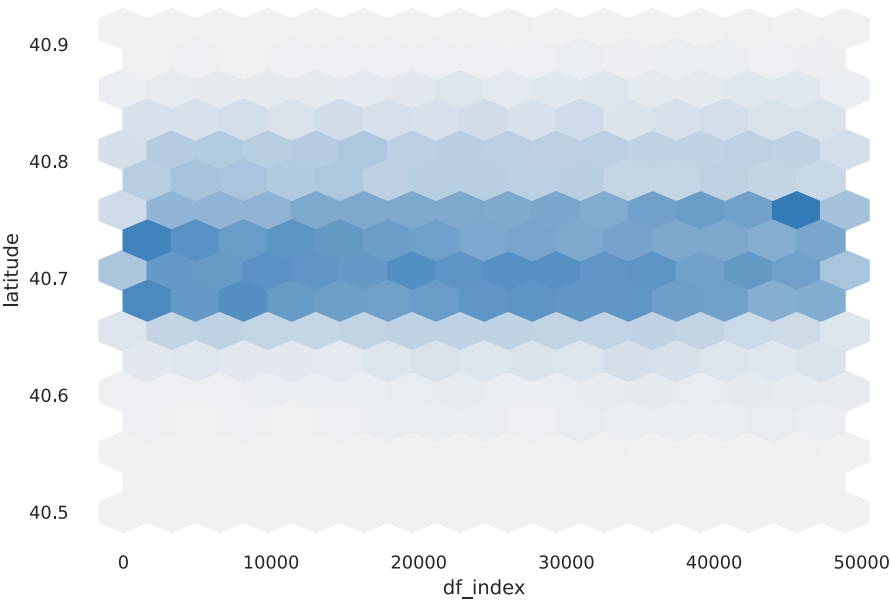
availability_365
df_index
id
host_id
latitude
longitude
price
minimum_nights
number_of_reviews
reviews_per_month
calculated_host_listing:

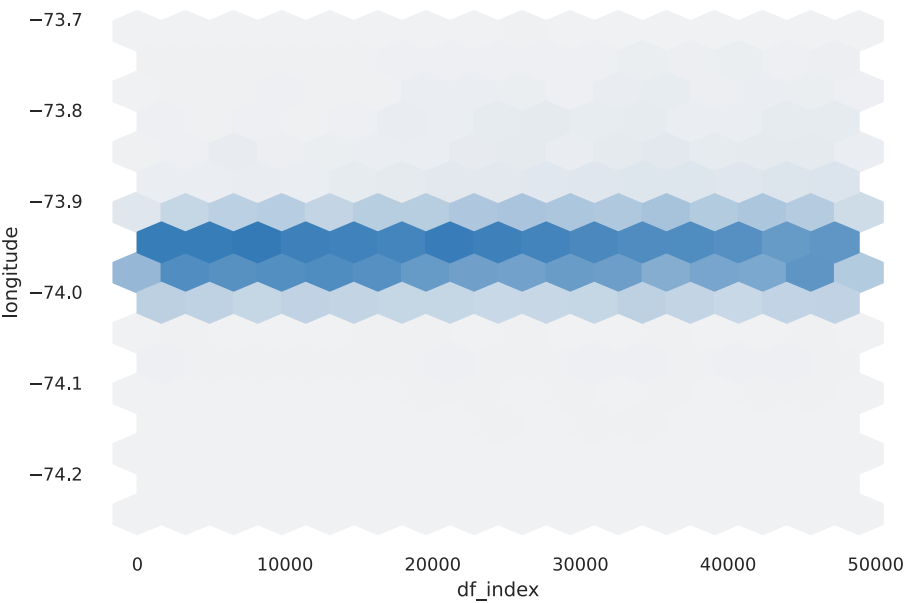


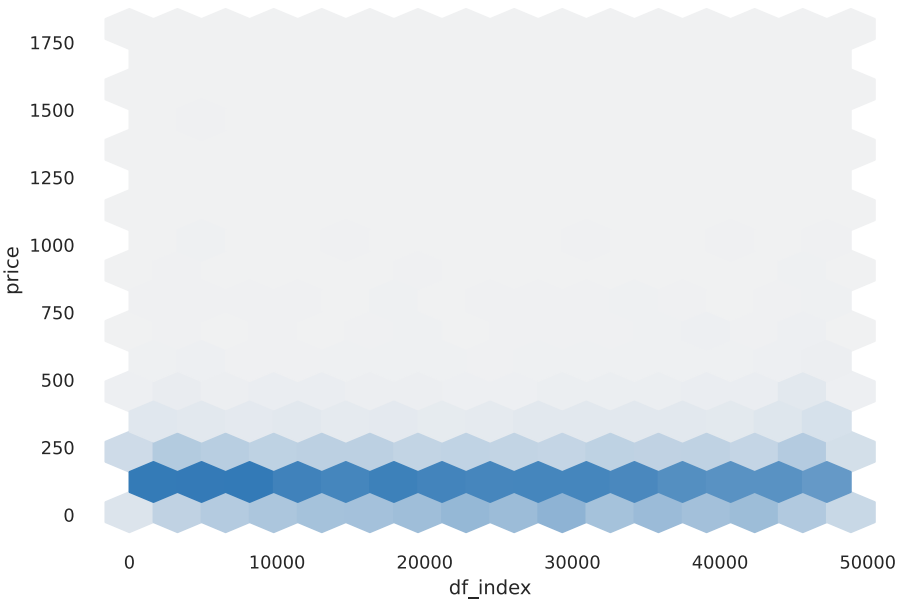


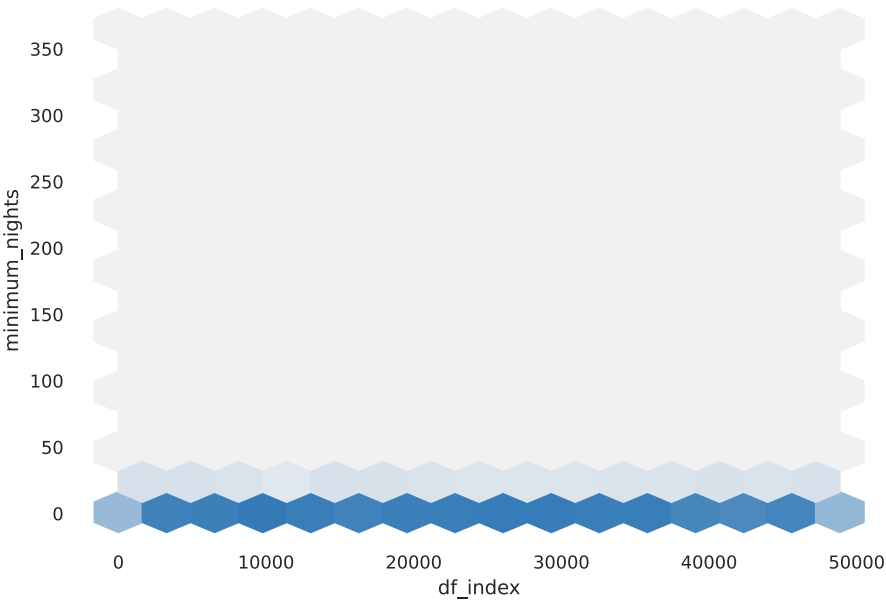


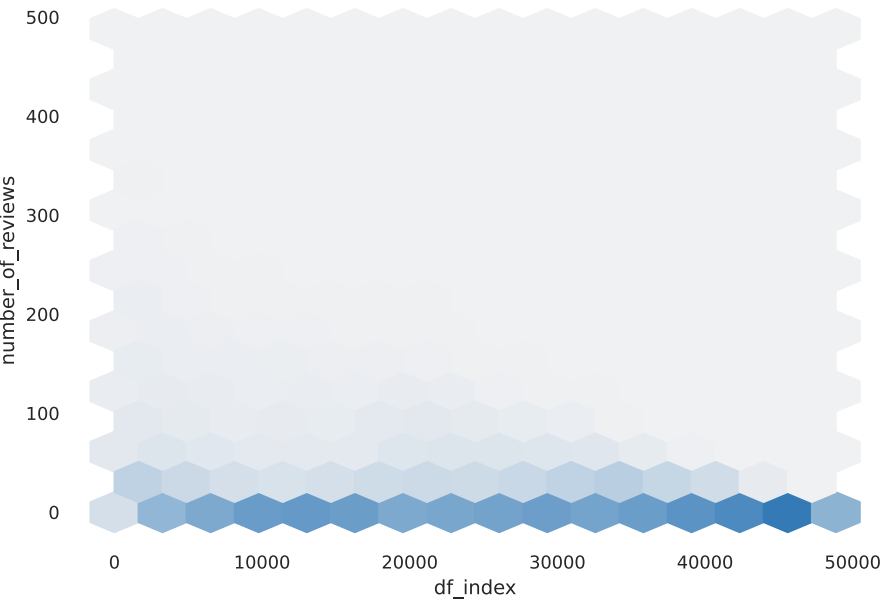


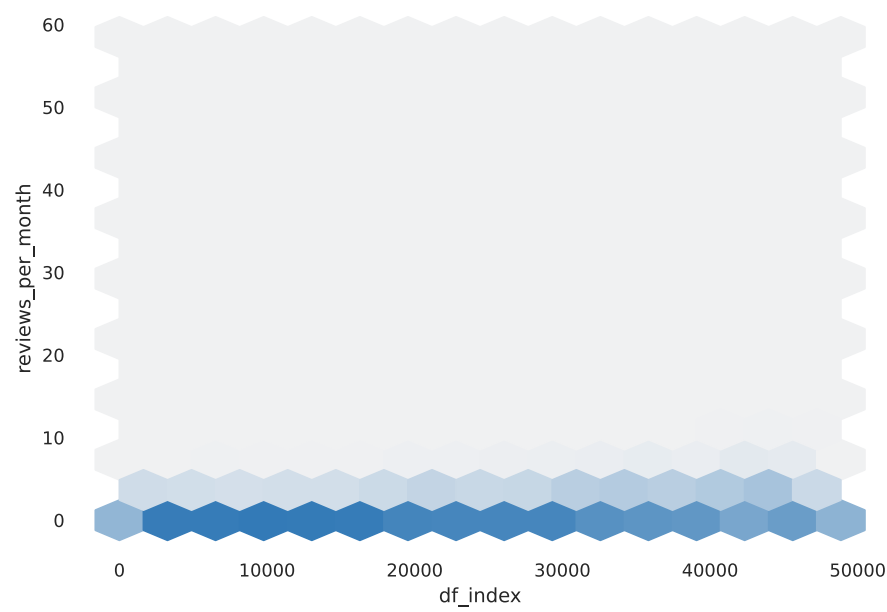


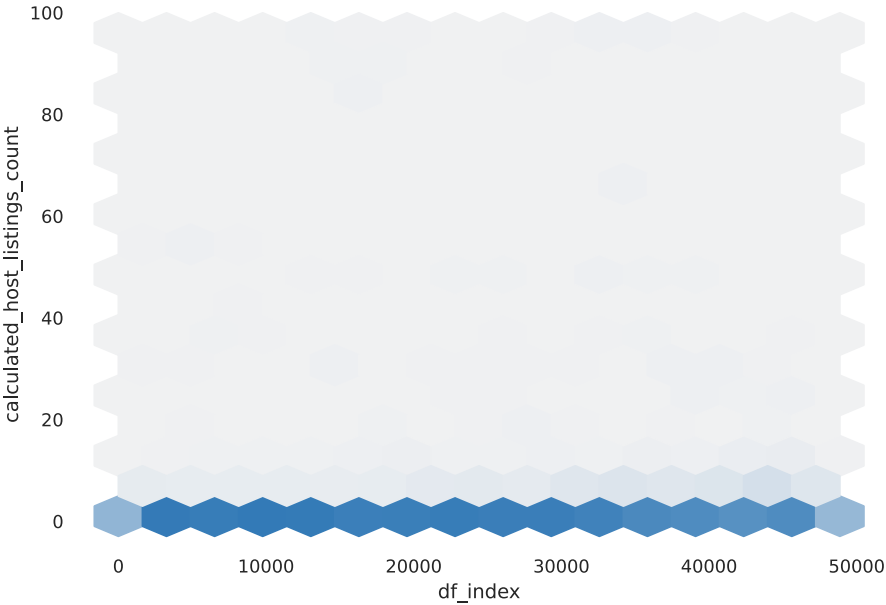












- availability_365

df_index

id

host_id

latitude

longitude

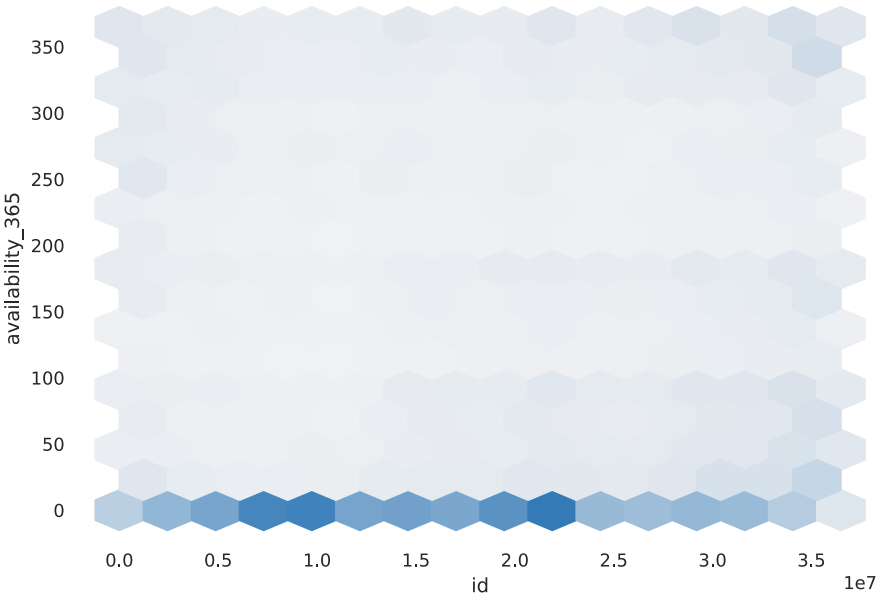
price

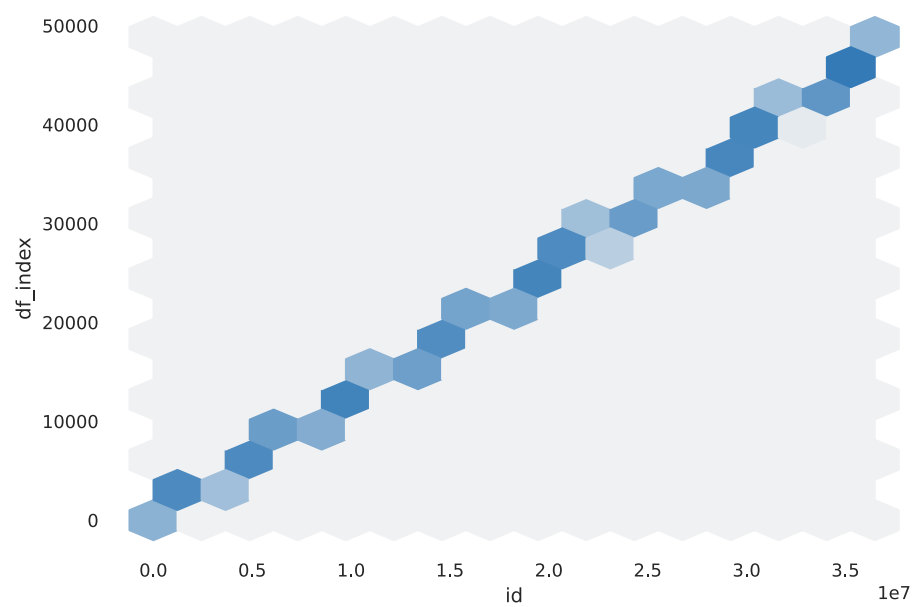
minimum_nights

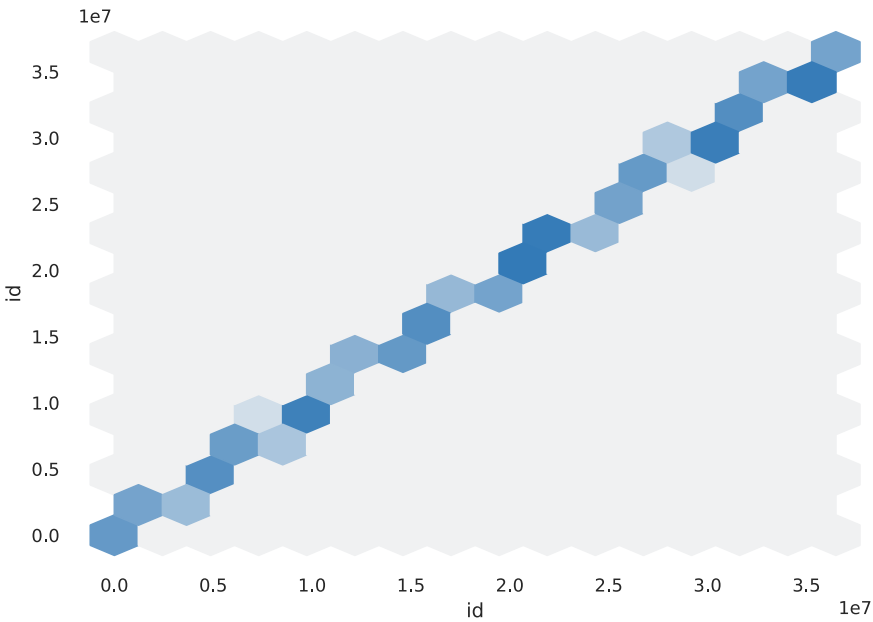
number_of_reviews

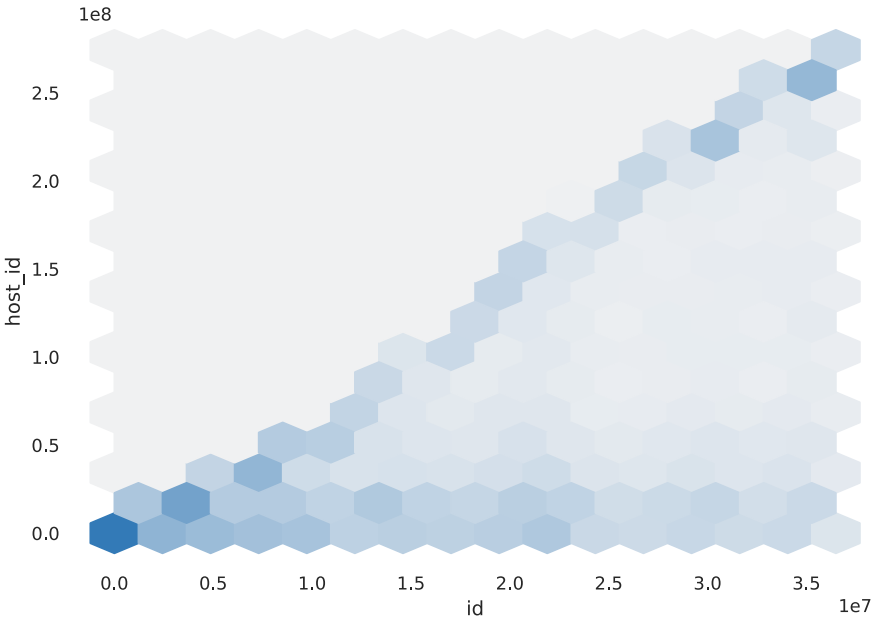
reviews_per_month

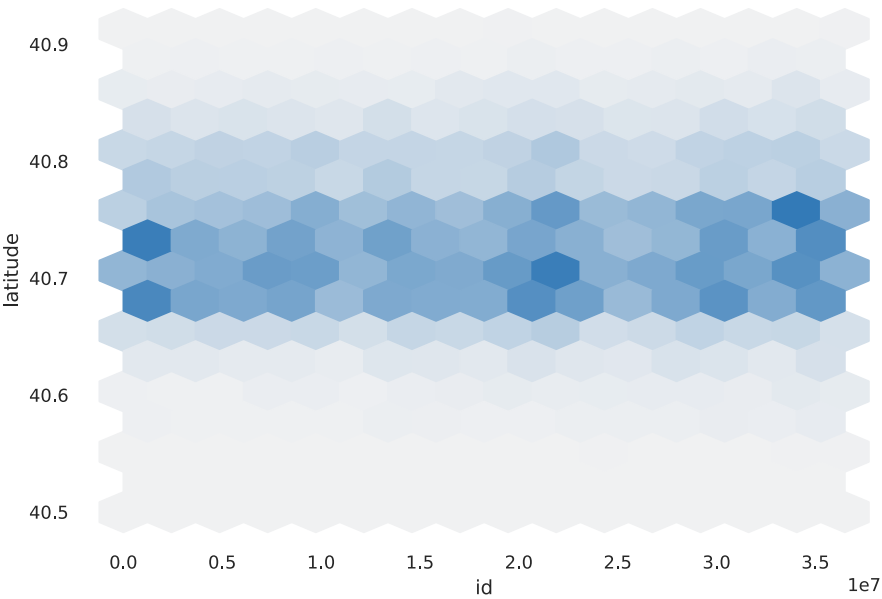
calculated_host_listing:

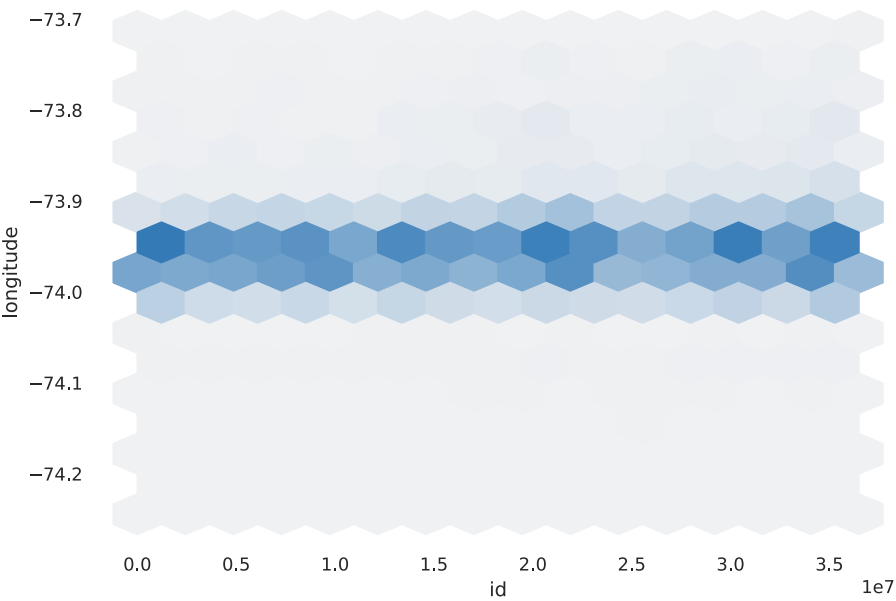


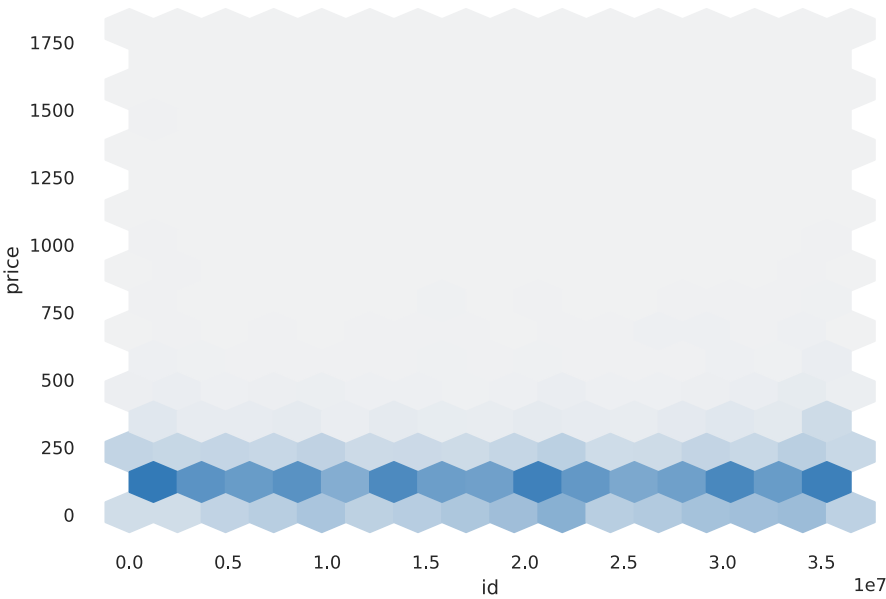


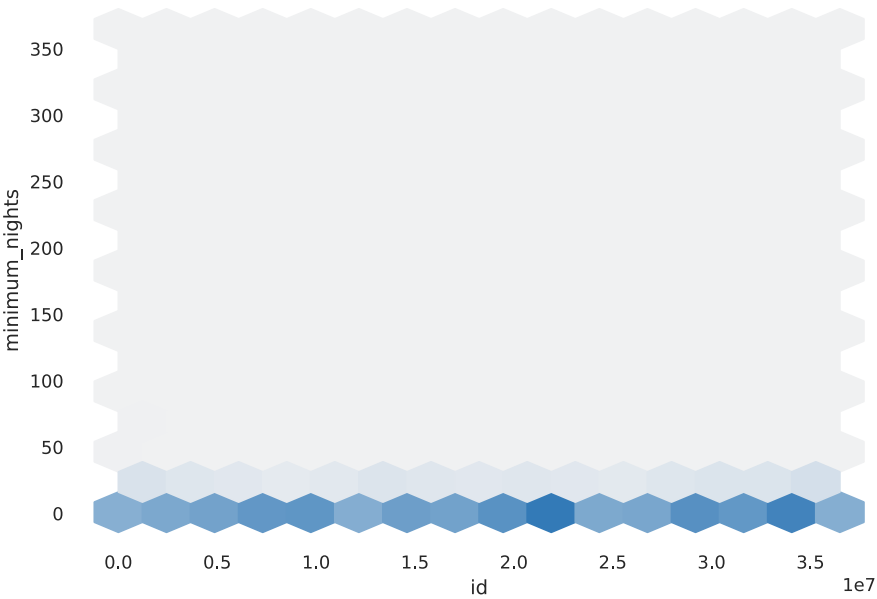


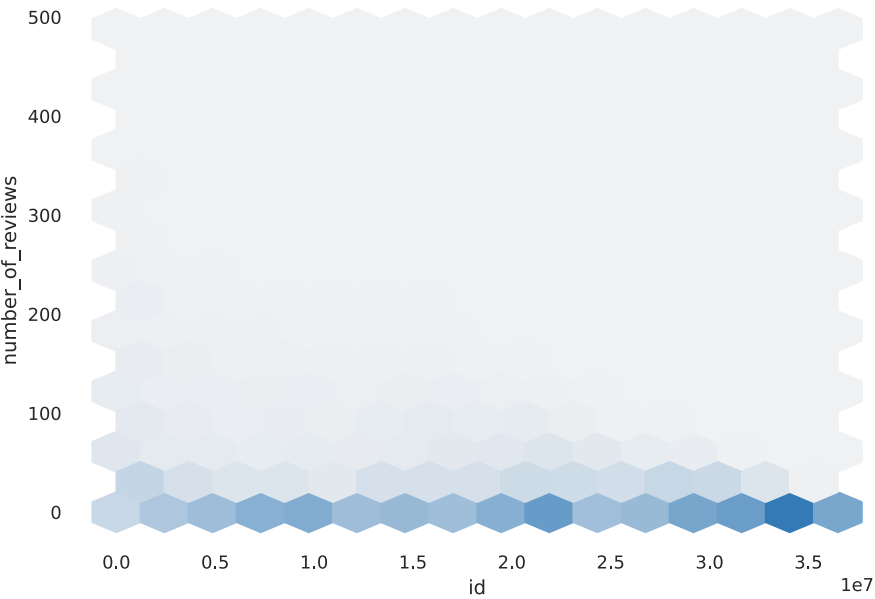


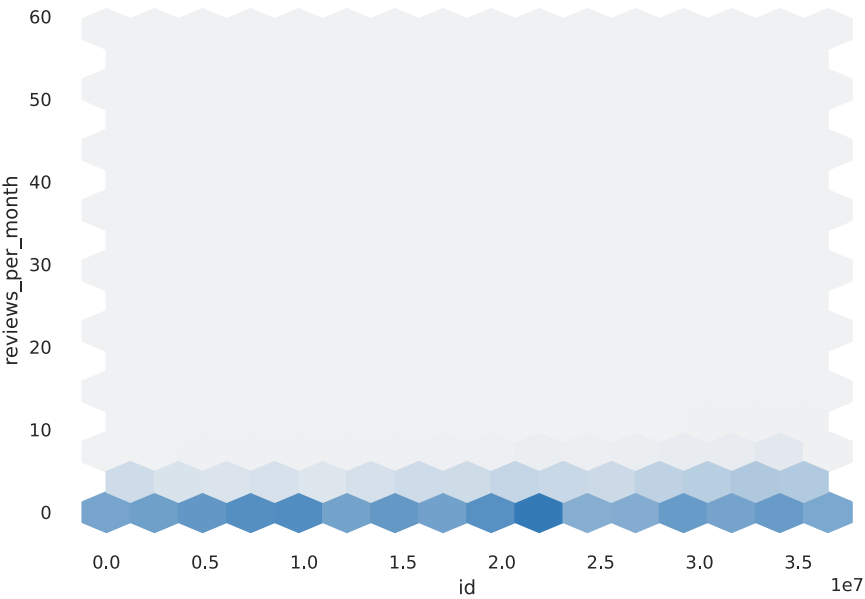


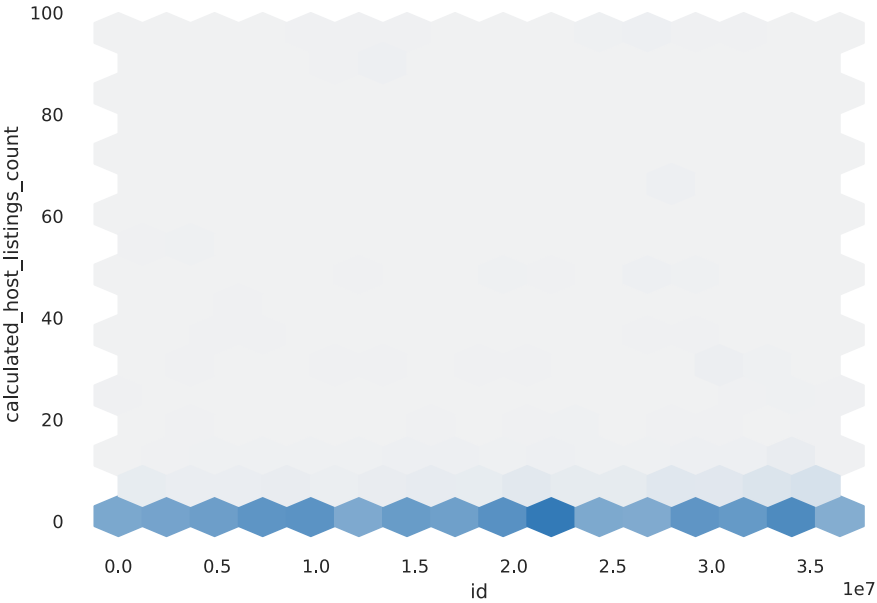




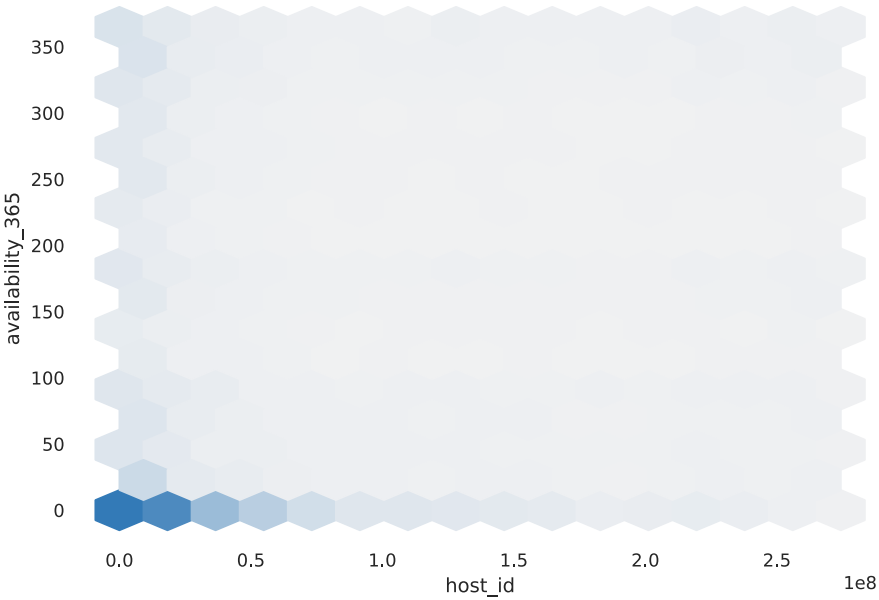


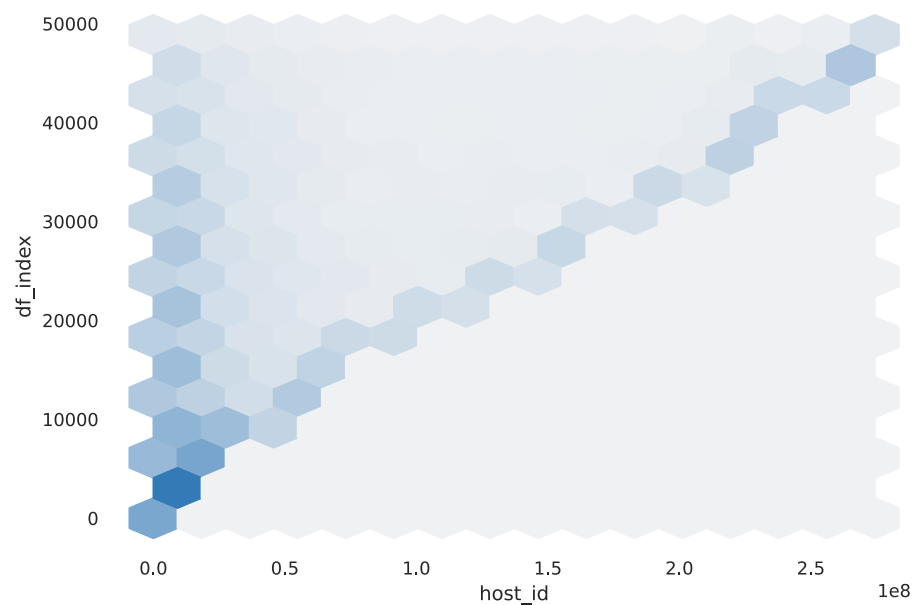


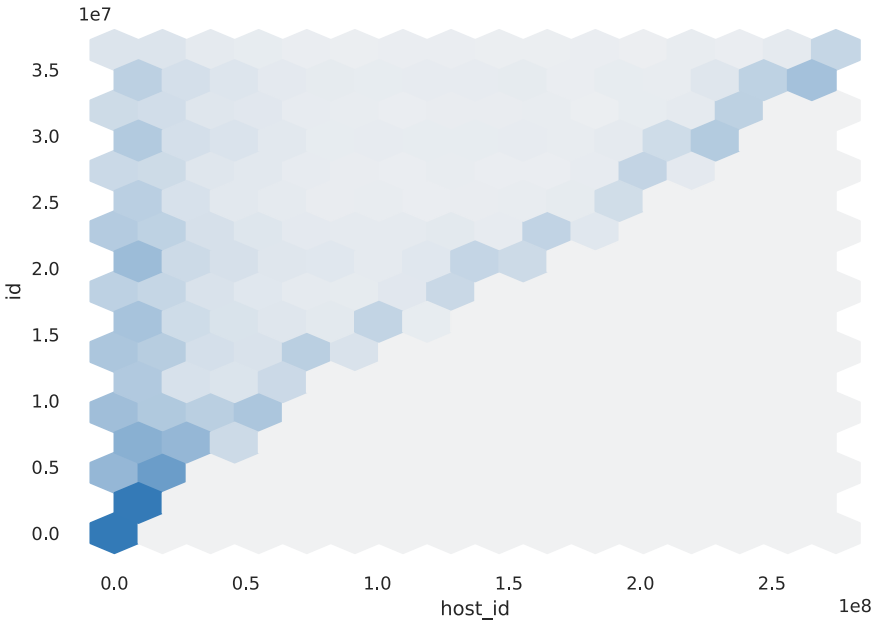


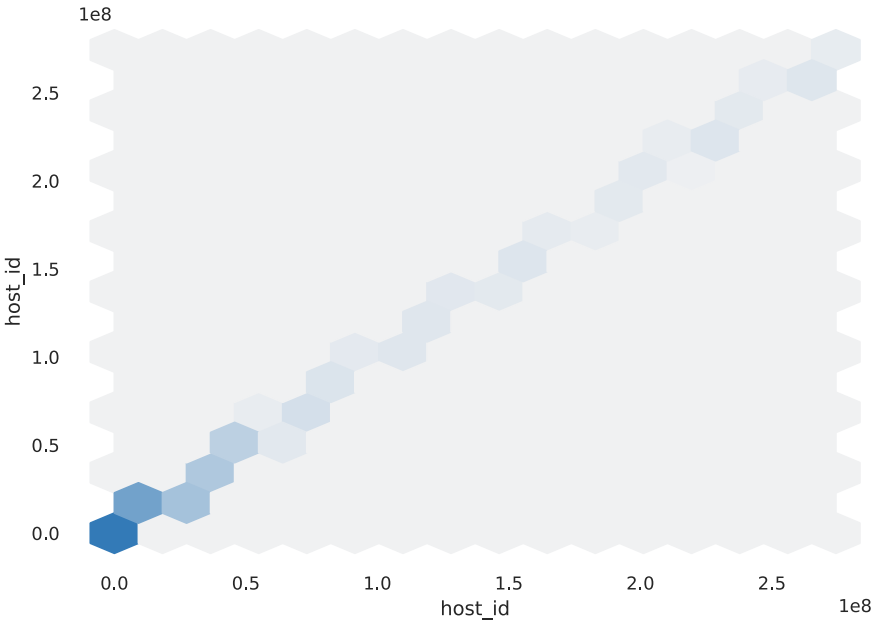


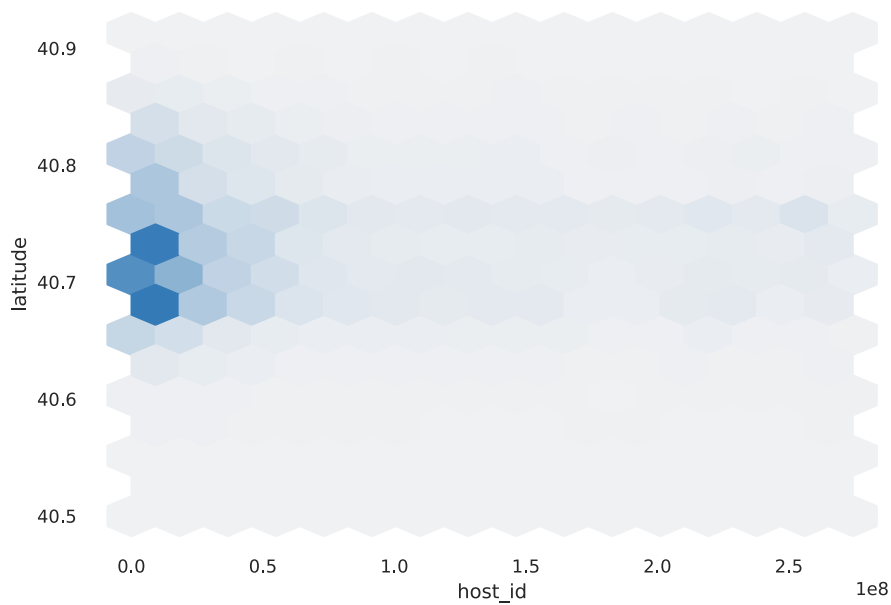
- availability_365
- df_index
- id
- host_id
- latitude
- longitude
- price
- minimum_nights
- number_of_reviews
- reviews_per_month
- calculated_host_listing:

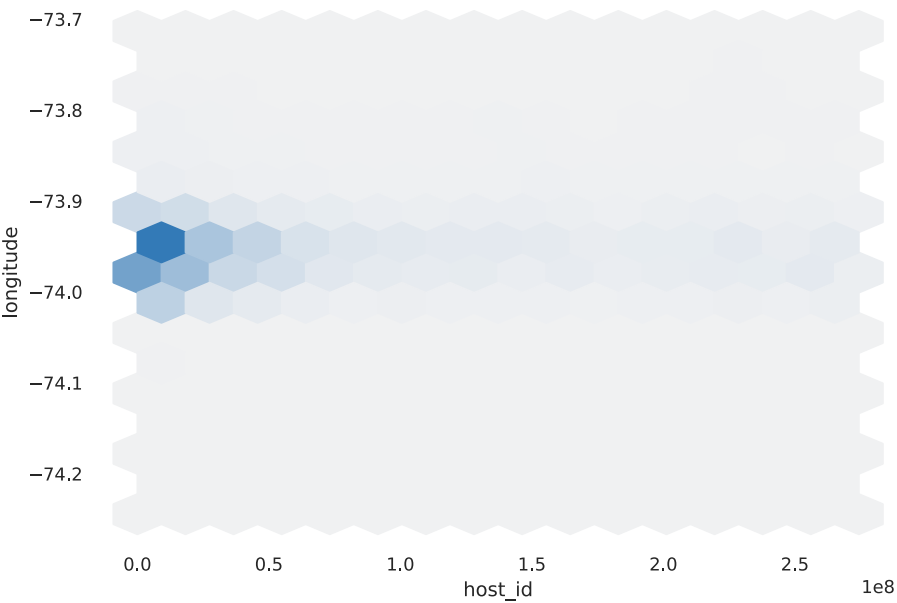


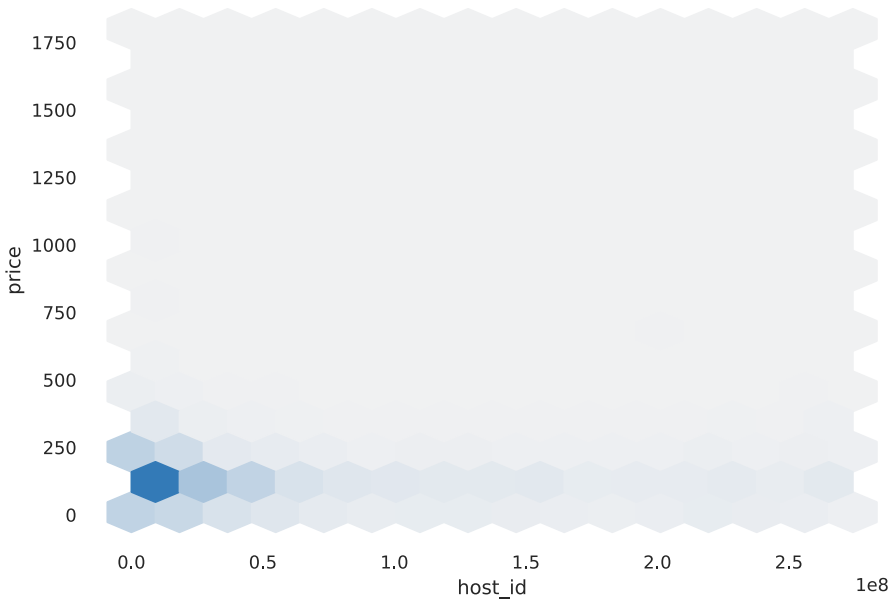


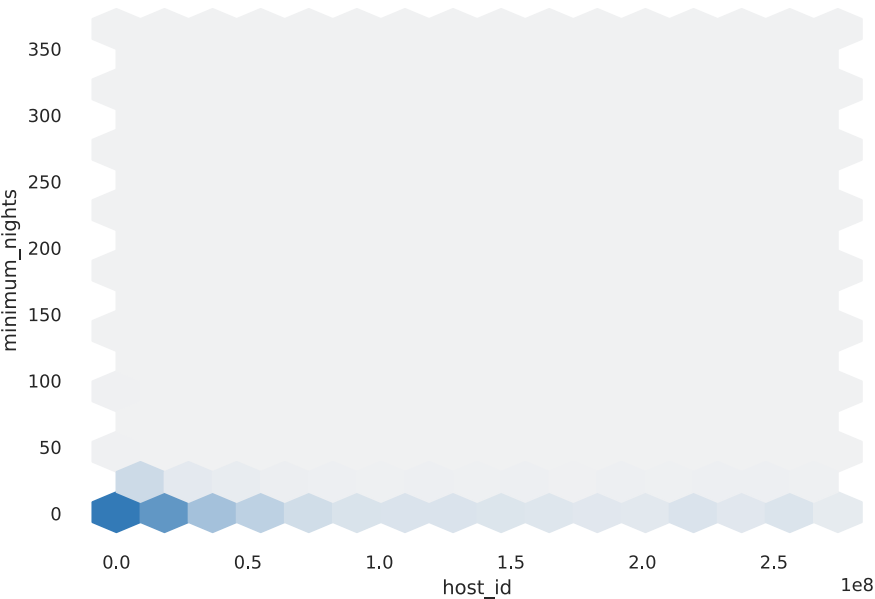


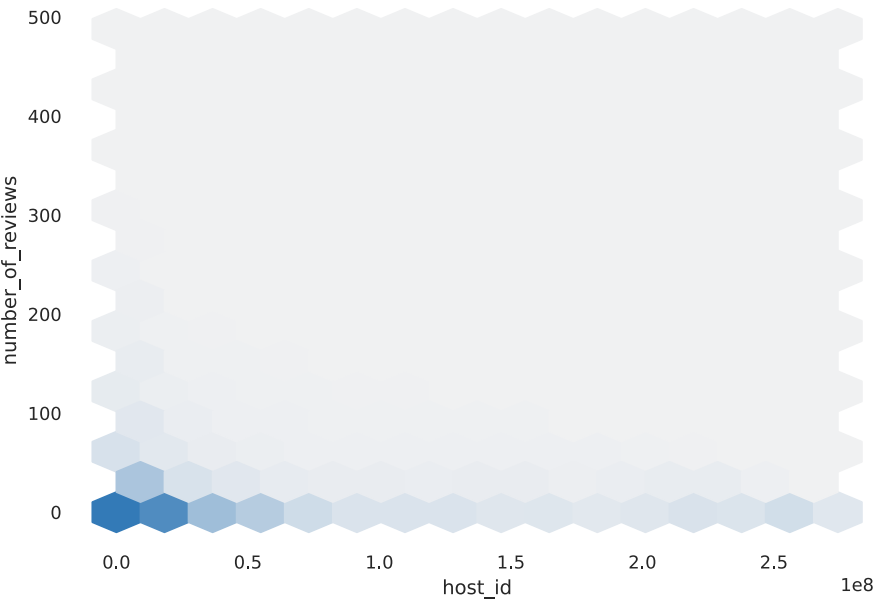


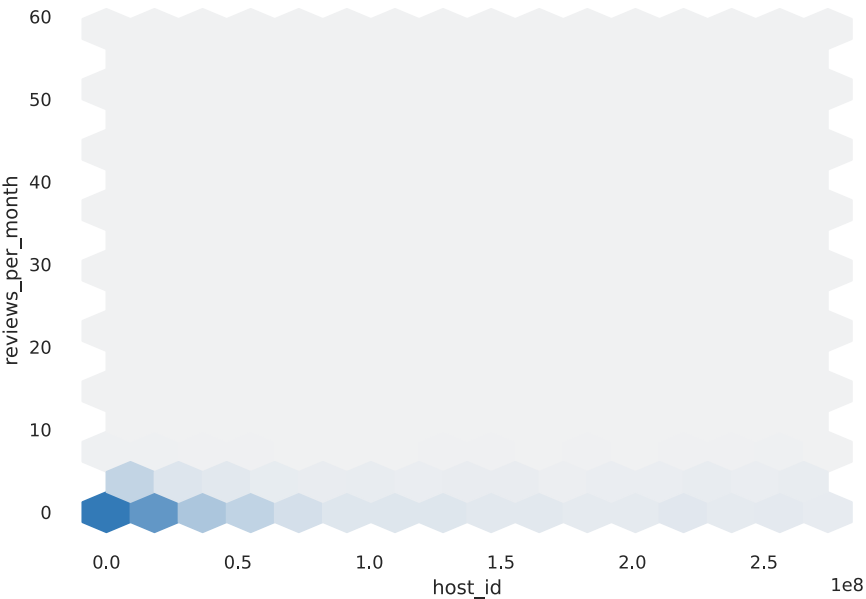


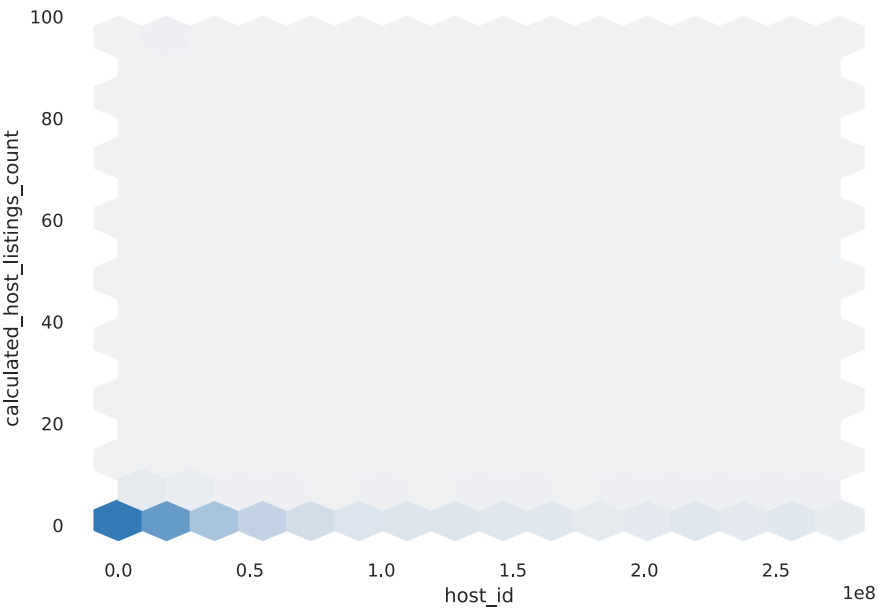




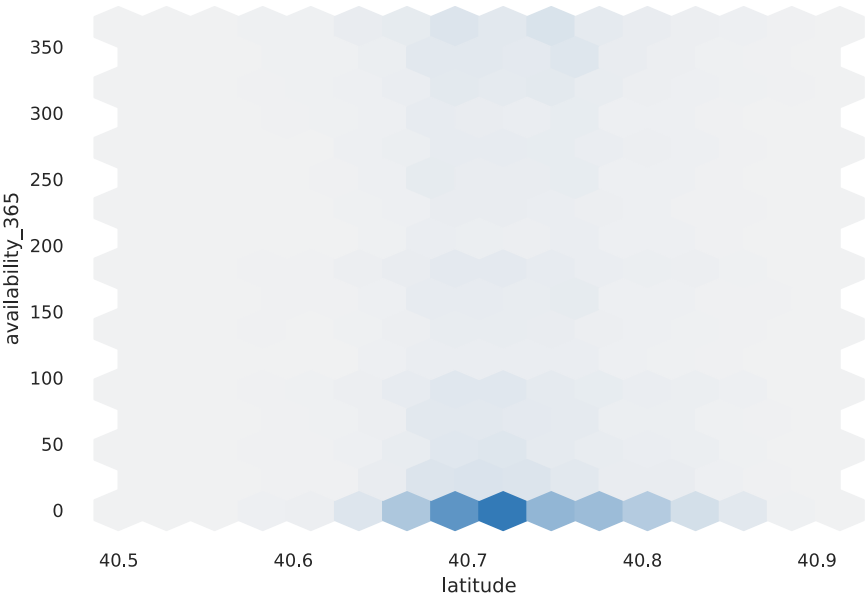


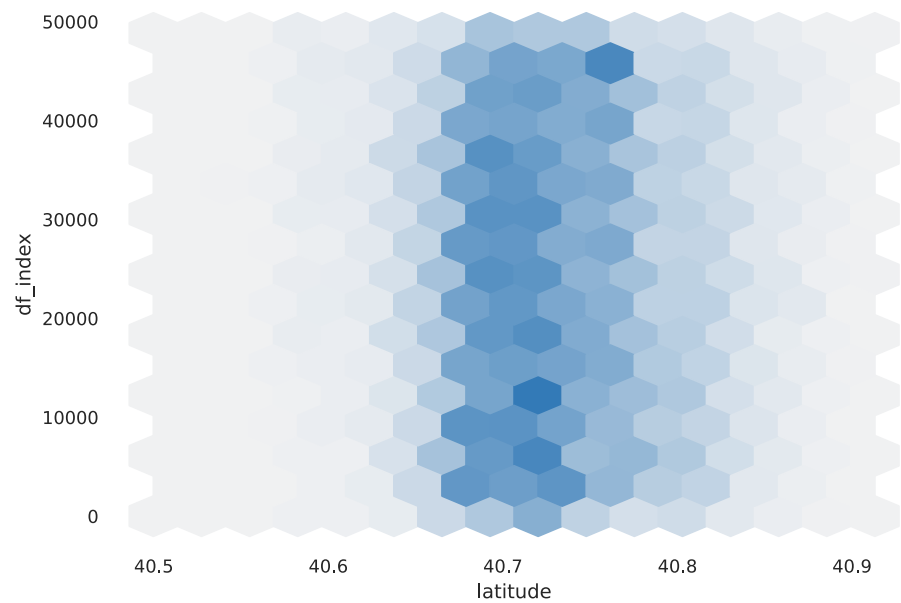


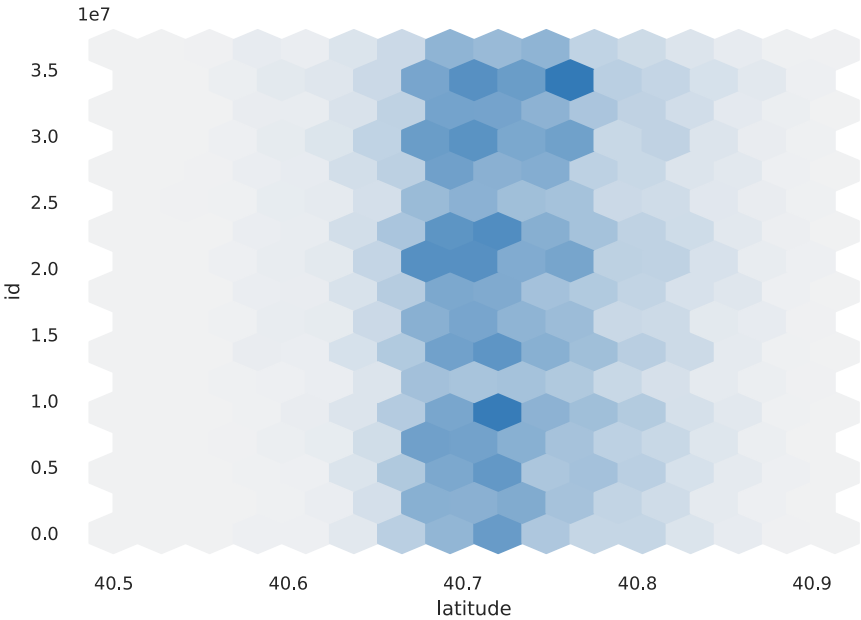


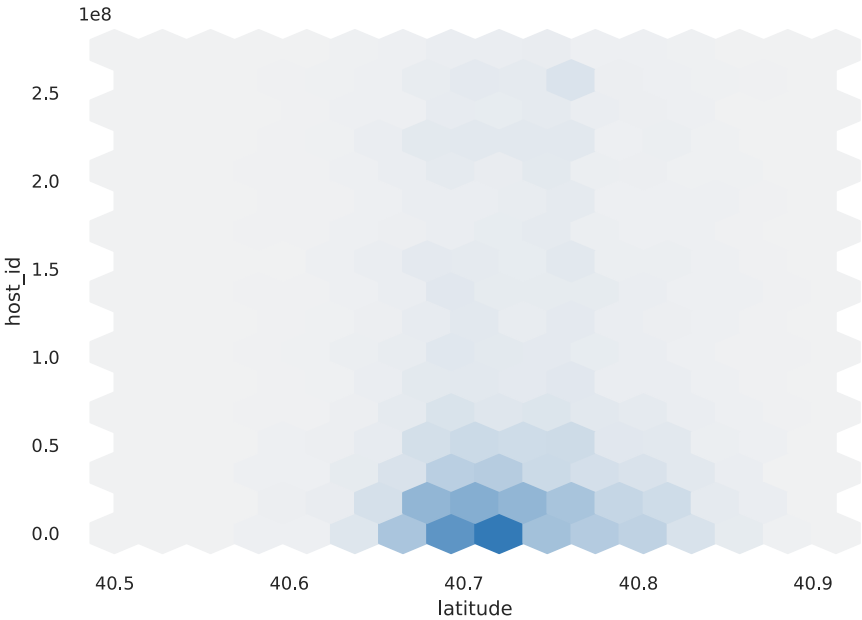


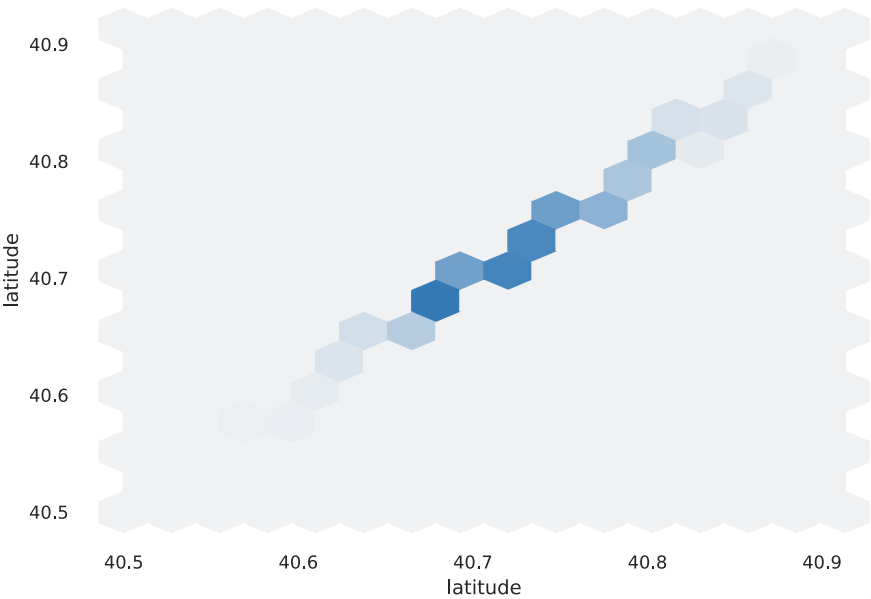
availability_365
df_index
id
host_id
latitude
longitude
price
minimum_nights
number_of_reviews
reviews_per_month
calculated_host_listing:

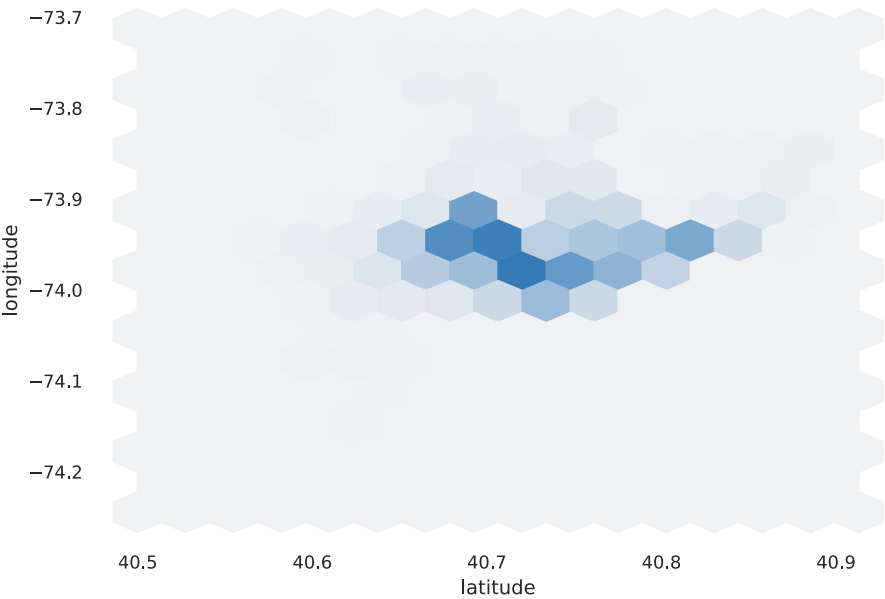


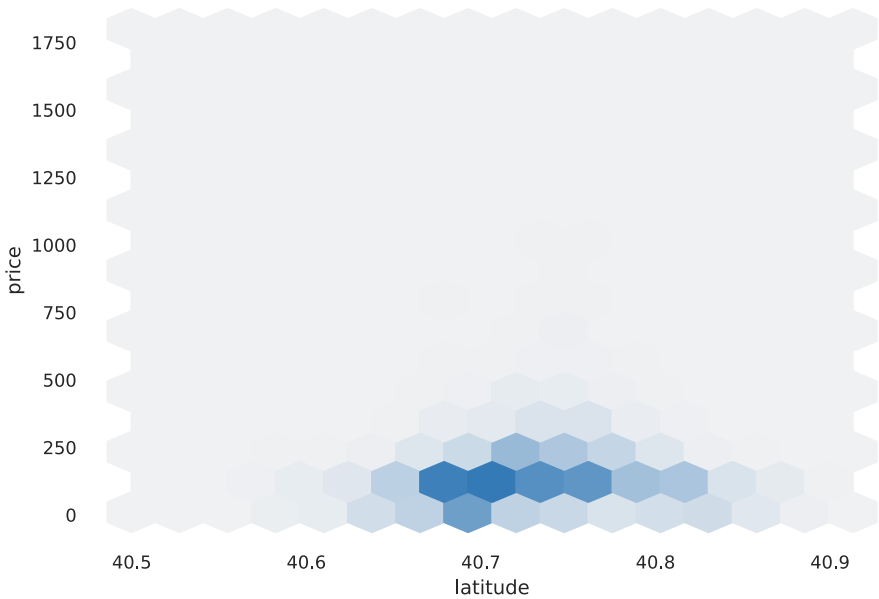


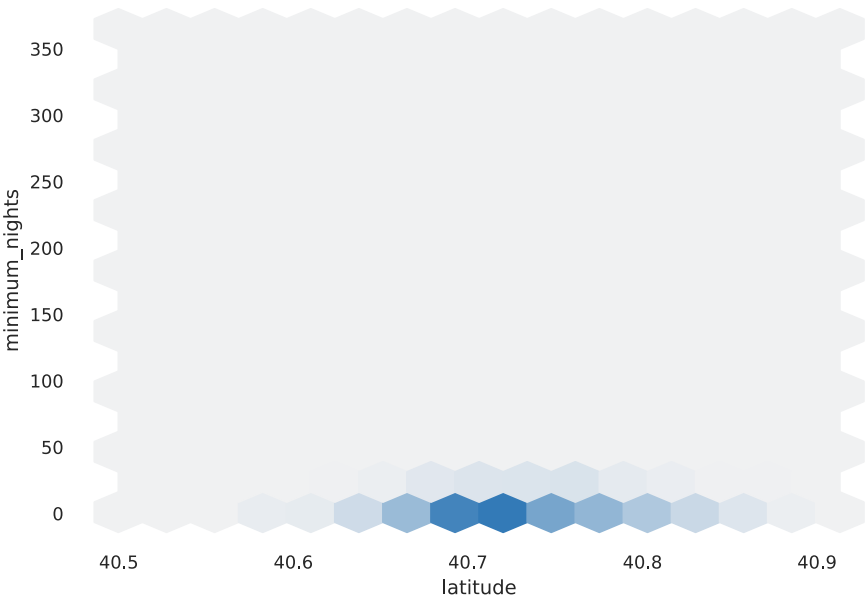


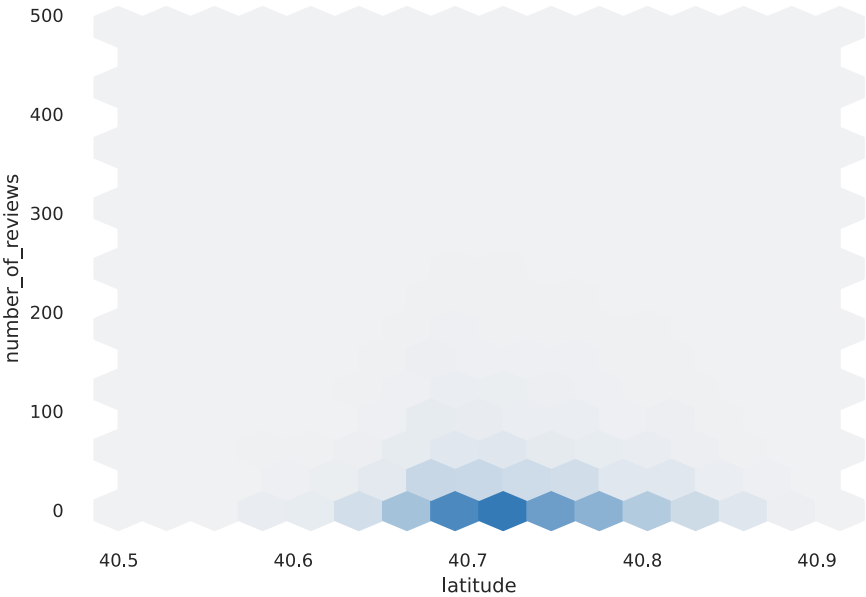


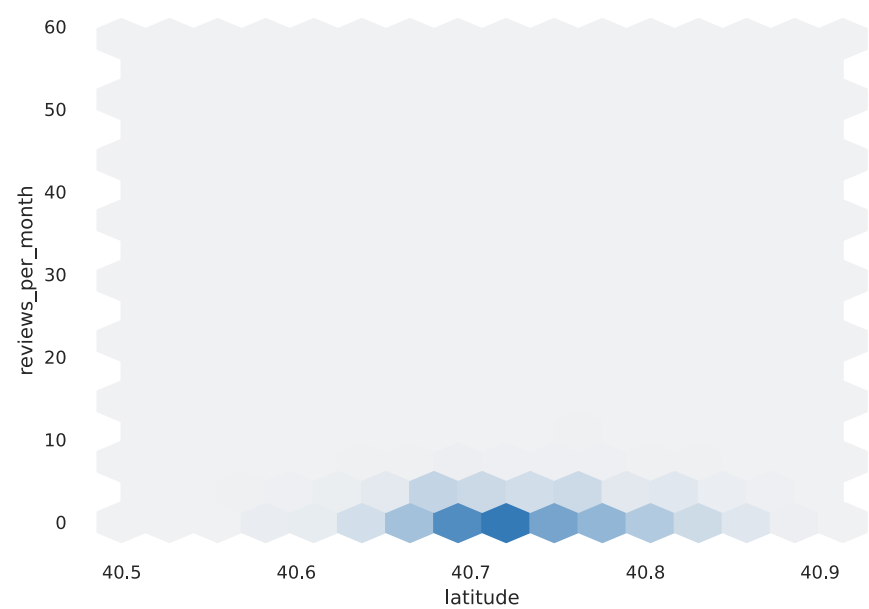


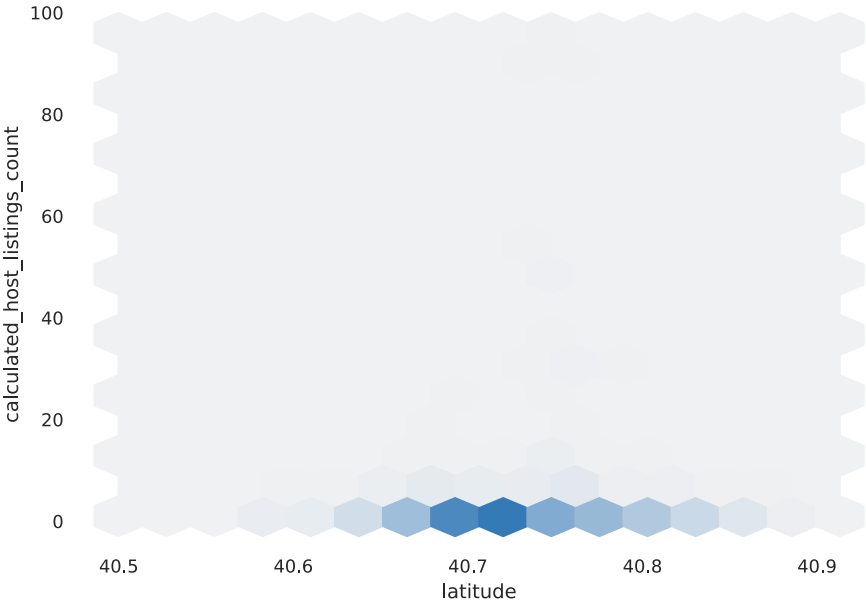




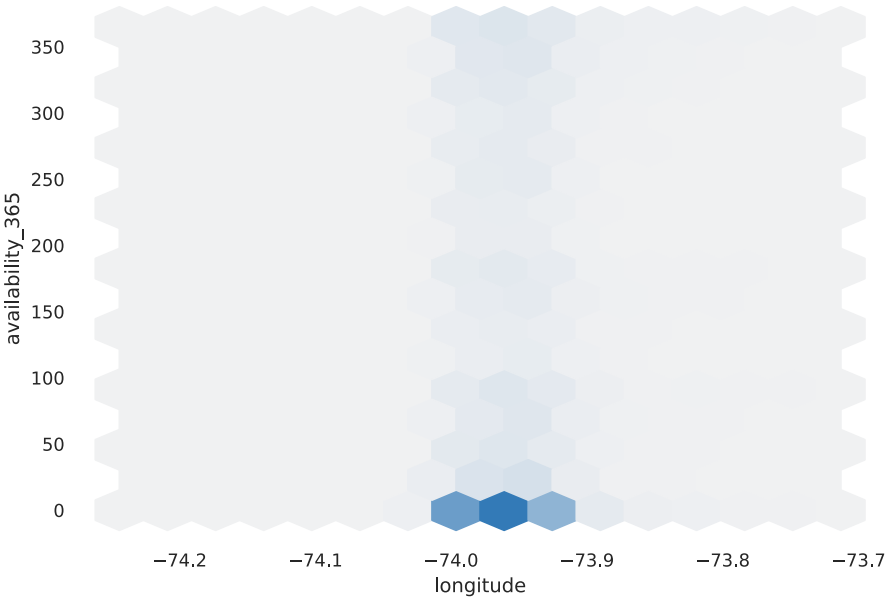


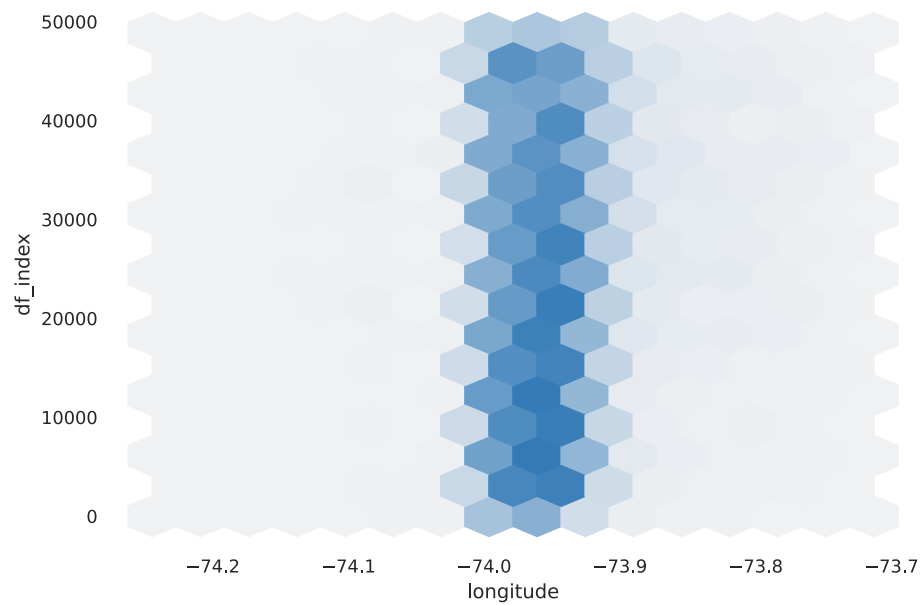


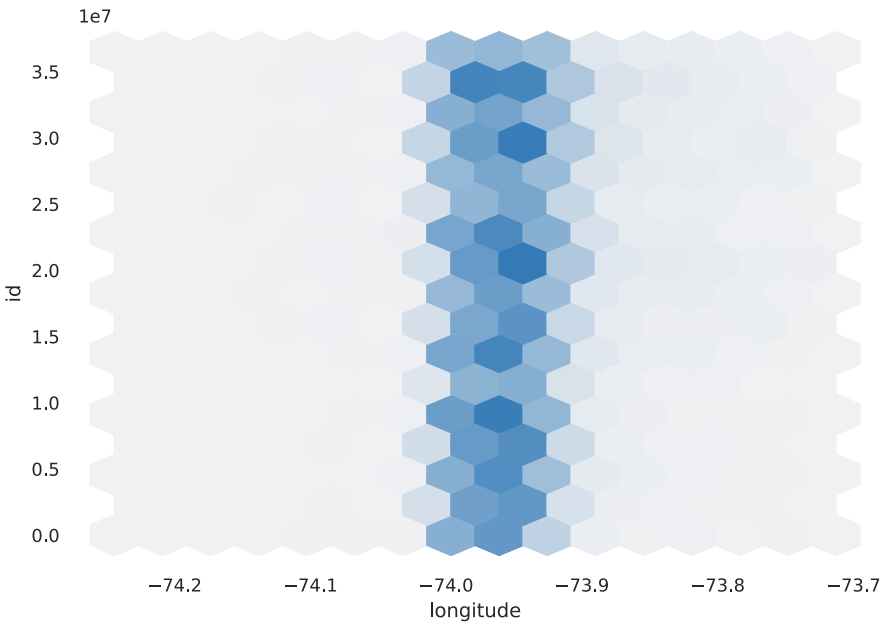


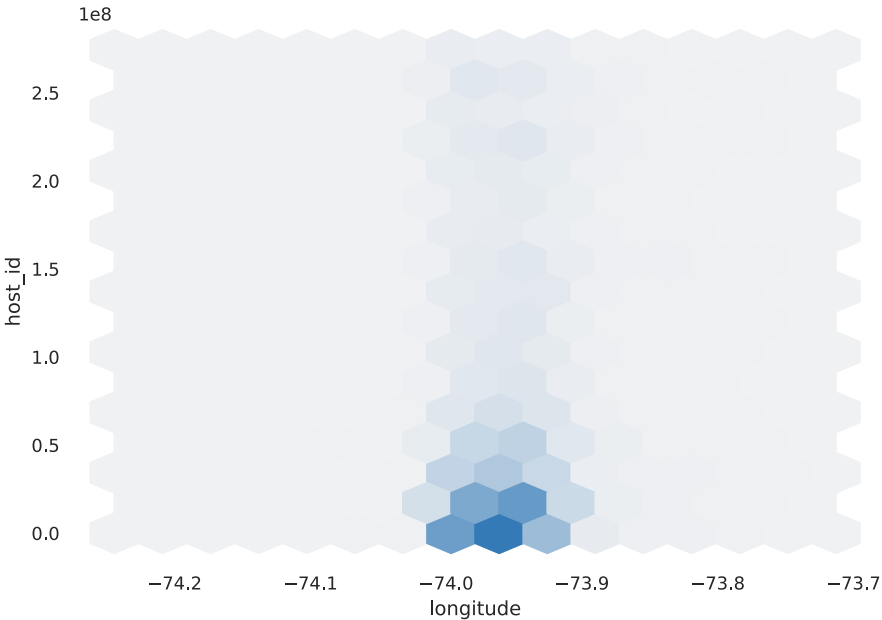


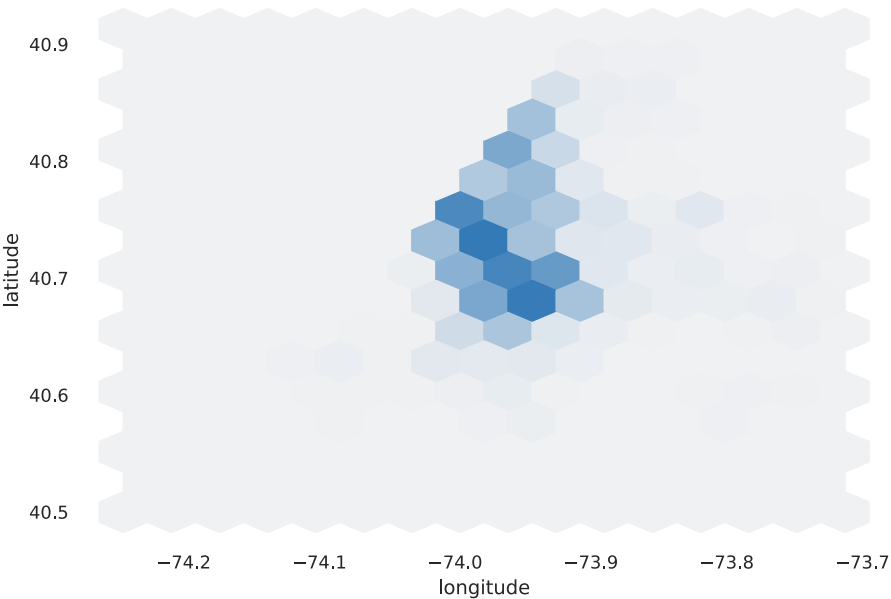
- availability_365
- df_index
- id
- host_id
- latitude
- longitude
- price
- minimum_nights
- number_of_reviews
- reviews_per_month
- calculated_host_listing:

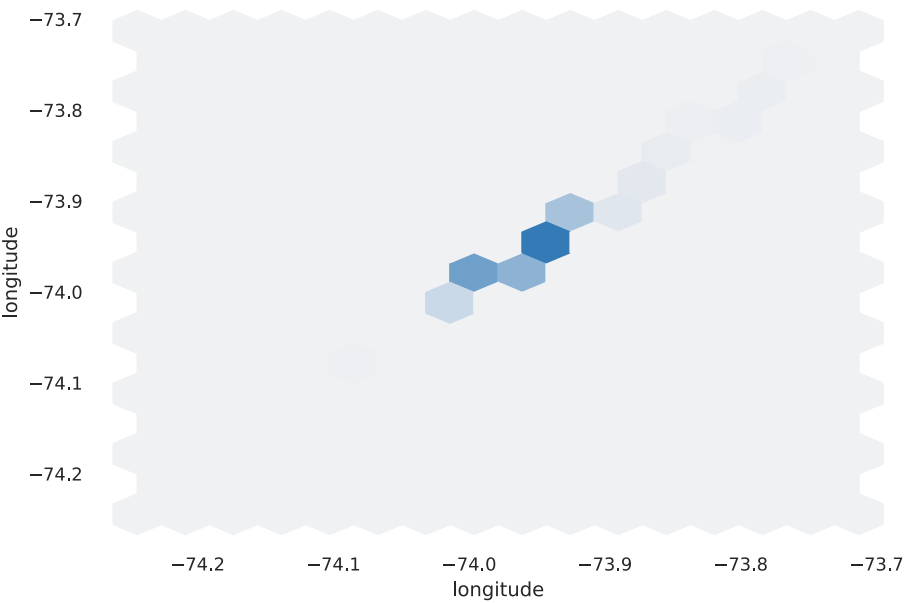


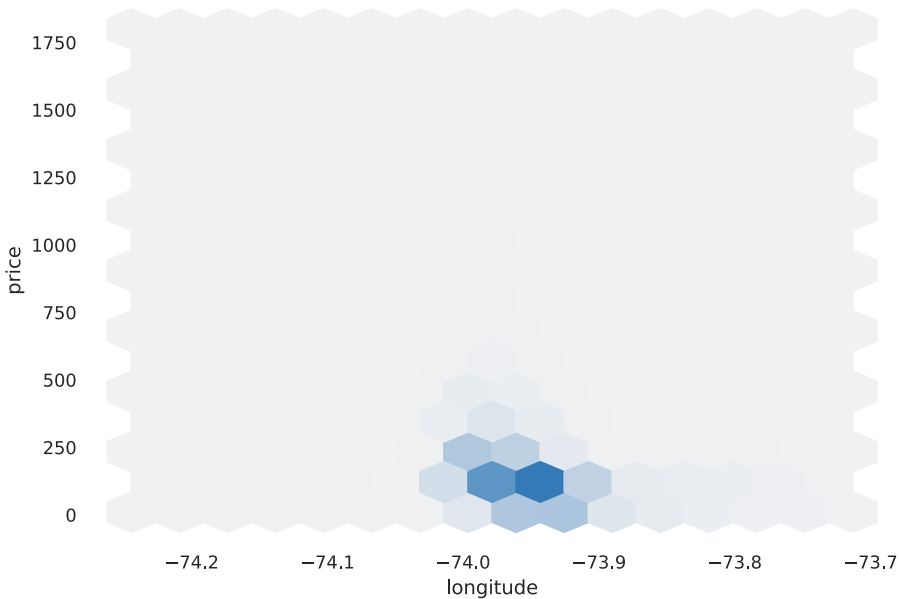


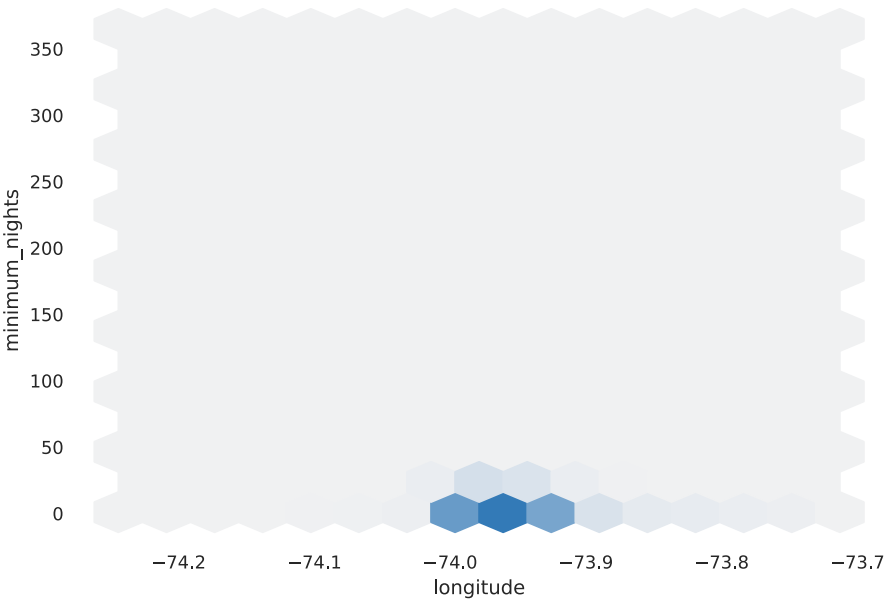


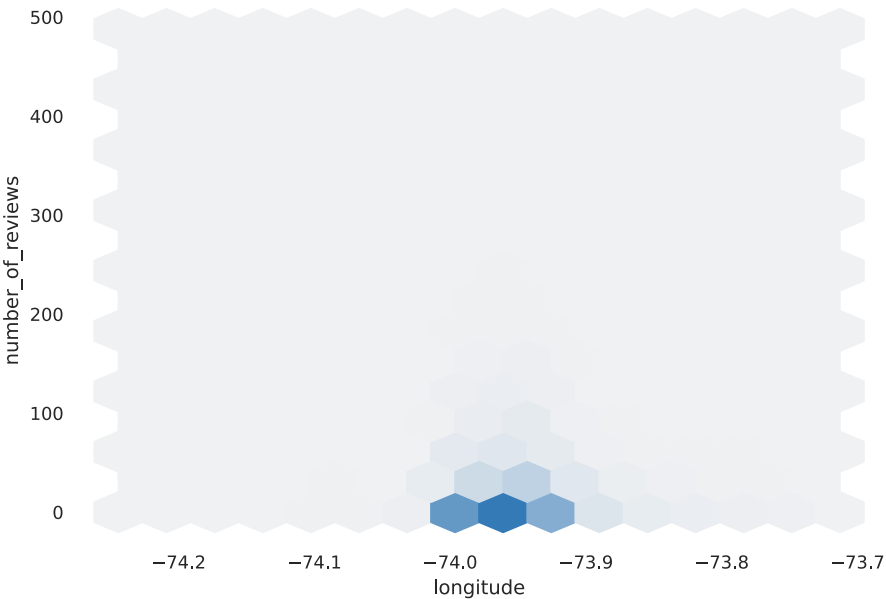


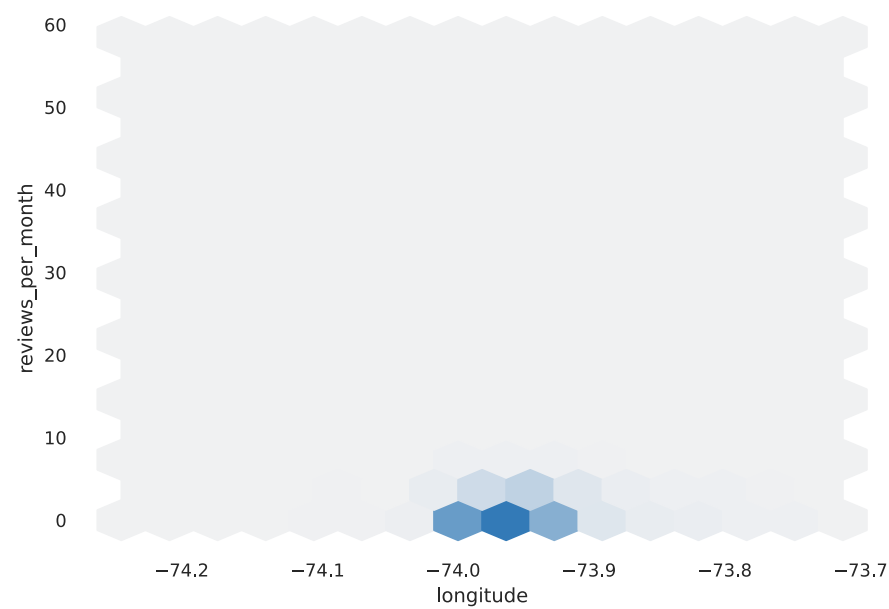


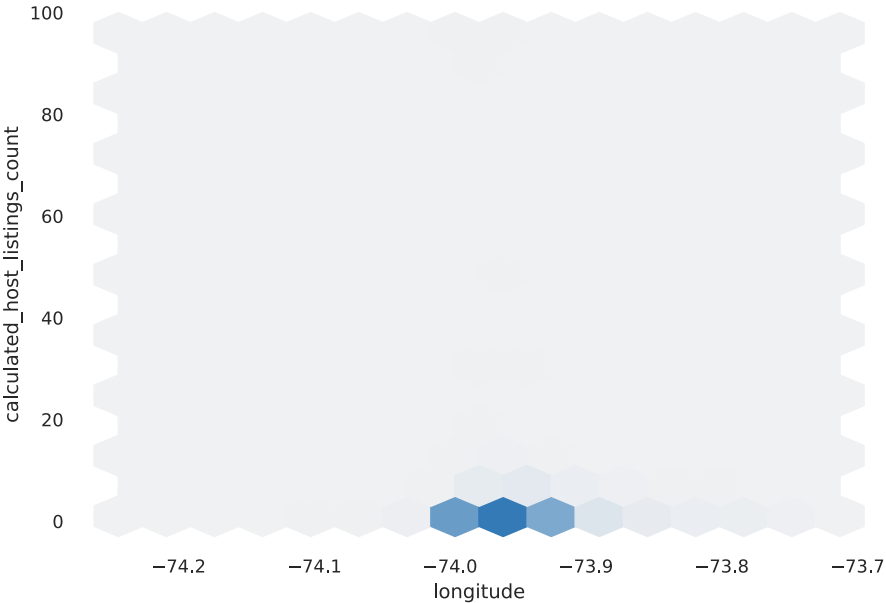












- availability_365
- df_index
- id
- host_id
- latitude
- longitude
- price
- minimum_nights
- number_of_reviews
- reviews_per_month
- calculated_host_listing:

