

Overview

Brought to you by YData (https://ydata.ai/?utm_source=opensource&utm_medium=ydataprofiling&utm_campaign=report)

Dataset statistics

Number of variables	18
Number of observations	37765
Missing cells	70824
Missing cells (%)	10.4%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	5.2 MiB
Average record size in memory	144.0 B

Variable types

Numeric	11
Text	4
Categorical	2
DateTime	1

Alerts

host_id is highly overall correlated with id	High correlation
id is highly overall correlated with host_id	High correlation
latitude is highly overall correlated with neighbourhood_group	High correlation
longitude is highly overall correlated with neighbourhood_group	High correlation
neighbourhood_group is highly overall correlated with latitude and 1 other fields (latitude, longitude)	High correlation
number_of_reviews is highly overall correlated with number_of_reviews_ltm and 1 other fields (number_of_reviews_ltm, reviews_per_month)	High correlation
number_of_reviews_ltm is highly overall correlated with number_of_reviews and 1 other fields (number_of_reviews, reviews_per_month)	High correlation
reviews_per_month is highly overall correlated with number_of_reviews and 1 other fields (number_of_reviews, number_of_reviews_ltm)	High correlation
price has 14721 (39.0%) missing values	Missing
last_review has 11751 (31.1%) missing values	Missing
reviews_per_month has 11751 (31.1%) missing values	Missing
license has 32594 (86.3%) missing values	Missing
price is highly skewed ($\gamma_1 = 99.07291859$)	Skewed
number_of_reviews_ltm is highly skewed ($\gamma_1 = 40.57284443$)	Skewed
id has unique values	Unique
number_of_reviews has 11751 (31.1%) zeros	Zeros
availability_365 has 12546 (33.2%) zeros	Zeros
number_of_reviews_ltm has 23090 (61.1%) zeros	Zeros

Reproduction

Analysis started	2024-09-14 23:42:00.299602
Analysis finished	2024-09-14 23:42:06.205034
Duration	5.91 seconds
Software version	ydata-profiling vv4.10.0 (https://github.com/ydataai/ydata-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22NYC%20Airbnb%20Data%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3

Variables

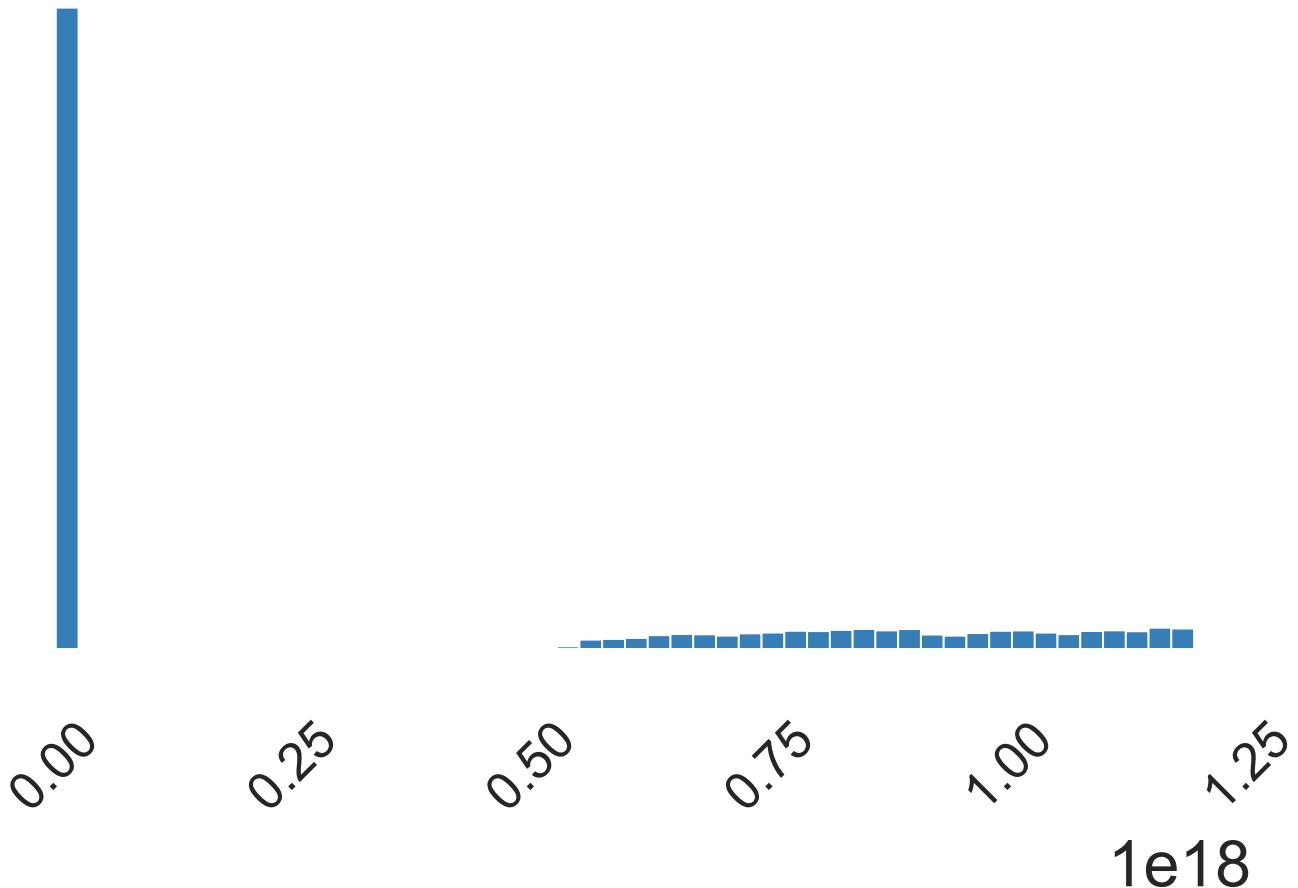
Select Columns 

id

Real number (\mathbb{R})

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

Distinct	37765
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.6530375×10^{17}
Minimum	2595
Maximum	1.1930861×10^{18}
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB



Quantile statistics

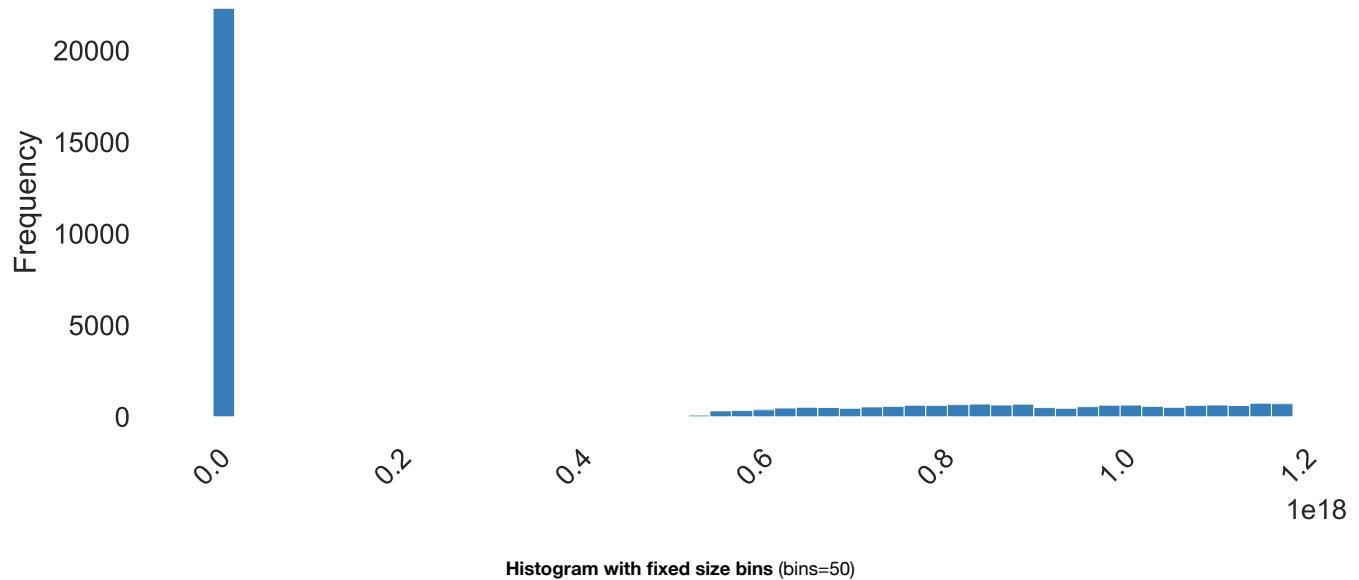
Minimum	2595
5-th percentile	3468623
Q1	20552544
median	48263817
Q3	8.2763805×10^{17}
95-th percentile	1.1282097×10^{18}
Maximum	1.1930861×10^{18}
Range	1.1930861×10^{18}
Interquartile range (IQR)	8.2763805×10^{17}

Descriptive statistics

Standard deviation	4.5399076×10^{17}
Coefficient of variation (CV)	1.2427761
Kurtosis	-1.467409
Mean	3.6530375×10^{17}
Median Absolute Deviation (MAD)	41674870
Skewness	0.55970973
Sum	$-2.4684305 \times 10^{18}$
Variance	2.0610761×10^{35}

Monotonicity

Strictly increasing



Value	Count	Frequency (%)
2595	1	< 0.1%
$7.006563344 \times 10^{17}$	1	< 0.1%
$6.998975458 \times 10^{17}$	1	< 0.1%
$6.999004723 \times 10^{17}$	1	< 0.1%
$6.999140631 \times 10^{17}$	1	< 0.1%
$7.000008124 \times 10^{17}$	1	< 0.1%
$7.000031192 \times 10^{17}$	1	< 0.1%
$7.000647756 \times 10^{17}$	1	< 0.1%
$7.000884189 \times 10^{17}$	1	< 0.1%
$7.001112691 \times 10^{17}$	1	< 0.1%
Other values (37755)	37755	> 99.9%

Value	Count	Frequency (%)
2595	1	< 0.1%
5136	1	< 0.1%
6848	1	< 0.1%
6872	1	< 0.1%
6990	1	< 0.1%
7064	1	< 0.1%
7097	1	< 0.1%
7801	1	< 0.1%
8490	1	< 0.1%
9357	1	< 0.1%

Value	Count	Frequency (%)
$1.193086059 \times 10^{18}$	1	< 0.1%
$1.193068036 \times 10^{18}$	1	< 0.1%
$1.19301613 \times 10^{18}$	1	< 0.1%
$1.193005543 \times 10^{18}$	1	< 0.1%
$1.192946306 \times 10^{18}$	1	< 0.1%
$1.192903502 \times 10^{18}$	1	< 0.1%
$1.192899751 \times 10^{18}$	1	< 0.1%
$1.192828516 \times 10^{18}$	1	< 0.1%
$1.192788528 \times 10^{18}$	1	< 0.1%
$1.192723875 \times 10^{18}$	1	< 0.1%

Unique

Unique	35175	?
Unique (%)	93.1%	

Sample

1st row	Skylit Midtown Castle
2nd row	Spacious Family Friendly Duplex w/ Patio + Yard
3rd row	Only 2 stops to Manhattan studio
4th row	Uptown Sanctuary w/ Private Bath (Month to Month)
5th row	UES Beautiful Blue Room

Most occurring characters

Value	Count	Frequency (%)
	195249	14.0%
e	101097	7.3%
o	96181	6.9%
t	80884	5.8%
a	78776	5.7%
r	76154	5.5%
n	73050	5.3%
i	72943	5.2%
l	40348	2.9%
s	37983	2.7%
Other values (708)	537660	38.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1390325	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
e	195249	14.0%
o	101097	7.3%
t	96181	6.9%
a	80884	5.8%
r	78776	5.7%
n	76154	5.5%
i	73050	5.3%
l	72943	5.2%
s	40348	2.9%
Other values (708)	37983	2.7%
	537660	38.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	1390325	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
e	195249	14.0%
o	101097	7.3%
t	96181	6.9%
a	80884	5.8%
r	78776	5.7%
n	76154	5.5%
i	73050	5.3%
l	72943	5.2%
s	40348	2.9%
Other values (708)	37983	2.7%
	537660	38.7%

Most occurring blocks

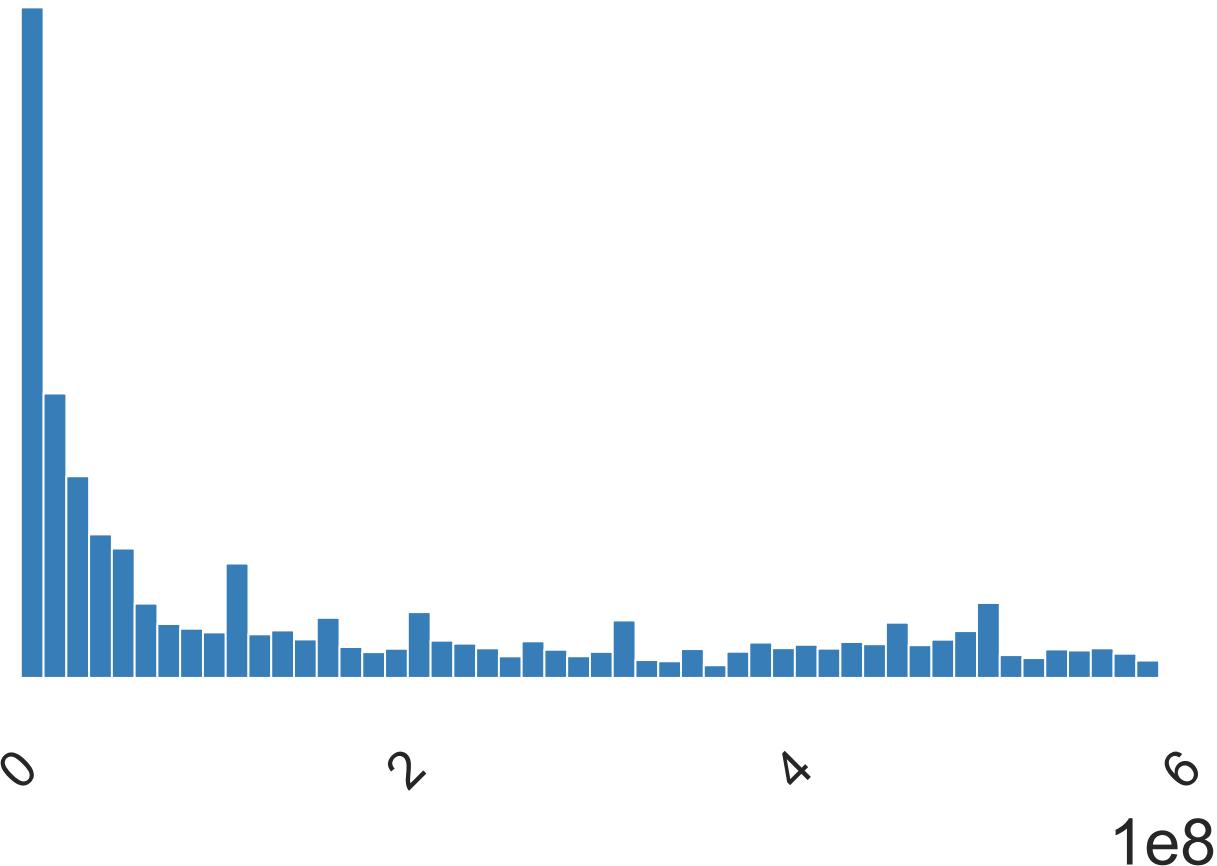
Value	Count	Frequency (%)
(unknown)	1390325	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
e	195249	14.0%
o	101097	7.3%
t	96181	6.9%
a	80884	5.8%
r	78776	5.7%
n	76154	5.5%
i	73050	5.3%
l	72943	5.2%
s	40348	2.9%
Other values (708)	37983	2.7%
	537660	38.7%

host_idReal number (\mathbb{R})**HIGH CORRELATION** (This variable has a high overall correlation with 1 fields: `id`)

Distinct	22663
Distinct (%)	60.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.6596842×10^8
Minimum	1678
Maximum	5.8691743×10^8
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

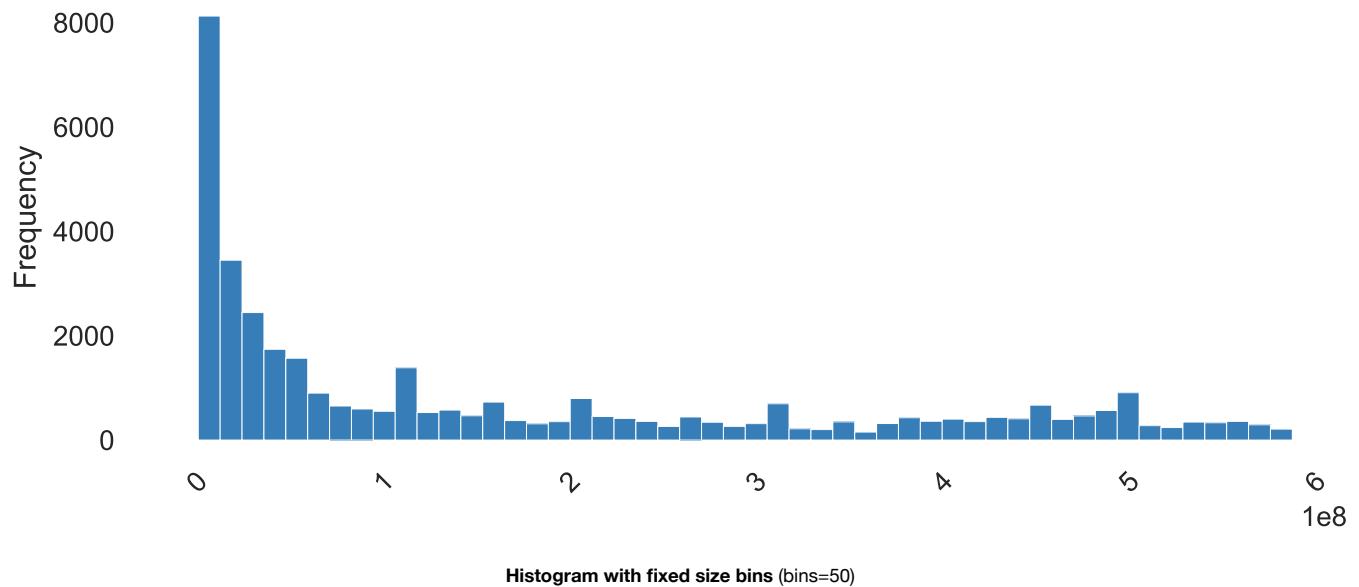


Quantile statistics

Minimum	1678
5-th percentile	1472922
Q1	16627758
median	82189528
Q3	3.0315693×10^8
95-th percentile	5.1260944×10^8
Maximum	5.8691743×10^8
Range	5.8691575×10^8
Interquartile range (IQR)	2.8652917×10^8

Descriptive statistics

Standard deviation	1.8005294×10^8
Coefficient of variation (CV)	1.0848626
Kurtosis	-0.67312507
Mean	1.6596842×10^8
Median Absolute Deviation (MAD)	78457267
Skewness	0.87760849
Sum	6.2677973×10^{12}
Variance	3.241906×10^{16}
Monotonicity	Not monotonic



Value	Count	Frequency (%)
107434423	842	2.2%
3223938	692	1.8%
305240193	351	0.9%
446820235	329	0.9%
19303369	262	0.7%
200239515	246	0.7%
501999278	245	0.6%
204704622	226	0.6%
162280872	212	0.6%
501499086	170	0.5%
Other values (22653)	34190	90.5%

Value	Count	Frequency (%)
1678	1	< 0.1%
2234	1	< 0.1%
2438	1	< 0.1%
2571	1	< 0.1%
2787	3	< 0.1%
2845	3	< 0.1%
2868	1	< 0.1%
3869	1	< 0.1%
4396	2	< 0.1%
5089	1	< 0.1%

Value	Count	Frequency (%)
586917430	2	< 0.1%
586900449	1	< 0.1%
586895343	1	< 0.1%
586836710	3	< 0.1%
586444653	1	< 0.1%
586424140	1	< 0.1%
586339781	1	< 0.1%
586297193	1	< 0.1%
585984692	1	< 0.1%
585973759	1	< 0.1%

host_name

Text

Distinct	8534
Distinct (%)	22.6%
Missing	5
Missing (%)	< 0.1%
Memory size	295.2 KiB

**Length**

Max length	45
Median length	42
Mean length	6.8542638
Min length	1

Characters and Unicode

Total characters	258817
Distinct characters	140
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	4967	?
Unique (%)	13.2%	

Sample

1st row	Jennifer
2nd row	Rebecca
3rd row	Allen & Irina
4th row	Kae
5th row	Cyn

Value	Count	Frequency (%)
blueground	842	1.9%
eugene	698	1.6%
roompicks	498	1.1%
furnished	401	0.9%
elena	390	0.9%
michael	368	0.8%
and	358	0.8%
kristina	354	0.8%
hotel	330	0.8%
luxurybookingsfze	329	0.8%
Other values (8060)	39149	89.6%



Most occurring characters

Value	Count	Frequency (%)
a	27528	10.6%
e	24449	9.4%
n	20030	7.7%
i	19809	7.7%
r	15310	5.9%
o	14266	5.5%
l	12504	4.8%
s	9269	3.6%
t	8407	3.2%
u	8345	3.2%
Other values (130)	98900	38.2%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	258817	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
a	27528	10.6%
e	24449	9.4%
n	20030	7.7%
i	19809	7.7%
r	15310	5.9%
o	14266	5.5%
l	12504	4.8%
s	9269	3.6%
t	8407	3.2%
u	8345	3.2%
Other values (130)	98900	38.2%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	258817	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
a	27528	10.6%
e	24449	9.4%
n	20030	7.7%
i	19809	7.7%
r	15310	5.9%
o	14266	5.5%
l	12504	4.8%
s	9269	3.6%
t	8407	3.2%
u	8345	3.2%
Other values (130)	98900	38.2%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	258817	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
a	27528	10.6%
e	24449	9.4%
n	20030	7.7%
i	19809	7.7%
r	15310	5.9%
o	14266	5.5%
l	12504	4.8%
s	9269	3.6%
t	8407	3.2%
u	8345	3.2%
Other values (130)	98900	38.2%

neighbourhood_group

Categorical

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

Distinct	5
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	295.2 KiB

Length

Max length	13
Median length	9
Mean length	8.0972064
Min length	5

Characters and Unicode

Total characters	305791	
Distinct characters	20	
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

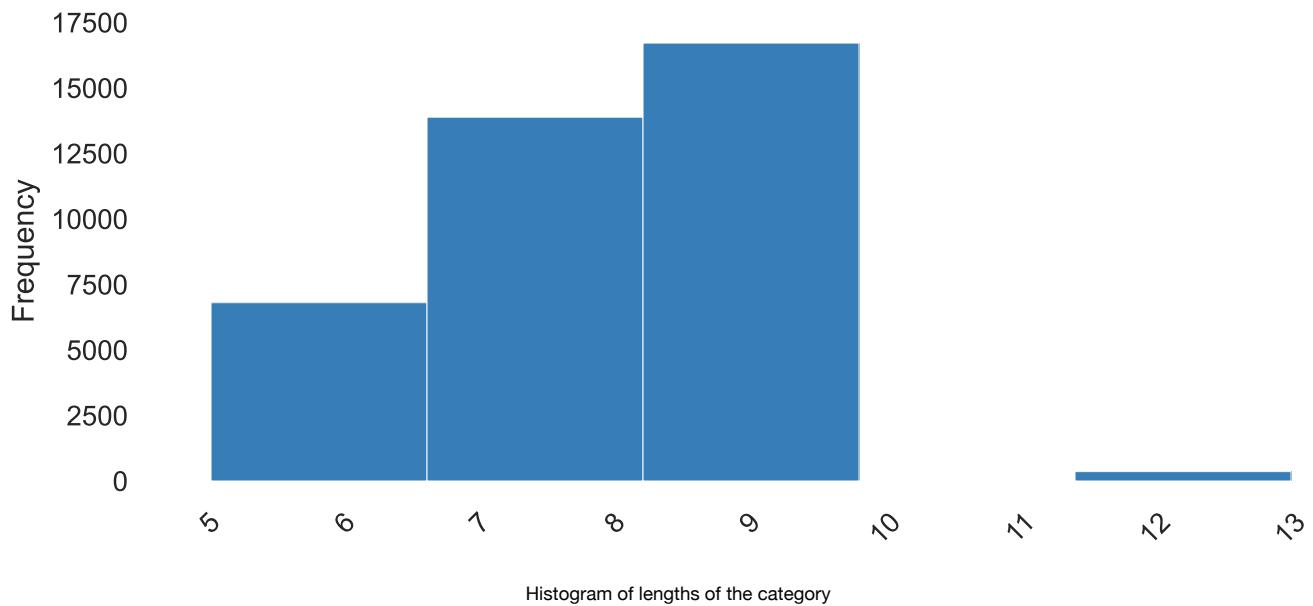
Unique	0	?
Unique (%)	0.0%	?

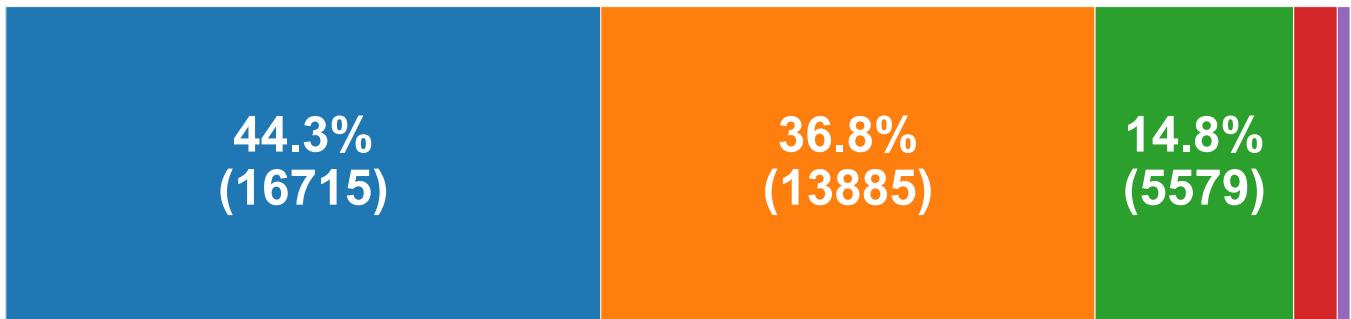
Sample

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

Common Values

Value	Count	Frequency (%)
Manhattan	16715	44.3%
Brooklyn	13885	36.8%
Queens	5579	14.8%
Bronx	1227	3.2%
Staten Island	359	1.0%

Length**Common Values (Plot)**



Value	Count	Frequency (%)
manhattan	16715	43.8%
brooklyn	13885	36.4%
queens	5579	14.6%
bronx	1227	3.2%
staten	359	0.9%
island	359	0.9%

Most occurring characters

Value	Count	Frequency (%)
n	54839	17.9%
a	50863	16.6%
t	34148	11.2%
o	28997	9.5%
M	16715	5.5%
h	16715	5.5%
B	15112	4.9%
r	15112	4.9%
l	14244	4.7%
y	13885	4.5%
Other values (10)	45161	14.8%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	305791	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
n	54839	17.9%
a	50863	16.6%
t	34148	11.2%
o	28997	9.5%
M	16715	5.5%
h	16715	5.5%
B	15112	4.9%
r	15112	4.9%
l	14244	4.7%
y	13885	4.5%
Other values (10)	45161	14.8%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	305791	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
n	54839	17.9%
a	50863	16.6%
t	34148	11.2%
o	28997	9.5%
M	16715	5.5%
h	16715	5.5%
B	15112	4.9%
r	15112	4.9%
l	14244	4.7%
y	13885	4.5%
Other values (10)	45161	14.8%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	305791	100.0%

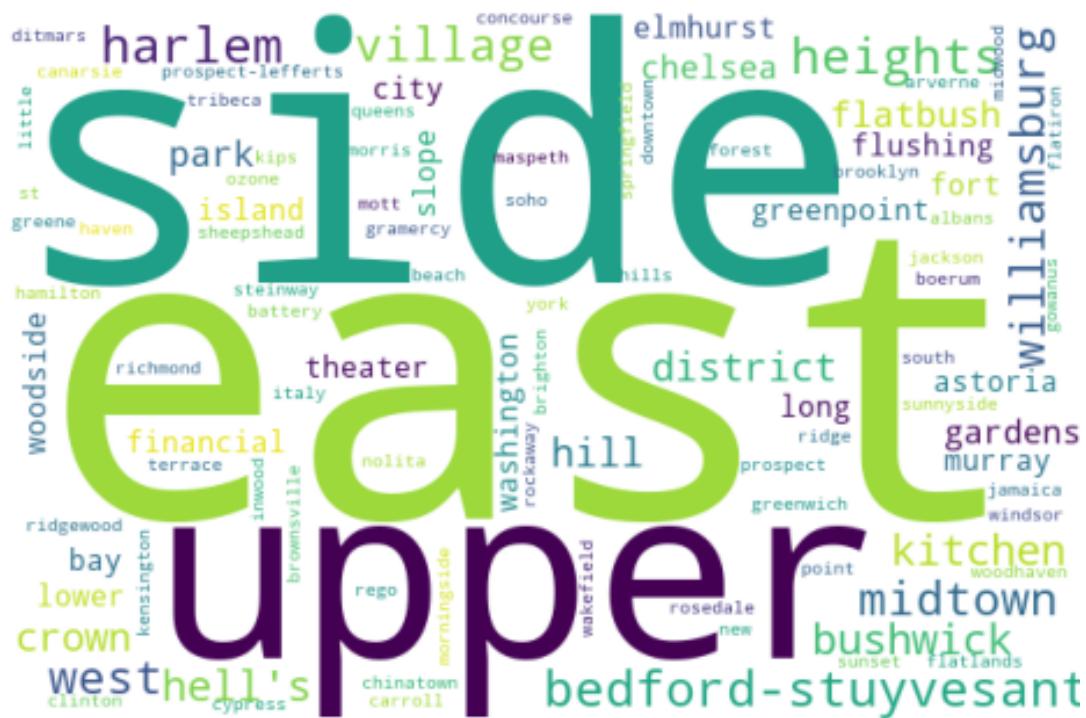
Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
n	54839	17.9%
a	50863	16.6%
t	34148	11.2%
o	28997	9.5%
M	16715	5.5%
h	16715	5.5%
B	15112	4.9%
r	15112	4.9%
l	14244	4.7%
y	13885	4.5%
Other values (10)	45161	14.8%

neighbourhood

Text

Distinct	225
Distinct (%)	0.6%
Missing	0
Missing (%)	0.0%
Memory size	295.2 KiB



Length

Max length	26
Median length	18
Mean length	11.788852
Min length	4

Characters and Unicode

Total characters	445206
Distinct characters	54
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	8	?
Unique (%)	< 0.1%	

Sample

1st row	Midtown
2nd row	Sunset Park
3rd row	Williamsburg
4th row	East Harlem
5th row	East Harlem

Value	Count	Frequency (%)
east	4828	7.9%
side	3605	5.9%
upper	2992	4.9%
bedford-stuyvesant	2709	4.4%
heights	2600	4.2%
harlem	2467	4.0%
williamsburg	2206	3.6%
west	2044	3.3%
midtown	2028	3.3%
village	1925	3.1%
Other values (236)	33973	55.4%



Most occurring characters

Value	Count	Frequency (%)
e	40581	9.1%
i	31904	7.2%
t	30777	6.9%
s	30140	6.8%
a	28615	6.4%
r	25285	5.7%
l	24743	5.6%
n	23612	5.3%
o	20714	4.7%
Other values (44)	19275	4.3%
	169560	38.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	445206	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
e	40581	9.1%
i	31904	7.2%
t	30777	6.9%
s	30140	6.8%
a	28615	6.4%
r	25285	5.7%
l	24743	5.6%
	23612	5.3%
n	20714	4.7%
o	19275	4.3%
Other values (44)	169560	38.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	445206	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
e	40581	9.1%
i	31904	7.2%
t	30777	6.9%
s	30140	6.8%
a	28615	6.4%
r	25285	5.7%
l	24743	5.6%
	23612	5.3%
n	20714	4.7%
o	19275	4.3%
Other values (44)	169560	38.1%

Most occurring blocks

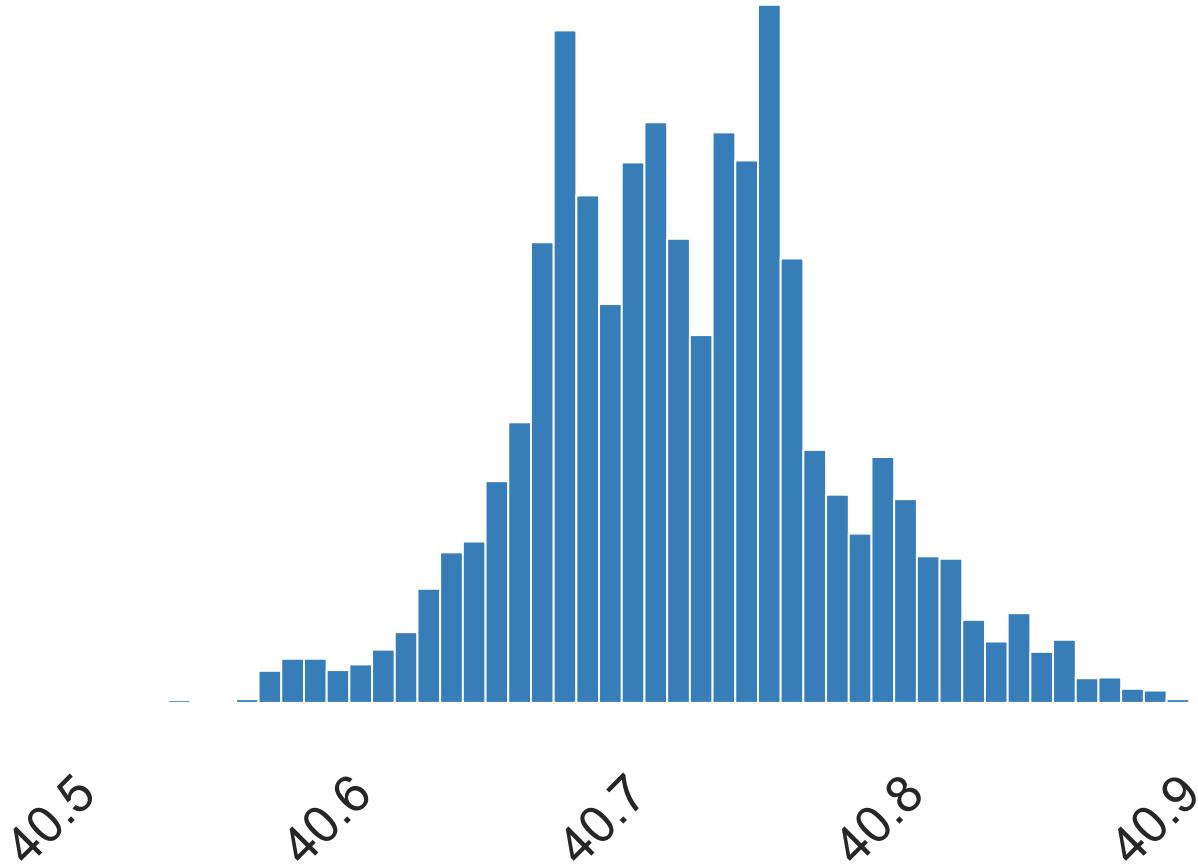
Value	Count	Frequency (%)
(unknown)	445206	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
e	40581	9.1%
i	31904	7.2%
t	30777	6.9%
s	30140	6.8%
a	28615	6.4%
r	25285	5.7%
l	24743	5.6%
	23612	5.3%
n	20714	4.7%
o	19275	4.3%
Other values (44)	169560	38.1%

latitudeReal number (\mathbb{R})HIGH CORRELATION (This variable has a high overall correlation with 1 fields: neighbourhood_group)

Distinct	23546
Distinct (%)	62.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	40.729075
Minimum	40.500366
Maximum	40.91139
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

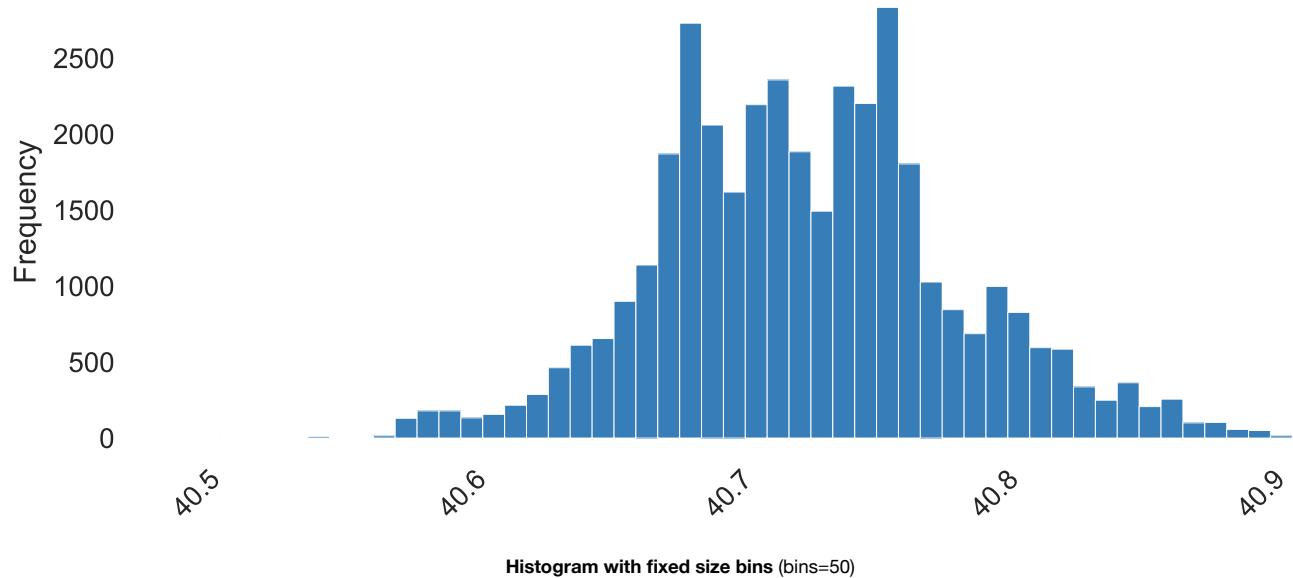


Quantile statistics

Minimum	40.500366
5-th percentile	40.641248
Q1	40.6888
median	40.72644
Q3	40.76247
95-th percentile	40.826338
Maximum	40.91139
Range	0.4110237
Interquartile range (IQR)	0.07367

Descriptive statistics

Standard deviation	0.05634298
Coefficient of variation (CV)	0.0013833602
Kurtosis	0.1803726
Mean	40.729075
Median Absolute Deviation (MAD)	0.036892
Skewness	0.15110474
Sum	1538133.5
Variance	0.0031745314
Monotonicity	Not monotonic



Value	Count	Frequency (%)
40.76153	54	0.1%
40.61881	46	0.1%
40.8024986	37	0.1%
40.7607614	36	0.1%
40.76411	28	0.1%
40.7092191	27	0.1%
40.7526556	25	0.1%
40.7456565	23	0.1%
40.6789456	23	0.1%
40.7056	23	0.1%
Other values (23536)	37443	99.1%

Value	Count	Frequency (%)
40.5003663	1	< 0.1%
40.50456	1	< 0.1%
40.507114	1	< 0.1%
40.50863	1	< 0.1%
40.51707	1	< 0.1%
40.52034	1	< 0.1%
40.52224	1	< 0.1%
40.52339	1	< 0.1%
40.52442553	1	< 0.1%
40.52498	1	< 0.1%

Value	Count	Frequency (%)
40.91139	1	< 0.1%
40.91138	1	< 0.1%
40.91114684	1	< 0.1%
40.91062	1	< 0.1%
40.90667	1	< 0.1%
40.906568	1	< 0.1%
40.90549	1	< 0.1%
40.90530196	1	< 0.1%
40.90505	1	< 0.1%
40.90421	1	< 0.1%

longitudeReal number (\mathbb{R})HIGH CORRELATION (This variable has a high overall correlation with 1 fields: neighbourhood_group)**Distinct**

21266

Distinct (%)

56.3%

Missing

0

Missing (%)

0.0%

Infinite

0

Infinite (%)

0.0%

Mean

-73.947084

Minimum

-74.251907

Maximum

-73.71365

Zeros

0

Zeros (%)

0.0%

Negative

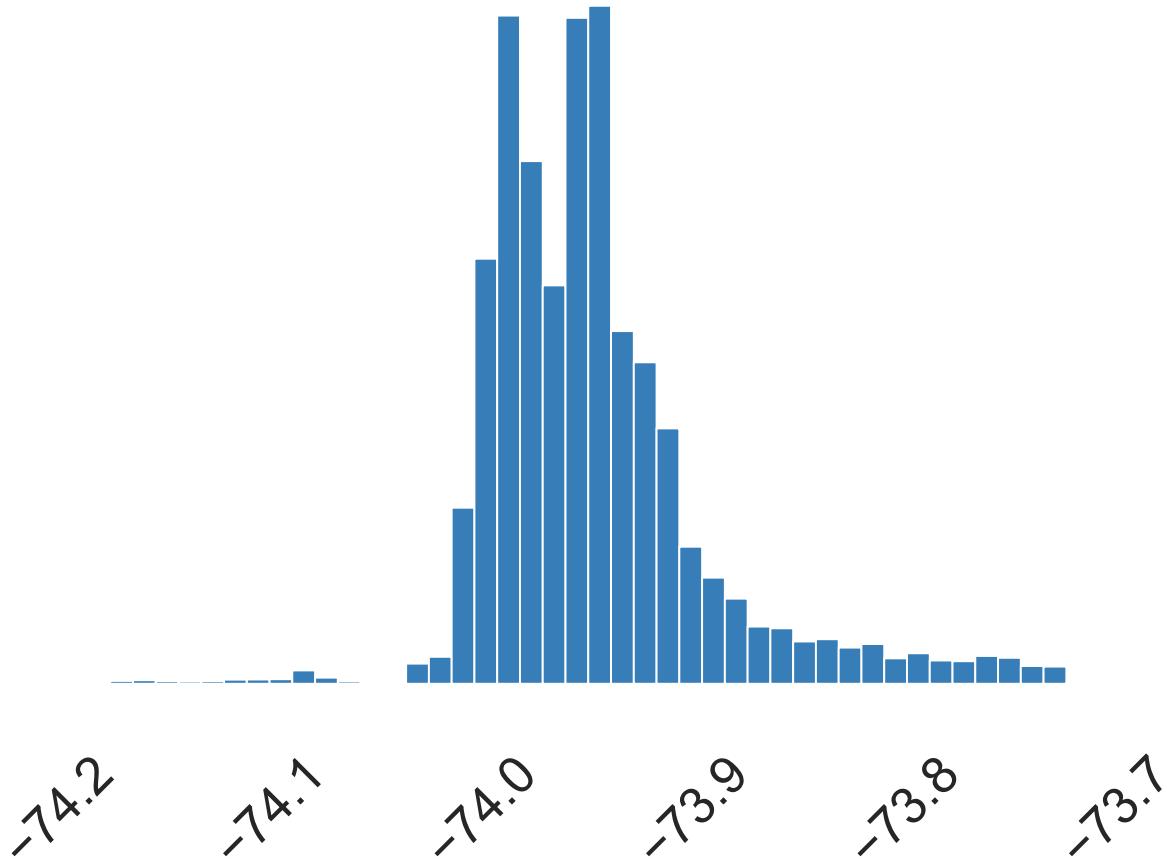
37765

Negative (%)

100.0%

Memory size

295.2 KiB

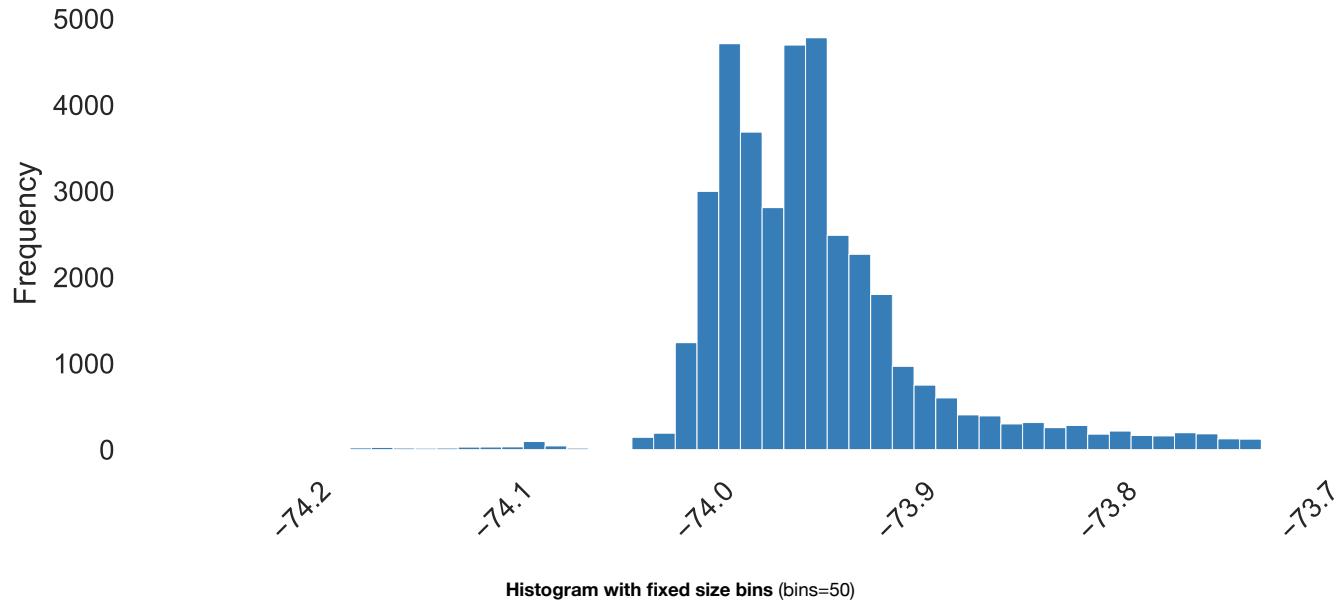


Quantile statistics

Minimum	-74.251907
5-th percentile	-74.004718
Q1	-73.98322
median	-73.95454
Q3	-73.928022
95-th percentile	-73.832298
Maximum	-73.71365
Range	0.538257
Interquartile range (IQR)	0.05519804

Descriptive statistics

Standard deviation	0.054457768
Coefficient of variation (CV)	-0.0007364424
Kurtosis	3.5505265
Mean	-73.947084
Median Absolute Deviation (MAD)	0.0280055
Skewness	1.2156205
Sum	-2792611.6
Variance	0.0029656485
Monotonicity	Not monotonic



Value	Count	Frequency (%)
-73.99878	53	0.1%
-73.9668444	37	0.1%
-73.9861055	36	0.1%
-74.0323824	31	0.1%
-73.99371	29	0.1%
-74.0137534	27	0.1%
-73.9724804	25	0.1%
-73.93061	24	0.1%
-73.9070911	23	0.1%
-73.9437591	23	0.1%
Other values (21256)	37457	99.2%

Value	Count	Frequency (%)
-74.251907	1	< 0.1%
-74.24984	1	< 0.1%
-74.24135	1	< 0.1%
-74.2394006	1	< 0.1%
-74.22263763	1	< 0.1%
-74.21514	1	< 0.1%
-74.20739	1	< 0.1%
-74.20517	1	< 0.1%
-74.20516	1	< 0.1%
-74.20178157	1	< 0.1%

Value	Count	Frequency (%)
-73.71365	1	< 0.1%
-73.71383551	1	< 0.1%
-73.7174099	1	< 0.1%
-73.7198118	1	< 0.1%
-73.71991	1	< 0.1%
-73.72408	1	< 0.1%
-73.72434	1	< 0.1%
-73.72543	1	< 0.1%
-73.72606	1	< 0.1%
-73.72614	1	< 0.1%

room_type

Categorical

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	295.2 KiB

Length

Max length	15
Median length	15
Mean length	13.546723
Min length	10

Characters and Unicode

Total characters	511592
Distinct characters	19
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

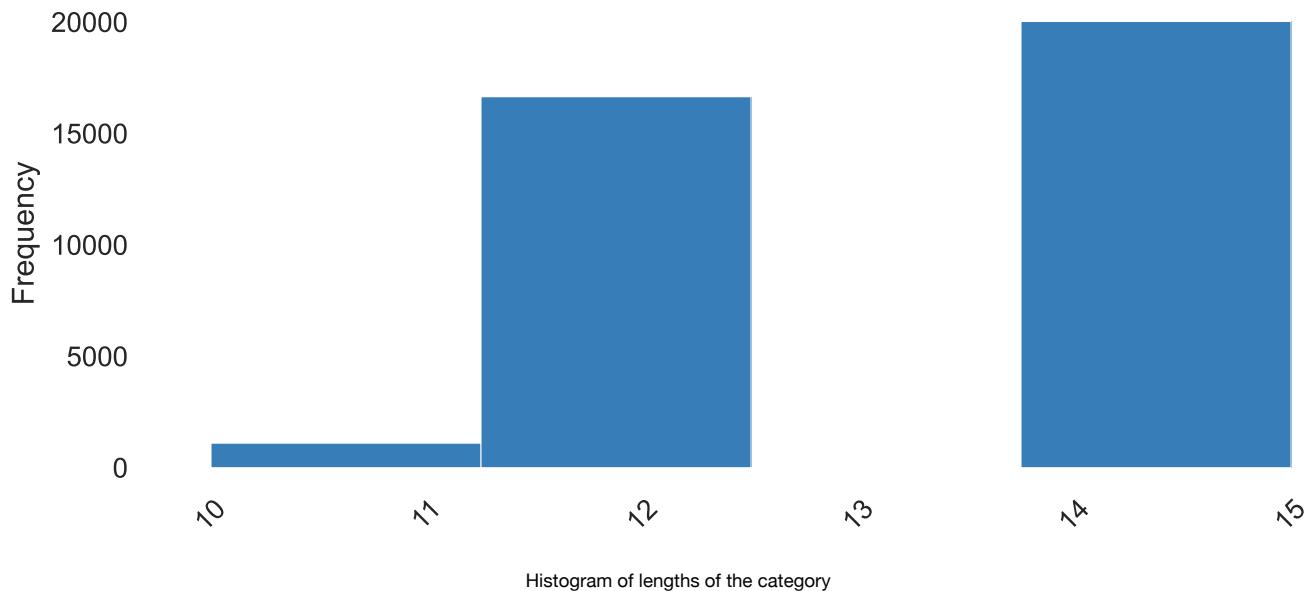
Unique	0	?
Unique (%)	0.0%	?

Sample

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

Common Values

Value	Count	Frequency (%)
Entire home/apt	20023	53.0%
Private room	16645	44.1%
Hotel room	560	1.5%
Shared room	537	1.4%

Length**Common Values (Plot)**

- █ Entire home/apt
- █ Private room
- █ Hotel room
- █ Shared room

Value	Count	Frequency (%)
entire	20023	26.5%
home/apt	20023	26.5%
room	17742	23.5%
private	16645	22.0%
hotel	560	0.7%
shared	537	0.7%

Most occurring characters

Value	Count	Frequency (%)
e	57788	11.3%
t	57251	11.2%
o	56067	11.0%
r	54947	10.7%
m	37765	7.4%
	37765	7.4%
a	37205	7.3%
i	36668	7.2%
h	20560	4.0%
p	20023	3.9%
Other values (9)	95553	18.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	511592	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
e	57788	11.3%
t	57251	11.2%
o	56067	11.0%
r	54947	10.7%
m	37765	7.4%
	37765	7.4%
a	37205	7.3%
i	36668	7.2%
h	20560	4.0%
p	20023	3.9%
Other values (9)	95553	18.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	511592	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
e	57788	11.3%
t	57251	11.2%
o	56067	11.0%
r	54947	10.7%
m	37765	7.4%
	37765	7.4%
a	37205	7.3%
i	36668	7.2%
h	20560	4.0%
p	20023	3.9%
Other values (9)	95553	18.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	511592	100.0%

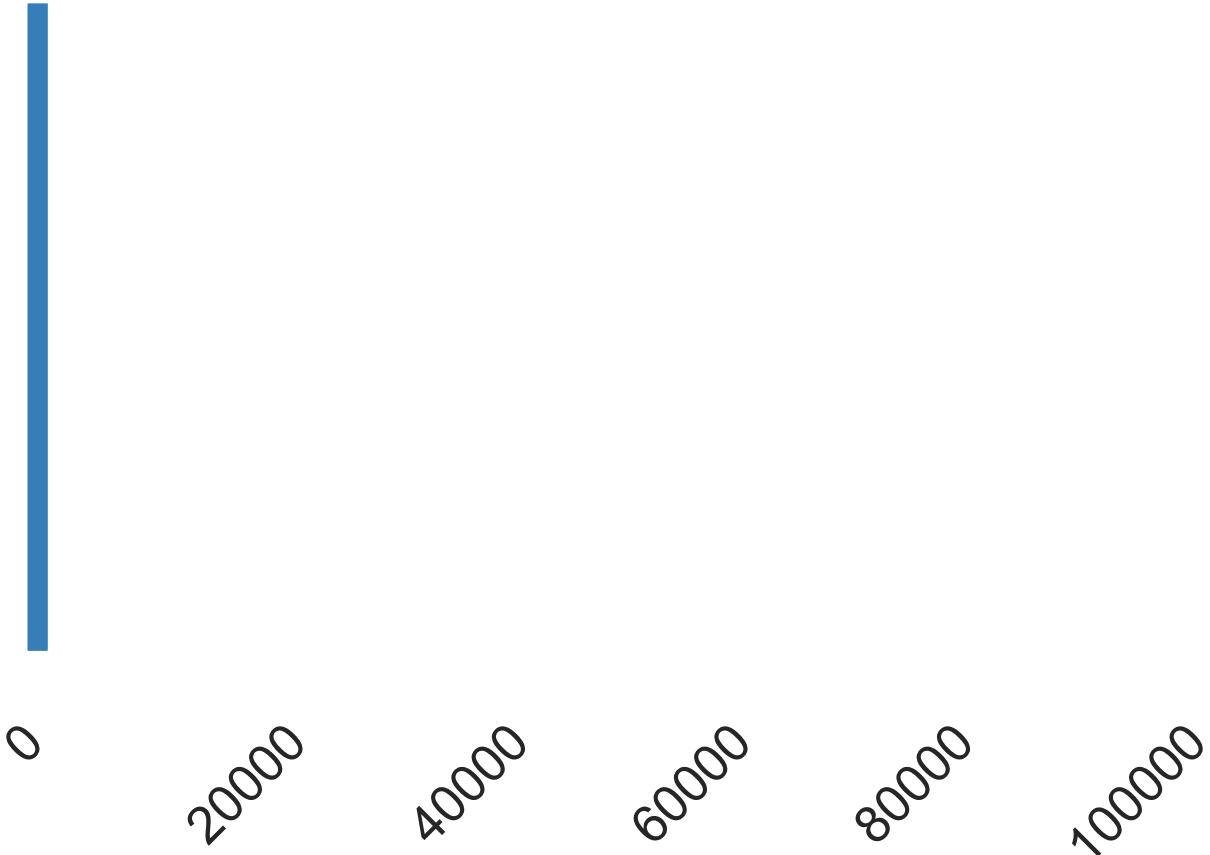
Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
e	57788	11.3%
t	57251	11.2%
o	56067	11.0%
r	54947	10.7%
m	37765	7.4%
	37765	7.4%
a	37205	7.3%
i	36668	7.2%
h	20560	4.0%
p	20023	3.9%
Other values (9)	95553	18.7%

priceReal number (\mathbb{R})

MISSING SKEWED

Distinct	982
Distinct (%)	4.3%
Missing	14721
Missing (%)	39.0%
Infinite	0
Infinite (%)	0.0%
Mean	221.0749
Minimum	8
Maximum	100000
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB



Quantile statistics

Minimum	8
5-th percentile	44
Q1	86
median	150
Q3	250
95-th percentile	550
Maximum	100000
Range	99992
Interquartile range (IQR)	164

Descriptive statistics

Standard deviation	766.62099
Coefficient of variation (CV)	3.467698
Kurtosis	12524.338
Mean	221.0749
Median Absolute Deviation (MAD)	75
Skewness	99.072919
Sum	5094450
Variance	587707.75
Monotonicity	Not monotonic



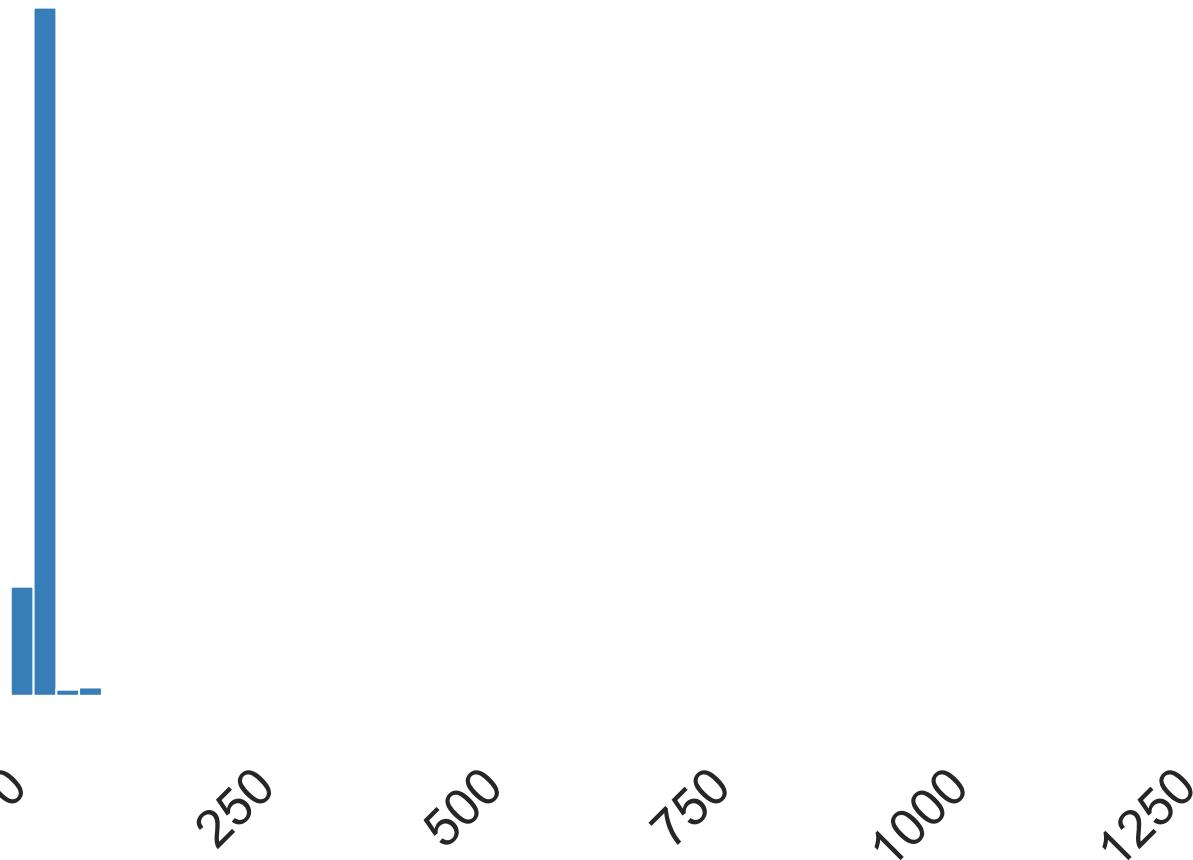
Value	Count	Frequency (%)
150	535	1.4%
200	441	1.2%
100	421	1.1%
120	346	0.9%
250	329	0.9%
125	303	0.8%
60	291	0.8%
80	289	0.8%
90	280	0.7%
70	265	0.7%
Other values (972)	19544	51.8%
(Missing)	14721	39.0%

Value	Count	Frequency (%)
8	1	< 0.1%
10	9	< 0.1%
13	1	< 0.1%
14	1	< 0.1%
15	1	< 0.1%
22	2	< 0.1%
23	2	< 0.1%
24	2	< 0.1%
25	3	< 0.1%
26	1	< 0.1%

Value	Count	Frequency (%)
100000	1	< 0.1%
20000	3	< 0.1%
12452	1	< 0.1%
10799	1	< 0.1%
10000	4	< 0.1%
9999	1	< 0.1%
8550	1	< 0.1%
8000	1	< 0.1%
7600	1	< 0.1%
7500	1	< 0.1%

minimum_nightsReal number (\mathbb{R})

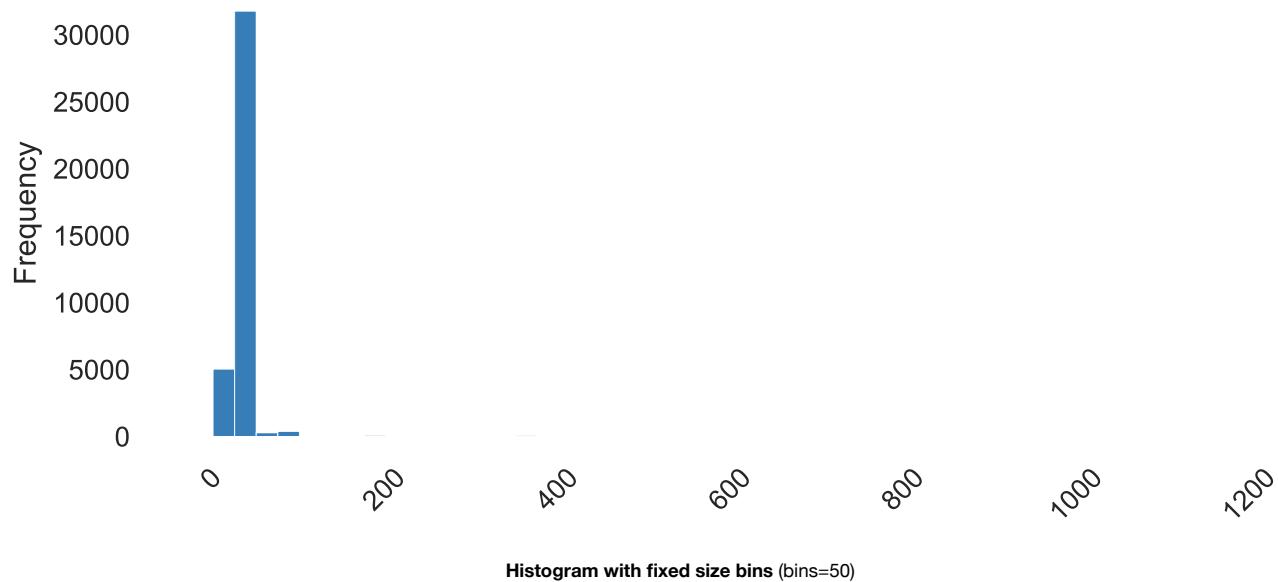
Distinct	118
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	29.164597
Minimum	1
Maximum	1250
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

**Quantile statistics**

Minimum	1
5-th percentile	1
Q1	30
median	30
Q3	30
95-th percentile	31
Maximum	1250
Range	1249
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	30.202538
Coefficient of variation (CV)	1.0355891
Kurtosis	415.97221
Mean	29.164597
Median Absolute Deviation (MAD)	0
Skewness	16.189945
Sum	1101401
Variance	912.19328
Monotonicity	Not monotonic



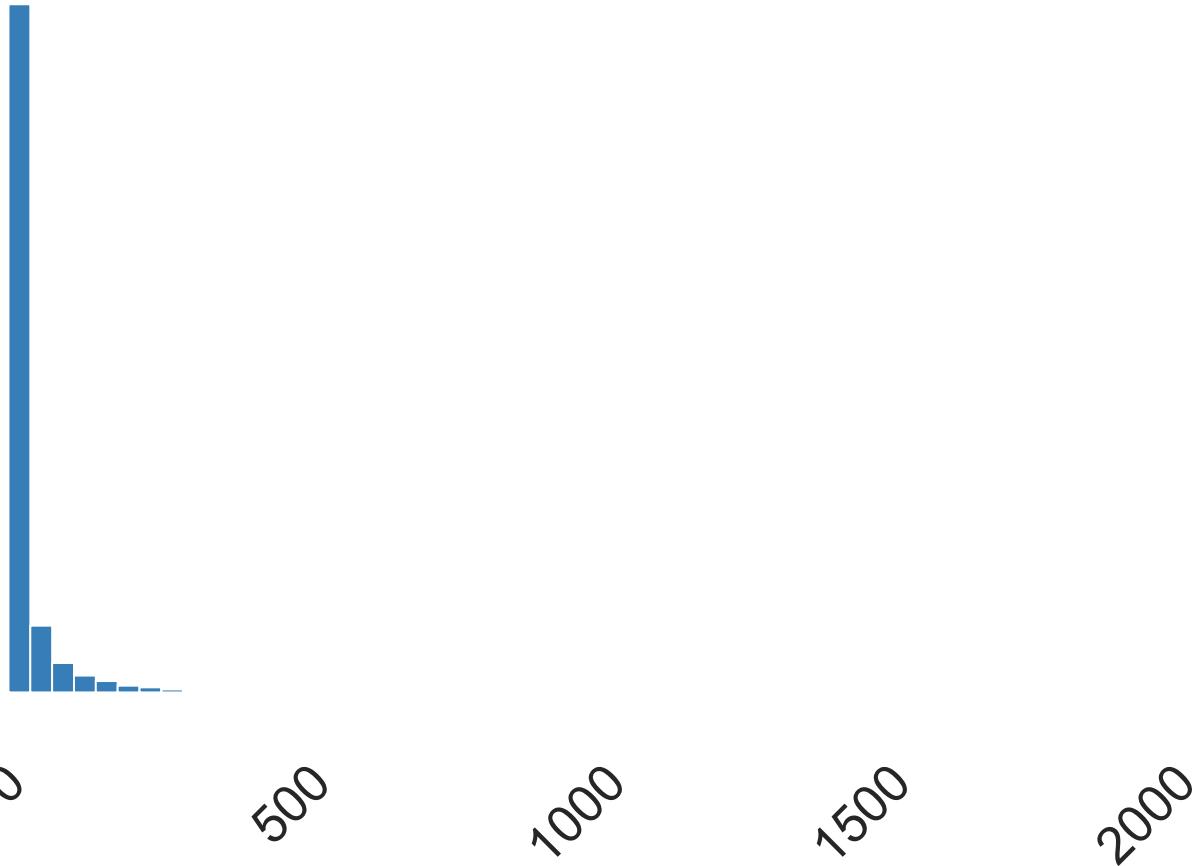
Value	Count	Frequency (%)
30	30448	80.6%
1	2913	7.7%
31	1190	3.2%
2	865	2.3%
3	653	1.7%
90	336	0.9%
4	216	0.6%
60	198	0.5%
5	195	0.5%
7	92	0.2%
Other values (108)	659	1.7%

Value	Count	Frequency (%)
1	2913	7.7%
2	865	2.3%
3	653	1.7%
4	216	0.6%
5	195	0.5%
6	31	0.1%
7	92	0.2%
8	1	< 0.1%
9	3	< 0.1%
10	9	< 0.1%

Value	Count	Frequency (%)
1250	1	< 0.1%
1124	1	< 0.1%
1000	6	< 0.1%
999	1	< 0.1%
729	6	< 0.1%
700	1	< 0.1%
500	6	< 0.1%
480	1	< 0.1%
400	3	< 0.1%
370	1	< 0.1%

number_of_reviewsReal number (\mathbb{R})HIGH CORRELATION (This variable has a high overall correlation with 2 fields: number_of_reviews_ltm, reviews_per_month) ZEROS

Distinct	475
Distinct (%)	1.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	24.898981
Minimum	0
Maximum	1915
Zeros	11751
Zeros (%)	31.1%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

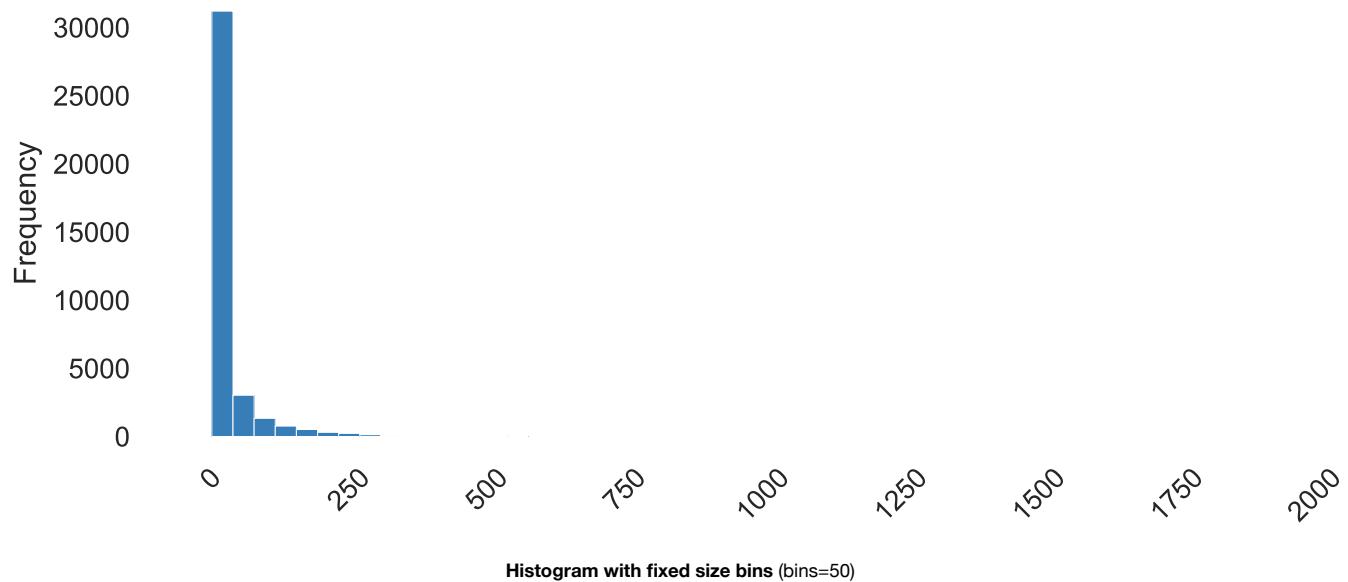


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	3
Q3	21
95-th percentile	128
Maximum	1915
Range	1915
Interquartile range (IQR)	21

Descriptive statistics

Standard deviation	58.802122
Coefficient of variation (CV)	2.3616277
Kurtosis	94.912428
Mean	24.898981
Median Absolute Deviation (MAD)	3
Skewness	6.5506417
Sum	940310
Variance	3457.6895
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	11751	31.1%
1	3586	9.5%
2	2195	5.8%
3	1548	4.1%
4	1216	3.2%
5	973	2.6%
6	848	2.2%
7	761	2.0%
8	677	1.8%
9	592	1.6%
Other values (465)	13618	36.1%

Value	Count	Frequency (%)
0	11751	31.1%
1	3586	9.5%
2	2195	5.8%
3	1548	4.1%
4	1216	3.2%
5	973	2.6%
6	848	2.2%
7	761	2.0%
8	677	1.8%
9	592	1.6%

Value	Count	Frequency (%)
1915	1	< 0.1%
1584	1	< 0.1%
1551	1	< 0.1%
1316	1	< 0.1%
1294	1	< 0.1%
1177	1	< 0.1%
979	1	< 0.1%
964	1	< 0.1%
927	1	< 0.1%
764	1	< 0.1%

last_review

Date

MISSING

Distinct

3059

Distinct (%)

11.8%

Missing

11751

Missing (%)

31.1%

Memory size

295.2 KiB

Minimum 2011-05-12 00:00:00

Maximum 2024-07-05 00:00:00

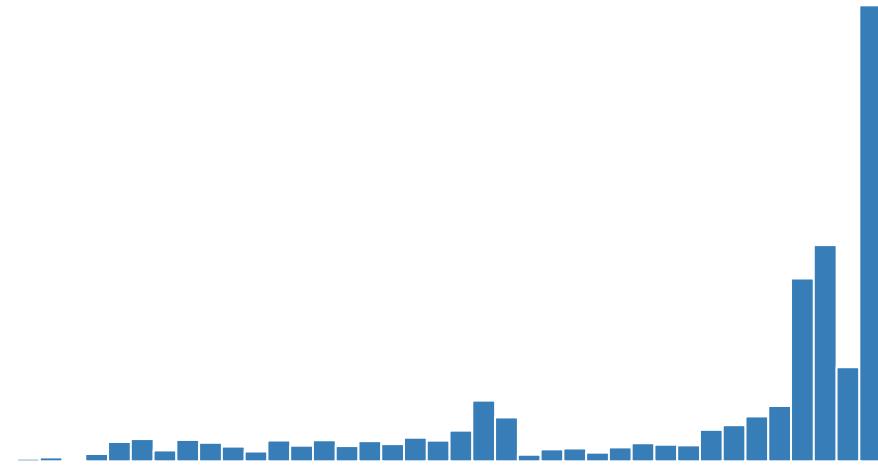
2011-03-13 03:06:40

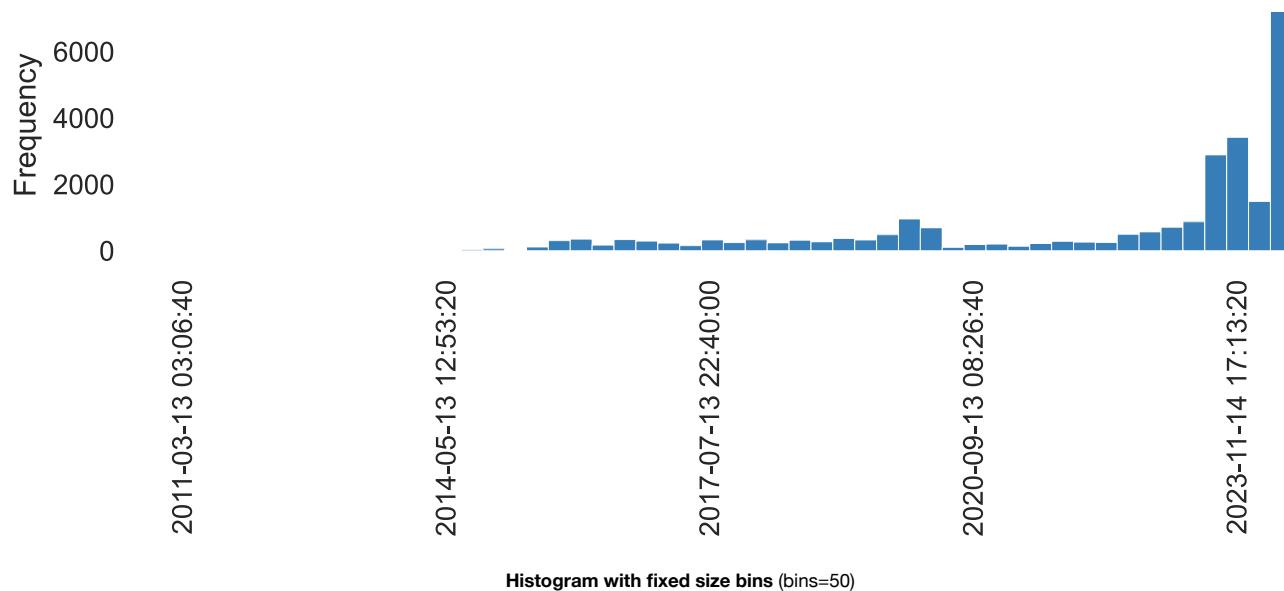
2014-05-13 12:53:20

2017-07-13 22:40:00

2020-09-13 08:26:40

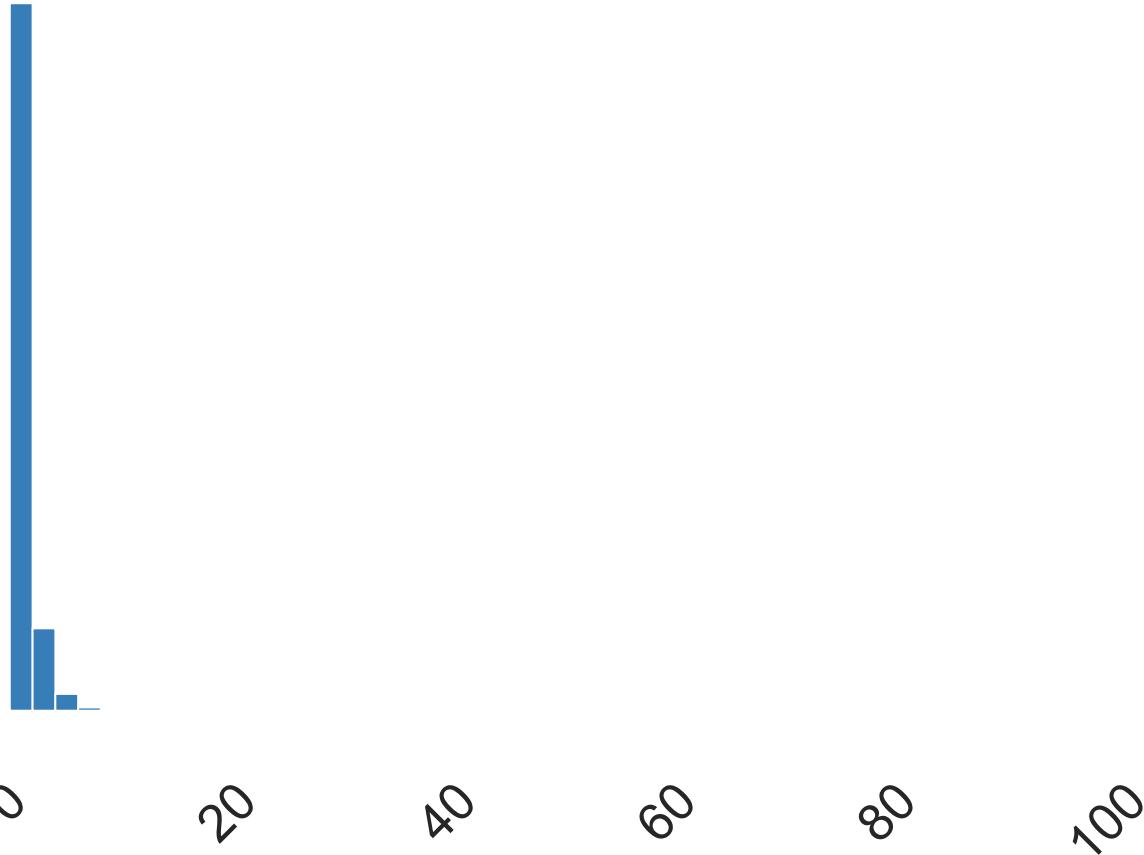
2023-11-14 17:13:20





reviews_per_monthReal number (\mathbb{R})~~HIGH CORRELATION (This variable has a high overall correlation with 2 fields: number_of_reviews, number_of_reviews_ltm). MISSING~~

Distinct	788
Distinct (%)	3.0%
Missing	11751
Missing (%)	31.1%
Infinite	0
Infinite (%)	0.0%
Mean	0.90554394
Minimum	0.01
Maximum	103.53
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

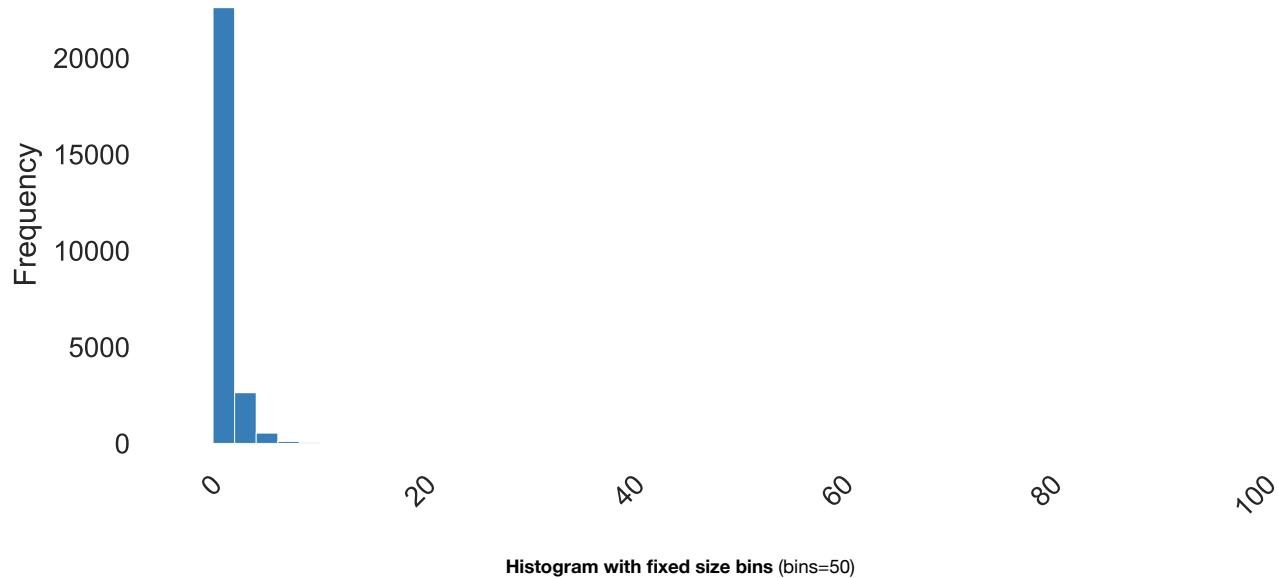


Quantile statistics

Minimum	0.01
5-th percentile	0.02
Q1	0.09
median	0.32
Q3	1.13
95-th percentile	3.38
Maximum	103.53
Range	103.52
Interquartile range (IQR)	1.04

Descriptive statistics

Standard deviation	1.8607836
Coefficient of variation (CV)	2.0548795
Kurtosis	749.32242
Mean	0.90554394
Median Absolute Deviation (MAD)	0.28
Skewness	18.49558
Sum	23556.82
Variance	3.4625158
Monotonicity	Not monotonic



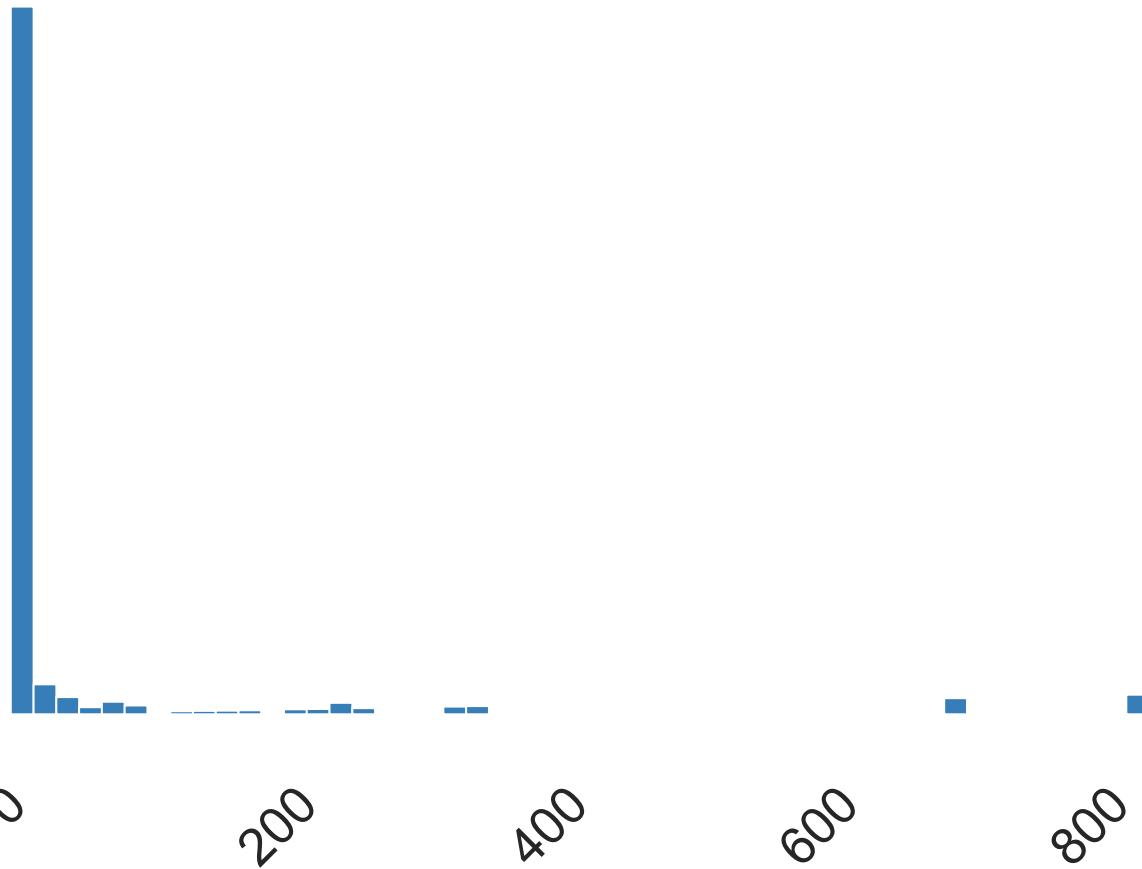
Value	Count	Frequency (%)
0.01	1081	2.9%
0.02	986	2.6%
0.03	847	2.2%
0.04	693	1.8%
0.05	686	1.8%
0.06	606	1.6%
0.07	542	1.4%
0.09	542	1.4%
0.08	524	1.4%
0.1	472	1.2%
Other values (778)	19035	50.4%
(Missing)	11751	31.1%

Value	Count	Frequency (%)
0.01	1081	2.9%
0.02	986	2.6%
0.03	847	2.2%
0.04	693	1.8%
0.05	686	1.8%
0.06	606	1.6%
0.07	542	1.4%
0.08	524	1.4%
0.09	542	1.4%
0.1	472	1.2%

Value	Count	Frequency (%)
103.53	1	< 0.1%
96.75	1	< 0.1%
57.05	1	< 0.1%
47	1	< 0.1%
45.35	1	< 0.1%
44.31	1	< 0.1%
42.44	1	< 0.1%
38.46	1	< 0.1%
35	1	< 0.1%
34.21	1	< 0.1%

calculated_host_listings_countReal number (\mathbb{R})

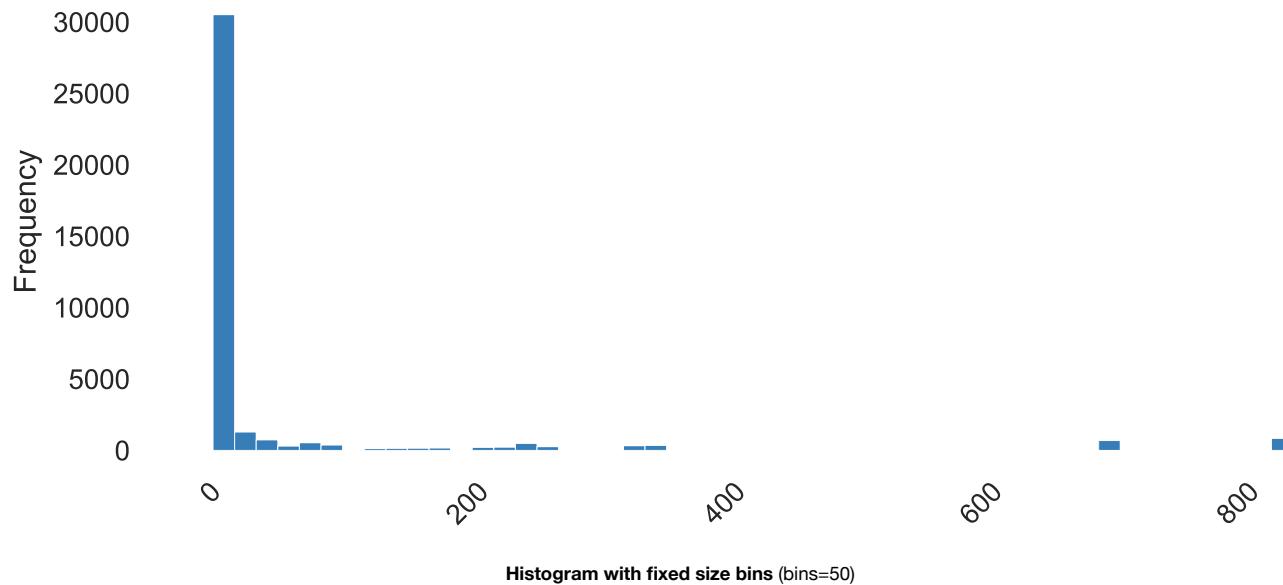
Distinct	73
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	53.763035
Minimum	1
Maximum	842
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

**Quantile statistics**

Minimum	1
5-th percentile	1
Q1	1
median	2
Q3	8
95-th percentile	329
Maximum	842
Range	841
Interquartile range (IQR)	7

Descriptive statistics

Standard deviation	161.97972
Coefficient of variation (CV)	3.0128456
Kurtosis	14.168405
Mean	53.763035
Median Absolute Deviation (MAD)	1
Skewness	3.8170847
Sum	2030361
Variance	26237.431
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1	18880	50.0%
2	4330	11.5%
3	2040	5.4%
4	1308	3.5%
842	842	2.2%
5	730	1.9%
692	692	1.8%
6	600	1.6%
8	448	1.2%
7	406	1.1%
Other values (63)	7489	19.8%

Value	Count	Frequency (%)
1	18880	50.0%
2	4330	11.5%
3	2040	5.4%
4	1308	3.5%
5	730	1.9%
6	600	1.6%
7	406	1.1%
8	448	1.2%
9	369	1.0%
10	330	0.9%

Value	Count	Frequency (%)
842	842	2.2%
692	692	1.8%
351	351	0.9%
329	329	0.9%
262	262	0.7%
246	246	0.7%
245	245	0.6%
226	226	0.6%
212	212	0.6%
170	170	0.5%

availability_365Real number (\mathbb{R})

ZEROS

Distinct

366

Distinct (%)

1.0%

Missing

0

Missing (%)

0.0%

Infinite

0

Infinite (%)

0.0%

Mean

152.4116

Minimum

0

Maximum

365

Zeros

12546

Zeros (%)

33.2%

Negative

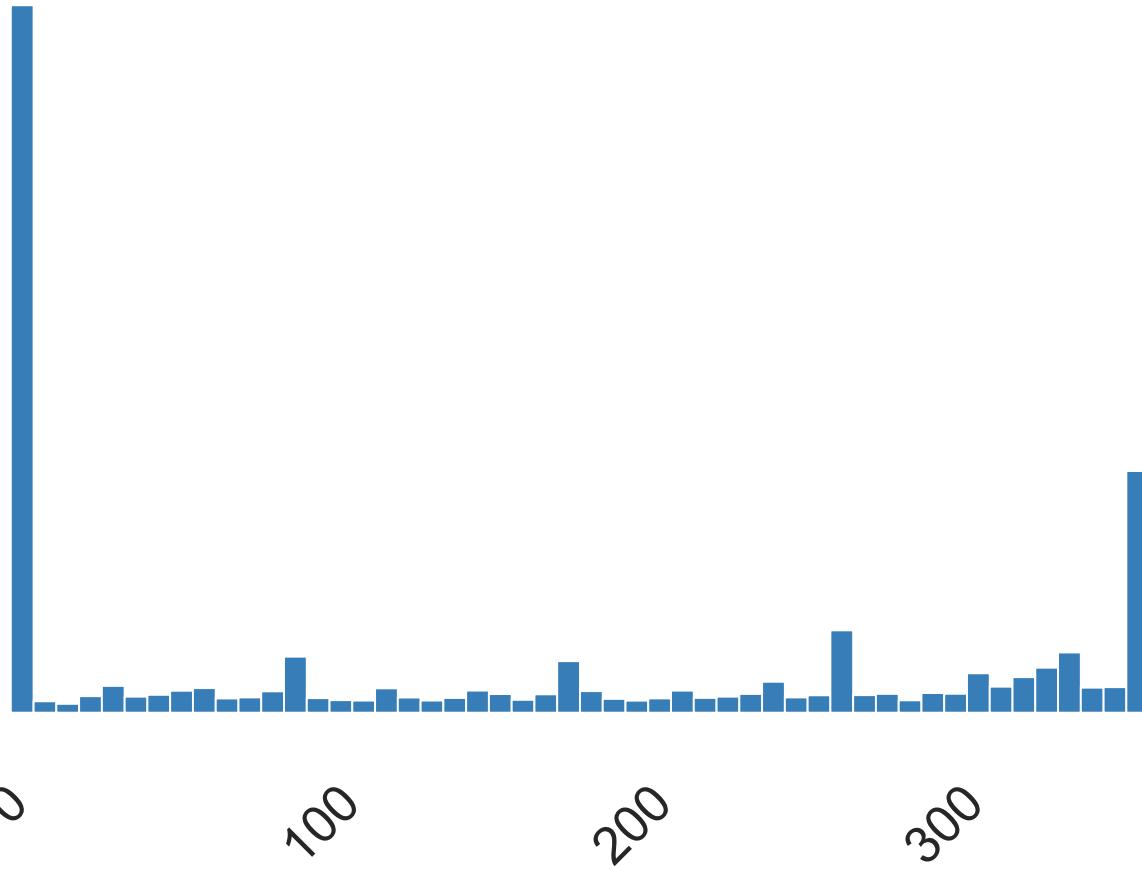
0

Negative (%)

0.0%

Memory size

295.2 KiB

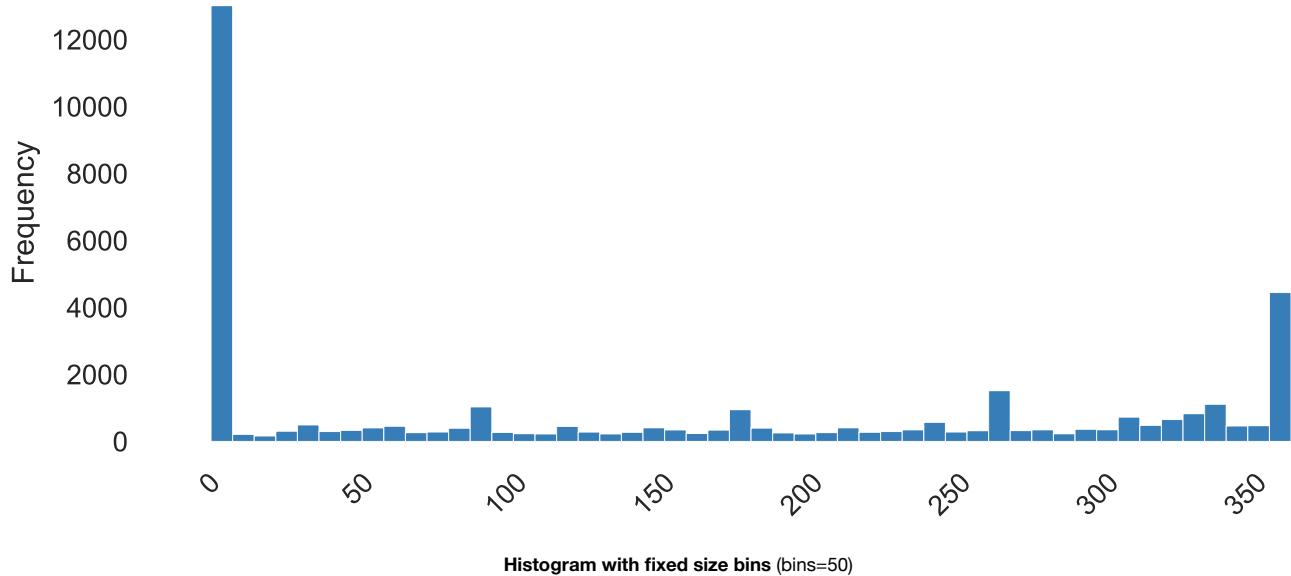


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	126
Q3	301
95-th percentile	364
Maximum	365
Range	365
Interquartile range (IQR)	301

Descriptive statistics

Standard deviation	142.79733
Coefficient of variation (CV)	0.93691907
Kurtosis	-1.5724648
Mean	152.4116
Median Absolute Deviation (MAD)	126
Skewness	0.24675532
Sum	5755824
Variance	20391.078
Monotonicity	Not monotonic



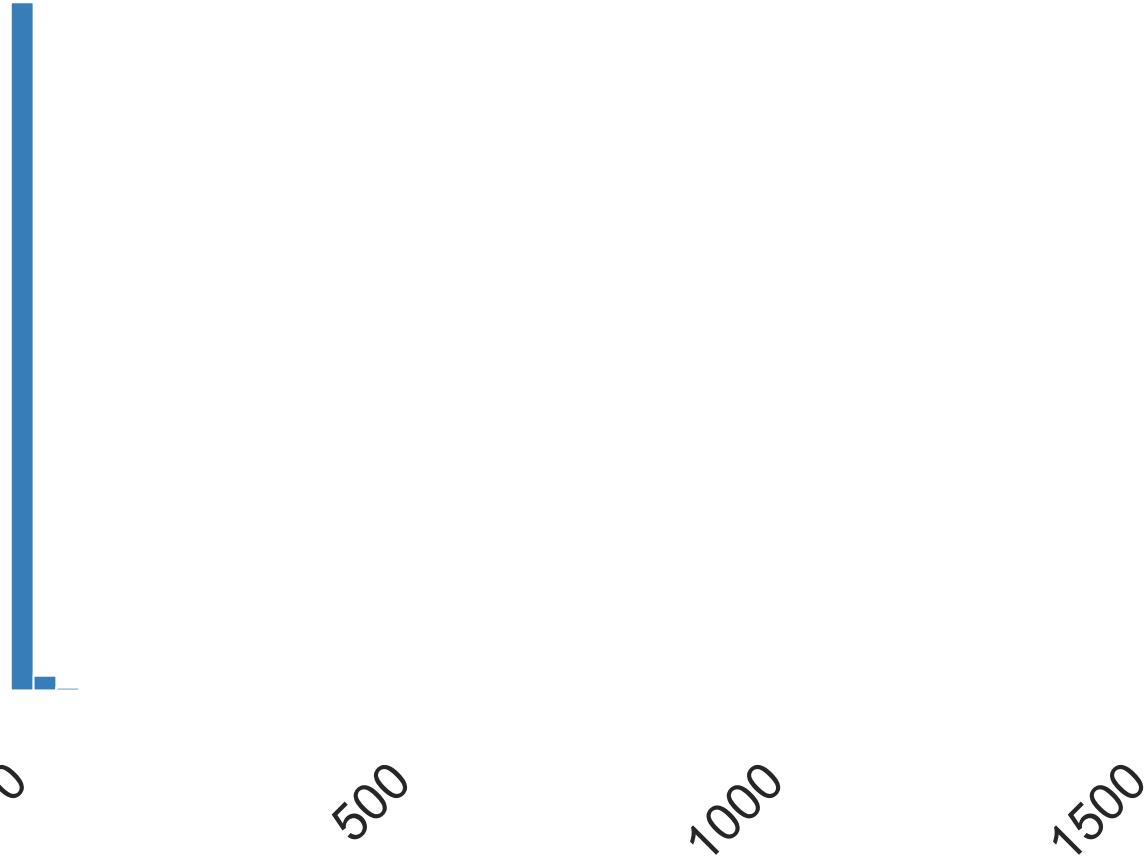
Value	Count	Frequency (%)
0	12546	33.2%
365	1784	4.7%
364	1359	3.6%
269	749	2.0%
363	674	1.8%
89	373	1.0%
270	351	0.9%
90	336	0.9%
179	316	0.8%
180	273	0.7%
Other values (356)	19004	50.3%

Value	Count	Frequency (%)
0	12546	33.2%
1	129	0.3%
2	63	0.2%
3	35	0.1%
4	79	0.2%
5	65	0.2%
6	54	0.1%
7	43	0.1%
8	40	0.1%
9	28	0.1%

Value	Count	Frequency (%)
365	1784	4.7%
364	1359	3.6%
363	674	1.8%
362	229	0.6%
361	51	0.1%
360	61	0.2%
359	99	0.3%
358	189	0.5%
357	77	0.2%
356	78	0.2%

number_of_reviews_ltmReal number (\mathbb{R})**HIGH CORRELATION** (This variable has a high overall correlation with 2 fields: `number_of_reviews`, `reviews_per_month`). SKEWED ZEROS

Distinct	145
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.9931683
Minimum	0
Maximum	1568
Zeros	23090
Zeros (%)	61.1%
Negative	0
Negative (%)	0.0%
Memory size	295.2 KiB

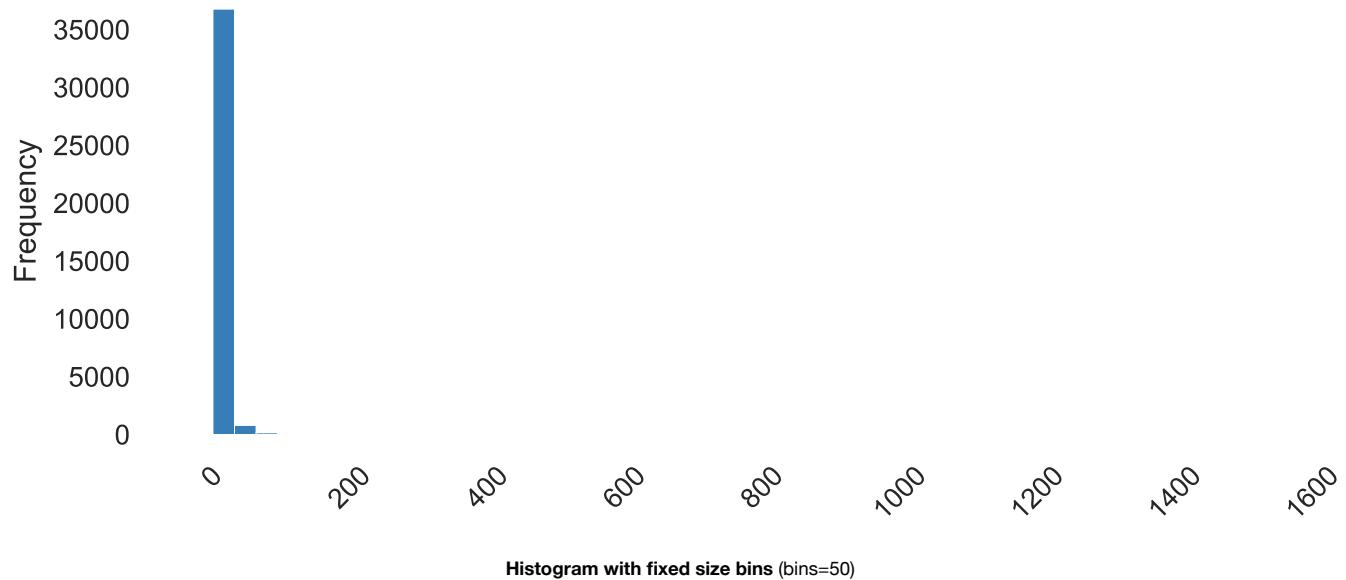


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	3
95-th percentile	21
Maximum	1568
Range	1568
Interquartile range (IQR)	3

Descriptive statistics

Standard deviation	15.762591
Coefficient of variation (CV)	3.9473897
Kurtosis	3207.7332
Mean	3.9931683
Median Absolute Deviation (MAD)	0
Skewness	40.572844
Sum	150802
Variance	248.45928
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	23090	61.1%
1	3164	8.4%
2	1947	5.2%
3	1358	3.6%
4	1027	2.7%
5	797	2.1%
6	584	1.5%
7	477	1.3%
8	407	1.1%
9	380	1.0%
Other values (135)	4534	12.0%

Value	Count	Frequency (%)
0	23090	61.1%
1	3164	8.4%
2	1947	5.2%
3	1358	3.6%
4	1027	2.7%
5	797	2.1%
6	584	1.5%
7	477	1.3%
8	407	1.1%
9	380	1.0%

Value	Count	Frequency (%)
1568	1	< 0.1%
976	1	< 0.1%
742	1	< 0.1%
608	1	< 0.1%
526	1	< 0.1%
390	1	< 0.1%
343	1	< 0.1%
332	1	< 0.1%
331	1	< 0.1%
280	1	< 0.1%

license

Text

MISSING

Distinct	1732
Distinct (%)	33.5%
Missing	32594
Missing (%)	86.3%
Memory size	295.2 KiB

**Length**

Max length	18
Median length	6
Mean length	11.335138
Min length	6

Characters and Unicode

Total characters	58614
Distinct characters	26
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	1449	?
Unique (%)	28.0%	

Sample

1st row	OSE-STRREG-0000008
2nd row	OSE-STRREG-0000923
3rd row	OSE-STRREG-0000108
4th row	OSE-STRREG-0001710
5th row	OSE-STRREG-0000041

Value	Count	Frequency (%)
exempt	2872	55.5%
ose-strreg-0000068	99	1.9%
ose-strreg-0041458	36	0.7%
ose-strreg-0001054	18	0.3%
ose-strreg-1081622	10	0.2%
ose-strreg-0091251	9	0.2%
ose-strreg-0001957	9	0.2%
ose-strreg-0000095	7	0.1%
ose-strreg-0000437	7	0.1%
ose-strreg-1044019	7	0.1%
Other values (1717)	2097	40.6%



Most occurring characters

Value	Count	Frequency (%)
0	8574	14.6%
E	7420	12.7%
-	4598	7.8%
S	4549	7.8%
R	4544	7.8%
e	2922	5.0%
t	2899	4.9%
m	2872	4.9%
p	2872	4.9%
x	2872	4.9%
Other values (16)	14492	24.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	58614	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
0	8574	14.6%
E	7420	12.7%
-	4598	7.8%
S	4549	7.8%
R	4544	7.8%
e	2922	5.0%
t	2899	4.9%
m	2872	4.9%
p	2872	4.9%
x	2872	4.9%
Other values (16)	14492	24.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	58614	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
0	8574	14.6%
E	7420	12.7%
-	4598	7.8%
S	4549	7.8%
R	4544	7.8%
e	2922	5.0%
t	2899	4.9%
m	2872	4.9%
p	2872	4.9%
x	2872	4.9%
Other values (16)	14492	24.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	58614	100.0%

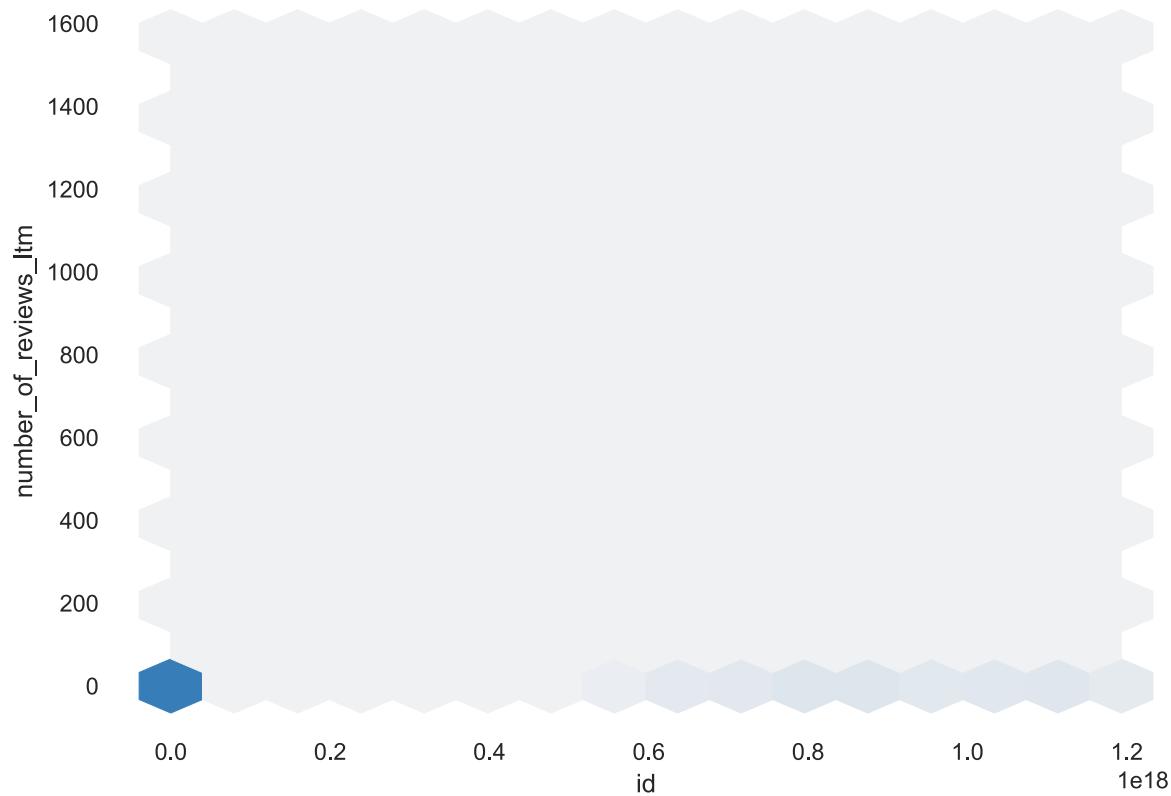
Most frequent character per block*(unknown)*

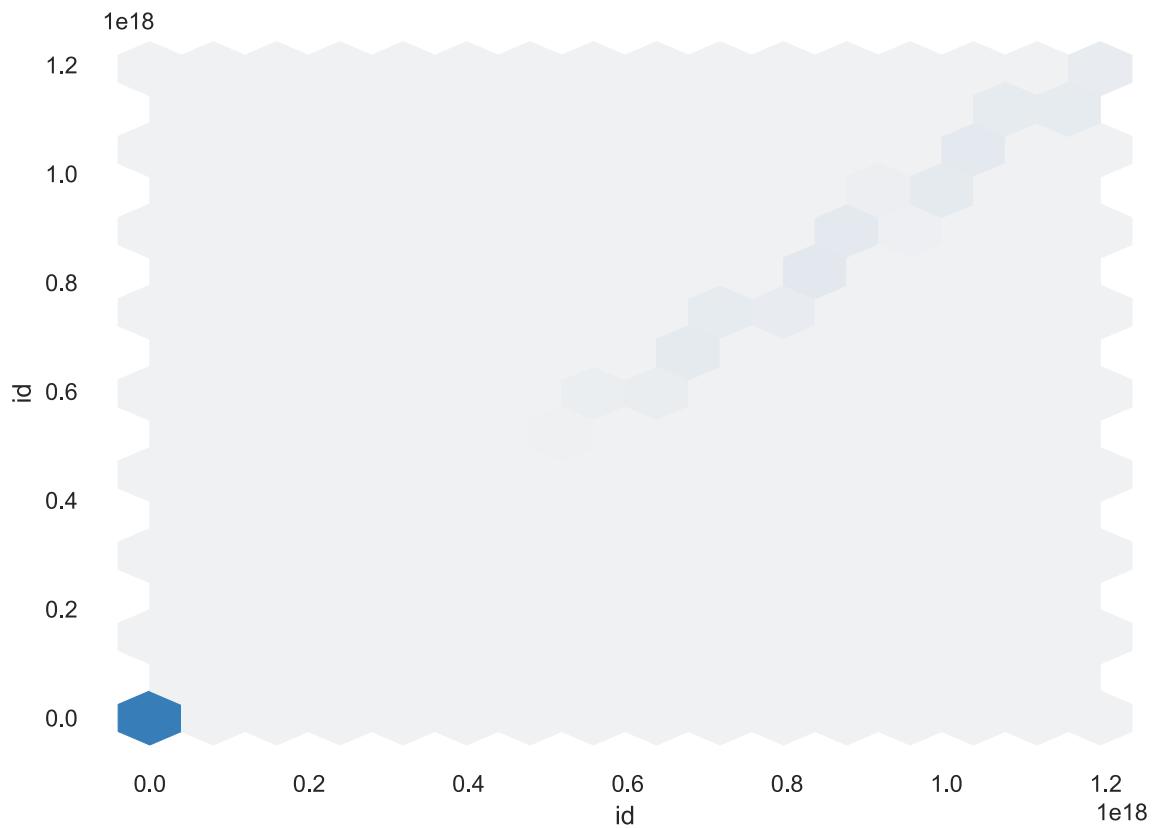
Value	Count	Frequency (%)
0	8574	14.6%
E	7420	12.7%
-	4598	7.8%
S	4549	7.8%
R	4544	7.8%
e	2922	5.0%
t	2899	4.9%
m	2872	4.9%
p	2872	4.9%
x	2872	4.9%
Other values (16)	14492	24.7%

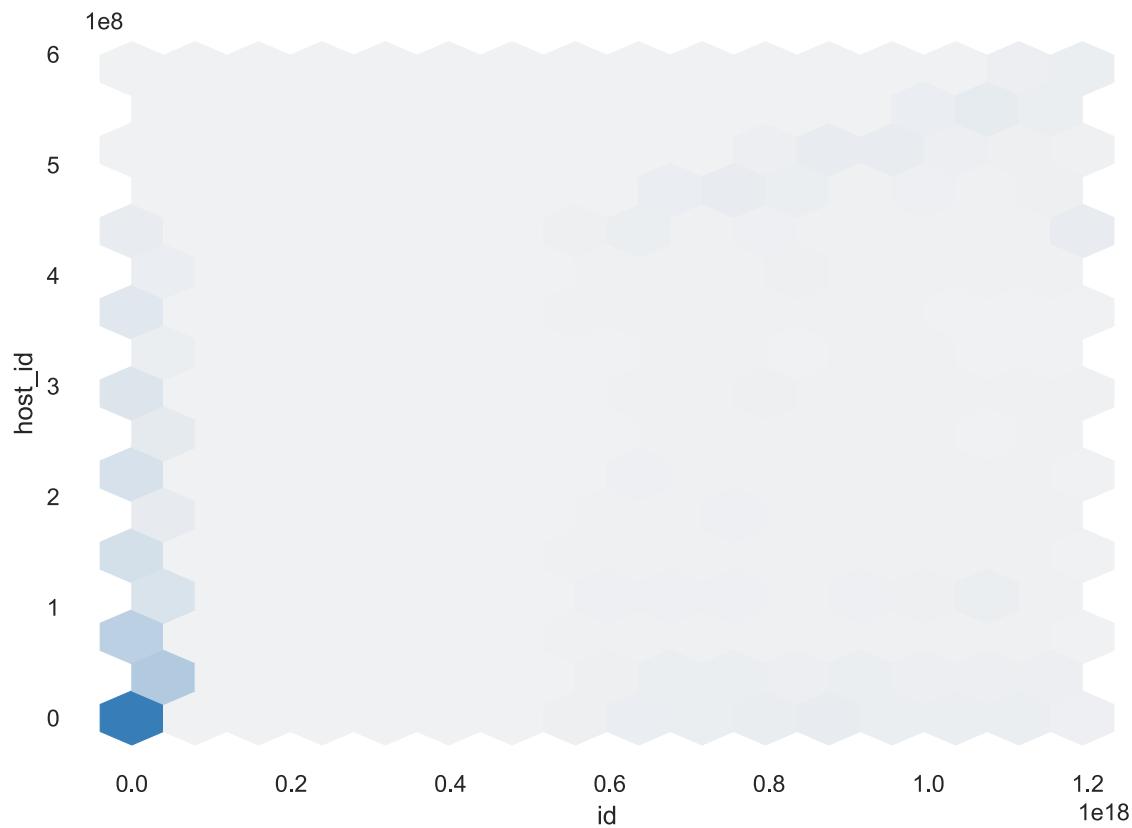
Interactions

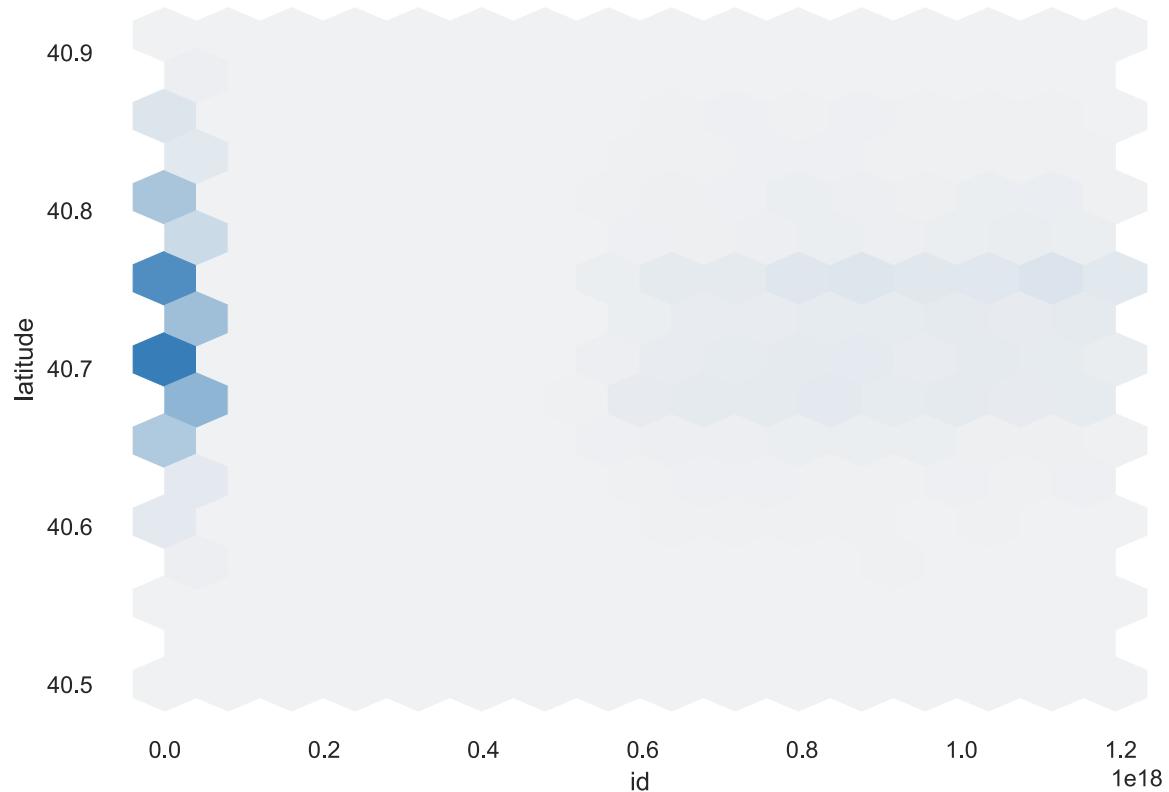
```
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365  
number_of_reviews_ltm
```

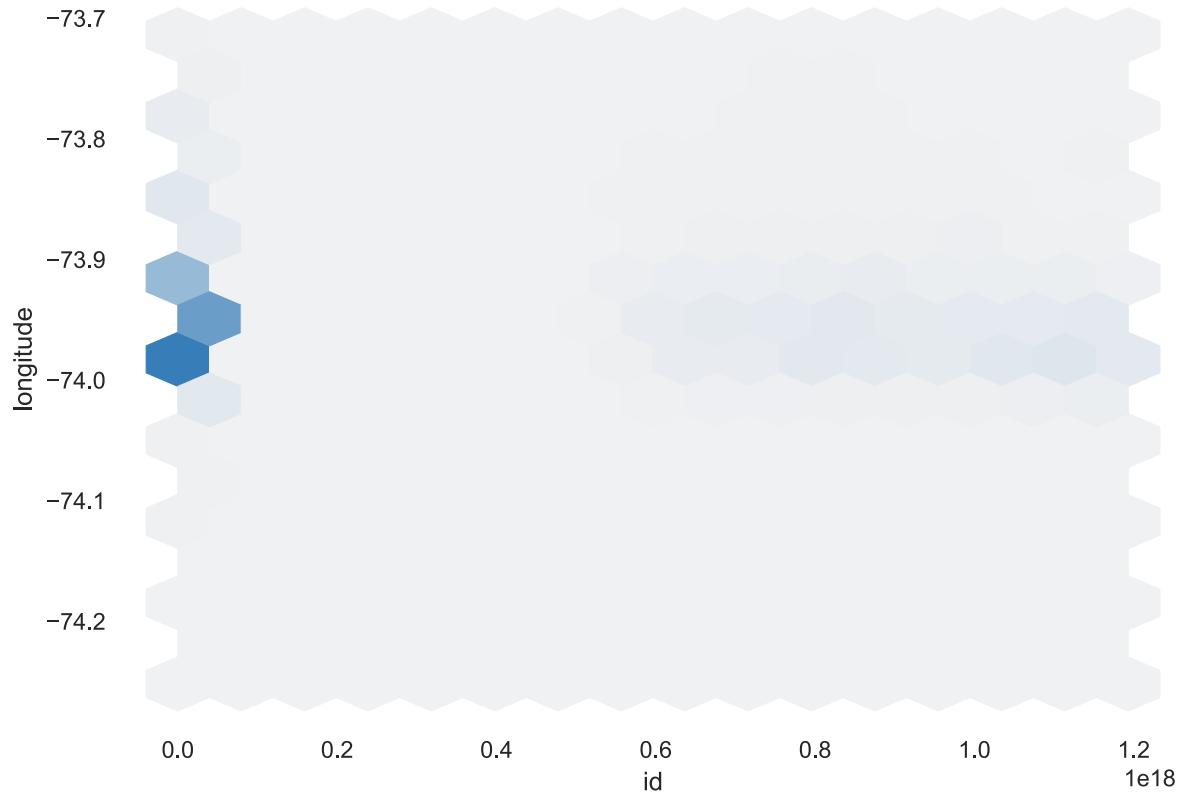
number_of_reviews_ltm
id
host_id
latitude
longitude
price
minimum_nights
number_of_reviews
reviews_per_month
calculated_host_listings_c
availability_365

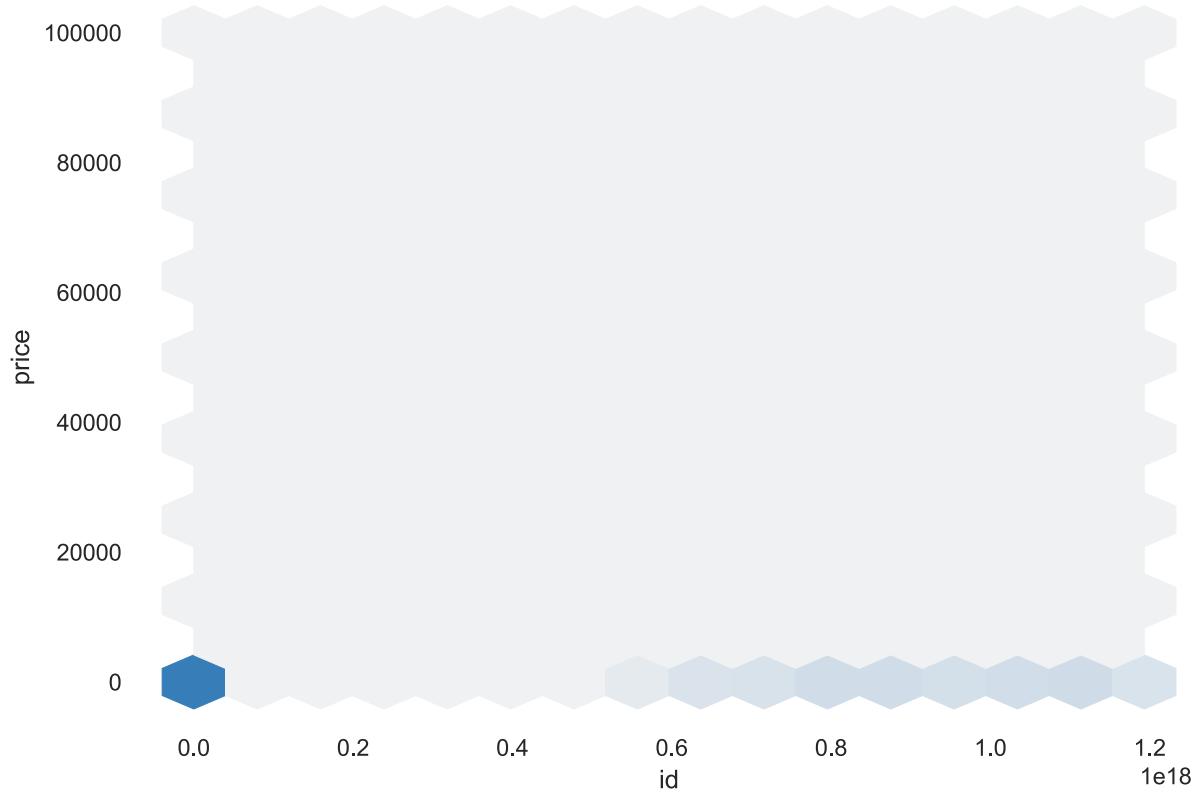


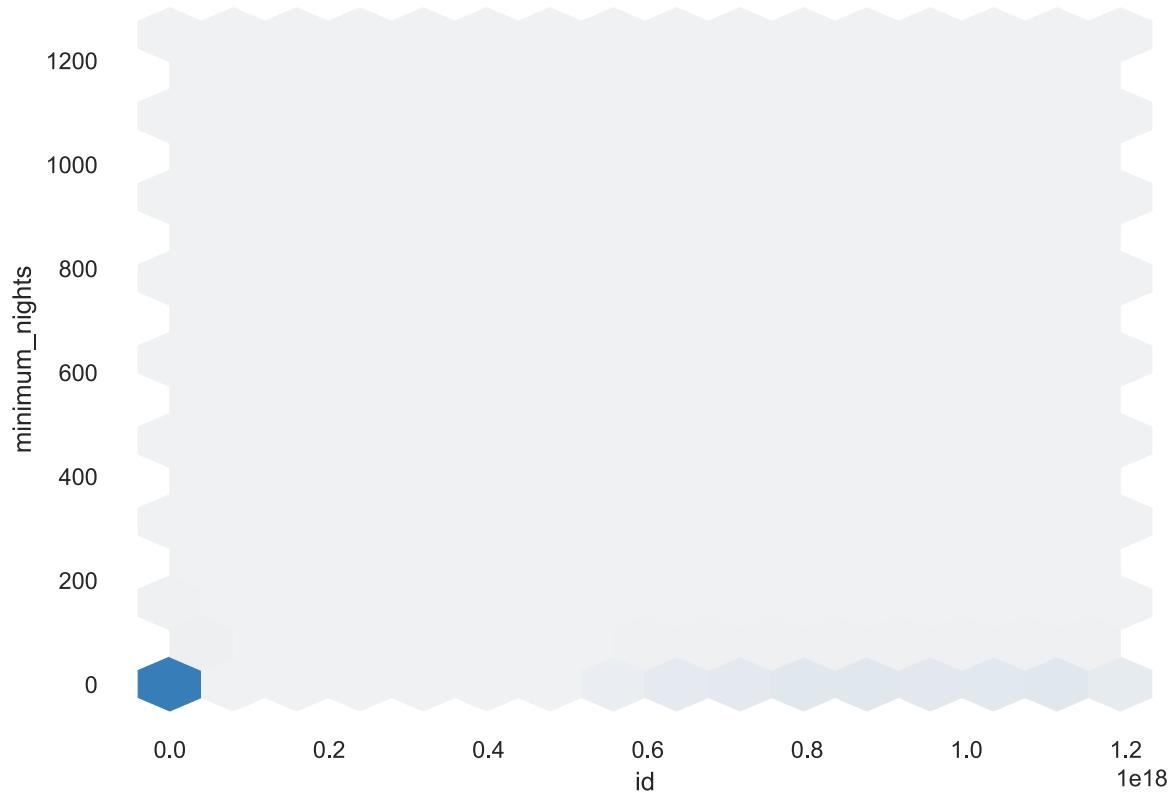


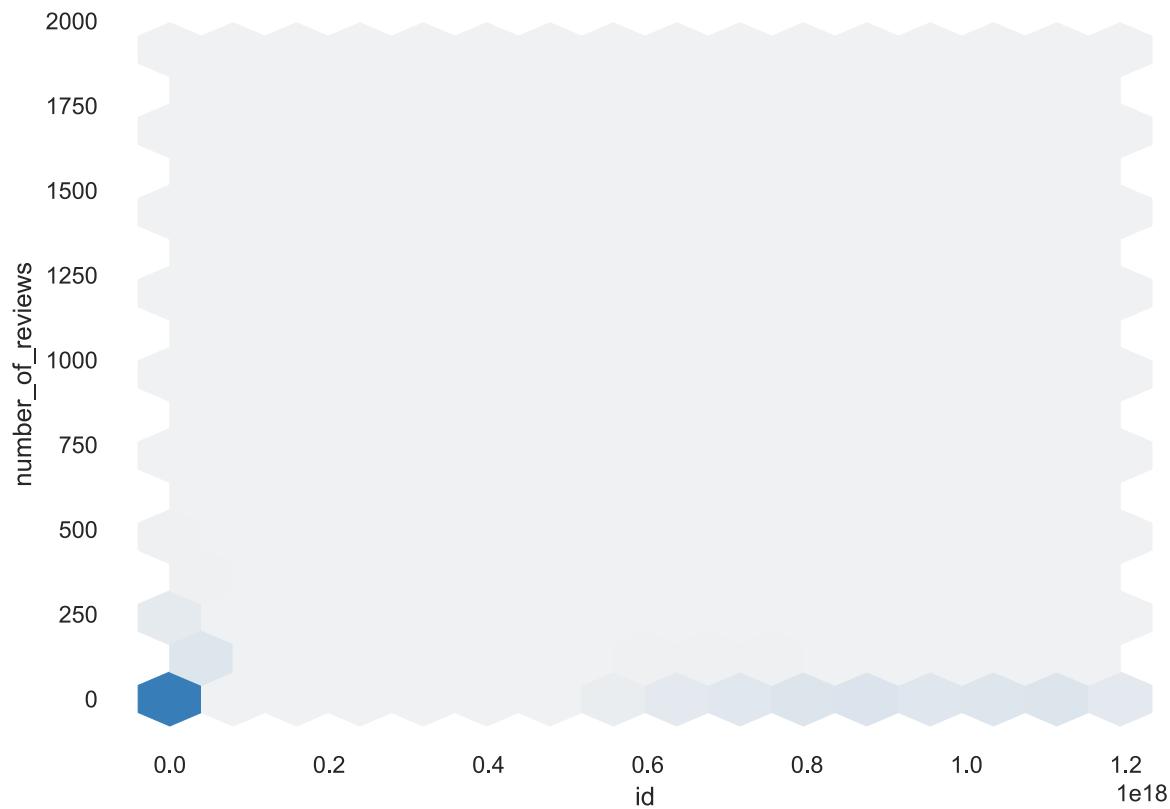


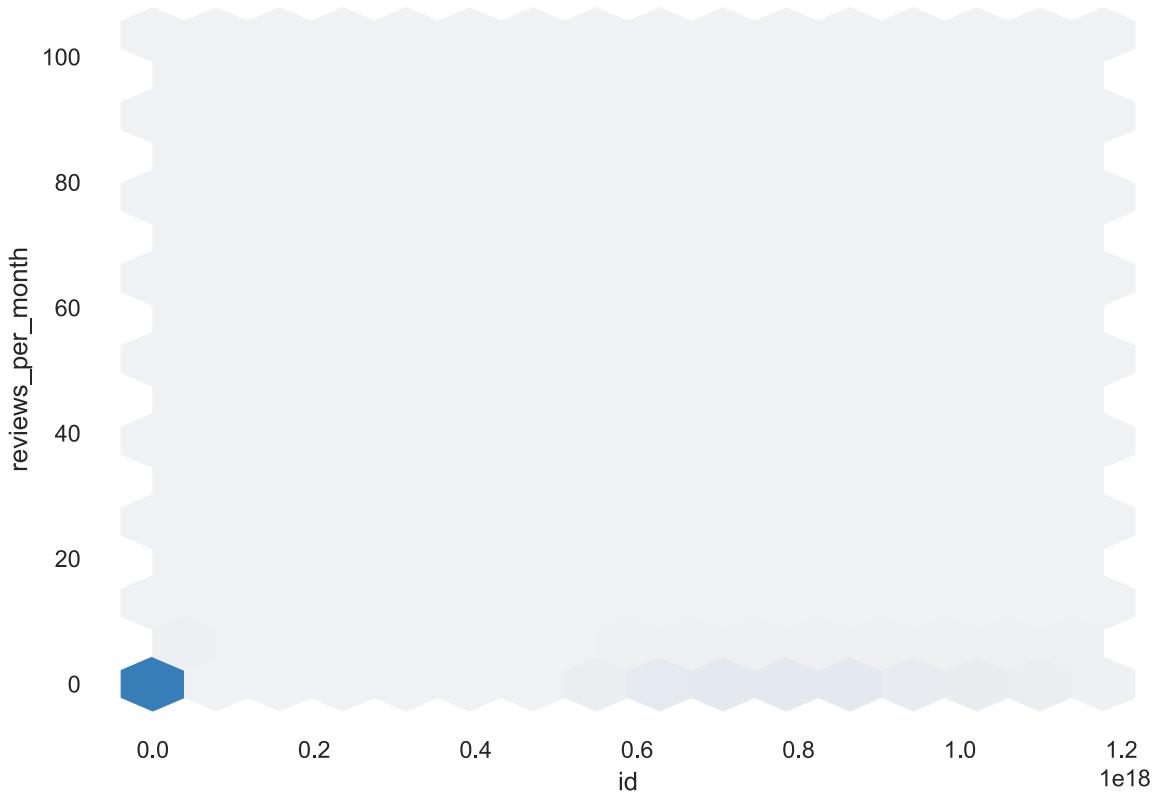


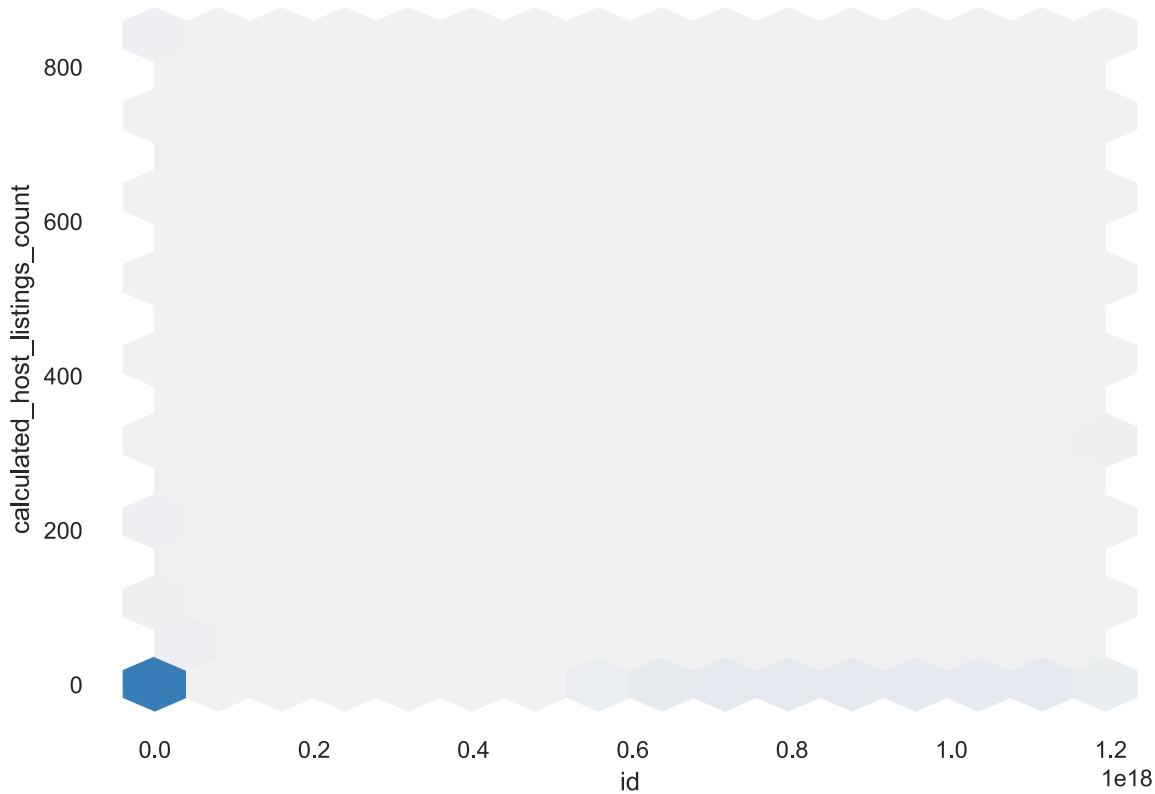


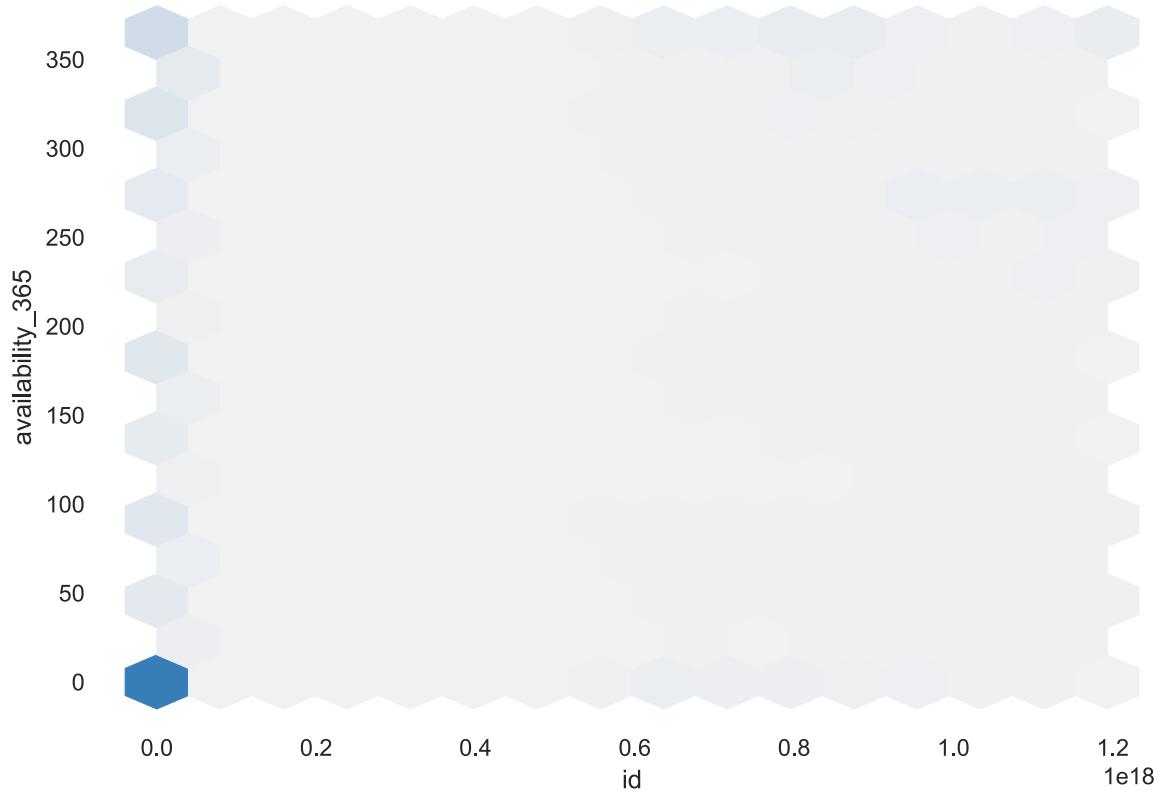




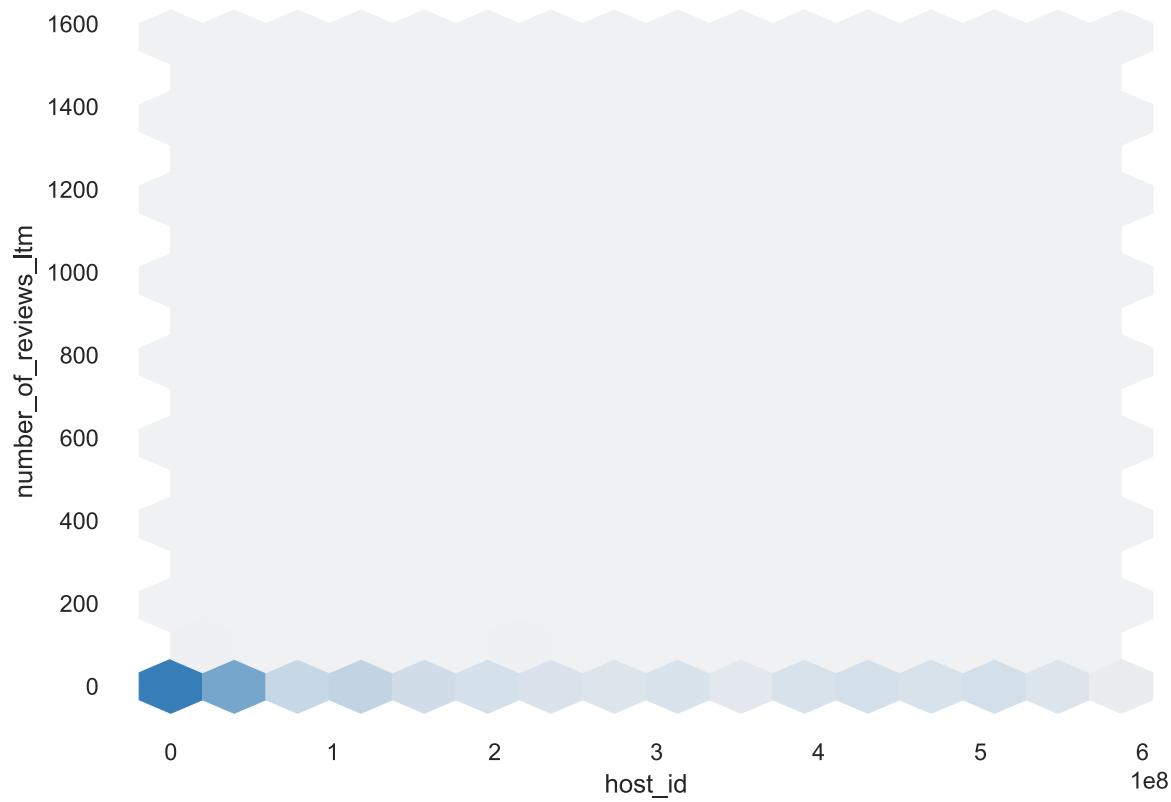


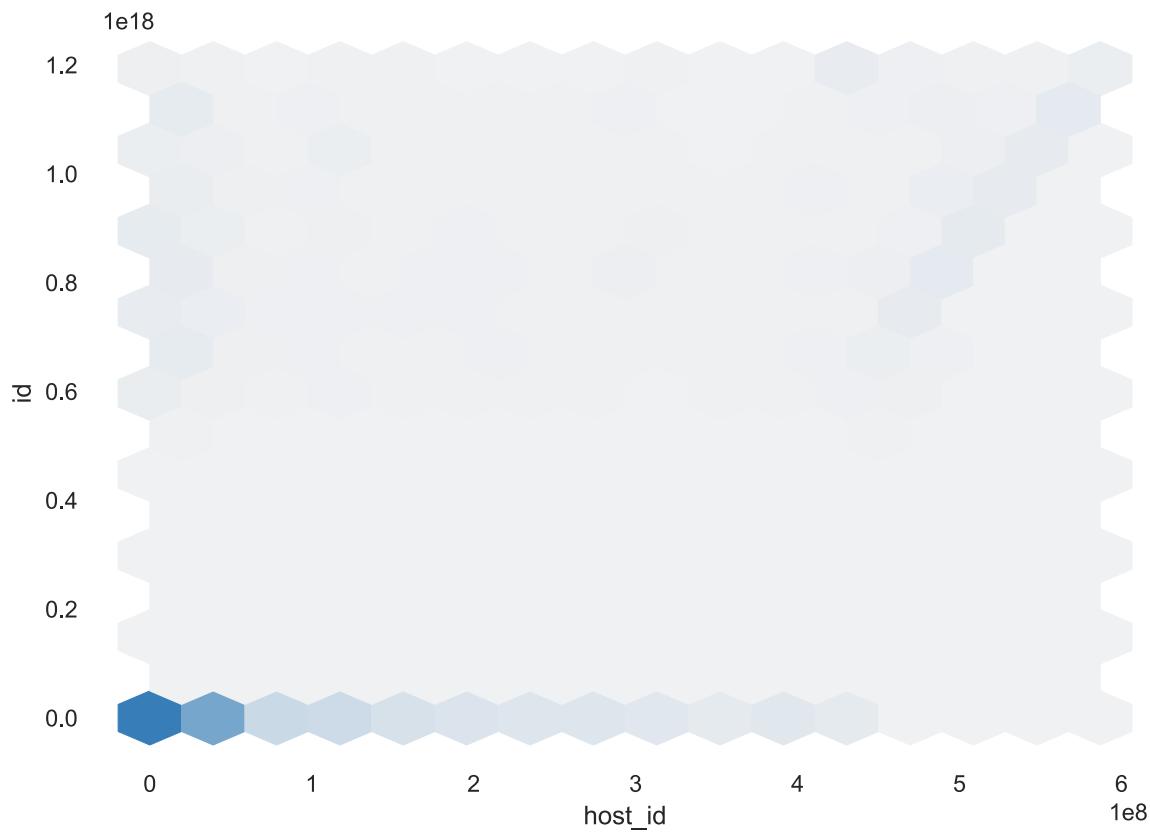


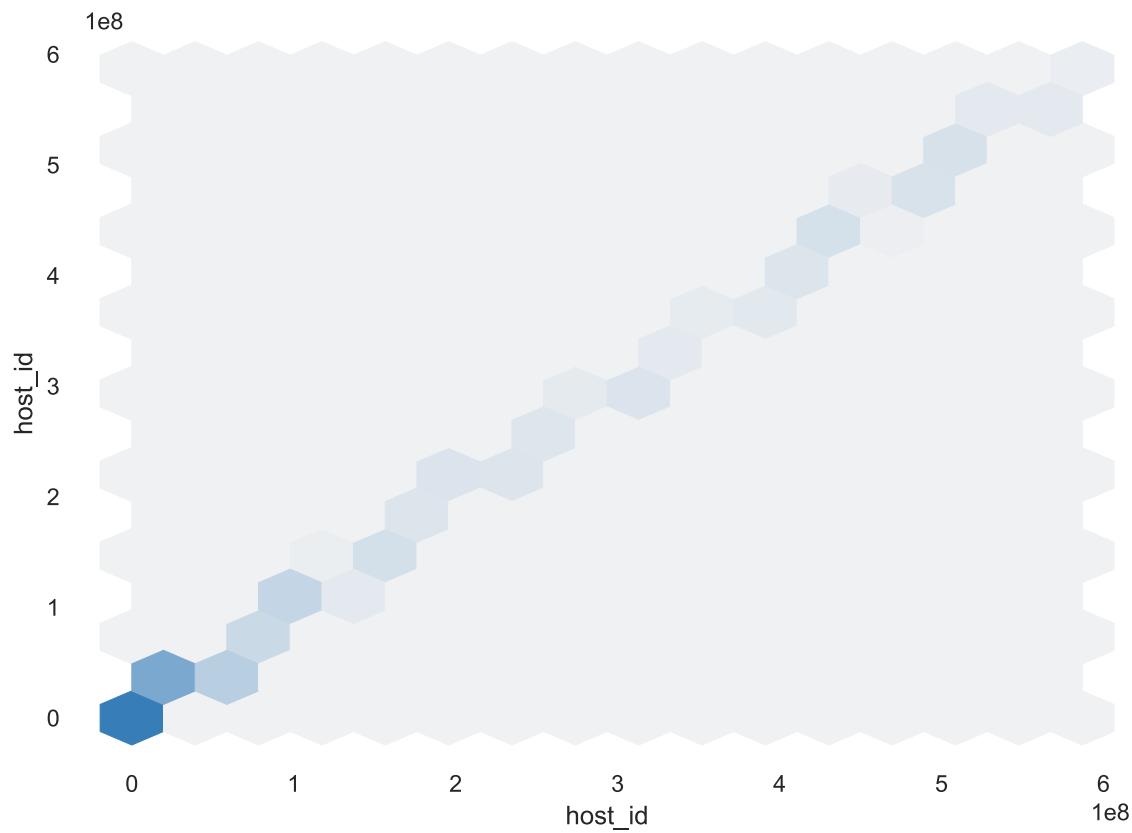


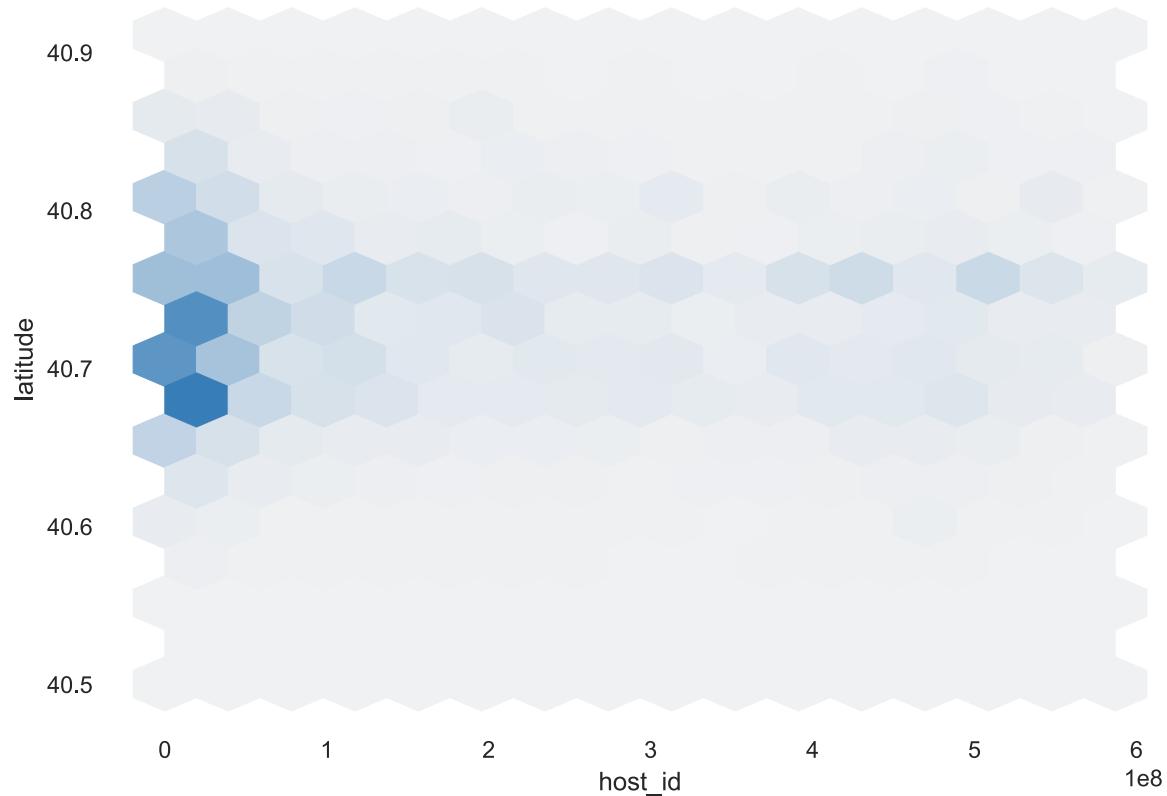


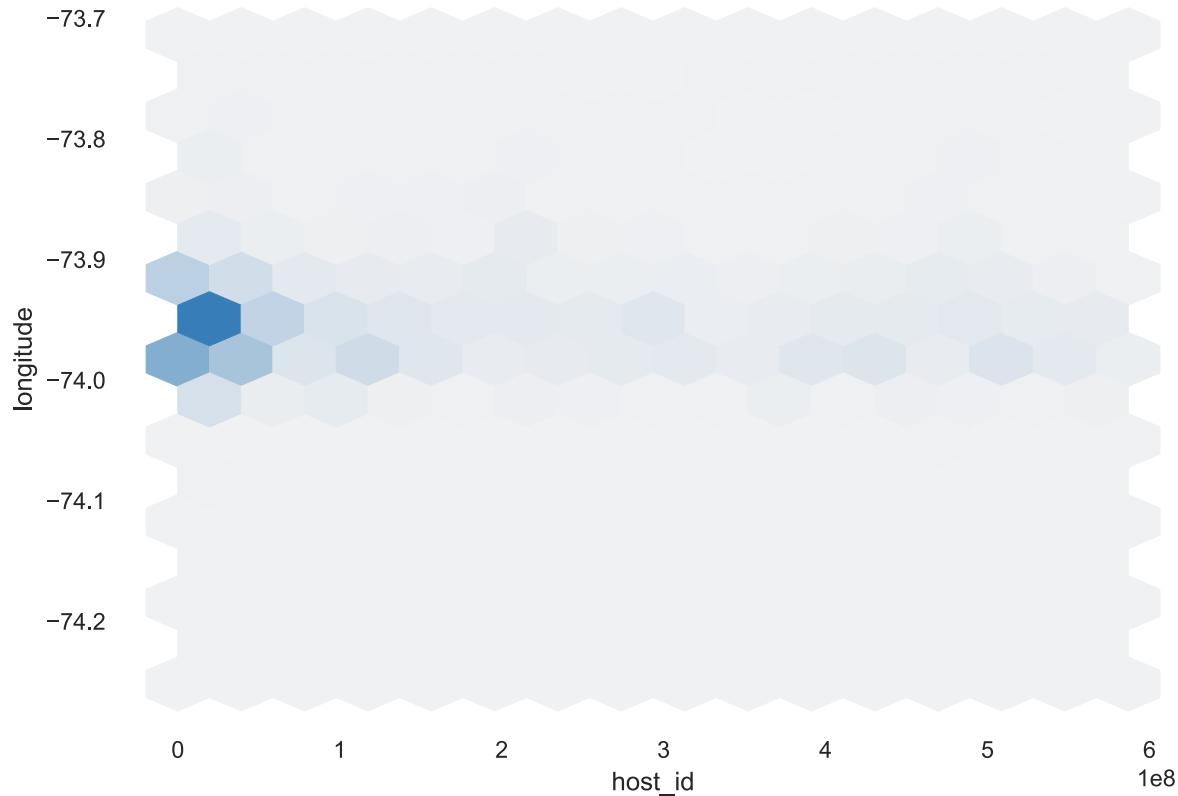

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

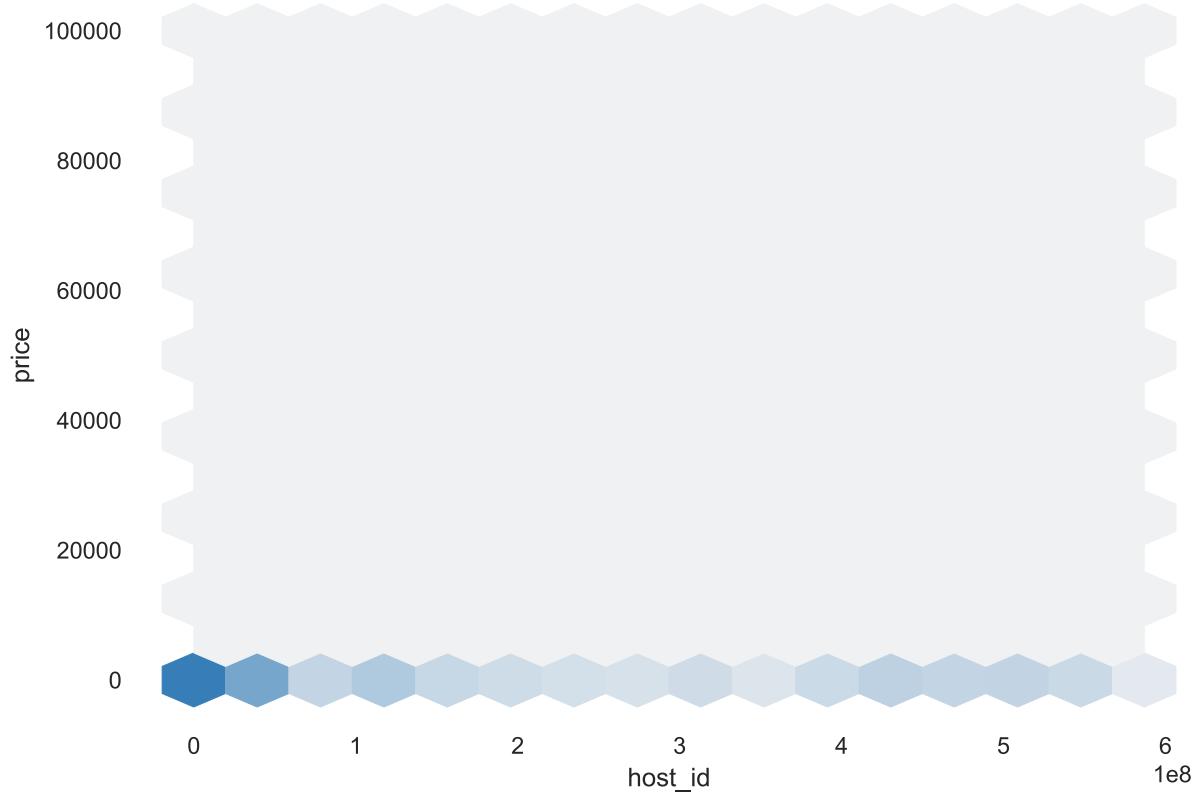


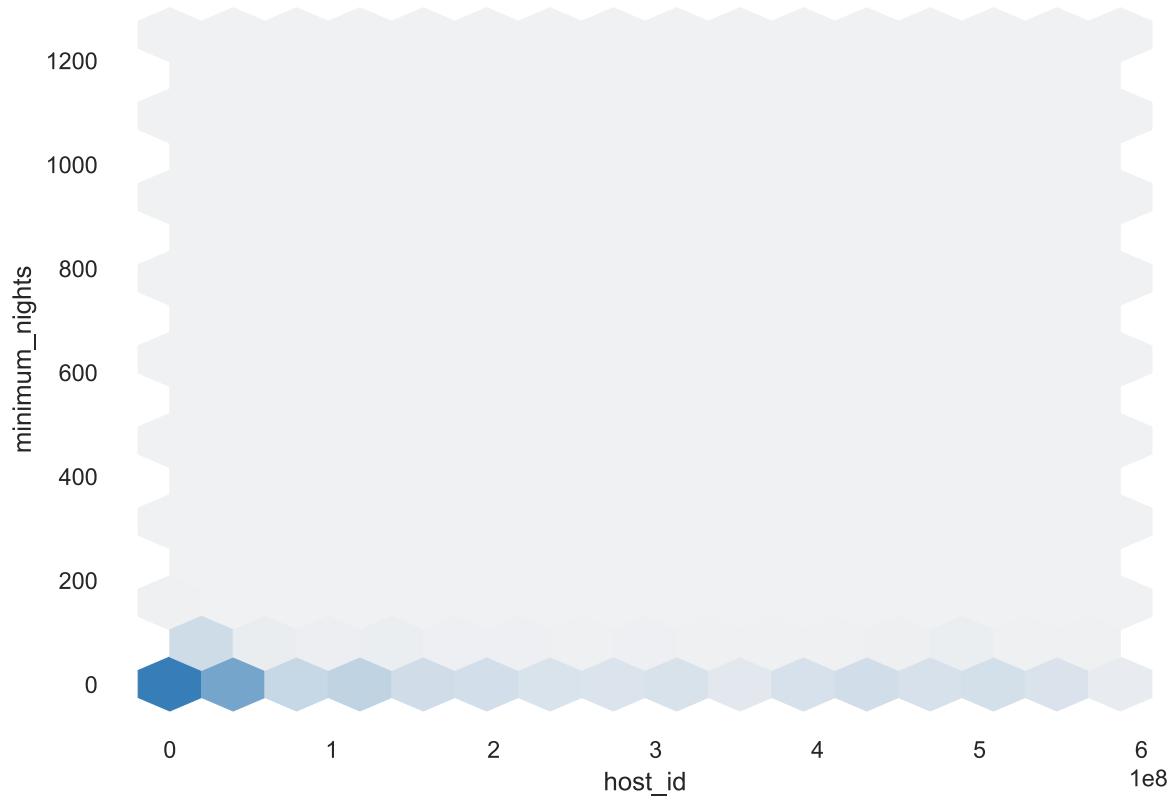


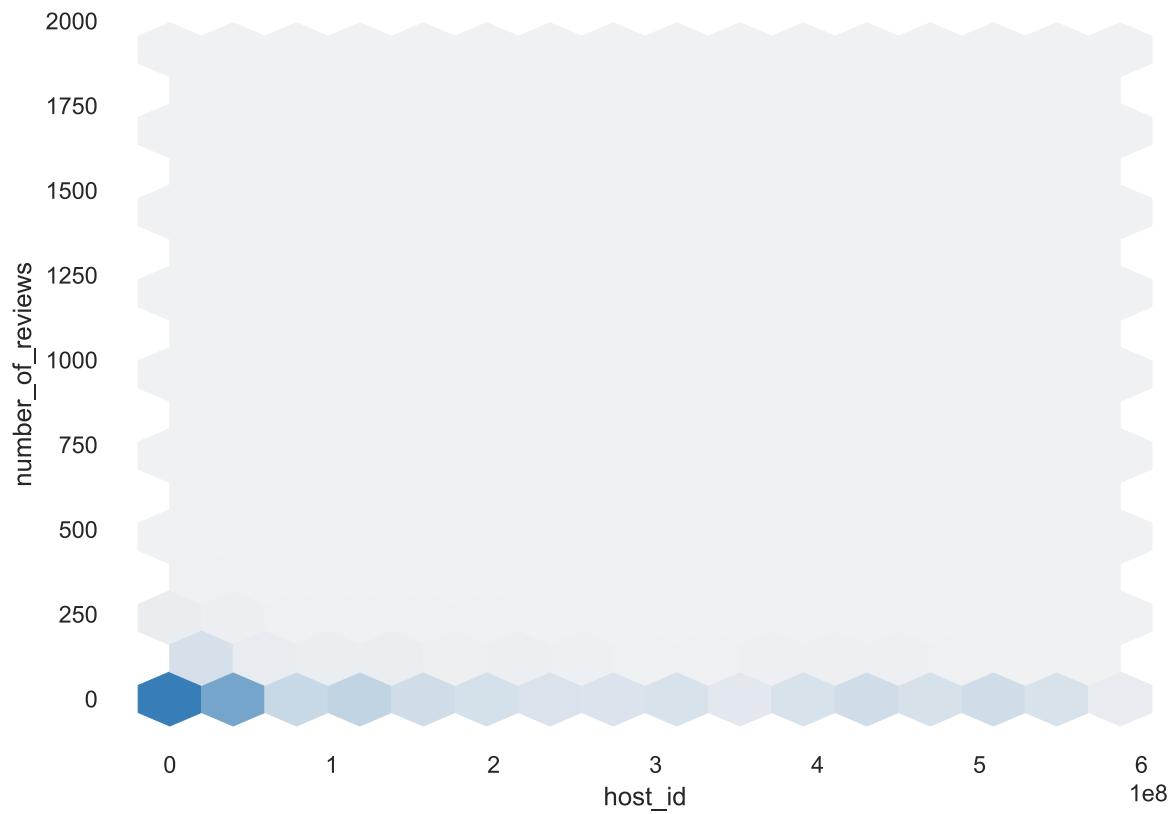


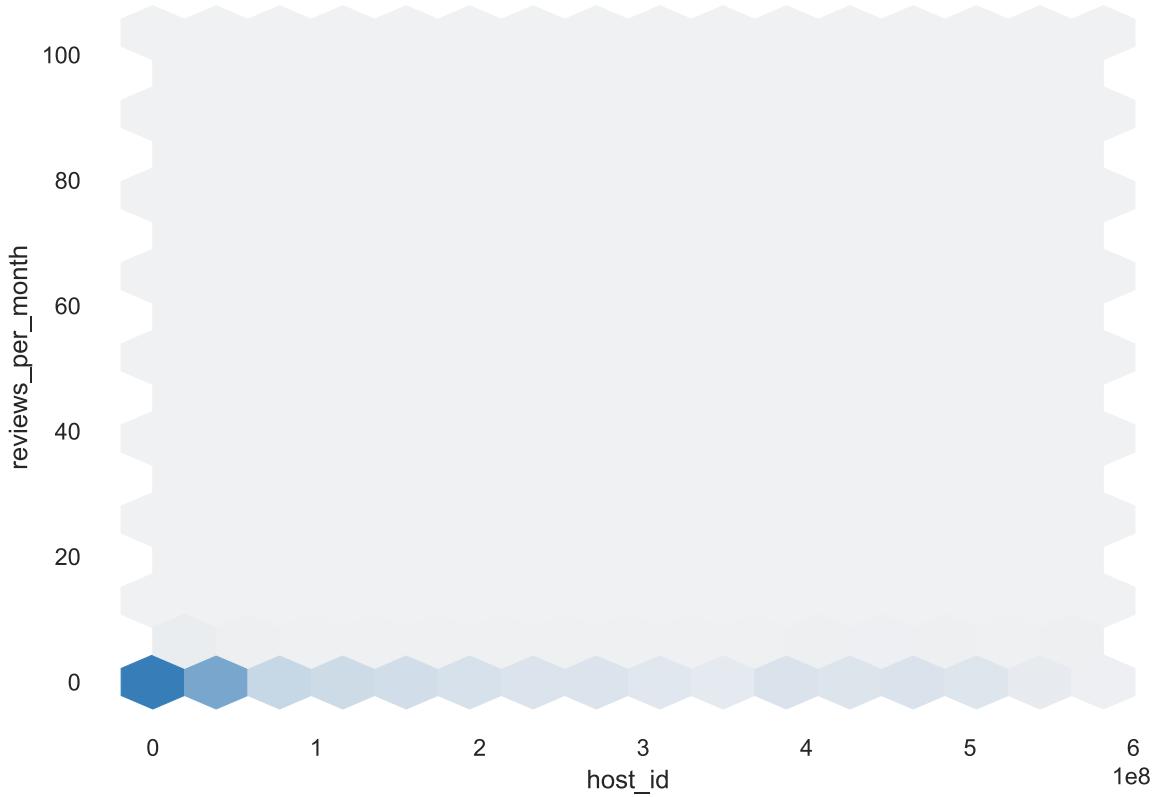


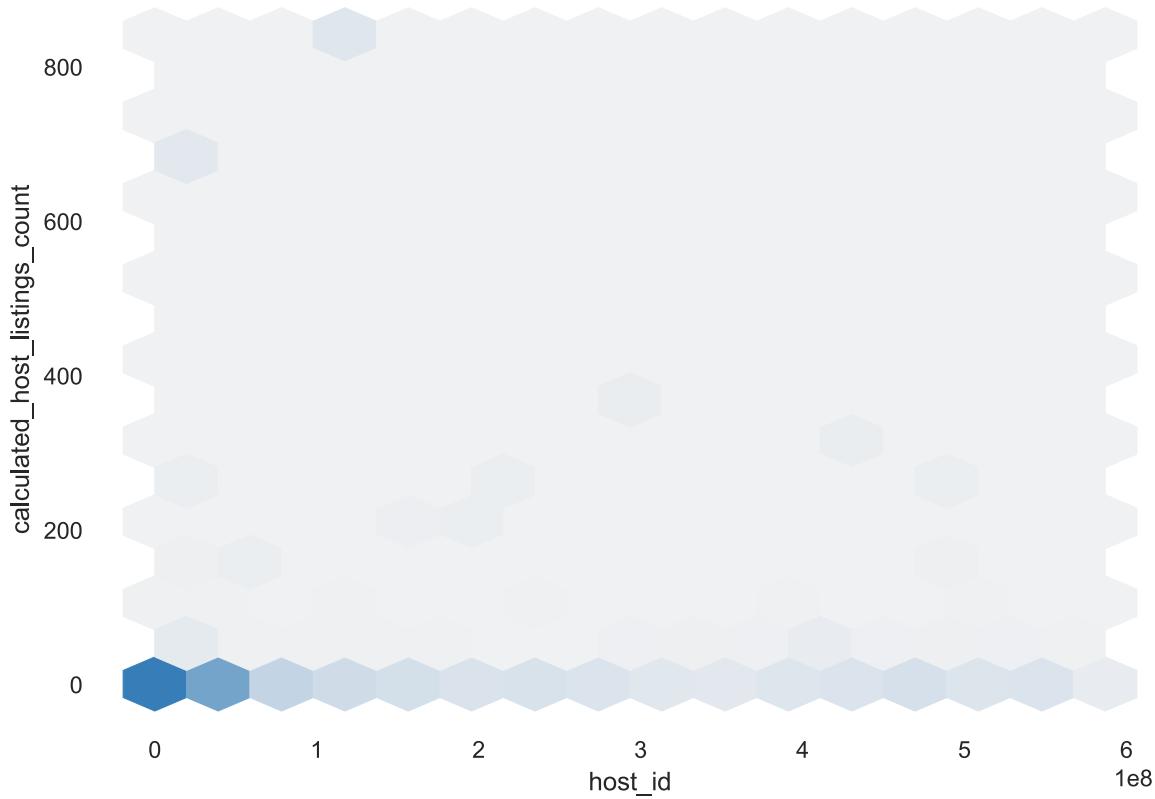


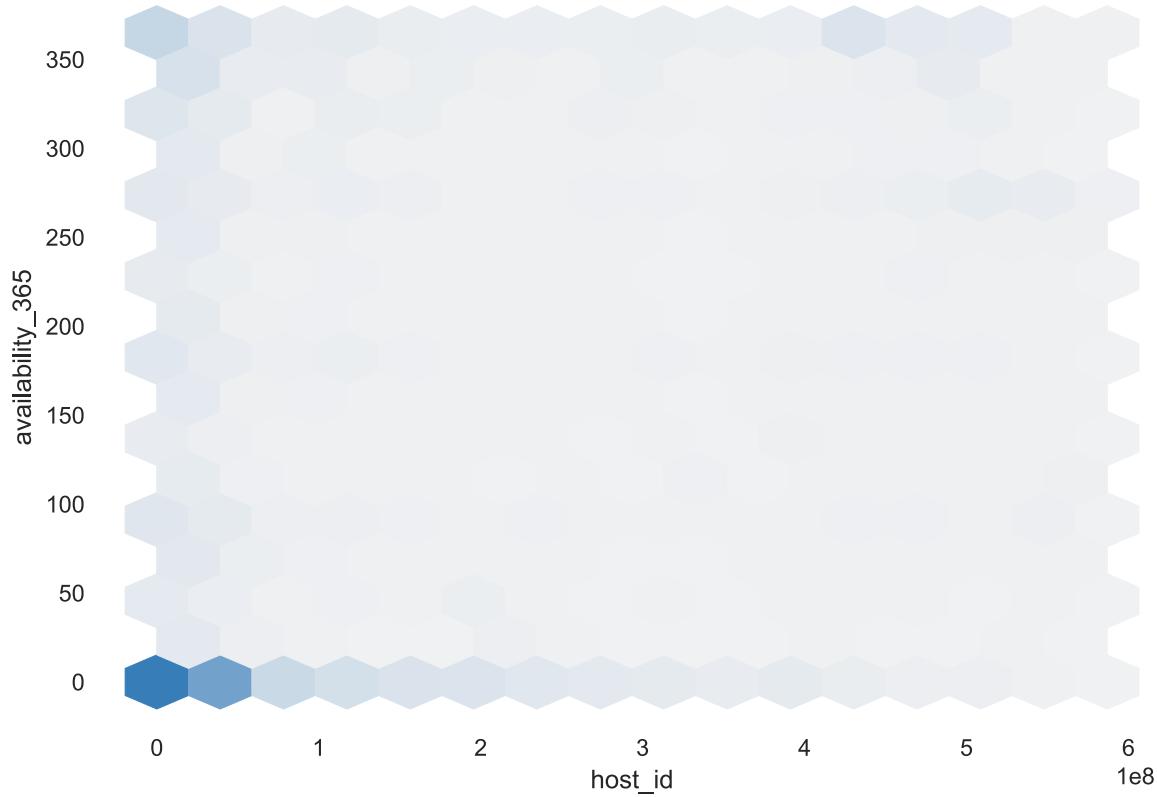




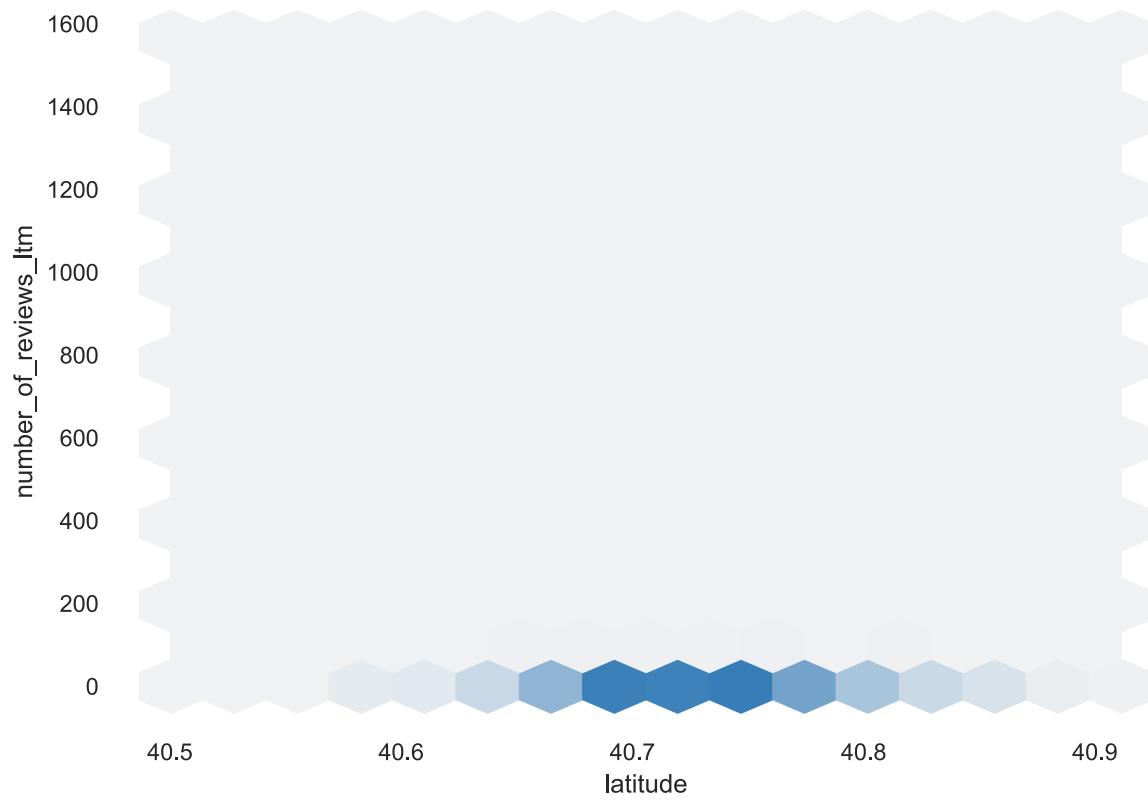


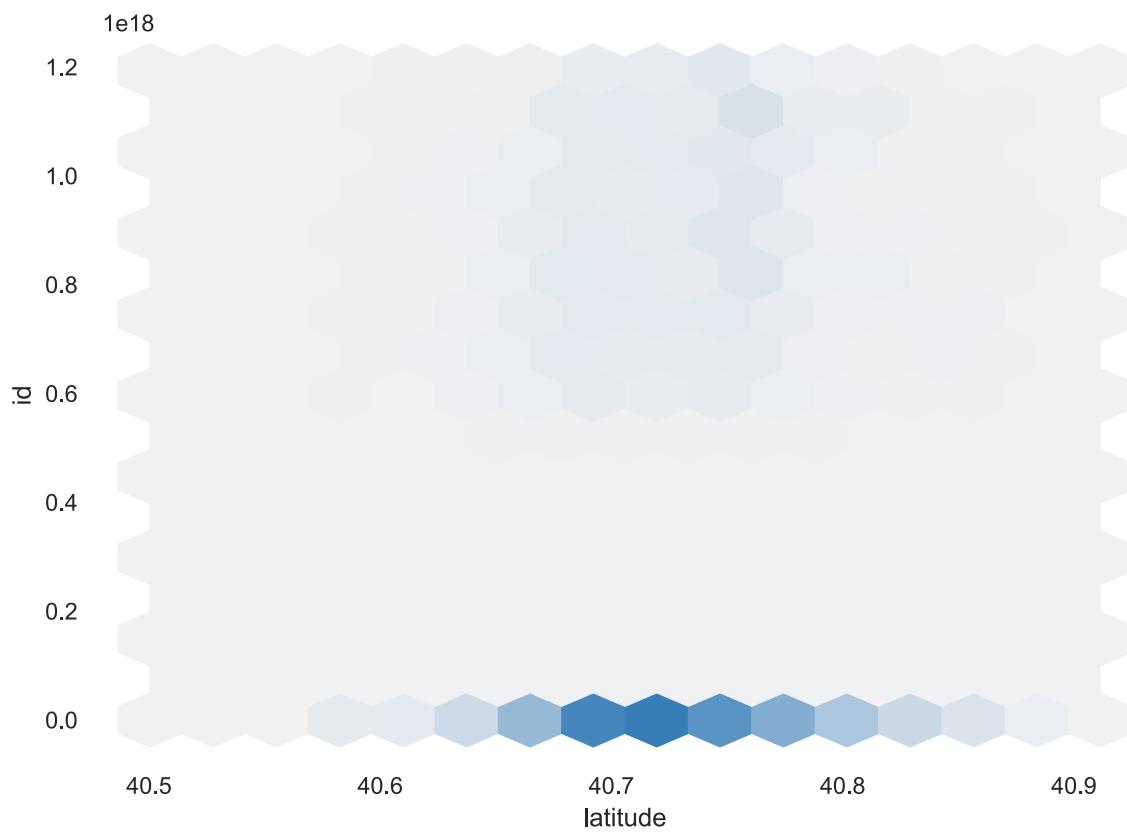


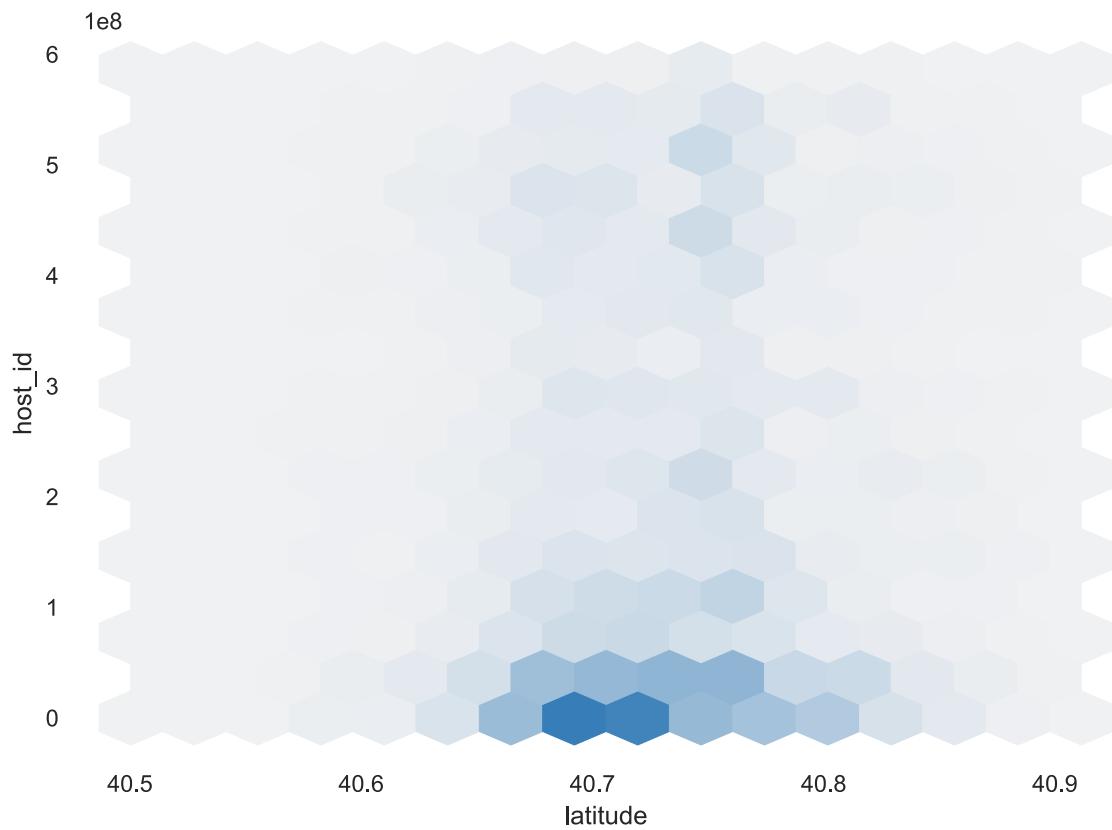


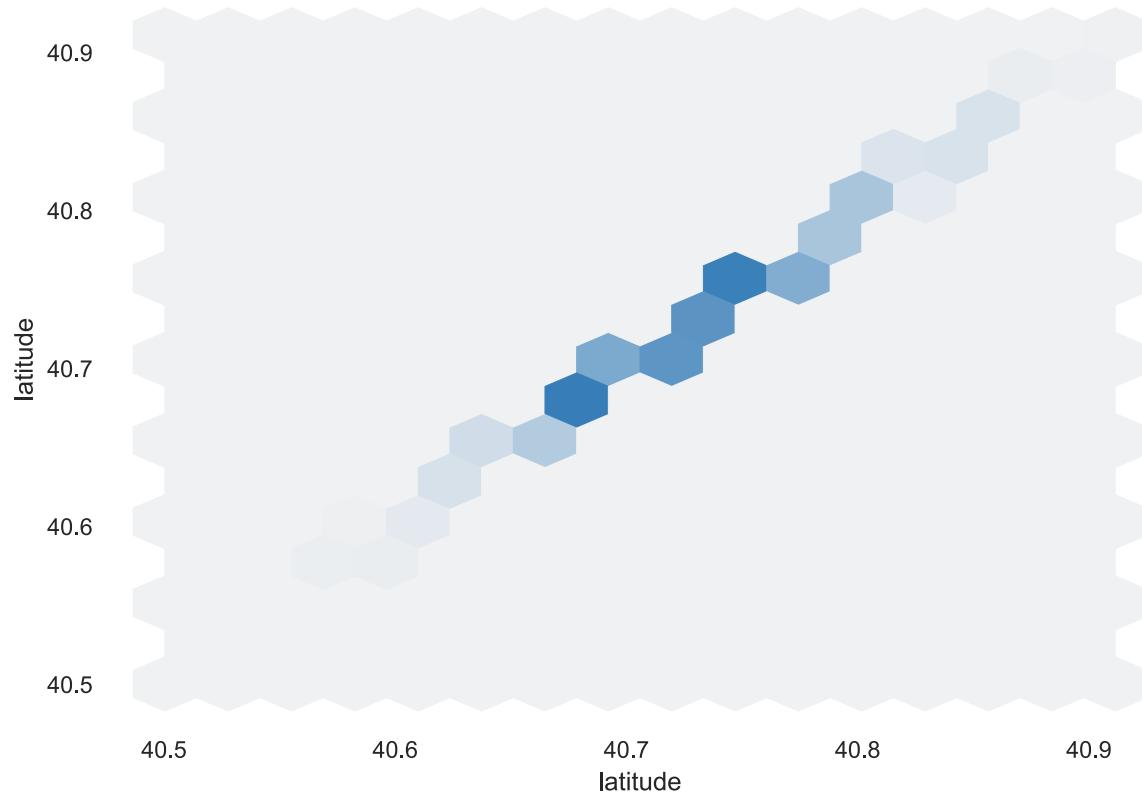


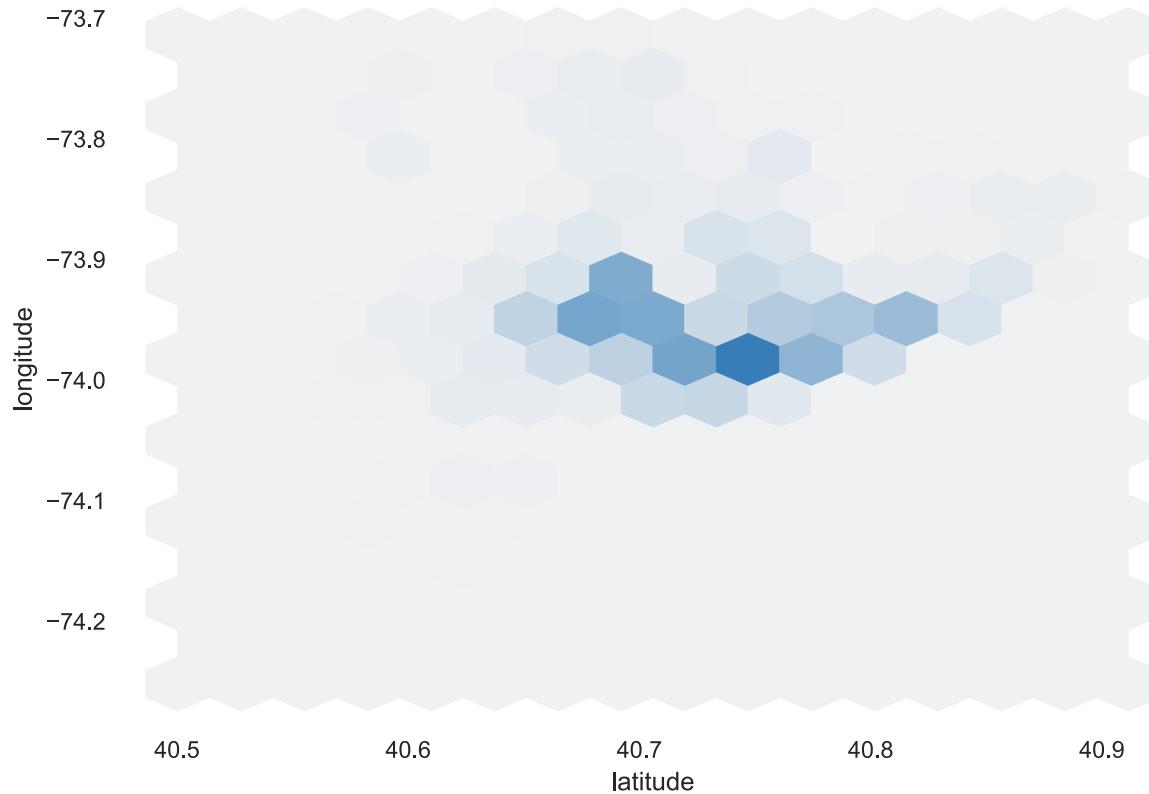

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

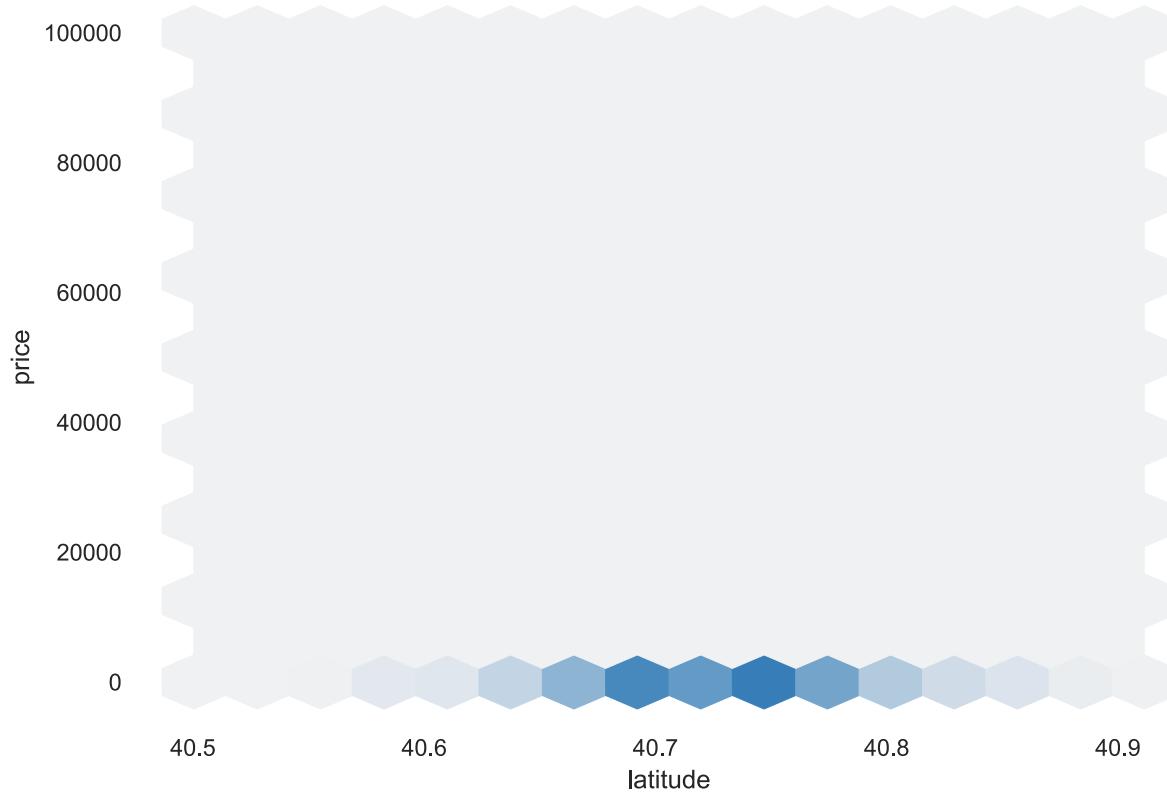


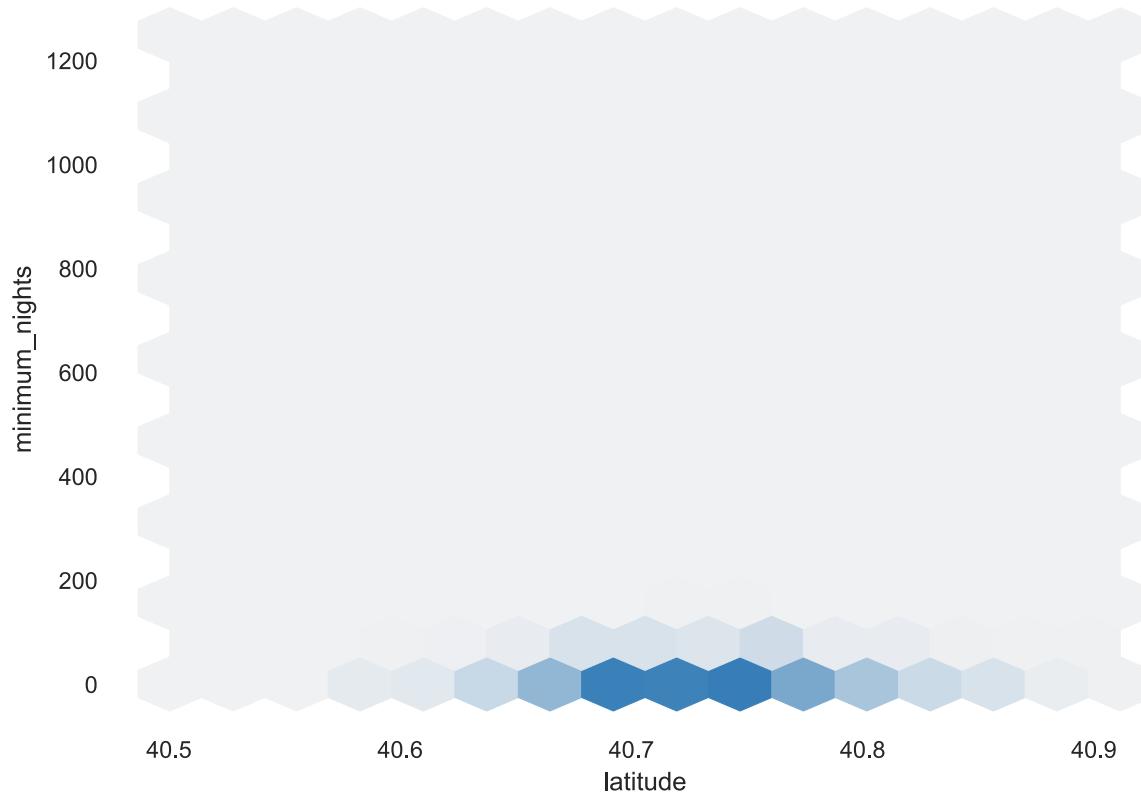


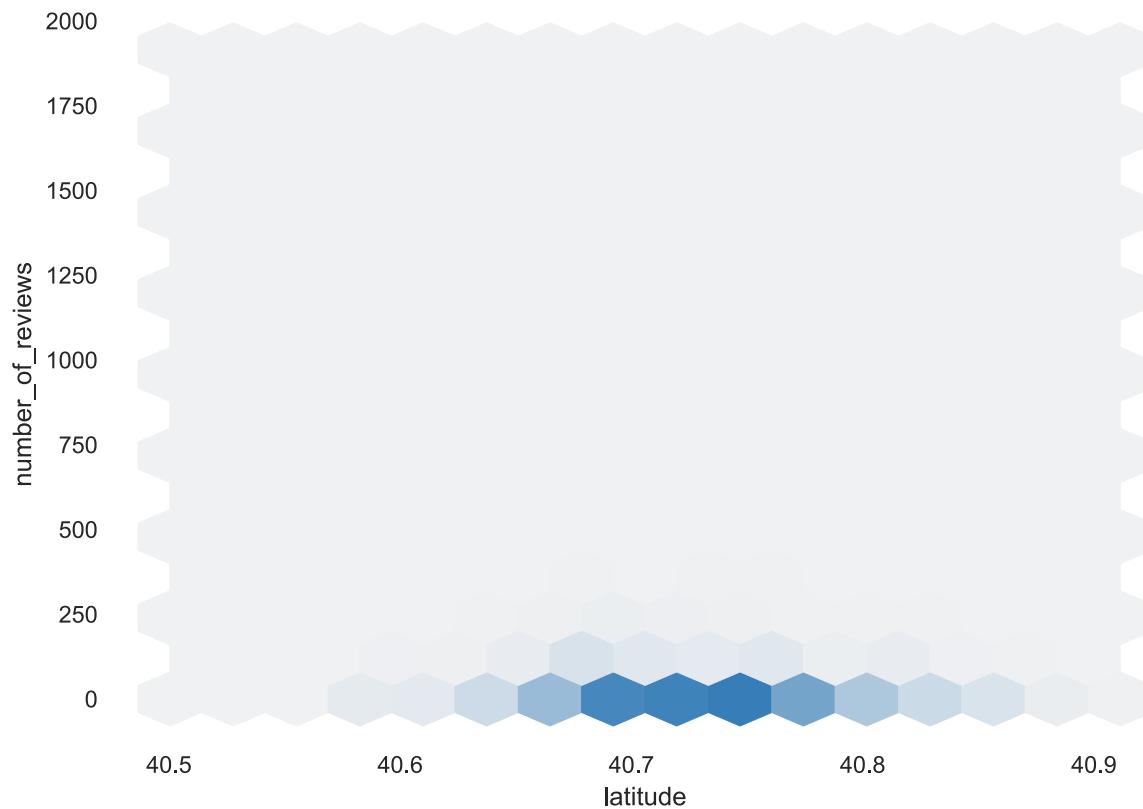


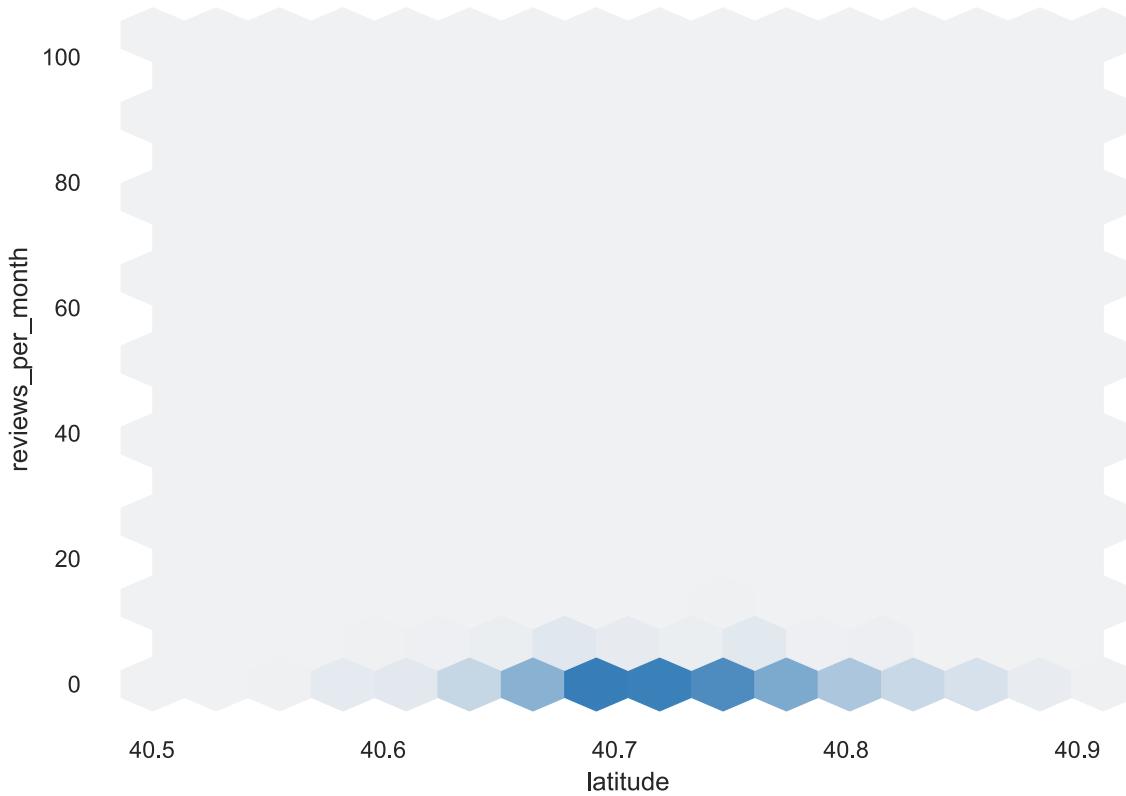


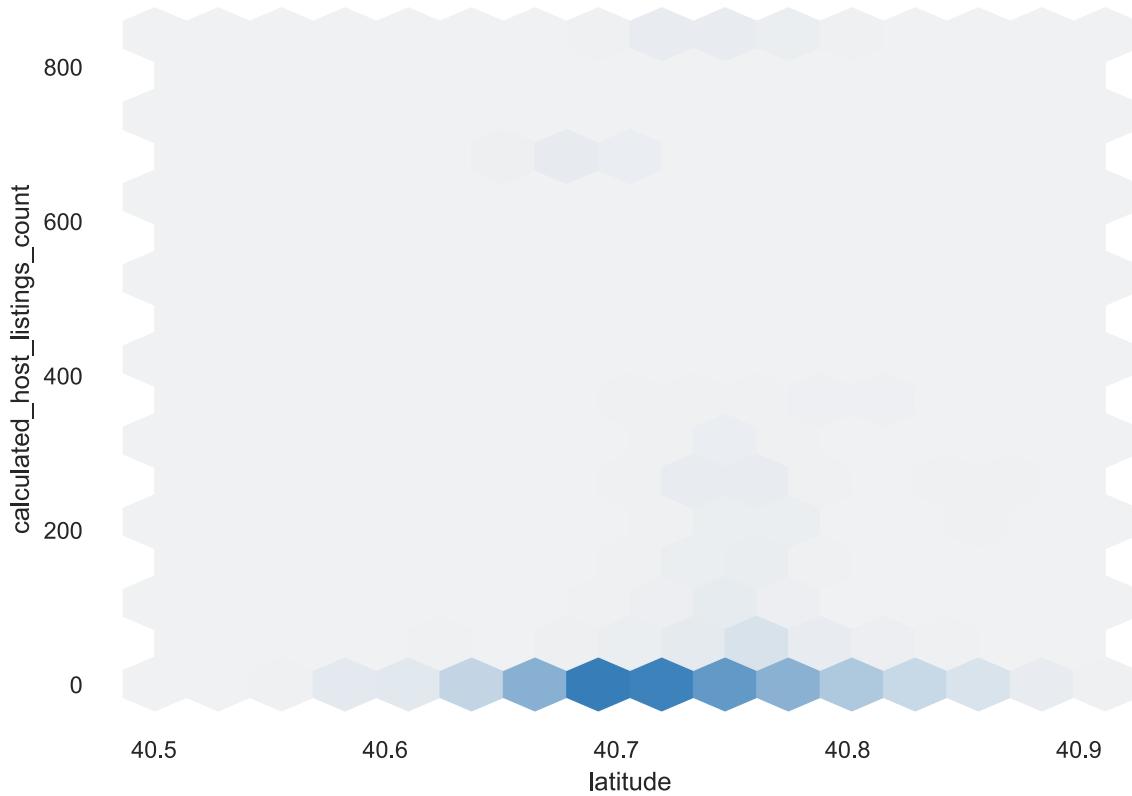


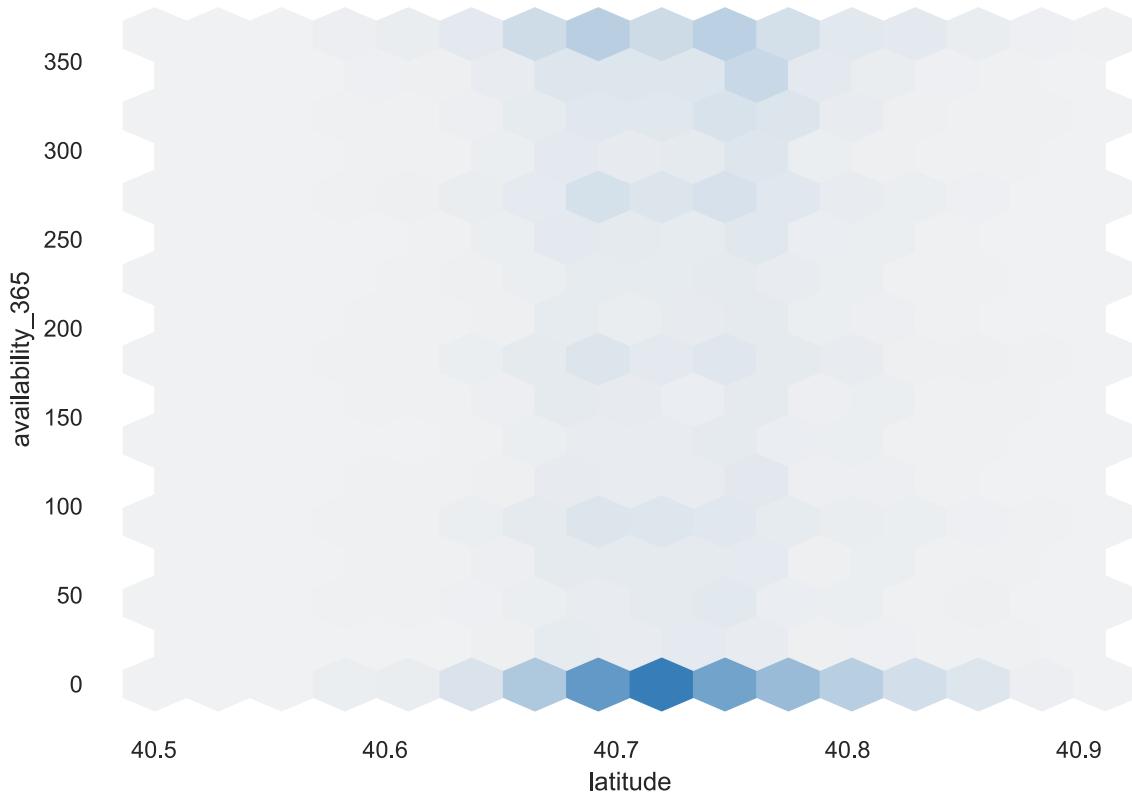




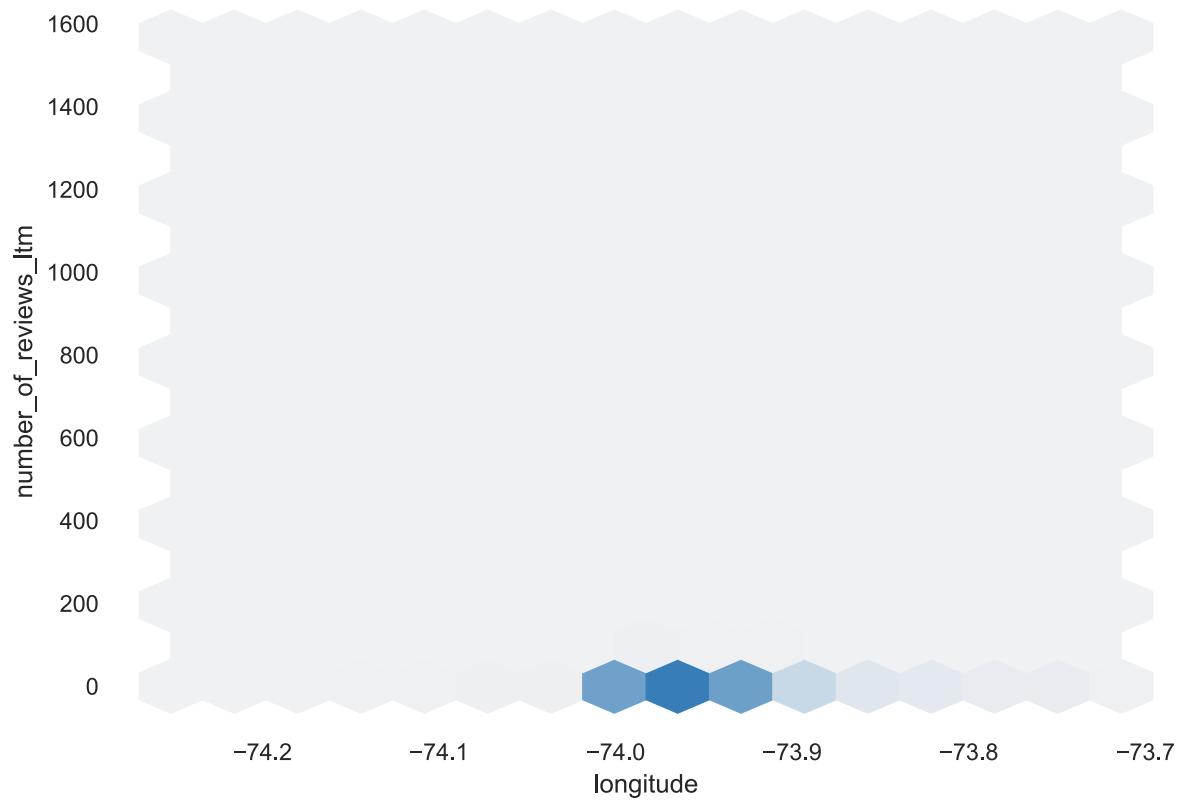


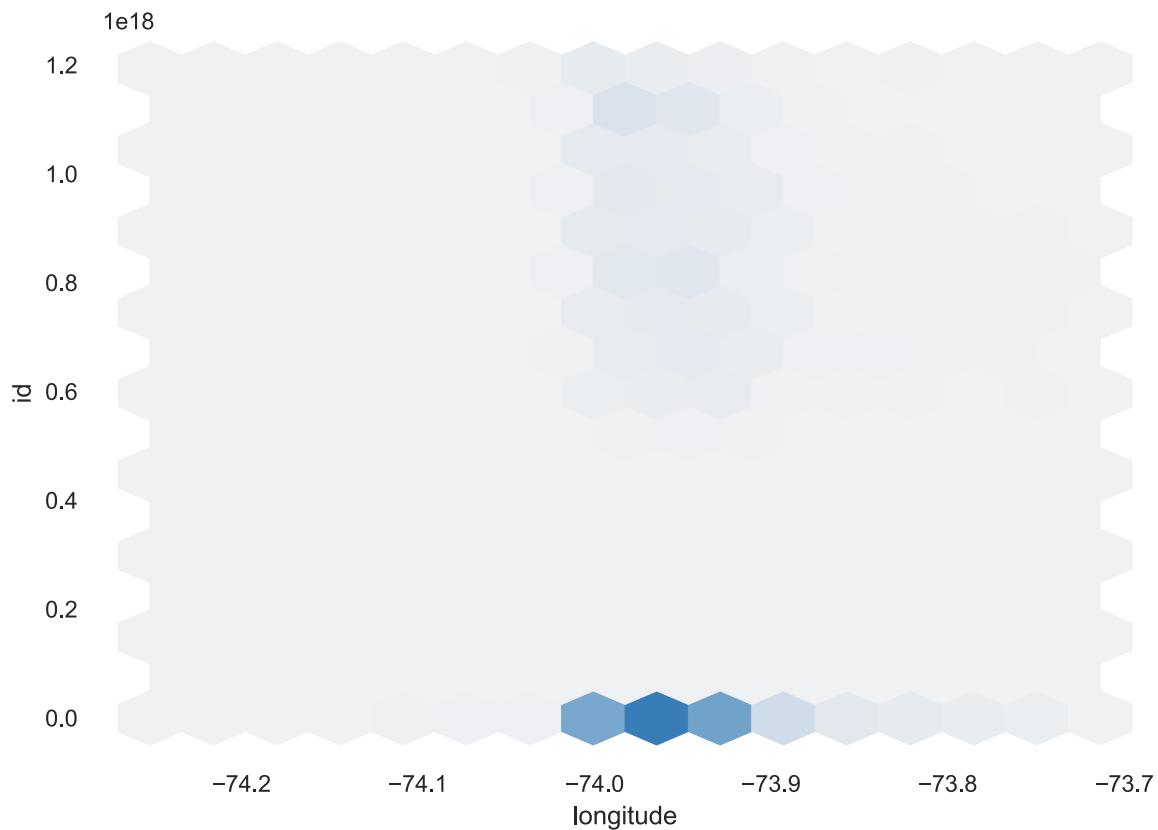


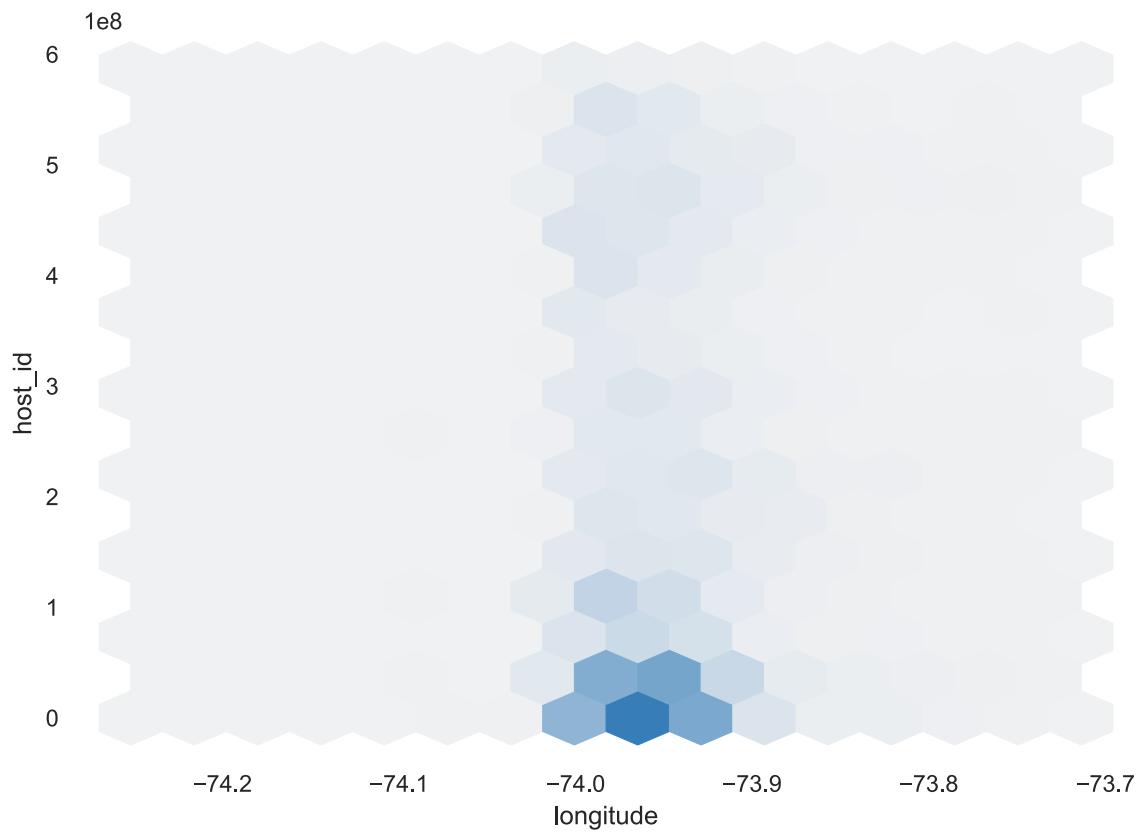


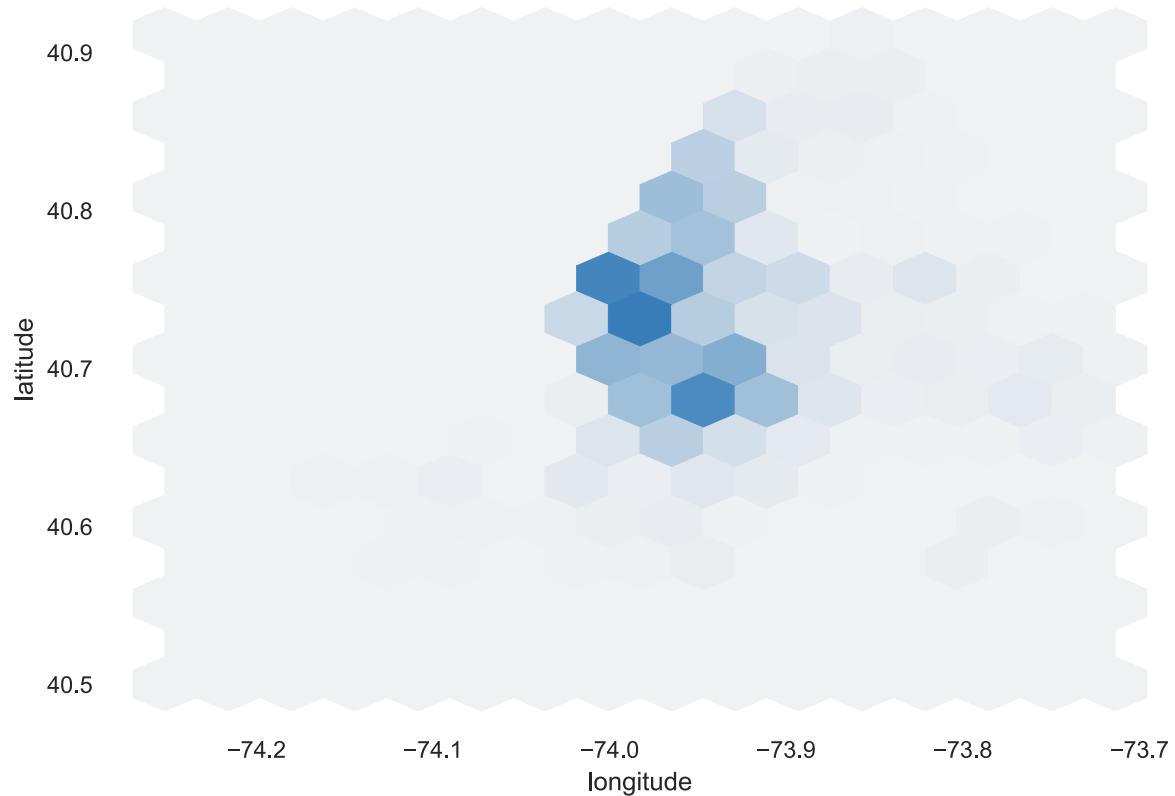


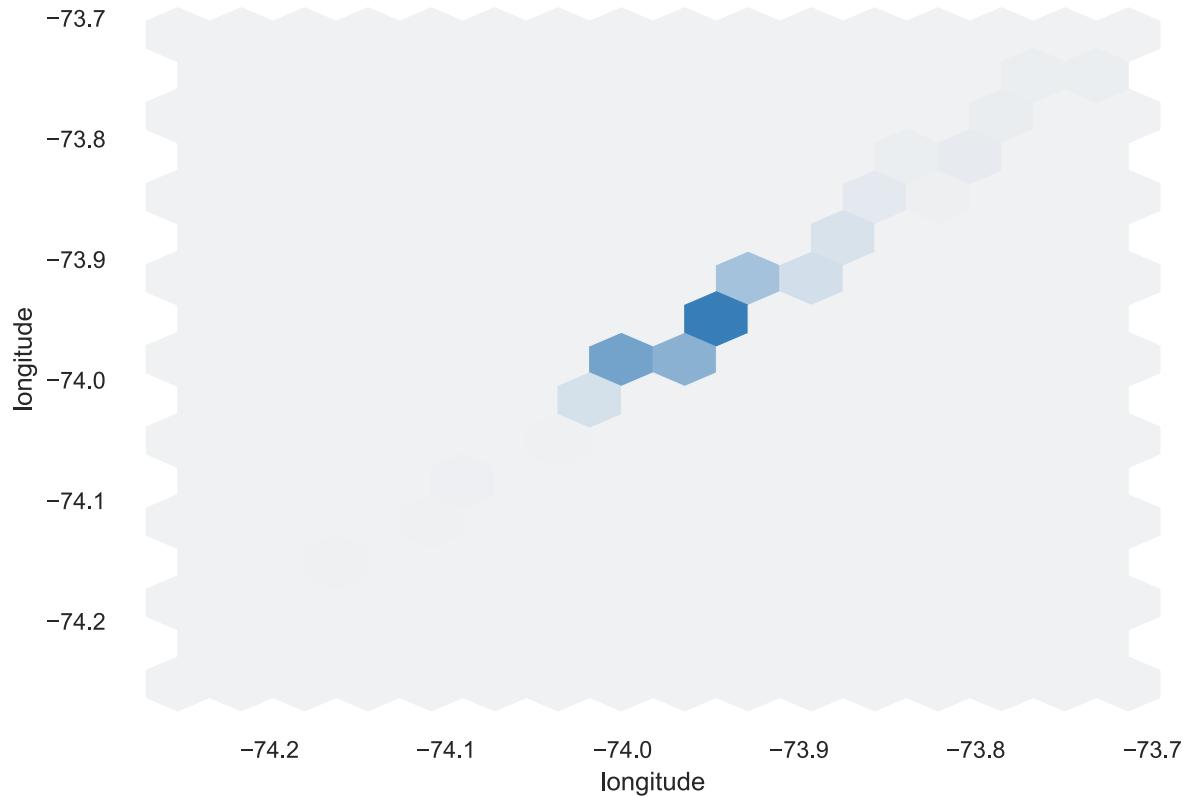

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

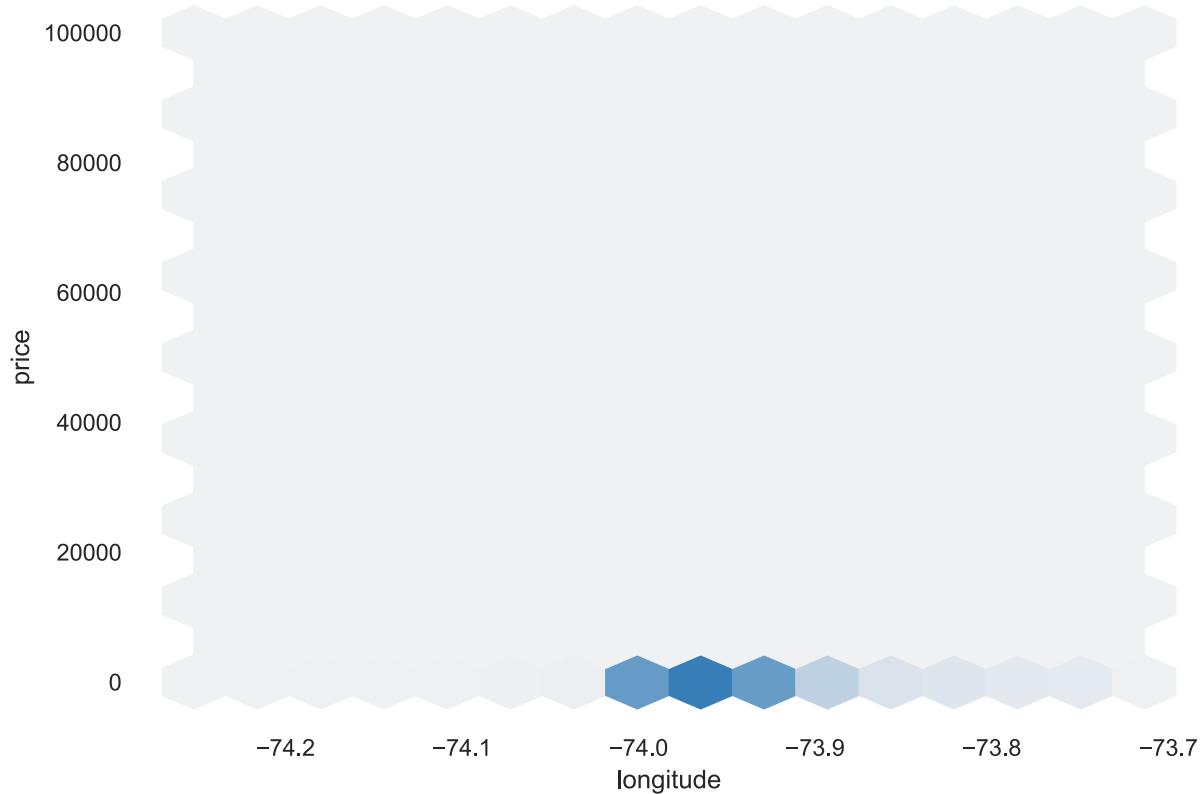


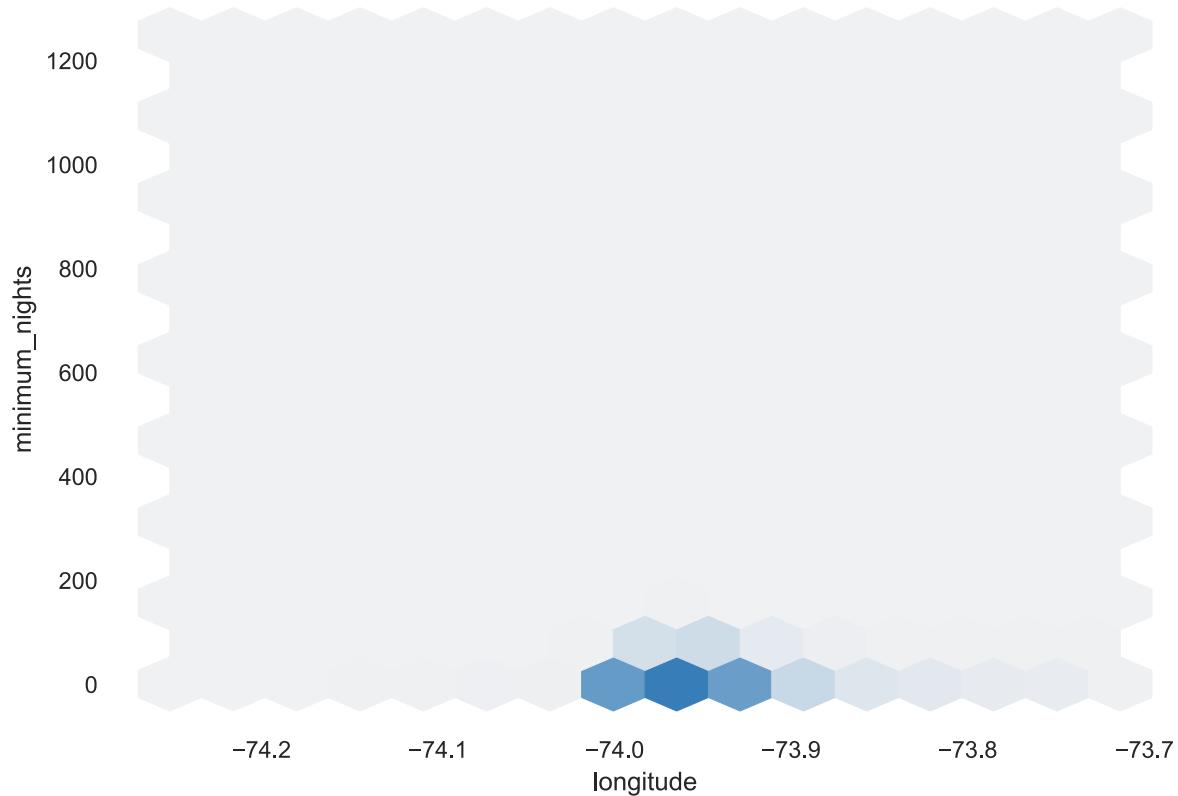


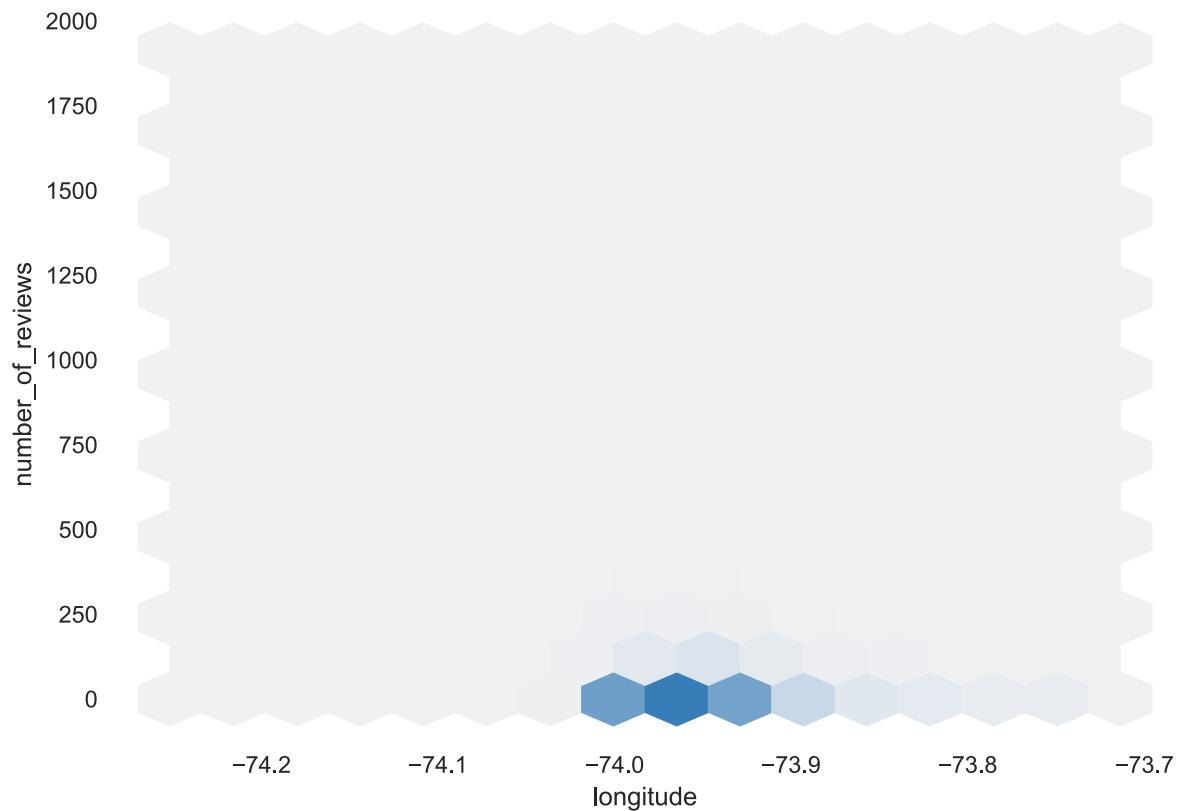


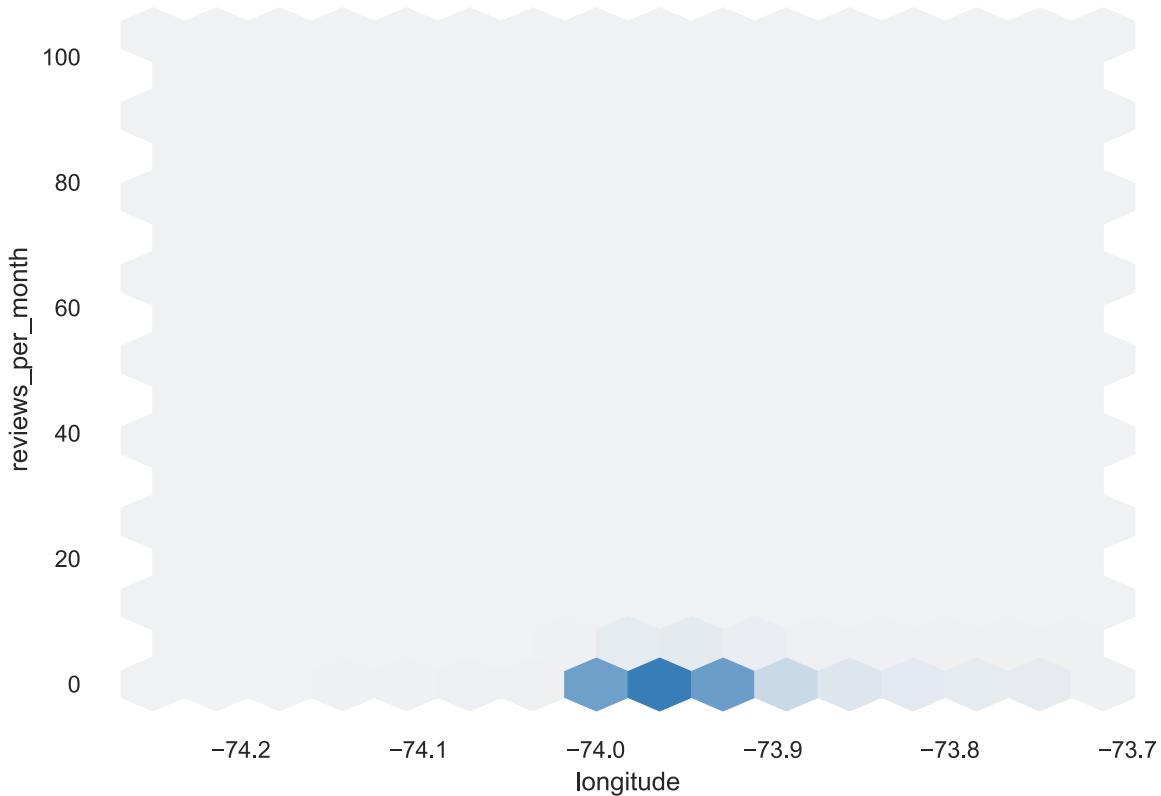


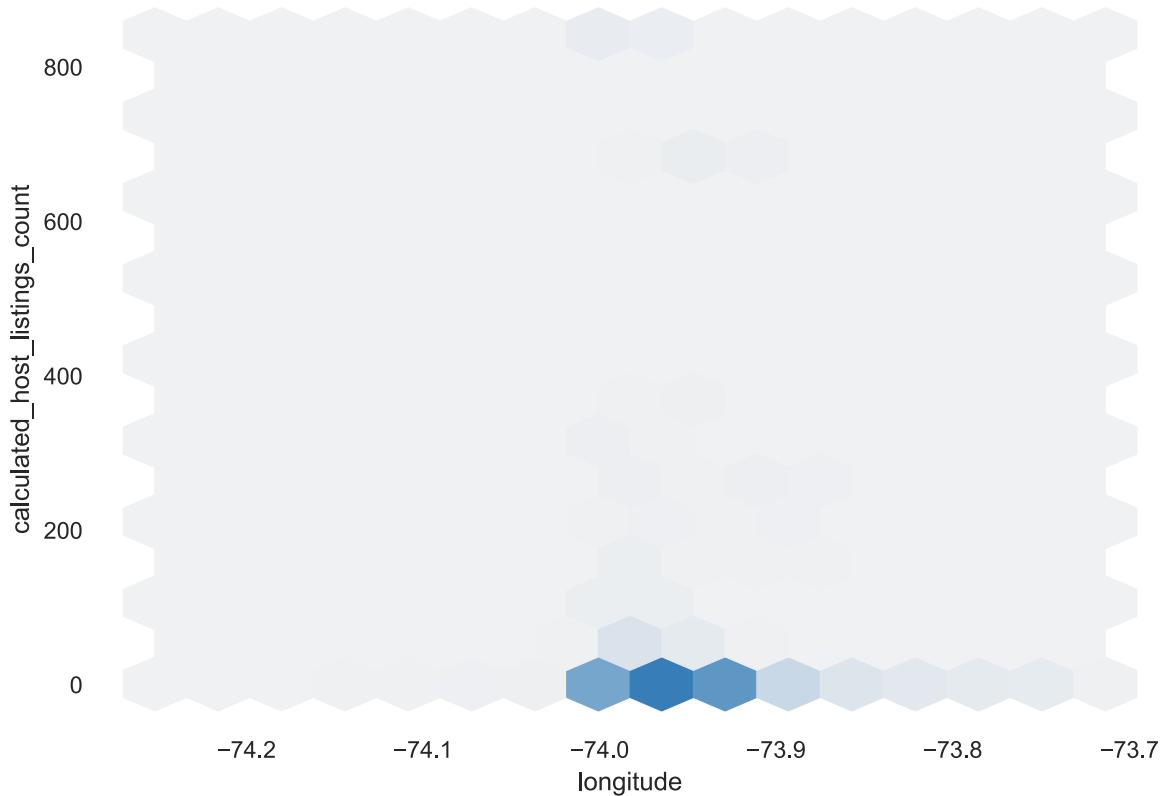


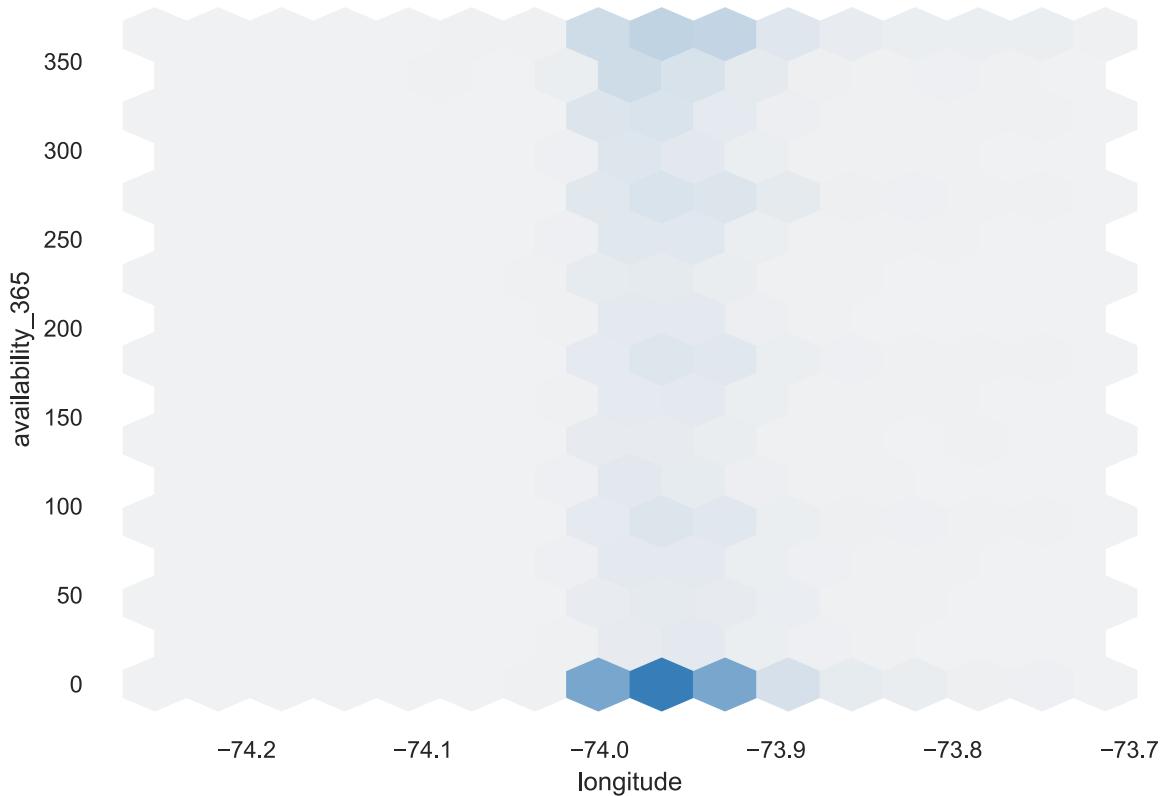




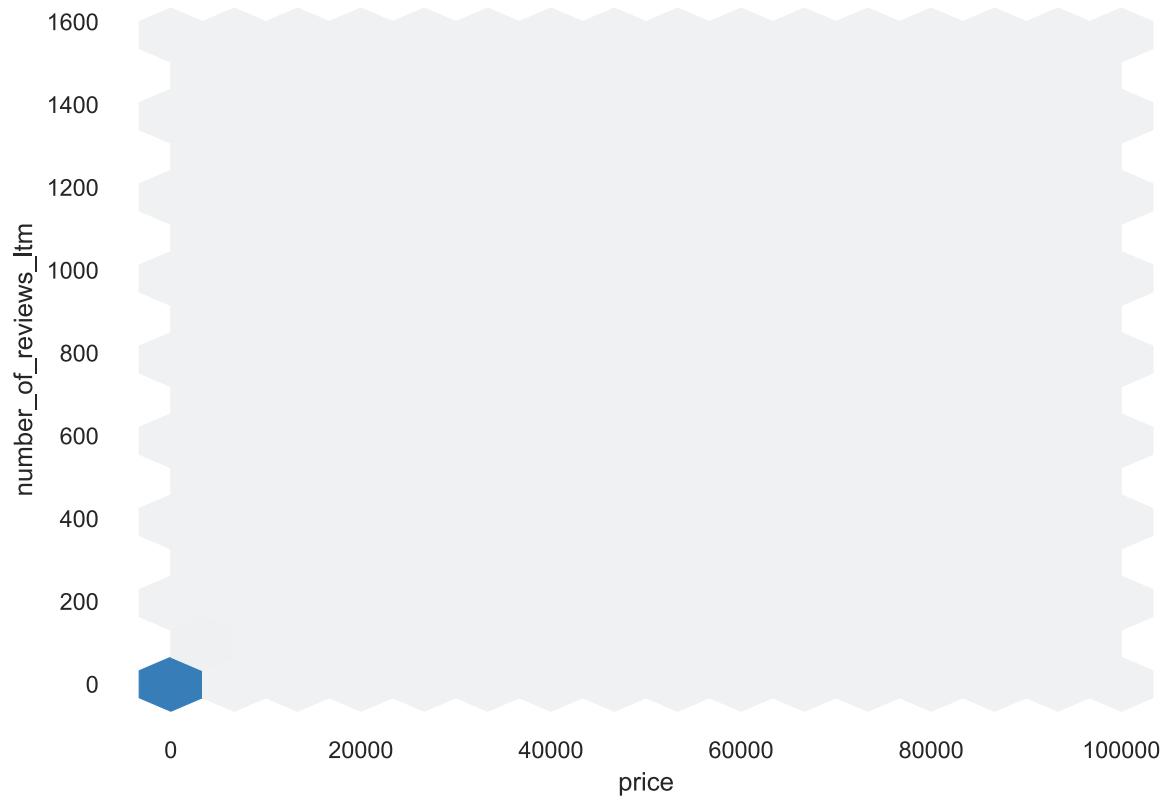


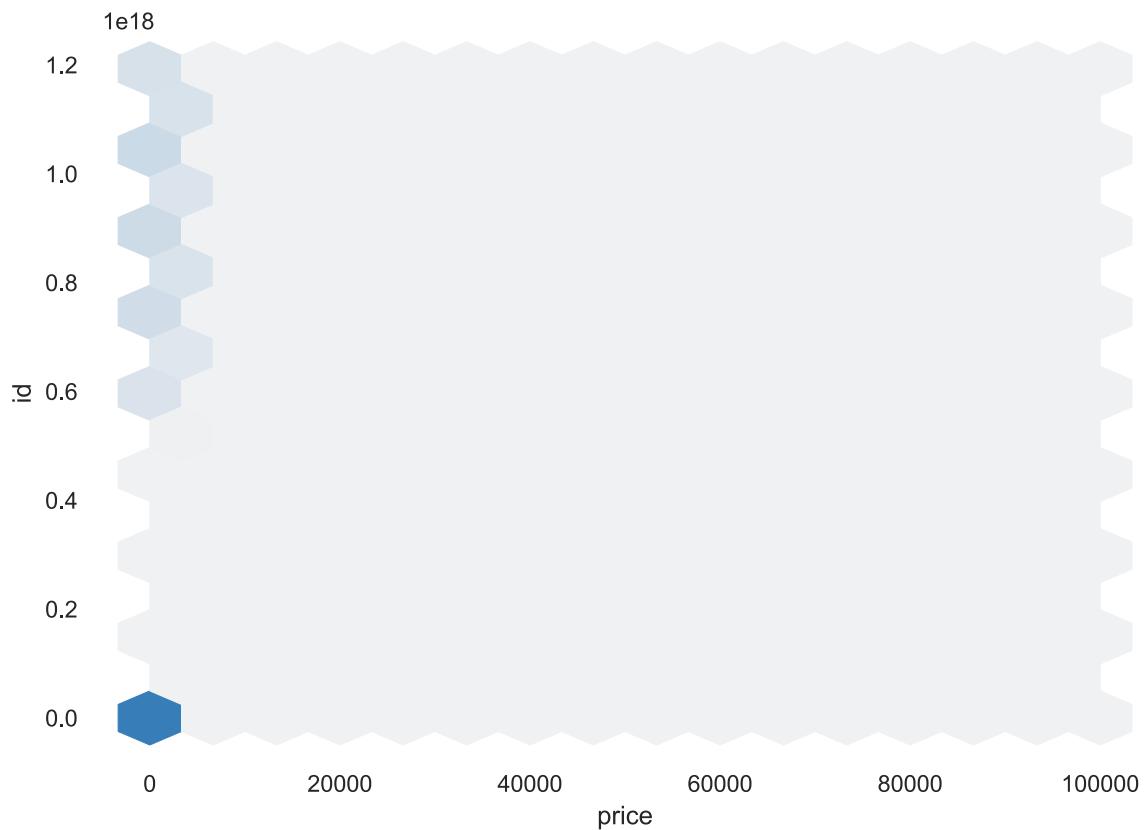


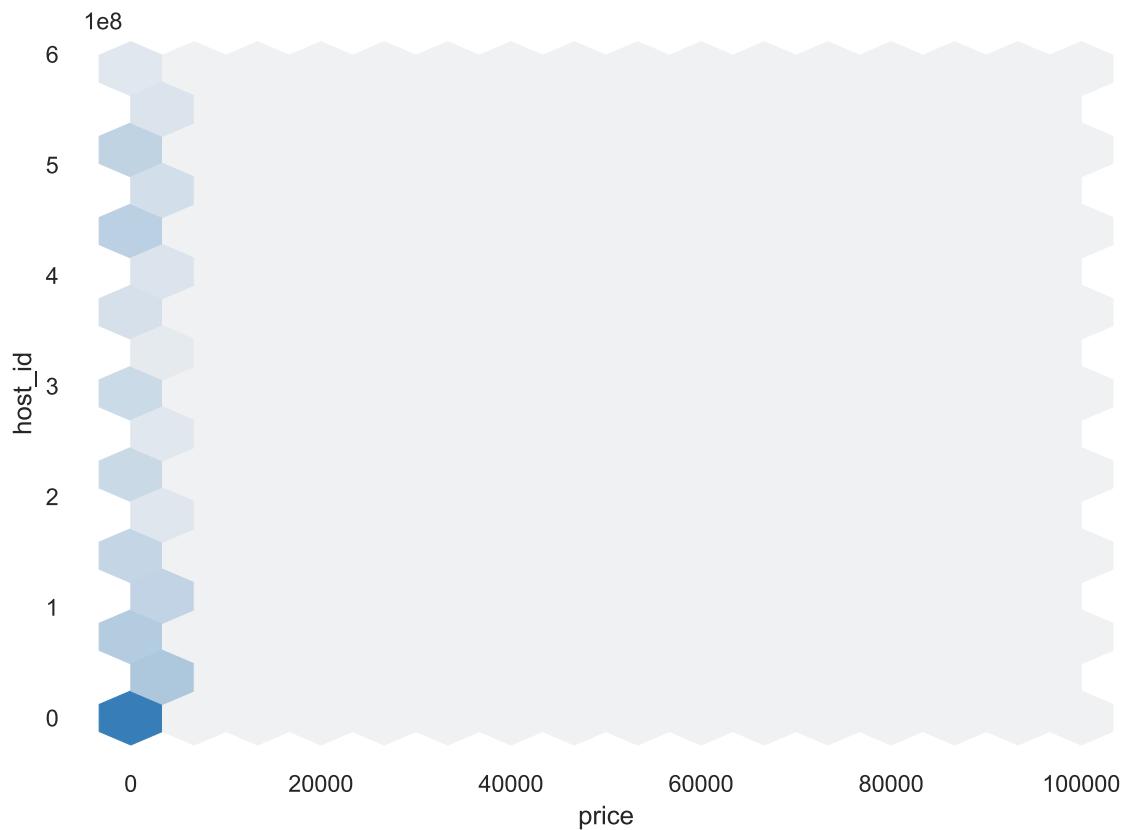


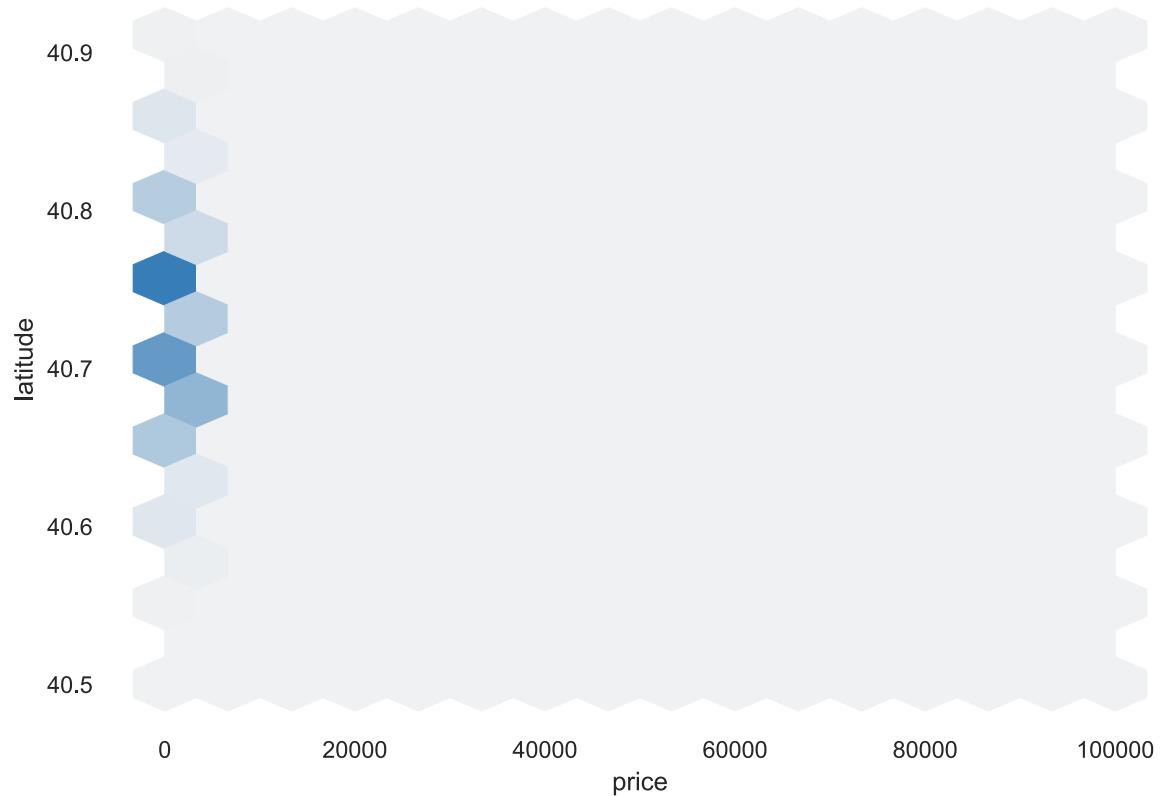


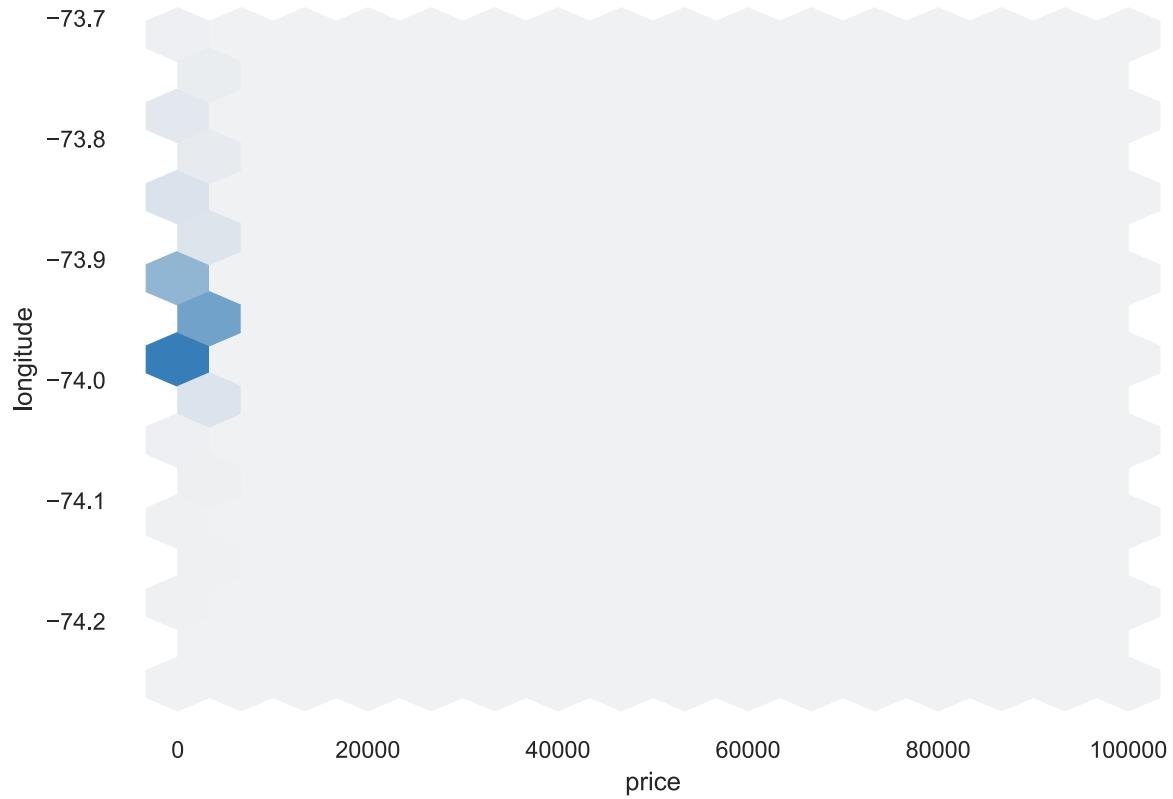
number_of_reviews_ltm
id
host_id
latitude
longitude
price
minimum_nights
number_of_reviews
reviews_per_month
calculated_host_listings_c
availability_365

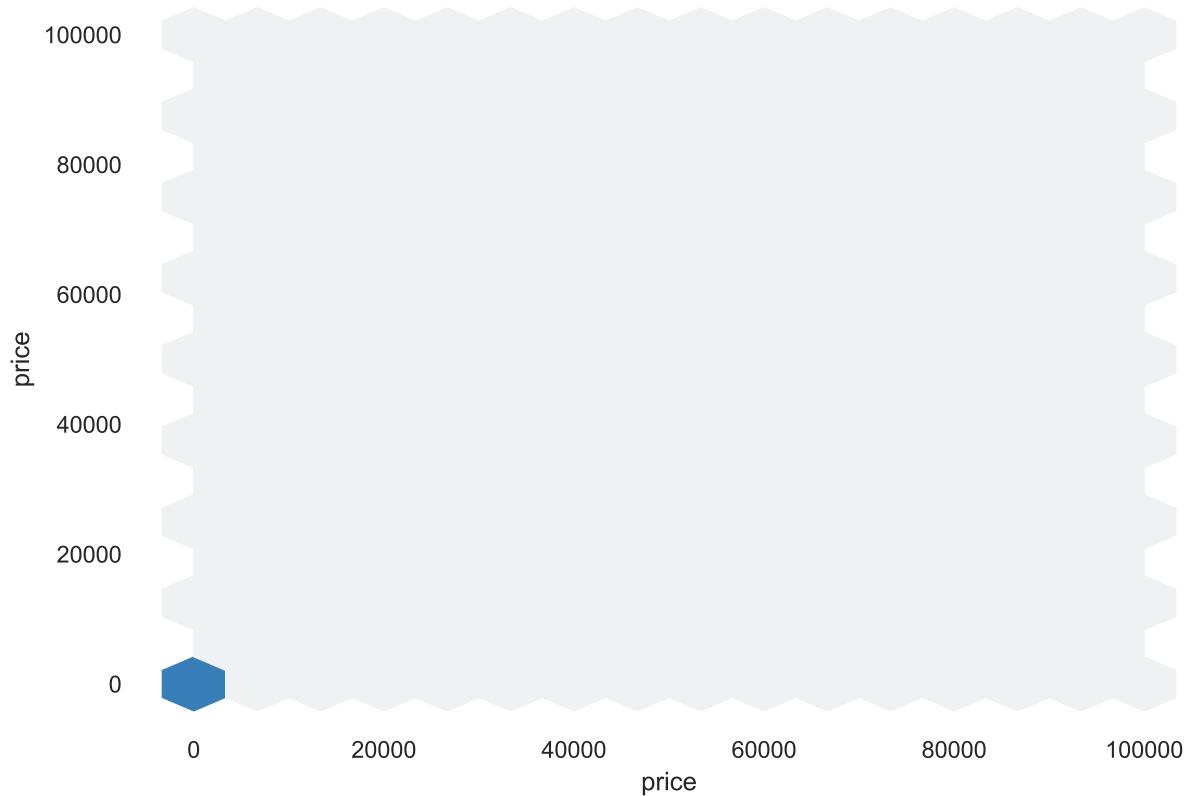


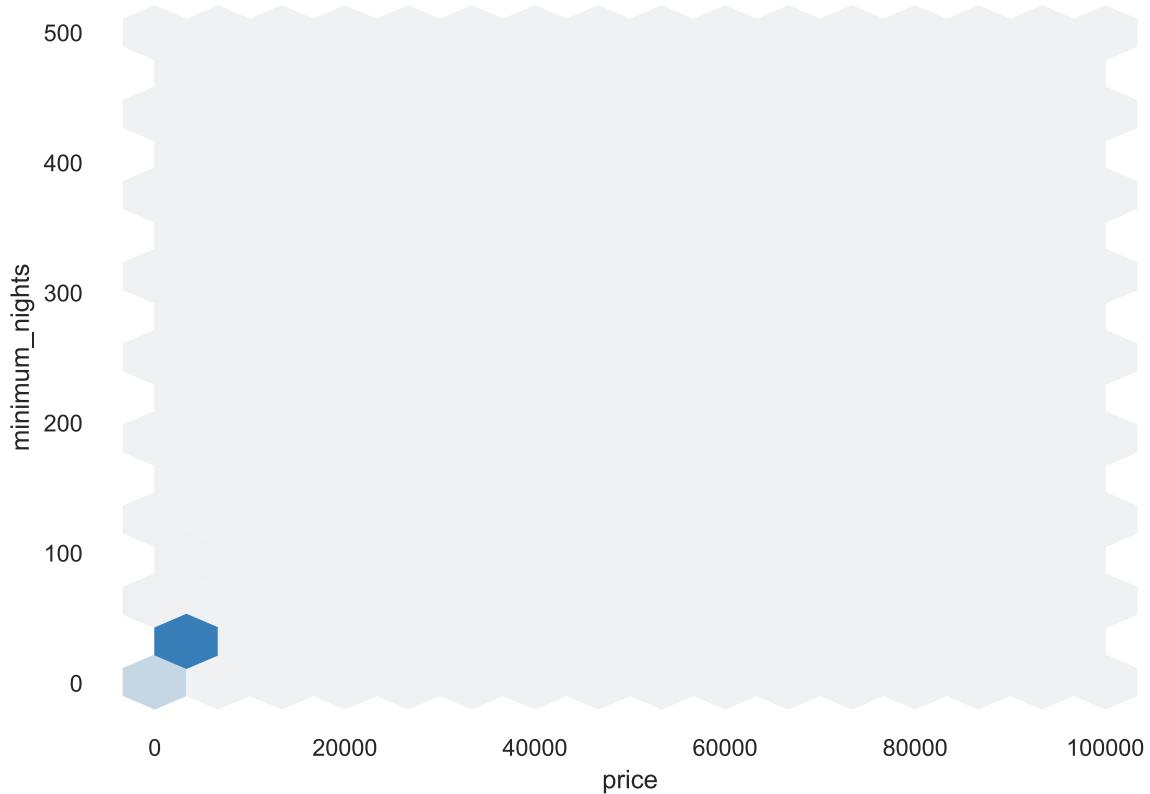


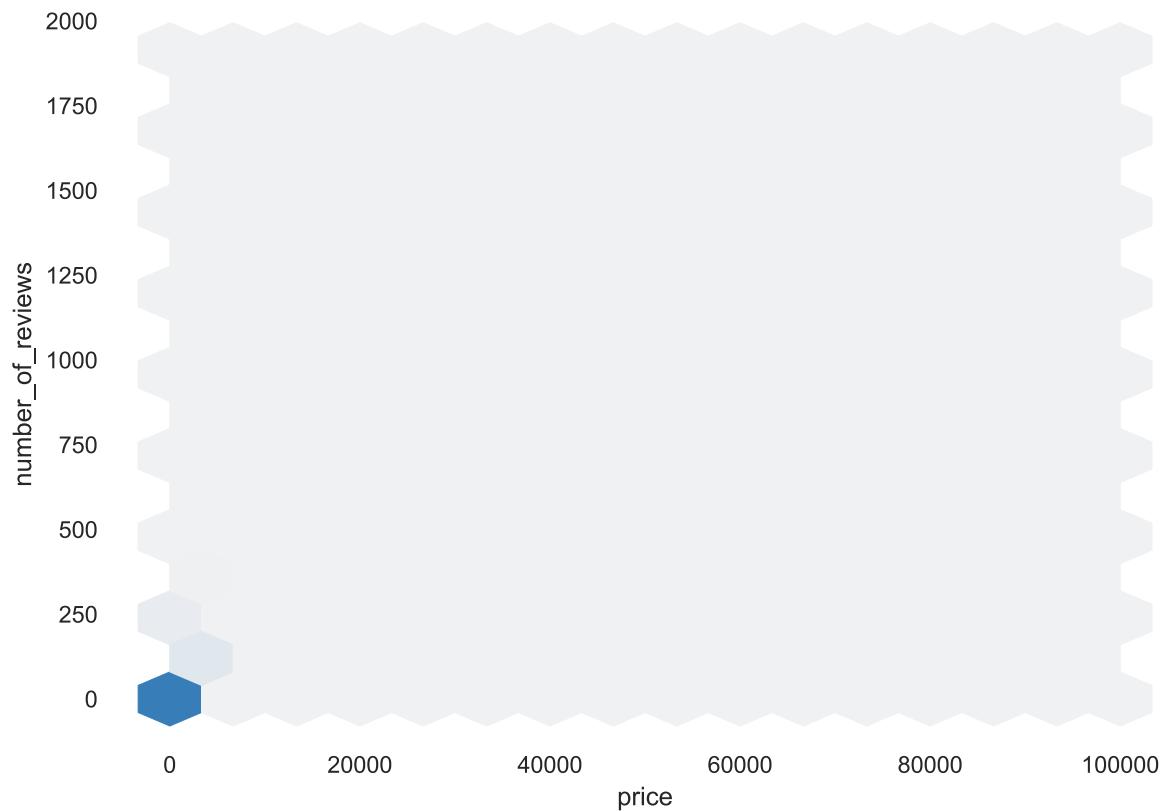


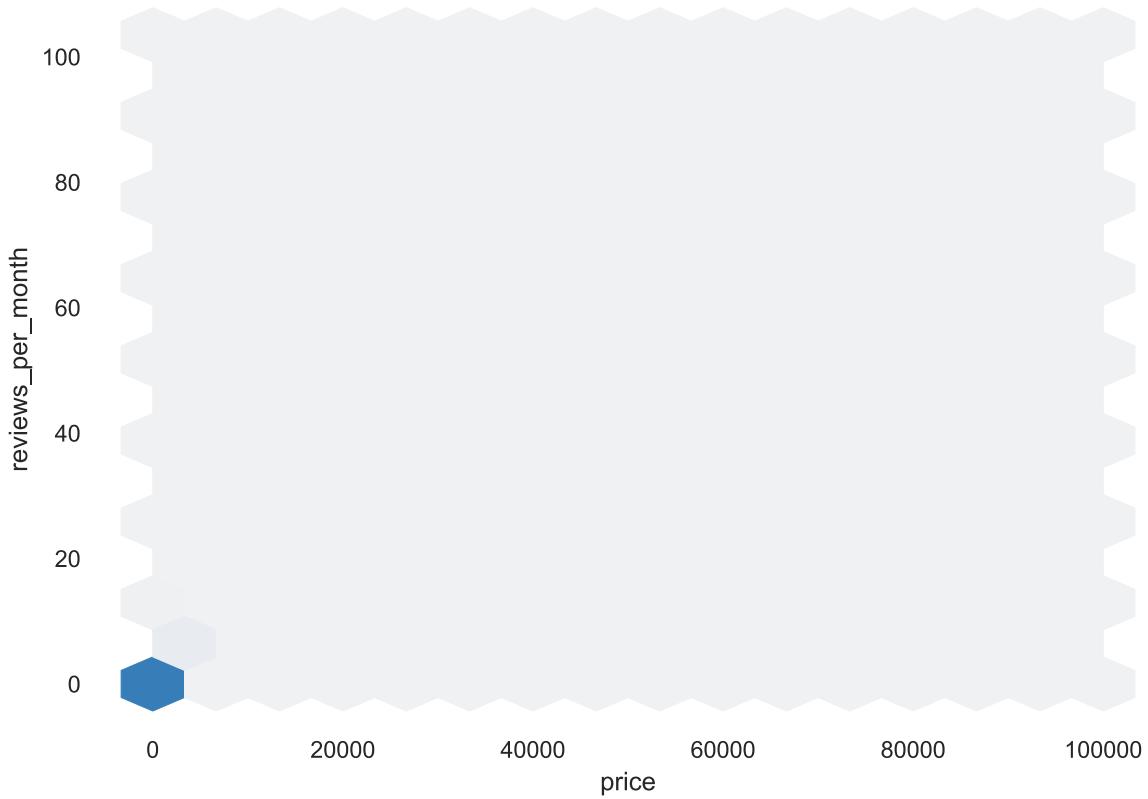


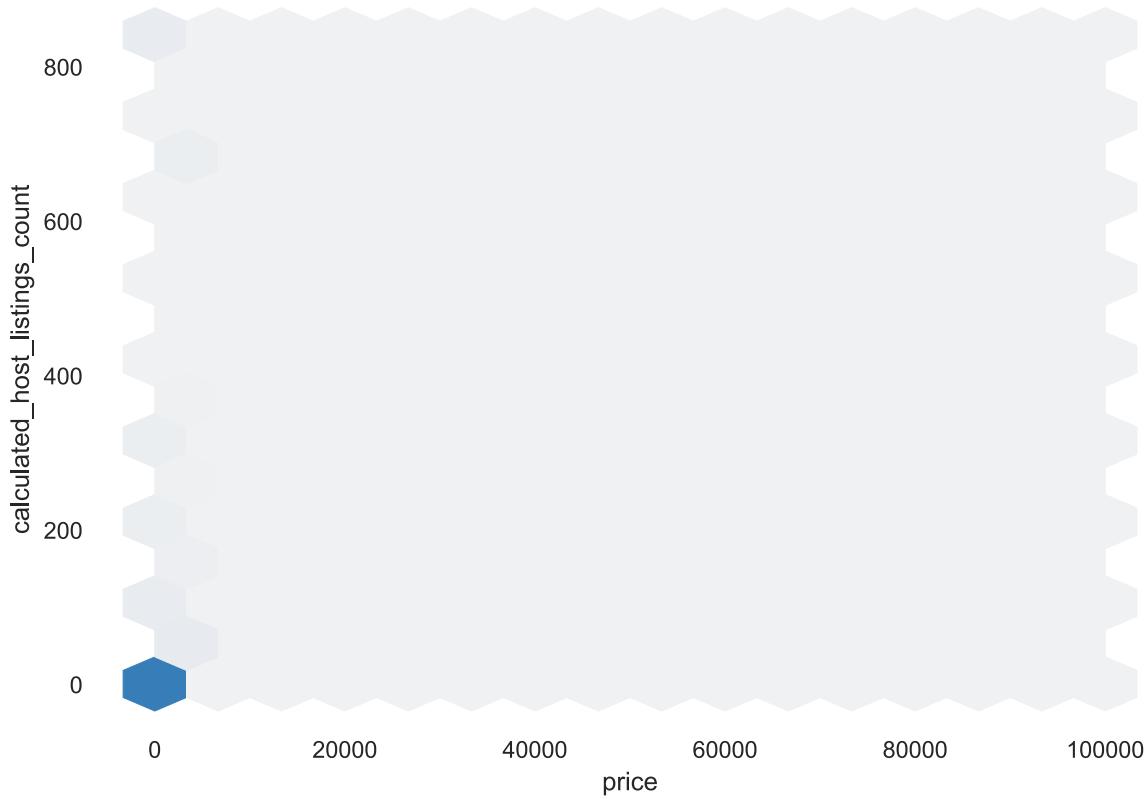


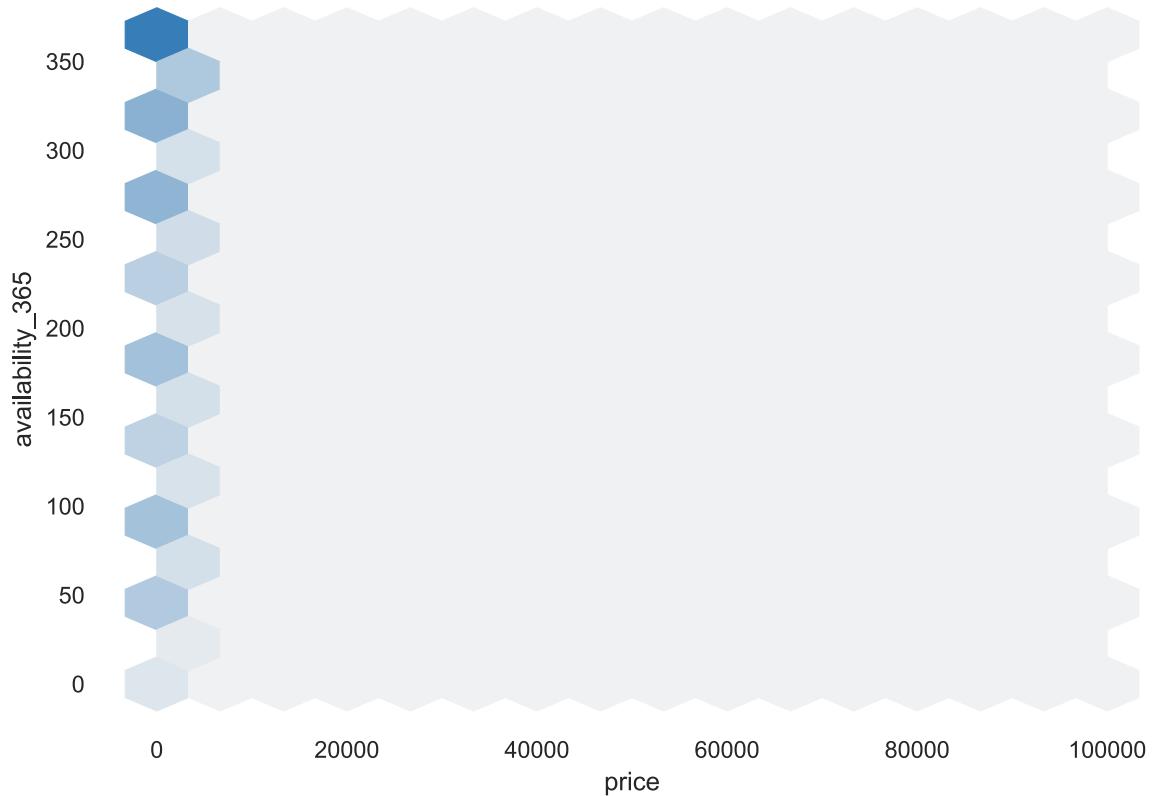




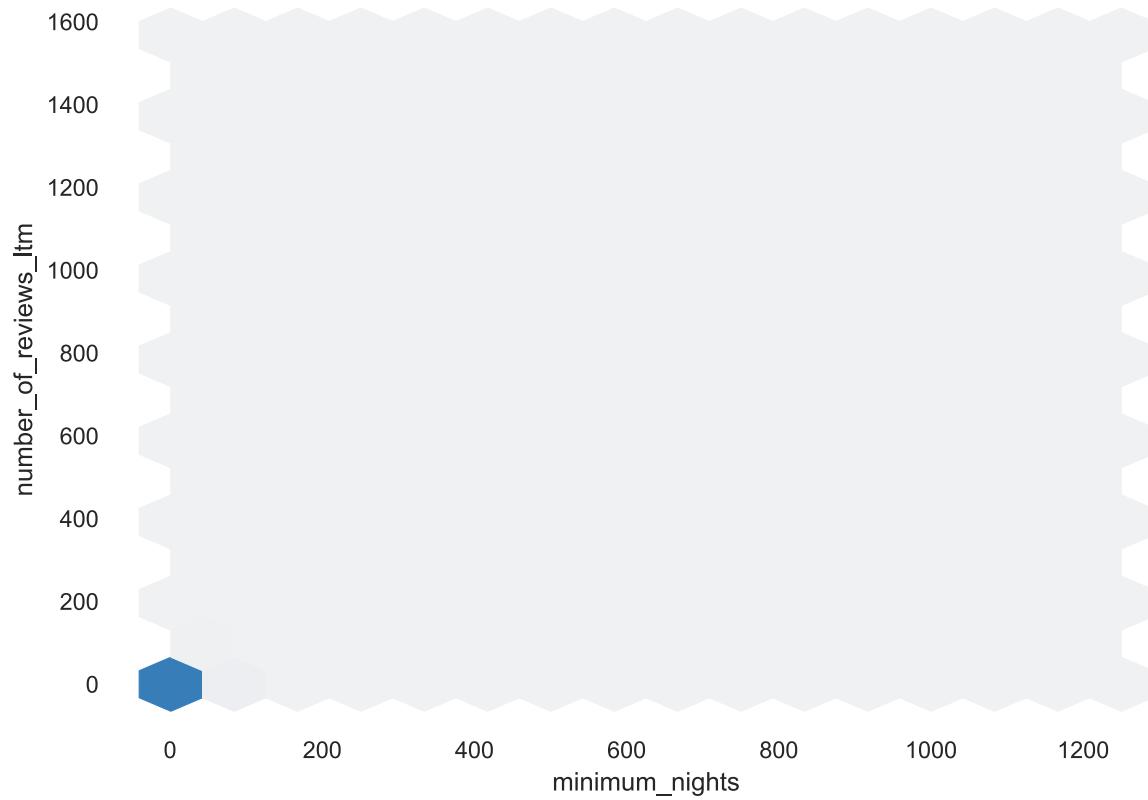


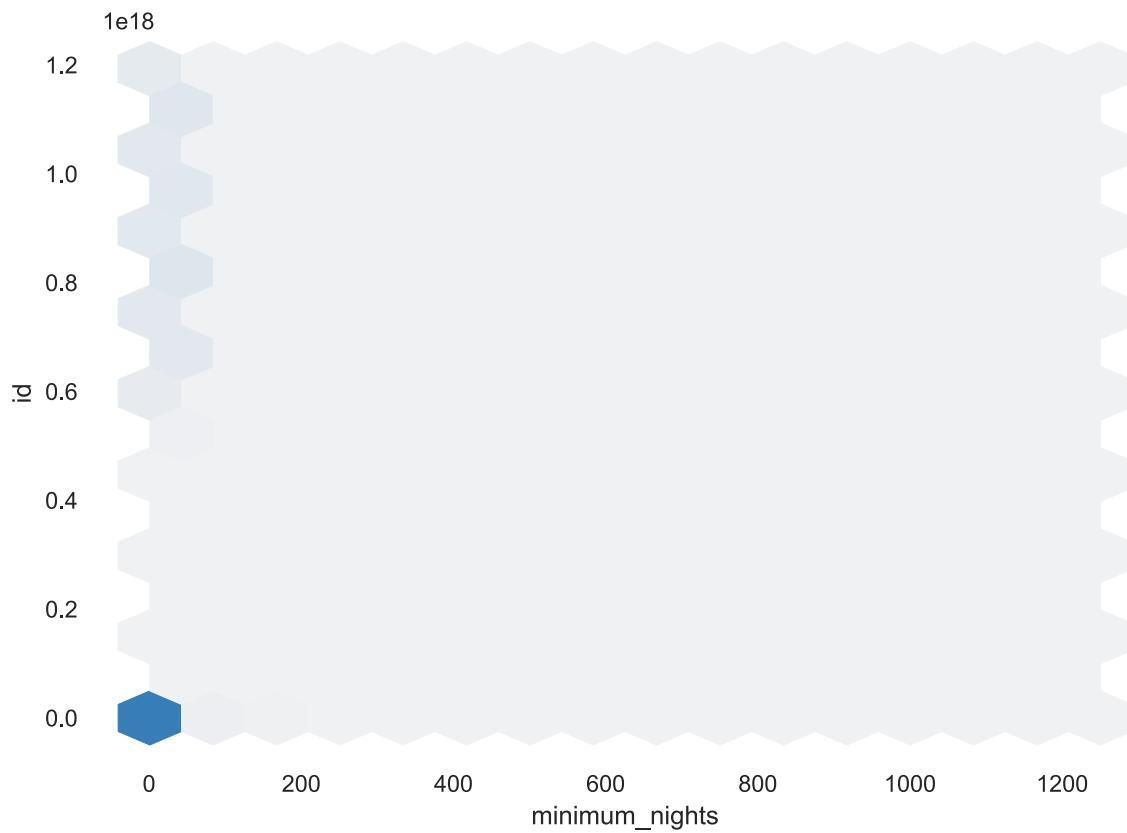


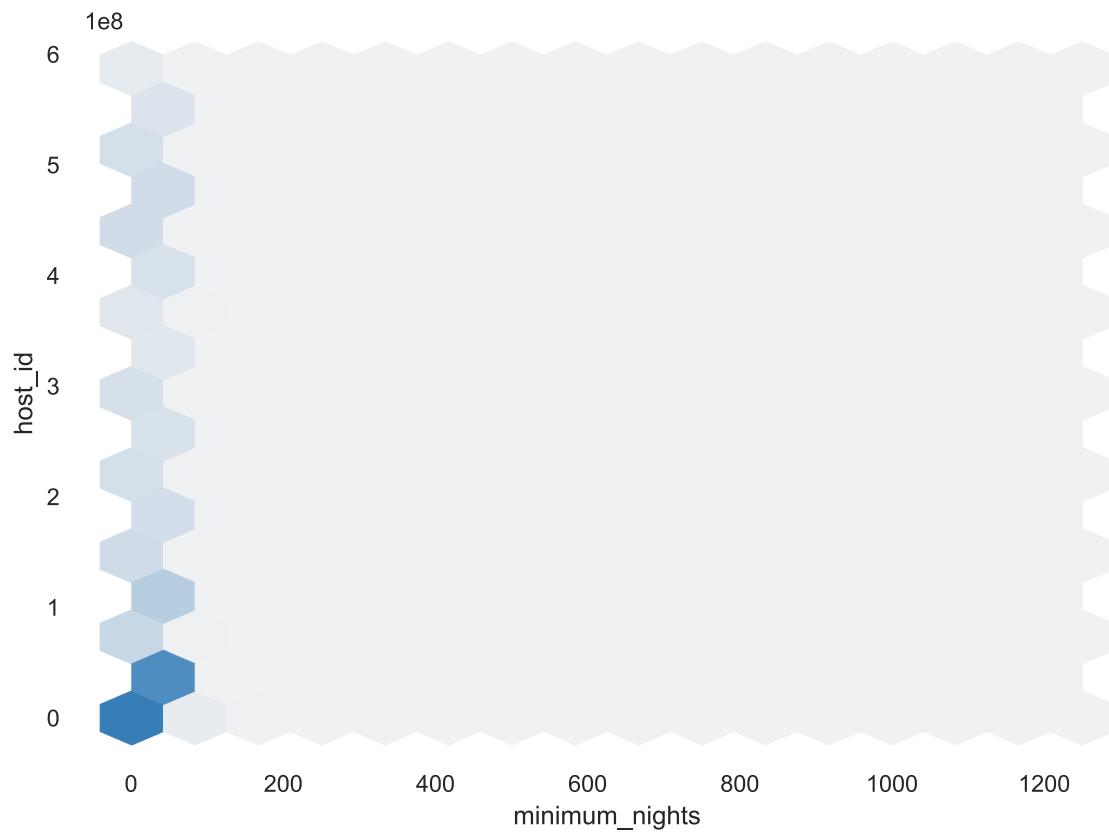


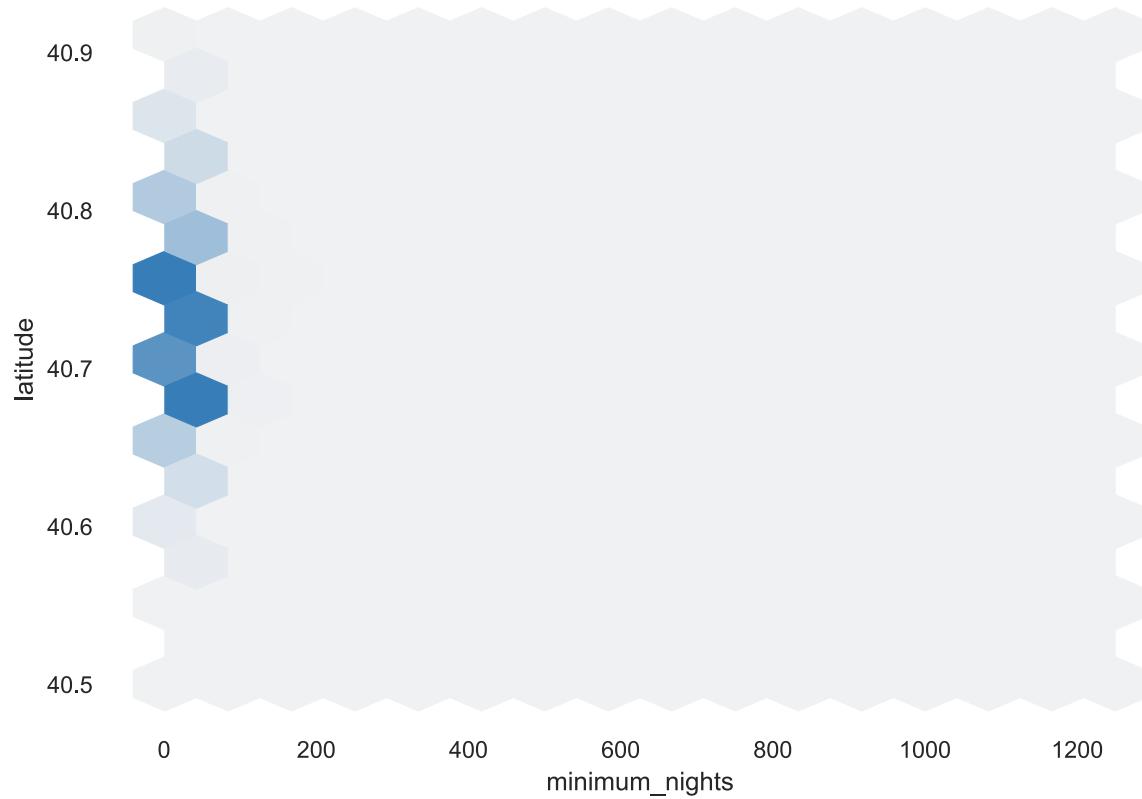


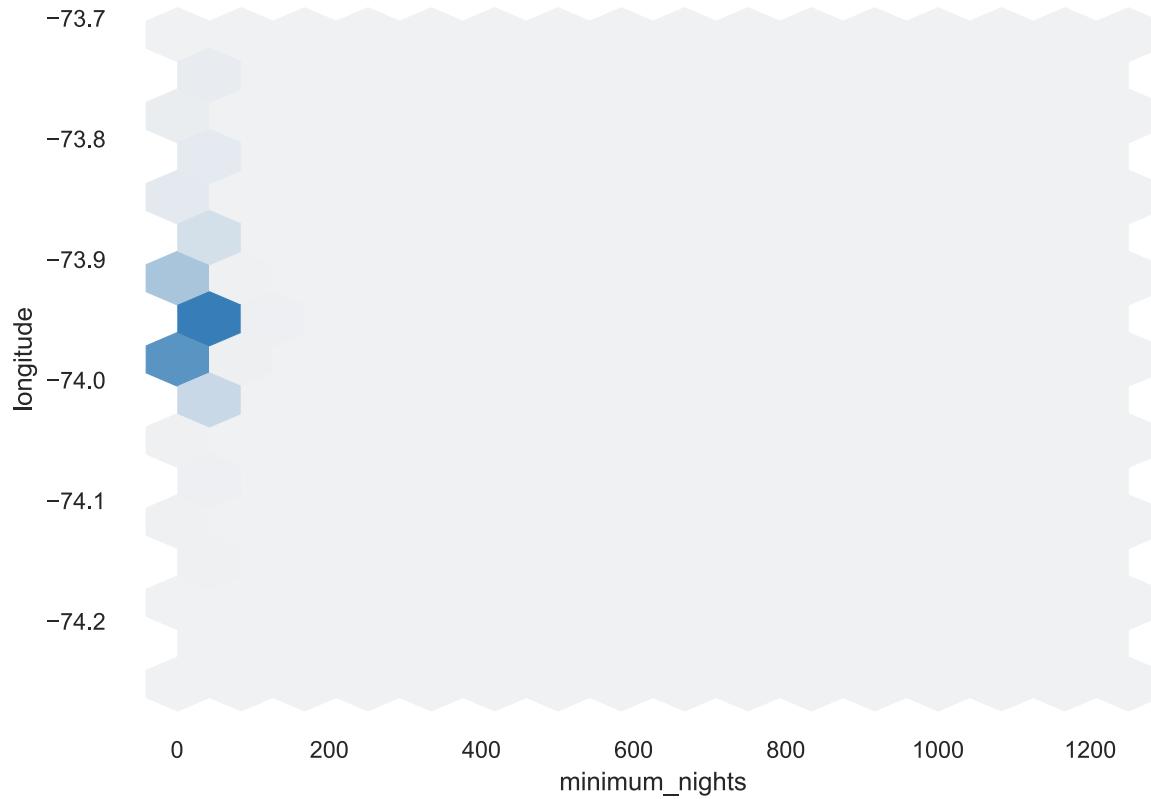

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

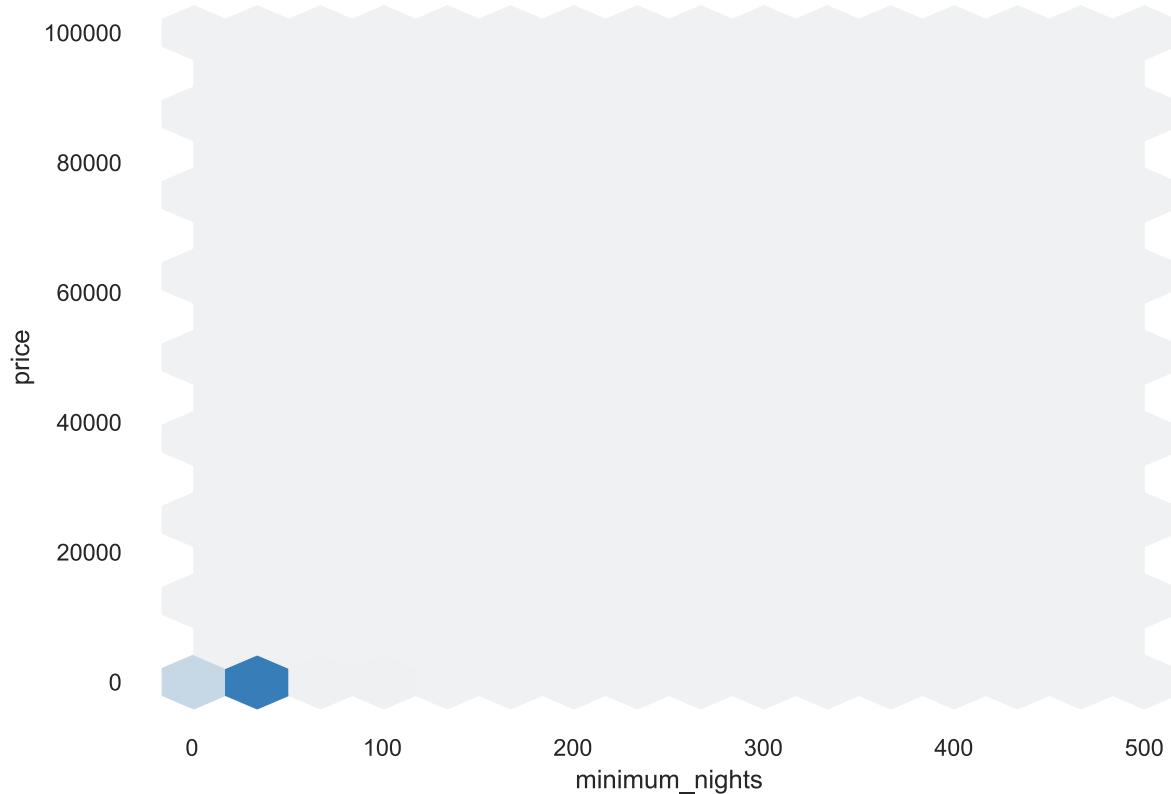


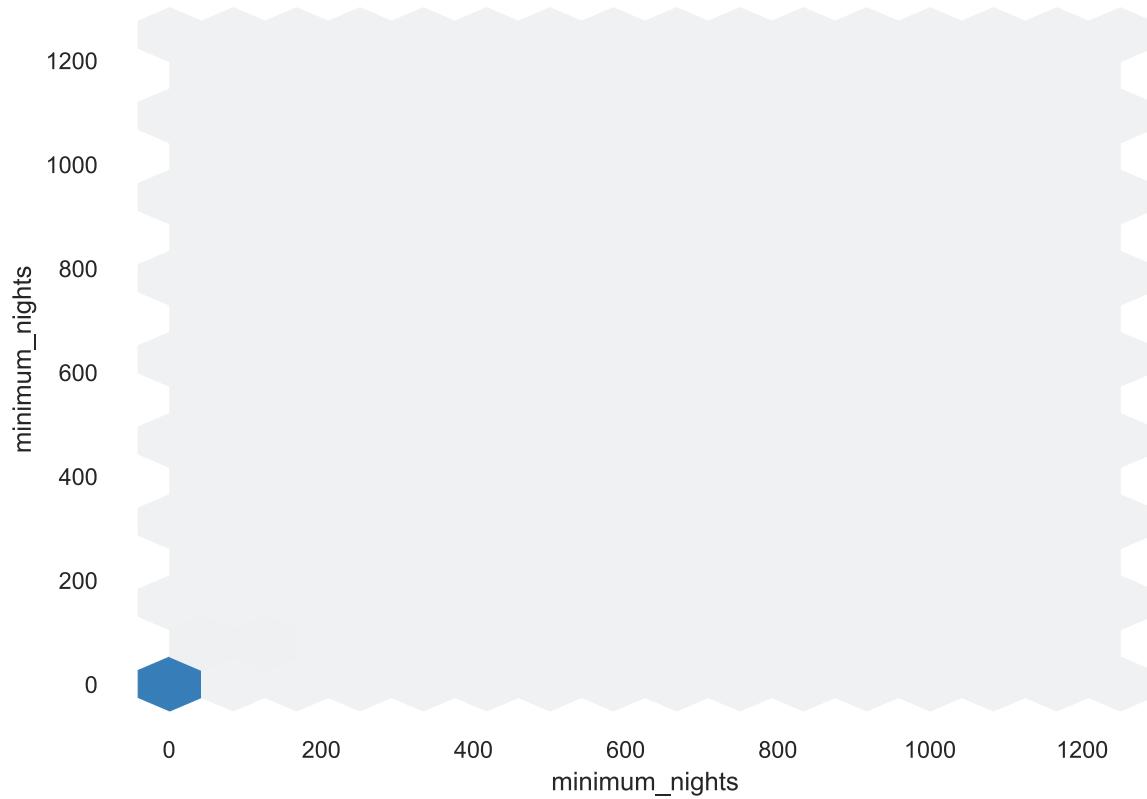


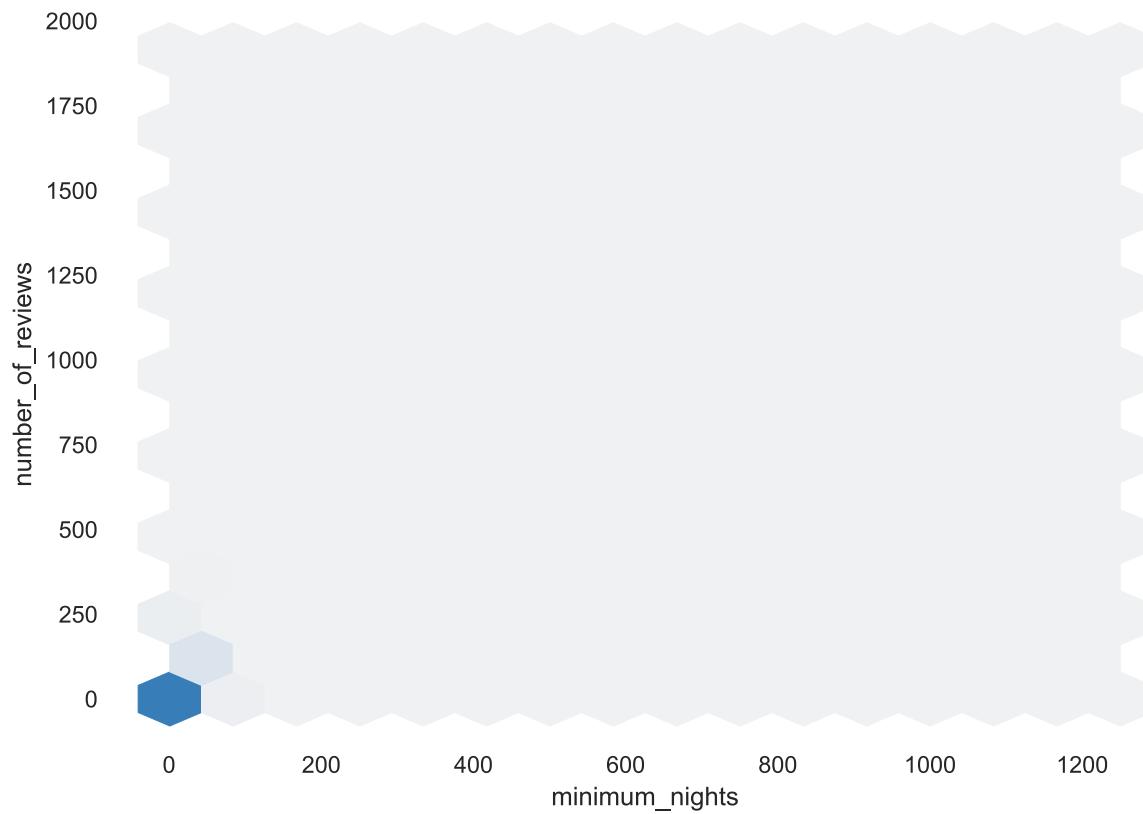


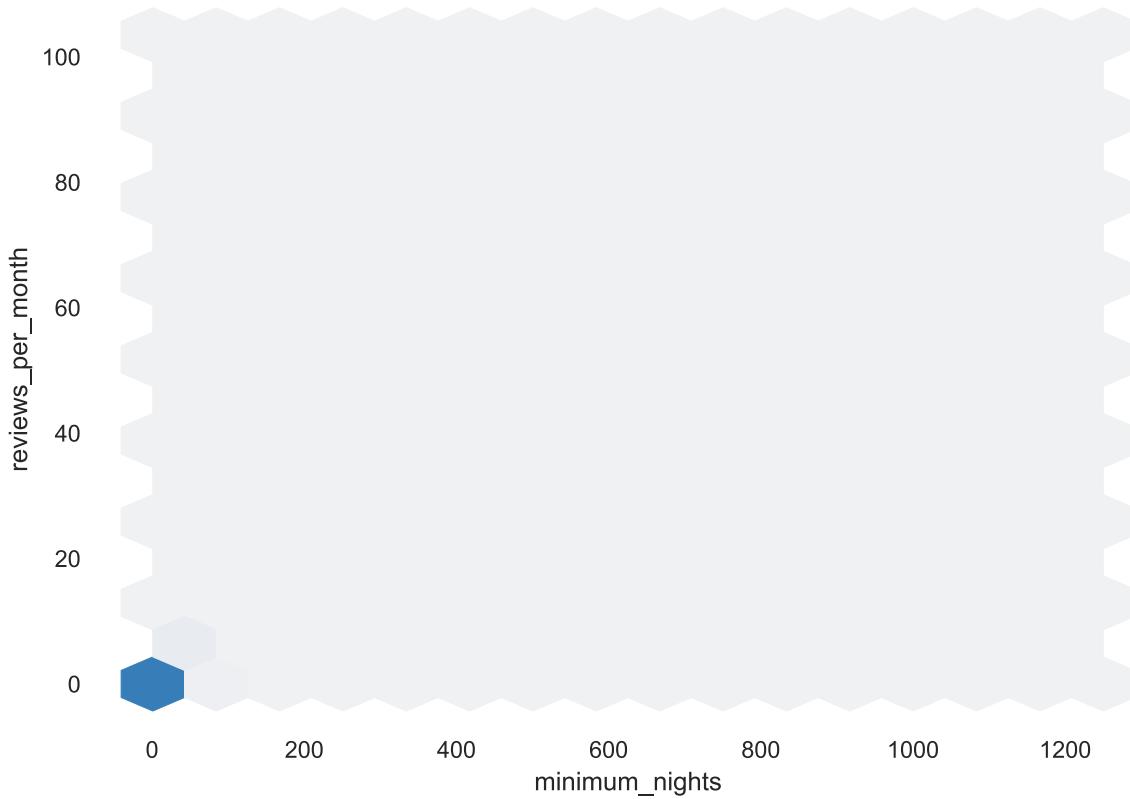


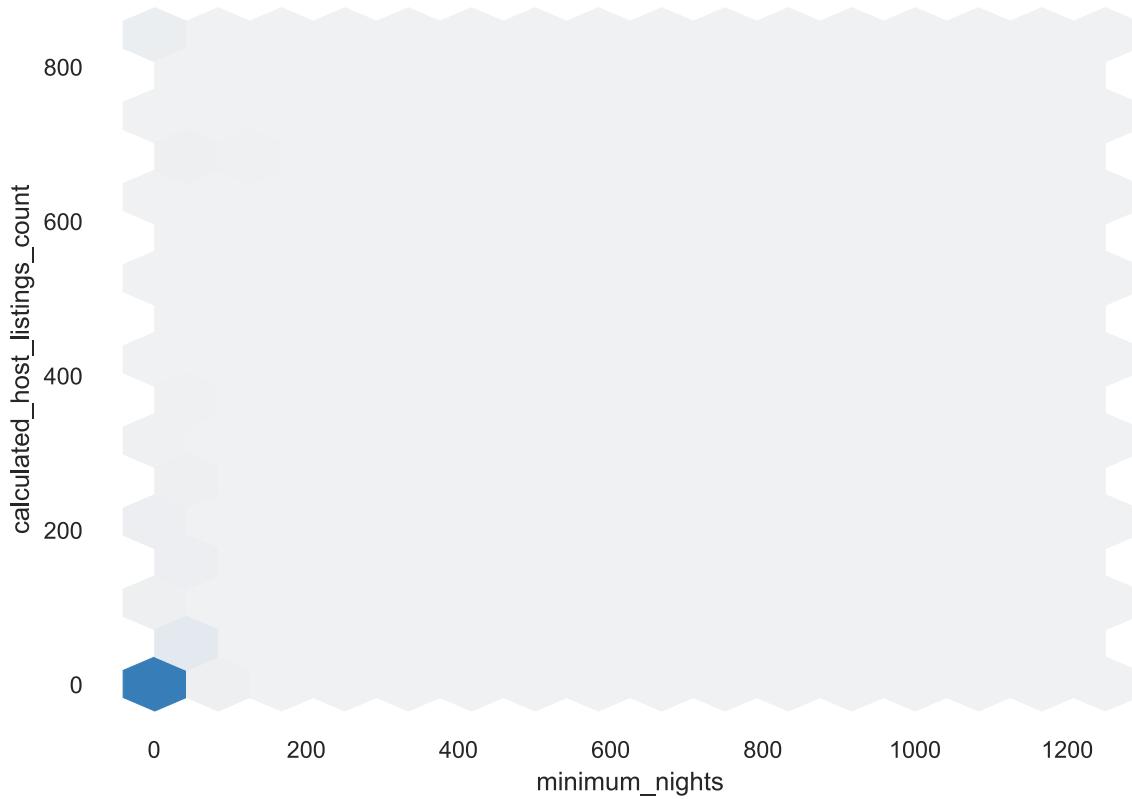


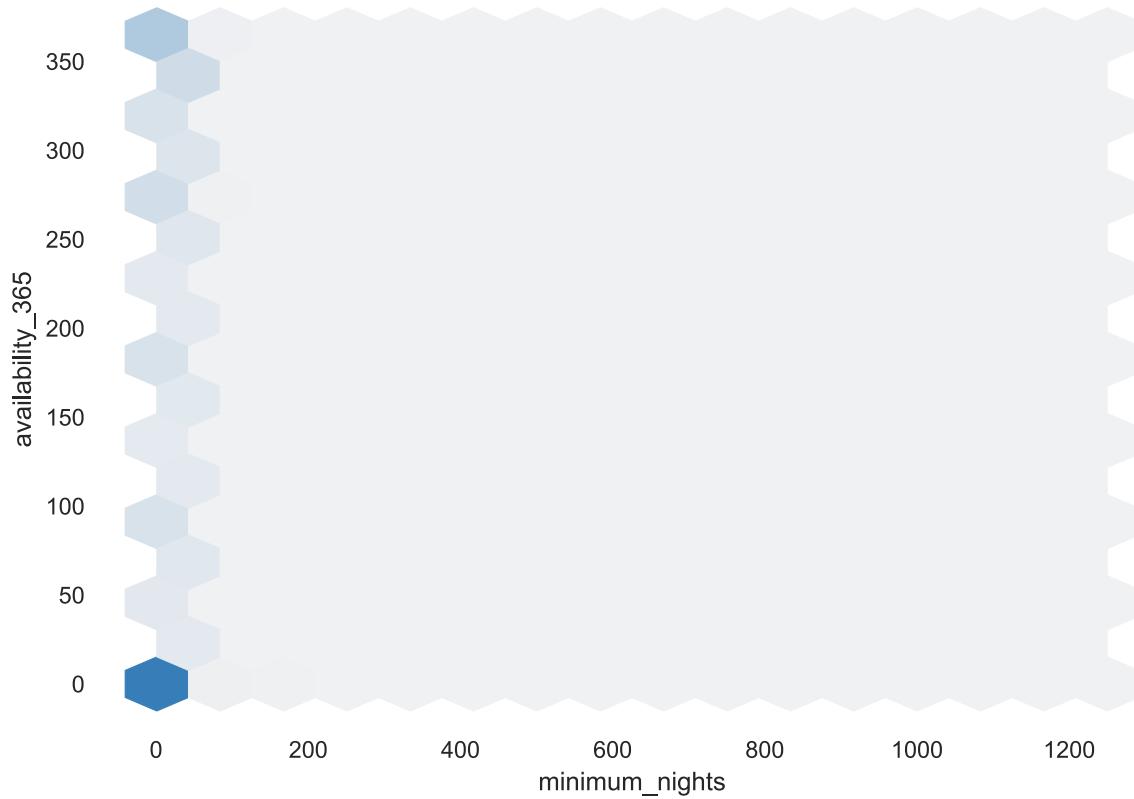




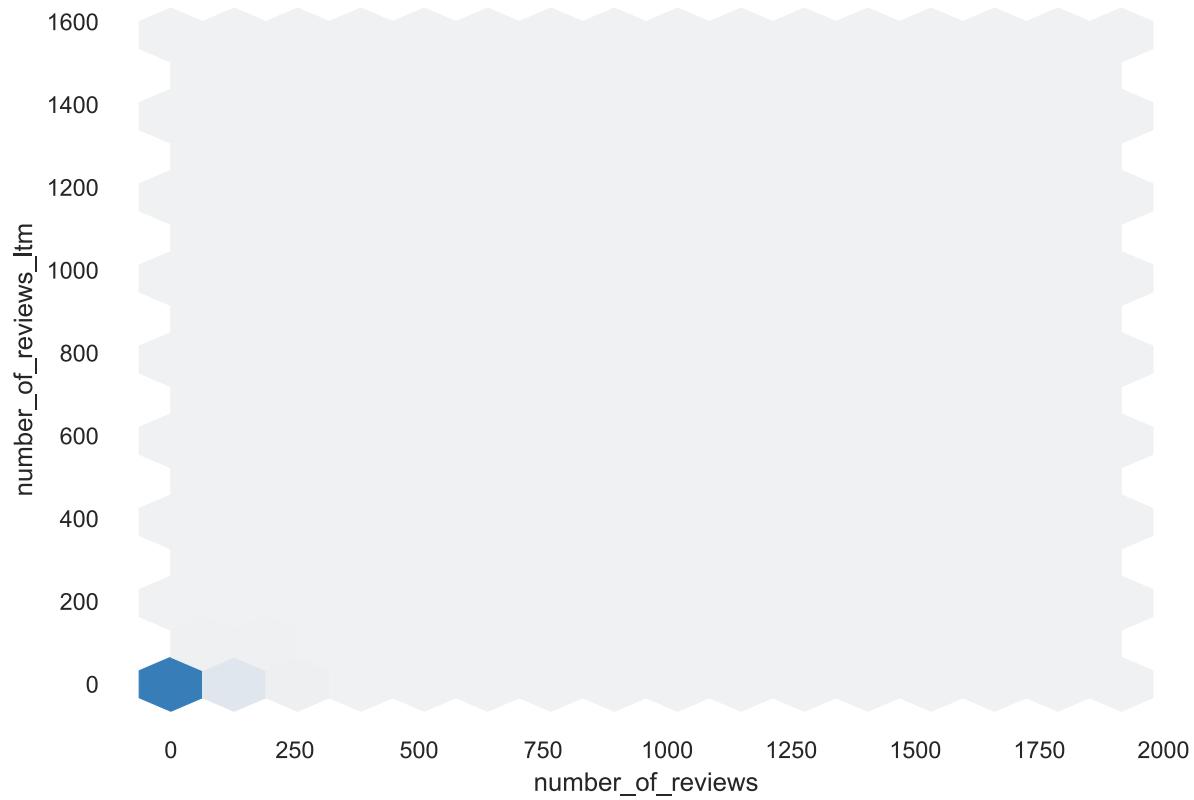


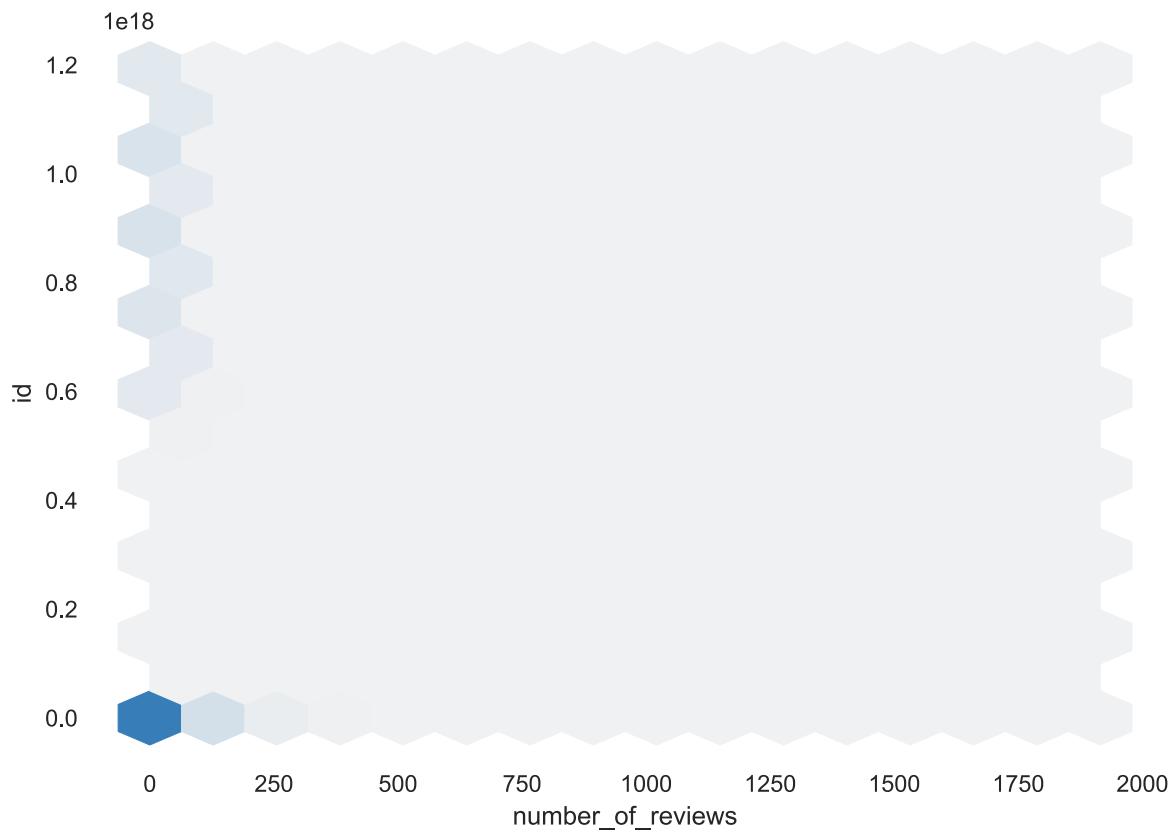


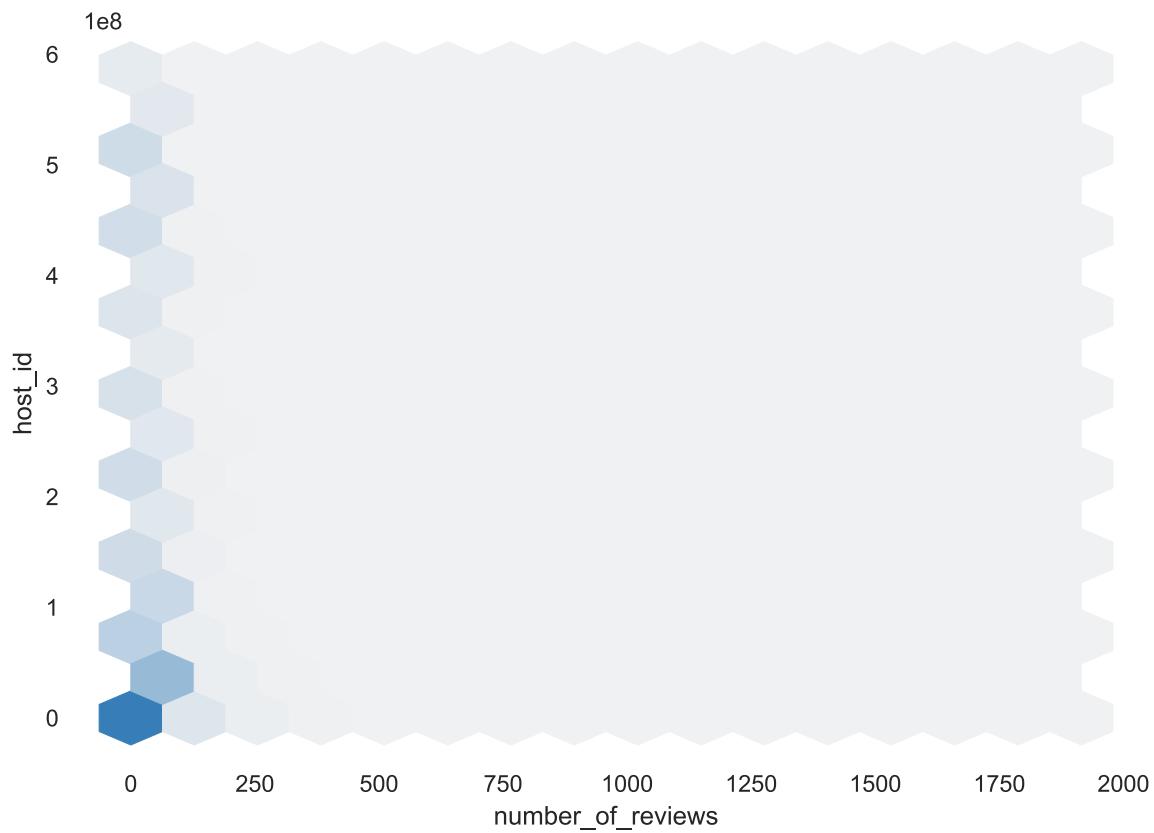


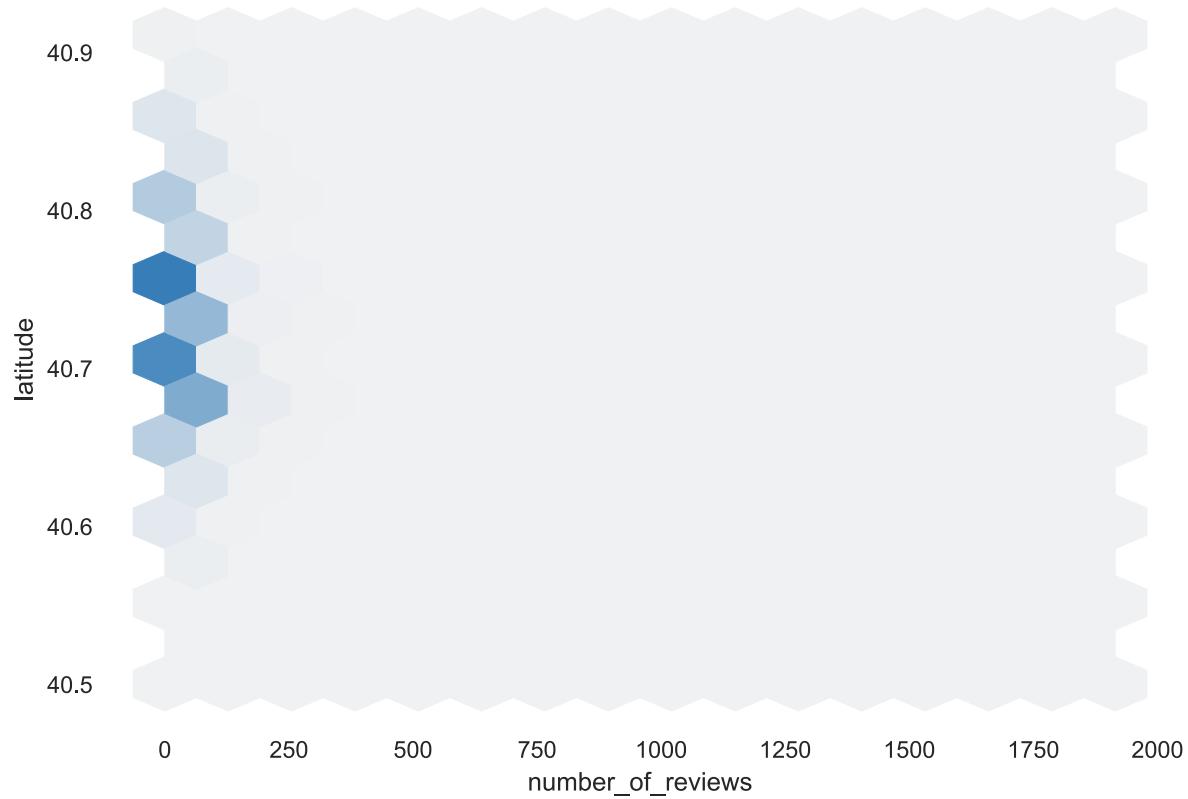


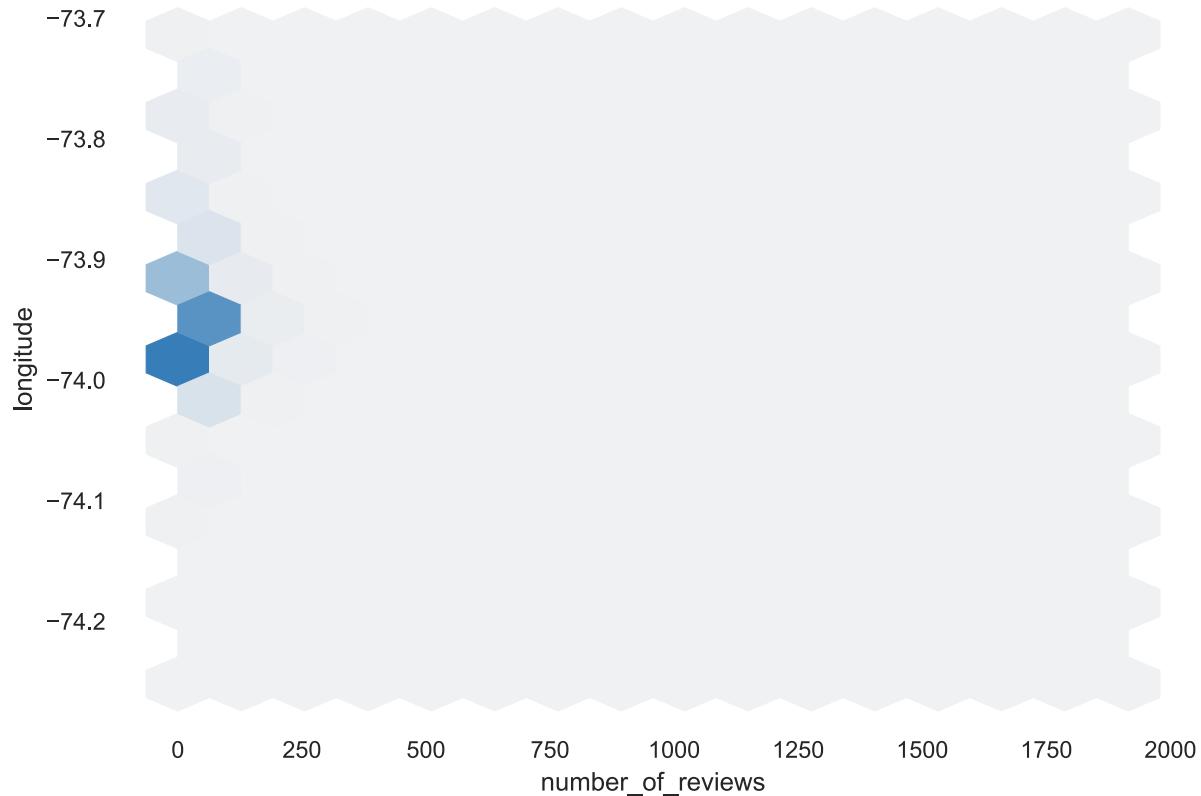

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

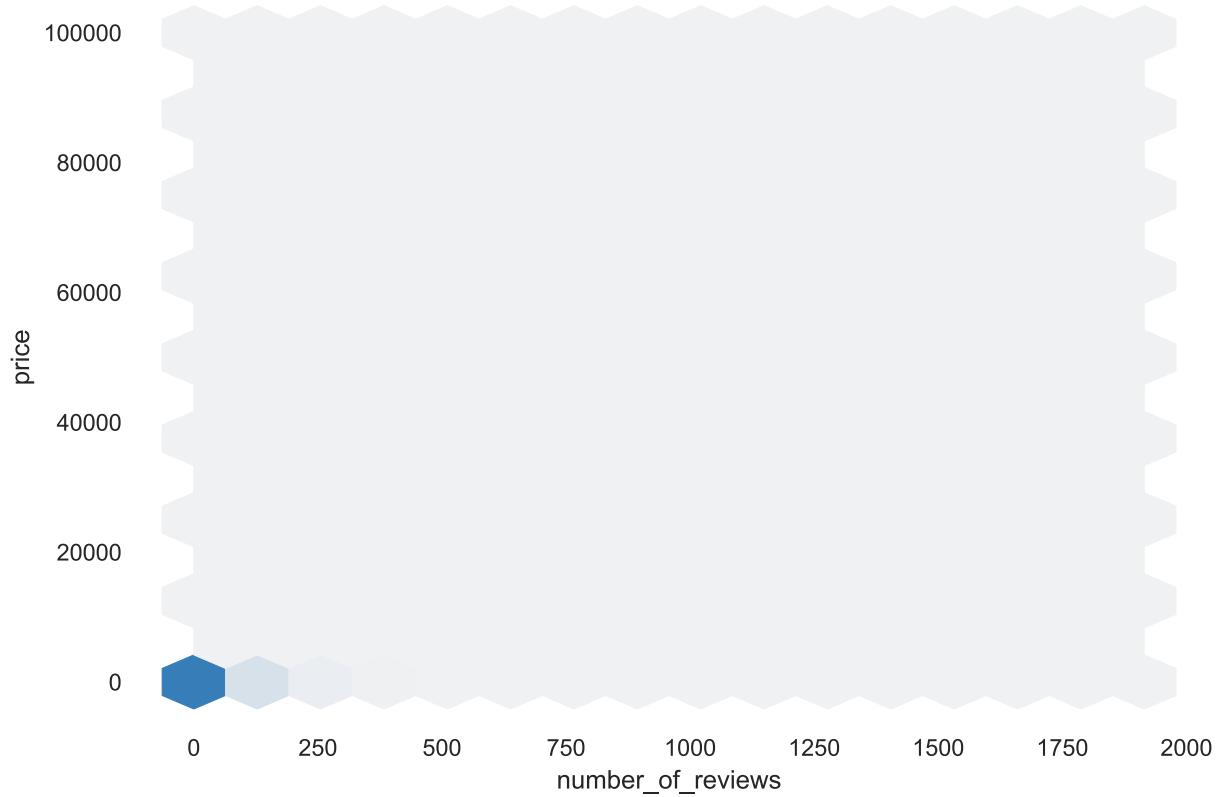


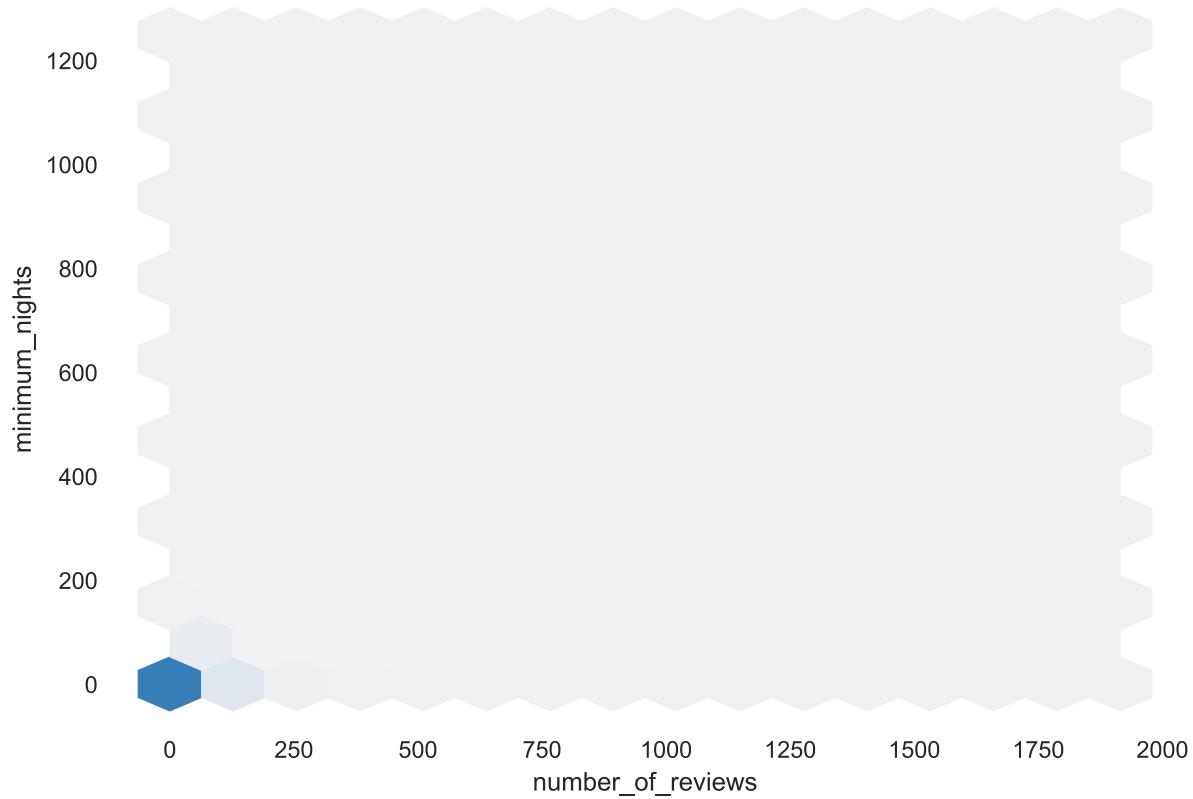


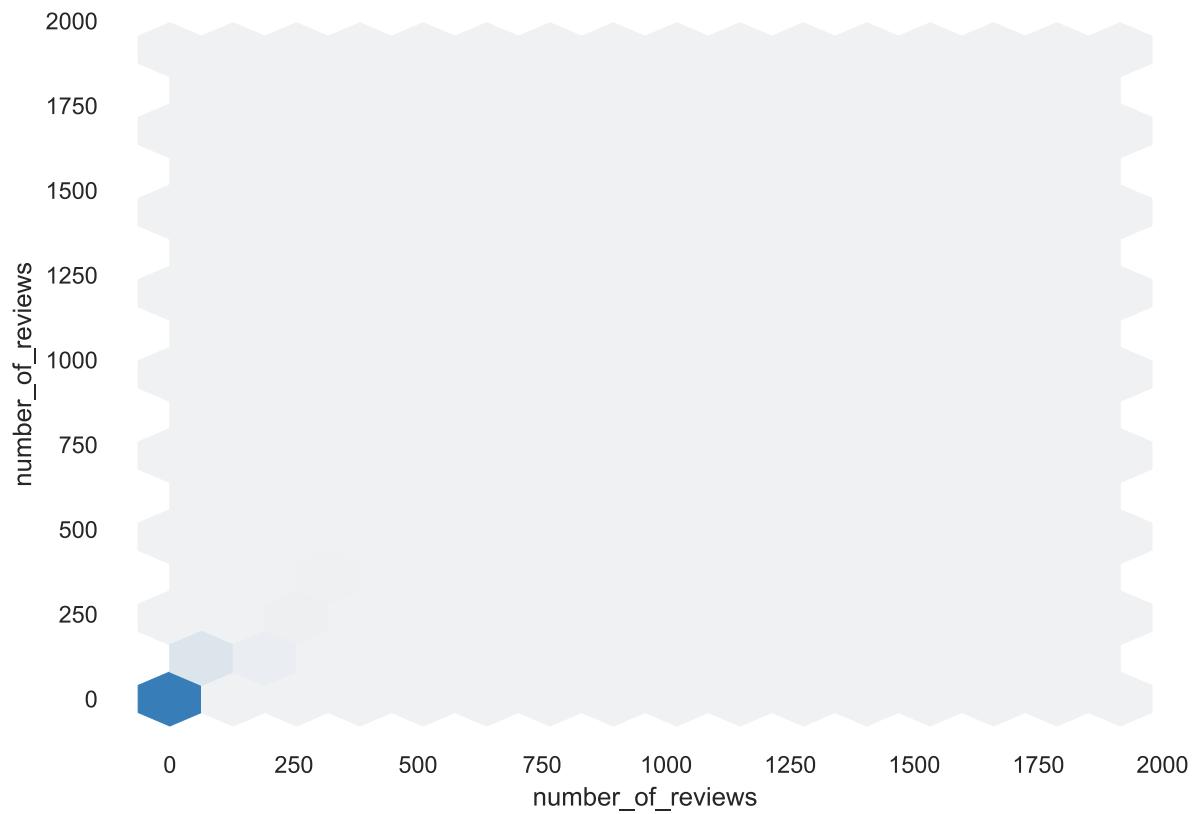


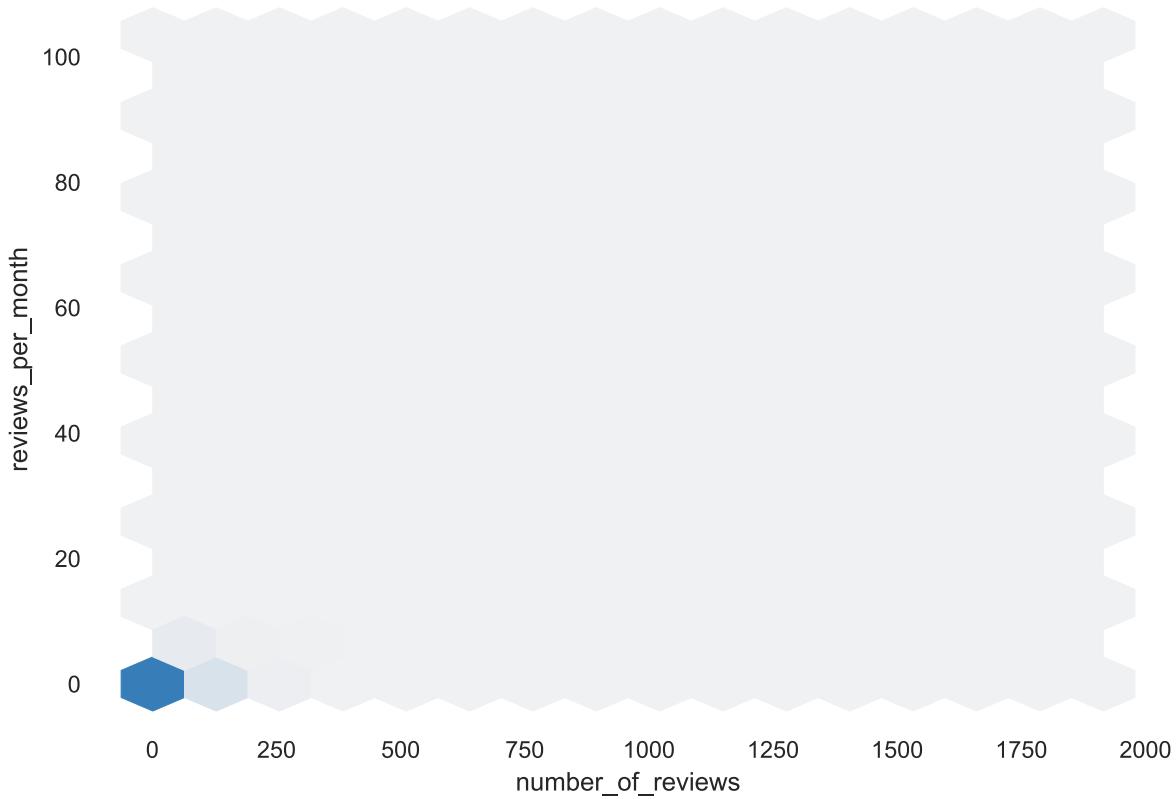


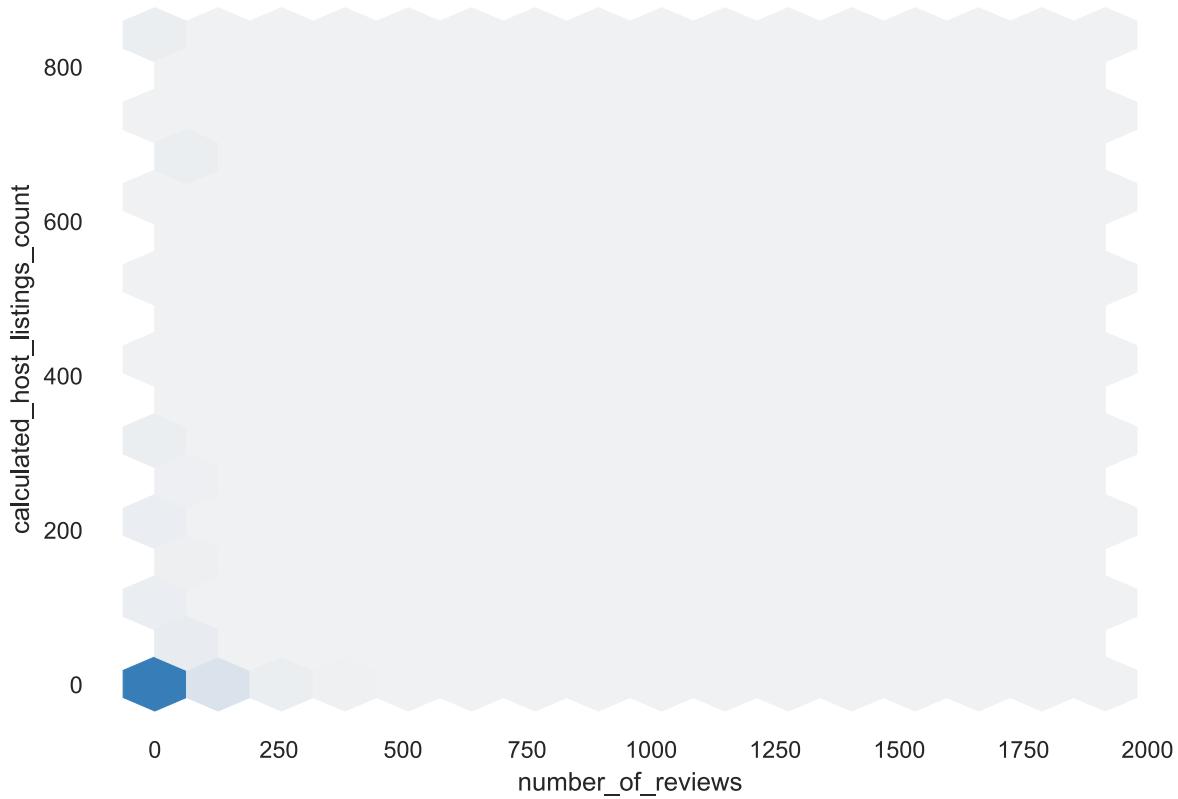


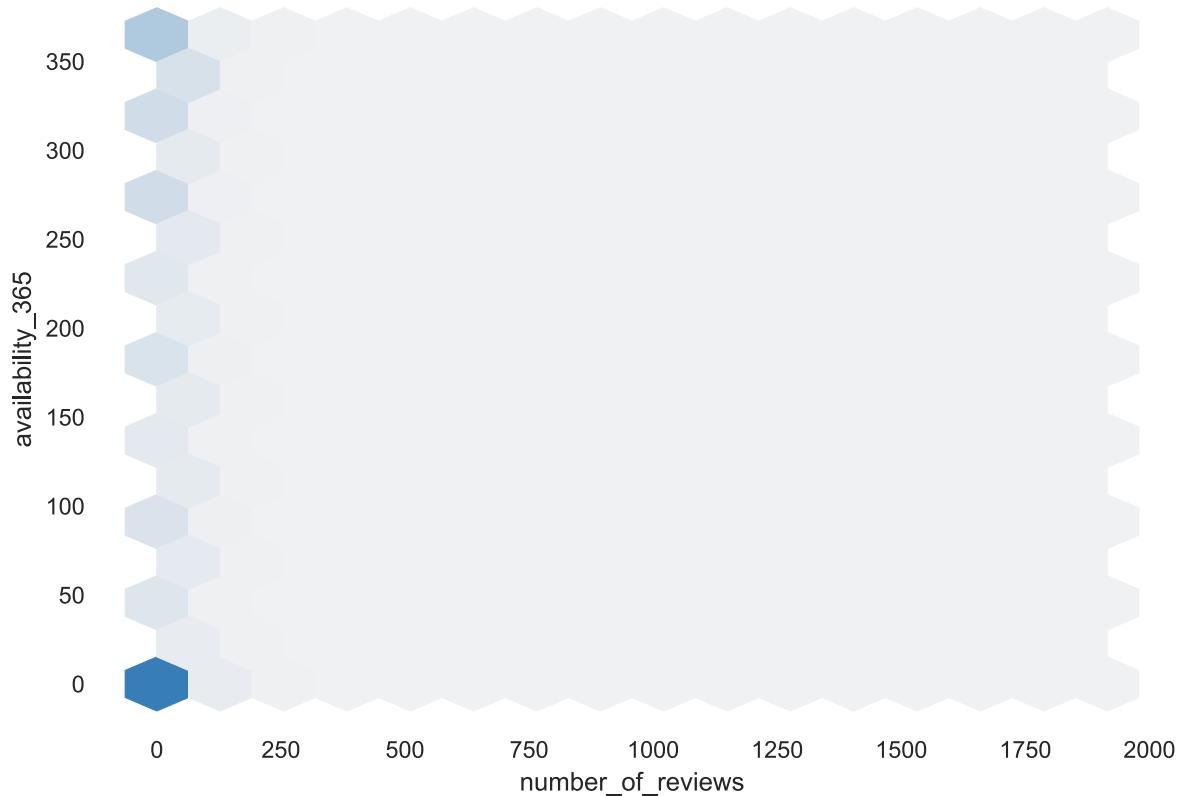




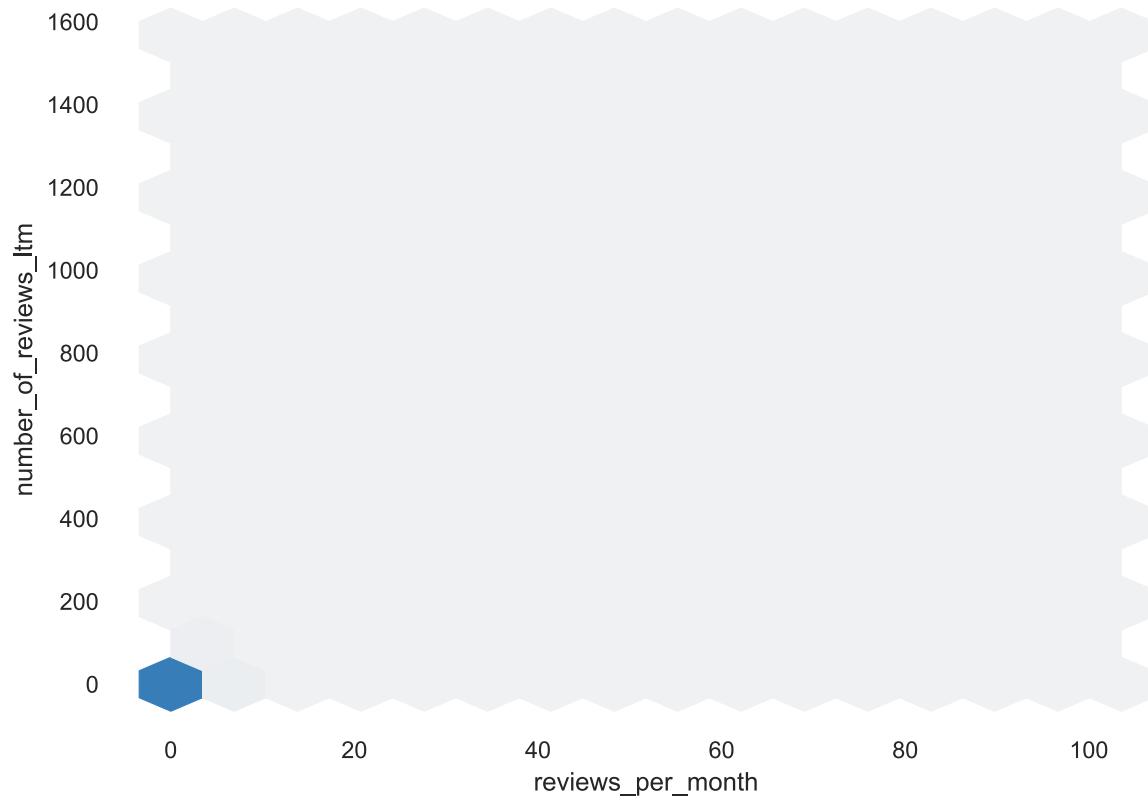


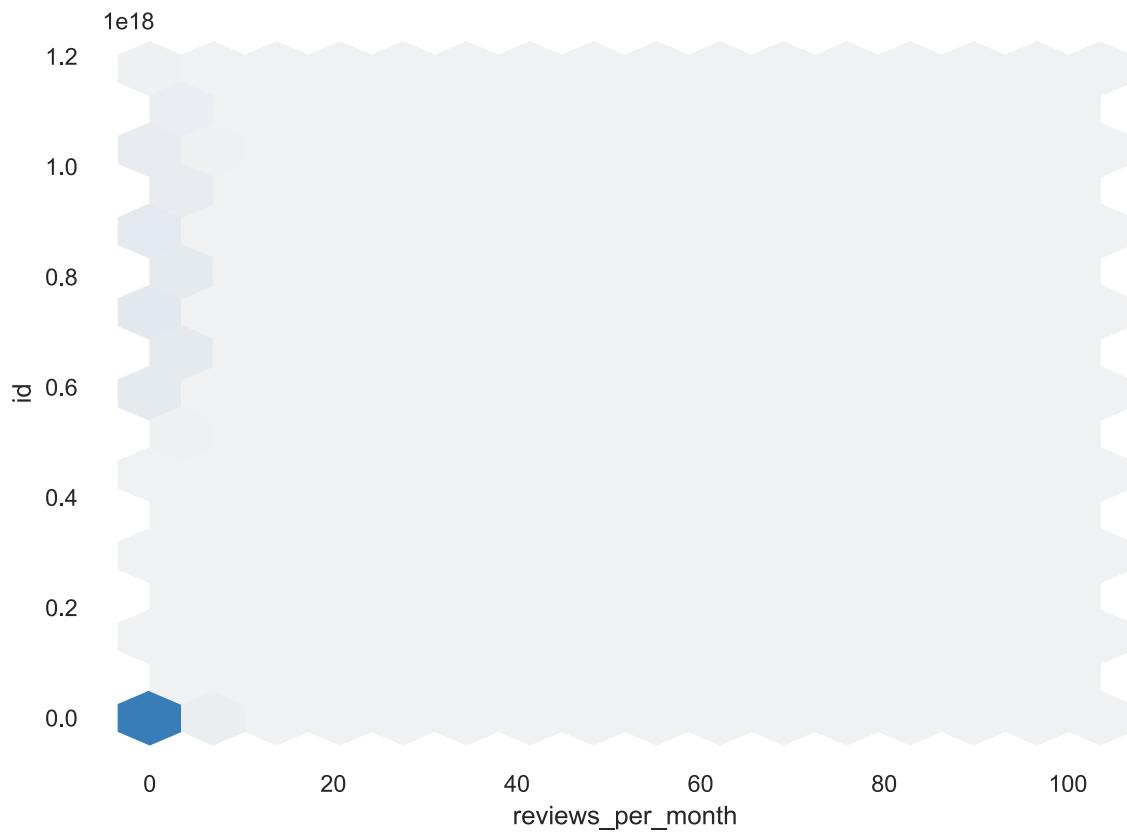


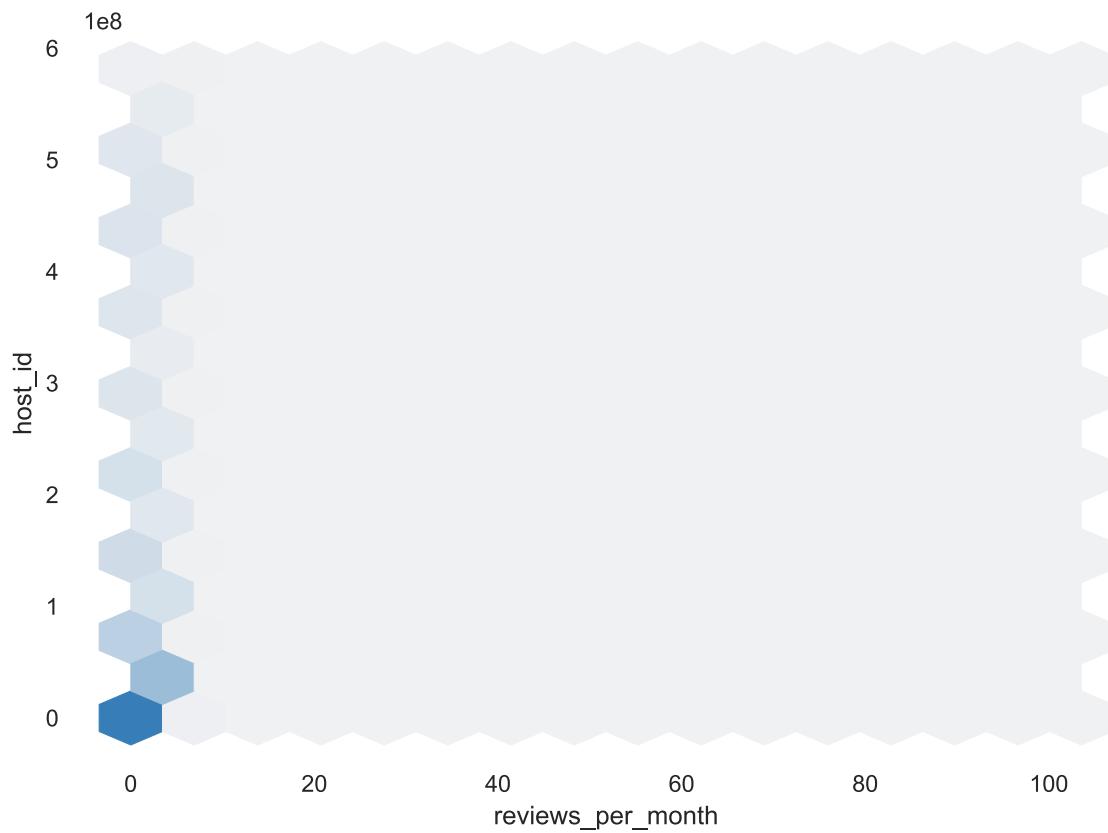


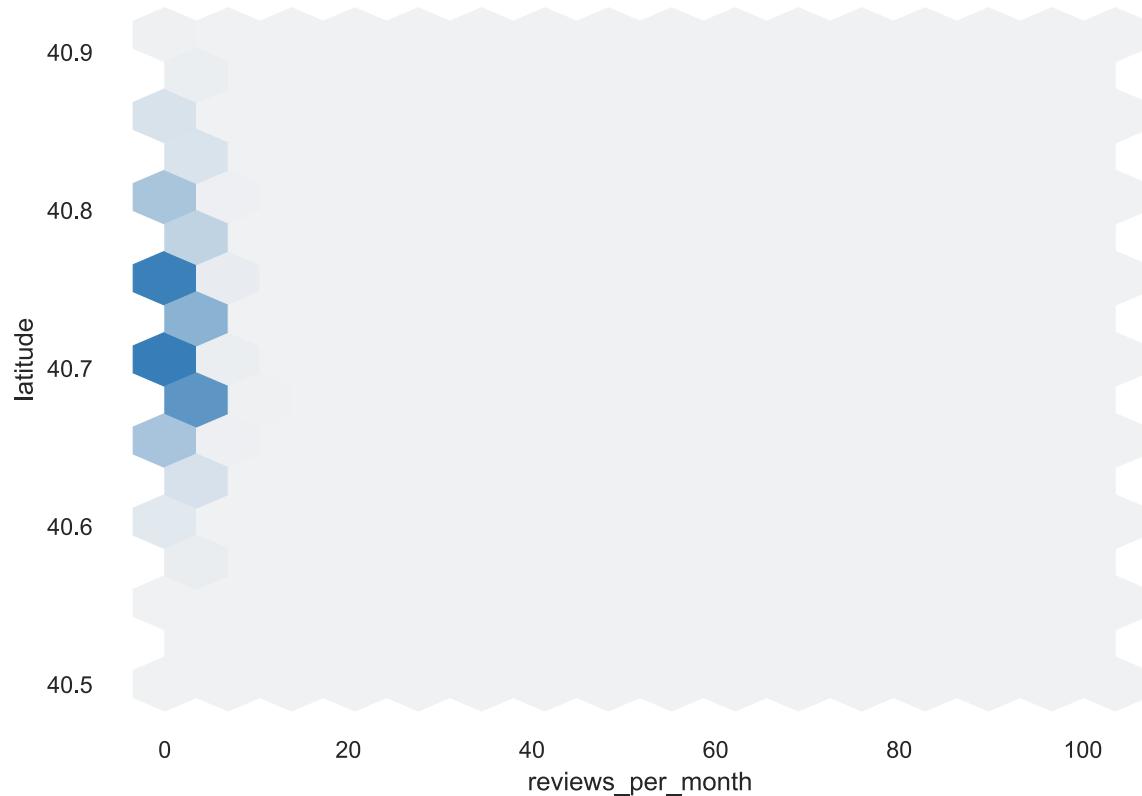


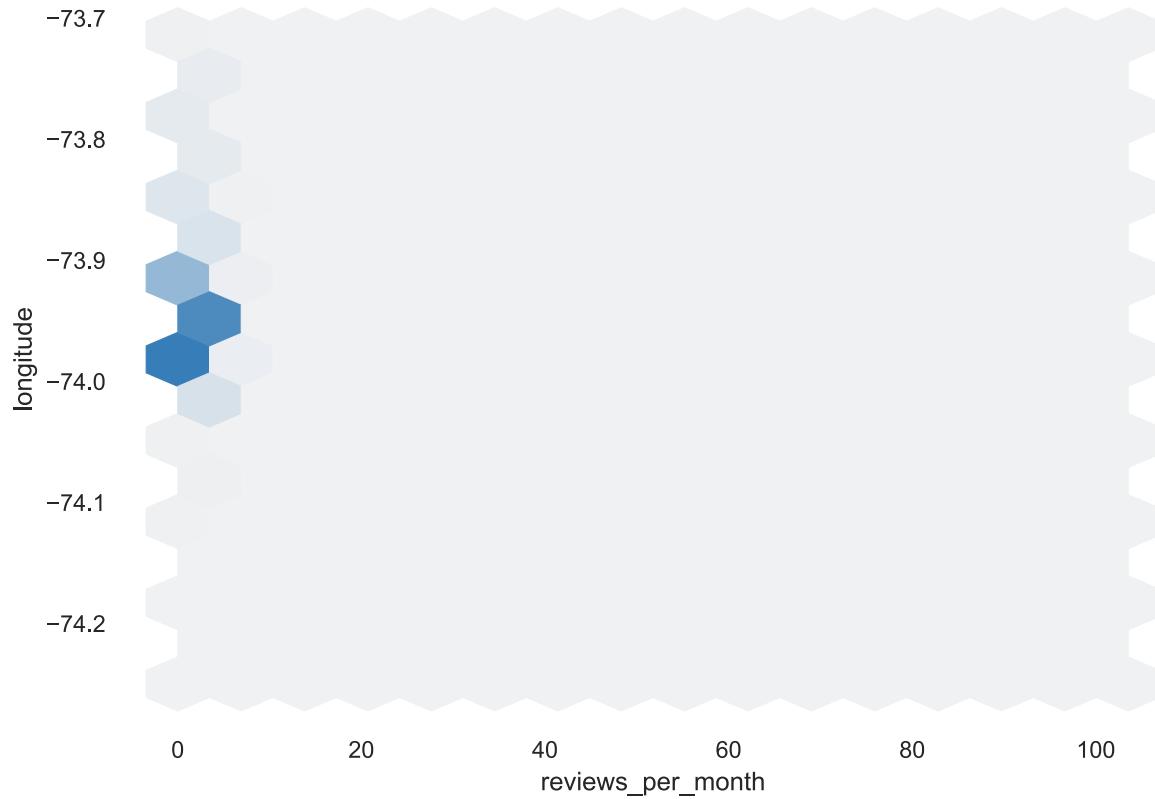

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

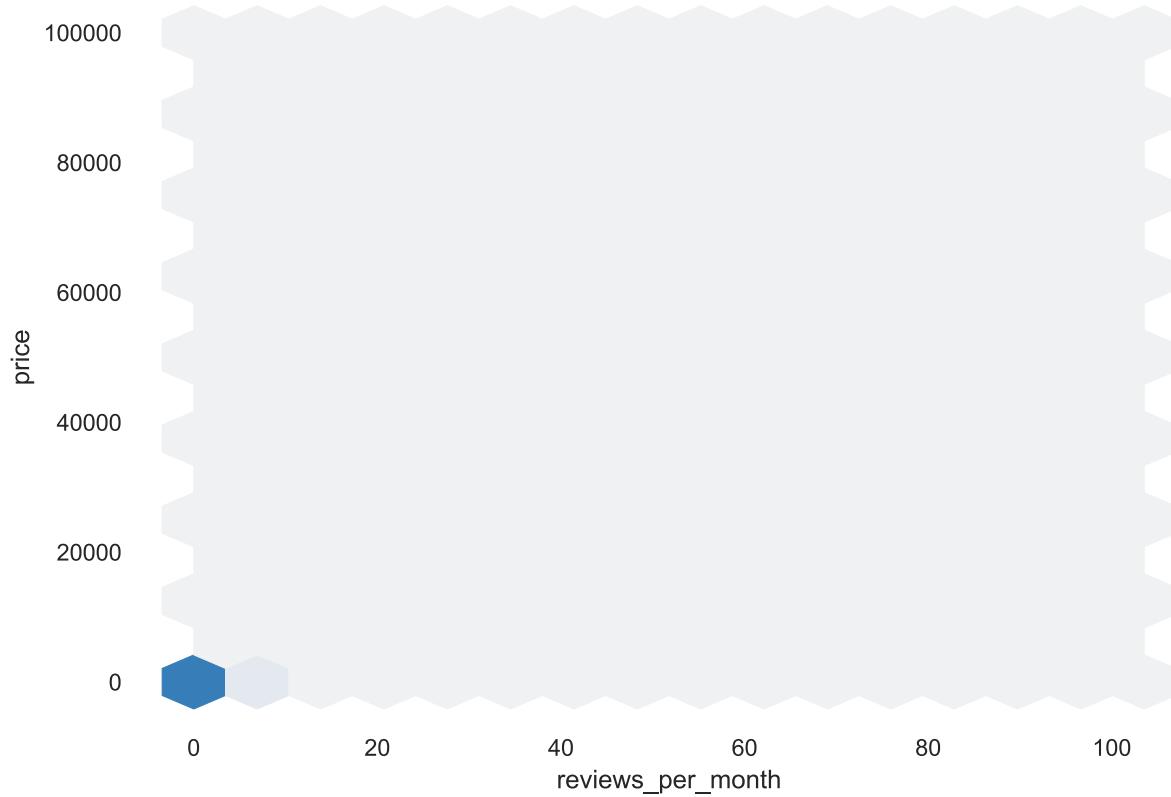


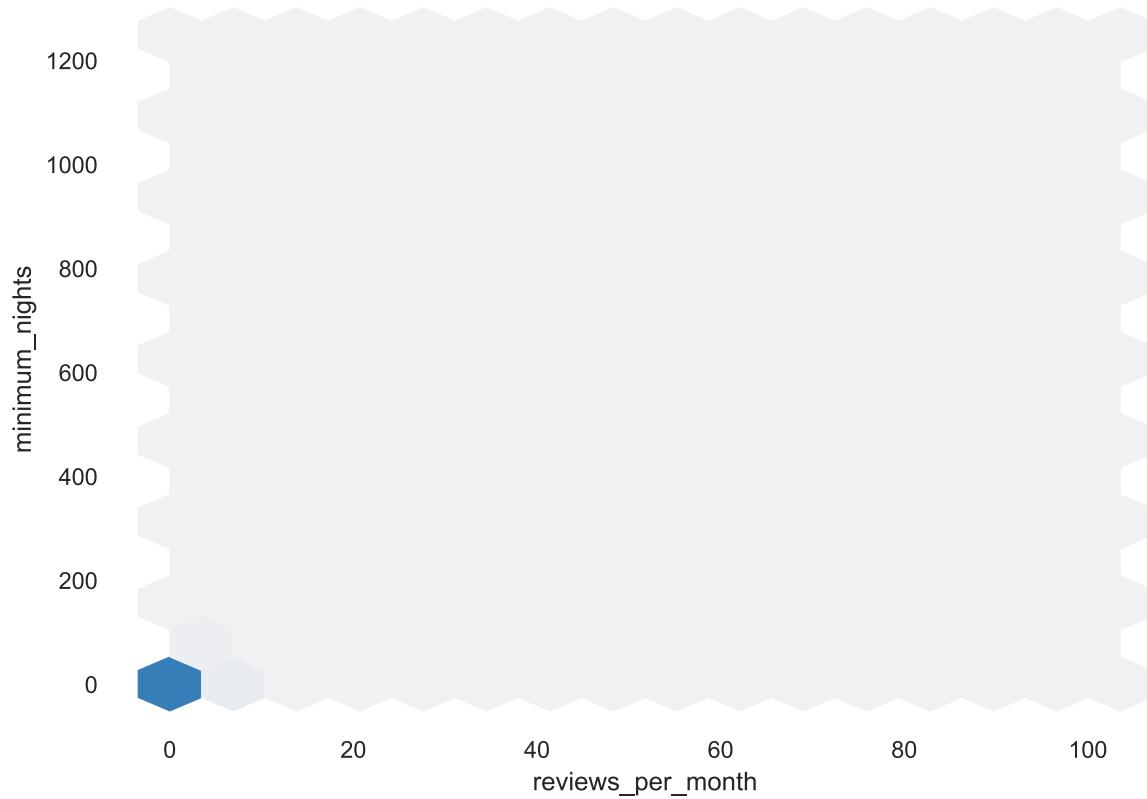


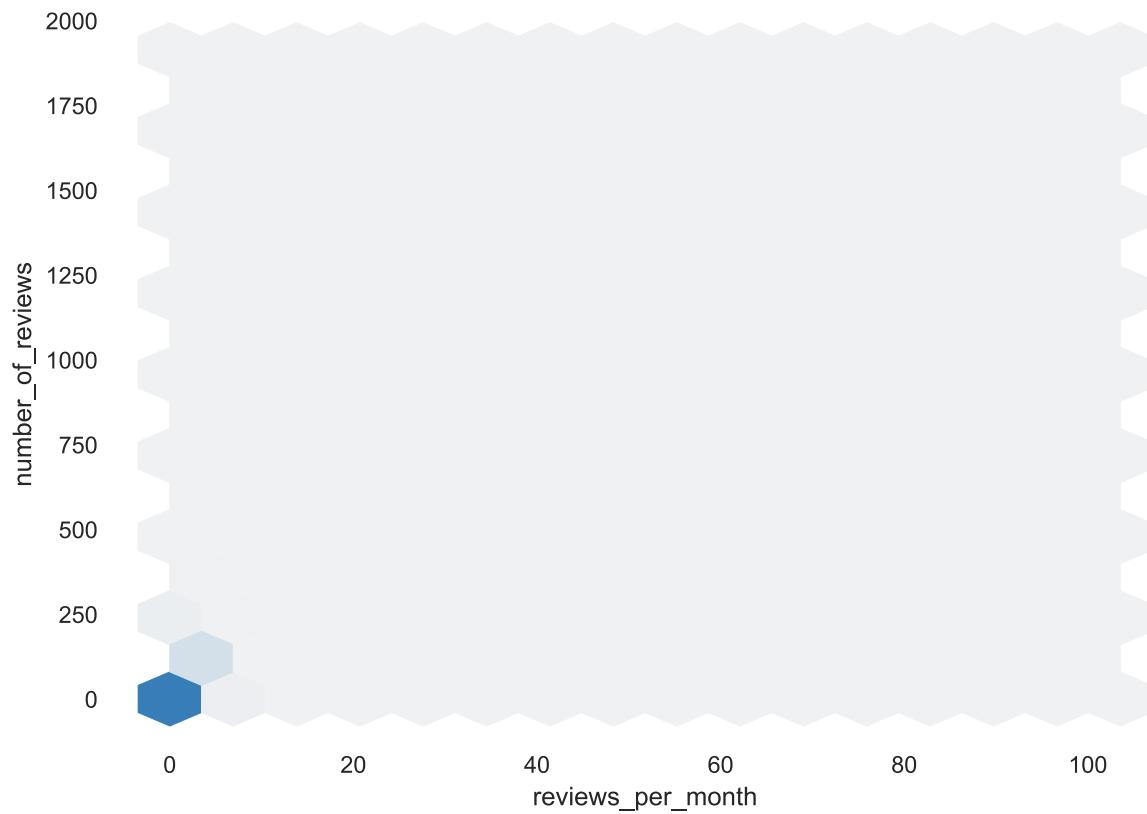


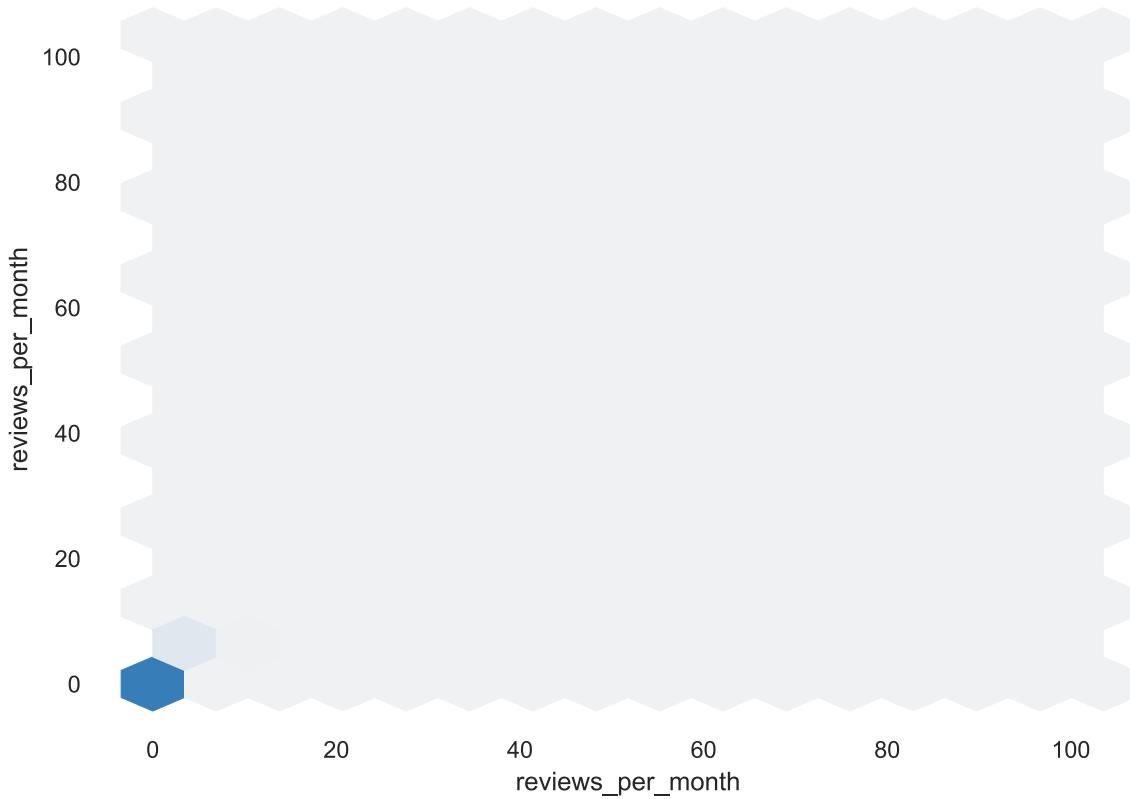


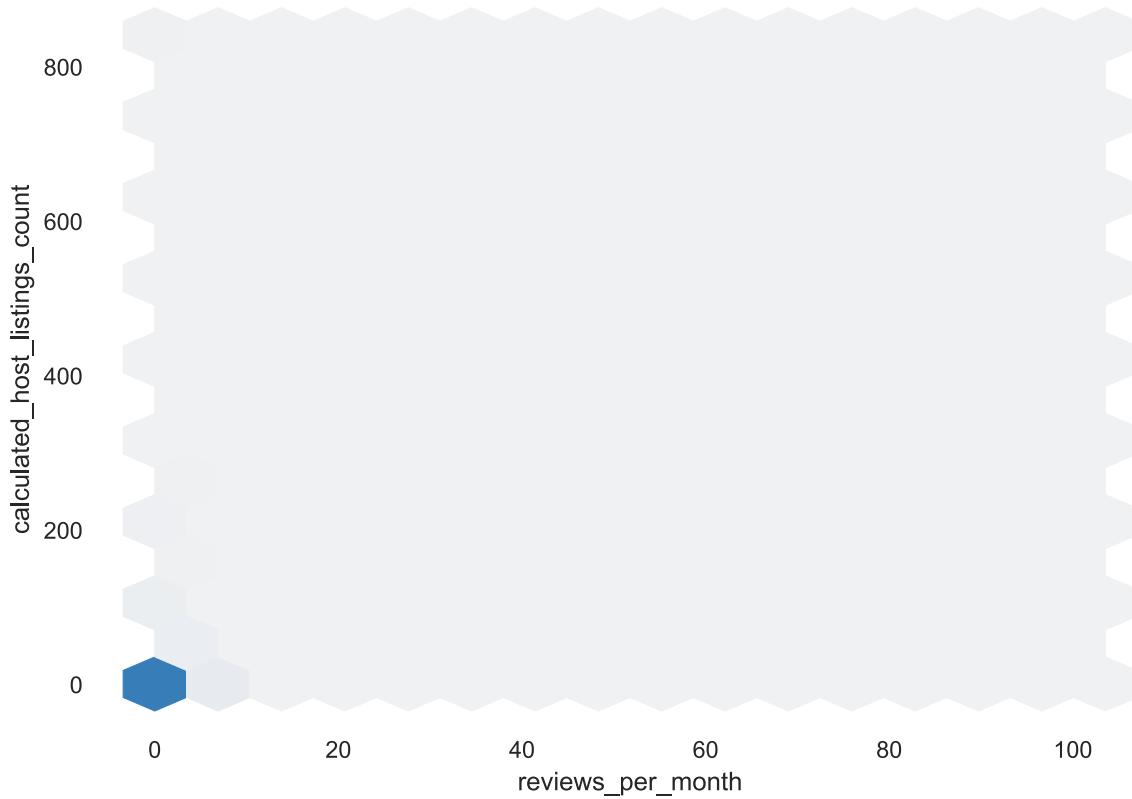


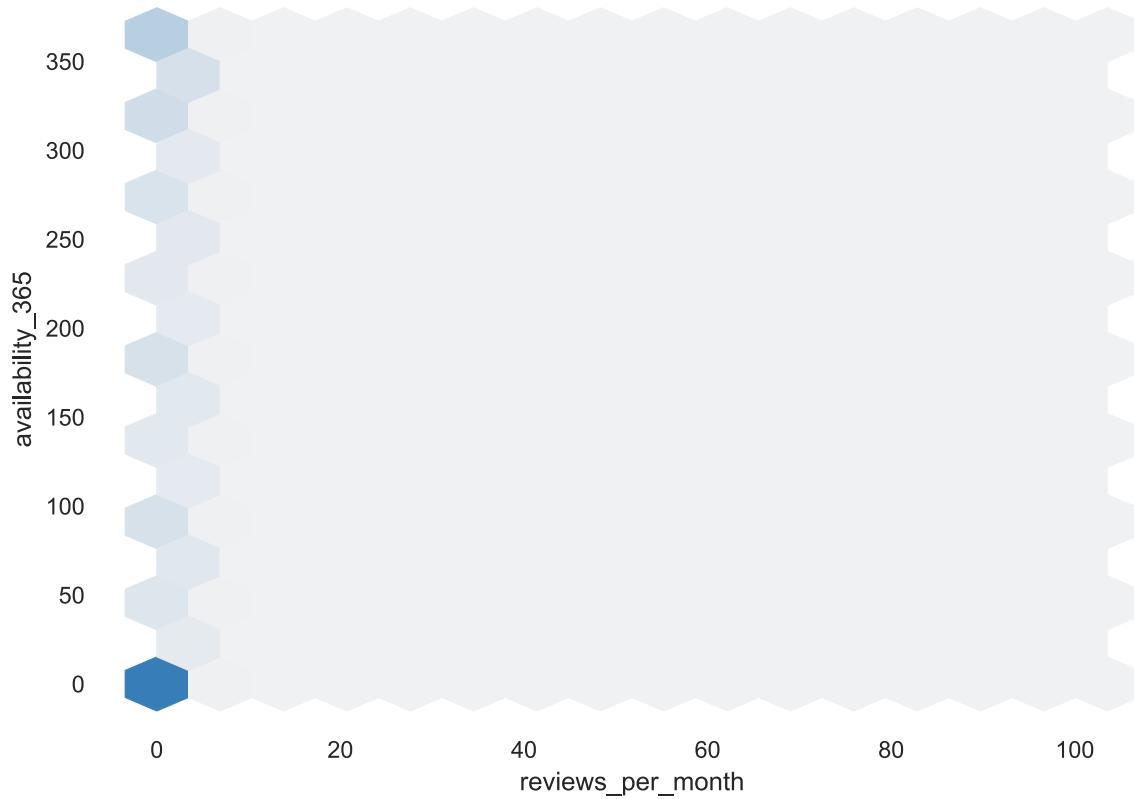




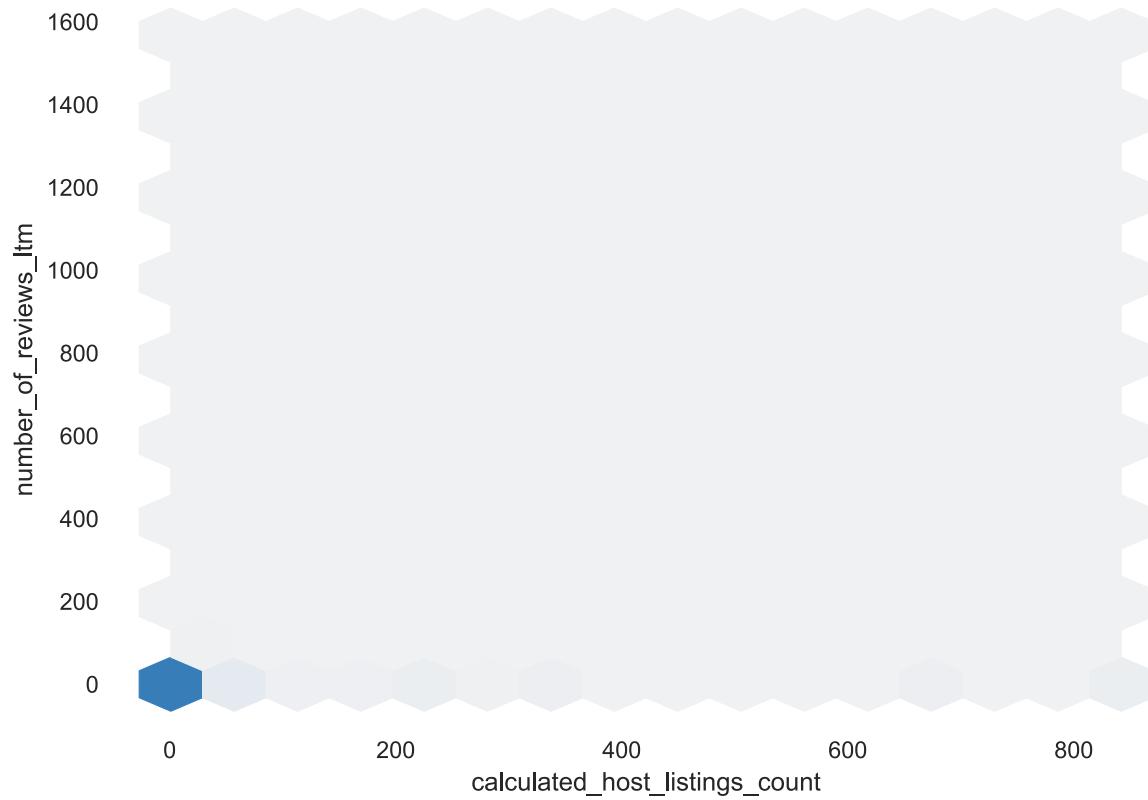


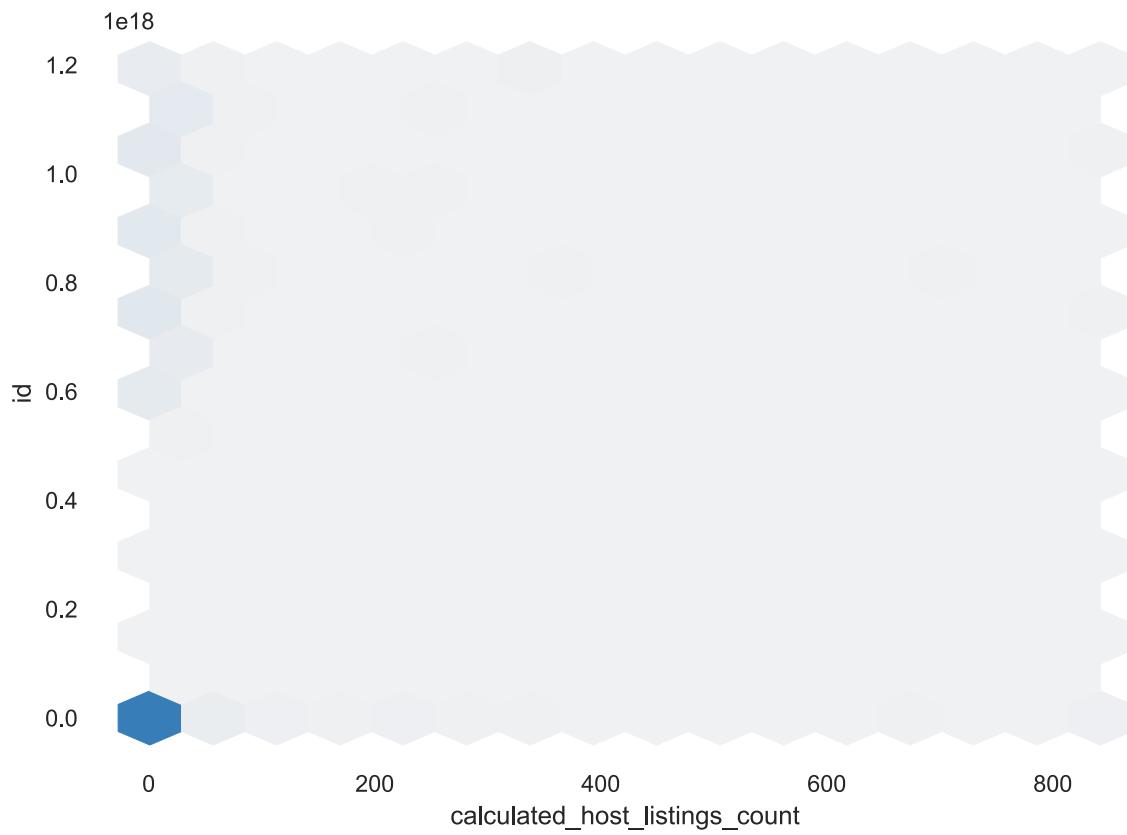


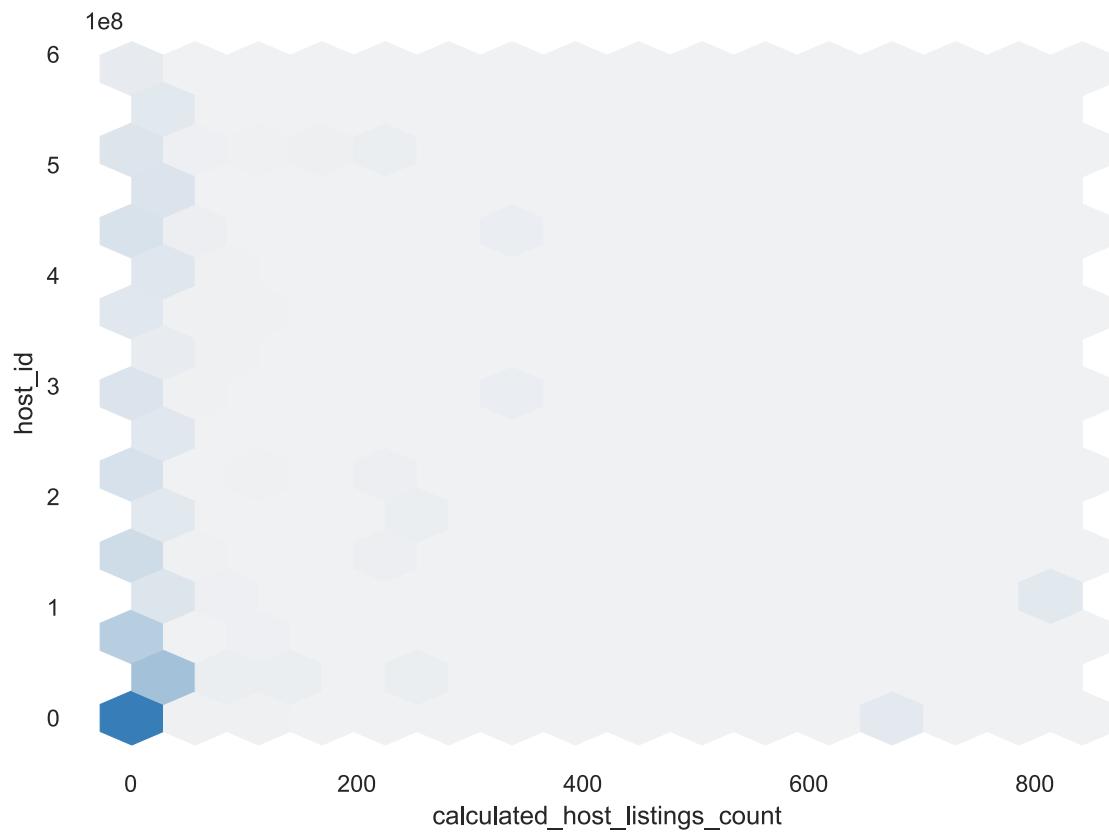


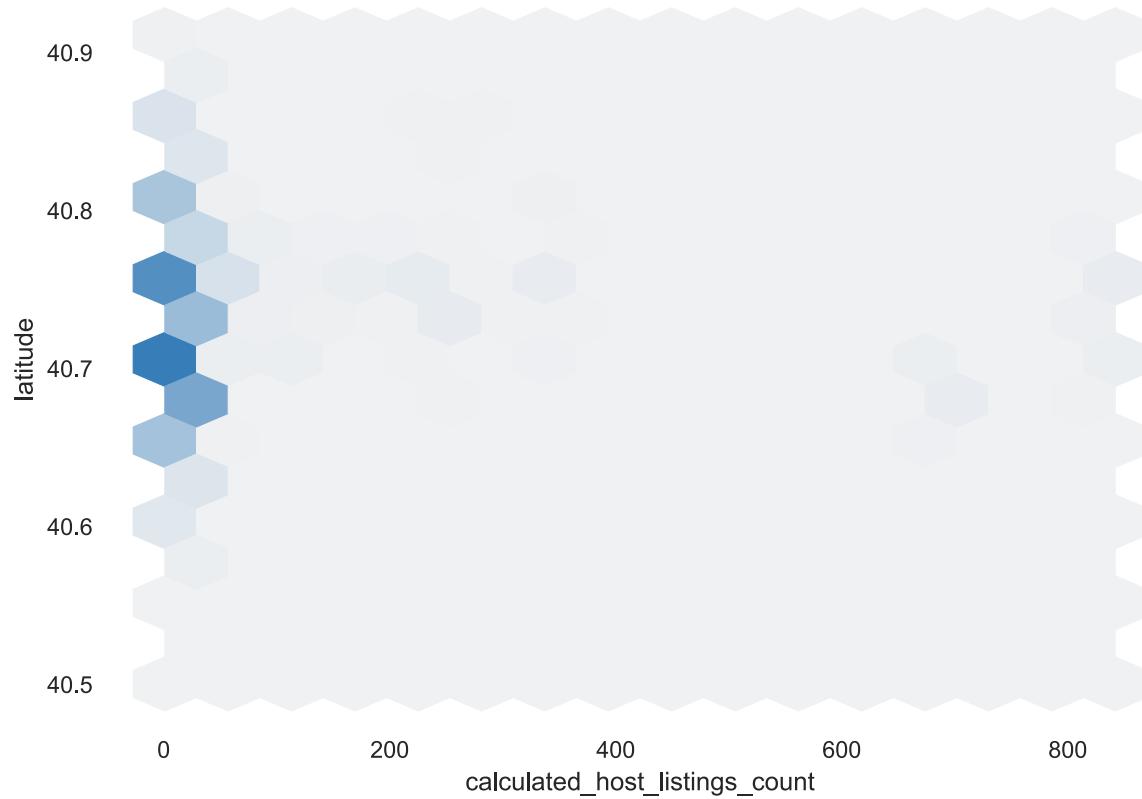


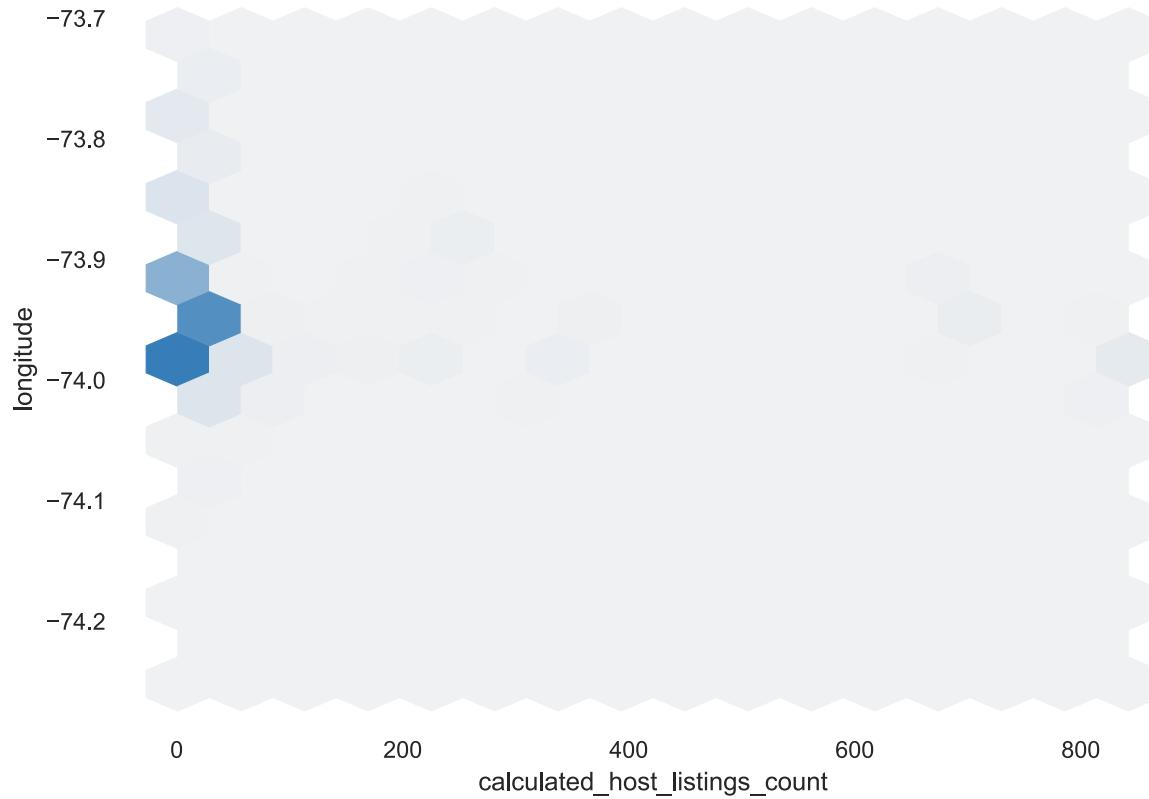

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_count  
availability_365
```

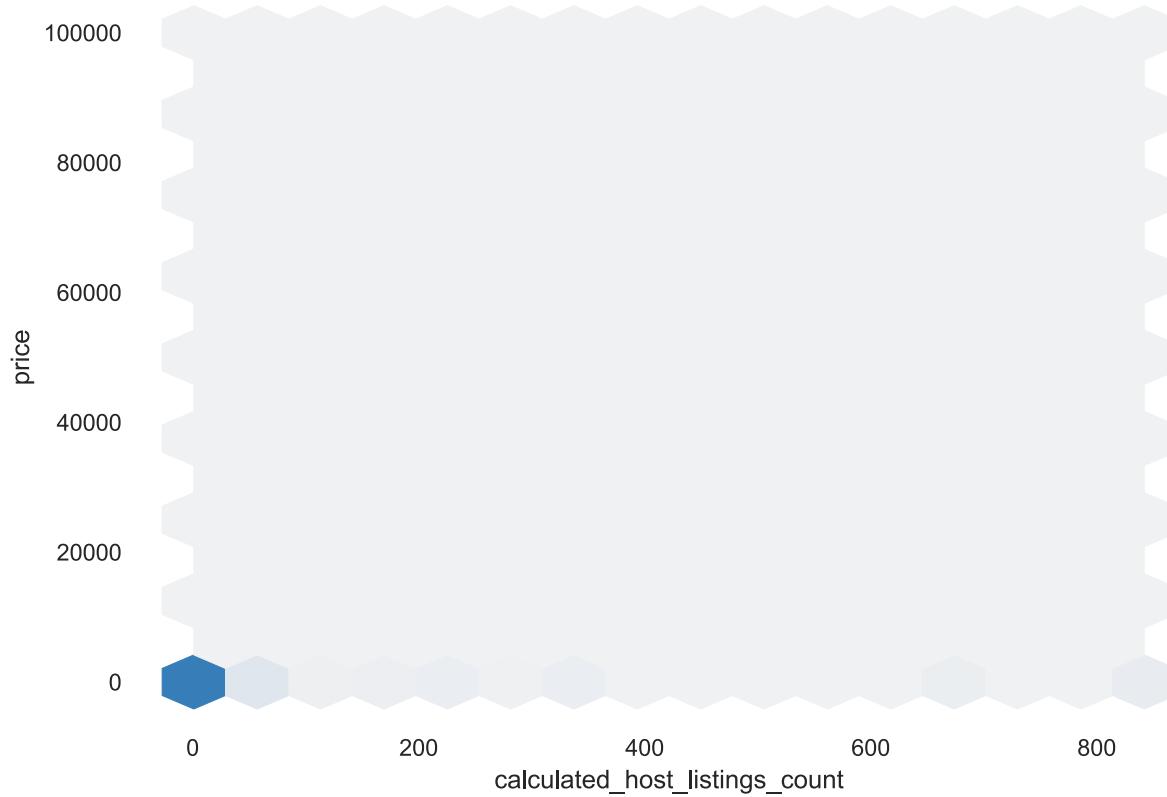




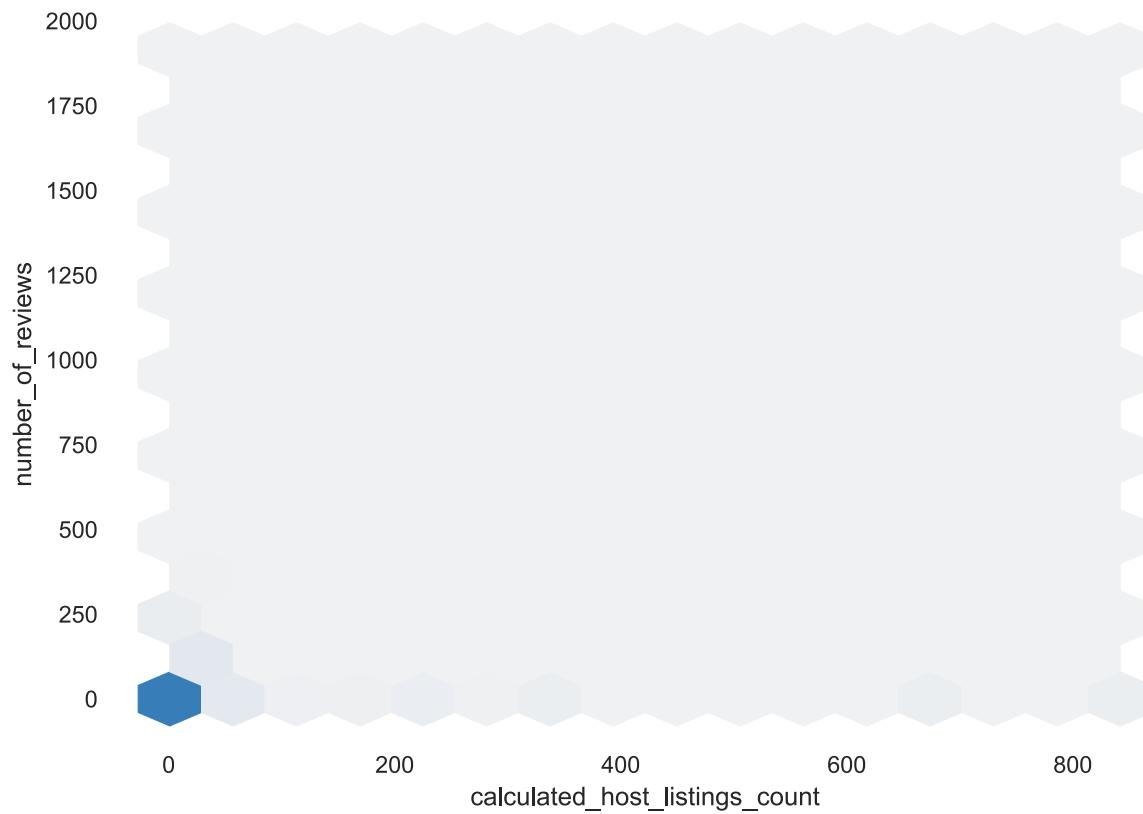


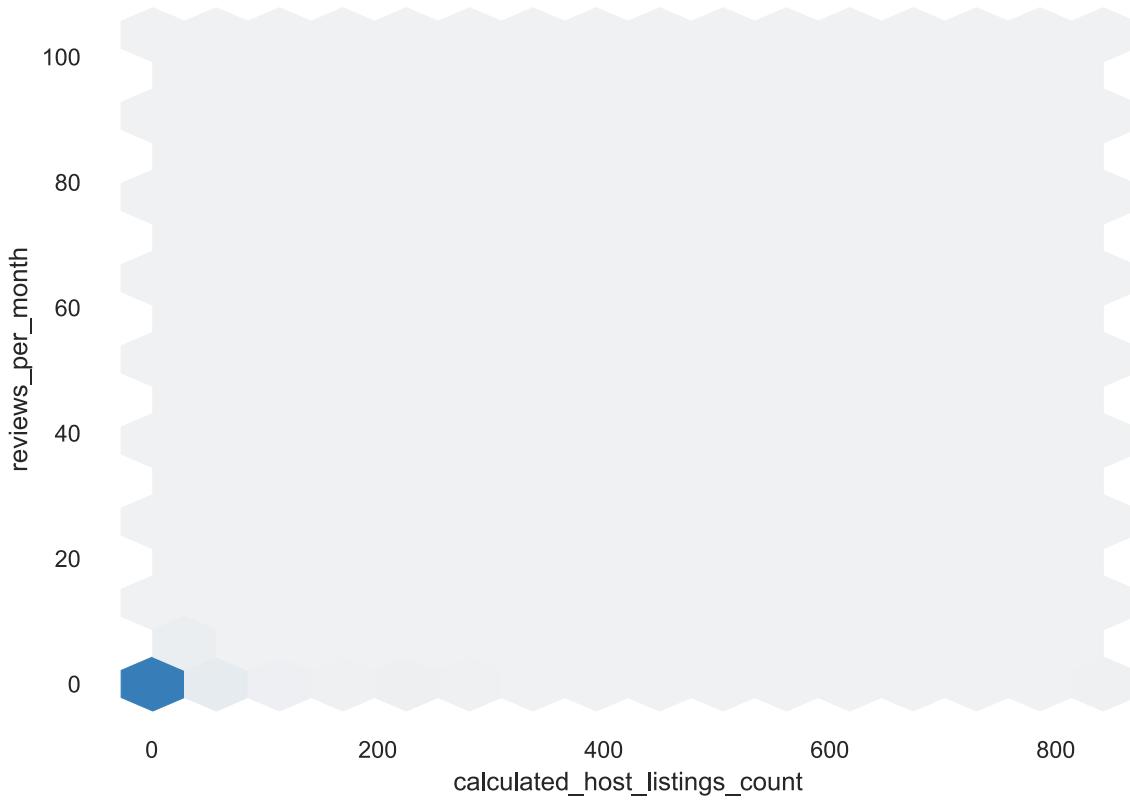








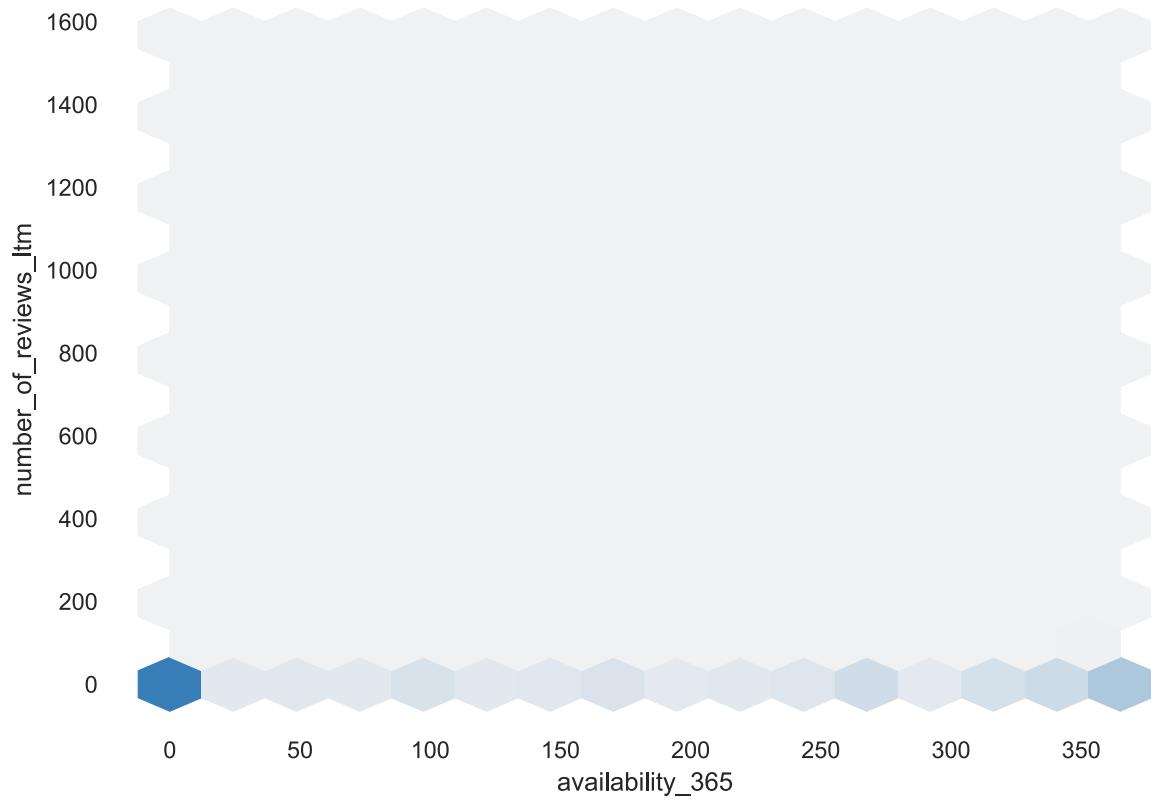


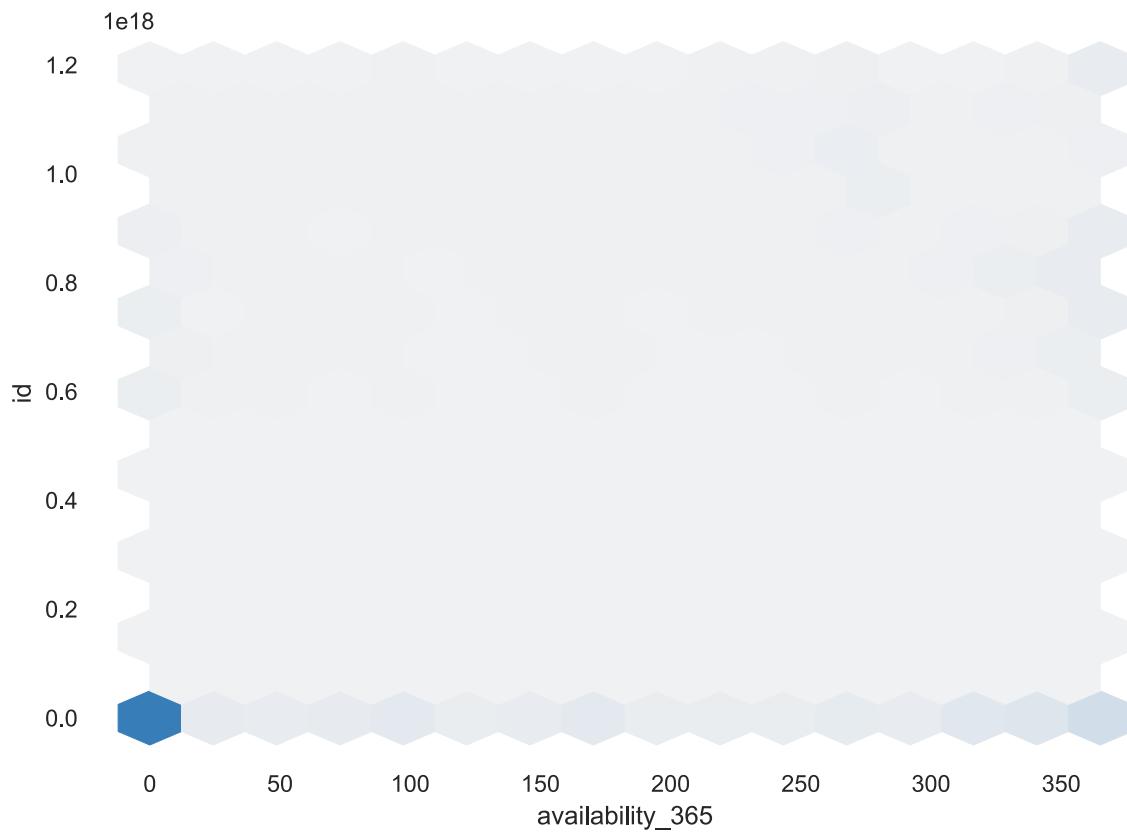


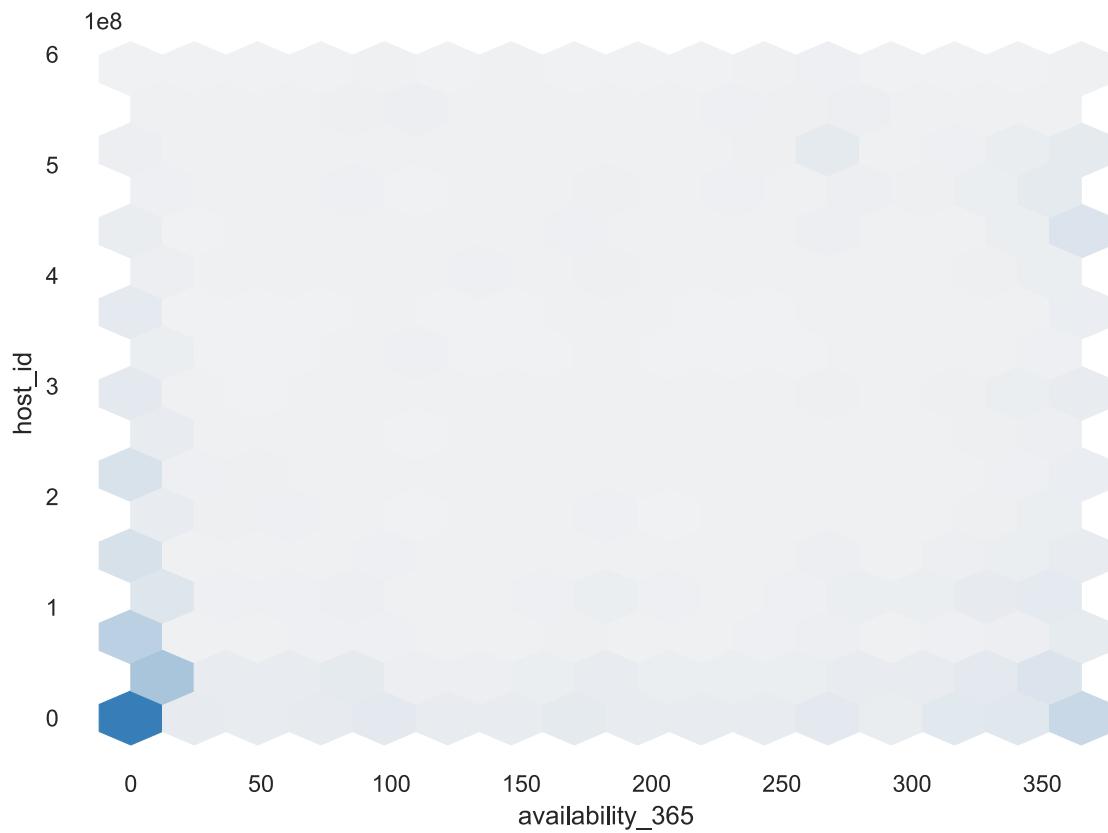


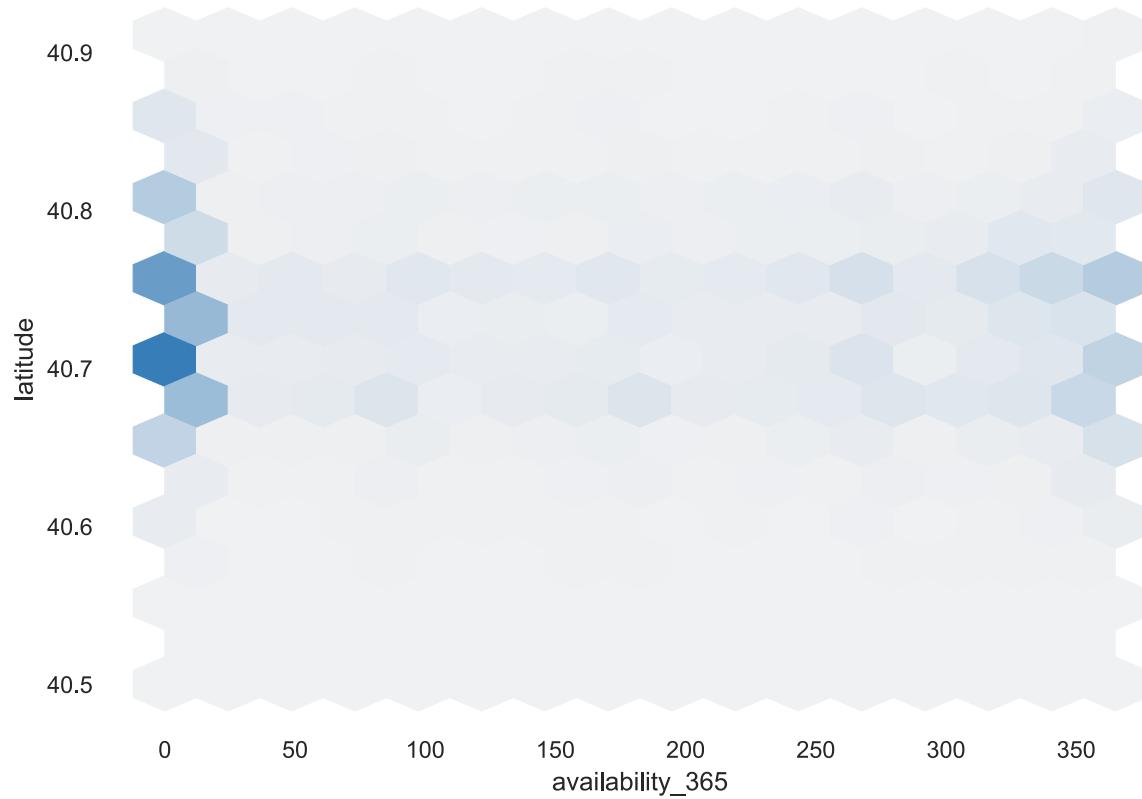


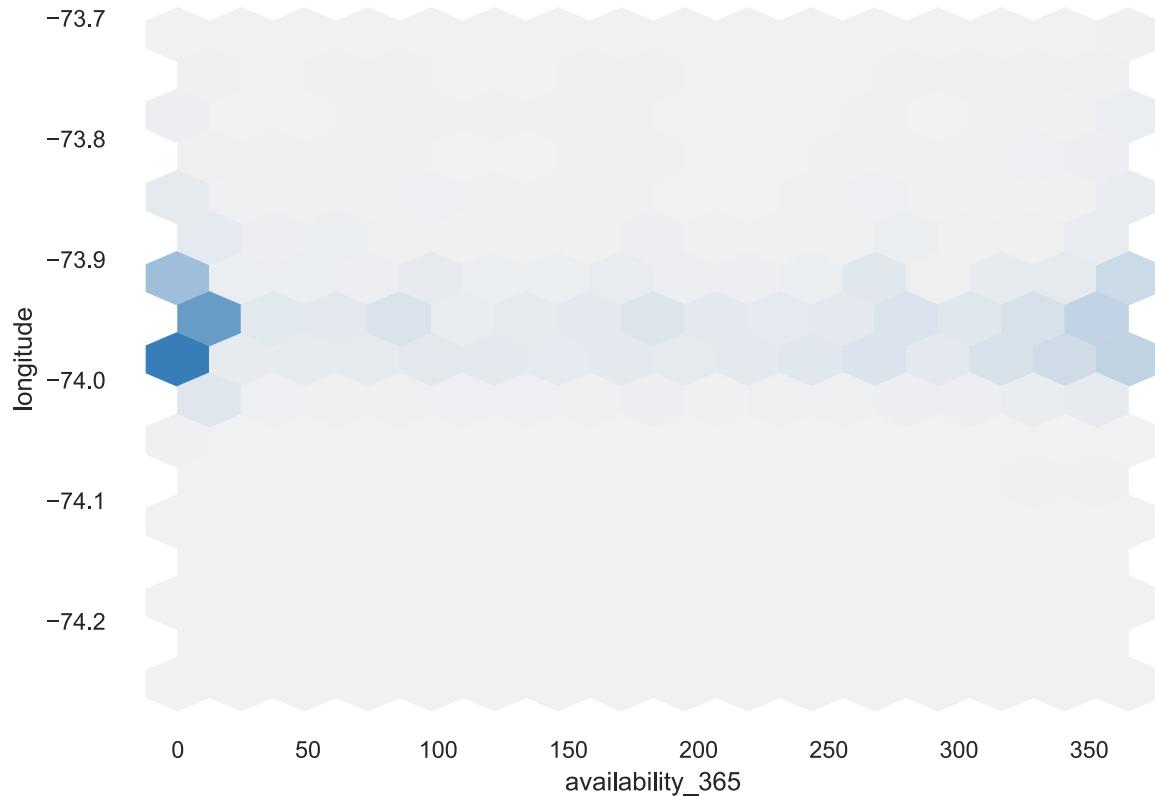

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

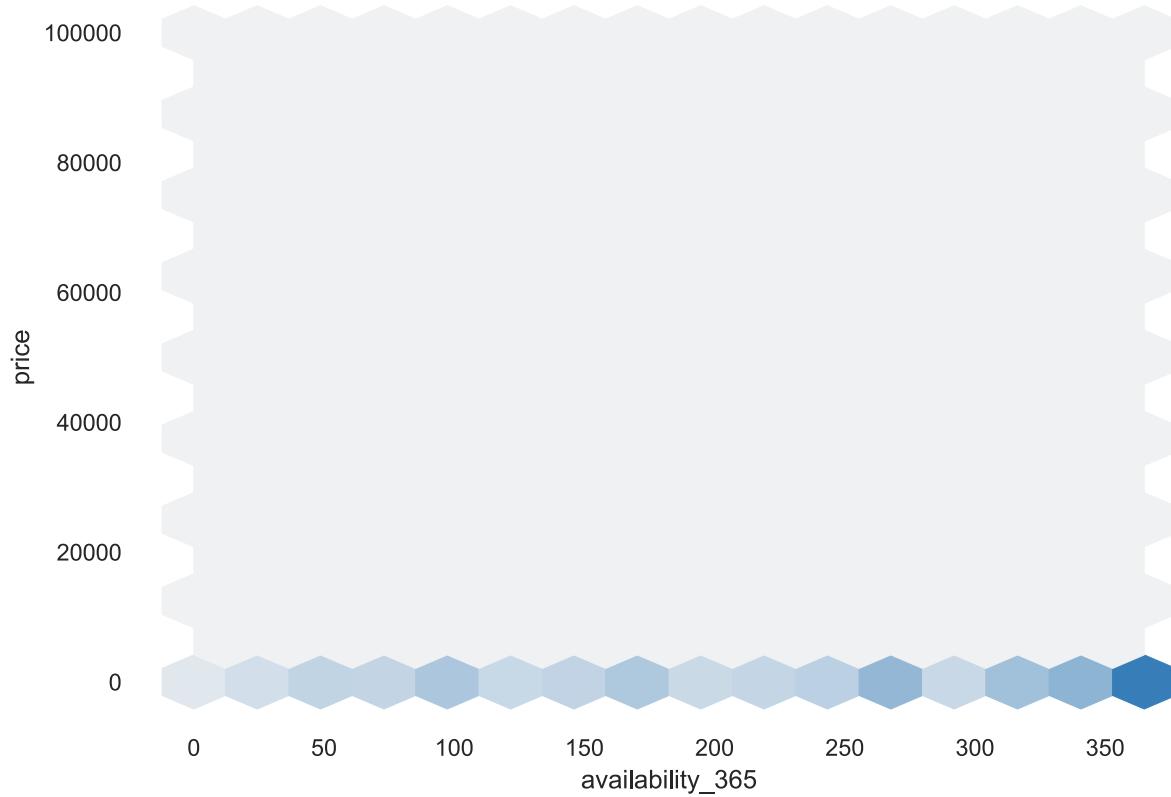


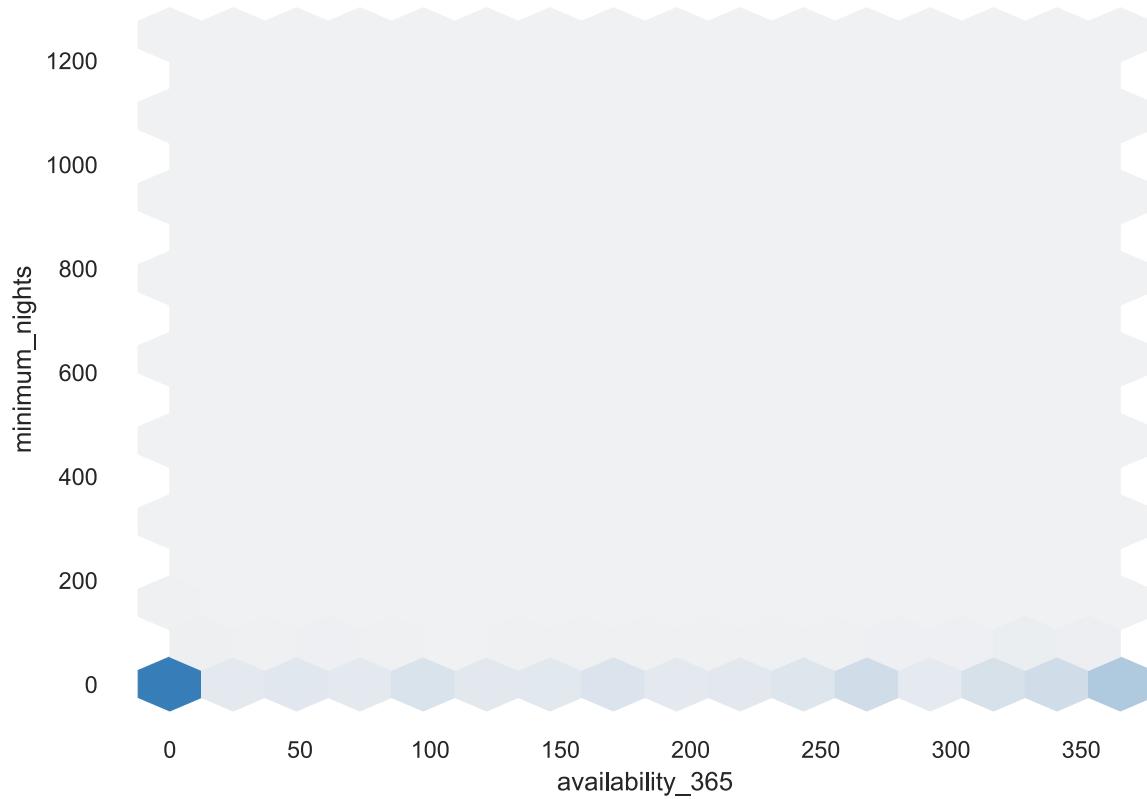


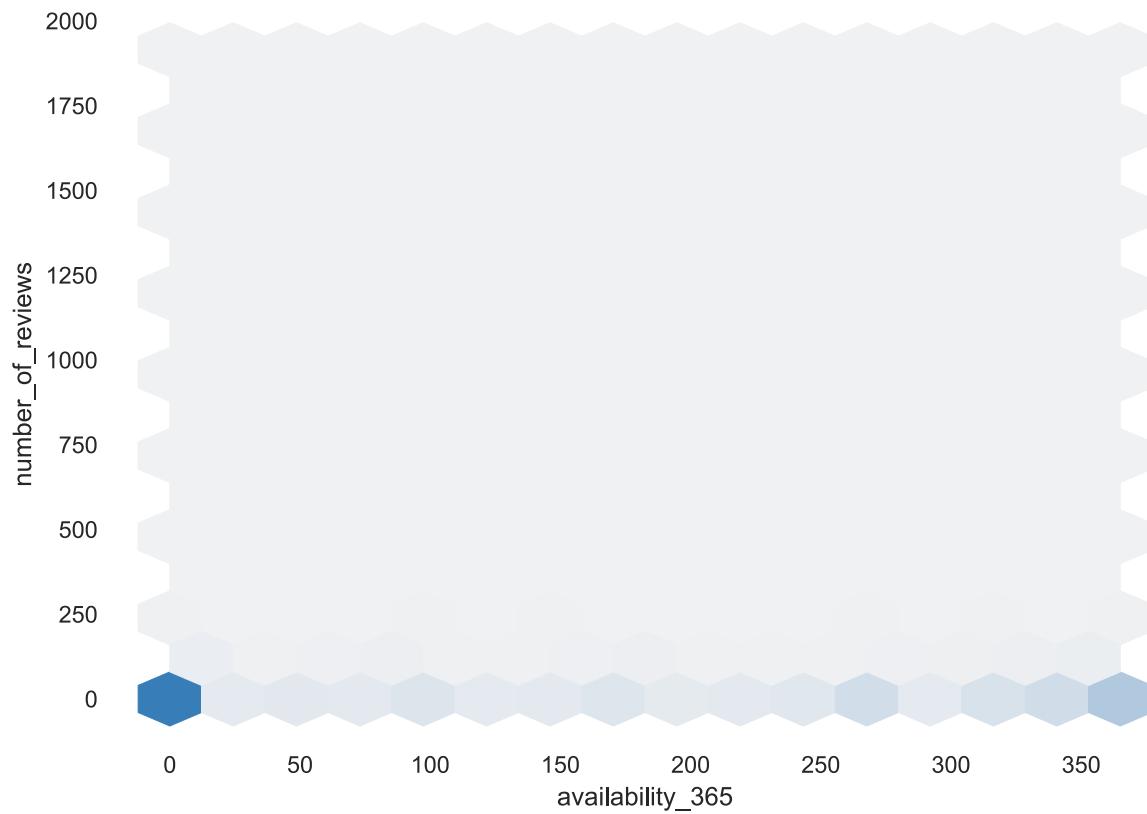


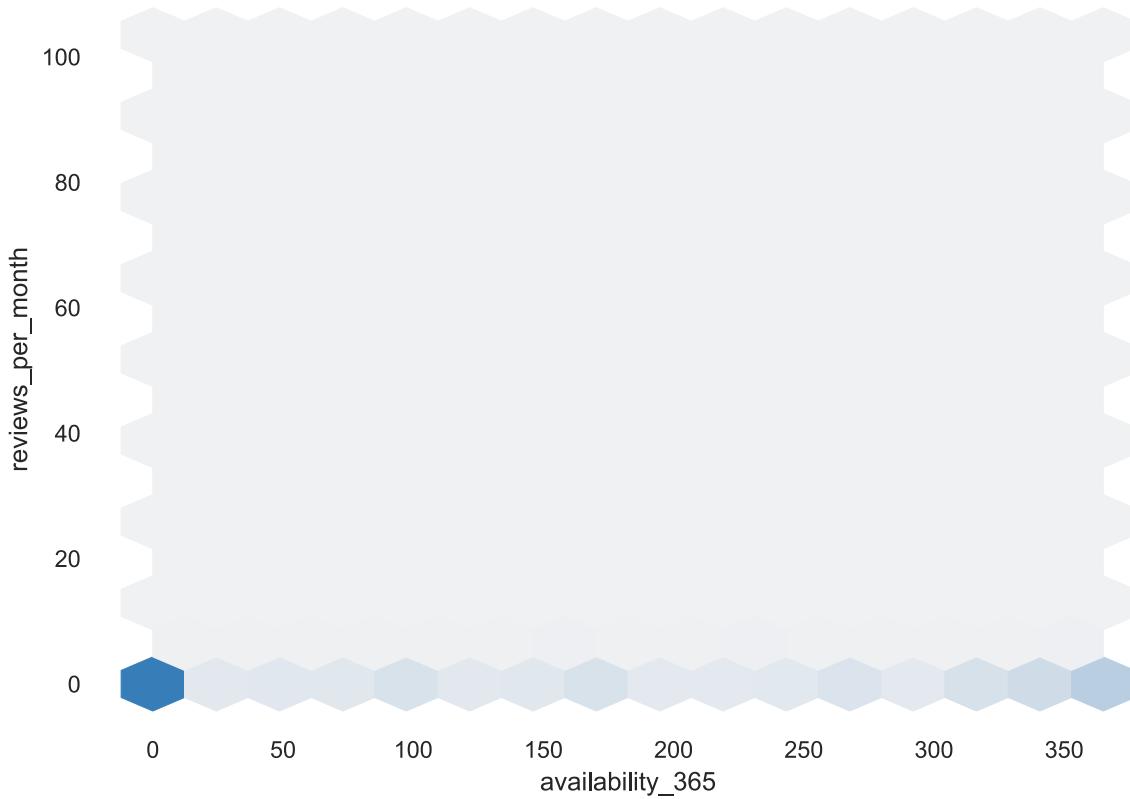


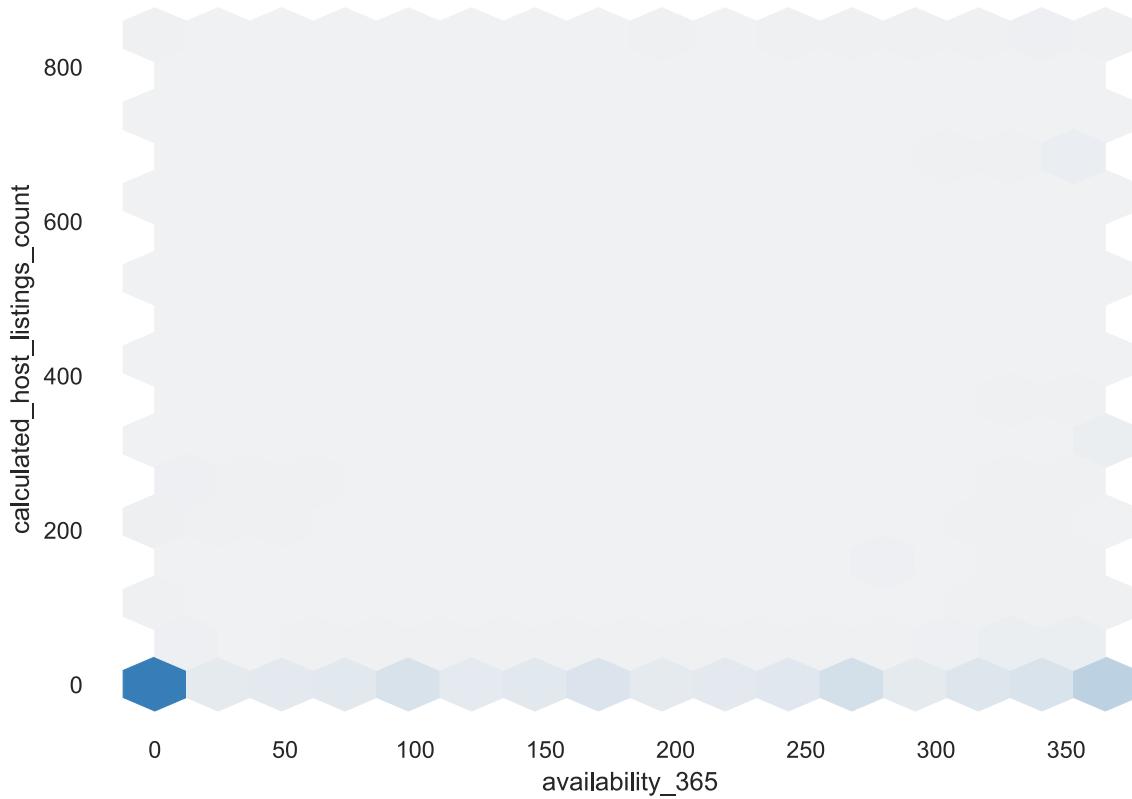


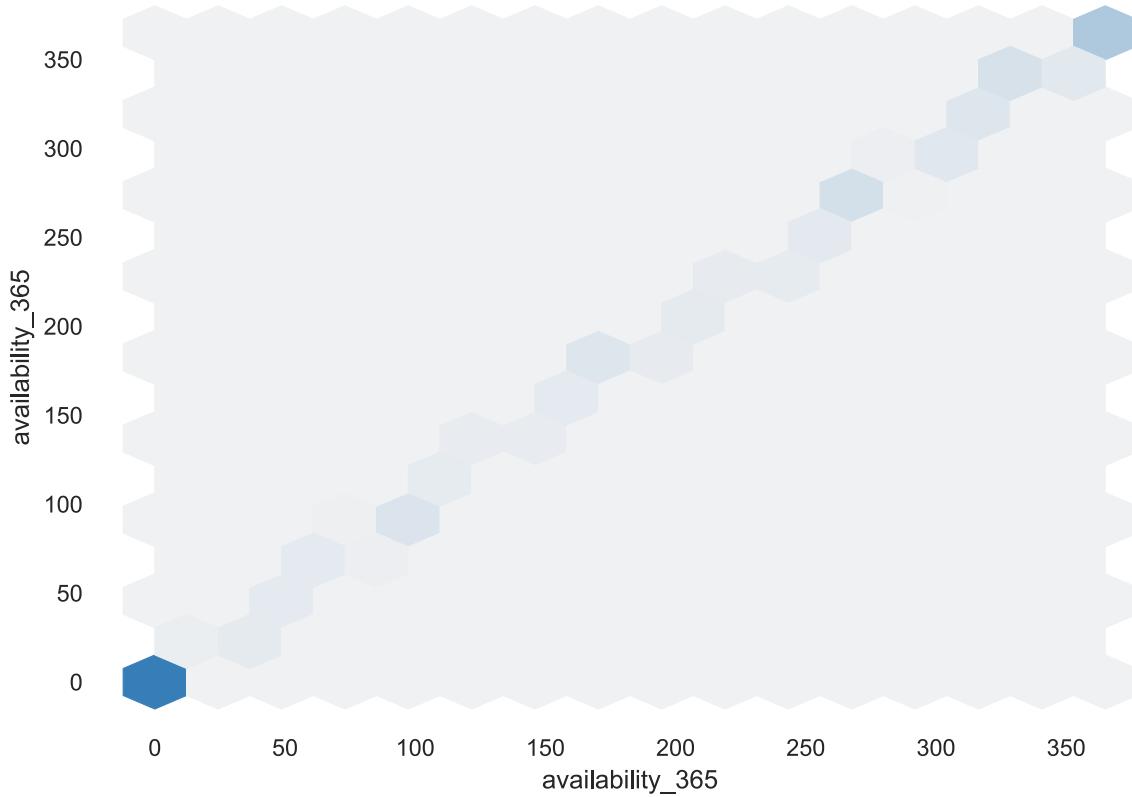




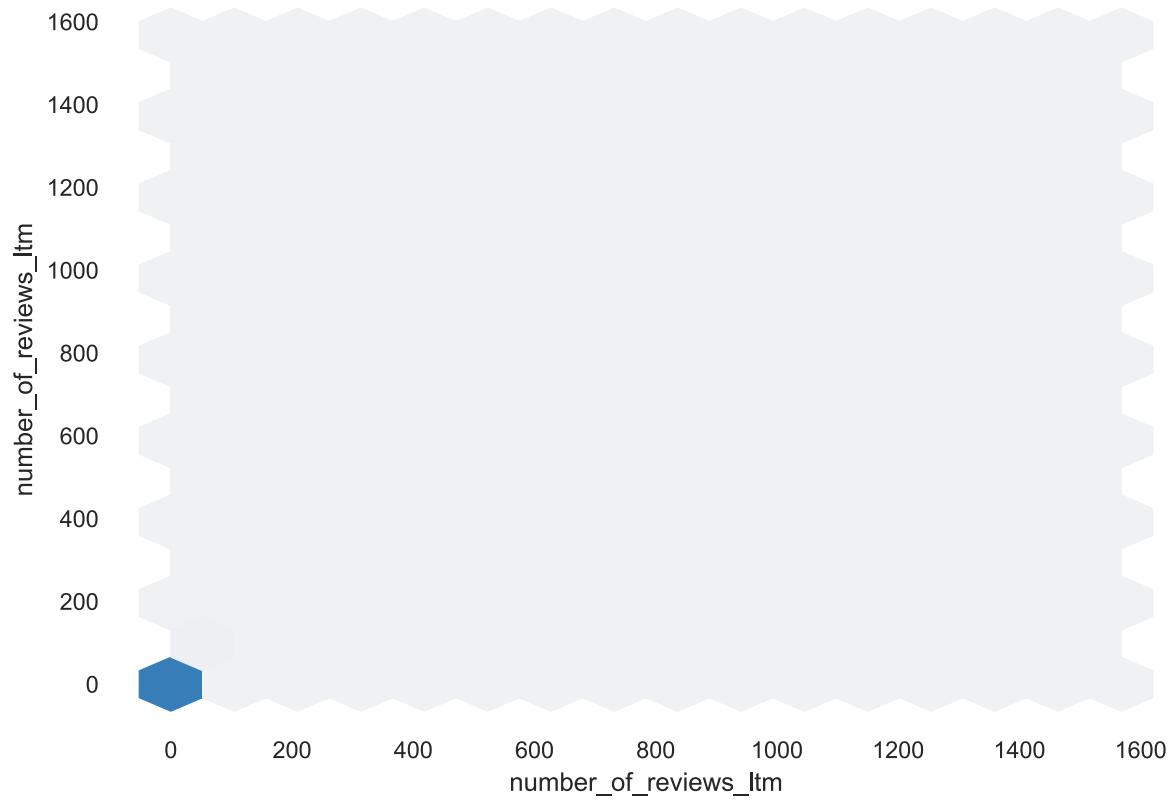


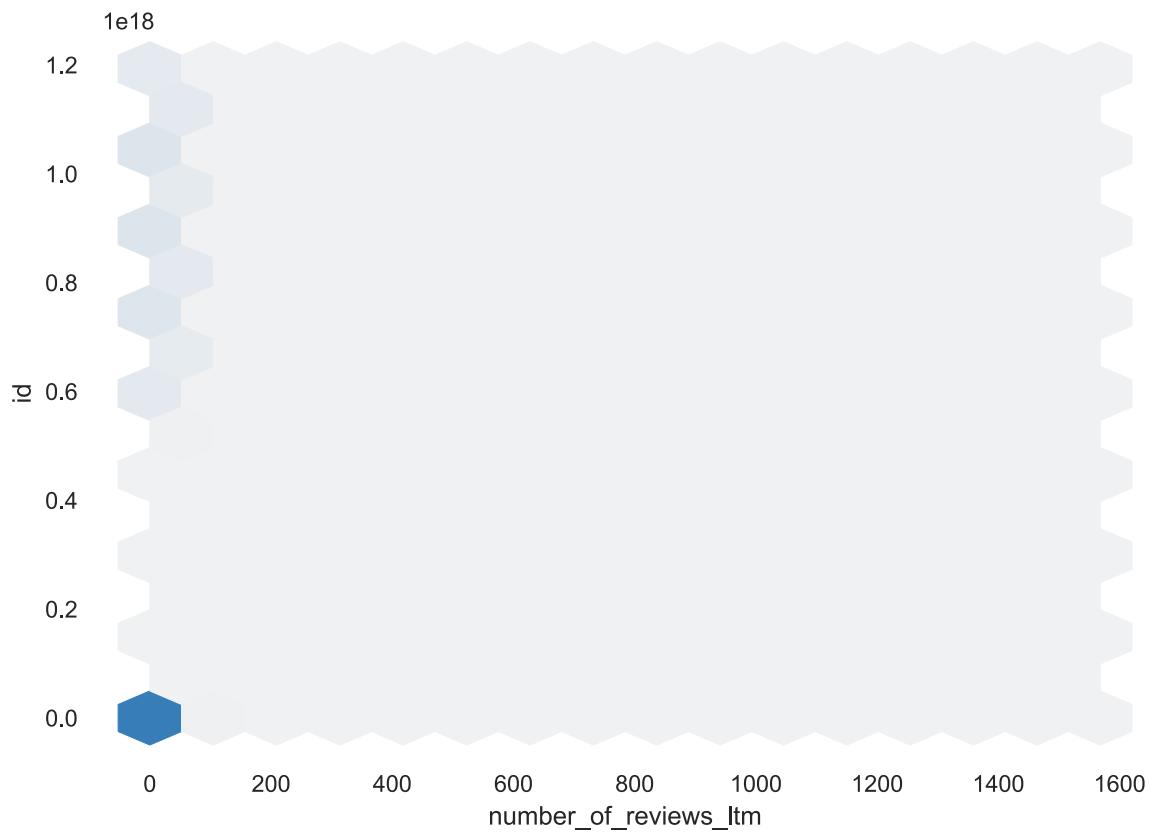


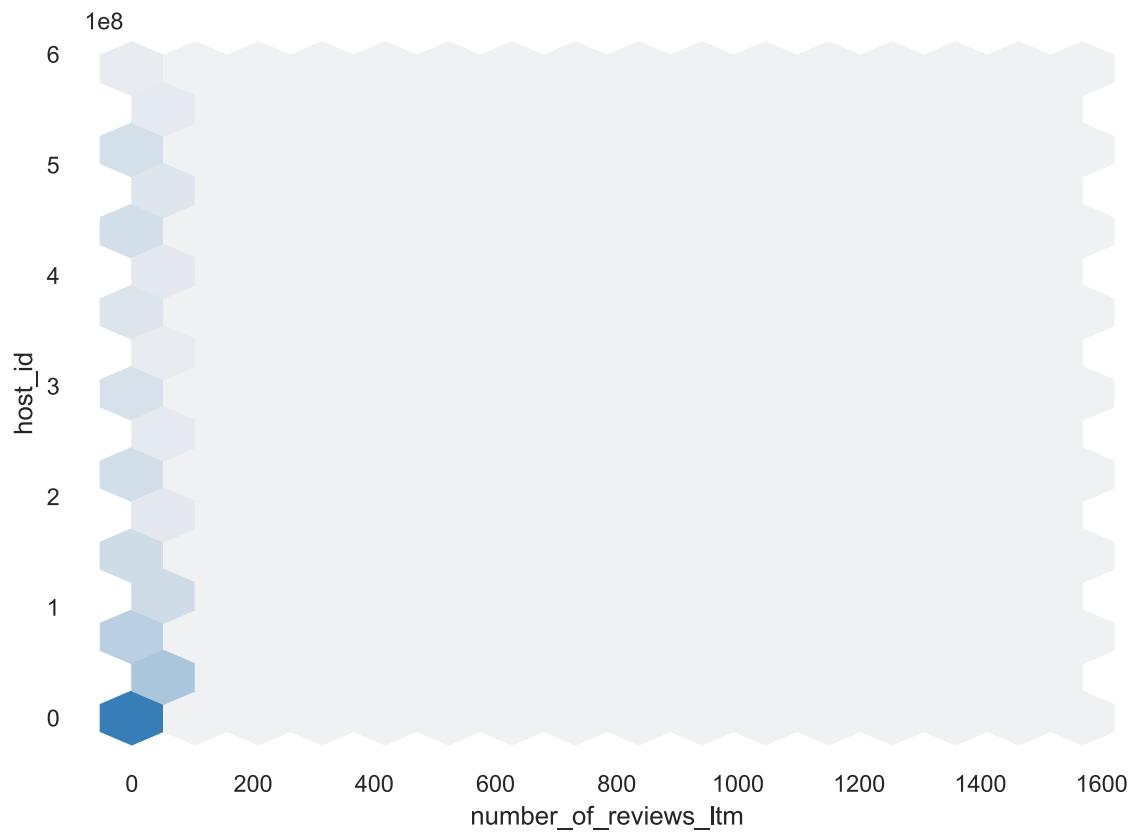


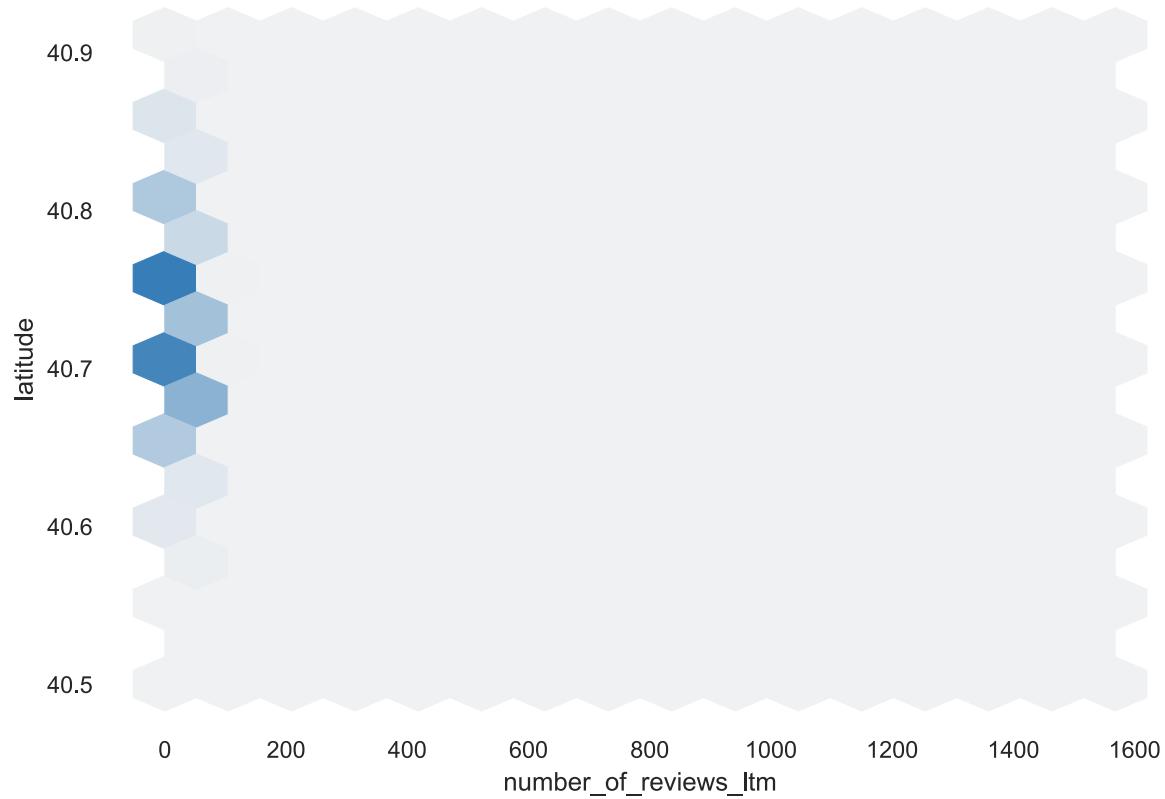


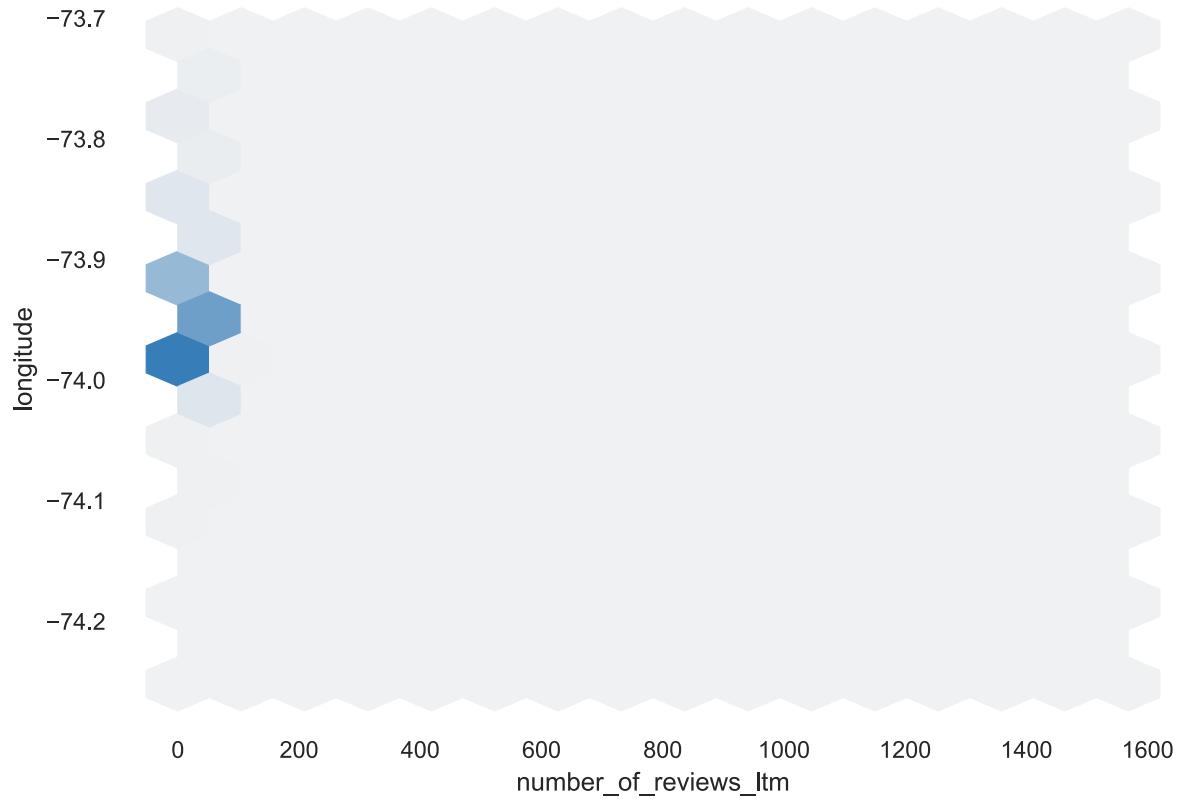

```
number_of_reviews_ltm  
id  
host_id  
latitude  
longitude  
price  
minimum_nights  
number_of_reviews  
reviews_per_month  
calculated_host_listings_c  
availability_365
```

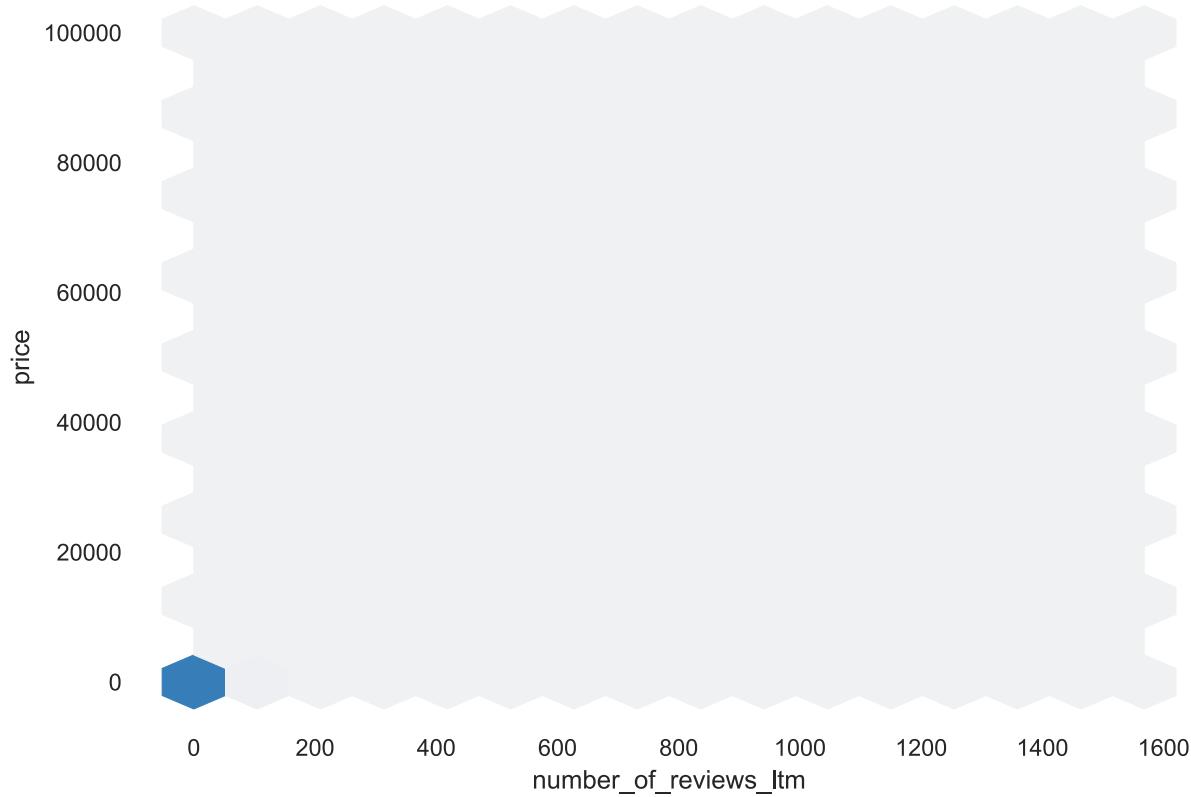


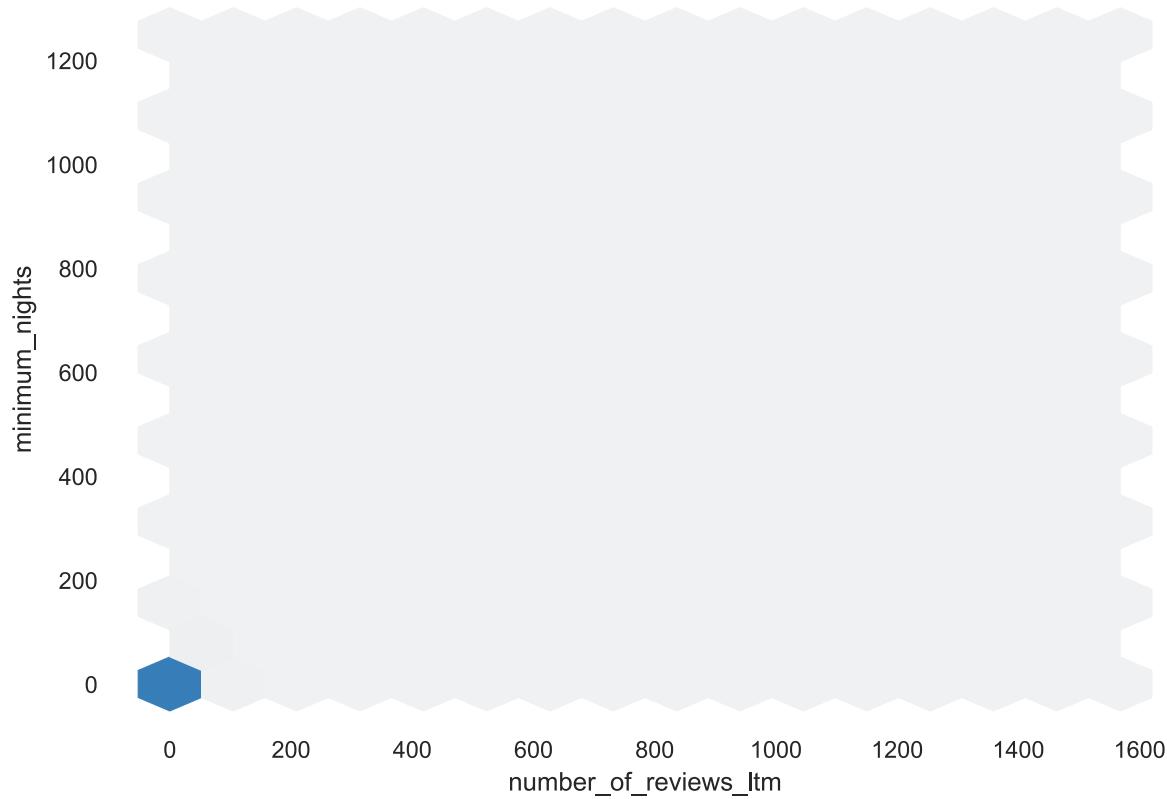


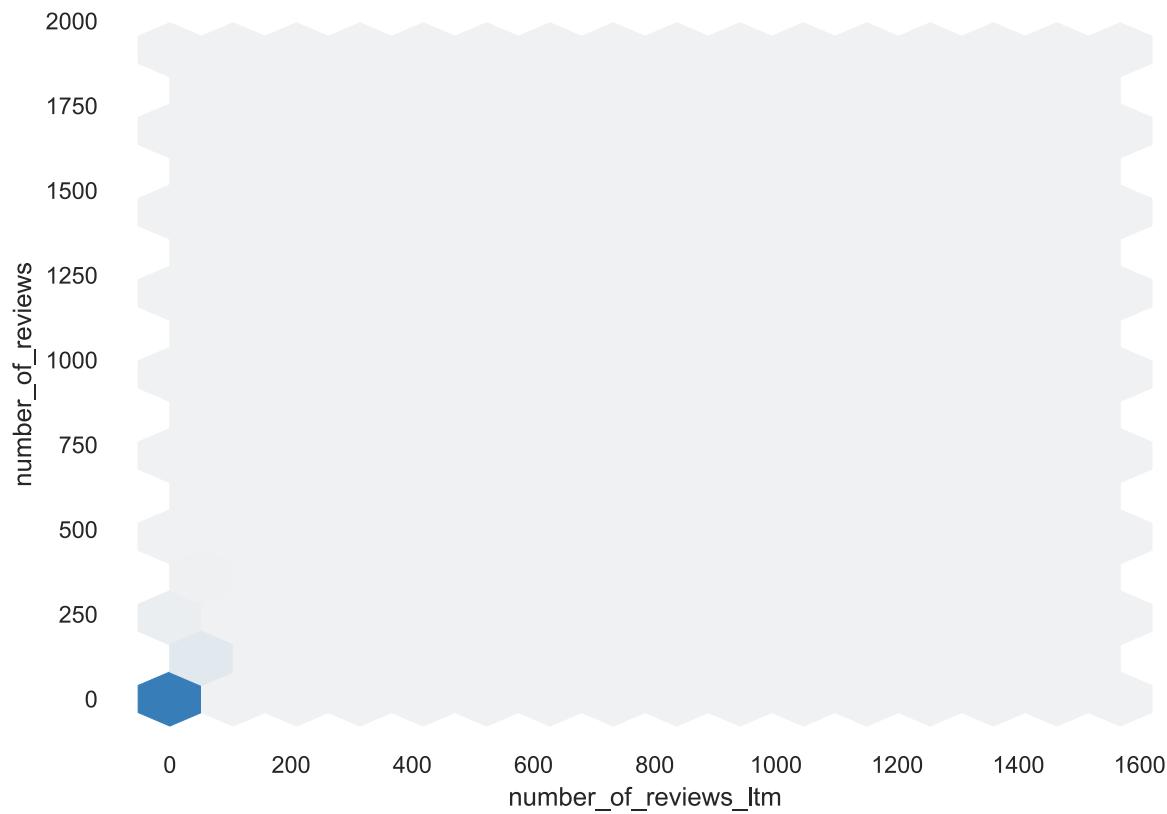


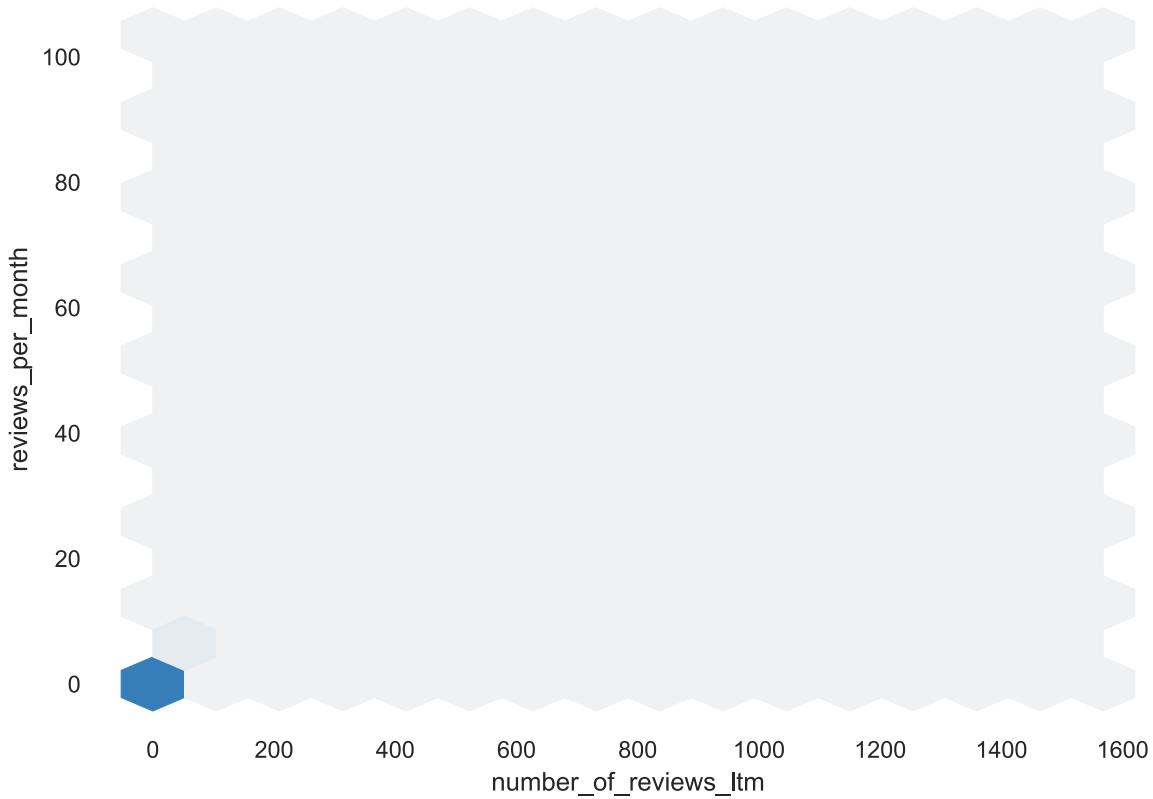


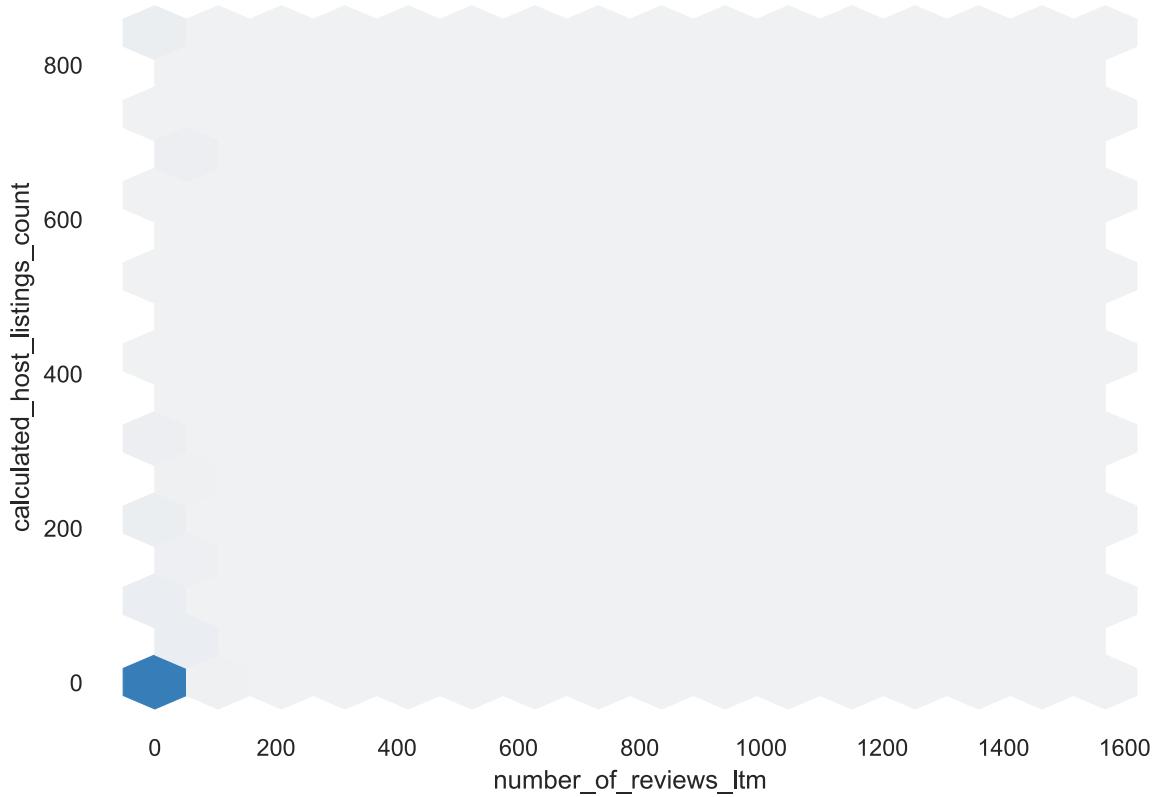


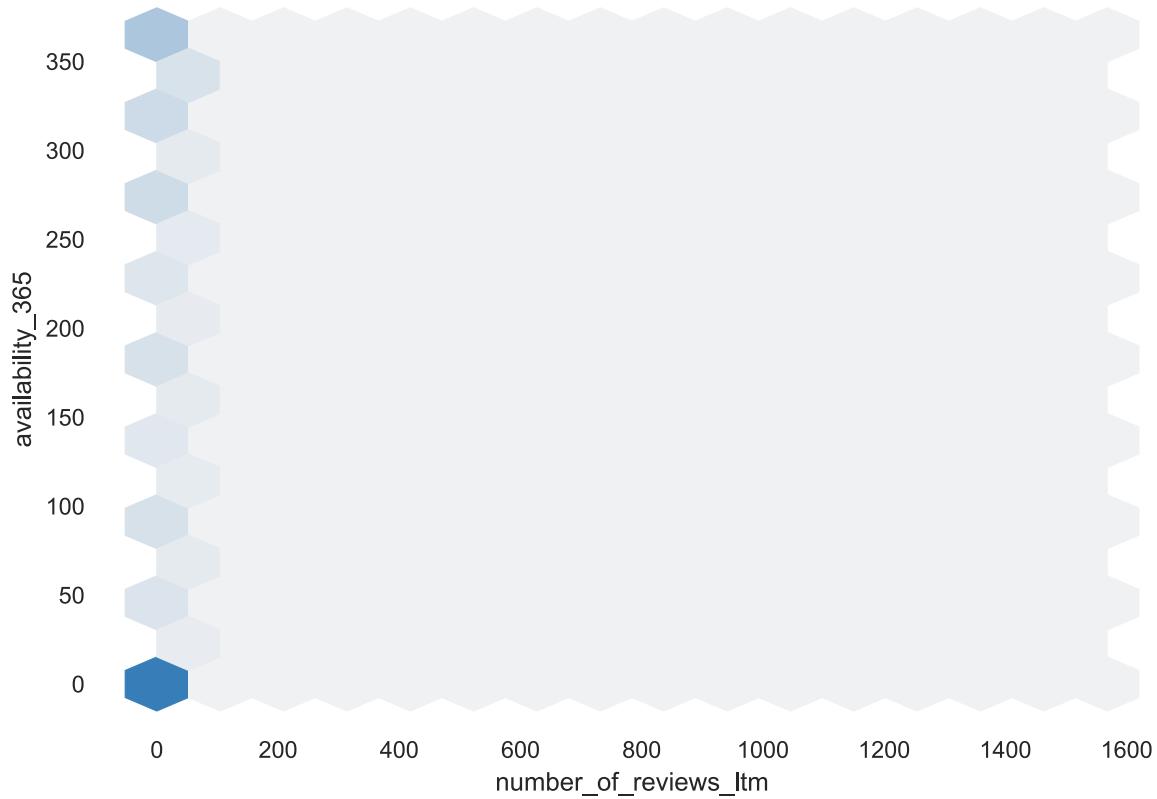




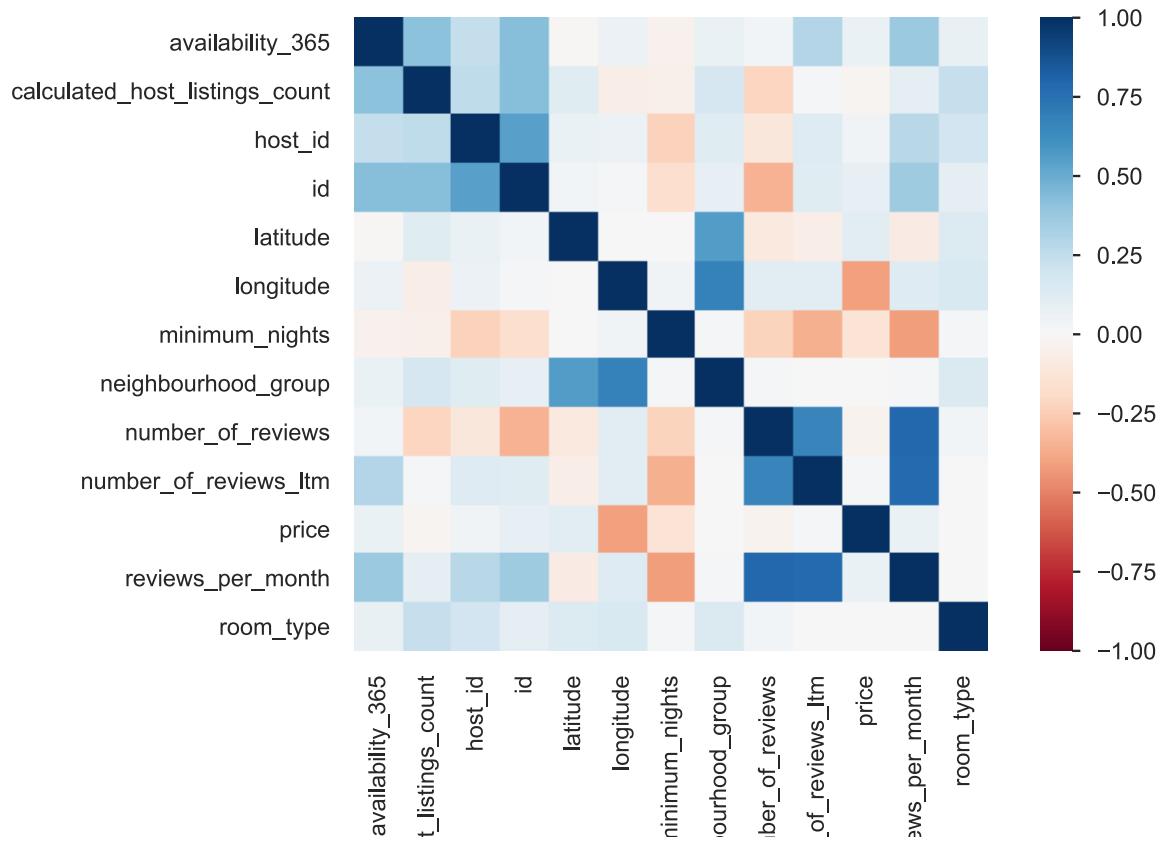






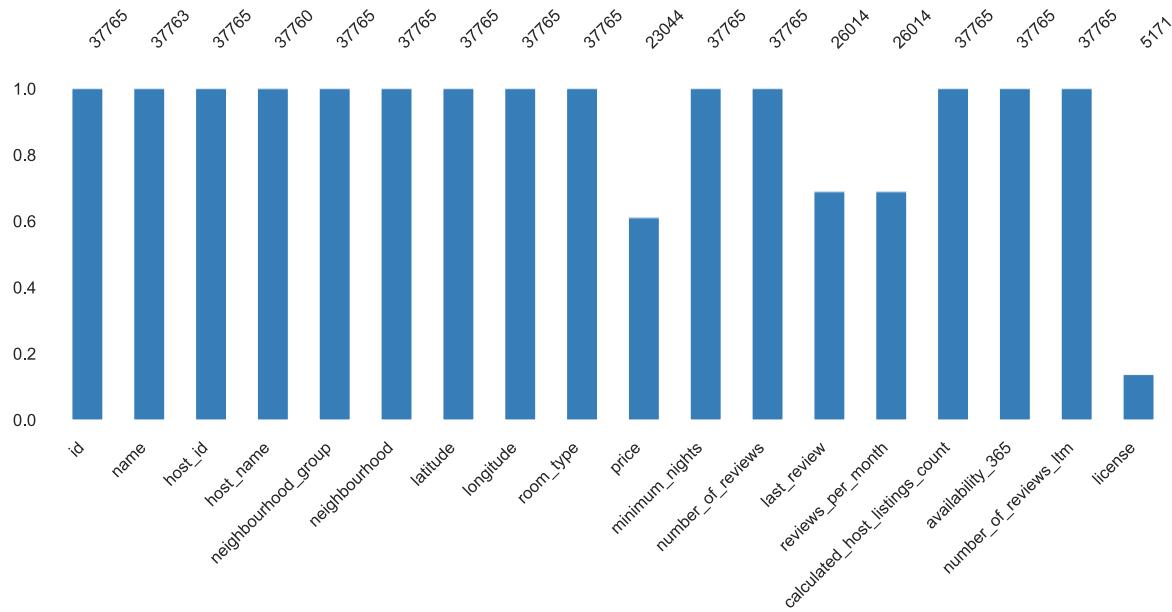


Correlations

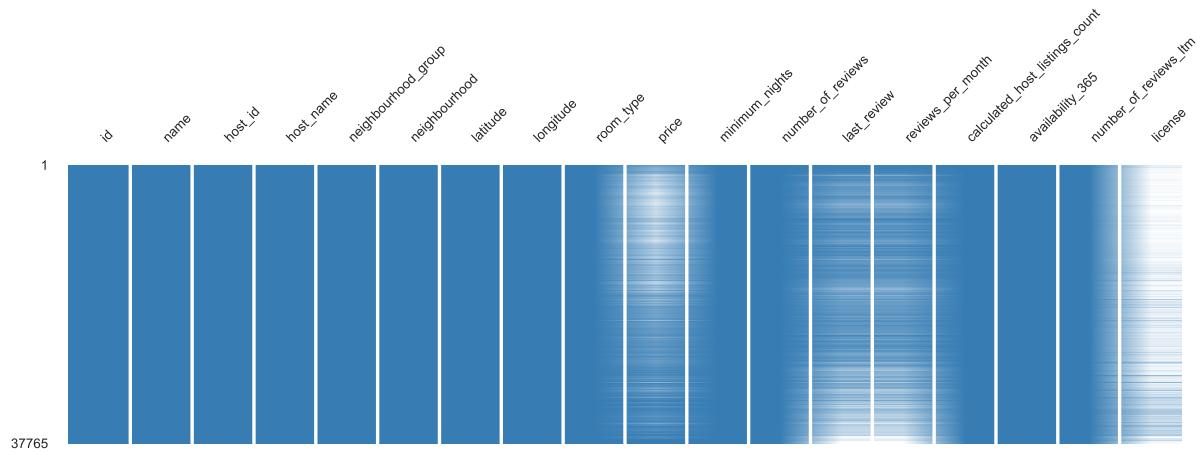


	availability_365	calculated_host_listings_count	host_id	id	latitude	longitude	minimum_nights	neighl
availability_365	1.000	0.409		0.241	0.422	-0.013	0.061	-0.040
calculated_host_listings_count	0.409	1.000		0.254	0.422	0.124	-0.070	-0.051
host_id	0.241	0.254		1.000	0.542	0.070	0.062	-0.233
id	0.422	0.422		0.542	1.000	0.024	0.022	-0.166
latitude	-0.013	0.124		0.070	0.024	1.000	0.001	-0.008
longitude	0.061	-0.070		0.062	0.022	0.001	1.000	0.036
minimum_nights	-0.040	-0.051		-0.233	-0.166	-0.008	0.036	1.000
neighbourhood_group	0.069	0.166		0.119	0.079	0.555	0.675	0.016
number_of_reviews	0.031	-0.213		-0.115	-0.350	-0.098	0.107	-0.226
number_of_reviews_ltm	0.292	0.014		0.129	0.123	-0.070	0.105	-0.358
								0.000

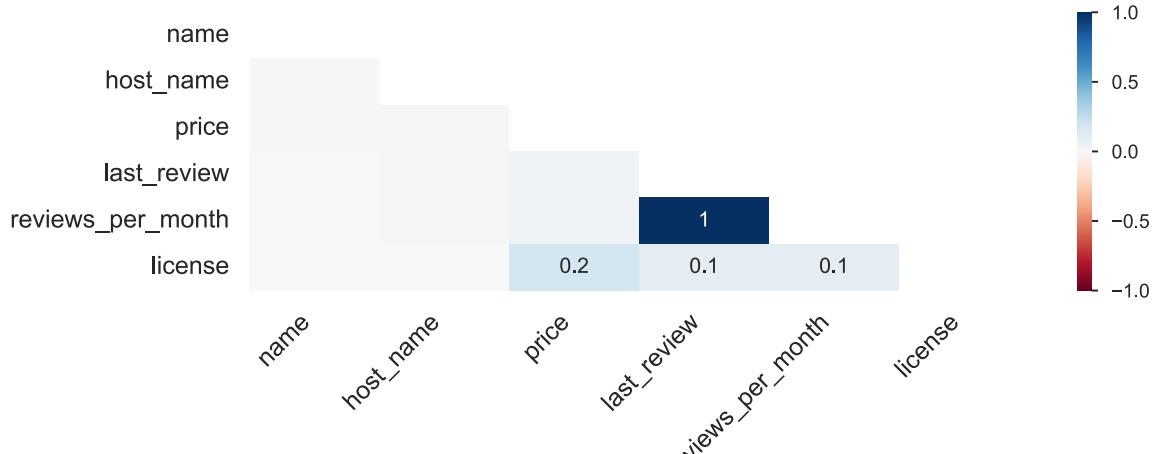
Missing values



A simple visualization of nullity by column.



Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.



The correlation heatmap measures nullity correlation: how strongly the presence or absence of one variable affects the presence of another.

Sample

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
0	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.753560	-73.98558
1	5136	Spacious Family Friendly Duplex w/ Patio + Yard	7378	Rebecca	Brooklyn	Sunset Park	40.662650	-73.99454
2	6848	Only 2 stops to Manhattan studio	15991	Allen & Irina	Brooklyn	Williamsburg	40.709350	-73.95342
3	6872	Uptown Sanctuary w/ Private Bath (Month to Month)	16104	Kae	Manhattan	East Harlem	40.801070	-73.94255
4	6990	UES Beautiful Blue Room	16800	Cyn	Manhattan	East Harlem	40.787780	-73.94758
5	7064	Amazing location! Wburg. Large, bright & tranquil	17297	Joelle	Brooklyn	Williamsburg	40.712480	-73.95881
6	7097	Perfect for Your Parents, With Garden & Patio	17571	Jane	Brooklyn	Fort Greene	40.691940	-73.97385
7	7801	Sunny Williamsburg Loft with Sauna	21207	Chaya	Brooklyn	Williamsburg	40.718807	-73.95617
8	8490	Maison des Sirenes1,bohemian, luminous apartment	25183	Nathalie	Brooklyn	Bedford-Stuyvesant	40.684556	-73.93963
9	9357	Midtown Pied-a-terre	30193	Tommi	Manhattan	Hell's Kitchen	40.767240	-73.98664

id	name		host_id	host_name	neighbourhood_group	neighbourhood
37755	1192723874961256642	Classic Private Room #1361 C	305240193	Elena	Manhattan	Lower East Side
37756	1192788527851447904	One-Bedroom Apartment in the Financial District	22541573	Furnished Quarters	Manhattan	Financial District
37757	1192828515848213902	Barrow by Rove Travel 1BR Loft w/Shared Roof	17775359	Rove	Manhattan	West Village
37758	1192899750875253463	Apollo by Rove Travel 3BR Duplex w/Garden	17775359	Rove	Manhattan	Harlem
37759	1192903501898923056	3 bed floor-through in Diplomatic District	683230	Thomas	Manhattan	Midtown
37760	1192946306020922085	Private Studio in Brooklyn	137493949	Sydney	Brooklyn	Clinton Hill
37761	1193005542763264121	Skyline Views from MSG Penthouse	559434422	Luis	Manhattan	Chelsea
37762	1193016130087171760	Stunning designer Chelsea studio on the best block	35491667	Nat	Manhattan	Chelsea
37763	1193068036013405081	Sunny & Spacious in Queens	20380663	Anastasios	Queens	Elmhurst
37764	1193086058992851452	Bright 2 Bedroom in Astoria	910709	Israel	Queens	Ditmars Street

Report generated by YData (https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report).