Overview

Dataset statistics

Number of variables	17
Number of observations	267751
Missing cells	84142
Missing cells (%)	1.8%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	34.7 MiB
Average record size in memory	136.0 B

Variable types

Numeric	5
Text	8
Categorical	3
DateTime	1

Alerts

`State` is highly imbalanced (> 99.9%)	Imbalance
`Facility Type` has 5119 (1.9%) missing values	Missing
`Violations` has 73407 (27.4%) missing values	Missing
`Zip` is highly skewed (γ1 = -111.2126359)	Skewed
`Inspection ID` has unique values	Unique

Reproduction

Analysis started	2024-02-22 02:19:04.607046
Analysis finished	2024-02-22 02:21:24.983656
Duration	2 minutes and 20.38 seconds
Software version	ydata-profiling v0.0.dev0 (https://github.com/ydataai/ydata-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Report%20for%20Food%20Inspection%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%

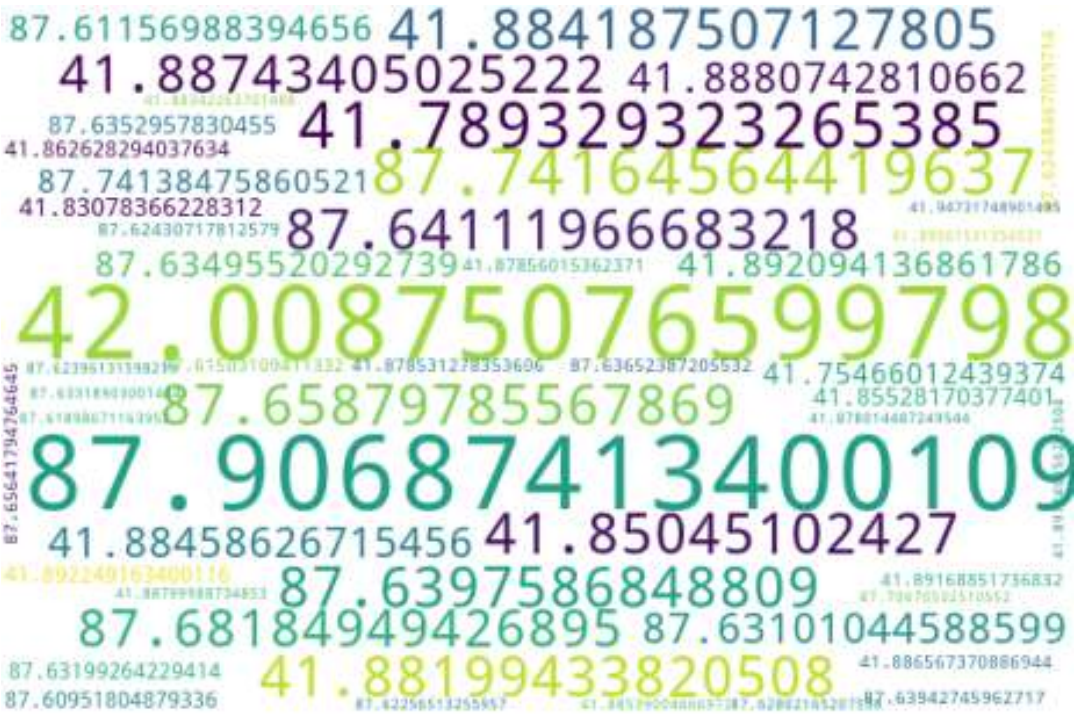
Variables

Location

Location

Text

Distinct	18215
Distinct (%)	6.8%
Missing	925
Missing (%)	0.3%
Memory size	2.0 MiB



Length

Max length	40
Median length	39
Mean length	39.059649
Min length	34

Characters and Unicode

Total characters	10422130	
Distinct characters	16	
Distinct categories	6 (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.

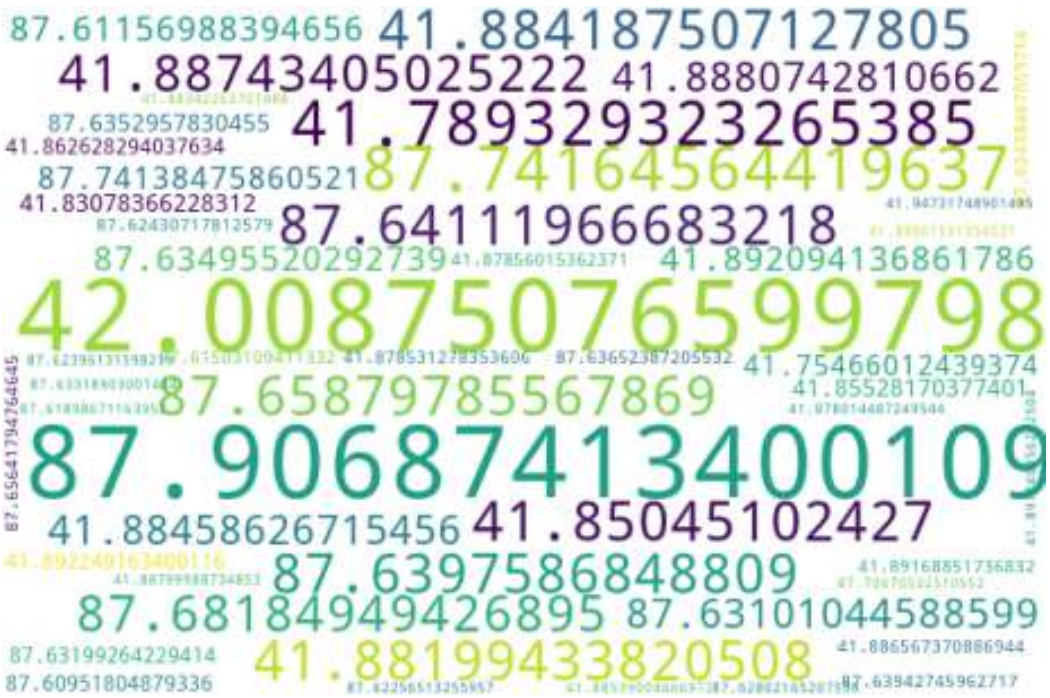
Unique

Unique	1595	?
Unique (%)	0.6%	

Sample

1st row	(41.84560861006466, -87.69538780161594)
2nd row	(41.799368015121125, -87.59468625684886)
3rd row	(41.894208809612294, -87.66738222845791)
4th row	(41.86629065519555, -87.72456177467829)
5th row	(41.7710280071688, -87.58641718666685)

Value	Count	Frequency (%)
42.00875076599798	3291	0.6%
87.90687413400109	3291	0.6%
41.789329323265385	768	0.1%
87.74164564419637	768	0.1%
41.884187507127805	597	0.1%
87.64111966683218	597	0.1%
41.88199433820508	586	0.1%
87.6397586848809	586	0.1%
41.85045102427	584	0.1%
87.65879785567869	584	0.1%
Other values (36420)	522000	97.8%



Most occurring characters

Value	Count	Frequency (%)
8	1079368	10.4%
7	1072631	10.3%
4	989479	9.5%
1	924788	8.9%
6	906312	8.7%
9	809228	7.8%
5	730762	7.0%
2	703163	6.7%
3	688983	6.6%
0	649634	6.2%
Other values (6)	1867782	17.9%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	8554348	82.1%
Other Punctuation	800478	7.7%
Open Punctuation	266826	2.6%
Space Separator	266826	2.6%
Dash Punctuation	266826	2.6%
Close Punctuation	266826	2.6%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
8	1079368	12.6%
7	1072631	12.5%
4	989479	11.6%
1	924788	10.8%
6	906312	10.6%
9	809228	9.5%
5	730762	8.5%
2	703163	8.2%
3	688983	8.1%
0	649634	7.6%

Other Punctuation

Value	Count	Frequency (%)
.	533652	66.7%
,	266826	33.3%

Open Punctuation

Value	Count	Frequency (%)
(	266826	100.0%

Space Separator

Value	Count	Frequency (%)
	266826	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	266826	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	266826	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	10422130	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
8	1079368	10.4%
7	1072631	10.3%
4	989479	9.5%
1	924788	8.9%
6	906312	8.7%
9	809228	7.8%
5	730762	7.0%
2	703163	6.7%
3	688983	6.6%
0	649634	6.2%
Other values (6)	1867782	17.9%

Most occurring blocks

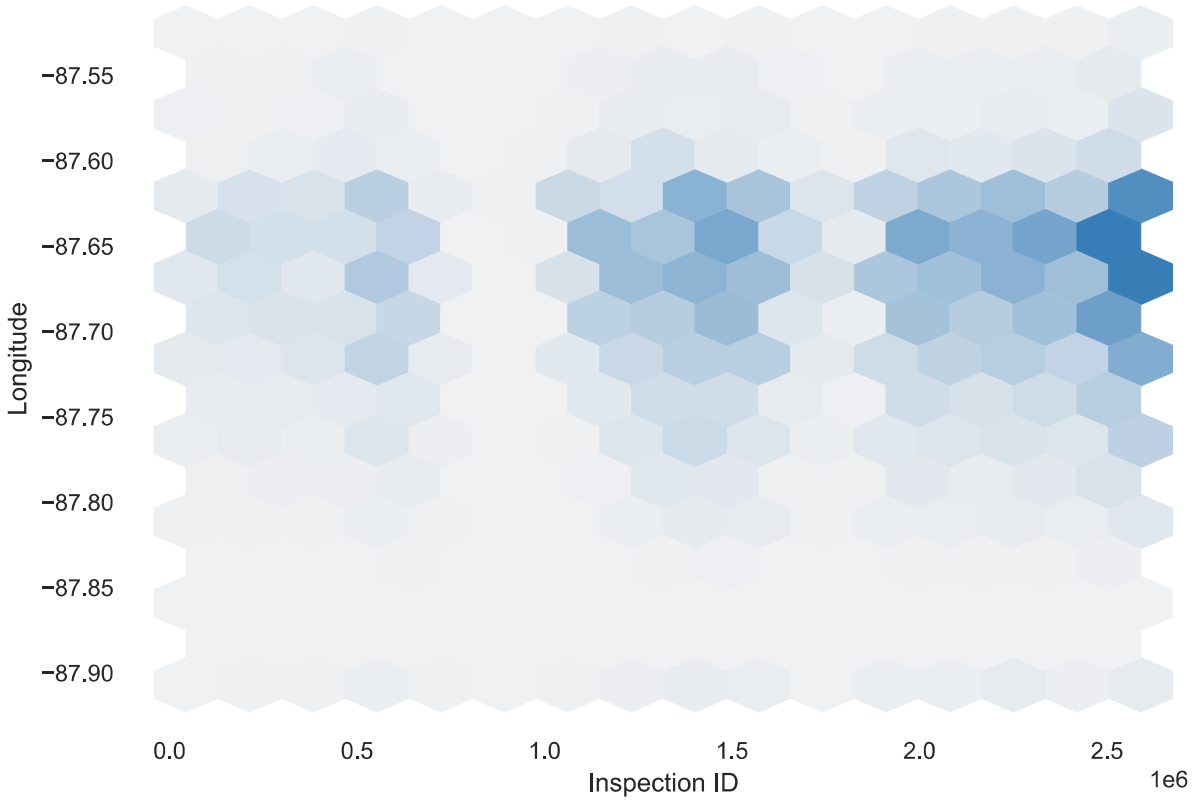
Value	Count	Frequency (%)
ASCII	10422130	100.0%

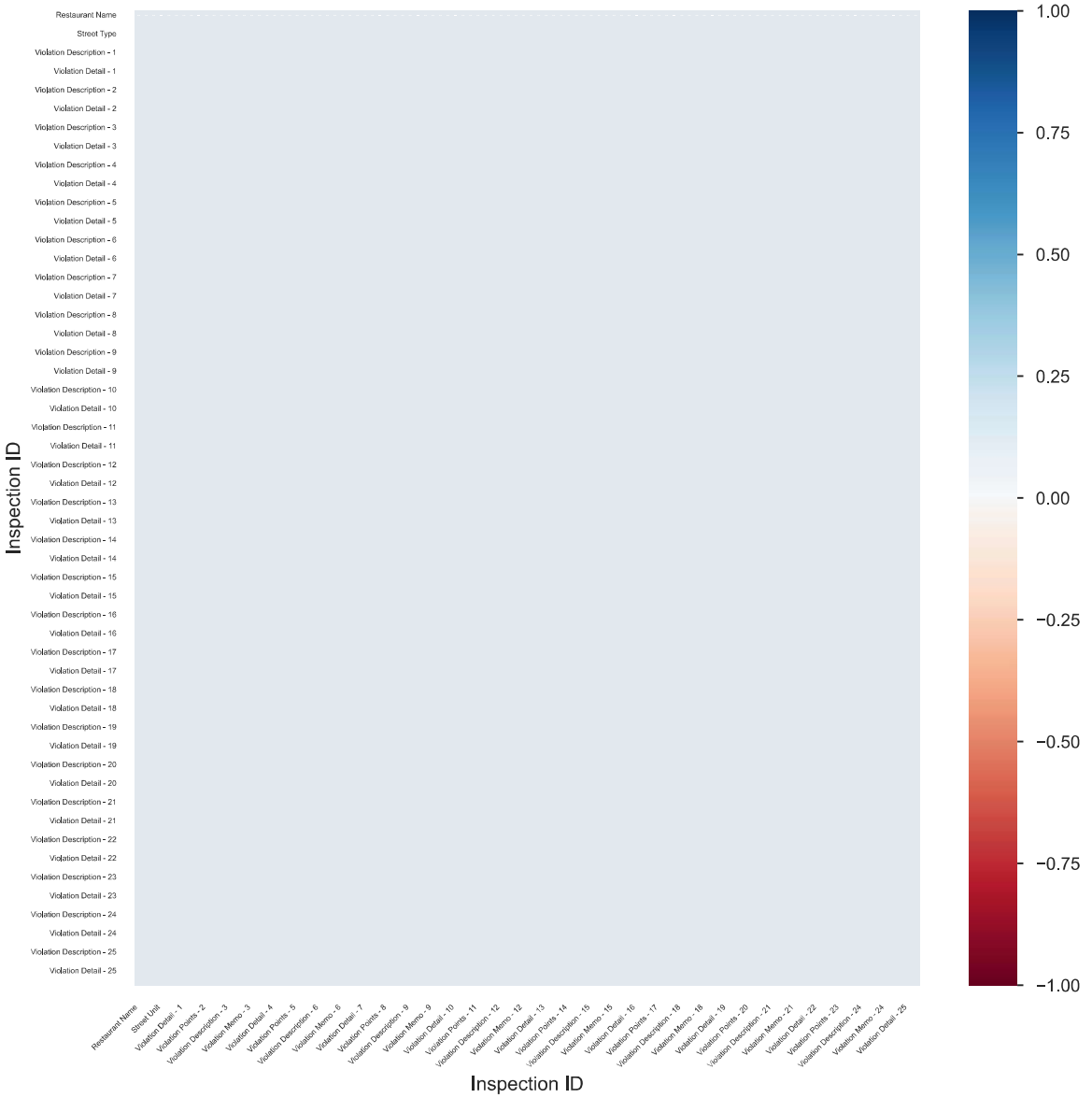
Most frequent character per block

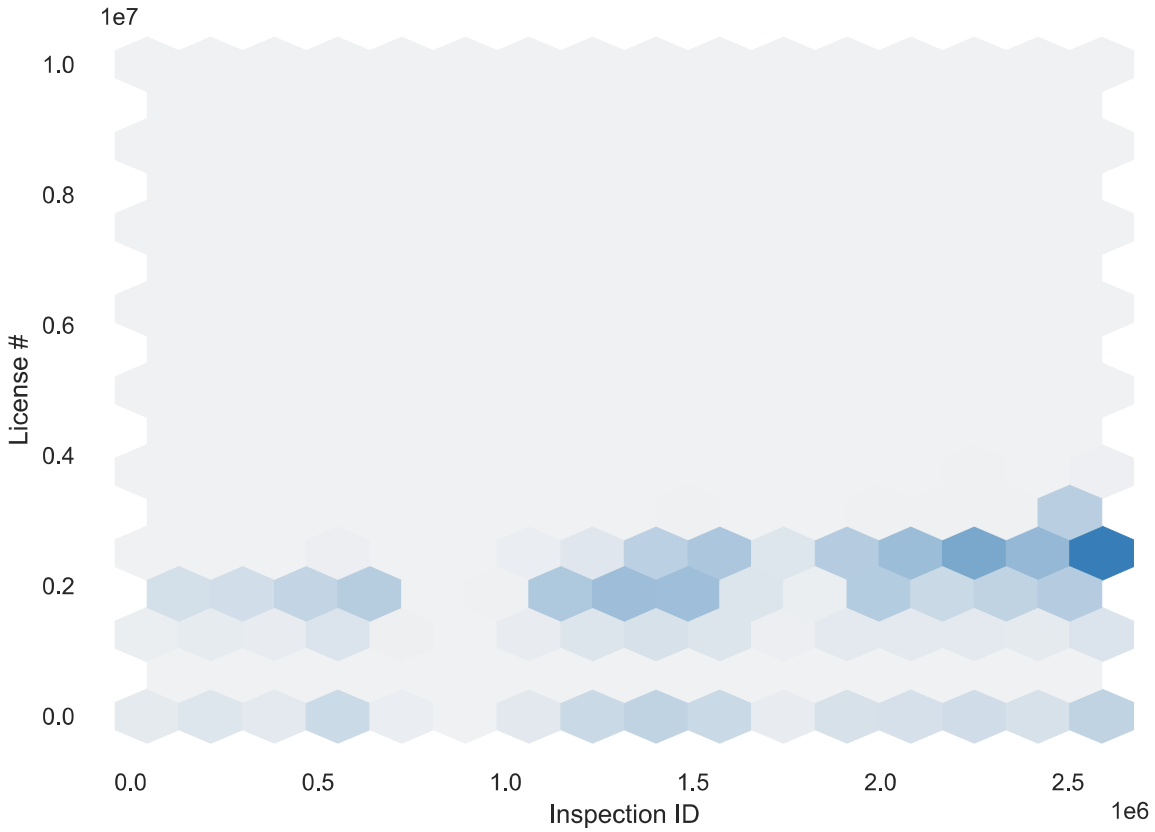
ASCII

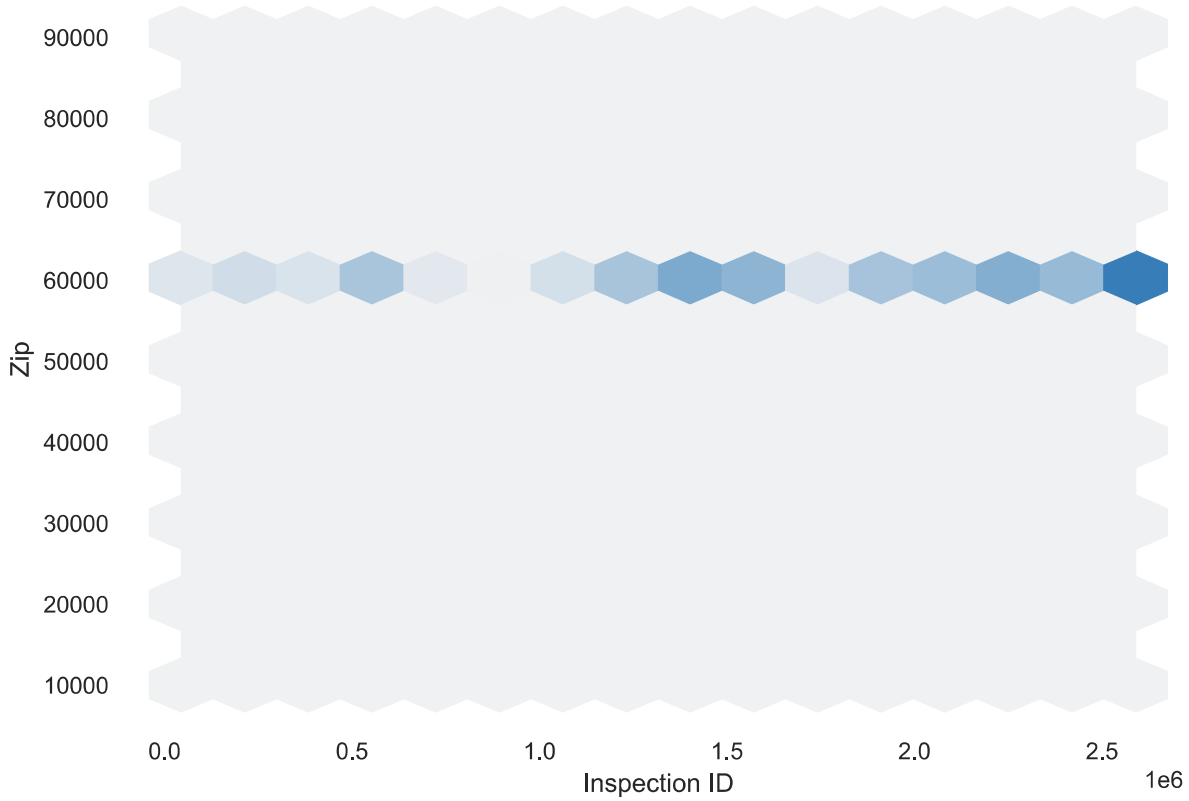
Value	Count	Frequency (%)
8	1079368	10.4%
7	1072631	10.3%
4	989479	9.5%
1	924788	8.9%
6	906312	8.7%
9	809228	7.8%
5	730762	7.0%
2	703163	6.7%
3	688983	6.6%
0	649634	6.2%
Other values (6)	1867782	17.9%

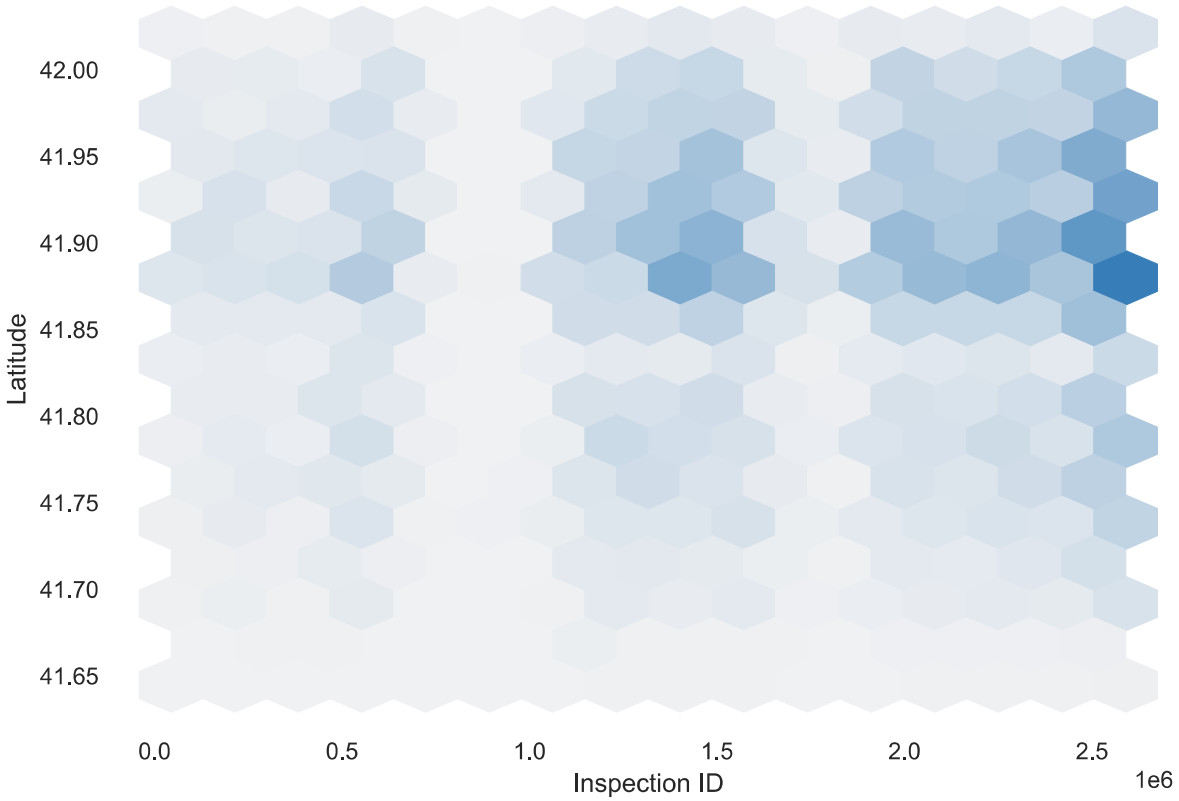
Interactions

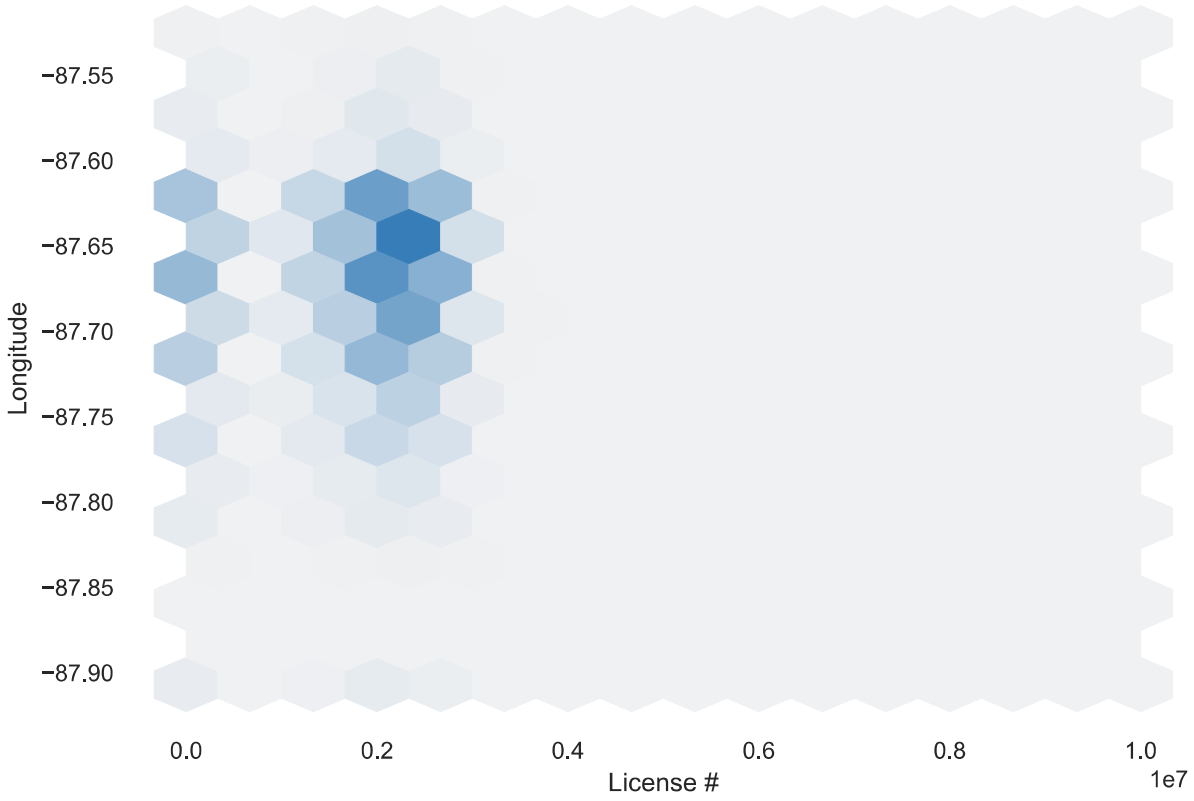


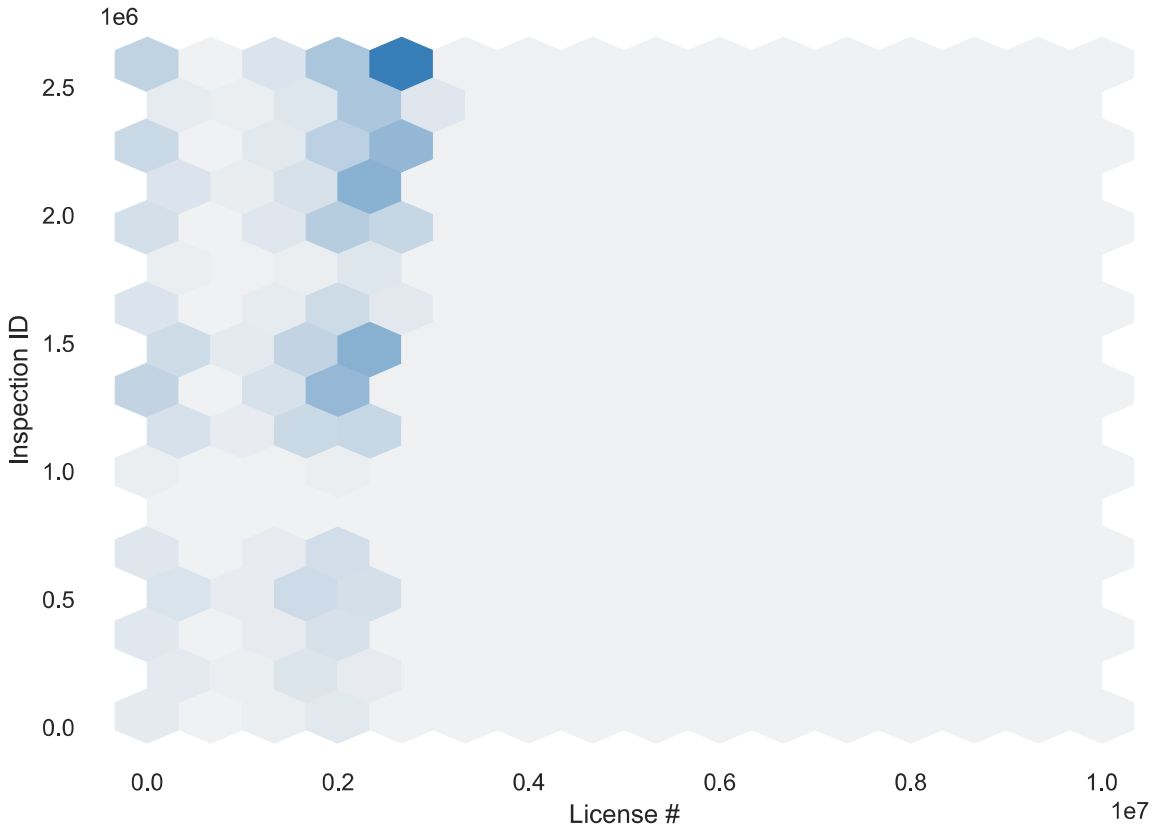

Restaurant Name
Street Unit
Violation Detail - 1
Violation Points - 2
Violation Description - 3
Violation Memo - 3
Violation Detail - 4
Violation Points - 5
Violation Description - 6
Violation Memo - 6
Violation Detail - 7
Violation Points - 8
Violation Description - 9
Violation Memo - 9
Violation Detail - 10
Violation Points - 11
Violation Description - 12
Violation Memo - 12
Violation Detail - 13
Violation Points - 14
Violation Description - 15
Violation Memo - 15
Violation Detail - 16
Violation Points - 17
Violation Description - 18
Violation Memo - 18
Violation Detail - 19
Violation Points - 20
Violation Description - 21
Violation Memo - 21
Violation Detail - 22
Violation Points - 23
Violation Description - 24
Violation Memo - 24
Violation Detail - 25
Inspection ID
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

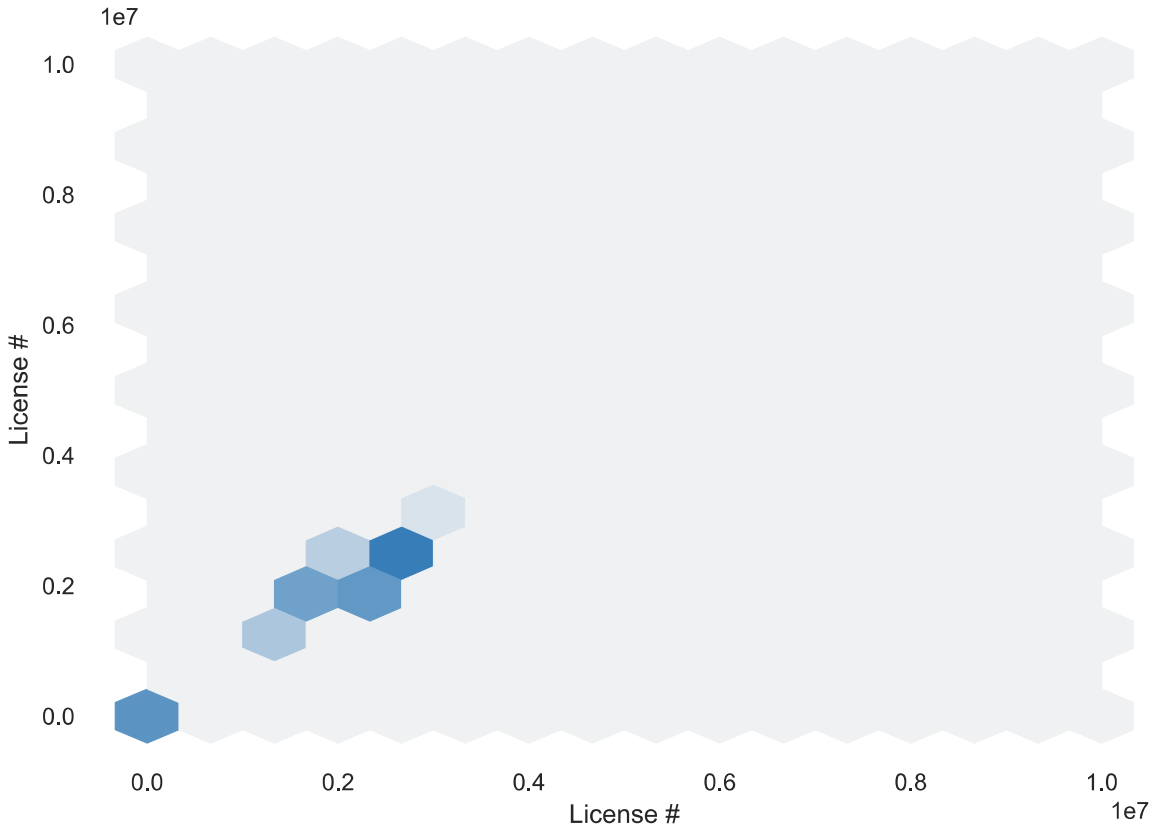

1e7
1.0
0.8
0.6
0.4
0.2
0.0
License #
0.0
0.5
1.0
1.5
2.0
2.5
1e6
Inspection ID

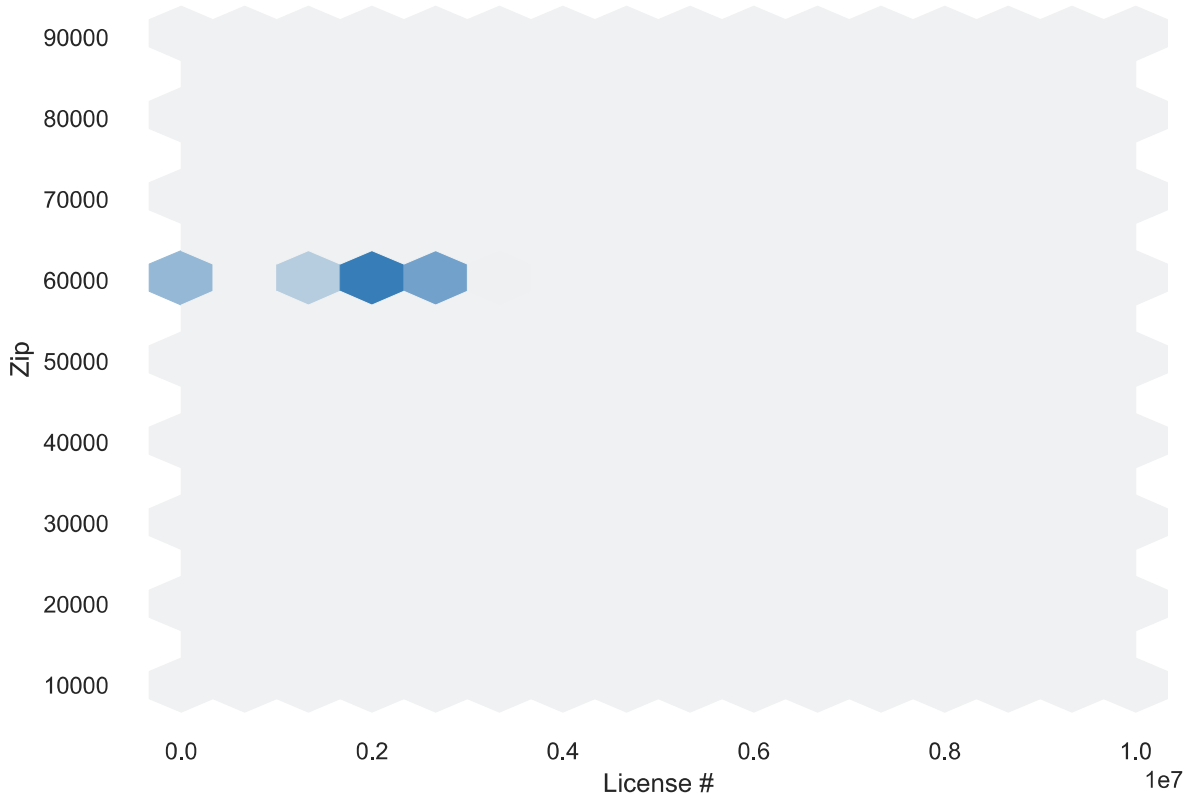

90000
80000
70000
60000
50000
40000
30000
20000
10000
Zip
0.0
0.5
1.0
1.5
2.0
2.5
1e6
Inspection ID

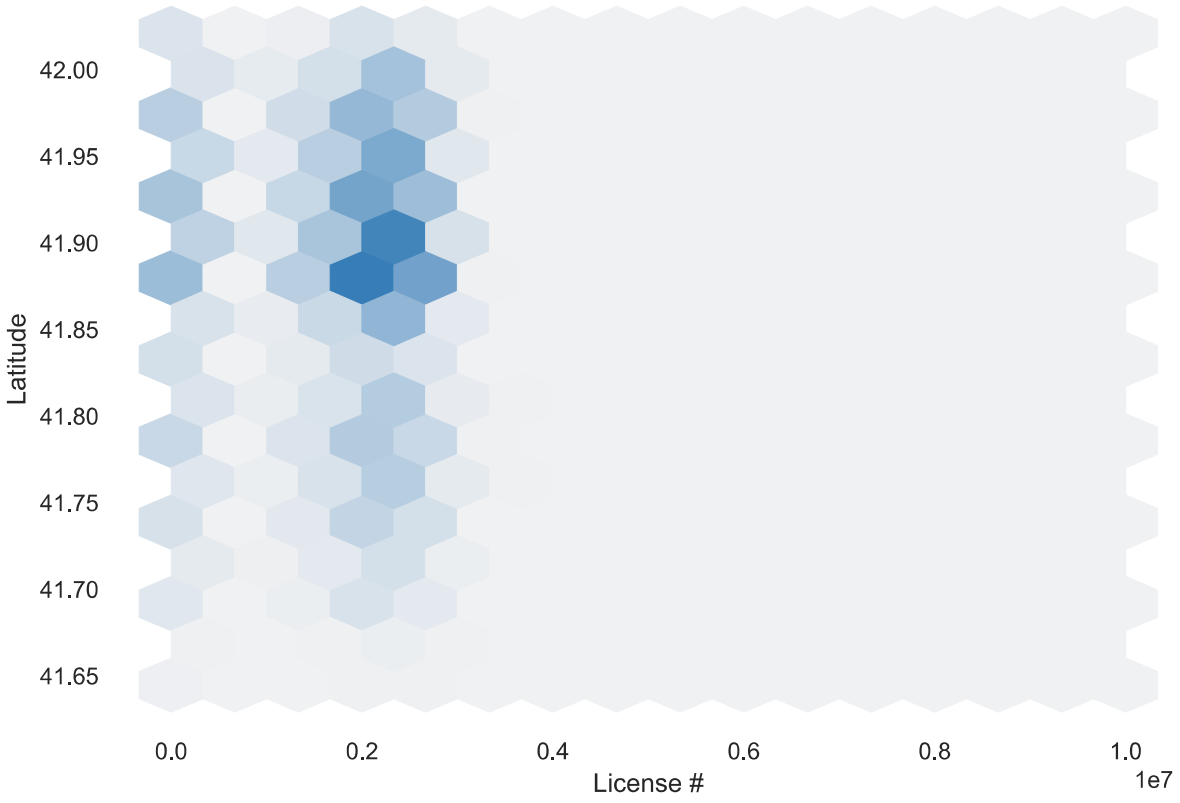

42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
Latitude
0.0
0.5
1.0
1.5
2.0
2.5
1e6
Inspection ID

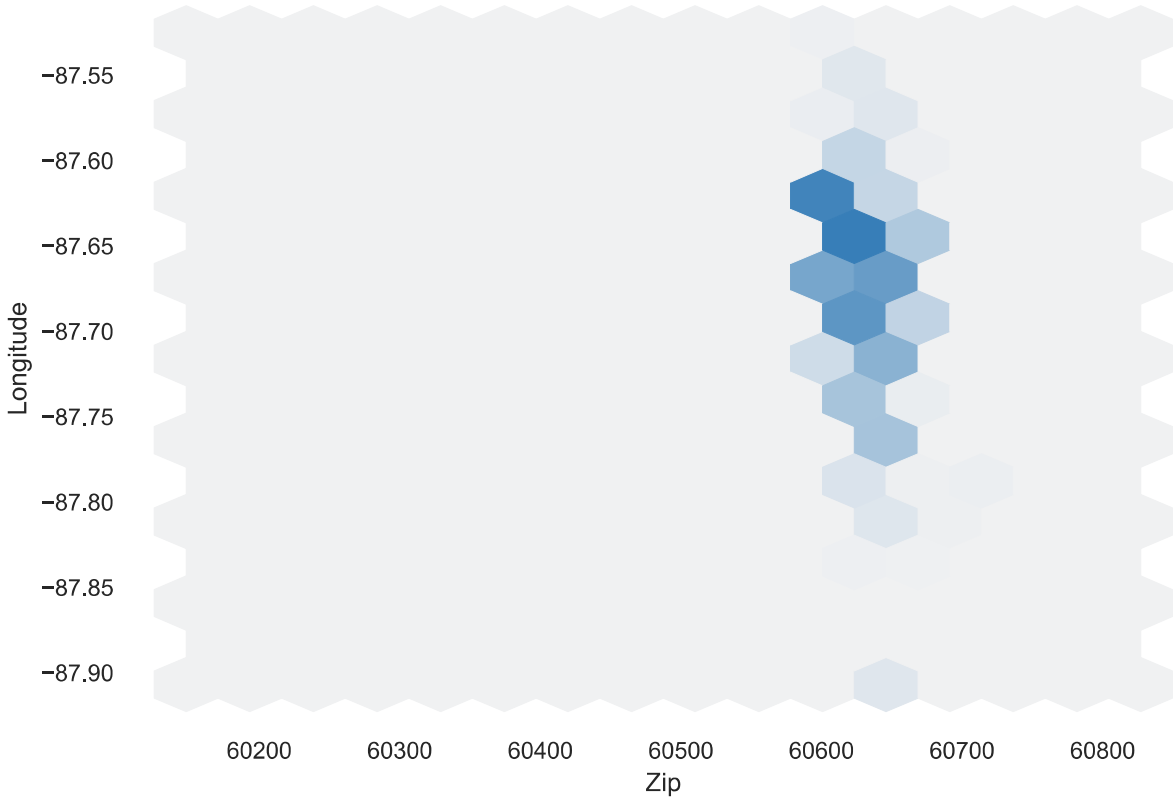

−87.55
−87.60
−87.65
−87.70
−87.75
−87.80
−87.85
−87.90
Longitude
0.0
0.2
0.4
0.6
0.8
1.0
1e7
License #

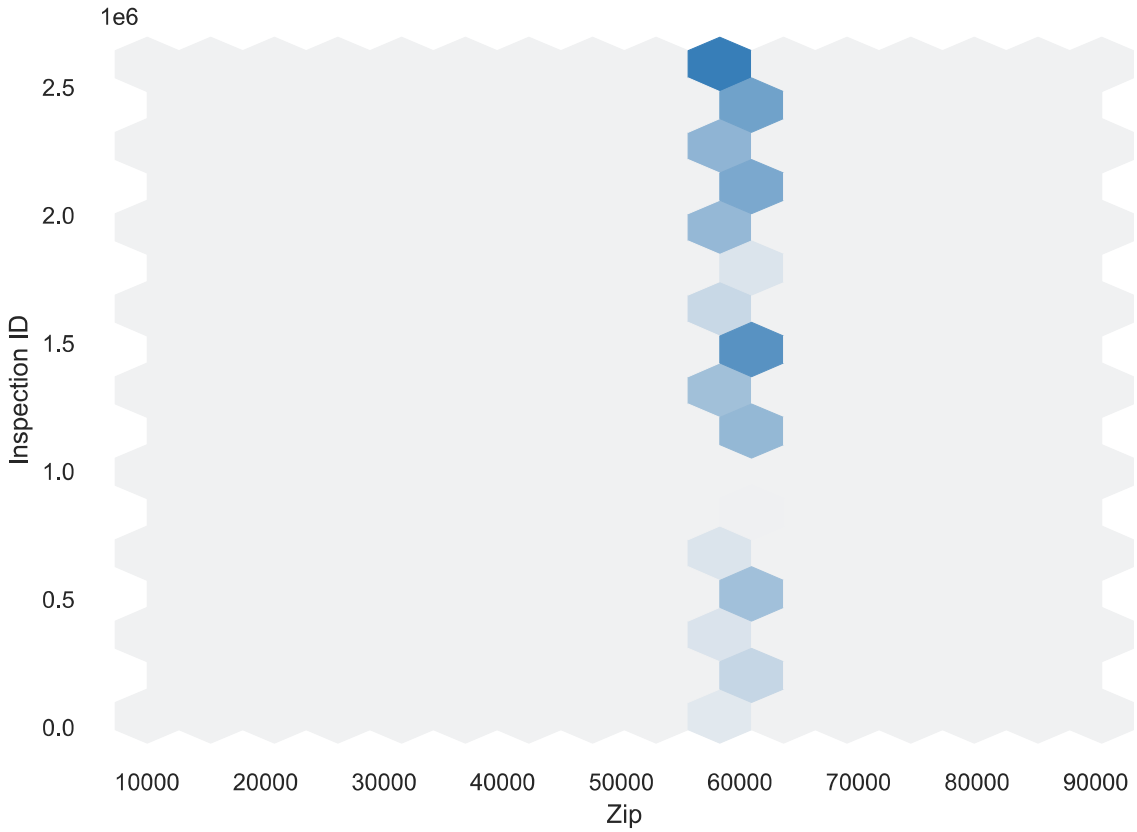

1e6
2.5
2.0
1.5
1.0
0.5
0.0
Inspection ID
0.0
0.2
0.4
0.6
0.8
1.0
1e7
License #

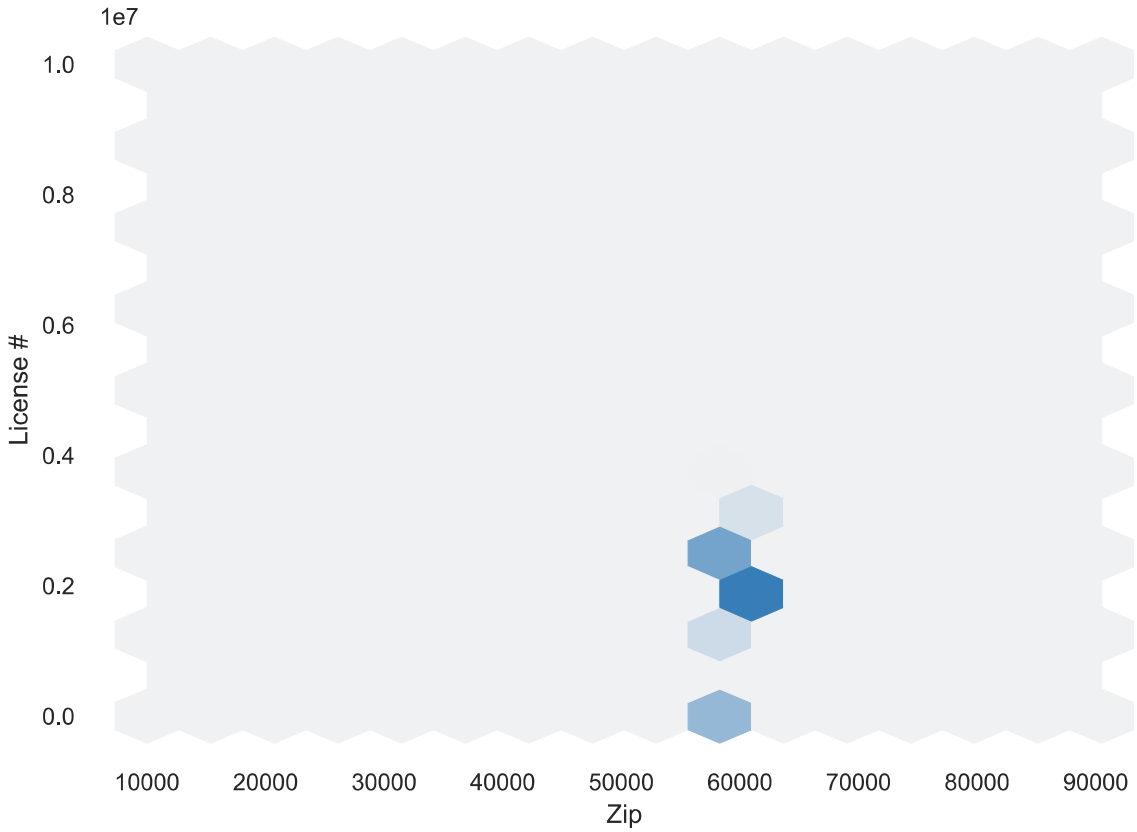

1e7
1.0
0.8
0.6
0.4
0.2
0.0
License #
0.0
0.2
0.4
0.6
0.8
1.0
1e7
License #

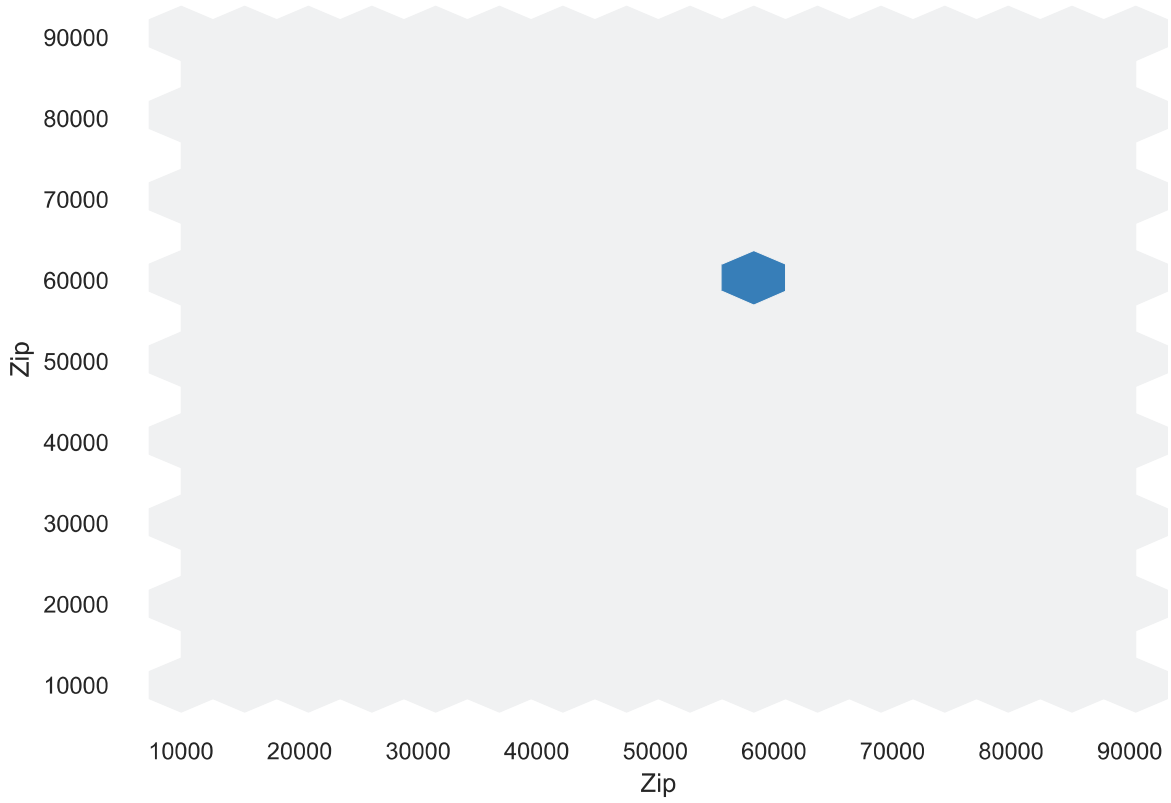

90000
80000
70000
60000
50000
40000
30000
20000
10000
Zip
0.0
0.2
0.4
0.6
0.8
1.0
1e7
License #

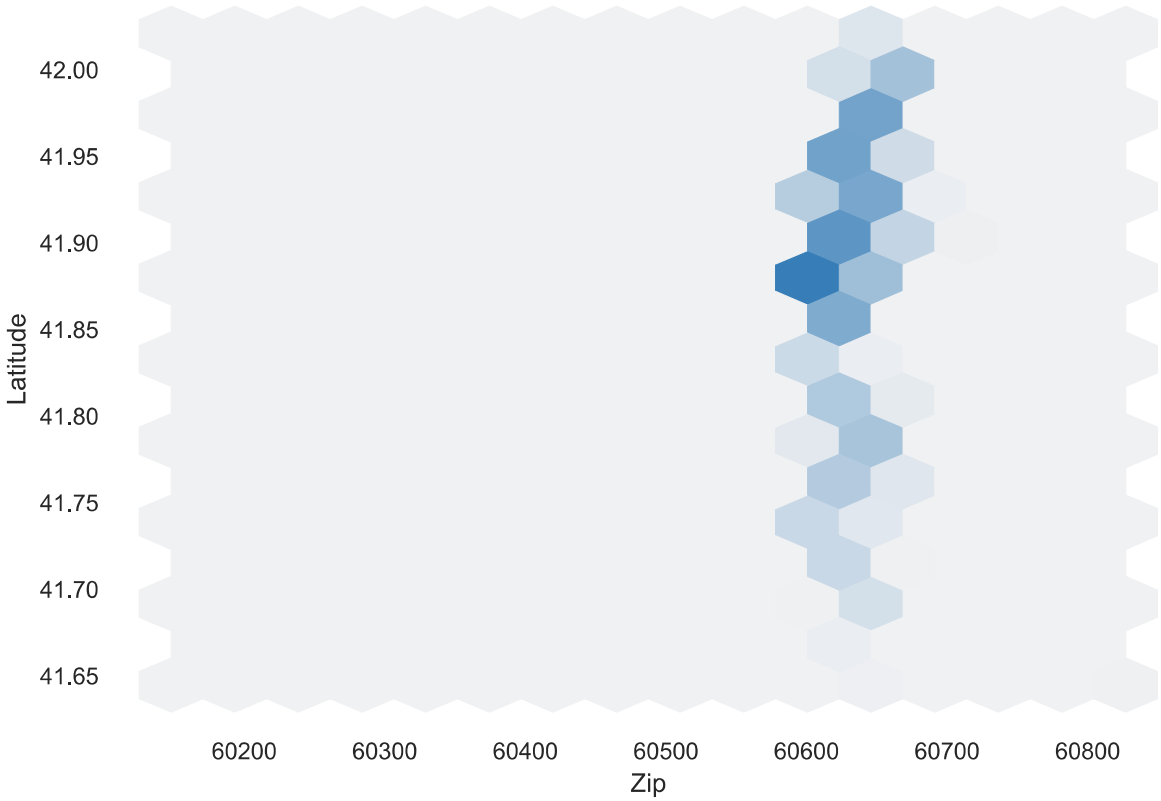

42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
Latitude
0.0
0.2
0.4
0.6
0.8
1.0
1e7
License #

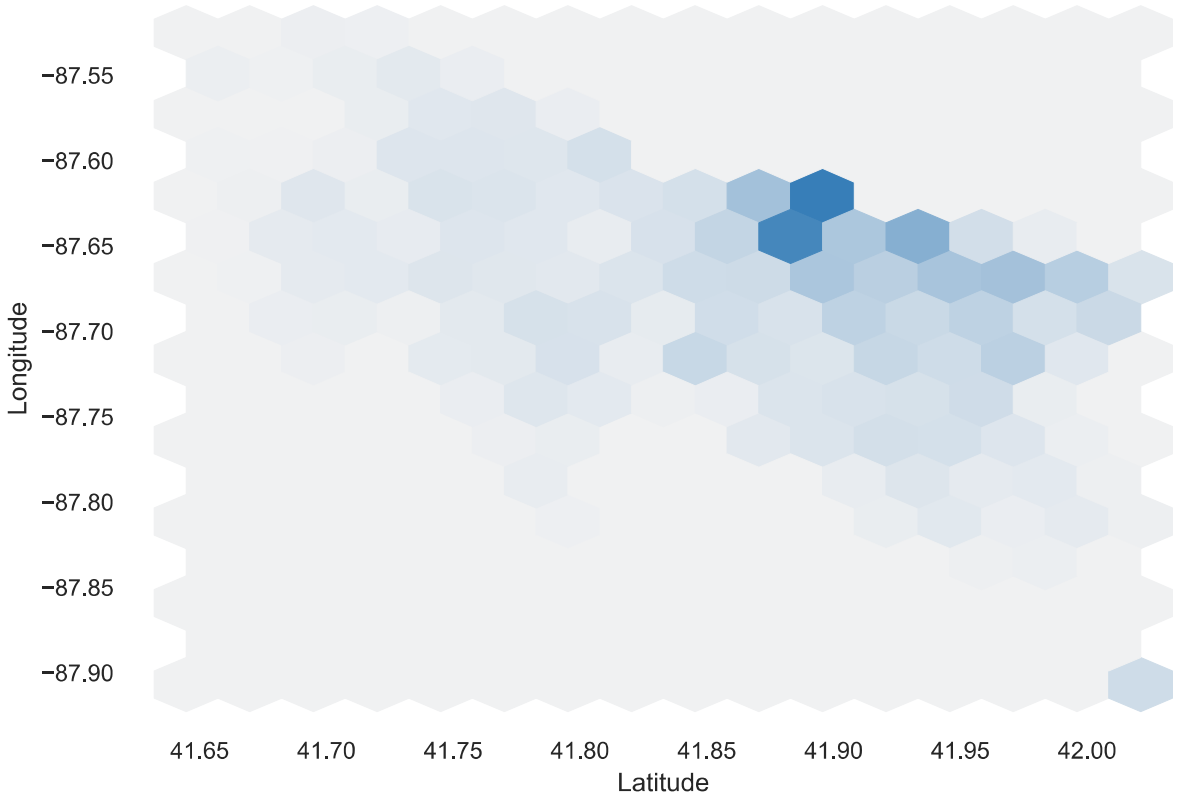

−87.55
−87.60
−87.65
−87.70
−87.75
−87.80
−87.85
−87.90
Longitude
60200
60300
60400
60500
60600
60700
60800
Zip

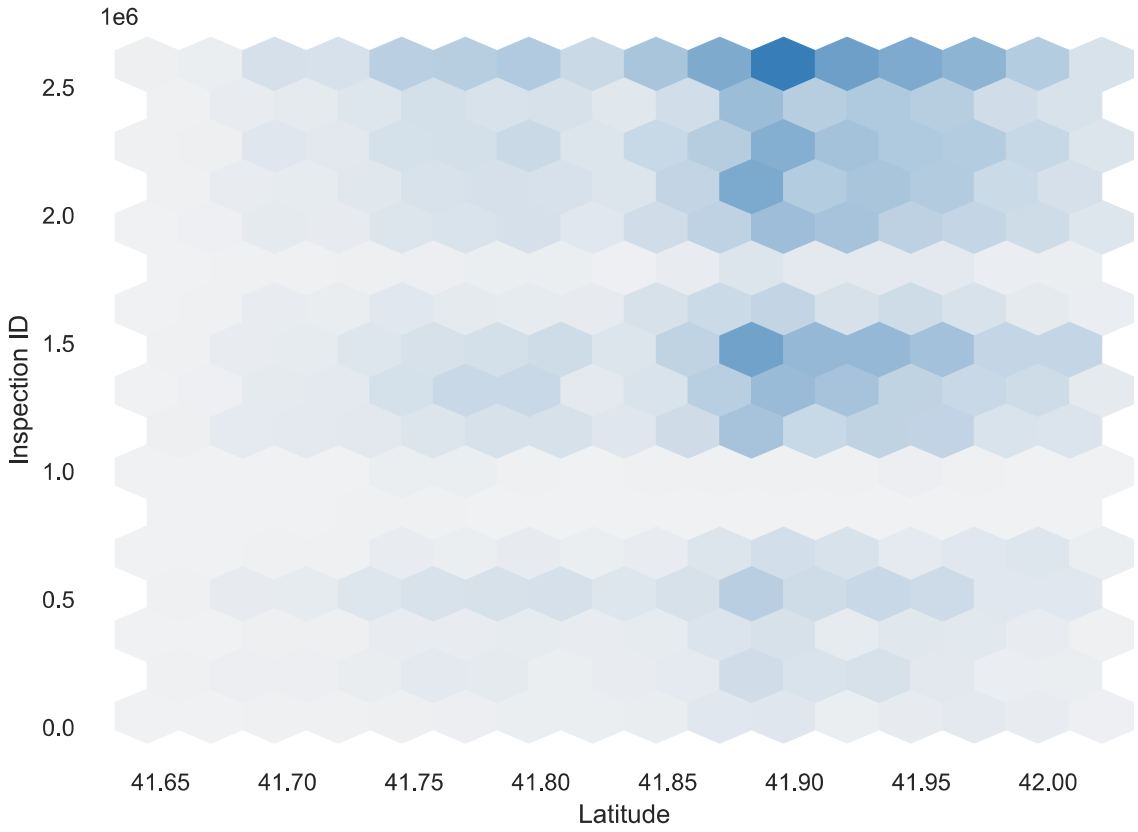

1e6
2.5
2.0
1.5
1.0
0.5
0.0
Inspection ID
10000
20000
30000
40000
50000
60000
70000
80000
90000
Zip

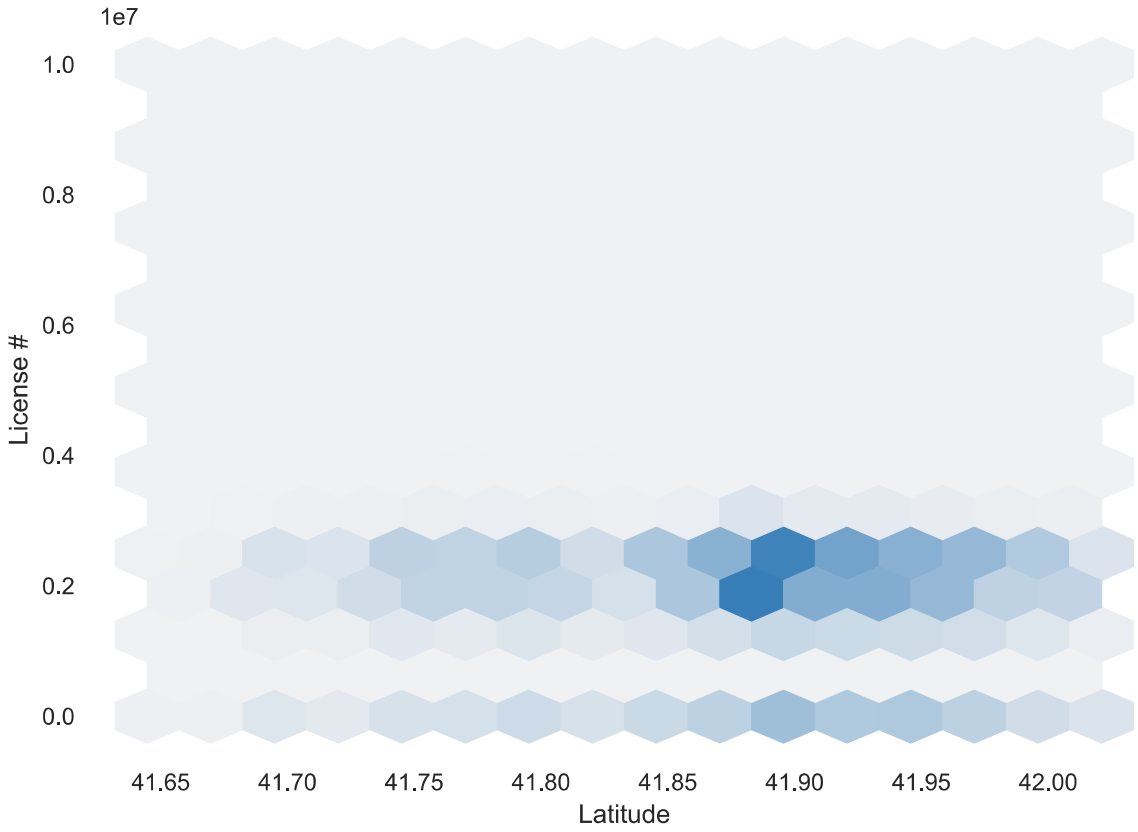

1e7
1.0
0.8
0.6
0.4
0.2
0.0
License #
10000
20000
30000
40000
50000
60000
70000
80000
90000
Zip

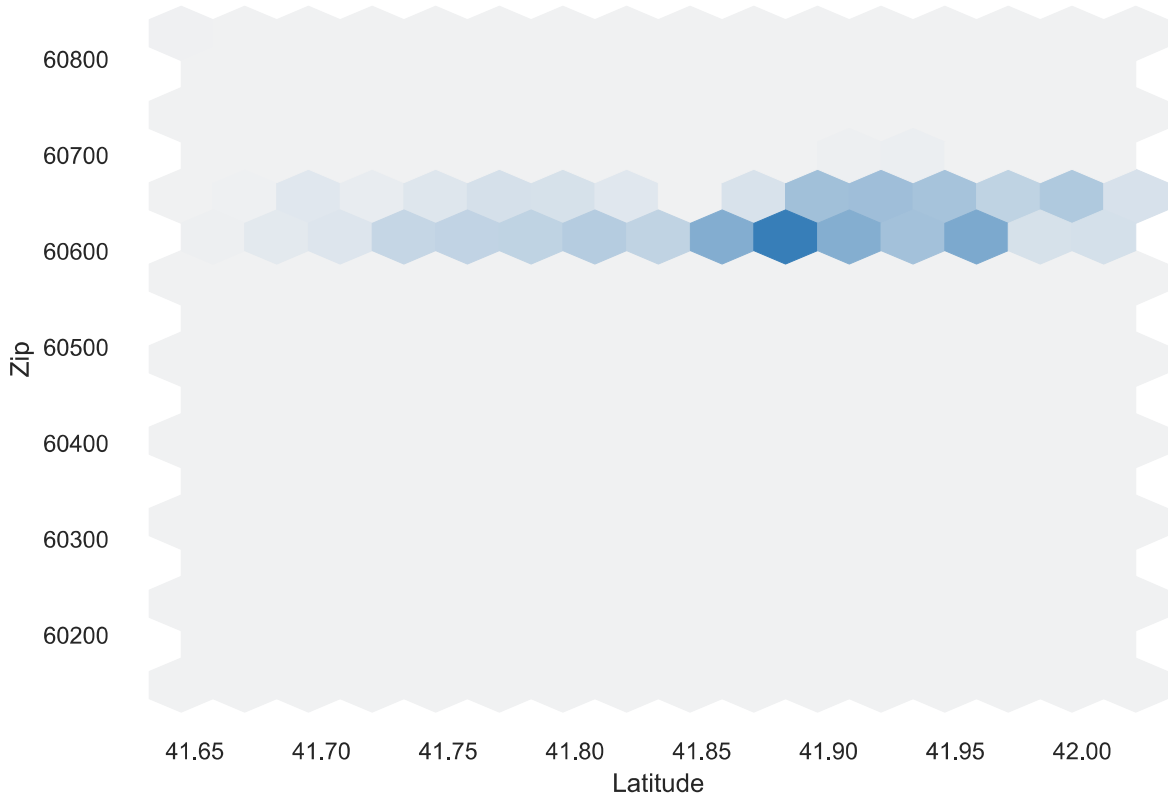

90000
80000
70000
60000
50000
40000
30000
20000
10000
Zip
10000
20000
30000
40000
50000
60000
70000
80000
90000
Zip

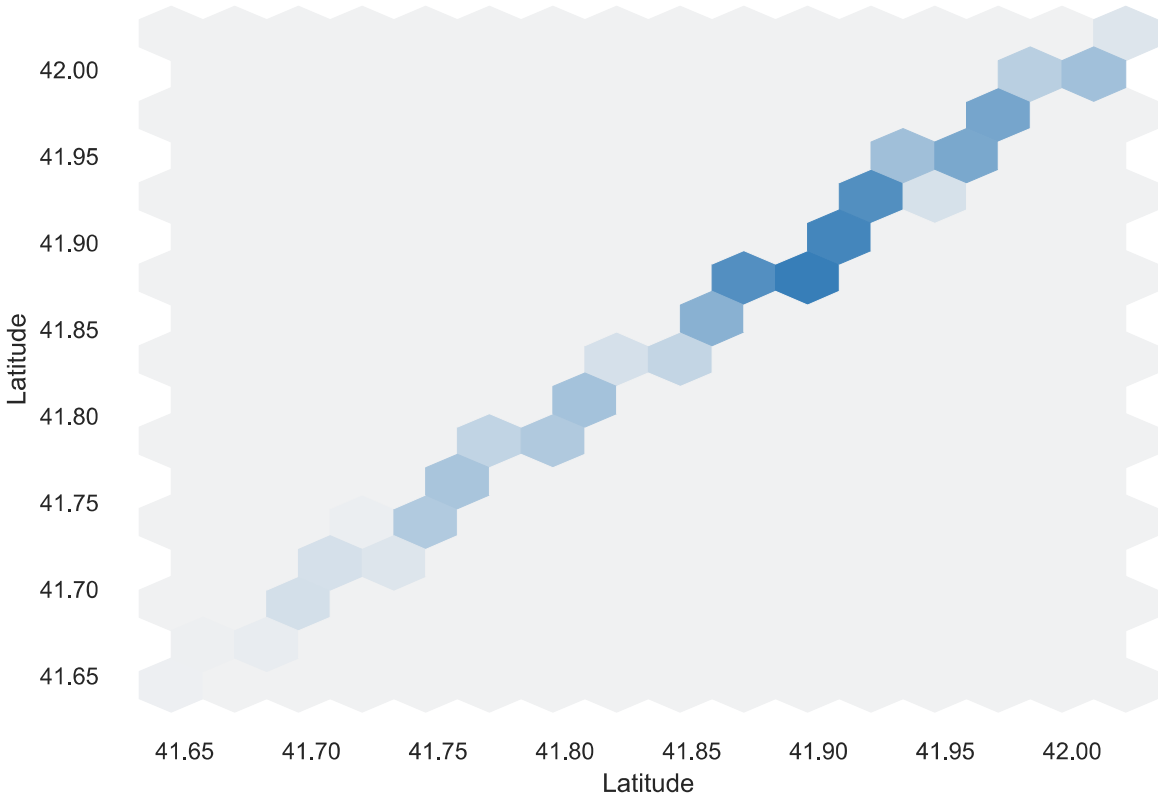

42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
Latitude
60200
60300
60400
60500
60600
60700
60800
Zip

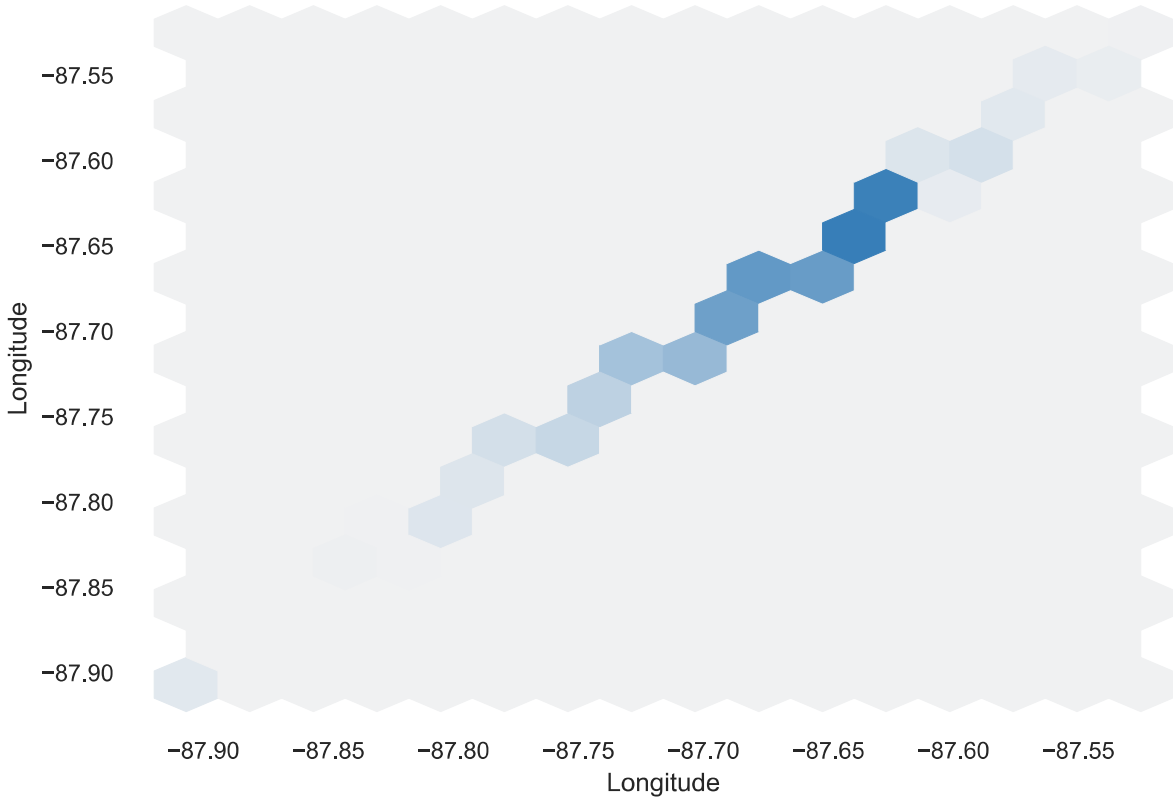

−87.55
−87.60
−87.65
−87.70
−87.75
−87.80
−87.85
−87.90
Longitude
41.65
41.70
41.75
41.80
41.85
41.90
41.95
42.00
Latitude

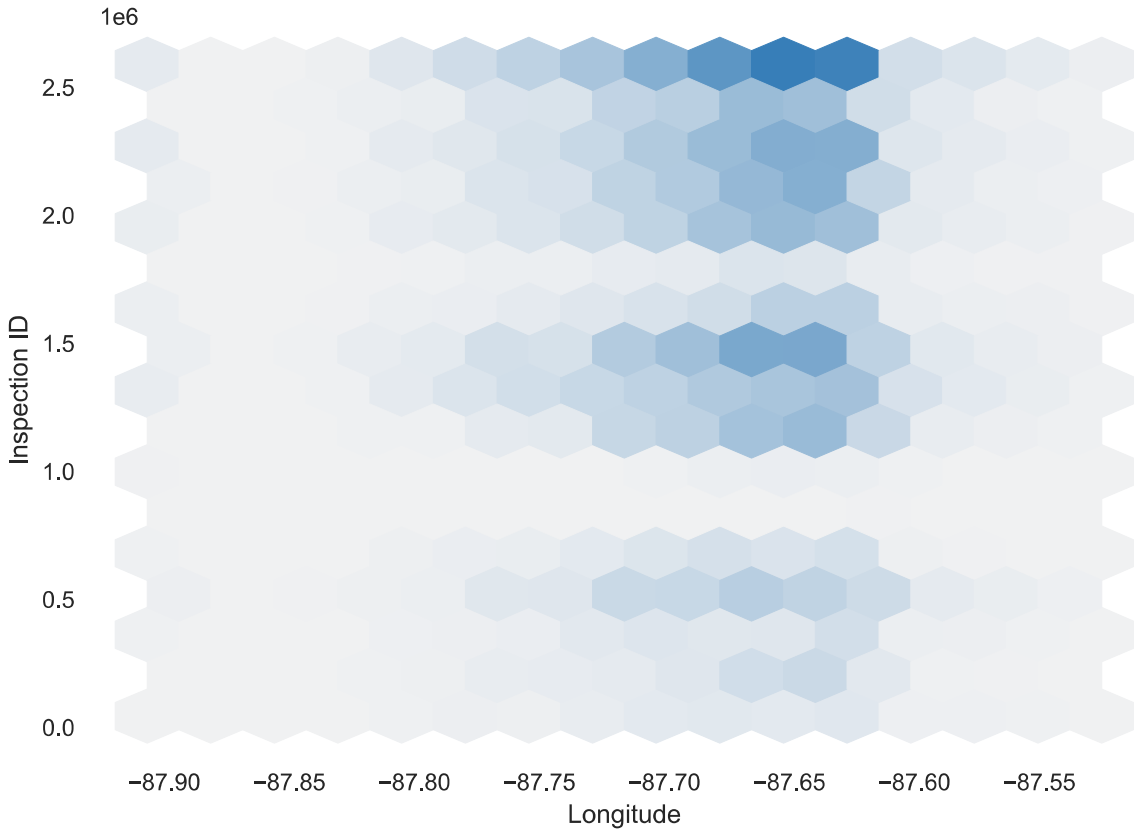

1e6
2.5
2.0
1.5
1.0
0.5
0.0
Inspection ID
41.65
41.70
41.75
41.80
41.85
41.90
41.95
42.00
Latitude

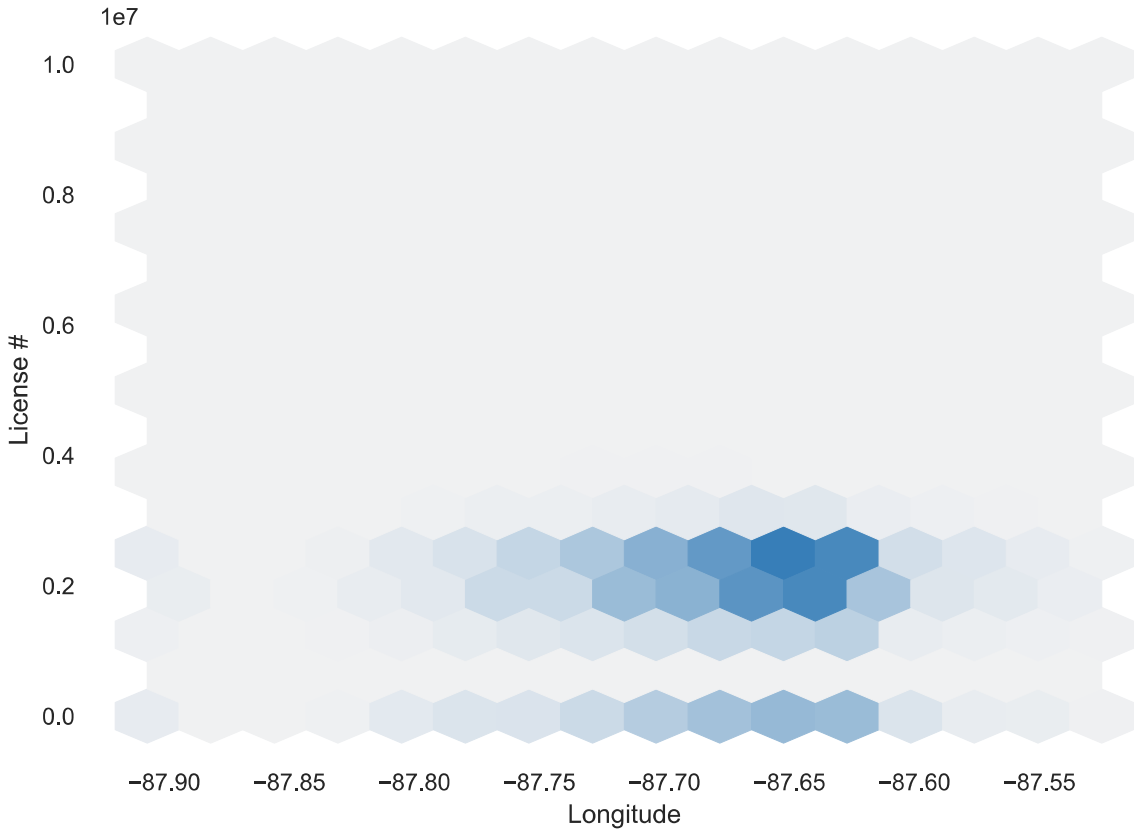

1e7
1.0
0.8
0.6
0.4
0.2
0.0
License #
41.65
41.70
41.75
41.80
41.85
41.90
41.95
42.00
Latitude

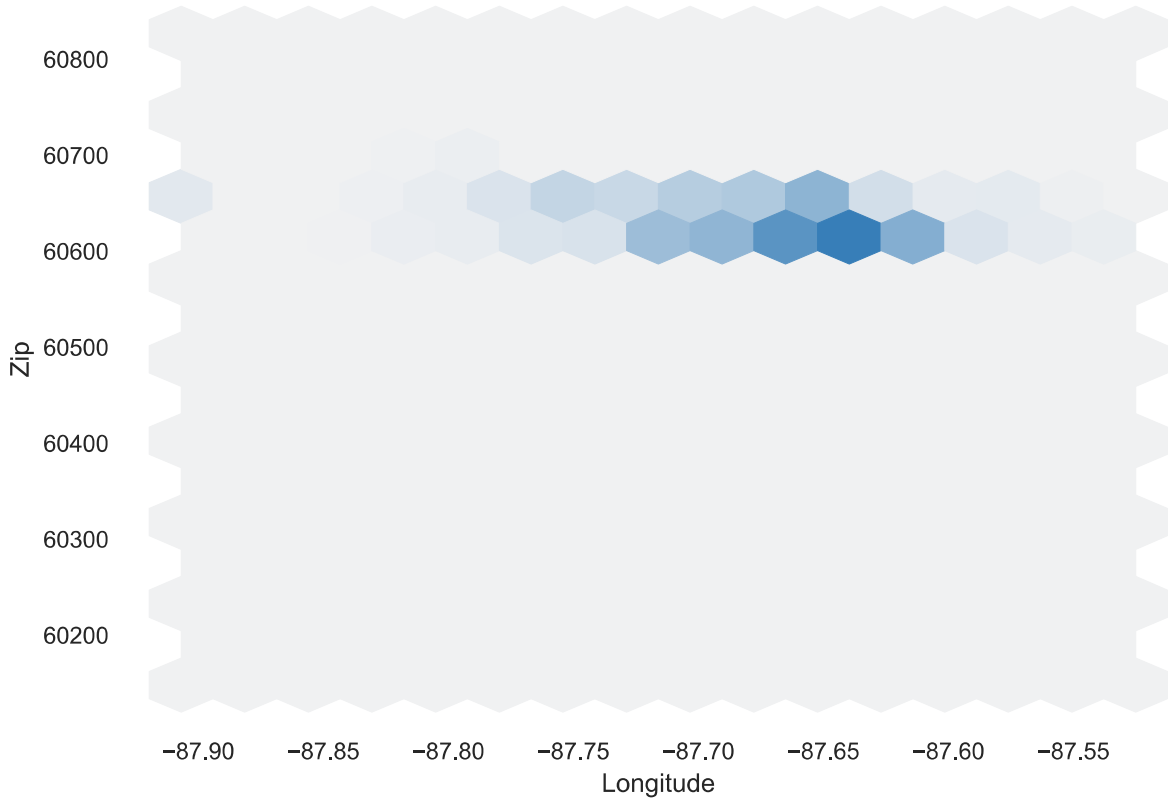

60800
60700
60600
60500
60400
60300
60200
Zip
41.65
41.70
41.75
41.80
41.85
41.90
41.95
42.00
Latitude

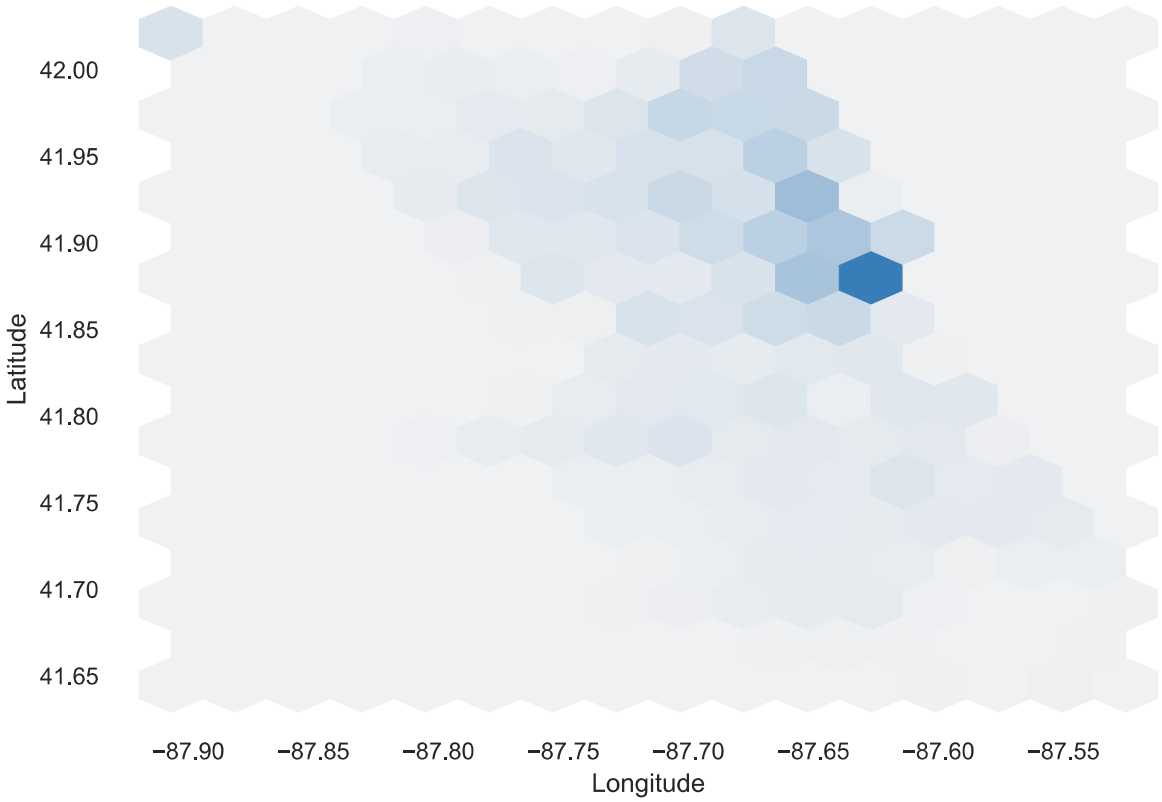

42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
Latitude
41.65
41.70
41.75
41.80
41.85
41.90
41.95
42.00
Latitude


−87.55
−87.60
−87.65
−87.70
−87.75
−87.80
−87.85
−87.90
Longitude
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55
Longitude

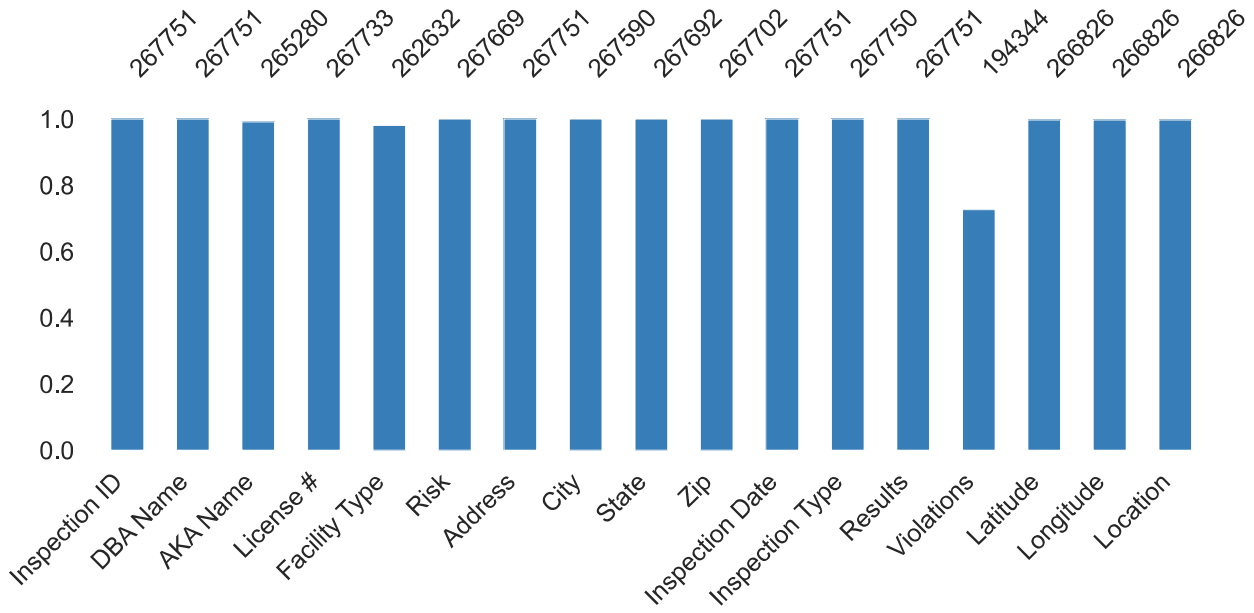

1e6
2.5
2.0
1.5
1.0
0.5
0.0
Inspection ID
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55
Longitude


1e7
1.0
0.8
0.6
0.4
0.2
0.0
License #
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55
Longitude


60800
60700
60600
60500
60400
60300
60200
Zip
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55
Longitude


42.00
41.95
41.90
41.85
41.80
41.75
41.70
41.65
Latitude
−87.90
−87.85
−87.80
−87.75
−87.70
−87.65
−87.60
−87.55
Longitude

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.

Sample

	Inspection ID	DBA Name	AKA Name	License #	Facility Type	Risk	Address
0	2589679	HOLY SMOQUES BBQ	HOLY SMOQUES BBQ	2948383.0	Restaurant	Risk 1 (High)	2524 S CALIFOI
1	2589673	SMALL CHEVAL	SMALL CHEVAL	2830796.0	Restaurant	Risk 1 (High)	1307-1309 E 53I
2	2589681	BARTOLI'S PIZZERIA	BARTOLI'S PIZZERIA	2709388.0	Restaurant	Risk 1 (High)	658 N ASHLANI
3	2589674	SEASHELL	SEASHELL	2551804.0	Restaurant	Risk 2 (Medium)	3944 W ROOSE
4	2589661	J & J FISH & CHICKEN	J & J FISH & CHICKEN	1938693.0	Restaurant	Risk 1 (High)	6814 S STONY I
5	2589645	NOORI CHICKEN CHICAGO	NOORI CHICKEN CHICAGO	2952358.0	NaN	All	3223 N CLARK S
6	2589643	CURRY ON FIRE	CURRY ON FIRE	2535924.0	Restaurant	Risk 1 (High)	1030 W TAYLOI
7	2589633	Clemente Comm (H/S)	Clemente Comm (H/S)	51091.0	School	Risk 1 (High)	1147 N Western
8	2589639	CHIPOTLE MEXICAN GRILL #3651	CHIPOTLE	2748958.0	Restaurant	Risk 1 (High)	5533 S HARLEM
9	2589628	BENITO JUAREZ COMMUNITY ACADEMY	JUAREZ (H/S)	2915819.0	School	Risk 1 (High)	1450-1510 W CE

	Inspection ID	DBA Name	AKA Name	License #	Facility Type	Risk
267741	158288	GREAT HUNNAN RESTAURANT	GREAT HUNNAN RESTAURANT	1193152.0	Restaurant	Risk 1
267742	80220	THAI OSCAR INC.	THAI OSCAR INC.	1873390.0	Restaurant	Risk 1
267743	197242	SOPRAFFINA	SOPRAFFINA	35095.0	Restaurant	Risk 1
267744	193209	APPLEBEE'S NEIGHBORHOOD GRILL & BAR	APPLEBEE'S NEIGHBORHOOD GRILL & BAR	1477462.0	Restaurant	Risk 1
267745	68128	DINER GRILL	DINER GRILL	57105.0	Restaurant	Risk 1
267746	164285	BARNARD	BARNARD	22131.0	School	Risk 1
267747	124314	Dany's Mini Mart	Danny's Mini Mart	2016778.0	Grocery Store	Risk 3
267748	98337	TAQUERIA EL RANCHITO	TAQUERIA EL RANCHITO	60184.0	Restaurant	Risk 1
267749	120283	EAST COAST PANTRY	EAST COAST PANTRY	1892217.0	Grocery Store	Risk 2
267750	124281	GAS CENTER	GAS CENTER	1958643.0	Grocery Store	Risk 3

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