

# Overview

## Dataset statistics

<b>Number of variables</b>	36
<b>Number of observations</b>	78033
<b>Missing cells</b>	649732
<b>Missing cells (%)</b>	23.1%
<b>Duplicate rows</b>	0
<b>Duplicate rows (%)</b>	0.0%
<b>Total size in memory</b>	21.4 MiB
<b>Average record size in memory</b>	288.0 B

## Variable types

<b>Numeric</b>	10
<b>Categorical</b>	24
<b>Unsupported</b>	2

## Alerts

Closed has constant value "0"	Constant
Name has a high cardinality: 22710 distinct values	High cardinality
Address has a high cardinality: 6618 distinct values	High cardinality
StreetName has a high cardinality: 669 distinct values	High cardinality
BldgNo has a high cardinality: 94 distinct values	High cardinality
UnitNo has a high cardinality: 3335 distinct values	High cardinality
PostalCode has a high cardinality: 2902 distinct values	High cardinality
Location has a high cardinality: 56 distinct values	High cardinality
NAICSDescr has a high cardinality: 1041 distinct values	High cardinality
Phone has a high cardinality: 25064 distinct values	High cardinality
Fax has a high cardinality: 15752 distinct values	High cardinality
TollFree has a high cardinality: 4117 distinct values	High cardinality
EMail has a high cardinality: 15058 distinct values	High cardinality
WebAddress has a high cardinality: 14200 distinct values	High cardinality
EmplUpdate has a high cardinality: 433 distinct values	High cardinality
Character has a high cardinality: 56 distinct values	High cardinality
CHArea has a high cardinality: 57 distinct values	High cardinality
Modified has a high cardinality: 189 distinct values	High cardinality
X is highly overall correlated with Y and 1 other fields (Y, Year)	High correlation
Y is highly overall correlated with X and 1 other fields (X, Year)	High correlation
BusinessID is highly overall correlated with FID and 2 other fields (FID, Location, RecordID)	High correlation
Ward is highly overall correlated with CENT_X	High correlation
CENT_X is highly overall correlated with Location and 1 other fields (Location, CENT_Y)	High correlation
CENT_Y is highly overall correlated with Location and 1 other fields (Location, CENT_X)	High correlation
Year is highly overall correlated with X and 3 other fields (X, Y, NAICSCat, EmplRange)	High correlation

RecordID is highly overall correlated with FID and 2 other fields (FID, BusinessID, Location)	High correlation
Character is highly overall correlated with FID and 3 other fields (FID, BldgNo, NAICSCat, PIN)	High correlation
BIA_NAME is highly overall correlated with FID and 2 other fields (FID, CHArea, BIAName)	High correlation
EmplRange is highly overall correlated with NAICSCat and 1 other fields (NAICSCat, Year)	High correlation
CHArea is highly overall correlated with FID and 5 other fields (FID, BldgNo, NAICSCat, PIN, BIA_NAME, BIAName)	High correlation
Sector_Des is highly overall correlated with NAICSCat	High correlation
BIAName is highly overall correlated with FID and 2 other fields (FID, CHArea, BIA_NAME)	High correlation
FID is highly overall correlated with BusinessID and 7 other fields (BusinessID, Location, PIN, Character, CHArea, BIA_NAME, BIAName, RecordID)	High correlation
BldgNo is highly overall correlated with Location and 2 other fields (Location, Character, CHArea)	High correlation
Location is highly overall correlated with FID and 6 other fields (FID, BusinessID, BldgNo, NAICSCat, CENT_X, CENT_Y, RecordID)	High correlation
NAICSCat is highly overall correlated with Location and 5 other fields (Location, EmplRange, Sector_Des, Year, Character, CHArea)	High correlation
PIN is highly overall correlated with FID and 2 other fields (FID, Character, CHArea)	High correlation
X has 48606 (62.3%) missing values	Missing
Y has 48606 (62.3%) missing values	Missing
Location has 47694 (61.1%) missing values	Missing
EmplRange has 2646 (3.4%) missing values	Missing
EmplUpdate has 15002 (19.2%) missing values	Missing
Sector_Des has 63431 (81.3%) missing values	Missing
CENT_X has 47694 (61.1%) missing values	Missing
CENT_Y has 47694 (61.1%) missing values	Missing
PIN has 30339 (38.9%) missing values	Missing
Character has 61682 (79.0%) missing values	Missing
CHArea has 46690 (59.8%) missing values	Missing
Modified has 63218 (81.0%) missing values	Missing
BIA_NAME has 63208 (81.0%) missing values	Missing
BIAName has 63208 (81.0%) missing values	Missing
StreetNo is highly skewed ( $\gamma_1 = 147.6519659$ )	Skewed
NAICSCode is an unsupported type, check if it needs cleaning or further analysis	Unsupported
isnew is an unsupported type, check if it needs cleaning or further analysis	Unsupported

## Reproduction

# Variables

**Select Columns** 

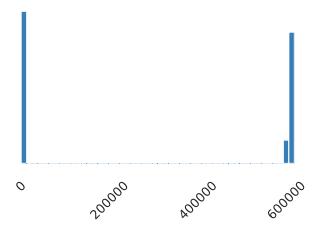
**X**  
Real number ( $\mathbb{R}$ )

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

MISSING

<b>Distinct</b>	8684
<b>Distinct (%)</b>	29.5%
<b>Missing</b>	48606
<b>Missing (%)</b>	62.3%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	306553.47

<b>Minimum</b>	-79.80298
<b>Maximum</b>	617060.11
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	14602
<b>Negative (%)</b>	18.7%
<b>Memory size</b>	609.8 KiB

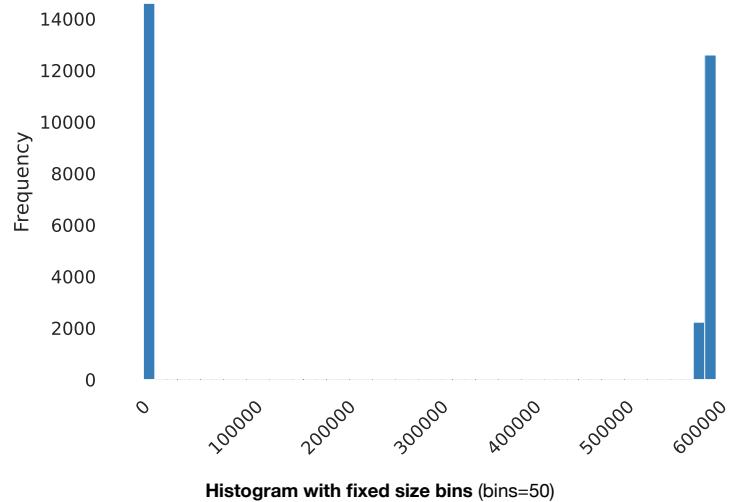


### Quantile statistics

<b>Minimum</b>	-79.80298
<b>5-th percentile</b>	-79.716419
<b>Q1</b>	-79.64992
<b>median</b>	598535.65
<b>Q3</b>	608829.52
<b>95-th percentile</b>	613567.3
<b>Maximum</b>	617060.11
<b>Range</b>	617139.91
<b>Interquartile range (IQR)</b>	608909.17

### Descriptive statistics

<b>Standard deviation</b>	304335.28
<b>Coefficient of variation (CV)</b>	0.99276409
<b>Kurtosis</b>	-1.9996012
<b>Mean</b>	306553.47
<b>Median Absolute Deviation (MAD)</b>	17202.025
<b>Skewness</b>	-0.014922506
<b>Sum</b>	$9.0209489 \times 10^9$
<b>Variance</b>	$9.261996 \times 10^{10}$
<b>Monotonicity</b>	Not monotonic



Value	Count	Frequency (%)
609566.1112	201	0.3%
-79.64275968	185	0.2%
-79.60364656	123	0.2%
607701.737	121	0.2%
-79.71222857	113	0.1%
-79.63864759	107	0.1%
604057.4854	101	0.1%
609718.3353	100	0.1%
-79.56936408	91	0.1%

615498.4771	66	0.1%
Other values (8674)	28219	36.2%
(Missing)	48606	62.3%

Value	Count	Frequency (%)
-79.80298035	1	< 0.1%
-79.8014612	1	< 0.1%
-79.79447393	1	< 0.1%
-79.79439767	1	< 0.1%
-79.78884298	1	< 0.1%
-79.78871792	20	< 0.1%
-79.78850259	1	< 0.1%
-79.78675536	5	< 0.1%
-79.78630211	12	< 0.1%
-79.78452433	11	< 0.1%

Value	Count	Frequency (%)
617060.1055	1	< 0.1%
616918.4738	1	< 0.1%
616839.6893	1	< 0.1%
616837.5953	1	< 0.1%
616769.3441	1	< 0.1%
616704.5391	1	< 0.1%
616692.2284	1	< 0.1%
616667.6043	1	< 0.1%
616657.8816	1	< 0.1%
616643.3766	1	< 0.1%

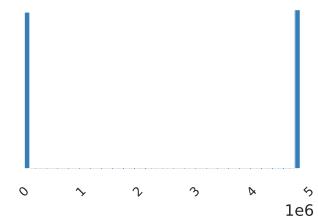
**Y**  
Real number ( $\mathbb{R}$ )

**HIGH CORRELATION** (This variable has a high overall correlation with 2 fields: X, Year)

MISSING

<b>Distinct</b>	8684
<b>Distinct (%)</b>	29.5%
<b>Missing</b>	48606
<b>Missing (%)</b>	62.3%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	2433290.7

<b>Minimum</b>	43.48517
<b>Maximum</b>	4843106.9
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

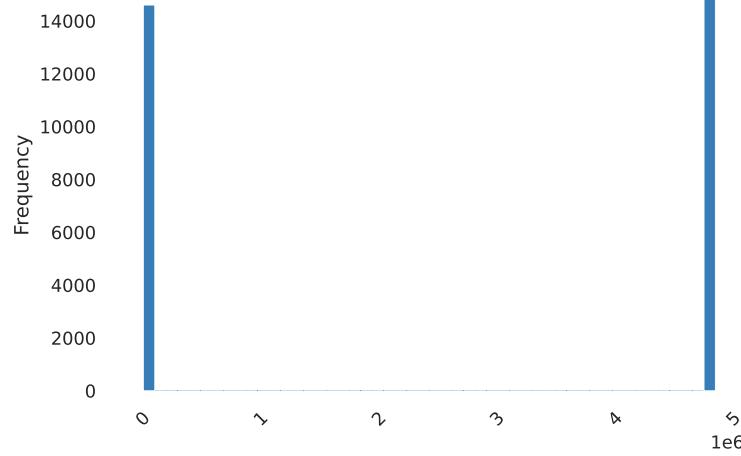


### Quantile statistics

<b>Minimum</b>	43.48517
<b>5-th percentile</b>	43.53859
<b>Q1</b>	43.608514
<b>median</b>	4818092
<b>Q3</b>	4829966.3
<b>95-th percentile</b>	4838021.6
<b>Maximum</b>	4843106.9
<b>Range</b>	4843063.4
<b>Interquartile range (IQR)</b>	4829922.6

### Descriptive statistics

<b>Standard deviation</b>	2414921.5
<b>Coefficient of variation (CV)</b>	0.99245088
<b>Kurtosis</b>	-1.9998953
<b>Mean</b>	2433290.7
<b>Median Absolute Deviation (MAD)</b>	23561.033
<b>Skewness</b>	-0.015148997
<b>Sum</b>	$7.1604446 \times 10^{10}$
<b>Variance</b>	$5.8318459 \times 10^{12}$
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
4827535.97	201	0.3%
43.59351505	185	0.2%
43.67999884	123	0.2%
4838234.833	121	0.2%
43.55837136	113	0.1%
43.72011759	107	0.1%
4823601.861	101	0.1%
4841653.08	100	0.1%

43.5935916	91	0.1%
4827677.175	66	0.1%
Other values (8674)	28219	36.2%
(Missing)	48606	62.3%

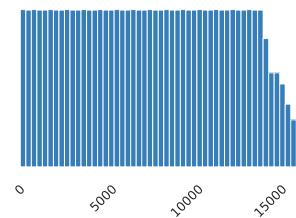
Value	Count	Frequency (%)
43.48517014	1	< 0.1%
43.48968489	1	< 0.1%
43.4915708	1	< 0.1%
43.49199992	2	< 0.1%
43.49224252	1	< 0.1%
43.49454092	1	< 0.1%
43.49517064	1	< 0.1%
43.49608236	1	< 0.1%
43.49636475	1	< 0.1%
43.49652992	2	< 0.1%

Value	Count	Frequency (%)
4843106.933	3	< 0.1%
4843045.912	1	< 0.1%
4842995.781	2	< 0.1%
4842852.901	1	< 0.1%
4842722.486	1	< 0.1%
4842531.982	2	< 0.1%
4842304.058	2	< 0.1%
4842274.717	1	< 0.1%
4842274.399	2	< 0.1%
4842200.556	2	< 0.1%

**FID**  
Real number ( $\mathbb{R}$ )

<b>Distinct</b>	16518
<b>Distinct (%)</b>	21.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	7823.163

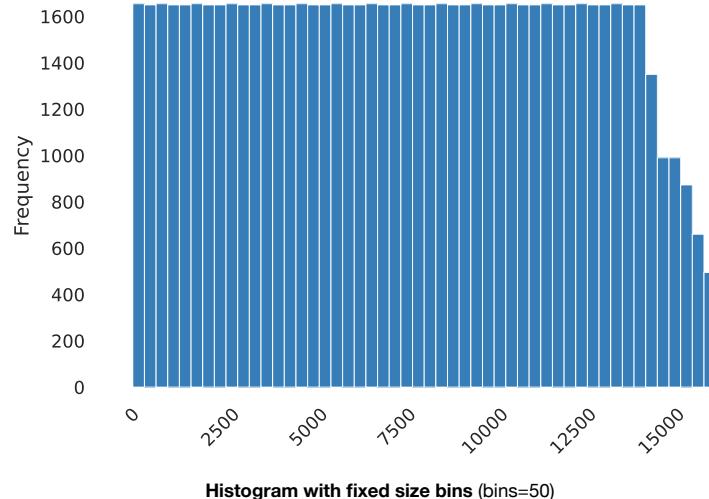
<b>Minimum</b>	1
<b>Maximum</b>	16518
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	1
<b>5-th percentile</b>	781
<b>Q1</b>	3902
<b>median</b>	7804
<b>Q3</b>	11705
<b>95-th percentile</b>	14902
<b>Maximum</b>	16518
<b>Range</b>	16517
<b>Interquartile range (IQR)</b>	7803

**Descriptive statistics**

<b>Standard deviation</b>	4538.4885
<b>Coefficient of variation (CV)</b>	0.58013472
<b>Kurtosis</b>	-1.1665313
<b>Mean</b>	7823.163
<b>Median Absolute Deviation (MAD)</b>	3902
<b>Skewness</b>	0.024778868
<b>Sum</b>	$6.1046488 \times 10^8$
<b>Variance</b>	20597878
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1	5	< 0.1%
9727	5	< 0.1%
9729	5	< 0.1%
9730	5	< 0.1%
9731	5	< 0.1%
9732	5	< 0.1%
9733	5	< 0.1%
9734	5	< 0.1%

9735	5	< 0.1%
9736	5	< 0.1%
Other values (16508)	77983	99.9%

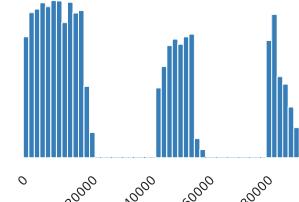
Value	Count	Frequency (%)
1	5	< 0.1%
2	5	< 0.1%
3	5	< 0.1%
4	5	< 0.1%
5	5	< 0.1%
6	5	< 0.1%
7	5	< 0.1%
8	5	< 0.1%
9	5	< 0.1%
10	5	< 0.1%

Value	Count	Frequency (%)
16518	1	< 0.1%
16517	1	< 0.1%
16516	1	< 0.1%
16515	1	< 0.1%
16514	1	< 0.1%
16513	1	< 0.1%
16512	1	< 0.1%
16511	1	< 0.1%
16510	1	< 0.1%
16509	1	< 0.1%

**BusinessID**  
Real number ( $\mathbb{R}$ )

<b>Distinct</b>	21240
<b>Distinct (%)</b>	27.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	34656.92

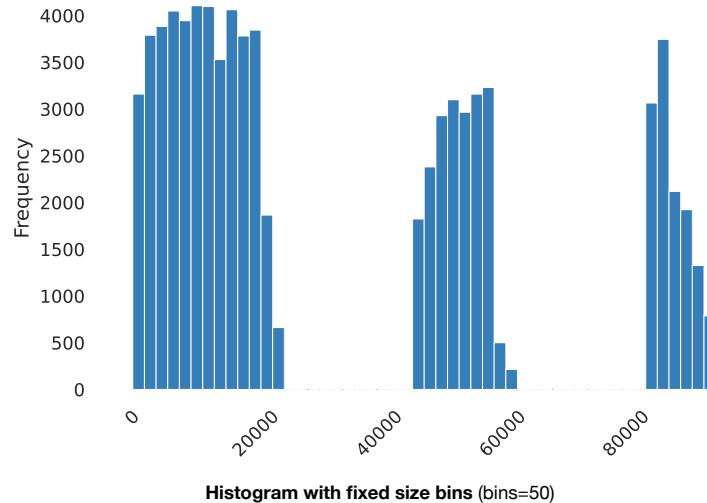
<b>Minimum</b>	2
<b>Maximum</b>	94424
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	2
<b>5-th percentile</b>	2230
<b>Q1</b>	9764
<b>median</b>	19183
<b>Q3</b>	55026
<b>95-th percentile</b>	88915
<b>Maximum</b>	94424
<b>Range</b>	94422
<b>Interquartile range (IQR)</b>	45262

**Descriptive statistics**

<b>Standard deviation</b>	29857.678
<b>Coefficient of variation (CV)</b>	0.8615214
<b>Kurtosis</b>	-0.9937126
<b>Mean</b>	34656.92
<b>Median Absolute Deviation (MAD)</b>	16020
<b>Skewness</b>	0.65053975
<b>Sum</b>	$2.7043834 \times 10^9$
<b>Variance</b>	$8.9148093 \times 10^8$
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
85606	6	< 0.1%
1055	5	< 0.1%
19338	5	< 0.1%
19580	5	< 0.1%
20871	5	< 0.1%
19831	5	< 0.1%
19332	5	< 0.1%
19583	5	< 0.1%

19832	5	< 0.1%
19584	5	< 0.1%
Other values (21230)	77982	99.9%

Value	Count	Frequency (%)
2	2	< 0.1%
7	5	< 0.1%
10	5	< 0.1%
12	3	< 0.1%
16	5	< 0.1%
18	5	< 0.1%
20	5	< 0.1%
21	5	< 0.1%
23	5	< 0.1%
26	4	< 0.1%

Value	Count	Frequency (%)
94424	1	< 0.1%
94423	1	< 0.1%
94419	1	< 0.1%
94371	1	< 0.1%
94321	1	< 0.1%
94319	1	< 0.1%
94318	1	< 0.1%
94317	1	< 0.1%
94313	1	< 0.1%
94293	1	< 0.1%

**Name**  
Categorical

<b>Distinct</b>	22710
<b>Distinct (%)</b>	29.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

Max length	118
Median length	76
Mean length	22.654351
Min length	1

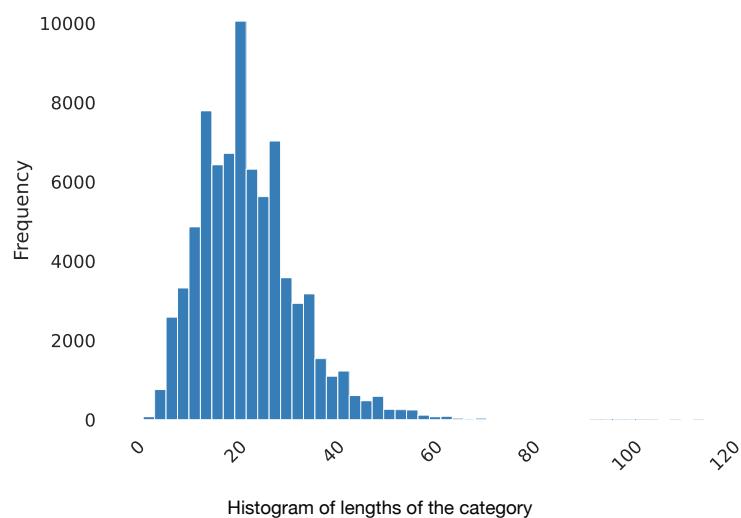
Characters and Unicode		Unique		Sample		
<b>Total characters</b>	1767787	<b>Unique</b>	5009	?	1st row	Golf Trends Inc.
<b>Distinct characters</b>	93	<b>Unique (%)</b>	6.4%		2nd row	Apex Graphics Inc.
<b>Distinct categories</b>	15 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )			?	3rd row	Sands, John & Associates Limited
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )			?	4th row	Printmedia-Tackaberry Times
<b>Distinct blocks</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )			?	5th row	S W R Industries Ltd.

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

**Common Values**

Value	Count	Frequency (%)
Subway	212	0.3%
Tim Hortons	181	0.2%
Petro Canada	123	0.2%
Shoppers Drug Mart	102	0.1%
Tim Horton's	97	0.1%
PLASP Child Care Centre	96	0.1%
Dollarama	92	0.1%
Starbucks	88	0.1%
Shell Canada	84	0.1%
Royal Bank of Canada	78	0.1%
Other values (22700)	76880	98.5%

**Length**



Value	Count	Frequency (%)
inc	15794	5.7%
ltd	9127	3.3%
canada	7946	2.9%
centre	4795	1.7%
and	2969	1.1%
services	2598	0.9%
the	2443	0.9%
a	2359	0.8%
of	2092	0.8%
Other values (16113)	2044	0.7%
	225480	81.2%

### Most occurring characters

Value	Count	Frequency (%)
e	199928	11.3%
a	132590	7.5%
n	128136	7.2%
i	115216	6.5%
r	104250	5.9%
o	101894	5.8%
t	97613	5.5%
s	94807	5.4%
l	77470	4.4%
	62777	3.6%

Other values (83)	653106	36.9%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1236773	70.0%
Uppercase Letter	275471	15.6%
Space Separator	199928	11.3%
Other Punctuation	44369	2.5%
Decimal Number	4222	0.2%
Dash Punctuation	4194	0.2%
Close Punctuation	1272	0.1%
Open Punctuation	1266	0.1%
Math Symbol	178	< 0.1%
Final Punctuation	99	< 0.1%
Other values (5)	15	< 0.1%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
e	132590	10.7%
a	128136	10.4%
n	115216	9.3%
i	104250	8.4%
r	101894	8.2%
o	97613	7.9%
t	94807	7.7%
s	77470	6.3%
l	62777	5.1%
c	60202	4.9%
Other values (20)	261818	21.2%

#### *Uppercase Letter*

Value	Count	Frequency (%)
C	35962	13.1%
S	28667	10.4%
I	23883	8.7%

M	18396	6.7%
L	18129	6.6%
A	17083	6.2%
P	16975	6.2%
T	15559	5.6%
D	13515	4.9%
B	11145	4.0%
Other values (17)	76157	27.6%

*Other Punctuation*

Value	Count	Frequency (%)
.	29522	66.5%
&	7166	16.2%
,	3463	7.8%
'	3108	7.0%
/	898	2.0%
:	88	0.2%
#	35	0.1%
@	29	0.1%
!	26	0.1%
"	16	< 0.1%
Other values (2)	18	< 0.1%

*Decimal Number*

Value	Count	Frequency (%)
1	906	21.5%
2	760	18.0%
0	712	16.9%
4	418	9.9%
3	334	7.9%
9	287	6.8%
8	245	5.8%
7	197	4.7%
5	184	4.4%
6	179	4.2%

*Math Symbol*

Value	Count	Frequency (%)

+	152	85.4%
	25	14.0%
>	1	0.6%

*Close Punctuation*

Value	Count	Frequency (%)
)	1264	99.4%
]	8	0.6%

*Space Separator*

Value	Count	Frequency (%)
	199928	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	4194	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	1266	100.0%

*Final Punctuation*

Value	Count	Frequency (%)
,	99	100.0%

*Control*

Value	Count	Frequency (%)
	6	100.0%

*Connector Punctuation*

Value	Count	Frequency (%)
-	3	100.0%

*Format*

Value	Count	Frequency (%)
	3	100.0%

*Currency Symbol*

Value	Count	Frequency (%)
\$	2	100.0%

*Other Symbol*

Value	Count	Frequency (%)
©	1	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	1512244	85.5%
Common	255543	14.5%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
e	132590	8.8%
a	128136	8.5%
n	115216	7.6%
i	104250	6.9%
r	101894	6.7%
o	97613	6.5%
t	94807	6.3%
s	77470	5.1%
l	62777	4.2%
c	60202	4.0%
Other values (47)	537289	35.5%

*Common*

Value	Count	Frequency (%)
.	199928	78.2%
,	29522	11.6%
&	7166	2.8%
-	4194	1.6%
'	3463	1.4%
(	3108	1.2%
)	1266	0.5%
1	1264	0.5%
/	906	0.4%
	898	0.4%

Other values (26)	3828	1.5%
-------------------	------	------

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	1767609	> 99.9%
Punctuation	102	< 0.1%
None	76	< 0.1%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
e	199928	11.3%
a	132590	7.5%
n	128136	7.2%
i	115216	6.5%
r	104250	5.9%
o	101894	5.8%
t	97613	5.5%
s	94807	5.4%
l	77470	4.4%
Other values (75)	62777	3.6%
	652928	36.9%

*Punctuation*

Value	Count	Frequency (%)
,	99	97.1%
	3	2.9%

*None*

Value	Count	Frequency (%)
é	67	88.2%
ü	4	5.3%
ē	2	2.6%
É	1	1.3%
ä	1	1.3%
©	1	1.3%



**Address**

Categorical

<b>Distinct</b>	6618
<b>Distinct (%)</b>	8.5%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	32
<b>Median length</b>	27
<b>Mean length</b>	16.625543
<b>Min length</b>	5

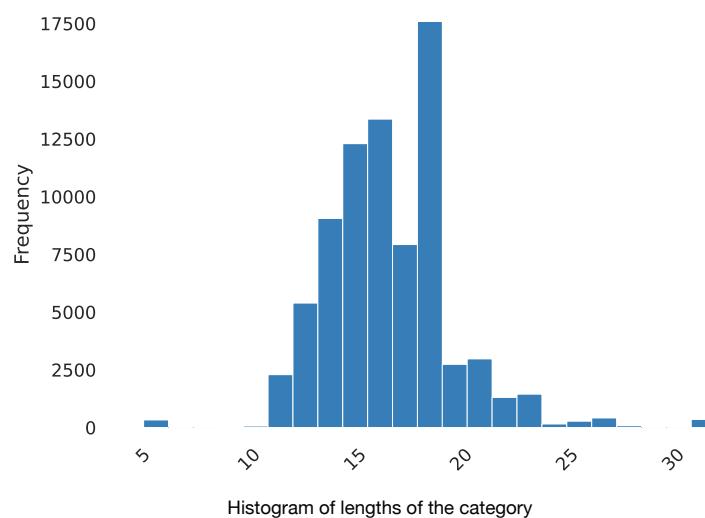
Characters and Unicode		Unique	Sample		
<b>Total characters</b>	1297341	<b>Unique</b>	292	?	<b>1st row</b> 300 Ambassador Dr
<b>Distinct characters</b>	64	<b>Unique (%)</b>	0.4%		<b>2nd row</b> 320 Ambassador Dr
<b>Distinct categories</b>	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				<b>3rd row</b> 320 Ambassador Dr
<b>Distinct scripts</b>	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?	<b>4th row</b> 320 Ambassador Dr
<b>Distinct blocks</b>	1 (https://en.wikipedia.org/wiki/Unicode_block)			?	<b>5th row</b> 321 Ambassador Dr

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

**Common Values**

Value	Count	Frequency (%)
100 City Centre Dr	954	1.2%
5100 Erin Mills Pky	523	0.7%
7205 Goreway Dr	483	0.6%
1250 South Service Rd	394	0.5%
1550 South Gateway Rd	284	0.4%
4141 Dixie Rd	248	0.3%
2225 Erin Mills Pky	238	0.3%
50 Burnhamthorpe Rd W	229	0.3%
2355 Derry Rd E	212	0.3%
2000 Credit Valley Rd	212	0.3%
Other values (6608)	74256	95.2%

**Length**



Value	Count	Frequency (%)
rd	28597	10.8%
dr	17908	6.8%
e	12047	4.6%
st	9954	3.8%
blvd	8013	3.0%
w	7245	2.7%
dundas	4805	1.8%
ave	3977	1.5%
matheson	2625	1.0%
pky	2579	1.0%
Other values (3761)	165839	62.9%

### Most occurring characters

Value	Count	Frequency (%)
r	185559	14.3%
e	77073	5.9%
a	71981	5.5%
d	58783	4.5%
o	55945	4.3%
n	51080	3.9%
s	49723	3.8%
t	48031	3.7%
i	47994	3.7%
l	45040	3.5%

Other values (54)	606132	46.7%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	636955	49.1%
Decimal Number	287143	22.1%
Uppercase Letter	187147	14.4%
Space Separator	185559	14.3%
Dash Punctuation	480	< 0.1%
Other Punctuation	54	< 0.1%
Modifier Symbol	3	< 0.1%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
r	77073	12.1%
e	71981	11.3%
a	58783	9.2%
d	55945	8.8%
n	49723	7.8%
t	47994	7.5%
i	45040	7.1%
o	36413	5.7%
l	32505	5.1%
s	27700	4.3%
Other values (15)	133798	21.0%

#### *Uppercase Letter*

Value	Count	Frequency (%)
R	31751	17.0%
D	29024	15.5%
S	18789	10.0%
E	16442	8.8%
B	14485	7.7%
C	13383	7.2%
W	11748	6.3%

M	9512	5.1%
A	9382	5.0%
T	6499	3.5%
Other values (14)	26132	14.0%

*Decimal Number*

Value	Count	Frequency (%)
0	51080	17.8%
5	48031	16.7%
1	41653	14.5%
2	31311	10.9%
3	25187	8.8%
6	23265	8.1%
7	20531	7.2%
4	17381	6.1%
9	14549	5.1%
8	14155	4.9%

*Other Punctuation*

Value	Count	Frequency (%)
'	46	85.2%
.	8	14.8%

*Space Separator*

Value	Count	Frequency (%)
	185559	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	480	100.0%

*Modifier Symbol*

Value	Count	Frequency (%)
'	3	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	824102	63.5%

Common

473239 36.5%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
r	77073	9.4%
e	71981	8.7%
a	58783	7.1%
d	55945	6.8%
n	49723	6.0%
t	47994	5.8%
i	45040	5.5%
o	36413	4.4%
l	32505	3.9%
R	31751	3.9%
Other values (39)	316894	38.5%

*Common*

Value	Count	Frequency (%)
	185559	39.2%
0	51080	10.8%
5	48031	10.1%
1	41653	8.8%
2	31311	6.6%
3	25187	5.3%
6	23265	4.9%
7	20531	4.3%
4	17381	3.7%
9	14549	3.1%
Other values (5)	14692	3.1%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	1297341	100.0%

**Most frequent character per block**

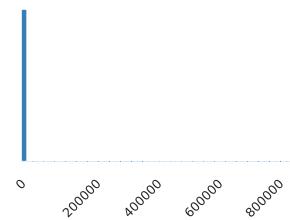
**ASCII**

Value	Count	Frequency (%)
r	185559	14.3%
e	77073	5.9%
a	71981	5.5%
d	58783	4.5%
o	55945	4.3%
n	51080	3.9%
s	49723	3.8%
t	48031	3.7%
i	47994	3.7%
Other values (54)	45040	3.5%
	606132	46.7%

**StreetNo**  
Real number (R)

<b>Distinct</b>	3090
<b>Distinct (%)</b>	4.0%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	2946.096

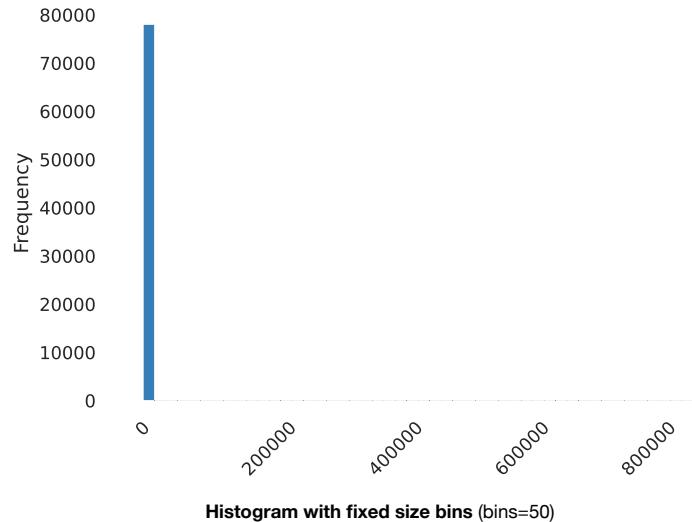
<b>Minimum</b>	1
<b>Maximum</b>	905629
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	1
<b>5-th percentile</b>	57
<b>Q1</b>	1050
<b>median</b>	2375
<b>Q3</b>	5100
<b>95-th percentile</b>	7070
<b>Maximum</b>	905629
<b>Range</b>	905628
<b>Interquartile range (IQR)</b>	4050

**Descriptive statistics**

<b>Standard deviation</b>	3997.6535
<b>Coefficient of variation (CV)</b>	1.3569325
<b>Kurtosis</b>	33315.386
<b>Mean</b>	2946.096
<b>Median Absolute Deviation (MAD)</b>	1655
<b>Skewness</b>	147.65197
<b>Sum</b>	$2.2989271 \times 10^8$
<b>Variance</b>	15981234
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
100	1102	1.4%
5100	601	0.8%
7205	520	0.7%
1250	448	0.6%
1	442	0.6%
2000	383	0.5%
1550	359	0.5%
50	313	0.4%

4141		310	0.4%
2425		304	0.4%
Other values (3080)		73251	93.9%

Value	Count	Frequency (%)
1	442	0.6%
2	198	0.3%
3	200	0.3%
4	154	0.2%
5	7	< 0.1%
6	33	< 0.1%
7	25	< 0.1%
8	21	< 0.1%
9	20	< 0.1%
10	154	0.2%

Value	Count	Frequency (%)
905629	1	< 0.1%
7895	138	0.2%
7890	7	< 0.1%
7885	79	0.1%
7880	6	< 0.1%
7875	30	< 0.1%
7860	5	< 0.1%
7855	5	< 0.1%
7850	4	< 0.1%
7840	1	< 0.1%

**StreetName**

Categorical

<b>Distinct</b>	669
<b>Distinct (%)</b>	0.9%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	26
<b>Median length</b>	22
<b>Mean length</b>	11.945062
<b>Min length</b>	3

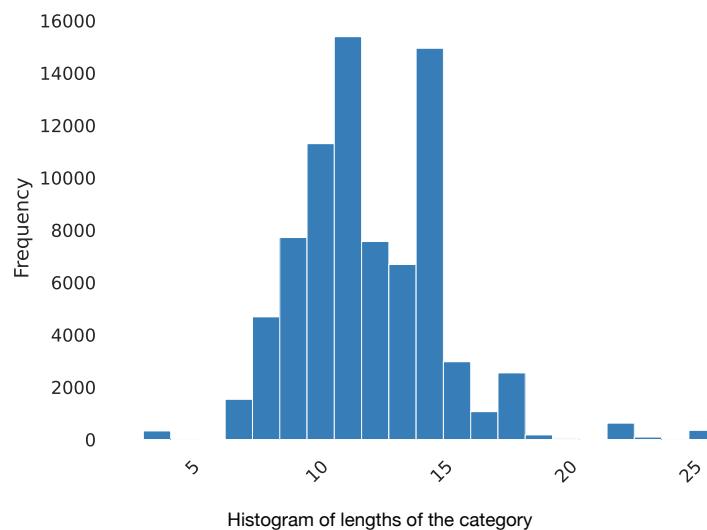
<b>Characters and Unicode</b>		<b>Unique</b>	<b>Sample</b>		
<b>Total characters</b>	932109	<b>Unique</b>	57	?	<b>1st row</b> Ambassador Dr
<b>Distinct characters</b>	53	<b>Unique (%)</b>	0.1%		<b>2nd row</b> Ambassador Dr
<b>Distinct categories</b>	5				<b>3rd row</b> Ambassador Dr
	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )				<b>4th row</b> Ambassador Dr
<b>Distinct scripts</b>	2			?	( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
<b>Distinct blocks</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )		?	Ambassador Dr

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

**Common Values**

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
Dundas St E	3202	4.1%
Matheson Blvd E	2125	2.7%
Dixie Rd	1982	2.5%
Hurontario St	1971	2.5%
Lakeshore Rd E	1628	2.1%
Dundas St W	1586	2.0%
City Centre Dr	1529	2.0%
Britannia Rd E	1441	1.8%
Tomken Rd	1416	1.8%
Argentia Rd	1400	1.8%
Other values (659)	59753	76.6%

**Length**



Value	Count	Frequency (%)
rd	28598	15.4%
dr	17907	9.7%
e	12045	6.5%
st	9954	5.4%
blvd	8011	4.3%
w	7247	3.9%
dundas	4805	2.6%
ave	3978	2.1%
matheson	2625	1.4%
pky	2575	1.4%
Other values (665)	87804	47.3%

### Most occurring characters

Value	Count	Frequency (%)
r	107517	11.5%
e	77033	8.3%
a	71982	7.7%
d	58785	6.3%
n	55948	6.0%
t	49726	5.3%
i	47988	5.1%
o	45032	4.8%
l	36410	3.9%
	32503	3.5%

Other values (43)	349185	37.5%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	636932	68.3%
Uppercase Letter	187129	20.1%
Space Separator	107517	11.5%
Dash Punctuation	480	0.1%
Other Punctuation	51	< 0.1%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
r	77033	12.1%
e	71982	11.3%
a	58785	9.2%
d	55948	8.8%
n	49726	7.8%
t	47988	7.5%
i	45032	7.1%
o	36410	5.7%
l	32503	5.1%
s	27702	4.3%
Other values (15)	133823	21.0%

#### *Uppercase Letter*

Value	Count	Frequency (%)
R	31747	17.0%
D	29018	15.5%
S	18788	10.0%
E	16439	8.8%
B	14481	7.7%
C	13376	7.1%
W	11747	6.3%
M	9514	5.1%
A	9382	5.0%

T	6500	3.5%
Other values (14)	26137	14.0%

*Other Punctuation*

Value	Count	Frequency (%)
'	45	88.2%
.	6	11.8%

*Space Separator*

Value	Count	Frequency (%)
	107517	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	480	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	824061	88.4%
Common	108048	11.6%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
r	77033	9.3%
e	71982	8.7%
a	58785	7.1%
d	55948	6.8%
n	49726	6.0%
t	47988	5.8%
i	45032	5.5%
o	36410	4.4%
l	32503	3.9%
R	31747	3.9%
Other values (39)	316907	38.5%

*Common*

Value	Count	Frequency (%)
-	107517	99.5%
'	480	0.4%
.	45	< 0.1%
.	6	< 0.1%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	932109	100.0%

**Most frequent character per block****ASCII**

Value	Count	Frequency (%)
r	107517	11.5%
e	77033	8.3%
a	71982	7.7%
d	58785	6.3%
n	55948	6.0%
t	49726	5.3%
i	47988	5.1%
o	45032	4.8%
l	36410	3.9%
Other values (43)	32503	3.5%
	349185	37.5%

<b>BldgNo</b>
Categorical
HIGH CARDINALITY
HIGH CORRELATION (This variable has a high overall correlation with 3 fields: Location, Character, CHArea)

<b>Distinct</b>	94
<b>Distinct (%)</b>	0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

<b>Length</b>
<b>Max length</b> 18
<b>Median length</b> 1
<b>Mean length</b> 1.2798303
<b>Min length</b> 1

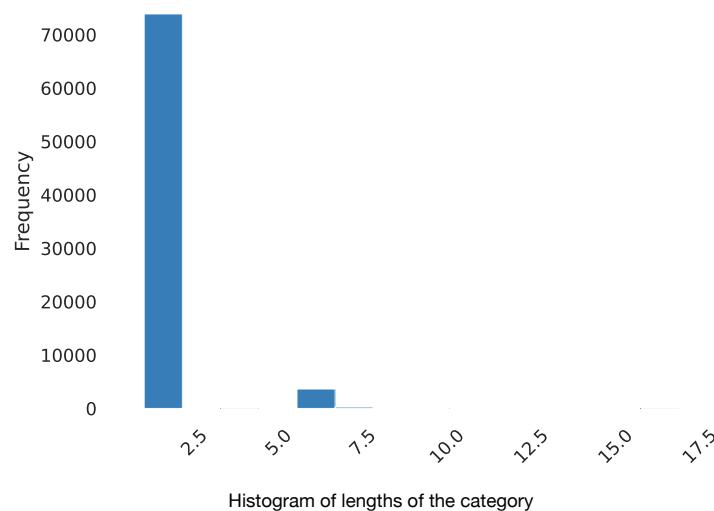
	Characters and Unicode	Unique		Sample
<b>Total characters</b>	99869	<b>Unique</b>	24	?
<b>Distinct characters</b>	53	<b>Unique (%)</b>	< 0.1%	1st row
<b>Distinct categories</b>	6			2nd row
<b>Distinct scripts</b>	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			3rd row
<b>Distinct blocks</b>	1 (https://en.wikipedia.org/wiki/Unicode_block)			4th row
				5th row

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

### Common Values

Value	Count	Frequency (%)
Bldg 2	73799	94.6%
Bldg 1	897	1.1%
Bldg A	858	1.1%
Bldg A	426	0.5%
Bldg B	348	0.4%
Bldg 3	292	0.4%
Bldg 4	221	0.3%
Bldg K	135	0.2%
Bldg C	97	0.1%
East Tower	67	0.1%
Other values (84)	893	1.1%

### Length



Value	Count	Frequency (%)
bldg	3720	44.4%
1	943	11.3%
2	941	11.2%
a	448	5.3%
b	372	4.4%
3	321	3.8%
4	276	3.3%
plaza	169	2.0%
k	135	1.6%
tower	118	1.4%
Other values (58)	931	11.1%

### Most occurring characters

Value	Count	Frequency (%)
B	77940	78.0%
I	4161	4.2%
g	3969	4.0%
d	3806	3.8%
1	3752	3.8%
2	1103	1.1%
a	1002	1.0%
A	514	0.5%
3	454	0.5%
3	326	0.3%

Other values (43)	2842	2.8%
-------------------	------	------

### Most occurring categories

Value	Count	Frequency (%)
Space Separator	77940	78.0%
Lowercase Letter	13394	13.4%
Uppercase Letter	5595	5.6%
Decimal Number	2933	2.9%
Other Punctuation	5	< 0.1%
Dash Punctuation	2	< 0.1%

### Most frequent character per category

#### *Uppercase Letter*

Value	Count	Frequency (%)
B	4161	74.4%
A	454	8.1%
P	170	3.0%
K	135	2.4%
E	119	2.1%
T	115	2.1%
C	106	1.9%
H	83	1.5%
D	57	1.0%
W	51	0.9%
Other values (10)	144	2.6%

#### *Lowercase Letter*

Value	Count	Frequency (%)
I	3969	29.6%
g	3806	28.4%
d	3752	28.0%
a	514	3.8%
e	269	2.0%
r	225	1.7%
z	169	1.3%
o	151	1.1%

t	149	1.1%
s	121	0.9%
Other values (10)	269	2.0%

*Decimal Number*

Value	Count	Frequency (%)
1	1103	37.6%
2	1002	34.2%
3	326	11.1%
4	279	9.5%
9	45	1.5%
6	43	1.5%
5	40	1.4%
7	39	1.3%
0	33	1.1%
8	23	0.8%

*Space Separator*

Value	Count	Frequency (%)
	77940	100.0%

*Other Punctuation*

Value	Count	Frequency (%)
&	5	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	2	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	80880	81.0%
Latin	18989	19.0%

**Most frequent character per script**

Latin	Value	Count	Frequency (%)

B	4161	21.9%
I	3969	20.9%
g	3806	20.0%
d	3752	19.8%
a	514	2.7%
A	454	2.4%
e	269	1.4%
r	225	1.2%
P	170	0.9%
z	169	0.9%
Other values (30)	1500	7.9%

**Common**

Value	Count	Frequency (%)
	77940	96.4%
1	1103	1.4%
2	1002	1.2%
3	326	0.4%
4	279	0.3%
9	45	0.1%
6	43	0.1%
5	40	< 0.1%
7	39	< 0.1%
0	33	< 0.1%
Other values (3)	30	< 0.1%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	99869	100.0%

**Most frequent character per block**

ASCII

Value	Count	Frequency (%)
	77940	78.0%
B	4161	4.2%
I	3969	4.0%

g	3806	3.8%
d	3752	3.8%
1	1103	1.1%
2	1002	1.0%
a	514	0.5%
A	454	0.5%
3	326	0.3%
Other values (43)	2842	2.8%

**UnitNo**  
Categorical

<b>Distinct</b>	3335
<b>Distinct (%)</b>	4.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

Max length	39
Median length	1
Mean length	2.2277626
Min length	1

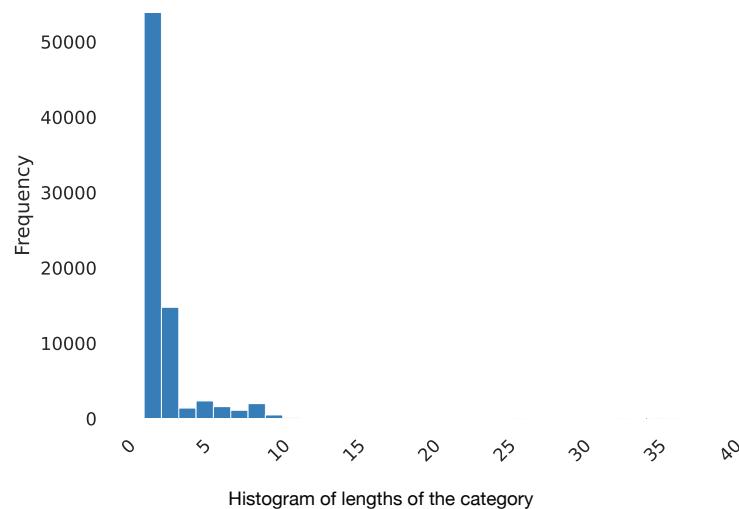
Characters and Unicode		Unique	Sample	
Total characters	173839	Unique	1153	?
Distinct characters	69	Unique (%)	1.5%	1st row
Distinct categories	10			2nd row
Distinct scripts	2			3rd row
Distinct blocks	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?	4th row
		(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)		5th row

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

**Common Values**

Value	Count	Frequency (%)
1	24368	31.2%
2	2762	3.5%
3	2226	2.9%
4	1941	2.5%
5	1823	2.3%
6	1597	2.0%
7	1483	1.9%
8	1286	1.6%
9	1182	1.5%
Other values (3325)	993	1.3%
	38372	49.2%

**Length**



Value	Count	Frequency (%)
1	3473	5.5%
to	2757	4.3%
2	2690	4.2%
3	2429	3.8%
4	2286	3.6%
5	2048	3.2%
6	1838	2.9%
7	1725	2.7%
8	1597	2.5%
	1350	2.1%
Other values (2124)	41504	65.2%

### Most occurring characters

Value	Count	Frequency (%)
	34555	19.9%
1	28490	16.4%
2	18424	10.6%
0	18069	10.4%
3	10194	5.9%
4	8347	4.8%
5	7059	4.1%
6	5947	3.4%
7	5021	2.9%
8	4667	2.7%

Other values (59)	33066	19.0%
-------------------	-------	-------

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	109924	63.2%
Space Separator	34555	19.9%
Lowercase Letter	12004	6.9%
Uppercase Letter	10431	6.0%
Other Punctuation	4953	2.8%
Dash Punctuation	1812	1.0%
Open Punctuation	70	< 0.1%
Close Punctuation	70	< 0.1%
Math Symbol	15	< 0.1%
Control	5	< 0.1%

### Most frequent character per category

#### *Uppercase Letter*

Value	Count	Frequency (%)
A	2910	27.9%
B	2399	23.0%
C	992	9.5%
F	772	7.4%
D	557	5.3%
E	547	5.2%
H	362	3.5%
L	333	3.2%
G	324	3.1%
K	170	1.6%
Other values (15)	1065	10.2%

#### *Lowercase Letter*

Value	Count	Frequency (%)
o	3773	31.4%
t	3290	27.4%
r	765	6.4%
l	743	6.2%

e	695	5.8%
s	410	3.4%
n	397	3.3%
a	329	2.7%
d	261	2.2%
p	236	2.0%
Other values (13)	1105	9.2%

*Decimal Number*

Value	Count	Frequency (%)
1	28490	25.9%
2	18424	16.8%
0	18069	16.4%
3	10194	9.3%
4	8347	7.6%
5	7059	6.4%
6	5947	5.4%
7	5021	4.6%
8	4667	4.2%
9	3706	3.4%

*Other Punctuation*

Value	Count	Frequency (%)
&	3862	78.0%
,	1058	21.4%
/	20	0.4%
.	12	0.2%
...	1	< 0.1%

*Space Separator*

Value	Count	Frequency (%)
	34555	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	1812	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
-------	-------	---------------

(	70	100.0%
---	----	--------

***Close Punctuation***

Value	Count	Frequency (%)
)	70	100.0%

***Math Symbol***

Value	Count	Frequency (%)
+	15	100.0%

***Control***

Value	Count	Frequency (%)
	5	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	151404	87.1%
Latin	22435	12.9%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
o	3773	16.8%
t	3290	14.7%
A	2910	13.0%
B	2399	10.7%
C	992	4.4%
F	772	3.4%
r	765	3.4%
I	743	3.3%
e	695	3.1%
D	557	2.5%
Other values (38)	5539	24.7%

*Common*

Value	Count	Frequency (%)
	34555	22.8%

1	28490	18.8%
2	18424	12.2%
0	18069	11.9%
3	10194	6.7%
4	8347	5.5%
5	7059	4.7%
6	5947	3.9%
7	5021	3.3%
8	4667	3.1%
Other values (11)	10631	7.0%

### Most occurring blocks

Value	Count	Frequency (%)
ASCII	173838	> 99.9%
Punctuation	1	< 0.1%

### Most frequent character per block

#### ASCII

Value	Count	Frequency (%)
1	34555	19.9%
2	28490	16.4%
0	18424	10.6%
3	18069	10.4%
4	10194	5.9%
5	8347	4.8%
6	7059	4.1%
7	5947	3.4%
8	5021	2.9%
Other values (58)	4667	2.7%
	33065	19.0%

#### Punctuation

Value	Count	Frequency (%)
...	1	100.0%



**PostalCode**

Categorical

<b>Distinct</b>	2902
<b>Distinct (%)</b>	3.7%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	33
<b>Median length</b>	7
<b>Mean length</b>	6.9953481
<b>Min length</b>	1

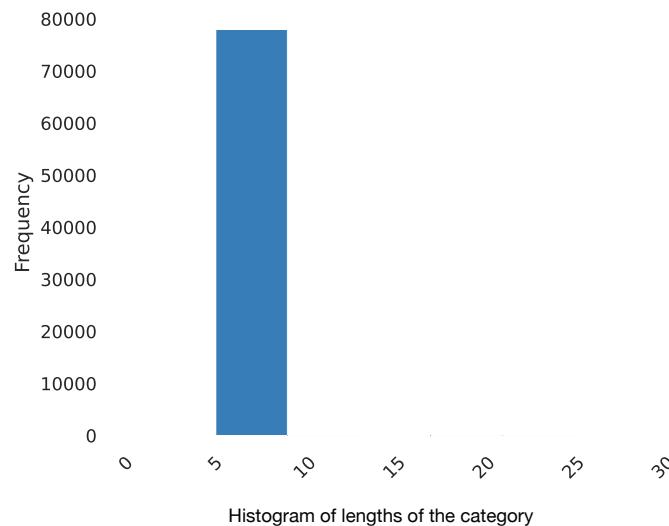
<b>Characters and Unicode</b>		<b>Unique</b>				<b>Sample</b>
<b>Total characters</b>	545868	<b>Unique</b>	139	?	<b>1st row</b>	L5T 2J3
<b>Distinct characters</b>	48	<b>Unique (%)</b>	0.2%		<b>2nd row</b>	L5T 2J3
<b>Distinct categories</b>	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				<b>3rd row</b>	L5T 2J3
<b>Distinct scripts</b>	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?	<b>4th row</b>	L5T 2J3
<b>Distinct blocks</b>	1 (https://en.wikipedia.org/wiki/Unicode_block)			?		L5T 2J3

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

**Common Values**

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
L5B 2C9	769	1.0%
L5M 4Z5	523	0.7%
L4T 2T9	477	0.6%
L5E 1V4	394	0.5%
L5P 1B2	386	0.5%
L5C 1V8	332	0.4%
L5J 1K5	296	0.4%
L4W 5G6	284	0.4%
L4X 1L4	249	0.3%
L5B 1M7	247	0.3%
Other values (2892)	74076	94.9%

**Length**



Value	Count	Frequency (%)
I4w	12403	8.0%
I5t	8317	5.3%
I5n	6069	3.9%
I4z	4948	3.2%
I5l	4693	3.0%
I5b	4589	2.9%
I5s	4258	2.7%
I5m	3801	2.4%
I4t	3311	2.1%
I5a	3290	2.1%
Other values (1078)	100200	64.3%

### Most occurring characters

Value	Count	Frequency (%)
L	86506	15.8%
	77968	14.3%
5	63752	11.7%
4	47369	8.7%
1	39205	7.2%
2	25914	4.7%
3	16424	3.0%
W	16127	3.0%
T	14622	2.7%
6	11449	2.1%

Other values (38)	146532	26.8%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	233941	42.9%
Decimal Number	233912	42.9%
Space Separator	77968	14.3%
Lowercase Letter	33	< 0.1%
Control	14	< 0.1%

### Most frequent character per category

#### *Uppercase Letter*

Value	Count	Frequency (%)
L	86506	37.0%
W	16127	6.9%
T	14622	6.3%
N	9608	4.1%
A	9326	4.0%
B	8749	3.7%
Z	8458	3.6%
M	7909	3.4%
C	7880	3.4%
V	7750	3.3%
Other values (12)	57006	24.4%

#### *Lowercase Letter*

Value	Count	Frequency (%)
k	9	27.3%
l	5	15.2%
c	5	15.2%
s	3	9.1%
t	2	6.1%
d	2	6.1%
g	1	3.0%
v	1	3.0%
h	1	3.0%

i	1	3.0%
Other values (3)	3	9.1%

*Decimal Number*

Value	Count	Frequency (%)
5	63752	27.3%
4	47369	20.3%
1	39205	16.8%
2	25914	11.1%
3	16424	7.0%
6	11449	4.9%
8	9658	4.1%
9	8879	3.8%
7	8525	3.6%
0	2737	1.2%

*Control*

Value	Count	Frequency (%)
	8	57.1%
	6	42.9%

*Space Separator*

Value	Count	Frequency (%)
	77968	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	311894	57.1%
Latin	233974	42.9%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
L	86506	37.0%
W	16127	6.9%
T	14622	6.2%
N	9608	4.1%

A	9326	4.0%
B	8749	3.7%
Z	8458	3.6%
M	7909	3.4%
C	7880	3.4%
V	7750	3.3%
Other values (25)	57039	24.4%

**Common**

Value	Count	Frequency (%)
	77968	25.0%
5	63752	20.4%
4	47369	15.2%
1	39205	12.6%
2	25914	8.3%
3	16424	5.3%
6	11449	3.7%
8	9658	3.1%
9	8879	2.8%
7	8525	2.7%
Other values (3)	2751	0.9%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	545868	100.0%

**Most frequent character per block****ASCII**

Value	Count	Frequency (%)
L	86506	15.8%
	77968	14.3%
5	63752	11.7%
4	47369	8.7%
1	39205	7.2%
2	25914	4.7%
3	16424	3.0%

W	16127	3.0%
T	14622	2.7%
6	11449	2.1%
Other values (38)	146532	26.8%

**Location**  
Categorical  
  
HIGH CARDINALITY  
HIGH CORRELATION (This variable has a high overall correlation with 7 fields: FID, BusinessID, BldgNo, NAICSCat, CENT\_X, CENT\_Y, RecordID)  
MISSING

<b>Distinct</b>	56
<b>Distinct (%)</b>	0.2%
<b>Missing</b>	47694
<b>Missing (%)</b>	61.1%

<b>Length</b>	
<b>Max length</b>	27
<b>Median length</b>	23
<b>Mean length</b>	16.483866
<b>Min length</b>	7

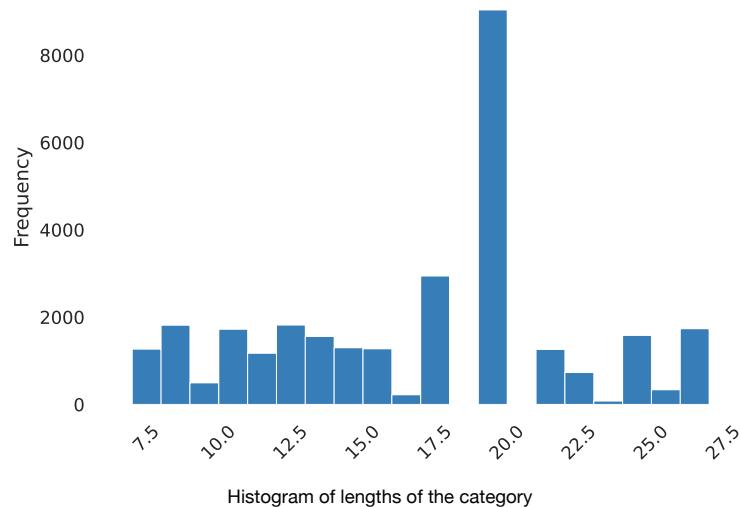
<b>Characters and Unicode</b>		<b>Unique</b>	<b>Sample</b>		
<b>Total characters</b>	500104	<b>Unique</b>	0	?	<b>1st row</b> Gateway EA (East)
<b>Distinct characters</b>	43	<b>Unique (%)</b>	0.0%		<b>2nd row</b> Gateway EA (East)
<b>Distinct categories</b>	6				<b>3rd row</b> Gateway EA (East)
<a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a>					<b>4th row</b> Gateway EA (East)
<b>Distinct scripts</b>	2			?	<b>5th row</b> Gateway EA (East)
<a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a>					
<b>Distinct blocks</b>	1	<a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a>		?	

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

### Common Values

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
Northeast EA (West)	8087	10.4%
Gateway EA (East)	1828	2.3%
Dixie EA	1814	2.3%
Meadowvale Business Park CC	1734	2.2%
Western Business Park EA	1580	2.0%
DT Core	1256	1.6%
DT Cooksville	931	1.2%
Airport CC	906	1.2%
Northeast EA (East)	738	0.9%
Mavis-Erindale EA	719	0.9%
Other values (46)	10746	13.8%
(Missing)	47694	61.1%

### Length



Value	Count	Frequency (%)
ea	15721	18.7%
northeast	8825	10.5%
west	8730	10.4%
nhd	5805	6.9%
park	3715	4.4%
east	3604	4.3%
business	3314	3.9%
cc	3101	3.7%
gateway	2618	3.1%
dt	2576	3.1%
Other values (45)	25930	30.9%

### Most occurring characters

Value	Count	Frequency (%)
e	53600	10.7%
t	44801	9.0%
s	42033	8.4%
a	38109	7.6%
r	32858	6.6%
o	25884	5.2%
E	23256	4.7%
i	21305	4.3%
A	18674	3.7%
A	17879	3.6%

Other values (33)	181705	36.3%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	300748	60.1%
Uppercase Letter	120982	24.2%
Space Separator	53600	10.7%
Open Punctuation	11741	2.3%
Close Punctuation	11741	2.3%
Dash Punctuation	1292	0.3%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
e	44801	14.9%
t	42033	14.0%
s	38109	12.7%
a	32858	10.9%
r	25884	8.6%
o	23256	7.7%
i	18674	6.2%
l	13559	4.5%
n	10785	3.6%
h	10586	3.5%
Other values (11)	40203	13.4%

#### *Uppercase Letter*

Value	Count	Frequency (%)
E	21305	17.6%
A	17879	14.8%
N	17487	14.5%
C	14057	11.6%
W	10310	8.5%
D	10195	8.4%
H	6417	5.3%
M	5583	4.6%

P	4710	3.9%
B	3314	2.7%
Other values (8)	9725	8.0%

*Space Separator*

Value	Count	Frequency (%)
	53600	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	11741	100.0%

*Close Punctuation*

Value	Count	Frequency (%)
)	11741	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	1292	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	421730	84.3%
Common	78374	15.7%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
e	44801	10.6%
t	42033	10.0%
s	38109	9.0%
a	32858	7.8%
r	25884	6.1%
o	23256	5.5%
E	21305	5.1%
i	18674	4.4%
A	17879	4.2%

N	17487	4.1%
Other values (29)	139444	33.1%

**Common**

Value	Count	Frequency (%)
	53600	68.4%
(	11741	15.0%
)	11741	15.0%
-	1292	1.6%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	500104	100.0%

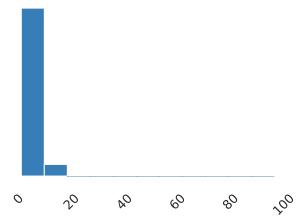
**Most frequent character per block****ASCII**

Value	Count	Frequency (%)
	53600	10.7%
e	44801	9.0%
t	42033	8.4%
s	38109	7.6%
a	32858	6.6%
r	25884	5.2%
o	23256	4.7%
E	21305	4.3%
i	18674	3.7%
A	17879	3.6%
Other values (33)	181705	36.3%

**Ward**  
Real number ( $\mathbb{R}$ )

<b>Distinct</b>	12
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	5.3925391

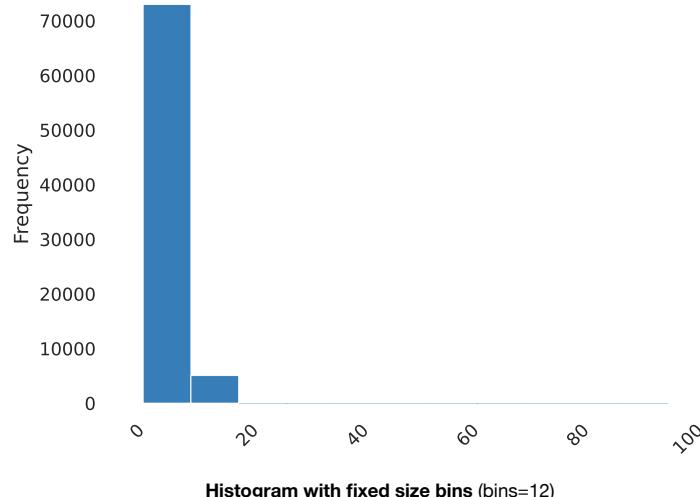
<b>Minimum</b>	1
<b>Maximum</b>	105
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	1
<b>5-th percentile</b>	1
<b>Q1</b>	5
<b>median</b>	5
<b>Q3</b>	7
<b>95-th percentile</b>	11
<b>Maximum</b>	105
<b>Range</b>	104
<b>Interquartile range (IQR)</b>	2

**Descriptive statistics**

<b>Standard deviation</b>	2.5013405
<b>Coefficient of variation (CV)</b>	0.46385208
<b>Kurtosis</b>	32.117201
<b>Mean</b>	5.3925391
<b>Median Absolute Deviation (MAD)</b>	1
<b>Skewness</b>	1.1404855
<b>Sum</b>	420796
<b>Variance</b>	6.2567041
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=12)

Value	Count	Frequency (%)
5	33956	43.5%
1	6772	8.7%
8	6086	7.8%
7	5561	7.1%
3	5005	6.4%
9	4687	6.0%
11	4300	5.5%
4	4164	5.3%

6		3584	4.6%
2		3163	4.1%
Other values (2)		755	1.0%

Value	Count	Frequency (%)
1	6772	8.7%
2	3163	4.1%
3	5005	6.4%
4	4164	5.3%
5	33956	43.5%
6	3584	4.6%
7	5561	7.1%
8	6086	7.8%
9	4687	6.0%
10	754	1.0%

Value	Count	Frequency (%)
105	1	< 0.1%
11	4300	5.5%
10	754	1.0%
9	4687	6.0%
8	6086	7.8%
7	5561	7.1%
6	3584	4.6%
5	33956	43.5%
4	4164	5.3%
3	5005	6.4%

<b>NAICSCode</b>	Missing	0
<b>Unsupported</b>	Missing (%)	0.0%
<b>REJECTED</b>	Memory size	609.8 KiB
<b>UNSUPPORTED</b>		

<b>NAICSCat</b>	Distinct	35
Categorical	Distinct (%)	< 0.1%
	Missing	14
	Missing (%)	< 0.1%
	Memory size	609.8 KiB

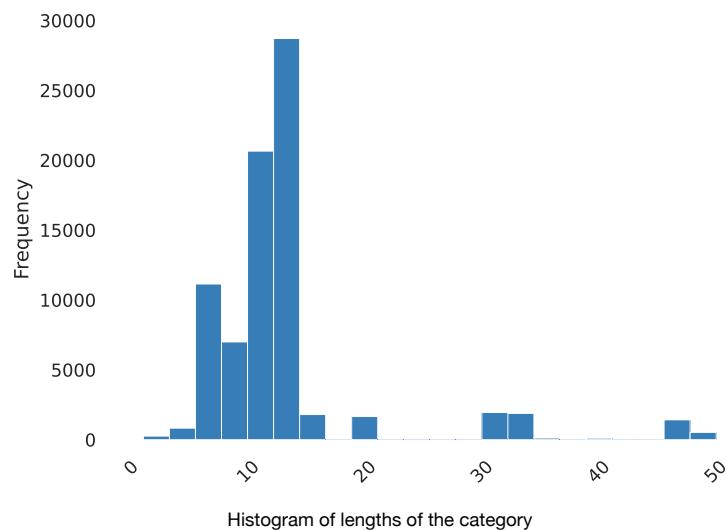
Length	Characters and Unicode		Unique	Sample
<b>Max length</b>	50	Total characters	Unique	0 ? 1st row Wholesale
<b>Median length</b>	39		Unique (%)	0.0% 2nd row Manufacturing
<b>Mean length</b>	13.393122	Distinct characters		3rd row Manufacturing
<b>Min length</b>	1	Distinct categories ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?	4th row Manufacturing
		Distinct scripts ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?	5th row Wholesale
		Distinct blocks ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?	

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

## Common Values

Value	Count	Frequency (%)
Manufacturing	9646	12.4%
Other Services	9030	11.6%
Retail	8746	11.2%
Wholesale	6933	8.9%
Professional	5654	7.2%
Health Care	5123	6.6%
Accommodation	4920	6.3%
Transportation	3039	3.9%
Construction	2778	3.6%
Educational	2430	3.1%
Other values (25)	19720	25.3%

## Length



Value	Count	Frequency (%)
services	12273	10.1%
retail	11036	9.0%
manufacturing	9646	7.9%
other	9030	7.4%
wholesale	8711	7.1%
and	7448	6.1%
professional	7076	5.8%
health	6436	5.3%
care	6436	5.3%
accommodation	6130	5.0%
Other values (37)	37833	31.0%

### Most occurring characters

Value	Count	Frequency (%)
a	109148	10.4%
e	105364	10.1%
i	79338	7.6%
n	76816	7.4%
t	76331	7.3%
r	66338	6.3%
o	64513	6.2%
s	54367	5.2%
c	52555	5.0%
l	50682	4.9%

Other values (27)	309466	29.6%
-------------------	--------	-------

**Most occurring categories**

Value	Count	Frequency (%)
Lowercase Letter	882975	84.5%
Uppercase Letter	115239	11.0%
Space Separator	44548	4.3%
Other Punctuation	2156	0.2%

**Most frequent character per category***Lowercase Letter*

Value	Count	Frequency (%)
a	109148	12.4%
e	105364	11.9%
i	79338	9.0%
n	76816	8.7%
t	76331	8.6%
r	66338	7.5%
o	64513	7.3%
s	54367	6.2%
c	52555	6.0%
l	50682	5.7%
Other values (10)	147523	16.7%

*Uppercase Letter*

Value	Count	Frequency (%)
S	15530	13.5%
R	13949	12.1%
A	12271	10.6%
M	10669	9.3%
W	9969	8.7%
C	9433	8.2%
T	9265	8.0%
O	9030	7.8%
P	7557	6.6%
H	6436	5.6%

Other values (5)	11130	9.7%
------------------	-------	------

**Space Separator**

Value	Count	Frequency (%)
	44548	100.0%

**Other Punctuation**

Value	Count	Frequency (%)
,	2156	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	998214	95.5%
Common	46704	4.5%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
a	109148	10.9%
e	105364	10.6%
i	79338	7.9%
n	76816	7.7%
t	76331	7.6%
r	66338	6.6%
o	64513	6.5%
s	54367	5.4%
c	52555	5.3%
l	50682	5.1%
Other values (25)	262762	26.3%

*Common*

Value	Count	Frequency (%)
	44548	95.4%
,	2156	4.6%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	1044918	100.0%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
a	109148	10.4%
e	105364	10.1%
i	79338	7.6%
n	76816	7.4%
t	76331	7.3%
r	66338	6.3%
o	64513	6.2%
s	54367	5.2%
c	52555	5.0%
l	50682	4.9%
Other values (27)	309466	29.6%

**NAICSDescr**

Categorical

<b>Distinct</b>	1041
<b>Distinct (%)</b>	1.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	175
<b>Median length</b>	80
<b>Mean length</b>	35.408122
<b>Min length</b>	1

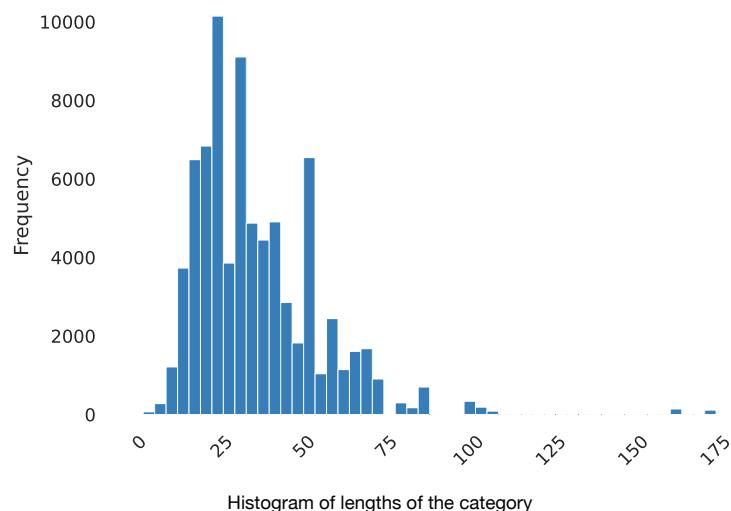
	Characters and Unicode	Unique		Sample
<b>Total characters</b>	2763002	<b>Unique</b>	125	?
<b>Distinct characters</b>	65	<b>Unique (%)</b>	0.2%	1st Amusement and Sporting Goods row Wholesaler-Distributors
<b>Distinct categories</b>	9 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )		?	2nd Support Activities for Printing row
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )		?	3rd Support Activities for Printing row
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )		?	4th Other Printing row
			?	5th Industrial Machinery, Equipment and Supplies Wholesaler-Distributors

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

**Common Values**

Value	Count	Frequency (%)
Limited-service eating places	3646	4.7%
General Automotive Repair	1991	2.6%
Full-service restaurants	1777	2.3%
Offices of Dentists	1603	2.1%
Offices of Physicians	1502	1.9%
Offices of Lawyers	1376	1.8%
Beauty Salons	1302	1.7%
Other Freight Transportation Arrangement	1253	1.6%
Elementary and Secondary Schools	1240	1.6%
Religious Organizations	1097	1.4%
Other values (1031)	61246	78.5%

**Length**



Value	Count	Frequency (%)
and	33328	10.0%
other	18664	5.6%
stores	9241	2.8%
offices	8690	2.6%
of	8401	2.5%
services	8309	2.5%
all	8269	2.5%
wholesaler-distributors	7172	2.1%
manufacturing	6726	2.0%
supplies	4484	1.3%
Other values (1055)	221527	66.2%

### Most occurring characters

Value	Count	Frequency (%)
e	278416	10.1%
	258010	9.3%
i	197847	7.2%
r	189118	6.8%
n	182948	6.6%
t	181610	6.6%
a	180904	6.5%
s	160054	5.8%
o	139237	5.0%
l	115454	4.2%

Other values (55)	879404	31.8%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	2191707	79.3%
Uppercase Letter	275883	10.0%
Space Separator	258451	9.4%
Dash Punctuation	17701	0.6%
Other Punctuation	11365	0.4%
Open Punctuation	4146	0.2%
Close Punctuation	3338	0.1%
Control	405	< 0.1%
Decimal Number	6	< 0.1%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
e	278416	12.7%
i	197847	9.0%
r	189118	8.6%
n	182948	8.3%
t	181610	8.3%
a	180904	8.3%
s	160054	7.3%
o	139237	6.4%
l	115454	5.3%
c	105554	4.8%
Other values (16)	460565	21.0%

#### *Uppercase Letter*

Value	Count	Frequency (%)
S	38630	14.0%
O	30834	11.2%
A	24801	9.0%
C	24420	8.9%
M	21757	7.9%

P	18971	6.9%
D	14639	5.3%
W	12579	4.6%
E	11730	4.3%
F	11257	4.1%
Other values (15)	66265	24.0%

*Other Punctuation*

Value	Count	Frequency (%)
,	9662	85.0%
'	803	7.1%
&	488	4.3%
-	412	3.6%

*Decimal Number*

Value	Count	Frequency (%)
1	2	33.3%
3	2	33.3%
8	1	16.7%
0	1	16.7%

*Space Separator*

Value	Count	Frequency (%)
	258010	99.8%
-	441	0.2%

*Dash Punctuation*

Value	Count	Frequency (%)
-	17701	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	4146	100.0%

*Close Punctuation*

Value	Count	Frequency (%)
)	3338	100.0%

*Control*

Value	Count	Frequency (%)
	405	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	2467590	89.3%
Common	295412	10.7%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
e	278416	11.3%
i	197847	8.0%
r	189118	7.7%
n	182948	7.4%
t	181610	7.4%
a	180904	7.3%
s	160054	6.5%
o	139237	5.6%
l	115454	4.7%
c	105554	4.3%
Other values (41)	736448	29.8%

*Common*

Value	Count	Frequency (%)
-	258010	87.3%
,	17701	6.0%
(	9662	3.3%
)	4146	1.4%
'	3338	1.1%
&	803	0.3%
.	488	0.2%
.	441	0.1%
.	412	0.1%
.	405	0.1%
Other values (4)	6	< 0.1%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	2762561	> 99.9%
None	441	< 0.1%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
e	278416	10.1%
i	258010	9.3%
r	197847	7.2%
n	189118	6.8%
t	182948	6.6%
a	181610	6.6%
s	180904	6.5%
o	160054	5.8%
l	139237	5.0%
Other values (54)	115454	4.2%
	878963	31.8%

*None*

Value	Count	Frequency (%)
	441	100.0%

**Phone**

Categorical

<b>Distinct</b>	25064
<b>Distinct (%)</b>	32.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	20
<b>Median length</b>	12
<b>Mean length</b>	11.666654
<b>Min length</b>	1

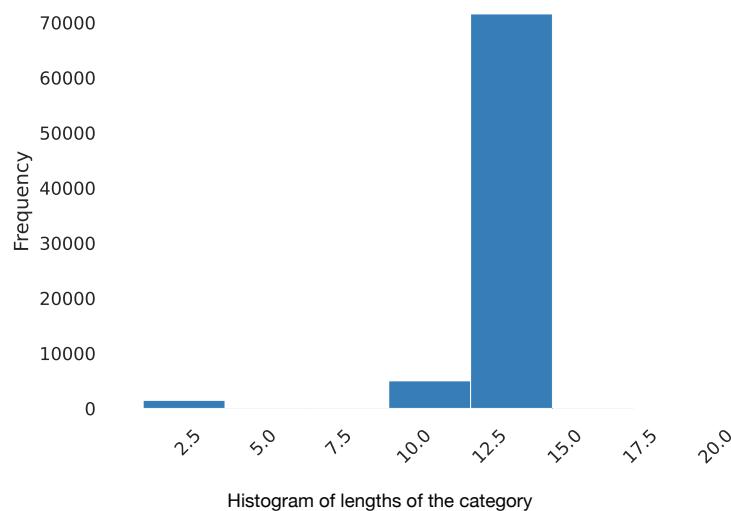
Characters and Unicode		Unique	Sample		
<b>Total characters</b>	910384	<b>Unique</b>	7404	?	<b>1st row</b> 905-795-8900
<b>Distinct characters</b>	21	<b>Unique (%)</b>	9.5%		<b>2nd row</b> 905-795-9575
<b>Distinct categories</b>	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				<b>3rd row</b> 905-795-9519
<b>Distinct scripts</b>	6 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?	<b>4th row</b> 905-564-8121
<b>Distinct blocks</b>	2 (https://en.wikipedia.org/wiki/Unicode_block)			?	<b>5th row</b> 905-564-8080

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

**Common Values**

Value	Count	Frequency (%)
905-615-3200	1457	1.9%
905-624-3811	40	0.1%
000-000-0000	35	< 0.1%
905-615-3777	35	< 0.1%
905-677-9354	24	< 0.1%
905-670-4070	21	< 0.1%
905-615-4640	20	< 0.1%
905-615-4750	20	< 0.1%
905-615-4653	18	< 0.1%
Other values (25054)	76343	97.8%

**Length**



Value	Count	Frequency (%)
905-615-3200	40	0.1%
000-000-0000	35	< 0.1%
905-624-3811	35	< 0.1%
905-615-3777	24	< 0.1%
905-677-9354	21	< 0.1%
905-670-4070	20	< 0.1%
905-615-4640	20	< 0.1%
905-615-4750	20	< 0.1%
905-615-4653	18	< 0.1%
905-949-2222	17	< 0.1%
Other values (25058)	76340	99.7%

### Most occurring characters

Value	Count	Frequency (%)
-	143128	15.7%
0	136709	15.0%
5	117588	12.9%
9	114776	12.6%
2	71079	7.8%
6	70911	7.8%
7	60428	6.6%
8	60294	6.6%
1	49065	5.4%
4	46596	5.1%

Other values (11)	39810	4.4%
-------------------	-------	------

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	765763	84.1%
Dash Punctuation	143132	15.7%
Space Separator	1471	0.2%
Other Punctuation	9	< 0.1%
Lowercase Letter	7	< 0.1%
Uppercase Letter	2	< 0.1%

### Most frequent character per category

#### *Decimal Number*

Value	Count	Frequency (%)
0	136709	17.9%
5	117588	15.4%
9	114776	15.0%
2	71079	9.3%
6	70911	9.3%
7	60428	7.9%
8	60294	7.9%
1	49065	6.4%
4	46596	6.1%
3	38317	5.0%

#### *Lowercase Letter*

Value	Count	Frequency (%)
o	2	28.6%
x	2	28.6%
t	2	28.6%
e	1	14.3%

#### *Dash Punctuation*

Value	Count	Frequency (%)
-	143128	> 99.9%
-	4	< 0.1%

*Other Punctuation*

Value	Count	Frequency (%)
.	6	66.7%
;	3	33.3%

*Uppercase Letter*

Value	Count	Frequency (%)
E	1	50.0%
B	1	50.0%

*Space Separator*

Value	Count	Frequency (%)
	1471	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	910375	> 99.9%
Latin	9	< 0.1%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
-	143128	15.7%
0	136709	15.0%
5	117588	12.9%
9	114776	12.6%
2	71079	7.8%
6	70911	7.8%
7	60428	6.6%
8	60294	6.6%
1	49065	5.4%
4	46596	5.1%
Other values (5)	39801	4.4%

*Latin*

Value	Count	Frequency (%)
0	2	22.2%

x	2	22.2%
t	2	22.2%
E	1	11.1%
e	1	11.1%
B	1	11.1%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	910380	> 99.9%
Punctuation	4	< 0.1%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
-	143128	15.7%
0	136709	15.0%
5	117588	12.9%
9	114776	12.6%
2	71079	7.8%
6	70911	7.8%
7	60428	6.6%
8	60294	6.6%
1	49065	5.4%
4	46596	5.1%
Other values (10)	39806	4.4%

*Punctuation*

Value	Count	Frequency (%)
-	4	100.0%

**Fax**  
Categorical

<b>Distinct</b>	15752
<b>Distinct (%)</b>	20.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	14
<b>Median length</b>	12
<b>Mean length</b>	7.7663296
<b>Min length</b>	1

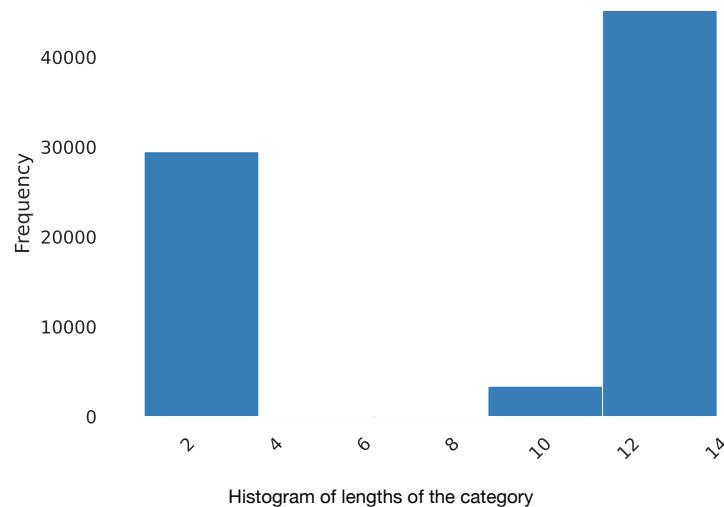
<b>Characters and Unicode</b>		<b>Unique</b>			<b>Sample</b>	
<b>Total characters</b>	606030	<b>Unique</b>	4752	?	<b>1st row</b>	905-795-8988
<b>Distinct characters</b>	12	<b>Unique (%)</b>	6.1%		<b>2nd row</b>	905-795-8775
<b>Distinct categories</b>	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )				<b>3rd row</b>	905-795-8775
<b>Distinct scripts</b>	3 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )			?	<b>4th row</b>	905-564-7395
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )			?	<b>5th row</b>	905-564-5003

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

**Common Values**

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
905-822-2673	29474	37.8%
905-361-6401	41	0.1%
905-896-9380	37	< 0.1%
905-502-6982	31	< 0.1%
905-625-4815	18	< 0.1%
905-542-0987	17	< 0.1%
905-607-9204	16	< 0.1%
905-625-8815	16	< 0.1%
905-403-8409	15	< 0.1%
Other values (15742)	14	< 0.1%
	48354	62.0%

**Length**



Value	Count	Frequency (%)
905-822-2673	41	0.1%
905-361-6401	37	0.1%
905-896-9380	31	0.1%
905-502-6982	18	< 0.1%
905-625-4815	17	< 0.1%
905-542-0987	16	< 0.1%
905-607-9204	16	< 0.1%
905-625-8815	15	< 0.1%
905-403-8409	14	< 0.1%
905-625-8245	13	< 0.1%
Other values (15742)	48342	99.6%

### Most occurring characters

Value	Count	Frequency (%)
-	90675	15.0%
0	79738	13.2%
5	78040	12.9%
9	75509	12.5%
6	47327	7.8%
2	44185	7.3%
8	39652	6.5%
7	37892	6.3%
1	30365	5.0%
	29475	4.9%

Other values (2)	53172	8.8%
------------------	-------	------

**Most occurring categories**

Value	Count	Frequency (%)
Decimal Number	485880	80.2%
Dash Punctuation	90675	15.0%
Space Separator	29475	4.9%

**Most frequent character per category***Decimal Number*

Value	Count	Frequency (%)
0	79738	16.4%
5	78040	16.1%
9	75509	15.5%
6	47327	9.7%
2	44185	9.1%
8	39652	8.2%
7	37892	7.8%
1	30365	6.2%
4	27785	5.7%
3	25387	5.2%

*Dash Punctuation*

Value	Count	Frequency (%)
-	90675	100.0%

*Space Separator*

Value	Count	Frequency (%)
	29475	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	606030	100.0%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
-	90675	15.0%
0	79738	13.2%
5	78040	12.9%
9	75509	12.5%
6	47327	7.8%
2	44185	7.3%
8	39652	6.5%
7	37892	6.3%
1	30365	5.0%
	29475	4.9%
Other values (2)	53172	8.8%

### Most occurring blocks

Value	Count	Frequency (%)
ASCII	606030	100.0%

### Most frequent character per block

ASCII

Value	Count	Frequency (%)
-	90675	15.0%
0	79738	13.2%
5	78040	12.9%
9	75509	12.5%
6	47327	7.8%
2	44185	7.3%
8	39652	6.5%
7	37892	6.3%
1	30365	5.0%
	29475	4.9%
Other values (2)	53172	8.8%

**TollFree**

Categorical

<b>Distinct</b>	4117
<b>Distinct (%)</b>	5.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	16
<b>Median length</b>	1
<b>Mean length</b>	2.8538695
<b>Min length</b>	1

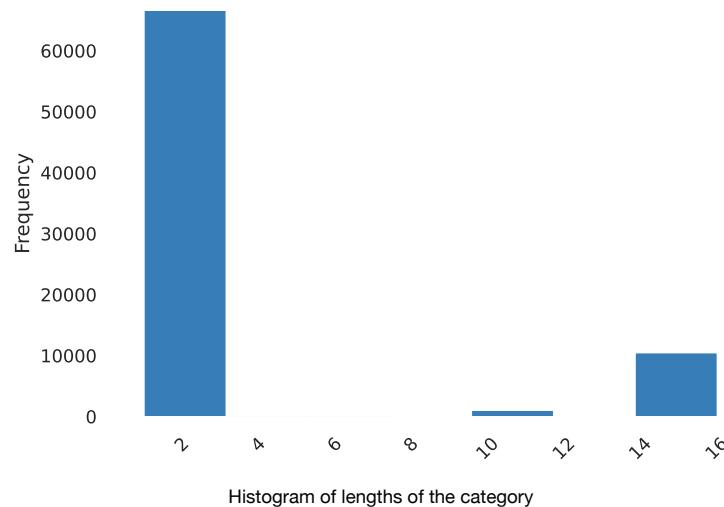
Characters and Unicode		Unique	Sample		
<b>Total characters</b>	222696	<b>Unique</b>	1434	?	1st row
<b>Distinct characters</b>	15	<b>Unique (%)</b>	1.8%		2nd row
<b>Distinct categories</b>	5				3rd row
( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )					4th row
<b>Distinct scripts</b>	2			?	
( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )					
<b>Distinct blocks</b>	2	( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )		?	

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

**Common Values**

Value	Count	Frequency (%)
1-800-769-2511	66597	85.3%
1-800-465-2422	32	< 0.1%
1-800-472-6842	23	< 0.1%
1-877-777-8672	16	< 0.1%
1-877-849-3637	16	< 0.1%
1-866-567-8888	13	< 0.1%
1-800-668-0414	10	< 0.1%
1-800-956-9543	10	< 0.1%
1-866-829-9433	10	< 0.1%
Other values (4107)	11274	14.4%

**Length**



Value	Count	Frequency (%)
1-800-769-2511	32	0.3%
1-800-465-2422	32	0.3%
1-800-472-6842	23	0.2%
1-877-777-8672	16	0.1%
1-877-849-3637	16	0.1%
1-866-567-8888	13	0.1%
1-877-526-6639	10	0.1%
1-800-254-0778	10	0.1%
1-800-563-4327	10	0.1%
1-866-829-9433	10	0.1%
Other values (4111)	11269	98.5%

### Most occurring characters

Value	Count	Frequency (%)
6	66602	29.9%
-	31297	14.1%
8	24221	10.9%
1	16130	7.2%
0	14466	6.5%
6	14461	6.5%
7	12782	5.7%
5	9818	4.4%
2	9799	4.4%
3	8526	3.8%

Other values (5)	14594	6.6%
------------------	-------	------

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	124793	56.0%
Space Separator	66602	29.9%
Dash Punctuation	31299	14.1%
Lowercase Letter	1	< 0.1%
Other Punctuation	1	< 0.1%

### Most frequent character per category

#### *Decimal Number*

Value	Count	Frequency (%)
8	24221	19.4%
1	16130	12.9%
0	14466	11.6%
6	14461	11.6%
7	12782	10.2%
5	9818	7.9%
2	9799	7.9%
3	8526	6.8%
4	7930	6.4%
9	6660	5.3%

#### *Dash Punctuation*

Value	Count	Frequency (%)
-	31297	> 99.9%
-	2	< 0.1%

#### *Space Separator*

Value	Count	Frequency (%)
	66602	100.0%

#### *Lowercase Letter*

Value	Count	Frequency (%)
x	1	100.0%

*Other Punctuation*

Value	Count	Frequency (%)
.	1	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	222695	> 99.9%
Latin	1	< 0.1%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
-	66602	29.9%
8	31297	14.1%
1	24221	10.9%
0	16130	7.2%
6	14466	6.5%
7	14461	6.5%
5	12782	5.7%
2	9818	4.4%
3	9799	4.4%
Other values (4)	8526	3.8%
	14593	6.6%

*Latin*

Value	Count	Frequency (%)
x	1	100.0%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	222694	> 99.9%
Punctuation	2	< 0.1%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
-	66602	29.9%
8	31297	14.1%
1	24221	10.9%
0	16130	7.2%
6	14466	6.5%
7	14461	6.5%
5	12782	5.7%
2	9818	4.4%
3	9799	4.4%
Other values (4)	8526	3.8%
	14592	6.6%

*Punctuation*

Value	Count	Frequency (%)
-	2	100.0%

**EMail**  
Categorical

<b>Distinct</b>	15058
<b>Distinct (%)</b>	19.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	97
<b>Median length</b>	55
<b>Mean length</b>	14.084964
<b>Min length</b>	1

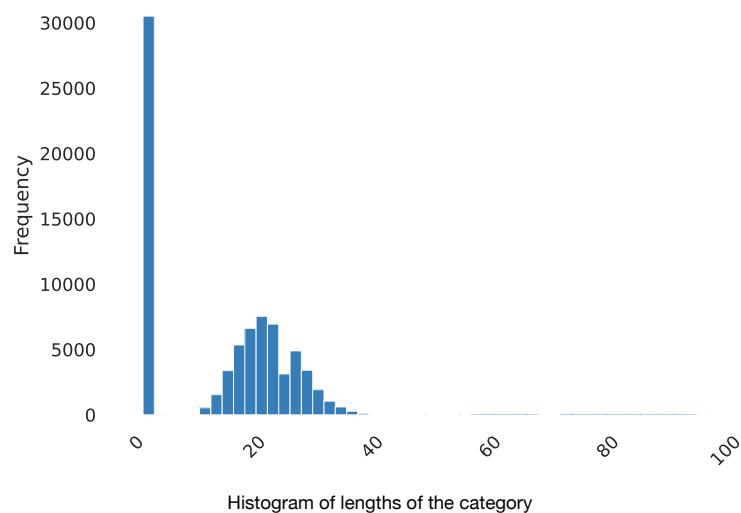
		<b>Characters and Unicode</b>	<b>Unique</b>			<b>Sample</b>
<b>Total characters</b>	1099092	<b>Unique</b>	3361	?	<b>1st row</b>	lfinch@golftrendsinc.com
<b>Distinct characters</b>	78	<b>Unique (%)</b>	4.3%		<b>2nd row</b>	prepress@apexgraphics.com
<b>Distinct categories</b>	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )				<b>3rd row</b>	
<b>Distinct scripts</b>	( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )			?	<b>4th row</b>	info@printmedia.ca
<b>Distinct blocks</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )			?	<b>5th row</b>	shsieh@swrltd.com

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

**Common Values**

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	30507	39.1%
info@publicstoragecanada.com	21	< 0.1%
info@taxwide.com	20	< 0.1%
info@ucmas.ca	13	< 0.1%
info@mississaugaschoolofmusic.ca	13	< 0.1%
cyclone@cyclonemfg.com	12	< 0.1%
millertrailers@rogers.com	12	< 0.1%
info@realfruitbubbletea.com	12	< 0.1%
info@akaloptical.com	12	< 0.1%
ktc.ca.info@kapsch.net	12	< 0.1%
Other values (15048)	47399	60.7%

**Length**



Value	Count	Frequency (%)
info@publicstoragecanada.com	21	< 0.1%
info@taxwide.com	20	< 0.1%
info@ucmas.ca	13	< 0.1%
info@mississaugaschoolofmusic.ca	13	< 0.1%
cyclone@cyclonemfg.com	12	< 0.1%
millertrailers@rogers.com	12	< 0.1%
info@realfruitbubbletea.com	12	< 0.1%
info@akaloptical.com	12	< 0.1%
ktc.ca.info@kapsch.net	12	< 0.1%
insure@all-risks.com	11	< 0.1%
Other values (15012)	47482	99.7%

### Most occurring characters

Value	Count	Frequency (%)
o	99086	9.0%
a	97080	8.8%
c	83214	7.6%
i	74076	6.7%
e	72811	6.6%
n	63754	5.8%
m	63062	5.7%
s	58432	5.3%
r	53466	4.9%
.	51798	4.7%

Other values (68)	382313	34.8%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	953466	86.8%
Other Punctuation	99332	9.0%
Space Separator	30708	2.8%
Decimal Number	11022	1.0%
Uppercase Letter	1925	0.2%
Dash Punctuation	1864	0.2%
Connector Punctuation	766	0.1%
Control	4	< 0.1%
Modifier Symbol	3	< 0.1%
Final Punctuation	1	< 0.1%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
o	99086	10.4%
a	97080	10.2%
c	83214	8.7%
i	74076	7.8%
e	72811	7.6%
n	63754	6.7%
m	63062	6.6%
s	58432	6.1%
r	53466	5.6%
t	50375	5.3%
Other values (16)	238110	25.0%

#### *Uppercase Letter*

Value	Count	Frequency (%)
I	281	14.6%
S	211	11.0%
M	203	10.5%
C	133	6.9%

A	122	6.3%
D	96	5.0%
P	88	4.6%
B	81	4.2%
J	79	4.1%
T	77	4.0%
Other values (16)	554	28.8%

*Decimal Number*

Value	Count	Frequency (%)
1	1932	17.5%
0	1824	16.5%
2	1678	15.2%
3	975	8.8%
5	873	7.9%
4	804	7.3%
7	764	6.9%
6	755	6.8%
8	753	6.8%
9	664	6.0%

*Other Punctuation*

Value	Count	Frequency (%)
.	51798	52.1%
@	47451	47.8%
/	35	< 0.1%
&	18	< 0.1%
,	8	< 0.1%
'	7	< 0.1%
#	5	< 0.1%
:	5	< 0.1%
-	5	< 0.1%

*Space Separator*

Value	Count	Frequency (%)
	30708	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	1864	100.0%

*Connector Punctuation*

Value	Count	Frequency (%)
-	766	100.0%

*Control*

Value	Count	Frequency (%)
	4	100.0%

*Modifier Symbol*

Value	Count	Frequency (%)
'	3	100.0%

*Final Punctuation*

Value	Count	Frequency (%)
,	1	100.0%

*Math Symbol*

Value	Count	Frequency (%)
+	1	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	955391	86.9%
Common	143701	13.1%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
o	99086	10.4%
a	97080	10.2%
c	83214	8.7%
i	74076	7.8%
e	72811	7.6%
n	63754	6.7%
m	63062	6.6%

s	58432	6.1%
r	53466	5.6%
t	50375	5.3%
Other values (42)	240035	25.1%

**Common**

Value	Count	Frequency (%)
.	51798	36.0%
@	47451	33.0%
	30708	21.4%
1	1932	1.3%
-	1864	1.3%
0	1824	1.3%
2	1678	1.2%
3	975	0.7%
5	873	0.6%
4	804	0.6%
Other values (16)	3794	2.6%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	1099086	> 99.9%
None	5	< 0.1%
Punctuation	1	< 0.1%

**Most frequent character per block**

ASCII

Value	Count	Frequency (%)
o	99086	9.0%
a	97080	8.8%
c	83214	7.6%
i	74076	6.7%
e	72811	6.6%
n	63754	5.8%
m	63062	5.7%
s	58432	5.3%

r	53466	4.9%
.	51798	4.7%
Other values (66)	382307	34.8%

*None*

Value	Count	Frequency (%)
.	5	100.0%

*Punctuation*

Value	Count	Frequency (%)
,	1	100.0%

**WebAddress**

Categorical

<b>Distinct</b>	14200
<b>Distinct (%)</b>	18.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	84
<b>Median length</b>	50
<b>Mean length</b>	14.52579
<b>Min length</b>	1

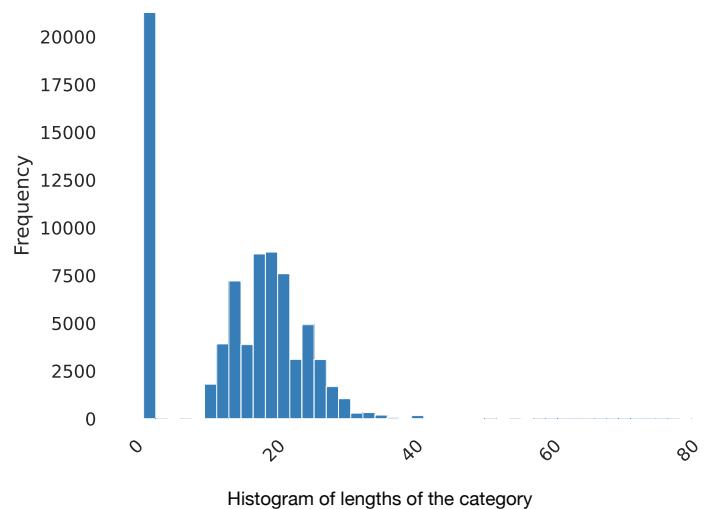
Characters and Unicode		Unique	Sample		
<b>Total characters</b>	1133491	<b>Unique</b>	2033	?	<b>1st row</b> www.golftrendsinc.com
<b>Distinct characters</b>	80	<b>Unique (%)</b>	2.6%		<b>2nd row</b> www.apexgraphics.com
<b>Distinct categories</b>	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				<b>3rd row</b>
<b>Distinct scripts</b>	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?	<b>4th row</b> www.printmedia.ca
<b>Distinct blocks</b>	2 (https://en.wikipedia.org/wiki/Unicode_block)			?	<b>5th row</b> www.swrltd.com

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

**Common Values**

Value	Count	Frequency (%)
www.dpcdsb.org	21267	27.3%
www.subway.com	221	0.3%
www.timhortons.com	215	0.3%
www.petro-canada.ca	211	0.3%
www.shoppersdrugmart.ca	115	0.1%
www.mississauga.ca/portal/residents/fire	107	0.1%
www.td.com	95	0.1%
www.dollarama.com	91	0.1%
www.shell.ca	88	0.1%
Other values (14190)	84	0.1%
	55539	71.2%

**Length**



Value	Count	Frequency (%)
www.dpcdsb.org	221	0.4%
www.subway.com	215	0.4%
www.timhortons.com	211	0.4%
www.petro-canada.ca	115	0.2%
www.shoppersdrugmart.ca	107	0.2%
www.mississauga.ca/portal/residents/fire	95	0.2%
www.td.com	91	0.2%
www.dollarama.com	88	0.2%
www.shell.ca	84	0.1%
www.starbucks.ca	83	0.1%
Other values (14093)	55517	97.7%

### Most occurring characters

Value	Count	Frequency (%)
w	178473	15.7%
.	114798	10.1%
c	90001	7.9%
a	87304	7.7%
o	81313	7.2%
e	65392	5.8%
m	55956	4.9%
s	50675	4.5%
i	50384	4.4%
r	49833	4.4%

Other values (70)	309362	27.3%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	989750	87.3%
Other Punctuation	116176	10.2%
Space Separator	21324	1.9%
Dash Punctuation	2684	0.2%
Decimal Number	2467	0.2%
Uppercase Letter	1007	0.1%
Math Symbol	52	< 0.1%
Control	10	< 0.1%
Connector Punctuation	10	< 0.1%
Modifier Symbol	8	< 0.1%
Other values (2)	3	< 0.1%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
w	178473	18.0%
c	90001	9.1%
a	87304	8.8%
o	81313	8.2%
e	65392	6.6%
m	55956	5.7%
s	50675	5.1%
i	50384	5.1%
r	49833	5.0%
t	47223	4.8%
Other values (17)	233196	23.6%

#### *Uppercase Letter*

Value	Count	Frequency (%)
C	108	10.7%
W	105	10.4%
S	71	7.1%

M	70	7.0%
T	59	5.9%
A	57	5.7%
L	57	5.7%
F	52	5.2%
R	51	5.1%
P	41	4.1%
Other values (16)	336	33.4%

*Decimal Number*

Value	Count	Frequency (%)
1	551	22.3%
2	475	19.3%
0	349	14.1%
4	324	13.1%
3	230	9.3%
6	129	5.2%
8	119	4.8%
9	119	4.8%
5	101	4.1%
7	70	2.8%

*Other Punctuation*

Value	Count	Frequency (%)
.	114798	98.8%
/	1297	1.1%
@	47	< 0.1%
&	18	< 0.1%
\	6	< 0.1%
,	4	< 0.1%
:	3	< 0.1%
'	2	< 0.1%
.	1	< 0.1%

*Space Separator*

Value	Count	Frequency (%)
	21324	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	2684	100.0%

*Math Symbol*

Value	Count	Frequency (%)
~	52	100.0%

*Control*

Value	Count	Frequency (%)
	10	100.0%

*Connector Punctuation*

Value	Count	Frequency (%)
-	10	100.0%

*Modifier Symbol*

Value	Count	Frequency (%)
'	8	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	2	100.0%

*Close Punctuation*

Value	Count	Frequency (%)
)	1	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	990757	87.4%
Common	142734	12.6%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
w	178473	18.0%
c	90001	9.1%

a	87304	8.8%
o	81313	8.2%
e	65392	6.6%
m	55956	5.6%
s	50675	5.1%
i	50384	5.1%
r	49833	5.0%
t	47223	4.8%
Other values (43)	234203	23.6%

**Common**

Value	Count	Frequency (%)
.	114798	80.4%
-	21324	14.9%
/	2684	1.9%
1	1297	0.9%
2	551	0.4%
0	475	0.3%
4	349	0.2%
3	324	0.2%
6	230	0.2%
Other values (17)	129	0.1%
	573	0.4%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	1133487	> 99.9%
None	4	< 0.1%

**Most frequent character per block**

ASCII	Value	Count	Frequency (%)
w	w	178473	15.7%
.	.	114798	10.1%
c	c	90001	7.9%
a	a	87304	7.7%

o	81313	7.2%
e	65392	5.8%
m	55956	4.9%
s	50675	4.5%
i	50384	4.4%
r	49833	4.4%
Other values (68)	309358	27.3%

***None***

Value	Count	Frequency (%)
é	3	75.0%
.	1	25.0%

**EmplRange**

Categorical

**HIGH CORRELATION** (This variable has a high overall correlation with 2 fields: NAICSCat, Year)

MISSING

<b>Distinct</b>	19
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	2646
<b>Missing (%)</b>	3.4%
<b>Memory size</b>	609.8 KiB

**Length**

Max length	10
Median length	6
Mean length	6.4964384
Min length	5

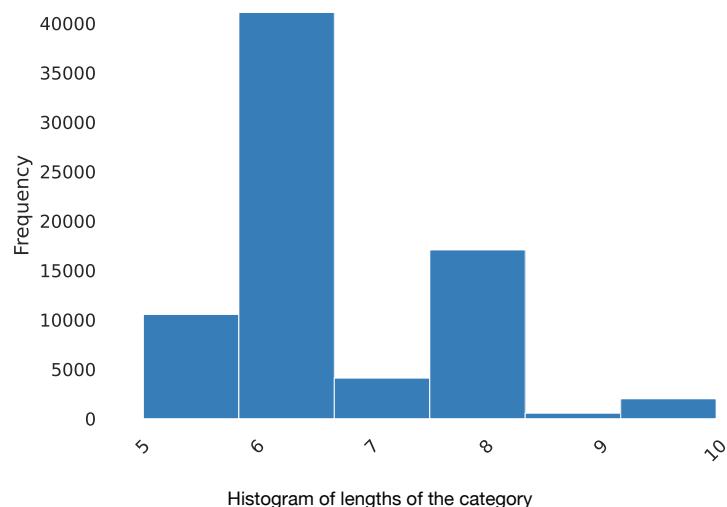
Characters and Unicode		Unique	Sample		
Total characters	489747	Unique	0	?	1st row
Distinct characters	16	Unique (%)	0.0%		2nd row
Distinct categories	5				3rd row
Distinct scripts	2			?	4th row
Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		5th row	5 to 9

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

**Common Values**

Value	Count	Frequency (%)
1 to 4	28587	36.6%
5 to 9	12508	16.0%
10 to 19	8204	10.5%
1 - 4	7498	9.6%
20 to 49	6290	8.1%
5 - 9	3032	3.9%
50 to 99	2582	3.3%
10 - 19	1967	2.5%
100 to 299	1640	2.1%
20 - 49	1527	2.0%
Other values (9)	1552	2.0%
(Missing)	2646	3.4%

**Length**

**Value****Count Frequency (%)**

to	60184	26.6%
1	36085	16.0%
4	36085	16.0%
5	15540	6.9%
9	15540	6.9%
	15116	6.7%
10	10171	4.5%
19	10171	4.5%
20	7817	3.5%
49	7817	3.5%
Other values (11)	11533	5.1%

**Most occurring characters****Value****Count Frequency (%)**

	150672	30.8%
t	60184	12.3%
o	60184	12.3%
1	58552	12.0%
9	45056	9.2%
4	44205	9.0%
0	26431	5.4%
5	18886	3.9%
-	15116	3.1%
2	9855	2.0%

Other values (6)	606	0.1%
------------------	-----	------

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	203288	41.5%
Space Separator	150672	30.8%
Lowercase Letter	120656	24.6%
Dash Punctuation	15116	3.1%
Math Symbol	15	< 0.1%

### Most frequent character per category

#### *Decimal Number*

Value	Count	Frequency (%)
1	58552	28.8%
9	45056	22.2%
4	44205	21.7%
0	26431	13.0%
5	18886	9.3%
2	9855	4.8%
3	303	0.1%

#### *Lowercase Letter*

Value	Count	Frequency (%)
t	60184	49.9%
o	60184	49.9%
p	72	0.1%
l	72	0.1%
u	72	0.1%
s	72	0.1%

#### *Space Separator*

Value	Count	Frequency (%)
-	150672	100.0%

#### *Dash Punctuation*

Value	Count	Frequency (%)
-	15116	100.0%

*Math Symbol*

Value	Count	Frequency (%)
+	15	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	369091	75.4%
Latin	120656	24.6%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
	150672	40.8%
1	58552	15.9%
9	45056	12.2%
4	44205	12.0%
0	26431	7.2%
5	18886	5.1%
-	15116	4.1%
2	9855	2.7%
3	303	0.1%
+	15	< 0.1%

*Latin*

Value	Count	Frequency (%)
t	60184	49.9%
o	60184	49.9%
p	72	0.1%
l	72	0.1%
u	72	0.1%
s	72	0.1%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	489747	100.0%

**Most frequent character per block**

ASC//

Value	Count	Frequency (%)
t	150672	30.8%
o	60184	12.3%
1	60184	12.3%
9	58552	12.0%
4	45056	9.2%
0	44205	9.0%
5	26431	5.4%
-	18886	3.9%
2	15116	3.1%
Other values (6)	9855	2.0%
	606	0.1%

**EmplUpdate**

Categorical

HIGH CARDINALITY  
MISSING

<b>Distinct</b>	433
<b>Distinct (%)</b>	0.7%
<b>Missing</b>	15002
<b>Missing (%)</b>	19.2%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	22
<b>Median length</b>	22
<b>Mean length</b>	22
<b>Min length</b>	22

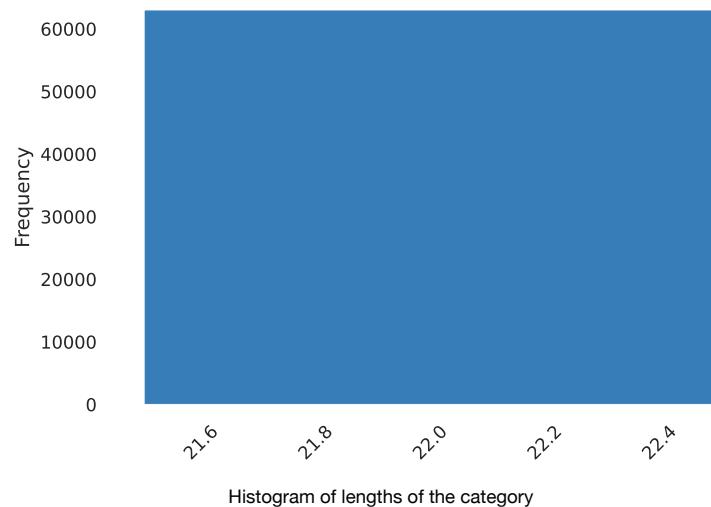
Characters and Unicode		Unique		Sample		
<b>Total characters</b>	1386682	<b>Unique</b>	111	?	<b>1st row</b>	2015/10/31 00:00:00+00
<b>Distinct characters</b>	14	<b>Unique (%)</b>	0.2%		<b>2nd row</b>	2016/10/31 00:00:00+00
<b>Distinct categories</b>	4				<b>3rd row</b>	2015/10/31 00:00:00+00
<b>Distinct scripts</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?		<b>4th row</b>	2015/10/31 00:00:00+00
<b>Distinct blocks</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?		<b>5th row</b>	2015/10/31 00:00:00+00

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

**Common Values**

Value	Count	Frequency (%)
2017/11/08 00:00:00+00	11038	14.1%
2018/12/30 00:00:00+00	9918	12.7%
2017/11/09 00:00:00+00	8042	10.3%
2015/10/31 00:00:00+00	4560	5.8%
2016/10/31 00:00:00+00	4499	5.8%
2019/12/12 00:00:00+00	3326	4.3%
2019/09/19 00:00:00+00	2718	3.5%
2018/09/30 00:00:00+00	849	1.1%
2017/06/08 00:00:00+00	726	0.9%
2017/05/24 00:00:00+00	646	0.8%
Other values (423)	16709	21.4%
(Missing)	15002	19.2%

**Length**



Value	Count	Frequency (%)
00:00:00+00	63031	50.0%
2017/11/08	11038	8.8%
2018/12/30	9918	7.9%
2017/11/09	8042	6.4%
2015/10/31	4560	3.6%
2016/10/31	4499	3.6%
2019/12/12	3326	2.6%
2019/09/19	2718	2.2%
2018/09/30	849	0.7%
2017/06/08	726	0.6%
Other values (424)	17355	13.8%

### Most occurring characters

Value	Count	Frequency (%)
0	635263	45.8%
1	147280	10.6%
/	126062	9.1%
:	126062	9.1%
2	85562	6.2%
	63031	4.5%
+	63031	4.5%
7	33727	2.4%
8	28229	2.0%
9	23888	1.7%

Other values (4)	54547	3.9%
------------------	-------	------

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	1008496	72.7%
Other Punctuation	252124	18.2%
Space Separator	63031	4.5%
Math Symbol	63031	4.5%

### Most frequent character per category

#### *Decimal Number*

Value	Count	Frequency (%)
0	635263	63.0%
1	147280	14.6%
2	85562	8.5%
7	33727	3.3%
8	28229	2.8%
9	23888	2.4%
3	22832	2.3%
5	16305	1.6%
6	13078	1.3%
4	2332	0.2%

#### *Other Punctuation*

Value	Count	Frequency (%)
/	126062	50.0%
:	126062	50.0%

#### *Space Separator*

Value	Count	Frequency (%)
	63031	100.0%

#### *Math Symbol*

Value	Count	Frequency (%)
+	63031	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	1386682	100.0%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
0	635263	45.8%
1	147280	10.6%
/	126062	9.1%
:	126062	9.1%
2	85562	6.2%
	63031	4.5%
+	63031	4.5%
7	33727	2.4%
8	28229	2.0%
9	23888	1.7%
Other values (4)	54547	3.9%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	1386682	100.0%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
0	635263	45.8%
1	147280	10.6%
/	126062	9.1%
:	126062	9.1%
2	85562	6.2%
	63031	4.5%
+	63031	4.5%
7	33727	2.4%
8	28229	2.0%

9		23888	1.7%
Other values (4)		54547	3.9%

**Sector\_Des**  
Categorical  
  
**HIGH CORRELATION** (This variable has a high overall correlation with 1 fields: NAICSCat)  
**MISSING**

<b>Distinct</b>	29
<b>Distinct (%)</b>	0.2%
<b>Missing</b>	63431
<b>Missing (%)</b>	81.3%
<b>Memory size</b>	609.8 KiB

<b>Length</b>	
<b>Max length</b>	57
<b>Median length</b>	1
<b>Mean length</b>	3.3009177
<b>Min length</b>	1

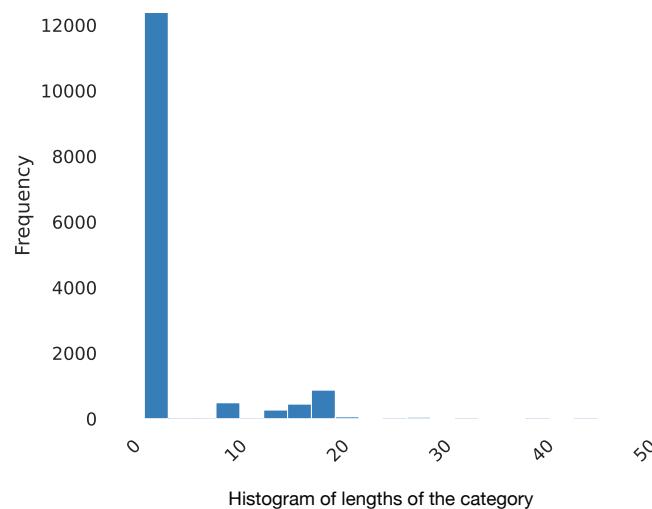
		<b>Characters and Unicode</b>	<b>Unique</b>	<b>Sample</b>		
<b>Total characters</b>	48200	<b>Unique</b>	6	?	1st row	
<b>Distinct characters</b>	26	<b>Unique (%)</b>	< 0.1%		2nd row	
<b>Distinct categories</b>	4				3rd row	
<b>Distinct scripts</b>	2			?	4th row	
<b>Distinct blocks</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )		?	5th row	

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

### Common Values

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	12383	15.9%
Financial Services	870	1.1%
Food and Beverage	444	0.6%
Automotive	329	0.4%
Life Sciences	263	0.3%
Aerospace	132	0.2%
Automotive,Aerospace	55	0.1%
Cleantech	24	< 0.1%
Automotive,Food and Beverage	24	< 0.1%
Automotive,Aerospace,Food and Beverage	15	< 0.1%
Other values (19)	63	0.1%
(Missing)	63431	81.3%

### Length



Value	Count	Frequency (%)
services	884	19.8%
financial	870	19.5%
and	528	11.9%
beverage	514	11.5%
food	452	10.1%
automotive	329	7.4%
life	281	6.3%
sciences	265	5.9%
aerospace	132	3.0%
automotive,aerospace	55	1.2%
Other values (15)	145	3.3%

### Most occurring characters

Value	Count	Frequency (%)
e	14623	30.3%
i	5221	10.8%
a	3691	7.7%
c	3091	6.4%
n	2627	5.5%
o	2183	5.4%
v	1859	4.5%
r	1645	3.9%
s	1413	3.4%

Other values (16)

9221 19.1%

**Most occurring categories**

Value	Count	Frequency (%)
Lowercase Letter	29243	60.7%
Space Separator	14623	30.3%
Uppercase Letter	4130	8.6%
Other Punctuation	204	0.4%

**Most frequent character per category***Lowercase Letter*

Value	Count	Frequency (%)
e	5221	17.9%
i	3691	12.6%
a	3091	10.6%
c	2627	9.0%
n	2626	9.0%
o	2183	7.5%
v	1859	6.4%
r	1645	5.6%
s	1413	4.8%
d	1056	3.6%
Other values (8)	3831	13.1%

*Uppercase Letter*

Value	Count	Frequency (%)
F	1412	34.2%
S	1180	28.6%
A	680	16.5%
B	528	12.8%
L	296	7.2%
C	34	0.8%

*Space Separator*

Value	Count	Frequency (%)
	14623	100.0%

*Other Punctuation*

Value	Count	Frequency (%)
,	204	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	33373	69.2%
Common	14827	30.8%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
e	5221	15.6%
i	3691	11.1%
a	3091	9.3%
c	2627	7.9%
n	2626	7.9%
o	2183	6.5%
v	1859	5.6%
r	1645	4.9%
s	1413	4.2%
F	1412	4.2%
Other values (14)	7605	22.8%

*Common*

Value	Count	Frequency (%)
,	14623	98.6%
	204	1.4%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	48200	100.0%

**Most frequent character per block***ASCII*

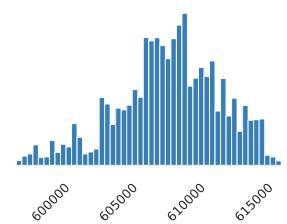
Value	Count	Frequency (%)
	14623	30.3%
e	5221	10.8%
i	3691	7.7%
a	3091	6.4%
c	2627	5.5%
n	2626	5.4%
o	2183	4.5%
v	1859	3.9%
r	1645	3.4%
s	1413	2.9%
Other values (16)	9221	19.1%

**CENT\_X**  
Real number ( $\mathbb{R}$ )

**HIGH CORRELATION** (This variable has a high overall correlation with 2 fields: Location, CENT\_Y)  
MISSING

<b>Distinct</b>	4685
<b>Distinct (%)</b>	15.4%
<b>Missing</b>	47694
<b>Missing (%)</b>	61.1%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	608659.35

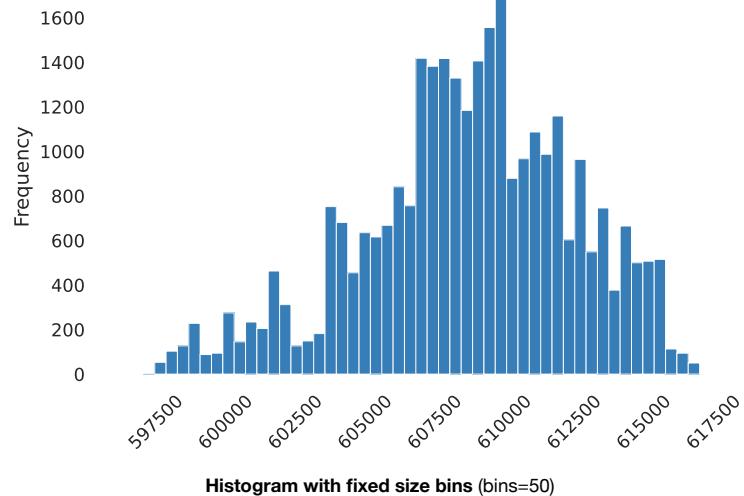
<b>Minimum</b>	596627.93
<b>Maximum</b>	616985.06
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	596627.93
<b>5-th percentile</b>	601465.65
<b>Q1</b>	606483.02
<b>median</b>	608923.98
<b>Q3</b>	611391.08
<b>95-th percentile</b>	614814.86
<b>Maximum</b>	616985.06
<b>Range</b>	20357.121
<b>Interquartile range (IQR)</b>	4908.0572

**Descriptive statistics**

<b>Standard deviation</b>	3852.0245
<b>Coefficient of variation (CV)</b>	0.0063287033
<b>Kurtosis</b>	-0.066028416
<b>Mean</b>	608659.35
<b>Median Absolute Deviation (MAD)</b>	2462.861
<b>Skewness</b>	-0.41317914
<b>Sum</b>	$1.8466116 \times 10^{10}$
<b>Variance</b>	14838093
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
609556.5032	367	0.5%
612552.1674	255	0.3%
604009.418	228	0.3%
609657.7584	205	0.3%
615480.8966	178	0.2%
604848.575	110	0.1%
608539.0792	107	0.1%
612581.1624	106	0.1%

608826.735	100	0.1%
600161.54	100	0.1%
Other values (4675)	28583	36.6%
(Missing)	47694	61.1%

Value	Count	Frequency (%)
596627.9342	2	< 0.1%
596752.9696	2	< 0.1%
597309.0542	3	< 0.1%
597312.632	2	< 0.1%
597772.3526	49	0.1%
597782.4012	2	< 0.1%
597812.404	2	< 0.1%
597933.2448	13	< 0.1%
597963.9396	25	< 0.1%
598104.1884	24	< 0.1%

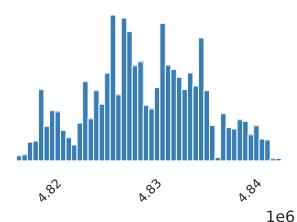
Value	Count	Frequency (%)
616985.0552	9	< 0.1%
616917.8604	1	< 0.1%
616879.86	1	< 0.1%
616836.9092	2	< 0.1%
616794.193	2	< 0.1%
616756.05	2	< 0.1%
616706.7026	2	< 0.1%
616695.363	4	< 0.1%
616668.1574	2	< 0.1%
616652.9546	1	< 0.1%

**CENT\_Y**  
Real number ( $\mathbb{R}$ )

**HIGH CORRELATION** (This variable has a high overall correlation with 2 fields: Location, CENT\_X)  
MISSING

<b>Distinct</b>	4686
<b>Distinct (%)</b>	15.4%
<b>Missing</b>	47694
<b>Missing (%)</b>	61.1%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	4829613.5

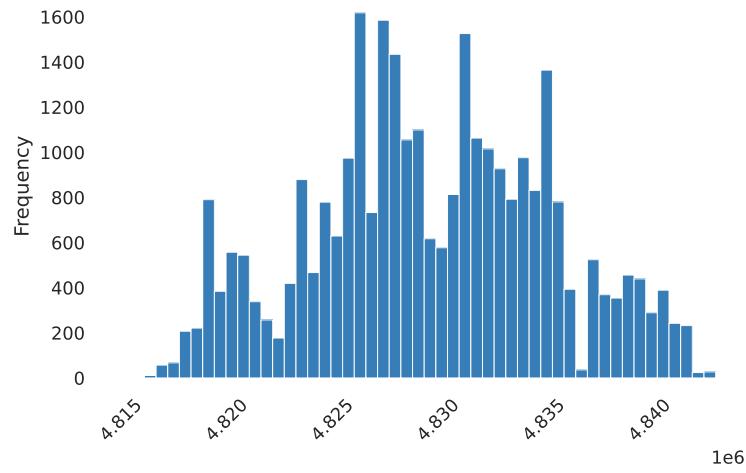
<b>Minimum</b>	4815546.6
<b>Maximum</b>	4843107.8
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	4815546.6
<b>5-th percentile</b>	4819703.7
<b>Q1</b>	4825956.9
<b>median</b>	4829277.7
<b>Q3</b>	4833786.4
<b>95-th percentile</b>	4839313.8
<b>Maximum</b>	4843107.8
<b>Range</b>	27561.198
<b>Interquartile range (IQR)</b>	7829.5472

**Descriptive statistics**

<b>Standard deviation</b>	5660.9074
<b>Coefficient of variation (CV)</b>	0.0011721243
<b>Kurtosis</b>	-0.58959864
<b>Mean</b>	4829613.5
<b>Median Absolute Deviation (MAD)</b>	3923.2536
<b>Skewness</b>	-0.0065033237
<b>Sum</b>	$1.4652564 \times 10^{11}$
<b>Variance</b>	32045872
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
4827620.949	367	0.5%
4837278.362	255	0.3%
4823628.592	228	0.3%
4841687.188	205	0.3%
4827728.859	178	0.2%
4824071.126	110	0.1%
4840485.574	107	0.1%
4831178.774	106	0.1%

4823713.954	100	0.1%
4826202.792	100	0.1%
Other values (4676)	28583	36.6%
(Missing)	47694	61.1%

Value	Count	Frequency (%)
4815546.641	1	< 0.1%
4815609.051	2	< 0.1%
4816109.607	2	< 0.1%
4816333.508	2	< 0.1%
4816381.801	4	< 0.1%
4816389.354	2	< 0.1%
4816462.515	1	< 0.1%
4816663.969	2	< 0.1%
4816718.415	2	< 0.1%
4816760.675	1	< 0.1%

Value	Count	Frequency (%)
4843107.84	19	< 0.1%
4843040.829	2	< 0.1%
4842998.68	2	< 0.1%
4842855.077	2	< 0.1%
4842717.945	2	< 0.1%
4842534.357	2	< 0.1%
4842303.169	5	< 0.1%
4842272.626	2	< 0.1%
4842238.75	2	< 0.1%
4842206.186	4	< 0.1%

**Year**  
Categorical

<b>Distinct</b>	5
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

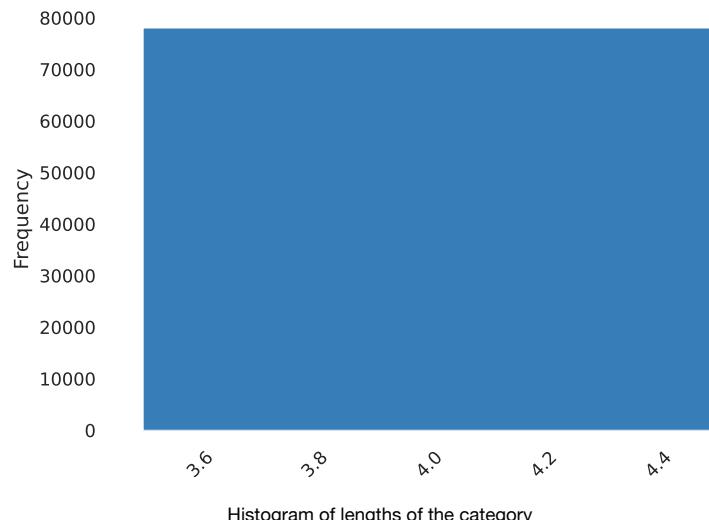
<b>Max length</b>	4
<b>Median length</b>	4
<b>Mean length</b>	4
<b>Min length</b>	4

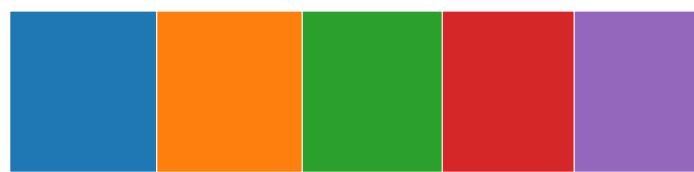
Characters and Unicode		Unique	Sample		
Total characters	312132	Unique	0	?	1st row 2016
Distinct characters	7	Unique (%)	0.0%	?	2nd row 2016
Distinct categories	1			?	3rd row 2016
Distinct scripts	1	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?	?	4th row 2016
Distinct blocks	1	( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?	?	5th row 2016

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

**Common Values**

Value	Count	Frequency (%)
2019	16518	21.2%
2018	16351	21.0%
2017	15737	20.2%
2021	14825	19.0%
2016	14602	18.7%

**Length****Common Values (Plot)**



Value	Count	Frequency (%)
2019	16518	21.2%
2018	16351	21.0%
2017	15737	20.2%
2021	14825	19.0%
2016	14602	18.7%

### Most occurring characters

Value	Count	Frequency (%)
2	92858	29.7%
0	78033	25.0%
1	78033	25.0%
9	16518	5.3%
8	16351	5.2%
7	15737	5.0%
6	14602	4.7%

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	312132	100.0%

### Most frequent character per category

*Decimal Number*

Value	Count	Frequency (%)
2	92858	29.7%

0	78033	25.0%
1	78033	25.0%
9	16518	5.3%
8	16351	5.2%
7	15737	5.0%
6	14602	4.7%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	312132	100.0%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
2	92858	29.7%
0	78033	25.0%
1	78033	25.0%
9	16518	5.3%
8	16351	5.2%
7	15737	5.0%
6	14602	4.7%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	312132	100.0%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
2	92858	29.7%
0	78033	25.0%
1	78033	25.0%
9	16518	5.3%
8	16351	5.2%
7	15737	5.0%

6

14602

4.7%

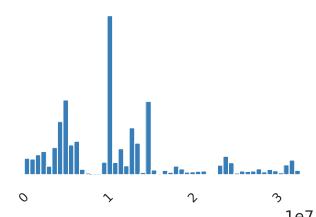
### PIN

Real number ( $\mathbb{R}$ )

**HIGH CORRELATION** (This variable has a high overall correlation with 3 fields: FID, Character, CHArea)  
MISSING

<b>Distinct</b>	4961
<b>Distinct (%)</b>	10.4%
<b>Missing</b>	30339
<b>Missing (%)</b>	38.9%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	11122766

<b>Minimum</b>	32500
<b>Maximum</b>	32656400
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

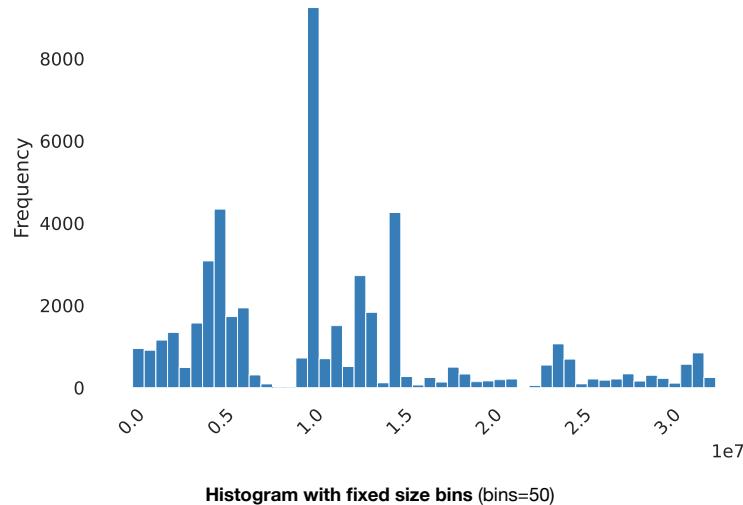


### Quantile statistics

<b>Minimum</b>	32500
<b>5-th percentile</b>	1878100
<b>Q1</b>	5158600
<b>median</b>	10172700
<b>Q3</b>	14774550
<b>95-th percentile</b>	28577700
<b>Maximum</b>	32656400
<b>Range</b>	32623900
<b>Interquartile range (IQR)</b>	9615950

### Descriptive statistics

<b>Standard deviation</b>	7579323.6
<b>Coefficient of variation (CV)</b>	0.68142438
<b>Kurtosis</b>	0.64247043
<b>Mean</b>	11122766
<b>Median Absolute Deviation (MAD)</b>	4630200
<b>Skewness</b>	1.0446214
<b>Sum</b>	$5.3048918 \times 10^{11}$
<b>Variance</b>	$5.7446147 \times 10^{13}$
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
6068300	587	0.8%
31141506	414	0.5%
4407700	328	0.4%
9663800	287	0.4%
12876900	216	0.3%
24265600	190	0.2%
14804200	186	0.2%
31381800	177	0.2%

17704200	161	0.2%
10173700	147	0.2%
Other values (4951)	45001	57.7%
(Missing)	30339	38.9%

Value	Count	Frequency (%)
32500	3	< 0.1%
37200	10	< 0.1%
37300	2	< 0.1%
37400	33	< 0.1%
38100	2	< 0.1%
38300	9	< 0.1%
38400	14	< 0.1%
38500	2	< 0.1%
38600	13	< 0.1%
38700	1	< 0.1%

Value	Count	Frequency (%)
32656400	1	< 0.1%
32646400	44	0.1%
32551400	1	< 0.1%
32526400	2	< 0.1%
32476400	11	< 0.1%
32442000	5	< 0.1%
32441600	2	< 0.1%
32436400	25	< 0.1%
32431500	43	0.1%
32371800	1	< 0.1%

**Character**

Categorical

HIGH CARDINALITY  
 HIGH CORRELATION (This variable has a high overall correlation with 4 fields: FID, BldgNo., NAICSCat, PIN)  
 MISSING

<b>Distinct</b>	56
<b>Distinct (%)</b>	0.3%
<b>Missing</b>	61682
<b>Missing (%)</b>	79.0%
<b>Memory size</b>	609.8 KiB

**Length**

Max length	27
Median length	23
Mean length	16.545777
Min length	7

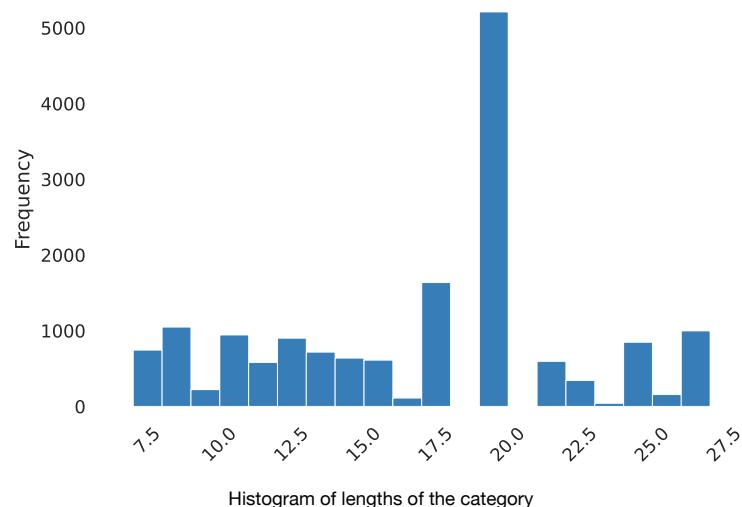
Characters and Unicode		Unique	Sample		
Total characters	270540	<b>Unique</b>	1	?	1st row Cooksville NHD (East)
Distinct characters	43	<b>Unique (%)</b>	< 0.1%		2nd row Rathwood NHD
Distinct categories	6				3rd row Cooksville NHD (East)
Distinct scripts	2			?	4th row Rathwood-Applewood CN
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )			?	5th row Cooksville NHD (East)

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

**Common Values**

Value	Count	Frequency (%)
Northeast EA (West)	4700	6.0%
Dixie EA	1048	1.3%
Gateway EA (East)	1034	1.3%
Meadowvale Business Park CC	998	1.3%
Western Business Park EA	847	1.1%
DT Core	739	0.9%
Airport CC	507	0.6%
Northeast EA (East)	411	0.5%
DT Cooksville	409	0.5%
Mavis-Erindale EA	392	0.5%
Other values (46)	5266	6.7%
(Missing)	61682	79.0%

**Length**



Value	Count	Frequency (%)
ea	8946	19.7%
northeast	5111	11.3%
west	5028	11.1%
nhd	2823	6.2%
park	2036	4.5%
east	1943	4.3%
business	1845	4.1%
cc	1768	3.9%
gateway	1473	3.2%
dt	1330	2.9%
Other values (45)	13072	28.8%

### Most occurring characters

Value	Count	Frequency (%)
e	29024	10.7%
t	24541	9.1%
s	23397	8.6%
a	21055	7.8%
r	17998	6.7%
o	14003	5.2%
E	12440	4.6%
A	11848	4.4%
i	10106	3.7%
	9697	3.6%

Other values (33)	96431	35.6%
-------------------	-------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	162451	60.0%
Uppercase Letter	65050	24.0%
Space Separator	29024	10.7%
Open Punctuation	6677	2.5%
Close Punctuation	6677	2.5%
Dash Punctuation	661	0.2%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
e	24541	15.1%
t	23397	14.4%
s	21055	13.0%
a	17998	11.1%
r	14003	8.6%
o	12440	7.7%
i	9697	6.0%
l	6578	4.0%
h	5996	3.7%
n	5527	3.4%
Other values (11)	21219	13.1%

#### *Uppercase Letter*

Value	Count	Frequency (%)
E	11848	18.2%
A	10106	15.5%
N	9262	14.2%
C	7360	11.3%
W	5875	9.0%
D	5201	8.0%
H	3127	4.8%
M	2865	4.4%

P	2537	3.9%
B	1845	2.8%
Other values (8)	5024	7.7%

*Space Separator*

Value	Count	Frequency (%)
	29024	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	6677	100.0%

*Close Punctuation*

Value	Count	Frequency (%)
)	6677	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	661	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	227501	84.1%
Common	43039	15.9%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
e	24541	10.8%
t	23397	10.3%
s	21055	9.3%
a	17998	7.9%
r	14003	6.2%
o	12440	5.5%
E	11848	5.2%
A	10106	4.4%
i	9697	4.3%

N	9262	4.1%
Other values (29)	73154	32.2%

**Common**

Value	Count	Frequency (%)
	29024	67.4%
(	6677	15.5%
)	6677	15.5%
-	661	1.5%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	270540	100.0%

**Most frequent character per block**

ASCII

Value	Count	Frequency (%)
e	29024	10.7%
t	24541	9.1%
s	23397	8.6%
a	21055	7.8%
r	17998	6.7%
o	14003	5.2%
E	12440	4.6%
A	11848	4.4%
i	10106	3.7%
Other values (33)	96431	35.6%

**CHArea**  
Categorical  
  
HIGH CARDINALITY  
HIGH CORRELATION (This variable has a high overall correlation with 6 fields: FID, BldgNo., NAICSCat, PIN, BIA\_NAME, BIAFullName).  
MISSING

<b>Distinct</b>	57
<b>Distinct (%)</b>	0.2%
<b>Missing</b>	46690
<b>Missing (%)</b>	59.8%
<b>Memory size</b>	609.8 KiB

<b>Length</b>	
<b>Max length</b>	27
<b>Median length</b>	23
<b>Mean length</b>	16.534633
<b>Min length</b>	7

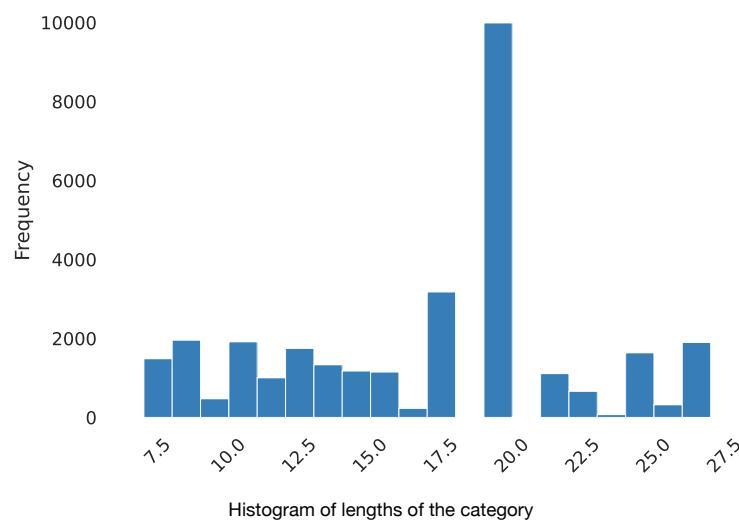
<b>Characters and Unicode</b>		<b>Unique</b>	<b>Sample</b>		
<b>Total characters</b>	518245	<b>Unique</b>	0	?	<b>1st row</b> Northeast EA (West)
<b>Distinct characters</b>	44	<b>Unique (%)</b>	0.0%		<b>2nd row</b> DT Core
<b>Distinct categories</b>	6				<b>3rd row</b> Northeast EA (West)
<b>Distinct scripts</b>	2			?	<b>4th row</b> DT Core
<b>Distinct blocks</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )			<b>5th row</b> DT Core

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

### Common Values

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
Northeast EA (West)	8989	11.5%
Gateway EA (East)	1975	2.5%
Dixie EA	1955	2.5%
Meadowvale Business Park CC	1898	2.4%
Western Business Park EA	1636	2.1%
DT Core	1477	1.9%
Airport CC	996	1.3%
Northeast EA (East)	804	1.0%
Mavis-Erindale EA	784	1.0%
DT Cooksville	724	0.9%
Other values (47)	10105	12.9%
(Missing)	46690	59.8%

### Length



Value	Count	Frequency (%)
ea	17070	19.6%
northeast	9793	11.3%
west	9630	11.1%
nhd	5337	6.1%
park	3923	4.5%
east	3694	4.2%
business	3534	4.1%
cc	3445	4.0%
gateway	2875	3.3%
dt	2519	2.9%
Other values (48)	25104	28.9%

### Most occurring characters

Value	Count	Frequency (%)
e	55581	10.7%
t	47046	9.1%
s	44934	8.7%
a	40159	7.7%
r	34746	6.7%
o	27014	5.2%
E	23860	4.6%
A	22566	4.4%
i	19328	3.7%
l	18277	3.5%

Other values (34)	184734	35.6%
-------------------	--------	-------

### Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	311221	60.1%
Uppercase Letter	124590	24.0%
Space Separator	55581	10.7%
Close Punctuation	12769	2.5%
Open Punctuation	12769	2.5%
Dash Punctuation	1315	0.3%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
e	47046	15.1%
t	44934	14.4%
s	40159	12.9%
a	34746	11.2%
r	27014	8.7%
o	23860	7.7%
i	18277	5.9%
l	12367	4.0%
h	11504	3.7%
n	10684	3.4%
Other values (12)	40630	13.1%

#### *Uppercase Letter*

Value	Count	Frequency (%)
E	22566	18.1%
A	19328	15.5%
N	17803	14.3%
C	14198	11.4%
W	11322	9.1%
D	9811	7.9%
H	5878	4.7%
M	5574	4.5%

P	4873	3.9%
B	3534	2.8%
Other values (8)	9703	7.8%

*Space Separator*

Value	Count	Frequency (%)
	55581	100.0%

*Close Punctuation*

Value	Count	Frequency (%)
)	12769	100.0%

*Open Punctuation*

Value	Count	Frequency (%)
(	12769	100.0%

*Dash Punctuation*

Value	Count	Frequency (%)
-	1315	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	435811	84.1%
Common	82434	15.9%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
e	47046	10.8%
t	44934	10.3%
s	40159	9.2%
a	34746	8.0%
r	27014	6.2%
o	23860	5.5%
E	22566	5.2%
A	19328	4.4%
i	18277	4.2%

N	17803	4.1%
Other values (30)	140078	32.1%

**Common**

Value	Count	Frequency (%)
)	55581	67.4%
(	12769	15.5%
-	12769	15.5%
-	1315	1.6%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	518245	100.0%

**Most frequent character per block****ASCII**

Value	Count	Frequency (%)
e	55581	10.7%
t	47046	9.1%
s	44934	8.7%
a	40159	7.7%
r	34746	6.7%
o	27014	5.2%
E	23860	4.6%
A	22566	4.4%
i	19328	3.7%
Other values (34)	184734	35.6%

**Modified**

Categorical

HIGH CARDINALITY  
MISSING

<b>Distinct</b>	189
<b>Distinct (%)</b>	1.3%
<b>Missing</b>	63218
<b>Missing (%)</b>	81.0%
<b>Memory size</b>	609.8 KiB

**Length**

<b>Max length</b>	22
<b>Median length</b>	22
<b>Mean length</b>	22
<b>Min length</b>	22

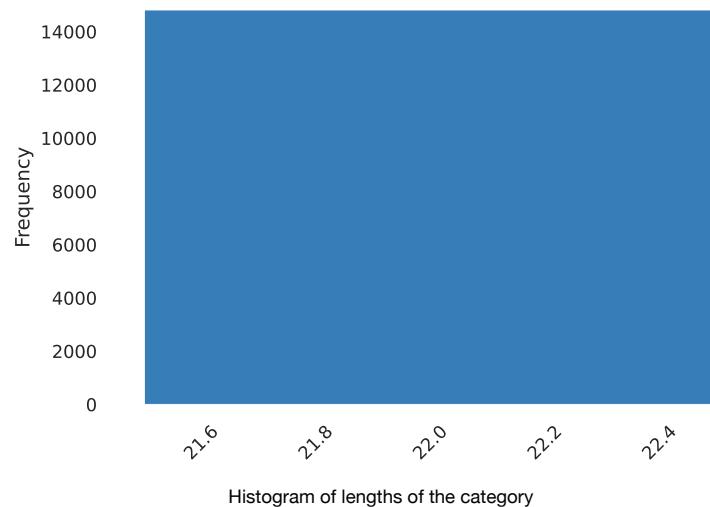
Characters and Unicode		Unique		Sample		
<b>Total characters</b>	325930	<b>Unique</b>	50	?	<b>1st row</b>	2021/06/25 00:00:00+00
<b>Distinct characters</b>	14	<b>Unique (%)</b>	0.3%		<b>2nd row</b>	2021/06/03 00:00:00+00
<b>Distinct categories</b>	4				<b>3rd row</b>	2021/07/15 00:00:00+00
<b>Distinct scripts</b>	1			?	<b>4th row</b>	2021/07/15 00:00:00+00
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )			?	<b>5th row</b>	2021/07/15 00:00:00+00

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

**Common Values**

Value	Count	Frequency (%)
2018/12/30 00:00:00+00	2771	3.6%
2019/12/12 00:00:00+00	1848	2.4%
2019/09/19 00:00:00+00	1586	2.0%
2017/11/09 00:00:00+00	1111	1.4%
2017/11/08 00:00:00+00	968	1.2%
2021/07/02 00:00:00+00	354	0.5%
2019/06/07 00:00:00+00	267	0.3%
2021/05/21 00:00:00+00	186	0.2%
2018/09/30 00:00:00+00	177	0.2%
2021/05/17 00:00:00+00	168	0.2%
Other values (179)	5379	6.9%
(Missing)	63218	81.0%

**Length**



Value	Count	Frequency (%)
00:00:00+00	14815	50.0%
2018/12/30	2771	9.4%
2019/12/12	1848	6.2%
2019/09/19	1586	5.4%
2017/11/09	1111	3.7%
2017/11/08	968	3.3%
2021/07/02	354	1.2%
2019/06/07	267	0.9%
2021/05/21	186	0.6%
2018/09/30	177	0.6%
Other values (180)	5547	18.7%

### Most occurring characters

Value	Count	Frequency (%)
0	148805	45.7%
1	29895	9.2%
/	29630	9.1%
:	29630	9.1%
2	29181	9.0%
	14815	4.5%
+	14815	4.5%
9	8963	2.7%
7	6006	1.8%
8	5090	1.6%

Other values (4)	9100	2.8%
------------------	------	------

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	237040	72.7%
Other Punctuation	59260	18.2%
Space Separator	14815	4.5%
Math Symbol	14815	4.5%

### Most frequent character per category

#### *Decimal Number*

Value	Count	Frequency (%)
0	148805	62.8%
1	29895	12.6%
2	29181	12.3%
9	8963	3.8%
7	6006	2.5%
8	5090	2.1%
3	3797	1.6%
6	2508	1.1%
5	2286	1.0%
4	509	0.2%

#### *Other Punctuation*

Value	Count	Frequency (%)
/	29630	50.0%
:	29630	50.0%

#### *Space Separator*

Value	Count	Frequency (%)
	14815	100.0%

#### *Math Symbol*

Value	Count	Frequency (%)
+	14815	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	325930	100.0%

**Most frequent character per script***Common*

Value	Count	Frequency (%)
0	148805	45.7%
1	29895	9.2%
/	29630	9.1%
:	29630	9.1%
2	29181	9.0%
	14815	4.5%
+	14815	4.5%
9	8963	2.7%
7	6006	1.8%
8	5090	1.6%
Other values (4)	9100	2.8%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	325930	100.0%

**Most frequent character per block***ASCII*

Value	Count	Frequency (%)
0	148805	45.7%
1	29895	9.2%
/	29630	9.1%
:	29630	9.1%
2	29181	9.0%
	14815	4.5%
+	14815	4.5%
9	8963	2.7%
7	6006	1.8%

8	5090	1.6%
Other values (4)	9100	2.8%

**BIA\_NAME**

Categorical

**HIGH CORRELATION** (This variable has a high overall correlation with 3 fields: FID, CHArea, BIAfName)  
MISSING

<b>Distinct</b>	6
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	63208
<b>Missing (%)</b>	81.0%
<b>Memory size</b>	609.8 KiB

**Length**

	Max length	3
	Median length	1
	Mean length	1.1399663
	Min length	1

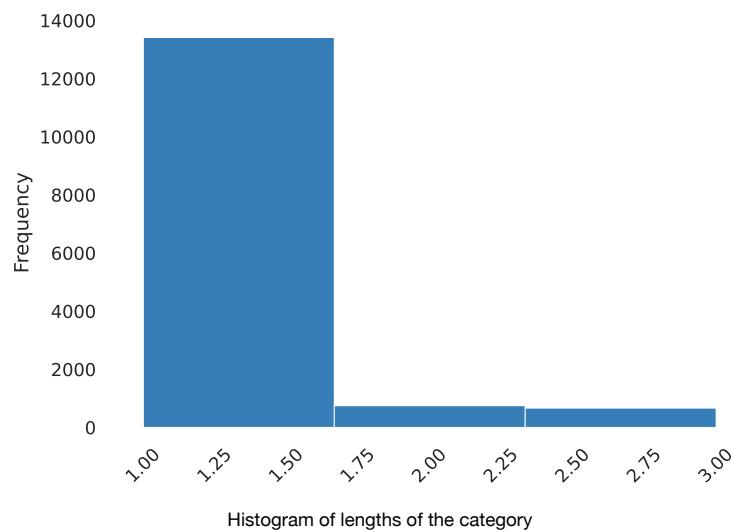
	Characters and Unicode	Unique		Sample
<b>Total characters</b>	16900	<b>Unique</b>	0 ?	<b>1st row</b>
<b>Distinct characters</b>	10	<b>Unique (%)</b>	0.0%	<b>2nd row</b>
<b>Distinct categories</b>	2			<b>3rd row</b>
<b>Distinct scripts</b>	2		?	<b>4th row</b>
<b>Distinct blocks</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?	<b>5th row</b>

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

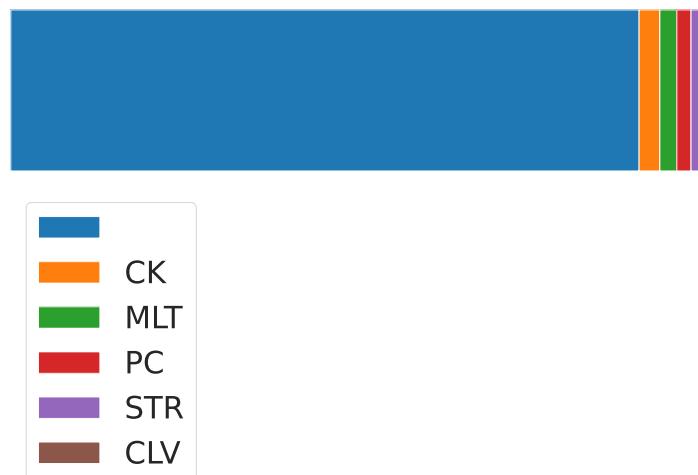
**Common Values**

Value	Count	Frequency (%)
CK	13414	17.2%
MLT	443	0.6%
PC	362	0.5%
STR	304	0.4%
CLV	215	0.3%
(Missing)	87	0.1%
	63208	81.0%

**Length**



### Common Values (Plot)



Value	Count	Frequency (%)
ck	443	31.4%
mlt	362	25.7%
pc	304	21.5%
str	215	15.2%
clv	87	6.2%

### Most occurring characters

Value	Count	Frequency (%)
C	13414	79.4%
T	834	4.9%
L	577	3.4%
K	449	2.7%
	443	2.6%

M	362	2.1%
P	304	1.8%
S	215	1.3%
R	215	1.3%
V	87	0.5%

**Most occurring categories**

Value	Count	Frequency (%)
Space Separator	13414	79.4%
Uppercase Letter	3486	20.6%

**Most frequent character per category***Uppercase Letter*

Value	Count	Frequency (%)
C	834	23.9%
T	577	16.6%
L	449	12.9%
K	443	12.7%
M	362	10.4%
P	304	8.7%
S	215	6.2%
R	215	6.2%
V	87	2.5%

*Space Separator*

Value	Count	Frequency (%)
	13414	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Common	13414	79.4%
Latin	3486	20.6%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
C	834	23.9%
T	577	16.6%
L	449	12.9%
K	443	12.7%
M	362	10.4%
P	304	8.7%
S	215	6.2%
R	215	6.2%
V	87	2.5%

**Common**

Value	Count	Frequency (%)
	13414	100.0%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	16900	100.0%

**Most frequent character per block**

ASCII

Value	Count	Frequency (%)
	13414	79.4%
C	834	4.9%
T	577	3.4%
L	449	2.7%
K	443	2.6%
M	362	2.1%
P	304	1.8%
S	215	1.3%
R	215	1.3%
V	87	0.5%

**BIAFullName**

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: FID, CHArea, BIA\_NAME)  
MISSING

<b>Distinct</b>	6
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	63208
<b>Missing (%)</b>	81.0%
<b>Memory size</b>	609.8 KiB

**Length**

	Max length	16
	Median length	1
	Mean length	2.177403
	Min length	1

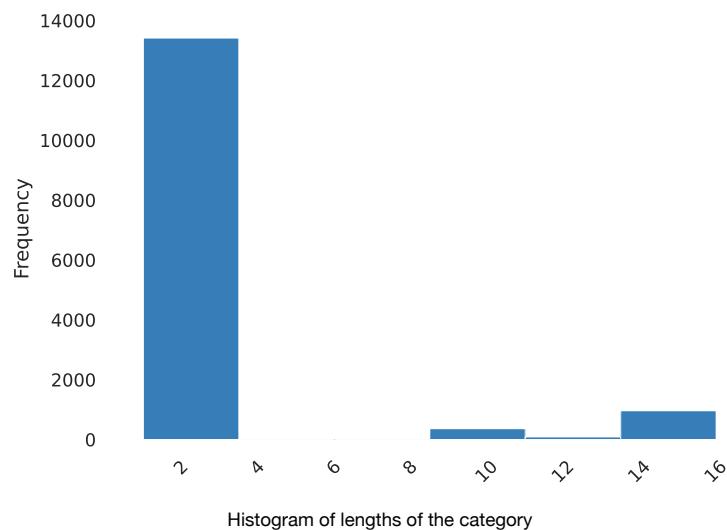
	Characters and Unicode	Unique		Sample
<b>Total characters</b>	32280	<b>Unique</b>	0 ?	<b>1st row</b>
<b>Distinct characters</b>	20	<b>Unique (%)</b>	0.0%	<b>2nd row</b>
<b>Distinct categories</b>	3			<b>3rd row</b>
<a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">Distinct categories (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)</a>				<b>4th row</b>
<b>Distinct scripts</b>	2		?	<a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">5th row</a>
<b>Distinct blocks</b>	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.

**Common Values**

Value	Count	Frequency (%)
	13414	17.2%
Cooksville BIA	443	0.6%
Malton BIA	362	0.5%
Port Credit BIA	304	0.4%
Streetsville BIA	215	0.3%
Clarkson BIA	87	0.1%
(Missing)	63208	81.0%

**Length**



### Common Values (Plot)



Value	Count	Frequency (%)
bia	1411	45.1%
cooksville	443	14.2%
malton	362	11.6%
port	304	9.7%
credit	304	9.7%
streetsville	215	6.9%
clarkson	87	2.8%

### Most occurring characters

Value	Count	Frequency (%)
I	15129	46.9%
O	1765	5.5%
o	1639	5.1%

A	1411	4.4%
B	1411	4.4%
I	1411	4.4%
t	1400	4.3%
e	1392	4.3%
i	962	3.0%
r	910	2.8%
Other values (10)	4850	15.0%

### Most occurring categories

Value	Count	Frequency (%)
Space Separator	15129	46.9%
Lowercase Letter	11203	34.7%
Uppercase Letter	5948	18.4%

### Most frequent character per category

#### *Lowercase Letter*

Value	Count	Frequency (%)
l	1765	15.8%
o	1639	14.6%
t	1400	12.5%
e	1392	12.4%
i	962	8.6%
r	910	8.1%
s	745	6.7%
v	658	5.9%
k	530	4.7%
a	449	4.0%
Other values (2)	753	6.7%

#### *Uppercase Letter*

Value	Count	Frequency (%)
A	1411	23.7%
B	1411	23.7%
I	1411	23.7%
C	834	14.0%

M	362	6.1%
P	304	5.1%
S	215	3.6%

**Space Separator**

Value	Count	Frequency (%)
	15129	100.0%

**Most occurring scripts**

Value	Count	Frequency (%)
Latin	17151	53.1%
Common	15129	46.9%

**Most frequent character per script***Latin*

Value	Count	Frequency (%)
I	1765	10.3%
O	1639	9.6%
A	1411	8.2%
B	1411	8.2%
L	1411	8.2%
T	1400	8.2%
E	1392	8.1%
I	962	5.6%
R	910	5.3%
C	834	4.9%
Other values (9)	4016	23.4%

*Common*

Value	Count	Frequency (%)
	15129	100.0%

**Most occurring blocks**

Value	Count	Frequency (%)
ASCII	32280	100.0%

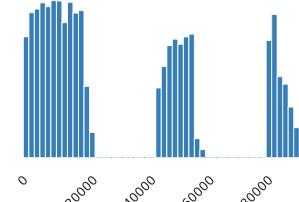
**Most frequent character per block****ASCII**

Value	Count	Frequency (%)
	15129	46.9%
I	1765	5.5%
O	1639	5.1%
A	1411	4.4%
B	1411	4.4%
I	1411	4.4%
t	1400	4.3%
e	1392	4.3%
i	962	3.0%
r	910	2.8%
Other values (10)	4850	15.0%

**RecordID**  
Real number ( $\mathbb{R}$ )

<b>Distinct</b>	21240
<b>Distinct (%)</b>	27.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	34656.92

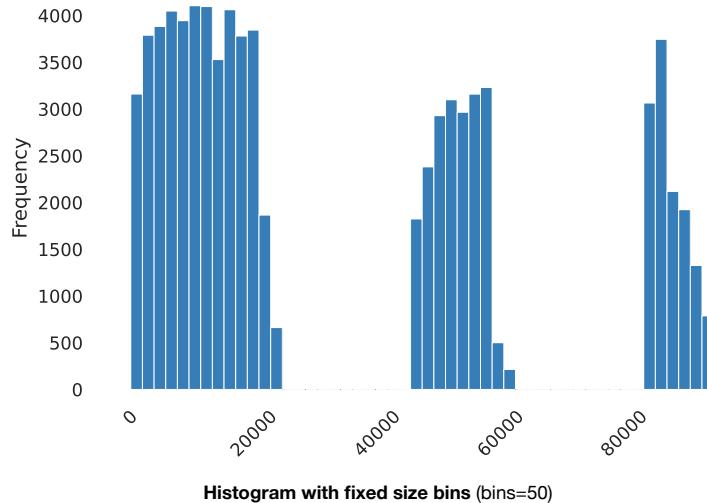
<b>Minimum</b>	2
<b>Maximum</b>	94424
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Quantile statistics**

<b>Minimum</b>	2
<b>5-th percentile</b>	2230
<b>Q1</b>	9764
<b>median</b>	19183
<b>Q3</b>	55026
<b>95-th percentile</b>	88915
<b>Maximum</b>	94424
<b>Range</b>	94422
<b>Interquartile range (IQR)</b>	45262

**Descriptive statistics**

<b>Standard deviation</b>	29857.678
<b>Coefficient of variation (CV)</b>	0.8615214
<b>Kurtosis</b>	-0.9937126
<b>Mean</b>	34656.92
<b>Median Absolute Deviation (MAD)</b>	16020
<b>Skewness</b>	0.65053975
<b>Sum</b>	$2.7043834 \times 10^9$
<b>Variance</b>	$8.9148093 \times 10^8$
<b>Monotonicity</b>	Not monotonic



Value	Count	Frequency (%)
85606	6	< 0.1%
1055	5	< 0.1%
19338	5	< 0.1%
19580	5	< 0.1%
20871	5	< 0.1%
19831	5	< 0.1%
19332	5	< 0.1%
19583	5	< 0.1%

19832	5	< 0.1%
19584	5	< 0.1%
Other values (21230)	77982	99.9%

Value	Count	Frequency (%)
2	2	< 0.1%
7	5	< 0.1%
10	5	< 0.1%
12	3	< 0.1%
16	5	< 0.1%
18	5	< 0.1%
20	5	< 0.1%
21	5	< 0.1%
23	5	< 0.1%
26	4	< 0.1%

Value	Count	Frequency (%)
94424	1	< 0.1%
94423	1	< 0.1%
94419	1	< 0.1%
94371	1	< 0.1%
94321	1	< 0.1%
94319	1	< 0.1%
94318	1	< 0.1%
94317	1	< 0.1%
94313	1	< 0.1%
94293	1	< 0.1%

**Closed**  
Categorical

<b>Distinct</b>	1
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	609.8 KiB

**Length**

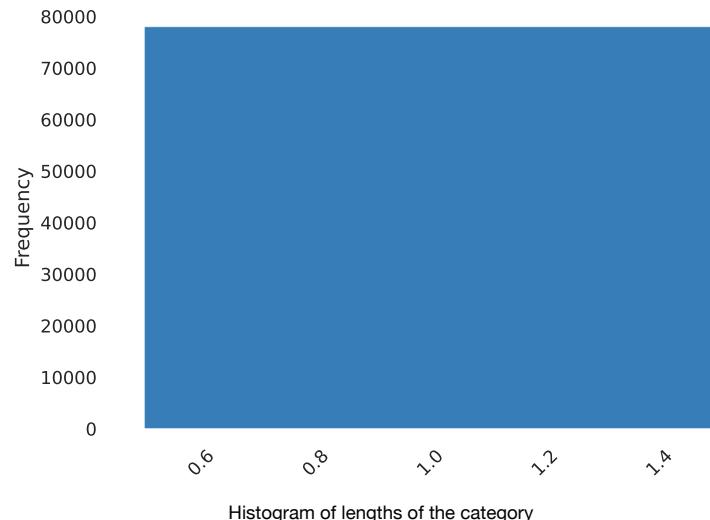
<b>Max length</b>	1
<b>Median length</b>	1
<b>Mean length</b>	1
<b>Min length</b>	1

	<b>Characters and Unicode</b>		<b>Unique</b>			<b>Sample</b>
<b>Total characters</b>	78033		<b>Unique</b>	0	?	<b>1st row</b>
<b>Distinct characters</b>	1		<b>Unique (%)</b>	0.0%		<b>2nd row</b>
<b>Distinct categories</b>	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)					<b>3rd row</b>
<b>Distinct scripts</b>	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)				?	<b>4th row</b>
<b>Distinct blocks</b>	1 (https://en.wikipedia.org/wiki/Unicode_block)				?	<b>5th row</b>

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

**Common Values**

		<b>Count</b>	<b>Frequency (%)</b>
0		78033	100.0%

**Length****Common Values (Plot)**



0

Value	Count	Frequency (%)
0	78033	100.0%

### Most occurring characters

Value	Count	Frequency (%)
0	78033	100.0%

### Most occurring categories

Value	Count	Frequency (%)
Decimal Number	78033	100.0%

### Most frequent character per category

Decimal Number	Value	Count	Frequency (%)
0	78033	100.0%	

### Most occurring scripts

Common	Value	Count	Frequency (%)
0	78033	100.0%	

### Most frequent character per script

Common	Value	Count	Frequency (%)
0	78033	100.0%	

### Most occurring blocks

Value	Count	Frequency (%)
0	78033	100.0%

ASCII

78033 100.0%

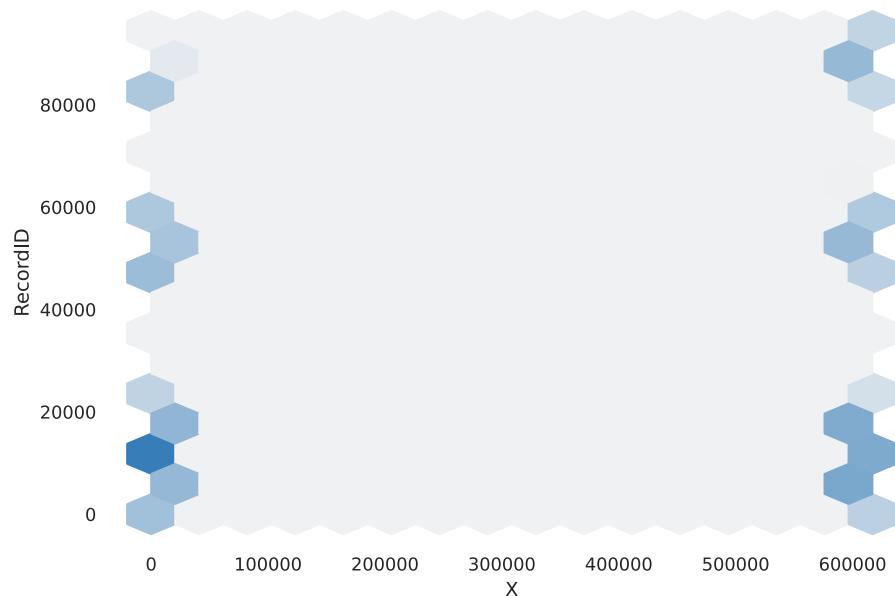
**Most frequent character per block***ASCII*

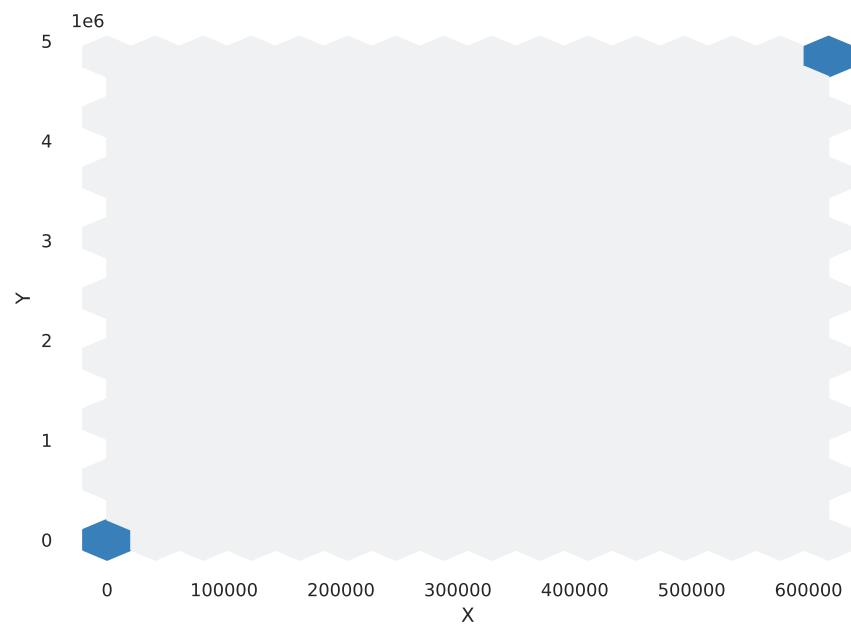
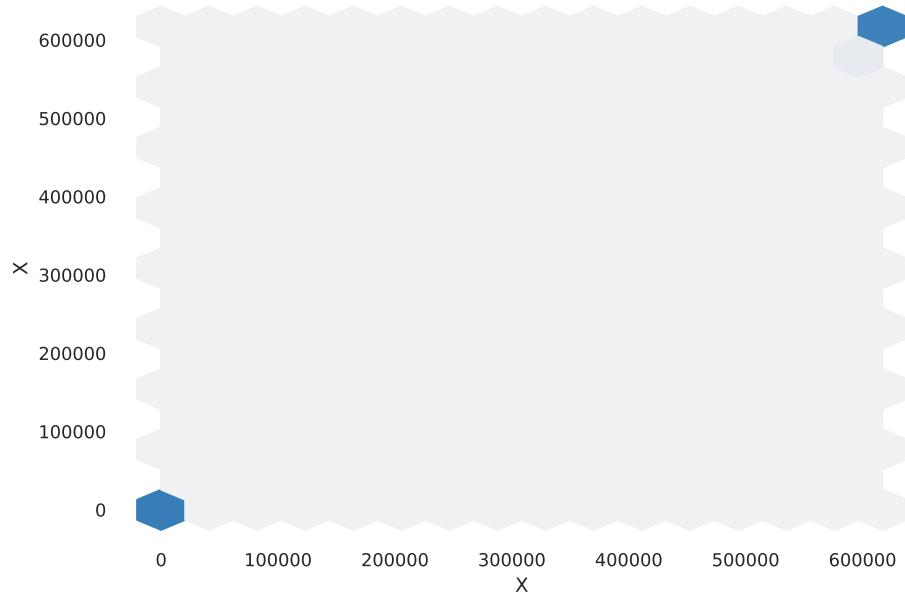
Value	Count	Frequency (%)
0	78033	100.0%

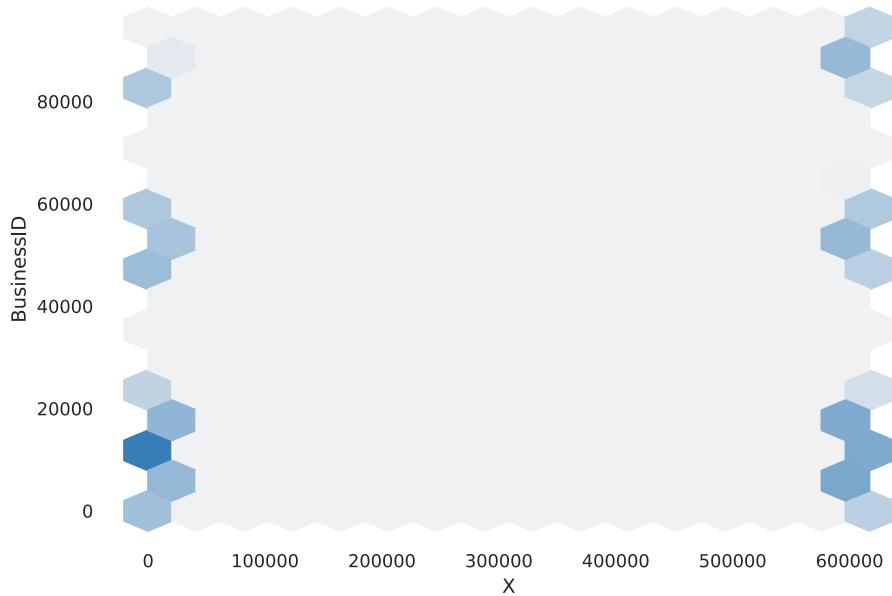
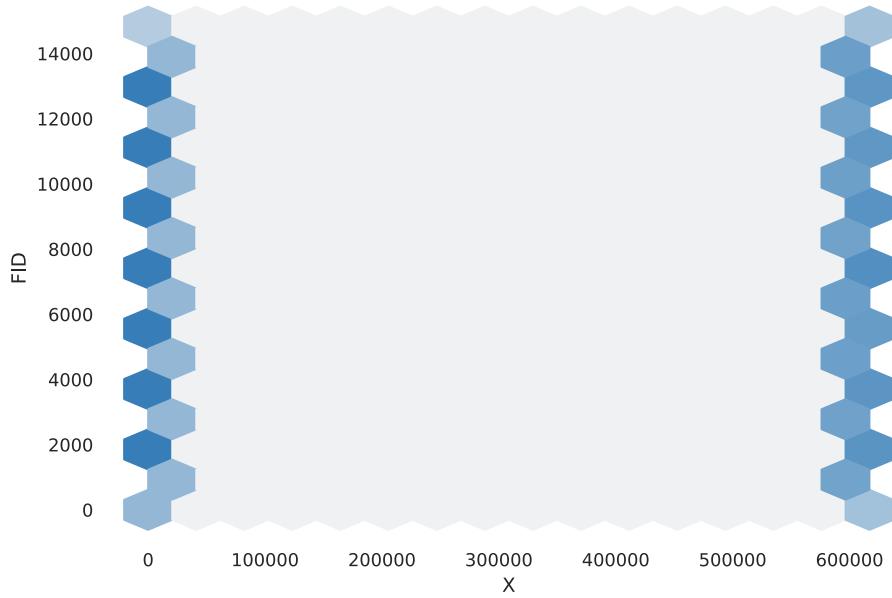
isnew  
Unsupported  
REJECTED  
UNSUPPORTED

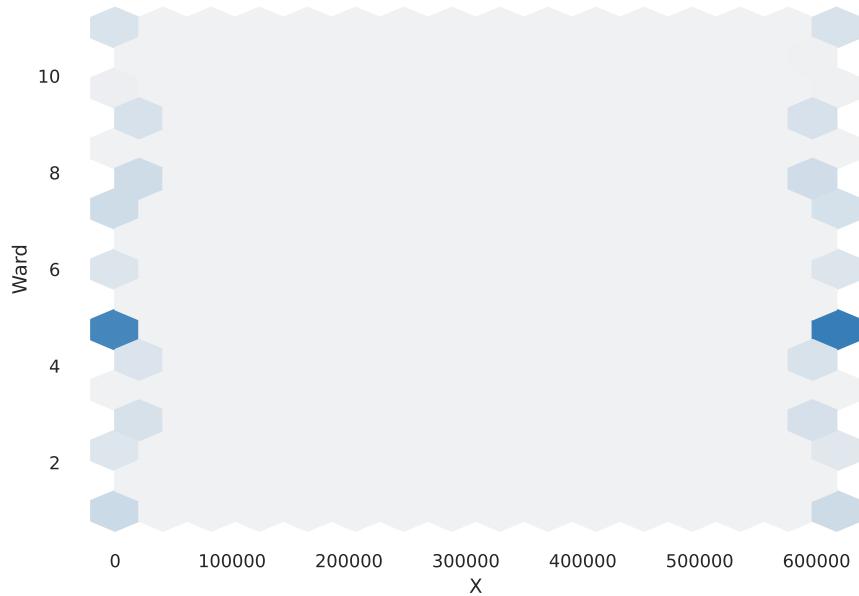
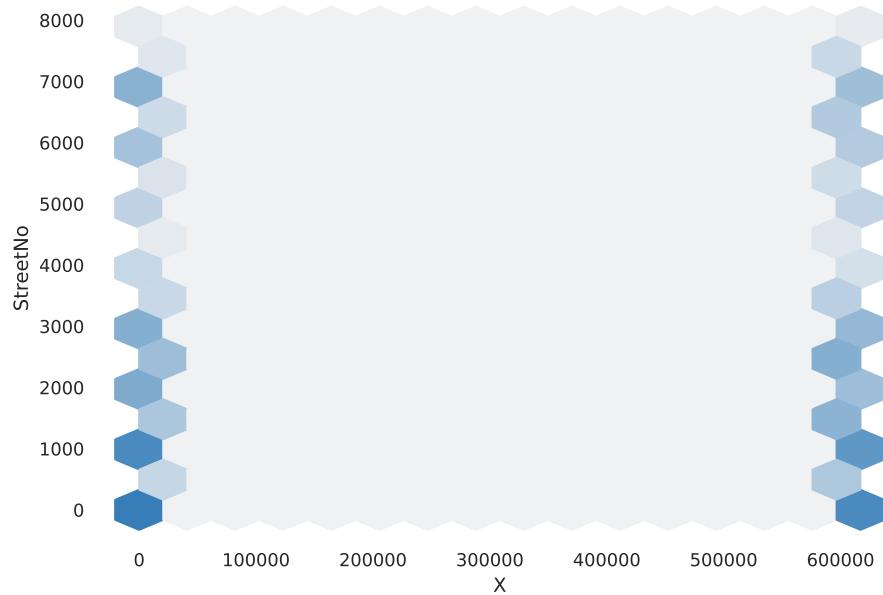
Missing	0
Missing (%)	0.0%
Memory size	609.8 KiB

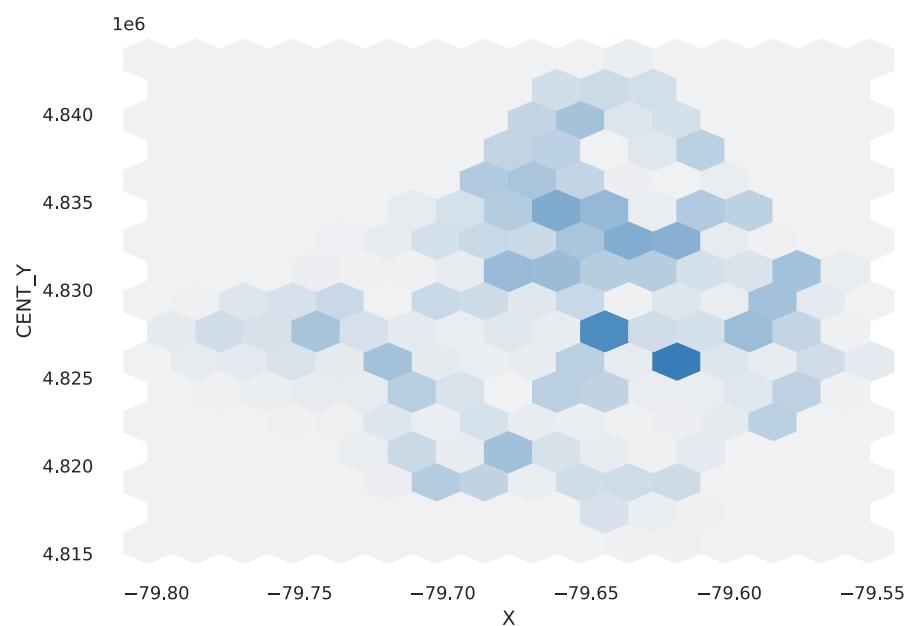
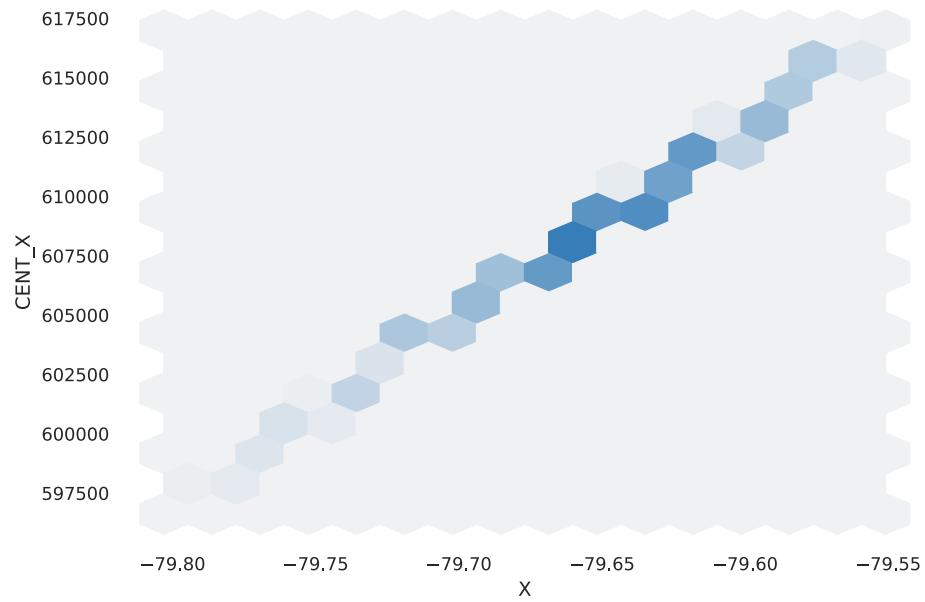
## Interactions

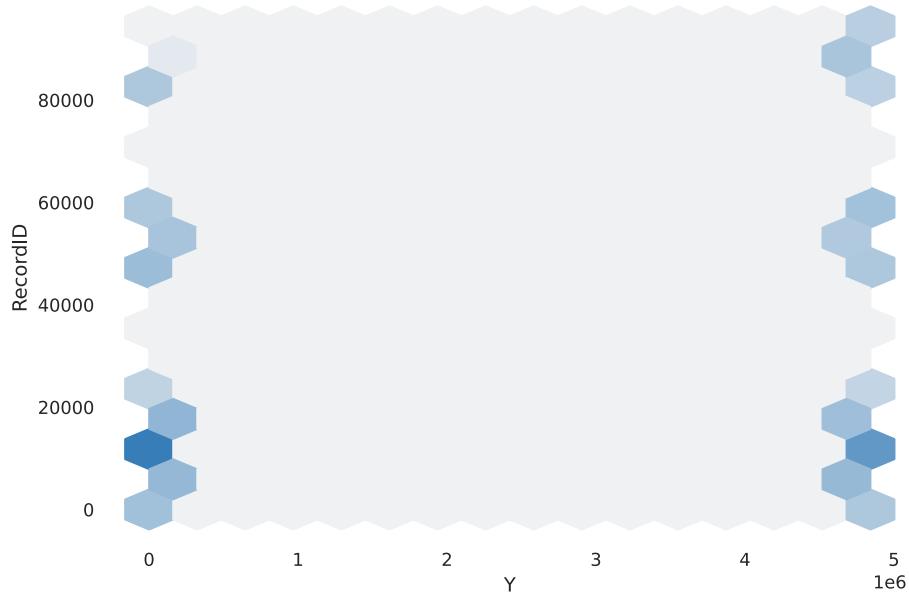
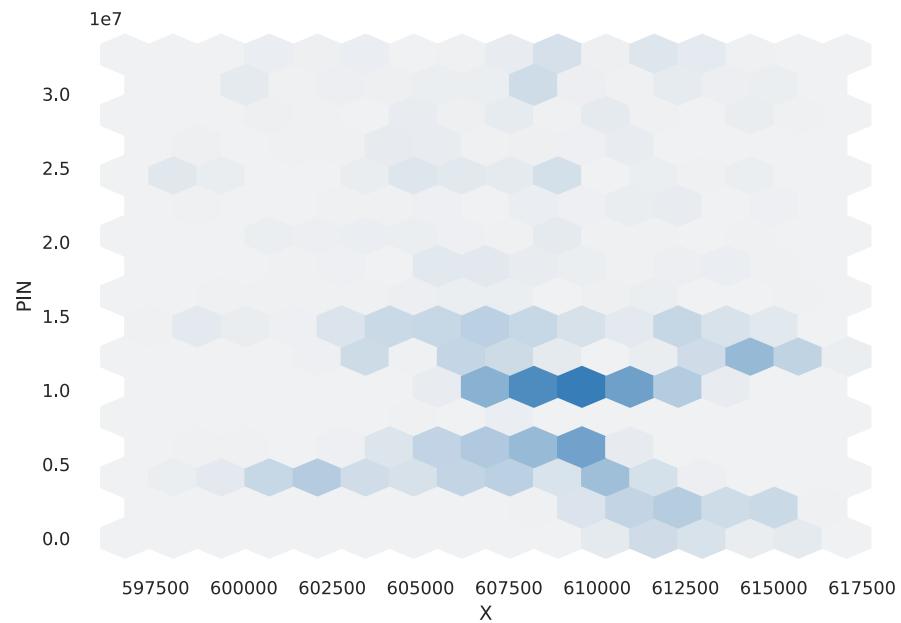




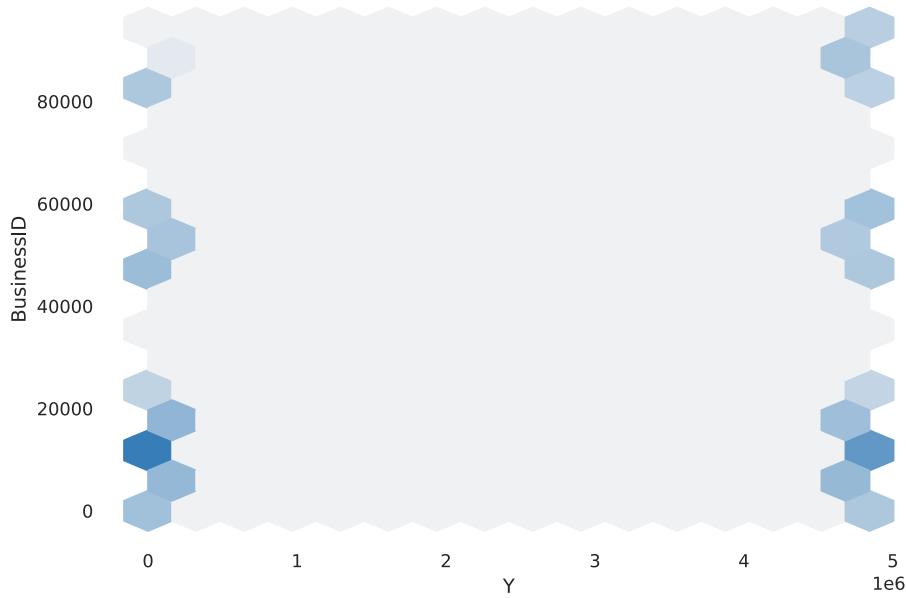
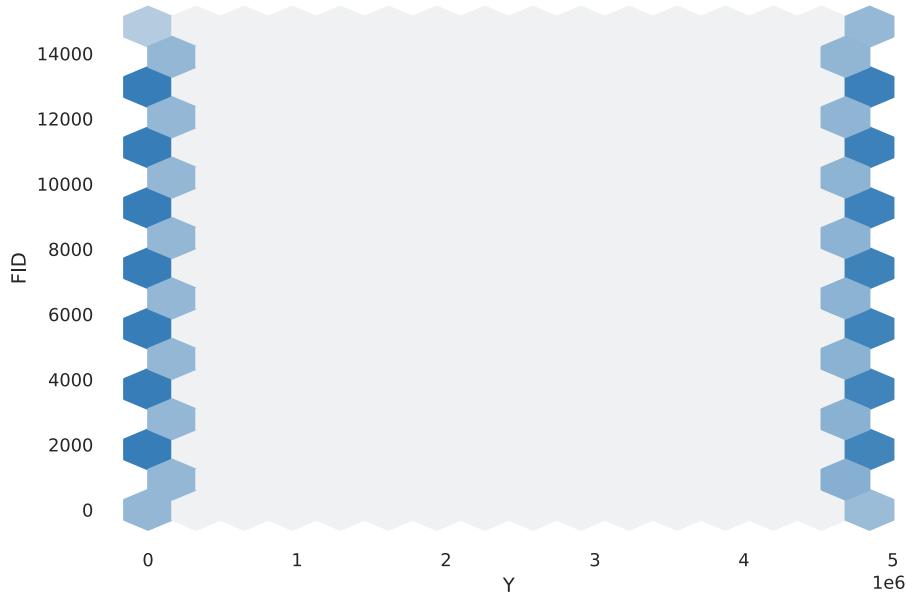


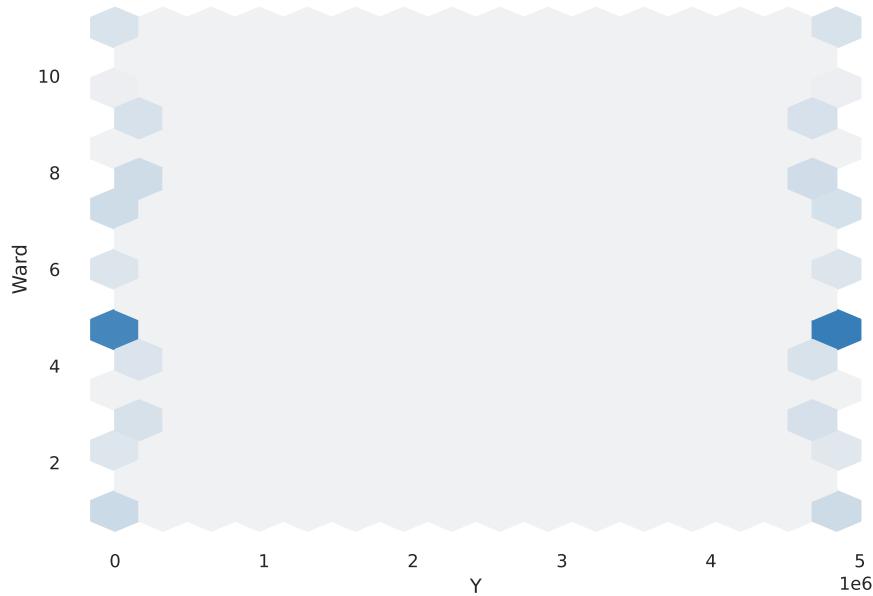
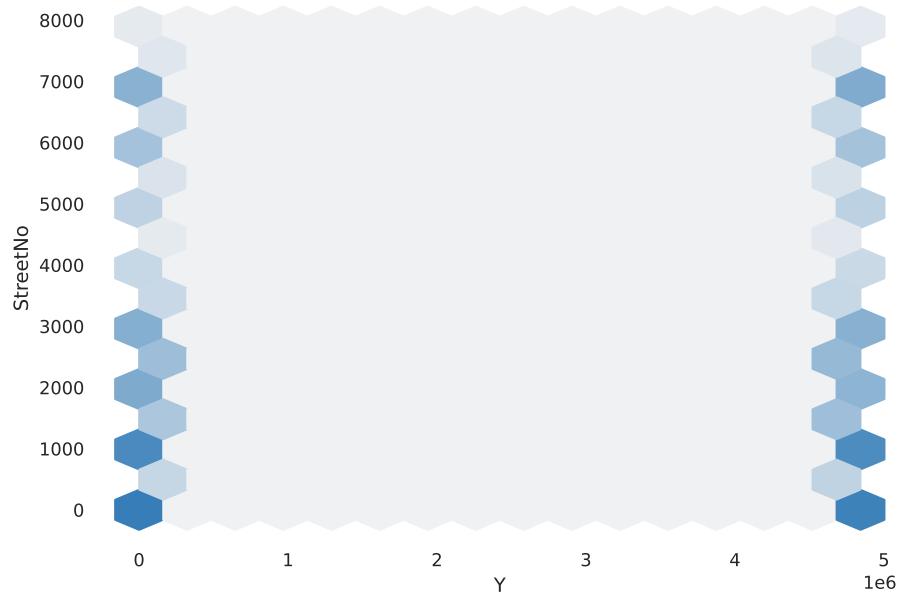


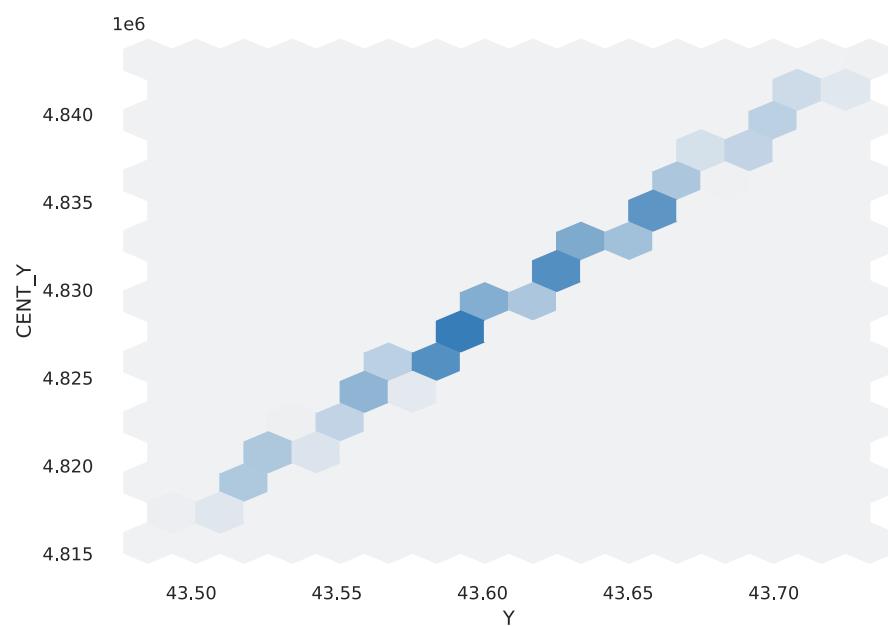
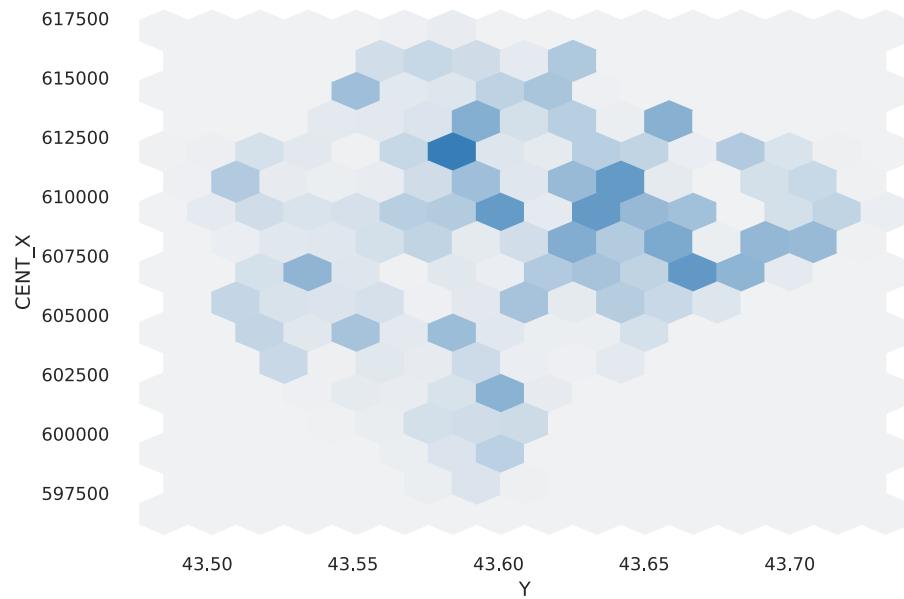


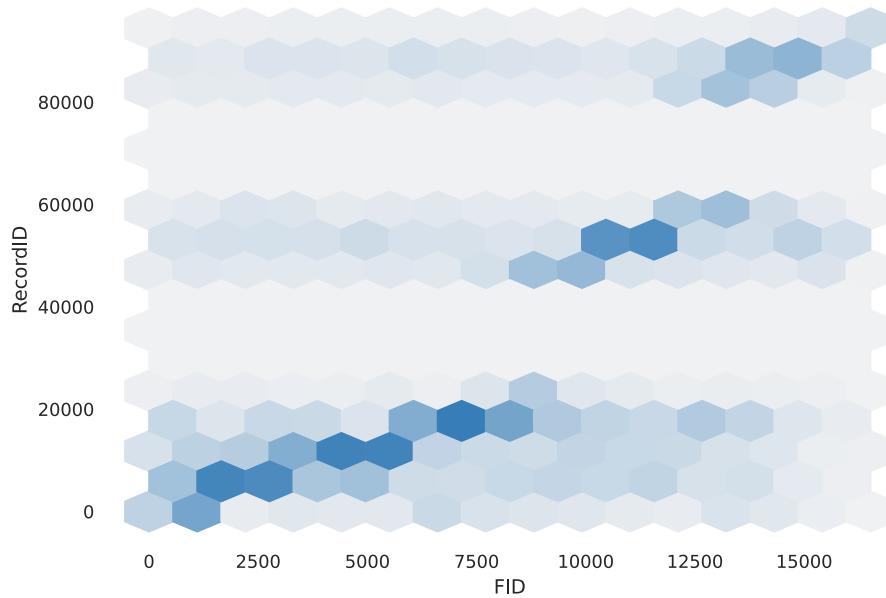
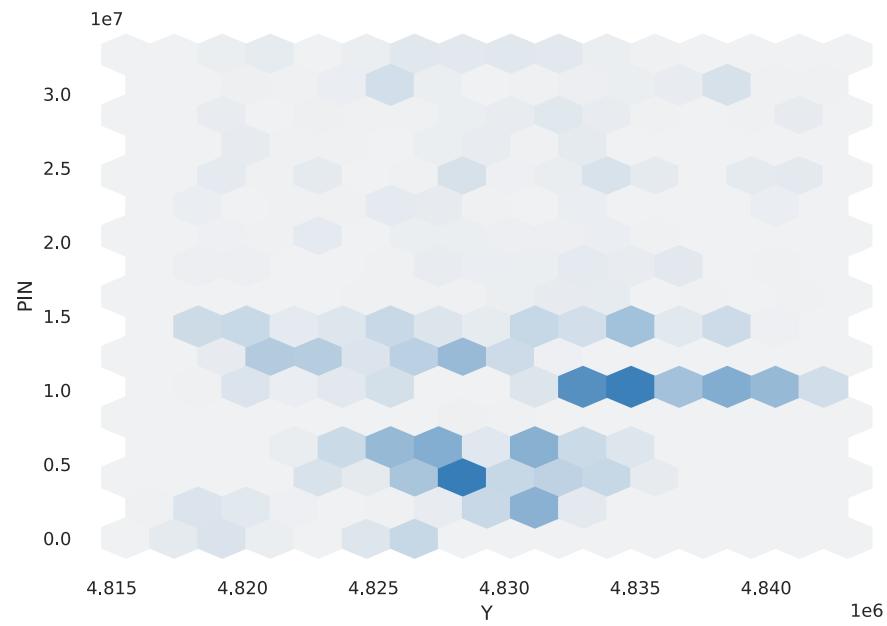


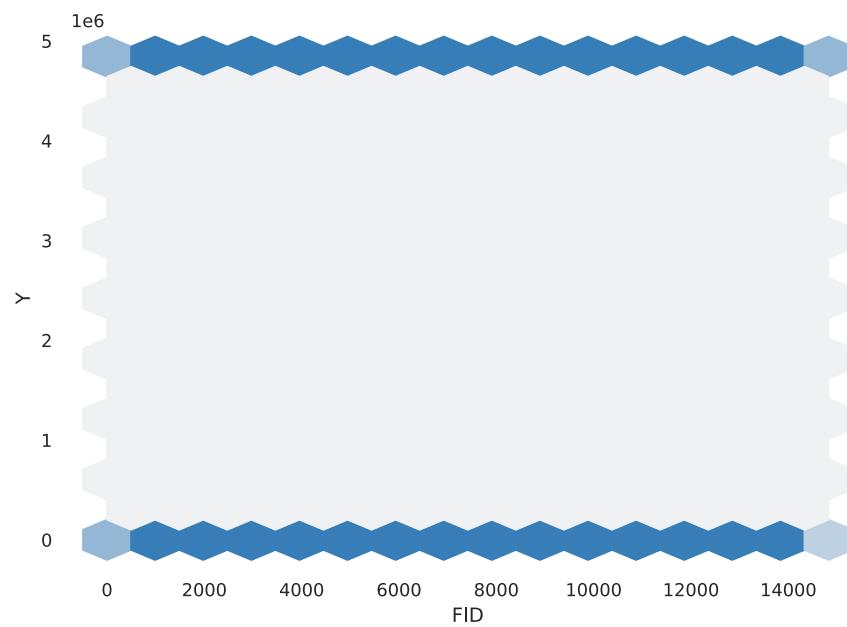
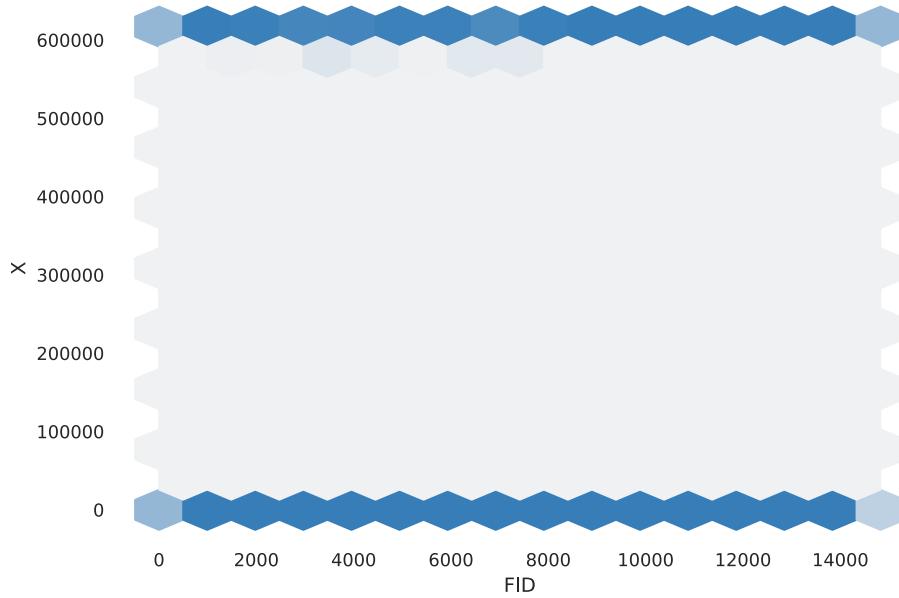


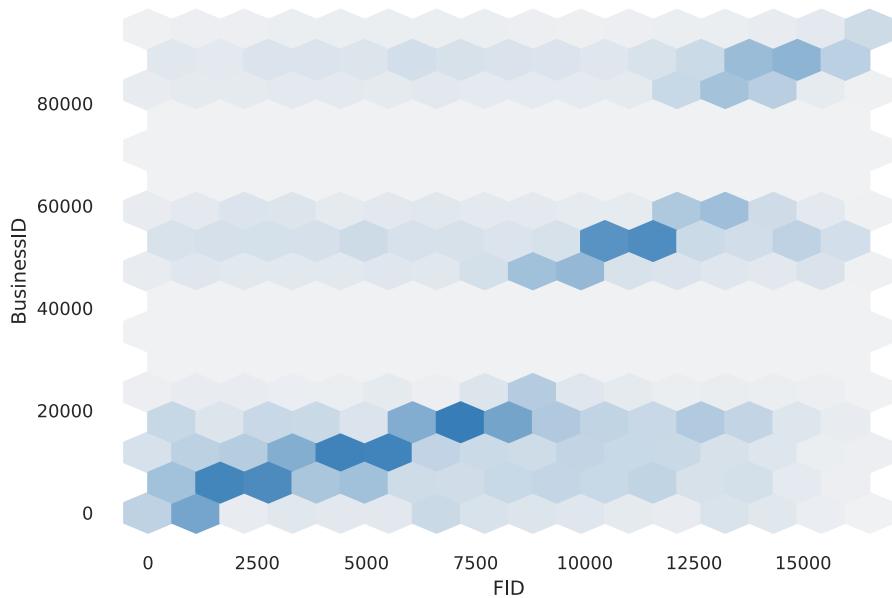
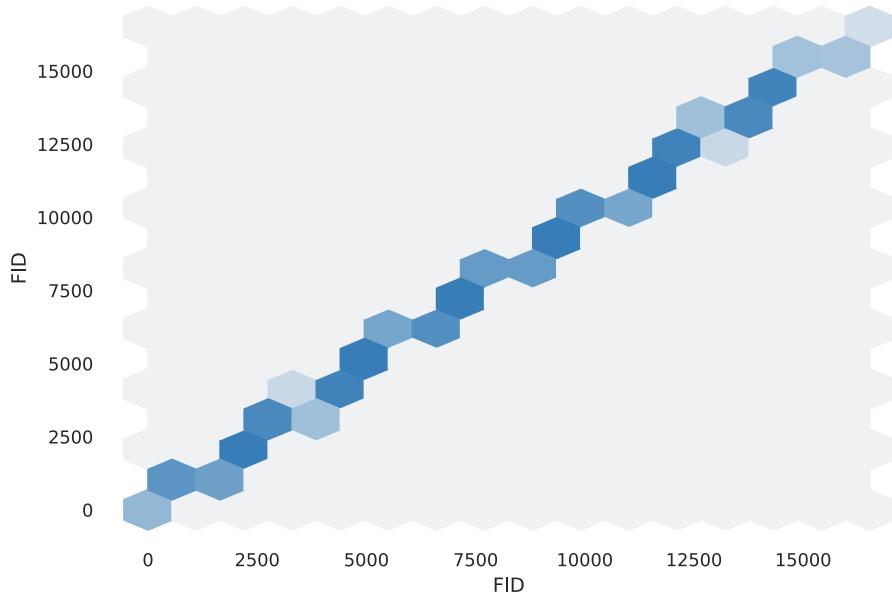


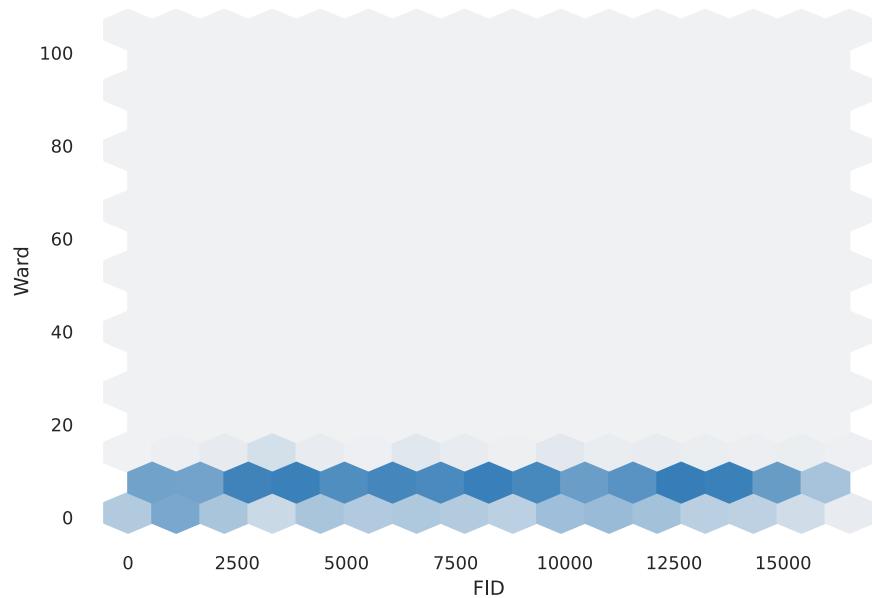
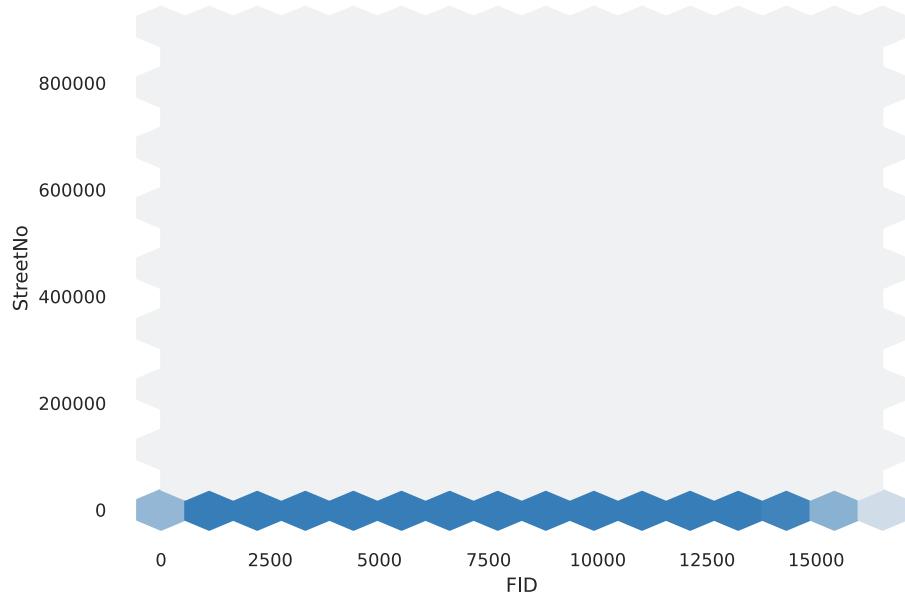


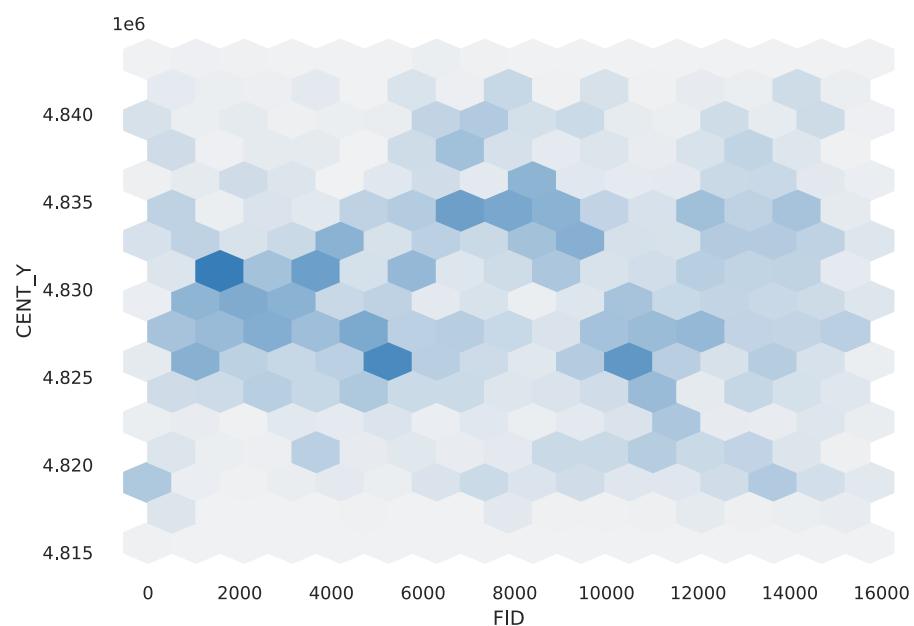
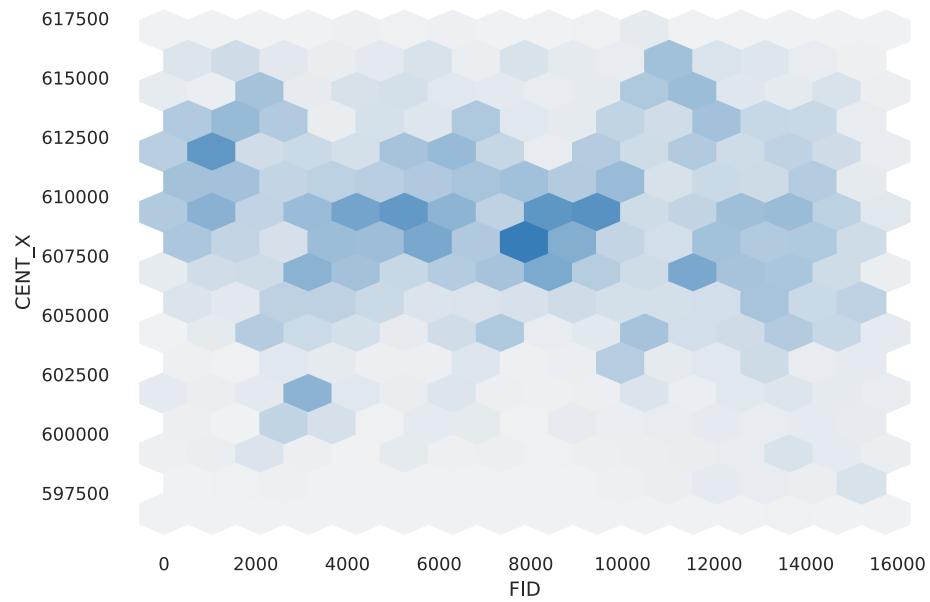


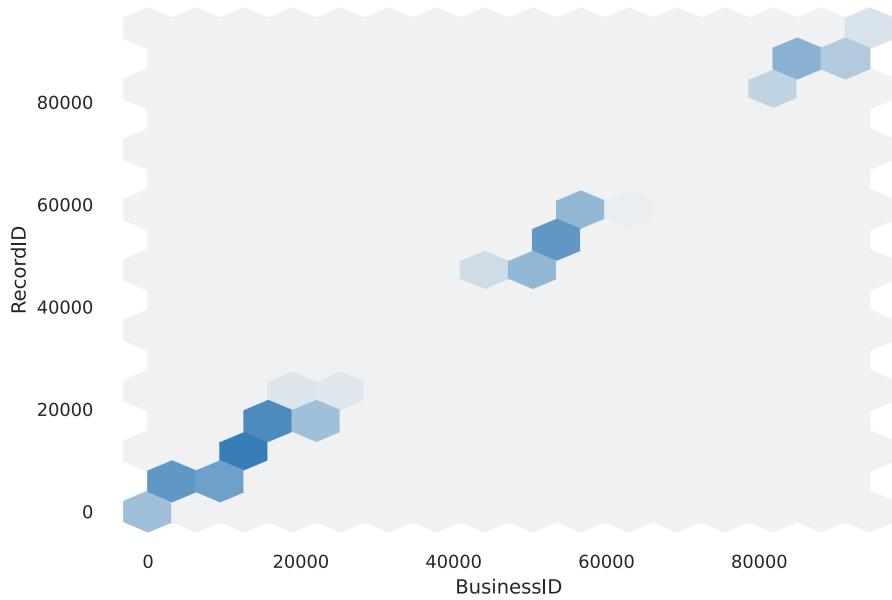
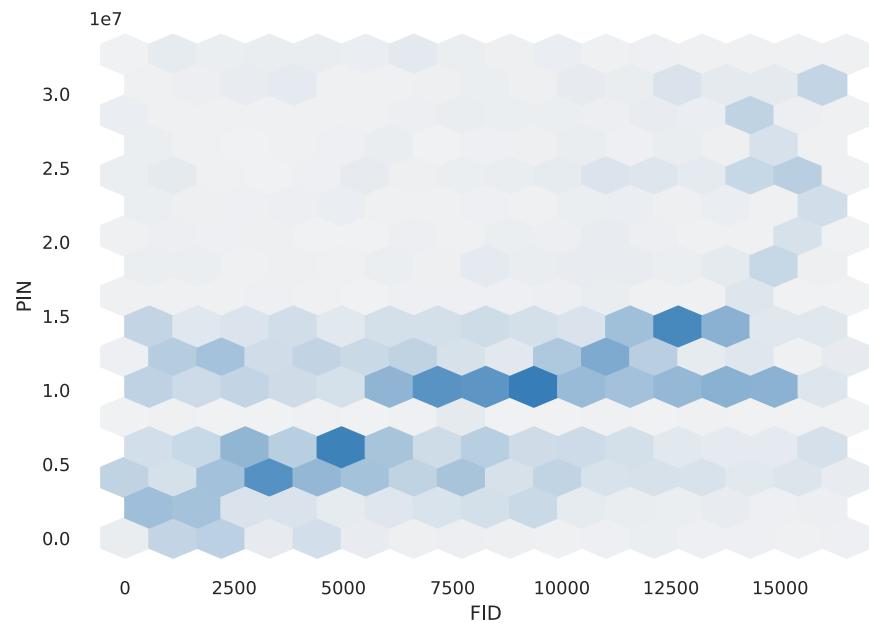


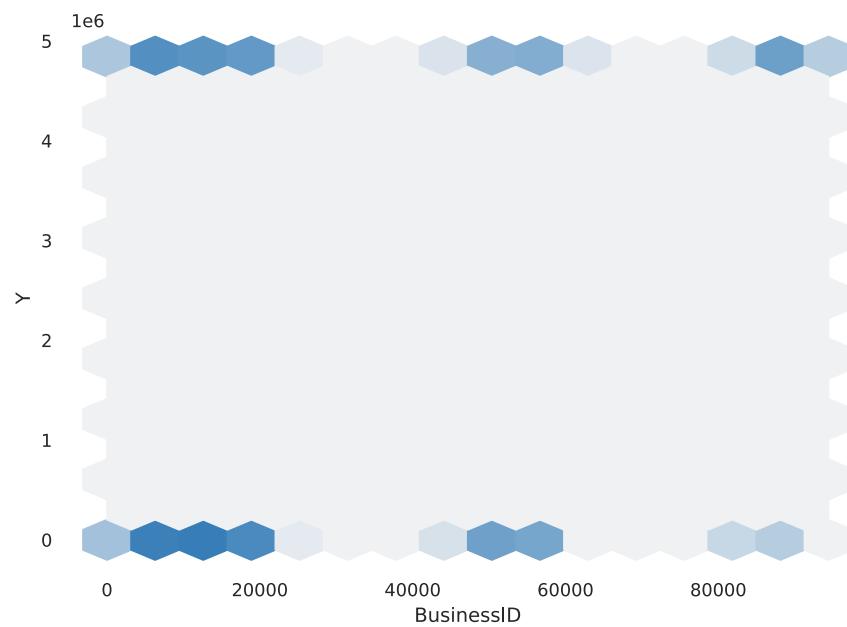
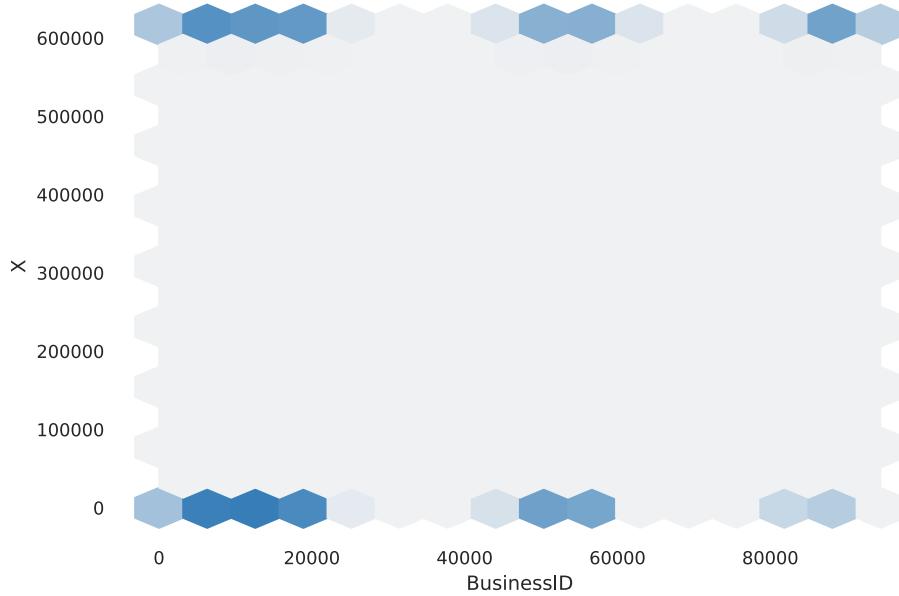


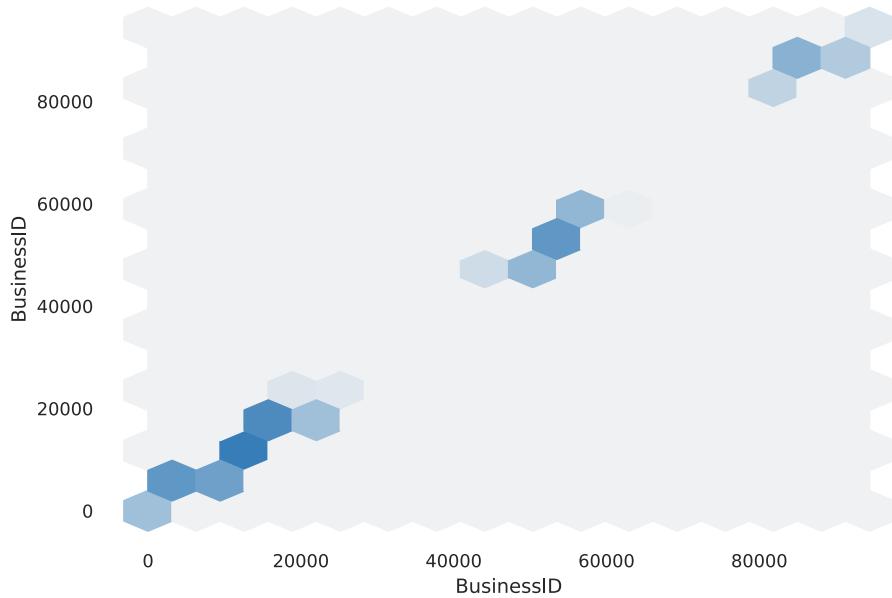
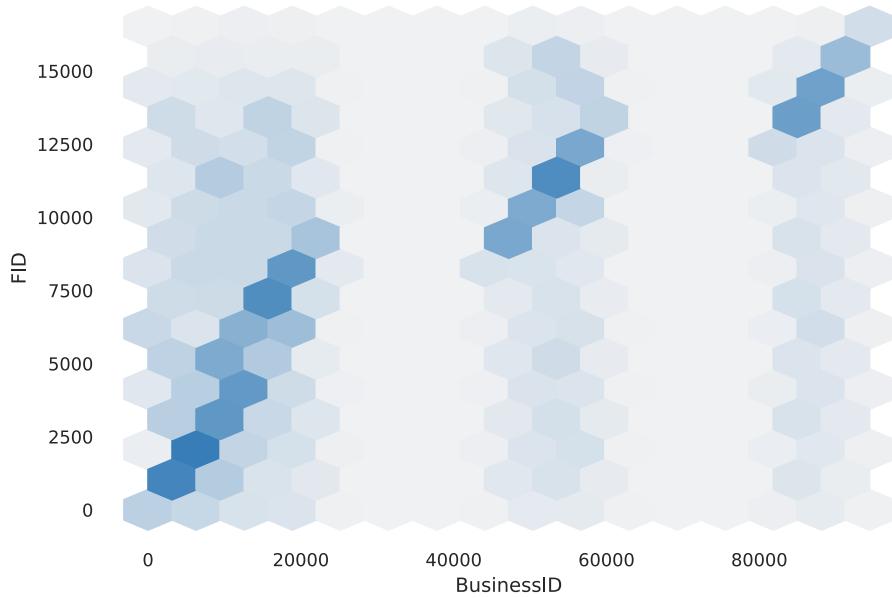


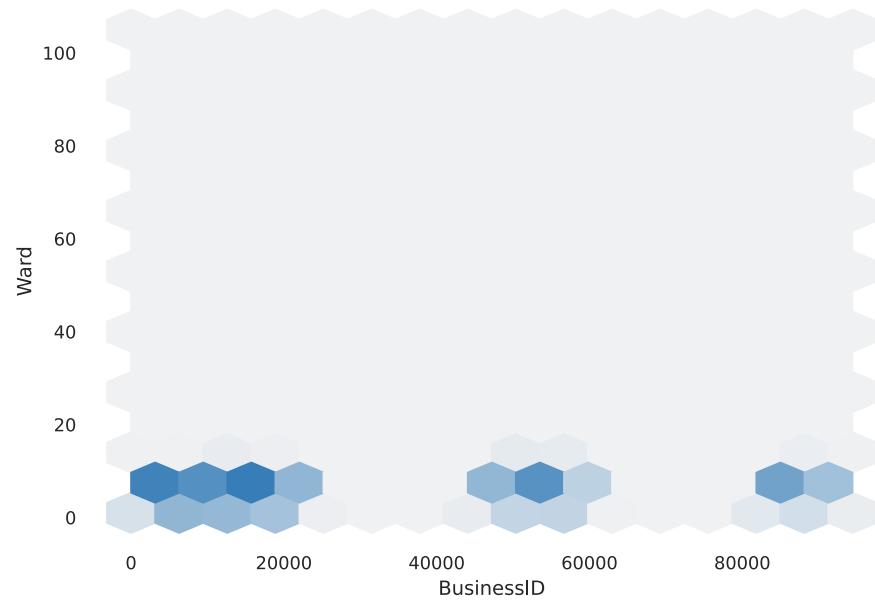
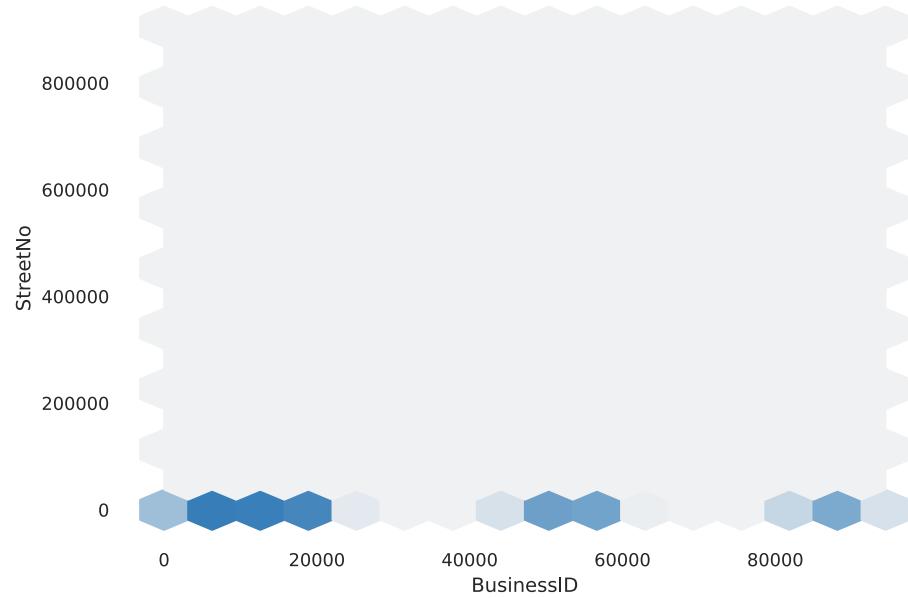


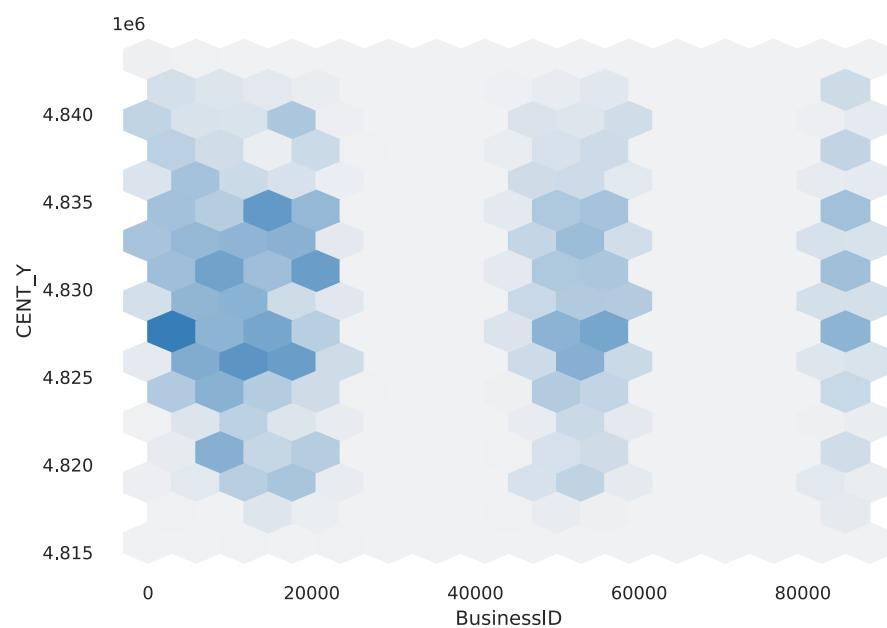
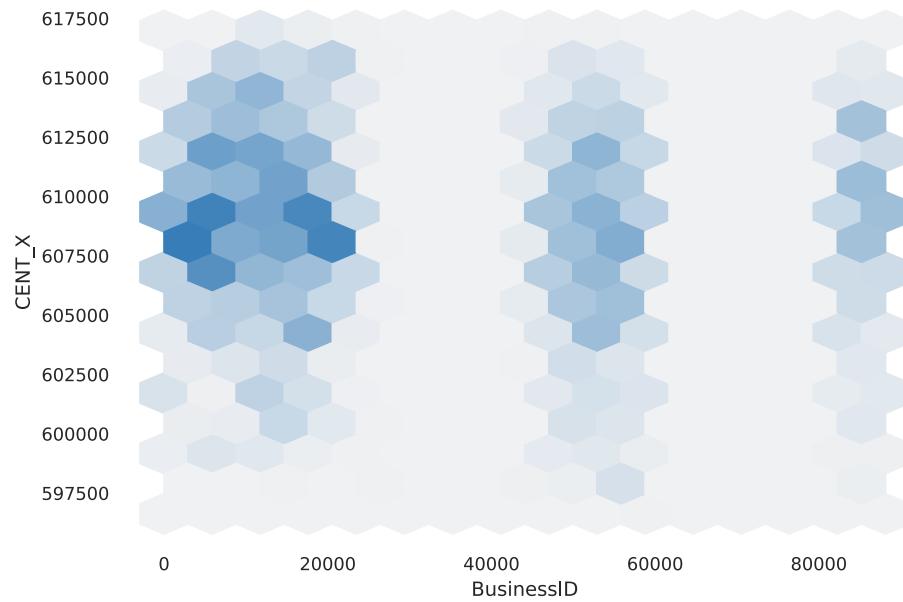


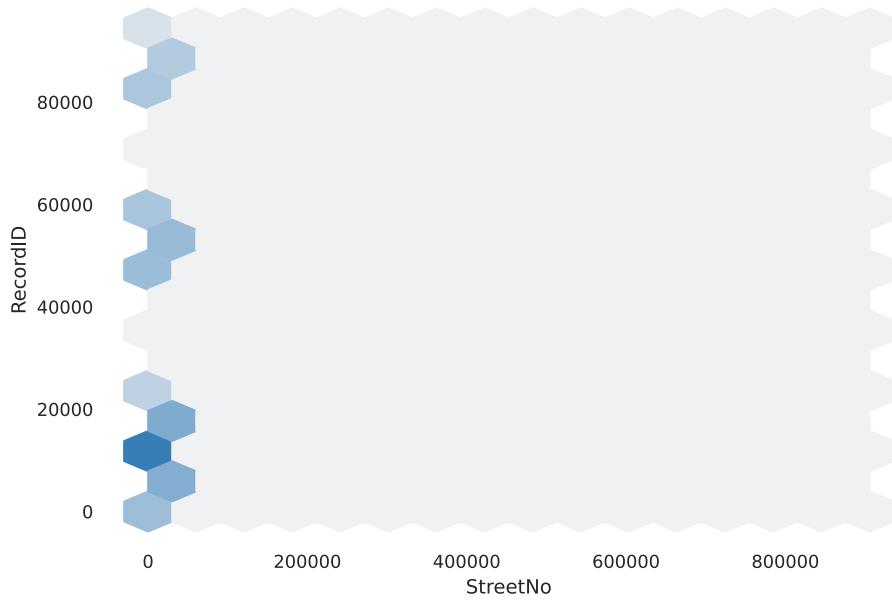
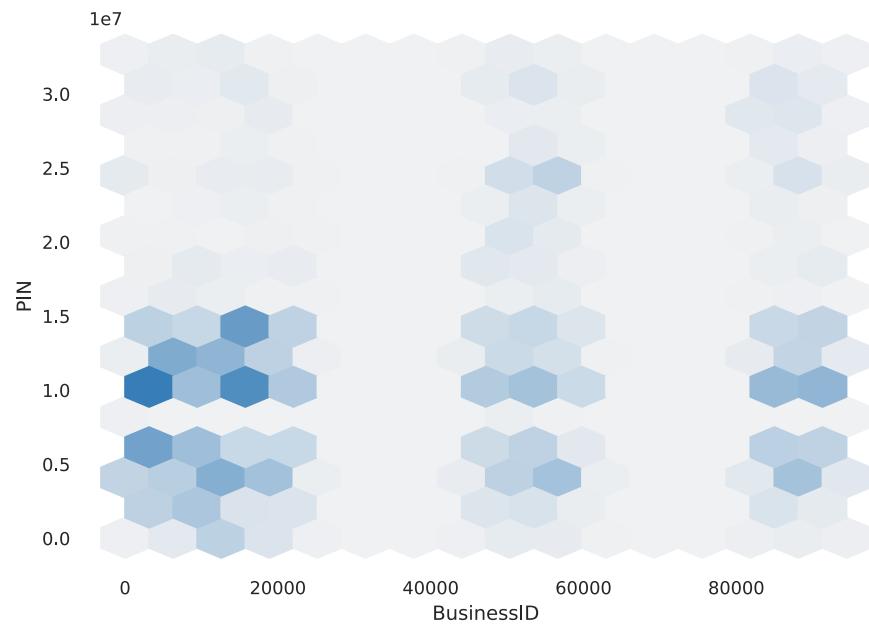


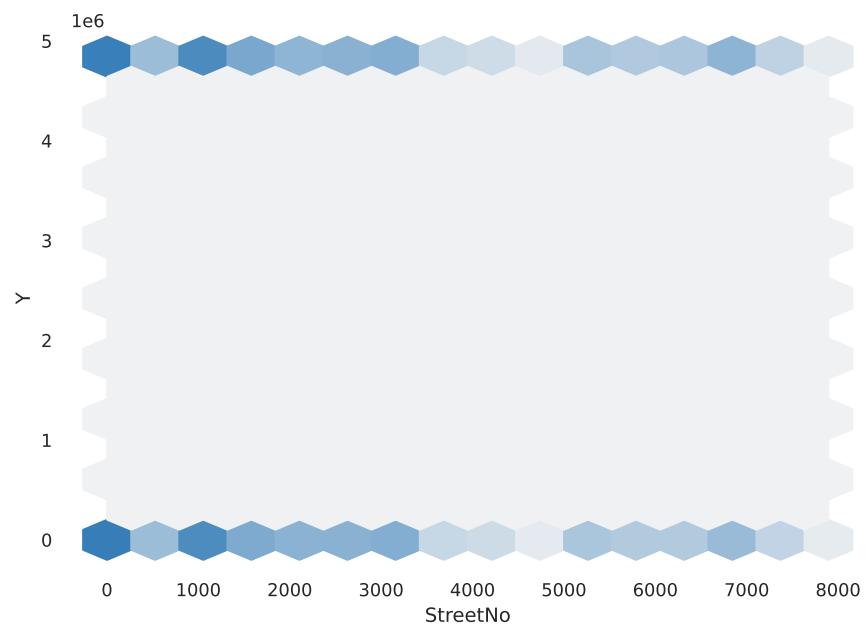
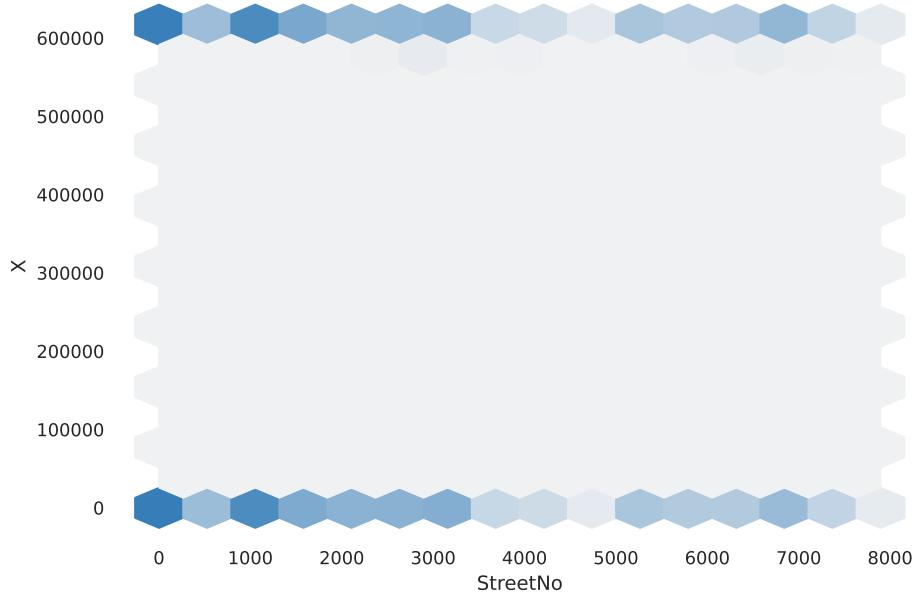


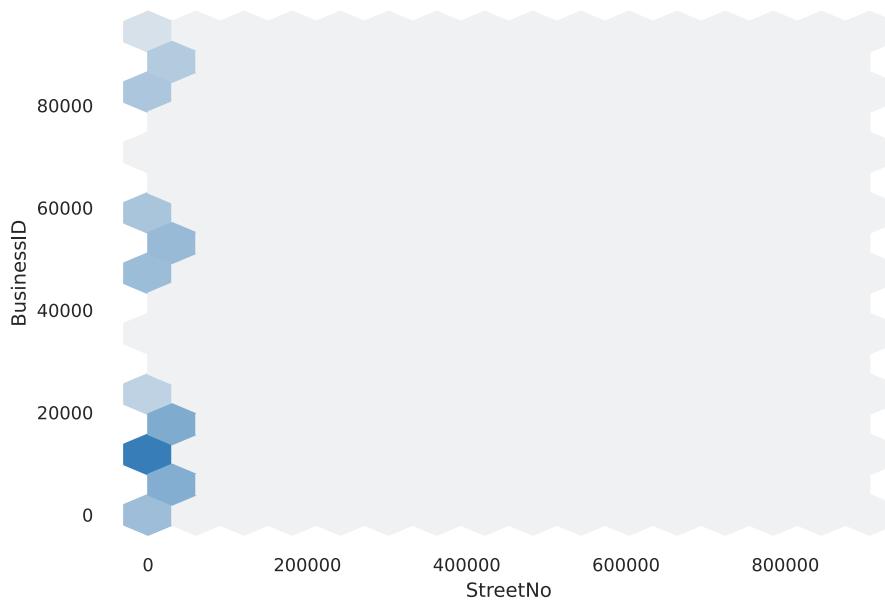
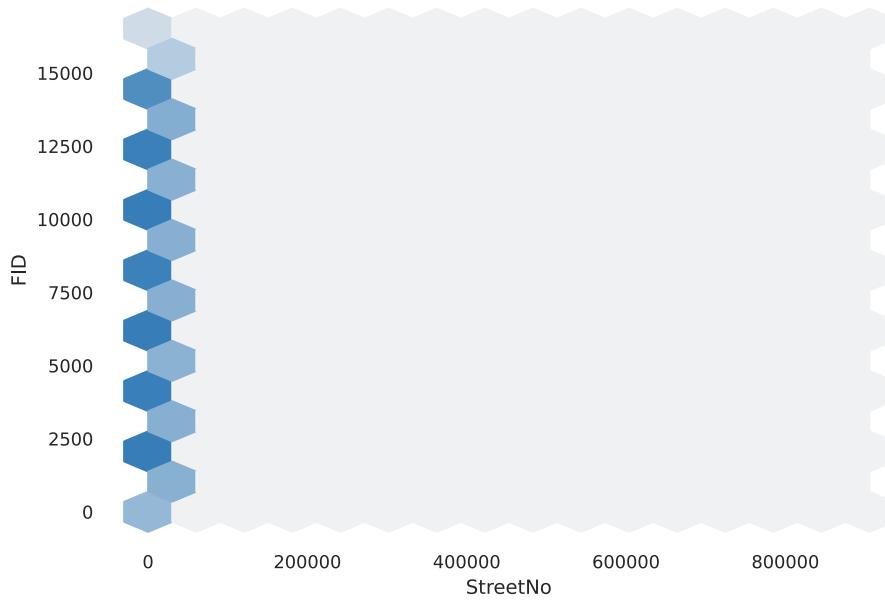


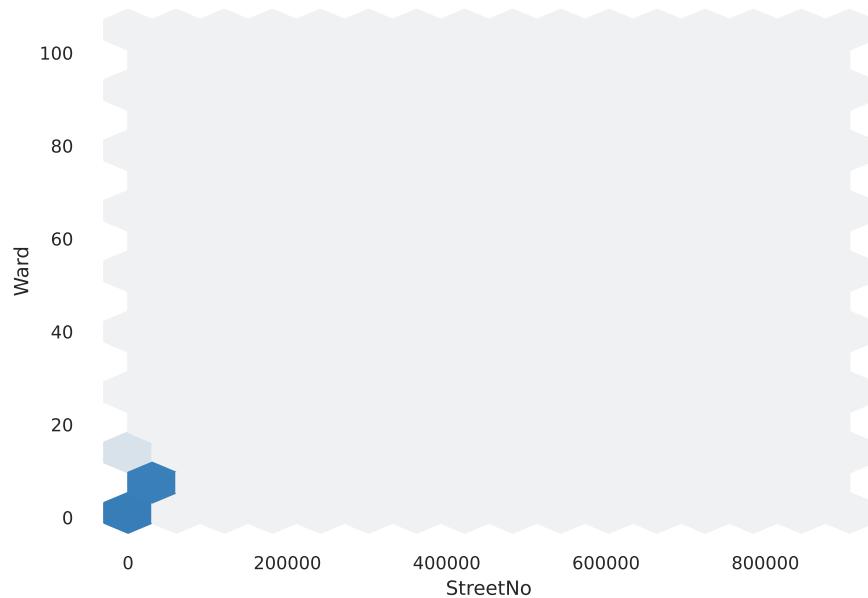
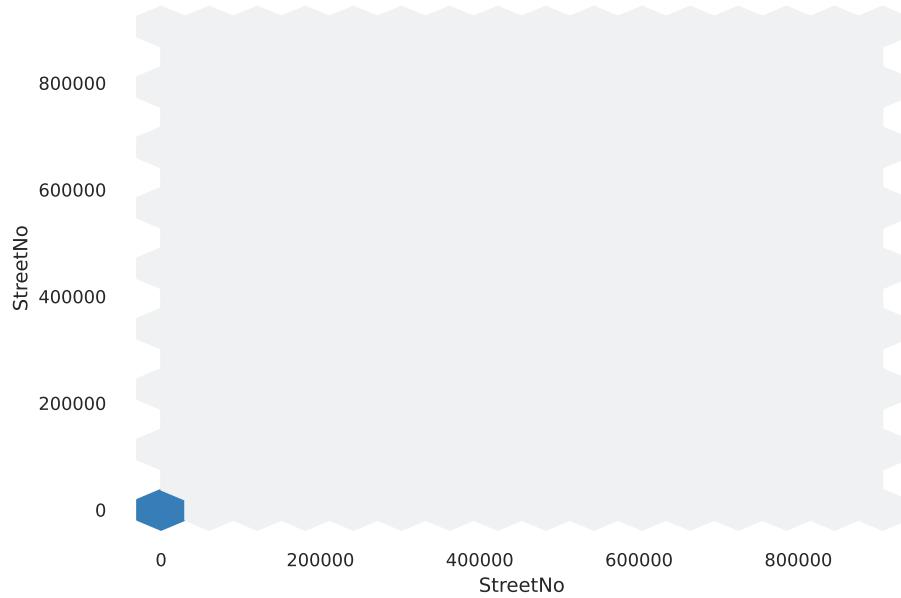


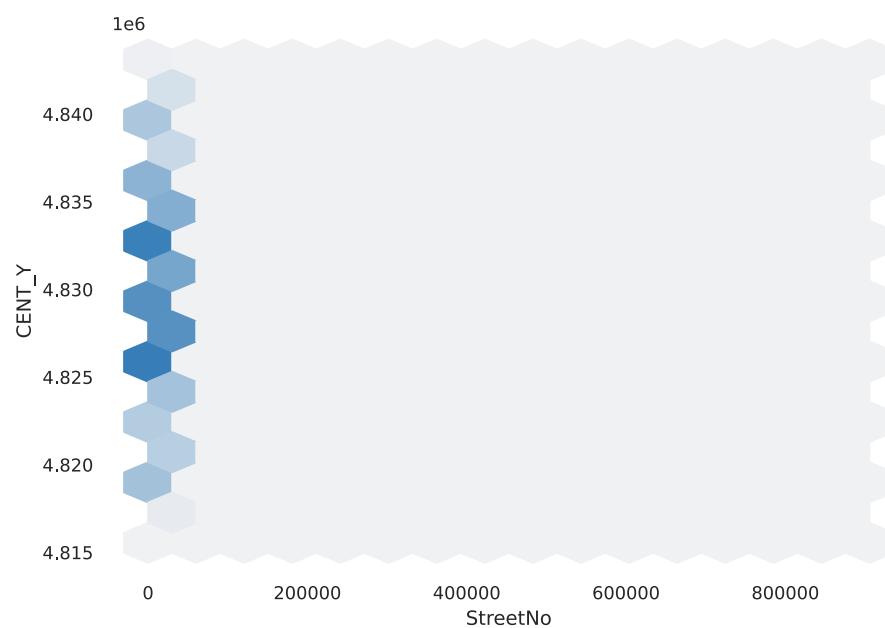
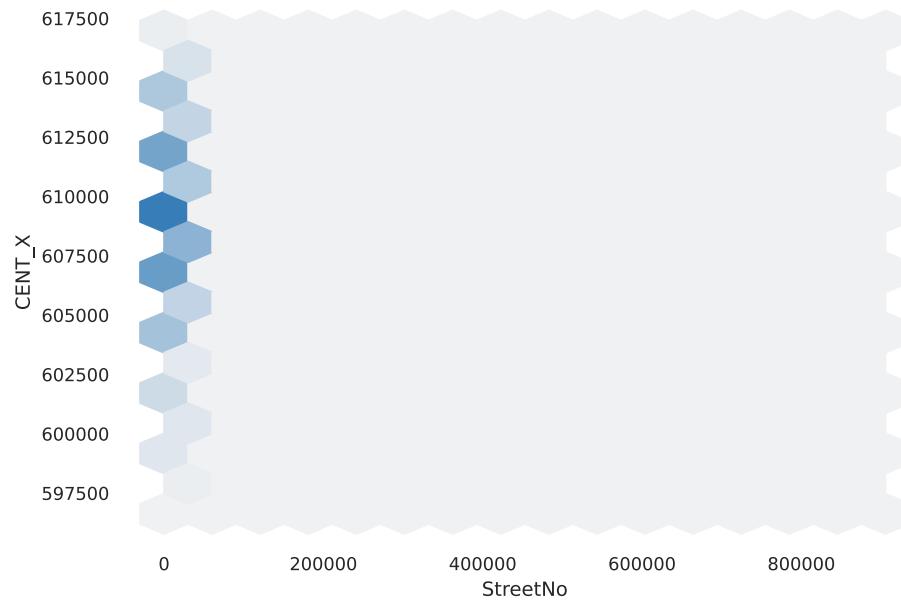


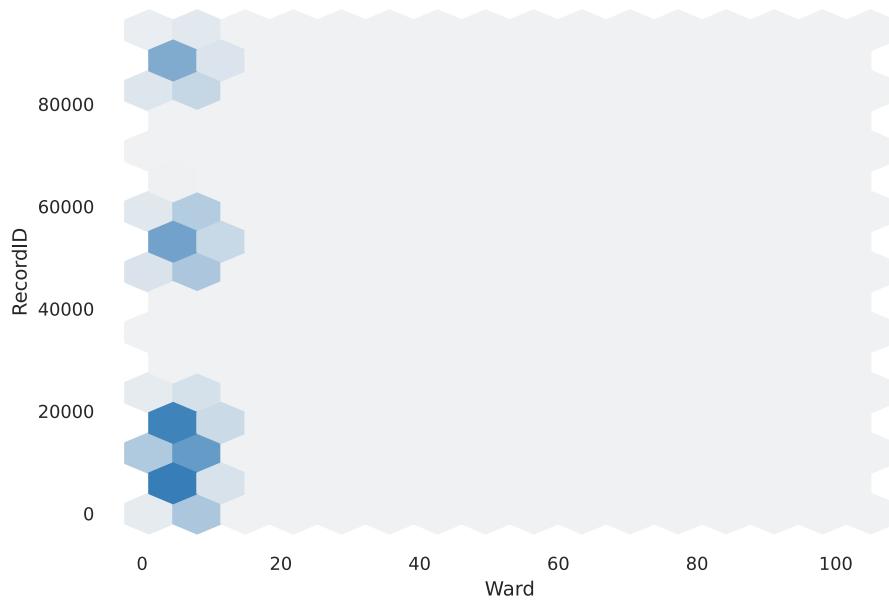
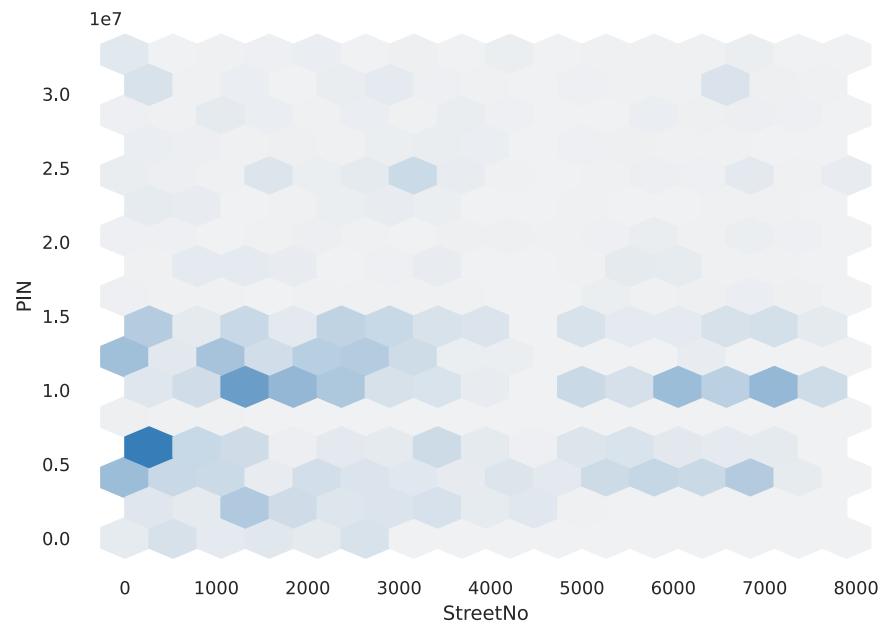


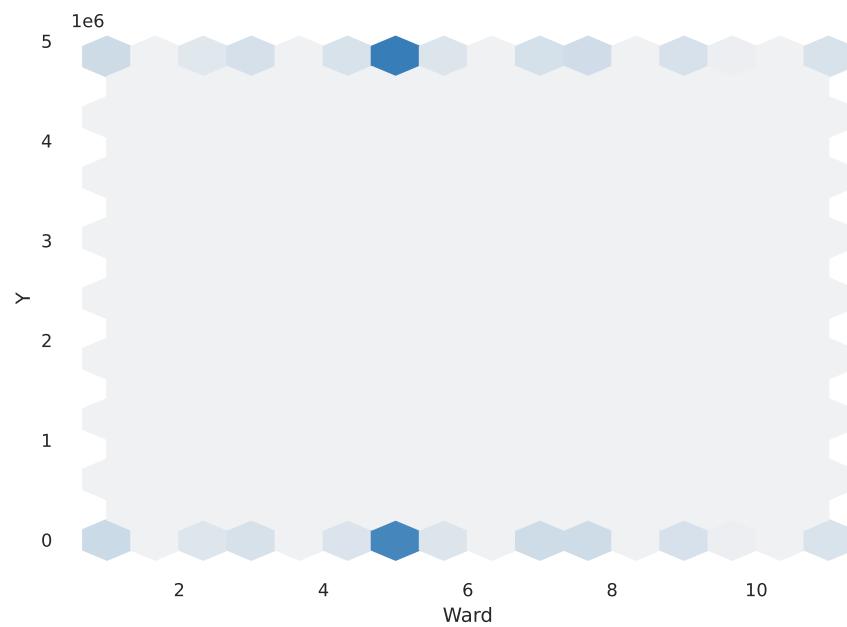
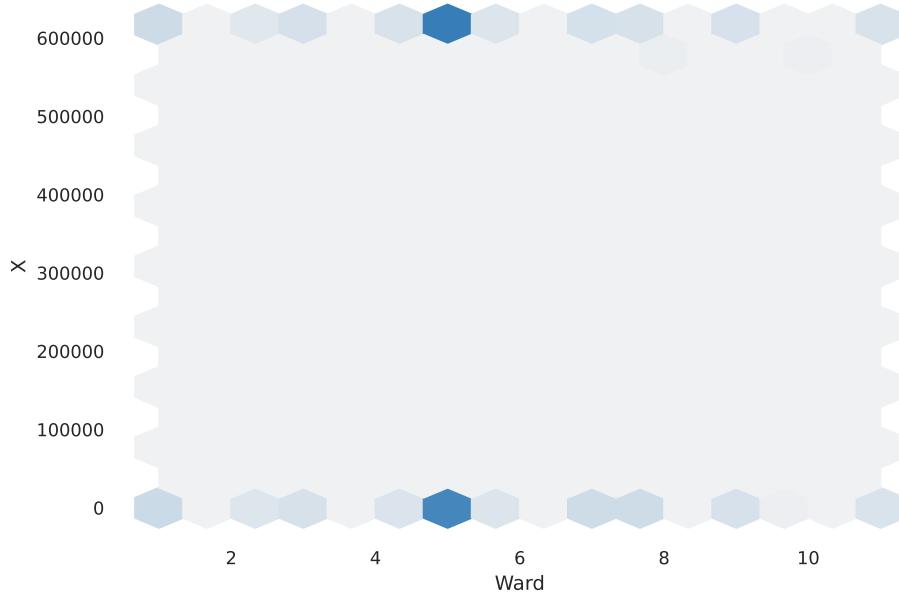


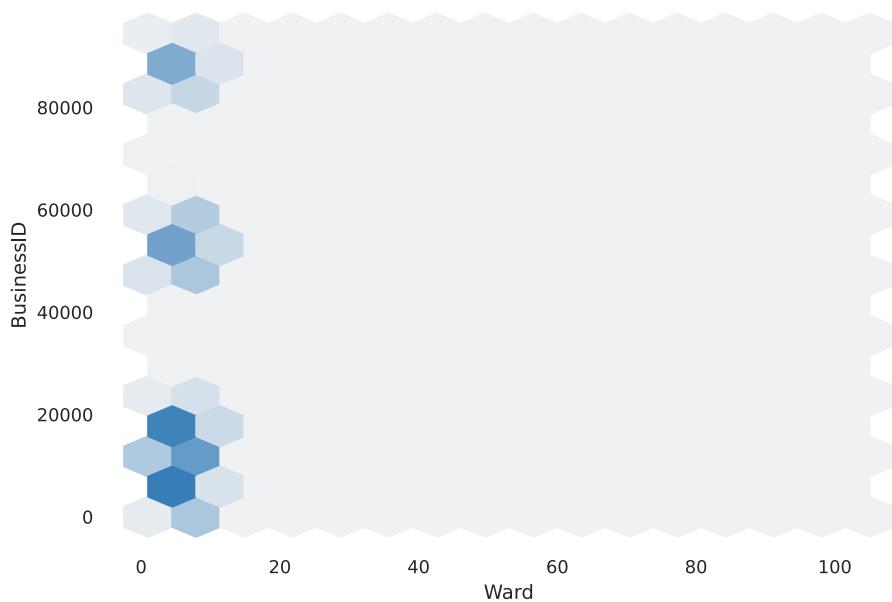
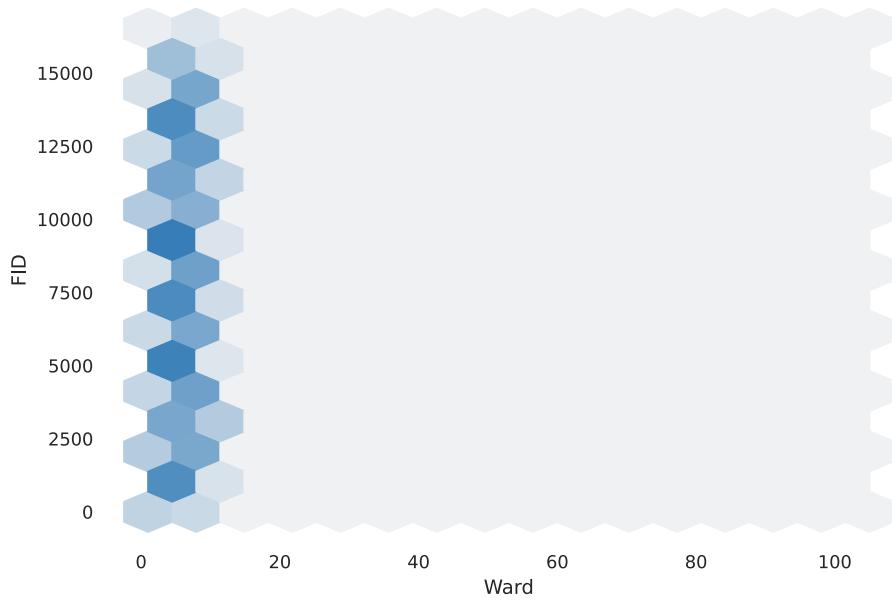


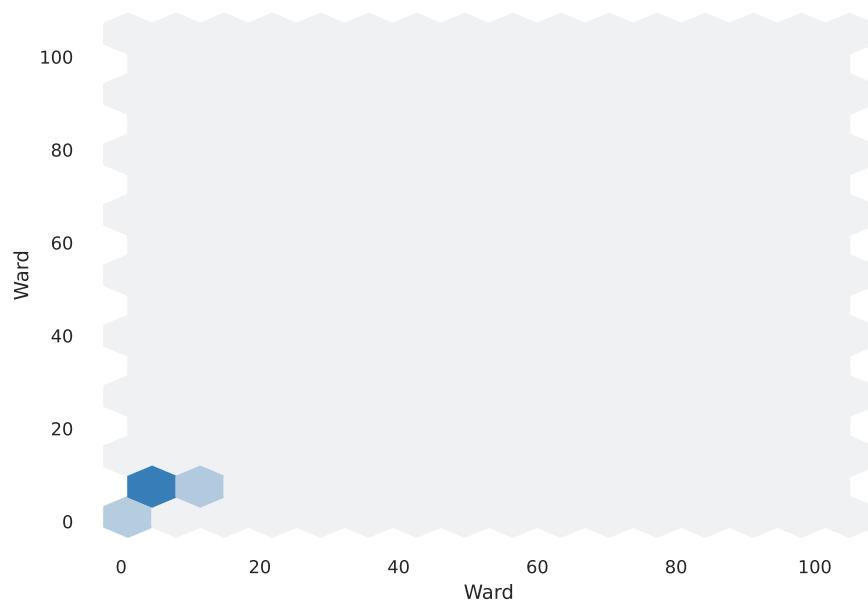
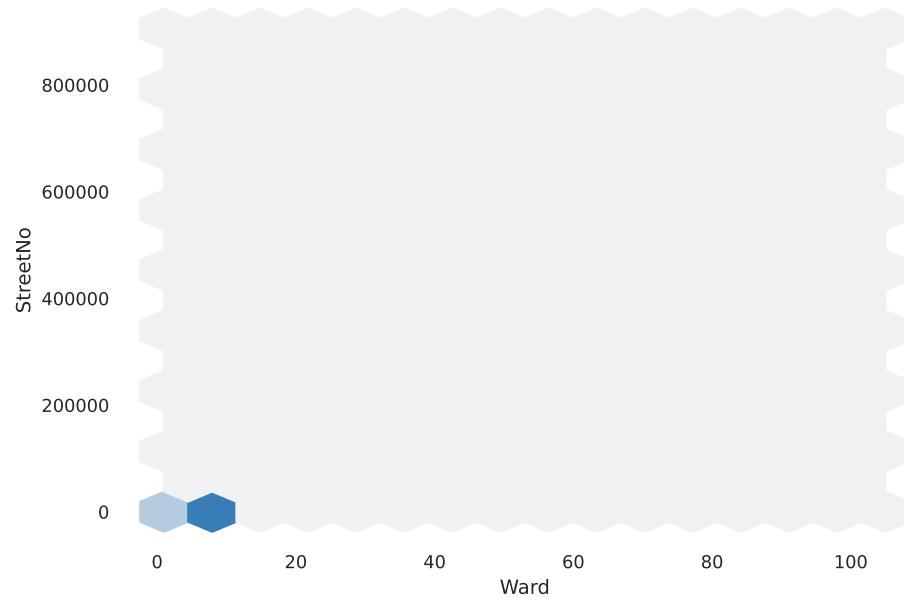


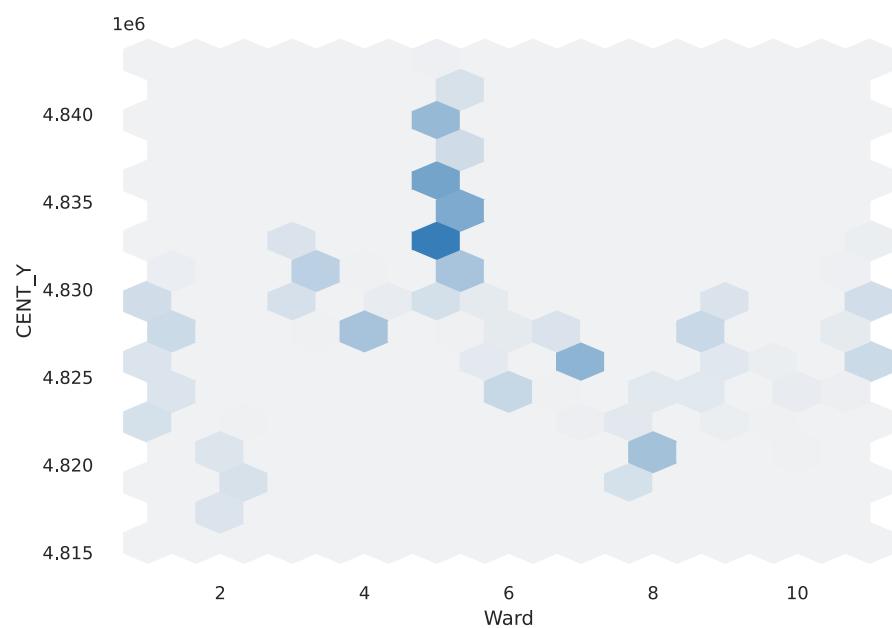
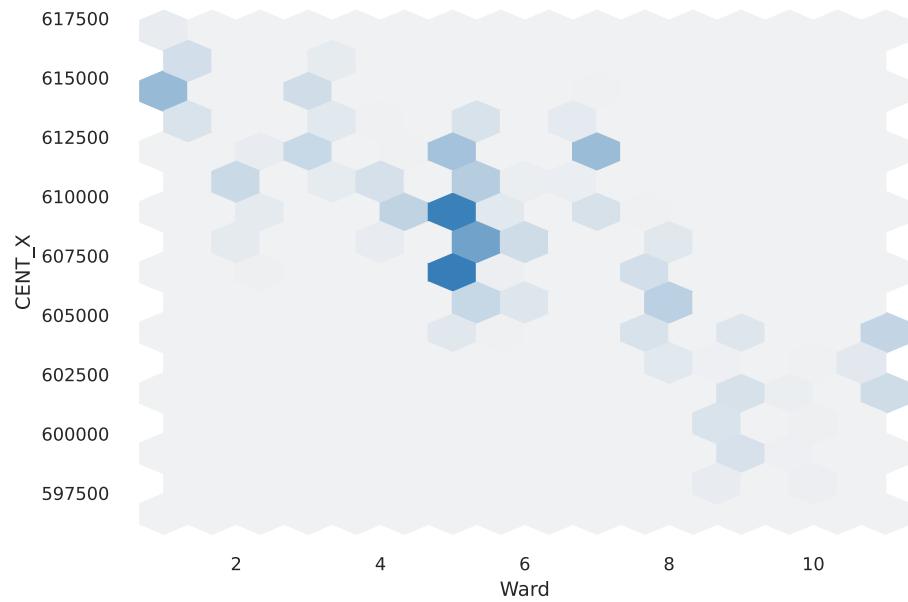


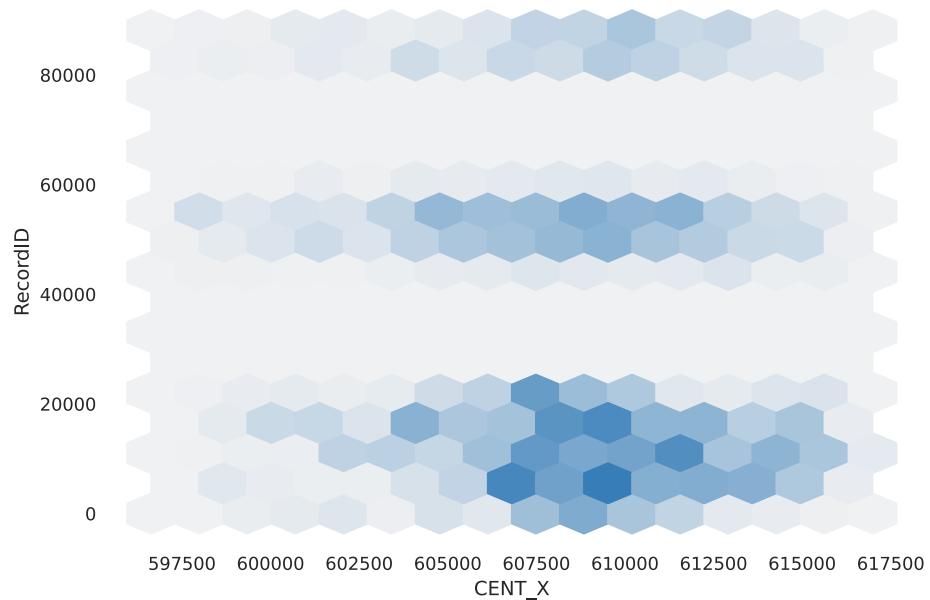
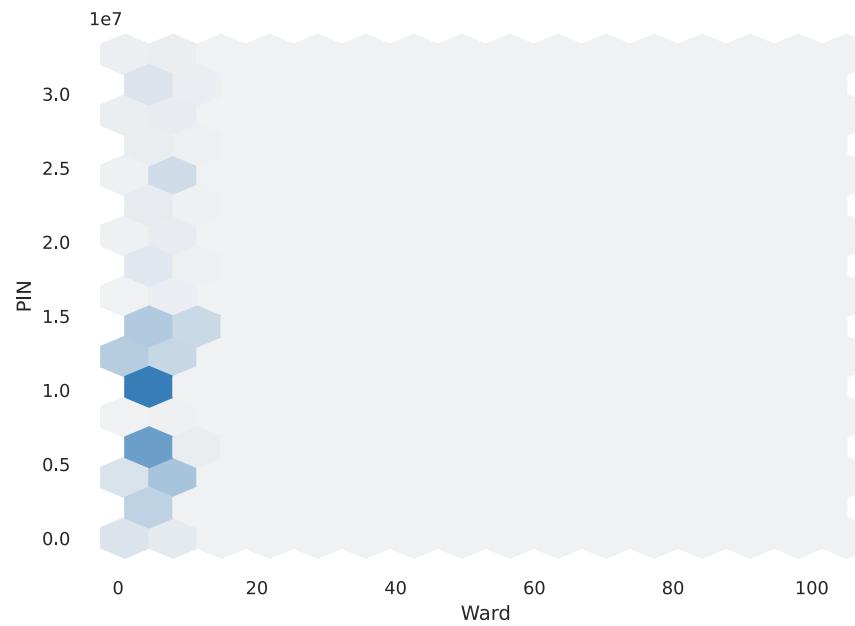


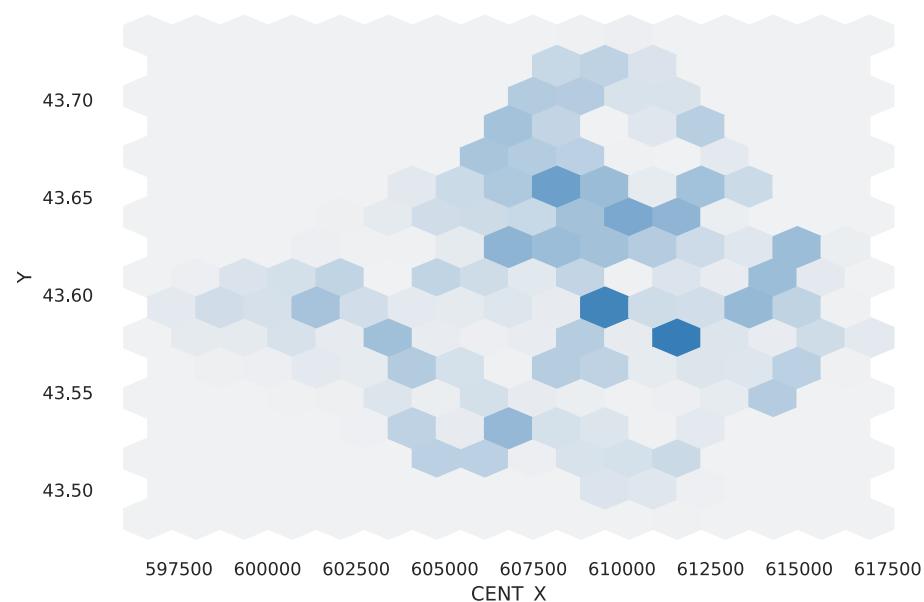
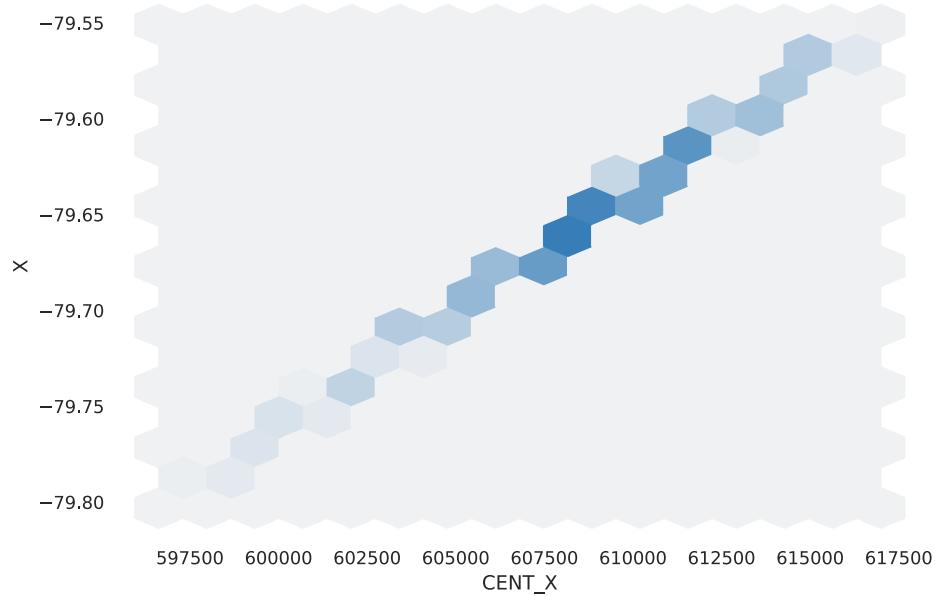


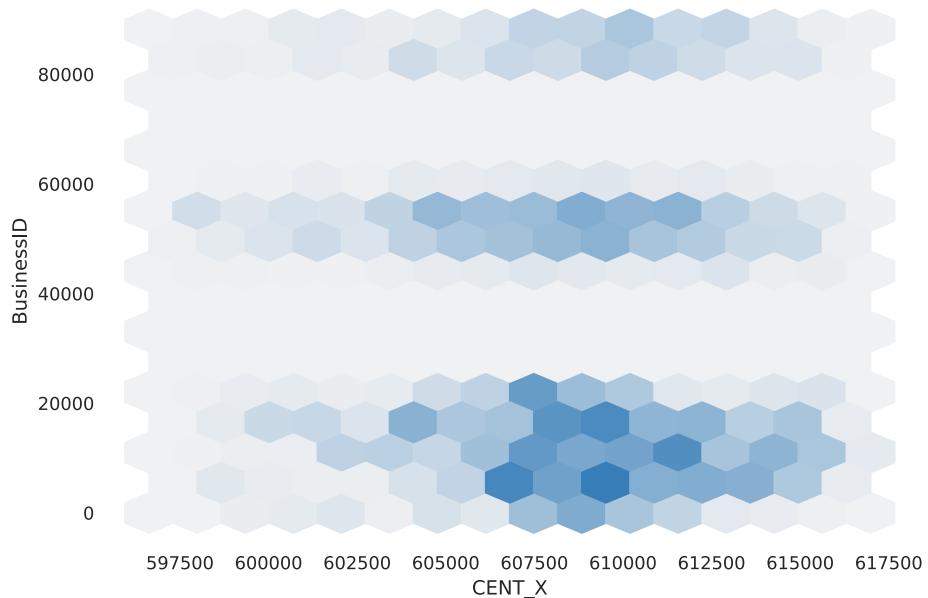
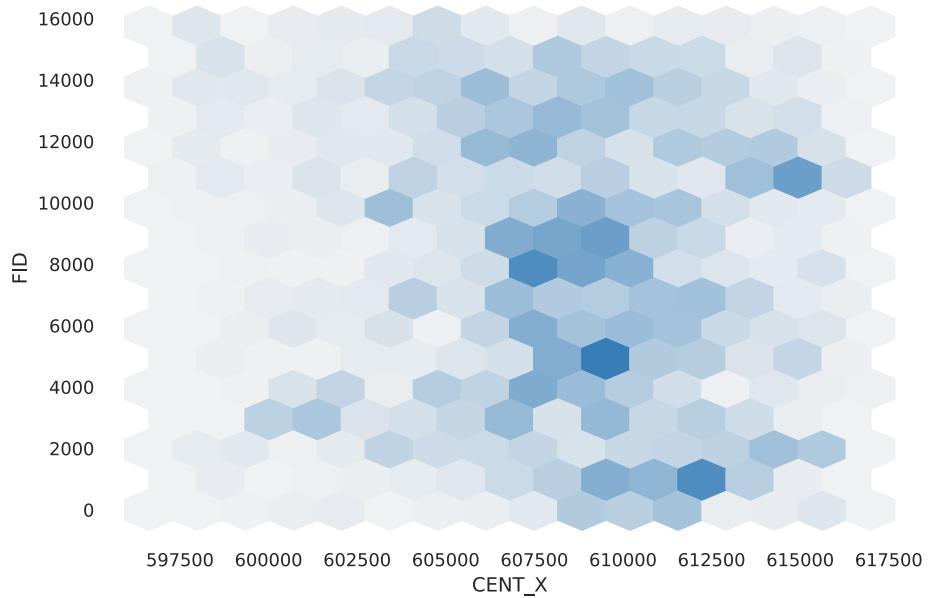


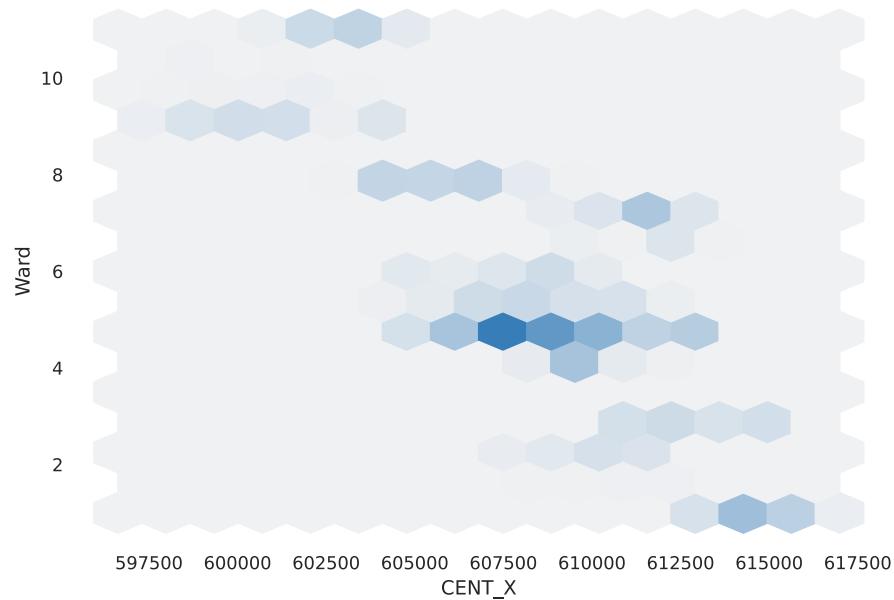
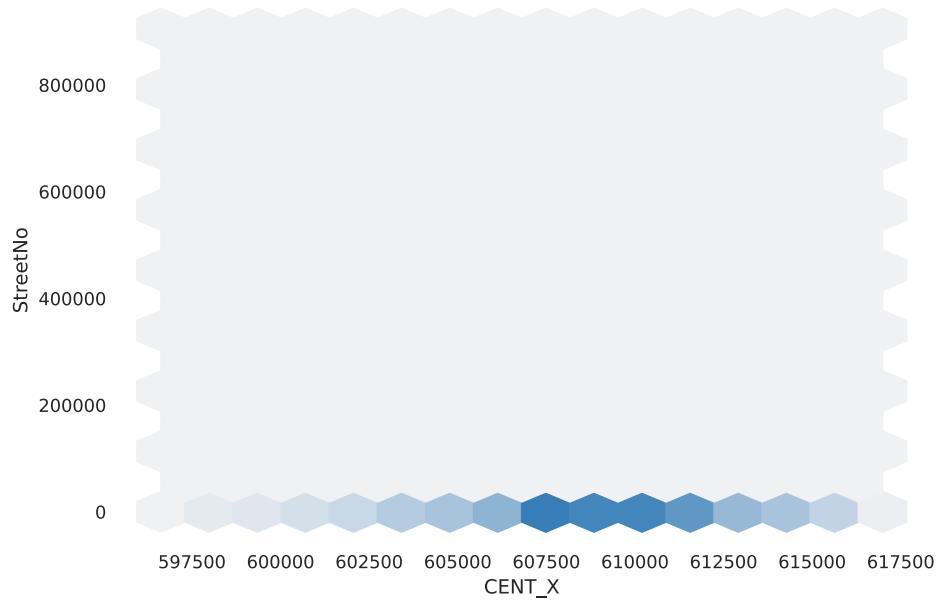


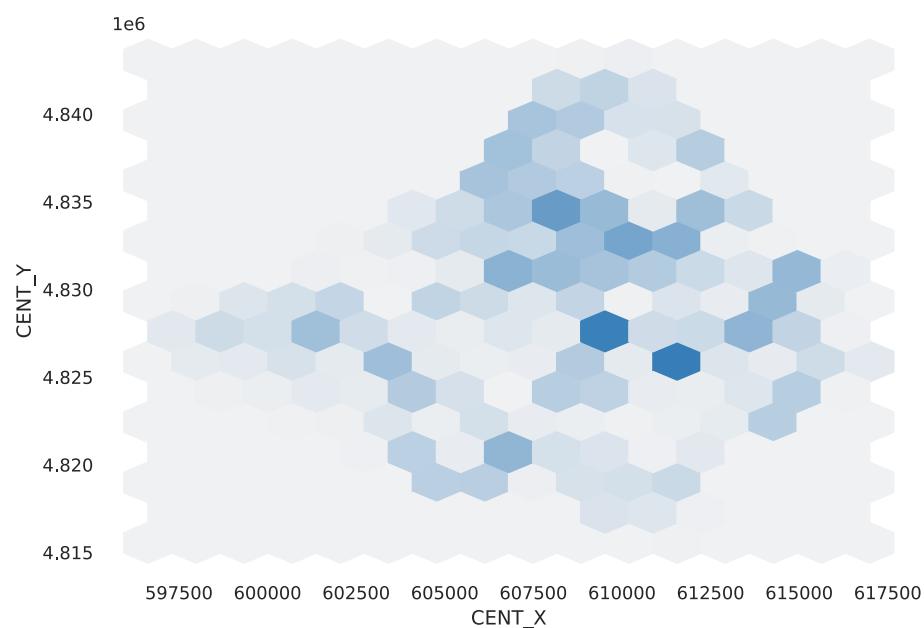
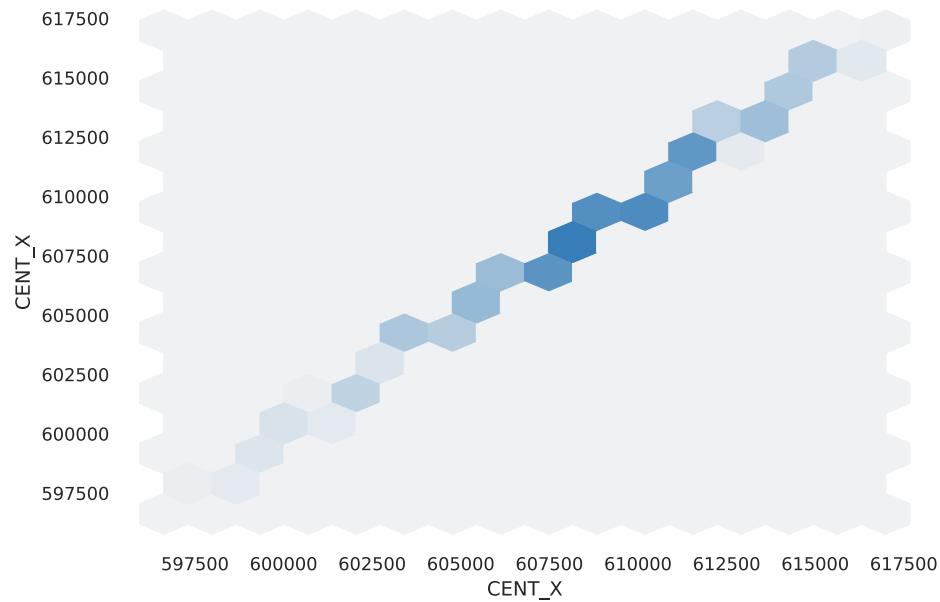


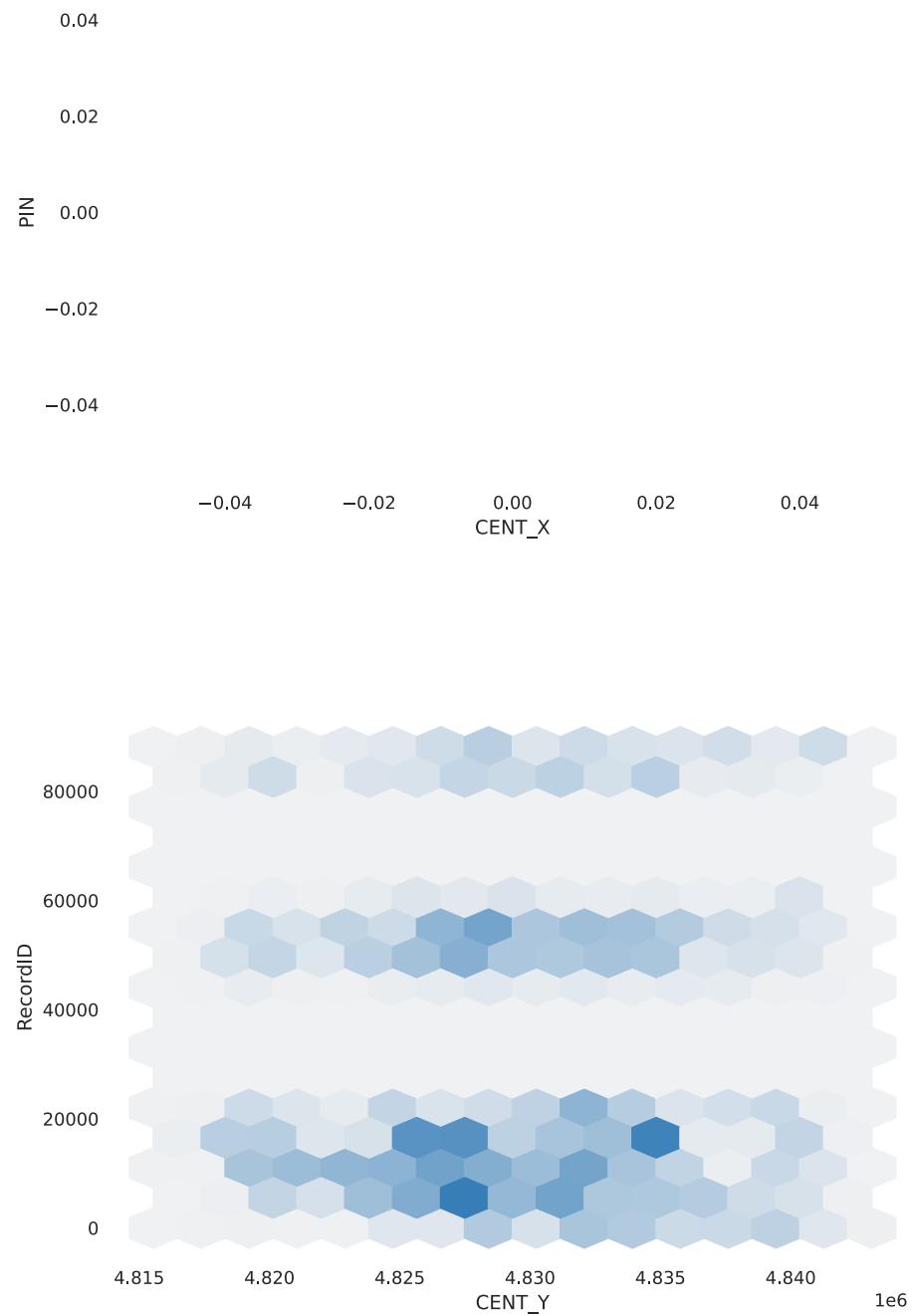


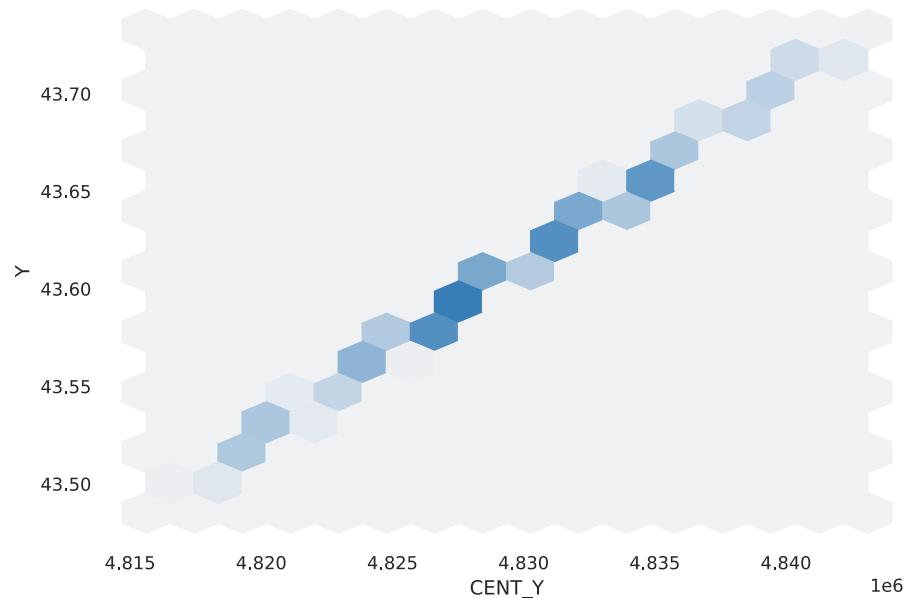
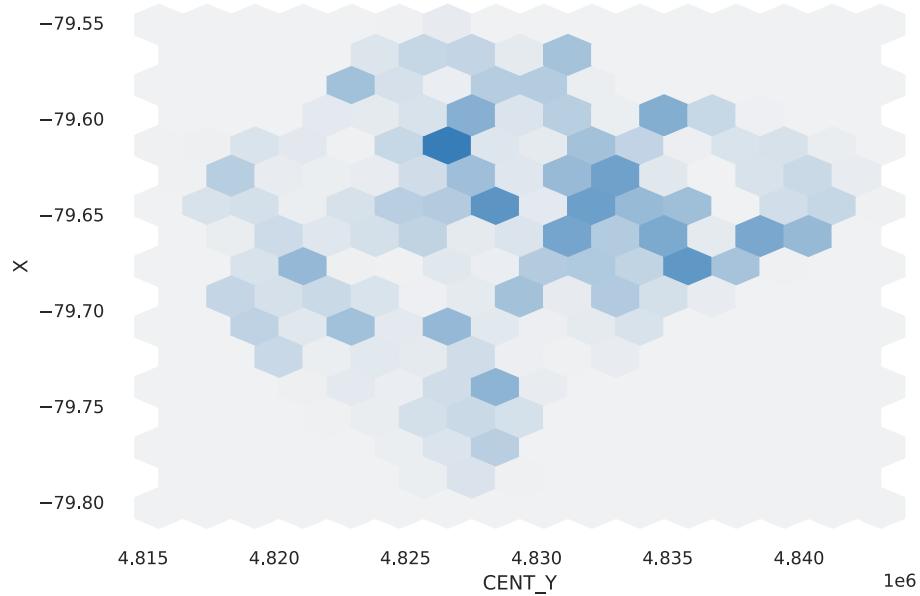


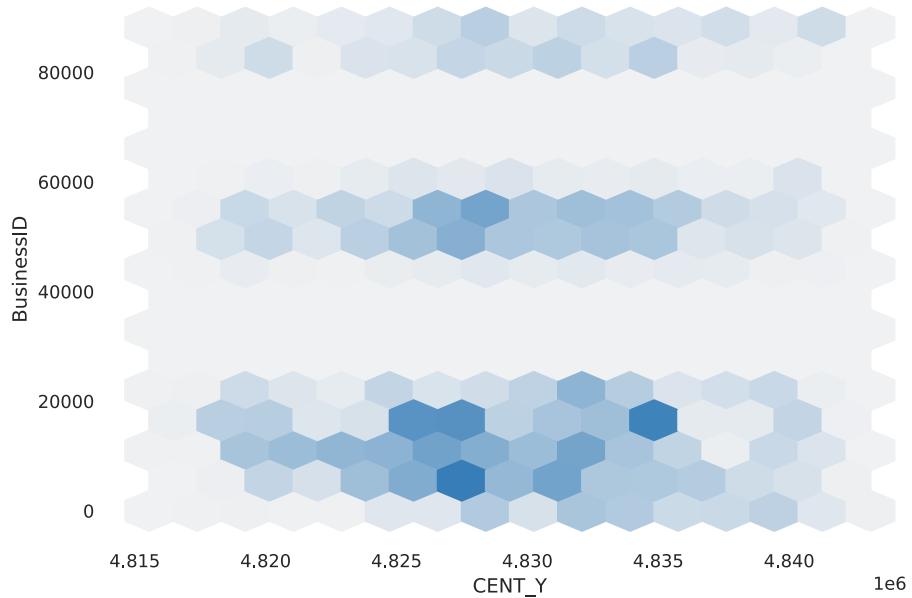
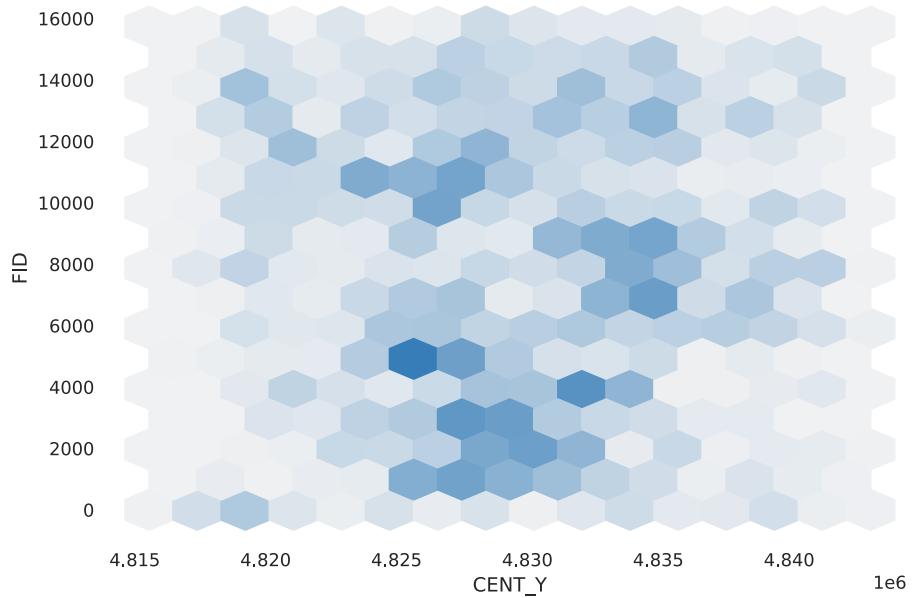


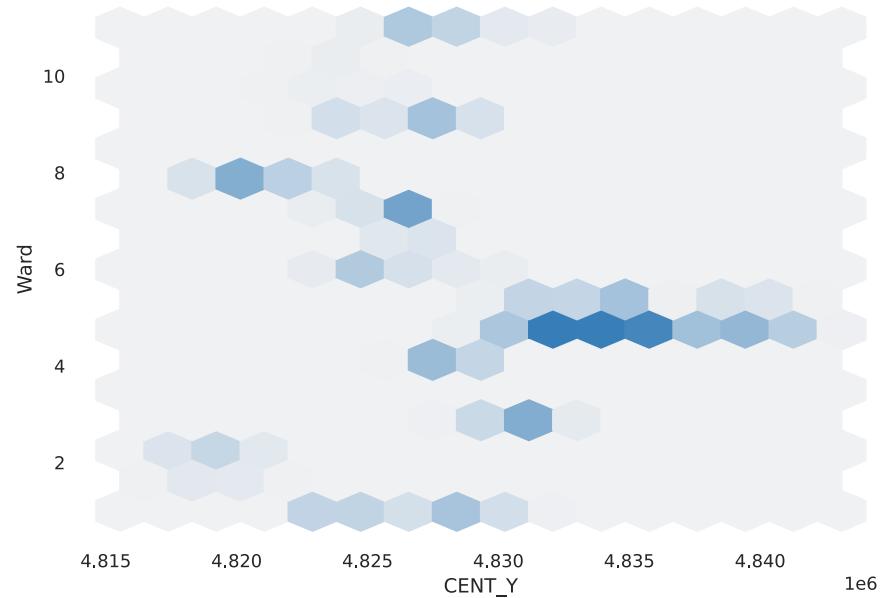
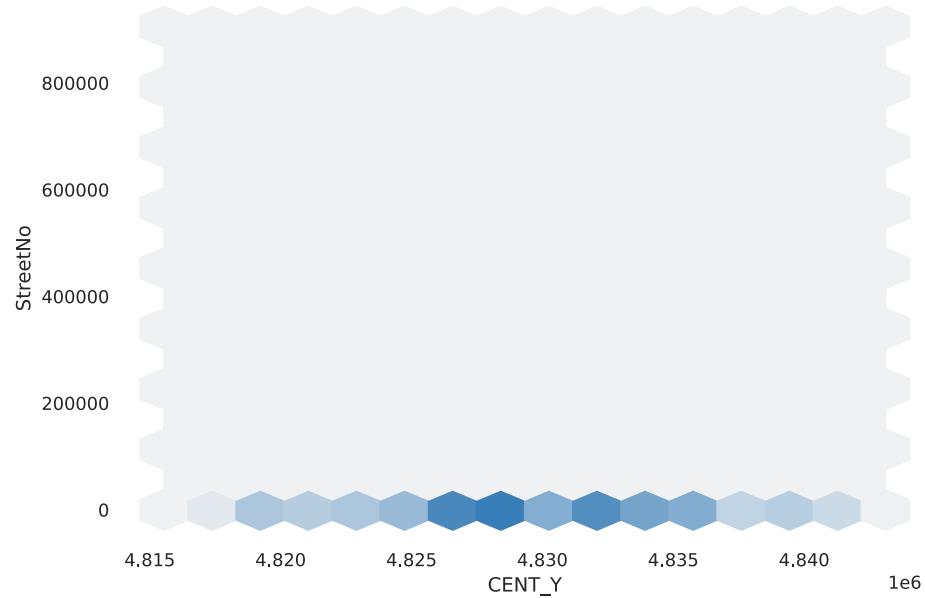


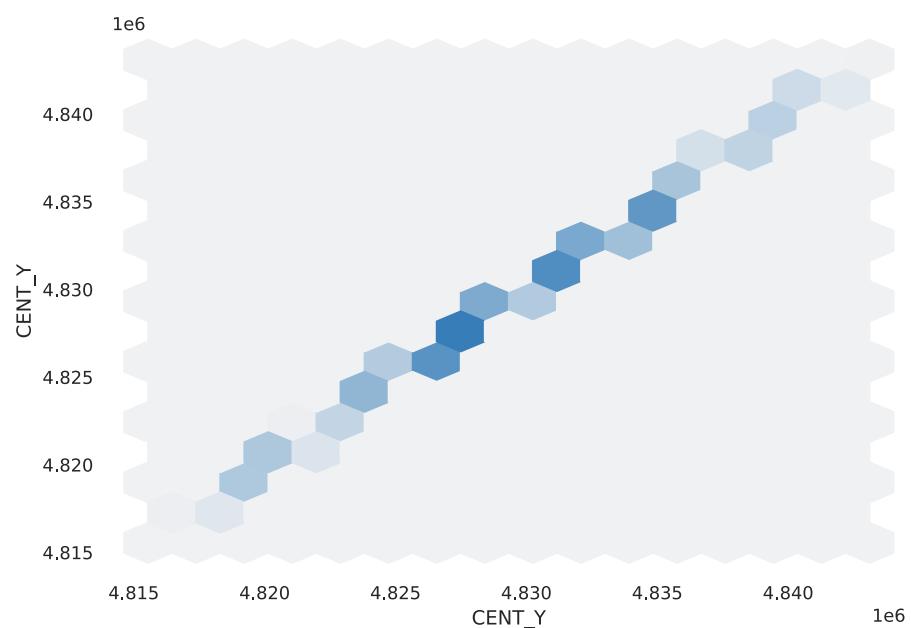
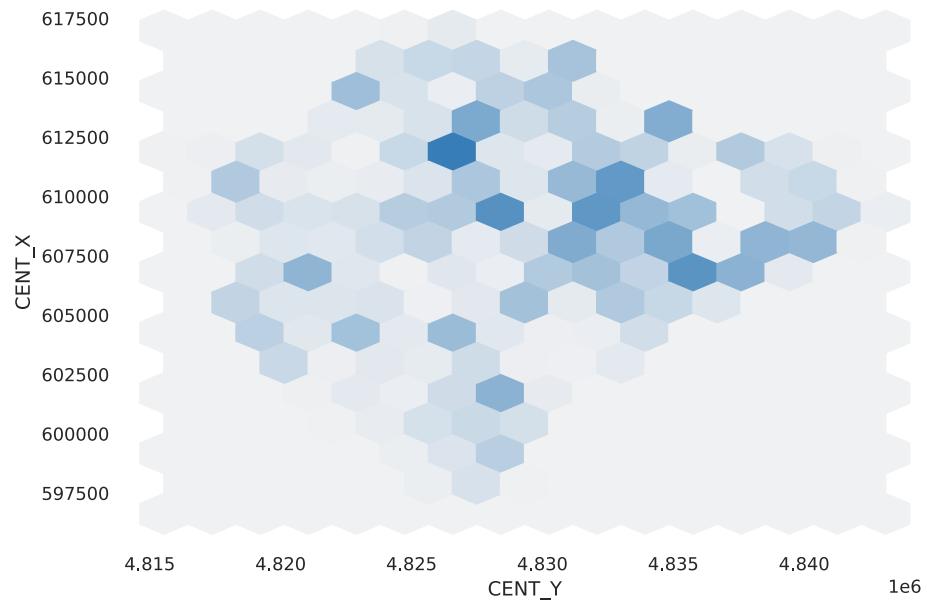


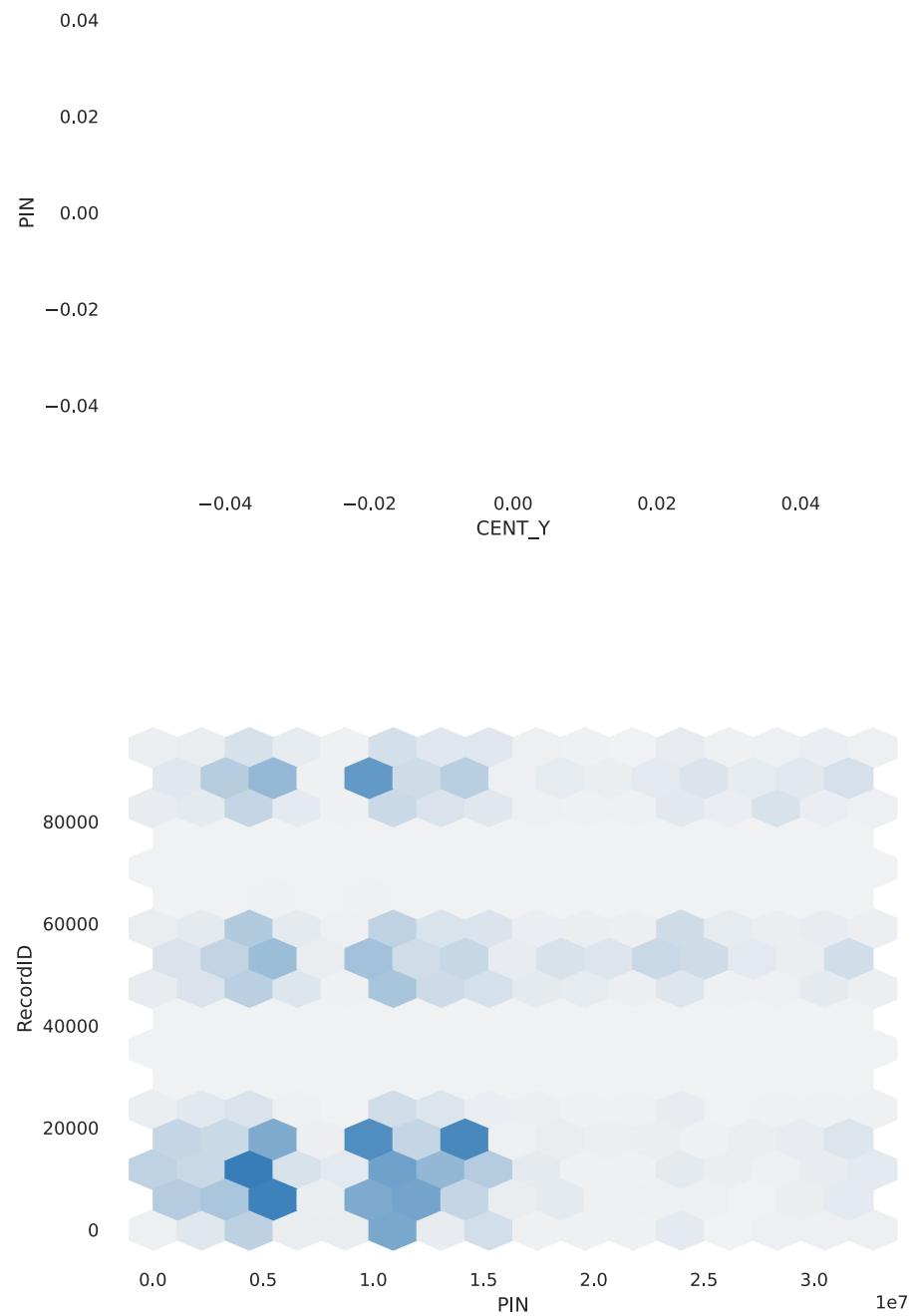


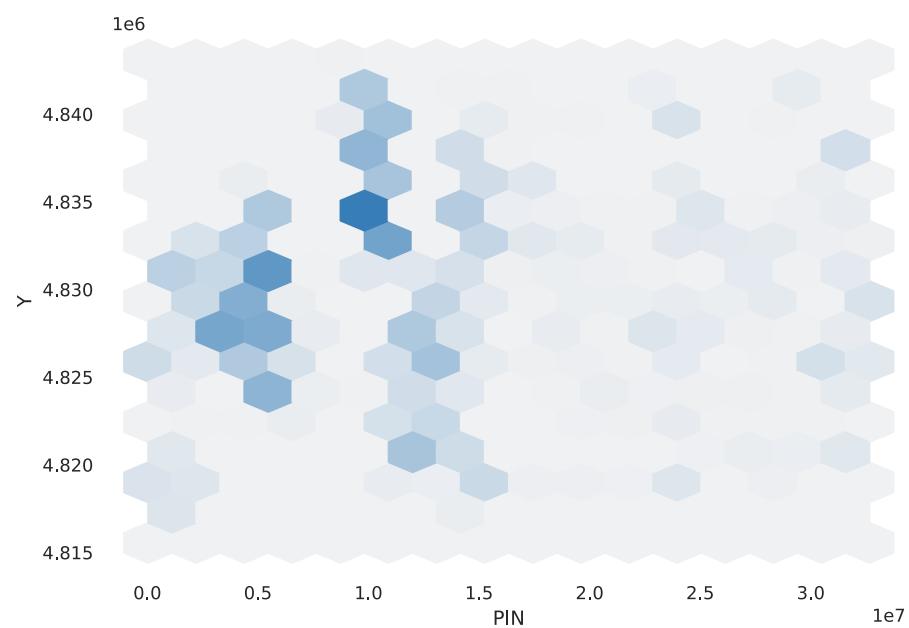
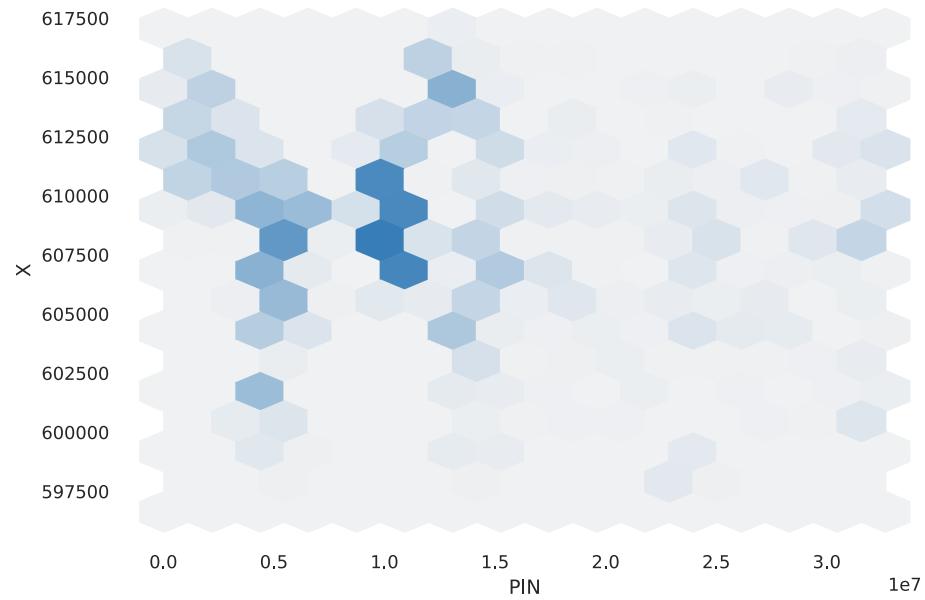


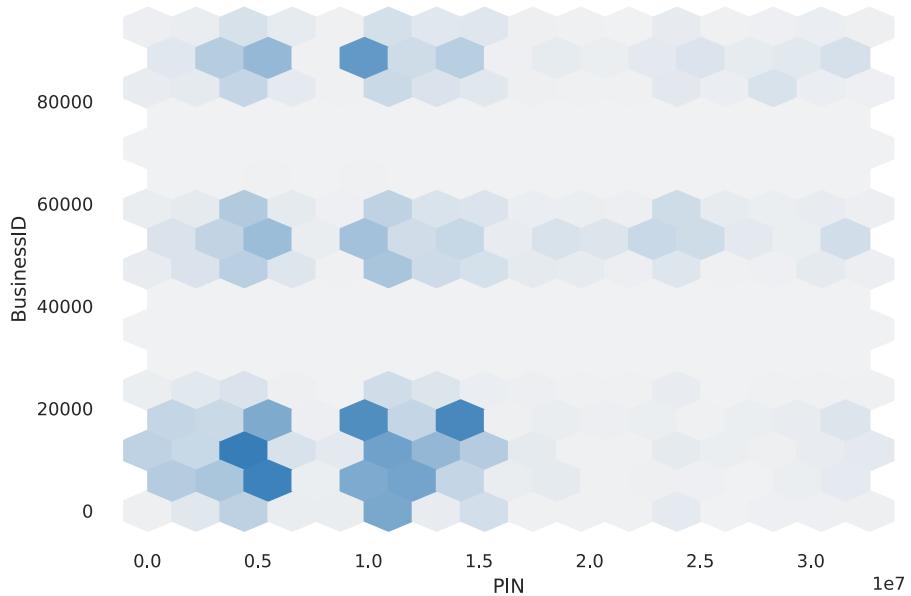
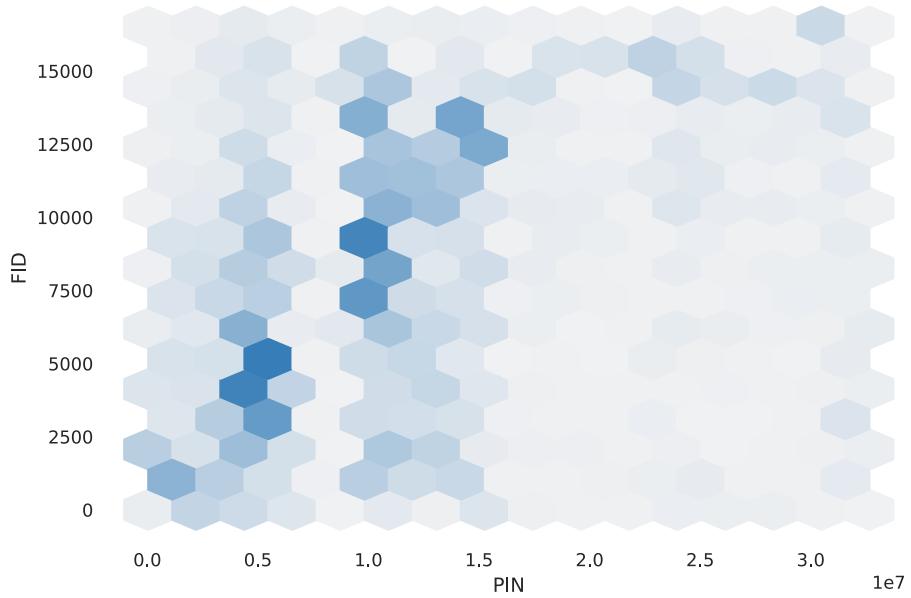


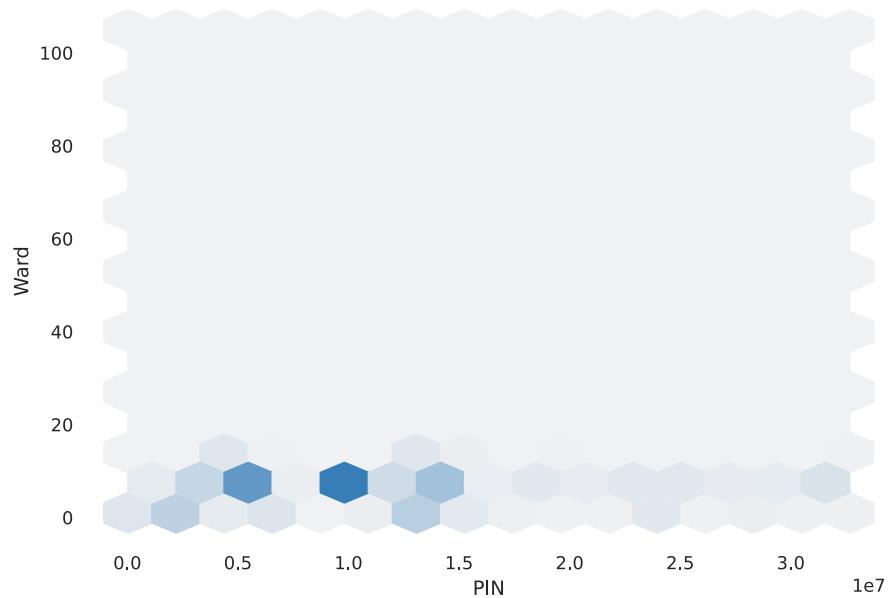
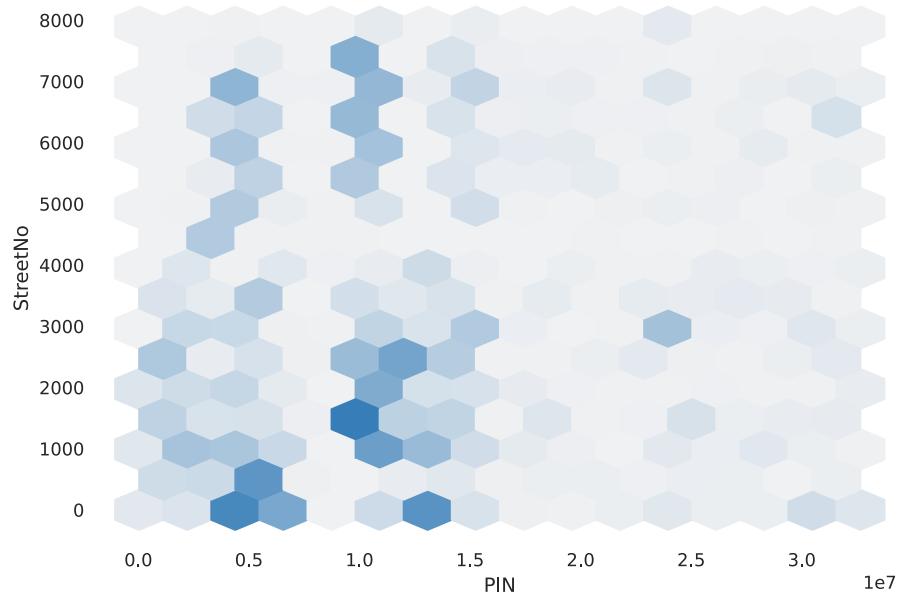


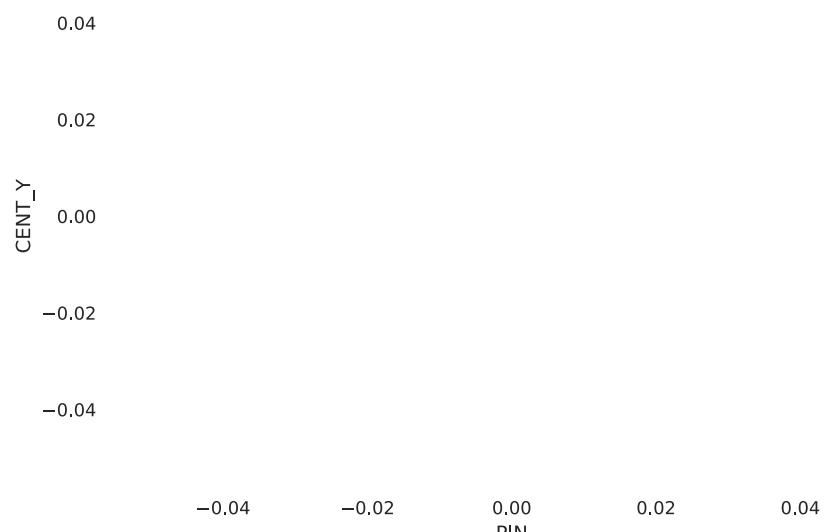
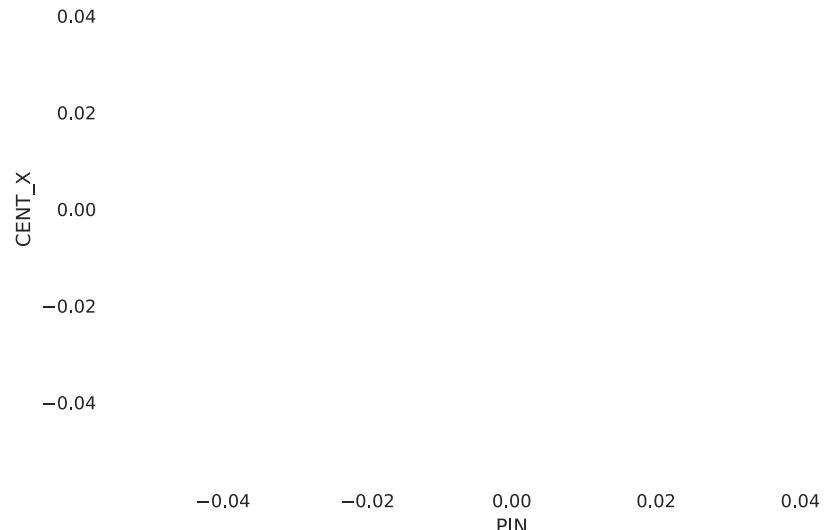


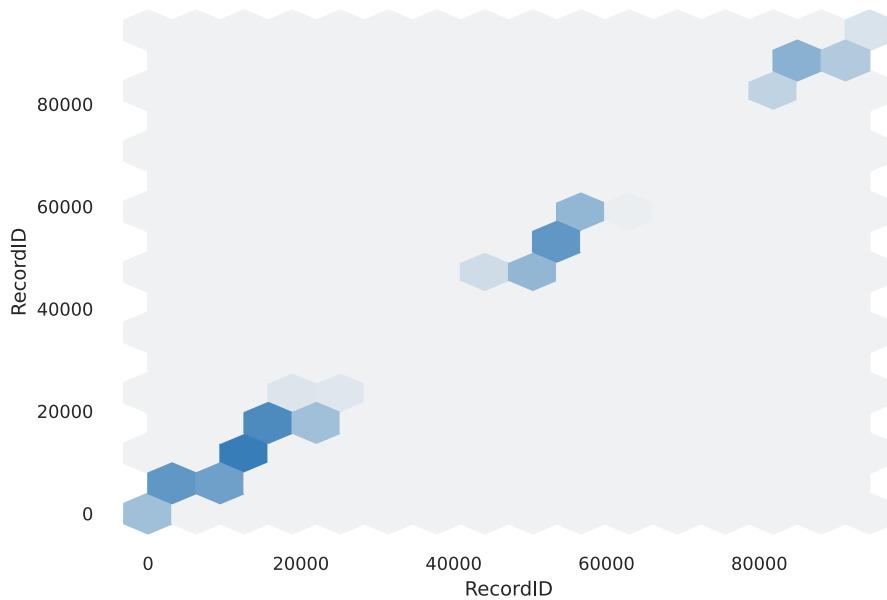
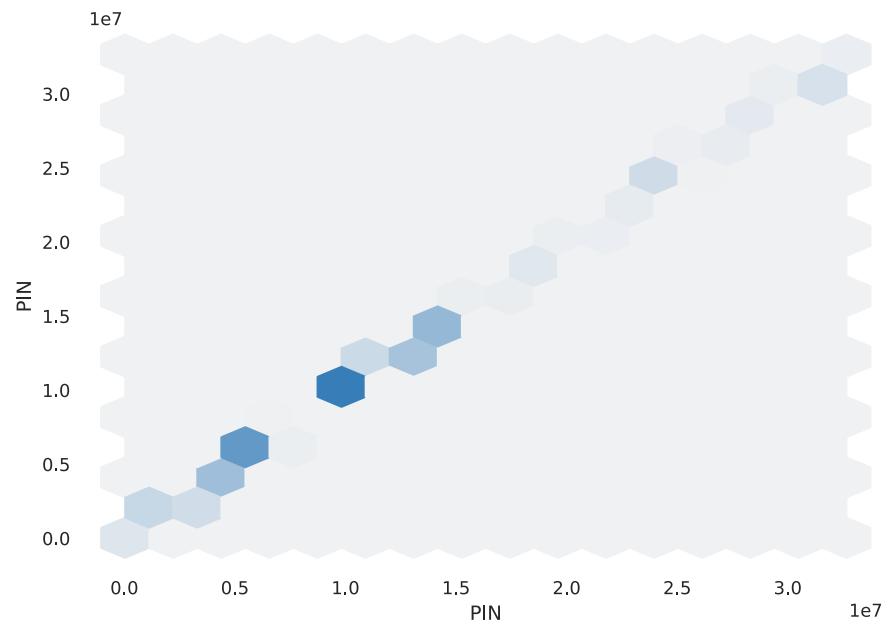


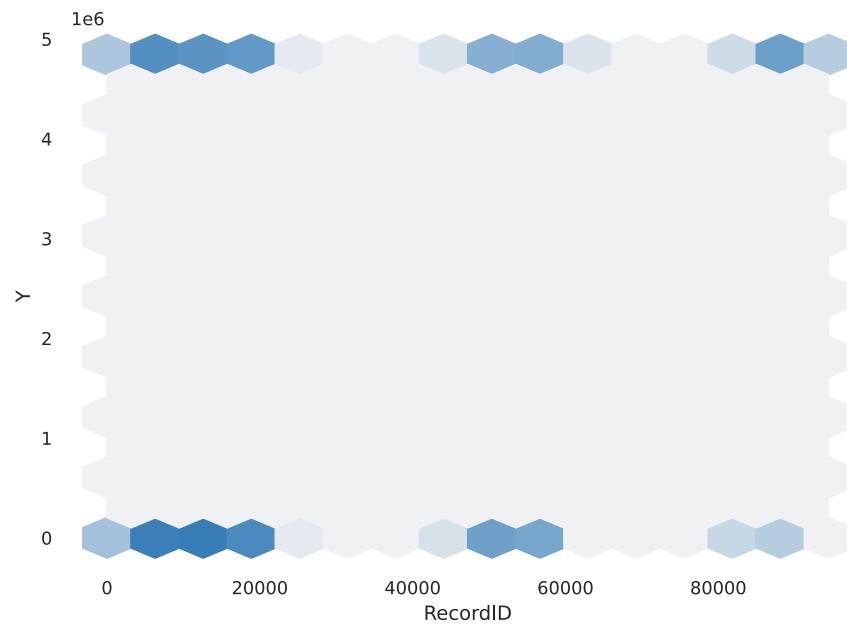
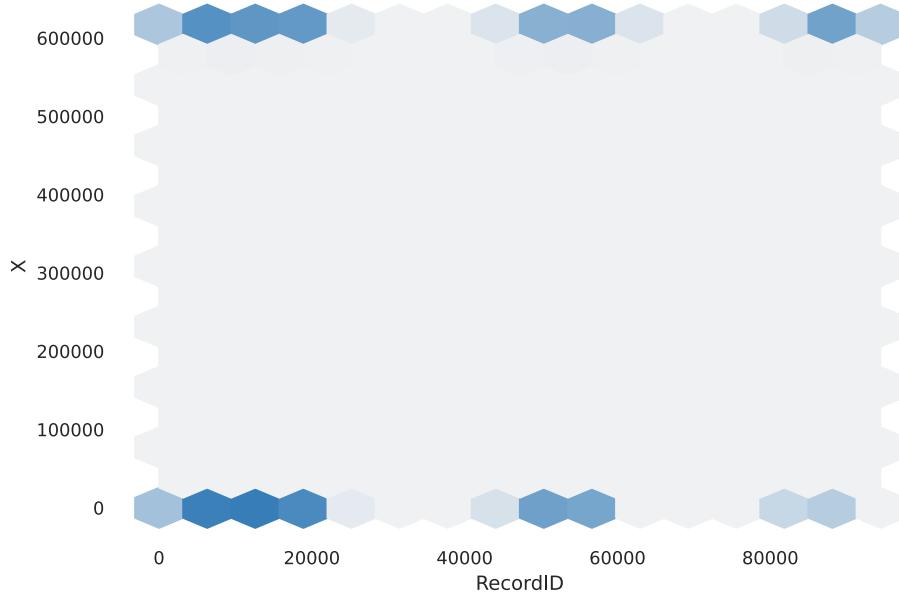


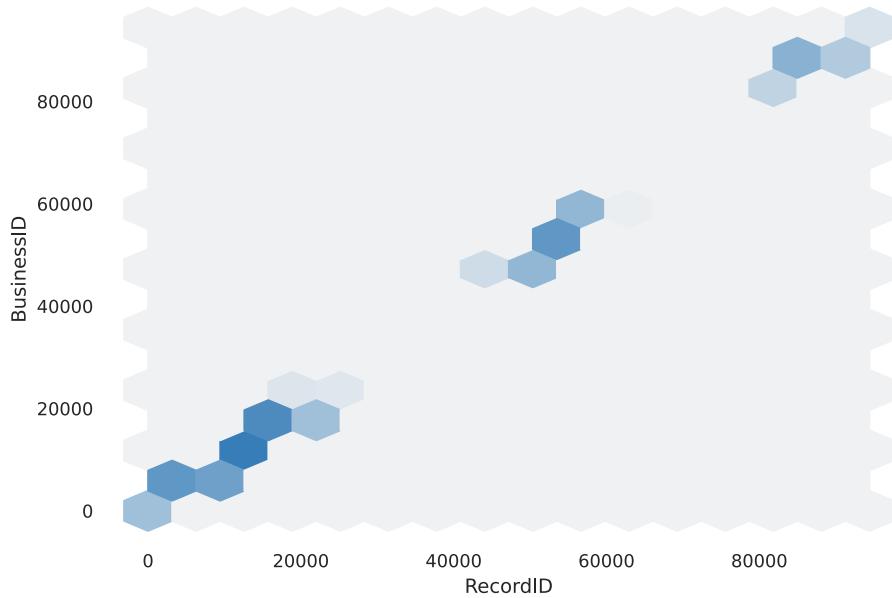
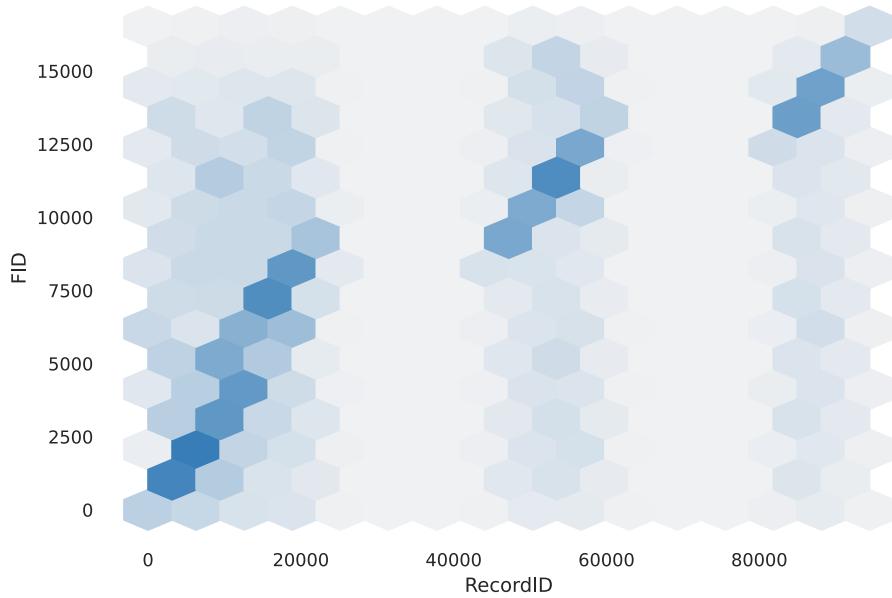


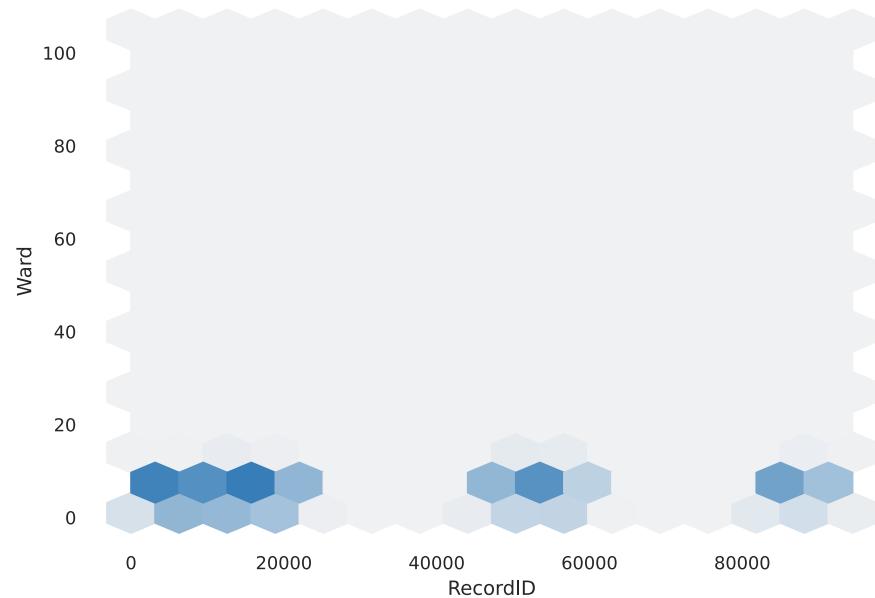
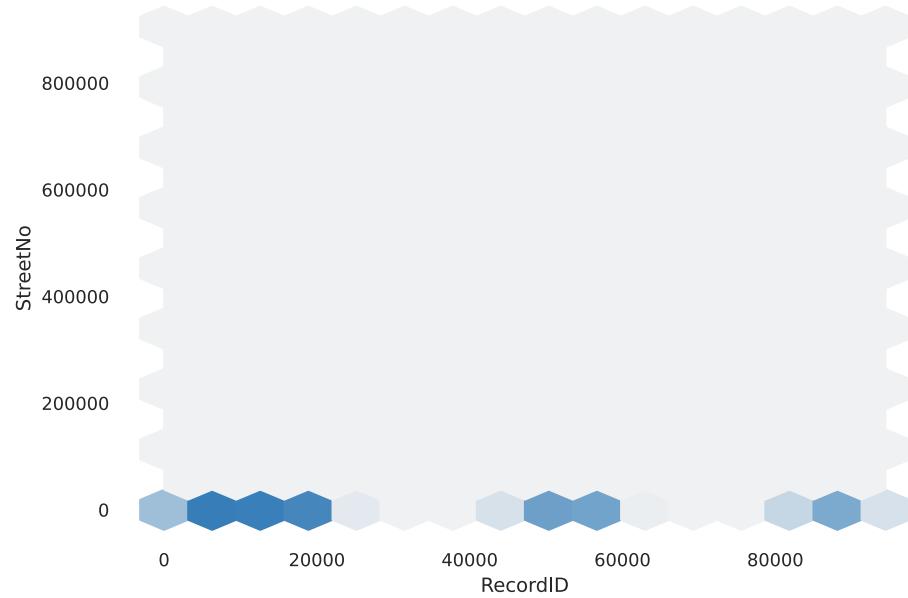


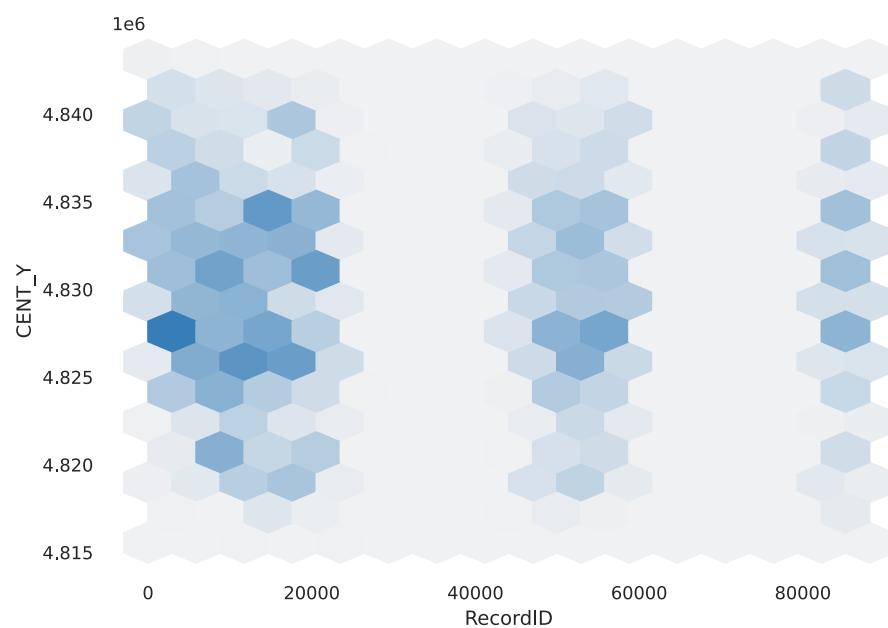
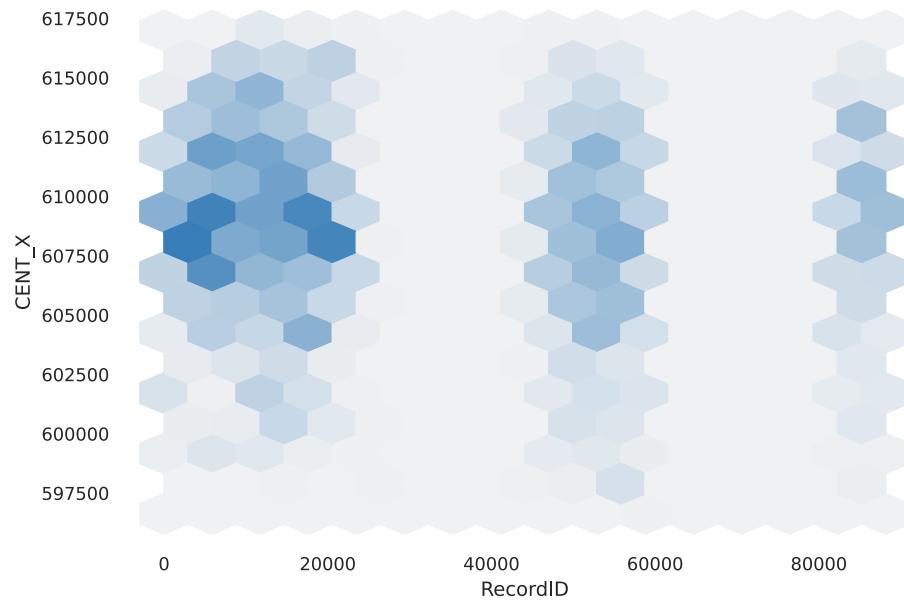


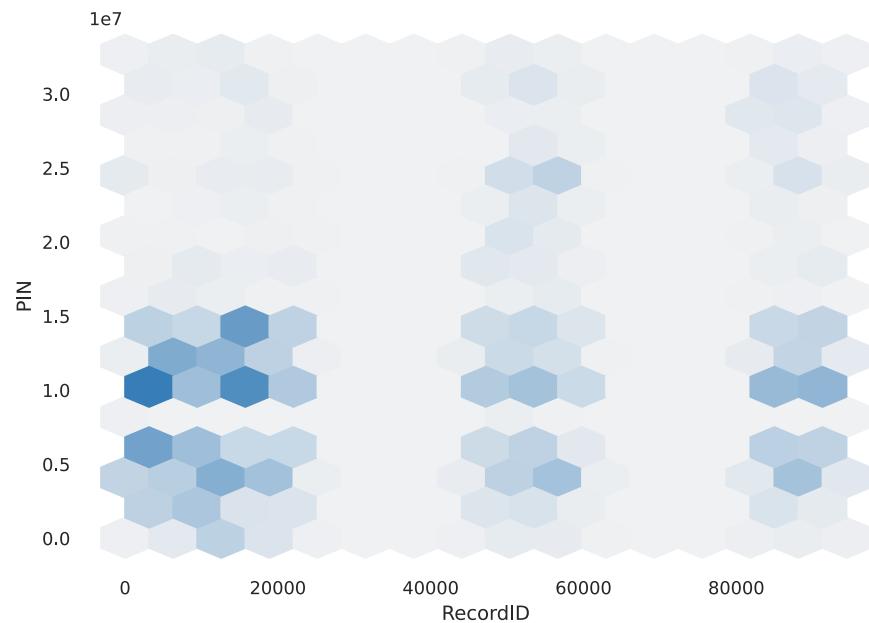




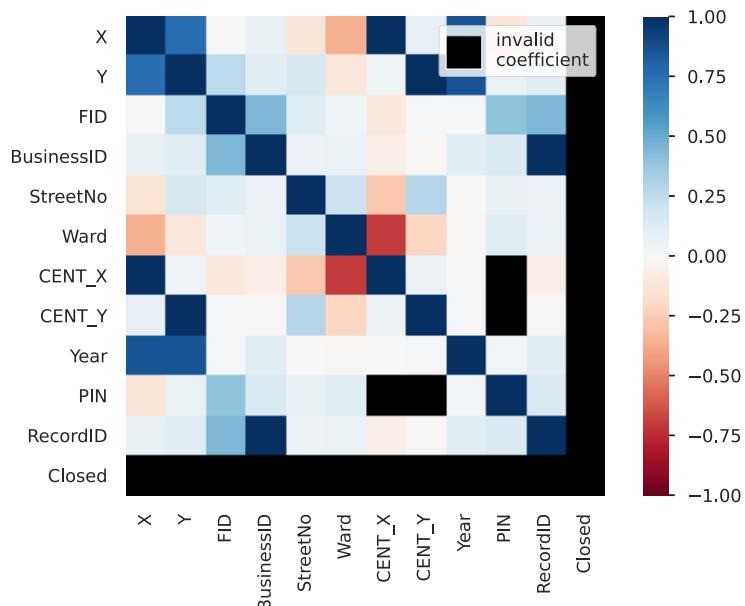


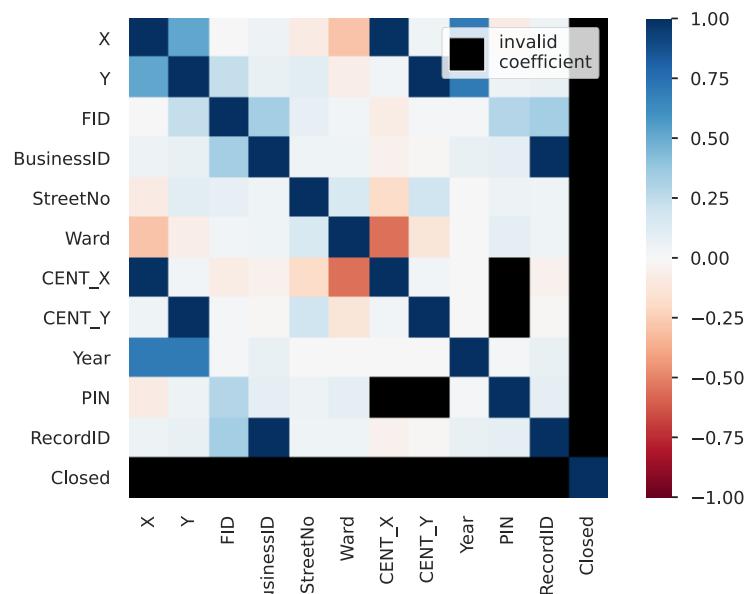
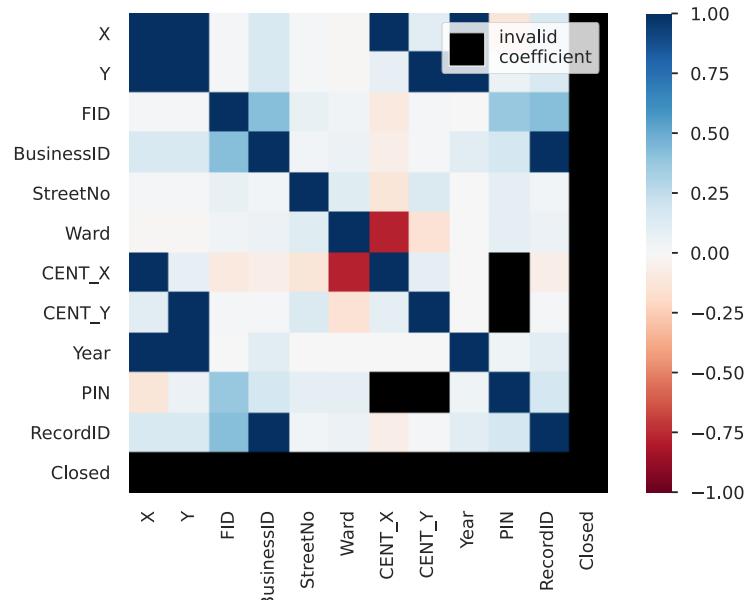


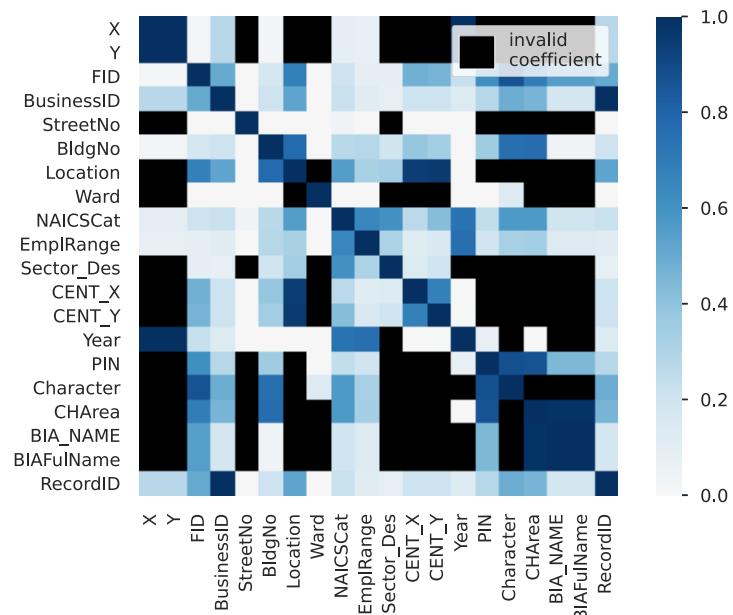
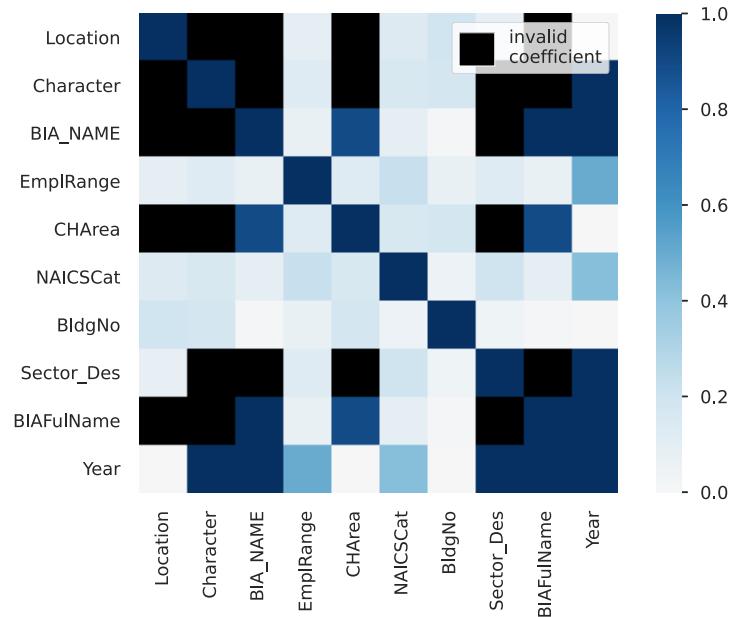




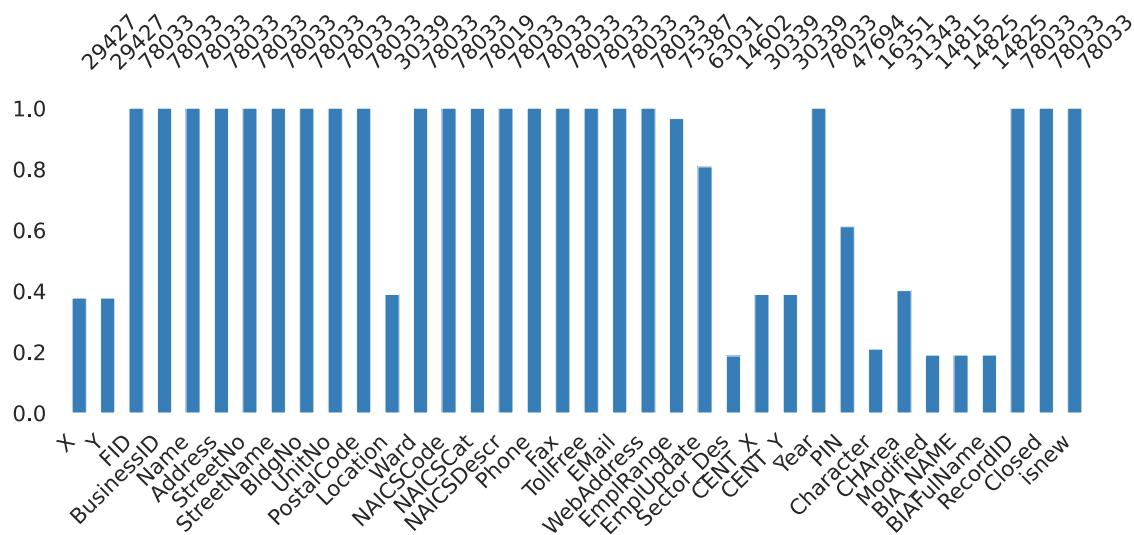
## Correlations



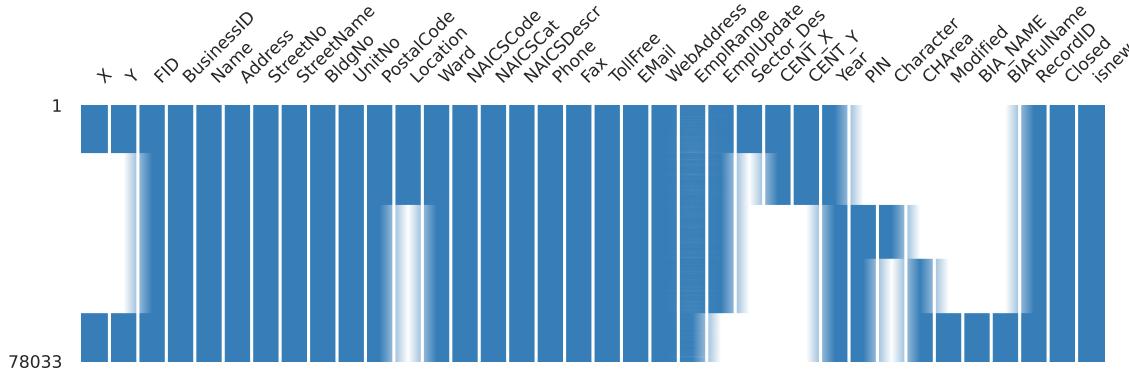




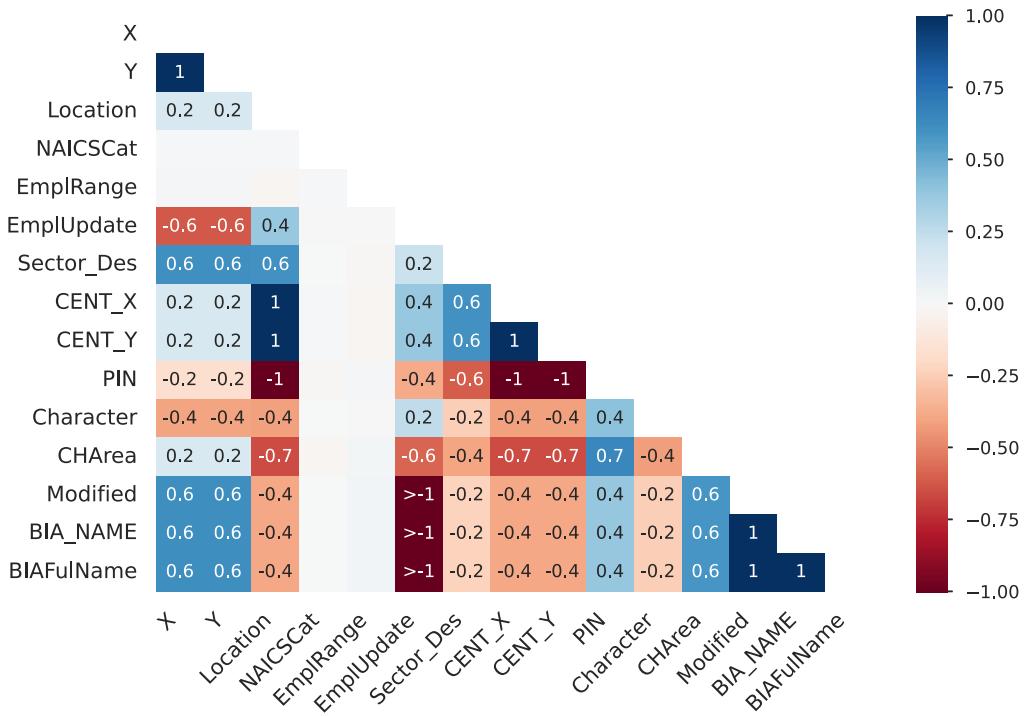
## Missing values



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

X	Y	FID	BusinessID	Name	Address	StreetNo	StreetName	BldgNo	UnitNo	PostalCo
0	-79.689829	43.644181	1	1055	Golf Trends Inc.	300 Ambassador Dr	300			L5T 2J3
1	-79.689419	43.644988	2	1057	Apex Graphics Inc.	320 Ambassador Dr	320			L5T 2J3
2	-79.689419	43.644988	3	1058	Sands, John & Associates Limited	320 Ambassador Dr	320			L5T 2J3
3	-79.689419	43.644988	4	1060	Printmedia-Tackaberry Times	320 Ambassador Dr	320			L5T 2J3
4	-79.690664	43.645493	5	1061	S W R Industries Ltd.	321 Ambassador Dr	321			L5T 2J3
5	-79.690277	43.646372	6	1063	Crossdock Freight Solutions	361 Ambassador Dr	361			L5T 2J3
6	-79.689877	43.646914	7	1065	Green Belting Industries Ltd.	381 Ambassador Dr	381			L5T 2J3

<b>7</b>	-79.634279	43.640404	8	1073	Dafco Filtration Group Corporation	5390 Ambler Dr	5390	Ambler Dr	B	L4W 1G9
<b>8</b>	-79.632844	43.641337	9	1074	Ace Trans Inc.	5391 Ambler Dr	5391	Ambler Dr	1	L4W 1H1
<b>9</b>	-79.637815	43.642638	10	1077	Petro Maxx	5510 Ambler Dr	5510	Ambler Dr	1 to 2	L4W 2V1

	X	Y	FID	BusinessID	Name	Address	StreetNo	StreetName	BldgNo	UnitNo	I
<b>78023</b>	608544.3664	4.840490e+06	14816	57550	Advance Car & Truck Rental	2960 Drew Rd	2960	Drew Rd		149	I
<b>78024</b>	608544.3664	4.840490e+06	14817	57551	Video Palace	2960 Drew Rd	2960	Drew Rd		150	I
<b>78025</b>	608544.3664	4.840490e+06	14818	57552	Secure Life Insurance Agency Inc.	2960 Drew Rd	2960	Drew Rd		151	I
<b>78026</b>	608544.3664	4.840490e+06	14819	57555	Skillman Flooring	2960 Drew Rd	2960	Drew Rd		155&157B	I
<b>78027</b>	608544.3664	4.840490e+06	14820	57557	Verma Vastar Manufacturing Inc.	2960 Drew Rd	2960	Drew Rd		160	I
<b>78028</b>	608544.3664	4.840490e+06	14821	60142	JobsForU	2960 Drew Rd	2960	Drew Rd		156	I
<b>78029</b>	608544.3664	4.840490e+06	14822	60159	Elite Source Solutions	2980 Drew Rd	2980	Drew Rd		133	I
<b>78030</b>	608544.3664	4.840490e+06	14823	60160	Indian Sweet Master	2980 Drew Rd	2980	Drew Rd		134	I
<b>78031</b>	608544.3664	4.840490e+06	14824	60161	Mississauga Flooring & Supplies Inc.	2980 Drew Rd	2980	Drew Rd		135 & 136	I
<b>78032</b>	608544.3664	4.840490e+06	14825	60162	Punjabi Textile Ltd.	2980 Drew Rd	2980	Drew Rd		132	I

Report generated by YData ([https://ydata.ai/?utm\\_source=opensource&utm\\_medium=pandasprofiling&utm\\_campaign=report](https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report)).