

Overview

Dataset statistics

Number of variables	16
Number of observations	354369
Missing cells	181077
Missing cells (%)	3.2%
Duplicate rows	4
Duplicate rows (%)	< 0.1%
Total size in memory	43.3 MiB
Average record size in memory	128.0 B

Variable types

Categorical	9
Numeric	6
Boolean	1

Alerts

NumberOfPictures has constant value "0"	Constant
Dataset has 4 (< 0.1%) duplicate rows	Duplicates
DateCrawled has a high cardinality: 271174 distinct values	High cardinality
Model has a high cardinality: 250 distinct values	High cardinality
DateCreated has a high cardinality: 109 distinct values	High cardinality
LastSeen has a high cardinality: 179150 distinct values	High cardinality
Price is highly overall correlated with RegistrationYear and 1 other fields (RegistrationYear, Power)	High correlation
RegistrationYear is highly overall correlated with Price and 1 other fields (Price, VehicleType)	High correlation
Power is highly overall correlated with Price	High correlation
VehicleType is highly overall correlated with RegistrationYear	High correlation
FuelType is highly imbalanced (63.0%)	Imbalance
VehicleType has 37490 (10.6%) missing values	Missing
Gearbox has 19833 (5.6%) missing values	Missing
Model has 19705 (5.6%) missing values	Missing
FuelType has 32895 (9.3%) missing values	Missing
Repaired has 71154 (20.1%) missing values	Missing
RegistrationYear is highly skewed ($\gamma_1 = 74.62100514$)	Skewed
Power is highly skewed ($\gamma_1 = 60.15719573$)	Skewed
DateCrawled is uniformly distributed	Uniform
Price has 10772 (3.0%) zeros	Zeros
Power has 40225 (11.4%) zeros	Zeros
RegistrationMonth has 37352 (10.5%) zeros	Zeros

Reproduction

[illegible]

Variables

Select Columns

▼

DateCrawled

Categorical

HIGH CARDINALITY UNIFORM

Distinct	271174
Distinct (%)	76.5%
Missing	0
Missing (%)	0.0%
Memory size	2.7 MiB

Length

Max length	19
Median length	19
Mean length	19
Min length	19

Characters and Unicode

Total characters	6733011	
Distinct characters	13	
Distinct categories	4 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	194743	?
Unique (%)	55.0%	

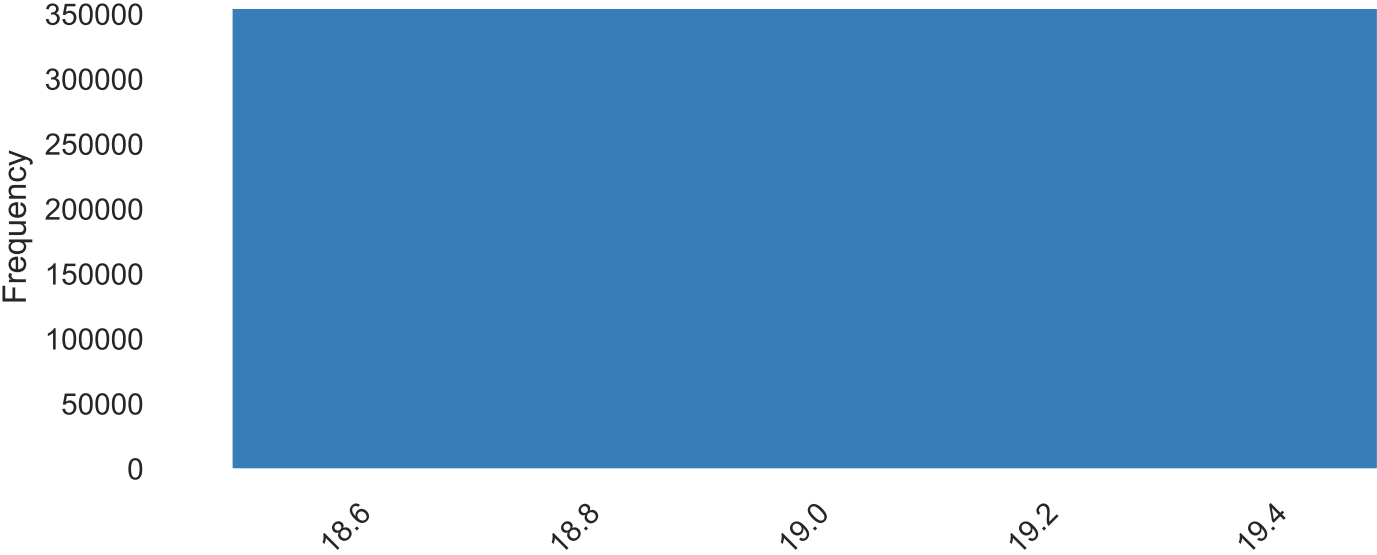
Sample

1st row	2016-03-24 11:52:17
2nd row	2016-03-24 10:58:45
3rd row	2016-03-14 12:52:21
4th row	2016-03-17 16:54:04
5th row	2016-03-31 17:25:20

Common Values

Value	Count	Frequency (%)
2016-03-24 14:49:47	7	< 0.1%
2016-03-26 22:57:31	6	< 0.1%
2016-03-19 21:49:56	6	< 0.1%
2016-03-19 12:50:39	5	< 0.1%
2016-03-31 18:50:16	5	< 0.1%
2016-03-28 10:36:47	5	< 0.1%
2016-04-01 19:25:23	5	< 0.1%
2016-03-31 17:57:07	5	< 0.1%
2016-03-20 16:50:22	5	< 0.1%
2016-03-29 22:50:49	5	< 0.1%
Other values (271164)	354315	> 99.9%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
2016-04-03	13669	1.9%
2016-04-04	13272	1.9%
2016-03-12	12902	1.8%
2016-03-20	12885	1.8%
2016-03-14	12847	1.8%
2016-03-21	12667	1.8%
2016-03-07	12662	1.8%
2016-03-19	12536	1.8%
2016-03-28	12438	1.8%
2016-04-02	12391	1.7%
Other values (32803)	580469	81.9%

Most occurring characters

Value	Count	Frequency (%)
0	1078846	16.0%
1	906579	13.5%
2	765912	11.4%
-	708738	10.5%
:	708738	10.5%
3	574992	8.5%
6	499889	7.4%
5	357379	5.3%
	354369	5.3%
4	343682	5.1%
Other values (3)	433887	6.4%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	4961166	73.7%
Dash Punctuation	708738	10.5%
Other Punctuation	708738	10.5%
Space Separator	354369	5.3%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	1078846	21.7%
1	906579	18.3%
2	765912	15.4%
3	574992	11.6%
6	499889	10.1%
5	357379	7.2%
4	343682	6.9%
7	150912	3.0%
8	142331	2.9%
9	140644	2.8%

Dash Punctuation

Value	Count	Frequency (%)
-	708738	100.0%

Other Punctuation

Value	Count	Frequency (%)
:	708738	100.0%

Space Separator

Value	Count	Frequency (%)
	354369	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	6733011	100.0%

Most frequent character per script

Common		
Value	Count	Frequency (%)
0	1078846	16.0%
1	906579	13.5%
2	765912	11.4%
-	708738	10.5%
:	708738	10.5%
3	574992	8.5%
6	499889	7.4%
5	357379	5.3%
	354369	5.3%
4	343682	5.1%
Other values (3)	433887	6.4%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	6733011	100.0%

Most frequent character per block

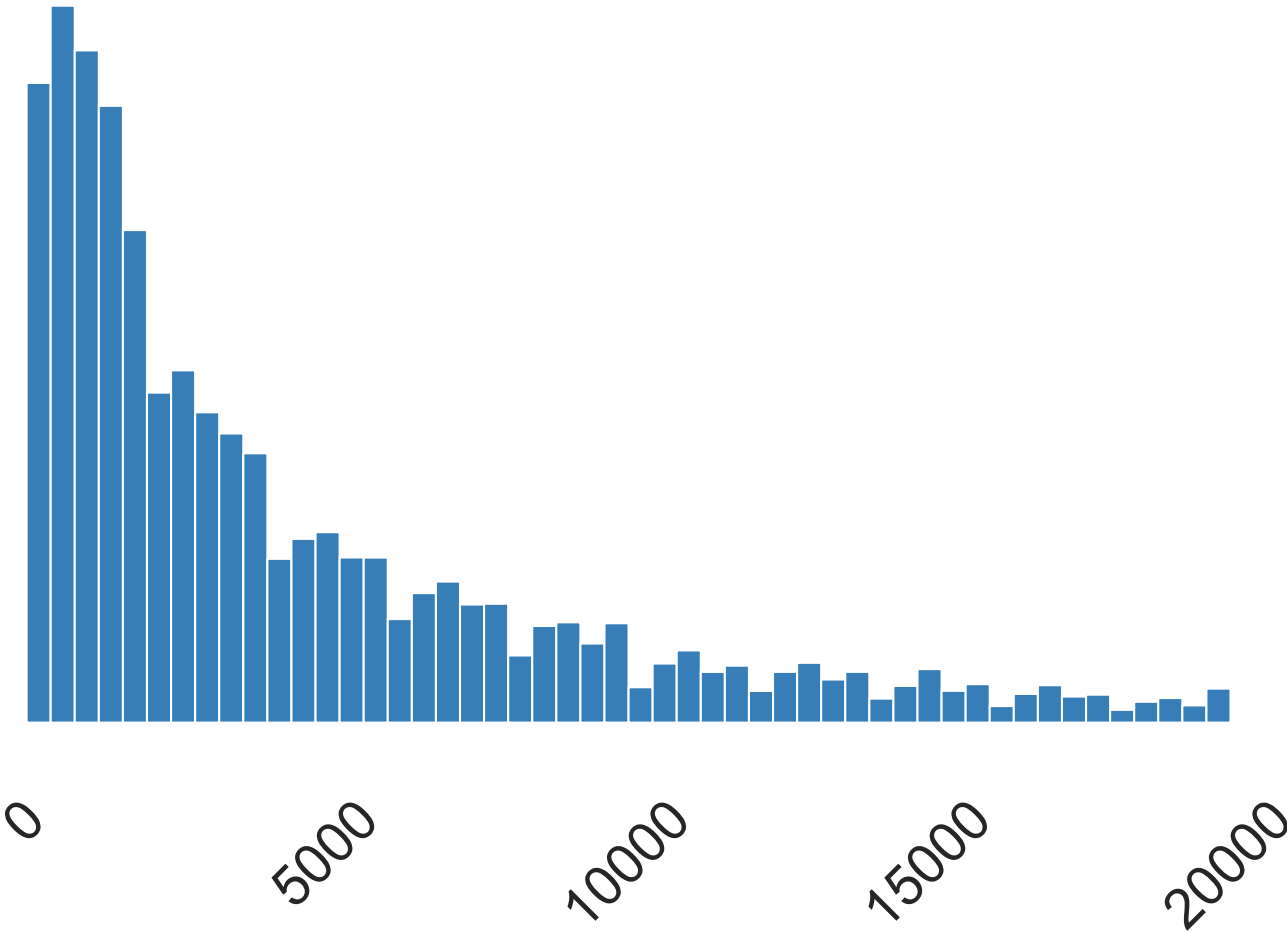
ASCII

Value	Count	Frequency (%)
0	1078846	16.0%
1	906579	13.5%
2	765912	11.4%
-	708738	10.5%
:	708738	10.5%
3	574992	8.5%
6	499889	7.4%
5	357379	5.3%
	354369	5.3%
4	343682	5.1%
Other values (3)	433887	6.4%

Price
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: RegistrationYear, Power) ZEROS

Distinct	3731
Distinct (%)	1.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4416.6568
Minimum	0
Maximum	20000
Zeros	10772
Zeros (%)	3.0%
Negative	0
Negative (%)	0.0%
Memory size	2.7 MiB



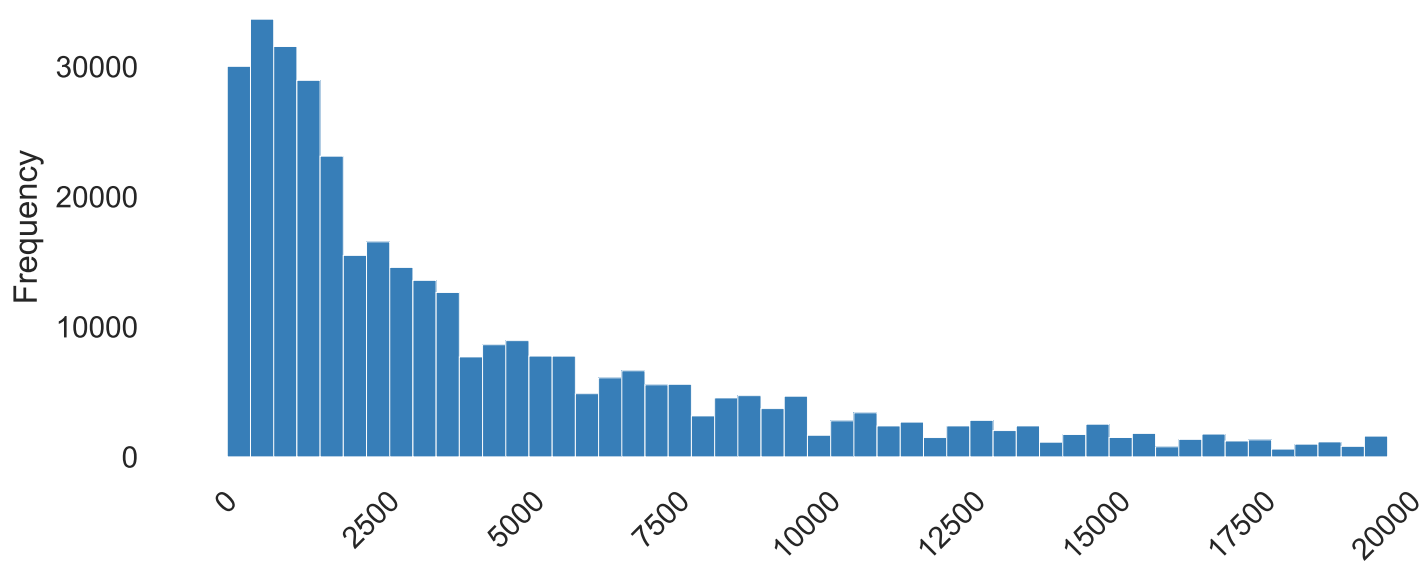
Quantile statistics

Minimum	0
5-th percentile	200
Q1	1050
median	2700

Q3	6400
95-th percentile	14600
Maximum	20000
Range	20000
Interquartile range (IQR)	5350

Descriptive statistics

Standard deviation	4514.1585
Coefficient of variation (CV)	1.0220759
Kurtosis	1.4106377
Mean	4416.6568
Median Absolute Deviation (MAD)	2000
Skewness	1.4306087
Sum	1.5651262×10^9
Variance	20377627
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	10772	3.0%
500	5670	1.6%
1500	5394	1.5%
1000	4649	1.3%
1200	4594	1.3%
2500	4438	1.3%
600	3819	1.1%
3500	3792	1.1%
800	3784	1.1%
2000	3431	1.0%
Other values (3721)	304026	85.8%

Value	Count	Frequency (%)
0	10772	3.0%
1	1189	0.3%
2	12	< 0.1%
3	8	< 0.1%
4	1	< 0.1%
5	26	< 0.1%
7	3	< 0.1%
8	9	< 0.1%
9	8	< 0.1%
10	84	< 0.1%

Value	Count	Frequency (%)
20000	268	0.1%
19999	278	0.1%
19998	6	< 0.1%
19997	1	< 0.1%
19995	10	< 0.1%
19990	178	0.1%
19989	1	< 0.1%
19987	2	< 0.1%
19986	1	< 0.1%
19985	2	< 0.1%

VehicleType

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: RegistrationYear)			MISSING
Distinct		8	
Distinct (%)		< 0.1%	
Missing		37490	
Missing (%)		10.6%	
Memory size		2.7 MiB	

Length

Max length	11
Median length	5
Mean length	5.1252087
Min length	3

Characters and Unicode

Total characters	1624071	
Distinct characters	19	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	0	?
Unique (%)	0.0%	

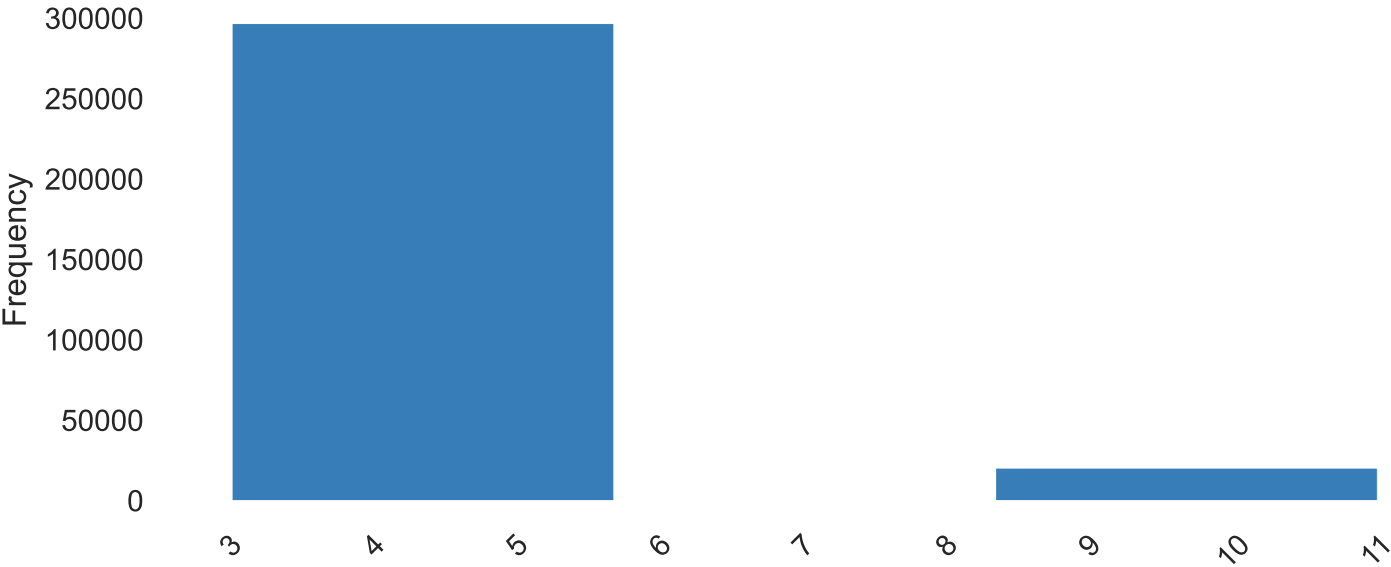
Sample

1st row	coupe
2nd row	suv
3rd row	small
4th row	small
5th row	sedan

Common Values

Value	Count	Frequency (%)
sedan	91457	25.8%
small	79831	22.5%
wagon	65166	18.4%
bus	28775	8.1%
convertible	20203	5.7%
coupe	16163	4.6%
suv	11996	3.4%
other	3288	0.9%
(Missing)	37490	10.6%

Length



Histogram of lengths of the category

Common Values (Plot)



-  sedan
-  small
-  wagon
-  bus
-  convertible
-  coupe
-  suv
-  other

Value	Count	Frequency (%)
sedan	91457	28.9%
small	79831	25.2%
wagon	65166	20.6%
bus	28775	9.1%
convertible	20203	6.4%
coupe	16163	5.1%
suv	11996	3.8%
other	3288	1.0%

Most occurring characters

Value	Count	Frequency (%)
a	236454	14.6%
s	212059	13.1%
l	179865	11.1%
n	176826	10.9%
e	151314	9.3%
o	104820	6.5%
d	91457	5.6%
m	79831	4.9%
w	65166	4.0%
g	65166	4.0%
Other values (9)	261113	16.1%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1624071	100.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	236454	14.6%
s	212059	13.1%
l	179865	11.1%
n	176826	10.9%
e	151314	9.3%
o	104820	6.5%
d	91457	5.6%
m	79831	4.9%
w	65166	4.0%
g	65166	4.0%
Other values (9)	261113	16.1%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1624071	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	236454	14.6%
s	212059	13.1%
l	179865	11.1%
n	176826	10.9%
e	151314	9.3%
o	104820	6.5%
d	91457	5.6%
m	79831	4.9%
w	65166	4.0%
g	65166	4.0%
Other values (9)	261113	16.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1624071	100.0%

Most frequent character per block

ASCII

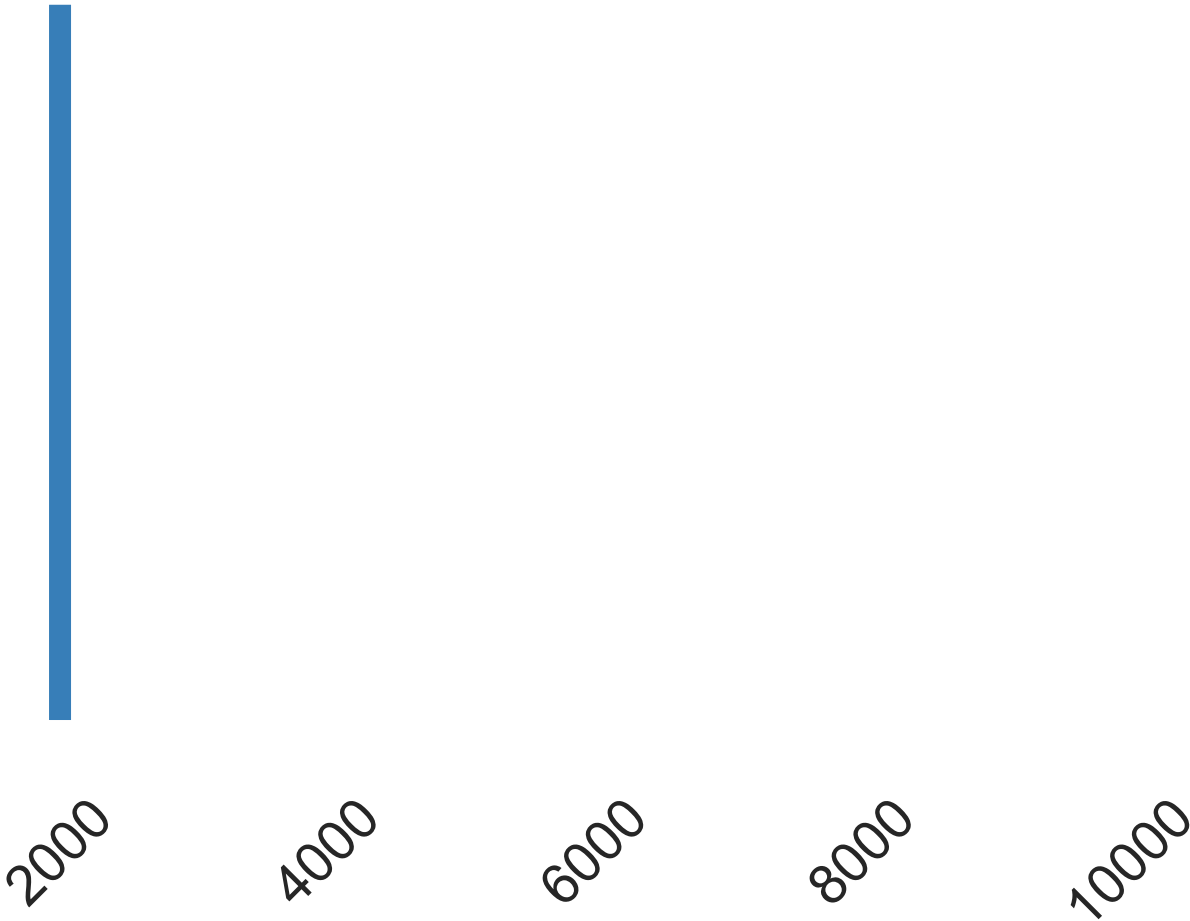
Value	Count	Frequency (%)
a	236454	14.6%
s	212059	13.1%
l	179865	11.1%
n	176826	10.9%
e	151314	9.3%
o	104820	6.5%
d	91457	5.6%
m	79831	4.9%
w	65166	4.0%
g	65166	4.0%
Other values (9)	261113	16.1%

RegistrationYear

Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: Price, VehicleType) SKEWED

Distinct	151
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2004.2344
Minimum	1000
Maximum	9999
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	2.7 MiB



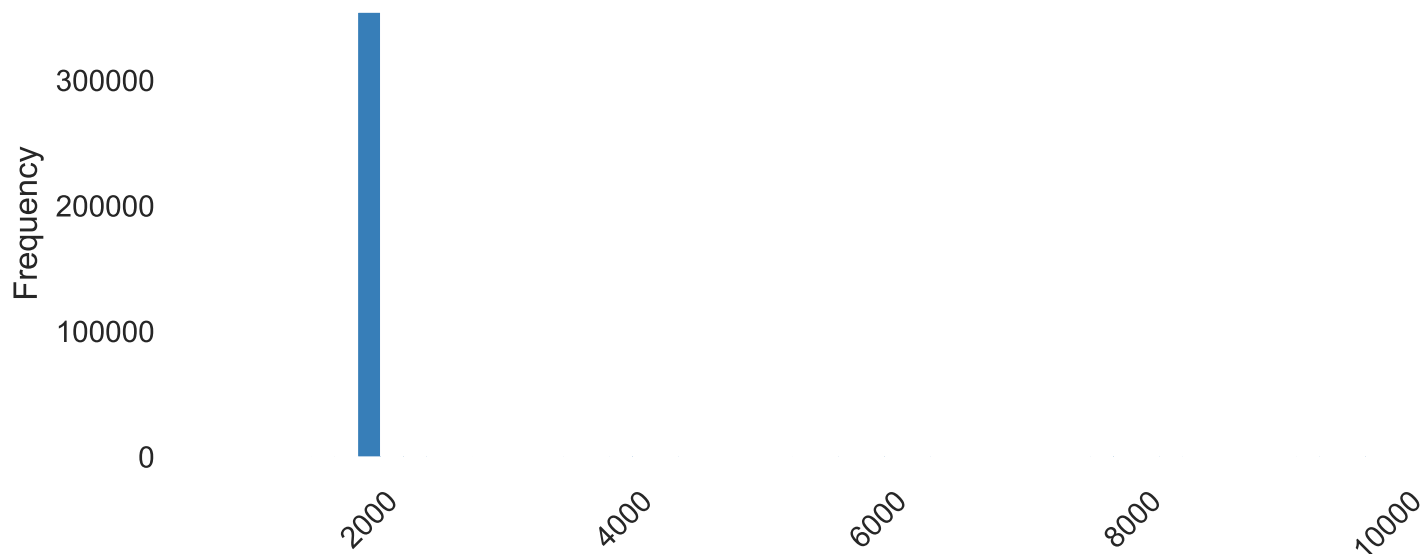
Quantile statistics

Minimum	1000
5-th percentile	1992
Q1	1999
median	2003

Q3	2008
95-th percentile	2016
Maximum	9999
Range	8999
Interquartile range (IQR)	9

Descriptive statistics

Standard deviation	90.227958
Coefficient of variation (CV)	0.045018664
Kurtosis	6083.102
Mean	2004.2344
Median Absolute Deviation (MAD)	4
Skewness	74.621005
Sum	7.1023856×10^8
Variance	8141.0844
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
2000	24490	6.9%
1999	22728	6.4%
2005	22109	6.2%
2001	20124	5.7%
2006	19900	5.6%
2003	19763	5.6%
2004	19611	5.5%
2002	19092	5.4%
1998	17899	5.1%
2007	17183	4.8%
Other values (141)	151470	42.7%

Value	Count	Frequency (%)
1000	37	< 0.1%
1001	1	< 0.1%
1039	1	< 0.1%
1111	3	< 0.1%
1200	1	< 0.1%
1234	4	< 0.1%
1253	1	< 0.1%
1255	1	< 0.1%
1300	2	< 0.1%
1400	1	< 0.1%

Value	Count	Frequency (%)
9999	26	< 0.1%
9996	1	< 0.1%
9450	1	< 0.1%
9229	1	< 0.1%
9000	3	< 0.1%
8888	1	< 0.1%
8500	1	< 0.1%
8455	1	< 0.1%
8200	1	< 0.1%
8000	2	< 0.1%

Gearbox

Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	19833
Missing (%)	5.6%
Memory size	2.7 MiB

Length

Max length	6
Median length	6
Mean length	5.6037198
Min length	4

Characters and Unicode

Total characters	1874646	
Distinct characters	7	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	0	?
Unique (%)	0.0%	

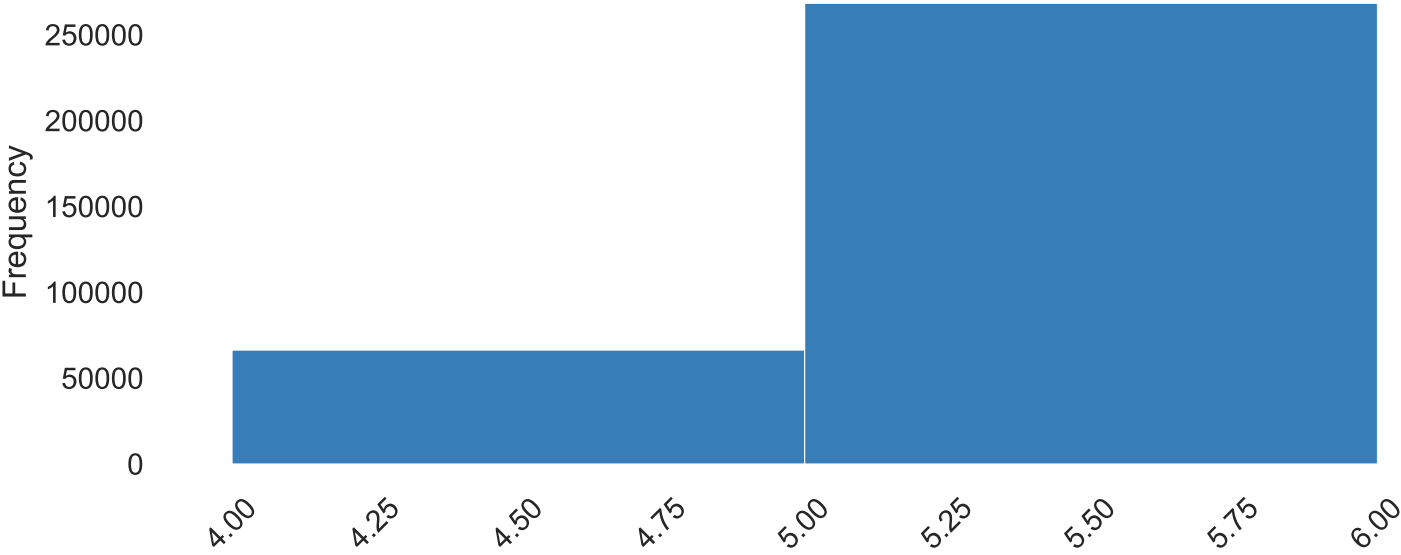
Sample

1st row	manual
2nd row	manual
3rd row	auto
4th row	manual
5th row	manual

Common Values

Value	Count	Frequency (%)
manual	268251	75.7%
auto	66285	18.7%
(Missing)	19833	5.6%

Length



Histogram of lengths of the category

Common Values (Plot)



manual

auto

Value	Count	Frequency (%)
manual	268251	80.2%
auto	66285	19.8%

Most occurring characters

Value	Count	Frequency (%)
a	602787	32.2%
u	334536	17.8%
m	268251	14.3%
n	268251	14.3%
l	268251	14.3%
t	66285	3.5%
o	66285	3.5%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1874646	100.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	602787	32.2%
u	334536	17.8%
m	268251	14.3%
n	268251	14.3%
l	268251	14.3%
t	66285	3.5%
o	66285	3.5%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1874646	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	602787	32.2%
u	334536	17.8%
m	268251	14.3%
n	268251	14.3%
l	268251	14.3%
t	66285	3.5%
o	66285	3.5%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1874646	100.0%

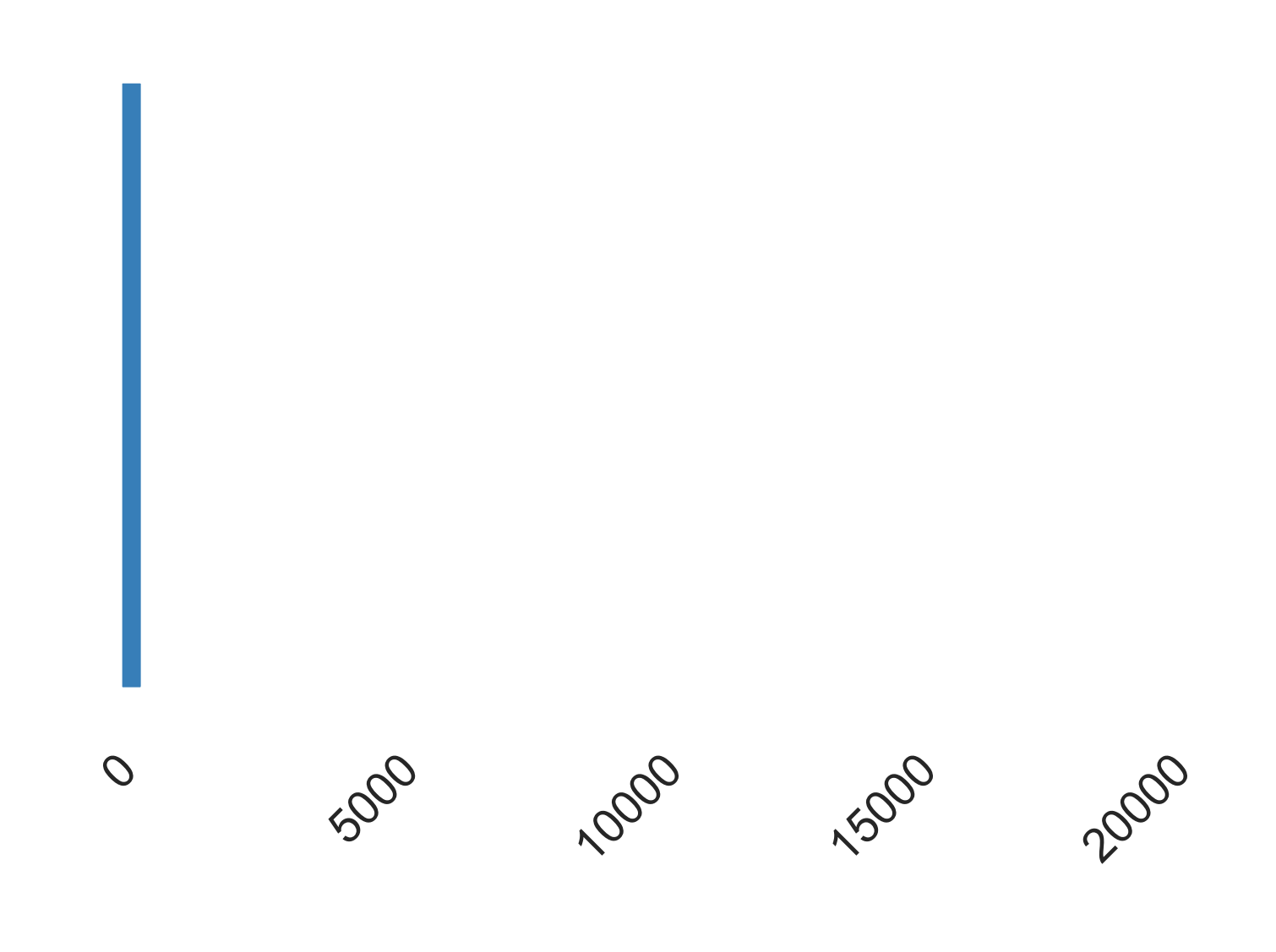
Most frequent character per block

ASCII

Value	Count	Frequency (%)
a	602787	32.2%
u	334536	17.8%
m	268251	14.3%
n	268251	14.3%
l	268251	14.3%
t	66285	3.5%
o	66285	3.5%

Power
Real number (\mathbb{R})

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Price)			SKEWED	ZEROS
Distinct	712			
Distinct (%)	0.2%			
Missing	0			
Missing (%)	0.0%			
Infinite	0			
Infinite (%)	0.0%			
Mean	110.09434			
Minimum		0		
Maximum		20000		
Zeros		40225		
Zeros (%)		11.4%		
Negative		0		
Negative (%)		0.0%		
Memory size		2.7 MiB		



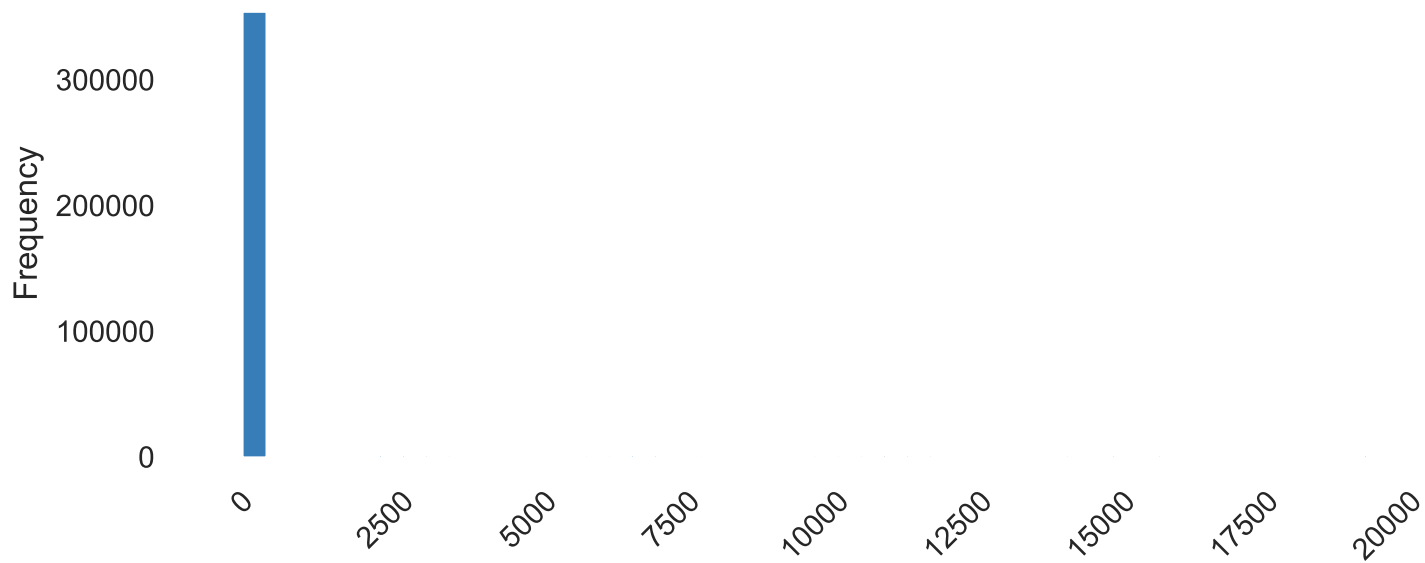
Quantile statistics

Minimum	0
5-th percentile	0
Q1	69
median	105

Q3	143
95-th percentile	218
Maximum	20000
Range	20000
Interquartile range (IQR)	74

Descriptive statistics

Standard deviation	189.85041
Coefficient of variation (CV)	1.7244339
Kurtosis	4624.1032
Mean	110.09434
Median Absolute Deviation (MAD)	36
Skewness	60.157196
Sum	39014020
Variance	36043.176
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
0	40225	11.4%
75	24023	6.8%
60	15897	4.5%
150	14590	4.1%
101	13298	3.8%
140	12852	3.6%
90	12734	3.6%
116	11918	3.4%
105	10355	2.9%
170	10121	2.9%
Other values (702)	188356	53.2%

Value	Count	Frequency (%)
0	40225	11.4%
1	34	< 0.1%
2	10	< 0.1%
3	9	< 0.1%
4	30	< 0.1%
5	103	< 0.1%
6	11	< 0.1%
7	11	< 0.1%
8	7	< 0.1%
9	5	< 0.1%

Value	Count	Frequency (%)
20000	1	< 0.1%
19312	1	< 0.1%
19211	1	< 0.1%
19208	1	< 0.1%
17932	1	< 0.1%
17700	1	< 0.1%
17410	1	< 0.1%
17019	1	< 0.1%
17011	1	< 0.1%
16312	1	< 0.1%

Model
Categorical

HIGH CARDINALITY MISSING

Distinct	250
Distinct (%)	0.1%
Missing	19705
Missing (%)	5.6%
Memory size	2.7 MiB

Length

Max length	18
Median length	11
Mean length	4.9828604
Min length	2

Characters and Unicode

Total characters	1667584	
Distinct characters	37	
Distinct categories	3 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	0	?
Unique (%)	0.0%	

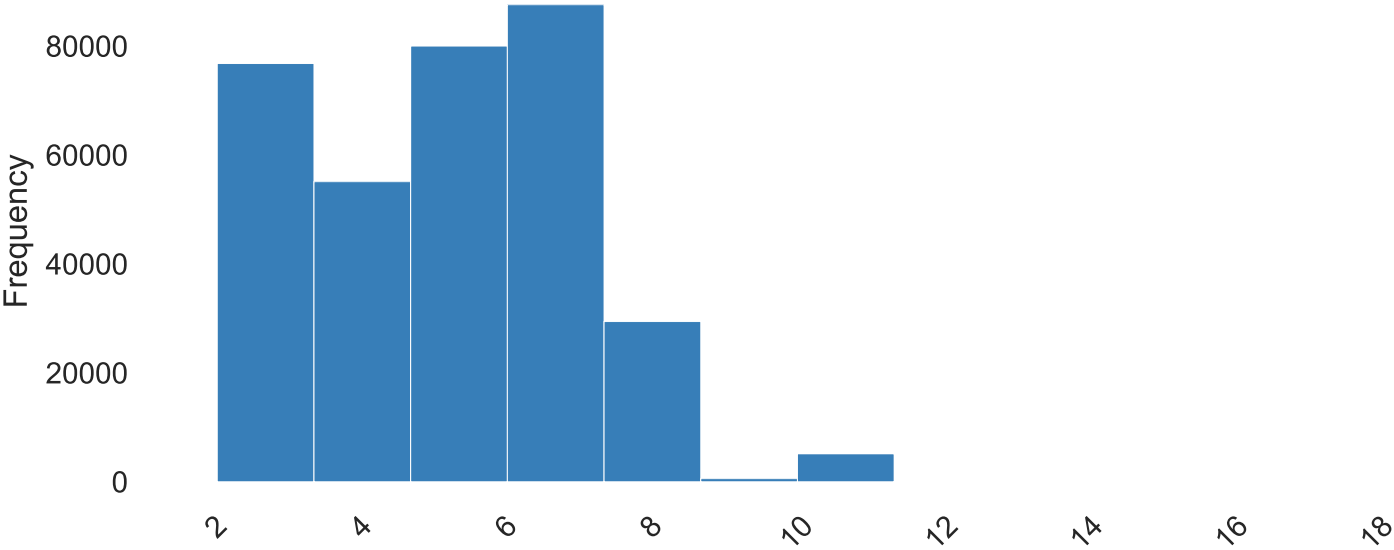
Sample

1st row	golf
2nd row	grand
3rd row	golf
4th row	fabia
5th row	3er

Common Values

Value	Count	Frequency (%)
golf	29232	8.2%
other	24421	6.9%
3er	19761	5.6%
polo	13066	3.7%
corsa	12570	3.5%
astra	10809	3.1%
passat	10146	2.9%
a4	9900	2.8%
c_klasse	8312	2.3%
5er	7868	2.2%
Other values (240)	188579	53.2%
(Missing)	19705	5.6%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
golf	29232	8.7%
other	24421	7.3%
3er	19761	5.9%
polo	13066	3.9%
corsa	12570	3.8%
astra	10809	3.2%
passat	10146	3.0%
a4	9900	3.0%
c_klasse	8312	2.5%
5er	7868	2.4%
Other values (240)	188579	56.3%

Most occurring characters

Value	Count	Frequency (%)
a	192434	11.5%
e	175795	10.5%
o	172661	10.4%
r	158911	9.5%
s	127888	7.7%
t	103731	6.2%
l	88012	5.3%
i	71561	4.3%
c	65216	3.9%
f	53803	3.2%
Other values (27)	457572	27.4%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1536484	92.1%
Decimal Number	89324	5.4%
Connector Punctuation	41776	2.5%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	192434	12.5%
e	175795	11.4%
o	172661	11.2%
r	158911	10.3%
s	127888	8.3%
t	103731	6.8%
l	88012	5.7%
i	71561	4.7%
c	65216	4.2%
f	53803	3.5%
Other values (16)	326472	21.2%

Decimal Number

Value	Count	Frequency (%)
3	30054	33.6%
4	12378	13.9%
5	11719	13.1%
1	8397	9.4%
6	8211	9.2%
0	7514	8.4%
2	5610	6.3%
7	2427	2.7%
8	2345	2.6%
9	669	0.7%

Connector Punctuation

Value	Count	Frequency (%)
—	41776	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1536484	92.1%
Common	131100	7.9%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	192434	12.5%
e	175795	11.4%
o	172661	11.2%
r	158911	10.3%
s	127888	8.3%
t	103731	6.8%
l	88012	5.7%
i	71561	4.7%
c	65216	4.2%
f	53803	3.5%
Other values (16)	326472	21.2%

Common

Value	Count	Frequency (%)
—	41776	31.9%
3	30054	22.9%
4	12378	9.4%
5	11719	8.9%
1	8397	6.4%
6	8211	6.3%
0	7514	5.7%
2	5610	4.3%
7	2427	1.9%
8	2345	1.8%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1667584	100.0%

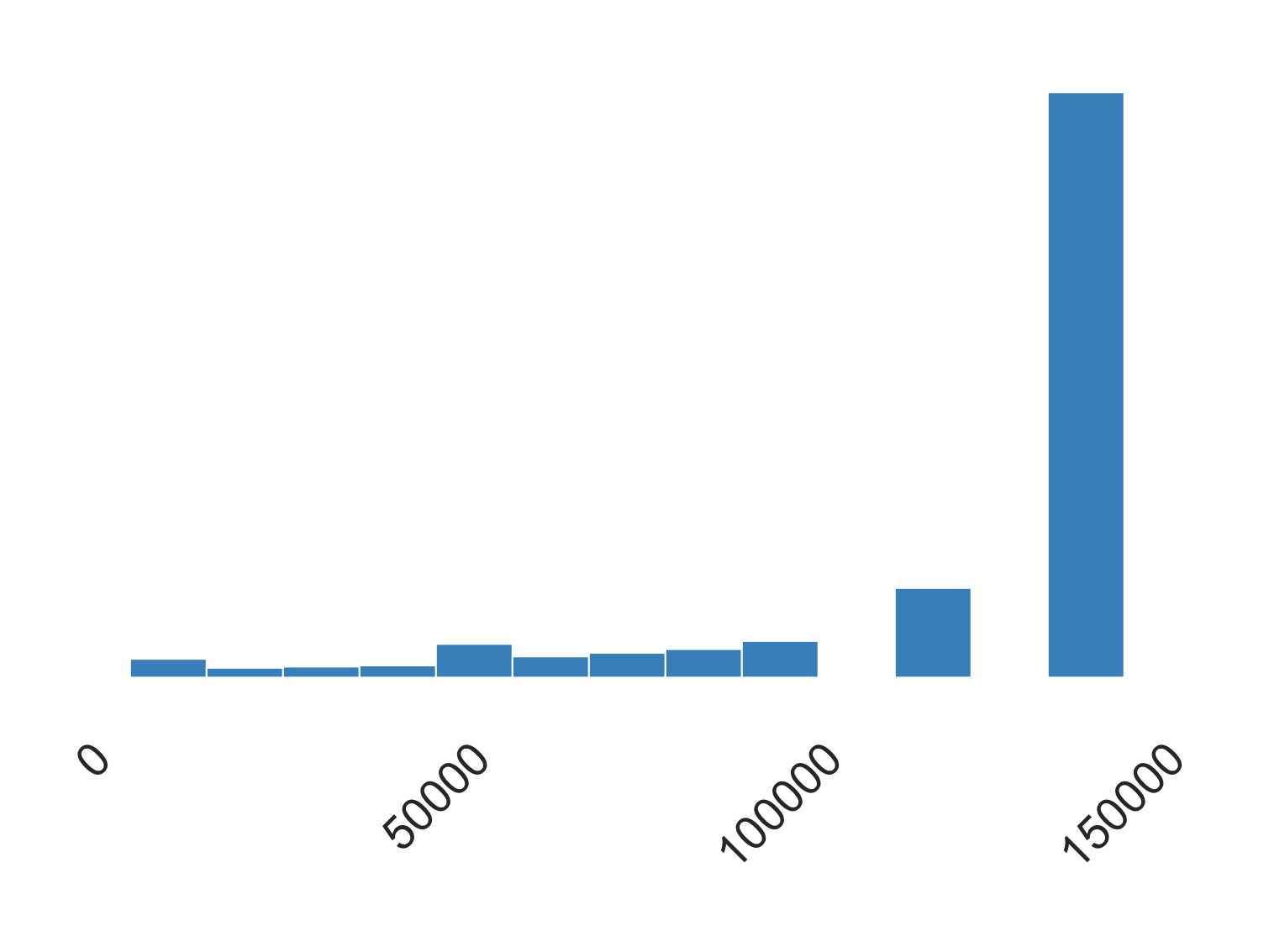
Most frequent character per block

ASCII

Value	Count	Frequency (%)
a	192434	11.5%
e	175795	10.5%
o	172661	10.4%
r	158911	9.5%
s	127888	7.7%
t	103731	6.2%
l	88012	5.3%
i	71561	4.3%
c	65216	3.9%
f	53803	3.2%
Other values (27)	457572	27.4%

Kilometer
Real number (\mathbb{R})

Distinct	13
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	128211.17
Minimum	5000
Maximum	150000
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	2.7 MiB



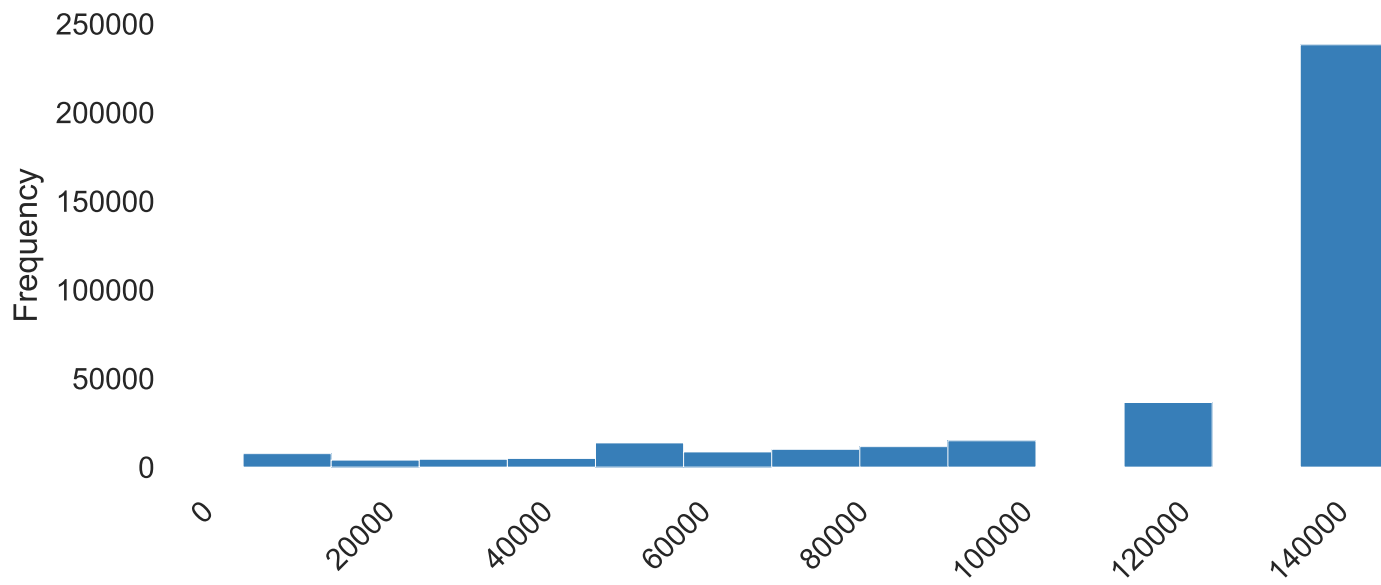
Quantile statistics

Minimum	5000
5-th percentile	40000
Q1	125000
median	150000
Q3	150000

95-th percentile	150000
Maximum	150000
Range	145000
Interquartile range (IQR)	25000

Descriptive statistics

Standard deviation	37905.342
Coefficient of variation (CV)	0.29564773
Kurtosis	1.925124
Mean	128211.17
Median Absolute Deviation (MAD)	0
Skewness	-1.7241192
Sum	4.5434065×10^{10}
Variance	1.4368149×10^9
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=13)

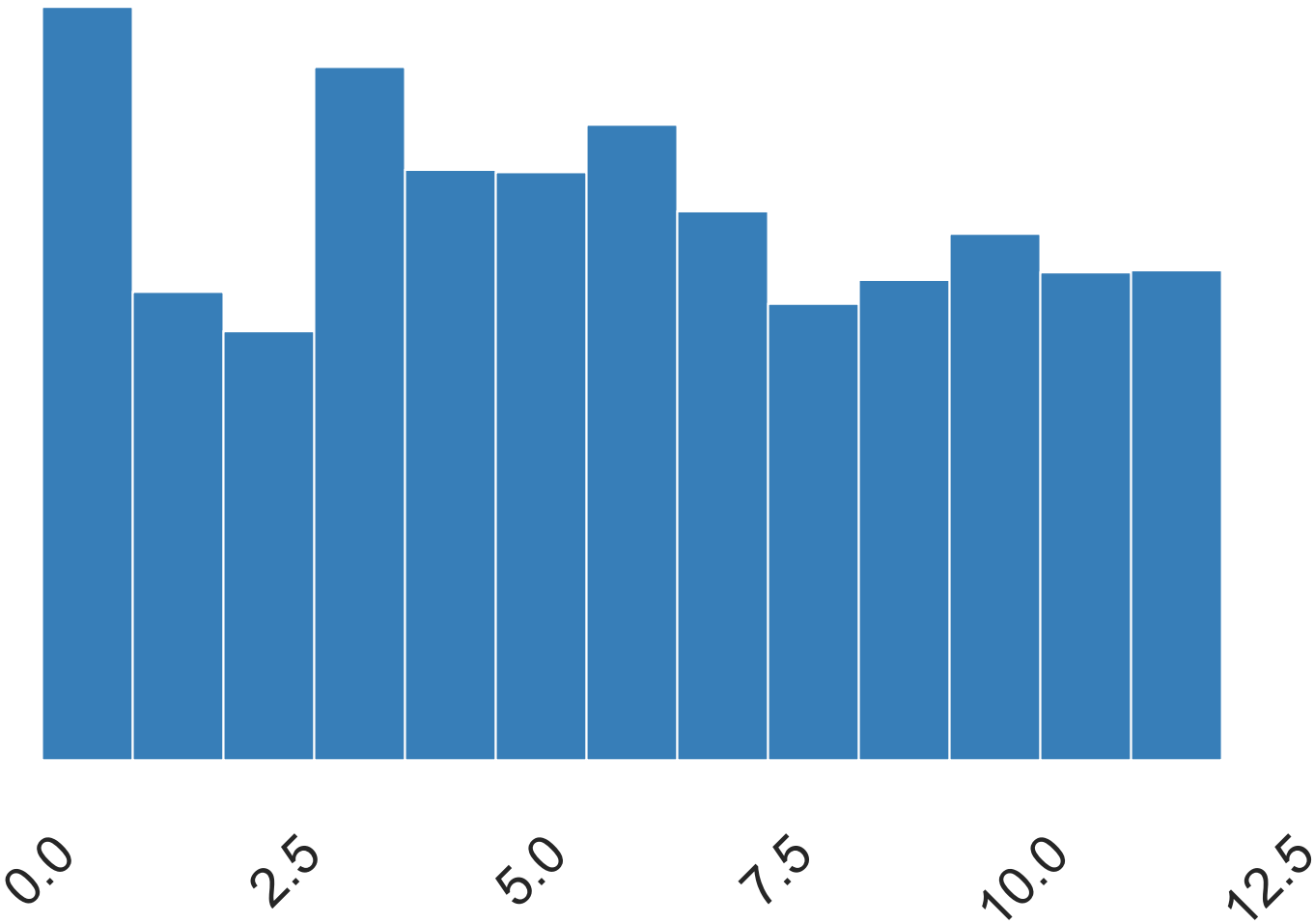
Value	Count	Frequency (%)
150000	238209	67.2%
125000	36454	10.3%
100000	14882	4.2%
90000	11567	3.3%
80000	10047	2.8%
70000	8593	2.4%
60000	7444	2.1%
5000	6397	1.8%
50000	6232	1.8%
40000	4911	1.4%
Other values (3)	9633	2.7%

Value	Count	Frequency (%)
5000	6397	1.8%
10000	1222	0.3%
20000	3975	1.1%
30000	4436	1.3%
40000	4911	1.4%
50000	6232	1.8%
60000	7444	2.1%
70000	8593	2.4%
80000	10047	2.8%
90000	11567	3.3%

Value	Count	Frequency (%)
150000	238209	67.2%
125000	36454	10.3%
100000	14882	4.2%
90000	11567	3.3%
80000	10047	2.8%
70000	8593	2.4%
60000	7444	2.1%
50000	6232	1.8%
40000	4911	1.4%
30000	4436	1.3%

RegistrationMonth
Real number (\mathbb{R})

Distinct	13
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	5.7146449
Minimum	0
Maximum	12
Zeros	37352
Zeros (%)	10.5%
Negative	0
Negative (%)	0.0%
Memory size	2.7 MiB



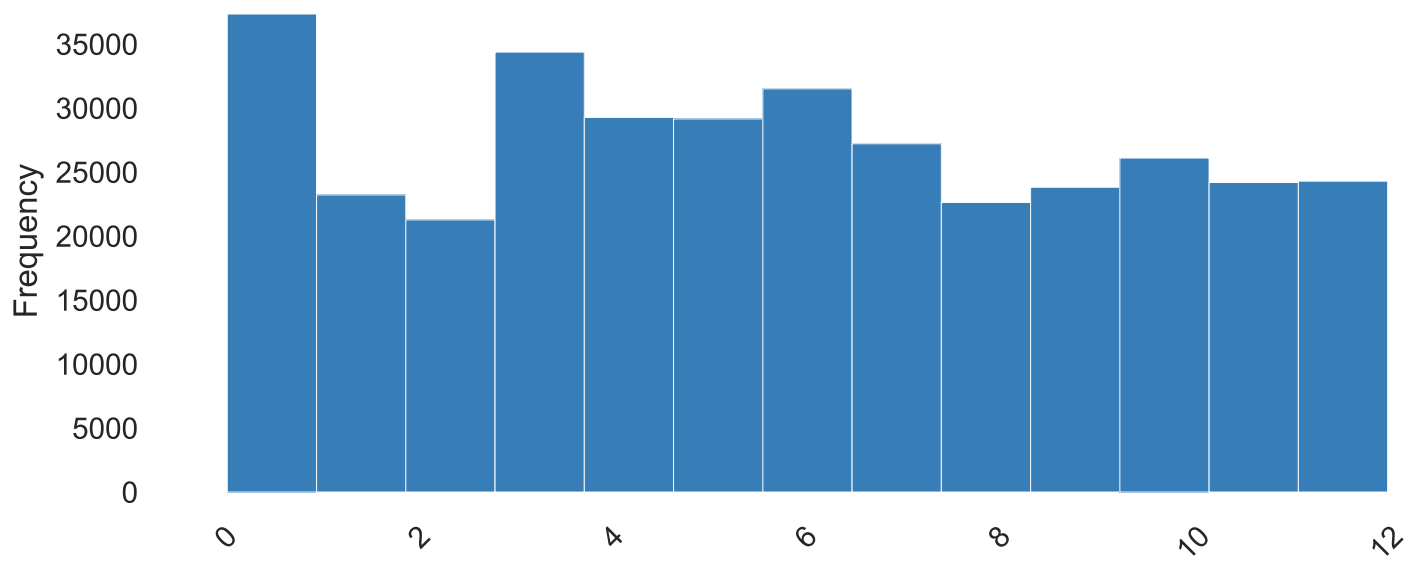
Quantile statistics

Minimum	0
5-th percentile	0
Q1	3
median	6
Q3	9

95-th percentile	12
Maximum	12
Range	12
Interquartile range (IQR)	6

Descriptive statistics

Standard deviation	3.726421
Coefficient of variation (CV)	0.65208268
Kurtosis	-1.1482393
Mean	5.7146449
Median Absolute Deviation (MAD)	3
Skewness	0.081207294
Sum	2025093
Variance	13.886213
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=13)

Value	Count	Frequency (%)
0	37352	10.5%
3	34373	9.7%
6	31508	8.9%
4	29270	8.3%
5	29153	8.2%
7	27213	7.7%
10	26099	7.4%
12	24289	6.9%
11	24186	6.8%
9	23813	6.7%
Other values (3)	67113	18.9%

Value	Count	Frequency (%)
0	37352	10.5%
1	23219	6.6%
2	21267	6.0%
3	34373	9.7%
4	29270	8.3%
5	29153	8.2%
6	31508	8.9%
7	27213	7.7%
8	22627	6.4%
9	23813	6.7%

Value	Count	Frequency (%)
12	24289	6.9%
11	24186	6.8%
10	26099	7.4%
9	23813	6.7%
8	22627	6.4%
7	27213	7.7%
6	31508	8.9%
5	29153	8.2%
4	29270	8.3%
3	34373	9.7%

FuelType

Categorical

IMBALANCE MISSING

Distinct	7
Distinct (%)	< 0.1%
Missing	32895
Missing (%)	9.3%
Memory size	2.7 MiB

Length

Max length	8
Median length	6
Mean length	6.5592707
Min length	3

Characters and Unicode

Total characters	2108635	
Distinct characters	16	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	0	?
Unique (%)	0.0%	

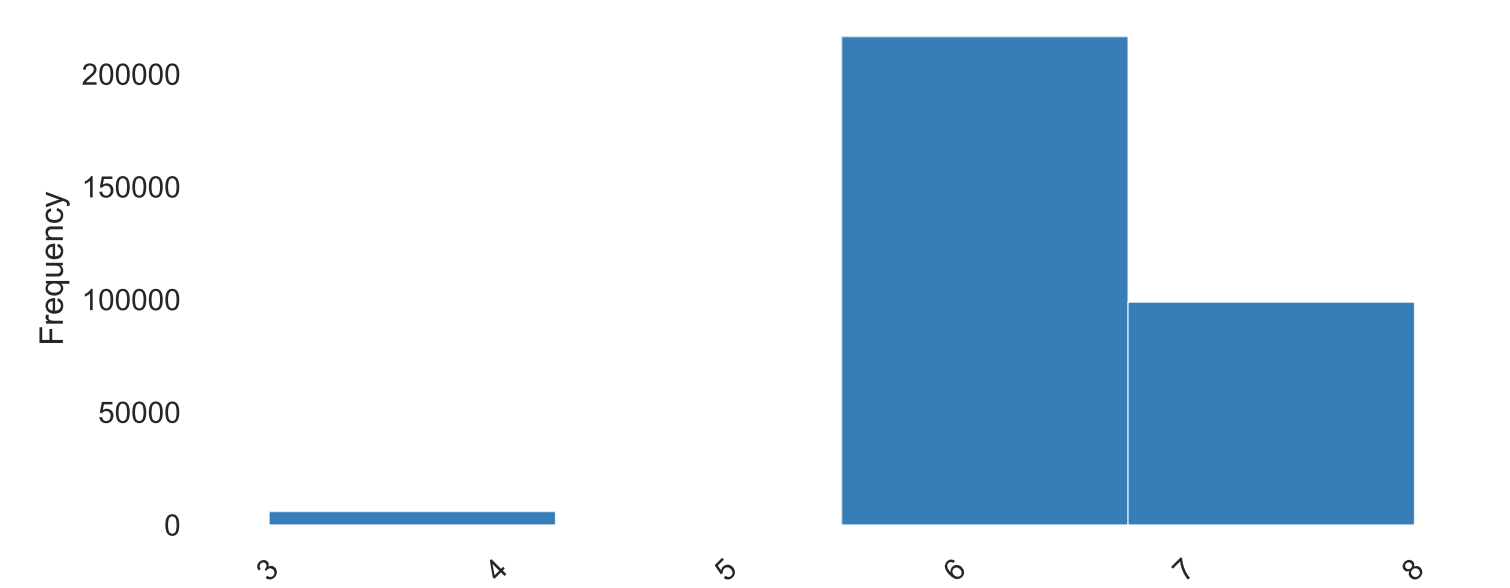
Sample

1st row	petrol
2nd row	gasoline
3rd row	gasoline
4th row	petrol
5th row	gasoline

Common Values

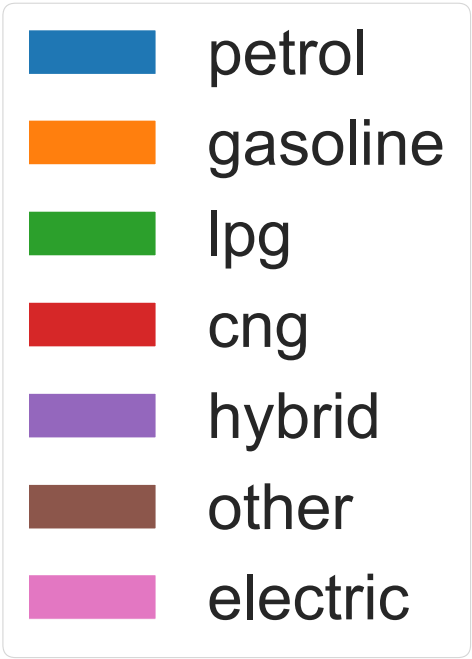
Value	Count	Frequency (%)
petrol	216352	61.1%
gasoline	98720	27.9%
lpg	5310	1.5%
cng	565	0.2%
hybrid	233	0.1%
other	204	0.1%
electric	90	< 0.1%
(Missing)	32895	9.3%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
petrol	216352	67.3%
gasoline	98720	30.7%
lpg	5310	1.7%
cng	565	0.2%
hybrid	233	0.1%
other	204	0.1%
electric	90	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
l	320472	15.2%
e	315456	15.0%
o	315276	15.0%
p	221662	10.5%
r	216879	10.3%
t	216646	10.3%
g	104595	5.0%
n	99285	4.7%
i	99043	4.7%
a	98720	4.7%
Other values (6)	100601	4.8%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	2108635	100.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
l	320472	15.2%
e	315456	15.0%
o	315276	15.0%
p	221662	10.5%
r	216879	10.3%
t	216646	10.3%
g	104595	5.0%
n	99285	4.7%
i	99043	4.7%
a	98720	4.7%
Other values (6)	100601	4.8%

Most occurring scripts

Value	Count	Frequency (%)
Latin	2108635	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
l	320472	15.2%
e	315456	15.0%
o	315276	15.0%
p	221662	10.5%
r	216879	10.3%
t	216646	10.3%
g	104595	5.0%
n	99285	4.7%
i	99043	4.7%
a	98720	4.7%
Other values (6)	100601	4.8%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	2108635	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
l	320472	15.2%
e	315456	15.0%
o	315276	15.0%
p	221662	10.5%
r	216879	10.3%
t	216646	10.3%
g	104595	5.0%
n	99285	4.7%
i	99043	4.7%
a	98720	4.7%
Other values (6)	100601	4.8%

Brand

Categorical

Distinct	40
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	2.7 MiB

Length

Max length	14
Median length	13
Mean length	6.7287488
Min length	3

Characters and Unicode

Total characters	2384460	
Distinct characters	25	
Distinct categories	2 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	0	?
Unique (%)	0.0%	

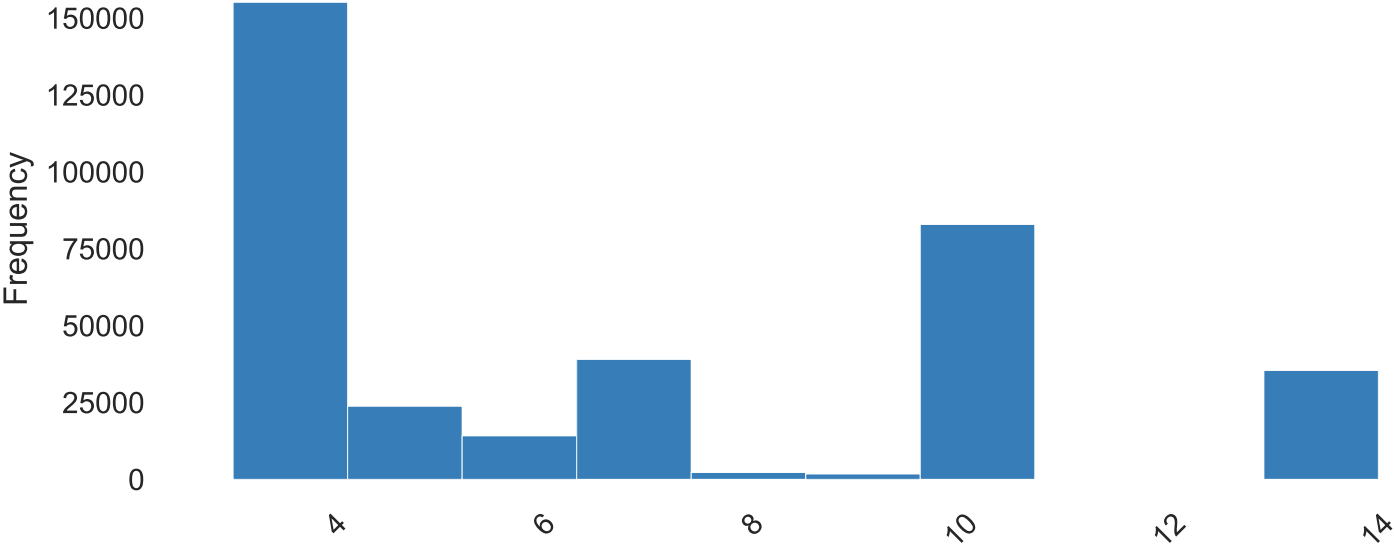
Sample

1st row	volkswagen
2nd row	audi
3rd row	jeep
4th row	volkswagen
5th row	skoda

Common Values

Value	Count	Frequency (%)
volkswagen	77013	21.7%
opel	39931	11.3%
bmw	36914	10.4%
mercedes_benz	32046	9.0%
audi	29456	8.3%
ford	25179	7.1%
renault	17927	5.1%
peugeot	10998	3.1%
fiat	9643	2.7%
seat	6907	1.9%
Other values (30)	68355	19.3%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
volkswagen	77013	21.7%
opel	39931	11.3%
bmw	36914	10.4%
mercedes_benz	32046	9.0%
audi	29456	8.3%
ford	25179	7.1%
renault	17927	5.1%
peugeot	10998	3.1%
fiat	9643	2.7%
seat	6907	1.9%
Other values (30)	68355	19.3%

Most occurring characters

Value	Count	Frequency (%)
e	311426	13.1%
a	199233	8.4%
o	198221	8.3%
s	159374	6.7%
n	156601	6.6%
l	144829	6.1%
w	114469	4.8%
d	107218	4.5%
r	97168	4.1%
g	91890	3.9%
Other values (15)	804031	33.7%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	2346181	98.4%
Connector Punctuation	38279	1.6%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	311426	13.3%
a	199233	8.5%
o	198221	8.4%
s	159374	6.8%
n	156601	6.7%
l	144829	6.2%
w	114469	4.9%
d	107218	4.6%
r	97168	4.1%
g	91890	3.9%
Other values (14)	765752	32.6%

Connector Punctuation

Value	Count	Frequency (%)
—	38279	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	2346181	98.4%
Common	38279	1.6%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	311426	13.3%
a	199233	8.5%
o	198221	8.4%
s	159374	6.8%
n	156601	6.7%
l	144829	6.2%
w	114469	4.9%
d	107218	4.6%
r	97168	4.1%
g	91890	3.9%
Other values (14)	765752	32.6%

Common

Value	Count	Frequency (%)
—	38279	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	2384460	100.0%

Most frequent character per block

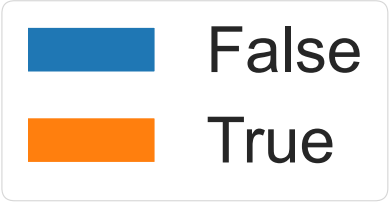
ASCII

Value	Count	Frequency (%)
e	311426	13.1%
a	199233	8.4%
o	198221	8.3%
s	159374	6.7%
n	156601	6.6%
l	144829	6.1%
w	114469	4.8%
d	107218	4.5%
r	97168	4.1%
g	91890	3.9%
Other values (15)	804031	33.7%

Repaired
Boolean

Distinct	2
Distinct (%)	< 0.1%
Missing	71154
Missing (%)	20.1%
Memory size	692.3 KiB

Value	Count	Frequency (%)
False	247161	69.7%
True	36054	10.2%
(Missing)	71154	20.1%



DateCreated
Categorical

Distinct	109
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	2.7 MiB

Length

Max length	19
Median length	19
Mean length	19
Min length	19

Characters and Unicode

Total characters	6733011	
Distinct characters	13	
Distinct categories	4 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	20	?
Unique (%)	< 0.1%	

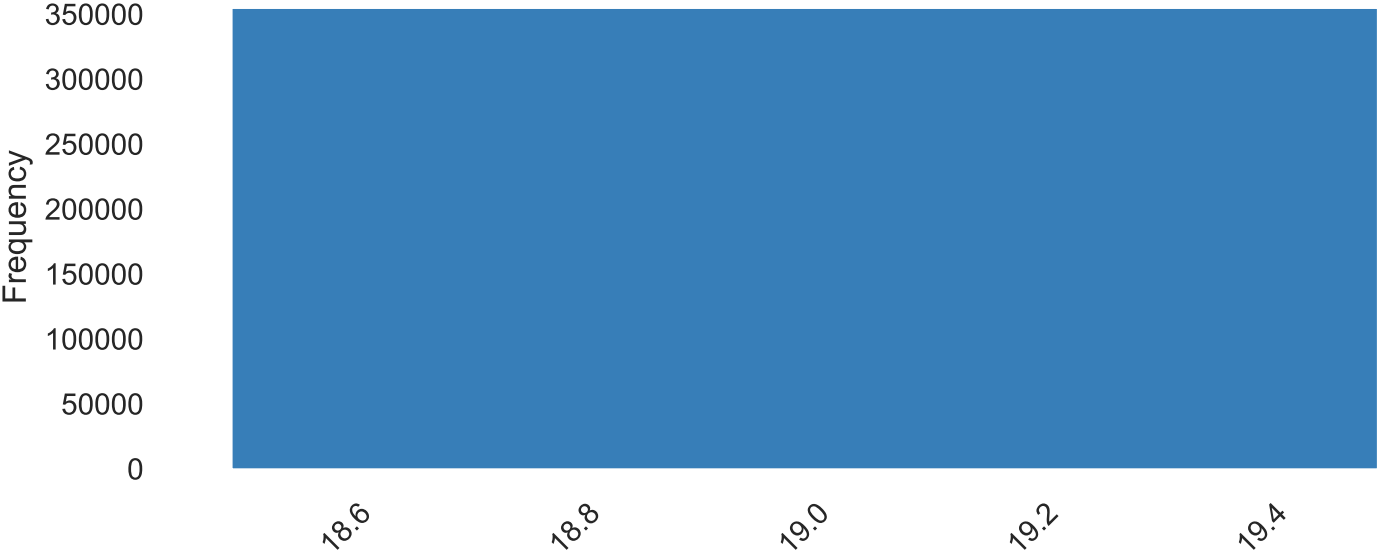
Sample

1st row	2016-03-24 00:00:00
2nd row	2016-03-24 00:00:00
3rd row	2016-03-14 00:00:00
4th row	2016-03-17 00:00:00
5th row	2016-03-31 00:00:00

Common Values

Value	Count	Frequency (%)
2016-04-03 00:00:00	13719	3.9%
2016-04-04 00:00:00	13317	3.8%
2016-03-20 00:00:00	12918	3.6%
2016-03-12 00:00:00	12840	3.6%
2016-03-21 00:00:00	12691	3.6%
2016-03-14 00:00:00	12478	3.5%
2016-03-28 00:00:00	12475	3.5%
2016-04-02 00:00:00	12362	3.5%
2016-03-07 00:00:00	12244	3.5%
2016-03-19 00:00:00	12197	3.4%
Other values (99)	227128	64.1%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
00:00:00	354369	50.0%
2016-04-03	13719	1.9%
2016-04-04	13317	1.9%
2016-03-20	12918	1.8%
2016-03-12	12840	1.8%
2016-03-21	12691	1.8%
2016-03-14	12478	1.8%
2016-03-28	12475	1.8%
2016-04-02	12362	1.7%
2016-03-07	12244	1.7%
Other values (100)	239325	33.8%

Most occurring characters

Value	Count	Frequency (%)
0	2979469	44.3%
-	708738	10.5%
:	708738	10.5%
2	508968	7.6%
1	507716	7.5%
6	383182	5.7%
	354369	5.3%
3	351358	5.2%
4	94151	1.4%
9	36527	0.5%
Other values (3)	99795	1.5%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	4961166	73.7%
Dash Punctuation	708738	10.5%
Other Punctuation	708738	10.5%
Space Separator	354369	5.3%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	2979469	60.1%
2	508968	10.3%
1	507716	10.2%
6	383182	7.7%
3	351358	7.1%
4	94151	1.9%
9	36527	0.7%
5	35852	0.7%
7	34540	0.7%
8	29403	0.6%

Dash Punctuation

Value	Count	Frequency (%)
-	708738	100.0%

Other Punctuation

Value	Count	Frequency (%)
:	708738	100.0%

Space Separator

Value	Count	Frequency (%)
	354369	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	6733011	100.0%

Most frequent character per script

Common		
Value	Count	Frequency (%)
0	2979469	44.3%
-	708738	10.5%
:	708738	10.5%
2	508968	7.6%
1	507716	7.5%
6	383182	5.7%
	354369	5.3%
3	351358	5.2%
4	94151	1.4%
9	36527	0.5%
Other values (3)	99795	1.5%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	6733011	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	2979469	44.3%
-	708738	10.5%
:	708738	10.5%
2	508968	7.6%
1	507716	7.5%
6	383182	5.7%
	354369	5.3%
3	351358	5.2%
4	94151	1.4%
9	36527	0.5%
Other values (3)	99795	1.5%

NumberOfPictures

Categorical

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	2.7 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	354369	
Distinct characters	1	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	0	?
Unique (%)	0.0%	

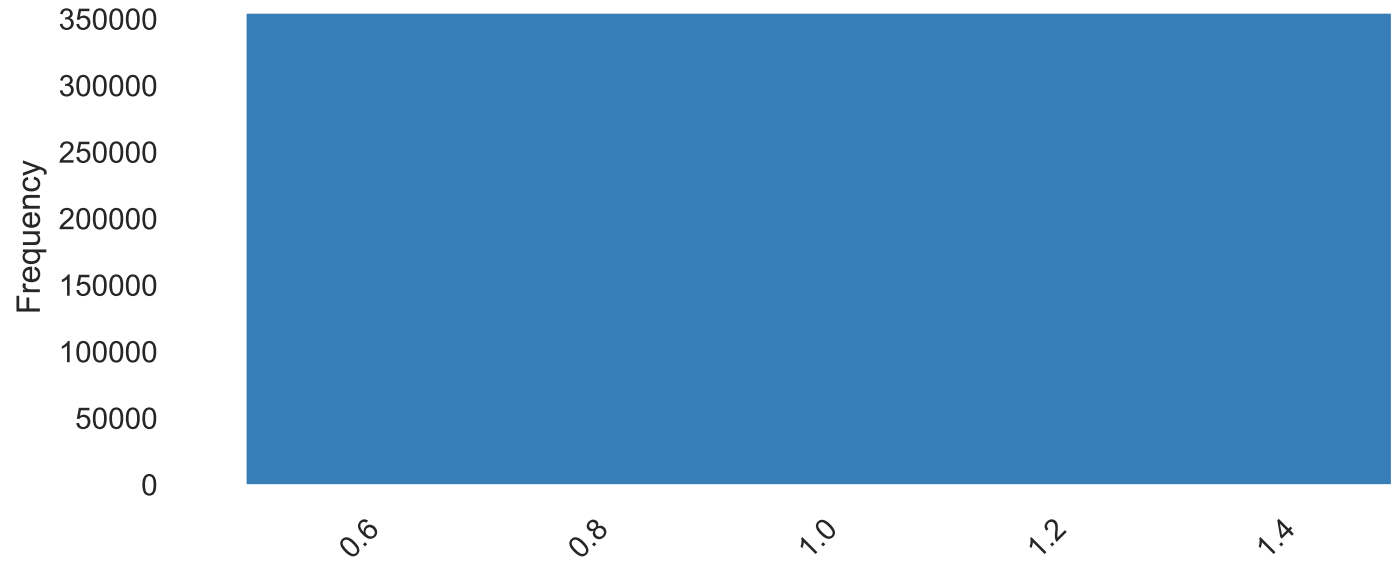
Sample

1st row	0
2nd row	0
3rd row	0
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	354369	100.0%

Length



Histogram of lengths of the category

Common Values (Plot)



0

Value	Count	Frequency (%)
0	354369	100.0%

Most occurring characters

Value	Count	Frequency (%)
0	354369	100.0%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	354369	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	354369	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	354369	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
0	354369	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	354369	100.0%

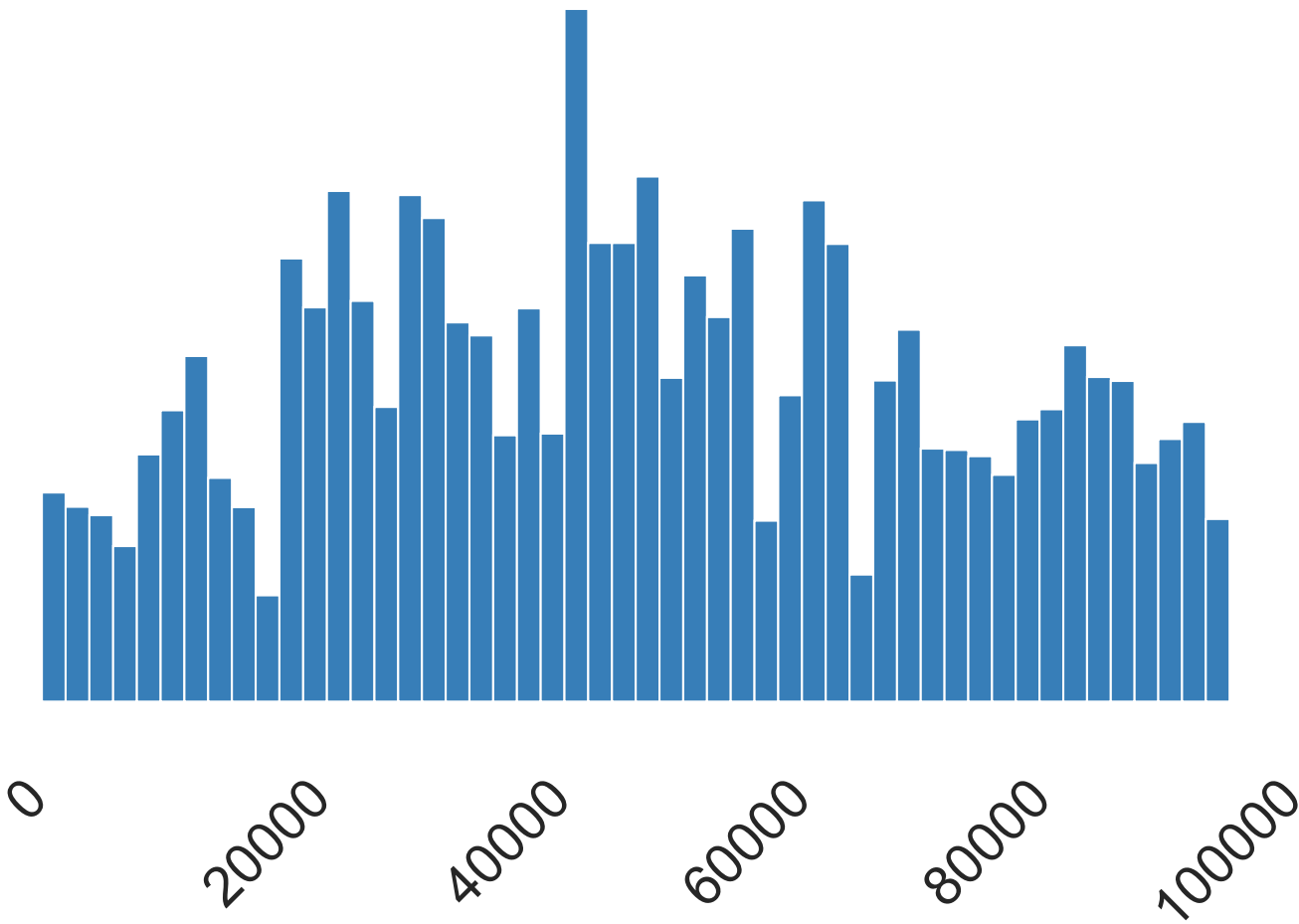
Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	354369	100.0%

PostalCode
Real number (\mathbb{R})

Distinct	8143
Distinct (%)	2.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	50508.689
Minimum	1067
Maximum	99998
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	2.7 MiB



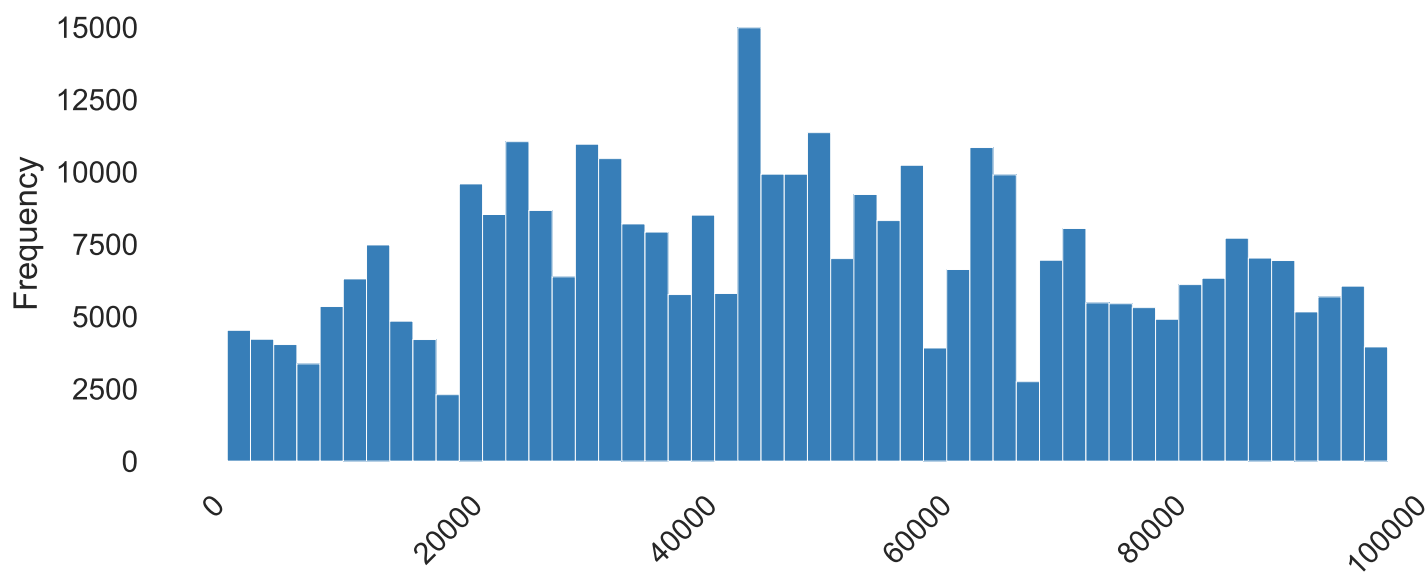
Quantile statistics

Minimum	1067
5-th percentile	9569
Q1	30165
median	49413
Q3	71083

95-th percentile	93109
Maximum	99998
Range	98931
Interquartile range (IQR)	40918

Descriptive statistics

Standard deviation	25783.096
Coefficient of variation (CV)	0.51046853
Kurtosis	-0.96529606
Mean	50508.689
Median Absolute Deviation (MAD)	20537
Skewness	0.077054275
Sum	1.7898714×10^{10}
Variance	6.6476805×10^8
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
10115	819	0.2%
65428	613	0.2%
66333	343	0.1%
32257	317	0.1%
44145	317	0.1%
52525	306	0.1%
26789	296	0.1%
78224	293	0.1%
38518	291	0.1%
13357	290	0.1%
Other values (8133)	350484	98.9%

Value	Count	Frequency (%)
1067	95	< 0.1%
1068	1	< 0.1%
1069	55	< 0.1%
1097	29	< 0.1%
1099	62	< 0.1%
1108	12	< 0.1%
1109	79	< 0.1%
1127	28	< 0.1%
1129	42	< 0.1%
1139	64	< 0.1%

Value	Count	Frequency (%)
99998	16	< 0.1%
99996	3	< 0.1%
99994	7	< 0.1%
99991	2	< 0.1%
99988	9	< 0.1%
99986	19	< 0.1%
99976	35	< 0.1%
99974	152	< 0.1%
99958	9	< 0.1%
99955	23	< 0.1%

LastSeen
Categorical

Distinct	179150
Distinct (%)	50.6%
Missing	0
Missing (%)	0.0%
Memory size	2.7 MiB

Length

Max length	19
Median length	19
Mean length	19
Min length	19

Characters and Unicode

Total characters	6733011	
Distinct characters	13	
Distinct categories	4 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	129415	?
Unique (%)	36.5%	

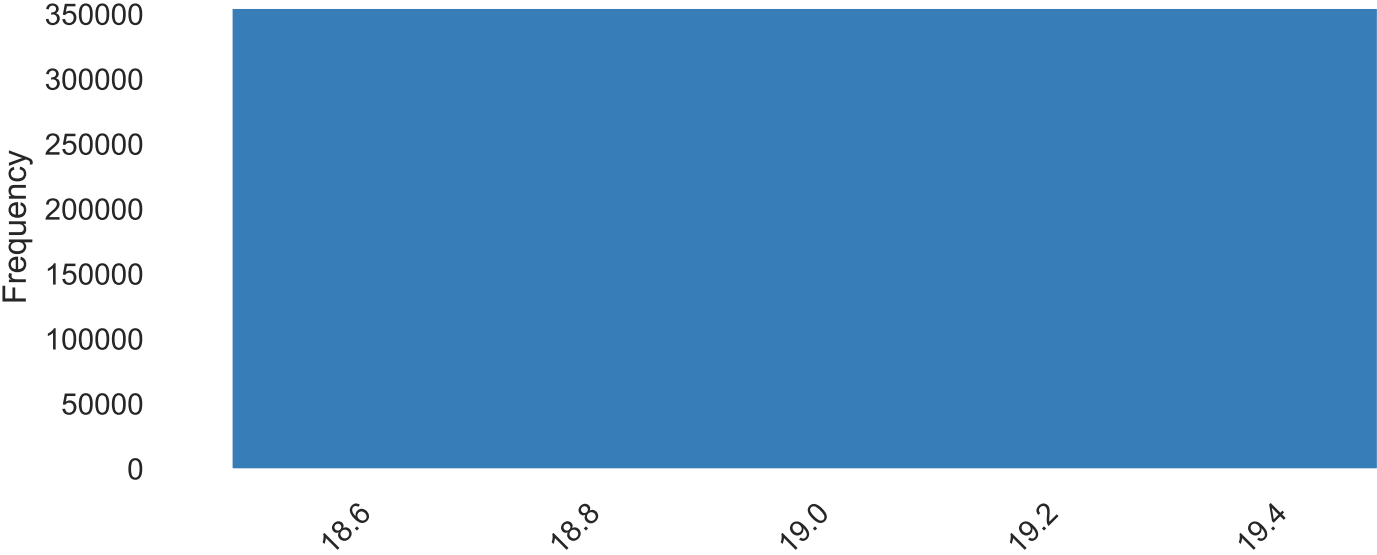
Sample

1st row	2016-04-07 03:16:57
2nd row	2016-04-07 01:46:50
3rd row	2016-04-05 12:47:46
4th row	2016-03-17 17:40:17
5th row	2016-04-06 10:17:21

Common Values

Value	Count	Frequency (%)
2016-04-06 13:45:54	17	< 0.1%
2016-04-06 12:15:45	16	< 0.1%
2016-04-07 05:44:30	16	< 0.1%
2016-04-06 01:15:23	16	< 0.1%
2016-04-07 09:44:27	16	< 0.1%
2016-04-06 09:46:00	16	< 0.1%
2016-04-06 10:15:26	16	< 0.1%
2016-04-06 04:15:59	16	< 0.1%
2016-04-07 08:16:54	16	< 0.1%
2016-04-07 13:17:48	16	< 0.1%
Other values (179140)	354208	> 99.9%

Length



Histogram of lengths of the category

Value	Count	Frequency (%)
2016-04-06	75616	10.7%
2016-04-07	44830	6.3%
2016-04-05	43901	6.2%
2016-03-17	10399	1.5%
2016-04-04	9172	1.3%
2016-04-03	9104	1.3%
2016-04-02	8968	1.3%
2016-03-31	8705	1.2%
2016-04-01	8623	1.2%
2016-03-30	8571	1.2%
Other values (36357)	480849	67.8%

Most occurring characters

Value	Count	Frequency (%)
0	1213504	18.0%
1	895221	13.3%
-	708738	10.5%
:	708738	10.5%
2	660646	9.8%
6	601012	8.9%
4	552813	8.2%
	354369	5.3%
3	348863	5.2%
5	282092	4.2%
Other values (3)	407015	6.0%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	4961166	73.7%
Dash Punctuation	708738	10.5%
Other Punctuation	708738	10.5%
Space Separator	354369	5.3%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	1213504	24.5%
1	895221	18.0%
2	660646	13.3%
6	601012	12.1%
4	552813	11.1%
3	348863	7.0%
5	282092	5.7%
7	200638	4.0%
8	110004	2.2%
9	96373	1.9%

Dash Punctuation

Value	Count	Frequency (%)
-	708738	100.0%

Other Punctuation

Value	Count	Frequency (%)
:	708738	100.0%

Space Separator

Value	Count	Frequency (%)
	354369	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	6733011	100.0%

Most frequent character per script

Common		
Value	Count	Frequency (%)
0	1213504	18.0%
1	895221	13.3%
-	708738	10.5%
:	708738	10.5%
2	660646	9.8%
6	601012	8.9%
4	552813	8.2%
	354369	5.3%
3	348863	5.2%
5	282092	4.2%
Other values (3)	407015	6.0%

Most occurring blocks

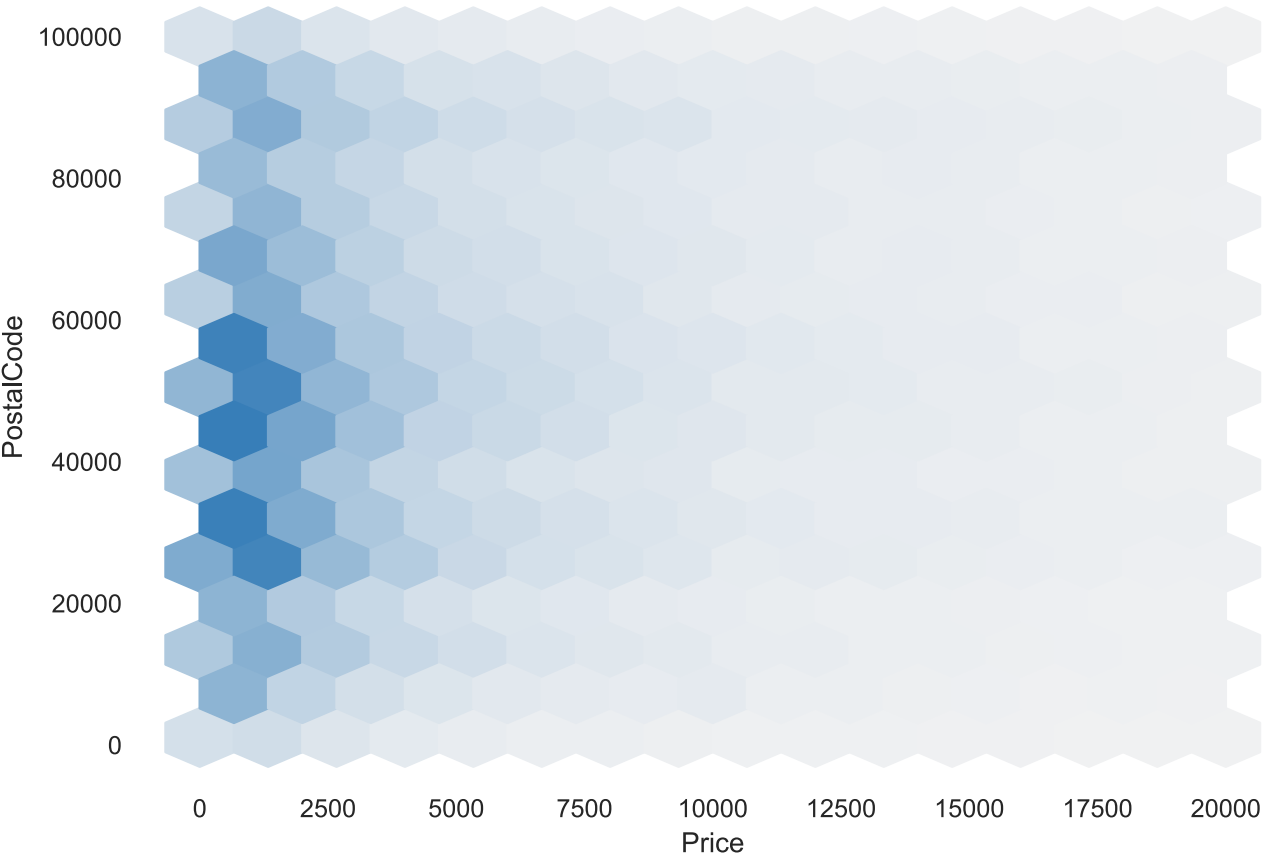
Value	Count	Frequency (%)
ASCII	6733011	100.0%

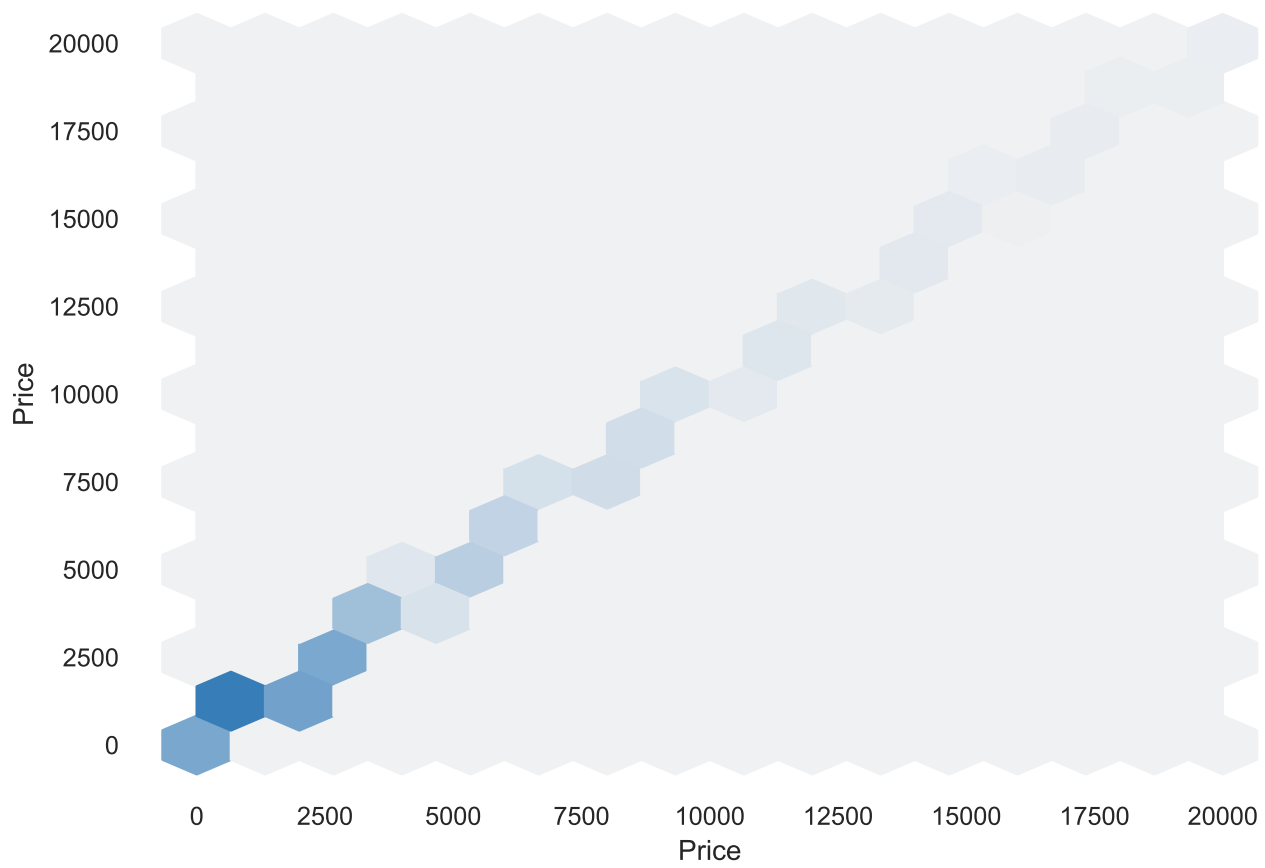
Most frequent character per block

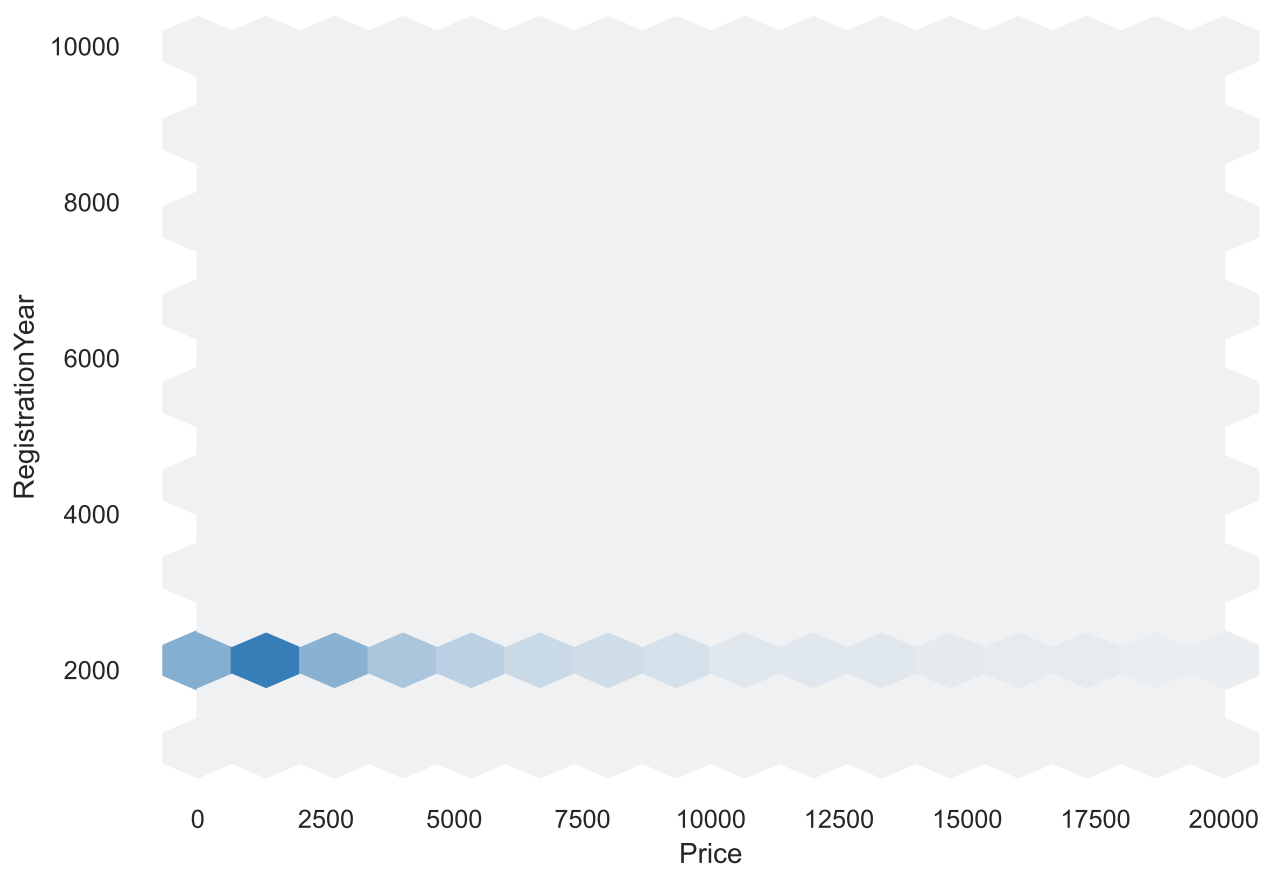
ASCII

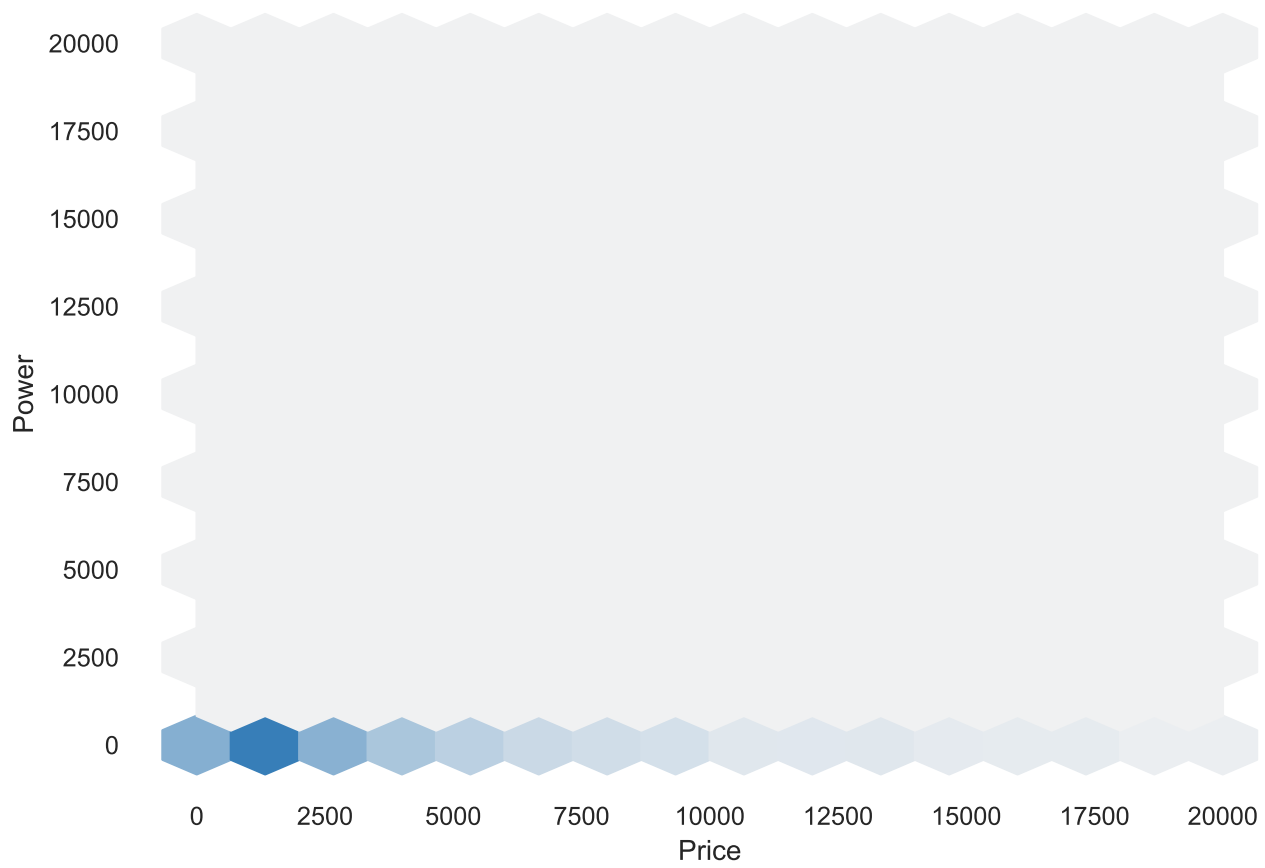
Value	Count	Frequency (%)
0	1213504	18.0%
1	895221	13.3%
-	708738	10.5%
:	708738	10.5%
2	660646	9.8%
6	601012	8.9%
4	552813	8.2%
	354369	5.3%
3	348863	5.2%
5	282092	4.2%
Other values (3)	407015	6.0%

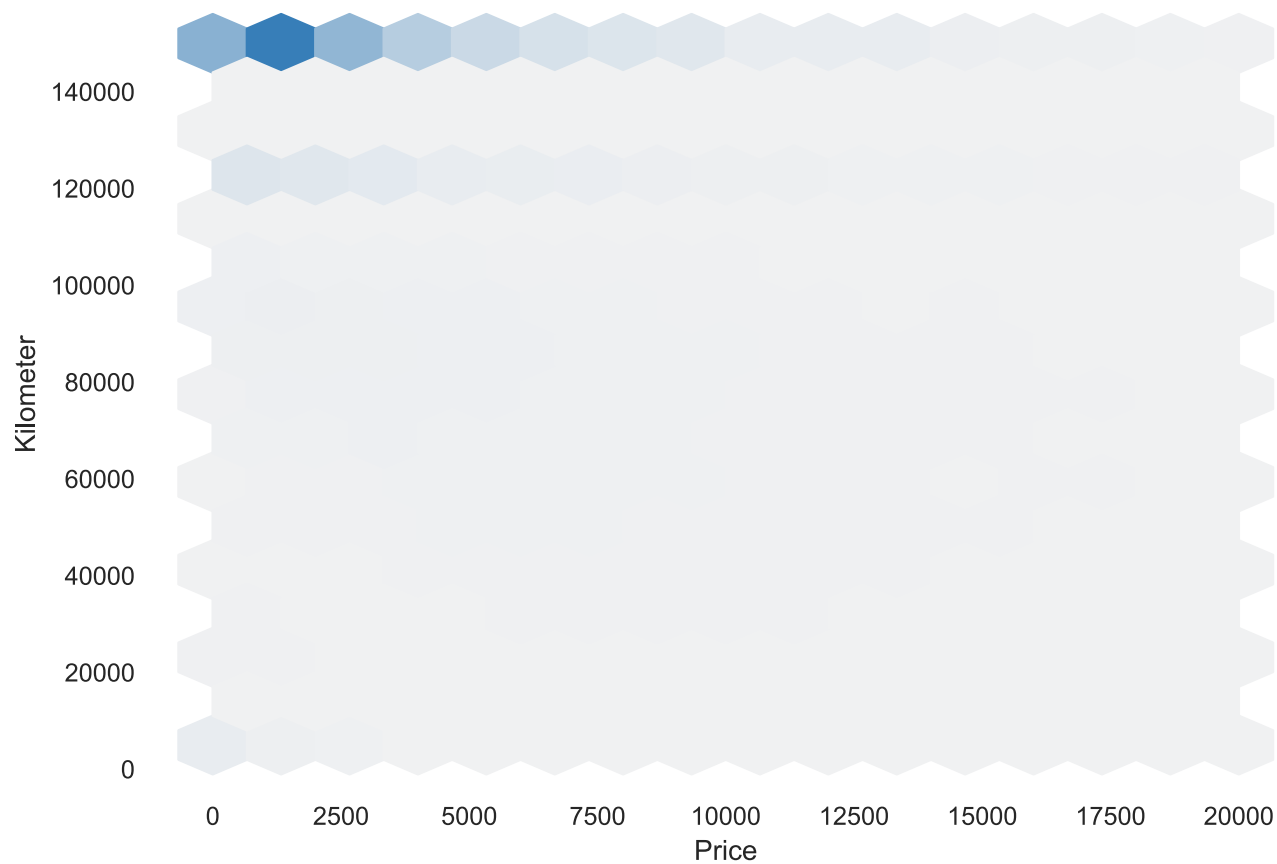
Interactions

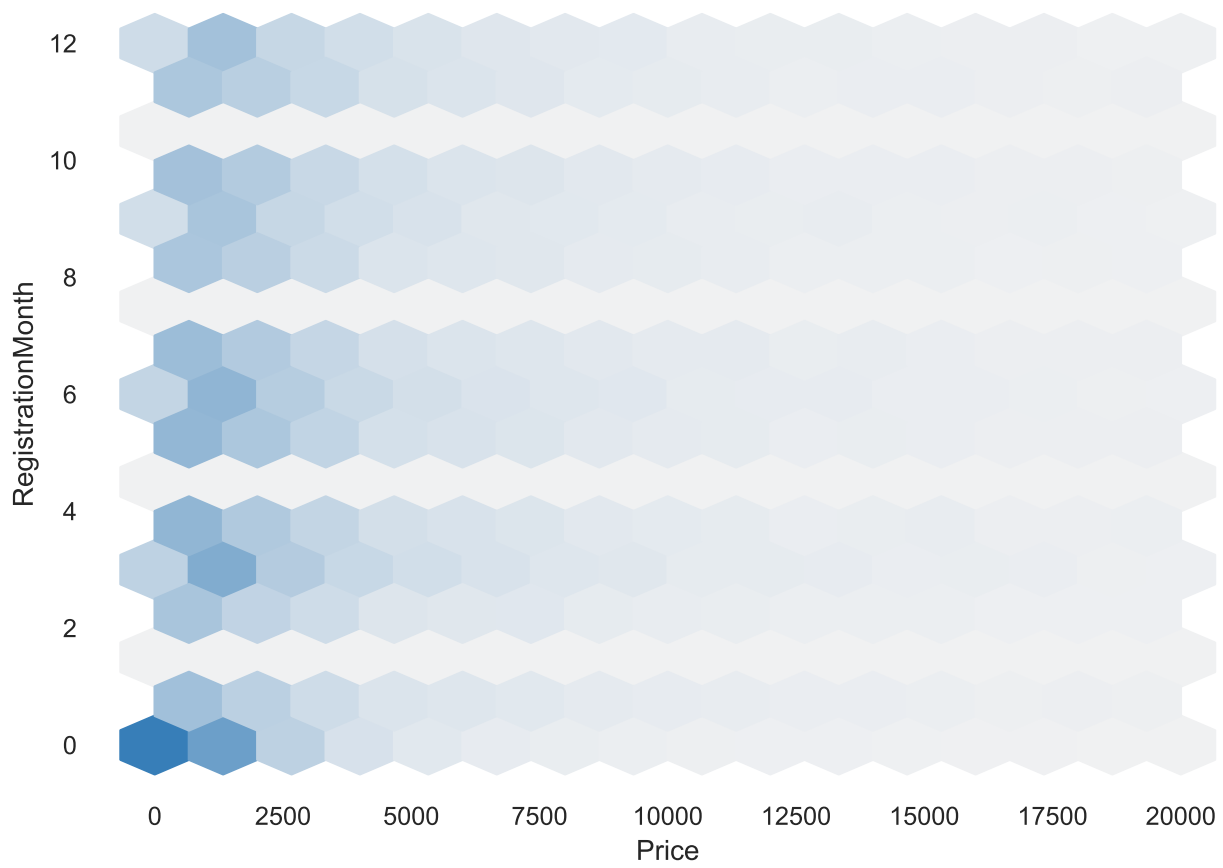


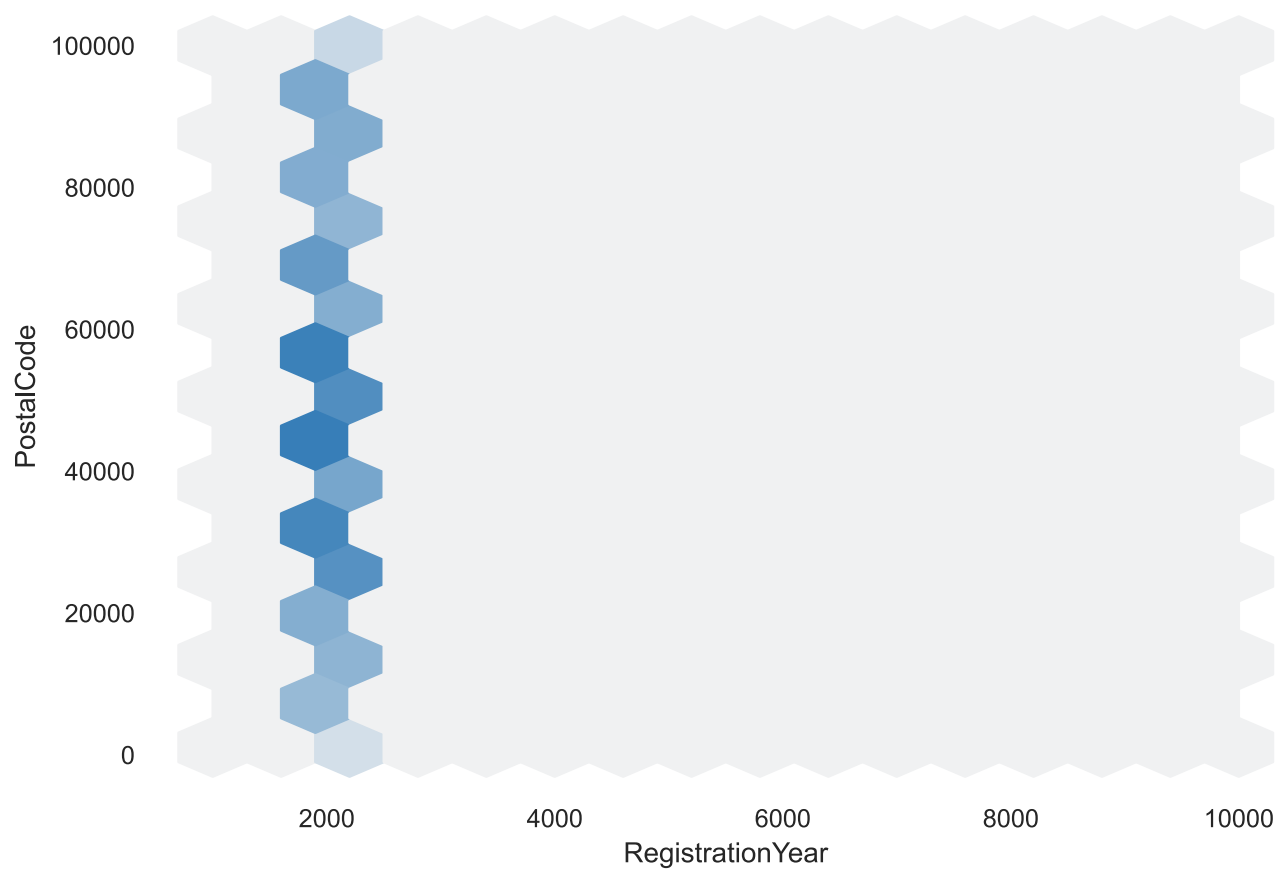


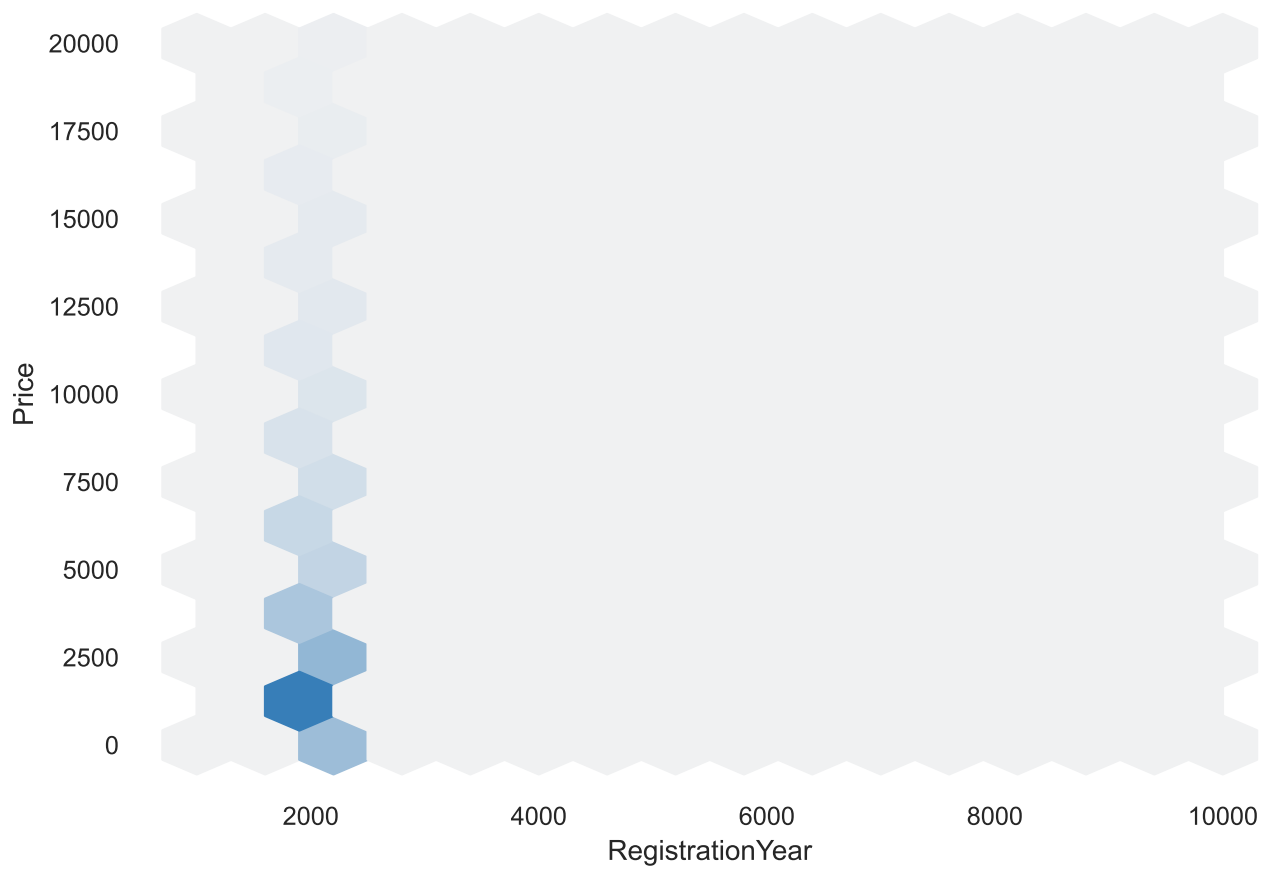


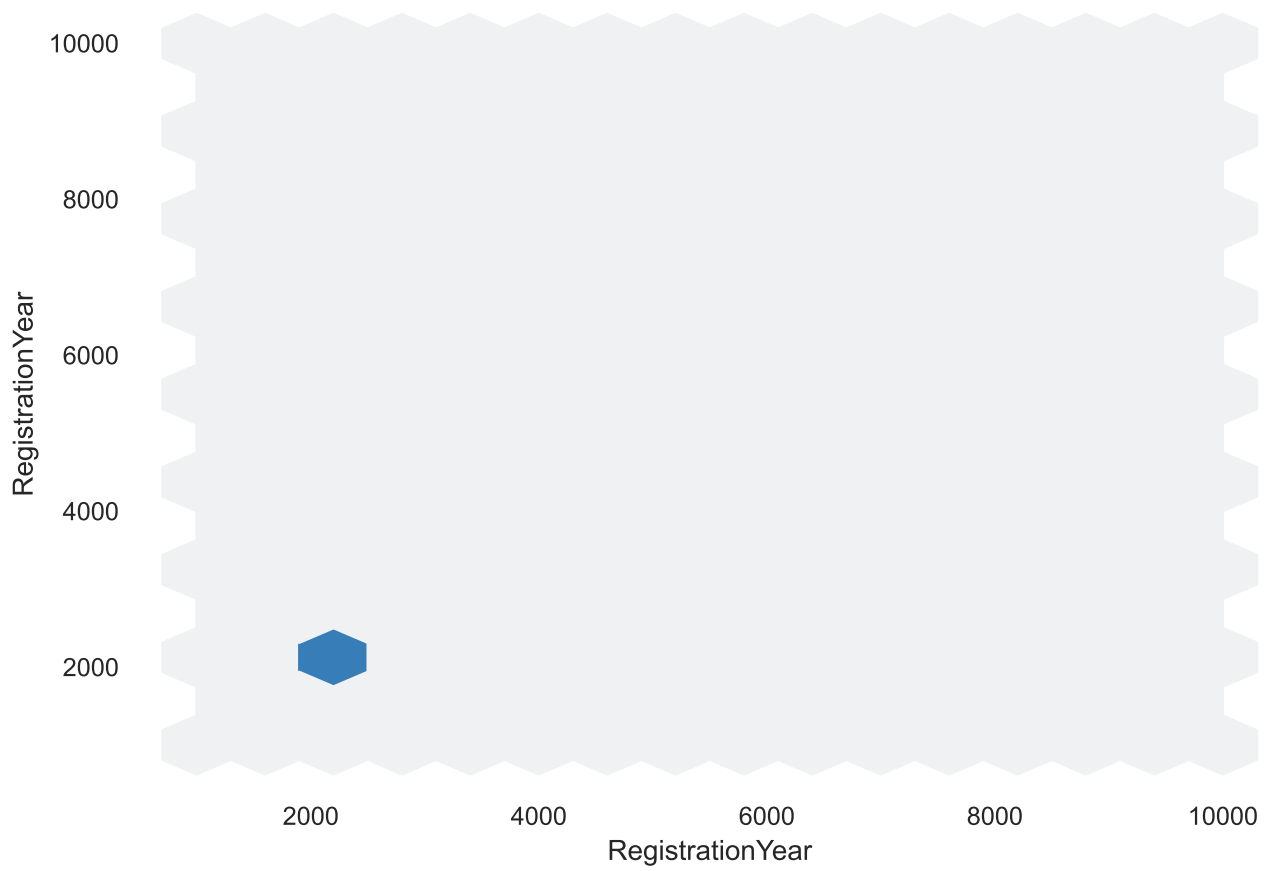


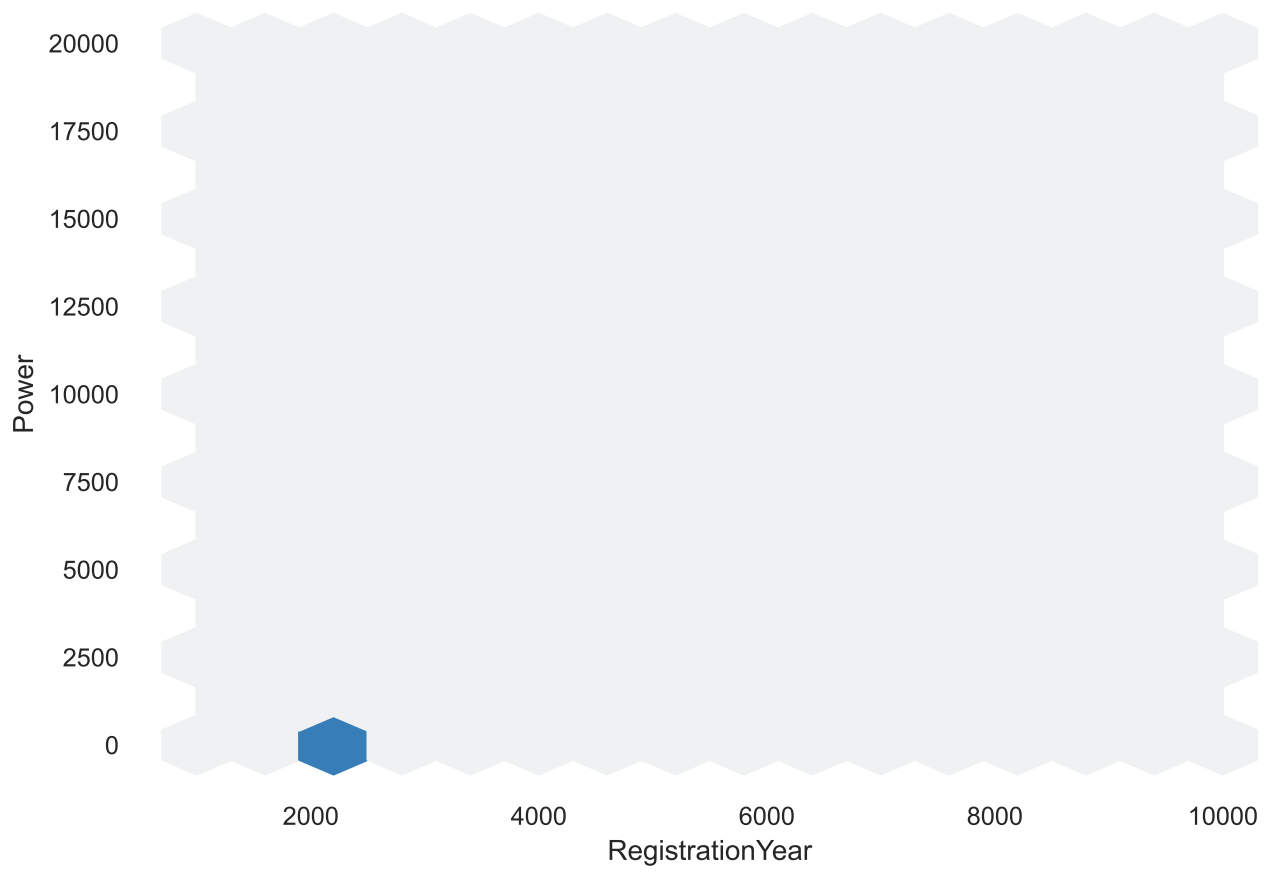


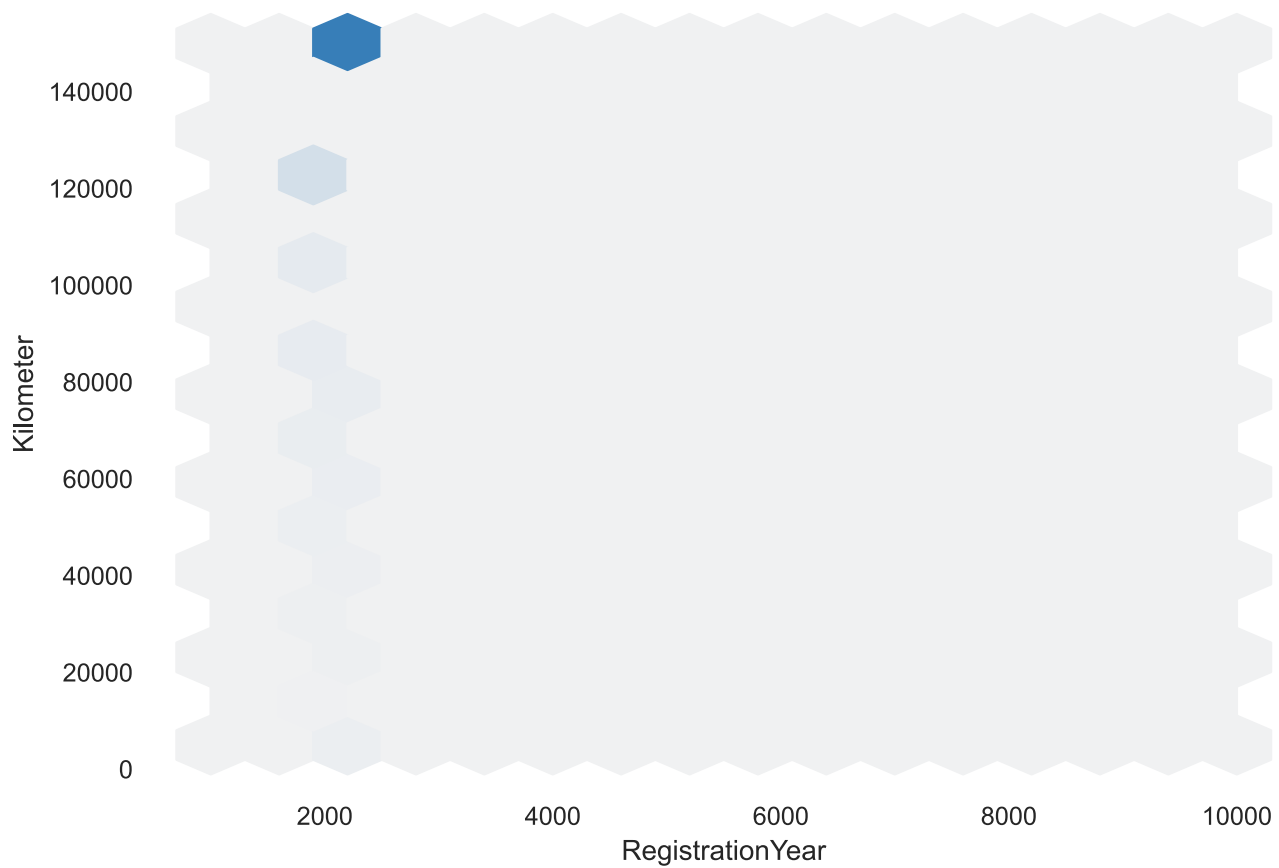


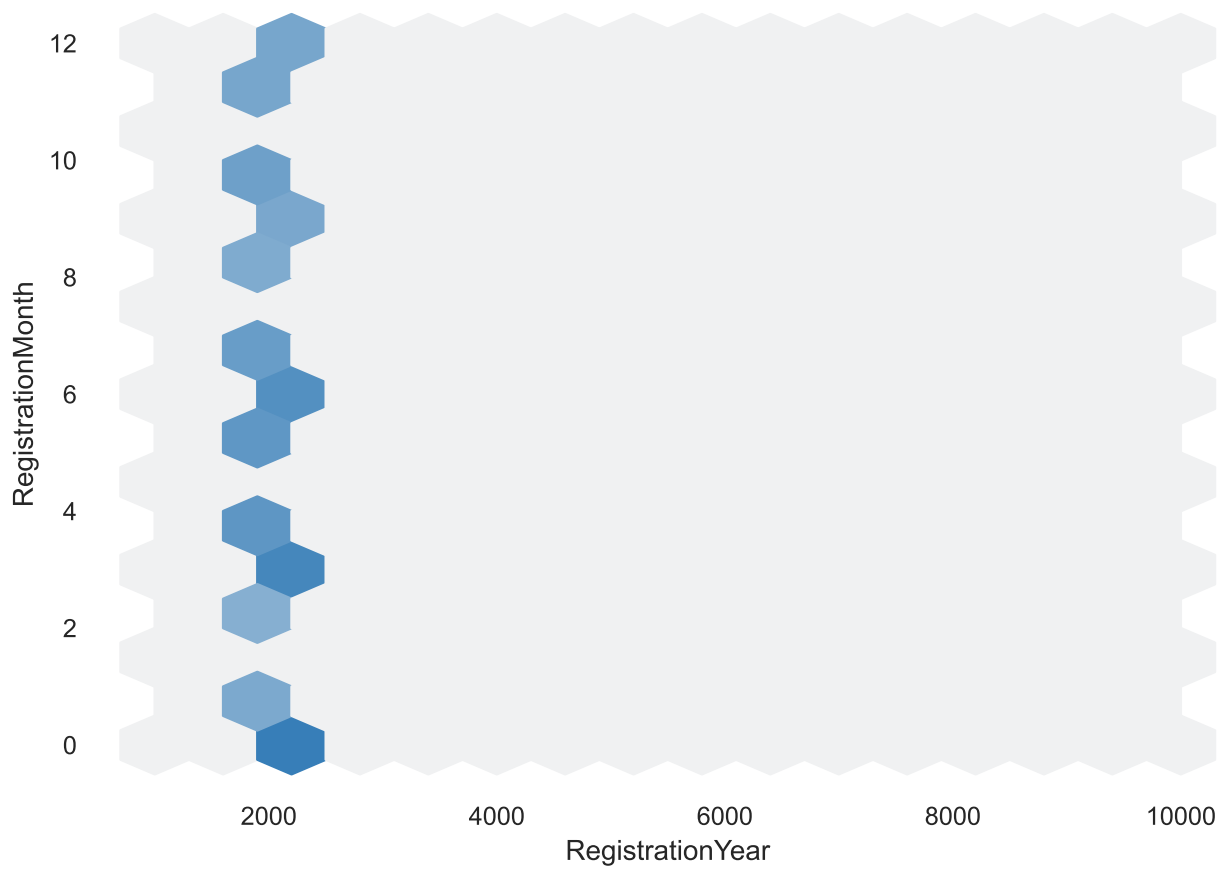


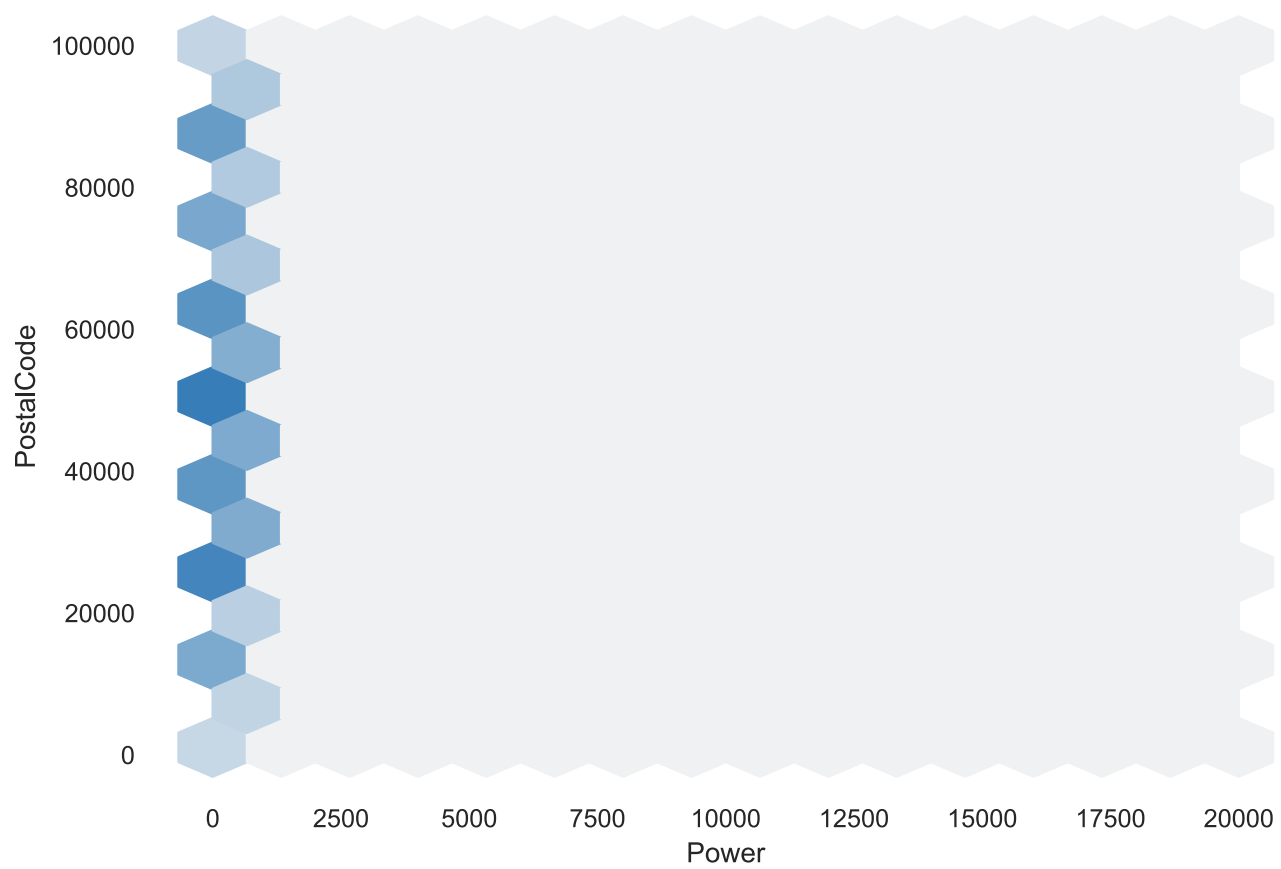


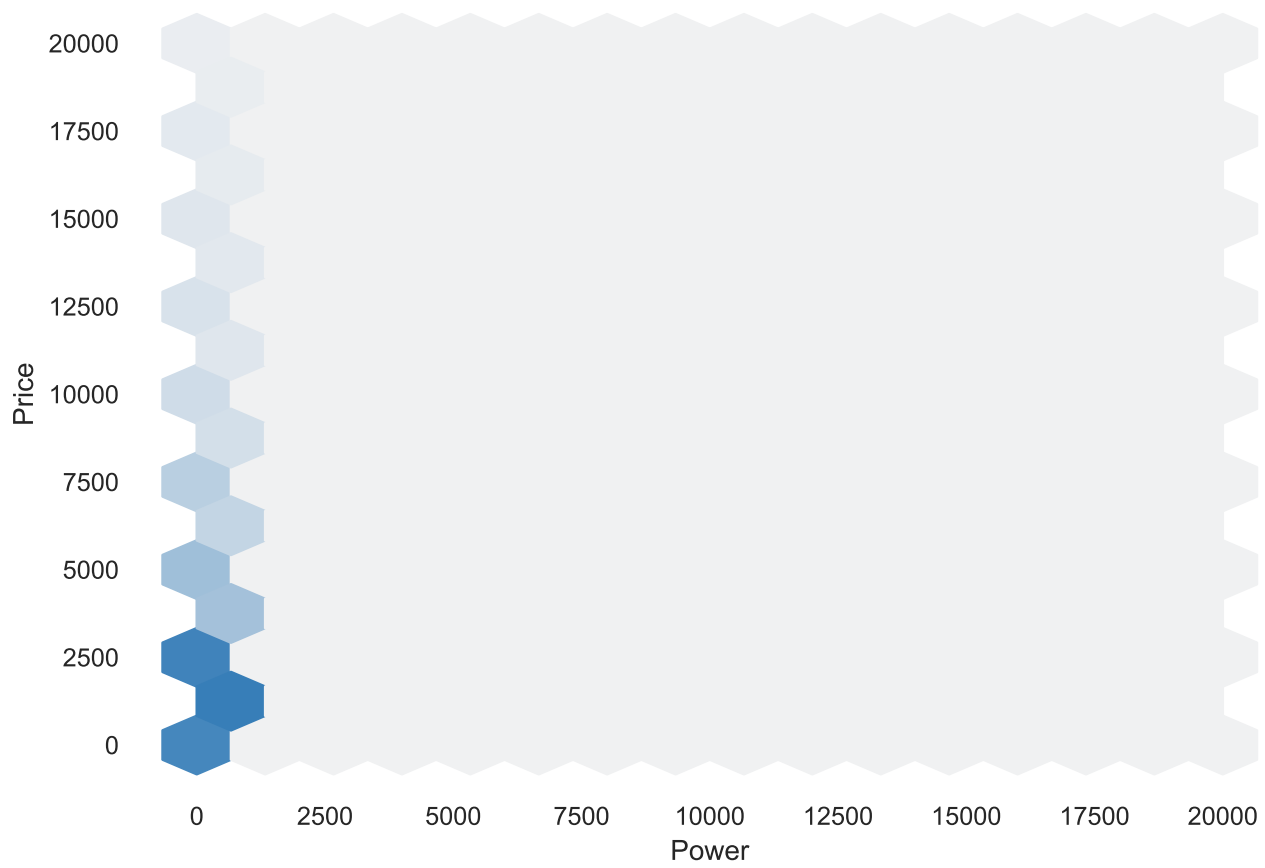


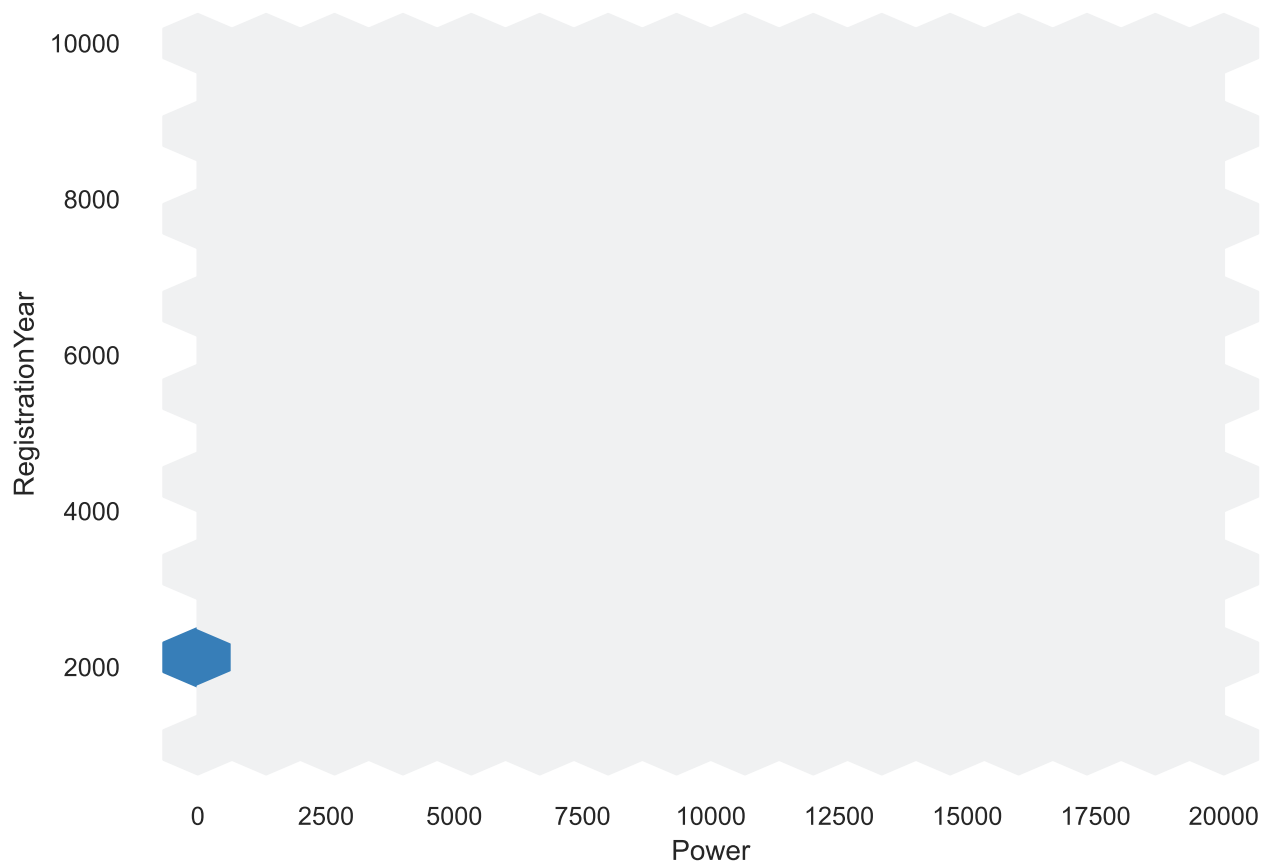


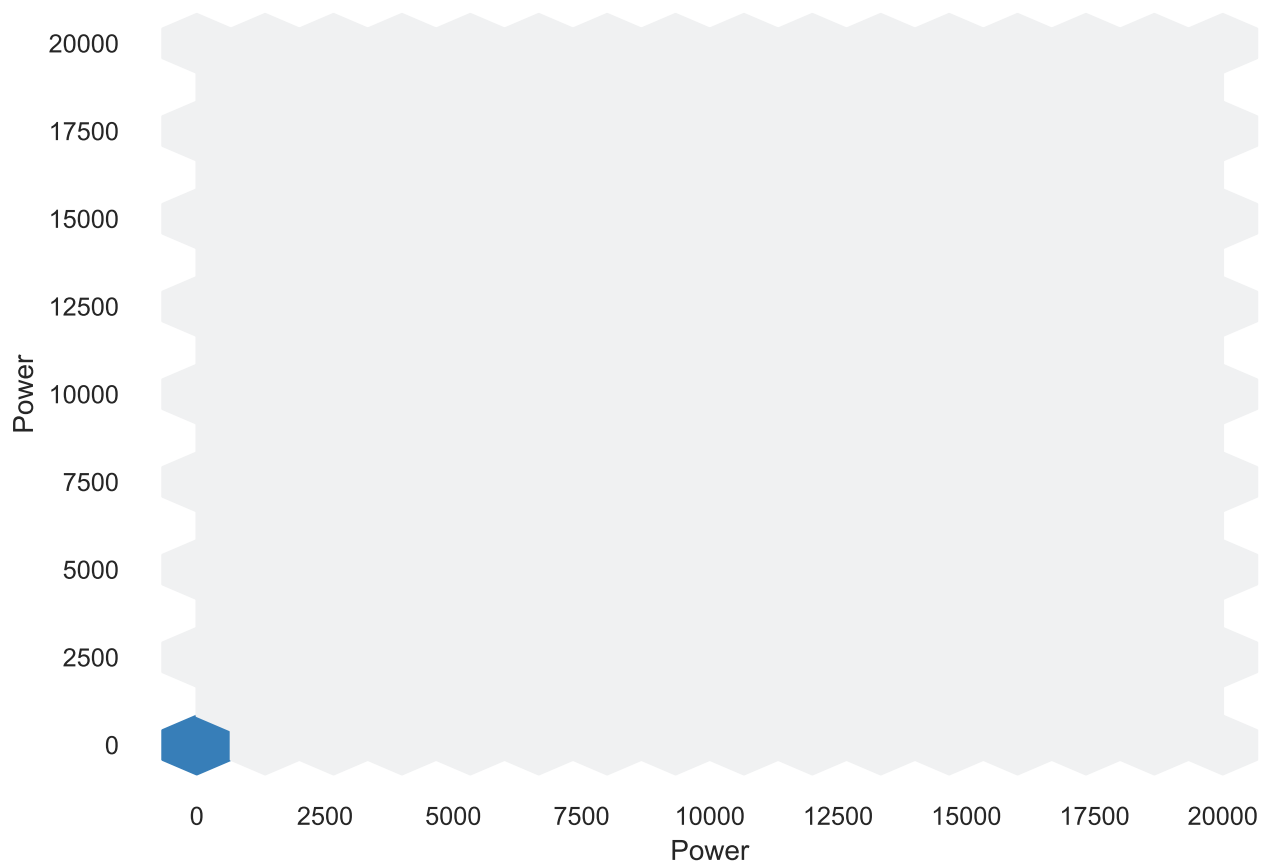


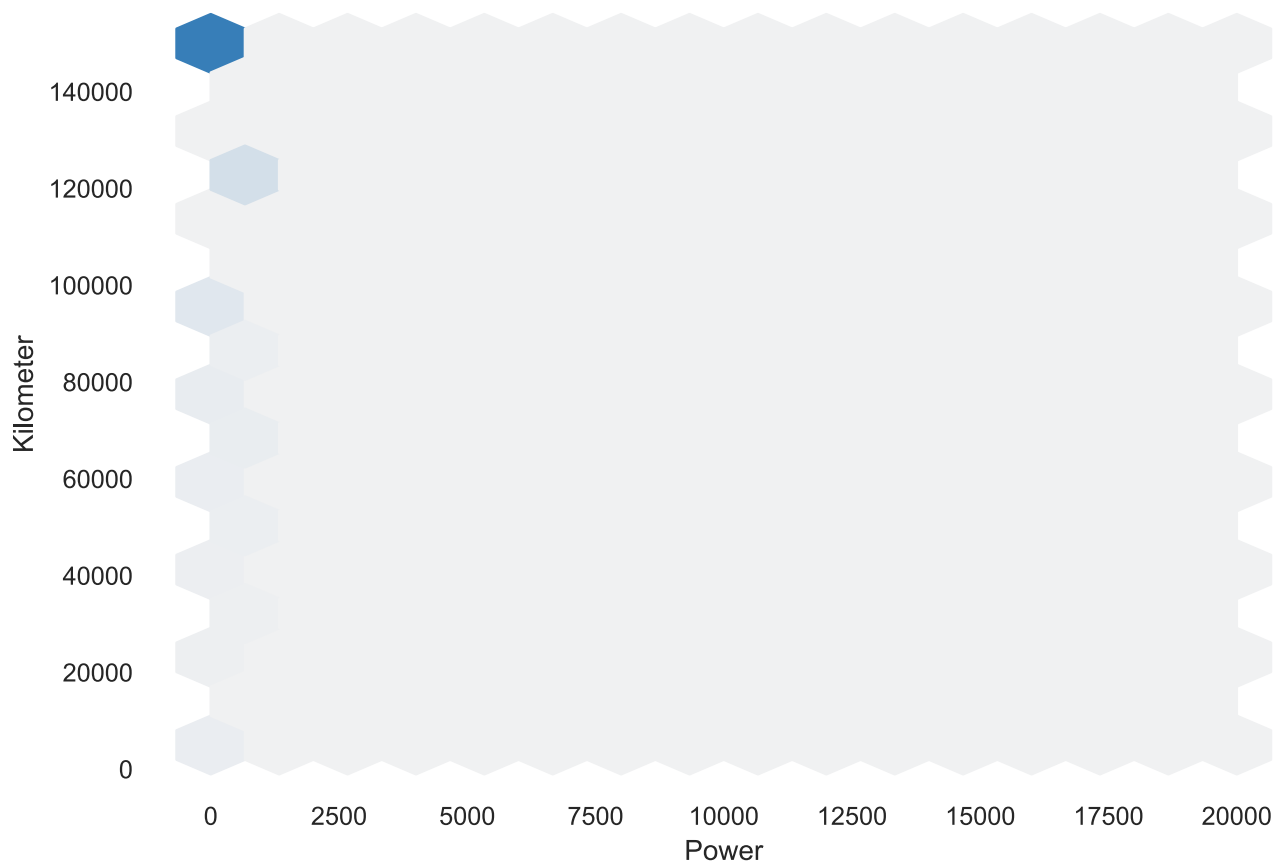


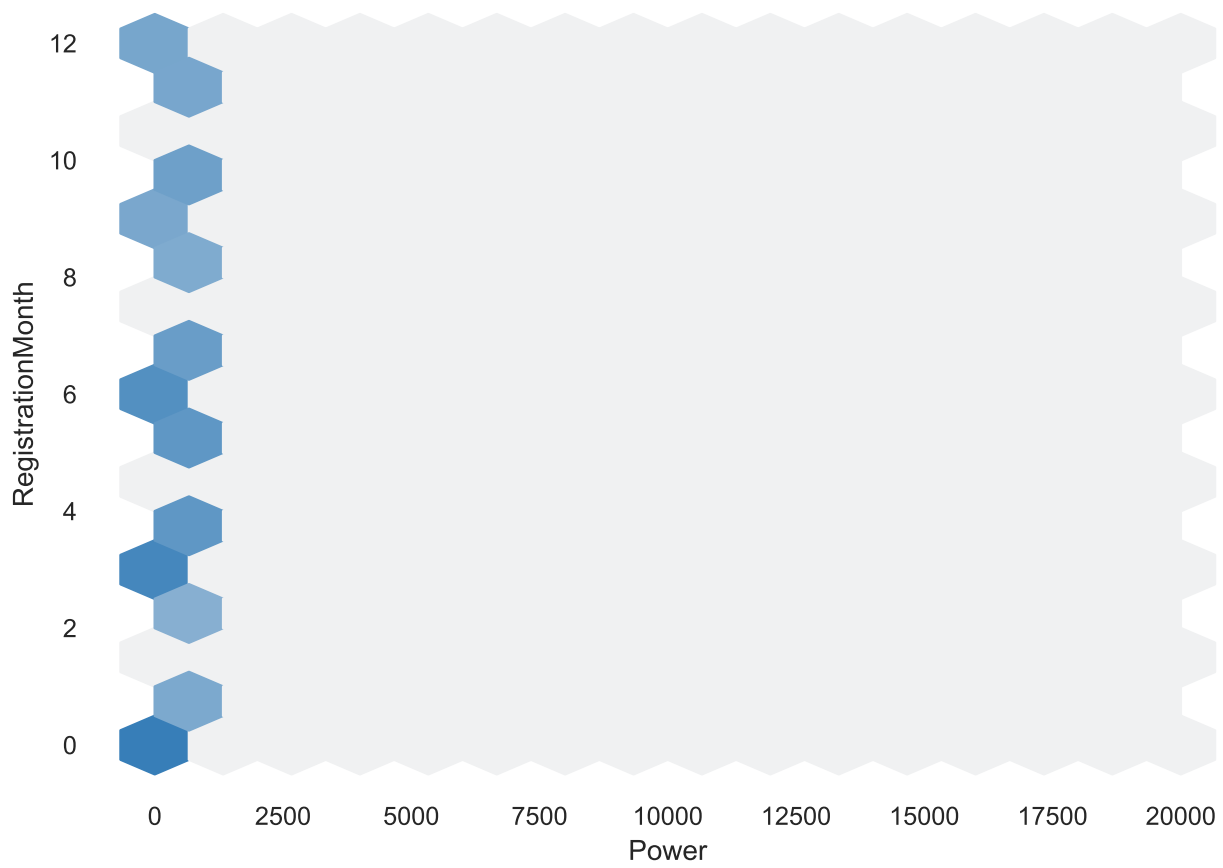


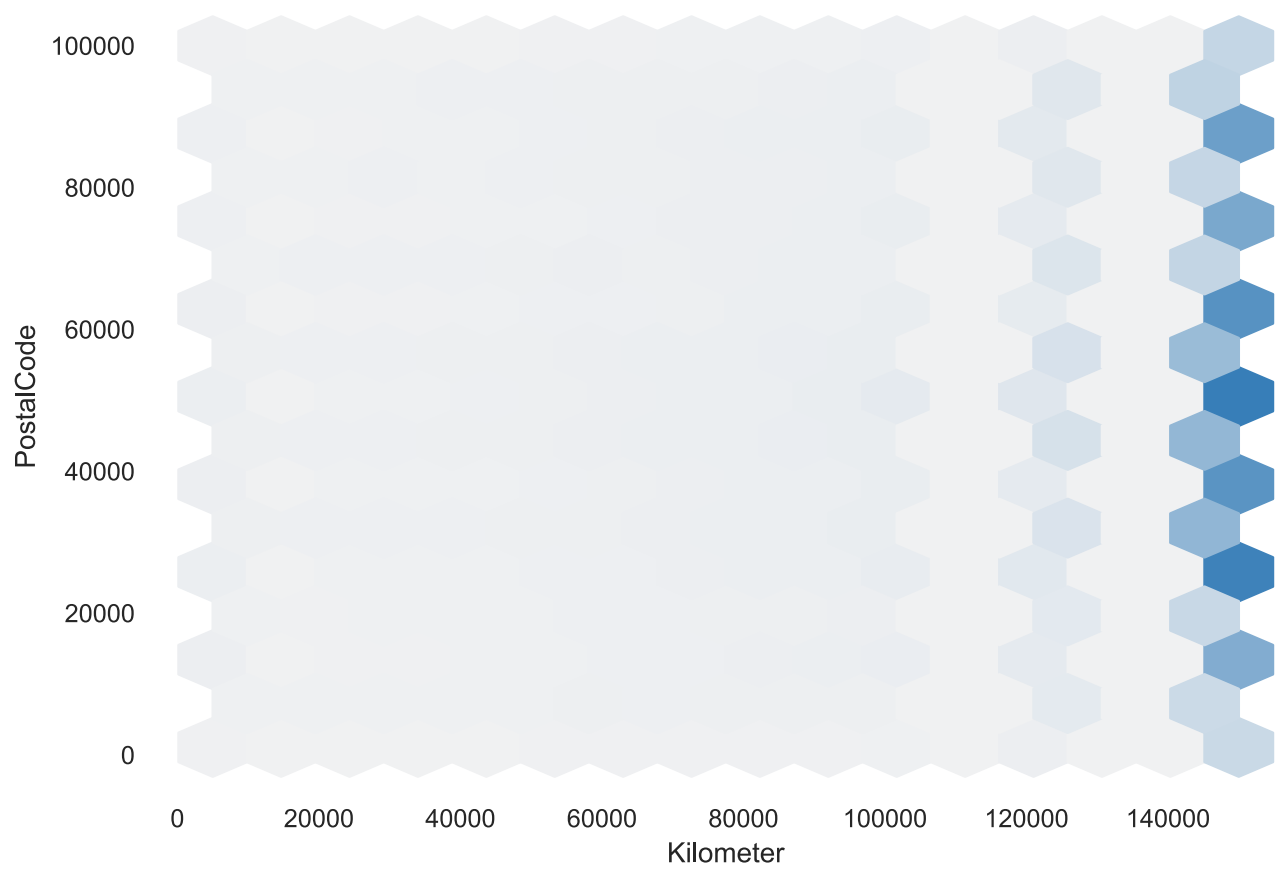


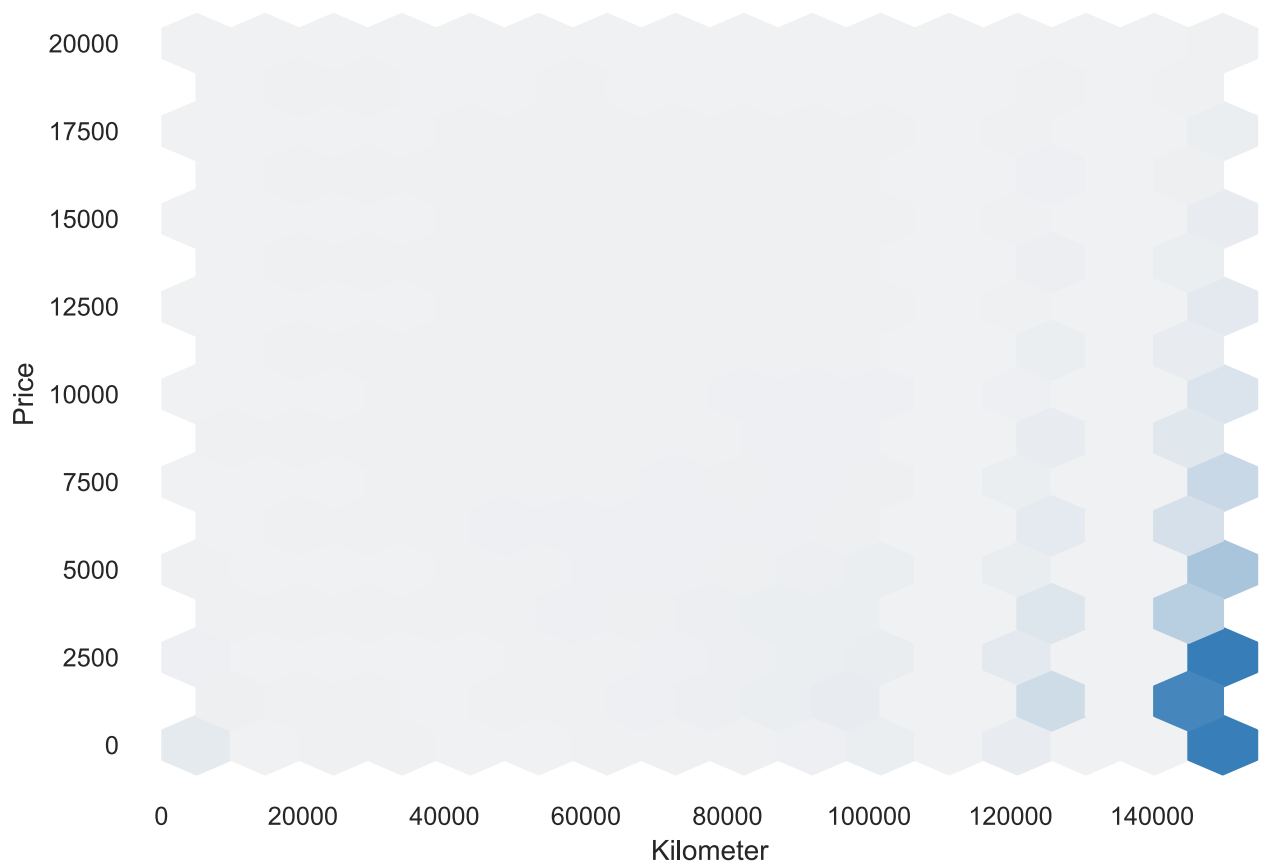


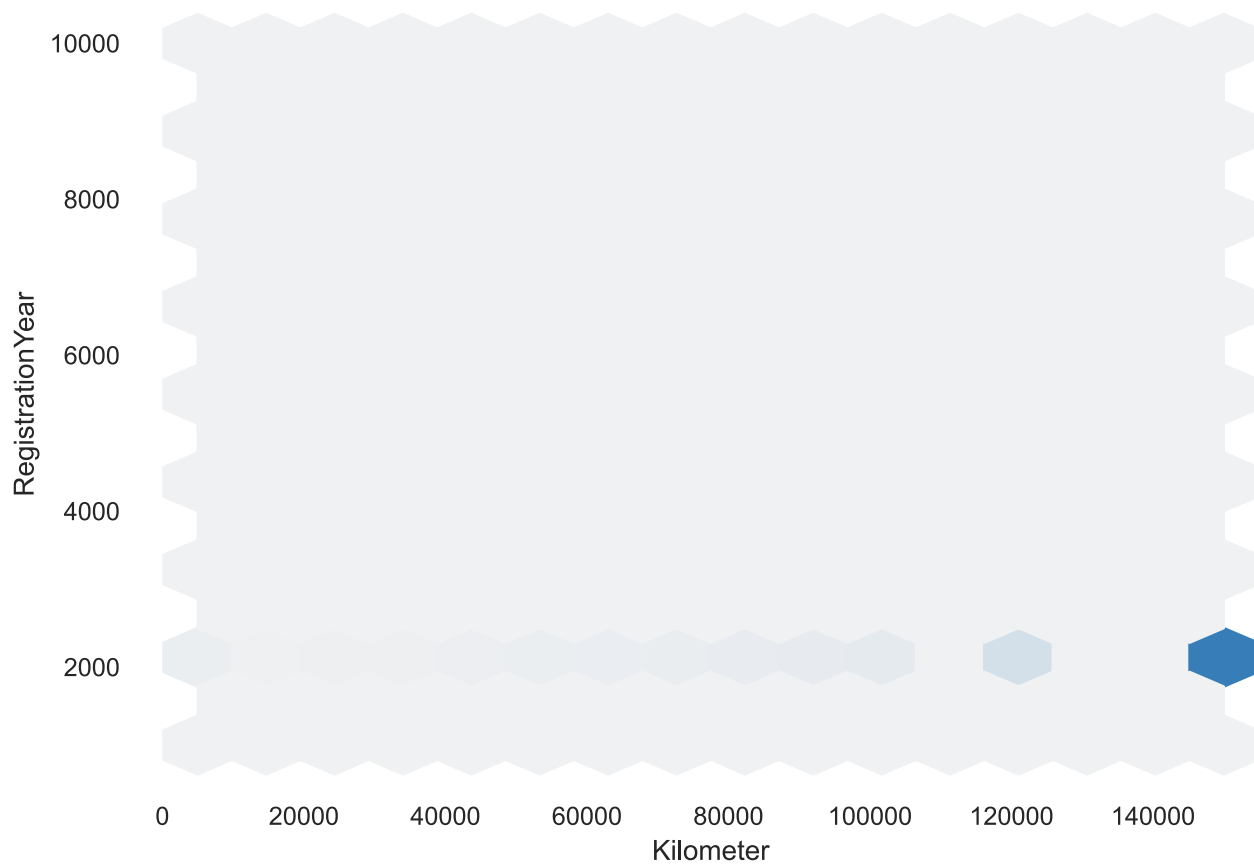


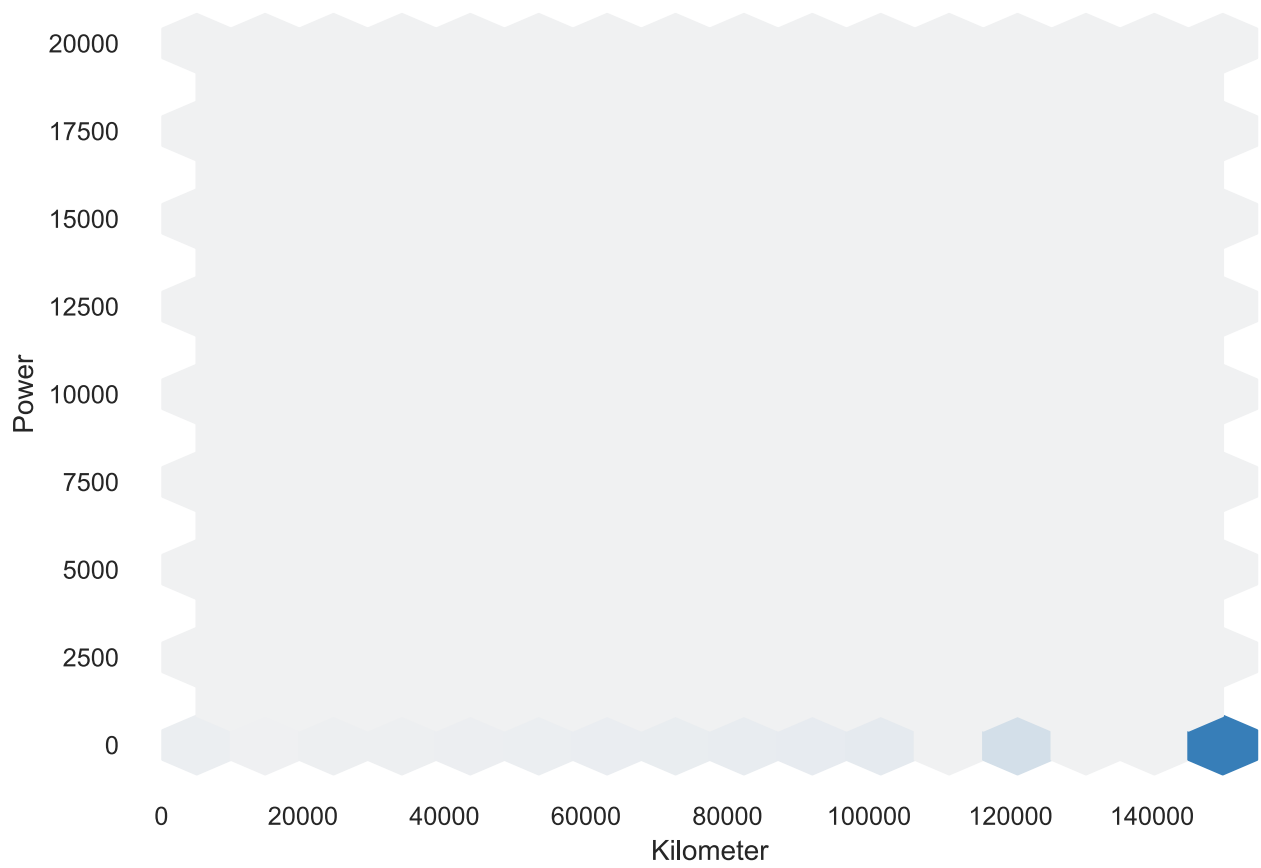


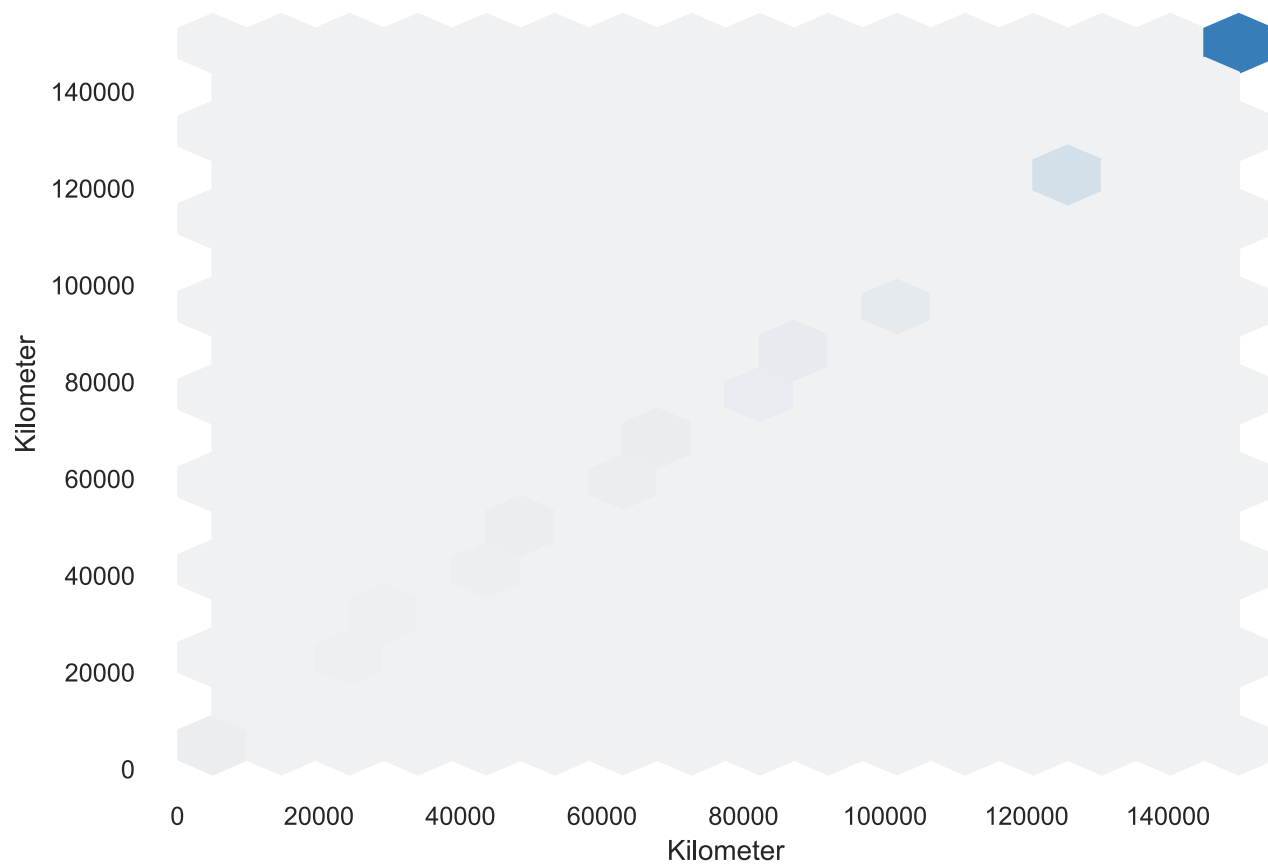


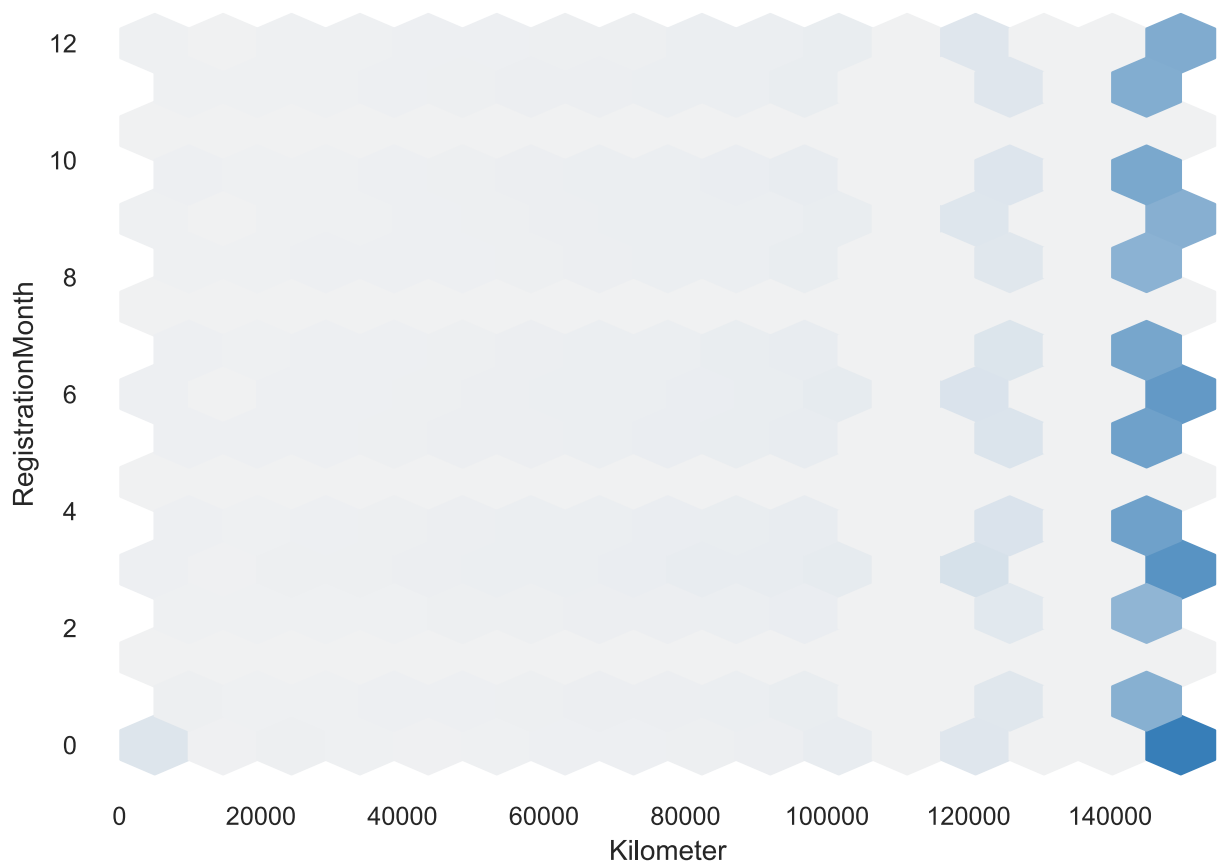


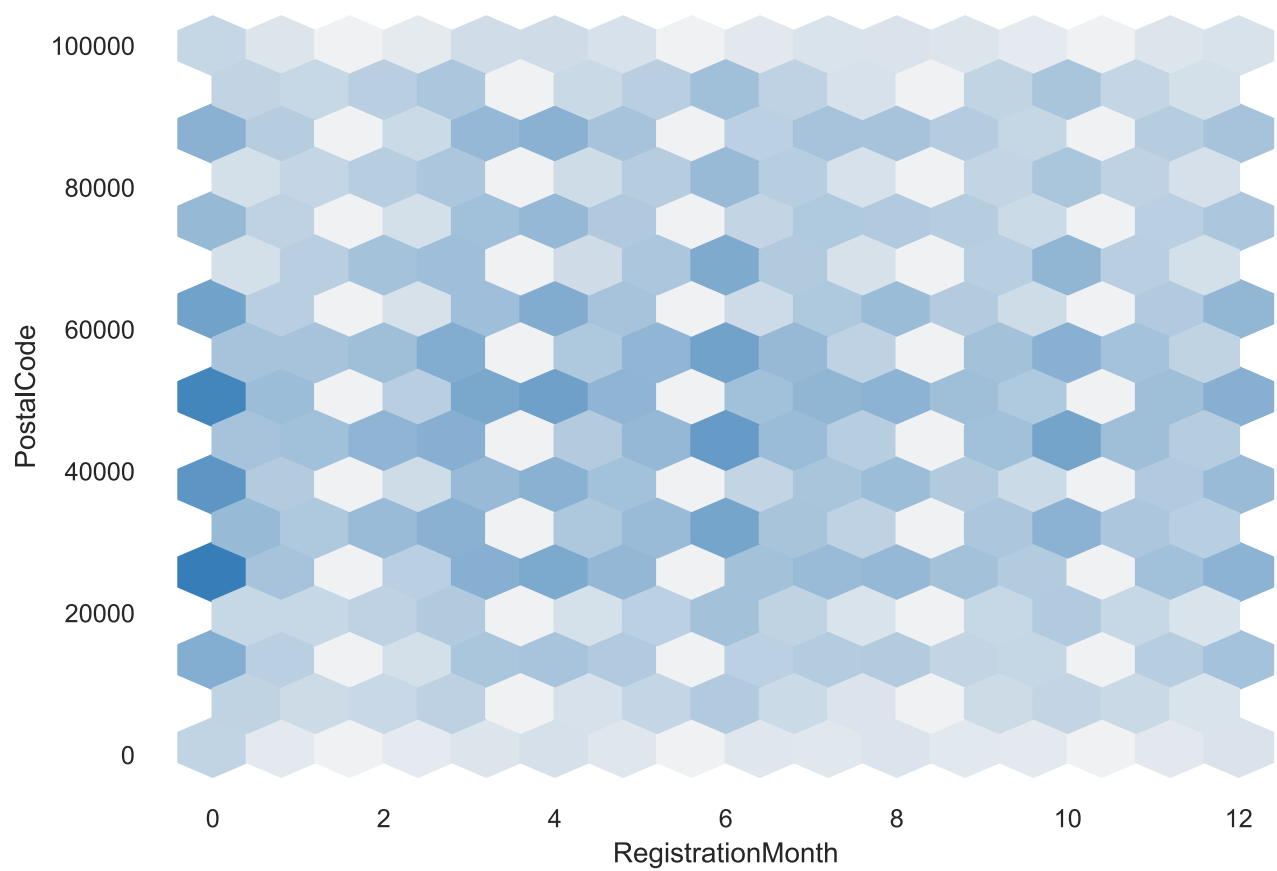


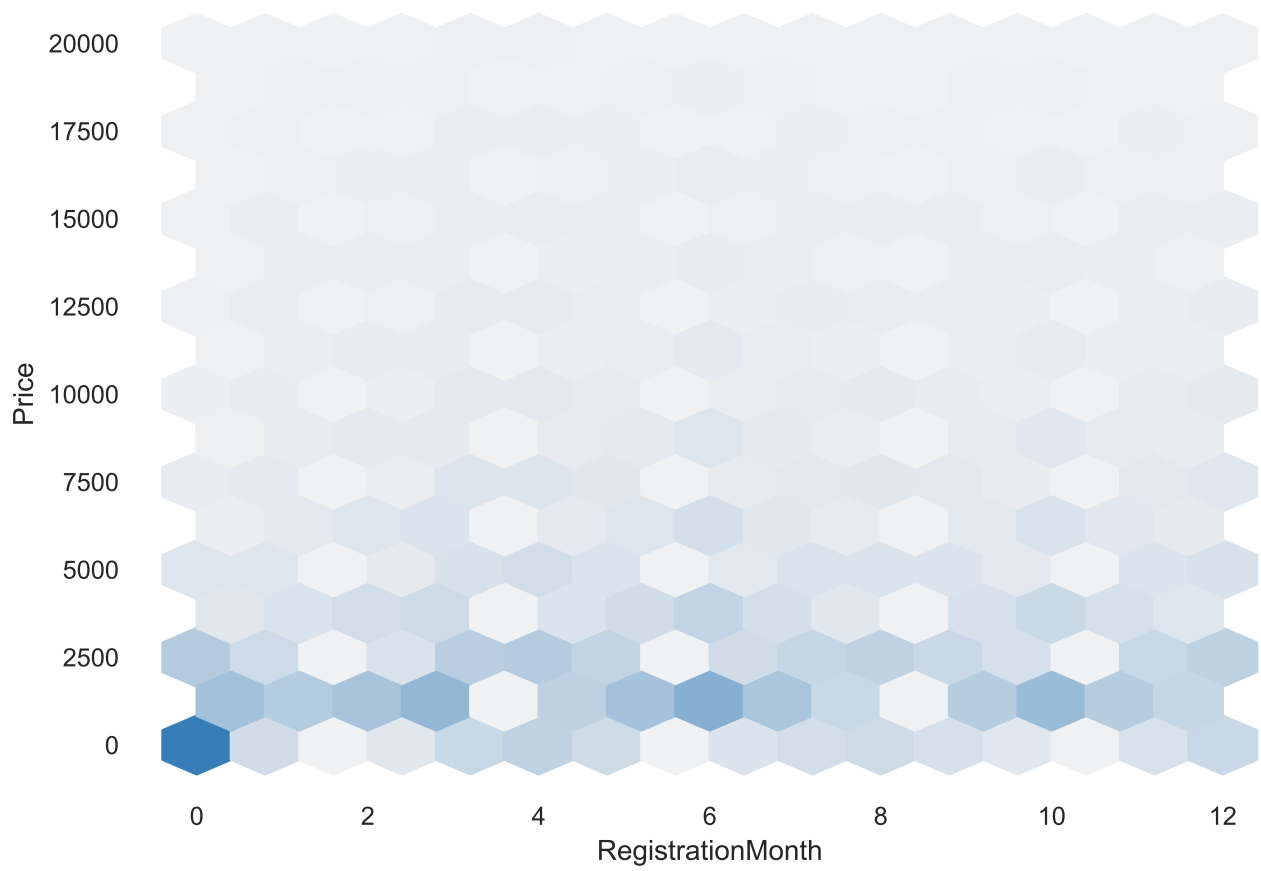


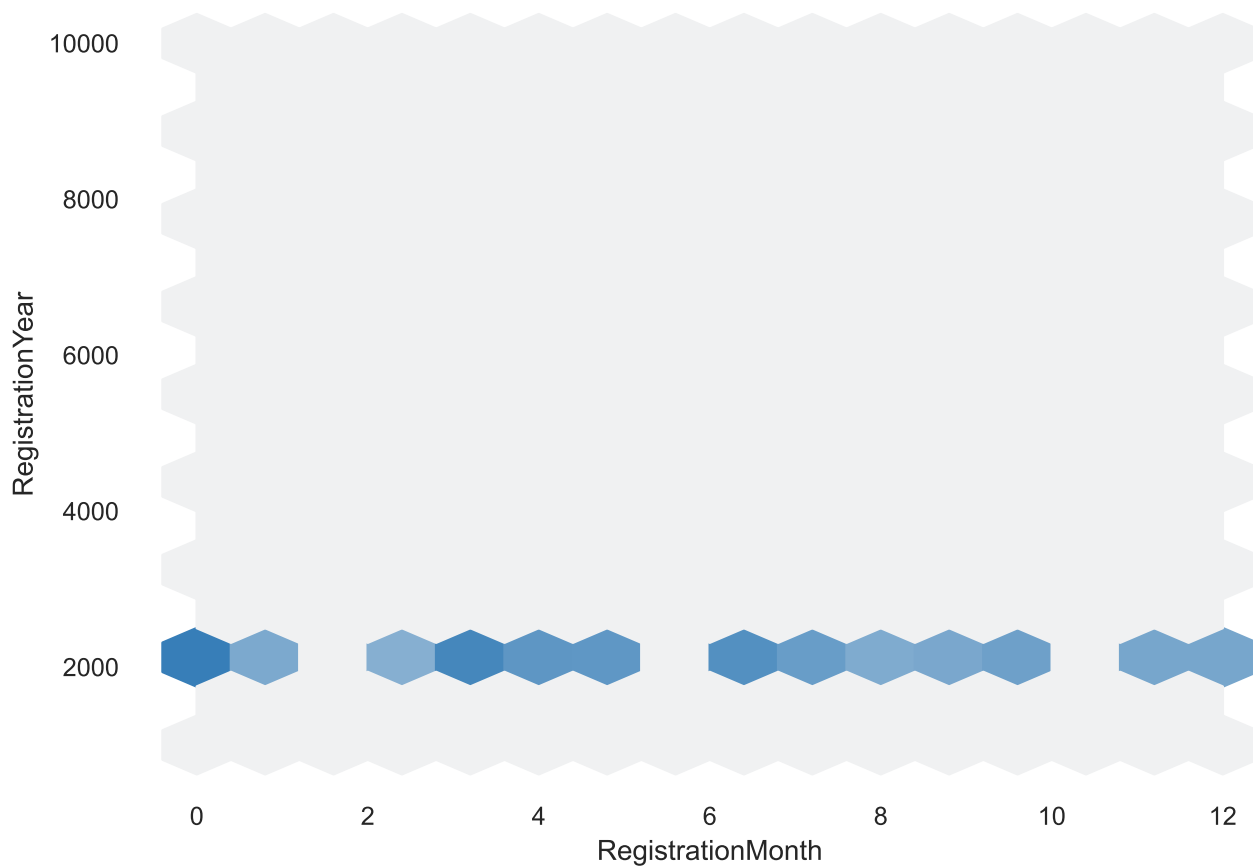


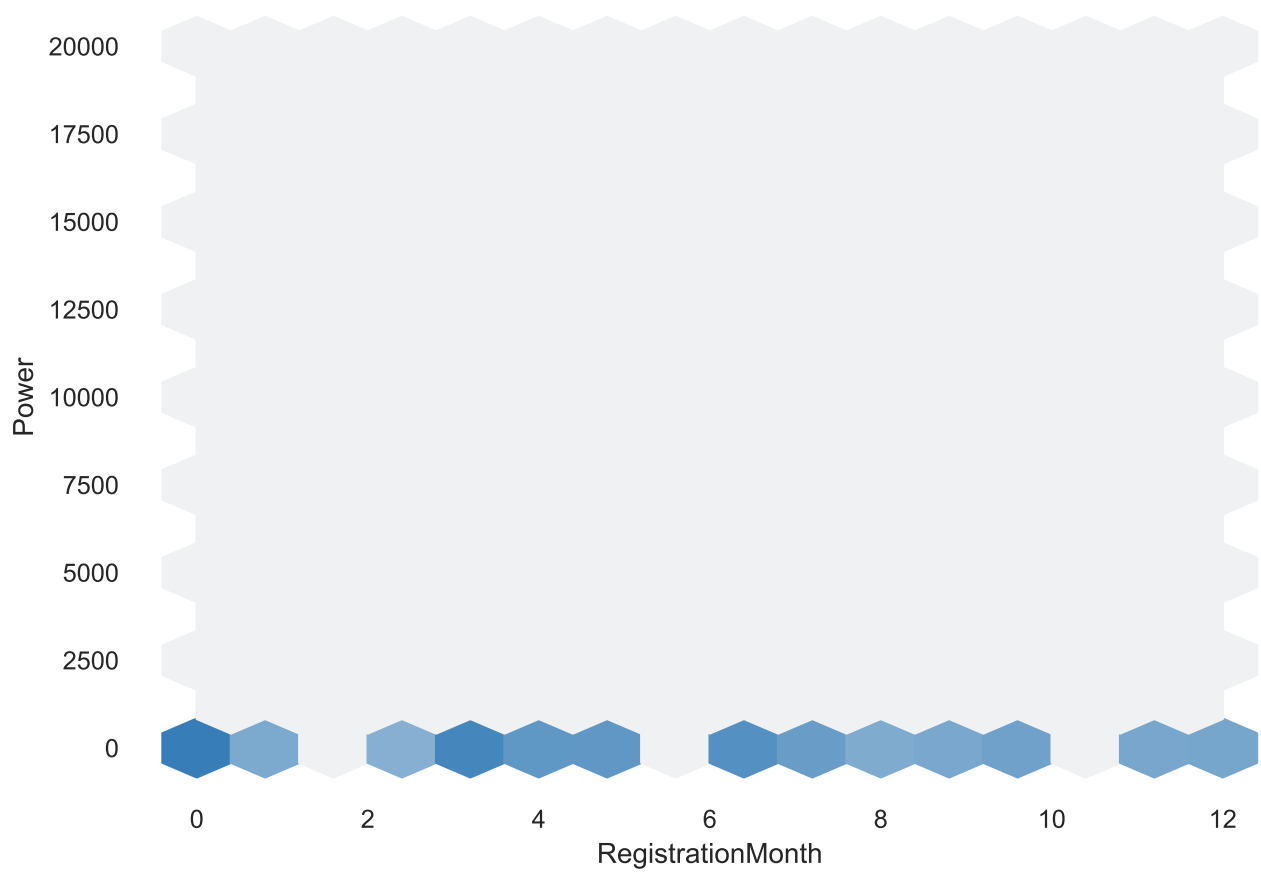


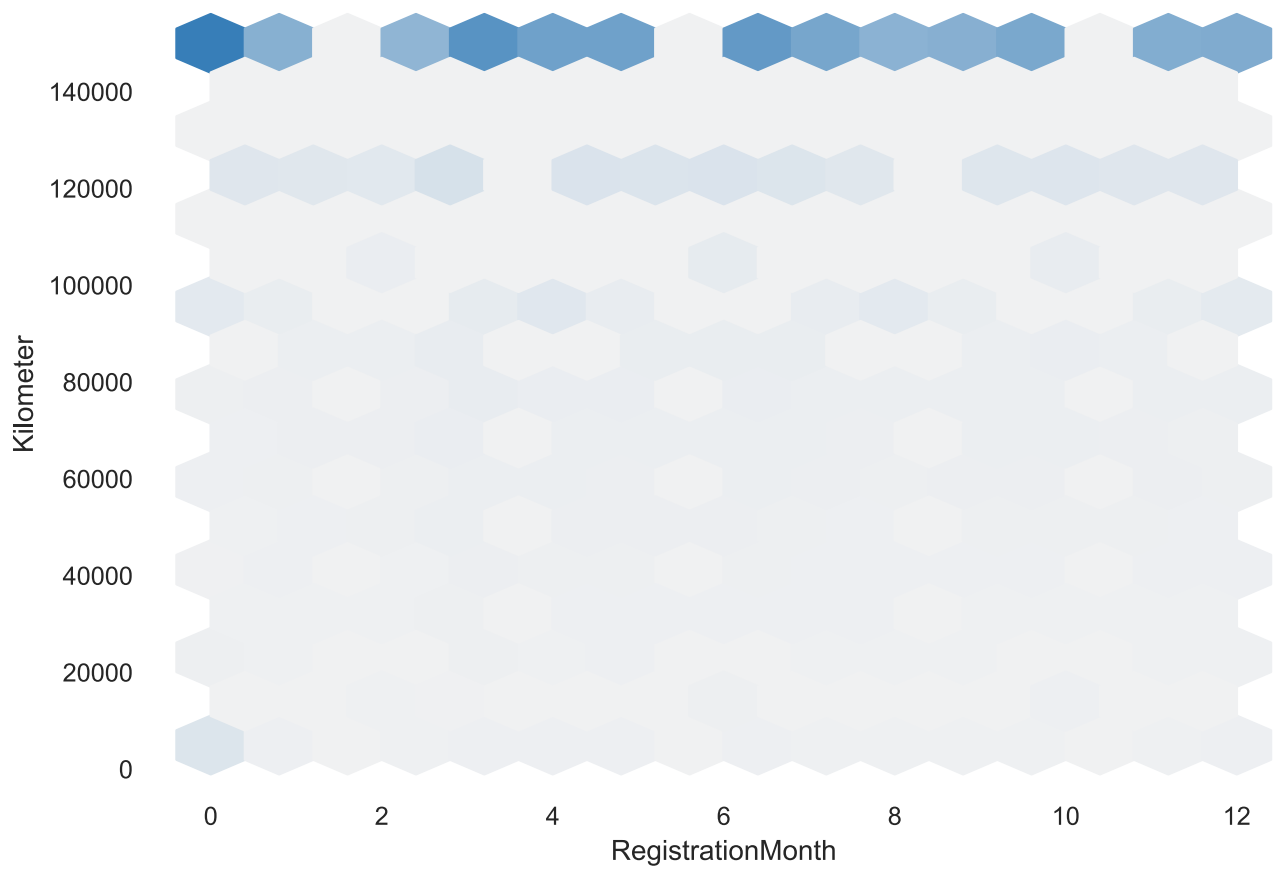


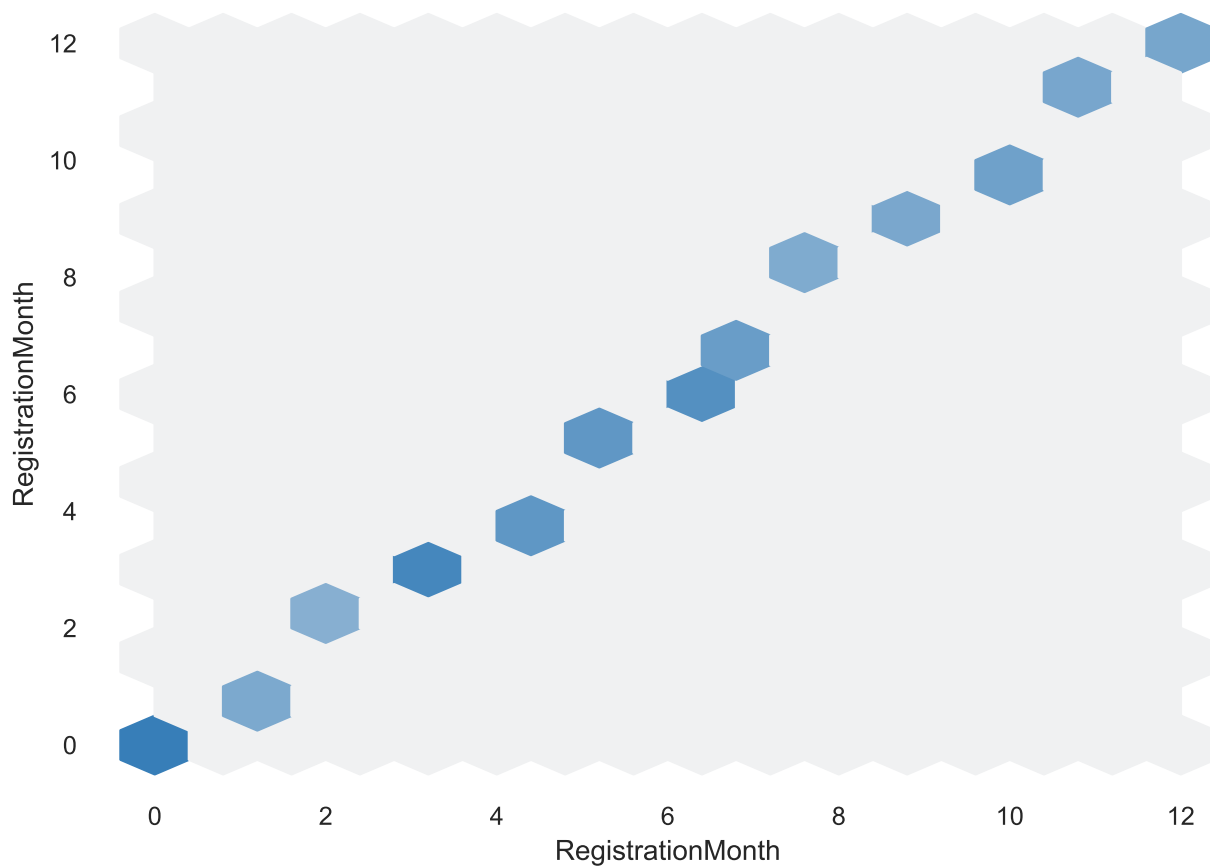


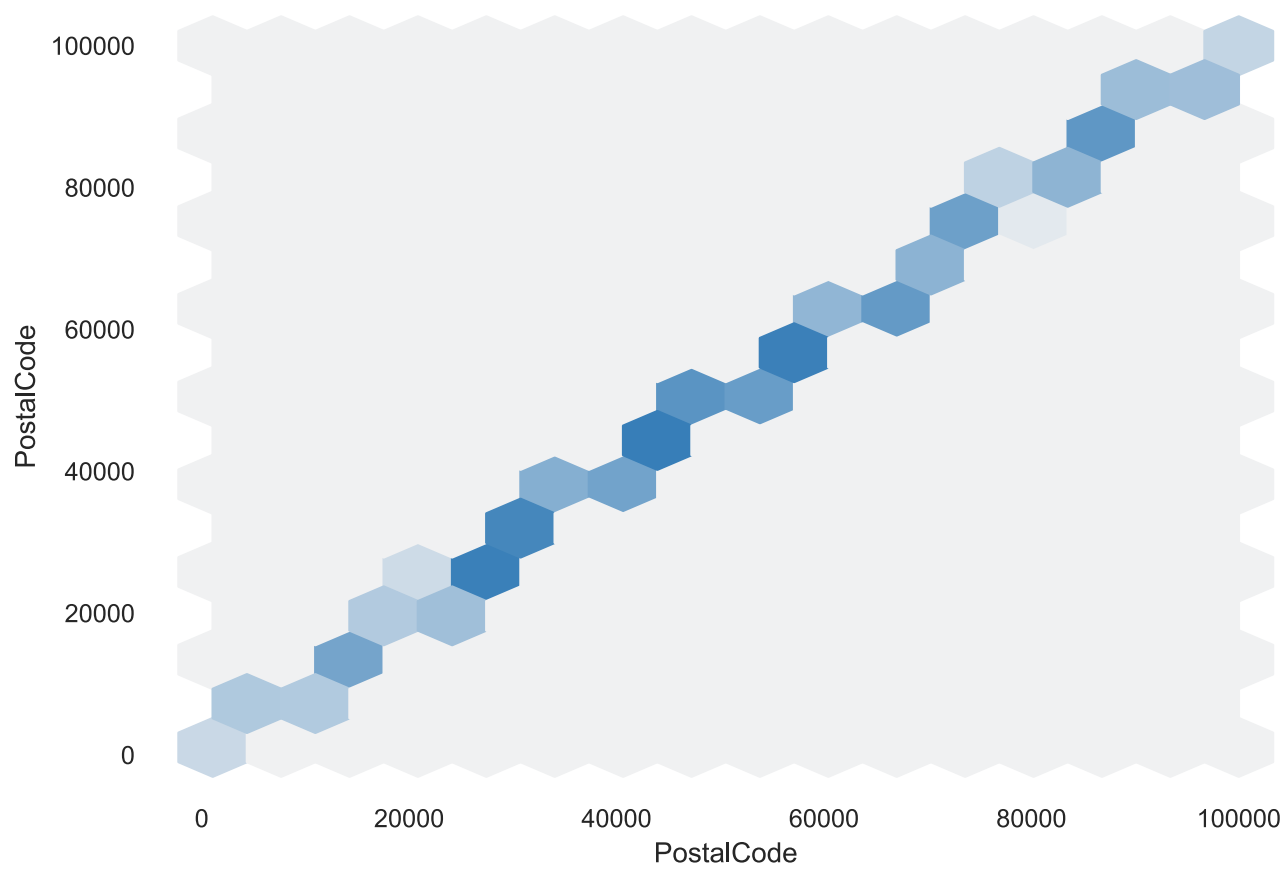


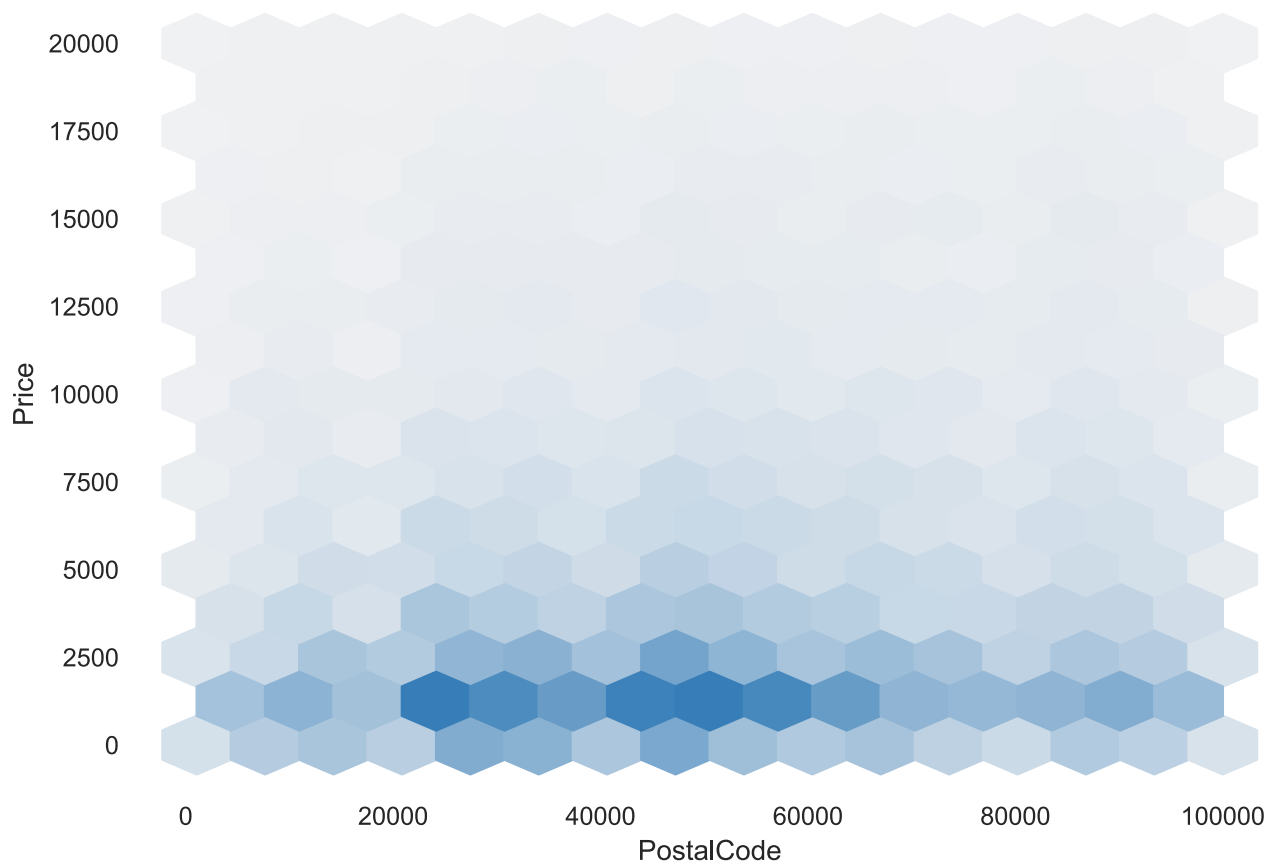


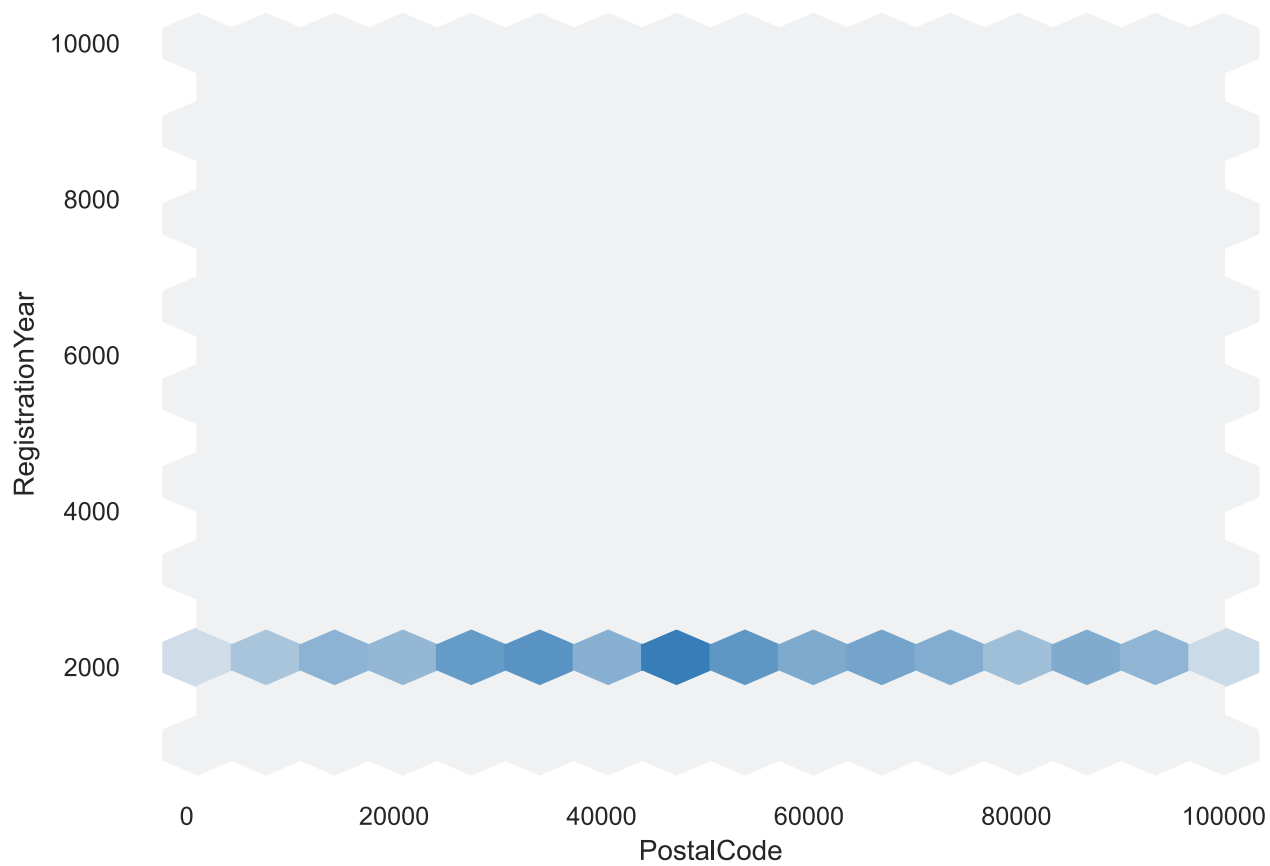


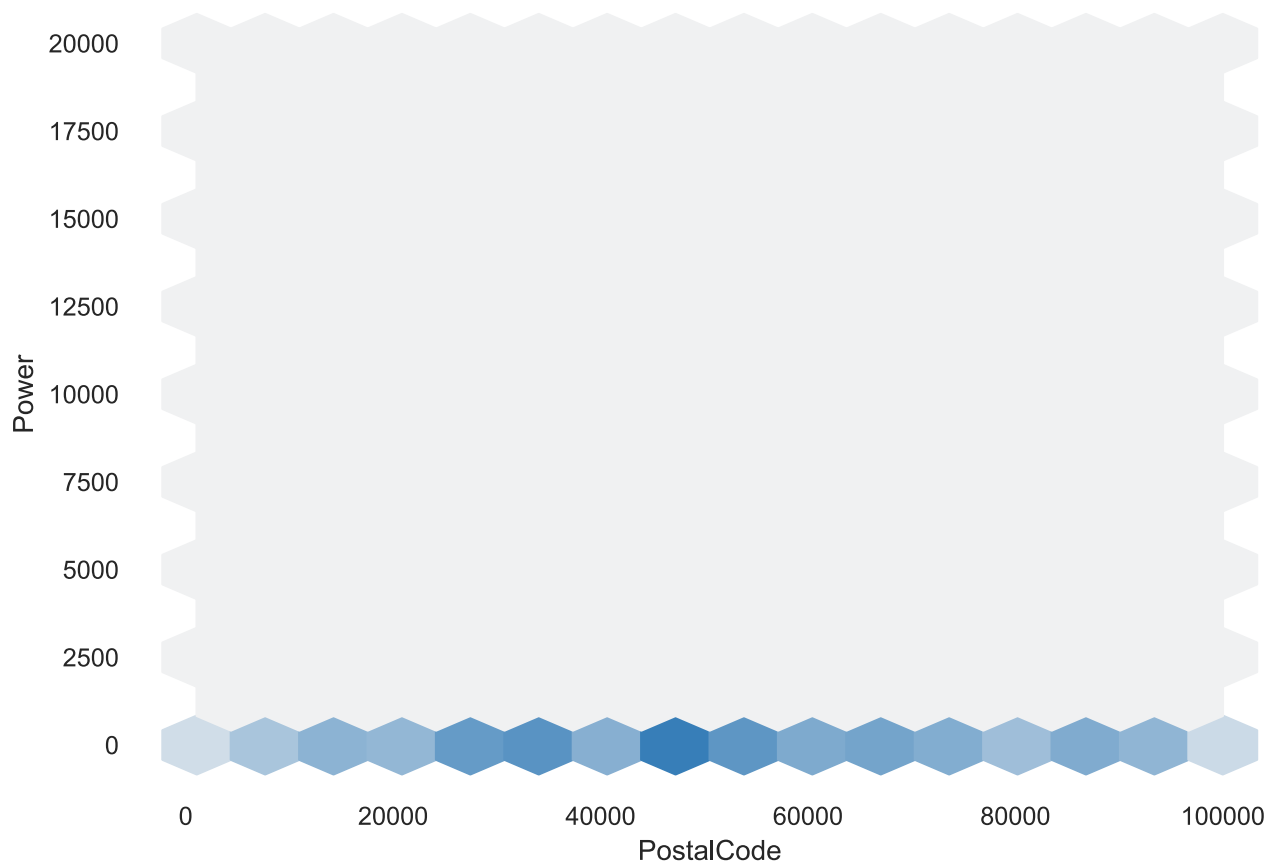


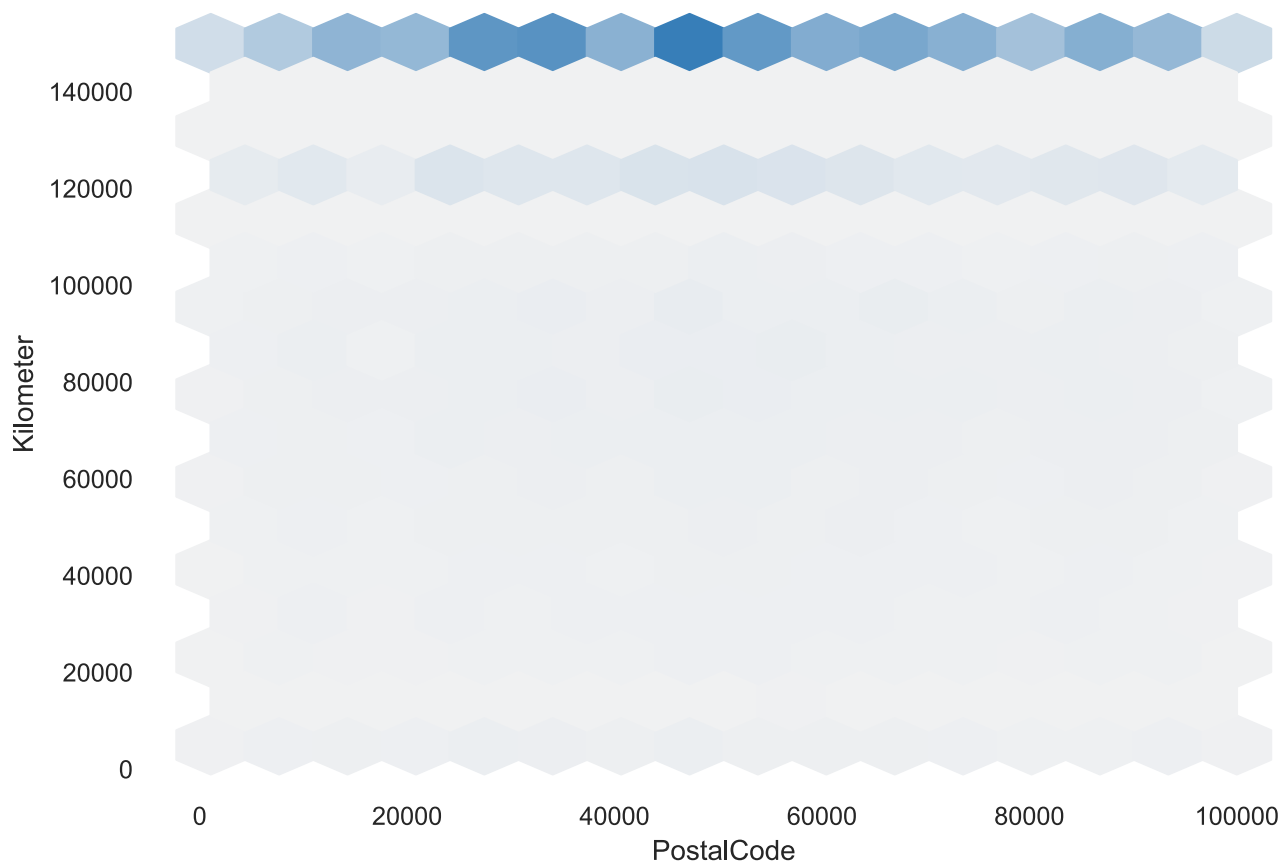


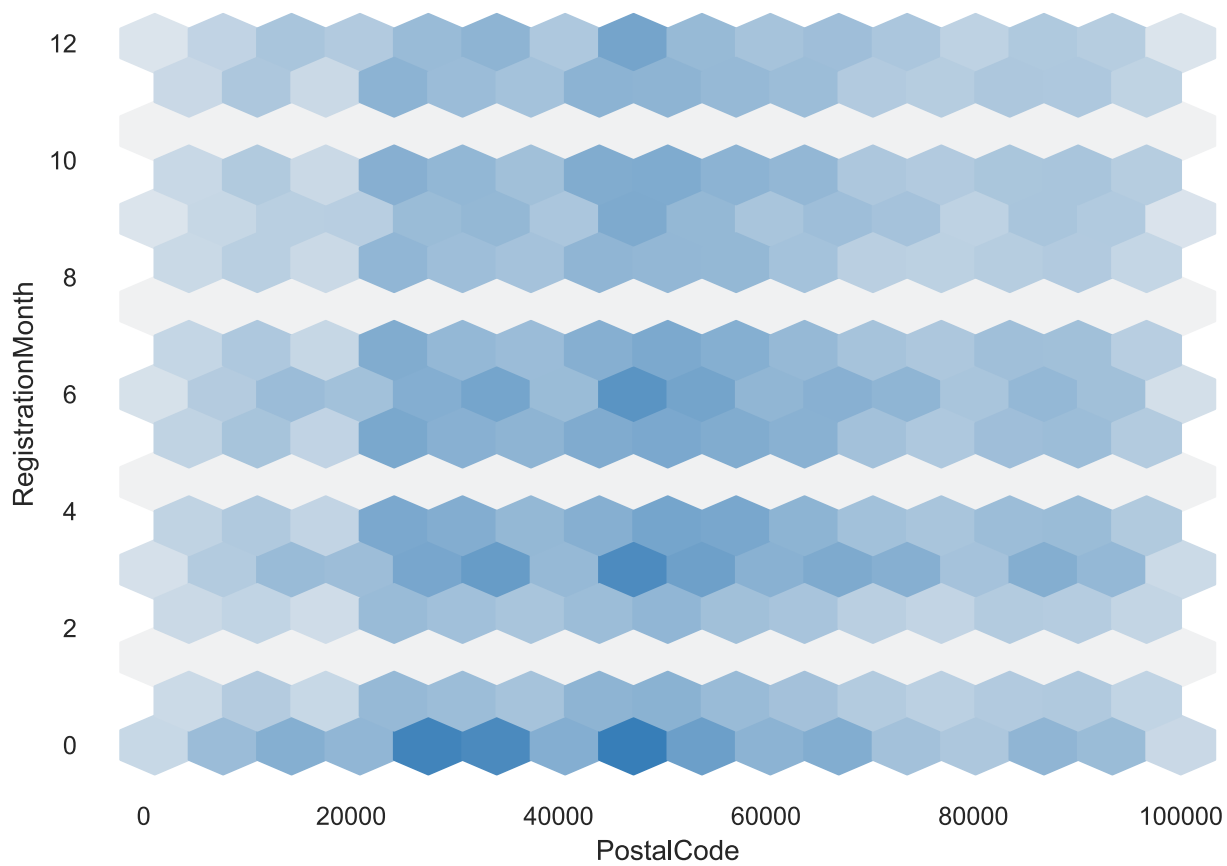




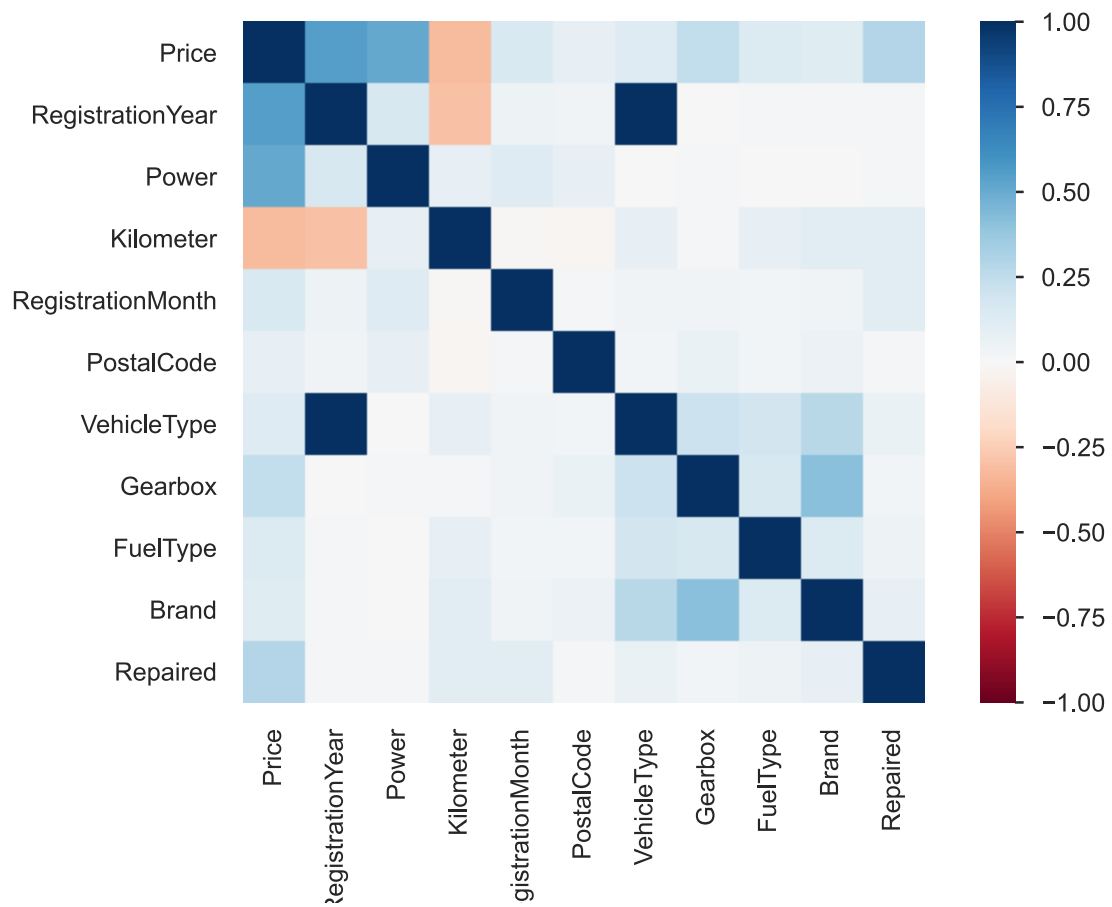






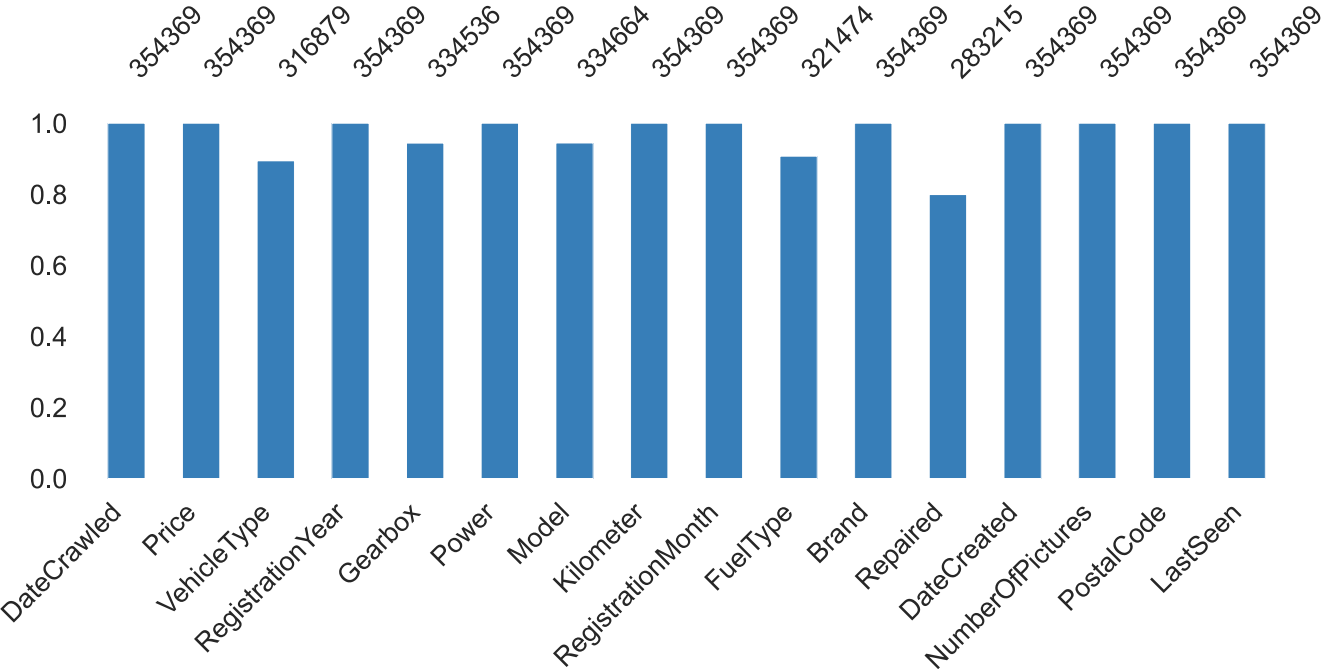


Correlations



	Price	RegistrationYear	Power	Kilometer	RegistrationMonth	PostalCode	VehicleType	Gearbox	FuelType	Brand	Repair
Price	1.000	0.551	0.513	-0.314	0.155	0.079	0.131	0.248	0.133	0.121	0.292
RegistrationYear	0.551	1.000	0.160	-0.295	0.053	0.035	1.000	0.003	0.020	0.018	0.009
Power	0.513	0.160	1.000	0.085	0.129	0.081	0.006	0.008	0.005	0.001	0.009
Kilometer	-0.314	-0.295	0.085	1.000	-0.011	-0.017	0.086	0.018	0.079	0.110	0.114
RegistrationMonth	0.155	0.053	0.129	-0.011	1.000	0.015	0.035	0.037	0.029	0.035	0.106
PostalCode	0.079	0.035	0.081	-0.017	0.015	1.000	0.031	0.070	0.025	0.059	0.014
VehicleType	0.131	1.000	0.006	0.086	0.035	0.031	1.000	0.217	0.184	0.280	0.063
Gearbox	0.248	0.003	0.008	0.018	0.037	0.070	0.217	1.000	0.162	0.417	0.028
FuelType	0.133	0.020	0.005	0.079	0.029	0.025	0.184	0.162	1.000	0.137	0.048

Missing values



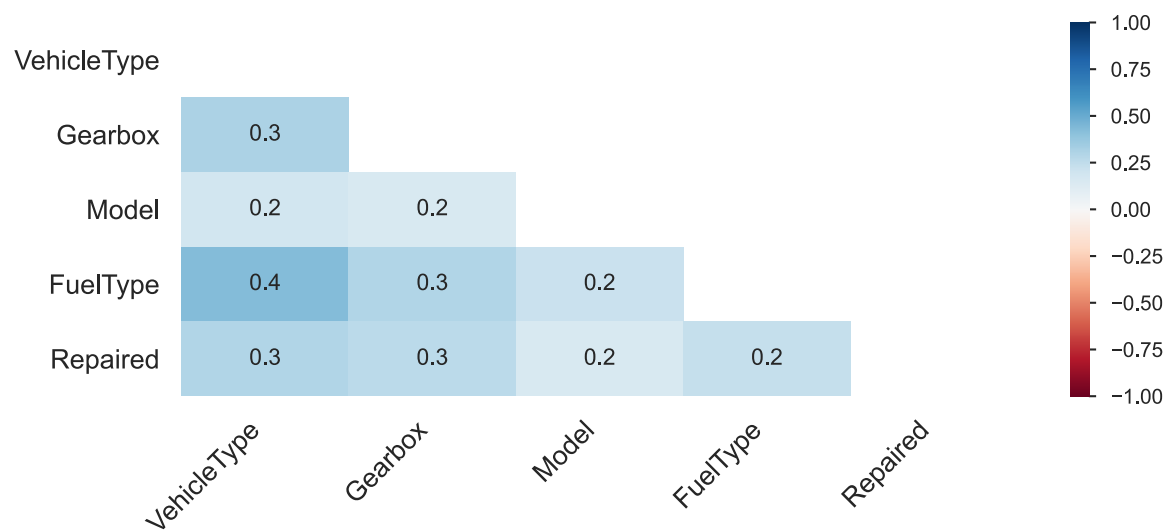
A simple visualization of nullity by column.

1

DateCrawled
Price
VehicleType
RegistrationYear
Gearbox
Power
Model
Kilometer
RegistrationMonth
FuelType
Brand
Repaired
DateCreated
NumberOfPictures
PostalCode
LastSeen

354369

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

	DateCrawled	Price	VehicleType	RegistrationYear	Gearbox	Power	Model	Kilometer	RegistrationMonth	FuelType	Brand	Repaired
0	2016-03-24 11:52:17	480	NaN	1993	manual	0	golf	150000	0	petrol	volkswagen	NaN
1	2016-03-24 10:58:45	18300	coupe	2011	manual	190	NaN	125000	5	gasoline	audi	yes
2	2016-03-14 12:52:21	9800	suv	2004	auto	163	grand	125000	8	gasoline	jeep	NaN
3	2016-03-17 16:54:04	1500	small	2001	manual	75	golf	150000	6	petrol	volkswagen	no
4	2016-03-31 17:25:20	3600	small	2008	manual	69	fabia	90000	7	gasoline	skoda	no
5	2016-04-04 17:36:23	650	sedan	1995	manual	102	3er	150000	10	petrol	bmw	yes
6	2016-04-01 20:48:51	2200	convertible	2004	manual	109	2_reihe	150000	8	petrol	peugeot	no
7	2016-03-21 18:54:38	0	sedan	1980	manual	50	other	40000	7	petrol	volkswagen	no
8	2016-04-04 23:42:13	14500	bus	2014	manual	125	c_max	30000	8	petrol	ford	NaN
9	2016-03-17 10:53:50	999	small	1998	manual	101	golf	150000	0	NaN	volkswagen	NaN



	DateCrawled	Price	VehicleType	RegistrationYear	Gearbox	Power	Model	Kilometer	RegistrationMonth	FuelType	Brand
354359	2016-03-28 13:48:07	7900	sedan	2010	manual	140	golf	150000	7	gasoline	volkswagen
354360	2016-04-02 20:37:03	3999	wagon	2005	manual	3	3er	150000	5	gasoline	bmw
354361	2016-03-09 13:37:43	5250	NaN	2016	auto	150	159	150000	12	NaN	alfa_romeo
354362	2016-03-19 19:53:49	3200	sedan	2004	manual	225	leon	150000	5	petrol	seat
354363	2016-03-27 20:36:20	1150	bus	2000	manual	0	zafira	150000	3	petrol	opel
354364	2016-03-21 09:50:58	0	NaN	2005	manual	0	colt	150000	7	petrol	mitsubishi
354365	2016-03-14 17:48:27	2200	NaN	2005	NaN	0	NaN	20000	1	NaN	sonstige_autos
354366	2016-03-05 19:56:21	1199	convertible	2000	auto	101	fortwo	125000	3	petrol	smart
354367	2016-03-19 18:57:12	9200	bus	1996	manual	102	transporter	150000	3	gasoline	volkswagen
354368	2016-03-20 19:41:08	3400	wagon	2002	manual	100	golf	150000	6	gasoline	volkswagen

Duplicate rows

Most frequently occurring

	DateCrawled	Price	VehicleType	RegistrationYear	Gearbox	Power	Model	Kilometer	RegistrationMonth	FuelType	Brand	Repaired	Index
0	2016-03-08 18:42:48	1799	coupe	1999	auto	193	clk	20000	7	petrol	mercedes_benz	no	2
1	2016-03-18 18:46:15	1999	wagon	2001	manual	131	passat	150000	7	gasoline	volkswagen	no	2
2	2016-03-28 00:56:10	1000	small	2002	manual	83	other	150000	1	petrol	suzuki	no	2
3	2016-04-03 09:01:15	4699	coupe	2003	auto	218	clk	125000	6	petrol	mercedes_benz	yes	2