Overview

Dataset statistics

Number of variables	17
Number of observations	166800
Missing cells	395
Missing cells (%)	< 0.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	21.6 MiB
Average record size in memory	136.0 B

Variable types

Text	6
Categorical	4
Numeric	7

Alerts

2020 Census Tract is highly overall correlated with State	**High correlation**
Clean Alternative Fuel Vehicle (CAFV) Eligibility is highly overall correlated with Electric Range and 2 other fields (Electric Range, Electric Vehicle Type, Make)	**High correlation**
Electric Range is highly overall correlated with Clean Alternative Fuel Vehicle (CAFV) Eligibility and 2 other fields (Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Vehicle Type, Model Year)	**High correlation**
Electric Vehicle Type is highly overall correlated with Clean Alternative Fuel Vehicle (CAFV) Eligibility and 2 other fields (Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Range, Make)	**High correlation**
Legislative District is highly overall correlated with State	**High correlation**
Make is highly overall correlated with Clean Alternative Fuel Vehicle (CAFV) Eligibility and 1 other fields (Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Vehicle Type)	**High correlation**
Model Year is highly overall correlated with Electric Range	**High correlation**
Postal Code is highly overall correlated with State	**High correlation**
State is highly overall correlated with 2020 Census Tract and 2 other fields (2020 Census Tract, Legislative District, Postal Code)	**High correlation**
State is highly imbalanced (99.4%)	**Imbalance**
Postal Code is highly skewed (γ1 = -30.09608692)	**Skewed**
2020 Census Tract is highly skewed (γ1 = -26.96677648)	**Skewed**
DOL Vehicle ID has unique values	**Unique**
Electric Range has 83517 (50.1%) zeros	**Zeros**
Base MSRP has 163437 (98.0%) zeros	**Zeros**

Reproduction

Analysis started	2024-02-04 09:22:58.579780
Analysis finished	2024-02-04 09:23:40.608845
Duration	42.03 seconds
Software version	ydata-profiling vv4.6.4 (https://github.com/ydataai/ydata-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Exploratory%20Data%20Analysis%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22creator%22%3A%20%22%22%2C%20%22author%

Variables

Select Columns ▾

VIN (1-10)
Text

Distinct	10316
Distinct (%)	6.2%
Missing	0
Missing (%)	0.0%
Memory size	1.3 MiB

Length

Max length	10
Median length	10
Mean length	10
Min length	10

Characters and Unicode

Total characters	1668000	
Distinct characters	34	
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	1969	?
Unique (%)	1.2%	

Sample

1st row	3C3CFFGE4E
2nd row	5YJXCBE40H
3rd row	3MW39FS03P
4th row	7PDSGABA8P
5th row	5YJ3E1EB8L

Value	Count	Frequency (%)
7saygdee6p	1114	0.7%
7saygdee7p	1090	0.7%
7saygdeexp	1071	0.6%
7saygdee8p	1071	0.6%
7saygdee0p	1052	0.6%
7saygdee9p	1041	0.6%
7saygdee5p	1037	0.6%
7saygdee2p	1029	0.6%
7saygdee1p	1022	0.6%
7saygdee3p	1014	0.6%
Other values (10306)	156259	93.7%

Most occurring characters

Value	Count	Frequency (%)
E	150671	9.0%
1	121417	7.3%
A	104815	6.3%
Y	94860	5.7%
J	84900	5.1%
P	82186	4.9%
5	79952	4.8%
3	69313	4.2%
D	64150	3.8%
G	64092	3.8%
Other values (24)	751644	45.1%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	1144298	68.6%
Decimal Number	523702	31.4%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
E	150671	13.2%
A	104815	9.2%
Y	94860	8.3%
J	84900	7.4%
P	82186	7.2%
D	64150	5.6%
G	64092	5.6%
C	56828	5.0%
N	55104	4.8%
S	53497	4.7%
Other values (14)	333195	29.1%

Decimal Number

Value	Count	Frequency (%)
1	121417	23.2%
5	79952	15.3%
3	69313	13.2%
7	50634	9.7%
4	47016	9.0%
0	43598	8.3%
6	38554	7.4%
2	34411	6.6%
8	22088	4.2%
9	16719	3.2%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1144298	68.6%
Common	523702	31.4%

Most frequent character per script

Latin

Value	Count	Frequency (%)
E	150671	13.2%
A	104815	9.2%
Y	94860	8.3%
J	84900	7.4%
P	82186	7.2%
D	64150	5.6%
G	64092	5.6%
C	56828	5.0%
N	55104	4.8%
S	53497	4.7%
Other values (14)	333195	29.1%

Common

Value	Count	Frequency (%)
1	121417	23.2%
5	79952	15.3%
3	69313	13.2%
7	50634	9.7%
4	47016	9.0%
0	43598	8.3%
6	38554	7.4%
2	34411	6.6%
8	22088	4.2%
9	16719	3.2%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1668000	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
E	150671	9.0%
1	121417	7.3%
A	104815	6.3%
Y	94860	5.7%
J	84900	5.1%
P	82186	4.9%
5	79952	4.8%
3	69313	4.2%
D	64150	3.8%
G	64092	3.8%
Other values (24)	751644	45.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1668000	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
E	150671	9.0%
1	121417	7.3%

County
Text

Distinct	187
Distinct (%)	0.1%
Missing	5
Missing (%)	< 0.1%
Memory size	1.3 MiB

Length

Max length	20
Median length	4
Mean length	5.4757697
Min length	3

Characters and Unicode

Total characters	913331	
Distinct characters	52	
Distinct categories	5 (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	86	?
Unique (%)	0.1%	

Sample

1st row	Yakima
2nd row	Thurston
3rd row	King
4th row	Snohomish
5th row	King

Value	Count	Frequency (%)
king	86594	51.3%
snohomish	19570	11.6%
pierce	12972	7.7%
clark	9847	5.8%
thurston	6042	3.6%
kitsap	5522	3.3%
spokane	4312	2.6%
whatcom	4039	2.4%
benton	2028	1.2%
skagit	1842	1.1%
Other values (192)	16096	9.5%

Most occurring characters

Value	Count	Frequency (%)
i	132573	14.5%
n	130137	14.2%
K	92994	10.2%
g	89098	9.8%
o	59669	6.5%
h	50618	5.5%
a	42008	4.6%
s	37326	4.1%
e	36776	4.0%
r	32990	3.6%
Other values (42)	209142	22.9%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	742390	81.3%
Uppercase Letter	168864	18.5%
Space Separator	2069	0.2%
Other Punctuation	7	< 0.1%
Dash Punctuation	1	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	132573	17.9%
n	130137	17.5%
g	89098	12.0%
o	59669	8.0%
h	50618	6.8%
a	42008	5.7%
s	37326	5.0%
e	36776	5.0%
r	32990	4.4%
m	26222	3.5%
Other values (15)	104973	14.1%

Uppercase Letter

Value	Count	Frequency (%)
K	92994	55.1%
S	27086	16.0%
P	13255	7.8%
C	12763	7.6%
T	6047	3.6%
W	5238	3.1%
B	2051	1.2%
J	1863	1.1%
I	1797	1.1%
G	1139	0.7%
Other values (13)	4631	2.7%

Other Punctuation

Value	Count	Frequency (%)
'	5	71.4%
.	2	28.6%

Space Separator

Value	Count	Frequency (%)
	2069	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	1	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	911254	99.8%
Common	2077	0.2%

Most frequent character per script

Latin

Value	Count	Frequency (%)
i	132573	14.5%
n	130137	14.3%
K	92994	10.2%
g	89098	9.8%
o	59669	6.5%
h	50618	5.6%
a	42008	4.6%
s	37326	4.1%
e	36776	4.0%
r	32990	3.6%
Other values (38)	207065	22.7%

Common

Value	Count	Frequency (%)
	2069	99.6%
'	5	0.2%
.	2	0.1%
-	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	913331	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	132573	14.5%
n	130137	14.2%
K	92994	10.2%
g	89098	9.8%
o	59669	6.5%
h	50618	5.5%
a	42008	4.6%
s	37326	4.1%
e	36776	4.0%
r	32990	3.6%
Other values (42)	209142	22.9%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	913331	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	132573	14.5%
n	130137	14.2%
K	92994	10.2%

City
Text

Distinct	704
Distinct (%)	0.4%
Missing	5
Missing (%)	< 0.1%
Memory size	1.3 MiB

Length

Max length	24
Median length	23
Mean length	8.215648
Min length	3

Characters and Unicode

Total characters	1370329	
Distinct characters	53	
Distinct categories	4 (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	227	?
Unique (%)	0.1%	

Sample

1st row	Yakima
2nd row	Olympia
3rd row	Renton
4th row	Bothell
5th row	Kent

Value	Count	Frequency (%)
seattle	27831	14.4%
bellevue	8364	4.3%
redmond	6032	3.1%
vancouver	5869	3.0%
bothell	5440	2.8%
kirkland	5028	2.6%
sammamish	4876	2.5%
island	4773	2.5%
renton	4617	2.4%
olympia	4058	2.1%
Other values (744)	116642	60.3%

Most occurring characters

Value	Count	Frequency (%)
e	186942	13.6%
a	133494	9.7%
l	121337	8.9%
t	95467	7.0%
n	90732	6.6%
o	80841	5.9%
r	56904	4.2%
i	54412	4.0%
S	47328	3.5%
d	45930	3.4%
Other values (43)	456942	33.3%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1149695	83.9%
Uppercase Letter	193715	14.1%
Space Separator	26735	2.0%
Dash Punctuation	184	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	186942	16.3%
a	133494	11.6%
l	121337	10.6%
t	95467	8.3%
n	90732	7.9%
o	80841	7.0%
r	56904	4.9%
i	54412	4.7%
d	45930	4.0%
m	44789	3.9%
Other values (16)	238847	20.8%

Uppercase Letter

Value	Count	Frequency (%)
S	47328	24.4%
B	24749	12.8%
R	13350	6.9%
L	10044	5.2%
K	10012	5.2%
M	9691	5.0%
V	9189	4.7%
T	8027	4.1%
I	7470	3.9%
P	7443	3.8%
Other values (15)	46412	24.0%

Space Separator

Value	Count	Frequency (%)
	26735	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	184	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1343410	98.0%
Common	26919	2.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	186942	13.9%
a	133494	9.9%
l	121337	9.0%
t	95467	7.1%
n	90732	6.8%
o	80841	6.0%
r	56904	4.2%
i	54412	4.1%
S	47328	3.5%
d	45930	3.4%
Other values (41)	430023	32.0%

Common

Value	Count	Frequency (%)
	26735	99.3%
-	184	0.7%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1370329	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	186942	13.6%
a	133494	9.7%
l	121337	8.9%
t	95467	7.0%
n	90732	6.6%
o	80841	5.9%
r	56904	4.2%
i	54412	4.0%
S	47328	3.5%
d	45930	3.4%
Other values (43)	456942	33.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1370329	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	186942	13.6%
a	133494	9.7%
l	121337	8.9%

State
Categorical

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: 2020 Census Tract, Legislative District, Postal Code)   IMBALANCE

Distinct	44
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.3 MiB

Length

Max length	2
Median length	2
Mean length	2
Min length	2

Characters and Unicode

Total characters	333600	
Distinct characters	25	
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	10	?
Unique (%)	< 0.1%	

Sample

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

Common Values

Value	Count	Frequency (%)
WA	166440	99.8%
CA	91	0.1%
VA	38	< 0.1%
MD	32	< 0.1%
TX	24	< 0.1%
NC	14	< 0.1%
IL	13	< 0.1%
CO	12	< 0.1%
FL	10	< 0.1%
HI	9	< 0.1%
Other values (34)	117	0.1%

Length
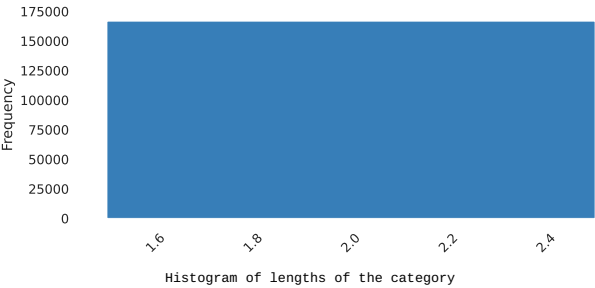
Histogram of lengths of the category

Value	Count	Frequency (%)
wa	166440	99.8%
ca	91	0.1%
va	38	< 0.1%
md	32	< 0.1%
tx	24	< 0.1%
nc	14	< 0.1%
il	13	< 0.1%
co	12	< 0.1%
fl	10	< 0.1%
hi	9	< 0.1%
Other values (34)	117	0.1%

Most occurring characters

Value	Count	Frequency (%)
A	166608	49.9%
W	166441	49.9%
C	138	< 0.1%
V	45	< 0.1%
M	43	< 0.1%
N	42	< 0.1%
D	39	< 0.1%
T	34	< 0.1%
L	33	< 0.1%
I	30	< 0.1%
Other values (15)	147	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	333600	100.0%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
A	166608	49.9%
W	166441	49.9%
C	138	< 0.1%
V	45	< 0.1%
M	43	< 0.1%
N	42	< 0.1%
D	39	< 0.1%
T	34	< 0.1%
L	33	< 0.1%
I	30	< 0.1%
Other values (15)	147	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
Latin	333600	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
A	166608	49.9%
W	166441	49.9%
C	138	< 0.1%
V	45	< 0.1%
M	43	< 0.1%
N	42	< 0.1%
D	39	< 0.1%
T	34	< 0.1%
L	33	< 0.1%
I	30	< 0.1%
Other values (15)	147	< 0.1%

Most frequent character per script

Latin

	Count	Frequency (%)
A	166608	49.9%
W	166441	49.9%
C	138	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	333600	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
A	166608	49.9%
W	166441	49.9%
C	138	< 0.1%
V	45	< 0.1%
M	43	< 0.1%
N	42	< 0.1%
D	39	< 0.1%
T	34	< 0.1%
L	33	< 0.1%
I	30	< 0.1%
Other values (15)	147	< 0.1%

Postal Code

Real number ()

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: State)   SKEWED

Distinct	836
Distinct (%)	0.5%
Missing	5
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	98173.714
Minimum	1730
Maximum	99577
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
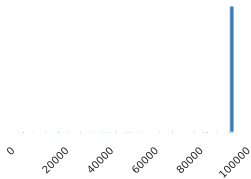
Quantile statistics

Minimum	1730
5-th percentile	98006
Q1	98052
median	98122
Q3	98371
95-th percentile	98942
Maximum	99577
Range	97847
Interquartile range (IQR)	319

Descriptive statistics

Standard deviation	2442.5844
Coefficient of variation (CV)	0.024880228
Kurtosis	954.3603
Mean	98173.714
Median Absolute Deviation (MAD)	101
Skewness	-30.096087
Sum	1.6374885×10^{10}
Variance	5966218.6
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
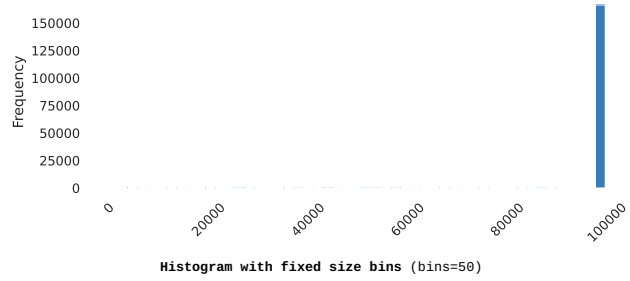
Value	Count	Frequency (%)
98052	4252	2.5%
98012	3115	1.9%
98033	2840	1.7%
98006	2663	1.6%
98004	2652	1.6%
98115	2553	1.5%
98074	2341	1.4%
98072	2312	1.4%
98188	2286	1.4%
98034	2237	1.3%
Other values (826)	139544	83.7%

Value	Count	Frequency (%)
1730	1	< 0.1%
1731	1	< 0.1%
1824	1	< 0.1%
1908	1	< 0.1%
2842	1	< 0.1%
3804	1	< 0.1%
6355	1	< 0.1%
6371	1	< 0.1%
6379	2	< 0.1%
6385	1	< 0.1%

Value	Count	Frequency (%)
99577	1	< 0.1%
99403	57	< 0.1%
99402	9	< 0.1%
99371	1	< 0.1%
99362	329	0.2%
99361	12	< 0.1%
99360	7	< 0.1%
99357	19	< 0.1%
99356	1	< 0.1%
99354	283	0.2%

Model Year
Real number ()

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Electric Range)

Distinct	22
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2020.3418
Minimum	1997
Maximum	2024
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
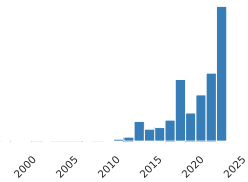
Quantile statistics

Minimum	1997
5-th percentile	2014
Q1	2018
median	2021
Q3	2023
95-th percentile	2023
Maximum	2024
Range	27
Interquartile range (IQR)	5

Descriptive statistics

Standard deviation	3.0014645
Coefficient of variation (CV)	0.0014856222
Kurtosis	0.45101207
Mean	2020.3418
Median Absolute Deviation (MAD)	2
Skewness	-1.086228
Sum	3.3699301×10^8
Variance	9.0087894
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=22)
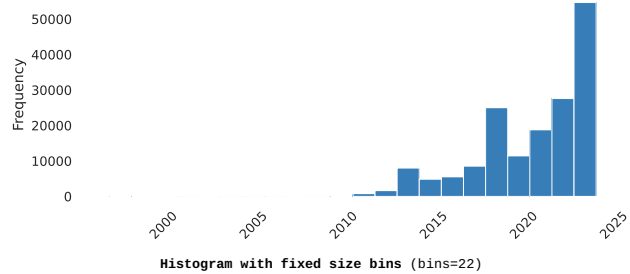
Value	Count	Frequency (%)
2023	51351	30.8%
2022	27592	16.5%
2021	18774	11.3%
2018	14151	8.5%
2020	11425	6.8%
2019	10860	6.5%
2017	8523	5.1%
2016	5518	3.3%
2015	4833	2.9%
2013	4455	2.7%
Other values (12)	9318	5.6%

Value	Count	Frequency (%)
1997	1	< 0.1%
1998	1	< 0.1%
1999	3	< 0.1%
2000	7	< 0.1%
2002	2	< 0.1%
2003	1	< 0.1%
2008	20	< 0.1%
2010	23	< 0.1%
2011	782	0.5%
2012	1630	1.0%

Value	Count	Frequency (%)
2024	3309	2.0%
2023	51351	30.8%
2022	27592	16.5%
2021	18774	11.3%
2020	11425	6.8%
2019	10860	6.5%
2018	14151	8.5%
2017	8523	5.1%
2016	5518	3.3%
2015	4833	2.9%

Make
Categorical

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Vehicle Type)

Distinct	39
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.3 MiB

Length

Max length	20
Median length	14
Mean length	5.557446
Min length	3

Characters and Unicode

Total characters	926982	
Distinct characters	28	
Distinct categories	4 (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	FIAT
2nd row	TESLA
3rd row	BMW
4th row	RIVIAN
5th row	TESLA

Common Values

Value	Count	Frequency (%)
TESLA	74834	44.9%
NISSAN	13848	8.3%
CHEVROLET	13072	7.8%
FORD	8577	5.1%
BMW	7196	4.3%
KIA	6995	4.2%
TOYOTA	5812	3.5%
VOLKSWAGEN	4717	2.8%
JEEP	4100	2.5%
HYUNDAI	4057	2.4%
Other values (29)	23592	14.1%

Length

Histogram of lengths of the category

Value	Count	Frequency (%)
TESLA	74834	44.9%
NISSAN	13848	8.3%
CHEVROLET	13072	7.8%
FORD	8577	5.1%
BMW	7196	4.3%
KIA	6995	4.2%
TOYOTA	5812	3.5%
VOLKSWAGEN	4717	2.8%
JEEP	4100	2.5%
HYUNDAI	4057	2.4%
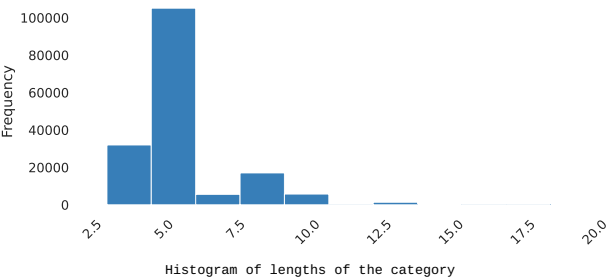
Value	Count	Frequency (%)
tesla	74834	44.8%
nissan	13848	8.3%
chevrolet	13072	7.8%
ford	8577	5.1%
bmw	7196	4.3%
kia	6995	4.2%
toyota	5812	3.5%
volkswagen	4717	2.8%
jeep	4100	2.5%
hyundai	4057	2.4%
Other values (34)	23688	14.2%
tesla	74834	44.8%
nissan	13848	8.3%
chevrolet	13072	7.8%
ford	8577	5.1%
bmw	7196	4.3%
kia	6995	4.2%
volkswagen	4717	2.8%

Most occurring characters

Value	Count	Frequency (%)
E	124884	13.5%
A	122649	13.2%
S	117030	12.6%
T	102411	11.0%
L	101940	11.0%
O	49081	5.3%
N	43779	4.7%
I	41734	4.5%
R	35712	3.9%
V	29320	3.2%
Other values (18)	158442	17.1%

Most occurring characters

Value	Count	Frequency (%)
	124884	13.5%
	122649	13.2%
	117030	12.6%
	102411	11.0%
	101940	11.0%
	49081	5.3%
	43779	4.7%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	925539	99.8%
Dash Punctuation	1342	0.1%
Space Separator	96	< 0.1%
Other Punctuation	5	< 0.1%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
E	124884	13.5%
A	122649	13.3%
S	117030	12.6%
T	102411	11.1%
L	101940	11.0%
O	49081	5.3%
N	43779	4.7%
I	41734	4.5%
R	35712	3.9%
V	29320	3.2%
Other values (15)	156999	17.0%

Dash Punctuation

Value	Count	Frequency (%)
-	1342	100.0%

Space Separator

Value	Count	Frequency (%)
	96	100.0%

Other Punctuation

Value	Count	Frequency (%)
!	5	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	925539	99.8%
Common	1443	0.2%

Most frequent character per script

Latin

Value	Count	Frequency (%)
E	124884	13.5%
A	122649	13.3%
S	117030	12.6%
T	102411	11.1%
L	101940	11.0%
O	49081	5.3%
N	43779	4.7%
I	41734	4.5%
R	35712	3.9%
V	29320	3.2%
Other values (15)	156999	17.0%

Common

Value	Count	Frequency (%)
-	1342	93.0%
	96	6.7%
!	5	0.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	926982	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
E	124884	13.5%
A	122649	13.2%
S	117030	12.6%
T	102411	11.0%
L	101940	11.0%
O	49081	5.3%
N	43779	4.7%
I	41734	4.5%
R	35712	3.9%
V	29320	3.2%
Other values (18)	158442	17.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	926982	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)

Model
Text

Distinct	138
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.3 MiB
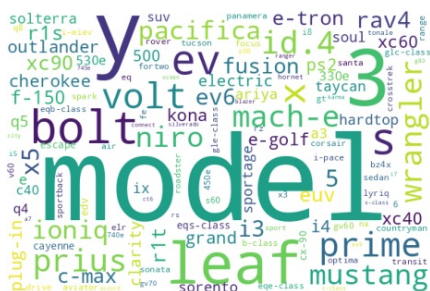
Length

Max length	24
Median length	7
Mean length	6.3818825
Min length	2

Characters and Unicode

Total characters	1064498	
Distinct characters	38	
Distinct categories	5 (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	4	?
Unique (%)	< 0.1%	

Sample

1st row	500
2nd row	MODEL X
3rd row	330E
4th row	R1S
5th row	MODEL 3

Value	Count	Frequency (%)
model	74788	28.2%
y	32822	12.4%
3	28926	10.9%
leaf	13274	5.0%
bolt	7983	3.0%
s	7611	2.9%
ev	6648	2.5%
x	5429	2.0%
volt	4825	1.8%
prime	4505	1.7%
Other values (135)	78207	29.5%

Most occurring characters

Value	Count	Frequency (%)
E	125149	11.8%
L	111748	10.5%
O	106152	10.0%
	98218	9.2%
M	87547	8.2%
D	81731	7.7%
A	45626	4.3%
I	36297	3.4%
Y	35554	3.3%
R	35303	3.3%
Other values (28)	301173	28.3%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	881972	82.9%
Space Separator	98218	9.2%
Decimal Number	69361	6.5%
Dash Punctuation	11300	1.1%
Other Punctuation	3647	0.3%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
E	125149	14.2%
L	111748	12.7%
O	106152	12.0%
M	87547	9.9%
D	81731	9.3%
A	45626	5.2%
I	36297	4.1%
Y	35554	4.0%
R	35303	4.0%
S	27137	3.1%
Other values (15)	189728	21.5%

Decimal Number

Value	Count	Frequency (%)
3	32990	47.6%
4	8794	12.7%
0	8390	12.1%
5	8009	11.5%
1	4467	6.4%
6	3647	5.3%
9	1789	2.6%
2	853	1.2%
8	336	0.5%
7	86	0.1%

Space Separator

Value	Count	Frequency (%)
	98218	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	11300	100.0%

Other Punctuation

Value	Count	Frequency (%)
.	3647	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	881972	82.9%
Common	182526	17.1%

Most frequent character per script

Latin

Value	Count	Frequency (%)
E	125149	14.2%
L	111748	12.7%
O	106152	12.0%
M	87547	9.9%
D	81731	9.3%
A	45626	5.2%
I	36297	4.1%
Y	35554	4.0%
R	35303	4.0%
S	27137	3.1%
Other values (15)	189728	21.5%

Common

Value	Count	Frequency (%)
	98218	53.8%
3	32990	18.1%
-	11300	6.2%
4	8794	4.8%
0	8390	4.6%
5	8009	4.4%
1	4467	2.4%
6	3647	2.0%
.	3647	2.0%
9	1789	1.0%
Other values (3)	1275	0.7%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1064498	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
E	125149	11.8%
L	111748	10.5%
O	106152	10.0%
	98218	9.2%
M	87547	8.2%
D	81731	7.7%
A	45626	4.3%
I	36297	3.4%
Y	35554	3.3%
R	35303	3.3%
Other values (28)	301173	28.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1064498	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
E	125149	11.8%
L	111748	10.5%
O	106152	10.0%

Electric Vehicle Type
Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.3 MiB

Length

Max length	38
Median length	30
Mean length	31.750935
Min length	30

Characters and Unicode

Total characters	5296056	
Distinct characters	23	
Distinct categories	6 (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

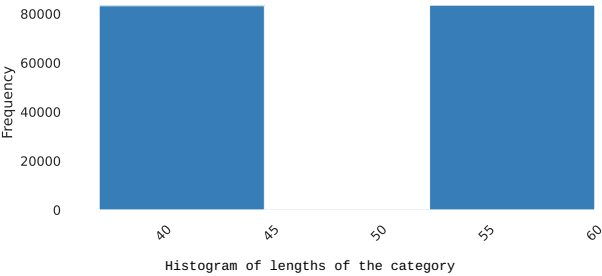|		
Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Battery Electric Vehicle (BEV)
2nd row	Battery Electric Vehicle (BEV)
3rd row	Plug-in Hybrid Electric Vehicle (PHEV)
4th row	Battery Electric Vehicle (BEV)
5th row	Battery Electric Vehicle (BEV)

Common Values

Value	Count	Frequency (%)
Battery Electric Vehicle (BEV)	130293	78.1%
Plug-in Hybrid Electric Vehicle (PHEV)	36507	21.9%

Length

Histogram of lengths of the category

Common Values (Plot)

- Battery Electric Vehicle (BEV)
- Plug-in Hybrid Electric Vehicle (PHEV)

Value	Count	Frequency (%)
Battery Electric Vehicle (BEV)	130293	78.1%
Plug-in Hybrid Electric Vehicle (PHEV)	36507	21.9%
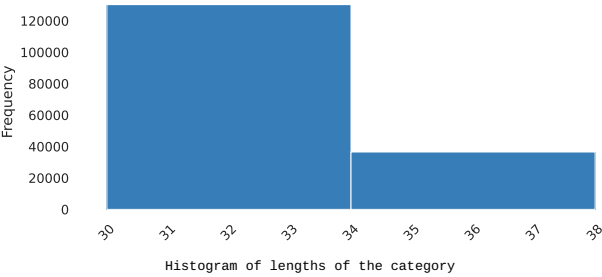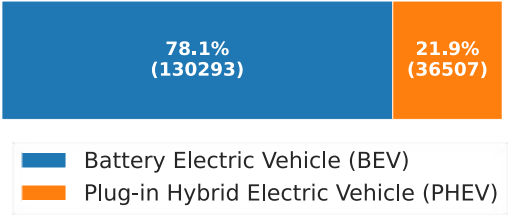
Value	Count	Frequency (%)
electric	166800	23.7%
vehicle	166800	23.7%
battery	130293	18.5%
bev	130293	18.5%
plug-in	36507	5.2%
hybrid	36507	5.2%
phev	36507	5.2%
electric	166800	23.7%
vehicle	166800	23.7%
battery	130293	18.5%
bev	130293	18.5%
plug-in	36507	5.2%
hybrid	36507	5.2%
phev	36507	5.2%

Most occurring characters

Value	Count	Frequency (%)
e	630693	11.9%
	536907	10.1%
c	500400	9.4%
t	427386	8.1%
i	406614	7.7%
l	370107	7.0%
V	333600	6.3%
r	333600	6.3%
E	333600	6.3%
B	260586	4.9%
Other values (13)	1162563	22.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	3315228	62.6%
Uppercase Letter	1073814	20.3%
Space Separator	536907	10.1%
Open Punctuation	166800	3.1%
Close Punctuation	166800	3.1%
Dash Punctuation	36507	0.7%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	630693	19.0%
c	500400	15.1%
t	427386	12.9%
i	406614	12.3%
l	370107	11.2%
r	333600	10.1%
y	166800	5.0%
h	166800	5.0%
a	130293	3.9%
u	36507	1.1%
Other values (4)	146028	4.4%

Uppercase Letter

Value	Count	Frequency (%)
V	333600	31.1%
E	333600	31.1%
B	260586	24.3%
P	73014	6.8%
H	73014	6.8%

Space Separator

Value	Count	Frequency (%)
	536907	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	166800	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	166800	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	36507	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	4389042	82.9%
Common	907014	17.1%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	630693	14.4%
c	500400	11.4%
t	427386	9.7%
i	406614	9.3%
l	370107	8.4%
V	333600	7.6%
r	333600	7.6%
E	333600	7.6%
B	260586	5.9%
y	166800	3.8%
Other values (9)	625656	14.3%

Common

Value	Count	Frequency (%)
	536907	59.2%
(	166800	18.4%
)	166800	18.4%
-	36507	4.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	5296056	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	630693	11.9%
	536907	10.1%
c	500400	9.4%
t	427386	8.1%
i	406614	7.7%
l	370107	7.0%
V	333600	6.3%
r	333600	6.3%
E	333600	6.3%
B	260586	4.9%
Other values (13)	1162563	22.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	5296056	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)

Clean Alternative Fuel Vehicle (CAFV) Eligibility
Categorical

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: Electric Range, Electric Vehicle Type, Make)

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.3 MiB

Length

Max length	60
Median length	60
Mean length	49.287104
Min length	37

Characters and Unicode

Total characters	8221089	
Distinct characters	26	
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	Clean Alternative Fuel Vehicle Eligible
2nd row	Clean Alternative Fuel Vehicle Eligible
3rd row	Not eligible due to low battery range
4th row	Eligibility unknown as battery range has not been researched
5th row	Clean Alternative Fuel Vehicle Eligible

Clean Alternative Fuel Vehicle (CAFV) Eligibility
Categorical

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: Electric Range, Electric Vehicle Type, Make)

Common Values

Value	Count	Frequency (%)
Eligibility unknown as battery range has not been researched	83517	50.1%
Clean Alternative Fuel Vehicle Eligible	64299	38.5%
Not eligible due to low battery range	18984	11.4%

Length

Histogram of lengths of the category

Common Values (Plot)

- Eligibility unknown as battery range has not been researched
- Clean Alternative Fuel Vehicle Eligible
- Not eligible due to low battery range

Common Values

Value	Count	Frequency (%)
Eligibility unknown as battery range has not been researched	83517	50.1%
Clean Alternative Fuel Vehicle Eligible	64299	38.5%
Not eligible due to low battery range	18984	11.4%

Value	Count	Frequency (%)
not	102501	8.5%
battery	102501	8.5%
range	102501	8.5%
eligibility	83517	6.9%
unknown	83517	6.9%
been	83517	6.9%
researched	83517	6.9%
has	83517	6.9%
as	83517	6.9%
eligible	83283	6.9%
Other values (7)	314148	26.0%

Most occurring characters

Value	Count	Frequency (%)
e	1129632	13.7%
	1039236	12.6%
n	648684	7.9%
i	629232	7.7%
l	609780	7.4%
a	584151	7.1%
t	538602	6.6%
r	436335	5.3%
b	352818	4.3%
g	269301	3.3%
Other values (16)	1983318	24.1%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	6757857	82.2%
Space Separator	1039236	12.6%
Uppercase Letter	423996	5.2%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	1129632	16.7%
n	648684	9.6%
i	629232	9.3%
l	609780	9.0%
a	584151	8.6%
t	538602	8.0%
r	436335	6.5%
b	352818	5.2%
g	269301	4.0%
s	250551	3.7%
Other values (9)	1308771	19.4%

Uppercase Letter

Value	Count	Frequency (%)
E	147816	34.9%
C	64299	15.2%
A	64299	15.2%
F	64299	15.2%
V	64299	15.2%
N	18984	4.5%

Space Separator

Value	Count	Frequency (%)
	1039236	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	7181853	87.4%
Common	1039236	12.6%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	1129632	15.7%
n	648684	9.0%
i	629232	8.8%
l	609780	8.5%
a	584151	8.1%
t	538602	7.5%
r	436335	6.1%
b	352818	4.9%
g	269301	3.7%
s	250551	3.5%
Other values (15)	1732767	24.1%

Common

Value	Count	Frequency (%)
	1039236	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	8221089	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	1129632	13.7%
	1039236	12.6%
n	648684	7.9%
i	629232	7.7%
l	609780	7.4%
a	584151	7.1%
t	538602	6.6%
r	436335	5.3%
b	352818	4.3%
g	269301	3.3%
Other values (16)	1983318	24.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	8221089	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	1129632	13.7%
	1039236	12.6%

Electric Range
Real number ()

HIGH CORRELATION (This variable has a high overall correlation with 3 fields: Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Vehicle Type, Model Year)   ZEROS

Distinct	102
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	61.508993
Minimum	0
Maximum	337
Zeros	83517
Zeros (%)	50.1%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
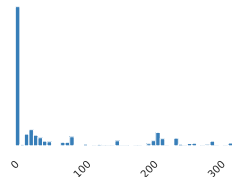
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	84
95-th percentile	259
Maximum	337
Range	337
Interquartile range (IQR)	84

Descriptive statistics

Standard deviation	93.271747
Coefficient of variation (CV)	1.516392
Kurtosis	0.38709605
Mean	61.508993
Median Absolute Deviation (MAD)	0
Skewness	1.3743532
Sum	10259700
Variance	8699.6188
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
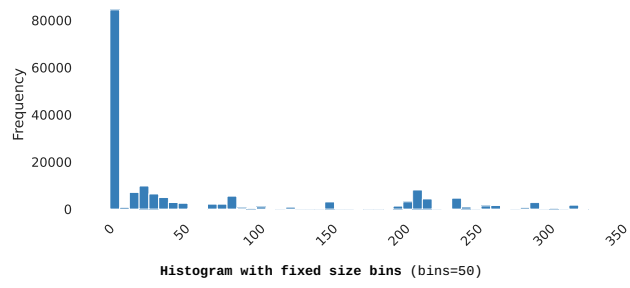
Value	Count	Frequency (%)
0	83517	50.1%
215	6272	3.8%
220	4103	2.5%
25	3918	2.3%
84	3918	2.3%
238	3790	2.3%
21	3298	2.0%
32	3182	1.9%
208	2472	1.5%
53	2466	1.5%
Other values (92)	49864	29.9%

Value	Count	Frequency (%)
0	83517	50.1%
6	935	0.6%
8	35	< 0.1%
9	21	< 0.1%
10	162	0.1%
11	3	< 0.1%
12	164	0.1%
13	358	0.2%
14	1109	0.7%
15	89	0.1%

Value	Count	Frequency (%)
337	74	< 0.1%
330	318	0.2%
322	1671	1.0%
308	485	0.3%
293	443	0.3%
291	2335	1.4%
289	646	0.4%
270	275	0.2%
266	1400	0.8%
265	124	0.1%

Base MSRP
Real number ()

ZEROS

Distinct	31
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1152.7232
Minimum	0
Maximum	845000
Zeros	163437
Zeros (%)	98.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
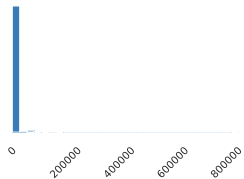
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	845000
Range	845000
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	8661.0811
Coefficient of variation (CV)	7.5135829
Kurtosis	602.62985
Mean	1152.7232
Median Absolute Deviation (MAD)	0
Skewness	12.846622
Sum	1.9227422×10^8
Variance	75014326
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=31)
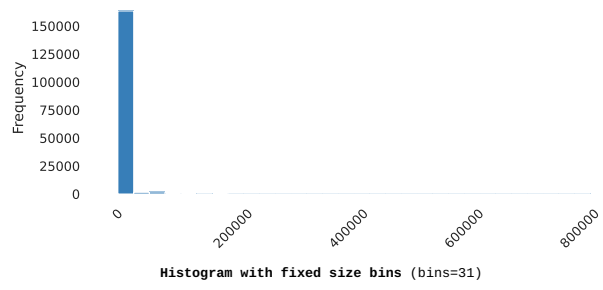
Value	Count	Frequency (%)
0	163437	98.0%
69900	1390	0.8%
31950	381	0.2%
52900	220	0.1%
32250	141	0.1%
54950	130	0.1%
59900	124	0.1%
39995	111	0.1%
36900	100	0.1%
44100	95	0.1%
Other values (21)	671	0.4%

Value	Count	Frequency (%)
0	163437	98.0%
31950	381	0.2%
32250	141	0.1%
32995	3	< 0.1%
33950	72	< 0.1%
34995	67	< 0.1%
36800	54	< 0.1%
36900	100	0.1%
39995	111	0.1%
43700	11	< 0.1%

Value	Count	Frequency (%)
845000	1	< 0.1%
184400	10	< 0.1%
110950	20	< 0.1%
109000	6	< 0.1%
102000	15	< 0.1%
98950	20	< 0.1%
91250	5	< 0.1%
90700	20	< 0.1%
89100	7	< 0.1%
81100	22	< 0.1%

Legislative District

Real number ()

Distinct	49
Distinct (%)	< 0.1%
Missing	360
Missing (%)	0.2%
Infinite	0
Infinite (%)	0.0%
Mean	29.178941
Minimum	1
Maximum	49
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
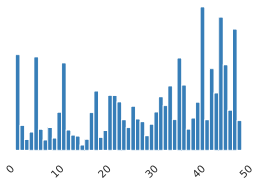
Quantile statistics

Minimum	1
5-th percentile	2
Q1	18
median	33
Q3	42
95-th percentile	48
Maximum	49
Range	48
Interquartile range (IQR)	24

Descriptive statistics

Standard deviation	14.853534
Coefficient of variation (CV)	0.50904978
Kurtosis	-1.0877132
Mean	29.178941
Median Absolute Deviation (MAD)	11
Skewness	-0.46529352
Sum	4856543
Variance	220.62746
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=49)
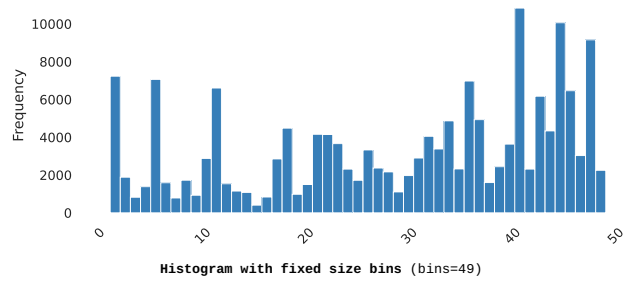
Value	Count	Frequency (%)
41	10837	6.5%
45	10062	6.0%
48	9158	5.5%
1	7231	4.3%
5	7062	4.2%
36	6977	4.2%
11	6607	4.0%
46	6464	3.9%
43	6172	3.7%
37	4933	3.0%
Other values (39)	90937	54.5%

Value	Count	Frequency (%)
1	7231	4.3%
2	1885	1.1%
3	824	0.5%
4	1385	0.8%
5	7062	4.2%
6	1593	1.0%
7	787	0.5%
8	1726	1.0%
9	930	0.6%
10	2879	1.7%
1	7231	4.3%
2	1885	1.1%
3	824	0.5%
4	1385	0.8%
5	7062	4.2%
6	1593	1.0%
7	787	0.5%
8	1726	1.0%
9	930	0.6%

Value	Count	Frequency (%)
49	2250	1.3%
48	9158	5.5%
47	3032	1.8%
46	6464	3.9%
45	10062	6.0%
44	4337	2.6%
43	6172	3.7%
42	2315	1.4%
41	10837	6.5%
40	3637	2.2%

DOL Vehicle ID
Real number ()

UNIQUE

Distinct	166800
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2.1724199×10^8
Minimum	4385
Maximum	4.7925477×10^8
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
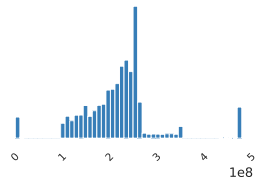
Quantile statistics

Minimum	4385
5-th percentile	1.0927191×10^8
Q1	1.7907406×10^8
median	2.2440453×10^8
Q3	2.5134213×10^8
95-th percentile	3.4421029×10^8
Maximum	4.7925477×10^8
Range	4.7925039×10^8
Interquartile range (IQR)	72268068

Descriptive statistics

Standard deviation	77274578
Coefficient of variation (CV)	0.35570737
Kurtosis	3.4968468
Mean	2.1724199×10^8
Median Absolute Deviation (MAD)	30844772
Skewness	0.7260005
Sum	3.6235965×10^{13}
Variance	5.9713604×10^{15}
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
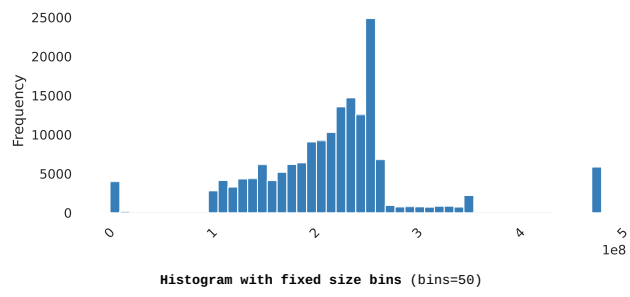
Value	Count	Frequency (%)
1593721	1	< 0.1%
166073470	1	< 0.1%
148905543	1	< 0.1%
249787583	1	< 0.1%
220480571	1	< 0.1%
249691865	1	< 0.1%
131525467	1	< 0.1%
233740279	1	< 0.1%
144426796	1	< 0.1%
118183838	1	< 0.1%
Other values (166790)	166790	> 99.9%

Value	Count	Frequency (%)
4385	1	< 0.1%
4777	1	< 0.1%
10286	1	< 0.1%
10734	1	< 0.1%
12050	1	< 0.1%
23145	1	< 0.1%
24629	1	< 0.1%
27702	1	< 0.1%
35325	1	< 0.1%
46112	1	< 0.1%

Value	Count	Frequency (%)
479254772	1	< 0.1%
479114996	1	< 0.1%
478935460	1	< 0.1%
478934571	1	< 0.1%
478926346	1	< 0.1%
478925947	1	< 0.1%
478925163	1	< 0.1%
478924358	1	< 0.1%
478916028	1	< 0.1%
478910428	1	< 0.1%

Vehicle Location
Text

Distinct	835
Distinct (%)	0.5%
Missing	10
Missing (%)	< 0.1%
Memory size	1.3 MiB
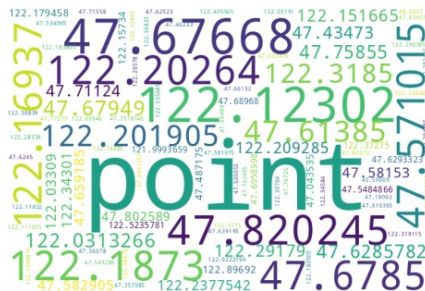
Length

Max length	33
Median length	32
Mean length	28.439799
Min length	24

Characters and Unicode

Total characters	4743474	
Distinct characters	20	
Distinct categories	7 (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	286	?
Unique (%)	0.2%	

Sample

1st row	POINT (-120.524012 46.5973939)
2nd row	POINT (-122.817545 46.98876)
3rd row	POINT (-122.1298876 47.4451257)
4th row	POINT (-122.1873 47.820245)
5th row	POINT (-122.2012521 47.3931814)

Value	Count	Frequency (%)
point	166790	33.3%
47.67668	4252	0.8%
122.12302	4252	0.8%
122.1873	3115	0.6%
47.820245	3115	0.6%
122.20264	2840	0.6%
47.6785	2840	0.6%
122.16937	2663	0.5%
47.571015	2663	0.5%
122.201905	2652	0.5%
Other values (1660)	305188	61.0%

Most occurring characters

Value	Count	Frequency (%)
2	492336	10.4%
1	377714	8.0%
4	335804	7.1%
.	333580	7.0%
	333580	7.0%
7	314789	6.6%
5	301633	6.4%
3	203552	4.3%
6	202005	4.3%
8	197371	4.2%
Other values (10)	1651110	34.8%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	2741994	57.8%
Uppercase Letter	833950	17.6%
Other Punctuation	333580	7.0%
Space Separator	333580	7.0%
Dash Punctuation	166790	3.5%
Open Punctuation	166790	3.5%
Close Punctuation	166790	3.5%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
2	492336	18.0%
1	377714	13.8%
4	335804	12.2%
7	314789	11.5%
5	301633	11.0%
3	203552	7.4%
6	202005	7.4%
8	197371	7.2%
9	172064	6.3%
0	144726	5.3%

Uppercase Letter

Value	Count	Frequency (%)
O	166790	20.0%
T	166790	20.0%
N	166790	20.0%
I	166790	20.0%
P	166790	20.0%

Other Punctuation

Value	Count	Frequency (%)
.	333580	100.0%

Space Separator

Value	Count	Frequency (%)
	333580	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	166790	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	166790	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	166790	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	3909524	82.4%
Latin	833950	17.6%

Most frequent character per script

Common

Value	Count	Frequency (%)
2	492336	12.6%
1	377714	9.7%
4	335804	8.6%
.	333580	8.5%
	333580	8.5%
7	314789	8.1%
5	301633	7.7%
3	203552	5.2%
6	202005	5.2%
8	197371	5.0%
Other values (5)	817160	20.9%

Latin

Value	Count	Frequency (%)
O	166790	20.0%
T	166790	20.0%
N	166790	20.0%
I	166790	20.0%
P	166790	20.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	4743474	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
2	492336	10.4%
1	377714	8.0%
4	335804	7.1%
.	333580	7.0%
	333580	7.0%
7	314789	6.6%
5	301633	6.4%
3	203552	4.3%
6	202005	4.3%
8	197371	4.2%
Other values (10)	1651110	34.8%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	4743474	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	333580	7.0%

Electric Utility
Text

Distinct	76
Distinct (%)	< 0.1%
Missing	5
Missing (%)	< 0.1%
Memory size	1.3 MiB

Length

Max length	112
Median length	110
Mean length	44.349777
Min length	10

Characters and Unicode

Total characters	7397321	
Distinct characters	36	
Distinct categories	8 (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	4	?
Unique (%)	< 0.1%	

Sample

1st row	PACIFICORP
2nd row	PUGET SOUND ENERGY INC
3rd row	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
4th row	PUGET SOUND ENERGY INC
5th row	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)

Value	Count	Frequency (%)
of	156743	12.6%
	146822	11.8%
wa	102355	8.2%
tacoma	101048	8.1%
sound	99739	8.0%
energy	99739	8.0%
puget	98811	7.9%
inc||city	61337	4.9%
power	36727	2.9%
inc	33917	2.7%
Other values (114)	310920	24.9%

Most occurring characters

Value	Count	Frequency (%)
	1081363	14.6%
O	543795	7.4%
N	527092	7.1%
T	513721	6.9%
A	499156	6.7%
E	483730	6.5%
I	404993	5.5%
C	402838	5.4%
Y	270111	3.7%
U	259031	3.5%
Other values (26)	2411491	32.6%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	5604845	75.8%
Space Separator	1081363	14.6%
Math Symbol	254096	3.4%
Close Punctuation	142278	1.9%
Dash Punctuation	142278	1.9%
Open Punctuation	142278	1.9%
Decimal Number	24893	0.3%
Other Punctuation	5290	0.1%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
O	543795	9.7%
N	527092	9.4%
T	513721	9.2%
A	499156	8.9%
E	483730	8.6%
I	404993	7.2%
C	402838	7.2%
Y	270111	4.8%
U	259031	4.6%
G	215415	3.8%
Other values (14)	1484963	26.5%

Decimal Number

Value	Count	Frequency (%)
1	23111	92.8%
2	729	2.9%
3	674	2.7%
5	379	1.5%

Other Punctuation

Value	Count	Frequency (%)
&	4544	85.9%
#	379	7.2%
,	367	6.9%

Space Separator

Value	Count	Frequency (%)
	1081363	100.0%

Math Symbol

Value	Count	Frequency (%)
|	254096	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	142278	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	142278	100.0%

Open Punctuation

Value	Count	Frequency (%)
(	142278	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	5604845	75.8%
Common	1792476	24.2%

Most frequent character per script

Latin

Value	Count	Frequency (%)
O	543795	9.7%
N	527092	9.4%
T	513721	9.2%
A	499156	8.9%
E	483730	8.6%
I	404993	7.2%
C	402838	7.2%
Y	270111	4.8%
U	259031	4.6%
G	215415	3.8%
Other values (14)	1484963	26.5%

Common

Value	Count	Frequency (%)
	1081363	60.3%
|	254096	14.2%
)	142278	7.9%
-	142278	7.9%
(	142278	7.9%
1	23111	1.3%
&	4544	0.3%
2	729	< 0.1%
3	674	< 0.1%
#	379	< 0.1%
Other values (2)	746	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	7397321	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	1081363	14.6%
O	543795	7.4%
N	527092	7.1%
T	513721	6.9%
A	499156	6.7%
E	483730	6.5%
I	404993	5.5%
C	402838	5.4%
Y	270111	3.7%
U	259031	3.5%
Other values (26)	2411491	32.6%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	7397321	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	1081363	14.6%
	543795	
	527092	

2020 Census Tract

Real number ()

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: State)   SKEWED

Distinct	2088
Distinct (%)	1.3%
Missing	5
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	5.2977092×10^{10}
Minimum	1.0010201×10^{9}
Maximum	5.6033×10^{10}
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.3 MiB
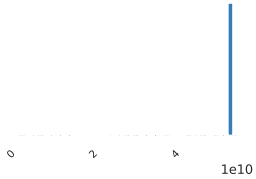
Quantile statistics

Minimum	1.0010201×10^{9}
5-th percentile	5.3011041×10^{10}
Q1	5.303301×10^{10}
median	5.303303×10^{10}
Q3	5.3053073×10^{10}
95-th percentile	5.3067012×10^{10}
Maximum	5.6033×10^{10}
Range	5.503198×10^{10}
Interquartile range (IQR)	20063300

Descriptive statistics

Standard deviation	1.5697544×10^{9}
Coefficient of variation (CV)	0.029630815
Kurtosis	751.92796
Mean	5.2977092×10^{10}
Median Absolute Deviation (MAD)	27702
Skewness	-26.966776
Sum	8.836314×10^{15}
Variance	2.4641288×10^{18}
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
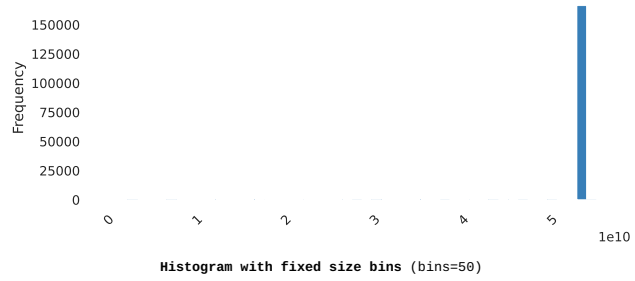
Value	Count	Frequency (%)
$5.30330282 \times 10^{10}$	1914	1.1%
$5.30330285 \times 10^{10}$	1001	0.6%
$5.303303232 \times 10^{10}$	763	0.5%
$5.30330262 \times 10^{10}$	713	0.4%
$5.30330093 \times 10^{10}$	647	0.4%
$5.30670112 \times 10^{10}$	624	0.4%
$5.303303232 \times 10^{10}$	564	0.3%
$5.303303222 \times 10^{10}$	545	0.3%
$5.303302501 \times 10^{10}$	532	0.3%
$5.306105211 \times 10^{10}$	529	0.3%
Other values (2078)	158963	95.3%

Value	Count	Frequency (%)
1001020100	2	< 0.1%
1081041901	1	< 0.1%
1097006803	1	< 0.1%
1117030352	1	< 0.1%
2020000206	1	< 0.1%
4013115900	1	< 0.1%
4013216901	1	< 0.1%
4013610301	1	< 0.1%
4013610302	1	< 0.1%
4013610500	1	< 0.1%

Value	Count	Frequency (%)
$5.60330001 \times 10^{10}$	1	< 0.1%
$5.307794001 \times 10^{10}$	5	< 0.1%
$5.307794001 \times 10^{10}$	3	< 0.1%
5.307794×10^{10}	2	< 0.1%
5.307794×10^{10}	7	< 0.1%
5.307794×10^{10}	7	< 0.1%
5.307794×10^{10}	3	< 0.1%
$5.30770034 \times 10^{10}$	33	< 0.1%
$5.30770032 \times 10^{10}$	40	< 0.1%
$5.30770031 \times 10^{10}$	20	< 0.1%

Interactions
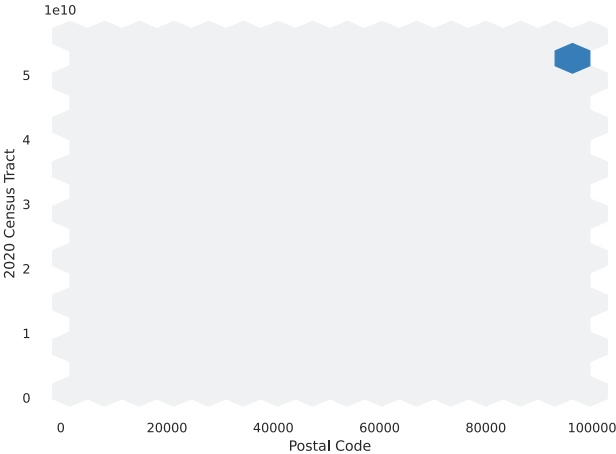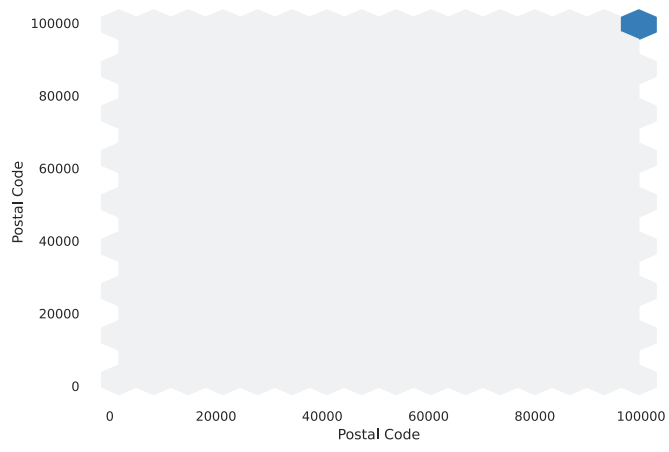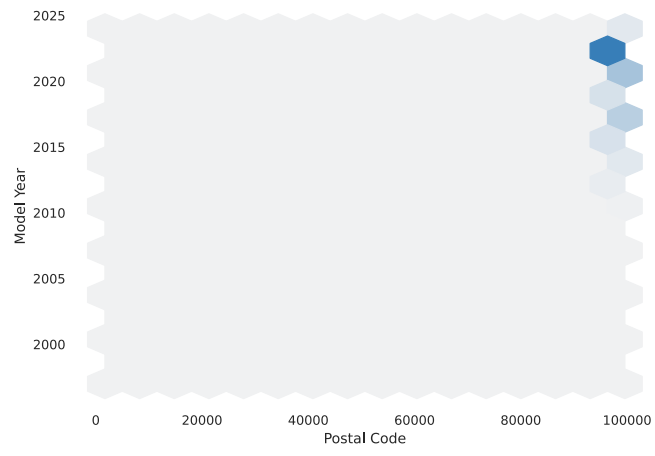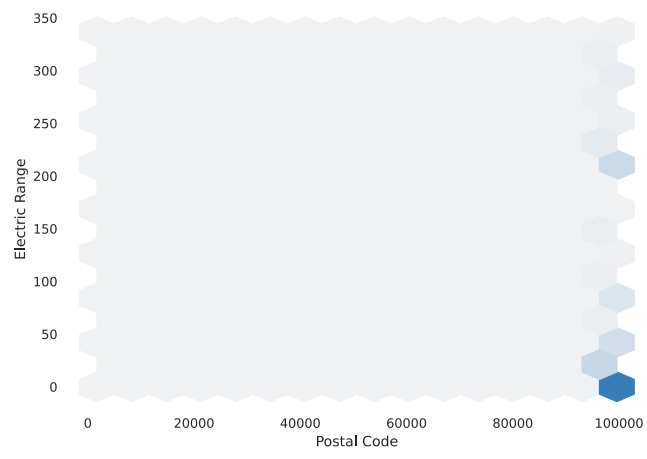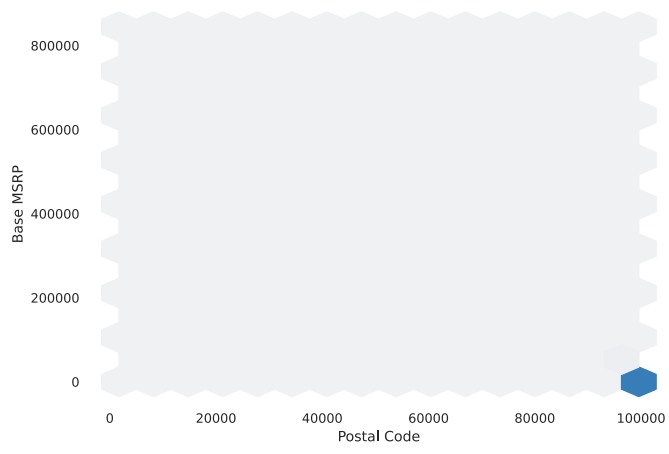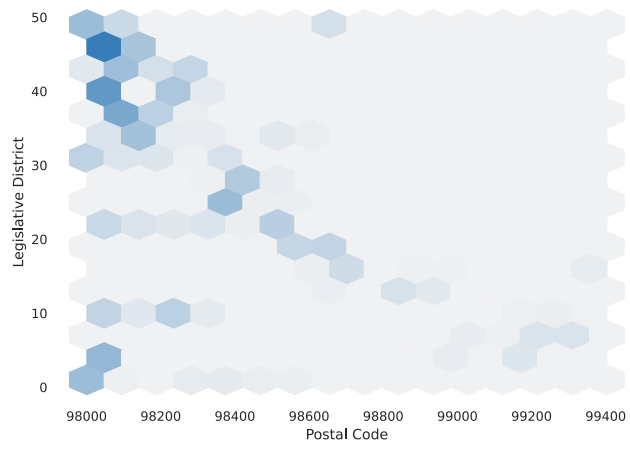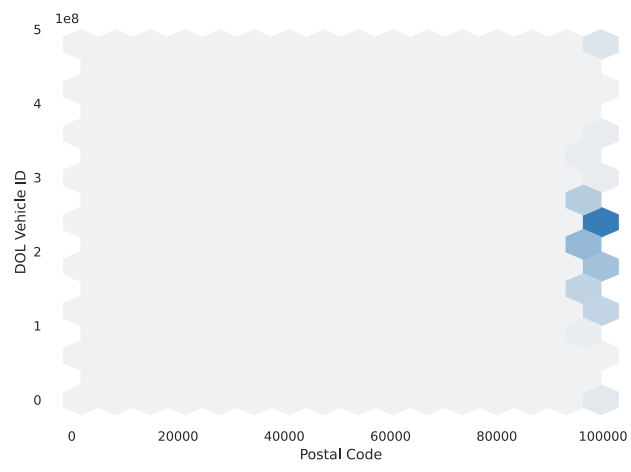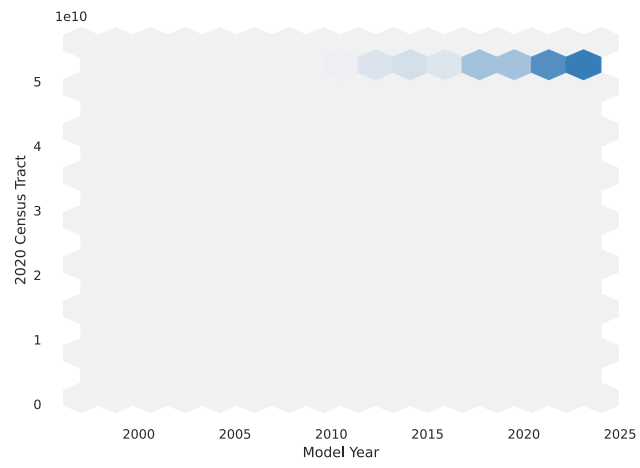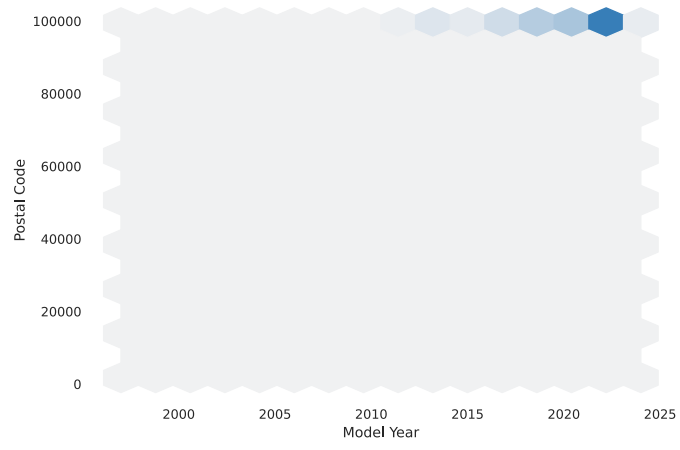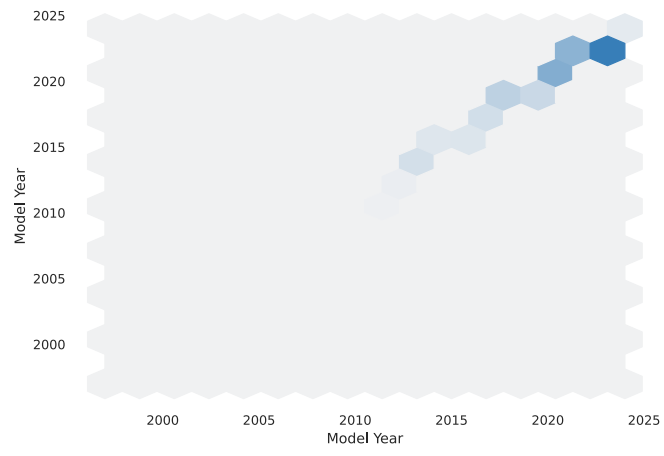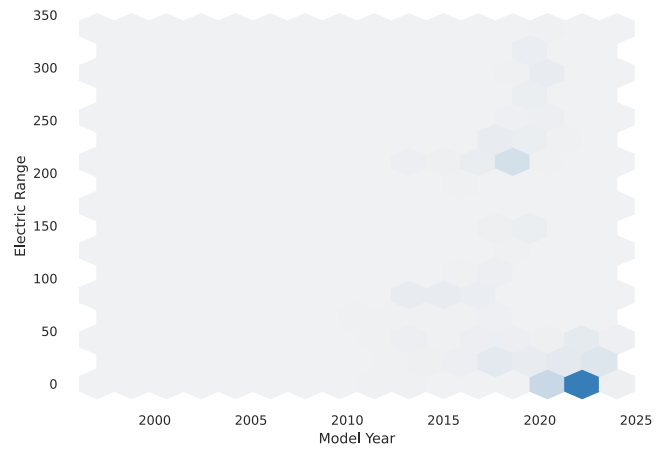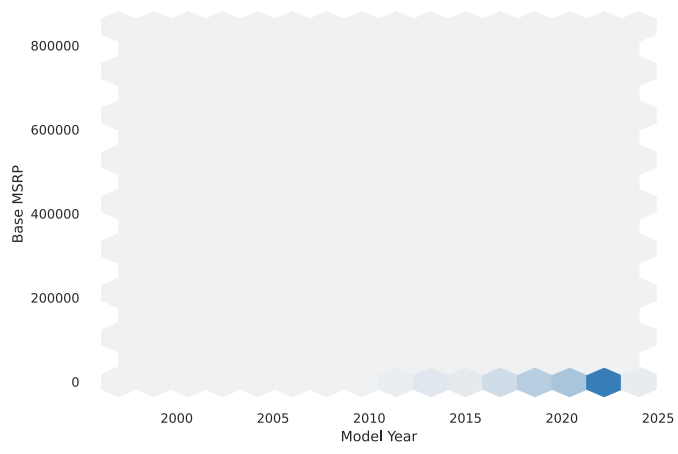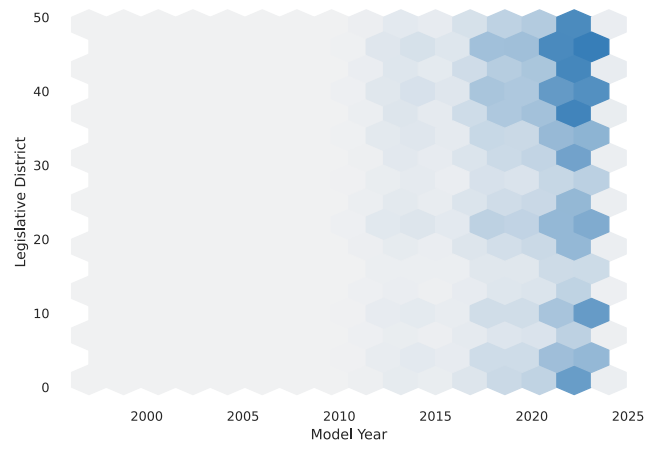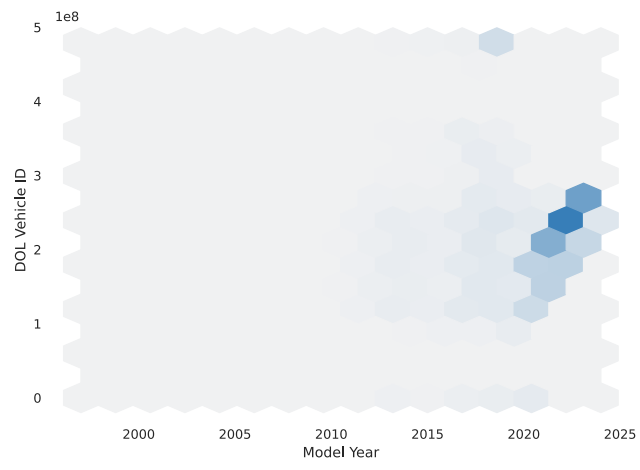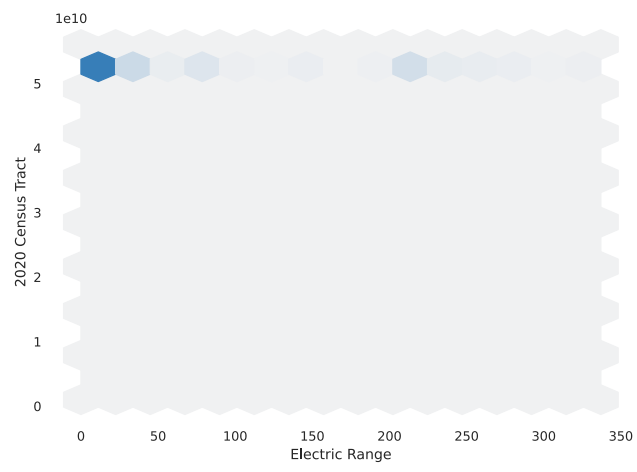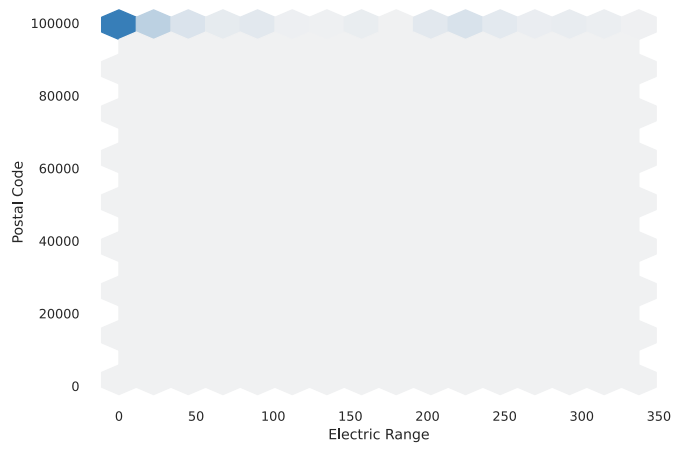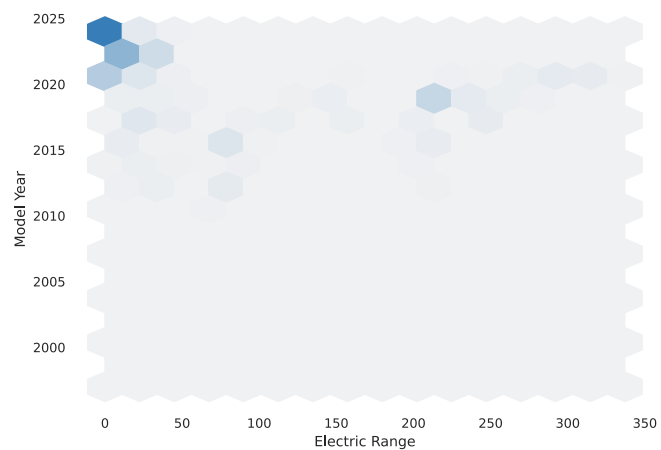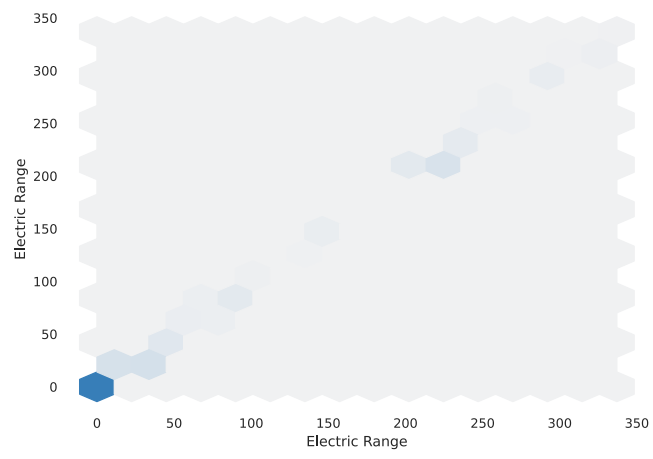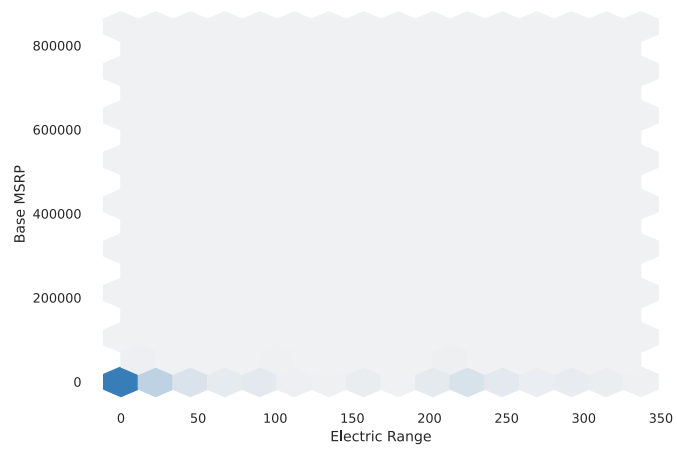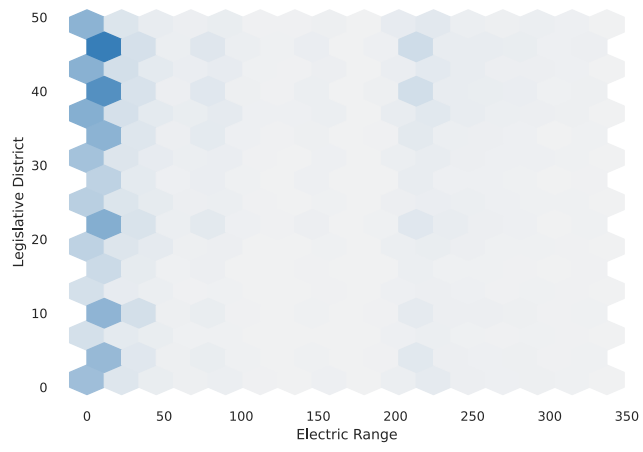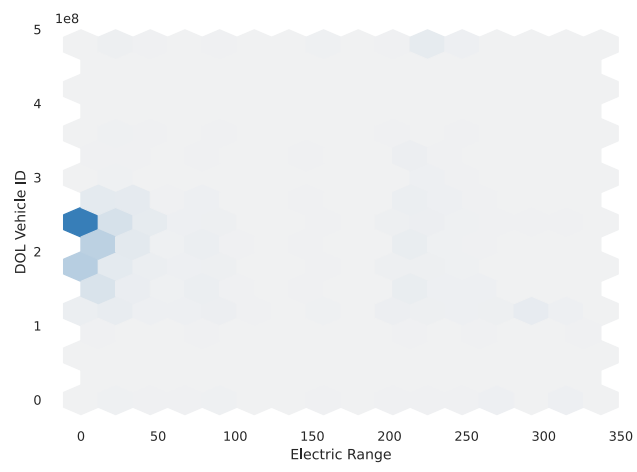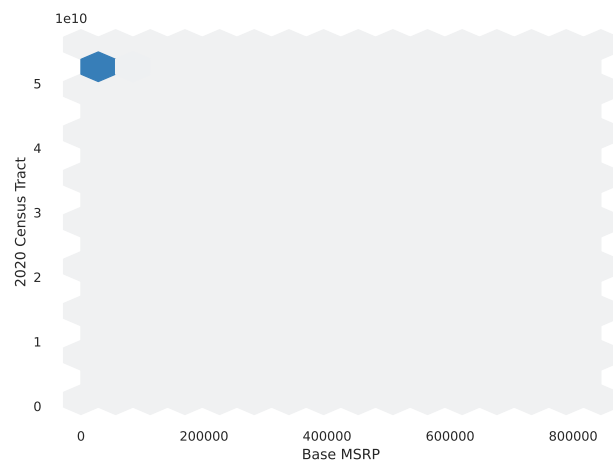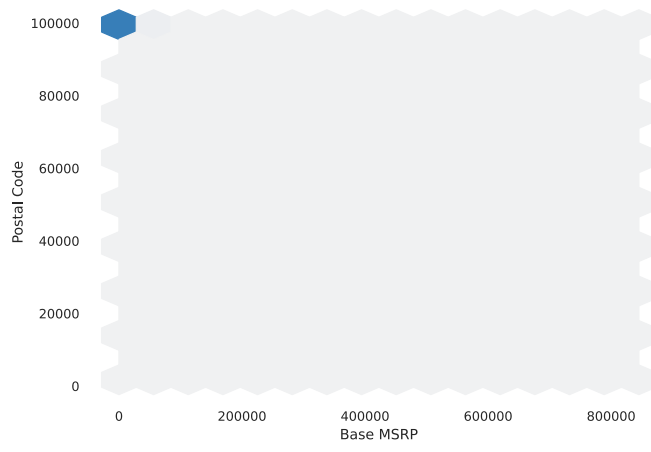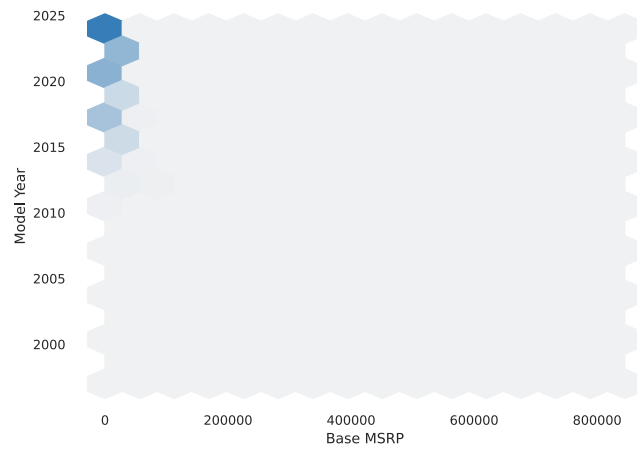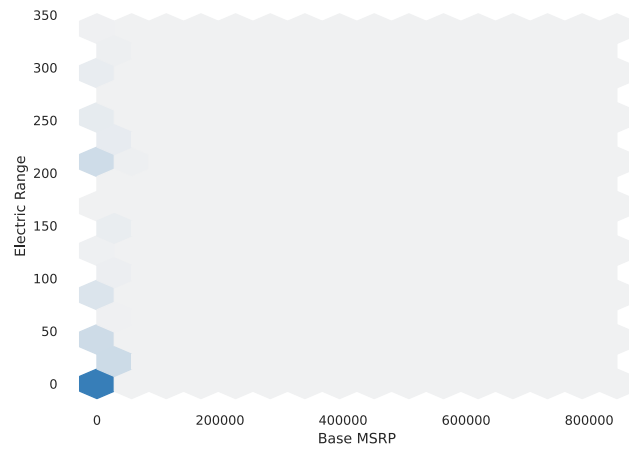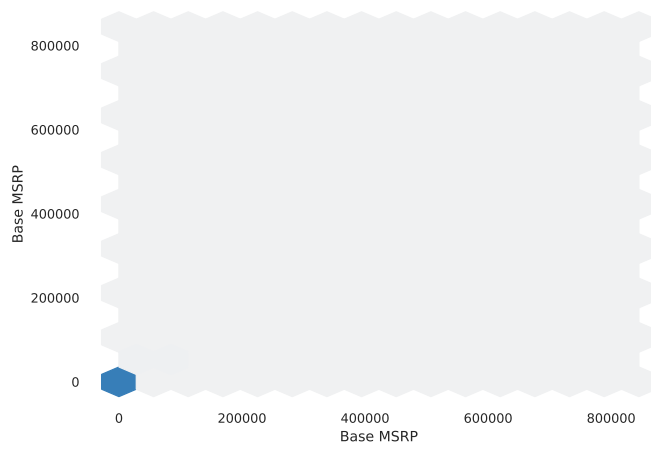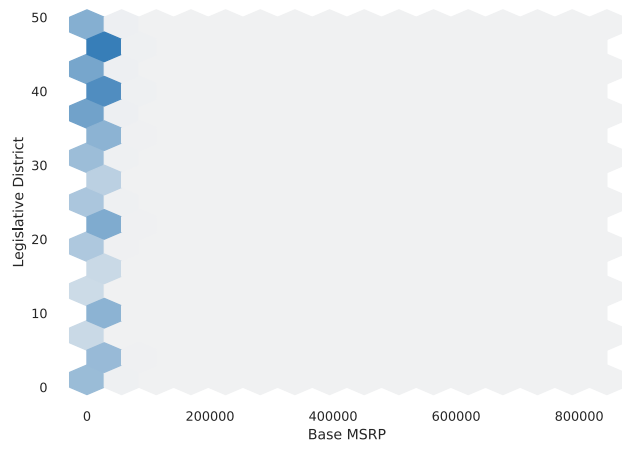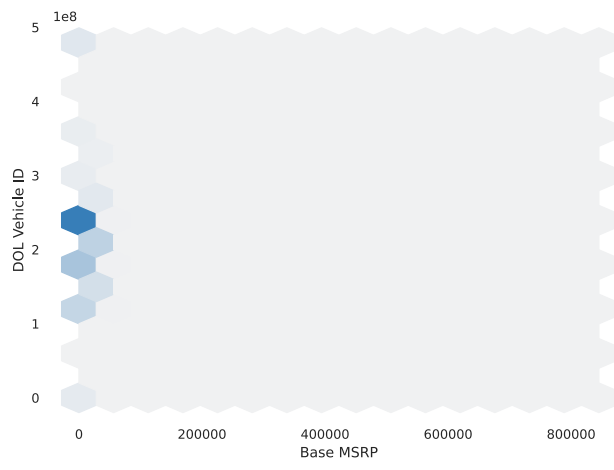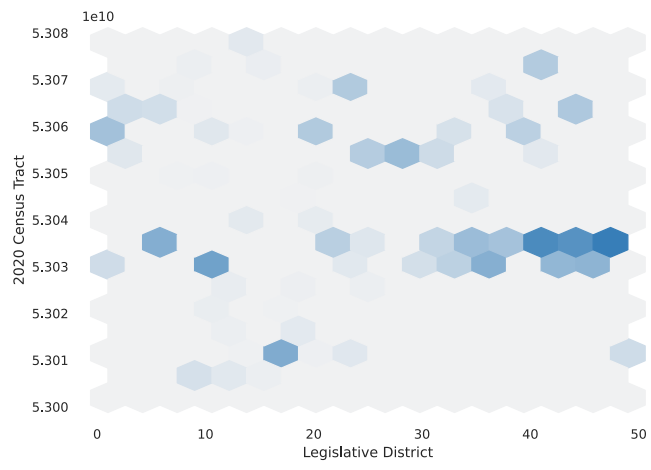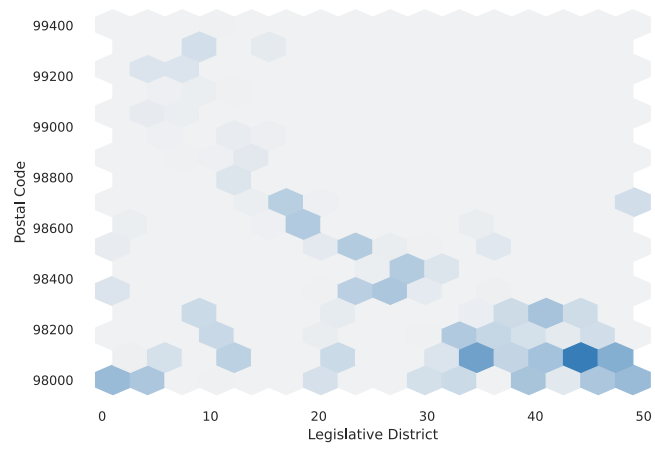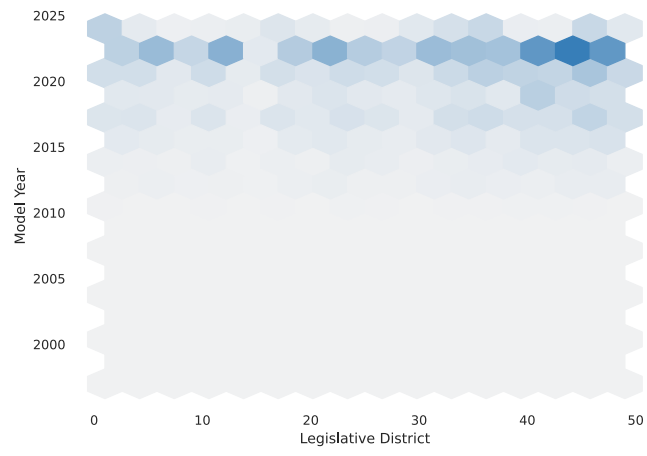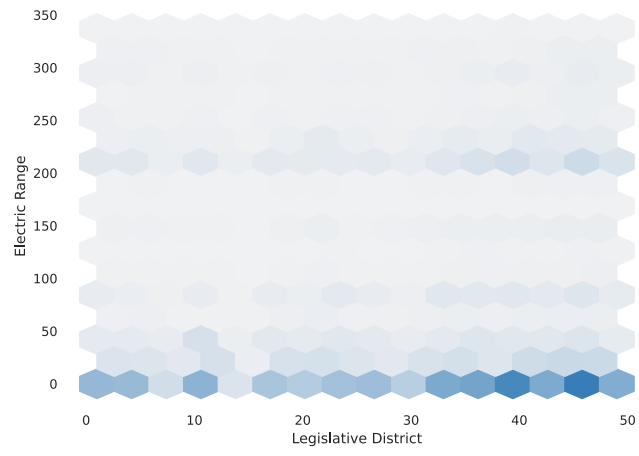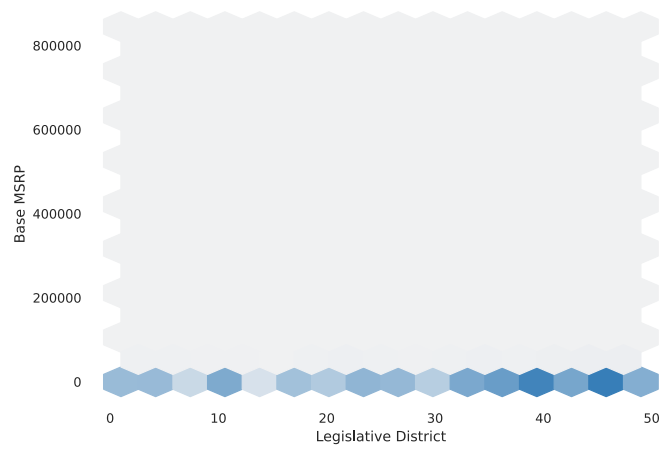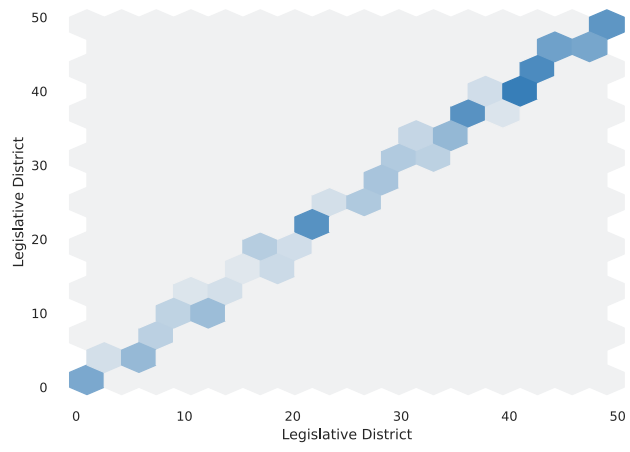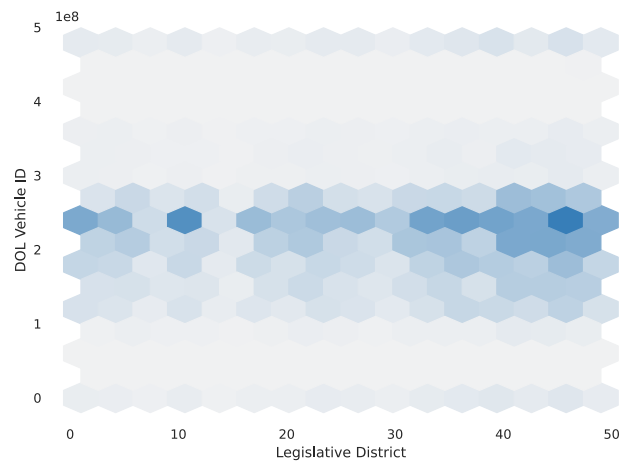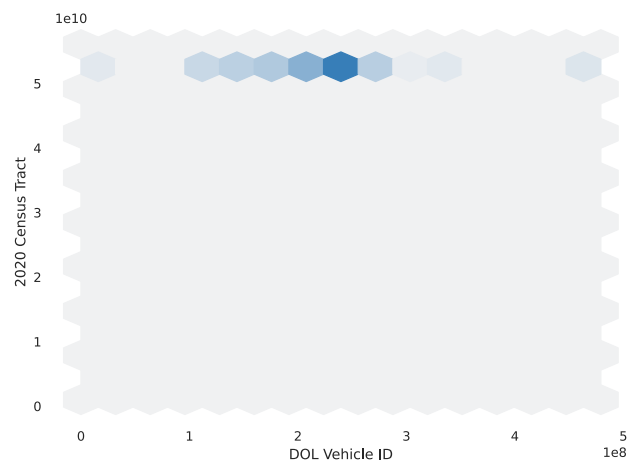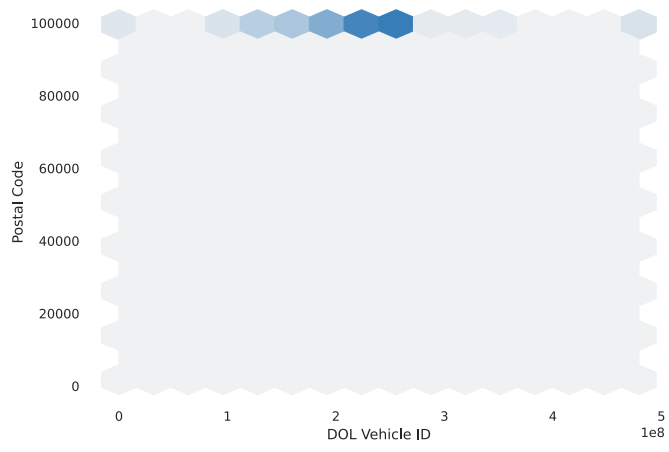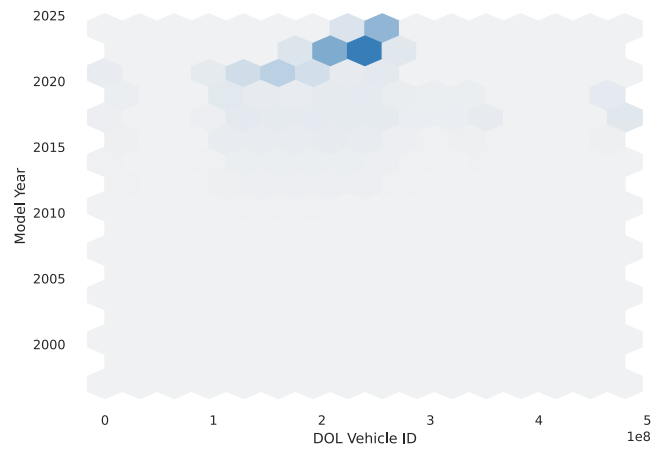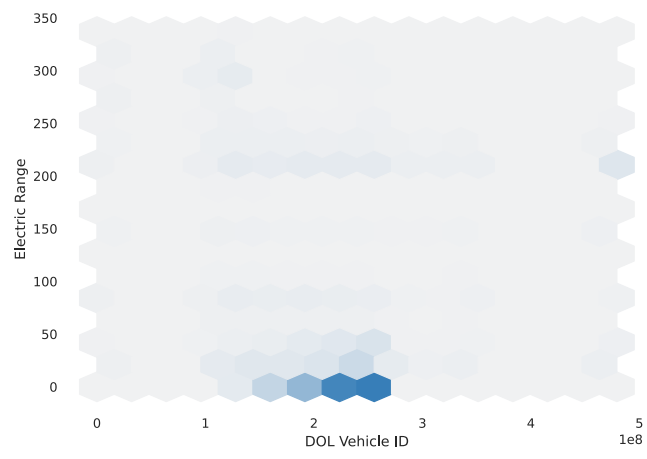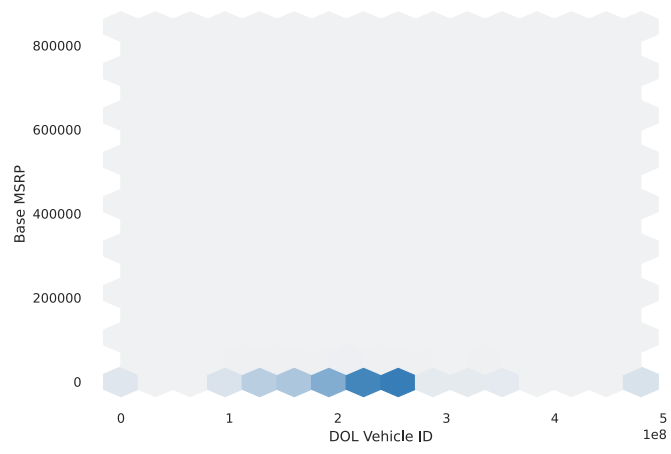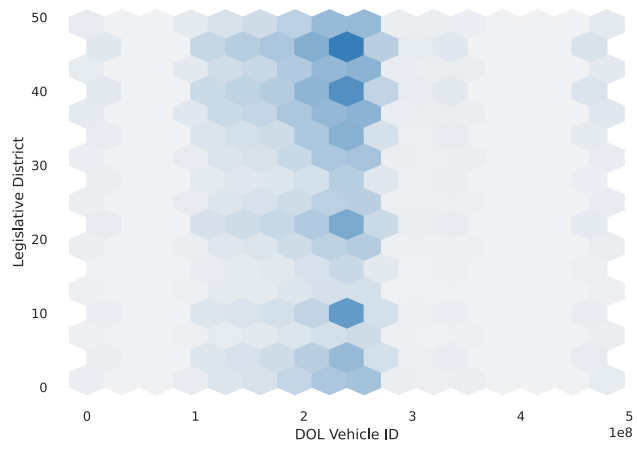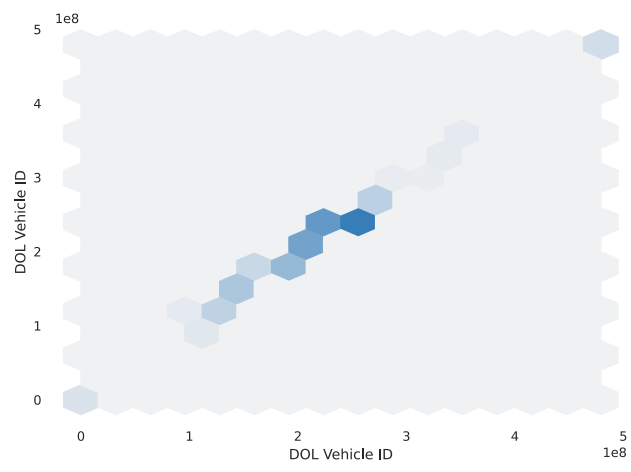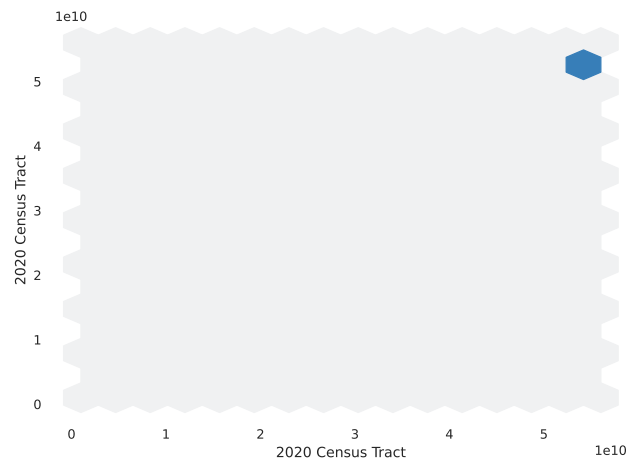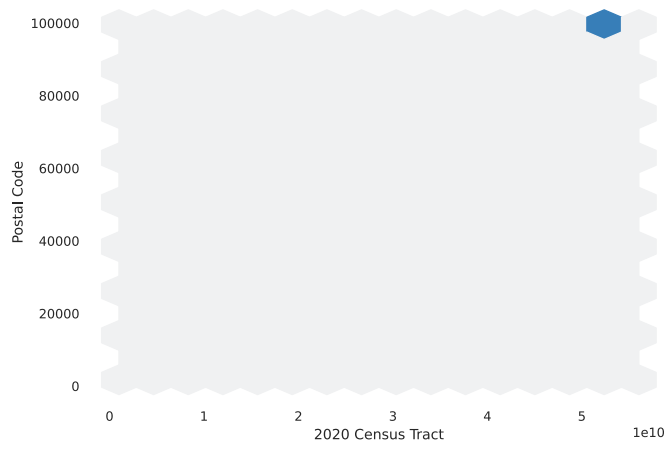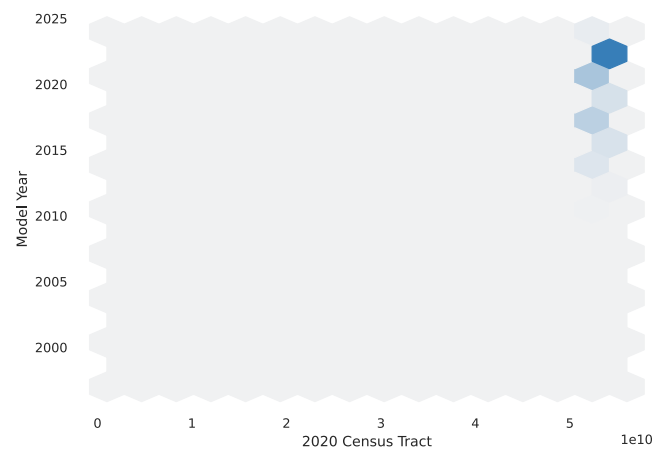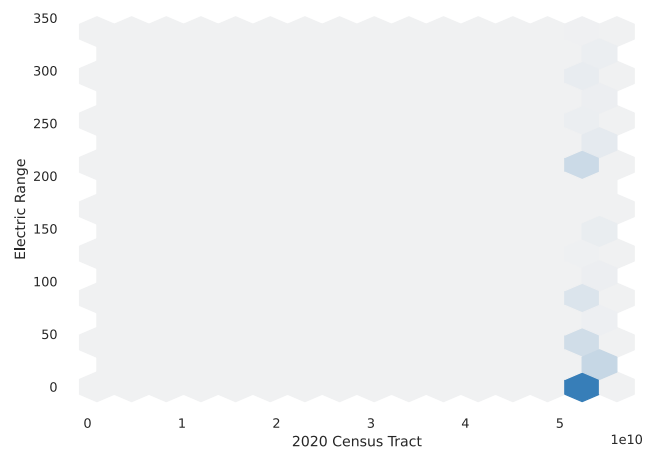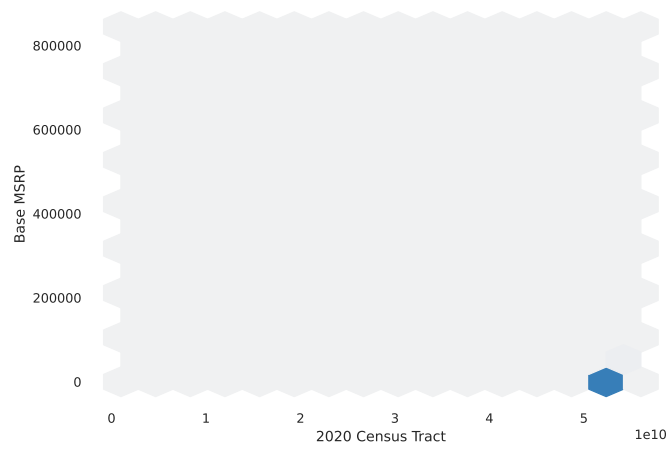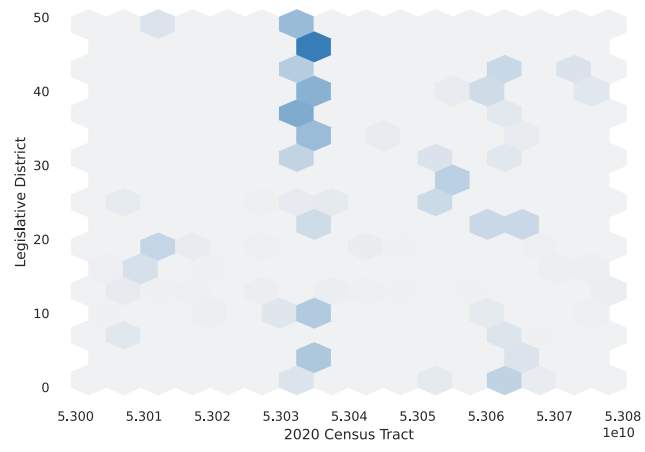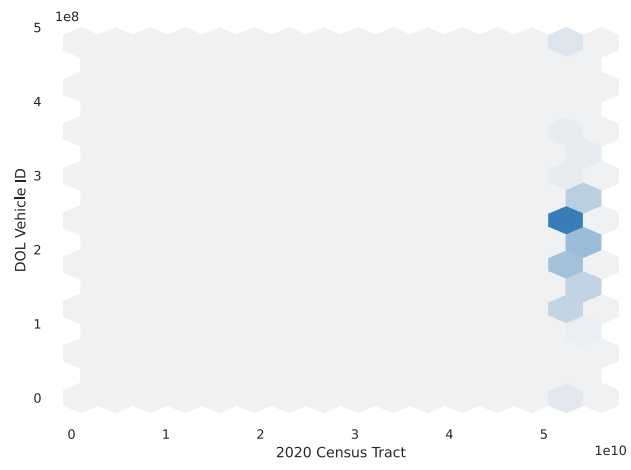
Correlations

	2020 Census Tract	Base MSRP	Clean Alternative Fuel Vehicle (CAFV) Eligibility	DOL Vehicle ID	Electric Range	Electric Vе
2020 Census Tract	1.000	0.000	0.009	0.011	-0.015	0.013
Base MSRP	0.000	1.000	0.024	-0.035	0.116	0.022
Clean Alternative Fuel Vehicle (CAFV) Eligibility	0.009	0.024	1.000	0.092	-0.717	0.743
DOL Vehicle ID	0.011	-0.035	0.092	1.000	-0.160	0.078
Electric Range	-0.015	0.116	-0.717	-0.160	1.000	0.525
Electric Vehicle Type	0.013	0.022	0.743	0.078	0.525	1.000
Legislative District	-0.187	0.011	-0.018	-0.017	-0.004	-0.068
Make	0.009	0.180	0.591	0.113	0.378	0.769
Model Year	0.011	-0.190	0.455	0.348	-0.697	-0.159

	2020 Census Tract	Base MSRP	Clean Alternative Fuel Vehicle (CAFV) Eligibility	DOL Vehicle ID	Electric Range	Electric Vе
2020 Census Tract	1.000	0.000	0.009	0.011	-0.015	0.013
Base MSRP	0.000	1.000	0.024	-0.035	0.116	0.022
Clean Alternative Fuel Vehicle (CAFV) Eligibility	0.009	0.024	1.000	0.092	-0.717	0.743
DOL Vehicle ID	0.011	-0.035	0.092	1.000	-0.160	0.078
Electric Range	-0.015	0.116	-0.717	-0.160	1.000	0.525
Electric Vehicle Type	0.013	0.022	0.743	0.078	0.525	1.000
Legislative District	-0.187	0.011	-0.018	-0.017	-0.004	-0.068
Make	0.009	0.180	0.591	0.113	0.378	0.769
Model Year	0.011	-0.190	0.455	0.348	-0.697	-0.159

Missing values

A simple visualization of nullity by column.

VIN (1-10)
County
City
State
Postal Code
Model Year
Make
Model
Electric Vehicle Type
Clean Alternative Fuel Vehicle (CAFV) Eligibility
Electric Range
Base MSRP
Legislative District
DOL Vehicle ID
Vehicle Location
Electric Utility
2020 Census Tract

1

166800

Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.
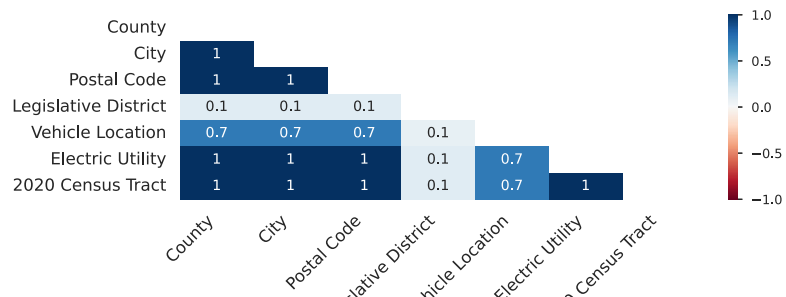
The correlation heatmap measures nullity correlation: how strongly the presence or absence of one variable affects the presence of another.

Sample

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibili
0	3C3CFFGE4E	Yakima	Yakima	WA	98902.0	2014	FIAT	500	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible
1	5YJXCBE40H	Thurston	Olympia	WA	98513.0	2017	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible
2	3MW39FS03P	King	Renton	WA	98058.0	2023	BMW	330E	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range
3	7PDSGABA8P	Snohomish	Bothell	WA	98012.0	2023	RIVIAN	R1S	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not be
4	5YJ3E1EB8L	King	Kent	WA	98031.0	2020	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible
5	5UX43EU02R	Kitsap	Poulsbo	WA	98370.0	2024	BMW	X5	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible
6	2C4RC1H7XJ	Kitsap	Port Orchard	WA	98367.0	2018	CHRYSLER	PACIFICA	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible
7	1G1FX6S01H	Kitsap	Poulsbo	WA	98370.0	2017	CHEVROLET	BOLT EV	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible
8	5YJ3E1EA2J	Kitsap	Port Orchard	WA	98366.0	2018	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible
9	WBY7Z6C59J	King	Duvall	WA	98019.0	2018	BMW	I3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligib
166790	KM8JBDD24R	King	Seattle	WA	98102.0	2024	HYUNDAI	TUCSON	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible
166791	1C4RJXN60R	Whatcom	Bellingham	WA	98225.0	2024	JEEP	WRANGLER	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range
166792	1N4AZ0CP7D	Skagit	Anacortes	WA	98221.0	2013	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible
166793	WP0AA2Y16N	King	Kirkland	WA	98033.0	2022	PORSCHE	TAYCAN	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not
166794	KNDC3DLC0N	Kittitas	Easton	WA	98925.0	2022	KIA	EV6	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not
166795	3FA6P0SU4D	Spokane	Spokane	WA	99223.0	2013	FORD	FUSION	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range
166796	5YJYGDEE5M	King	Sammamish	WA	98074.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not
166797	7SAYGDEE5N	Snohomish	Mukilteo	WA	98275.0	2022	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not
166798	1G1RH6E43D	Lewis	Mossyrock	WA	98564.0	2013	CHEVROLET	VOLT	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible
166799	5YJSA1E27H	Pierce	Gig Harbor	WA	98332.0	2017	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible