

DATA QUALITY REPORT

File Description

File Name: NY property 1 million.xlsx

Description: Data about property in New York.

Number of records: 1,048,575

Number of Features: 30 features

Time Frame: 2010-11

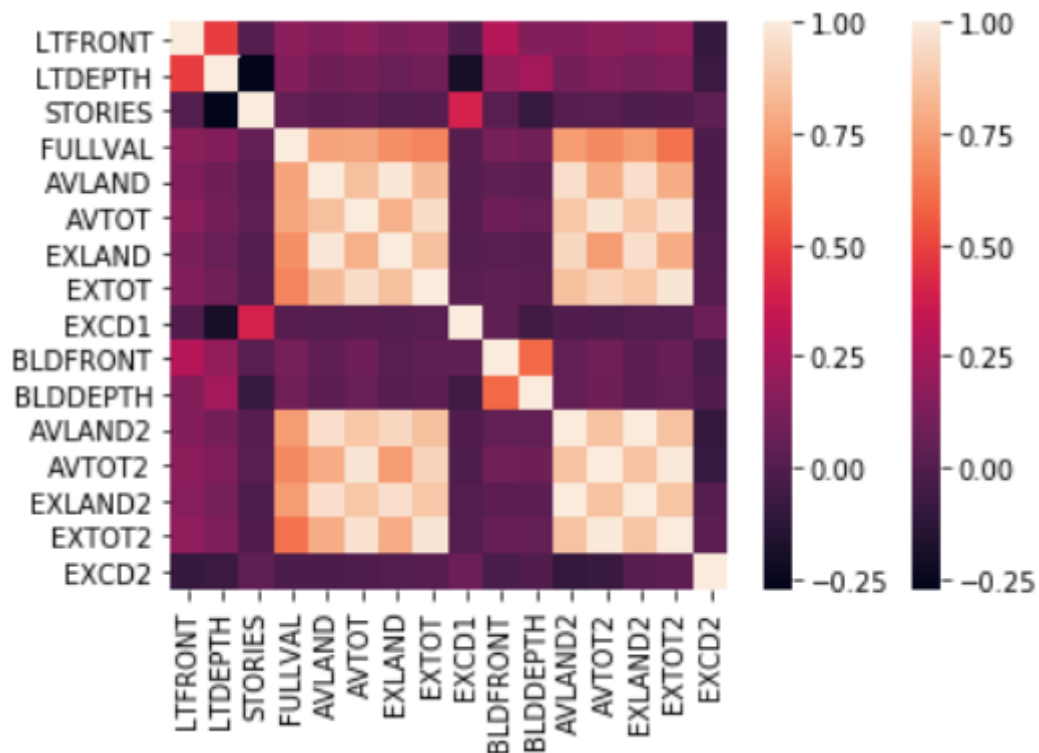
Data Description

Summary Table

This table gives basic summary for numerical features of the data.

	count	mean	std	min	25%	50%	75%	max
RECORD	1048575.0	524288.000000	3.026977e+05	1.0	262144.5	524288.0	786431.50	1.048575e+06
BLOCK	1048575.0	4708.867421	3.699547e+03	1.0	1534.0	3944.0	6797.00	1.635000e+04
LOT	1048575.0	370.092395	8.605382e+02	1.0	23.0	49.0	146.00	9.978000e+03
LTFRONT	1048575.0	36.174248	7.373356e+01	0.0	19.0	25.0	40.00	9.999000e+03
LTDEPTH	1048575.0	88.276429	7.547885e+01	0.0	80.0	100.0	100.00	9.999000e+03
STORIES	996433.0	5.063363	8.431372e+00	1.0	2.0	2.0	3.00	1.190000e+02
FULLVAL	1048575.0	880487.657847	1.170293e+07	0.0	303000.0	446000.0	619000.00	6.150000e+09
AVLAND	1048575.0	85995.027083	4.100755e+06	0.0	9160.0	13646.0	19706.00	2.668500e+09
AVTOT	1048575.0	230758.183174	6.951206e+06	0.0	18385.0	25339.0	46095.00	4.668309e+09
EXLAND	1048575.0	36811.788682	4.024330e+06	0.0	0.0	1620.0	1620.00	2.668500e+09
EXTOT	1048575.0	92543.814625	6.578281e+06	0.0	0.0	1620.0	2090.00	4.668309e+09
EXCD1	622642.0	1604.500100	1.388132e+03	1010.0	1017.0	1017.0	1017.00	7.170000e+03
ZIP	1022219.0	10935.315946	5.265759e+02	10001.0	10453.0	11215.0	11364.00	3.380300e+04
BLDFRONT	1048575.0	23.018722	3.578847e+01	0.0	15.0	20.0	24.00	7.575000e+03
BLDDEPTH	1048575.0	40.074211	4.303640e+01	0.0	26.0	39.0	51.00	9.393000e+03
AVLAND2	280966.0	246365.484475	6.199390e+06	3.0	5705.0	20059.0	62338.75	2.371005e+09
AVTOT2	280972.0	716078.713584	1.169017e+07	3.0	34013.5	80010.0	240792.00	4.501180e+09
EXLAND2	86675.0	351802.210545	1.085248e+07	1.0	2090.0	3053.0	31419.00	2.371005e+09
EXTOT2	129933.0	658114.779009	1.612981e+07	7.0	2889.0	37116.0	106629.00	4.501180e+09
EXCD2	90941.0	1371.659098	1.105490e+03	1011.0	1017.0	1017.0	1017.00	7.160000e+03

Heatmap showing correlation of numerical variables:



Feature Description

1. RECORD

Description: Manually added field to uniquely identify each observation.

Percent Populated: 100%

Number of unique values: 1,048,575

2. BBLE

Description: Combination of boro, block, lot and easement.

Percent Populated: 100%

Number of unique values: 1,048,575

3. BLOCK

Description: Block in which the property is located

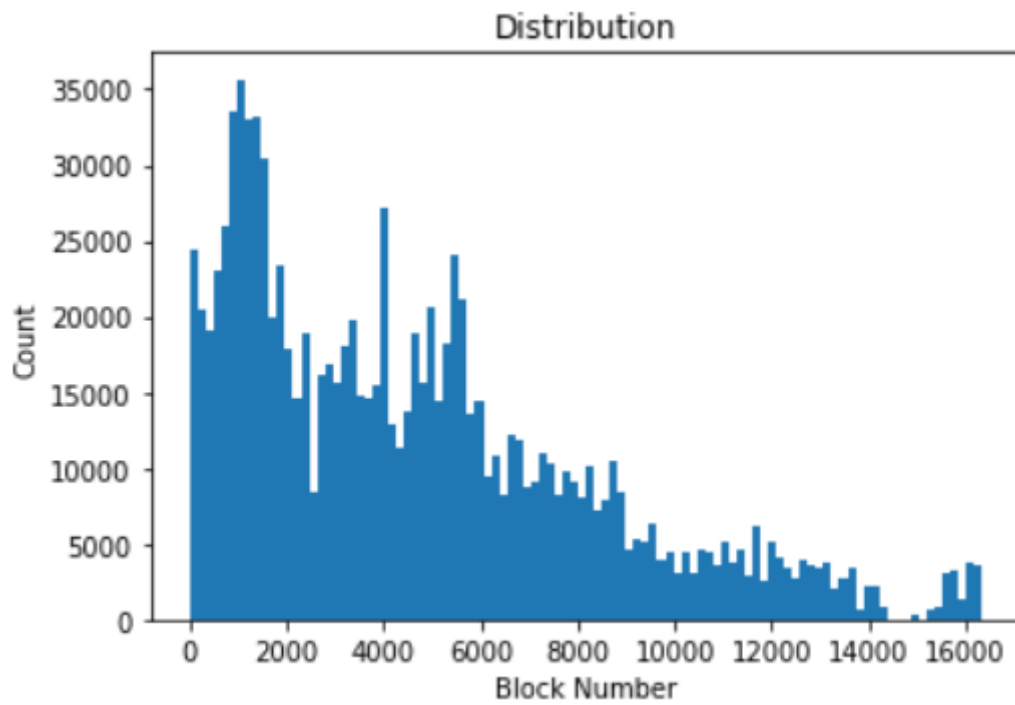
Percent Populated: 100%

Number of Unique Values: 13949

Minimum: 1

Maximum: 16350

Distribution:



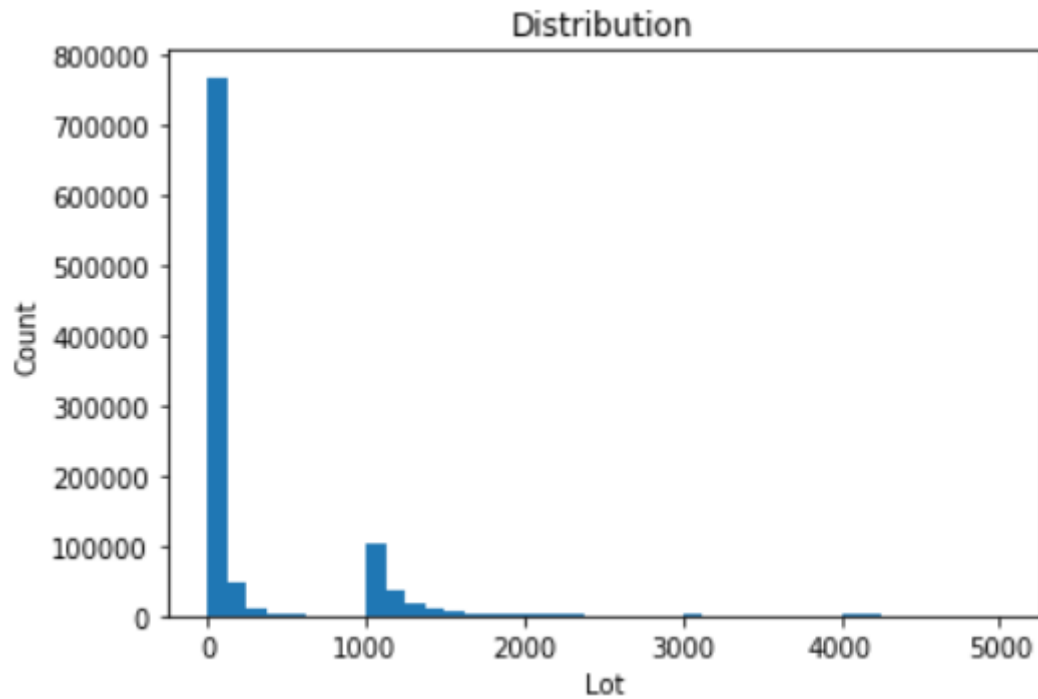
4. LOT

Description: Lot in which the property is located within the boro/block

Percent Populated: 100%

Number of Unique Values: 6366

Distribution:



5. Easement

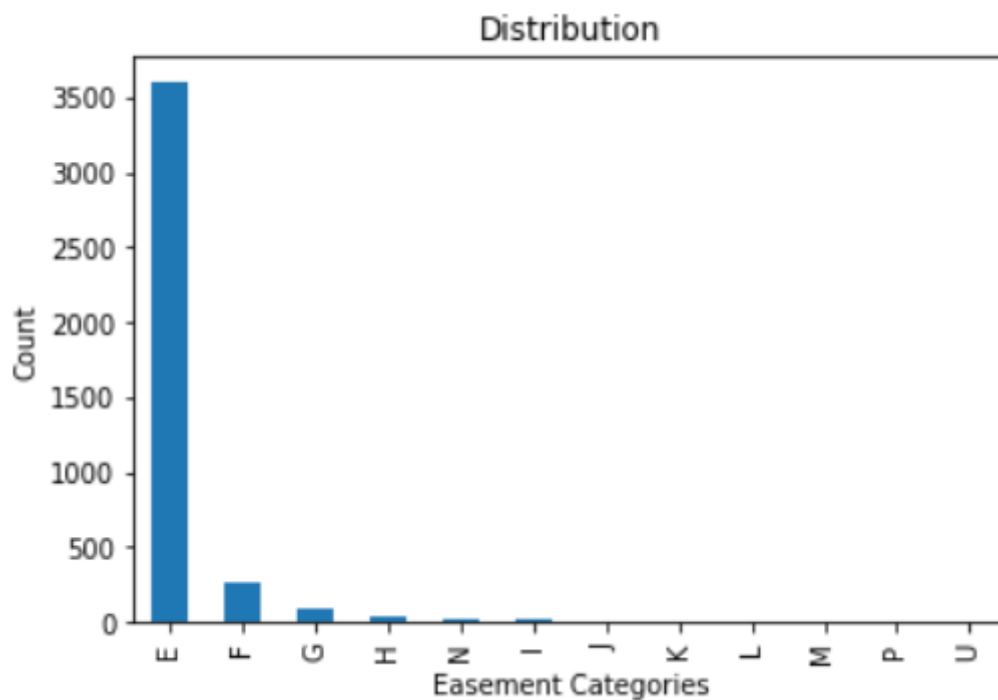
Description: Categorical variable related to property

- i. SPACE Indicates the lot has no Easement
- ii. 'A' Indicates the portion of the Lot that has an Air
- iii. 'B' Indicates Non-Air Rights
- iv. 'E' Indicates the portion of the lot that has a Land
- v. 'F' through 'M' are duplicates of 'E'.
- vi. 'N' indicates Non-Transit
- vii. 'P' indicates Piers
- viii. 'R' Indicates Railroads
- ix. 'S' Indicates Street
- x. 'U' Indicates U.S. Government

Percent Populated: 38.55%

Number of Unique Values: 13

Distribution:



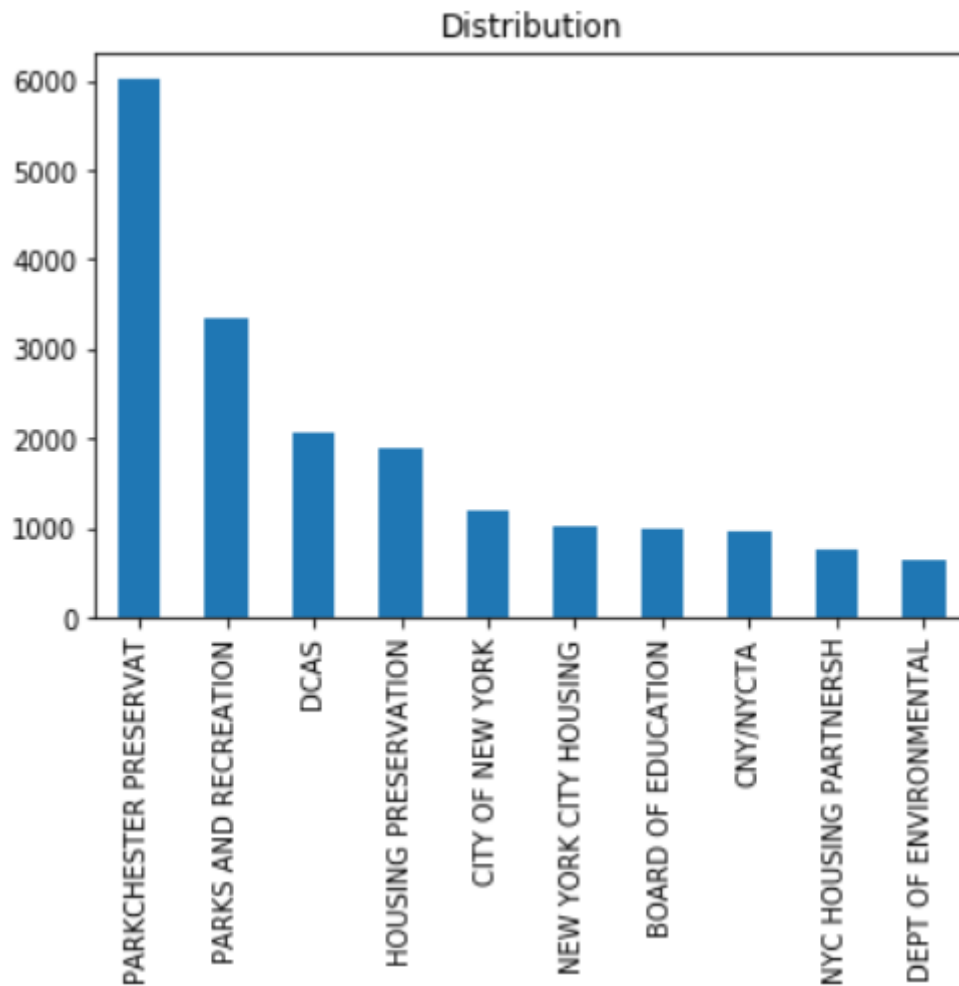
6. Owner

Description: Owner Details

Percent Populated: 97.03%

Number of Unique Values: 847054

Distribution:



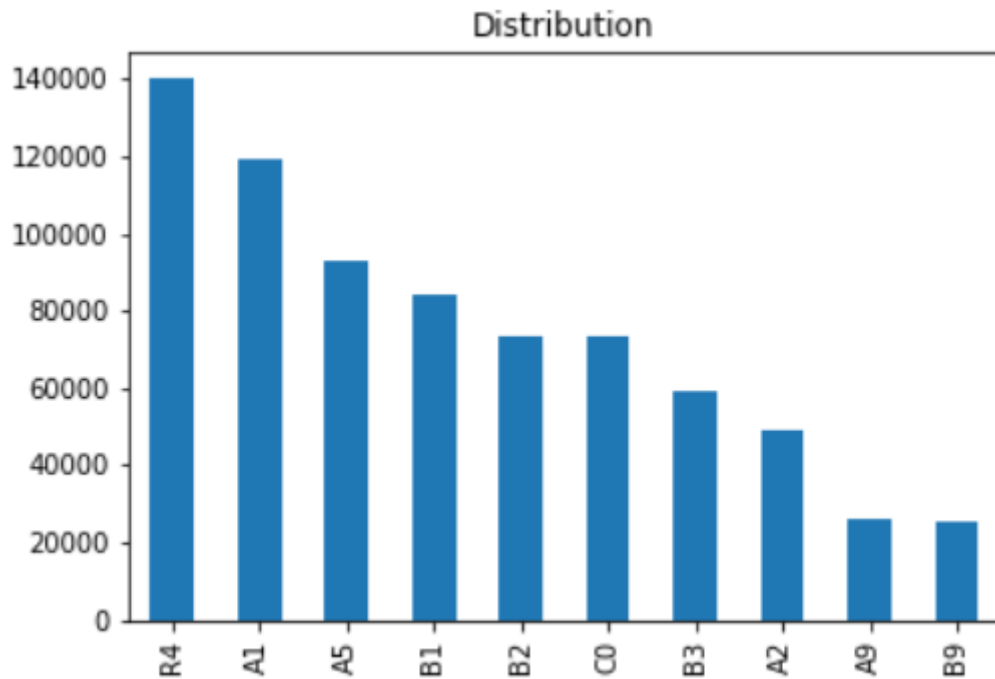
7. BLDGCL

Description: Class of building based on tax laws.

Percent Populated: 100%

Number of Unique Values: 200

Distribution:



8. Tax Class

Description: Categorical Variable about tax class to which property belongs.

Percent Populated: 100%

Number of Unique Values: 11

9. LTFRONT

Description: Measure of front lot of a property

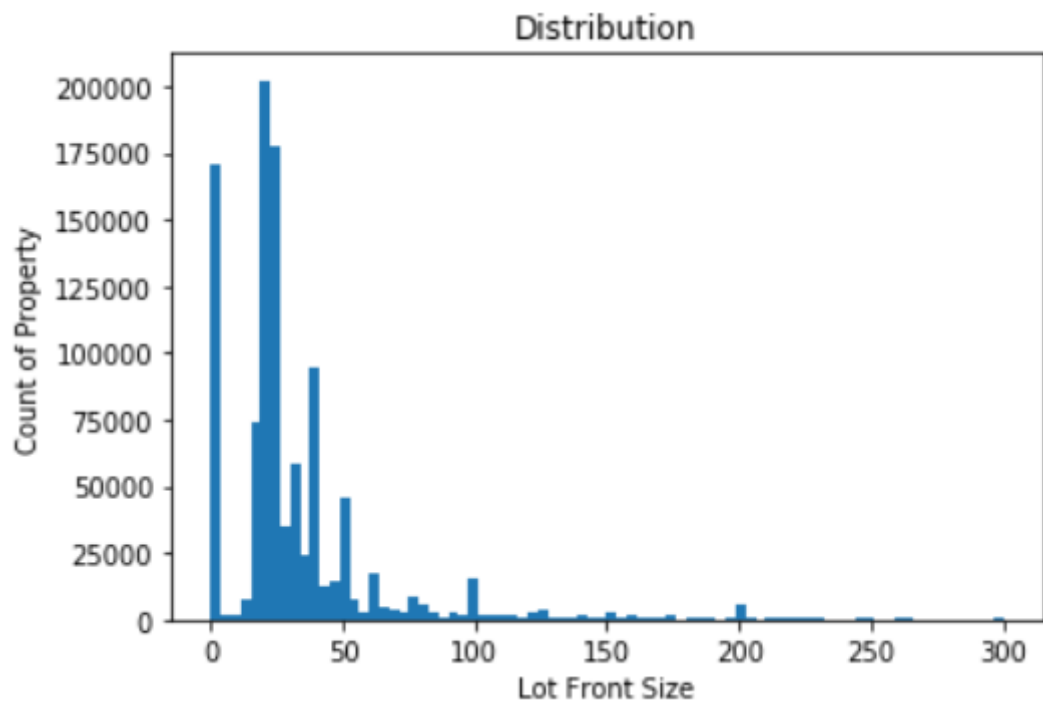
Percent Populated: 100%

Number of Unique Values: 1277

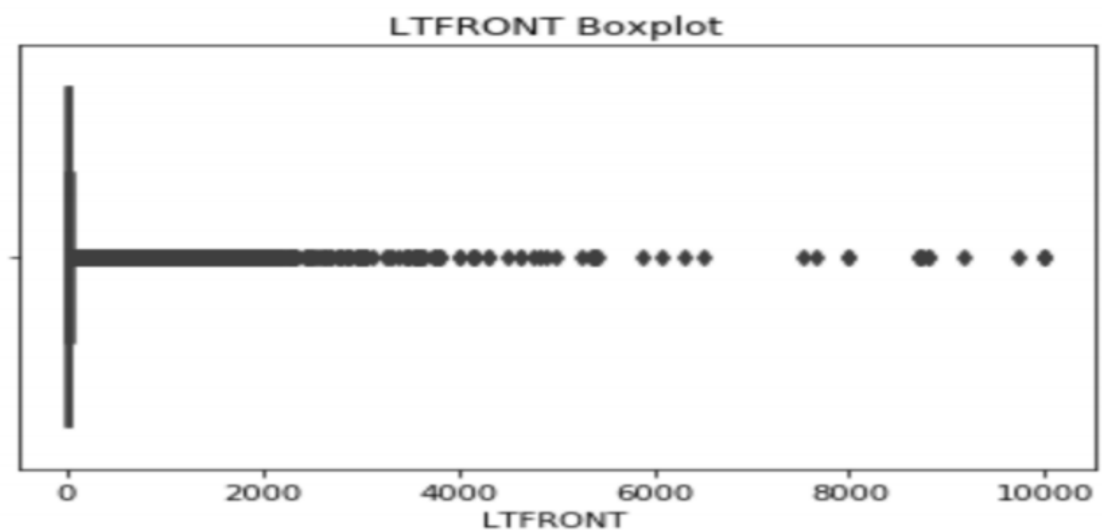
Minimum: 0

Maximum: 9999

Distribution:



Boxplot:



10. LTDEPTH

Description: Measure of front lot of a property

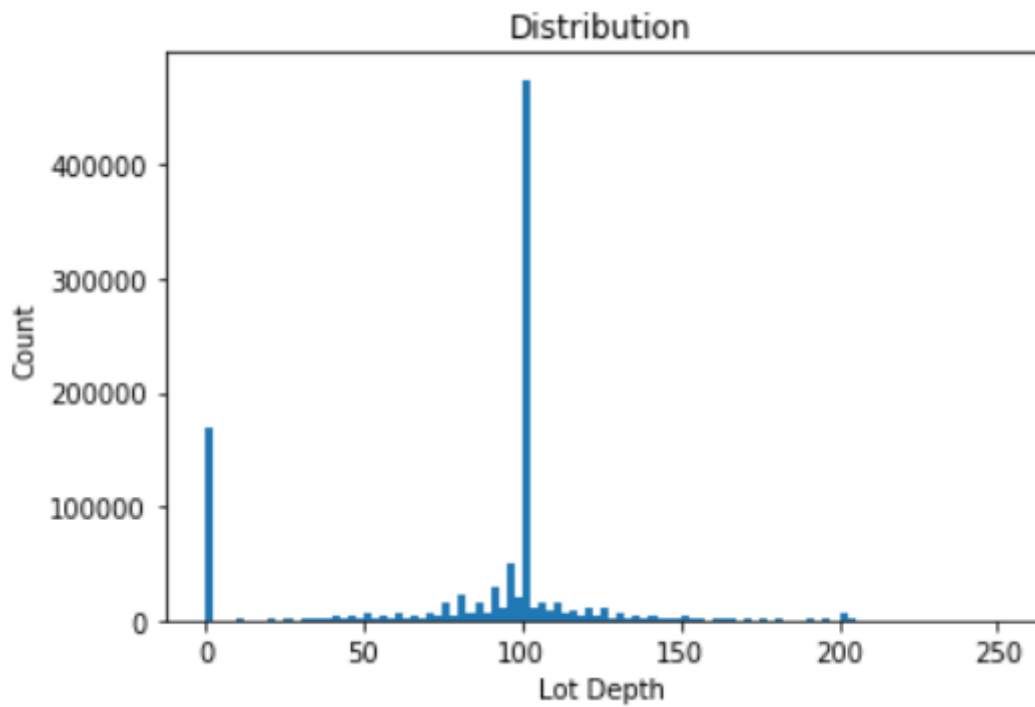
Percent Populated: 100%

Number of Unique Values: 1336

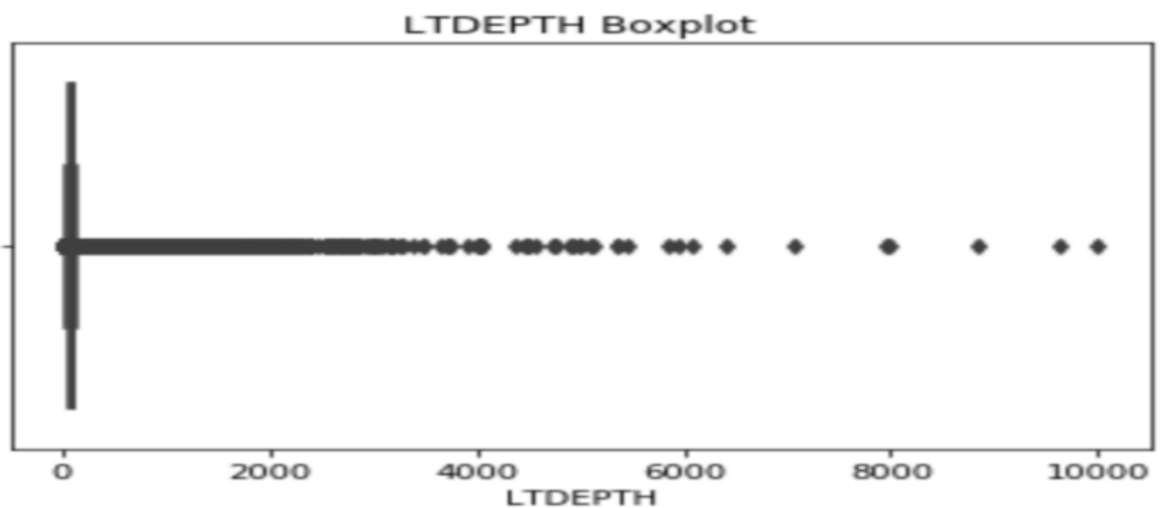
Minimum: 0

Maximum: 9999

Distribution:



Boxplot:



11. STORIES

Description: Number of levels in a building

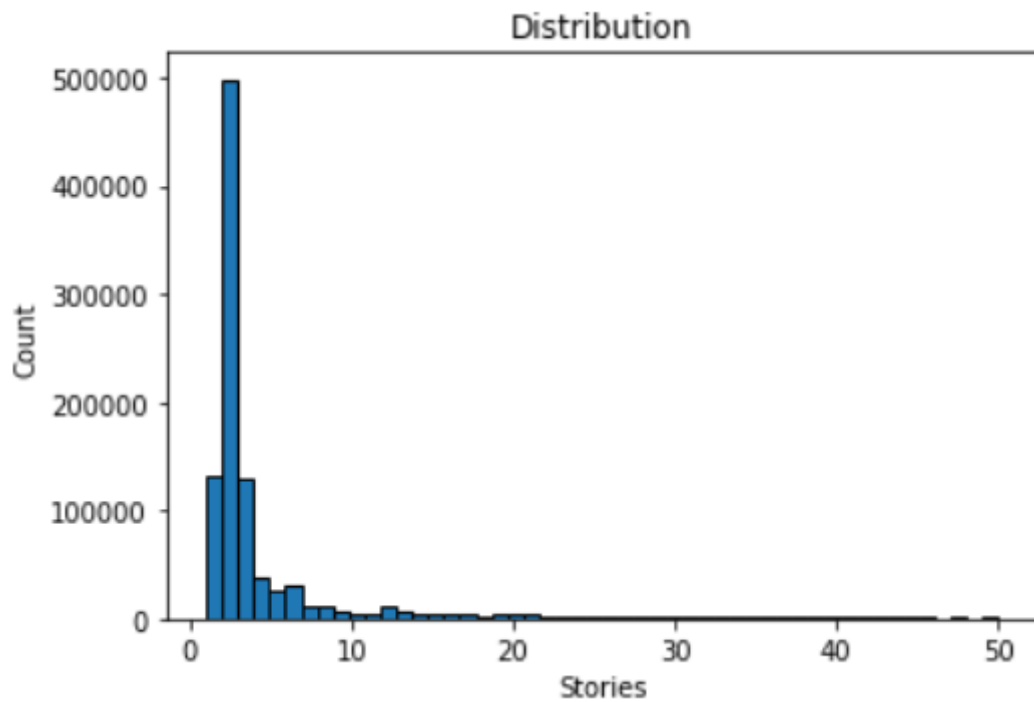
Percent Populated: 95.02%

Number of Unique Values: 112

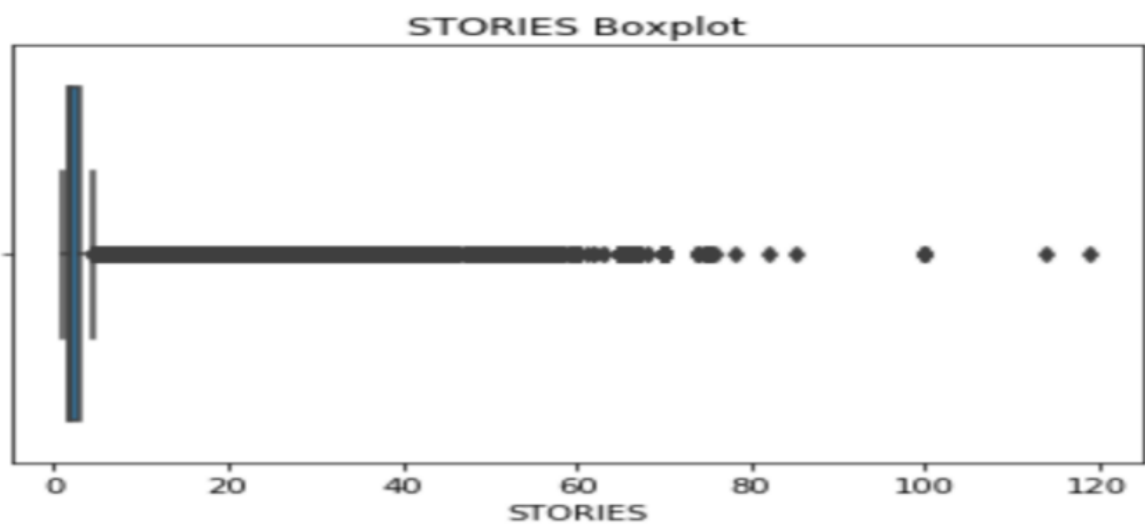
Minimum: 1

Maximum: 119

Distribution:



Boxplot:



12. FULLVAL

Description: Numerical variable related to value of the property

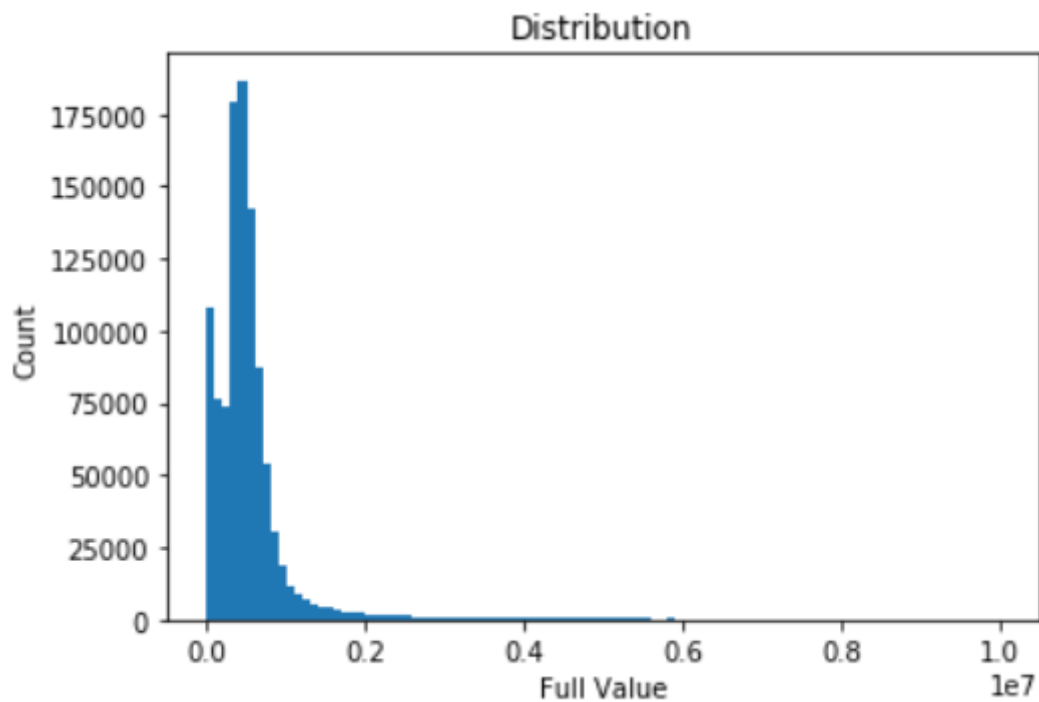
Percent Populated: 95.02%

Number of Unique Values: 108277

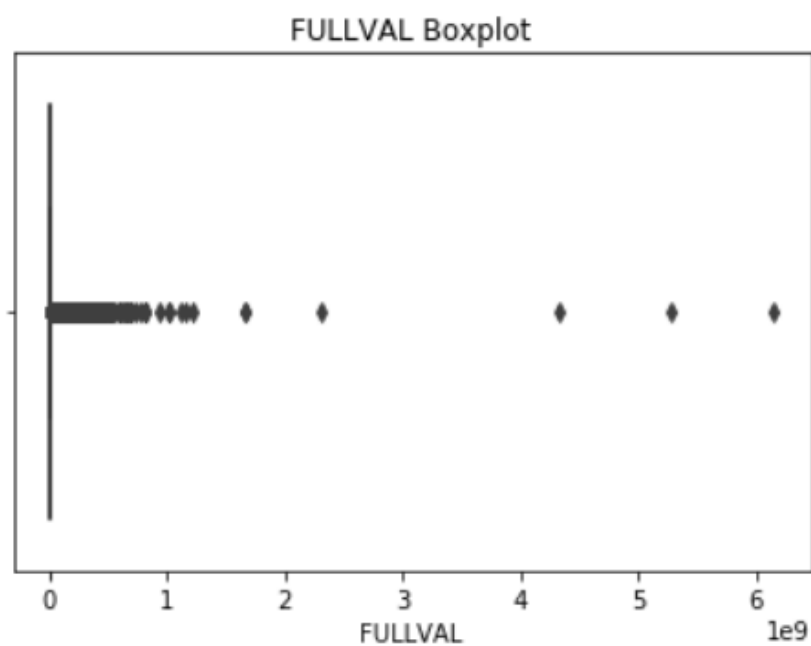
Minimum: 0

Maximum: 6,150,000,000

Distribution:



Boxplot:



13. AVLAND

Description: Numerical variable related to value of the property

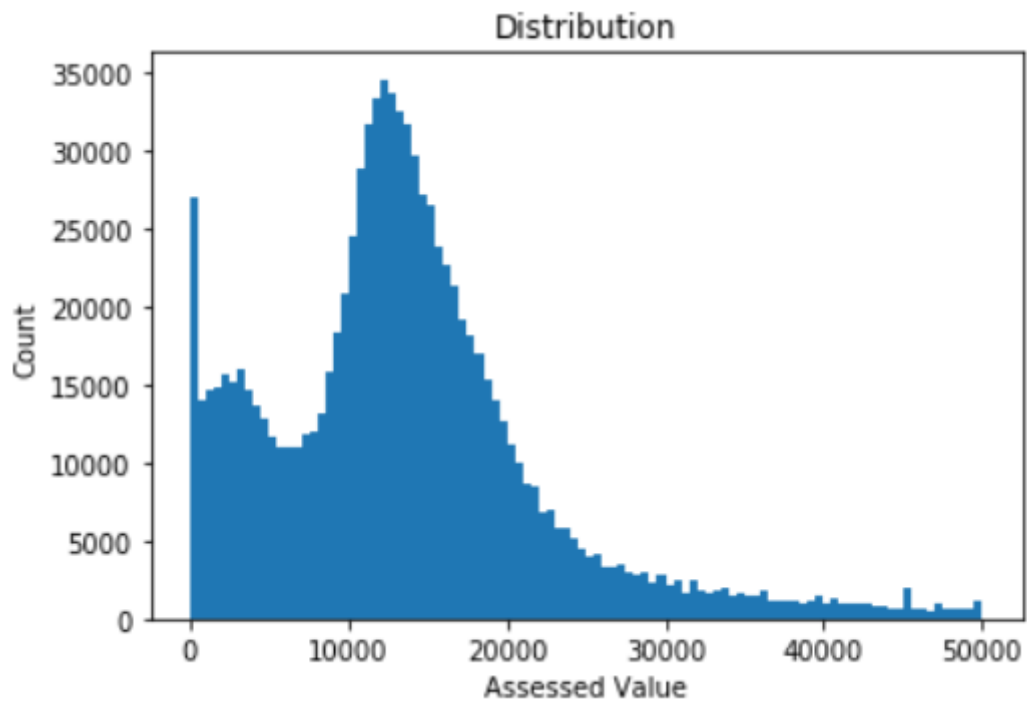
Percent Populated: 95.02%

Number of Unique Values: 70529

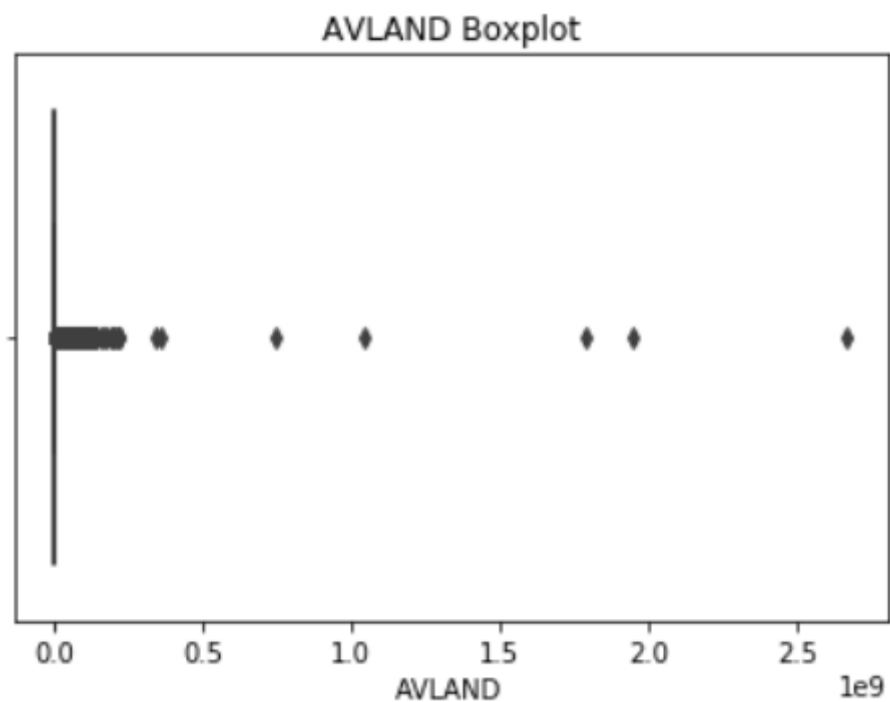
Minimum: 0

Maximum: 2,668,500,000

Distribution:



Boxplot:



14. AVTOT

Description: Numerical variable related to value of the property

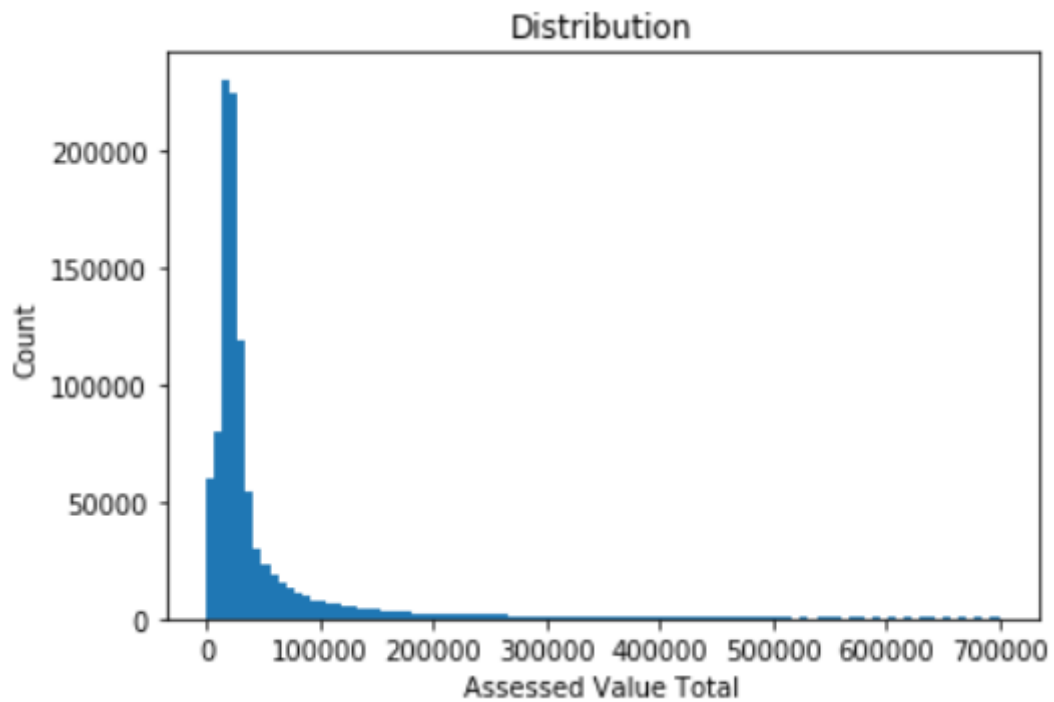
Percent Populated: 100%

Number of Unique Values: 112294

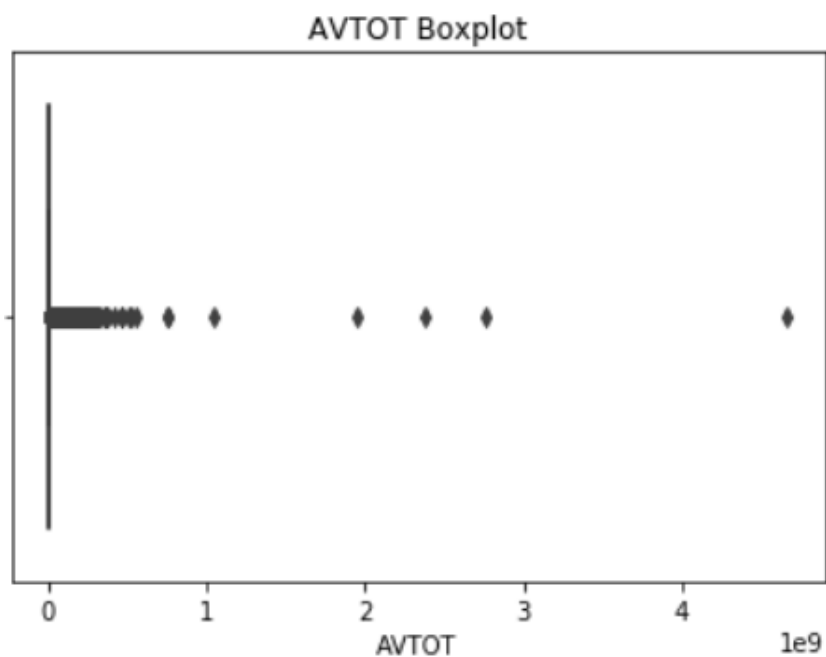
Minimum: 0

Maximum: 4,668,308,947

Distribution:



Boxplot:



15. EXLAND

Description: Numerical variable

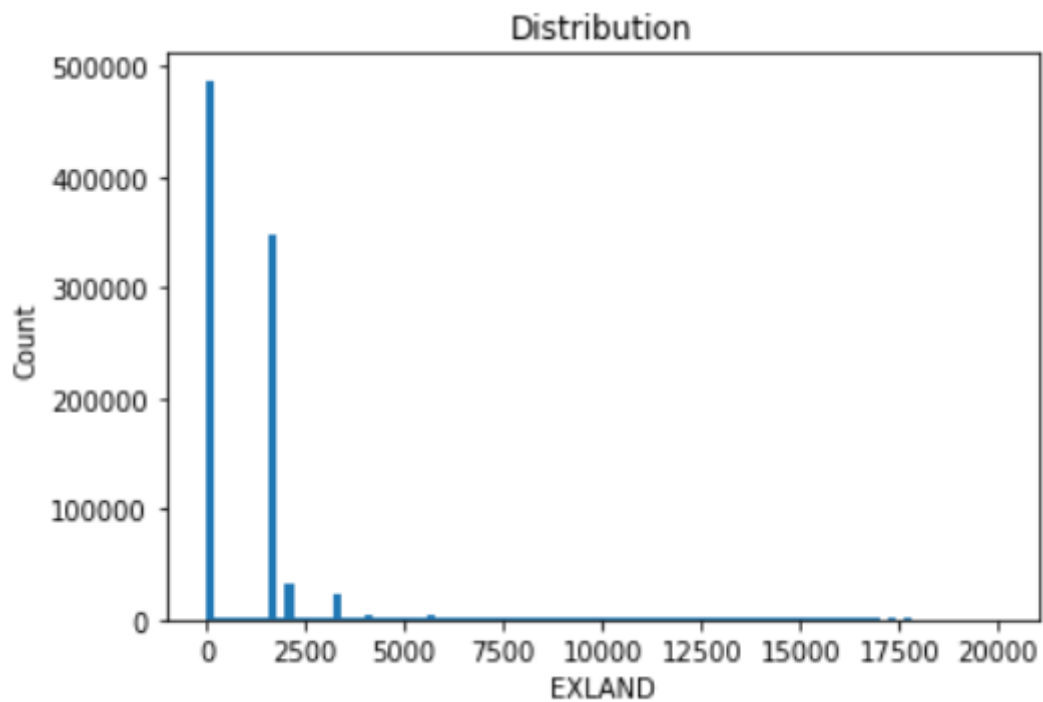
Percent Populated: 100%

Number of Unique Values: 33186

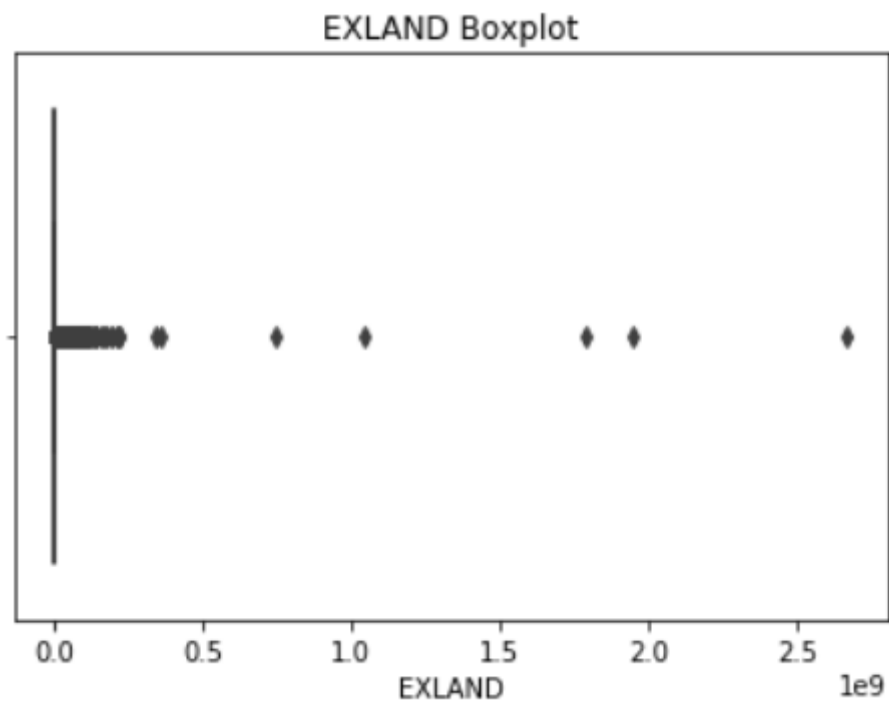
Minimum: 0

Maximum: 2,668,500,000

Distribution:



Boxplot:



16. EXTOT

Description: Numerical variable

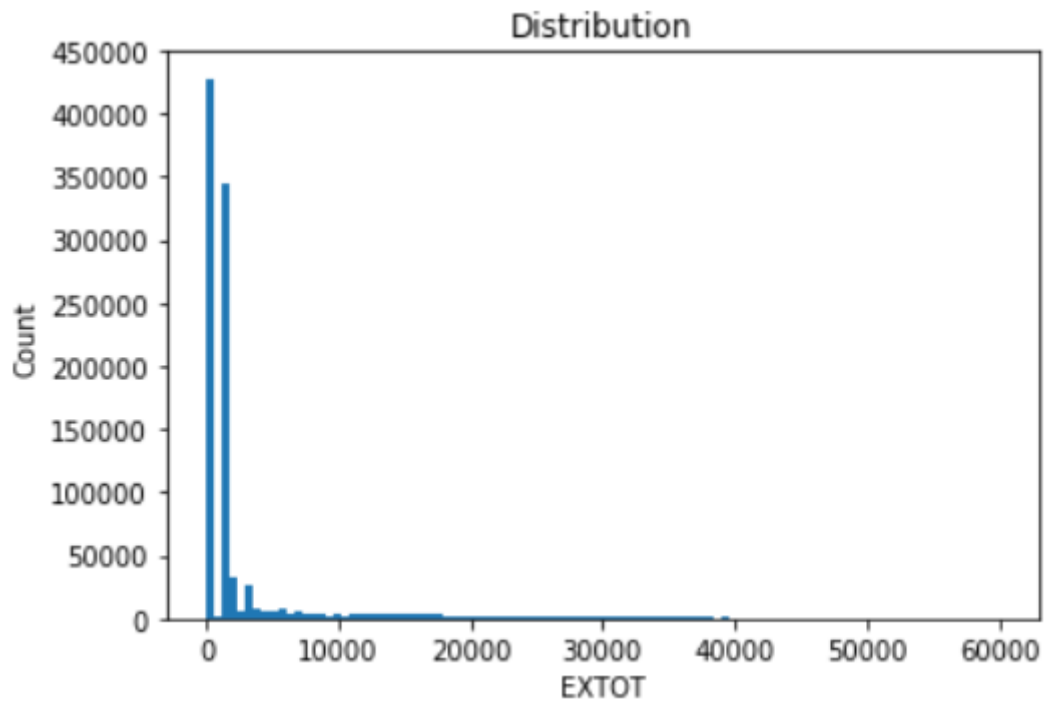
Percent Populated: 100%

Number of Unique Values: 63805

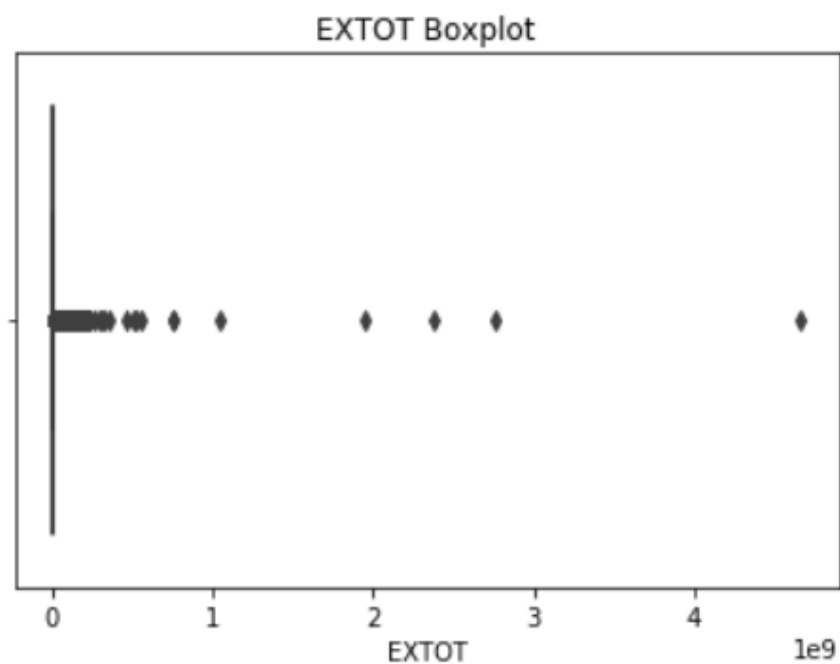
Minimum: 0

Maximum: 4,668,308,947

Distribution:



Boxplot:

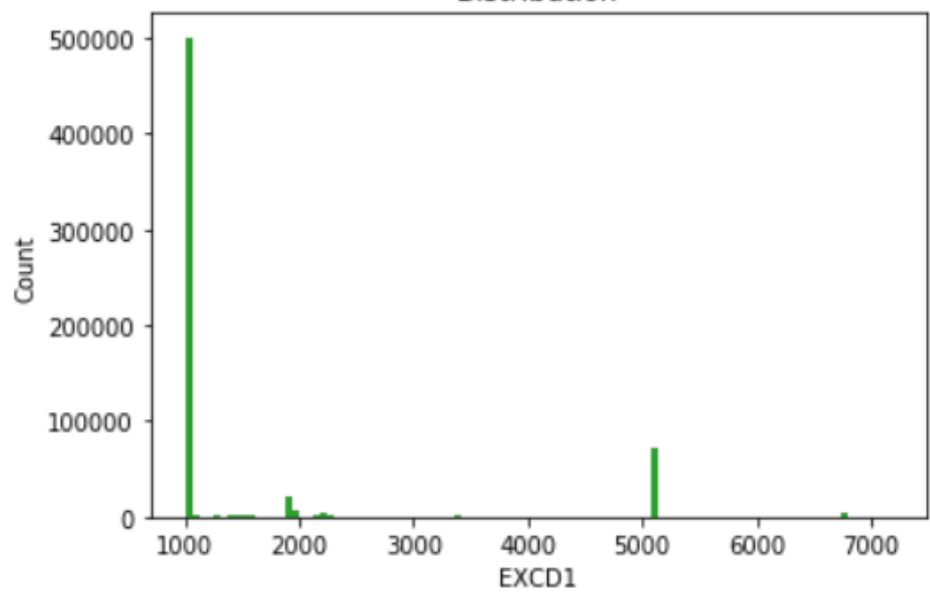


A noticeable thing here is how maximum for AVTOT and EXTOT are the same and EXLAND and AVLAND are the same.

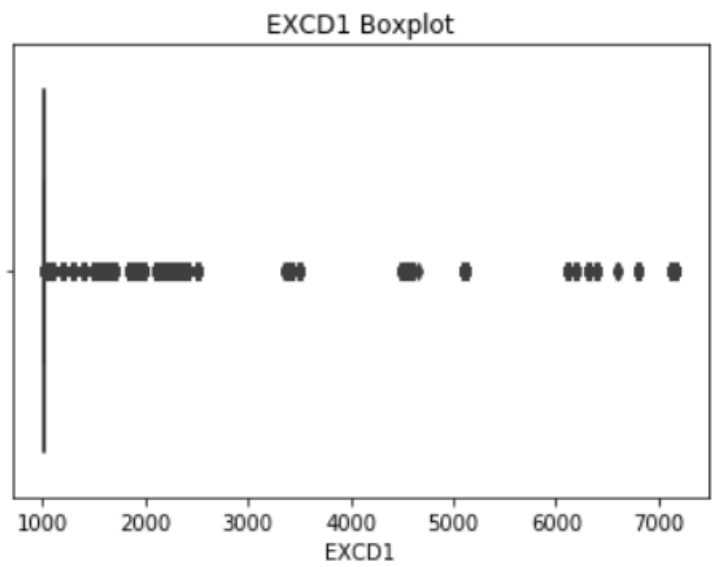
17. EXCD1

Description: Numerical variable
Percent Populated: 59.37%
Number of Unique Values: 130
Minimum: 1010
Maximum: 7170

Distribution:
Distribution



Boxplot:



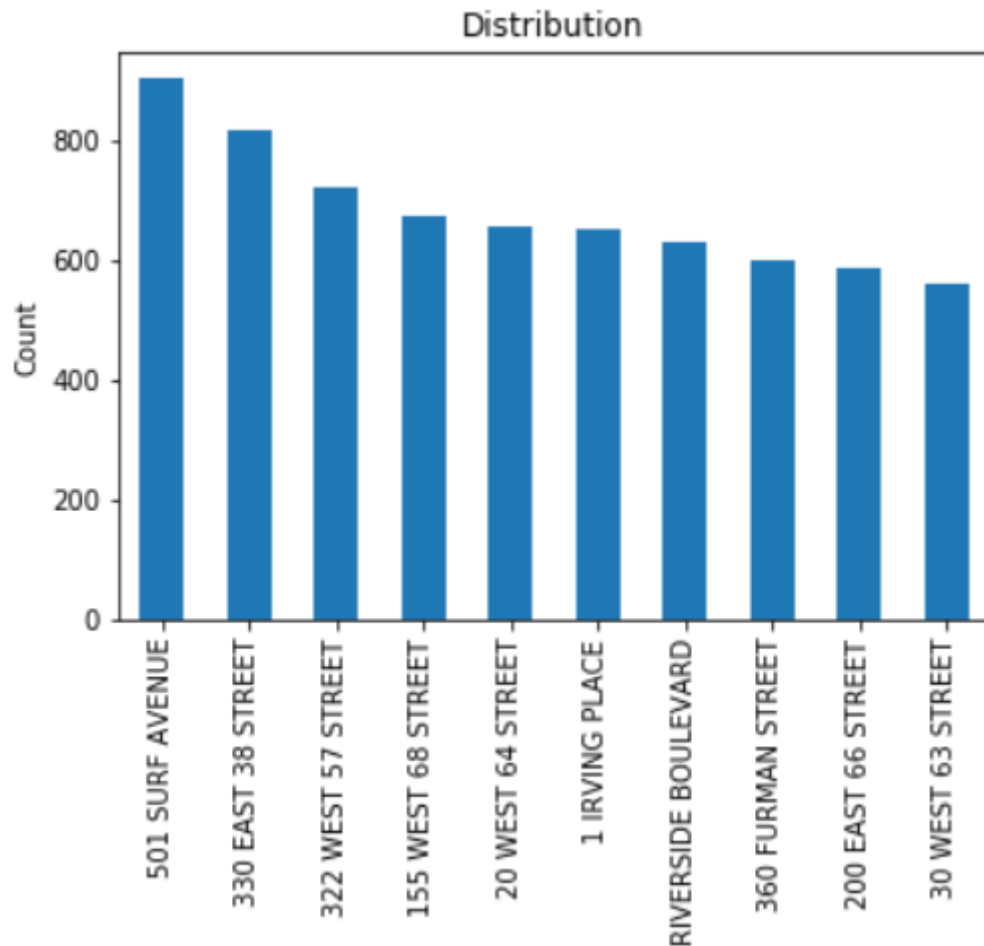
18. STADDR

Description: Street address of the property

Percent Populated: 99.93%

Number of Unique Values: 820638

Distribution: Top 10 addresses by property count -



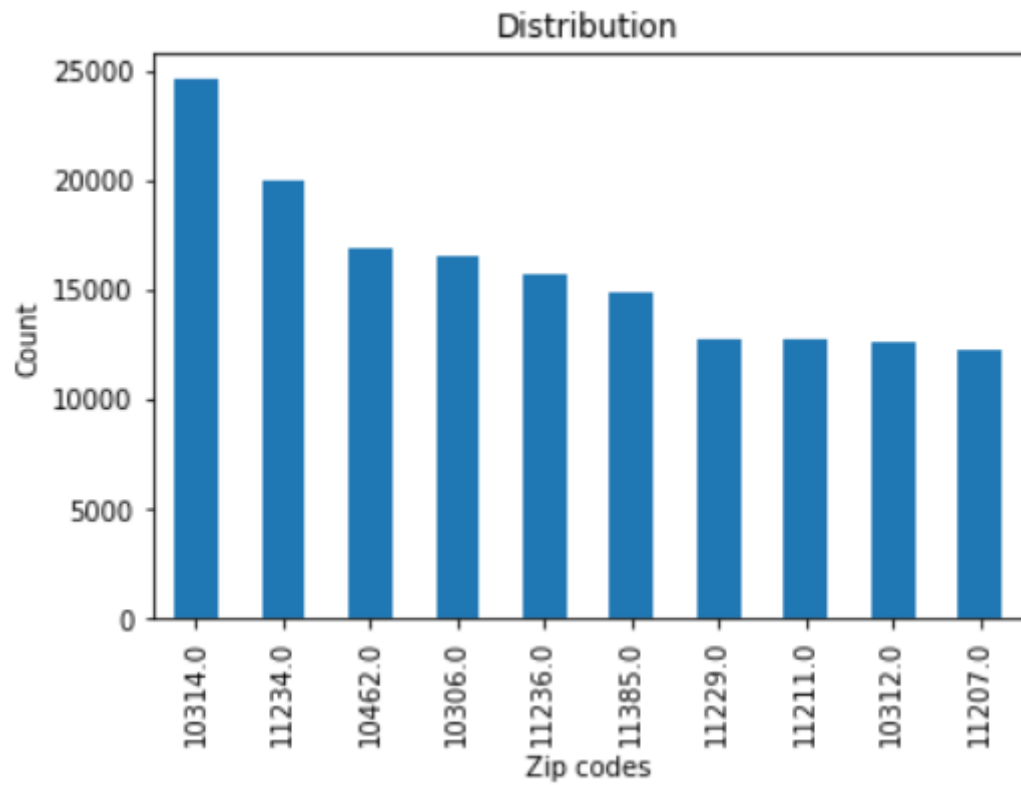
19. ZIP

Description: Zip codes of the property

Percent Populated: 97.48%

Number of Unique Values: 197

Distribution: Top 10 zip codes by property count -



20. EXMPTCL

Description: Categorical variable

Percent Populated: 1.42%

Number of Unique Values: 15

21. BLDFRONT

Description: Numerical variable describing building's front area in feet

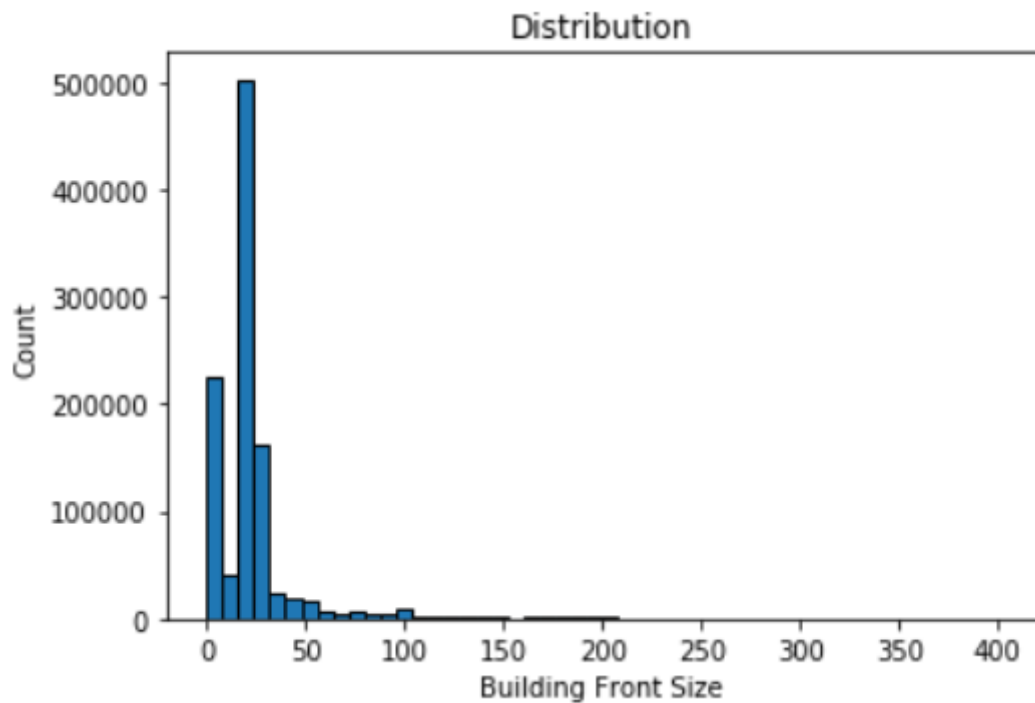
Percent Populated: 100%

Number of Unique Values: 610

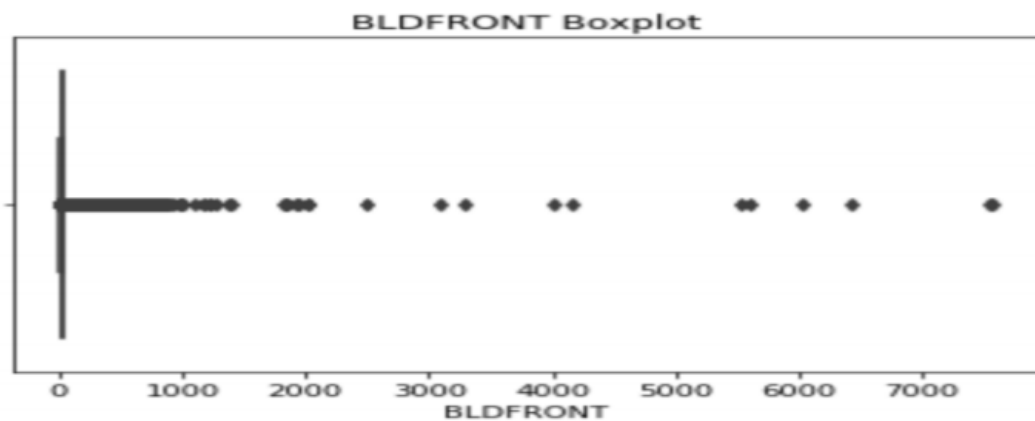
Minimum: 0

Maximum: 7575

Distribution:



Boxplot:



22. BLDDEPTH

Description: Numerical variable describing building's depth in feet

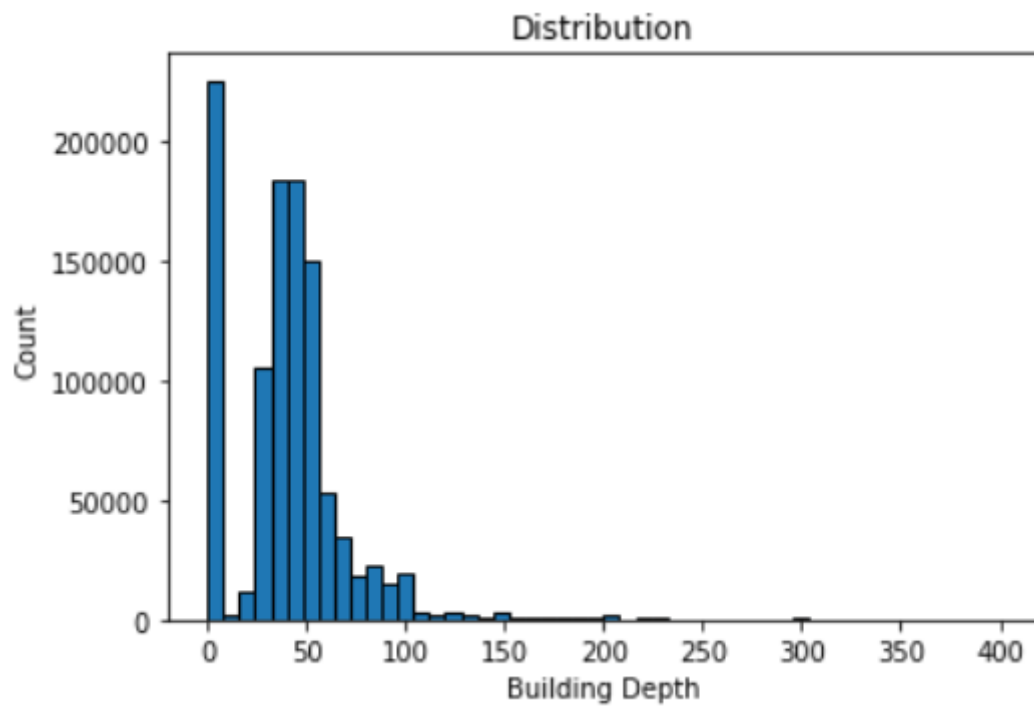
Percent Populated: 100%

Number of Unique Values: 620

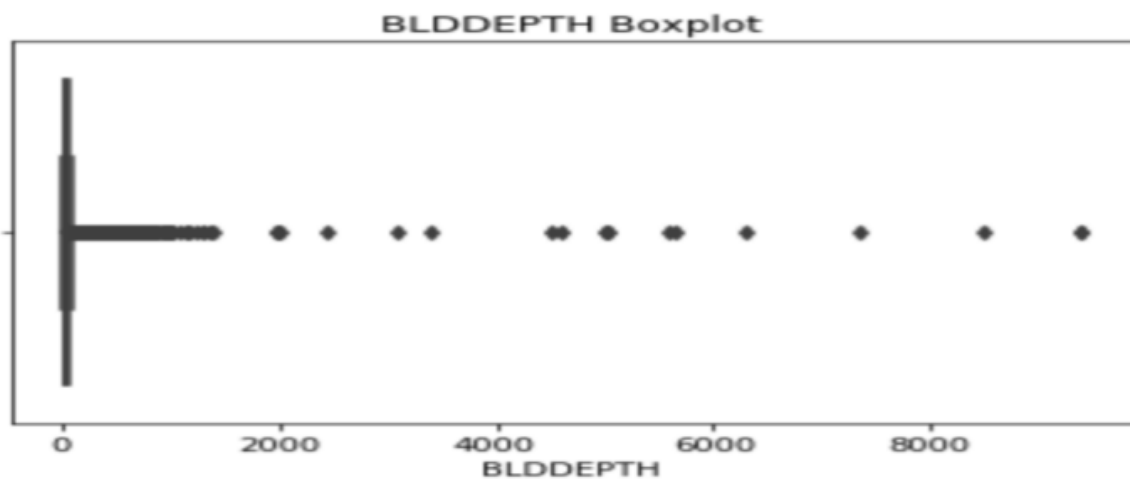
Minimum: 0

Maximum: 9393

Distribution:



Boxplot:



23. AVLAND2

Description: Numerical variable

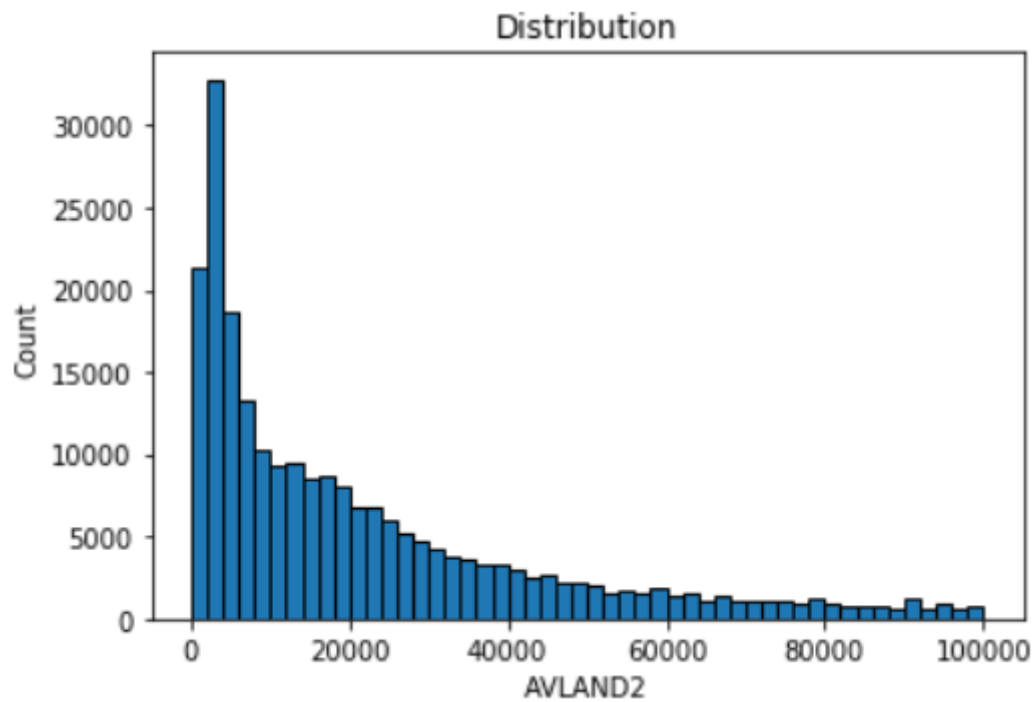
Percent Populated: 26.79%

Number of Unique Values: 58170

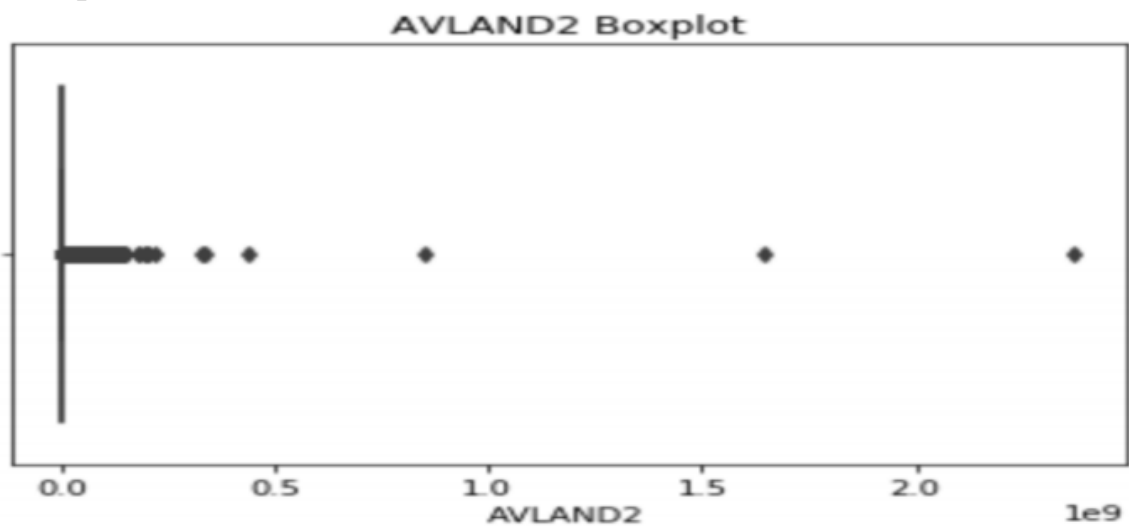
Minimum: 3

Maximum: 2,371,005,000

Distribution:



Boxplot:



24. AVTOT2

Description: Numerical variable

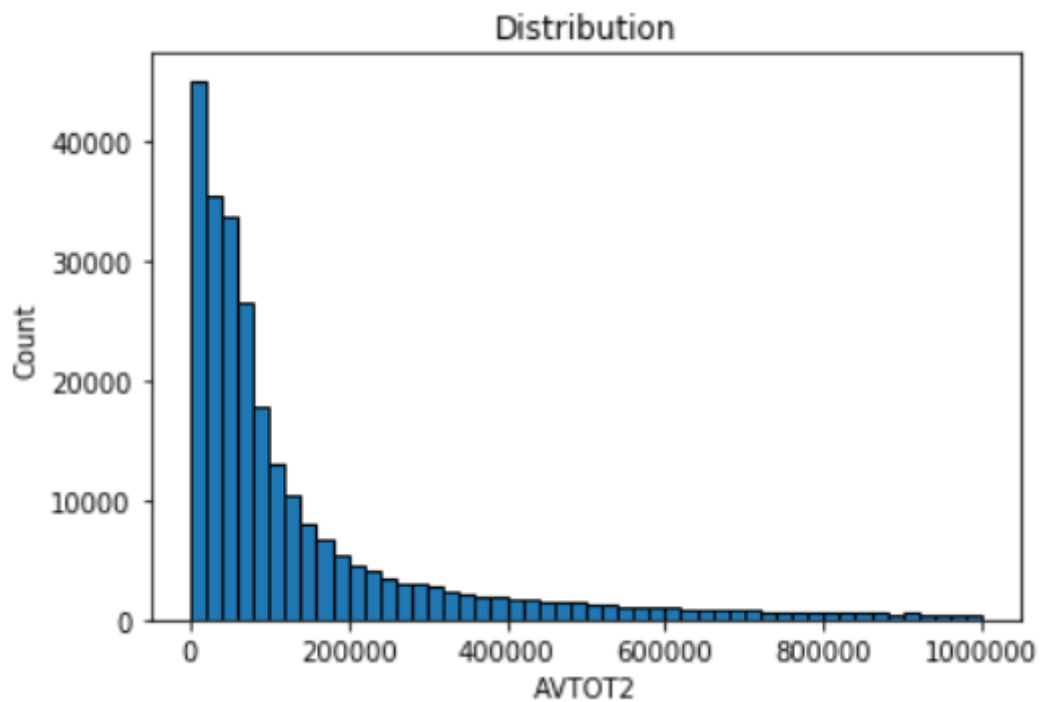
Percent Populated: 26.79%

Number of Unique Values: 110891

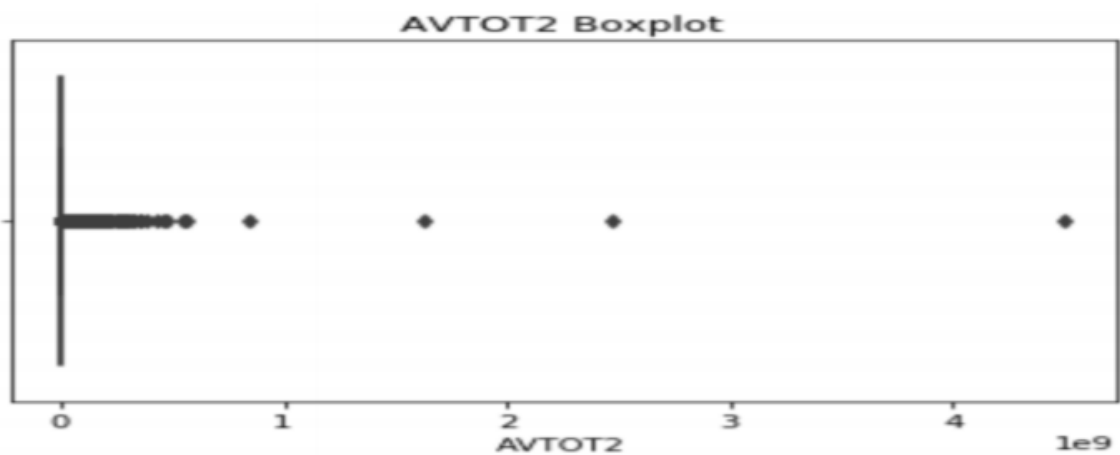
Minimum: 3

Maximum: 4,501,180,002

Distribution:



Boxplot:



25. EXLAND2

Description: Numerical variable

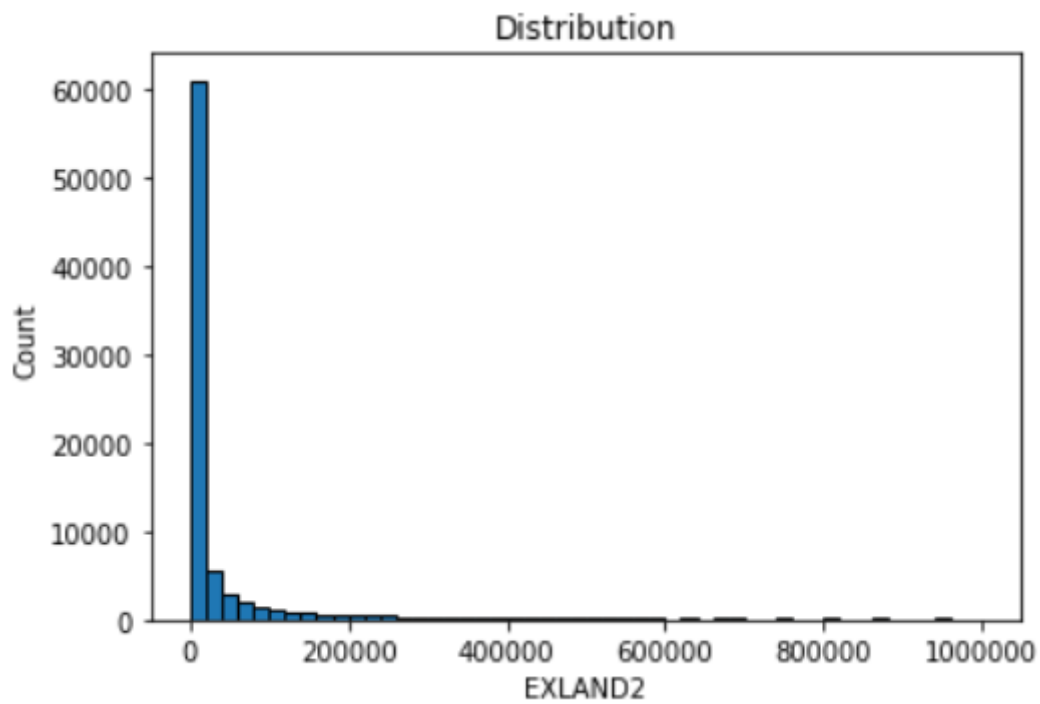
Percent Populated: 8.26%

Number of Unique Values: 21997

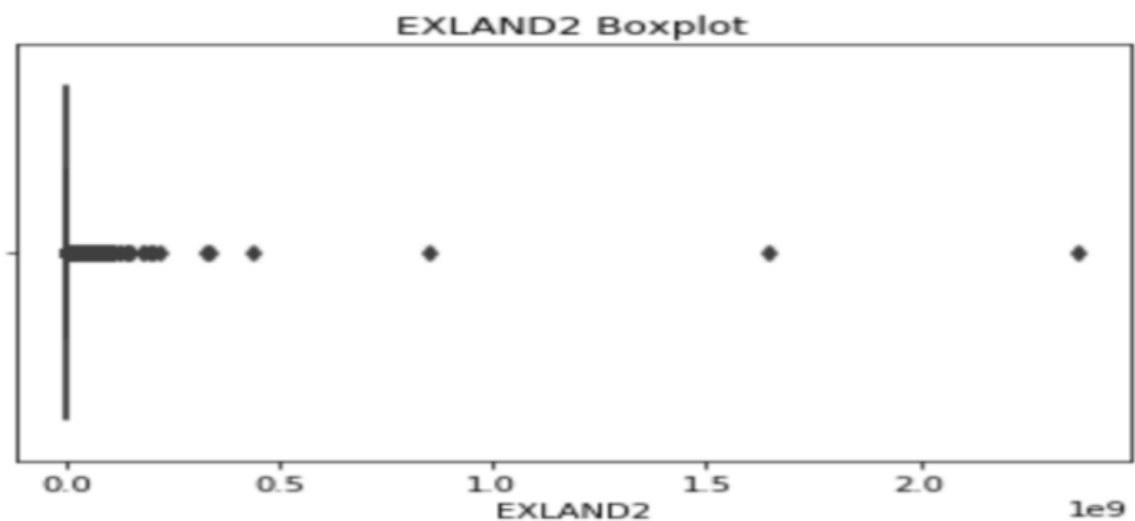
Minimum: 1

Maximum: 2,371,005,000

Distribution:



Boxplot:



26. EXTOT2

Description: Numerical variable

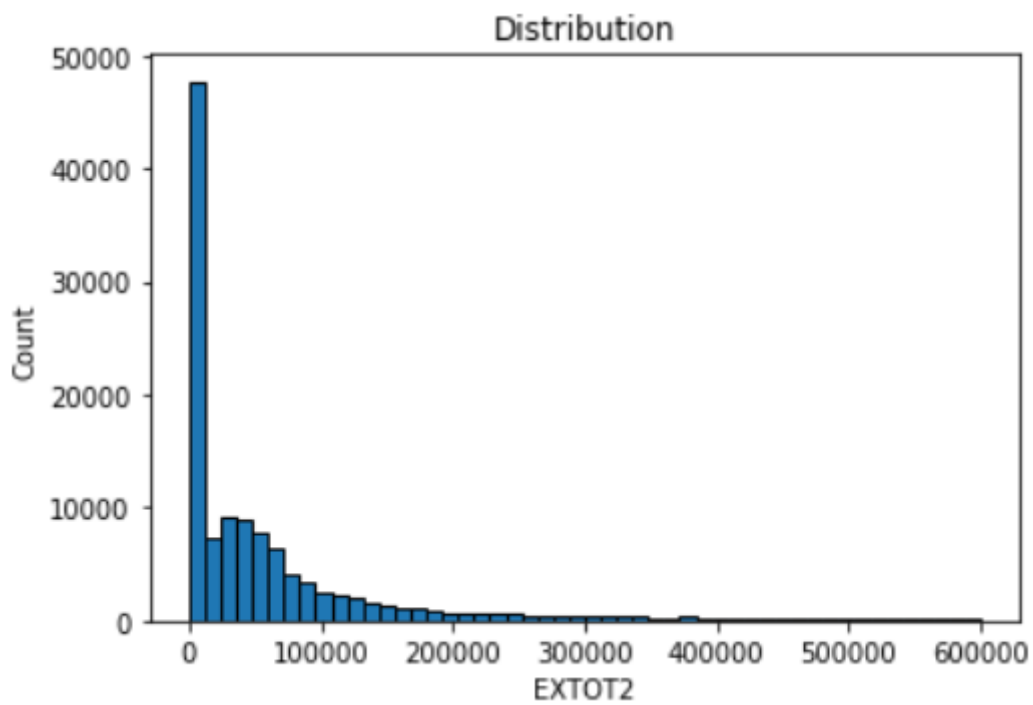
Percent Populated: 12.39%

Number of Unique Values: 48107

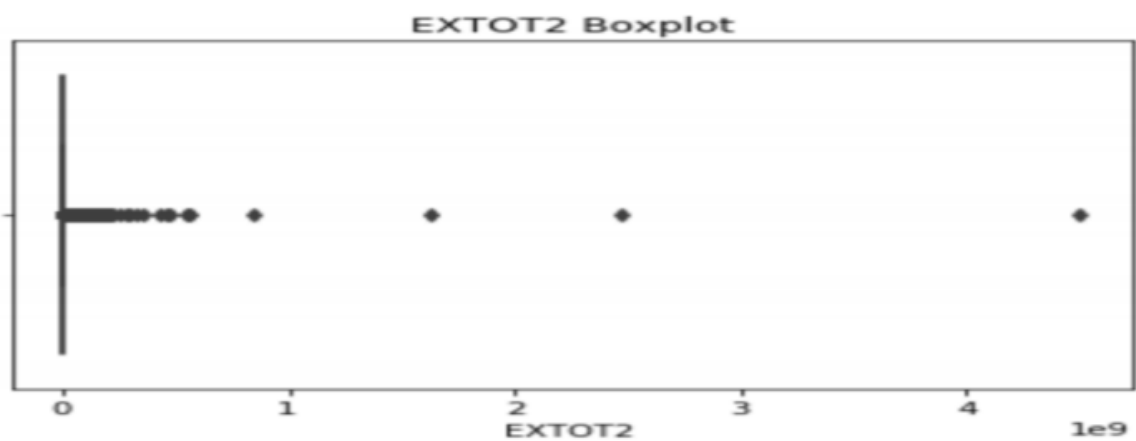
Minimum: 7

Maximum: 4,501,180,002

Distribution:



Boxplot:



27. EXCD2

Description: Numerical variable

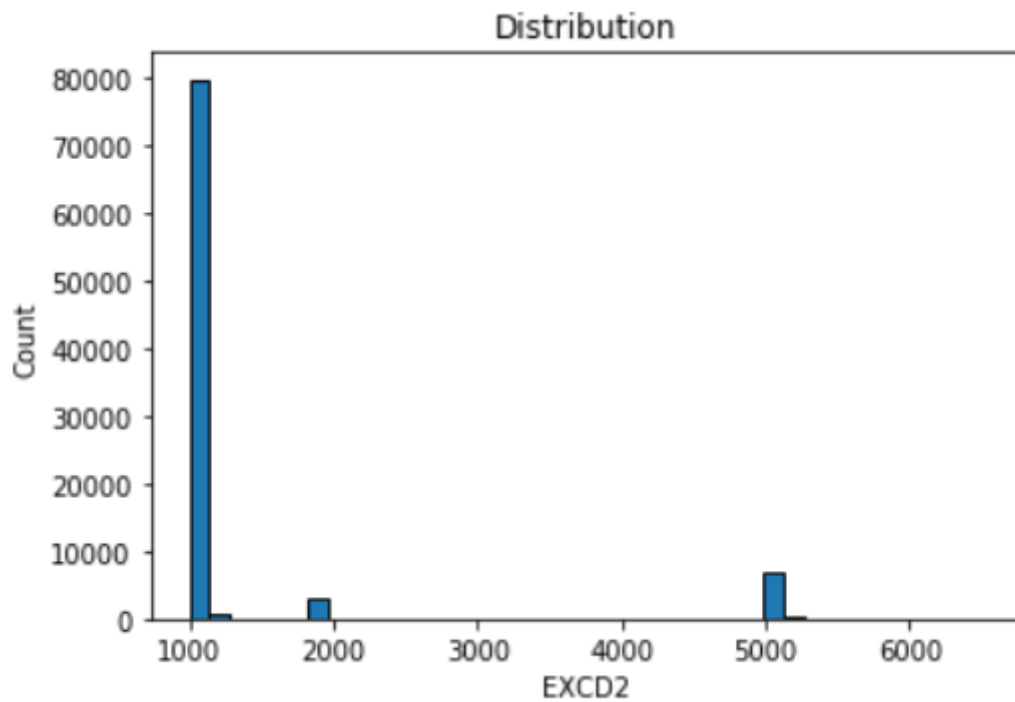
Percent Populated: 8.67%

Number of Unique Values: 62

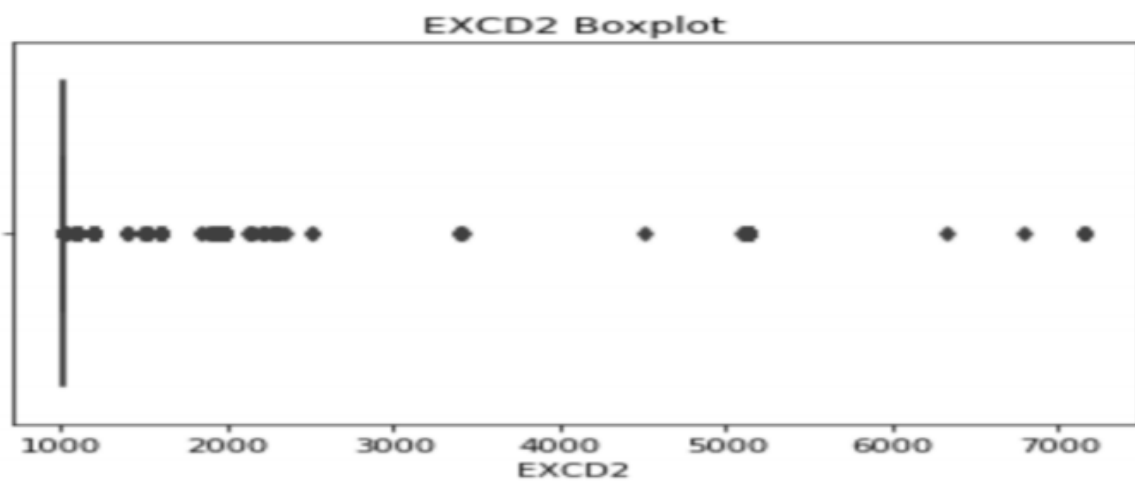
Minimum: 1011

Maximum: 7160

Distribution:



Boxplot:



28. PERIOD

Description: Categorical

Percent Populated: 100%

Number of Unique Values: 1

29. YEAR

Description: Categorical

Percent Populated: 100%

Number of Unique Values: 1

30. VALTYPE

Description: Categorical

Percent Populated: 100%

Number of Unique Values: 1