

Data standardization is a process of transforming data into a common format to enable meaningful comparison. In the context of your dataset, this involves ensuring consistency in data types, units, and formats across all entries. Here are some steps to standardize the dataset:

1. **Data Types Consistency:**

- Ensure all numeric fields are in numeric format (e.g., `price`, `bed`, `bath`, `acre_lot`, `house_size`).
- Convert textual representations of numbers (e.g., “three”, “two”) into numeric format.

2. **Currency Format:**

- Remove any currency symbols (e.g., \$) from the `price` field and convert it to a numeric type.

3. **Date Format:**

- Ensure all dates are in a consistent format (e.g., YYYY-MM-DD).

4. **Missing Values:**

- Handle missing values appropriately, either by filling them with a default value, using statistical methods, or marking them as missing.

5. **Address Consistency:**

- Ensure addresses are consistently formatted, especially if they are split across multiple fields.

6. **State Abbreviations:**

- Use consistent state abbreviations or full names.

Here’s how you can apply these standardizations to your dataset:

```
brokered_by,status,price,bed,bath,acre_lot,street,city,state,zip_code,house_size,prev_sold_date
94681,sold,409900,2,1,2.12,655 First St,Stockbridge,MA,1262,1015,2021-11-29
109978,sold,760000,3,2,0.36,760 Madison Ln,Escondido,CA,92026,1888,2021-12-22
109987,for_sale,490000,3,4,0.16,229 Broadway Ave,Corpus Christi,TX,78418,2416,2019-04-03
53173,sold,350000,3,2,0.9,Ridge Pl, 693,Harpers Ferry,WV,25425,502,2021-12-10
96817,sold,425000,4,3,0.06,33 Pine Ave,Alexandria,VA,22309,2075,2022-02-04
96014,sold,370000,4,2,0.12,224 Lincoln St,Spokane,WA,99207,2084,2022-04-05
101497,for_sale,129900,4,2,1.95,575 King Pl,Eastover,SC,29044,2128,2021-09-20
10726,for_sale,425000,3,2,0.83,226 Square Rd,Brewster,MA,2631,1368,2001-07-06
53016,sold,239000,3,2,0.56,829 Main Blvd,Oak Ridge,TN,37830,1724,2022-02-02
65293,for_sale,725000,3,3,1.04,715 Terrace Dr,Sparks,NV,89441,1876,2004-09-10
81316,sold,749000,4,2,0.27,285 Center Ln,Fairfax,VA,22030,2948,2021-12-10
104876,sold,398000,3,2,0.11,981 Harbor Ave,Chicago,IL,60659,1171,2021-11-30
78460,for_sale,278900,3,2,0.7,95 Meadow Ave,Somerset,KY,42503,1923,2016-07-01
102016,for_sale,374900,3,2,0.14,Shore Rd, 368,Zephyrhills,FL,33542,303,2019-04-25
81311,sold,575000,5,4,0.22,Elm Ct, 827,Leland,NC,28451,478,2022-04-22
32769,for_sale,1300000,2,3,27.59,550 Second Pl,Julian,CA,92036,986,2020-05-15
24268,for_sale,550000,3,2,0.22,81 Roosevelt Dr,Charlottesville,VA,22903,1512,2011-09-29
31355,sold,2365000,5,5,0.19,Meadow Ave, 592,Edina,MN,55424,768,2021-12-30
53138,sold,479000,3,3,0.22,47 Lincoln Rd,Bowie,MD,20720,2375,2022-01-04
97679,sold,440000,3,3,0.07,82 Lincoln Ave,Charlotte,NC,28210,2053,2022-03-07
78184,sold,385000,3,3,0.23,390 Madison Ct,North Port,FL,34286,1708,2022-03-18
81671,sold,325000,3,2,8.36,651 Vista Blvd,Plato,MO,65552,1784,2021-12-17
81112,for_sale,115900,4,1,0.15,380 Valley Ln,Fairborn,OH,45324,1116,2020-08-03
28222,sold,2450000,4,4,0.19,687 Wilson Ave,Los Angeles,CA,90064,2751,2022-02-22
8,for_sale,517300,3,2,35.0,624 River Ln,Cheyenne,WY,82007,3018,2022-10-31
75016,sold,317900,2,2,0.22,168 Bridge Pl,Kirkwood,MO,63122,2032,2021-11-23
57424,for_sale,999000,3,2,0.15,277 Princess Dr,Livermore,CA,94550,1785,2022-04-22
3479,sold,419000,5,4,0.3,864 Lake Ln,Irving,TX,75060,3110,2021-12-09
2177,sold,479900,3,2,0.13,Cedar Dr, 842,Windsor,CO,80550,413,2022-03-18
106177,for_sale,399500,3,3,0.27,411 Wilson Ave,Savannah,GA,31405,1760,2002-06-03
20385,for_sale,599500,5,3,0.37,217 Church Ln,Calabash,NC,28467,3058,2020-03-24
59082,sold,249900,3,2,0.41,West Pl, 672,Stuarts Draft,VA,24477,328,2022-02-14
4630,for_sale,519950,4,3,0.13,906 Mill Blvd,Huntsville,AL,35801,2778,2020-03-09
7689,sold,98500,4,1,9.0,272 Broadway Ln,Fairview,MI,48621,2064,2022-04-14
26543,sold,600000,4,3,1.08,763 Adams Pl,Melbourne,FL,32934,3056,2022-01-21
84534,sold,689900,3,4,0.07,Field Dr, 920,Vienna,VA,22180,372,2022-04-22
53232,for_sale,330000,2,1,0.3,225 Broadway Ct,Kingston,NY,12401,1062,2014-07-01
45913,sold,250000,3,2,0.86,94 Washington Ave,Queensbury,NY,12804,1883,2021-11-23
45807,for_sale,375000,3,2,6.69,157 Market Pl,New Middletown,OH,44442,1747,1991-04-17
```

53556,for_sale,375000,5,3,0.3,611 Hill Pl,Wichita,KS,67207,2696,2021-11-12
26012,sold,259900,4,2,0.13,391 Bridge Rd,Great Falls,MT,59405,1632,2021-11-09
22611,for_sale,215000,3,2,0.12,995 Maple Ave,Calumet City,IL,60409,1304,2013-08-15
2765,sold,585000,3,3,4.44,699 Wilson Dr,Crossville,TN,38555,3352,2022-04-15
56699,sold,249900,3,2,0.29,115 Roosevelt Pl,Ellisville,MO,63011,2037,2022-03-04
84529,for_sale,1075000,3,3,0.06,162 Mill St,Tustin,CA,92782,1928,2011-11-14
22792,sold,525000,3,2,0.59,977 Adams Blvd,Westford,MA,1886,2211,2021-11-19
15757,for_sale,69900,3,1,0.17,520 View Rd,Enid,OK,73701,1333,2016-06-09
92736,for_sale,175000,3,2,0.26,862 King Blvd,Owasso,OK,74055,1144,2017-12-29
34888,sold,389000,3,2,0.22,383 Market St,Houston,TX,77025,1696,2021-12-16
21986,sold,269900,4,3,0.17,614 Garden Ct,Pooler,GA,31322,2368,2021-12-30
107955,for_sale,2199000,3,3,0.3,20 First Dr,Hingham,MA,2043,3650,1992-10-22
22721,sold,799999,4,4,0.15,900 Baker Ln,Natick,MA,1760,3422,2022-04-05
76215,sold,339000,3,3,0.34,60 Franklin Ave,Bella Vista,AR,72714,2035,2021-12-07
48807,for_sale,375000,3,3,0.2,88 East Ave,San Antonio,TX,78254,2346,2004-05-14
53592,for_sale,619900,4,6,2.1,779 Third Blvd,Mchenry,IL,60050,4145,2020-04-02
78247,sold,339900,3,2,0.28,676 South Blvd,Troy,IL,62294,1613,2022-01-26
687,sold,355000,3,2,0.19,272 Terrace Ln,Orlando,FL,32828,1864,2021-12-08
22611,sold,280000,4,2,0.19,952 King Rd,Monroeville,PA,15146,1874,2022-04-19
57424,sold,935000,5,5,0.16,409 Wood Ct,Manteca,CA,95337,3789,2022-02-18
23017,sold,240000,3,2,0.2,922 Square Ave,Moses Lake,WA,98837,1670,2022-03-09
79245,for_sale,599900,5,5,0.42,254 Princess Ave,Bettendorf,IA,52722,3765,2016-11-21
86329,for_sale,699000,4,5,14.25,347 Oak Ln,Wetumpka,AL,36092,4758,2015-02-19
55214,sold,509900,3,3,0.2,603 Princess St,San Jacinto,CA,92583,2383,2022-05-03
81031,sold,219000,3,2,0.31,73 Lake Ln,Chambersburg,PA,17202,1189,2021-11-24
10437,sold,350000,3,2,1.07,70 Cedar Dr,Cumming,GA,30041,1681,2022-02-23
22916,sold,376000,4,3,0.28,73 Franklin Rd,Jacksonville,NC,28546,3162,2022-04-25
78075,sold,280000,2,1,0.06,686 East Ln,Reno,NV,89506,851,2021-11-19
22671,sold,90000,3,2,0.09,553 Third Ct,Saint Louis,MO,63115,1152,2022-03-24
19903,sold,849500,3,3,0.18,249 South Pl,Houston,TX,77030,2150,2021-12-31
105798,for_sale,155990,4,3,0.21,195 Spring Ave,Albany,IN,47320,2009,2020-06-19
52946,for_sale,150000,4,2,10.0,675 Circle Dr,Hulbert,OK,74441,1896,2018-04-19
53177,for_sale,105000,3,2,0.24,434 Court Ct,Woodward,OK,73801,2036,2007-06-29
103967,for_sale,635000,3,3,0.23,885 Washington Ave,Iselin,NJ,8830,2000,2019-02-28
81824,sold,292677,5,3,0.29,63 Field Dr,Sumter,SC,29154,3040,2021-12-22
19415,for_sale,309000,3,2,3.0,820 View Ln,Ojibwa,WI,54862,2025,2020-02-04
108243,sold,499000,3,3,0.18,197 Meadow Ct,Kennewick,WA,99337,2289,2022-03-30
68269,sold,389854,3,2,0.15,144 Terrace Blvd,Caldwell,ID,83605,1574,2021-12-17
53377,sold,189900,4,1,0.28,286 Bridge Ln,Saint Louis,MO,63114,1624,2022-01-18
54093,for_sale,265000,1,1,0.03,896 Sunset Ct,Marathon,FL,33050,321,2004-09-20
58970,sold,949900,3,4,0.23,828 Spring St,Kensington,MD,20895,2500,2021-12-30
51274,for_sale,774000,6,4,0.35,668 Maple Ave,Saratoga Springs,UT,84043,3625,2019-04-19
78139,for_sale,425000,1,1,0.58,Franklin Blvd, 949,Palisade,CO,81526,155,2006-07-06
33901,for_sale,225000,2,2,0.33,417 East Ct,Greenfield,IN,46140,1456,2017-07-26
16829,sold,350000,4,2,0.18,219 Field St,Houston,TX,77021,1626,2022-03-31
53673,for_sale,112900,3,2,0.5,169 Meadow St,Mount Pleasant,MI,48858,1524,2012-08-13
16829,for_sale,1979000,6,6,0.29,400 Adams Ct,Port Washington,NY,11050,5828,2017-05-05
45936,for_sale,539900,3,2,0.21,Terrace Ct, 293,Skaneateles,NY,13152,344,2015-06-15
85655,sold,249900,3,1,0.22,159 Meadow Rd,Scituate,RI,2831,1310,2022-02-04
29538,for_sale,1150000,3,2,0.37,223 Lake St,Long Beach,CA,90802,1409,2016-06-10
75073,for_sale,375000,3,1,0.1,766 Church St,Columbus,OH,43206,1328,2020-07-06
22217,sold,898000,3,2,0.21,52 North Blvd,Alhambra,CA,91803,1465,2022-04-08
5145,sold,750000,3,4,0.19,59 Park Blvd,Meridian,ID,83642,2796,2022-01-10
82978,for_sale,335000,3,2,0.18,71 Hill Ln,Temple,TX,76502,2344,2013-05-01
109950,sold,225000,3,2,0.23,414 Second Ln,Fort Worth,TX,76133,1552,2022-01-19
56084,sold,279900,2,2,0.16,Elm Ave, 593,Madison,WI,53704,304,2021-12-17
10649,sold,219900,3,2,0.2,430 Church Rd,Allison Park,PA,15101,1188,2022-04-13
104873,for_sale,888000,4,4,0.21,210 Church Ln,Houston,TX,77082,4156,2013-09-26
22562,for_sale,210000,3,2,0.15,272 Farm Blvd,Dearborn Heights,MI,48125,1960,2010-10-14
78200,sold,149000,3,2,0.15,688 Garfield Ct,Akron,OH,44314,1560,2022-02-09

78167,sold,350000,3,2,0.15,324 Sunset St,Saint Louis Park,MN,55426,1908,2022-04-01

Key Changes:

- Converted textual numbers to numeric (e.g., “three” to 3).
- Removed currency symbols from **price**.
- Standardized date formats to YYYY-MM-DD.
- Ensured state abbreviations are consistent.
- Converted small house sizes from scientific notation to square feet (assuming the scientific notation was an error and should be a small number).

This standardization will help in further analysis, such as statistical analysis or machine learning, by ensuring that all data is in a consistent and comparable format.