```
-- This SQL program demonstrates data cleaning using SQL
queries.

select * from housing;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/database/sql&db=SQLProject2

Showing rows 0 – 24 (1997 total, Query took 0.0003 seconds.)

UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	O
2045	007 00 0 125.00	SINGLE FAMILY	1808 FOX CHASE DR, GOODLETTSVILLE	9-Apr-13	240000	20130412-0036474	No	F
16918	007 00 0 130.00	SINGLE FAMILY	1832 FOX CHASE DR, GOODLETTSVILLE	10-Jun-14	366000	20140619-0053768	No	B
54582	007 00 0 138.00	SINGLE FAMILY	1864 FOX CHASE DR, GOODLETTSVILLE	26-Sep-16	435000	20160927-0101718	No	W
43070	007 00 0 143.00	SINGLE FAMILY	1853 FOX CHASE DR, GOODLETTSVILLE	29-Jan-16	255000	20160129-0008913	No	B
22714	007 00 0 149.00	SINGLE FAMILY	1829 FOX CHASE DR, GOODLETTSVILLE	10-Oct-14	278000	20141015-0095255	No	P
18367	007 00 0 151.00	SINGLE FAMILY	1821 FOX CHASE DR, GOODLETTSVILLE	16-Jul-14	267000	20140718-0063802	No	F
19804	007 14 0 002.00	SINGLE FAMILY	2005 SADIE LN, GOODLETTSVILLE	28-Aug-14	171000	20140903-0080214	No	H
54583	007 14 0 024.00	SINGLE FAMILY	1917 GRACELAND DR, GOODLETTSVILLE	27-Sep-16	262000	20161005-0105441	No	B
36500	007 14 0 026.00	SINGLE FAMILY	1428 SPRINGFIELD HWY, GOODLETTSVILLE	14-Aug-15	285000	20150819-0083440	No	F
19805	007 14 0 034.00	SINGLE FAMILY	1420 SPRINGFIELD HWY, GOODLETTSVILLE	29-Aug-14	340000	20140909-0082348	No	L
29467	007 14 0A 024.00	SINGLE FAMILY	2209 KAYLA DR, GOODLETTSVILLE	14-Apr-15	425000	20150415-0033442	No	
10754	007 14 0A 027.00	SINGLE FAMILY	109 BAILEY VIEW CT, GOODLETTSVILLE	12-Dec-13	585000	20131227-0130352	No	
34751	007 14 0B 010.00	RESIDENTIAL CONDO	1900 TINNIN RD, GOODLETTSVILLE	13-Jul-15	190000	20150717-0069947	No	
4512	007 15 0 002.00	SINGLE FAMILY	629 GAYLEMORE DR, GOODLETTSVILLE	7-Jun-13	189900	20130612-0059715	No	U
16919	007 15 0 003.00	SINGLE FAMILY	633 GAYLEMORE DR, GOODLETTSVILLE	30-Jun-14	157500	20140702-0058050	No	S
16920	007 15 0 004.00	SINGLE FAMILY	637 GAYLEMORE DR, GOODLETTSVILLE	30-Jun-14	247400	20140630-0057267	No	M
51967	007 15 0 008.00	SINGLE FAMILY	1976 SADIE LN, GOODLETTSVILLE	15-Jul-16	211500	20160720-0074793	No	M
28155	007 15 0 014.00	SINGLE FAMILY	644 GAYLEMORE DR, GOODLETTSVILLE	31-Mar-15	185900	20150402-0029022	No	S
8899	007 15 0 020.00	SINGLE FAMILY	1921 NORMERLE DR, GOODLETTSVILLE	11-Oct-13	349900	20131018-0109102	No	V
4513	007 15 0 031.00	SINGLE FAMILY	1916 NORMERLE DR, GOODLETTSVILLE	28-Jun-13	192500	20130711-0071698	No	P
27161	007 15 0 044.00	SINGLE FAMILY	2050 GRACELAND DR, GOODLETTSVILLE	10-Feb-15	279900	20150212-0012993	No	A
46859	007 15 0 048.00	SINGLE FAMILY	2034 GRACELAND DR, GOODLETTSVILLE	14-Apr-16	379900	20160418-0036715	No	D
5802	007 15 0 052.00	SINGLE FAMILY	811 BENTON CT, GOODLETTSVILLE	8-Jul-13	192500	20130712-0072376	No	B
36501	010 00 0 045.00	SINGLE FAMILY	331 VIEW RIDGE DR, GOODLETTSVILLE	31-Aug-15	193500	20150903-0090036	No	B
8900	010 00 0 052.00	SINGLE FAMILY	361 VIEW RIDGE DR, GOODLETTSVILLE	21-Oct-13	172400	20131030-0112723	No	W

```
-- Standardize date formate
select SaleDate, str_to_date(SaleDate,'%d-%M-%Y') from housing;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 24 (1997 total, Query took 0.0003 seconds.)

SaleDate	str_to_date(SaleDate,'%d-%M-%Y')
9-Apr-13	2013-04-09
10-Jun-14	2014-06-10
26-Sep-16	2016-09-26
29-Jan-16	2016-01-29
10-Oct-14	2014-10-10
16-Jul-14	2014-07-16
28-Aug-14	2014-08-28
27-Sep-16	2016-09-27
14-Aug-15	2015-08-14
29-Aug-14	2014-08-29
14-Apr-15	2015-04-14
12-Dec-13	2013-12-12
13-Jul-15	2015-07-13
7-Jun-13	2013-06-07
30-Jun-14	2014-06-30
30-Jun-14	2014-06-30
15-Jul-16	2016-07-15
31-Mar-15	2015-03-31
11-Oct-13	2013-10-11
28-Jun-13	2013-06-28
10-Feb-15	2015-02-10
14-Apr-16	2016-04-14
8-Jul-13	2013-07-08
31-Aug-15	2015-08-31
21-Oct-13	2013-10-21

```sql
update housing
set SaleDate = str_to_date(SaleDate,'%d-%M-%Y');

-- Add standardize date formate to a new column
alter table housing
add SaleDateConverted date;

update housing
set SaleDateConverted = str_to_date(SaleDate,'%d-%M-%Y');
```

```sql
-- Populate property address data
select * from housing
where PropertyAddress = ''
order by ParcelID;

select a.ParcelID, a.PropertyAddress, b.ParcelID,
b.PropertyAddress
from housing as a
join housing as b
on a.ParcelID = b.ParcelID
and a.UniqueID != b.UniqueID
where a.PropertyAddress = '';
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 16 (17 total, Query took 1.1071 seconds.)

ParcelID	PropertyAddress	ParcelID	PropertyAddress
025 07 0 031.00		025 07 0 031.00	410 ROSEHILL CT, GOODLETTSVILLE
026 01 0 069.00		026 01 0 069.00	141 TWO MILE PIKE, GOODLETTSVILLE
026 05 0 017.00		026 05 0 017.00	208 EAST AVE, GOODLETTSVILLE
026 06 0A 038.00		026 06 0A 038.00	109 CANTON CT, GOODLETTSVILLE
033 06 0 041.00		033 06 0 041.00	1129 CAMPBELL RD, GOODLETTSVILLE
033 06 0A 002.00		033 06 0A 002.00	1116 CAMPBELL RD, GOODLETTSVILLE
033 15 0 123.00		033 15 0 123.00	438 W CAMPBELL RD, GOODLETTSVILLE
034 03 0 059.00		034 03 0 059.00	2117 PAULA DR, MADISON
034 03 0 059.00		034 03 0 059.00	2117 PAULA DR, MADISON
034 07 0B 015.00		034 07 0B 015.00	2524 VAL MARIE DR, MADISON
034 07 0B 015.00		034 07 0B 015.00	2524 VAL MARIE DR, MADISON
034 07 0B 015.00		034 07 0B 015.00	2524 VAL MARIE DR, MADISON
034 16 0A 004.00		034 16 0A 004.00	213 WARREN CT, OLD HICKORY
041 03 0A 100.00		041 03 0A 100.00	1289 GOODMORNING DR, NASHVILLE
042 13 0 075.00		042 13 0 075.00	222 FOXBORO DR, MADISON
043 04 0 014.00		043 04 0 014.00	112 HILLER DR, OLD HICKORY
043 09 0 074.00		043 09 0 074.00	213 B LOVELL ST, MADISON

```
update housing as a
join housing as b
on a.ParcelID = b.ParcelID
and a.UniqueID != b.UniqueID
set a.PropertyAddress = b.PropertyAddress
where a.PropertyAddress = '';
```

```
-- Break out property address into individual columns (address,
city)
select PropertyAddress from housing;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 24 (1997 total, Query took 0.0002 seconds.)

PropertyAddress
1808 FOX CHASE DR, GOODLETTSVILLE
1832 FOX CHASE DR, GOODLETTSVILLE
1864 FOX CHASE DR, GOODLETTSVILLE
1853 FOX CHASE DR, GOODLETTSVILLE
1829 FOX CHASE DR, GOODLETTSVILLE
1821 FOX CHASE DR, GOODLETTSVILLE
2005 SADIE LN, GOODLETTSVILLE
1917 GRACELAND DR, GOODLETTSVILLE
1428 SPRINGFIELD HWY, GOODLETTSVILLE
1420 SPRINGFIELD HWY, GOODLETTSVILLE
2209 KAYLA DR, GOODLETTSVILLE
109 BAILEY VIEW CT, GOODLETTSVILLE
1900 TINNIN RD, GOODLETTSVILLE
629 GAYLEMORE DR, GOODLETTSVILLE
633 GAYLEMORE DR, GOODLETTSVILLE
637 GAYLEMORE DR, GOODLETTSVILLE
1976 SADIE LN, GOODLETTSVILLE
644 GAYLEMORE DR, GOODLETTSVILLE
1921 NORMERLE DR, GOODLETTSVILLE
1916 NORMERLE DR, GOODLETTSVILLE
2050 GRACELAND DR, GOODLETTSVILLE
2034 GRACELAND DR, GOODLETTSVILLE
811 BENTON CT, GOODLETTSVILLE
331 VIEW RIDGE DR, GOODLETTSVILLE
361 VIEW RIDGE DR, GOODLETTSVILLE

```
select substring(PropertyAddress, 1, instr(PropertyAddress, ',')
- 1) as address,
substring(PropertyAddress, instr(PropertyAddress, ',') + 1) as
city
from housing;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 24 (1997 total, Query took 0.0003 seconds.)

address	city
1808 FOX CHASE DR	GOODLETTSVILLE
1832 FOX CHASE DR	GOODLETTSVILLE
1864 FOX CHASE DR	GOODLETTSVILLE
1853 FOX CHASE DR	GOODLETTSVILLE
1829 FOX CHASE DR	GOODLETTSVILLE
1821 FOX CHASE DR	GOODLETTSVILLE
2005 SADIE LN	GOODLETTSVILLE
1917 GRACELAND DR	GOODLETTSVILLE
1428 SPRINGFIELD HWY	GOODLETTSVILLE
1420 SPRINGFIELD HWY	GOODLETTSVILLE
2209 KAYLA DR	GOODLETTSVILLE
109 BAILEY VIEW CT	GOODLETTSVILLE
1900 TINNIN RD	GOODLETTSVILLE
629 GAYLEMORE DR	GOODLETTSVILLE
633 GAYLEMORE DR	GOODLETTSVILLE
637 GAYLEMORE DR	GOODLETTSVILLE
1976 SADIE LN	GOODLETTSVILLE
644 GAYLEMORE DR	GOODLETTSVILLE
1921 NORMERLE DR	GOODLETTSVILLE
1916 NORMERLE DR	GOODLETTSVILLE
2050 GRACELAND DR	GOODLETTSVILLE
2034 GRACELAND DR	GOODLETTSVILLE
811 BENTON CT	GOODLETTSVILLE
331 VIEW RIDGE DR	GOODLETTSVILLE
361 VIEW RIDGE DR	GOODLETTSVILLE

```
alter table housing
add PropertySplitAddress varchar(255);

update housing
set PropertySplitAddress = substring(PropertyAddress, 1,
instr(PropertyAddress, ',') - 1);

alter table housing
add PropertySplitCity varchar(255);

update housing
set PropertySplitCity = substring(PropertyAddress,
instr(PropertyAddress, ',') + 1);
```

```
-- Break out owner address into individual columns (address,
city, state)
select OwnerAddress from housing;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 24 (1997 total, Query took 0.0002 seconds.)

OwnerAddress
1808 FOX CHASE DR, GOODLETTSVILLE, TN
1832 FOX CHASE DR, GOODLETTSVILLE, TN
1864 FOX CHASE DR, GOODLETTSVILLE, TN
1853 FOX CHASE DR, GOODLETTSVILLE, TN
1829 FOX CHASE DR, GOODLETTSVILLE, TN
1821 FOX CHASE DR, GOODLETTSVILLE, TN
2005 SADIE LN, GOODLETTSVILLE, TN
1917 GRACELAND DR, GOODLETTSVILLE, TN
1428 SPRINGFIELD HWY, GOODLETTSVILLE, TN
1420 SPRINGFIELD HWY, GOODLETTSVILLE, TN
629 GAYLEMORE DR, GOODLETTSVILLE, TN
633 GAYLEMORE DR, GOODLETTSVILLE, TN
637 GAYLEMORE DR, GOODLETTSVILLE, TN
2001 SADIE LN, GOODLETTSVILLE, TN
644 GAYLEMORE DR, GOODLETTSVILLE, TN
1921 NORMERLE DR, GOODLETTSVILLE, TN
1916 NORMERLE DR, GOODLETTSVILLE, TN
2050 GRACELAND DR, GOODLETTSVILLE, TN
2034 GRACELAND DR, GOODLETTSVILLE, TN
811 BENTON CT, GOODLETTSVILLE, TN
331 VIEW RIDGE DR, GOODLETTSVILLE, TN
361 VIEW RIDGE DR, GOODLETTSVILLE, TN

```
select substring_index(OwnerAddress, ',', 1),
substring_index(substring_index(OwnerAddress, ',', -2), ',', 1),
substring_index(OwnerAddress, ',', -1)
from housing;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 – 24 (1997 total, Query took 0.0978 seconds.)

substring_index(OwnerAddress, ',', 1)	substring_index(substring_index(OwnerAddress, ',', -2), ',', 1)	substring_index(OwnerAddress, ',', -1)
1808 FOX CHASE DR	GOODLETTSVILLE	TN
1832 FOX CHASE DR	GOODLETTSVILLE	TN
1864 FOX CHASE DR	GOODLETTSVILLE	TN
1853 FOX CHASE DR	GOODLETTSVILLE	TN
1829 FOX CHASE DR	GOODLETTSVILLE	TN
1821 FOX CHASE DR	GOODLETTSVILLE	TN
2005 SADIE LN	GOODLETTSVILLE	TN
1917 GRACELAND DR	GOODLETTSVILLE	TN
1428 SPRINGFIELD HWY	GOODLETTSVILLE	TN
1420 SPRINGFIELD HWY	GOODLETTSVILLE	TN
629 GAYLEMORE DR	GOODLETTSVILLE	TN
633 GAYLEMORE DR	GOODLETTSVILLE	TN
637 GAYLEMORE DR	GOODLETTSVILLE	TN
2001 SADIE LN	GOODLETTSVILLE	TN
644 GAYLEMORE DR	GOODLETTSVILLE	TN
1921 NORMERLE DR	GOODLETTSVILLE	TN
1916 NORMERLE DR	GOODLETTSVILLE	TN
2050 GRACELAND DR	GOODLETTSVILLE	TN
2034 GRACELAND DR	GOODLETTSVILLE	TN
811 BENTON CT	GOODLETTSVILLE	TN
331 VIEW RIDGE DR	GOODLETTSVILLE	TN
361 VIEW RIDGE DR	GOODLETTSVILLE	TN

```
alter table housing
add OwnerSplitAddress varchar(255);

update housing
set OwnerSplitAddress = substring_index(OwnerAddress, ',', 1);

alter table housing
add OwnerSplitCity varchar(255);

update housing
set OwnerSplitCity =
substring_index(substring_index(OwnerAddress, ',', -2), ',', 1);

alter table housing
add OwnerSplitState varchar(255);

update housing
set OwnerSplitState = substring_index(OwnerAddress, ',', -1);
```

```
-- Change Y and N to Yes and No in SoldAsVacant field
select distinct(SoldAsVacant), count(SoldAsVacant) from housing
group by SoldAsVacant
order by 2;
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 5 (6 total, Query took 0.0034 seconds.)

SoldAsVacant	count(SoldAsVacant)
144 SCENIC VIEW RD, OLD HICKORY, TN	1
142 SCENIC VIEW RD, OLD HICKORY, TN	1
Y	2
N	26
Yes	265
No	1702

```
select SoldAsVacant,
    case when SoldAsVacant = 'Y' then 'Yes'
         when SoldAsVacant = 'N' then 'No'
         else SoldAsVacant
    end
from housing
where SoldAsVacant = 'Y' or SoldAsVacant = 'N';
```

---

localhost/SQLProject2/housing/ http://localhost/phpmyadmin/index.php?route=/table/sql&db=SQLProject2&table=housing

Showing rows 0 - 24 (28 total, Query took 0.0301 seconds.)

SoldAsVacant	case when SoldAsVacant = 'Y' then 'Yes' when SoldAsVacant = 'N' then 'No' else SoldAsVacant end
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
N	No
Y	Yes
N	No
N	No
N	No
Y	Yes
N	No

```sql
update housing
set SoldAsVacant = case when SoldAsVacant = 'Y' then 'Yes'
        when SoldAsVacant = 'N' then 'No'
        else SoldAsVacant
    end;

-- Remove unused columns
alter table housing
drop column PropertyAddress;
```