



SQL

Data Cleaning

Nashville Housing Dataset

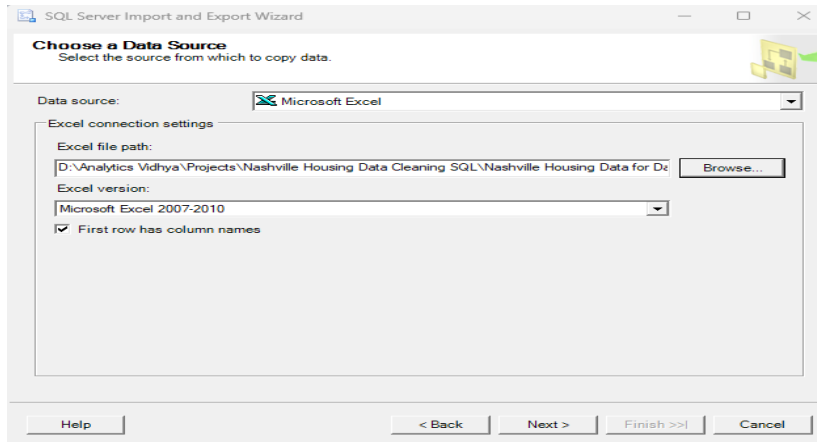
Steps Performed

- Standardised Date Format
- Populating Address NULLs
- Extracted Address, City & State from Single Column
 - Standardised Abbreviations used
 - Removed Duplicates

Developed & programmed by:

Rithik Tripathi

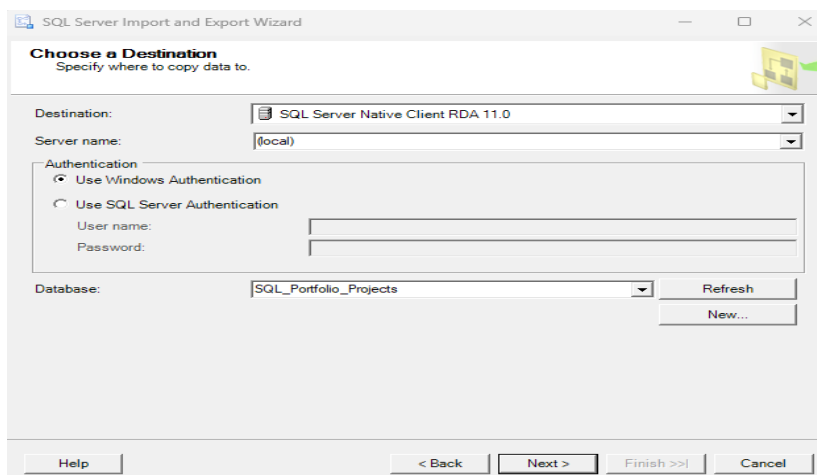
Importing Data in SQL Studio from Excel



Open *SQL Server Import and Export Wizard*

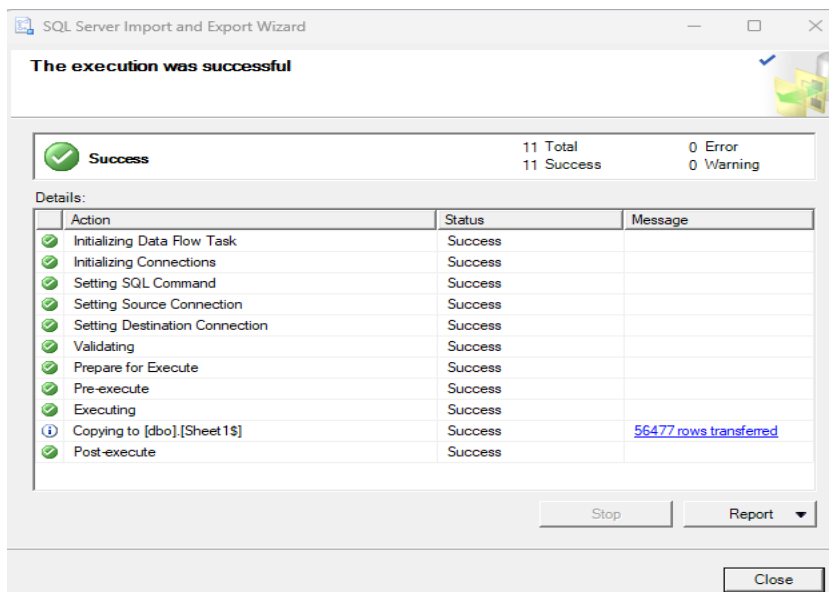
Select *Data Source* as Microsoft Excel

Select *File Path* & *Proceed*



Choose Destination as *SQL Server Native Client* & *Server Name*

Select *Database* in which Data is to be loaded & *Proceed*



Wait for Process to complete and Data will be Loaded Successfully.

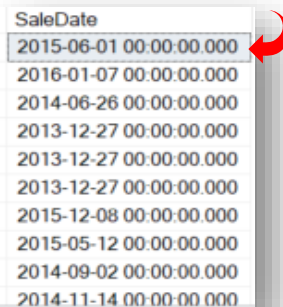
Glance at Data

UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant
33159	063 00 0 012.00	SINGLE FAMILY	1604 NEELYS BEND RD, MADISON	2015-06-01 00:00:00.000	133500	20150604-0052251	No
43175	063 00 0 017.00	CHURCH	1700 NEELYS BEND RD, MADISON	2016-01-07 00:00:00.000	152500	20160108-0002037	No
17007	063 00 0 045.00	SINGLE FAMILY	1606 HUDSON RD, MADISON	2014-06-26 00:00:00.000	130000	20140701-0057509	No
10836	063 00 0 060.00	GREENBELT	1833 HUDSON RD, MADISON	2013-12-27 00:00:00.000	309535	20131231-0131263	No
10837	063 00 0 061.00	GREENBELT	1829 HUDSON RD, MADISON	2013-12-27 00:00:00.000	309535	20131231-0131263	No
10838	063 00 0 062.00	GREENBELT	1823 HUDSON RD, MADISON	2013-12-27 00:00:00.000	309535	20131231-0131263	No
42021	063 00 0 067.00	SINGLE FAMILY	1725 HUDSON RD, MADISON	2015-12-08 00:00:00.000	90000	20151209-0124105	No
31298	063 00 0 069.00	SINGLE FAMILY	1717 HUDSON RD, MADISON	2015-05-12 00:00:00.000	229900	20150522-0047605	No
21392	063 00 0 072.00	SINGLE FAMILY	1633 PAWNEE TRL, MADISON	2014-09-02 00:00:00.000	200000	20140904-0080899	No
23968	063 00 0 112.00	SINGLE FAMILY	1610 PAWNEE TRL, MADISON	2014-11-14 00:00:00.000	295000	20141117-0105769	No
42022	063 00 0 114.00	VACANT RESIDENTIAL LAND	1805 HUDSON RD, MADISON	2015-12-18 00:00:00.000	107000	20151222-0128630	No
4601	063 00 0 114.00	RESIDENTIAL COMBO/MISC	1805 HUDSON RD, MADISON	2013-06-18 00:00:00.000	126000	20130619-0062825	No
4602	063 00 0 167.00	VACANT RES LAND	1430 NEELYS BEND RD, MADISON	2013-06-24 00:00:00.000	89900	20130702-0068068	No
38183	063 00 0 174.00	SINGLE FAMILY	653 HIDDEN ACRES DR, MADISON	2015-09-28 00:00:00.000	360000	20151001-0099846	No
9860	063 00 0 178.00	SINGLE FAMILY	514 MENEES LN, MADISON	2013-11-08 00:00:00.000	250000	20131114-0117632	No
50729	063 00 0 224.00	SINGLE FAMILY	1548 NEELYS BEND RD, MADISON	2016-06-22 00:00:00.000	425000	20160722-0075958	No
4603	063 00 0 229.00	SINGLE FAMILY	1731 HUDSON RD, MADISON	2013-06-14 00:00:00.000	475000	20130618-0062386	No

OwnerName	OwnerAddress	Acreage	TaxDistrict	LandValue	BuildingValue	TotalValue	YearBuilt	Bedrooms
BLIER, DAVID A. & JANICE C.	1604 NEELYS BEND RD, MADISON, TN	5.01	GENERAL SERVICES DISTRICT	99100	31100	142900	1935	1
LASKA, LEWIS L.	1700 NEELYS BEND RD, MADISON, TN	5.15	GENERAL SERVICES DISTRICT	101000	260400	361400	1999	0
GREENE, FRANKLIN J. & PETERSON, PAMELA J.	1606 HUDSON RD, MADISON, TN	2.2	GENERAL SERVICES DISTRICT	61200	81500	142700	1935	3
MANNING, DAVID W. & BETSY B.	1831 HUDSON RD, MADISON, TN	7.46	GENERAL SERVICES DISTRICT	9933	0	9933	NULL	NULL
MANNING, DAVID W. & BETSY B.	1825 HUDSON RD, MADISON, TN	7.7	GENERAL SERVICES DISTRICT	11024	0	75024	NULL	NULL
MANNING, DAVID W. & BETSY B.	1819 HUDSON RD, MADISON, TN	7.75	GENERAL SERVICES DISTRICT	12819	0	12819	NULL	NULL
LANIER, DAVID LUDWELL & YERUSALEM ABEBE	1725 HUDSON RD, MADISON, TN	1.65	GENERAL SERVICES DISTRICT	56300	116000	172300	1962	4
THURSTON, LOUIS J. & RITA K.	1717 HUDSON RD, MADISON, TN	2	GENERAL SERVICES DISTRICT	63500	160000	243400	1971	3
BROOKS, FREDRICK R. & ELIZABETH A.	1633 PAWNEE TRL, MADISON, TN	1.38	GENERAL SERVICES DISTRICT	45000	169000	216900	1969	4
SHAW, IAN A. & NIGRELLI, GINA N.	1610 PAWNEE TRL, MADISON, TN	3	GENERAL SERVICES DISTRICT	56100	172600	236800	1976	3
ANDERSON, WILLIAM & VICKI	1805 HUDSON RD, MADISON, TN	7	GENERAL SERVICES DISTRICT	117000	0	121400	NULL	NULL
ANDERSON, WILLIAM & VICKI	1805 HUDSON RD, MADISON, TN	7	GENERAL SERVICES DISTRICT	117000	0	121400	NULL	NULL
GRADDY, ANTHONY PAUL	1430 NEELYS BEND RD, MADISON, TN	8.36	GENERAL SERVICES DISTRICT	87800	0	87800	NULL	NULL
RICHARDSON, BONNIE	653 HIDDEN ACRES DR, MADISON, TN	1.24	GENERAL SERVICES DISTRICT	56300	189200	290900	1976	4
DELLA MORTE, RACHEL & ROBERT	514 MENEES LN, MADISON, TN	1.38	GENERAL SERVICES DISTRICT	56300	173900	230900	1987	3
ASPEN, CRAIG & FRAZZINI, CYNTHIA A.	1548 NEELYS BEND RD, MADISON, TN	3.27	GENERAL SERVICES DISTRICT	75600	242400	337800	1996	3
ARMSTRONG, MARK A. & TRINA B.	1731 HUDSON RD, MADISON, TN	11.74	GENERAL SERVICES DISTRICT	136300	417100	553400	1997	4

Standardising Date Format

```
SELECT SaleDate
FROM NashvilleHousing
```

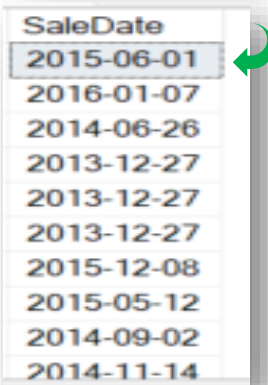


SaleDate
2015-06-01 00:00:00.000
2016-01-07 00:00:00.000
2014-06-26 00:00:00.000
2013-12-27 00:00:00.000
2013-12-27 00:00:00.000
2013-12-27 00:00:00.000
2015-12-08 00:00:00.000
2015-05-12 00:00:00.000
2014-09-02 00:00:00.000
2014-11-14 00:00:00.000

Data contains SaleDates in DateTime Format where Time is Not Required

So Altering and Updating Datatype to DATE

```
ALTER TABLE NashvilleHousing
ALTER COLUMN SaleDate DATE
```



SaleDate
2015-06-01
2016-01-07
2014-06-26
2013-12-27
2013-12-27
2013-12-27
2015-12-08
2015-05-12
2014-09-02
2014-11-14

Populating NULLs

```
SELECT *
FROM NashvilleHousing
WHERE PropertyAddress IS NULL
```

UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant
43076	025 07 0 031.00	SINGLE FAMILY	NULL	2016-01-15	179900	20160120-0005776	No
39432	026 01 0 069.00	VACANT RESIDENTIAL LAND	NULL	2015-10-23	153000	20151028-0109602	No
45290	026 05 0 017.00	SINGLE FAMILY	NULL	2016-03-29	155000	20160330-0029941	No
53147	026 06 0A 038.00	RESIDENTIAL CONDO	NULL	2016-08-25	144900	20160831-0091567	No
43080	033 06 0 041.00	SINGLE FAMILY	NULL	2016-01-04	170000	20160107-0001526	No
45295	033 06 0A 002.00	SINGLE FAMILY	NULL	2016-03-29	210000	20160331-0030709	No
48731	033 15 0 123.00	SINGLE FAMILY	NULL	2016-05-05	199900	20160506-0045368	No
8126	093 08 0 054.00	SINGLE FAMILY	NULL	2013-09-20	25000	20130923-0099666	N
36531	034 03 0 059.00	SINGLE FAMILY	NULL	2015-08-13	245000	20150819-0083759	No
46919	034 07 0B 015.00	VACANT RESIDENTIAL LAND	NULL	2016-04-27	40000	20160304-0020905	Yes
44264	034 16 0A 004.00	VACANT RESIDENTIAL LAND	NULL	2016-02-04	130000	20160205-0011327	Yes
45298	041 03 0A 100.00	SINGLE FAMILY	NULL	2016-03-18	170000	20160322-0027048	No

We observed that there are 29 NULL Values present.

We can populate the NULLs present if we can have some point of Reference to generate the address

```
SELECT *
FROM NashvilleHousing
ORDER BY ParcelID
```

UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	OwnerName
48717	015 14 0 045.00	SINGLE FAMILY	7515 GARY RD, JOELTON	2016-05-27	160000	20160606-0056507	No	PORTER, BRIAN KELLY
53117	015 14 0 053.00	SINGLE FAMILY	7620 GERALD DR, JOELTON	2016-08-08	195900	20160810-0083381	No	TEMPLE, LORENE
39427	015 14 0 054.00	SINGLE FAMILY	7624 GERALD DR, JOELTON	2015-10-09	122431	20151013-0103951	No	STEWART, DONNA D.
5803	015 14 0 055.00	SINGLE FAMILY	7628 GERALD DR, JOELTON	2013-07-12	119900	20130718-0074201	No	PROULX, JENNIFER L. & JESSICA
6993	015 14 0 060.00	SINGLE FAMILY	3113 MILLIKEN DR, JOELTON	2013-08-20	145000	20130821-0088382	No	SMITH, HOLLI NICHOLE
43071	015 14 0 060.00	SINGLE FAMILY	3113 MILLIKEN DR, JOELTON	2016-01-15	160000	20160119-0004923	No	SMITH, HOLLI NICHOLE
45286	015 14 0 061.00	SINGLE FAMILY	3117 MILLIKEN DR, JOELTON	2016-03-18	169900	20160413-0035454	No	DISMANG, JEREMY
3267	015 14 0 071.00	SINGLE FAMILY	3107 MARGIE DR, JOELTON	2013-05-10	103500	20130513-0047997	No	DOWNS, JESSICA
34753	015 14 0 075.00	SINGLE FAMILY	3124 MARGIE DR, JOELTON	2015-07-31	184000	20150803-0076727	No	MURRY, SHAWN
39428	016 00 0 093.00	SINGLE FAMILY	2466 UNION HILL RD, GOODLETTSVILLE	2015-10-13	165000	20151014-0104536	No	HARRISON, JASON C. & AMANDA D.
21298	016 00 0 110.00	SINGLE FAMILY	7221 APPLE VIEW DR, GOODLETTSVILLE	2014-09-24	290000	20141009-0093665	No	MILLER, WILLIAM B. & RETA A.
34754	016 00 0 143.00	SINGLE FAMILY	7104 LAMA TERRA DR, GOODLETTSVILLE	2015-07-06	260000	20150708-0066261	No	GIGEAR, DAVID & DESIREE
39429	016 00 0 184.00	SINGLE FAMILY	7143 LAMA TERRA DR, GOODLETTSVILLE	2015-10-09	200000	20151013-0103879	No	MORELOCK, WALTER L. JR & LISA D.

Notice here, ParcelId 015 14 0 060.00 is Occurring **more than once** & all respective details are varying but the **Address is same**.

With this analysis we can conclude that we can populate address if NULL from address of same other ParcelID

We will Join the table with *itself* in order to achieve the respective addresses where one is Null

```
SELECT a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress
FROM NashvilleHousing a
JOIN NashvilleHousing b
ON a.ParcelID = b.ParcelID
AND a.[UniqueID ] <> b.[UniqueID ] -- to avoid getting the same row again & again
WHERE a.PropertyAddress IS NULL
```

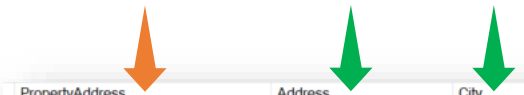
ParcelID	PropertyAddress	ParcelID	PropertyAddress
025 07 0 031.00	NULL	025 07 0 031.00	410 ROSEHILL CT, GOODLETTSVILLE
026 01 0 069.00	NULL	026 01 0 069.00	141 TWO MILE PIKE, GOODLETTSVILLE
026 05 0 017.00	NULL	026 05 0 017.00	208 EAST AVE, GOODLETTSVILLE
026 06 0A 038.00	NULL	026 06 0A 038.00	109 CANTON CT, GOODLETTSVILLE
033 06 0 041.00	NULL	033 06 0 041.00	1129 CAMPBELL RD, GOODLETTSVILLE
033 06 0A 002.00	NULL	033 06 0A 002.00	1116 CAMPBELL RD, GOODLETTSVILLE
033 15 0 123.00	NULL	033 15 0 123.00	438 W CAMPBELL RD, GOODLETTSVILLE
093 08 0 054.00	NULL	093 08 0 054.00	700 GLENVIEW DR, NASHVILLE
093 08 0 054.00	NULL	093 08 0 054.00	700 GLENVIEW DR, NASHVILLE
034 03 0 059.00	NULL	034 03 0 059.00	2117 PAULA DR, MADISON
034 03 0 059.00	NULL	034 03 0 059.00	2117 PAULA DR, MADISON
034 07 0B 015.00	NULL	034 07 0B 015.00	2524 VAL MARIE DR, MADISON
034 07 0B 015.00	NULL	034 07 0B 015.00	2524 VAL MARIE DR, MADISON
034 07 0B 015.00	NULL	034 07 0B 015.00	2524 VAL MARIE DR, MADISON
034 16 0A 004.00	NULL	034 16 0A 004.00	213 WARREN CT, OLD HICKORY
041 03 0A 100.00	NULL	041 03 0A 100.00	1289 GOODMORNING DR, NASHVILLE

Now Respective address of each parcel id is available & we just need to *Impute* them now.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

100


```
SELECT PropertyAddress, SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1) AS Address,
SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress)+1, LEN(PropertyAddress)) AS City
FROM NashvilleHousing
```



PropertyAddress	Address	City
1604 NEELYS BEND RD, MADISON	1604 NEELYS BEND RD	MADISON
1700 NEELYS BEND RD, MADISON	1700 NEELYS BEND RD	MADISON
1606 HUDSON RD, MADISON	1606 HUDSON RD	MADISON
1833 HUDSON RD, MADISON	1833 HUDSON RD	MADISON
1829 HUDSON RD, MADISON	1829 HUDSON RD	MADISON
1823 HUDSON RD, MADISON	1823 HUDSON RD	MADISON
1725 HUDSON RD, MADISON	1725 HUDSON RD	MADISON
1717 HUDSON RD, MADISON	1717 HUDSON RD	MADISON
1633 PAWNEE TRL, MADISON	1633 PAWNEE TRL	MADISON
1610 PAWNEE TRL, MADISON	1610 PAWNEE TRL	MADISON
1805 HUDSON RD, MADISON	1805 HUDSON RD	MADISON
1805 HUDSON RD, MADISON	1805 HUDSON RD	MADISON
1430 NEELYS BEND RD, MADISON	1430 NEELYS BEND RD	MADISON
653 HIDDEN ACRES DR, MADISON	653 HIDDEN ACRES DR	MADISON
514 MENEES LN, MADISON	514 MENEES LN	MADISON
1548 NEELYS BEND RD, MADISON	1548 NEELYS BEND RD	MADISON

So now we have Extracted Address & City from Property Address.

We need to Create Respective Columns & Update the Values

```
-- Adding Desired Column
ALTER TABLE NashvilleHousing
ADD PropertySplitAddress NVARCHAR(255)

UPDATE NashvilleHousing
SET PropertySplitAddress = SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1)

-- Adding Desired Column
ALTER TABLE NashvilleHousing
ADD PropertySplitCity NVARCHAR(255)

UPDATE NashvilleHousing
SET PropertySplitCity = SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress)+1, LEN(PropertyAddress))
```

If we observe, OwnerAddress contains 3 parts :

- Address
- City
- State

And all are separated by a ' , ' comma as a *Delimiter*. So using *PARSENAME()* we will extract both parts using the index positions dynamically.

But as since *PARSENAME()* used ' . ' by Default as a delimiter, so we have to replace ' , ' with ' . ' First.




OwnerAddress
1604 NEELYS BEND RD, MADISON, TN
1700 NEELYS BEND RD, MADISON, TN
1606 HUDSON RD, MADISON, TN
1831 HUDSON RD, MADISON, TN
1825 HUDSON RD, MADISON, TN
1819 HUDSON RD, MADISON, TN
1725 HUDSON RD, MADISON, TN
1717 HUDSON RD, MADISON, TN
1633 PAWNEE TRL, MADISON, TN
1610 PAWNEE TRL, MADISON, TN
1805 HUDSON RD, MADISON, TN
1805 HUDSON RD, MADISON, TN

```

SELECT OwnerAddress,
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1) AS State,
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2) AS City,
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3) AS address
FROM NashvilleHousing

```



OwnerAddress	State	City	address
1604 NEELYS BEND RD, MADISON, TN	TN	MADISON	1604 NEELYS BEND RD
1700 NEELYS BEND RD, MADISON, TN	TN	MADISON	1700 NEELYS BEND RD
1606 HUDSON RD, MADISON, TN	TN	MADISON	1606 HUDSON RD
1831 HUDSON RD, MADISON, TN	TN	MADISON	1831 HUDSON RD
1825 HUDSON RD, MADISON, TN	TN	MADISON	1825 HUDSON RD
1819 HUDSON RD, MADISON, TN	TN	MADISON	1819 HUDSON RD
1725 HUDSON RD, MADISON, TN	TN	MADISON	1725 HUDSON RD
1717 HUDSON RD, MADISON, TN	TN	MADISON	1717 HUDSON RD
1633 PAWNEE TRL, MADISON, TN	TN	MADISON	1633 PAWNEE TRL
1610 PAWNEE TRL, MADISON, TN	TN	MADISON	1610 PAWNEE TRL
1805 HUDSON RD, MADISON, TN	TN	MADISON	1805 HUDSON RD
1805 HUDSON RD, MADISON, TN	TN	MADISON	1805 HUDSON RD
1430 NEELYS BEND RD, MADISON, TN	TN	MADISON	1430 NEELYS BEND RD

So now we have Extracted Address, City & State from Owner Address.

We need to Create Respective Columns & Update the Values

```

ALTER TABLE NashvilleHousing
ADD OwnerSplitAddress NVARCHAR(255)

UPDATE NashvilleHousing
SET OwnerSplitAddress = PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3)

ALTER TABLE NashvilleHousing
ADD OwnerSplitCity NVARCHAR(255)

UPDATE NashvilleHousing
SET OwnerSplitCity = PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2)

ALTER TABLE NashvilleHousing
ADD OwnerSplitState NVARCHAR(255)

UPDATE NashvilleHousing
SET OwnerSplitState = PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1)

```

PropertySplitAddress	PropertySplitCity	OwnerSplitAddress	OwnerSplitCity	OwnerSplitState
1604 NEELYS BEND RD	MADISON	1604 NEELYS BEND RD	MADISON	TN
1700 NEELYS BEND RD	MADISON	1700 NEELYS BEND RD	MADISON	TN
1606 HUDSON RD	MADISON	1606 HUDSON RD	MADISON	TN
1833 HUDSON RD	MADISON	1831 HUDSON RD	MADISON	TN
1829 HUDSON RD	MADISON	1825 HUDSON RD	MADISON	TN
1823 HUDSON RD	MADISON	1725 HUDSON RD	MADISON	TN
1725 HUDSON RD	MADISON	1717 HUDSON RD	MADISON	TN
1717 HUDSON RD	MADISON	1633 PAWNEE TRL	MADISON	TN
1633 PAWNEE TRL	MADISON	1610 PAWNEE TRL	MADISON	TN
1610 PAWNEE TRL	MADISON	1805 HUDSON RD	MADISON	TN
1805 HUDSON RD	MADISON	1805 HUDSON RD	MADISON	TN
1805 HUDSON RD	MADISON	1430 NEELYS BEND RD	MADISON	TN
1430 NEELYS BEND RD	MADISON	653 HIDDEN ACRES DR	MADISON	TN
653 HIDDEN ACRES DR	MADISON	514 MENEES LN	MADISON	TN
514 MENEES LN	MADISON	1548 NEELYS BEND RD	MADISON	TN
1548 NEELYS BEND RD	MADISON	1731 HUDSON RD	MADISON	TN
1731 HUDSON RD	MADISON	1835 HUDSON RD	MADISON	TN
1835 HUDSON RD	MADISON	1542 HUDSON RD	MADISON	TN
1542 HUDSON RD	MADISON			

Standardising Abbreviations Used

We have observed that in SoldAsVacant column, Yes is sometimes mentioned as 'Y'
& No is sometimes mentioned as 'N'.

```
SELECT DISTINCT(SoldAsVacant), COUNT(SoldAsVacant) AS Count
FROM NashvilleHousing
GROUP BY SoldAsVacant
```

SoldAsVacant	Count
N	399
Yes	4623
Y	52
No	51403

We will obtain right values with a *CASE* statement

```
SELECT SoldAsVacant,
CASE
    WHEN SoldAsVacant = 'Y' THEN 'Yes'
    WHEN SoldAsVacant = 'N' THEN 'No'
    ELSE SoldAsVacant
END
FROM NashvilleHousing
```

SoldAsVacant	UpdatedValues
No	No
No	No
N	No
No	No
N	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No

Now, We need to update them in the Data.

```

UPDATE NashvilleHousing
SET SoldAsVacant =
CASE
    WHEN SoldAsVacant = 'Y' THEN 'Yes'
    WHEN SoldAsVacant = 'N' THEN 'No'
    ELSE SoldAsVacant
END

```

Reverifying the changes.

```

SELECT DISTINCT(SoldAsVacant), COUNT(SoldAsVacant) AS Count
FROM NashvilleHousing
GROUP BY SoldAsVacant

```

SoldAsVacant	Count
Yes	4675
No	51802

Removing Duplicates

Here, We will use *Row_NUMBER()* to allocate a unique row number to each record in the data.

Then we will create a CTE by partitioning over multiple columns and hence wherever row number comes more than 1, that record would be considered as a Duplicate.

```

WITH RowNum_CTE AS
(
    SELECT *, ROW_NUMBER() OVER(PARTITION BY ParcelID, PropertyAddress, SalePrice, SaleDate, LegalReference
                                ORDER BY UniqueID ) Rn
    FROM NashvilleHousing
)
SELECT * FROM RowNum_CTE
WHERE Rn > 1
ORDER BY PropertyAddress

```

UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant
27354	107 14 0 157.00	SINGLE FAMILY	1003 BRILEY PKWY, NASHVILLE	2015-02-12	80000	20150219-0014347	No
27383	150 04 0 144.00	SINGLE FAMILY	1005 ESTURBRIDGE CT, ANTIOCH	2015-02-13	180000	20150213-0013547	No
27362	119 16 0 084.00	DUPLEX	1006 DRUMMOND DR, NASHVILLE	2015-02-17	80000	20150218-0014086	No
27132	091 08 0 056.00	SINGLE FAMILY	1011 43RD AVE N, NASHVILLE	2015-02-23	69900	20150227-0017102	No
27368	135 10 0E 020.00	RESIDENTIAL CONDO	1023 CAPITAL FUNDS CT, NASHVILLE	2015-02-06	118981	20150210-0012380	No
27369	135 10 0E 021.00	RESIDENTIAL CONDO	1025 CAPITAL FUNDS CT, NASHVILLE	2015-02-06	118981	20150210-0012389	No
27123	091 04 0 046.00	SINGLE FAMILY	1038 40TH AVE N, NASHVILLE	2015-02-23	40000	20150227-0016978	No
27351	098 09 0B 009.00	SINGLE FAMILY	108 WINDCHASE RUN, HERMITAGE	2015-02-18	210000	20150225-0016199	No
27367	135 10 0E 014.00	RESIDENTIAL CONDO	1129 RANSOM WAY, NASHVILLE	2015-02-13	118900	20150224-0015807	No
27350	098 06 0B 137.00	RESIDENTIAL CONDO	1207 CHICKADEE CIR, HERMITAGE	2015-02-26	297000	20150227-0017053	No

These are the Duplucates found. We will go ahead and delete them from data. And this marks our end for the Nashville Data cleaning in SQL.

```

WITH RowNum_CTE AS
(
    SELECT *, ROW_NUMBER() OVER(PARTITION BY ParcelID, PropertyAddress, SalePrice, SaleDate, LegalReference
                                ORDER BY UniqueID ) Rn
    FROM NashvilleHousing
)
DELETE FROM RowNum_CTE
WHERE Rn > 1

```

Thank You