

# Data Cleaning Using SQL

# Standardize Date Format

PropertyAddress	SaleDate
1208 3RD AVE S, NASHVILLE	2013-01-24 00:00:00.000
1802 STEWART PL, NASHVILLE	2013-01-11 00:00:00.000
2761 ROSEDALE PL, NASHVILLE	2013-01-18 00:00:00.000
224 PEACHTREE ST, NASHVILLE	2013-01-18 00:00:00.000
316 LUTIE ST, NASHVILLE	2013-01-23 00:00:00.000
2626 FOSTER AVE, NASHVILLE	2013-01-04 00:00:00.000
104 PRESCOTT PL, NASHVILLE	2013-01-07 00:00:00.000
501 MORTON AVE, NASHVILLE	2013-01-15 00:00:00.000
501 MORTON AVE, NASHVILLE	2013-01-25 00:00:00.000
113 NEESE DR, NASHVILLE	2013-01-09 00:00:00.000
184 WHEELER AVE, NASHVILLE	2013-01-18 00:00:00.000
238 ELYSIAN FIELDS RD, NASHVILLE	2013-01-11 00:00:00.000
3801 ECKHART DR, NASHVILLE	2013-01-04 00:00:00.000
4832 CORNING DR, NASHVILLE	2013-01-23 00:00:00.000
285 PARAGON MILLS RD, NASHVILLE	2013-01-18 00:00:00.000
259 EISENHOWER DR, NASHVILLE	2013-01-17 00:00:00.000
5100 OVERTON RD, NASHVILLE	2013-01-04 00:00:00.000
370 WALLACE RD, NASHVILLE	2013-01-29 00:00:00.000



SaleDateConverted
2013-01-24
2013-01-11
2013-01-18
2013-01-18
2013-01-23
2013-01-04
2013-01-07
2013-01-15
2013-01-25
2013-01-09
2013-01-18
2013-01-11
2013-01-04
2013-01-23
2013-01-18
2013-01-17
2013-01-04
2013-01-29

## Code

```
alter table [Project 2].dbo.Nashville
add SaleDateConverted Date;

Update [Project 2].dbo.Nashville
set SaleDateConverted = convert(date, SaleDate)
```

# Breaking Address into Individual Columns(Address, State, City)

	PropertyAddress
1	1208 3RD AVE S, NASHVILLE
2	1802 STEWART PL, NASHVILLE
3	2761 ROSEDALE PL, NASHVILLE
4	224 PEACHTREE ST, NASHVILLE
5	316 LUTIE ST, NASHVILLE
6	2626 FOSTER AVE, NASHVILLE
7	104 PRESCOTT PL, NASHVILLE
8	501 MORTON AVE, NASHVILLE
9	501 MORTON AVE, NASHVILLE
10	113 NEESE DR, NASHVILLE
11	184 WHEELER AVE, NASHVILLE
12	238 ELYSIAN FIELDS RD, NASHVILLE
13	3801 ECKHART DR, NASHVILLE
14	4832 CORNING DR, NASHVILLE
15	285 PARAGON MILLS RD, NASHVILLE
16	259 EISENHOWER DR, NASHVILLE
17	5100 OVERTON RD, NASHVILLE
18	370 WALLACE RD, NASHVILLE
19	109 NORTHCREST COMMONS CIR, NASHVILLE

	propertyAddressSplit	propertyAddressCity
1	1208 3RD AVE S	NASHVILLE
2	1802 STEWART PL	NASHVILLE
3	2761 ROSEDALE PL	NASHVILLE
4	224 PEACHTREE ST	NASHVILLE
5	316 LUTIE ST	NASHVILLE
6	2626 FOSTER AVE	NASHVILLE
7	104 PRESCOTT PL	NASHVILLE
8	501 MORTON AVE	NASHVILLE
9	501 MORTON AVE	NASHVILLE
10	113 NEESE DR	NASHVILLE
11	184 WHEELER AVE	NASHVILLE
12	238 ELYSIAN FIELDS RD	NASHVILLE
13	3801 ECKHART DR	NASHVILLE
14	4832 CORNING DR	NASHVILLE
15	285 PARAGON MILLS RD	NASHVILLE
16	259 EISENHOWER DR	NASHVILLE
17	5100 OVERTON RD	NASHVILLE
18	370 WALLACE RD	NASHVILLE
19	109 NORTHCREST COMMONS CIR	NASHVILLE

## Code(Using SubString)

```
Select
SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1) as Address,
SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress)+1, LEN(PropertyAddress)) as City
from [Project 2].dbo.Nashville

Alter table [Project 2].dbo.Nashville
add propertyAddressSplit Nvarchar(255);

update [Project 2].dbo.Nashville
set propertyAddressSplit = SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1)

Alter table [Project 2].dbo.Nashville
add propertyAddressCity Nvarchar(255);

update [Project 2].dbo.Nashville
set propertyAddressCity = SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress)+1, LEN(PropertyAddress))
```

# Breaking Address into Individual Columns(Address, State, City)

	OwnerAddress
1	NULL
2	1802 STEWART PL, NASHVILLE, TN
3	2761 ROSEDALE PL, NASHVILLE, TN
4	224 PEACHTREE ST, NASHVILLE, TN
5	316 LUTIE ST, NASHVILLE, TN
6	2626 FOSTER AVE, NASHVILLE, TN
7	NULL
8	501 MORTON AVE, NASHVILLE, TN
9	501 MORTON AVE, NASHVILLE, TN
10	113 NEESE DR, NASHVILLE, TN
11	184 WHEELER AVE, NASHVILLE, TN
12	238 ELYSIAN FIELDS RD, NASHVILLE, TN
13	3801 ECKHART DR, NASHVILLE, TN
14	4832 CORNING DR, NASHVILLE, TN
15	285 PARAGON MILLS RD, NASHVILLE, TN
16	259 EISENHOWER DR, NASHVILLE, TN
17	5100 OVERTON RD, NASHVILLE, TN
18	NULL
19	NULL

	OwnerAddressSplit	OwnerCitySplit	OwnerStateSplit
1	NULL	NULL	NULL
2	1802 STEWART PL	NASHVILLE	TN
3	2761 ROSEDALE PL	NASHVILLE	TN
4	224 PEACHTREE ST	NASHVILLE	TN
5	316 LUTIE ST	NASHVILLE	TN
6	2626 FOSTER AVE	NASHVILLE	TN
7	NULL	NULL	NULL
8	501 MORTON AVE	NASHVILLE	TN
9	501 MORTON AVE	NASHVILLE	TN
10	113 NEESE DR	NASHVILLE	TN
11	184 WHEELER AVE	NASHVILLE	TN
12	238 ELYSIAN FIELDS RD	NASHVILLE	TN
13	3801 ECKHART DR	NASHVILLE	TN
14	4832 CORNING DR	NASHVILLE	TN
15	285 PARAGON MILLS RD	NASHVILLE	TN
16	259 EISENHOWER DR	NASHVILLE	TN
17	5100 OVERTON RD	NASHVILLE	TN
18	NULL	NULL	NULL
19	NULL	NULL	NULL

## Code(Using ParseName)

```

Select
PARSENAME(Replace(OwnerAddress, ',', '.'), 3) as Address, --ParseName
PARSENAME(Replace(OwnerAddress, ',', '.'), 2) as city,    -- ParseName
PARSENAME(Replace(OwnerAddress, ',', '.'), 1) as State
from [Project 2].dbo.Nashville

Alter table [Project 2].dbo.Nashville
add OwnerAddressSplit nvarchar(255)

update [Project 2].dbo.Nashville
set OwnerAddressSplit = PARSENAME(Replace(OwnerAddress, ',', '.'), 3)

Alter table [Project 2].dbo.Nashville
add OwnerCitySplit nvarchar(255)

update [Project 2].dbo.Nashville
set OwnerCitySplit = PARSENAME(Replace(OwnerAddress, ',', '.'), 2)

Alter table [Project 2].dbo.Nashville
add OwnerStateSplit nvarchar(255)

update [Project 2].dbo.Nashville
set OwnerStateSplit = PARSENAME(Replace(OwnerAddress, ',', '.'), 1)

```

# Removing Duplicates(Using CTE Method)

Code(Using CTE Method)

```
}with rownumcte as(  
_Select *,  
    ROW_NUMBER() over (  
        PARTITION BY ParcelID,  
                    PropertyAddressSplit,  
                    PropertyAddressCity,  
                    SalePrice,  
                    LegalReference  
        ORDER BY  
            UniqueID  
        ) Row_num  
_from [Project 2].dbo.Nashville)  
--order by ParcelId  
  
}delete  
_from rownumcte  
_where row_num > 1
```