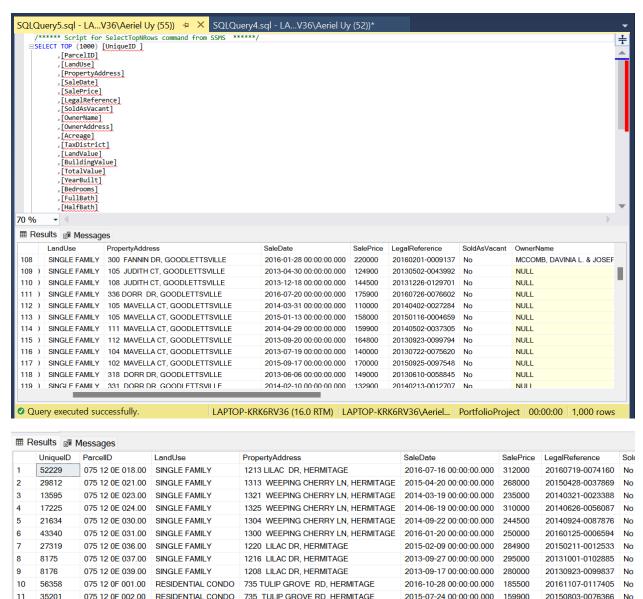
Nashville Housing Data Cleaning

Imported Data to Microsoft SQL

Objective: Data Cleaning

12

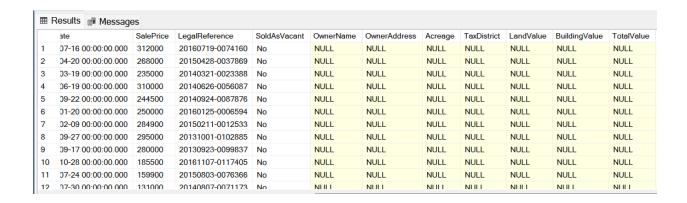
18702



2014-07-30 00:00:00 000 131000

20140807-0071173 No

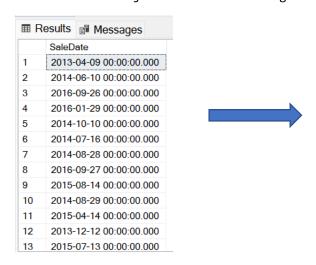
075 12 0F 005 00 RESIDENTIAL CONDO 735 TULIP GROVE RD HERMITAGE

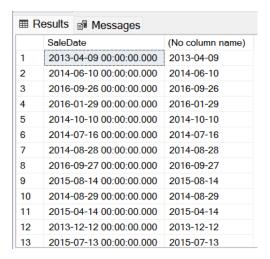


Tasks

Standardized Date Format on SaleDate Column showing the date and time

Select SaleDate, CONVERT(Date, SaleDate)
From PortfolioProject..NashvilleHousing





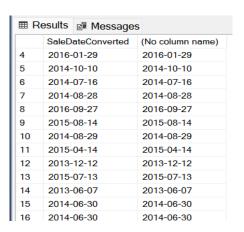
Update the data with New Column named 'SaleDateConverted'

ALTER TABLE was used as the previous **UPDATE** did not work and changed the column name to SaleDateConverted to be the new column. SaleDate column to be deleted.

```
Update NashvilleHousing
SET SaleDate = CONVERT(Date, SaleDate)

ALTER TABLE NashvilleHousing
Add SaleDateConverted Date;

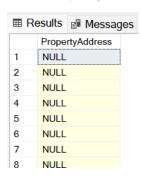
Update NashvilleHousing
SET SaleDateConverted = CONVERT(Date, SaleDate)
```



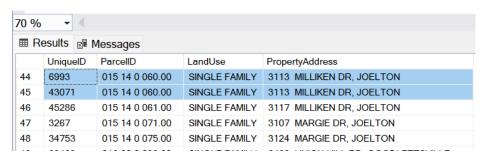
Second Task

Populate Property Address Data

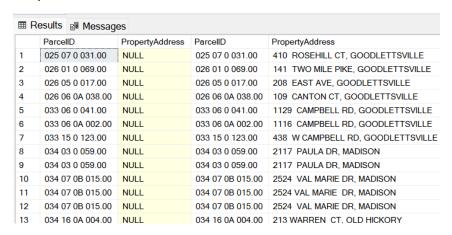
Select PropertyAddress
From PortfolioProject.dbo.NashvilleHousing
Where PropertyAddress is null



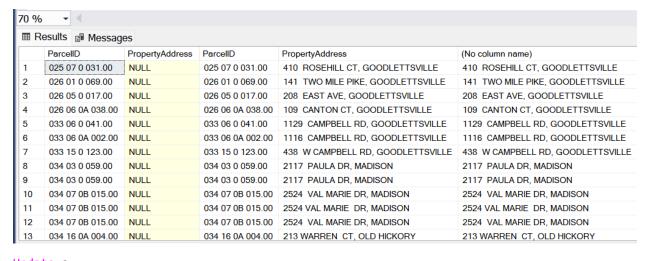
To populate the Null values, the ParcellD was used as it shows that if it is the same ParcellD then the PropertyAddress is the same.



We use JOIN to populate the property address. Alias were used such as 'a' and 'b' for the table name. Query shows to populate the address with the same parcel ID but not the same unique ID.





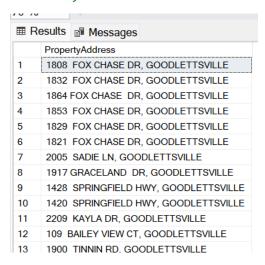


Where a.PropertyAddress is null

Third Task

Separate the addresses to individual columns (Address, City, State)

Select PropertyAddress
From PortfolioProject.dbo.NashvilleHousing
--Where PropertyAddress is null
--order by ParcelID

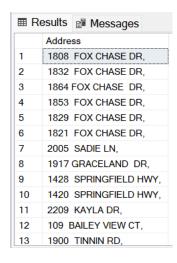


SELECT

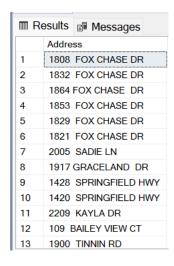
SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)) as Address From PortfolioProject..NashvilleHousing

SELECT

SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1) as Address From PortfolioProject..NashvilleHousing







SELECT

SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress) -1) as Address
, SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress) + 1 , LEN(PropertyAddress))
as Address

From PortfolioProject.dbo.NashvilleHousing

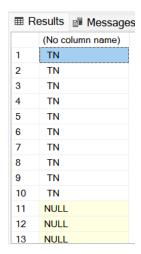
+1 was removed to remove the comma from the result and 'LEN' to separate the address even with different lengths



Select

```
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1)
From PortfolioProject.dbo.NashvilleHousing
```

--Another easier way of splitting the addresses. Commas were replaced with periods as PARSENAME is useful only for periods.

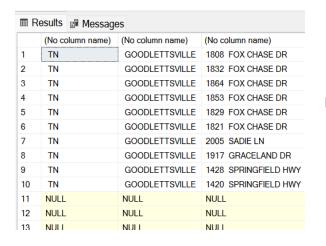


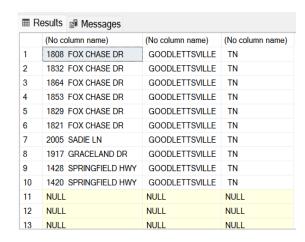
Select

```
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1)
,PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2)
,PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3)
From PortfolioProject.dbo.NashvilleHousing
```

Select

```
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1)
,PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2)
,PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3)
From PortfolioProject.dbo.NashvilleHousing
```





```
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)
 OwnerSplitAddress
                    OwnerSplitCity
                                   OwnerSplitState
 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
 1821 FOX CHASE DR
                     GOODLETTSVILLE TN
```

TN

TN

TN

NULL

NHILI

Fourth Task

2005 SADIE LN

NULL

NULL

1917 GRACELAND DR

1420 SPRINGFIELD HWY

1428 SPRINGFIELD HWY GOODLETTSVILLE

NULL

NHL

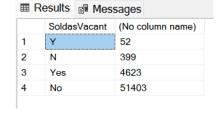
Replacing the Y and N to Yes and No in "Sold As Vacant" field

GOODLETTSVILLE TN

GOODLETTSVILLE

GOODLETTSVILLE

```
Select Distinct (SoldasVacant), COUNT(SoldasVacant)
From PortfolioProject..NashvilleHousing
Group by SoldasVacant
Order by 2
```



■ R	esults 🖹	■ Mes	sages
	SoldAsVacant		(No column name)
163	No		No
164	No		No
165	No		No
166	No		No
167	No		No
168	No		No
169	No		No
170	No		No
171	No		No
172	No		No
173	No		No
174	No		No
175	No		No

Fifth Task

--Remove Duplicates

Identify Duplicates

```
WITH ROWNUMCTE AS(
Select *,

ROW_NUMBER() OVER (
       PARTITION BY ParcelID,
                             PropertyAddress,
                             SalePrice,
                             SaleDate,
                             LegalReference
                             Order by
                                   UniqueID
                                   ) as row_num
From PortfolioProject.dbo.NashvilleHousing
--order by ParcelID
Select *
From RowNumCTE
Where row_num > 1
order by PropertyAddress
```

■ R	esults 📳	Messages			
	UniqueID	ParcellD	LandUse	PropertyAddress	SaleDate
46	27380	149 15 0A 418.00	SINGLE FAMILY	3044 BROMLEY WAY, ANTIOCH	2015-02-12 00:00:00.000
47	12692	163 11 0B 081.00	RESIDENTIAL CONDO	305 ASH FORGE DR, ANTIOCH	2014-02-03 00:00:00.000
48	27143	114 14 0A 022.00	SINGLE FAMILY	309 DOGWOOD TRL, NASHVILLE	2015-02-18 00:00:00.000
49	27392	164 07 0A 139.00	SINGLE FAMILY	312 GROVEDALE TRCE, ANTIOCH	2015-02-20 00:00:00.000
50	27355	108 08 0 114.00	SINGLE FAMILY	3121 E LAKE DR, NASHVILLE	2015-02-06 00:00:00.000
51	27356	108 12 0B 031.00	SINGLE FAMILY	3124 HARBORWOOD DR, NASHVILLE	2015-02-13 00:00:00.000
52	27387	150 07 0 181.00	SINGLE FAMILY	3140 COUNTRY MEADOW RD, ANTI	2015-02-18 00:00:00.000
53	27388	150 09 0A 028.00	SINGLE FAMILY	3313 CALAIS CIR, ANTIOCH	2015-02-13 00:00:00.000
54	27382	150 03 0 442.00	SINGLE FAMILY	3412 WESTCLIFFE CT, ANTIOCH	2015-02-02 00:00:00.000
55	27353	098 13 0C 200.00	SINGLE FAMILY	3420 EARHART RD, MOUNT JULIET	2015-02-06 00:00:00.000
56	27386	150 05 0 353.00	SINGLE FAMILY	365 CEDARCLIFF RD, ANTIOCH	2015-02-10 00:00:00.000
57	27402	176 01 0 017 00	SINGLE FAMILY	3848 ROUND ROCK DR ANTIOCH	2015-02-13 00:00:00 000

then Delete the duplicates



```
Results Messages

UniqueID ParceIID LandUse PropertyAddress SaleDate SalePrice LegalReference SoldAsVacant OwnerName O
```

Sixth Task

Delete Unused Columns

ALTER TABLE PortfolioProject.dbo.NashvilleHousing
DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress



ш.		Messages					
	UniqueID	ParcelID	LandUse	SaleDate	SalePrice	LegalReference	SoldAsVaca
1	2045	007 00 0 125.00	SINGLE FAMILY	2013-04-09 00:00:00.000	240000	20130412-0036474	No
2	16918	007 00 0 130.00	SINGLE FAMILY	2014-06-10 00:00:00.000	366000	20140619-0053768	No
3	54582	007 00 0 138.00	SINGLE FAMILY	2016-09-26 00:00:00.000	435000	20160927-0101718	No
4	43070	007 00 0 143.00	SINGLE FAMILY	2016-01-29 00:00:00.000	255000	20160129-0008913	No
5	22714	007 00 0 149.00	SINGLE FAMILY	2014-10-10 00:00:00.000	278000	20141015-0095255	No
6	18367	007 00 0 151.00	SINGLE FAMILY	2014-07-16 00:00:00.000	267000	20140718-0063802	No
7	19804	007 14 0 002.00	SINGLE FAMILY	2014-08-28 00:00:00.000	171000	20140903-0080214	No
8	54583	007 14 0 024.00	SINGLE FAMILY	2016-09-27 00:00:00.000	262000	20161005-0105441	No
9	36500	007 14 0 026.00	SINGLE FAMILY	2015-08-14 00:00:00.000	285000	20150819-0083440	No
10	19805	007 14 0 034.00	SINGLE FAMILY	2014-08-29 00:00:00.000	340000	20140909-0082348	No
11	29467	007 14 0A 024.00	SINGLE FAMILY	2015-04-14 00:00:00.000	425000	20150415-0033442	No
12	10754	007 14 0A 027 00	SINGLE FAMILY	2013-12-12 00:00:00 000	585000	20131227-0130352	No

Dataset: Alex the Analyst Guided Project