SQL Project: Nashville Housing Data Analysis

Overview of the data

SELECT * FROM PortfolioProject..NashvilleHousing;

	UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	OwnerName
1	2045	007 00 0 125.00	SINGLE FAMILY	1808 FOX CHASE DR, GOODLETTSVILLE	2013-04-09 00:00:00.000	240000	20130412-0036474	No	FRAZIER, CYR
2	16918	007 00 0 130.00	SINGLE FAMILY	1832 FOX CHASE DR, GOODLETTSVILLE	2014-06-10 00:00:00.000	366000	20140619-0053768	No	BONER, CHAR
3	54582	007 00 0 138.00	SINGLE FAMILY	1864 FOX CHASE DR, GOODLETTSVILLE	2016-09-26 00:00:00.000	435000	20160927-0101718	No	WILSON, JAME
4	43070	007 00 0 143.00	SINGLE FAMILY	1853 FOX CHASE DR, GOODLETTSVILLE	2016-01-29 00:00:00.000	255000	20160129-0008913	No	BAKER, JAY K.
5	22714	007 00 0 149.00	SINGLE FAMILY	1829 FOX CHASE DR, GOODLETTSVILLE	2014-10-10 00:00:00.000	278000	20141015-0095255	No	POST, CHRIST
6	18367	007 00 0 151.00	SINGLE FAMILY	1821 FOX CHASE DR, GOODLETTSVILLE	2014-07-16 00:00:00.000	267000	20140718-0063802	No	FIELDS, KAREN
7	19804	007 14 0 002.00	SINGLE FAMILY	2005 SADIE LN, GOODLETTSVILLE	2014-08-28 00:00:00.000	171000	20140903-0080214	No	HINTON, MICH
8	54583	007 14 0 024.00	SINGLE FAMILY	1917 GRACELAND DR, GOODLETTSVILLE	2016-09-27 00:00:00.000	262000	20161005-0105441	No	BAILOR, DARR
9	36500	007 14 0 026.00	SINGLE FAMILY	1428 SPRINGFIELD HWY, GOODLETTSVILLE	2015-08-14 00:00:00.000	285000	20150819-0083440	No	ROBERTS, MIS
10	19805	007 14 0 034.00	SINGLE FAMILY	1420 SPRINGFIELD HWY, GOODLETTSVILLE	2014-08-29 00:00:00.000	340000	20140909-0082348	No	LEE, JEFFREY

Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16....   SATYAJEETDC\Admin (54)   PortfolioProject   00:00:01   56,477 rows

Standardize Date Format

ALTER TABLE NashvilleHousing
ADD SaleDateConverted Date;

UPDATE NashvilleHousing SET SaleDateConverted = CONVERT(Date,SaleDate);

SELECT SaleDate,SaleDateConverted FROM PortfolioProject..NashvilleHousing;

	SaleDate	SaleDateConverted
1	2013-04-09 00:00:00.000	2013-04-09
2	2014-06-10 00:00:00.000	2014-06-10
3	2016-09-26 00:00:00.000	2016-09-26
4	2016-01-29 00:00:00.000	2016-01-29
5	2014-10-10 00:00:00.000	2014-10-10
6	2014-07-16 00:00:00.000	2014-07-16
7	2014-08-28 00:00:00.000	2014-08-28
8	2016-09-27 00:00:00.000	2016-09-27

Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16....   SATYAJEETDC\Admin (54)   PortfolioProject   00:00:00   56,477 rows

Populate Property Address data

SELECT * FROM PortfolioProject..NashvilleHousing WHERE PropertyAddress IS NULL;

	UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	OwnerName
1	43076	025 07 0 031.00	SINGLE FAMILY	NULL	2016-01-15 00:00:00.000	179900	20160120-0005776	No	COSTNER, FRED & CAROLYN
2	39432	026 01 0 069.00	VACANT RESIDENTIAL LAND	NULL	2015-10-23 00:00:00.000	153000	20151028-0109602	No	SHACKLEFORD, MICHAEL C., JF
3	45290	026 05 0 017.00	SINGLE FAMILY	NULL	2016-03-29 00:00:00.000	155000	20160330-0029941	No	TRIPP, MARVIN S. & DEBORAH
4	53147	026 06 0A 038.00	RESIDENTIAL CONDO	NULL	2016-08-25 00:00:00.000	144900	20160831-0091567	No	NULL
5	43080	033 06 0 041.00	SINGLE FAMILY	NULL	2016-01-04 00:00:00.000	170000	20160107-0001526	No	FRANK, ZACHARY & NIKI
6	45295	033 06 0A 002.00	SINGLE FAMILY	NULL	2016-03-29 00:00:00.000	210000	20160331-0030709	No	NULL
7	48731	033 15 0 123.00	SINGLE FAMILY	NULL	2016-05-06 00:00:00.000	199900	20160506-0045368	No	COLEMAN, AARON A. & CECIL, (
8	50927	044 05 0 135.00	SINGLE FAMILY	NULL	2016-06-15 00:00:00.000	160000	20160617-0061987	No	SMITH, WARREN & MARTHA

Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16....   SATYAJEETDC\Admin (54)   PortfolioProject   00:00:00   29 rows

Where PropertyAddress is null we are going to populate using same ParcelID
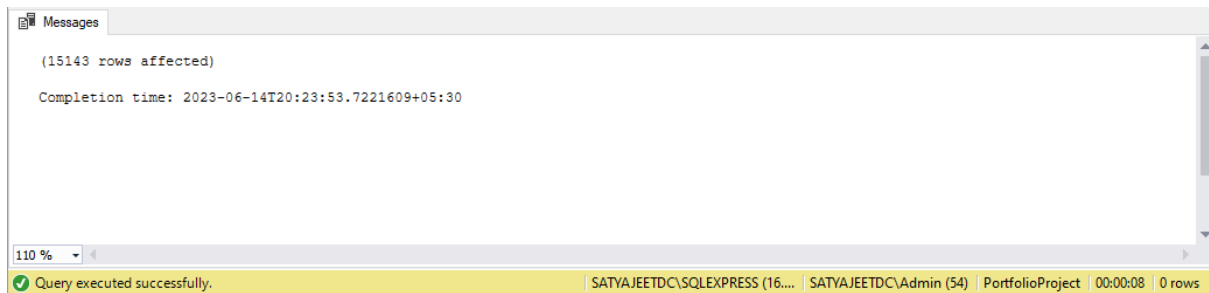
UPDATE A

SET PropertyAddress=ISNULL(A.PropertyAddress,B.PropertyAddress)
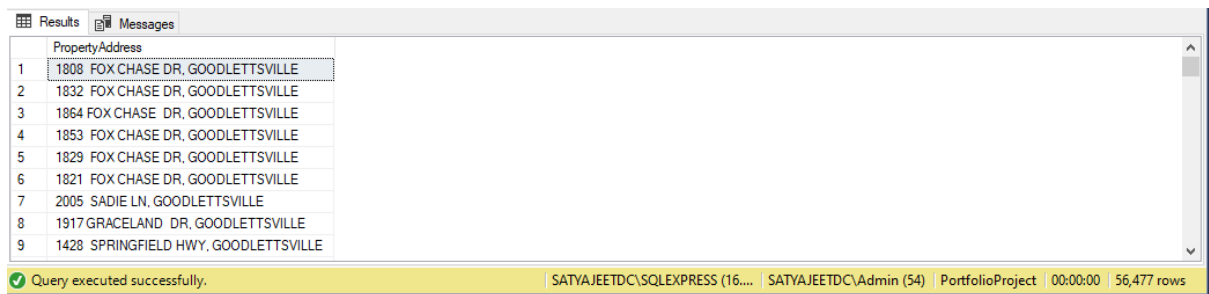
FROM PortfolioProject..NashvilleHousing A

JOIN PortfolioProject..NashvilleHousing B

ON A.ParcelID=B.parcelID AND A.UniqueID != B.UniqueID

```
Messages

 (15143 rows affected)

 Completion time: 2023-06-14T20:23:53.7221609+05:30
```

110 %   ◄
✔ Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16.... SATYAJEETDC\Admin (54)   PortfolioProject   00:00:08   0 rows

Breaking out Address into Individual Columns (Address, City, State)-PropertyAddress

Initially: SELECT PropertyAddress FROM PortfolioProject..NashvilleHousing;

	PropertyAddress
1	1808 FOX CHASE DR, GOODLETTSVILLE
2	1832 FOX CHASE DR, GOODLETTSVILLE
3	1864 FOX CHASE DR, GOODLETTSVILLE
4	1853 FOX CHASE DR, GOODLETTSVILLE
5	1829 FOX CHASE DR, GOODLETTSVILLE
6	1821 FOX CHASE DR, GOODLETTSVILLE
7	2005 SADIE LN, GOODLETTSVILLE
8	1917 GRACELAND DR, GOODLETTSVILLE
9	1428 SPRINGFIELD HWY, GOODLETTSVILLE

✔ Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16.... SATYAJEETDC\Admin (54)   PortfolioProject   00:00:00   56,477 rows

ALTER TABLE NashvilleHousing

ADD PropertySplitAddress nvarchar(255);

UPDATE NashvilleHousing

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

ALTER TABLE NashvilleHousing

ADD PropertySplitCity nvarchar(255);

UPDATE NashvilleHousing

SET PropertySplitCity=SUBSTRING(PropertyAddress,CHARINDEX(',',PropertyAddress)+1,LEN(PropertyAddress));

SELECT PropertyAddress,PropertySplitAddress,PropertySplitCity FROM
PortfolioProject..NashvilleHousing;

	PropertyAddress	PropertySplitAddress	PropertySplitCity
1	1808 FOX CHASE DR, GOODLETTSVILLE	1808 FOX CHASE DR	GOODLETTSVILLE
2	1832 FOX CHASE DR, GOODLETTSVILLE	1832 FOX CHASE DR	GOODLETTSVILLE
3	1864 FOX CHASE DR, GOODLETTSVILLE	1864 FOX CHASE DR	GOODLETTSVILLE
4	1853 FOX CHASE DR, GOODLETTSVILLE	1853 FOX CHASE DR	GOODLETTSVILLE
5	1829 FOX CHASE DR, GOODLETTSVILLE	1829 FOX CHASE DR	GOODLETTSVILLE
6	1821 FOX CHASE DR, GOODLETTSVILLE	1821 FOX CHASE DR	GOODLETTSVILLE
7	2005 SADIE LN, GOODLETTSVILLE	2005 SADIE LN	GOODLETTSVILLE
8	1917 GRACELAND DR, GOODLETTSVILLE	1917 GRACELAND DR	GOODLETTSVILLE
9	1428 SPRINGFIELD HWY, GOODLETTSVILLE	1428 SPRINGFIELD HWY	GOODLETTSVILLE

Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16.... | SATYAJEETDC\Admin (54) | PortfolioProject | 00:00:01 | 56,477 rows

Breaking out Address into Individual Columns (Address, City, State)-OwnerAddress

Initially: SELECT OwnerAddress FROM PortfolioProject..NashvilleHousing;

	OwnerAddress
1	1808 FOX CHASE DR, GOODLETTSVILLE, TN
2	1832 FOX CHASE DR, GOODLETTSVILLE, TN
3	1864 FOX CHASE DR, GOODLETTSVILLE, TN
4	1853 FOX CHASE DR, GOODLETTSVILLE, TN
5	1829 FOX CHASE DR, GOODLETTSVILLE, TN
6	1821 FOX CHASE DR, GOODLETTSVILLE, TN
7	2005 SADIE LN, GOODLETTSVILLE, TN
8	1917 GRACELAND DR, GOODLETTSVILLE, TN
9	1428 SPRINGFIELD HWY, GOODLETTSVILLE, TN
10	1420 SPRINGFIELD HWY, GOODLETTSVILLE, TN
11	NULL

Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16.... | SATYAJEETDC\Admin (52) | master | 00:00:00 | 56,477 rows

ALTER TABLE NashvilleHousing ADD OwnerSplitAddress Nvarchar(255);

ALTER TABLE NashvilleHousing ADD OwnerSplitCity nvarchar(255);

ALTER TABLE NashvilleHousing ADD OwnerSplitState nvarchar(255);

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

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

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

SELECT OwnerAddress,OwnerSplitAddress,OwnerSplitCity,OwnerSplitState
FROM PortfolioProject..NashvilleHousing;

	OwnerAddress	OwnerSplitAddress	OwnerSplitCity	OwnerSplitState
1	1808 FOX CHASE DR, GOODLETTSVILLE, TN	1808 FOX CHASE DR	GOODLETTSVILLE	TN
2	1832 FOX CHASE DR, GOODLETTSVILLE, TN	1832 FOX CHASE DR	GOODLETTSVILLE	TN
3	1864 FOX CHASE DR, GOODLETTSVILLE, TN	1864 FOX CHASE DR	GOODLETTSVILLE	TN
4	1853 FOX CHASE DR, GOODLETTSVILLE, TN	1853 FOX CHASE DR	GOODLETTSVILLE	TN
5	1829 FOX CHASE DR, GOODLETTSVILLE, TN	1829 FOX CHASE DR	GOODLETTSVILLE	TN
6	1821 FOX CHASE DR, GOODLETTSVILLE, TN	1821 FOX CHASE DR	GOODLETTSVILLE	TN
7	2005 SADIE LN, GOODLETTSVILLE, TN	2005 SADIE LN	GOODLETTSVILLE	TN
8	1917 GRACELAND DR, GOODLETTSVILLE, TN	1917 GRACELAND DR	GOODLETTSVILLE	TN
9	1428 SPRINGFIELD HWY, GOODLETTSVILLE, TN	1428 SPRINGFIELD HWY	GOODLETTSVILLE	TN
10	1420 SPRINGFIELD HWY, GOODLETTSVILLE, TN	1420 SPRINGFIELD HWY	GOODLETTSVILLE	TN
11	NULL	NULL	NULL	NULL

Query executed successfully.   SATYAJEETDC\SQLEXPRESS (16.... | SATYAJEETDC\Admin (60) | PortfolioProject | 00:00:01 | 56,477 rows

Change Y and N in the column "Sold as Vacant" to Yes and No

SELECT DISTINCT SoldAsVacant, COUNT(SoldAsVacant) FROM PortfolioProject..NashvilleHousing

GROUP BY SoldAsVacant ORDER BY 2;

	SoldAsVacant	(No column name)
1	Y	52
2	N	399
3	Yes	4623
4	No	51403

Query executed successfully.　　　　SATYAJEETDC\SQLEXPRESS (16....　SATYAJEETDC\Admin (60)　PortfolioProject　00:00:00　4 rows

UPDATE NashvilleHousing

SET SoldAsVacant = CASE WHEN SoldAsVacant='Y' THEN 'Yes'

WHEN SoldAsVacant='N' THEN 'No'

ELSE SoldAsVacant

END;

SELECT DISTINCT SoldAsVacant, COUNT(SoldAsVacant) FROM PortfolioProject..NashvilleHousing

GROUP BY SoldAsVacant

ORDER BY 2;

	SoldAsVacant	(No column name)
1	Yes	4675
2	No	51802

Query executed successfully.　　　　SATYAJEETDC\SQLEXPRESS (16....　SATYAJEETDC\Admin (60)　PortfolioProject　00:00:00　2 rows

Remove Duplicates

Initially:

WITH RowNumCTE AS (

SELECT * ,

        ROW_NUMBER() OVER (

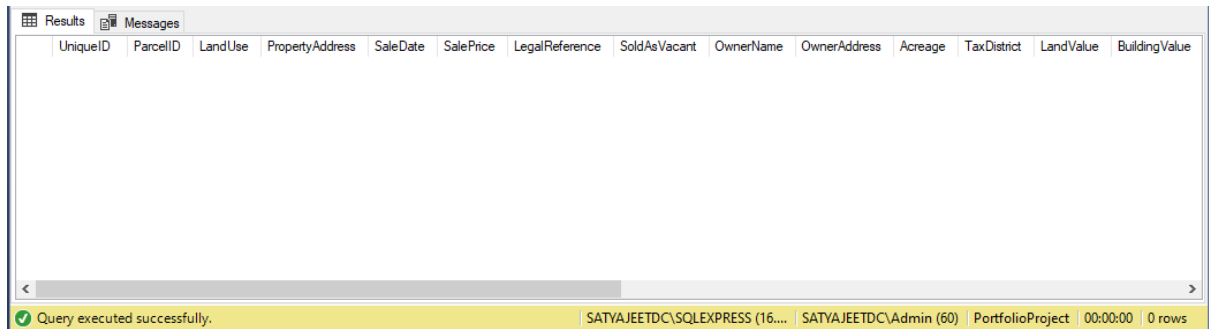   PARTITION BY ParcelID,PropertyAddress,SalePrice,SaleDate,LegalReference ORDER BY UniqueID

) row_num

FROM PortfolioProject.dbo.NashvilleHousing
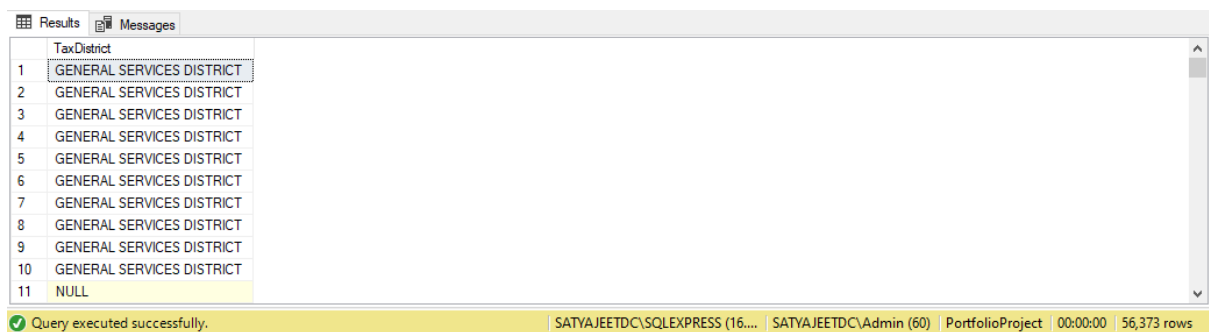
--ORDER BY ParcelID

)

SELECT * FROM RowNumCTE WHERE row_num>1;

WITH RowNumCTE AS (

SELECT * ,

    ROW_NUMBER() OVER (

    PARTITION BY ParcelID,PropertyAddress,SalePrice,SaleDate,LegalReference ORDER BY UniqueID

) row_num

FROM PortfolioProject.dbo.NashvilleHousing

--ORDER BY ParcelID

)

DELETE FROM RowNumCTE WHERE row_num>1;

Now if we check for duplicates :

WITH RowNumCTE AS (

SELECT * ,

    ROW_NUMBER() OVER (

    PARTITION BY ParcelID,PropertyAddress,SalePrice,SaleDate,LegalReference ORDER BY UniqueID

) row_num

FROM PortfolioProject.dbo.NashvilleHousing

--ORDER BY ParcelID

)

SELECT * FROM RowNumCTE WHERE row_num>1;

DELETE Unused Columns

Initially: SELECT TaxDistrict FROM PortfolioProject.dbo.NashvilleHousing;
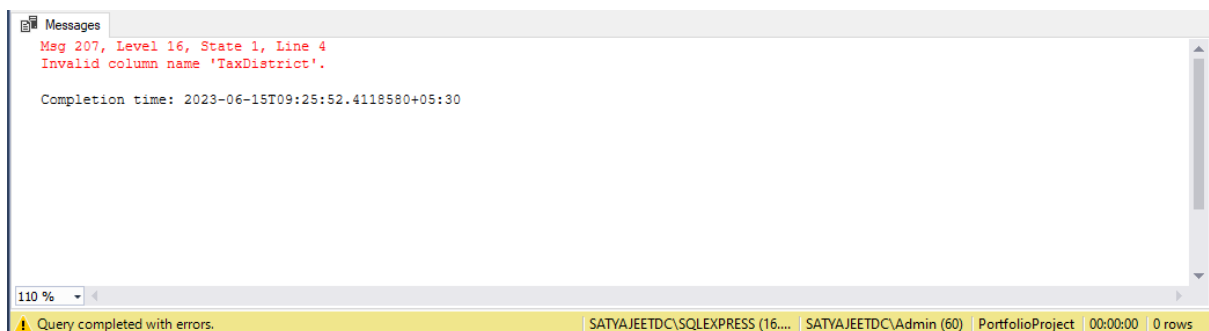
ALTER TABLE PortfolioProject.dbo.NashvilleHousing

DROP COLUMN TaxDistrict;

SELECT TaxDistrict FROM PortfolioProject.dbo.NashvilleHousing;