Data cleaning in MS SQL Server

In this project, we will be using MS SQL Server to do some Data Cleaning from an excel file that contains fictitious information about owners in Nashville area (Tennessee, US).

The database contains one table (NashvilleHousingData). Let's understand it first to be able to modify it later!

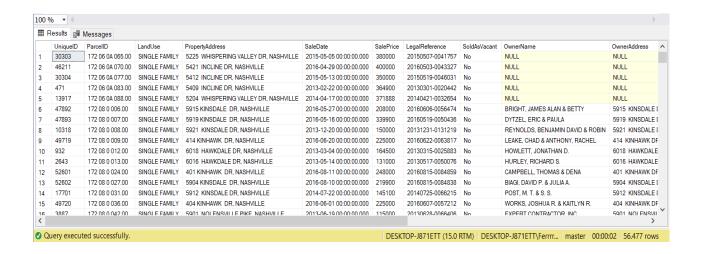
Skills used (SQL):

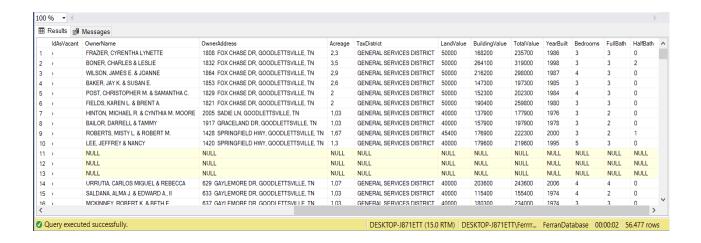
Convert, Joins, Substring, Charindex, Parsename, Replace, Case Statements, Remove Duplicates, Drop column.

--We start by exploring the data:

Select *

From FerranDatabase.dbo.NashvilleHousing





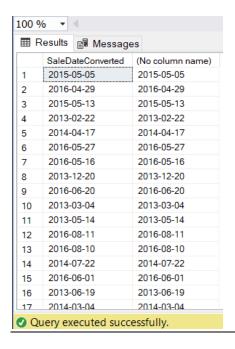
--Standarizing Date Format

Select SaleDateConverted, CONVERT(Date, SaleDate)
FROM FerranDatabase.dbo.NashvilleHousing

Update FerranDatabase.dbo.NashvilleHousing
SET SaleDate = CONVERT(Date, SaleDate)

ALTER TABLE FerranDatabase.dbo.NashvilleHousing Add SaleDateConverted Date;

Update FerranDatabase.dbo.NashvilleHousing
SET SaleDateConverted = CONVERT(Date, SaleDate)

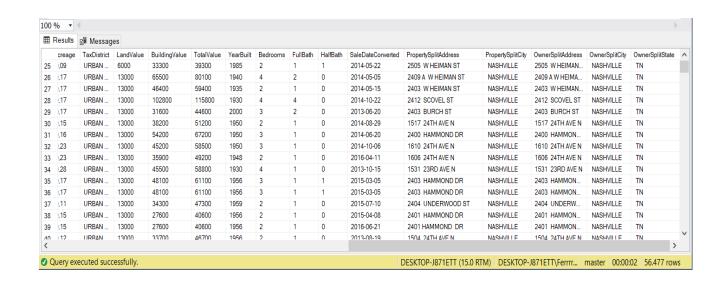


```
--Populating the Property Address data
Select *
From FerranDatabase.dbo.NashvilleHousing
--Where PropertyAddress is null
Order by ParcelID
Select a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress,
ISNULL(a.PropertyAddress,b.PropertyAddress)
From FerranDatabase.dbo.NashvilleHousing a
JOIN FerranDatabase.dbo.NashvilleHousing b
      on a.ParcelID = b.ParcelID
      AND a. [UniqueID ] <> b. [UniqueID ]
Where a.PropertyAddress is null
Update a
SET PropertyAddress = ISNULL(a.PropertyAddress,b.PropertyAddress)
From FerranDatabase.dbo.NashvilleHousing a
JOIN FerranDatabase.dbo.NashvilleHousing b
      on a.ParcelID = b.ParcelID
      AND a.[UniqueID ] <> b.[UniqueID ]
Where a.PropertyAddress is null
```



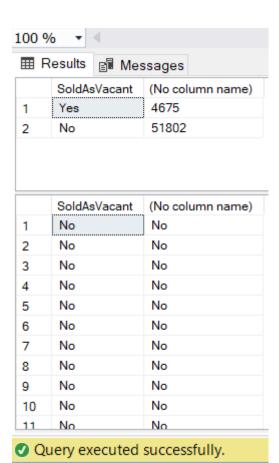
-- Breaking out Address into Individual Columns (Address, City, State) Select PropertyAddress From FerranDatabase.dbo.NashvilleHousing **SELECT** SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress) -1) as Address , SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress) + 1 , LEN(PropertyAddress)) as Address From FerranDatabase.dbo.NashvilleHousing **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 *
From FerranDatabase.dbo.NashvilleHousing
Select OwnerAddress
From FerranDatabase.dbo.NashvilleHousing
Select
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3)
,PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2)
,PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1)
From FerranDatabase.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)
Select *
From FerranDatabase.dbo.NashvilleHousing
```



-- Change Y and N to Yes and No in "Sold as Vacant" field

Order by 2



```
-- Remove Duplicates
WITH ROWNUMCTE AS(
Select *,
      ROW NUMBER() OVER (
      PARTITION BY ParcelID,
                        PropertyAddress,
                        SalePrice,
                        SaleDate,
                        LegalReference
                        ORDER BY
                              UniqueID
                              ) row_num
From FerranDatabase.dbo.NashvilleHousing
Select *
From RowNumCTE
Where row_num > 1
--Delete Unused Columns
Select *
From FerranDatabase.dbo.NashvilleHousing
ALTER TABLE FerranDatabase.dbo.NashvilleHousing
DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress
ALTER TABLE FerranDatabase.dbo.NashvilleHousing
DROP COLUMN SaleDate
```

