/\*

Cleaning Data in SQL Queries

\*/

--------------------------------------------------------------------------------------------------------------------------

-- Standardize Date Format

Select SaleDateConverted, CONVERT(date, SaleDate)

From PortfolioProject..NashvilleHousing

Update NashvilleHousing

SET SaleDate = CONVERT(date, SaleDate)

ALTER TABLE NashvilleHousing

Add SaleDateConverted Date;

Update NashvilleHousing

SET SaleDateConverted = CONVERT(date, SaleDate)

--------------------------------------------------------------------------------------------------------------------------

-- Populate Property Address data ~ uses a self join

Select \*

From PortfolioProject..NashvilleHousing

--Where PropertyAddress is NULL

Order By ParcelID

Select a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress, ISNULL(a.propertyaddress, b.PropertyAddress)

From PortfolioProject..NashvilleHousing a

JOIN PortfolioProject..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 PortfolioProject..NashvilleHousing a

JOIN PortfolioProject..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) ~ using substring and character index

Select PropertyAddress

From PortfolioProject..NashvilleHousing

--Where PropertyAddress is NULL

Select

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

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

From PortfolioProject..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 PortfolioProject..NashvilleHousing

Select OwnerAddress

From PortfolioProject..NashvilleHousing

--PARSENAME CHECKS FOR PERIODS, REPLACE COMMAS WITH A PERIOD

Select

PARSENAME(REPLACE(OwnerAddress,',','.'),3)

,PARSENAME(REPLACE(OwnerAddress,',','.'),2)

,PARSENAME(REPLACE(OwnerAddress,',','.'),1)

From PortfolioProject..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)

--------------------------------------------------------------------------------------------------------------------------

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

Select Distinct(SoldAsVacant), Count(SoldAsVacant)

From PortfolioProject..NashvilleHousing

Group By SoldAsVacant

Order by 2

Select SoldAsVacant

, CASE When SoldAsVacant = 'Y' THEN 'Yes'

When SoldAsVacant = 'N' THEN 'No'

ELSE SoldAsVacant

END

From PortfolioProject..NashvilleHousing

Update NashvilleHousing

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

When SoldAsVacant = 'N' THEN 'No'

ELSE SoldAsVacant

END

-----------------------------------------------------------------------------------------------------------------------------------------------------------

-- Remove Duplicates

WITH RowNumCTE AS(

Select \*,

ROW\_NUMBER() OVER (

PARTITION BY ParcelID,

PropertyAddress,

SalePrice,

SaleDate,

LegalReference

ORDER BY

UniqueID

) row\_num

From PortfolioProject..NashvilleHousing

)

Delete

FROM RowNumCTE

WHERE row\_num >1

---------------------------------------------------------------------------------------------------------

-- Delete Unused Columns

Select \*

From PortfolioProject..NashvilleHousing

ALTER TABLE PortfolioProject..NashvilleHousing

DROP COLUMN SaleDate