**Cleaning Data in SQL Queries**

Select \*

From .dbo.HousingData

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

**-- Standardize Date Format**

Select saleDateConverted, CONVERT(Date,SaleDate)

From .dbo.HousingData

Update HousingData

SET SaleDate = CONVERT(Date,SaleDate)

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

**-- Populate Property Address data**

Select \*

From dbo.HousingData

Where PropertyAddress is null

order by ParcelID

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

From dbo.HousingData a

JOIN dbo.HousingData 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 dbo.HousingData a

JOIN dbo.HousingData 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 dbo.HousingData

Where PropertyAddress is null

order by ParcelID

SELECT

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

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

From dbo.HousingData

ALTER TABLE HousingData

Add PropertySplitAddress Nvarchar(255);

Update HousingData

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

ALTER TABLE HousingData

Add PropertySplitCity Nvarchar(255);

Update HousingData

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

Select \*

From dbo.HousingData

Select OwnerAddress

From dbo.HousingData

Select

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

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

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

From dbo.HousingData

ALTER TABLE HousingData

Add OwnerSplitAddress Nvarchar(255);

Update HousingData

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

ALTER TABLE HousingData

Add OwnerSplitCity Nvarchar(255);

Update HousingData

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

ALTER TABLE HousingData

Add OwnerSplitState Nvarchar(255);

Update HousingData

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

Select \*

From dbo.HousingData

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

Select Distinct(SoldAsVacant), Count(SoldAsVacant)

From dbo.HousingData

Group by SoldAsVacant

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 dbo.HousingData

--order by ParcelID

)

Select \*

From RowNumCTE

Where row\_num > 1

Order by PropertyAddress

Select \*

From .dbo.HousingData

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

-- Delete Unused Columns

Select \*

From dbo.HousingData

ALTER TABLE dbo.HousingData

DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress, SaleDate