--Cleaning Data

SELECT \* FROM HousingData

--Sale Date correction

SELECT SaleDateConverted,CONVERT(Date,SaleDate) as date

FROM HousingData

UPDATE HousingData

SET SaleDate = CONVERT(Date,SaleDate)

ALTER TABLE HousingData

ADD SaleDateConverted Date

UPDATE HousingData

SET SaleDateConverted= CONVERT(Date,SaleDate)

--Property Address (Null values)

SELECT \*

FROM HousingData

--Where PropertyAddress is null

ORDER BY ParcelID

SELECT a.ParcelId,a.PropertyAddress,b.ParcelId,b.PropertyAddress, ISNULL(a.PropertyAddress,b.PropertyAddress)

FROM HousingData a

JOIN 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 HousingData a

JOIN HousingData b

ON a.ParcelID=b.ParcelID

and a.[UniqueID] <> b.[UniqueID]

WHERE a.PropertyAddress is null

--Address/City/State

SELECT PropertyAddress

FROM HousingData

SELECT SUBSTRING(PropertyAddress,1,

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

SUBSTRING(PropertyAddress,

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

FROM HousingData

ALTER TABLE HousingData

ADD PropertyAddressSplit Nvarchar(255)

UPDATE HousingData

SET PropertyAddressSplit= SUBSTRING(PropertyAddress,1,

CHARINDEX(',',PropertyAddress)-1)

ALTER TABLE HousingData

ADD PropertyCitySplit Nvarchar(255)

UPDATE HousingData

SET PropertyCitySplit = SUBSTRING(PropertyAddress,

CHARINDEX(',',PropertyAddress)+1,LEN(PropertyAddress))

SELECT \* FROM HousingData

--Owner Address For splitting Address City State

SELECT ParseName(Replace(OwnerAddress,',','.'),3),

ParseName(Replace(OwnerAddress,',','.'),2),

ParseName(Replace(OwnerAddress,',','.'),1)

FROM 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 ParcelID,OwnerAddress FROM HousingData

WHERE OwnerAddress is not null

SELECT a.ParcelID,a.OwnerAddress, b.ParcelID,b.OwnerAddress, ISNULL(a.OwnerAddress,b.OwnerAddress)

FROM HousingData a

JOIN HousingData b

on a.ParcelID=b.ParcelID

and a.[UniqueID] <> b.[UniqueID]

WHERE a.OwnerAddress is null

--Change Y and N to Yes and No on SoldVacant field

SELECT DISTINCT SoldasVacant,COUNT(SoldasVacant)

FROM HousingData

GROUP BY SoldasVacant

ORDER BY 2

SELECT SoldAsVacant,

CASE WHEN SoldasVacant = 'Y' THEN 'Yes'

WHEN SoldasVacant = 'N' THEN 'No'

ELSE SoldAsVacant

END

FROM HousingData

UPDATE HousingData

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 HousingData)

DELETE FROM RownumCTE

WHERE row\_num > 1

--ORDER BY PropertyAddress

ALTER TABLE HousingData

DROP COLUMN TaxDistrict

SELECT \* FROM HousingData

ALTER TABLE HousingData

DROP COLUMN SaleDate