

```

/*
NAME: Tayo Obafaiye
Cleaning Data in SQL Queries
Housing Data
*/

SELECT *
FROM PortfolioProject.dbo.NashvilleHousing

-- Standardize Date Format
SELECT SaleDate, CONVERT(DATE, SaleDate)
FROM PortfolioProject.dbo.NashvilleHousing

UPDATE NashvilleHousing
SET SaleDate = CONVERT(DATE, SaleDate)

SELECT SaleDate
FROM PortfolioProject.dbo.NashvilleHousing

ALTER TABLE NashvilleHousing ADD SaleDateConverted DATE

UPDATE NashvilleHousing
SET SaleDateConverted = CONVERT(DATE, SaleDate)

SELECT SaleDateConverted
FROM PortfolioProject.dbo.NashvilleHousing

-- Populate Property Address Data
SELECT PropertyAddress
FROM PortfolioProject.dbo.NashvilleHousing

SELECT PropertyAddress
FROM PortfolioProject.dbo.NashvilleHousing
WHERE PropertyAddress IS NULL

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

SELECT *
FROM PortfolioProject.dbo.NashvilleHousing
--WHERE PropertyAddress IS NULL
ORDER BY ParcelID

SELECT a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress,
ISNULL(a.PropertyAddress, b.PropertyAddress)
FROM PortfolioProject.dbo.NashvilleHousing a
JOIN PortfolioProject.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 PortfolioProject.dbo.NashvilleHousing a
JOIN PortfolioProject.dbo.NashvilleHousing b
    ON a.ParcelID = b.ParcelID
    AND a.[UniqueID ] <> b.[UniqueID ]

```

```

WHERE a.PropertyAddress IS NULL

-- Breaking out address into separate columns (Address, City & State)
SELECT PropertyAddress
FROM PortfolioProject.dbo.NashvilleHousing

SELECT
SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1) AS Address,
SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress)+1, LEN(PropertyAddress)) AS
City
FROM PortfolioProject.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 PortfolioProject.dbo.NashvilleHousing

-- OwnerAddress
SELECT OwnerAddress
FROM NashvilleHousing

SELECT
PARSENAME(OwnerAddress, 1)
FROM NashvilleHousing

SELECT
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1),
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2),
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3)
FROM NashvilleHousing

SELECT
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 3),
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 2),
PARSENAME(REPLACE(OwnerAddress, ',', '.'), 1)
FROM NashvilleHousing

ALTER TABLE NashvilleHousing ADD OwnerSplitAddress NVARCHAR(255)

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

ALTER TABLE NashvilleHousing ADD
OwnerSplitCity NVARCHAR(255),
OwnerSplitState NVARCHAR(255)

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

```

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

```
-- Change Y and N to Yes and No, respectively in "SoldAsVacant" field
```

```
SELECT DISTINCT SoldAsVacant, COUNT(SoldAsVacant)
FROM NashvilleHousing
GROUP BY SoldAsVacant
```

```
SELECT DISTINCT SoldAsVacant, COUNT(SoldAsVacant)
FROM NashvilleHousing
GROUP BY SoldAsVacant
ORDER BY 2
```

```
SELECT SoldAsVacant,
CASE
    WHEN SoldAsVacant = 'Y' THEN 'Yes'
    WHEN SoldAsVacant = 'N' THEN 'No'
    ELSE SoldAsVacant
END
FROM NashvilleHousing
```

```
UPDATE NashvilleHousing
SET SoldAsVacant = CASE
    WHEN SoldAsVacant = 'Y' THEN 'Yes'
    WHEN SoldAsVacant = 'N' THEN 'No'
    ELSE SoldAsVacant
END
```

```
-- Remove Duplicates
```

```
SELECT *, ROW_NUMBER() OVER (PARTITION BY ParcelID,
```

```
PropertyAddress,
SalePrice,
SaleDate,
LegalReference
ORDER BY UniqueID) row_num
```

```
FROM NashvilleHousing
ORDER BY ParcelID
```

```
WITH RowNumCTE AS (
SELECT *, ROW_NUMBER() OVER (PARTITION BY ParcelID,
```

```
PropertyAddress,
SalePrice,
SaleDate,
LegalReference
ORDER BY UniqueID) row_num
```

```
FROM NashvilleHousing
)
SELECT *
FROM RowNumCTE
WHERE row_num > 1
ORDER BY PropertyAddress
```

```
WITH RowNumCTE AS (
SELECT *, ROW_NUMBER() OVER (PARTITION BY ParcelID,
```

```
PropertyAddress,
SalePrice,
SaleDate,
```

```
LegalReference
ORDER BY UniqueID) row_num
```

```
FROM NashvilleHousing
)
```

```
DELETE
```

```
FROM RowNumCTE
```

```
WHERE row_num > 1
```

```
-- Delete Unused Columns
```

```
ALTER TABLE NashvilleHousing
```

```
DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress
```

```
SELECT *
```

```
FROM NashvilleHousing
```

```
ALTER TABLE NashvilleHousing
```

```
DROP COLUMN SaleDate
```

```
SELECT *
```

```
FROM NashvilleHousing
```

```
/*
```

```
Credit to Alex The Analyst https://www.youtube.com/watch?v=8r07ztF4NtU&t=1709s
```

```
*/
```