HOUSING SQL DATA CLEANING PROJECT

Ximena Bustamante

INTRODUCTION

- This project employs Nashville housing data to be explored and clean in SQL Server Management Studio
- This is a guided project led by @Alex The Analyst

DATASET

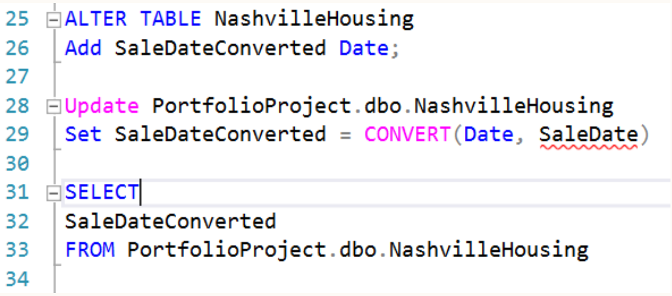
- This dataset consists of the following 19 variables:
  - UniqueID
  - ParcelID
  - LandUse
  - PropertyAddress
  - SaleDate
  - SalePrice
  - LegalReference
  - SoldAsVacant
  - OwnerName
  - OwnerAddress
  - Acreage
  - TaxDistrict
  - LandValue
  - BuildingValue
  - TotalValue
  - YearBuilt
  - Bedrooms
  - FullBath
  - HalfBath

DATA CLEANING AND MANIPULATION SQL

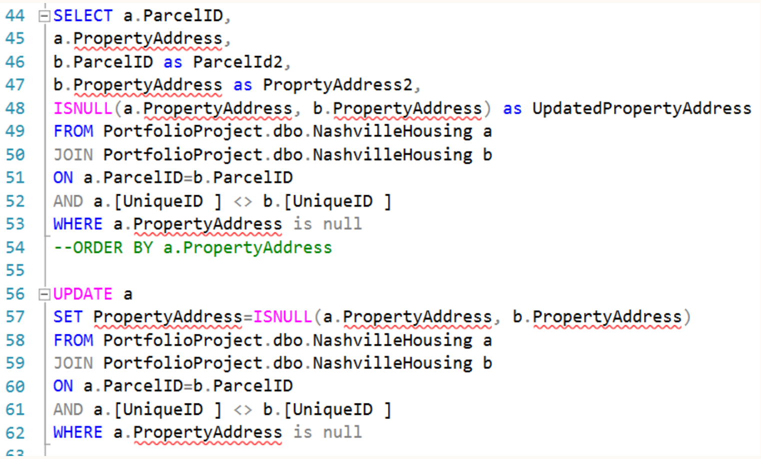
DATA CLEANING AND MANIPULATION

SQL

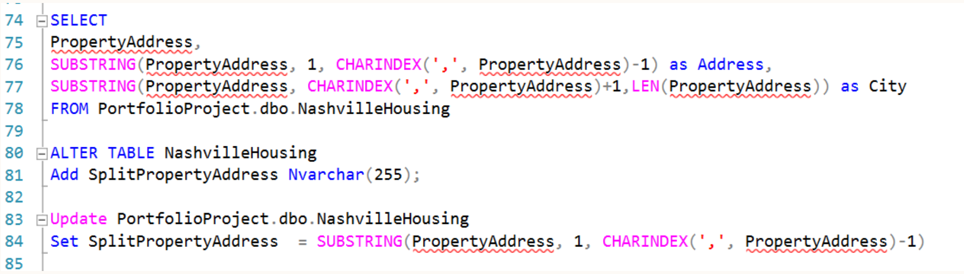
- SQL Server Management Studio was used to clean and manipulate the data
- Alter table, update and convert were used to change data type permanently
- ISNULL and self joins were used to populate missing addresses
- Substring and Charindex were used to break apart cities and zip codes from addresses



```sql
ALTER TABLE NashvilleHousing
Add SaleDateConverted Date;

Update PortfolioProject.dbo.NashvilleHousing
Set SaleDateConverted = CONVERT(Date, SaleDate)

SELECT
SaleDateConverted
FROM PortfolioProject.dbo.NashvilleHousing

```



```sql
SELECT a.ParcelID,
a.PropertyAddress,
b.ParcelID as ParcelId2,
b.PropertyAddress as ProprtyAddress2,
ISNULL(a.PropertyAddress, b.PropertyAddress) as UpdatedPropertyAddress
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
--ORDER BY a.PropertyAddress

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
```

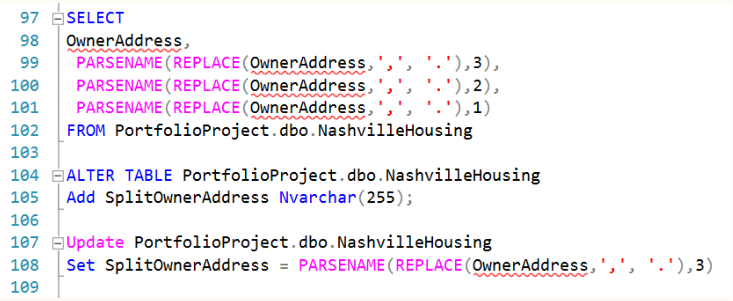


```sql
SELECT
PropertyAddress,
SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1) as Address,
SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress)+1,LEN(PropertyAddress)) as City
FROM PortfolioProject.dbo.NashvilleHousing

ALTER TABLE NashvilleHousing
Add SplitPropertyAddress Nvarchar(255);

Update PortfolioProject.dbo.NashvilleHousing
Set SplitPropertyAddress  = SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress)-1)

```

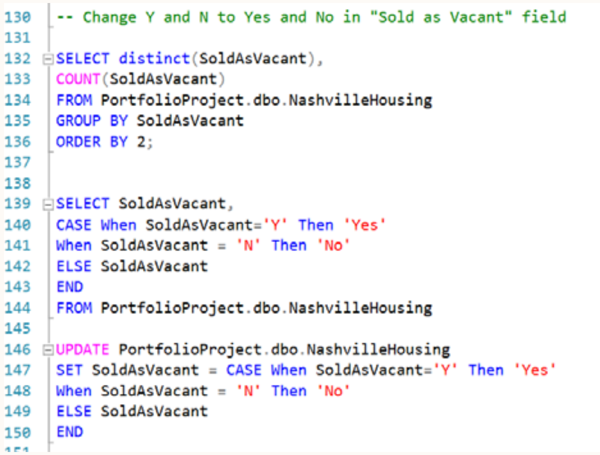
DATA CLEANING AND MANIPULATION

SQL

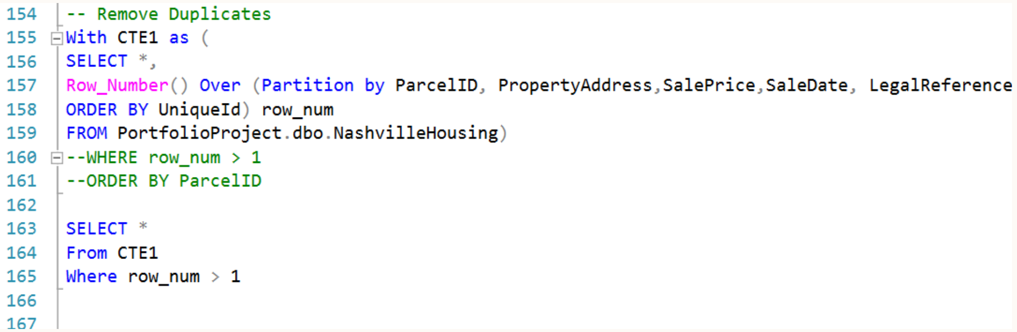
- Parsename and replace were used to break into columns based on a delimiter
- Count, case when and update were used to create a yes/no column to identify specific categories
- CTEs were used to remove duplicates
- Drop column and alter table were used to drop unused columns



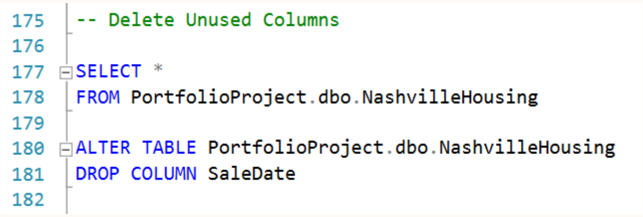
```sql
SELECT
OwnerAddress,
 PARSENAME(REPLACE(OwnerAddress,',', '.'),3),
 PARSENAME(REPLACE(OwnerAddress,',', '.'),2),
 PARSENAME(REPLACE(OwnerAddress,',', '.'),1)
FROM PortfolioProject.dbo.NashvilleHousing

ALTER TABLE PortfolioProject.dbo.NashvilleHousing
Add SplitOwnerAddress Nvarchar(255);

Update PortfolioProject.dbo.NashvilleHousing
Set SplitOwnerAddress = PARSENAME(REPLACE(OwnerAddress,',', '.'),3)

```



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

SELECT distinct(SoldAsVacant),
COUNT(SoldAsVacant)
FROM PortfolioProject.dbo.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.dbo.NashvilleHousing

UPDATE PortfolioProject.dbo.NashvilleHousing
SET SoldAsVacant = CASE When SoldAsVacant='Y' Then 'Yes'
When SoldAsVacant = 'N' Then 'No'
ELSE SoldAsVacant
END
```



```sql
-- Remove Duplicates
With CTE1 as (
SELECT *,
Row_Number() Over (Partition by ParcelID, PropertyAddress,SalePrice,SaleDate, LegalReference
ORDER BY UniqueId) row_num
FROM PortfolioProject.dbo.NashvilleHousing)
--WHERE row_num > 1
--ORDER BY ParcelID

SELECT *
From CTE1
Where row_num > 1


```



```sql
-- Delete Unused Columns

SELECT *
FROM PortfolioProject.dbo.NashvilleHousing

ALTER TABLE PortfolioProject.dbo.NashvilleHousing
DROP COLUMN SaleDate

```

THANK YOU FOR CHECKING OUT MY PROJECT!

- Follow me for more project ideas
- If you have any questions, comments, feedback, feel free to DM me!