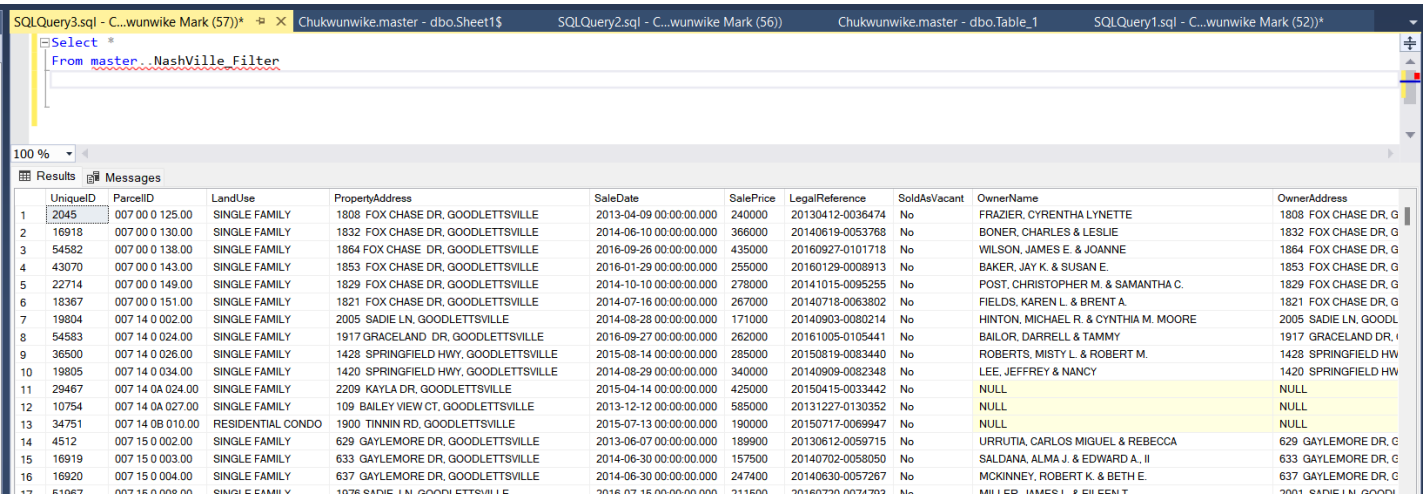


# SQL ‘Data Cleaning’ Housing Portfolio Project

## Data Cleaning Process:

In the realm of real estate and housing projects, data integrity is paramount for making informed decisions. As a data analyst tasked with cleaning and organizing housing data for a project, I utilized various SQL queries to ensure the accuracy and consistency of the dataset. The following report outlines the process undertaken to clean and refine the housing data, highlighting the methods employed and the resulting improvements in data quality.

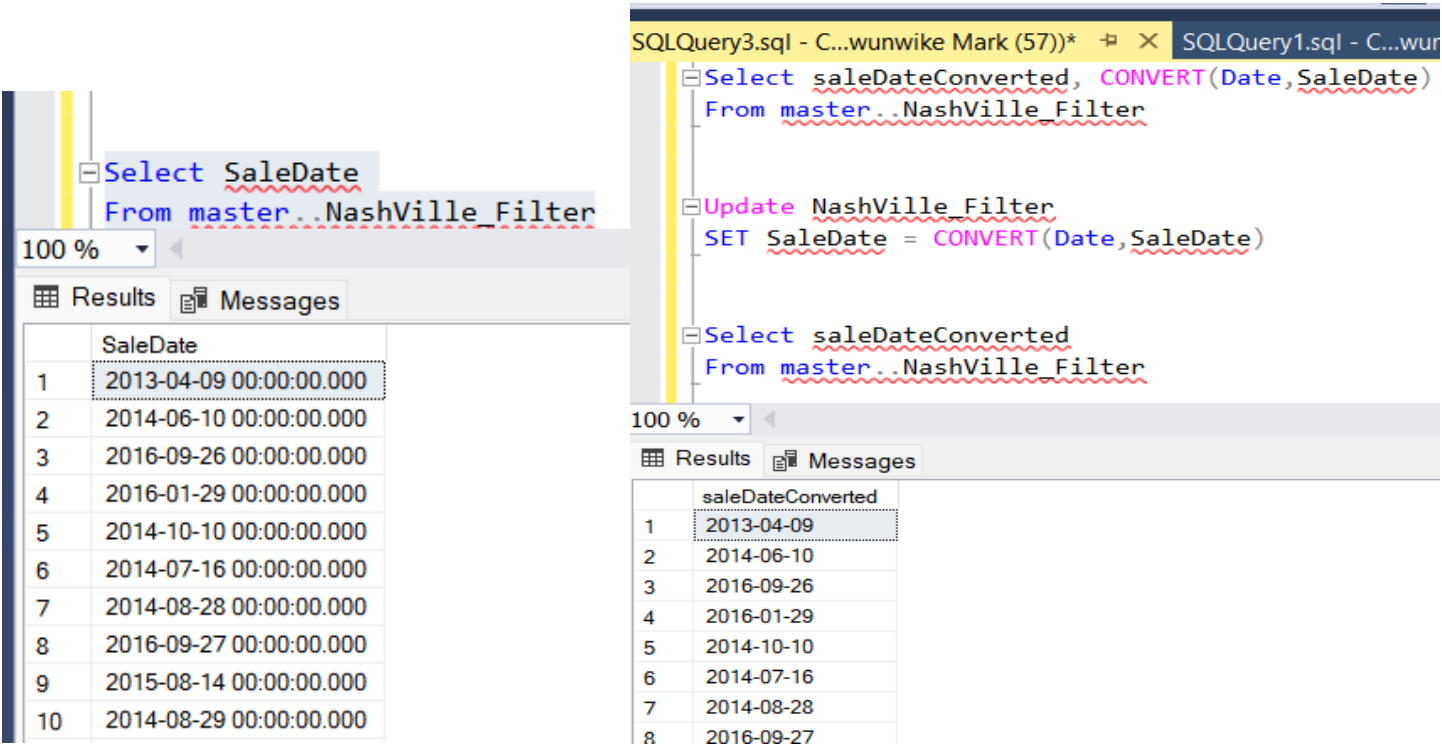
### 1. Selecting Data :



	UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	OwnerName	OwnerAddress
1	2045	007 00 0 125.00	SINGLE FAMILY	1808 FOX CHASE DR, GOODLETTSVILLE	2013-04-09 00:00:00.000	240000	20130412-0036474	No	FRAZIER, CYRENTHA LYNETTE	1808 FOX CHASE DR, G
2	16918	007 00 0 130.00	SINGLE FAMILY	1832 FOX CHASE DR, GOODLETTSVILLE	2014-06-10 00:00:00.000	366000	20140619-0053768	No	BONER, CHARLES & LESLIE	1832 FOX CHASE DR, G
3	54582	007 00 0 138.00	SINGLE FAMILY	1864 FOX CHASE DR, GOODLETTSVILLE	2016-09-26 00:00:00.000	435000	20160927-0101718	No	WILSON, JAMES E & JOANNE	1864 FOX CHASE DR, G
4	43070	007 00 0 143.00	SINGLE FAMILY	1853 FOX CHASE DR, GOODLETTSVILLE	2016-01-29 00:00:00.000	255000	20160129-0008913	No	BAKER, JAY K & SUSAN E.	1853 FOX CHASE DR, G
5	22714	007 00 0 149.00	SINGLE FAMILY	1829 FOX CHASE DR, GOODLETTSVILLE	2014-10-10 00:00:00.000	278000	20141015-0095255	No	POST, CHRISTOPHER M. & SAMANTHA C.	1829 FOX CHASE DR, G
6	18367	007 00 0 151.00	SINGLE FAMILY	1821 FOX CHASE DR, GOODLETTSVILLE	2014-07-16 00:00:00.000	267000	20140718-0063802	No	FIELDS, KAREN L. & BRENT A.	1821 FOX CHASE DR, G
7	19804	007 14 0 002.00	SINGLE FAMILY	2005 SADIE LN, GOODLETTSVILLE	2014-08-28 00:00:00.000	171000	20140903-0080214	No	HINTON, MICHAEL R. & CYNTHIA M. MOORE	2005 SADIE LN, GOODL
8	54583	007 14 0 024.00	SINGLE FAMILY	1917 GRACELAND DR, GOODLETTSVILLE	2016-09-27 00:00:00.000	262000	20161005-0105441	No	BAILOR, DARRELL & TAMMY	1917 GRACELAND DR, I
9	36500	007 14 0 026.00	SINGLE FAMILY	1428 SPRINGFIELD HWY, GOODLETTSVILLE	2015-08-14 00:00:00.000	285000	20150819-0083440	No	ROBERTS, MISTY L. & ROBERT M.	1428 SPRINGFIELD HW
10	19805	007 14 0 034.00	SINGLE FAMILY	1420 SPRINGFIELD HWY, GOODLETTSVILLE	2014-08-29 00:00:00.000	340000	20140909-0082348	No	LEE, JEFFREY & NANCY	1420 SPRINGFIELD HW
11	29467	007 14 0A 024.00	SINGLE FAMILY	2209 KAYLA DR, GOODLETTSVILLE	2015-04-14 00:00:00.000	425000	20150415-0033442	No	NULL	NULL
12	10754	007 14 0A 027.00	SINGLE FAMILY	109 BAILEY VIEW CT, GOODLETTSVILLE	2013-12-12 00:00:00.000	585000	20131227-0130352	No	NULL	NULL
13	34751	007 14 0B 010.00	RESIDENTIAL CONDO	1900 TINNIN RD, GOODLETTSVILLE	2015-07-13 00:00:00.000	190000	20150717-0069947	No	NULL	NULL
14	4512	007 15 0 002.00	SINGLE FAMILY	629 GAYLEMORE DR, GOODLETTSVILLE	2013-06-07 00:00:00.000	189900	20130612-0059715	No	URRUTIA, CARLOS MIGUEL & REBECCA	629 GAYLEMORE DR, G
15	16919	007 15 0 003.00	SINGLE FAMILY	633 GAYLEMORE DR, GOODLETTSVILLE	2014-06-30 00:00:00.000	157500	20140702-0058050	No	SALDANA, ALMA J. & EDWARD A., II	633 GAYLEMORE DR, G
16	16920	007 15 0 004.00	SINGLE FAMILY	637 GAYLEMORE DR, GOODLETTSVILLE	2014-06-30 00:00:00.000	247400	20140630-0057267	No	MCKINNEY, ROBERT K. & BETH E.	637 GAYLEMORE DR, G
17	51967	007 15 0 009.00	SINGLE FAMILY	1076 SADIE LN, GOODLETTSVILLE	2016-07-16 00:00:00.000	211500	20160719-0074703	No	MILLER, JAMES I. & DEBENT	1076 SADIE LN, GOODL

As I embark upon the exploration of Nashville's housing landscape, akin to a scholarly archaeologist, I initiate my inquiry with a fundamental SQL query. This initial step serves as the gateway to a rich repository of housing data, promising invaluable insights awaiting meticulous examination. I start by checking out all the housing info in Nashville.

### 2. Standardize Date Format :



```
Select SaleDate
From master..NashVille_Filter
```

	SaleDate
1	2013-04-09 00:00:00.000
2	2014-06-10 00:00:00.000
3	2016-09-26 00:00:00.000
4	2016-01-29 00:00:00.000
5	2014-10-10 00:00:00.000
6	2014-07-16 00:00:00.000
7	2014-08-28 00:00:00.000
8	2016-09-27 00:00:00.000
9	2015-08-14 00:00:00.000
10	2014-08-29 00:00:00.000

```
Select saleDateConverted, CONVERT(Date, SaleDate)
From master..NashVille_Filter

Update Nashville_Filter
SET SaleDate = CONVERT(Date, SaleDate)

Select saleDateConverted
From master..NashVille_Filter
```

	saleDateConverted
1	2013-04-09
2	2014-06-10
3	2016-09-26
4	2016-01-29
5	2014-10-10
6	2014-07-16
7	2014-08-28
8	2016-09-27

With a conscientious approach reminiscent of a meticulous curator, I endeavour to standardize the myriad dates within the dataset. The initial dataset contained date values that needed conversion to a standardized format. I utilized the `CONVERT` function to transform the `SaleDate` column into a

consistent date format. I make sure all the dates in the data look the same. It's like tidying up a messy room, making everything neat and easy to understand.

### 3. Populate Property Address Data :

```
Update a
SET PropertyAddress = ISNULL(a.PropertyAddress,b.PropertyAddress)
From master..NashVille_Filter a
JOIN master..NashVille_Filter b
on a.ParcelID = b.ParcelID
AND a.[UniqueID ] <> b.[UniqueID ]
Where a.PropertyAddress is null

Select PropertyAddress
From NashVille_Filter
```

```
Select a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress, ISNULL(a.PropertyAddress, b.PropertyAddress)
From master..NashVille_Filter a
JOIN master..NashVille_Filter b
on a.ParcelID = b.ParcelID
AND a.[UniqueID ] <> b.[UniqueID ]
Where a.PropertyAddress is null
```

ParcelID	PropertyAddress	ParcelID	PropertyAddress	(No column name)
025 07 0 031.00	NULL	025 07 0 031.00	410 ROSEHILL CT, GOODLETTSVILLE	410 ROSEHILL CT, GOODLETTSVILLE
026 01 0 069.00	NULL	026 01 0 069.00	141 TWO MILE PIKE, GOODLETTSVILLE	141 TWO MILE PIKE, GOODLETTSVILLE
026 05 0 017.00	NULL	026 05 0 017.00	208 EAST AVE, GOODLETTSVILLE	208 EAST AVE, GOODLETTSVILLE
026 06 0A 038.00	NULL	026 06 0A 038.00	109 CANTON CT, GOODLETTSVILLE	109 CANTON CT, GOODLETTSVILLE
033 06 0 041.00	NULL	033 06 0 041.00	1129 CAMPBELL RD, GOODLETTSVILLE	1129 CAMPBELL RD, GOODLETTSVILLE
033 06 0A 002.00	NULL	033 06 0A 002.00	1116 CAMPBELL RD, GOODLETTSVILLE	1116 CAMPBELL RD, GOODLETTSVILLE
033 15 0 123.00	NULL	033 15 0 123.00	438 W CAMPBELL RD, GOODLETTSVILLE	438 W CAMPBELL RD, GOODLETTSVILLE
114 15 0A 030.00	NULL	114 15 0A 030.00	109 CEDAR PLACE BND, NASHVILLE	109 CEDAR PLACE BND, NASHVILLE
034 03 0 059.00	NULL	034 03 0 059.00	2117 PAULA DR, MADISON	2117 PAULA DR, MADISON

```
Update a
SET PropertyAddress = ISNULL(a.PropertyAddress,b.PropertyAddress)
From master..NashVille_Filter a
JOIN master..NashVille_Filter b
on a.ParcelID = b.ParcelID
AND a.[UniqueID ] <> b.[UniqueID ]
Where a.PropertyAddress is null

Select PropertyAddress
From NashVille_Filter
```

PropertyAddress
1808 FOX CHASE DR, GOODLETTSVILLE
1832 FOX CHASE DR, GOODLETTSVILLE
1864 FOX CHASE DR, GOODLETTSVILLE
1853 FOX CHASE DR, GOODLETTSVILLE
1829 FOX CHASE DR, GOODLETTSVILLE
1821 FOX CHASE DR, GOODLETTSVILLE
2005 SADIE LN, GOODLETTSVILLE
1917 GRACELAND DR, GOODLETTSVILLE
1428 SPRINGFIELD HWY, GOODLETTSVILLE
1420 SPRINGFIELD HWY, GOODLETTSVILLE
2209 KAYLA DR, GOODLETTSVILLE
109 BAILEY VIEW CT, GOODLETTSVILLE
1900 TINNIN RD, GOODLETTSVILLE
629 GAYLEMORE DR, GOODLETTSVILLE
633 GAYLEMORE DR, GOODLETTSVILLE
637 GAYLEMORE DR, GOODLETTSVILLE
1976 SADIE LN, GOODLETTSVILLE
644 GAYLEMORE DR, GOODLETTSVILLE
1921 NORMERLE DR, GOODLETTSVILLE
1916 NORMERLE DR, GOODLETTSVILLE
2050 GRACELAND DR, GOODLETTSVILLE
2034 GRACELAND DR, GOODLETTSVILLE
811 BENTON CT, GOODLETTSVILLE
331 VIEW RIDGE DR, GOODLETTSVILLE

My scholarly pursuit of completeness and accuracy compels me to address gaps within the dataset with a sense of scholarly rigour, I meticulously populate missing property addresses, thus fostering a comprehensive understanding of Nashville's housing landscape. Some entries in the dataset lacked property addresses. To address this issue, I performed a self-join operation on the dataset, comparing entries with the same `ParcelID` but different `UniqueID`. Using the `ISNULL` function, I updated the missing `PropertyAddress` values with existing ones from the joined dataset.

### 4. Breaking out Addresses into Individual Columns:

```
Select
SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress) - 1) as Address
From NashVille_Filter

Select
SUBSTRING(PropertyAddress, 1, CHARINDEX(',', PropertyAddress) - 1) as Address
, SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress) +1, LEN(PropertyAddress)) as Address
From master..NashVille_Filter
```

```
ALTER TABLE NashVille_Filter
Add PropertySplitAddress Nvarchar(255);

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

ALTER TABLE NashVille_Filter
Add PropertySplitCity Nvarchar(255);

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

Address	Address
1808 FOX CHASE DR	GOODLETTSVILLE
1832 FOX CHASE DR	GOODLETTSVILLE
1864 FOX CHASE DR	GOODLETTSVILLE
1853 FOX CHASE DR	GOODLETTSVILLE
1829 FOX CHASE DR	GOODLETTSVILLE
1821 FOX CHASE DR	GOODLETTSVILLE
2005 SADIE LN	GOODLETTSVILLE
1917 GRACELAND DR	GOODLETTSVILLE
1428 SPRINGFIELD HWY	GOODLETTSVILLE
1420 SPRINGFIELD HWY	GOODLETTSVILLE

Land/Value	Building/Value	Total/Value	Year/Built	Bedrooms	FullBath	HalfBath	SaleDate/Converted	PropertySplitAddress	PropertySplitCity	
RVICES DISTRICT	50000	168200	235700	1986	3	3	0	2013-04-09	1808 FOX CHASE DR	GOODLETTSVILLE
RVICES DISTRICT	50000	264100	319000	1998	3	3	2	2014-06-10	1832 FOX CHASE DR	GOODLETTSVILLE
RVICES DISTRICT	50000	216200	298000	1987	4	3	0	2016-09-26	1864 FOX CHASE DR	GOODLETTSVILLE
RVICES DISTRICT	50000	147300	197300	1985	3	3	0	2016-01-29	1853 FOX CHASE DR	GOODLETTSVILLE
RVICES DISTRICT	50000	152300	202300	1984	4	3	0	2014-10-10	1829 FOX CHASE DR	GOODLETTSVILLE
RVICES DISTRICT	50000	190400	259800	1980	3	3	0	2014-07-16	1821 FOX CHASE DR	GOODLETTSVILLE
RVICES DISTRICT	40000	137900	177900	1976	3	2	0	2014-08-28	2005 SADIE LN	GOODLETTSVILLE
RVICES DISTRICT	40000	157900	197900	1978	3	2	0	2016-09-27	1917 GRACELAND DR	GOODLETTSVILLE
RVICES DISTRICT	45400	176900	222300	2000	3	2	1	2015-08-14	1428 SPRINGFIELD HWY	GOODLETTSVILLE
RVICES DISTRICT	40000	179600	219600	1995	5	3	0	2015-08-29	1420 SPRINGFIELD HWY	GOODLETTSVILLE
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2015-04-14	2209 KAYLA DR	GOODLETTSVILLE
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2013-12-12	109 BAILEY VIEW CT	GOODLETTSVILLE
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2015-07-13	1900 TINNIN RD	GOODLETTSVILLE
RVICES DISTRICT	40000	203600	243600	2006	4	4	0	2013-06-07	629 GAYLEMORE DR	GOODLETTSVILLE
RVICES DISTRICT	40000	115400	156400	1974	4	2	0	2014-06-30	633 GAYLEMORE DR	GOODLETTSVILLE
RVICES DISTRICT	40000	180300	234000	1974	3	3	0	2014-06-30	637 GAYLEMORE DR	GOODLETTSVILLE
RVICES DISTRICT	40000	129200	169200	1974	4	2	0	2016-07-15	1976 SADIE LN	GOODLETTSVILLE

To further refine the dataset, I split the `PropertyAddress` column into separate components for address and city. This was achieved using the `SUBSTRING` function along with `CHARINDEX` to identify and extract the desired segments of the address. It's like breaking down a big problem into smaller, more manageable pieces.

5. Changing Y and N to Yes and No:

SQLQuery3.sql - C...wunwike Mark (57))\*

SQLQuery1.sql - C...wunwike Mark (52)

Update Nashville\_Filter

SET SoldAsVacant = CASE When SoldAsVacant = 'Y' THEN 'Yes'  
When SoldAsVacant = 'N' THEN 'No'  
ELSE SoldAsVacant  
END

100 %

Results Messages

	Price	LegalReference	SoldAsVacant	OwnerName
29...	000	20160825-0089001	No	DRINKWATER, KATHERINE & WILLIAM WOODS
29...	000	20141218-0115774	No	JONES, WADE L.
29...	000	20131023-0110373	No	DONKIN, GEOFFREY & MOE, JULIE
29...	000	20130828-0090677	No	NULL
29...	000	20130910-0095391	No	MOORE, REMICK F.
29...	000	20140715-0062553	No	SIMI, MICHAEL L. & ROBIN S.
29...	500	20141119-0106684	No	COOK, JASON
29...	000	20160201-0009689	No	URBANGATE DEVELOPMENT GROUP, LLC
29...	000	20151117-0116713	No	BUDA, NICHOLAS & BOSSHART, ASHLEY
29...	000	20140610-0050412	No	HOLTE, RYAN THOMAS
29...	500	20150130-0008864	No	JAMI, NAJMEH
29...	00	20130709-0070780	Yes	NULL
29...	000	20150406-0029403	No	FAVRET, ANDRE G. & LACIE HURST
29...	500	20150331-0027456	No	HAHN, THOMAS & TEEPLE, CHRISTINA

SQLQuery3.sql - C...wunwike Mark (57))\*

SQLQuery1.sql - C...wunwike Mark (52)

Select Distinct(SoldAsVacant), Count(SoldAsVacant)

From Nashville\_Filter

Group by SoldAsVacant

order by 2

100 %

Results Messages

	SoldAsVacant	(No column name)
1	Y	52
2	N	399
3	Yes	4623
4	No	51403

Employing linguistic precision reminiscent of a semantic scholar, I transmute coded values into universally understandable terms. ` The 'SoldAsVacant' column contained values denoted by 'Y' and 'N', which I standardized to 'Yes' and 'No', respectively, using a `CASE` statement. I foster clarity and comprehension within the dataset, akin to translating ancient inscriptions for scholarly interpretation.

6. Remove Duplicates:

SQLQuery3.sql - C...wunwike Mark (57))\*

SQLQuery1.sql - C...wunwike Mark (52))\*

WITH RowNumCTE AS(  
Select \*,  
ROW\_NUMBER() OVER (  
PARTITION BY ParcelID,  
PropertyAddress,  
SalePrice,  
SaleDate,  
LegalReference  
ORDER BY  
UniqueID  
) row\_num  
  
From Nashville\_Filter  
--order by ParcelID  
)  
Select \*  
From RowNumCTE  
Where row\_num > 1  
Order by PropertyAddress

100 %

Results Messages

slue	TotalValue	YearBuilt	Bedrooms	FullBath	HalfBath	SaleDateConverted	PropertySplitAddress	PropertySplitCity	row_num
91	NULL	NULL	NULL	NULL	NULL	2015-02-19	6203 NEW YORK AVE	NASHVILLE	2
92	NULL	NULL	NULL	NULL	NULL	2015-02-20	6829 HICKORY RIM CT	ANTIOCH	2
93	NULL	NULL	NULL	NULL	NULL	2015-02-17	713 44TH AVE N	NASHVILLE	2
94	120800	1969	2	1	0	2015-02-13	743 CROLEY DR	NASHVILLE	2
95	NULL	NULL	NULL	NULL	NULL	2015-02-13	813 POINT BREAK CR S	ANTIOCH	2
96	62400	1960	2	1	0	2015-02-20	815 31ST AVE N	NASHVILLE	2
97	129500	1987	3	2	0	2015-02-26	820 DRESDEN CT	ANTIOCH	2
98	NULL	NULL	NULL	NULL	NULL	2015-02-13	8528 BEAUTIFUL VALLEY ...	NASHVILLE	2
99	NULL	NULL	NULL	NULL	NULL	2015-02-19	8553 BEAUTIFUL VALLEY ...	NASHVILLE	2
100	NULL	NULL	NULL	NULL	NULL	2014-02-03	904 BIRCHMILL PT S	ANTIOCH	2
101	95500	1954	2	1	0	2015-02-13	906 DELRAY DR	NASHVILLE	2
102	NULL	NULL	NULL	NULL	NULL	2015-02-06	908 TOWNVIEW PL	NASHVILLE	2
103	NULL	NULL	NULL	NULL	NULL	2014-02-05	926 COARSEY DR	NASHVILLE	2
104	111600	1958	2	1	0	2015-02-13	934 DRUMMOND DR	NASHVILLE	2

SQLQuery3.sql - C...wunwike Mark (57))\*

SQLQuery1.sql - C...wunwike Mark (52))\*

WITH RowNumCTE AS(  
Select \*,  
ROW\_NUMBER() OVER (  
PARTITION BY ParcelID,  
PropertyAddress,  
SalePrice,  
SaleDate,  
LegalReference  
ORDER BY  
UniqueID  
) row\_num  
  
From Nashville\_Filter  
--order by ParcelID  
)  
Delete  
From RowNumCTE  
Where row\_num > 1  
--Order by PropertyAddress

100 %

Messages

(104 rows affected)

Completion time: 2024-03-12T15:44:07.3580006+01:00

With the discerning eye of a scholarly critic, I discern and expunge duplicative entries within the dataset. Duplicate entries can skew analysis results. I identified and removed duplicates based on a combination of key columns (`ParcelID`, `PropertyAddress`, `SalePrice`, `SaleDate`, and `LegalReference`) using the `ROW\_NUMBER` function within a Common Table Expression (CTE).

8. Delete Unused Columns:

SQLQuery3.sql - C...wunwike Mark (57))\*SQLQuery1.sql - C...wunwike Mark (52))\*

```
--order by ParcelID
)
Delete
From RowNumCTE
Where row_num > 1
--Order by PropertyAddress

Select *
From master..NashVille_Filter

ALTER TABLE NashVille_Filter
DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress, SaleDate
```

100 %

ResultsMessages

		OwnerAddress	Acreage	TaxDistrict	LandValue	BuildingValue
1	LYNETTE	1808 FOX CHASE DR, GOODLETTSVILLE, TN	2.3	GENERAL SERVICES DISTRICT	50000	168200
2	LESLIE	1832 FOX CHASE DR, GOODLETTSVILLE, TN	3.5	GENERAL SERVICES DISTRICT	50000	264100
3	JOANNE	1864 FOX CHASE DR, GOODLETTSVILLE, TN	2.9	GENERAL SERVICES DISTRICT	50000	216200
4	AN E.	1853 FOX CHASE DR, GOODLETTSVILLE, TN	2.6	GENERAL SERVICES DISTRICT	50000	147300
5	3 M. & SAMANTHA C.	1829 FOX CHASE DR, GOODLETTSVILLE, TN	2	GENERAL SERVICES DISTRICT	50000	152300
6	3RENT A.	1821 FOX CHASE DR, GOODLETTSVILLE, TN	2	GENERAL SERVICES DISTRICT	50000	190400
7	& CYNTHIA M. MOORE	2005 SADIE LN, GOODLETTSVILLE, TN	1.03	GENERAL SERVICES DISTRICT	40000	137900
8	TAMMY	1917 GRACELAND DR, GOODLETTSVILLE, TN	1.03	GENERAL SERVICES DISTRICT	40000	157900
9	3 ROBERT M.	1428 SPRINGFIELD HWY, GOODLETTSVILLE, TN	1.67	GENERAL SERVICES DISTRICT	45400	176900
10	CY	1420 SPRINGFIELD HWY, GOODLETTSVILLE, TN	1.3	GENERAL SERVICES DISTRICT	40000	179600
11		NULL	NULL	NULL	NULL	NULL
12		NULL	NULL	NULL	NULL	NULL

```
Select *
From master..NashVille_Filter

ALTER TABLE NashVille_Filter
DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress, SaleDate
```

%

ResultsMessages

	OwnerName	Acreage	LandValue	BuildingValue	TotalValue	YearBuilt	Bedrooms	FullBath	Half
	FRAZIER, CYRENTHA LYNETTE	2.3	50000	168200	235700	1986	3	3	0
	BONER, CHARLES & LESLIE	3.5	50000	264100	319000	1998	3	3	2
	WILSON, JAMES E. & JOANNE	2.9	50000	216200	298000	1987	4	3	0
	BAKER, JAY K. & SUSAN E.	2.6	50000	147300	197300	1985	3	3	0
	POST, CHRISTOPHER M. & SAMANTHA C.	2	50000	152300	202300	1984	4	3	0
	FIELDS, KAREN L. & BRENT A.	2	50000	190400	259800	1980	3	3	0

Finally, unnecessary columns ( `OwnerAddress`, `TaxDistrict`, `PropertyAddress`, `SaleDate` ) were dropped from the dataset using the `ALTER TABLE` statement.

Conclusion

Through systematic data cleaning procedures, I successfully enhanced the quality and reliability of the housing dataset. By standardizing formats, handling missing values, splitting addresses, and removing duplicates, the dataset now provides a solid foundation for meaningful analysis and decision-making in the housing project. These cleaning tasks demonstrate my proficiency in SQL and data manipulation, ensuring accurate insights for informed decision-making.