NASHVILLE HOUSING – SQL DATA CLEANING #1

PROJECT PURPOSE: *To clean and prepare the Nashville Housing dataset so that it can be ready for analysis. Cleaning and manipulation included handling duplicates and null values, creating new columns to break up multi-part fields and handling dates.*

What is shown below is a portion of this project, specifically dealing with the handling of strings (names) in SQL. The full Nashville Housing SQL project (SQL Data Cleaning #1) can be found on my Github.

Once completed a Tableau dashboard was created using a series of SQL queries from the cleaned data.

Objectives:

- Handling Names (Data Manipulation)
- Final clean-up

Handling Names

Looking at the OwnerName column, it was determined that each owner name is separated by the ampersand ('&') character. Since this is the case, the '&' will be used as the primary delimiter to separate each owner in the OwnerName column. It is also seen that the full name of each owner is structured like (***Last Name, First Name***). This information will be important later when constructing each individual owner name later.

To make this process more manageable, I will divide the work based on the number of owners for each property (whether a property has one owner, two owners, etc.). Since the '&' separates each owner in the OwnerName field, the number of instances the '&' shows up in each field will determine how many owners each property has (e.g., 1 instance of '&' means two owners, 2 instances of '&' means three owners, and so on).

OwnerName
FIELDS, KAREN L. & BRENT A.
BAILOR, DARRELL & TAMMY
LEE, JEFFREY & NANCY
NULL
URRUTIA, CARLOS MIGUEL & REBECCA
SALDANA, ALMA J. & EDWARD A., II
SANDAGE, LEAH S.
BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE
WILLIAMS, JOSEPH D. & AUDREY
ADAMS, ALEXANDER J. & JENNIFER

only first 10 rows shown

Finding the number of owners for each property based on number of instances '&' is present

Selecting OwnerName where '&' does not occur

```
SELECT OwnerName
FROM housingcopy
WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 0;
```

OwnerName
SANDAGE, LEAH S.
BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE
TUDERS, KEVIN
PORTER, BRIAN KELLY
TEMPLE, LORENE
STEWART, DONNA D.
SMITH, HOLLI NICHOLE
DAHLHAUSER, ANDREW FRANCIS
MILOM, KATHEY
CONNER, BRENDA

only first 10 rows shown

Notice how '&' does not show and there is only one owner showing.

Selecting OwnerName where '&' occurs once (indicating 2 owners)

```
SELECT OwnerName
FROM housingcopy
WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 1;
```

OwnerName
FIELDS, KAREN L. & BRENT A.
BAILOR, DARRELL & TAMMY
LEE, JEFFREY & NANCY
URRUTIA, CARLOS MIGUEL & REBECCA
SALDANA, ALMA J. & EDWARD A., II
WILLIAMS, JOSEPH D. & AUDREY
ADAMS, ALEXANDER J. & JENNIFER
O'CONNELL, JAMES L. & VANESSA R.
BOISE, LINDSEY J. & BURDEN, GINGER E.
PROULX, JENNIFER L. & JESSICA

only first 10 rows shown

Notice how '&' (highlighted in red text) appears once and that it separates two names (owners).

Selecting OwnerName where '&' occurs twice (indicating 3 owners)

```
SELECT OwnerName
FROM housingcopy
WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 2;
```

OwnerName
GAINS, JEFFREY A. & LISA M. & SNYDER, BARBARA
BYERS, FRANK & KATHY & WHITEHORN, CLAUDIA
PYATT, JUDSON E. & LAURA & MAUS, NANCY S.
MILLER, DALE R. & CORA L. & SHELTON, BRENDA
RUTHERFORD, STEVEN & GIA A. & ROBERT
RESIDE, ANDREA R. & KAREN & DONALD B.
APPLETON, ZAKAIA S. & RANSOM, MELLNEE & DILLARD, SHAMAR
LAVENDER, KAREN & RAVENSCROFT, GARY & MORSHEAD, ROBYN
ARNOLD, RACHEL & SPARKMAN, JIMMY & SANDRA
HILL, RAYMOND & KILPATRICK, MARK W. & DAWN H.

only first 10 rows shown

Two instances of '&' that separates three names (owners).

Selecting OwnerName where '&' occurs three times (indicating 4 owners)

```
SELECT OwnerName
FROM housingcopy
WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 3;
```

OwnerName
WINGO, ROBERT J. & JUDY S. & ROBERT J., JR. & AMY GIPSON
BREWSTER, ANTHONY JR & VICTORIA & FLACK, JONATHON & JENNIFER
SAM, CHHOEUN & JODECI & HAS, KOU & SOUNG, SINAK
REIGLE, DAN & TINA (LE) & REIGLE, STEVEN & STACY
HAMPTON, JUSTIN W. & ANSLEY P. & ERNEST W. & DOROTHY
REEVES, CLEDUE S. & SHERRY & DION & CHRISTY
REEVES, SHERMAN & SHERRY & DION & CHRISTY
PETTY, BRADLEY T. & BRIAN D. & K & T ASSOCIATES, L.P.
PETTY, BRADLEY L. & BRIAN D. & K & T ASSOCIATIES, L.P.
HOLL, BRIAN D. & GERALD R. & LYNN M. & LOEHRER, KATHERINE

only first 10 rows shown

Three instances of '&' that separates four names (owners). The '&' did not exceed 3 times in any row, so four owners are the max for any property in the dataset. Note: Owner names that are like LLC, L.P., TRUST, etc. will be treated as one owner.

Plan of Execution

1) Create queries that find OwnerName rows that have one, two, three and four owners separately and divide the names into their own columns.
2) Run the results of each of those CTEs into temp tables.
3) Add owner1, owner2, owner3 and owner4 columns to the main housing table.
4) Join the CTEs together into a single table and then use that table to join on the main housing table to fill the owner columns where the OwnerName columns match.

Creating one owner temp table

```
/* FINDING THE OWNERS WHERE '&' DOES NOT OCCUR ONCE -- INDICATES ONLY ONE OWNER THE FOR THE PROPERTY */

SELECT OwnerName, OwnerName AS 'owner1' INTO #one_owner
FROM housingcopy
WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 0 OR
OwnerName IN ('%ASSOCIATE%','%LLC%','%INVESTMENTS%','%L.P.%','%TRUST%','%INC%','%GOV''T%');

-- INCLUDING ANY REFERENCES TO COMPANY OWNERSHIP (INC, LLC, ETC.) --
```

OwnerName	owner1
SANDAGE, LEAH S.	SANDAGE, LEAH S.
BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE	BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE
TUDERS, KEVIN	TUDERS, KEVIN
PORTER, BRIAN KELLY	PORTER, BRIAN KELLY
TEMPLE, LORENE	TEMPLE, LORENE
STEWART, DONNA D.	STEWART, DONNA D.
SMITH, HOLLI NICHOLE	SMITH, HOLLI NICHOLE
DAHLHAUSER, ANDREW FRANCIS	DAHLHAUSER, ANDREW FRANCIS
MILOM, KATHEY	MILOM, KATHEY
CONNER, BRENDA	CONNER, BRENDA

first 10 rows shown above

Creating two owner temp table

```
/* FINDING THE FIRST AND SECOND OWNERS WHERE '&' OCCURS ONLY ONCE -- INDICATES TWO OWNERS FOR THE
PROPERTY */

WITH owners AS
      (SELECT OwnerName,
      REPLACE(LEFT(OwnerName, PATINDEX('%,%',OwnerName)),',','') AS 'last_name',
      SUBSTRING(OwnerName, 1, PATINDEX('%&%', OwnerName) - 1) AS 'first_owner',
      LTRIM(REPLACE(RIGHT(OwnerName, PATINDEX('%&%', REVERSE(OwnerName))),'&','')) AS 'second_name'
      FROM housingcopy
      WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 1
      AND OwnerName NOT LIKE '%ASSOCIATE%' AND OwnerName NOT LIKE '%LLC%' AND OwnerName NOT LIKE
'%INVESTMENTS%' AND OwnerName NOT LIKE '%L.P.%' AND OwnerName NOT LIKE '%TRUST%' AND OwnerName NOT LIKE
'%INC%' AND OwnerName NOT LIKE '%GOV''T%'),
      owners2 AS
            (SELECT OwnerName, first_owner AS 'owner1',
            CASE WHEN second_name NOT LIKE '%,%' THEN CONCAT(last_name,', ',second_name)
                  ELSE second_name
            END AS 'owner2' FROM owners)
SELECT * INTO #two_owners FROM owners2;
```

Breakdown of CTE (part 1):

- OwnerName from the main housing table (*housingcopy*)
- Everything left of the first ',' as the last name of the first owner
- Everything left of the '&' as the first owner (==*last name, first name*==)
- Everything right of the '&' as the second name

OwnerName	last_name	first_owner	second_name
FIELDS, KAREN L. & BRENT A.	FIELDS	FIELDS, KAREN L.	BRENT A.
BAILOR, DARRELL & TAMMY	BAILOR	BAILOR, DARRELL	TAMMY
LEE, JEFFREY & NANCY	LEE	LEE, JEFFREY	NANCY
URRUTIA, CARLOS MIGUEL & REBECCA	URRUTIA	URRUTIA, CARLOS MIGUEL	REBECCA
SALDANA, ALMA J. & EDWARD A., II	SALDANA	SALDANA, ALMA J.	EDWARD A., II
WILLIAMS, JOSEPH D. & AUDREY	WILLIAMS	WILLIAMS, JOSEPH D.	AUDREY
ADAMS, ALEXANDER J. & JENNIFER	ADAMS	ADAMS, ALEXANDER J.	JENNIFER
O'CONNELL, JAMES L. & VANESSA R.	O'CONNELL	O'CONNELL, JAMES L.	VANESSA R.
BOISE, LINDSEY J. & BURDEN, GINGER E.	BOISE	BOISE, LINDSEY J.	BURDEN, GINGER E.
PROULX, JENNIFER L. & JESSICA	PROULX	PROULX, JENNIFER L.	JESSICA

Breakdown of CTE (part 2):

- Select OwnerName from first part of CTE
- Select the first owner from first part of CTE
- Assigning the second owner by checking the second name field from the first part of CTE for an instance of a ','
  using a case statement. Logic: If a ',' is not present in the second name field then it only contains the second
  owner's first name. If this is true, then the value will be concatenated to the value showing in the last name
  field. Anything else leave as is as the second owner. *Not perfect but works with most names. Results of final
  CTE below.

OwnerName	owner1	owner2
FIELDS, KAREN L. & BRENT A.	FIELDS, KAREN L.	FIELDS, BRENT A.
BAILOR, DARRELL & TAMMY	BAILOR, DARRELL	BAILOR, TAMMY
LEE, JEFFREY & NANCY	LEE, JEFFREY	LEE, NANCY
URRUTIA, CARLOS MIGUEL & REBECCA	URRUTIA, CARLOS MIGUEL	URRUTIA, REBECCA
SALDANA, ALMA J. & EDWARD A., II	SALDANA, ALMA J.	EDWARD A., II
WILLIAMS, JOSEPH D. & AUDREY	WILLIAMS, JOSEPH D.	WILLIAMS, AUDREY
ADAMS, ALEXANDER J. & JENNIFER	ADAMS, ALEXANDER J.	ADAMS, JENNIFER
O'CONNELL, JAMES L. & VANESSA R.	O'CONNELL, JAMES L.	O'CONNELL, VANESSA R.
BOISE, LINDSEY J. & BURDEN, GINGER E.	BOISE, LINDSEY J.	BURDEN, GINGER E.
PROULX, JENNIFER L. & JESSICA	PROULX, JENNIFER L.	PROULX, JESSICA

first 10 rows shown above

Creating three owner temp table

```sql
/* FINDING THE FIRST, SECOND, AND THIRD OWNERS WHERE '&' OCCURS TWICE -- INDICATES THREE OWNERS FOR THE
PROPERTY */ --

WITH owners3 AS
        (SELECT OwnerName,
        REPLACE(LEFT(OwnerName, PATINDEX('%,%',OwnerName)),',','') AS 'last_name',
        SUBSTRING(OwnerName, 1, PATINDEX('%&%', OwnerName) - 1) AS 'first_owner',
        LTRIM(SUBSTRING(OwnerName, CHARINDEX('&',OwnerName)+1, CHARINDEX('&',OwnerName,
CHARINDEX('&',OwnerName) +1)-CHARINDEX('&',OwnerName)-1)) AS 'second_owner_fname',
        LTRIM(REPLACE(RIGHT(OwnerName, PATINDEX('%&%', REVERSE(OwnerName))),'&','')) AS 'third_name'
        FROM housingcopy
        WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 2
        AND OwnerName NOT LIKE '%ASSOCIATE%' AND OwnerName NOT LIKE '%LLC%' AND OwnerName NOT LIKE
'%INVESTMENTS%' AND OwnerName NOT LIKE '%L.P.%' AND OwnerName NOT LIKE '%TRUST%' AND OwnerName NOT LIKE
'%INC%' AND OwnerName NOT LIKE '%GOV''T%'),
owners4 AS
        (SELECT OwnerName, first_owner AS 'owner1',
        CASE WHEN second_owner_fname NOT LIKE '%,%' THEN CONCAT(last_name,', ',second_owner_fname)
                ELSE second_owner_fname
                END AS 'owner2',
        CASE WHEN third_name NOT LIKE '%,%' AND second_owner_fname NOT LIKE '%,%' THEN
CONCAT(last_name,', ',third_name)
                WHEN third_name NOT LIKE '%,%' AND second_owner_fname LIKE '%,%' THEN
CONCAT(REPLACE(LEFT(second_owner_fname, PATINDEX('%,%',second_owner_fname)),',',''),', ',third_name)
                ELSE third_name
                END AS 'owner3'
                FROM owners3)
SELECT * INTO #three_owners FROM owners4;
```

Breakdown of CTE (part 1):

- OwnerName from the main housing table (*housingcopy*)
- Everything left of the first comma (',') as the last name of the first owner
- Everything left of the first position of '&' as the first owner (==*last name, first name*==)
- Extract the string between the 1st and 2nd instance of '&' in the OwnerName as the second owners first name
- Extract the string right of the 2nd instance of '&' in the OwnerName as third name

OwnerName	last_name	first_owner	second_owner_fname	third_name
GAINS, JEFFREY A. & LISA M. & SNYDER, BARBARA	GAINS	GAINS, JEFFREY A.	LISA M.	SNYDER, BARBARA
BYERS, FRANK & KATHY & WHITEHORN, CLAUDIA	BYERS	BYERS, FRANK	KATHY	WHITEHORN, CLAUDIA
PYATT, JUDSON E. & LAURA & MAUS, NANCY S.	PYATT	PYATT, JUDSON E.	LAURA	MAUS, NANCY S.
MILLER, DALE R. & CORA L. & SHELTON, BRENDA	MILLER	MILLER, DALE R.	CORA L.	SHELTON, BRENDA
RUTHERFORD, STEVEN & GIA A. & ROBERT	RUTHERFORD	RUTHERFORD, STEVEN	GIA A.	ROBERT
RESIDE, ANDREA R. & KAREN & DONALD B.	RESIDE	RESIDE, ANDREA R.	KAREN	DONALD B.
APPLETON, ZAKAIA S. & RANSOM, MELLNEE & DILLARD, SHAMAR	APPLETON	APPLETON, ZAKAIA S.	RANSOM, MELLNEE	DILLARD, SHAMAR
LAVENDER, KAREN & RAVENSCROFT, GARY & MORSHEAD, ROBYN	LAVENDER	LAVENDER, KAREN	RAVENSCROFT, GARY	MORSHEAD, ROBYN
ARNOLD, RACHEL & SPARKMAN, JIMMY & SANDRA	ARNOLD	ARNOLD, RACHEL	SPARKMAN, JIMMY	SANDRA
HILL, RAYMOND & KILPATRICK, MARK W. & DAWN H.	HILL	HILL, RAYMOND	KILPATRICK, MARK W.	DAWN H.

Breakdown of CTE (part 2):

- Select OwnerName from the first part of the CTE
- Select the first owner from the first part of the CTE
- Assigning the second owner by checking the second name field from the first part of the CTE for an instance of a ','. **Logic:** If a ',' is not present in the second name field then it only contains the second owner's first name. If this is true, then the value will be concatenated to the value showing in the last name field. Anything else will be left as the second owner.
- Assigning the third owner by checking the third name field and second owner first name from the first part of the CTE for an instance of a ',' using a case statement. **Logic:** Should the third owner field not contain a ',' it means there is no last name. The first case statement checks for a ',' in both the third owner and second owner first name column, if none are present, then the value in the third owner field will be concatenated to the value showing in the last name.
- If there is a ',' present in the second owner first name field, then the string left of the ',' (indicating the second owners last name) will be concatenated to the value in the third name. **Logic:** If there is not a last name present in the third owner field or second owner first name field, then the third owner shares the same last name as the first owner. If the is a last name present in the second owner first name field, then the third owner shares the last name with the second owner. Note: Not perfect but works with most names. Results of final CTE below.

OwnerName	owner1	owner2	owner3
GAINS, JEFFREY A. & LISA M. & SNYDER, BARBARA	GAINS, JEFFREY A.	GAINS, LISA M.	SNYDER, BARBARA
BYERS, FRANK & KATHY & WHITEHORN, CLAUDIA	BYERS, FRANK	BYERS, KATHY	WHITEHORN, CLAUDIA
PYATT, JUDSON E. & LAURA & MAUS, NANCY S.	PYATT, JUDSON E.	PYATT, LAURA	MAUS, NANCY S.
MILLER, DALE R. & CORA L. & SHELTON, BRENDA	MILLER, DALE R.	MILLER, CORA L.	SHELTON, BRENDA
RUTHERFORD, STEVEN & GIA A. & ROBERT	RUTHERFORD, STEVEN	RUTHERFORD, GIA A.	RUTHERFORD, ROBERT
RESIDE, ANDREA R. & KAREN & DONALD B.	RESIDE, ANDREA R.	RESIDE, KAREN	RESIDE, DONALD B.
APPLETON, ZAKAIA S. & RANSOM, MELLNEE & DILLARD, SHAMAR	APPLETON, ZAKAIA S.	RANSOM, MELLNEE	DILLARD, SHAMAR
LAVENDER, KAREN & RAVENSCROFT, GARY & MORSHEAD, ROBYN	LAVENDER, KAREN	RAVENSCROFT, GARY	MORSHEAD, ROBYN
ARNOLD, RACHEL & SPARKMAN, JIMMY & SANDRA	ARNOLD, RACHEL	SPARKMAN, JIMMY	SPARKMAN, SANDRA
HILL, RAYMOND & KILPATRICK, MARK W. & DAWN H.	HILL, RAYMOND	KILPATRICK, MARK W.	KILPATRICK, DAWN H.

*first 10 rows shown above

Finding OwnerName rows with four owners

```sql
/* FINDING THE FIRST, SECOND, THIRD, FOURTH OWNERS WHERE '&' OCCURS 3 TIMES -- INDICATES FOUR OWNERS
FOR THE PROPERTY */

WITH owners5 AS
        (SELECT OwnerName,
        REPLACE(LEFT(OwnerName, PATINDEX('%,%',OwnerName)),',','') AS 'last_name',
        REPLACE(LEFT(OwnerName, PATINDEX('%&%',OwnerName)),'&','') AS 'first_owner',
        LTRIM(SUBSTRING(OwnerName, CHARINDEX('&',OwnerName)+1, CHARINDEX('&',OwnerName,
CHARINDEX('&',OwnerName) +1)-CHARINDEX('&',OwnerName)-1)) AS 'second_owner_fname',
        LTRIM(REPLACE(RIGHT(OwnerName, PATINDEX('%&%', REVERSE(OwnerName))),'&','')) AS 'fourth_owner'
        FROM housingcopy
        WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 3
        AND OwnerName NOT LIKE '%ASSOCIATE%' AND OwnerName NOT LIKE '%LLC%' AND OwnerName NOT LIKE
'%INVESTMENTS%' AND OwnerName NOT LIKE '%L.P.%' AND OwnerName NOT LIKE '%TRUST%' AND OwnerName NOT LIKE
'%INC%' AND OwnerName NOT LIKE '%GOV''T%'),
        third_name_position AS
                (SELECT OwnerName, CHARINDEX('&',OwnerName, (CHARINDEX('&',OwnerName,1))+1) AS 'position'
                FROM housingcopy
                WHERE LEN(OwnerName) - LEN(REPLACE(OwnerName,'&','')) = 3),
        third_name_position2 AS
                (SELECT OwnerName, SUBSTRING(OwnerName, position+1,LEN(OwnerName)-
CHARINDEX('&',OwnerName)) AS 'third_name'
                FROM third_name_position
                WHERE OwnerName NOT LIKE '%L.P.%'),
        third_name_position3 AS
                (SELECT OwnerName, LTRIM(REPLACE(LEFT(third_name, PATINDEX('%,%',third_name)),',','')) AS
'third_owner_lname', third_name
                FROM third_name_position2),
        owners6 AS
                (SELECT o.OwnerName, o.last_name, o.first_owner, REPLACE(LEFT(o.second_owner_fname,
PATINDEX('%,%',o.second_owner_fname)),',','') AS 'second_owner_lname',
                o.second_owner_fname, n.third_owner_lname, REPLACE(LEFT(n.third_name,
PATINDEX('%&%',n.third_name)),'&','') AS 'third_owner', o.fourth_owner
                FROM owners5 AS o
                JOIN third_name_position3 AS n
                ON o.OwnerName = n.OwnerName),
        owners7 AS
                (SELECT OwnerName, first_owner AS 'owner1',
        CASE WHEN second_owner_fname NOT LIKE '%,%' THEN CONCAT(last_name,', ',second_owner_fname)
                ELSE second_owner_fname
        END AS 'owner2',
        CASE WHEN third_owner NOT LIKE '%,%' AND third_owner_lname NOT LIKE '' AND third_owner_lname NOT
LIKE '%&%' THEN CONCAT(third_owner_lname,', ',third_owner)
                WHEN third_owner NOT LIKE '%,%' AND second_owner_lname NOT LIKE '' THEN
CONCAT(second_owner_lname,', ',third_owner)
                WHEN third_owner NOT LIKE '%,%' THEN CONCAT(last_name,', ',third_owner)
                ELSE third_owner
                END AS 'owner3',
        CASE WHEN fourth_owner NOT LIKE '%,%' AND third_owner_lname NOT LIKE '' AND third_owner_lname
NOT LIKE '%.%' AND third_owner_lname NOT LIKE '%&%' THEN CONCAT(third_owner_lname,', ',fourth_owner)
                WHEN fourth_owner NOT LIKE '%,%' AND second_owner_lname NOT LIKE '' THEN
CONCAT(second_owner_lname,', ',fourth_owner)
                WHEN fourth_owner NOT LIKE '%,%' THEN CONCAT(last_name,', ',fourth_owner)
                ELSE fourth_owner
                END AS 'owner4'
                FROM owners6)
SELECT * INTO #four_owners FROM owners7;
```

OwnerName	owner1	owner2	owner3	owner4
ADKINS, GREG & KIM & WILLIAMS, DALLAS & KEALIE	ADKINS, GREG	ADKINS, KIM	WILLIAMS, DALLAS	WILLIAMS, KEALIE
ALLEN, NEWTON & BURKLEY & MARTIN, ADAM & SARAH	ALLEN, NEWTON	ALLEN, BURKLEY	MARTIN, ADAM	MARTIN, SARAH
BEAM, STACY W. & AMANDA B. & GARY W. & OLETHIA B.	BEAM, STACY W.	BEAM, AMANDA B.	BEAM, GARY W.	BEAM, OLETHIA B.
BERRY, JOHN CHARLES & LINDA & BERRY, JOHN C. JR. & CORRIE	BERRY, JOHN CHARLES	BERRY, LINDA	BERRY, JOHN C. JR.	BERRY, CORRIE
BOUDREAUX, DANIELLE M, & MARK & LISA & KNOX, CALEB R.	BOUDREAUX, DANIELLE M,	BOUDREAUX, MARK	BOUDREAUX, LISA	KNOX, CALEB R.
BOWDEN, LESLIE M. & EARLINE & BASS, WILLIAM S. & SARA C.	BOWDEN, LESLIE M.	BOWDEN, EARLINE	BASS, WILLIAM S.	BASS, SARA C.
BREWSTER, ANTHONY JR & VICTORIA & FLACK, JONATHON & JENNIFER	BREWSTER, ANTHONY JR	BREWSTER, VICTORIA	FLACK, JONATHON	FLACK, JENNIFER
BROWN, RICKEY L. & LOUANN D. & ROBERT & REBECCA	BROWN, RICKEY L.	BROWN, LOUANN D.	BROWN, ROBERT	BROWN, REBECCA
COOK, LINDSEY & WILLIAM K. & TRACY L. & REINFORD, KYLE	COOK, LINDSEY	COOK, WILLIAM K.	COOK, TRACY L.	REINFORD, KYLE
DAVIS, RICHARD W. & AMANDA & JONATHAN & KATHERINE	DAVIS, RICHARD W.	DAVIS, AMANDA	DAVIS, JONATHAN	DAVIS, KATHERINE

first 10 rows shown above

Breakdown of this CTE can be seen in the notes on my Github.

Joining the temp tables together

```
SELECT * INTO #owners
FROM
(SELECT OwnerName, owner1, NULL AS owner2, NULL AS owner3, NULL AS owner4 FROM #one_owner
UNION
SELECT OwnerName, owner1, owner2, NULL AS owner3, NULL AS owner4 FROM #two_owners
UNION
SELECT OwnerName, owner1, owner2, owner3, NULL AS owner4 FROM #three_owners
UNION
SELECT * FROM #four_owners) AS #owners;
```

Adding owner columns to housing table

```
ALTER TABLE housingcopy
ADD owner1 NVARCHAR(255);

ALTER TABLE housingcopy
ADD owner2 NVARCHAR(255);

ALTER TABLE housingcopy
ADD owner3 NVARCHAR(255);

ALTER TABLE housingcopy
ADD owner4 NVARCHAR(255);
```

Checking the owner columns in the housing table

```sql
SELECT OwnerName, owner1, owner2, owner3, owner4 FROM housingcopy
```

OwnerName	owner1	owner2	owner3	owner4
FIELDS, KAREN L. & BRENT A.	NULL	NULL	NULL	NULL
BAILOR, DARRELL & TAMMY	NULL	NULL	NULL	NULL
LEE, JEFFREY & NANCY	NULL	NULL	NULL	NULL
NULL	NULL	NULL	NULL	NULL
URRUTIA, CARLOS MIGUEL & REBECCA	NULL	NULL	NULL	NULL
SALDANA, ALMA J. & EDWARD A., II	NULL	NULL	NULL	NULL
SANDAGE, LEAH S.	NULL	NULL	NULL	NULL
BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE	NULL	NULL	NULL	NULL
WILLIAMS, JOSEPH D. & AUDREY	NULL	NULL	NULL	NULL
ADAMS, ALEXANDER J. & JENNIFER	NULL	NULL	NULL	NULL

Updating the owner columns to housing table

```sql
-- UPDATE FOR ONE OWNER --
UPDATE housingcopy
SET owner1 = h1.owner1 FROM #owners AS h1
                                JOIN housingcopy AS o
                                ON h1.OwnerName = o.OwnerName;



-- UPDATE FOR TWO OWNERS --
UPDATE housingcopy
SET owner2 = h2.owner2 FROM #owners AS h2
                                JOIN housingcopy AS o
                                ON h2.OwnerName = o.OwnerName;



-- UPDATE FOR THREE OWNERS --
UPDATE housingcopy
SET owner3 = h3.owner3 FROM #owners AS h3
                                JOIN housingcopy AS o
                                ON h3.OwnerName = o.OwnerName;



-- UPDATE FOR FOUR OWNERS --

UPDATE housingcopy
SET owner4 = h4.owner4 FROM #owners AS h4
                                JOIN housingcopy AS o
                                ON h4.OwnerName = o.OwnerName;
```

Checking updates to the owner columns in the housing table

```sql
SELECT OwnerName, owner1, owner2, owner3, owner4 FROM housingcopy;
```

OwnerName	owner1	owner2	owner3	owner4
FIELDS, KAREN L. & BRENT A.	FIELDS, KAREN L.	FIELDS, BRENT A.	NULL	NULL
BAILOR, DARRELL & TAMMY	BAILOR, DARRELL	BAILOR, TAMMY	NULL	NULL
LEE, JEFFREY & NANCY	LEE, JEFFREY	LEE, NANCY	NULL	NULL
NULL	NULL	NULL	NULL	NULL
URRUTIA, CARLOS MIGUEL & REBECCA	URRUTIA, CARLOS MIGUEL	URRUTIA, REBECCA	NULL	NULL
SALDANA, ALMA J. & EDWARD A., II	SALDANA, ALMA J.	EDWARD A., II	NULL	NULL
SANDAGE, LEAH S.	SANDAGE, LEAH S.	NULL	NULL	NULL
BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE	BENJAMIN, DAVID AND LINDA REVOCABLE LIVING TRUST, THE	NULL	NULL	NULL
WILLIAMS, JOSEPH D. & AUDREY	WILLIAMS, JOSEPH D.	WILLIAMS, AUDREY	NULL	NULL
ADAMS, ALEXANDER J. & JENNIFER	ADAMS, ALEXANDER J.	ADAMS, JENNIFER	NULL	NULL

first 10 rows shown above

Final Clean-up

Deleting temp tables

```sql
DROP TABLE #one_owner
DROP TABLE #two_owners
DROP TABLE #three_owners
DROP TABLE #four_owners
DROP TABLE #owners
```

NASHVILLE HOUSING – ANALYSIS

Once cleaned, the below queries were used to pull information from the dataset (i.e., Avg. Land Value of each Property Type, Total Value of Each Property Type, etc.).

```sql
-------------------------------------------------------------------------------
-------------------------- NASHVILLE HOUSING - ANALYSIS --------------------------------------
-------------------------------------------------------------------------

-- ONCE CLEANED, THE BELOW QUERIES CAN BE USED PULL INFORMATION FROM THE DATASET --

-- FINDING THE TOTAL NUMBER OF EACH PROPERTY TYPE --

SELECT LandUse AS 'PropertyType', COUNT(LandUse) AS 'Number of Properties'
FROM housingcopy
GROUP BY LandUse
ORDER BY 'Number of Properties' DESC;

-- FINDING THE AVERAGE LAND VALUE OF EACH PROPERTY TYPE --

SELECT LandUse AS 'PropertyType', ROUND(AVG(LandValue), 0) AS 'Avg. Land Value'
FROM housingcopy
GROUP BY LandUse
ORDER BY 'Avg. Land Value' DESC;

-- FINDING THE TOTAL ACREAGE OF EACH PROPERTY TYPE --

SELECT LandUse AS 'PropertyType', ROUND(SUM(Acreage), 0) AS 'Total'
FROM housingcopy
WHERE Acreage IS NOT NULL
GROUP BY LandUse
ORDER BY 'Total' DESC;

-- FINDING TOTAL VALUE OF EACH PROPERTY TYPE --

SELECT LandUse AS 'PropertyType', SUM(TotalValue) AS 'TotalValue'
FROM housingcopy
WHERE TotalValue IS NOT NULL
GROUP BY LandUse
ORDER BY 'TotalValue' DESC;

-- FINDING THE TOTAL NUMBER OF PROPERTIES BUILT IN EACH CITY --

SELECT OwnerAddressCity AS 'City', COUNT(UniqueID) AS 'NumberofProperties'
FROM housingcopy
GROUP BY OwnerAddressCity
ORDER BY 'NumberofProperties' DESC;
```

* query used in Tableau dashboard

```sql
-- FINDING THE AVERAGE LAND VALUES OF EACH CITY --

SELECT OwnerAddressCity AS 'City', ROUND(AVG(LandValue), 0) AS 'Avg. LandValue'
FROM housingcopy
WHERE LandValue IS NOT NULL
GROUP BY OwnerAddressCity
ORDER BY 'Avg. LandValue' DESC;
```

* query used in Tableau dashboard

```sql
-- FINDING TOTAL ACREAGE BY CITY --

SELECT OwnerAddressCity AS 'City', ROUND(SUM(Acreage), 0) AS 'Total Acreage'
FROM housingcopy
WHERE Acreage IS NOT NULL
GROUP BY OwnerAddressCity
ORDER BY 'Total Acreage' DESC;

-- FINDING TOTAL PROPERTY VALUE BY CITY --

SELECT OwnerAddressCity AS 'City', SUM(TotalValue) AS 'Total Property Value'
FROM housingcopy
WHERE TotalValue IS NOT NULL
GROUP BY OwnerAddressCity
ORDER BY 'Total Property Value' DESC;
```

* query used in Tableau dashboard

```sql
-- FINDING THE NUMBER OF PROPERTIES BUILT BY YEAR --

SELECT YearBuilt AS 'Year', COUNT(UniqueID) AS 'Number of Properties Built'
FROM housingcopy
WHERE YearBuilt IS NOT NULL
GROUP BY YearBuilt
ORDER BY YearBuilt;
```

* query used in Tableau dashboard

```sql
-- FINDING THE NUMBER OF PROPERTIES TYPES BUILT EACH YEAR --

SELECT YearBuilt AS 'Year', LandUse AS 'PropertyType', COUNT(UniqueID) AS 'Number of Properties Built'
FROM housingcopy
WHERE YearBuilt IS NOT NULL
GROUP BY YearBuilt, LandUse
ORDER BY 'Year', 'PropertyType';
```