Nashville Housing Project
November, 2023
By Kidus Adaye

Step 1 – Ask

In this step we will identify the problem and objective of our case study and present our results.

Background –

Nashville, Tennessee has become a very popular destination for people looking for a relatively low cost of living, large job market and affordable housing. For this reason, the city has emerged as one of the fastest destinations for net domestic migration since 1990.

This project is intended to analyze and discover trends in the Nashville housing market. This analysis will help inform people with who are considering moving into the Nashville, Tennessee area with data-driven results.

Business objectives –

- What are the trends in sale price?
- What are the trends in housing properties and quality in respect to sales?
- How do different variables affect sale price?

Deliverables –

- Summary of the business objectives
- Description of data source
- Summary of the analysis
- Visualizations and key findings
- Recommendations

Stakeholders –

- Anyone who is interested in moving into the Nashville, Tennessee area and want to understand the housing market.

Step 2 – Prepare

In this step we will identify the data being used.

Information on data source –

This data was collected and publicly posted by TMTHYJAMES from Kaggle, found here, Kaggle .

This data contains 31 fields with over 56,000 number of records dating from 2013 to 2016.

Data includes columns on sale date, price, land use, city, etc.

Is data ROCCC?

A good data source is ROCCC which stands for Reliable, Original, Comprehensive, Current, and Cited.

- Reliable – MED – Dataset has over 56,00 records, but contains a lot of null values
- Original – LOW – Third party provider
- Comprehensive – MED – Just the right amount of information to gather insightful analysis.
- Current – LOW – Older than 5 years
- Cited – LOW – Third party provider

Even though the dataset is old and not as reliable, it can be cleaned and formatted correctly to analyze trends within that time frame.

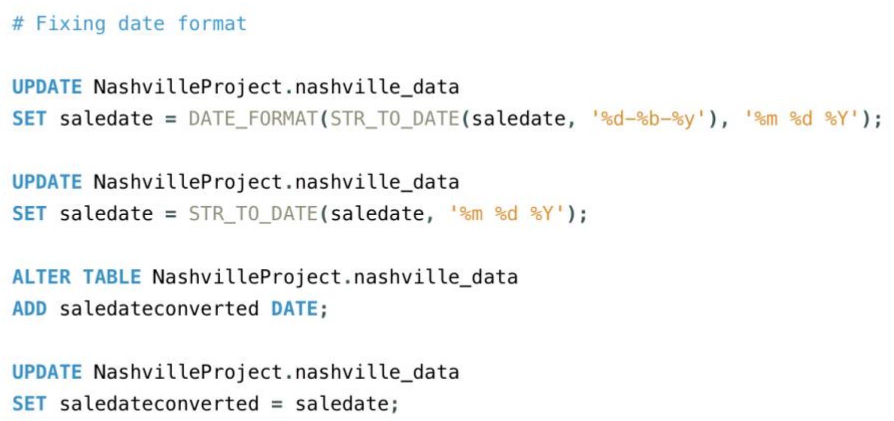
Step 3 - Process

In this step we will clean the data so that it is useable and contains no errors while we perform our analysis.

- Explore and observe the dataset
- Fix the sale date format
- Populate the missing property address data
- Separate the property and owner address into individual columns (address, city, and state)
- Renaming and grouping columns
- Fixing ‘Y’ and ‘N’ to ‘Yes’ and ‘No’ in sold as vacant field
- Removing duplicates
- Deleting unused columns

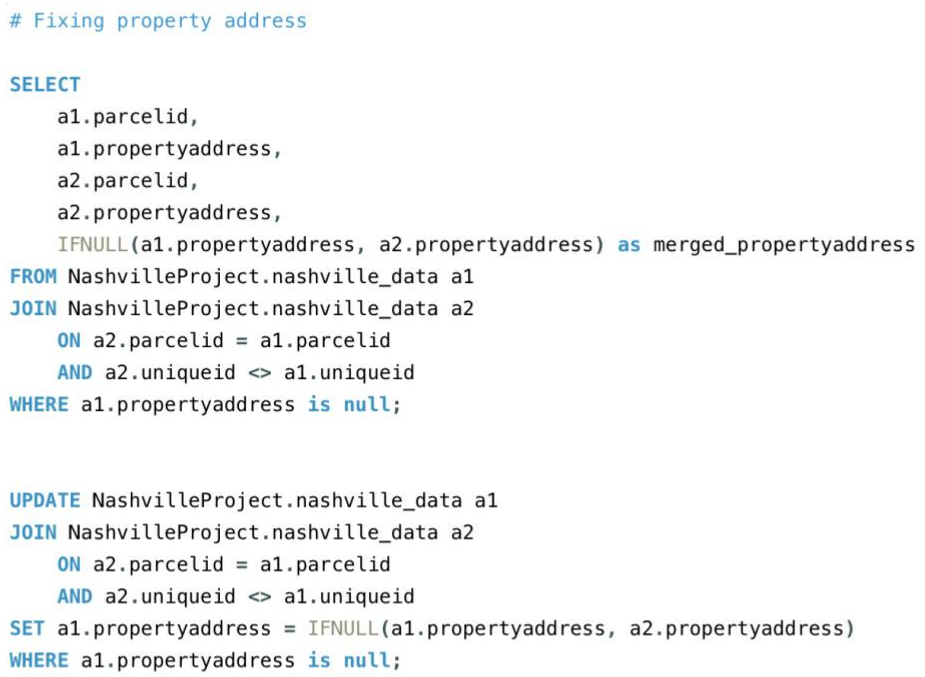
Data Cleaning and Manipulation

Previously, sale date had the times set to ’00:00:00’, so we set it to ‘YYYY-MM-DD’, so it can be usable in our analysis.



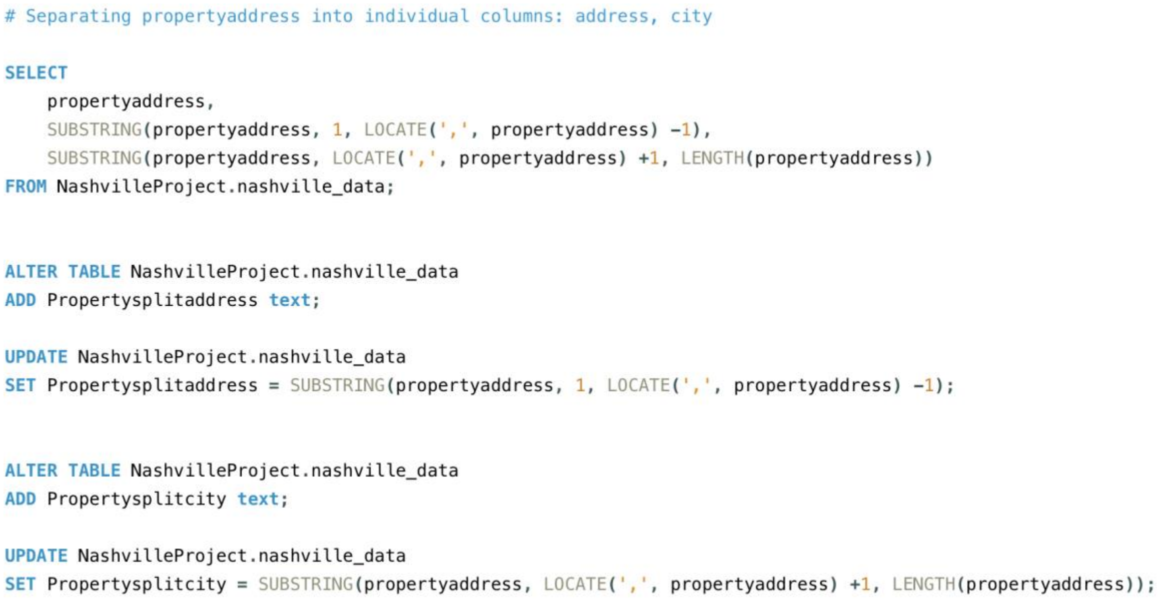
```
# Fixing date format

UPDATE NashvilleProject.nashville_data
SET saledate = DATE_FORMAT(STR_TO_DATE(saledate, '%d-%b-%y'), '%m %d %Y');

UPDATE NashvilleProject.nashville_data
SET saledate = STR_TO_DATE(saledate, '%m %d %Y');

ALTER TABLE NashvilleProject.nashville_data
ADD saledateconverted DATE;

UPDATE NashvilleProject.nashville_data
SET saledateconverted = saledate;
```

Here, we populated the null property address by doing a self-join that checks for similar ‘parcelid’, but different ‘uniqueid’.

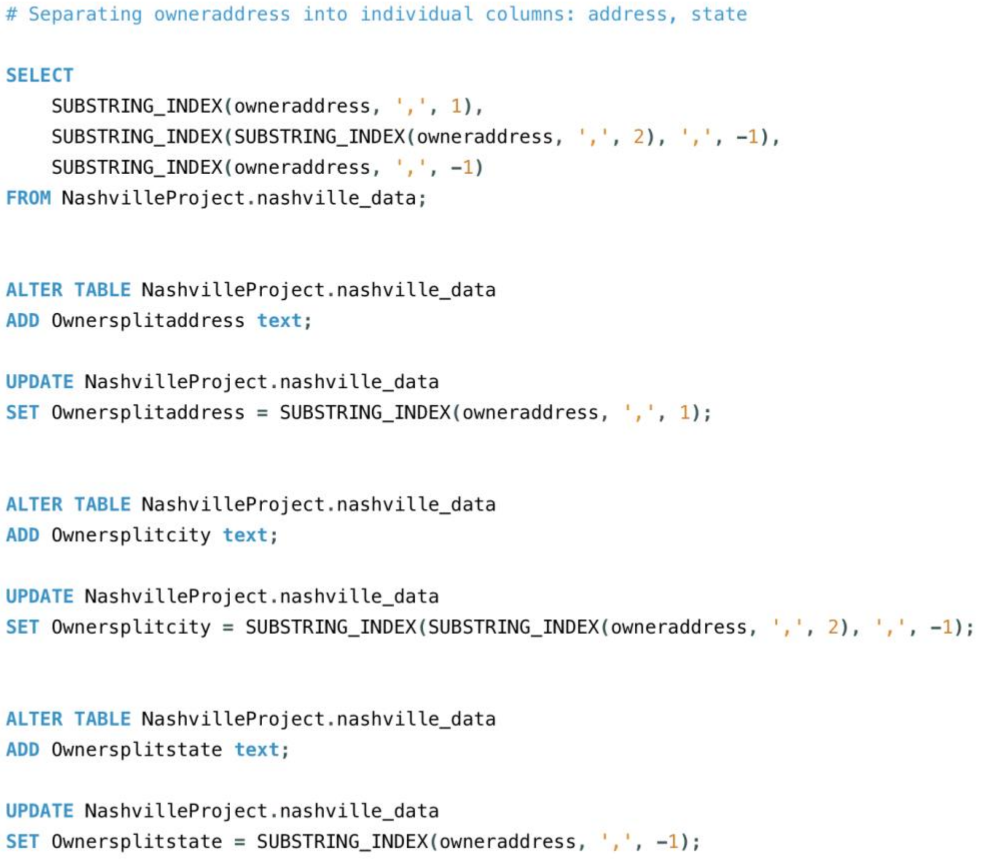


```
# Fixing property address

SELECT
    a1.parcelid,
    a1.propertyaddress,
    a2.parcelid,
    a2.propertyaddress,
    IFNULL(a1.propertyaddress, a2.propertyaddress) as merged_propertyaddress
FROM NashvilleProject.nashville_data a1
JOIN NashvilleProject.nashville_data a2
    ON a2.parcelid = a1.parcelid
    AND a2.uniqueid <> a1.uniqueid
WHERE a1.propertyaddress is null;


UPDATE NashvilleProject.nashville_data a1
JOIN NashvilleProject.nashville_data a2
    ON a2.parcelid = a1.parcelid
    AND a2.uniqueid <> a1.uniqueid
SET a1.propertyaddress = IFNULL(a1.propertyaddress, a2.propertyaddress)
WHERE a1.propertyaddress is null;
```

Now, we will break the ‘propertyaddress’ and ‘owneraddress’ into individual columns consisting of (address, city, and state)

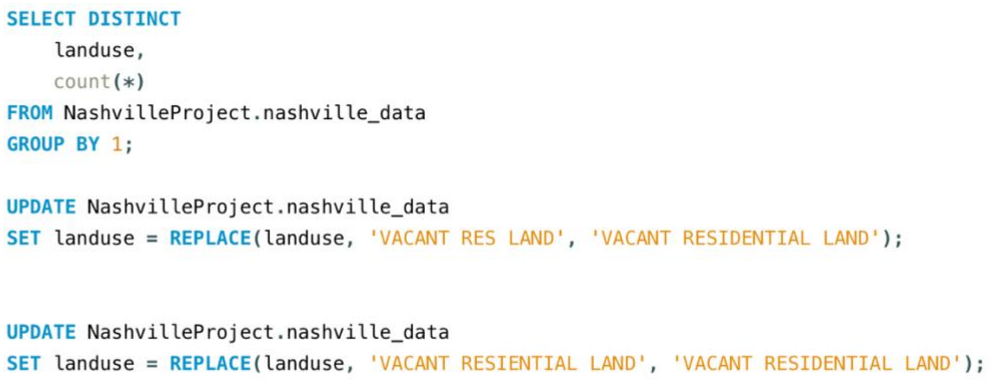


```
# Separating propertyaddress into individual columns: address, city

SELECT
    propertyaddress,
    SUBSTRING(propertyaddress, 1, LOCATE(',', propertyaddress) -1),
    SUBSTRING(propertyaddress, LOCATE(',', propertyaddress) +1, LENGTH(propertyaddress))
FROM NashvilleProject.nashville_data;


ALTER TABLE NashvilleProject.nashville_data
ADD Propertysplitaddress text;

UPDATE NashvilleProject.nashville_data
SET Propertysplitaddress = SUBSTRING(propertyaddress, 1, LOCATE(',', propertyaddress) -1);


ALTER TABLE NashvilleProject.nashville_data
ADD Propertysplitcity text;

UPDATE NashvilleProject.nashville_data
SET Propertysplitcity = SUBSTRING(propertyaddress, LOCATE(',', propertyaddress) +1, LENGTH(propertyaddress));
```



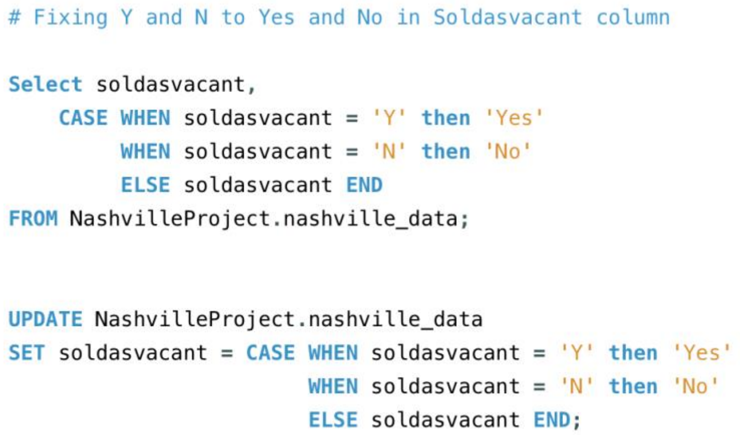
```
# Separating owneraddress into individual columns: address, state

SELECT
    SUBSTRING_INDEX(owneraddress, ',', 1),
    SUBSTRING_INDEX(SUBSTRING_INDEX(owneraddress, ',', 2), ',', -1),
    SUBSTRING_INDEX(owneraddress, ',', -1)
FROM NashvilleProject.nashville_data;


ALTER TABLE NashvilleProject.nashville_data
ADD Ownersplitaddress text;

UPDATE NashvilleProject.nashville_data
SET Ownersplitaddress = SUBSTRING_INDEX(owneraddress, ',', 1);


ALTER TABLE NashvilleProject.nashville_data
ADD Ownersplitcity text;

UPDATE NashvilleProject.nashville_data
SET Ownersplitcity = SUBSTRING_INDEX(SUBSTRING_INDEX(owneraddress, ',', 2), ',', -1);


ALTER TABLE NashvilleProject.nashville_data
ADD Ownersplitstate text;

UPDATE NashvilleProject.nashville_data
SET Ownersplitstate = SUBSTRING_INDEX(owneraddress, ',', -1);
```

Next, we will rename and group misspelled values together



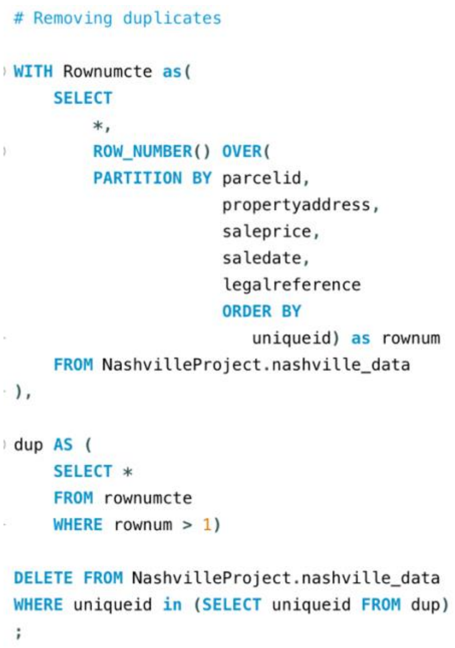
```
SELECT DISTINCT
    landuse,
    count(*)
FROM NashvilleProject.nashville_data
GROUP BY 1;

UPDATE NashvilleProject.nashville_data
SET landuse = REPLACE(landuse, 'VACANT RES LAND', 'VACANT RESIDENTIAL LAND');


UPDATE NashvilleProject.nashville_data
SET landuse = REPLACE(landuse, 'VACANT RESIENTIAL LAND', 'VACANT RESIDENTIAL LAND');
```

We then converted the ‘Y’ and ‘N’ values in sold as vacant field to ‘Yes’ and ‘No’



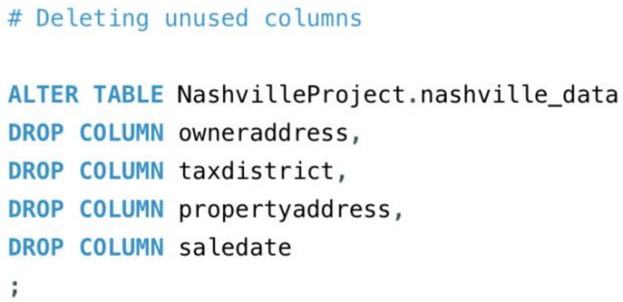
```
# Fixing Y and N to Yes and No in Soldasvacant column

Select soldasvacant,
    CASE WHEN soldasvacant = 'Y' then 'Yes'
         WHEN soldasvacant = 'N' then 'No'
         ELSE soldasvacant END
FROM NashvilleProject.nashville_data;


UPDATE NashvilleProject.nashville_data
SET soldasvacant = CASE WHEN soldasvacant = 'Y' then 'Yes'
                        WHEN soldasvacant = 'N' then 'No'
                        ELSE soldasvacant END;
```

Removing duplicates so we do not have inconsistencies



```
# Removing duplicates

WITH Rownumcte as(
    SELECT
        *,
        ROW_NUMBER() OVER(
        PARTITION BY parcelid,
                     propertyaddress,
                     saleprice,
                     saledate,
                     legalreference
                     ORDER BY
                        uniqueid) as rownum
    FROM NashvilleProject.nashville_data
),

dup AS (
    SELECT *
    FROM rownumcte
    WHERE rownum > 1)

DELETE FROM NashvilleProject.nashville_data
WHERE uniqueid in (SELECT uniqueid FROM dup)
;
```

Deleting unused columns (Demonstration purposes)



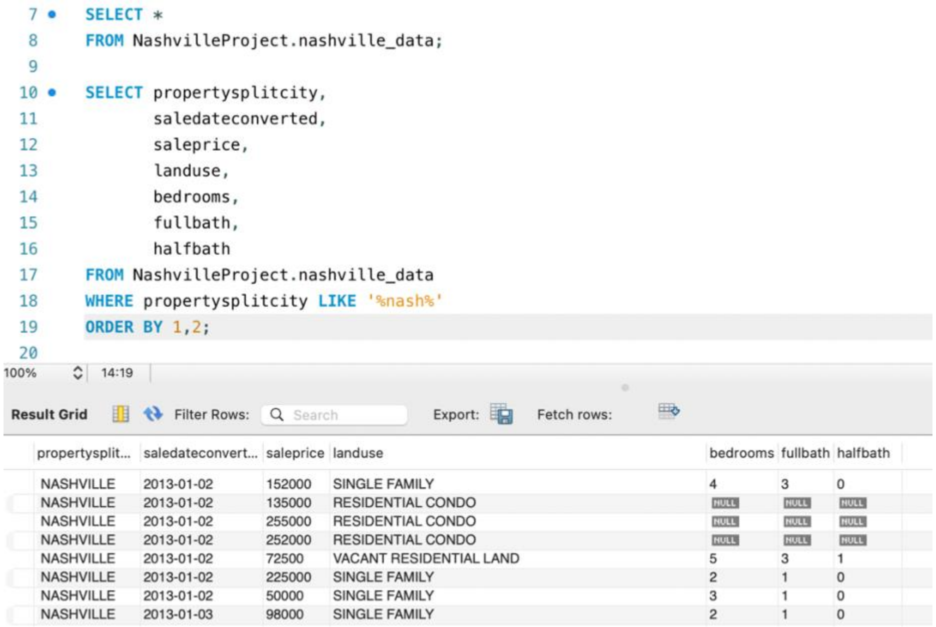
```
# Deleting unused columns

ALTER TABLE NashvilleProject.nashville_data
DROP COLUMN owneraddress,
DROP COLUMN taxdistrict,
DROP COLUMN propertyaddress,
DROP COLUMN saledate
;
```

Step 4 - Analyze

Key tasks for analysis –

1. Selecting and going over the data used for data analysis
2. Finding the sales distribution of Nashville housing prices
3. Calculating the daily average sale price over time
4. Calculating running sales over time segmented by city
5. Finding the average sale price of every property
6. Finding the average sale price and count of distinct combinations of bedroom, fullbath, and halfbath
7. Distribution of properties across Nashville metro area
8. Distribution of sales per sale date

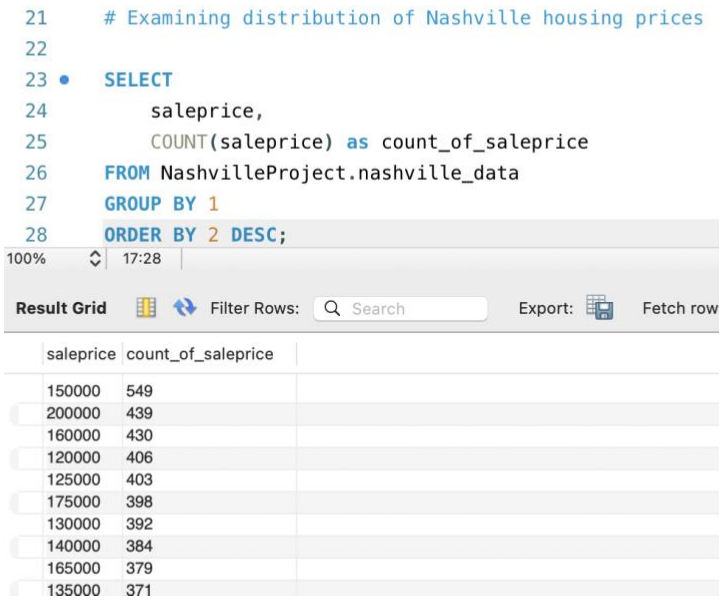
Query 1 Results –



```sql
SELECT *
FROM NashvilleProject.nashville_data;

SELECT propertysplitcity,
       saledateconverted,
       saleprice,
       landuse,
       bedrooms,
       fullbath,
       halfbath
FROM NashvilleProject.nashville_data
WHERE propertysplitcity LIKE '%nash%'
ORDER BY 1,2;
```

propertysplit...	saledateconvert...	saleprice	landuse	bedrooms	fullbath	halfbath
NASHVILLE	2013-01-02	152000	SINGLE FAMILY	4	3	0
NASHVILLE	2013-01-02	135000	RESIDENTIAL CONDO	NULL	NULL	NULL
NASHVILLE	2013-01-02	255000	RESIDENTIAL CONDO	NULL	NULL	NULL
NASHVILLE	2013-01-02	252000	RESIDENTIAL CONDO	NULL	NULL	NULL
NASHVILLE	2013-01-02	72500	VACANT RESIDENTIAL LAND	5	3	1
NASHVILLE	2013-01-02	225000	SINGLE FAMILY	2	1	0
NASHVILLE	2013-01-02	50000	SINGLE FAMILY	3	1	0
NASHVILLE	2013-01-03	98000	SINGLE FAMILY	2	1	0

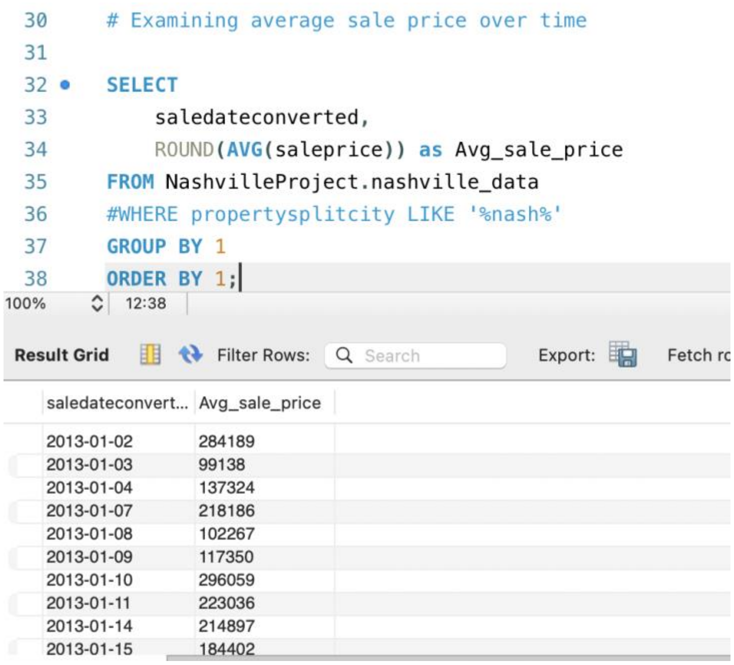
Query 2 Results –



```sql
# Examining distribution of Nashville housing prices

SELECT
    saleprice,
    COUNT(saleprice) as count_of_saleprice
FROM NashvilleProject.nashville_data
GROUP BY 1
ORDER BY 2 DESC;
```

saleprice	count_of_saleprice
150000	549
200000	439
160000	430
120000	406
125000	403
175000	398
130000	392
140000	384
165000	379
135000	371

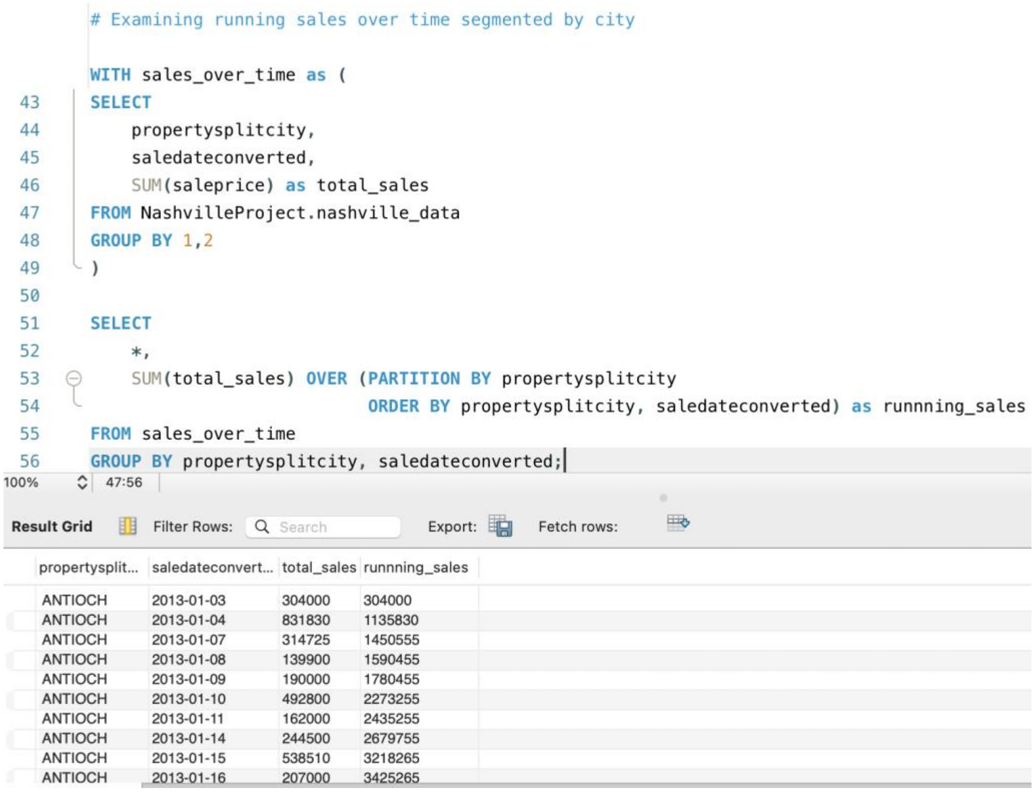
Query 3 Results –



```
# Examining average sale price over time

SELECT
    saledateconverted,
    ROUND(AVG(saleprice)) as Avg_sale_price
FROM NashvilleProject.nashville_data
#WHERE propertysplitcity LIKE '%nash%'
GROUP BY 1
ORDER BY 1;
```

saledateconvert...	Avg_sale_price
2013-01-02	284189
2013-01-03	99138
2013-01-04	137324
2013-01-07	218186
2013-01-08	102267
2013-01-09	117350
2013-01-10	296059
2013-01-11	223036
2013-01-14	214897
2013-01-15	184402

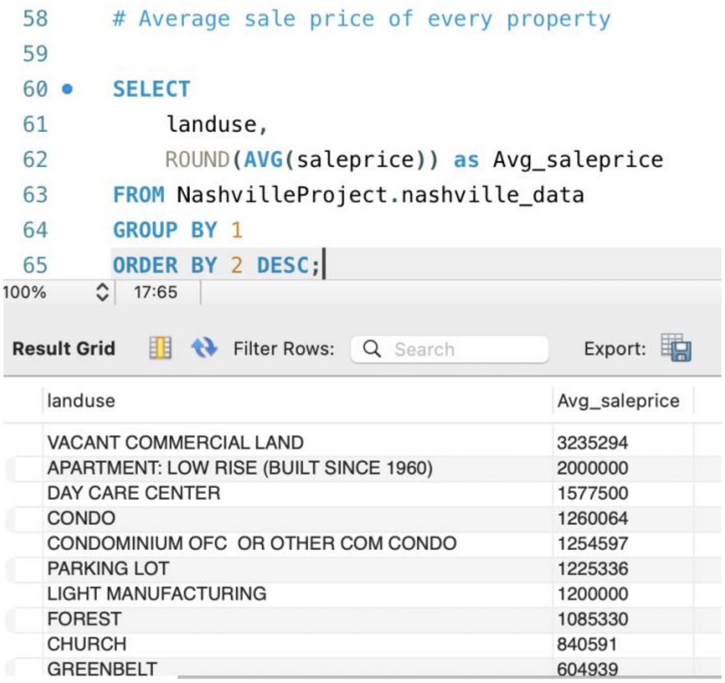
Query 4 Results –



```
# Examining running sales over time segmented by city

WITH sales_over_time as (
SELECT
    propertysplitcity,
    saledateconverted,
    SUM(saleprice) as total_sales
FROM NashvilleProject.nashville_data
GROUP BY 1,2
)

SELECT
    *,
    SUM(total_sales) OVER (PARTITION BY propertysplitcity
                           ORDER BY propertysplitcity, saledateconverted) as runnning_sales
FROM sales_over_time
GROUP BY propertysplitcity, saledateconverted;
```

propertysplit...	saledateconvert...	total_sales	runnning_sales
ANTIOCH	2013-01-03	304000	304000
ANTIOCH	2013-01-04	831830	1135830
ANTIOCH	2013-01-07	314725	1450555
ANTIOCH	2013-01-08	139900	1590455
ANTIOCH	2013-01-09	190000	1780455
ANTIOCH	2013-01-10	492800	2273255
ANTIOCH	2013-01-11	162000	2435255
ANTIOCH	2013-01-14	244500	2679755
ANTIOCH	2013-01-15	538510	3218265
ANTIOCH	2013-01-16	207000	3425265

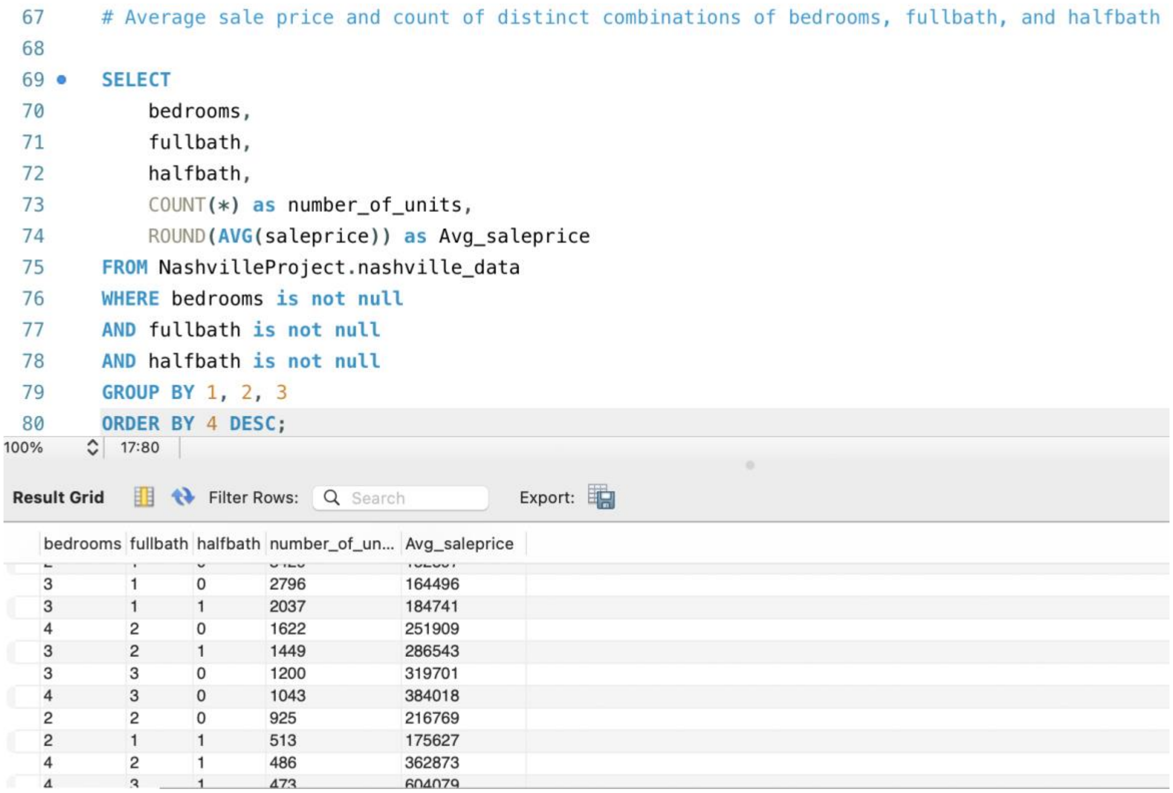
Query 5 Results –



```
# Average sale price of every property

SELECT
    landuse,
    ROUND(AVG(saleprice)) as Avg_saleprice
FROM NashvilleProject.nashville_data
GROUP BY 1
ORDER BY 2 DESC;
```

landuse	Avg_saleprice
VACANT COMMERCIAL LAND	3235294
APARTMENT: LOW RISE (BUILT SINCE 1960)	2000000
DAY CARE CENTER	1577500
CONDO	1260064
CONDOMINIUM OFC OR OTHER COM CONDO	1254597
PARKING LOT	1225336
LIGHT MANUFACTURING	1200000
FOREST	1085330
CHURCH	840591
GREENBELT	604939

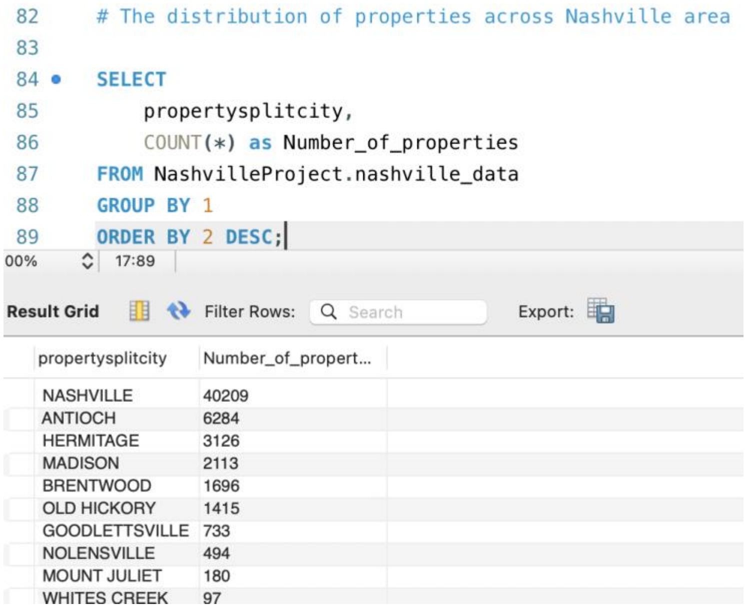
Query 6 Results –



```
# Average sale price and count of distinct combinations of bedrooms, fullbath, and halfbath

SELECT
    bedrooms,
    fullbath,
    halfbath,
    COUNT(*) as number_of_units,
    ROUND(AVG(saleprice)) as Avg_saleprice
FROM NashvilleProject.nashville_data
WHERE bedrooms is not null
AND fullbath is not null
AND halfbath is not null
GROUP BY 1, 2, 3
ORDER BY 4 DESC;
```

bedrooms	fullbath	halfbath	number_of_un...	Avg_saleprice
3	1	0	2796	164496
3	1	1	2037	184741
4	2	0	1622	251909
3	2	1	1449	286543
3	3	0	1200	319701
4	3	0	1043	384018
2	2	0	925	216769
2	1	1	513	175627
4	2	1	486	362873
4	3	1	473	604079

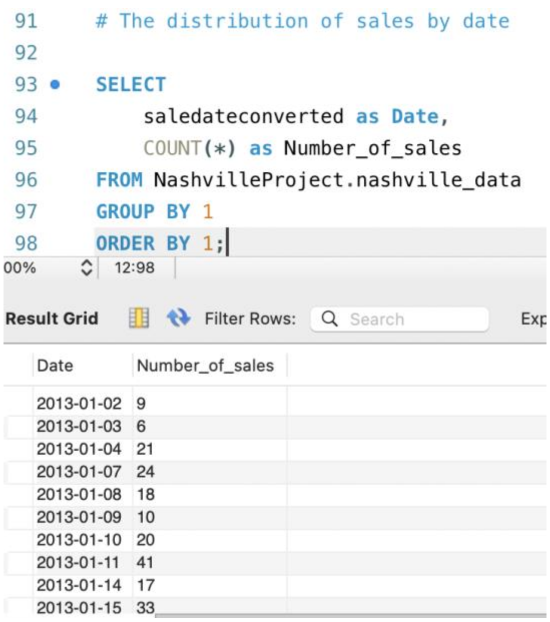
Query 7 Results –



```
# The distribution of properties across Nashville area

SELECT
    propertysplitcity,
    COUNT(*) as Number_of_properties
FROM NashvilleProject.nashville_data
GROUP BY 1
ORDER BY 2 DESC;
```

propertysplitcity	Number_of_propert...
NASHVILLE	40209
ANTIOCH	6284
HERMITAGE	3126
MADISON	2113
BRENTWOOD	1696
OLD HICKORY	1415
GOODLETTSVILLE	733
NOLENSVILLE	494
MOUNT JULIET	180
WHITES CREEK	97

Query 8 Results –



```
# The distribution of sales by date

SELECT
    saledateconverted as Date,
    COUNT(*) as Number_of_sales
FROM NashvilleProject.nashville_data
GROUP BY 1
ORDER BY 1;
```

```
Date        Number_of_sales
2013-01-02  9
2013-01-03  6
2013-01-04  21
2013-01-07  24
2013-01-08  18
2013-01-09  10
2013-01-10  20
2013-01-11  41
2013-01-14  17
2013-01-15  33
```

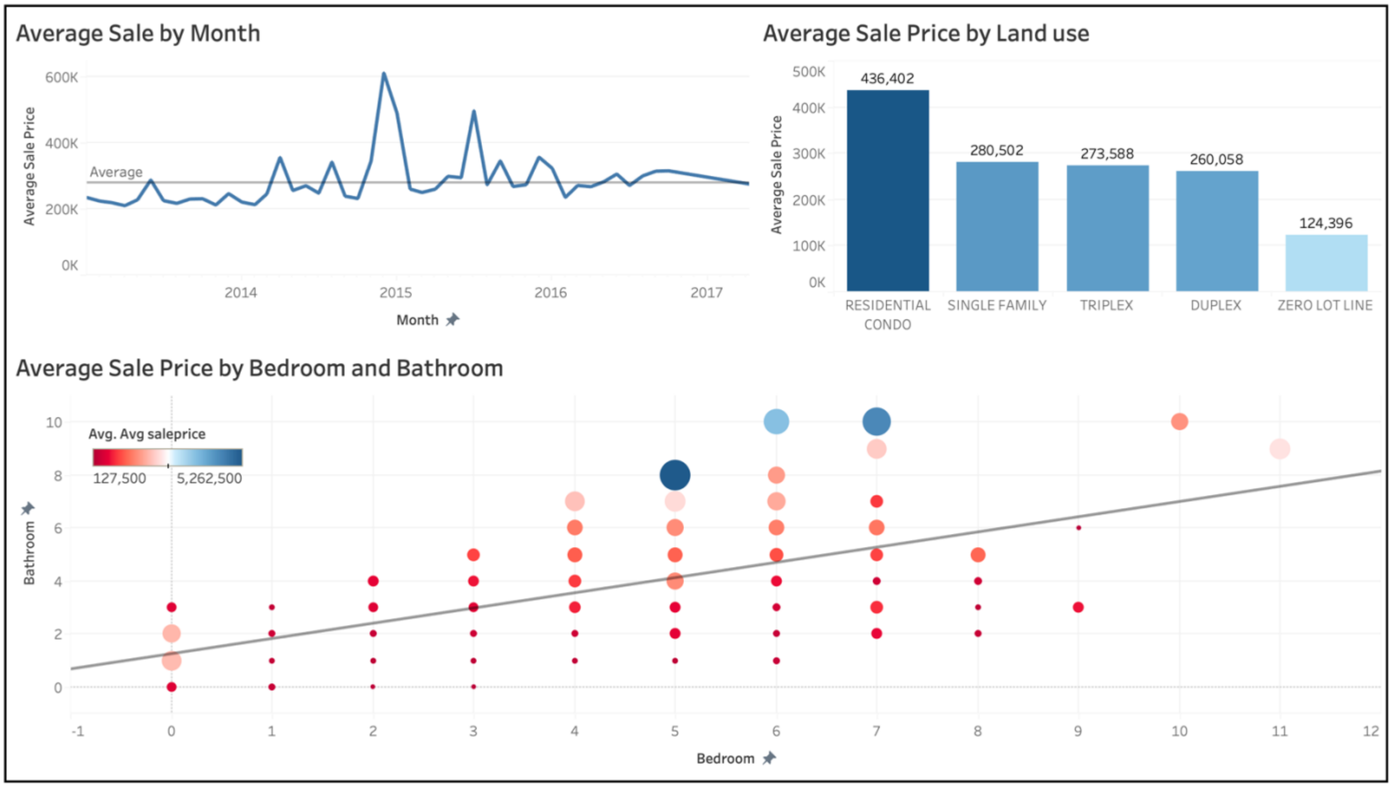
Interpreting findings –

1. The information required for our analysis is going to be columns: propertysplitcity, saledateconverted, saleprice, landuse, bedrooms, fullbath, and halfbath.
2. The distribution of Nashville housing prices is priced at around **100K – 250K** according to these findings.
3. In January 2013 – January 2016, the average sale price over time has grown by **11%** per year. That is faster than the national average during that time which was around **6%**.
4. From 2013 to 2016, Nashville has accounted for **$14.5B (79%)** in total sales compared to the other cities that are around **$4B** in total sales.
5. The average sale price of most housing properties were: Condos **$452K,** Vacant residential land **$333K,** single family **$280K,** and Duplex **$259K.**
6. The average sale for the top 3 distinct bed/bath housing were: **3bed2bath** $215K, **2bed1bath** $152K, and **3bed1bath** $164K.
7. Out of the entire 56K properties, the Nashville area holds the most amount of properties accounting for over **70% (40K)**of the market.
8. In 2013 to 2016 the average count of sales per month was **50**. The average tends to be higher during May – July and lower in the first couple months of the year.

Step 5 – Share

In this step, we will create visualizations and present our findings based on our analysis using tableau.

Tableau Dashboard(s) –

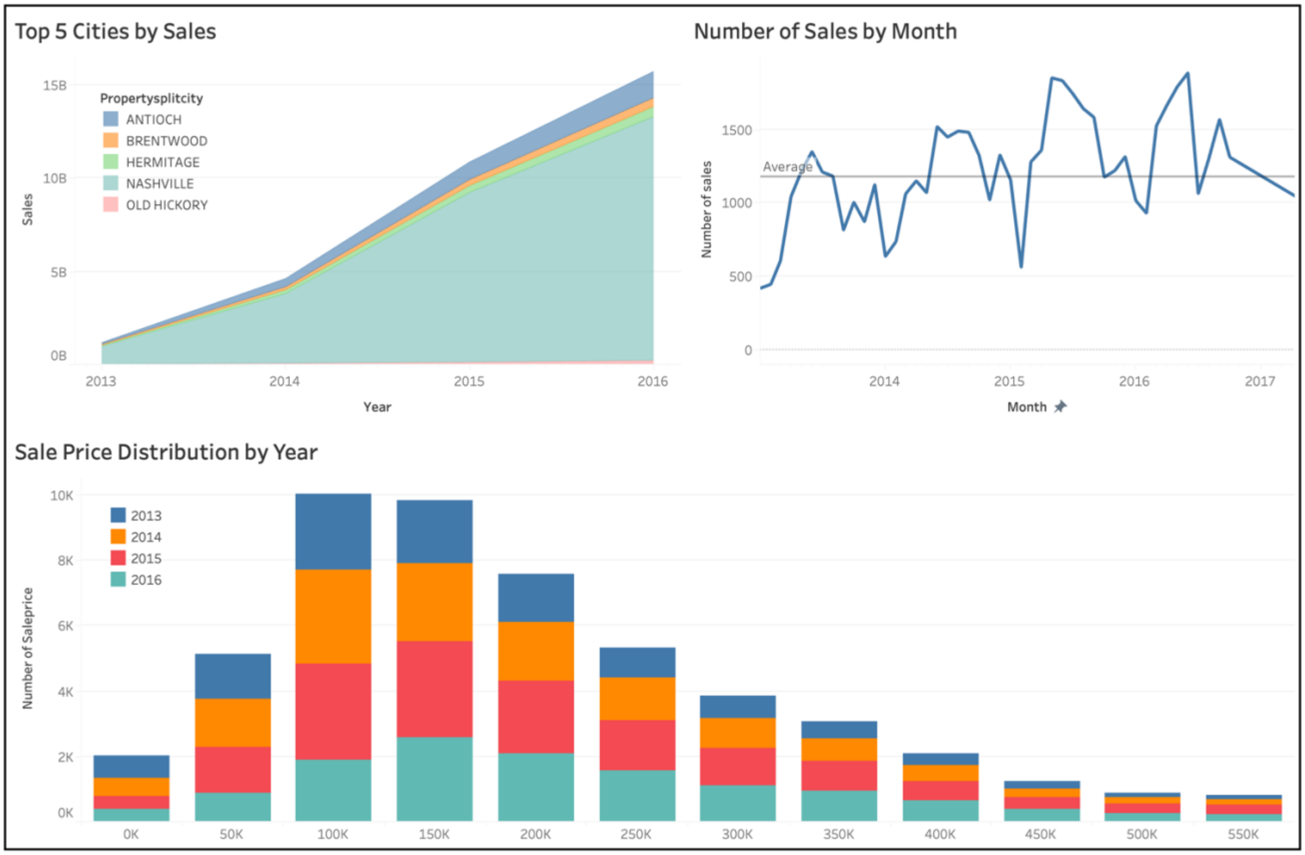


Average Sale Price Trends –

This dashboard displays the average sales price of different variables. The average sales broken down by each month, average sale price of property land use, and average sale price by bedroom and bathroom.

1. Average sale per month progressed in an upward direction during the 2013-2016 years with an average of **278K.** The highest averages came during the middle of the year with the highest being **613K** and the lowest average being **210K** near the beginning of the year.

2. Condos are going to be priced the highest going for about **436K** on average. Single family, triplex, and duplex properties are going to be around the same price range and zero lot line properties will be the lowest going for about **124K** on average. The supply of these properties will have a direct correlation to their price as there are more single-family homes than condos.
3. The average sale price of properties is **positively correlated** to the amount of bathrooms supplied, while there seems to not be as strong of a correlation to bedrooms supplied. This may vary due to all the various property types and quantity available.



Sale Price volume/distribution Trends -

This dashboard displays the sale price distribution by year, month to month basis, and running total sales of the top 5 cities in the Nashville metro area.

1. Nashville dominates all other areas in terms of sales generated. In the spam of 2013-2016, they have reached a total of over **14B** in sales. Antioch comes in second at around **1.4B** in sales. It is important to keep in mind Nashville has a much larger housing supply compared to its neighboring areas.
2. The sales volume of properties went way up with June-August being peak months as they reached well over **1,500** during that time. The sales volume also dropped to as low as **418** as that might be the effects of a new year/quarter.
3. The overall sale price distribution was well centered around the **$50K-$250K** range during the span of 2013 to 2016. The volume of sales also increases very closely with respect to time.

Step 6 – Act

In this final step we will be delivering our insights and provide recommendations based on our analysis.

Here, we will revisit our business questions and share our data driven business recommendations.

Recommendations -

- Nashville's home prices are increasing and rising over time. This is due to the housing market demand and the new infrastructure being built. The majority of prices are going to be around $50K to $250K. This allows flexibility and availability to buyers that are looking for a specific price range.
- Keep in mind that as more and more homes are being sold over time, the Nashville housing market will continue growing as the total sales revenue has doubled each year. This means that higher priced homes are being sold and that number will only continue to rise. If you are planning to relocate in the Nashville area, it is better to start now rather than later.
- The variables that directly influence sale price are going to be land use and how many bathrooms and bedrooms there are. For people looking to relocate, condos are around $436K, $280K for single family, $273K for triplex, $260K for duplex, and $124K for zero lot line. It is also important to

note that sale prices tend to increase as the number of bathrooms increase, but that does not hold true for bedrooms.