

**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](https://www.kaggle.com/datasets/tmthyjames/nashville-housing-data) .

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.

A computer code with text

Description automatically generated

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

A screen shot of a computer

Description automatically generated

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

A screen shot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

Next, we will rename and group misspelled values together

A screenshot of a computer code

Description automatically generated

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

A screenshot of a computer

Description automatically generated

Removing duplicates so we do not have inconsistencies

A screenshot of a computer code

Description automatically generated

Deleting unused columns (Demonstration purposes)

A screenshot of a computer

Description automatically generated

**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 –**

**A screenshot of a computer

Description automatically generated**

**Query 2 Results –**

**A screenshot of a computer

Description automatically generated**

**Query 3 Results –**

**A screenshot of a computer

Description automatically generated**

**Query 4 Results –**

**A screenshot of a computer

Description automatically generated**

**Query 5 Results –**

**A screenshot of a computer

Description automatically generated**

**Query 6 Results –**

**A screenshot of a computer

Description automatically generated**

**Query 7 Results –**

**A screenshot of a computer

Description automatically generated**

**Query 8 Results –**

**A screenshot of a computer

Description automatically generated**

**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) –**

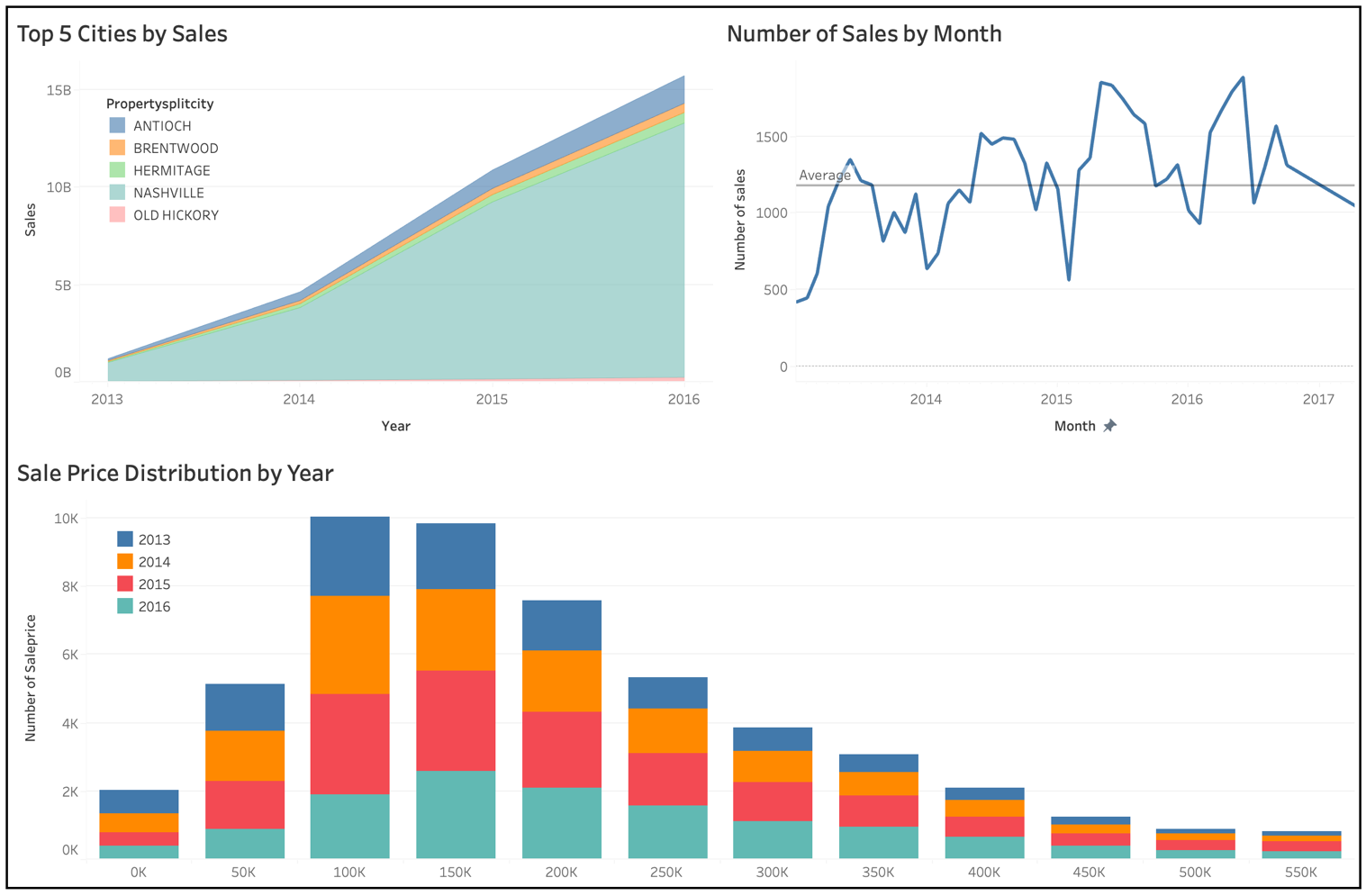
**A graph and diagram of a house

Description automatically generated**

**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.