Nashville Housing Market Data Analysis Project

Executive Summary

The aim of this analysis is to focus on Nashville housing boom from 2013 to 2016. More specifically, the analysis should help answer the following questions:

- How did Nashville housing market evolve from 2013 to 2016?
- Where are most demanded properties located?
- What type of homes are mostly demanded?
- What is the average sale price of a property in Nashville?
- How much did property agencies generate in 2013, 2014, 2015 and 2016

Methodology

Before performing the analysis, the data was collected through a public domain (http://www.kaggle.com/datasets/tmthyjames/nashville-housing-data), then wrangled to make sure it's cleaned, reliable and error-free by finding and filling missing values, removing duplicates and normalizing data. After that, I explored and gained insights through variables, and finally proceeded to data visualization.

DATA CLEANING PROCESS

```
-- Standardize Date Format

SELECT*
FROM [Portfolio Project]..NashvilleHousing

ALTER TABLE NashvilleHousing
ADD SaleDateConverted DATE

UPDATE NashvilleHousing
SET SaleDateConverted = CONVERT(DATE,SaleDate)


-- Add column SaleYear

ALTER TABLE [Portfolio Project]..NashvilleHousing
ADD SaleYear FLOAT

UPDATE [Portfolio Project]..NashvilleHousing
SET SaleYear = PARSENAME(REPLACE(SaleDateConverted, '-', '.') , 3)


-- Add column SaleMonth

ALTER TABLE [Portfolio Project]..NashvilleHousing
ADD SaleMonth FLOAT

UPDATE [Portfolio Project]..NashvilleHousing
SET SaleMonth = PARSENAME(REPLACE(SaleDateConverted, '-', '.') , 2)
```

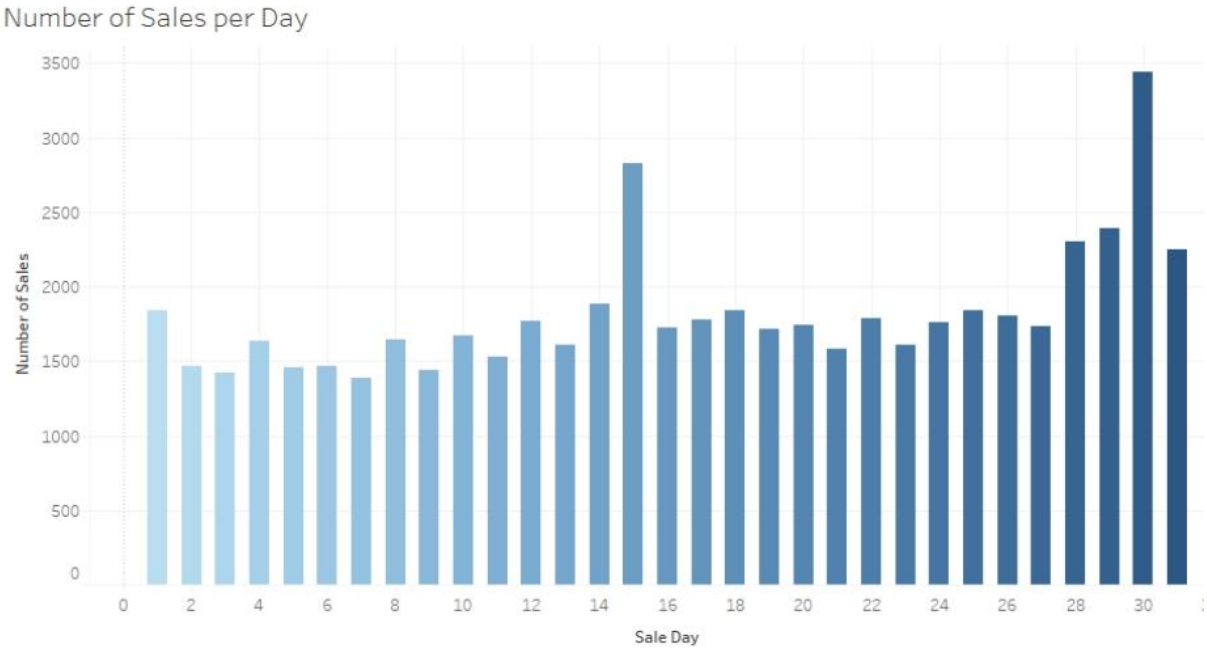
```sql
-- Add column SaleMonthName

ALTER TABLE [Portfolio Project]..NashvilleHousing
ADD SaleMonthName nvarchar(50)

UPDATE [Portfolio Project]..NashvilleHousing
SET SaleMonthName = DATENAME(MONTH, DATEADD(MONTH,SaleMonth,'2022-12-01'))


-- Add column SaleDay

ALTER TABLE [Portfolio Project]..NashvilleHousing
ADD SaleDay FLOAT

UPDATE [Portfolio Project]..NashvilleHousing
SET SaleDay = PARSENAME(REPLACE(SaleDateConverted, '-', '.') , 1)


-- Populate Property Address column

SELECT a.ParcelId, a.PropertyAddress, b.ParcelId, b.PropertyAddress,
ISNULL(a.PropertyAddress, b.PropertyAddress)
FROM [Portfolio Project]..NashvilleHousing a
JOIN [Portfolio Project]..NashvilleHousing b
      ON a.ParcelId = b.ParcelId
      AND a.[UniqueID] <> b.[UniqueID]
WHERE a.PropertyAddress is null

UPDATE a
SET PropertyAddress = ISNULL(a.PropertyAddress, b.PropertyAddress)
FROM [Portfolio Project]..NashvilleHousing a
JOIN [Portfolio Project]..NashvilleHousing b
      ON a.ParcelId = b.ParcelId
      AND a.[UniqueID] <> b.[UniqueID]
WHERE a.PropertyAddress is null


-- Break out Address into Individual Columns (Address, City, State)

SELECT PropertyAddress
FROM [Portfolio Project]..NashvilleHousing


ALTER TABLE NashvilleHousing
ADD PropertySplitAddress Nvarchar(255)

UPDATE NashvilleHousing
SET PropertySplitAddress = SUBSTRING(PropertyAddress,1, CHARINDEX(',',
PropertyAddress)-1)


ALTER TABLE NashvilleHousing
ADD PropertySplitCity Nvarchar(255)

UPDATE NashvilleHousing
SET PropertySplitCity = SUBSTRING(PropertyAddress, CHARINDEX(',', PropertyAddress) +
1, LEN(PropertyAddress))


ALTER TABLE NashvilleHousing
ADD OwnerSplitAddress Nvarchar(255)
```

```sql
UPDATE NashvilleHousing
SET OwnerSplitAddress = PARSENAME(REPLACE(OwnerAddress, ',', '.') , 3)


ALTER TABLE NashvilleHousing
ADD OwnerSplitCity Nvarchar(255)

UPDATE NashvilleHousing
SET OwnerSplitCity = PARSENAME(REPLACE(OwnerAddress, ',', '.') , 2)

ALTER TABLE NashvilleHousing
ADD OwnerSplitState Nvarchar(255)

UPDATE NashvilleHousing
SET OwnerSplitState = PARSENAME(REPLACE(OwnerAddress, ',', '.') , 1)

-- Rename inconsistent names in LandUse column

UPDATE [Portfolio Project]..NashvilleHousing
SET LandUse = REPLACE(LandUse, 'VACANT RESIENTIAL LAND', 'VACANT RESIDENTIAL LAND')

UPDATE [Portfolio Project]..NashvilleHousing
SET LandUse = REPLACE(LandUse, 'VACANT RES LAND', 'VACANT RESIDENTIAL LAND')


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

SELECT DISTINCT (SoldAsVacant), COUNT(SoldAsVacant)
FROM [Portfolio Project]..NashvilleHousing
GROUP BY SoldAsVacant


UPDATE NashvilleHousing
SET SoldAsVacant = CASE WHEN SoldAsVacant = 'Y' THEN 'Yes'
                WHEN SoldAsVacant = 'N' THEN 'No'
                ELSE SoldAsVacant
                END


-- Remove Duplicates

WITH RowNumCTE AS(
SELECT *,
       ROW_NUMBER() OVER (
       PARTITION BY ParcelID,
                            PropertyAddress,
                            SalePrice,
                            SaleDate,
                            LegalReference
                            ORDER BY
                                  UniqueID
                                  ) row_num

FROM [Portfolio Project]..NashvilleHousing
)
DELETE
FROM RowNumCTE
WHERE row_num > 1


-- Delete Unused Columns
```

```
Alter Table [Portfolio Project]..NashvilleHousing
DROP COLUMN SaleMonth, OwnerAddress, TaxDistrict, PropertyAddress, SaleDate


-- Checking new columns

SELECT *
FROM [Portfolio Project]..NashvilleHousing
```

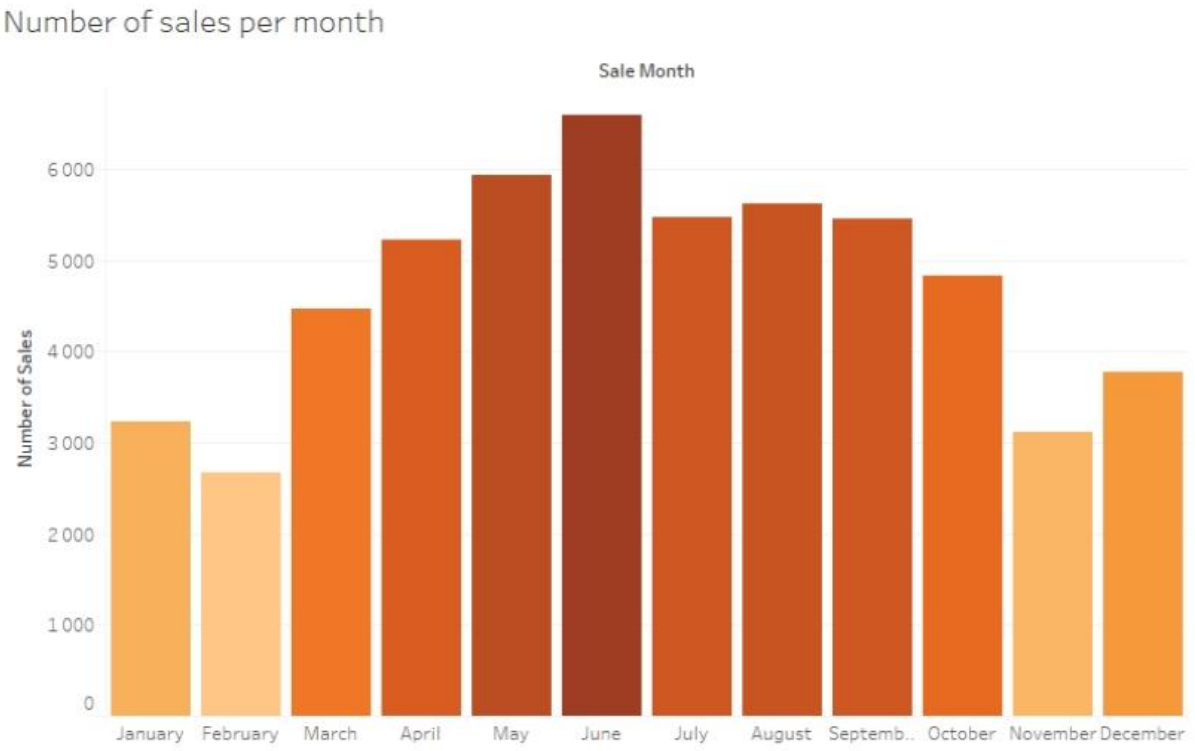
DATA ANALYSIS AND VISUALIZATION

```
-- Showing number of sales per day

SELECT SaleDay, COUNT(*) AS Daily_Numb_Sales
FROM [Portfolio Project]..NashvilleHousing
GROUP BY SaleDay
ORDER BY Daily_Numb_Sales DESC
```



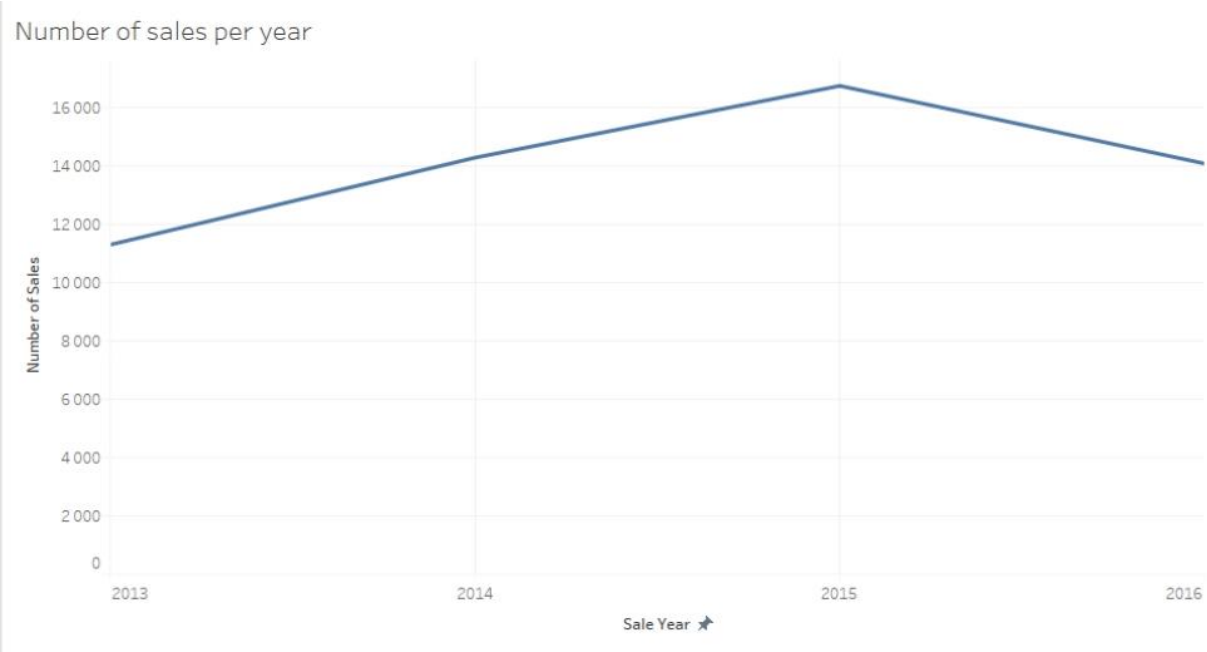
It appears that more sales (3478) are generated the 30th of month, followed by the 15th.

```
-- Showing number of sales per month

SELECT SaleMonthName, COUNT(*) AS Monthly_Numb_Sales
FROM [Portfolio Project]..NashvilleHousing
GROUP BY SaleMonthName
ORDER BY  Monthly_Numb_Sales DESC
```



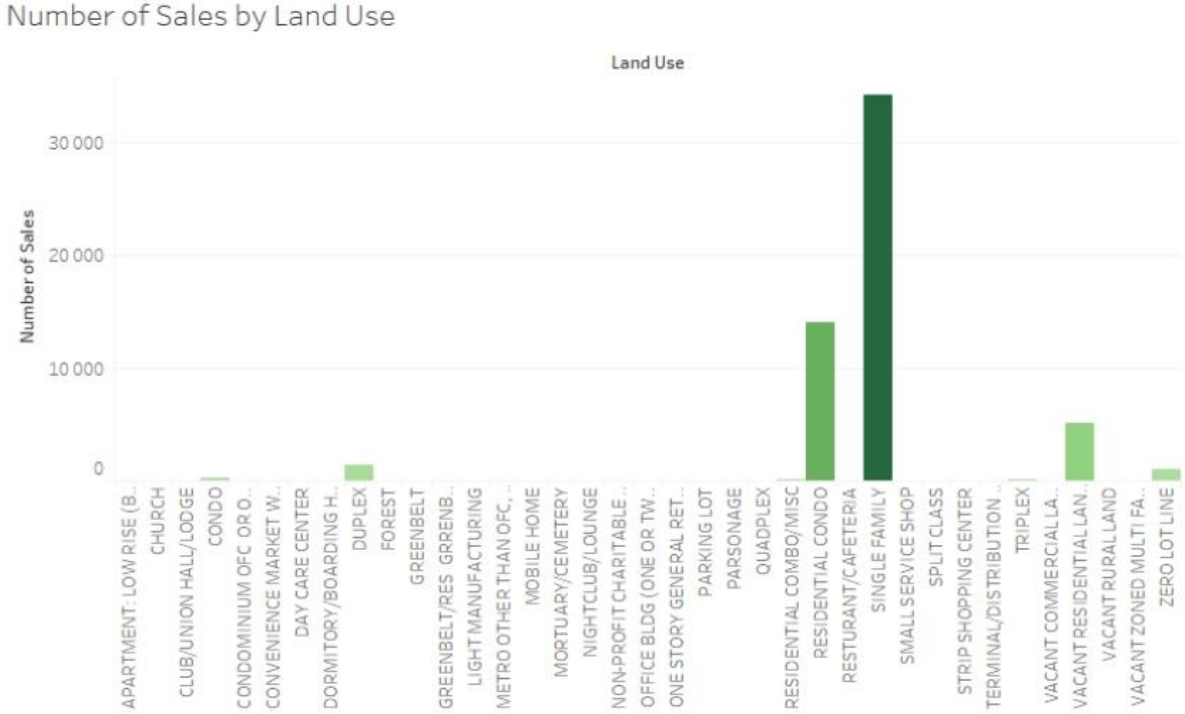
We see that June has the highest number of sales (6593) followed by May (5931) and August (5620). This shows that highest sales are generated in the middle of year.

```
-- Showing number of sales per year

SELECT SaleYear, COUNT(*) AS Yearly_Numb_Sales
FROM [Portfolio Project]..NashvilleHousing
GROUP BY SaleYear
ORDER BY SaleYear DESC
```



The Nashville housing boom started in 2013 with 11292 as total sales. This figure increased and reached 14274 in 2014, and 16734 in 2015 and finally decreased slightly in 2016 (14071).

```
-- Showing number of sales by land use

SELECT LandUse, COUNT(*) AS LandUse_Num_Sales
FROM [Portfolio Project]..NashvilleHousing
GROUP BY LandUse
ORDER BY LandUse_Num_Sales DESC
```



We notice that the top 3 land use are: "Single Family" having the highest number of sales (34119), followed by "Residential Condo" (14064) and Vacant Residential Land (5092).

```
-- Showing number of sales by property location

SELECT PropertySplitCity, COUNT(*) AS Numb_Sales_by_location
FROM [Portfolio Project]..NashvilleHousing
GROUP BY PropertySplitCity
ORDER BY Numb_Sales_by_location DESC
```

It appears that Nashville has the highest demand for homes with a total sale of 40246, followed by other locations Antioch (6284) and Hermitage (3216).

```
-- Showing homes with highest sale prices

SELECT SaleDateConverted, LandUse, PropertySplitCity, MAX(SalePrice) AS
Highest_Sale_Price
, (SELECT ROUND(AVG(SalePrice),0) FROM [Portfolio Project]..NashvilleHousing) AS
AllAvgSalePrice
FROM [Portfolio Project]..NashvilleHousing
GROUP BY SaleDateConverted, LandUse, PropertySplitCity, SalePrice
ORDER BY SalePrice DESC
```

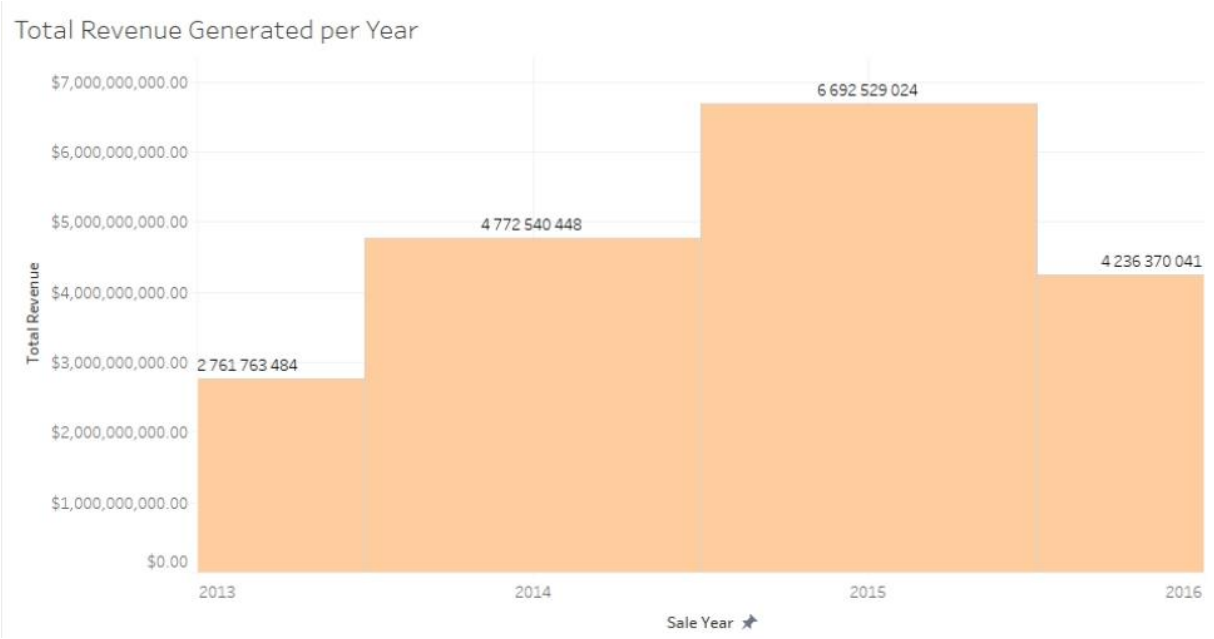
Here we notice that the top three expensive homes (respectively $54,278,060, $14,100,000 and $13,156,000) have been sold for Residential Condo purposes and are located in Nashville.

```
-- Showing home prices higher than all average sale price in relation to their
locations

SELECT LandUse, PropertySplitCity, SalePrice
, COUNT(LandUse) OVER (PARTITION BY LandUse) AS LandUse_Num_Sales
, AVG(SalePrice) OVER (PARTITION BY LandUse) AS AvgSalePrice
FROM [Portfolio Project]..NashvilleHousing
WHERE SalePrice > 327530
GROUP BY LandUse, PropertySplitCity, SalePrice
```

It appears that most expensive homes higher than all average sale price ($327,530) that have been sold are mainly located in the following cities: Whites Creek, Old Hickory, Nolensville, Nashville, Mount Juliet, Madison, Hermitage, Goodlettsville, Brentwood , and Antioch.

```
-- Showing Total revenue generated per year

WITH CTE_Revenue (SaleYear, Yearly_Numb_Sales, Yearly_AvPrice)
AS
(SELECT SaleYear, COUNT(*) AS Yearly_Numb_Sales
, ROUND(AVG(SalePrice),0) AS Yearly_AvPrice
```

```
FROM [Portfolio Project]..NashvilleHousing
GROUP BY SaleYear
)
SELECT SaleYear
, Yearly_Numb_Sales
,FORMAT (Yearly_AvPrice,'C','en-us') AS Yearly_AvPrice
,FORMAT (Yearly_Numb_Sales * Yearly_AvPrice,'C','en-us') AS Yearly_Tot_Revenue
FROM CTE_Revenue
```



It appears that 2015 is the winning year in Nashville housing market with a total revenue of $6,692,529,024 generated. Next, comes 2014 at 2nd position ($4,772,540,448), followed by 2016 ($4,236,370,041) and 2013 ($2,761,763,484).

Analysis Summary

- The year 2015 was particularly important in Nashville housing market with the highest demand for homes (16734) and highest revenue generated ($6,692,529,024)
- The middle of year, especially June has the highest number of sales (6593)
- Sales (3478) are mostly generated the 30th of each month
- Homes located in Nashville are highly demanded (40246)
- People buy homes firstly for family purposes (34119), followed by Residential Condo purposes (14064).