Nashville Housing Data

Stanislav Fatkhutdinov
University of Maryland Global Campus

Assignment 2
DATA 610 Section 9040
Decision Management Systems
Vahe Heboyan

Nashville Housing Data

The Nashville housing data contains the information from Nashville housing market from 2013 to 2016. The data is consisting of Parcel Id, Land Use, Property Address, Suit/Condo Number, Property City, Sale Date, Sale Price, Legal Reference, Sold As Vacant and Multiple Parcels Involved In Sale Information, Owner Name, Address, City, State, Acreage, Tax District, Neighborhood, Land Value, Building Value, Total Value, Finished Area, Foundation Type, Year Built, Exterior Wall, Grade, Bedrooms, Full and Half Baths (Dobbins, 2017). Some of the data points are missing and are replaced by the Cognos Analytics software to null values once converted into a data module. Provided data can be used to gain insights into relationship between the aforementioned features, however, some of the data appears to be duplicated. For example, Property Address and Address columns are superfluous and some of the features like Total Value are derived from Land and Building Values. Some of the features are so niche that they can be omitted from the evaluation right away, for example Neighborhood and Grade features are irrelevant due to their ambiguity of purpose. I've decided to investigate which data relationships are affecting the sale and total prices of the properties in Nashville, TN.

The data preparation begins with handling out irrelevant data. Looking at the data, I've decided to remove suite/condo number column since it is full of null values and only applicable to a certain Land Use types. I also removed Legal Reference column, since it doesn't have any relationships besides identification of the property in question, which is already covered by the Parcel Id feature. Owner Name column is irrelevant to the analysis of data, so I removed it. In addition, I've removed Address and City columns because their information is a duplicate of Property Address and City columns. The last two columns I removed were the Neighborhood and Grade, I did not saw any value in looking for their relationships to the data due to their specificity and applicability in determining the sale price trends.

After removing all the irrelevant data I've concentrated on tackling the Null values. The easiest place to start was changing the State Null values to TN, since it is assumed that all the entries are from the same state of Tennessee. In order to do that I've filtered out the Null values and replaced them with 'TN'. After changing all the null values to TN I've decided that the State value has no analytical bearing and hid it from users since it is assumed that the data is from the same state.

The next problem was the Null values in the Total Value field, I chose it because at the first glance if the data was null in that column, majority of the other columns associated with the same record were also null. In order to filter it out I used an expression editor to create a filter for the null values in the Total Value column. This step cleaned up a majority of the records with a single filter.

I kept investigating null values in the other features and replaced Foundation Type null values with 'EMPTY PLOT' since I realized that this was the data for vacant plots of land. I also changed the exterior wall for those kinds of vacant properties to 'NOT BUILT'. It was also natural to change the null values in the Finished Area columns to 0 since there were no building

on that land. I applied the same approach to the number of full and half baths as well as the number of bedrooms, for all the 'EMPTY PLOT' I changed those null values of to 0.

The last null values I had left to tackle was the Year Built value. I realized that it doesn't make sense for the vacant undeveloped property to have a built value, but I didn't want to lose that data from my calculations. In order to mitigate this problem, I've decided to change all the null values in the Year Built for the empty plots of land to '2018'. All these steps left me with some clean data to explore.

Once the cleaning was done, I've created the data exploration based on the Nashville housing data model. Cognos Analytics 'Starting Point' feature (Figure 8) lets me see how sale price data is related to other features of the data (IBM Cognos Analytics).

After looking at the Figure 1 demonstrates that the number of sales in Nashville is considerably higher than in any other city in the given data. This is most likely due to the fact that this data is targeting Nashville and that makes all the other cities outliers that shouldn't be included in the analysis.

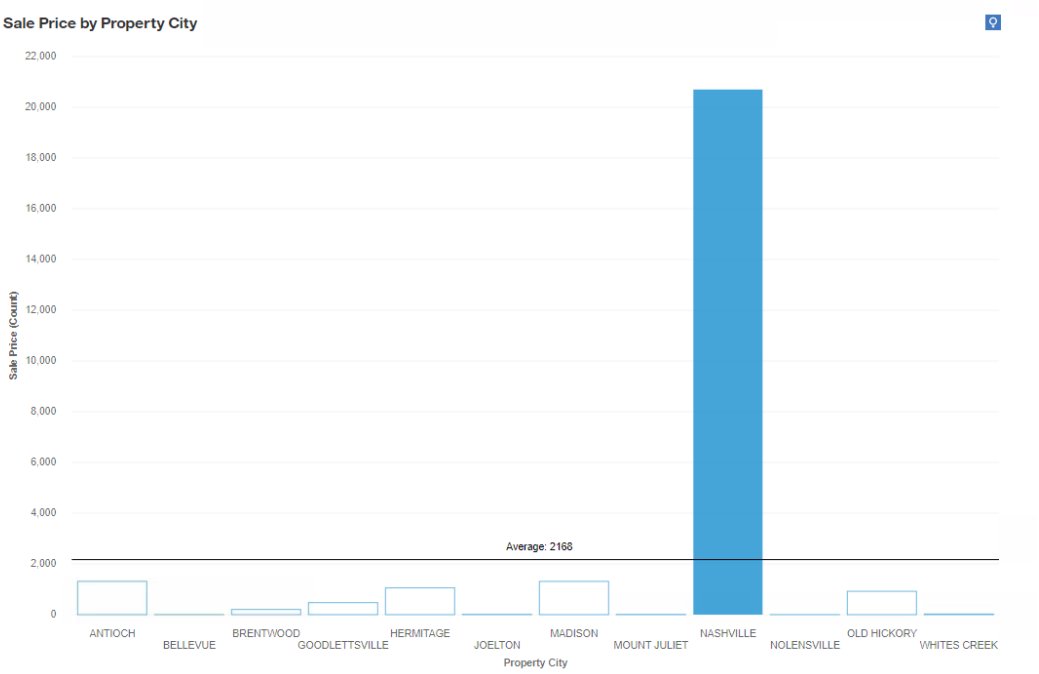


Figure 1

Once realized that the all the other city data was an outlier, I wanted to know what the most common land type was sold in Nashville during the given time frame. I found that Single Family homes took the lion share of the market in terms of the quantity of sales. The visualization in Figure 3 demonstrates this disparity.

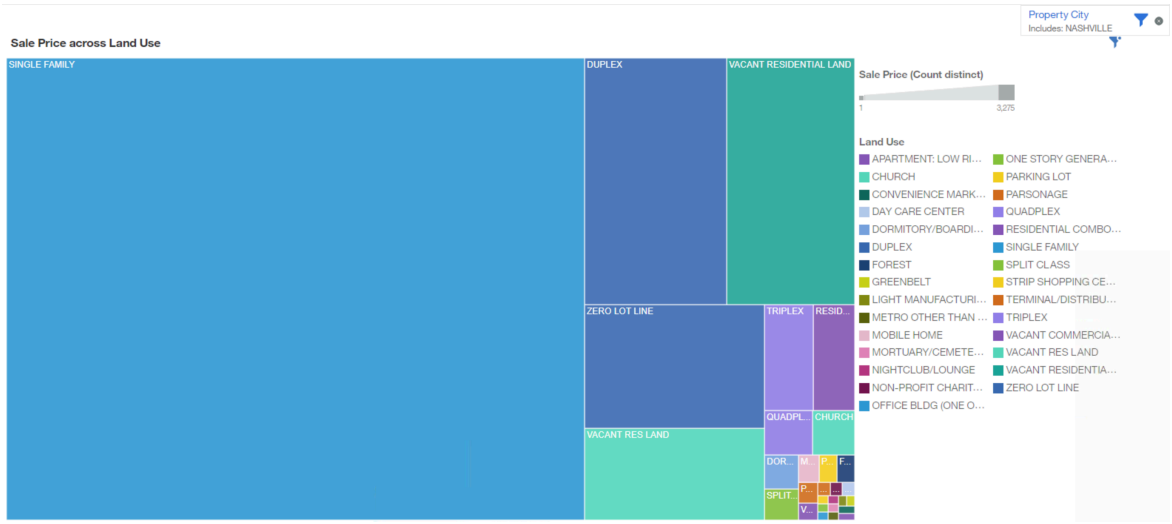


Figure 2

Then I wanted to find out what was the most expensive property sold in Nashville and was surprised to find out that it was a church and a vacant residential plot, however, there was something off about this finding because both of the values had the exact same price and might have been the outliers. I was able to find it with the visualization in Figure 3 that presents the price value for the property based on the year this property was built.

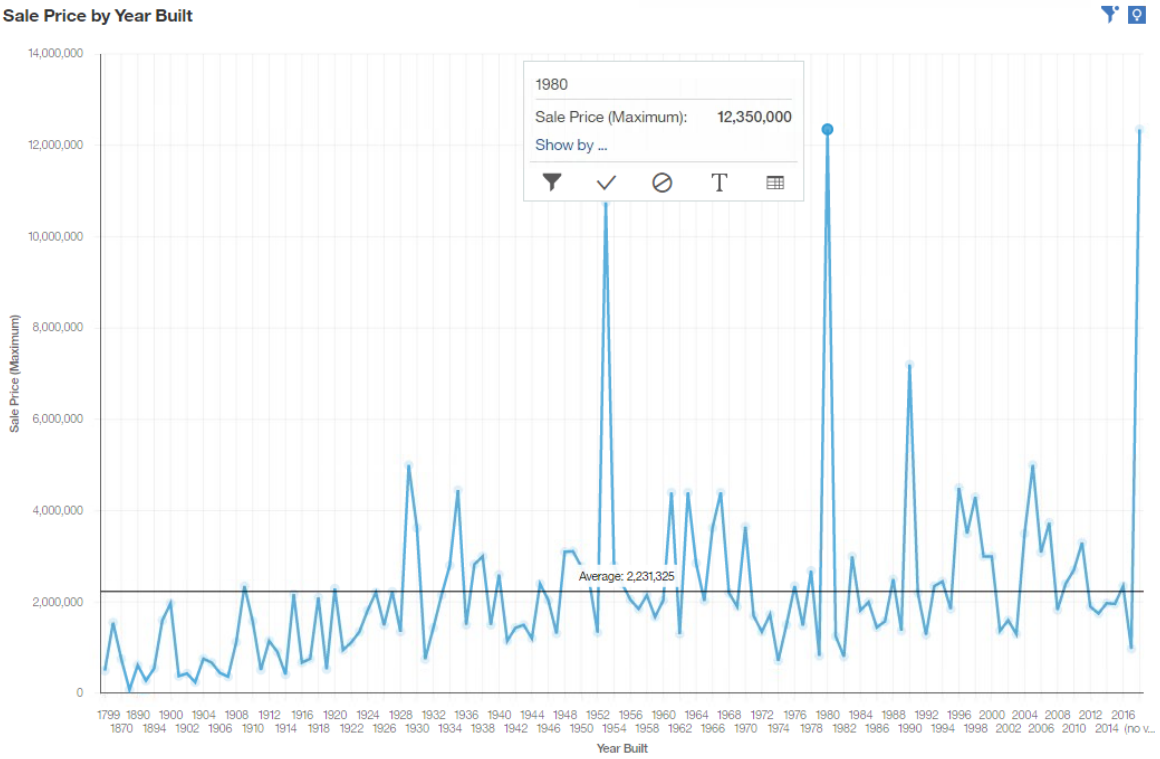


Figure 3

Continuing to dig around, I was wondering what I found that the vacant commercial property had the highest mark than any other land type sold in Nashville. Figure 4 visualizes the

just how much of the difference it presents. I used the total value of the land and compared it against the sales price and sorted by the land use. As you can see in the figure below the vacant commercial land has the highest mark up on average, even though it is not the most expensive property that was sold.

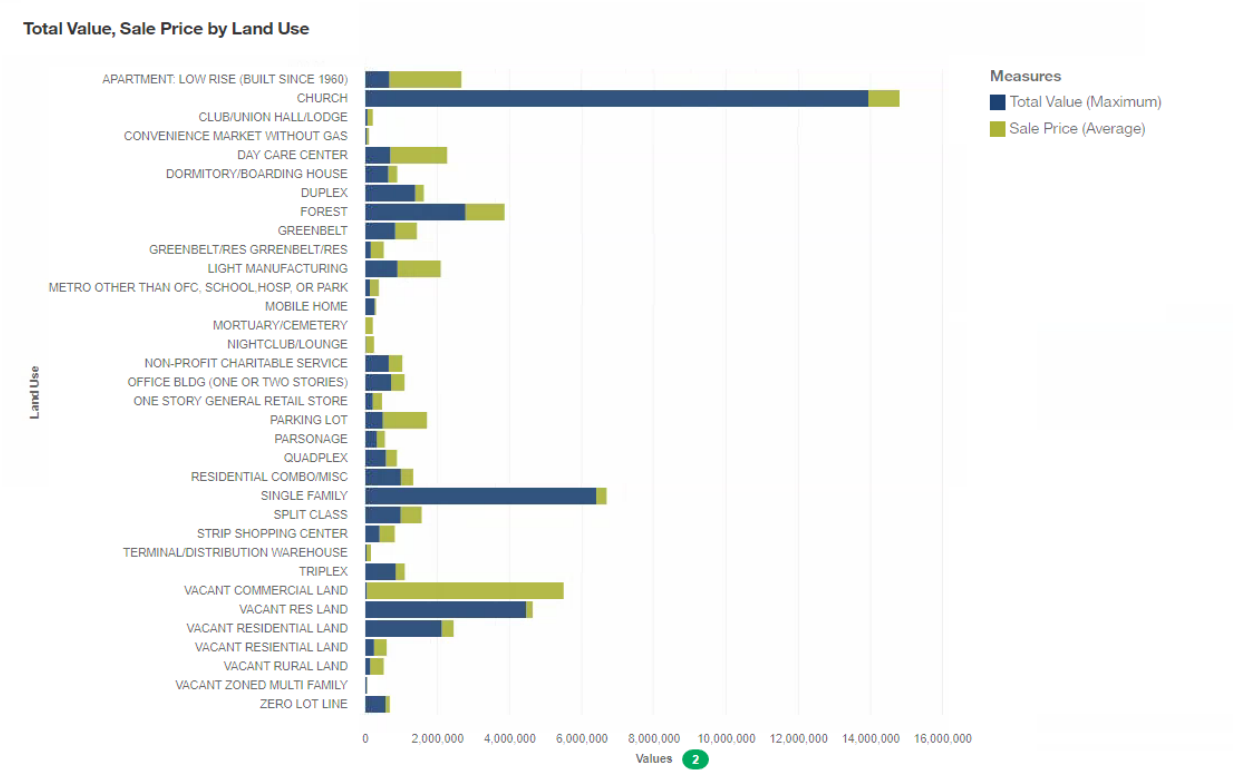


Figure 4

I also wanted to find out what was the quantity of sales by the year and month. In order to do that I had to separate the Sales Date field into two separate fields: Sales Month and Year. I used the to graph the relationship between the distinct sales in the Figure 5. It can be noticed that more property was sold during June for almost all years.

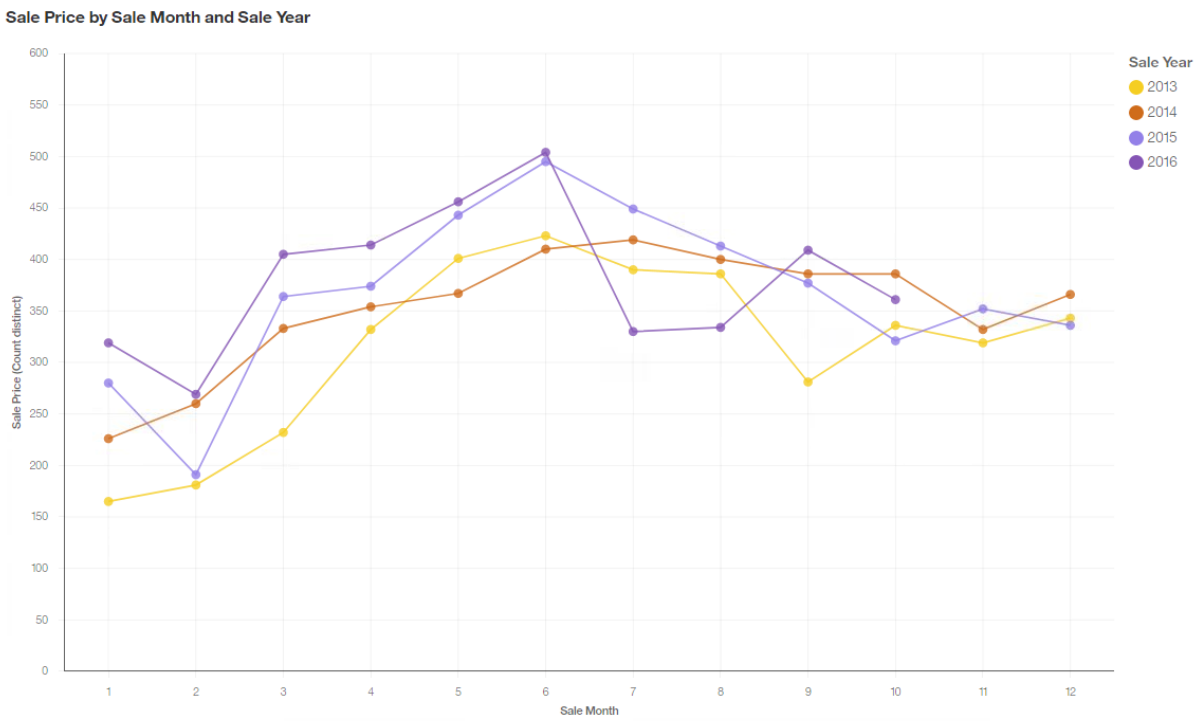
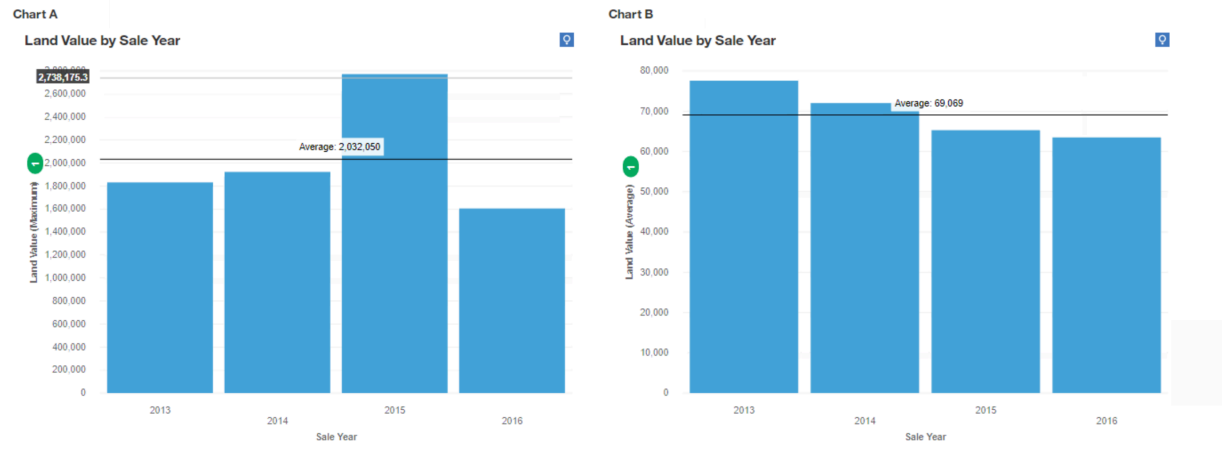


Figure 5

Another observation I made when comparing the land value between the years. In 2015 the maximum land value was much higher than any other year as seen in Figure 6. However when comparing the average land values the value is on the decline. This might due to a single large sale rather than the overall increase of the land value.



Summary	Chart A : Land Value	Chart B : Land Value	Combined
Chart total	-	-	-
Chart percent of data set	-	-	-
Minimum	1,603,800	63,474.37	-
Maximum	2,772,000	77,598.62	-

Figure 6

To determine if taxation district had anything to do with a land values, I've compared the taxation districts of Nashville and realized that on average the City of Belle Meade had more expensive land, however when comparing the maximum values of land, the Urban Service District has significantly more expensive land. The comparison is clear in the Figure 7 below.

Figure 7

The value of the visualization exploration allowed me to gain insight into the Nashville housing data. I was able to determine that the most sales were conducted in the city of Nashville as well as the fact that Single Family homes were the largest share of the market of all the sold properties. Using powerful Cognos Analytics visualization constructors, I was able to see the most expensive property and determine the outliers in the data. One of the most interesting insights the software allowed me to do is to find the mark up ratio of some of the land types and their value. Another realization was that some of the most expensive land types are churches and vacant commercial land, as well as the fact that majority of the market movements were happening in the summer months.

References

Dobbins, T. J. (2017, January 30). *Nashville Housing Data*. Retrieved from www.kagle.com: https://www.kaggle.com/tmthyjames/nashville-housing-data/metadata

IBM Cognos Analytics. (n.d.). *Explorations User Guide.* Retrieved September 2020, from www.ibm.com: https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ca_explorations.doc/ca_explorations.pdf?view=kc

Appendix A

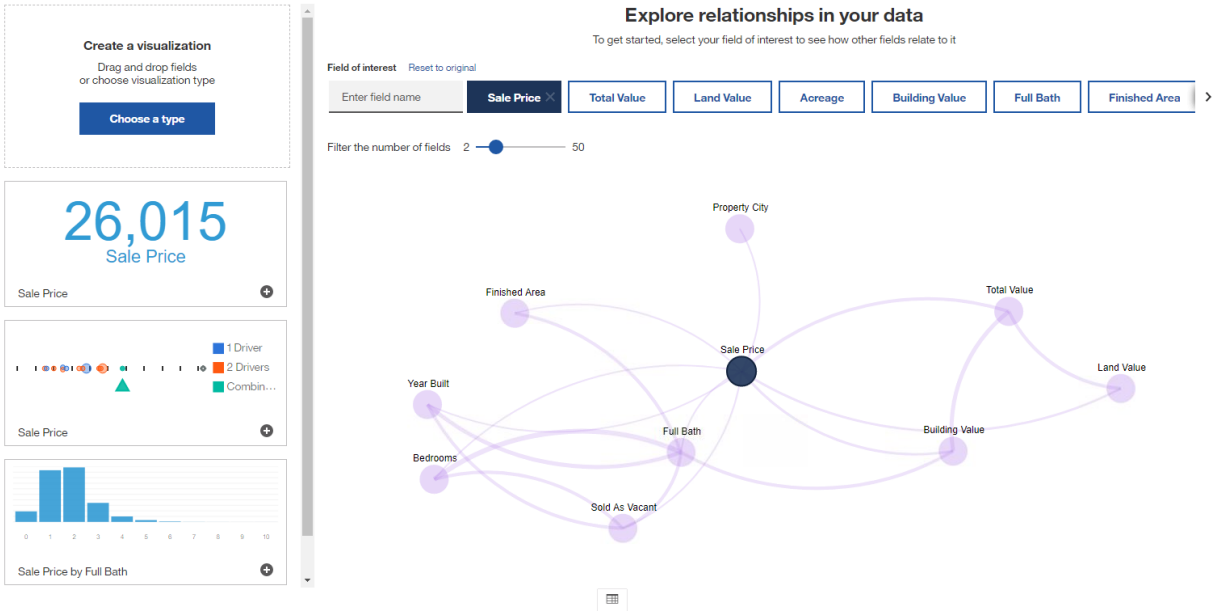


Figure 8