DATA 610-9080 Fall 2020

Assignment 2

10/23/20

Nashville Housing Data Analysis

Prepared by Kierra Young

Email: [kyoung93@student.umgc.edu](mailto:kyoung93@student.umgc.edu)

Prepared for Laila Moretto

**Introduction**

The is purpose of this report will be to analyze Nashville, TN Housing Data. This data contains 56,000 cases of data collected from the years of 1799 to 2017. The inputs vary between numerical and text. Some of the data consist of homes categories such as single homes, condos, or apartments. Other data inputs are numerical and measure the sales prices, land value, or acreage. With all the information this data set has some of the variables could be used to develop predictive models for the future of sales prices in the immensely popular hometown of Nashville, TN.

**Data Preparation**

Handling Missing Values

This data set had numerous missing values labeled as null. In order to handle the missing values, there were a couple of methods used. The first method was to make all the numerical values zero and to change the missing text values to empty. The second method was data amputation. In “Statistics for Analytics”, chapter five power point video, Professor mentioned that the more data you remove the less accurate the data will be since you are working with many missing values. However, Professor did mention data amputation at time (7:49) which will reduce the size of the data, while removing the missing values. A visual of the variables that were removed is in Table 1. model results in Appendix A.

Removing Irrelevant Variables

Furthermore, the first method did not aid because the data needed to be narrowed down. For this report, the second method that was used was amputation of the irrelevant variables. Before the amputation of irrelevant variables, the data was carefully examined. While the data was being examined it needed to be determined what information would be valuable and what information would not. The judgment that was used in order to determine the removal of the irrelevant values was by forming an initial set of questions to asked. If the variables in the original data set served no purpose in answering those questions, then the column itself was removed.

**Data Exploration Process**

Outliers

In the data exploration process, there were a few outliers. From the power point “Statistics to Analytics” Professor spoke about outliers and how they can overly influence a model. Some of the outliers that were in this dataset were the land use, sale date, and tax district. On IBM Cognos Analytics in the “Explore the Relationship” section those were the only variables without a strong connection to the main variables sales price, year built, exterior walls, land value, total value, property city, and acreage. Dr. Taylor mentioned that the outliers can be measured with standard deviation or range. For this report, range was used so it could measure the relevance of the outliers. After the outliers were examined it was determined that they were natural outliers that define how the land is characterized( single family homes, condos, apartments), the time frame of when the houses were sold, and the tax district which indicated the specific tax community each house is under. Since these outliers were of importance when making visuals most were still used. A recommendation is to adjusted the outliers would be to filter the specific type of home that needs analyzing in the field of land use, adjusting the needed time frame in the field of sale date, and specifying which tax district to be analyzed in the field of tax district.

Initial Questions Thought Process

While observing the exploration of the relationships on Cognos Analytics. One relationship was an unusual and there was an exceedingly small correlation between them. The sales price and the year built had a relationship strength of only 10%, as seen in Table 2. model results. The first initial question was “Why is there such a small correlation between the sales price and the year the house was built in?” The first thought would seem that the relationship strength would be rather strong, because the older the house the cheaper the price vs the newer the more expensive. However, with deeper analysis there were other driver factors that contribute to correlation of the year built and sales price. IBM Cognos showed the other driver factors that also contribute to the small relationship of the sales price and year built. An example is in Table 2. model results. Another initial question was “What year had the highest sales prices by tax district?” When observing the data, it could be noted that all of it came from the years of 1799- 2017. Instead of using all the years I adjusted the data to the years of 1917-2017 to get a range of one-hundred years of sales prices and tax district to see which year sold the best. On a line and column chart made on IBM Cognos Analytics it showed that very precise data range in Table 3. model results. It turned out that the year 1960 unusually had the highest rank in sales according to IBM Cognos Analytics.

Specific Questions/ Ideas

While observing the data there were a couple of specific questions that were gathered from the initial analysis. The first specific question was “Is renting out a single-family home in Nashville a good idea for property rental owners?” On IBM Cognos a visual was made to show the bestselling properties in Nashville. This type of information can be used by property owners that are interested in buying homes in Nashville to rent them out to the public. “‘Nashville has one of the most active SFR REIT sectors in the country.’ ‘In 2017, the average sales price of housing rose by 8% to $275,000 [27], this continued a growth trend that began in the previous two years. Data from the U.S. Federal Housing Finance Agency shows that the housing price index for the Nashville region has steadily increased over time without a major drop during the recession.’ ”[[1]](#footnote-1) From the data set it could be seen that the single-family homes have been the most popular homes in Nashville. Because of this information, Nashville would seem like the best place for people to rent single-family homes in. In addition, information such as this would be helpful to my organization because it shows which property city has more people renting homes. Since my organization was a hotel, they would be able to see Nashville as a possible tourism market. They could buy land there and make great profit from tourism and people visiting their families. Table 4. Model Results compares the single-family homes market in the area.

Move over, the second specific question was “Which type of housing walls (brick or frame) were more likely to get the highest sales price?”. This question was chosen because the data showed that brick and frame were the most used exterior walls for housing in Nashville area. “Everyone assumes that a brick and mortar house is more expensive to build, but that’s not necessarily true. In fact, the two methods are generally considered to be remarkably similar in terms of total material and labor costs. Time is probably the element which has the largest influence on the difference in costs between the two methods. Time plays a large role in building costs, and delays can be caused by the number of different jobs that need to be done to get a building erected.”[[2]](#footnote-2) This article explains that cost of the house can change depending on how much time it takes to build it with using either brick or frame(mortar). The idea is that knowing which type of exterior walls sells for more would not only save time, but also add more value in the total value of the house. Property owners would benefit greatly with that information before building a house to rent in Nashville. Furthermore, my organization would be able to benefit from this information because they would see what kind of community is in Nashville. Since there are more single-family homes, they would know what type of people would be more inclined to book with them. Table 5. model results show a visual of brick vs. frame exteriors and which one has the highest selling price in the year range of 2014- 2015.

Lastly, an additional specific question occurred “Does the tax district influence the total value of a house price? What would be the best tax district to live in for a person that wants to own a single-family home”? The data visualization on IBM Cognos shows the sales price by tax district by the year built. There were only six different tax districts that the data showed. The visualization shows the district with the highest sales price and the years the houses were built. The Urban Services District had the highest sales price from the years 1799-2017. IBM Cognos Analytics mentioned they their tax district is 45.8% of the all the districts combined. This information would improve relevancy when it comes to home buyers looking for a good tax district to live. The district with the least sales should be alarming because those districts also have the least sold homes as well. Likewise, my organization would be able to benefit from this information because they would be able to see which tax district would be best to build a hotel in. They can look at each tax districts sales and see which one has the most business. Table 6. model results show a visual of the sales price by tax district.

**Visualizations Created**

The visualizations that will be focused on are Tables 3-6. However, to briefly review, Table 1 as discussed in the introduction is a data set for Nashville, TN that was modified and cleaned for analysis of sales prices and home values. Likewise, Table 2, as discussed in the Initial Thoughts section was a representation of the relationships that were connected to sales prices on Nashville homes.

Moreover, Table 3 is a line graph visualization of 100 years (1917-2017) of the “Highest Sales Prices on Houses in Nashville considering the Tax District”. This visualization is valuable because it shows all the single-family home sales prices from the past 100 years. The line shows when the prices were at its highest and lowest. IBM Cognos summary was able to identify three knowledgeable aspects from this visual. First the sale price, 1960 is the most important category of year built with a total value of 25,668,951 dollars, which is 6.7% of the total. Second, the tax district Urban Services District is the most important category of Tax district with a total value of 175,746,581 which is 45.8% of the total. This information is important because around 1960 is when “The U.S. Congress authorized the creation of real estate investment trusts (REITs) in 1960 so companies could develop publicly traded real estate investment portfolios. REITs focus on commercial property, retail property, and rental property.”[[3]](#footnote-3) This piece of information is particularly important, because it shows how Congressional decisions can influence the future of real estate in the U.S. Before the authorization, data shows that housing sales prices were not as high before the 1960s. If a decision such this happened again it could greatly affect the housing prices for years to come.

Likewise, Table 4 shows “Single-Family Home Sales”. However, the sales are not in only Nashville, but also the suburb of Brentwood. Brentwood was chosen as competition city in the area. The goal was to observe which area sold the most houses between those two cities. From the visualization Nashville suburbs has higher sales prices and has sold more than Brentwood has in the past one-hundred years. Brentwood is built on old Nashville wealth. However, an article mentions that “Change in Brentwood seems to occur at a more glacial pace, and the reluctance to change funding structures, invest in transit, or address the growing economic divide feels like a holdover from the old money, old guard of power in Nashville.” [[4]](#footnote-4) The economic divide between Nashville and Brentwood may be the reason why Brentwood’s housing sales price over the years have been extremely low compared to other areas of Nashville. A suggestion would be a change property ownership. A change in ownership might be able to increase Brentwood’s house sale prices in the future.

To continue, Table 5 presents the “Total Value and Sales Price for the Year Built”. The two years that were analyzed were 2014-2015. These two years were chosen because they are very recent. The goal was to see which exterior wall has the best sales prices and total values. These houses were based in Nashville city. The types of exterior walls chosen were brick or frame. This visualization is important because it can display what type of exterior walls were more popular in the singe-family housing markets in Nashville. If an architect wanted to build a home, they would know that brick walls guaranteed higher sales than frame walls do. In the year 2014, frame wall sales were higher than brick wall sales, but according to IBM Cognos Analytics that is because most brick walls were not built/finished yet. It takes more time to complete a brick wall exterior than frame and that information can be greatly beneficial.

Lastly, the visualization in Table 6 was to compare “Sales Price by the Tax District of Nashville”. This comparison model compares the sales prices by the tax district. The goal was to view which tax district in Nashville has the stronger sales prices. The sales price was adjusted from 200,000 to 2,200,000 dollars. This is the range for single-family homes in Nashville. “An examination of the population trends in the Nashville metro-politan area since 1900 reveals that the external city has exceeded the internal city in rate of increase in five of the past six census reports.”[[5]](#footnote-5) The original tax district was the General Services District and that is why the sales prices for that district is very low. However, the Urban Services District expands beyond Nashville’s inner city and that is why their sales price was higher. Property owners would benefit greatly off knowing that the Urban Services District sales prices are 82.3% higher than the General Services District with a sales rate of 17.7%.

**Calculation**

For the calculation, I decided to use Tableau and the same Nashville Housing Data from Kaggle.com. The visuals are in Table 7. Part 1 & 2. Out of curiosity I decided to edit the dataset on Tableau to show a prediction analysis of sales price by the dimension of bedrooms. I made a linear regression model for my calculation. The results show that as the number of rooms increase as the prices fluctuate. I used a trend line model to minimize the error. The linear regression was negative which means the sales price on bedrooms declines by 9.56098 when you multiply it by 7 bedrooms plus 1.07895e+09(intercept). The R-squared value can only explain the error in the data set only by 13%. Since R- squared is not high or 80% it is not good at explaining the error. However, the P-value is 0.244038 which greater than 0.05 so that means a linear regression model such as the one I made can be used to make future predictions on sales prices by number bedrooms. Since there was an error, I also added an upper and lower confidence band just to show how much the sales prices by bedrooms can vary.

**References**

Chilton, K.; Silverman, R.M.; Chaudhry, R.; Wang, C. The Impact of Single-Family Rental REITs on Regional Housing Markets: A Case Study of Nashville, TN. *Societies* **2018**, *8*, 93.

Grant, D. (1955). Urban and Suburban Nashville: A Case Study in Metropolitanism. *The Journal of Politics,* *17*(1), 82-99. Retrieved October 27, 2020, from http://www.jstor.org/stable/2126405

Fyffe, M. (2018, December 18). Timber Frame Homes vs Brick Homes: Pros and Cons for Self-build. Retrieved October 26, 2020, from [https://cl-pm.com/timber-frame-homes-vs brickhomes/](https://cl-pm.com/timber-frame-homes-vs%20brickhomes/)

Lockman, E. Janney (2019). "Old Money, New Nashville: A Tale of Changing Wealth in Music City," Agora Journal of Urban Planning and Design, 62-67.

**Appendix** A.

Table 1. Models Results Data Set

Table

Description automatically generated

Table 2. Model Results “Data Relationships”

Graphical user interface, application

Description automatically generated

Table 3. Model Results “Years of the Highest Sales Prices on Houses in Nashville by Tax District”

Chart, histogram

Description automatically generated

Table 4. Model Results Single-Family Homes in Nashville Vs Brentwood

Chart, bar chart

Description automatically generated

Table 5. Model Results “Brick vs Frame Exterior”

Chart

Description automatically generated

Table 6. Sales Price by Tax District Comparison

Graphical user interface, application, table

Description automatically generated

**Table 7 Part 1. Model Results for the Sum Price by Bedrooms in Nashville, TN**

**Chart, line chart

Description automatically generated**

**Table 7 Part 2. Calculation Model Results for The Sales Price given Bedrooms**

**Trend Lines Model**

A linear trend model is computed for sum of Sale Price given Bedrooms.

|  |  |
| --- | --- |
| **Model formula:** | (Bedrooms + intercept) |
| **Number of modeled observations:** | 12 |
| **Number of filtered observations:** | 0 |
| **Model degrees of freedom:** | 2 |
| **Residual degrees of freedom (DF):** | 10 |
| **SSE (sum squared error):** | 8.53076e+18 |
| **MSE (mean squared error):** | 8.53076e+17 |
| **R-Squared:** | 0.132873 |
| **Standard error:** | 9.23621e+08 |
| **p-value (significance):** | 0.244038 |

**Individual trend lines:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Panes** | | **Line** | | **Coefficients** | | | | |
| **Row** | **Column** | **p-value** | **DF** | **Term** | **Value** | **StdErr** | **t-value** | **p-value** |
| Sale Price | Bedrooms | 0.244038 | 10 | Bedrooms | -9.56098e+07 | 7.72371e+07 | -1.23787 | 0.244038 |
|  | | | | intercept | 1.07895e+09 | 5.01546e+08 | 2.15126 | 0.0569434 |

1. Chilton, K.; Silverman, R.M.; Chaudhry, R.; Wang, C. The Impact of Single-Family Rental REITs on Regional Housing Markets: A Case Study of Nashville, TN. *Societies* **2018**, *8*, 93 [↑](#footnote-ref-1)
2. Fyffe, M. (2018, December 18). Timber Frame Homes vs Brick Homes: Pros and Cons for Self-build. Retrieved October 26, 2020, from https://cl-pm.com/timber-frame-homes-vs-brick-homes/ [↑](#footnote-ref-2)
3. Chilton, K.; Silverman, R.M.; Chaudhry, R.; Wang, C. The Impact of Single-Family Rental REITs on Regional Housing Markets: A Case Study of Nashville, TN. *Societies* **2018**, *8*, 93 [↑](#footnote-ref-3)
4. Lockman, E. Janney (2019). "Old Money, New Nashville: A Tale of Changing Wealth in Music City," Agora Journal of Urban Planning and Design, 62-67. [↑](#footnote-ref-4)
5. Grant, D. (1955). Urban and Suburban Nashville: A Case Study in Metropolitanism. *The Journal of Politics,* *17*(1), 82-99. Retrieved October 27, 2020, from http://www.jstor.org/stable/2126405 [↑](#footnote-ref-5)