STAT 15 - Final Project

Chelsea Lee

Pasadena City College

Introduction

For my final project's topic, I have chosen to identify property features that most significantly impact property valuation to guide the Real Estate Investment Trust (REIT) in property acquisition and capital improvement decisions. This topic is intriguing because it seems more complex in comparison to the others due to the numerous variables in the dataset. It is an opportunity to explore more complex relationships and interactions between variables as well as improve my statistical and Excel skills. This calls for stronger analytical skills and interpretations, producing a greater challenge against my skills. In addition, researching this topic can be a valuable experience for future projects in data analysis and finance, which is the career path I would like to venture into. All in all, I believe this research topic will provide a more comprehensive learning experience that can transfer onto real-world issues.

Data Analysis

Numerical Description

In this ReMax Real Estate dataset, I will be analyzing the following variables: the number of bedrooms, the total square footage of the property, the presence or absence of a pool, the distance from the nearest emergency room in miles, the distance from the nearest school in miles, the presence or absence of a garage, and the number of bathrooms on their impact on home prices in hundreds of thousands. The average price of these homes is $221.10, the average number of bedrooms is 3.8, the average square footage is 2,223.81, the average presence or absence of a pool is 0.36, the average distance to the nearest emergency room is 14.63 miles, the average distance to the nearest school is 3.1 miles, the average presence or absence of a garage is 0.68, and the average number of bathrooms is 2.08. The median for home prices is $213.60, for

the number of bedrooms it is 4, for square footage it is 2,200, for the presence or absence of a pool it is 0, for the distance to the nearest emergency room it is 15 miles, for the distance to the nearest school it is 3 miles, for the presence or absence of a garage it is 1, and for the number of bathrooms it is 2. The most common price is $188.30, for the number of bedrooms it is 4 and 3, for square footage it is 2,100, for the presence or absence of a pool it is 0, for the distance to the nearest emergency room it is 16 miles, for the distance to the nearest school it is 4 miles, for the presence or absence of a garage it is 1, and for the number of bathrooms it is 2. Because the presence or absence of a pool and garage is a categorical variable, I believe the mode is the best evaluator since it indicates whether most homes have or do not have a pool and garage. In this case, it is more common for a home to not have a pool and more common for it to have a garage. The average amount of variability in the prices and square footage is $47.12 and 248.67. These large values indicate a greater distribution of home prices and square footage. The average amount of variability in the distance to the nearest emergency room and school is 4.87 miles and 1.29 miles. Since their averages are 14.63 miles and 3.1 miles, I would say this standard deviation is relatively modest indicating a moderate spread. The average amount of variability in the number of bedrooms and bathrooms is 1.50 and 0.39. Since the averages are 3.8 bedrooms and 2.08 bathrooms, I would say the standard deviation for bedrooms is quite large indicating a greater distribution. However, the standard deviation for bathrooms is relatively small, indicating a clustered distribution. The average amount of variability in the presence or absence of a pool and garage is 0.48 and 0.47. Since their values range from 0 to 1, this is a moderate distribution. The cheapest home is listed at $125 while the most expensive home is listed at $345.30. The lowest number of bedrooms is 2 and the largest number is 8. The smallest square footage is 1,600 and the largest is 2,900. The closest distance to the nearest emergency room is 6 miles

while the farthest distance is 28 miles. The closest distance to the nearest school is 1 mile while the farthest distance is 5 miles. The lowest number of bathrooms is 1.5 and the largest number is 3. Due to these significant ranges, we must keep in mind that there is a possibility of extreme outliers that can affect the mean and standard deviation values.

Confidence Interval

I do not have the average cost of homes, square footage, the number of bedrooms and bathrooms, the distance to the nearest emergency room and school, or the presence of a pool and garage in the entire Oxnard area. However, using this sample of 105 homes I can conclude with a 95% level of confidence the averages of these variables. The average cost of homes in this area lies between $212.09 and $230.11. The average number of bedrooms lies between 3.51 and 4.08. The average square footage lies between 2,176.25 and 2,271.37. The average number of bathrooms lies between 2.01 and 2.16. The average presence or absence of a pool lies between 0.27 and 0.45. The average presence or absence of a garage lies between 0.59 and 0.77. The average distance to the nearest emergency room lies between 13.70 and 15.56. The average distance to the nearest school lies between 2.86 and 3.35.

Regression

The variables: the number of bedrooms, the total square footage, the presence or absence of a pool, the distance to the nearest emergency room, the distance to the nearest school, the presence or absence of a garage, and the number of bathrooms, directly determine about 73.05% of a home's listed price. Using a 5% level of significance, I have sufficient evidence to conclude that there is a relationship between the price of a home and at least one of the variables. Furthermore,

I believe that the number of bedrooms, the square footage, the presence or absence of a pool, the presence or absence of a garage, and the number of bathrooms play the most significant role in a home's pricing.

Conclusion

My original research question is as follows: which property features most significantly impact a home's value? My research has shown the averages, medians, modes, standard deviations, and ranges of the number of bedrooms and bathrooms, the total square footage, the distance to the nearest emergency room and school, and the presence or absence of a pool and garage. I have also found the average population range of these variables in the surrounding Oxnard area. Finally, through these calculations, I have found that the number of bedrooms and bathrooms, the total square footage, and the presence or absence of a pool and garage most significantly impact a home's value. If I were to do this project again, I would incorporate additional variables, such as the neighborhood's quality and the age of the property and evaluate the interaction between variables to better understand the effect of these features on property price. In my opinion, working alone is more efficient than working with others for this project because the calculations and interpretations were not too intensive or challenging to require assistance.