



Forecast and Plan the sales - A model of sales day's prediction on real estate market



Tae-ei Park and Dr. Jun Huang

ABSTRACT Since statisticians invented the statistical model, and many scholars have used various statistically logic models to prove theories for business, science, and engineering. Like them, this research will create a specific real estate model, using the sample city of Dallas. In this article, a researcher will infer Dallas house price matrix, crime rate, number of bathrooms and marketing days to make statistical modeling. To do this, the researcher will collect Dallas real estate data from 2016 to 2017. After researcher receiving the data and annotate concisely, the researcher will separate the independent and dependent data, and the standard will decide by the statistical model. At the end of the project, the researcher will make the logistic model to explain which house has more than 60 business days and less than 60 business days. Also, the researcher sure that this real estate statistical model will be beneficial will contribute to someone or an organization.

INTRODUCTION

One of the fastest growing city in the United States of America is Dallas. Recently a lot of real estate companies and institutions are looking for forecasting. Texas Real Estate Center is one of them, and they mention "Total Texas housing sales managed 4 percent annual growth, outpacing the national rate for the second straight year" (Gary, Texas A&M REAL ESTATE CENTER). There is another Texas real estate information for you. Here is the graph (Figure 01). The under graph shows us about average days on the market from 2008 to 2018. When the time is going to pass by the year, the average days on the market is going to decrease that mention to us Texas housing economy is growing fast. Therefore, this research help to forecast how fast the house can be sold on the market. The researcher has been collecting 15 variable and 363 data to make the prediction. In this research, the researchers used logistic regression, we use accuracy rate and AUC to measure the performance of our model. The dependent variable is a binary variable which measures how many days they require in the market. The Independent variable has the 15 variables, such as Number of bathroom and bathroom, Square foot, lot, Price per square, Median listing price, Median days on market, School ranking, Built of renovation, Location, House price, Crime rate, Type of family, Types of school around by house, and Zip code. In logistic regression, there are two values existed. One is 0, and another one is 1. If the house has less than or equal to 60 days business time required, it denoted with 0. However, if the house needs more than 60 days, it is called 1. Which originates from how many days the house is on the market before it has been sold

10 Year Charts of Average Days on Market of Single Family Homes in the City of Dallas, TX

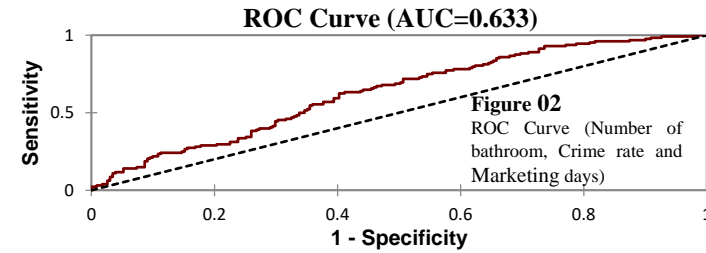


Figure 01
10 Year chart of Average Days on Market (2008 - 2018)

METHOD

PREPARATION AND PLANNING.

Before I started the statistical analysis, I needed to collect the data. To this end, I used Zillow, Trulia, Realtor.com and real estate market report, and I have been compiling this data by taking into account 15 variable and a pool of 359 dataset. Before feeding the data into statistical software, junk data was removed. In this research, the marketing days are dependent variables. For logistic regression, we require less than 60 days to be 0 and more than 60 days to be 1. However, some data was not available because of open housing. Therefore, I deleted 43 datasets from the pool. For software, Excel's logistic regression function was used. In logistic regression, I have numerical data and categorical data. Finally, I create the ROC Curve(AUC=0.633), (Figure 02). The definition of ROC is a fundamental tool for diagnostic test evaluation. This graph is used a lot in forecasting areas, and the value of the range is from 0 to 1. Normally, ROC curve used to assess the validity of a model in any of several different ways. The dashed line in the diagonal presents the ROC curve of a random predictor. It has an AUC of 0.5. It is used as a baseline to show whether the model is useful and 0.633 is just slightly greater than 0.5. The result can be improved if we can increase the sample size and include more variables, such as the distance to closet hospital, supermarket, and feature of the house style such as the architectural style. In the future study, we can also interview people about the determinants features of buying a house for them. Limitations to this research include constraints of budget and time, which prevented me from collecting more diverse data or verifying the reliability of my data.



RESULT AND DISCUSSION

If we apply this logistic curve, we should forecast the prediction days on the market. Here is a real-life example for the audience: If someone flips a coin, they have a 50% chance of either heads or tails. In the ROC curve, a probability value of less than 50 % is useless. We already have 50% guaranteed for the result; therefore, in this research, the value 0.633 is not a meaningless value. Also, the researcher used the marketing days, the number of bathrooms and the crime rate to make the logistic graph. All of these factors are helpful in predicting the future. Therefore, if the customers check the number of bathrooms and the crime rate, they will forecast their marketing days.

(Table 01)
Classification table for the training sample (Variable Y1):

from \ to	0	1	Total	% correct
0	229	2	231	99.13%
1	125	3	128	2.34%
Total	354	5	359	64.62%

Table 01 (Accuracy rate of the data classification)

Reference

- John, J. (April 10, 2018) 10 Year Charts of Median and Average Sales Prices of Single Family Homes in the City of Dallas, TX. Retrieved from <http://www.homesourcedallas.com/reports/dallas-tx-real-estate-market-report/>
- MedCalc Software byba, MEDCALC easy-to-use statistical software. (2018). Retrieved from <https://www.medcalc.org/manual/roc-curves.php>
- Torres, L., Miller, W., Guadra, B. Outlook for the Texas Economy, Real Estate Texas A&M Center. Retrieved from <https://www.recenter.tamu.edu/articles/technical-report/outlook-for-the-texas-economy>