Final Project PART I: Predicting Housing Prices

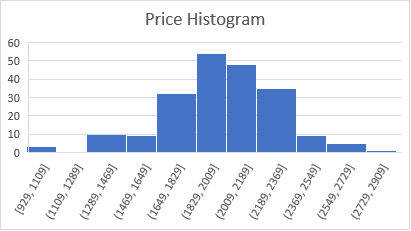
Blake Rosenhagen

Concordia University Texas

Intro

The goal of this model is to predict the value of certain real-estate locations, to reduce the man power needed to search for “highly undervalued apartments”. By eliminating locations that are not worthy of further prospect, employees will be able to use their time more efficiently.

With the data gathered, I have determined that there were four factors that may affect the price of the real-estate location. These are the distance from the sea, house square footage, number of cafes and convenient stores in a 1-mile radius, and finally, whether there was a shopping mall within a 3-mile radius to the apartment. A snap shot of the data used is posted below.



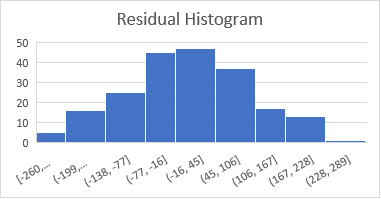
Stage 1

To begin the process, I created a linear model that took into account all of the factors listed above. Next, I checked to see which factor were significant, meaning which ones actually helped determine the price of the location. It turns out that only the distance to the sea and house square footage were significant in determining the price. AS a rule of thumb, for every 100 square feet added to a apartment, you can expect the price to rise around 220 dollars. Related ANOVA table and data plots for the four factors below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 4 | 16031671 | 4007918 | 381.1398 | 1.3E-92 |  |  |  |
| Residual | 201 | 2113638 | 10515.61 |  |  |  |  |  |
| Total | 205 | 18145309 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 45.12404 | 66.89504 | 0.67455 | 0.500737 | -86.782 | 177.0301 | -86.782 | 177.0301 |
| DISTANCE | -0.07515 | 0.008896 | -8.44754 | 5.94E-15 | -0.09269 | -0.05761 | -0.09269 | -0.05761 |
| SQRFT | 2.171585 | 0.057098 | 38.03253 | 9.09E-94 | 2.058997 | 2.284173 | 2.058997 | 2.284173 |
| CAFES# | -7.62079 | 6.588901 | -1.15661 | 0.248804 | -20.613 | 5.371445 | -20.613 | 5.371445 |
| SHOPMLL | -20.2066 | 16.28268 | -1.24098 | 0.216058 | -52.3133 | 11.90022 | -52.3133 | 11.90022 |

At this stage a ran a more specific model that only took into account the factors of distance of the apartment to the sea and Square footage. This way I am able to more accurately determine if the two factors chosen out of the four, helped with the prediction of price.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 2 | 16002133 | 8001067 | 757.854 | 6.88E-95 |  |  |  |
| Residual | 203 | 2143176 | 10557.5 |  |  |  |  |  |
| Total | 205 | 18145309 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 14.70595 | 64.44078 | 0.22820 | 0.81971 | -112.353 | 141.765 | -112.353 | 141.765 |
| DISTANCE | -0.07563 | 0.008825 | -8.57013 | 2.61E-15 | -0.09303 | -0.05823 | -0.09303 | -0.05823 |
| SQRFT | 2.180478 | 0.056935 | 38.2977 | 7.73E-95 | 2.068218 | 2.292737 | 2.068218 | 2.292737 |



Indeed, the second model has a lower “significant f”, this means that it has a greater “accuracy in prediction”. The data plots for both factors will be the same as what is stated illustrated previously.

Stage 2

At this point in the process, we use the models generated with a test data to simulate the prediction ability of said models. We obtain the predicted prices for each location and find how far we are off by looking at the mean squared error (MSE).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mdl4MSE | 8627.75 |  |  | Mdl2MSE | 8042.95 |

The output of this process is shown above. The information we gain from this is that the model is that the second model has a lower MSE, which explains how close the predictions were to the actual price.

From the analysis done, as well following the Parsimony Principle, it is decided that the two-factor model is our best choice for application, because it better predicts and is the simpler model.

Stage 3

Conducting our final analysis, we are given random sample of 10 out of 9100 apartments in the same region as the locations used to create the model. The data inputs as well as the predicted price are labelled below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ID | Intercept | DISTANCE | SQRFT | CAFES# | SHOPMLL | PredTotalprice |
| 1 | 1 | 2499 | 738 | 1 | 1 | 1434.896 |
| 2 | 1 | 3795 | 1152 | 0 | 1 | 2239.596 |
| 3 | 1 | 3810 | 693 | 0 | 1 | 1237.622 |
| 4 | 1 | 3258 | 1035 | 1 | 1 | 2025.094 |
| 5 | 1 | 3414 | 819 | 1 | 0 | 1542.313 |
| 6 | 1 | 2340 | 1008 | 2 | 1 | 2035.651 |
| 7 | 1 | 2022 | 1026 | 0 | 1 | 2098.95 |
| 8 | 1 | 3351 | 1026 | 1 | 0 | 1998.436 |
| 9 | 1 | 2949 | 1035 | 0 | 1 | 2048.464 |
| 10 | 1 | 3807 | 972 | 2 | 1 | 1846.203 |

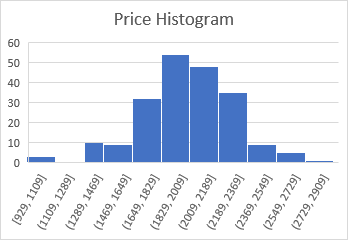
What we collect from this data in summary is that waterfront and high square footage locations hold the most value. This is also confirmed by SFGate, a real-estate website that hosts guides, living near or on the water is an attractive environment ("Why Does Waterfront Property Cost More?", 2019). My recommendation for the real-estate company is to invest in and market towards waterfront locations, this will increase the chances of purchasing a home that is undervalued, and can be turned over for a greater profit. and or market their to families, since more people require more square footage ("How Big Should Your Home Be?", 2019).

Final Project PART II: Housing Prices comparison

Blake Rosenhagen

Concordia University Texas

The goal of this testing was to figure out if there was a difference in price of an apartment location whether the locations was within a shopping mall or outside of one, meaning at least 3 miles away form one. By using data from a total of 206 different locations, I constructed a couple of statistical test to address this question. Here is a chart illustrating the different price points for locations.



The results I received for the first test concluded that there is indeed a price difference for apartments that were inside compared to farther away with the locations within 3 miles of a shopping mall costing about $61 more than locations outside that range. The framework and output is posted below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mean of 0 | 2022.7818 |  |  |  |  |  |  |
| SD of 0 | 265.3682 |  |  |  |  |  |  |
| n of 0 | 55 |  |  |  |  |  |  |
|  |  |  | S^2 | 296.4378 |  | t score | 22.33219 |
| Mean of 1 | 1962.2252 |  | df | 204 |  | p value | 0 |
| SD of 1 | 307.6228 |  | SE | 2.711631 |  |  |  |
| n of 1 | 151 |  |  |  |  | Less than .05 | |
|  |  |  |  |  |  |  |  |

Another statistical test was conducted to ensure a thorough examination. This time the goal was to create a range that we could decipher if it would be reasonable to assume that the shopping mall did determine the price. This test was done through the information below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Point Estimate | | 60.5567 |  |  | Lower | Upper |  |
| df |  | 204 |  |  | 55.20959 | 65.90372 |  |
| t\* |  | 1.9719 |  |  |  |  |  |
| SE |  | 2.711631 |  |  | Confidence Level: 95% | | |

Again, we are greeted with the same result that apartments near shopping malls are generally more expensive then one that aren’t. The advice that can be given hear is if a real-estate agent is looking for an investment that could potential yield higher payout, he/she should choose to buying apartment near shopping malls.

Something that is worth noting is the possibility of shopping malls being more expensive places since they tend to be placed in highly populated areas. Below there is a map of the greater Austin area where warming colors are higher priced real-estate, whereas cooler colors are cheaper. I have also placed down a couple of markers that depict mall/large shopping centers with a blue x. You can see that the price around the malls are higher, but also due because they are closer to the center of the city (Trulia, 2019).

A close up of a map

Description automatically generated

References

(2019). Retrieved from https://www.trulia.com/for\_sale/Austin,TX/12\_zm/

How Big Should Your Home Be?. (2019). Retrieved from <https://www.newhomesource.com/guide/articles/how-much->square-footage-fits-your-family

Why Does Waterfront Property Cost More?. (2019). Retrieved from <https://homeguides.sfgate.com/waterfront-property> cost-more-34701.html