Hello Homes Expansion Project: Builders Edition

By Monique Hercules

Summary

Our stakeholder is a local real estate agency that help homeowners buy/sell homes. After years in the industry they have decided to expand to include the building and flipping of homes as well to better meet there consumer needs and increase revenue.

Today we will use multiple linear regression models to analyze house sales in the Washington state King county area.

Outline

- Business Problem
- The Data
- Methods
- Modeling
- Results
- Conclusion
- Next Steps

1. Business Problem

- Hello Homes would like to find out what types of homes they would need to build based off of there past sales to meet there consumers needs and increase revenue,
- They are targeting the most popular price ranges, sqft living, and accommodation preferences like bathrooms, bedrooms and number of floors.
- Today we will provide evidential advice on various aspects of a home to fit the builders needs. By targeting this information the real estate agency will have more homes available to sell to their consumers. With the use of multiple linear regression we will be able to accurately see the relationship between the various attributes of a home and and how they reflect upon the homes price.

Data

The dataset in use will be from King County in Washington State.

2.Methods

Chose multiple linear regression to accurately study the narrative between the various aspects of a home and how they influence the price.This method will allow for manual manipulations to individual variables and how they affect the price. Before analyzing the data for Hello Homes to begin modeling the data was first prepared in with various methods:.

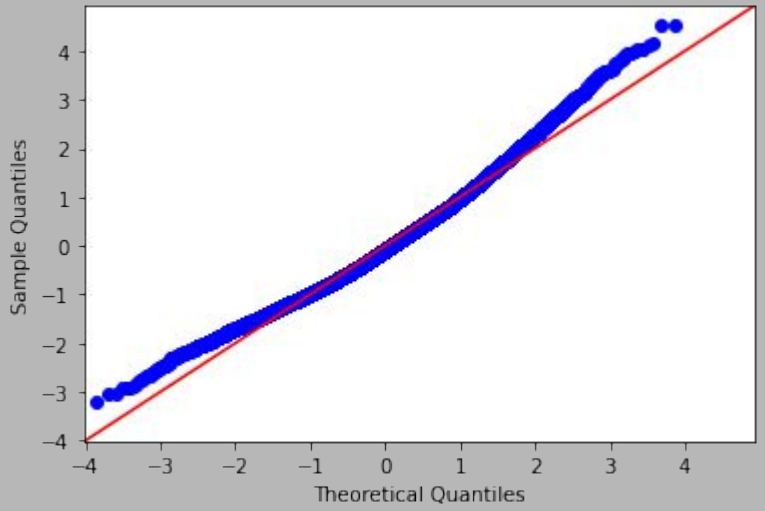
- **Data Cleaning**
  Dropped unnecessary columns to make them easier to work with. Cleaned through the various data by removing Nan variables (when applicable), commas and the symbols and while also converting data types as well as removing outliers. By doing this made made the model more accurate overall.

- **Verifying Assumptions for Linear Regression**
  Ensured linearity, normality and multicollinearity are met before modeling

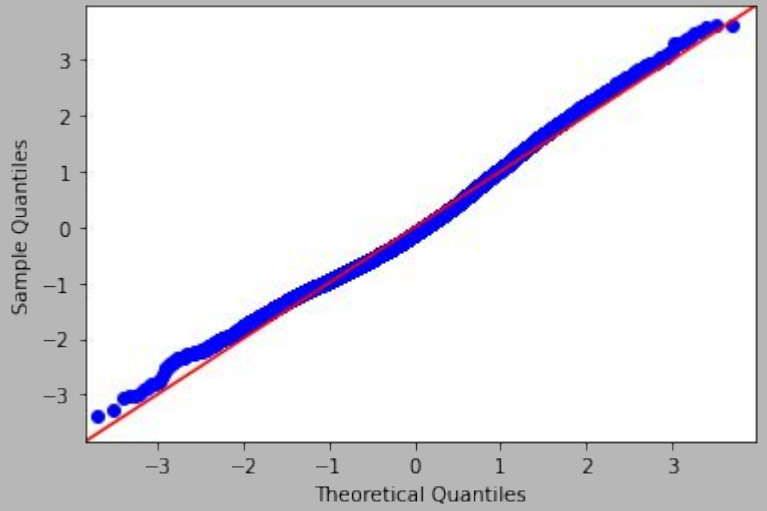
- **Preprocessing**
  One Hot Encoded chosen categorical variables. Decided against feature scaling to work with numbers in a more natural setting. Normalized my chosen numerical variables to reduce multicollinearity.

- **Modeling**
  Used Statsmodel with OLS to model data, with the use of p values, variance check and homoscedasticity to refine model.

Modeling

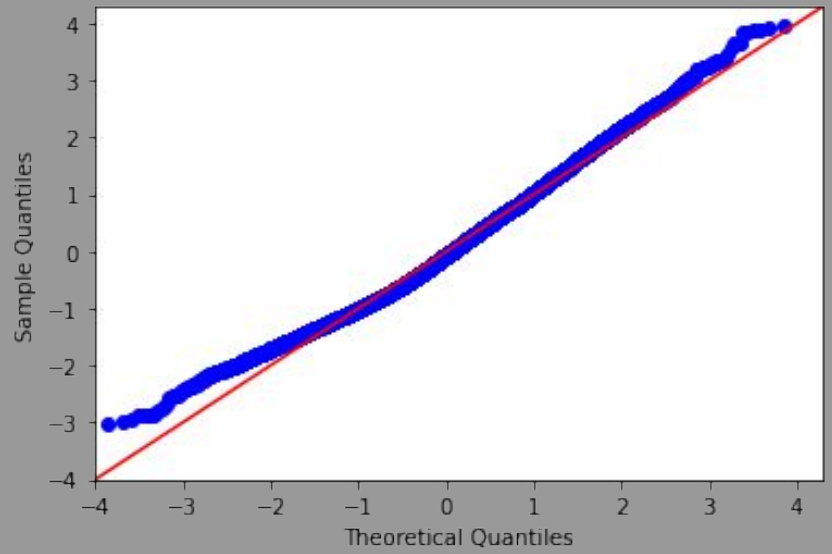


Model 1 R-squared = 0.44



Model 4 R-squared = 0.268

Modeling



Model 3 R-squared 0.415

3. Results

Decided to use model 3 to draw results from:

- **SQFT**
  The ideal price per sqft living being around $95
- **Conditions**
  Homes that had conditions fair degrades price range by $32,000 while homes with condition very good increase home prices by $70,000 and homes with the condition good increases the price by $25,000
- **Floors**
  Lofts, townhouses and three stories homes are the most profitable with lofts being the best by increasing the price by $77,000. Avoid two story homes which degraded the price of a home by 23,000

Condition Status

Fair

Brings down home value by
$32,000

Very Good

Increases home value by
$70,000

Good

Increases home value by
$25,000

Conclusion

Target price per a sqft at $95

All houses flipped/made need to meet the condition good to ensure the best price value.

Focus on townhomes and 3 story homes while avoiding 2 story homes but lofts are the best value.

Recommendations

3. Next Steps - Considerations

Based on Model 3

➔ **Limitations**

The model is on the weaker side with a R-squared value of 0.415 preferable we would have a model with a higher R-squared conveying a more accurate model.

➔ **Limited Data**

Bedrooms in model 3 is negative as well as lacking options in their respective groups to draw conclusions from. The grading status seems to be unreliable with even positive grading scale decreasing the price of a home. More aspects of a home should be explored to be a able to convey a model diverse model.

➔ **Issues with multicollinearity**

Even after verifying assumptions for multicollinearity the error still appeared under the models final results, for future use multicollinearity need to be reduced.

THE END

email : moniquehercules101@gmail.com

Github @moniquehercules

LinkedIn LinkedIn/in/moniquehercules