

🔗 main ▾

🔗 2 branches

🏷 0 tags

Go to file

About ▾

Code ▾

⚙

Your main branch isn't protected

Protect this branch from force pushing or deletion, or require status checks before merging.
[Learn more](#)

Protect this branch

No description, website, or topics provided.


- 📖 Readme
- ★ 0 stars
- 👁 1 watching
- 🔗 1 fork

Releases

No releases published
[Create a new release](#)

Packages




No packages published
[Publish your first package](#)

	71bb6b1	5 hours ago	🕒 56 commits
📁 data	removed imag...	5 hours ago	
📁 imag...	updated visual...	5 hours ago	
📁 note...	organization of...	5 hours ago	
📄 .giti...	Adding .gitign...	last week	
📄 Final...	fixed visuals	5 hours ago	
📄 REA...	fixed visuals	5 hours ago	

Contributors 2

-  albertcchen Albert Chen
-  anniezhengaz Annie Zheng

Languages

	data...	Created final n...	2 days ago
	phas...	added present...	7 hours ago
	pres...	added present...	7 hours ago

● Jupyter Notebook 100.0%

 README.md 



Recommendations for First Time Single-Family Home Buyers in King County, WA

Authors: Annie Zheng & Albert Chen

Overview & Business Problem

Our stakeholders are a King County real estate agency and families looking to purchase their very first home. They are looking to find what underlying factors are critical to the perfect home and how these factors may affect the sale price.

Purchasing a home is a huge step in one's life and a big undertaking that requires significant forethought, especially if the home is for a family. Our analysis will provide recommendations and provide assistance to first time single-family home buyers by looking into critically values home features and factors they should focus on in their home journey. Livable square footage, building grade, and condition grade were the focal points of this analysis. By analyzing the correlation of these factors to the price of the home, families can use these results to make an informed purchase with confidence for their very first home to plant their roots within King County, Washington.



The Bottom Line

1. Settle down *out* of Seattle.
2. Bigger homes available at a lower price out of Seattle.
3. Better construction and quality comes at a higher price. Better prices out of Seattle.

Data Understanding

The data used in this analysis was taken from [King County Department of Assessments](#), whereas the `address`, `lat`, and `long` fields have been retrieved using a third-party [geocoding API](#). The data provided information such as sale price, sale and renovation years, square footage of specific home features, as well as number of bedrooms and bathrooms. The target variable in our analysis was the sale price of the home which laid a foundation for the rest of the analysis and modeling. The dataset contains numerical data, with the instances of categorical data which is then converted into numerical data. Our final dataset had numerous fixed factors as well as the removal of price outliers.

Data Cleaning

The dataset appeared to have one row of duplicate data as well as three rows with null values in the `sewer_system` and `heat_source`. We have elect to drop those rows instead of imputing it because the number of rows was considerably small compared to the dataset.

In the cleaning process, it was discovered that a considerable number of homes were not within King County, 911 rows to be precise. King County postal codes begin with the prefix '98' and those with postal codes without that prefix were dropped from the dataset as well.

There were numerous categorical variables that will cause problems when fitting a linear regression line. To ensure the dataset contained numeric continuous data, we had created new columns based off of the `condition` and `grade` columns with their respective numeric code values.

To weed out extreme outliers, we dropped homes whose sale price was beyond 1.5 times the InterQuartile Range. This data limitation may not capture the price range of all single families.

The following factors were considered as fixed:

- bedrooms : at least 2 bedrooms
- bathrooms : at least 1 bathroom
- floors : only one floor
- yr_built : houses built after 1977 due to the asbestos ban
- grade_code : building grade of 6 and above
- condition_code : condition code of 3 and above

The following variables were excluded from our model exploration as they were not of interest to our business problem and the stakeholder: date , condition , grade , heat_source , sewer_system , address , yr_renovated , sqft_above , sqft_basement , waterfront , greenbelt , nuisance , view

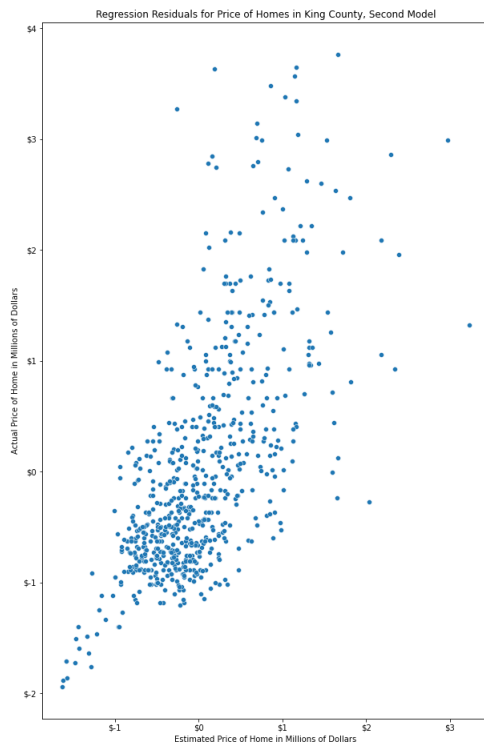
Modeling with the Dataset

With the cleaned dataset, 2,713 homes remained and scikit-learn was utilized to model the data. Firstly, it was essential that we scaled the data utilized StandardScaler as variables such as number of bedrooms and price were not on the same scale.

Dummy Regressor Model

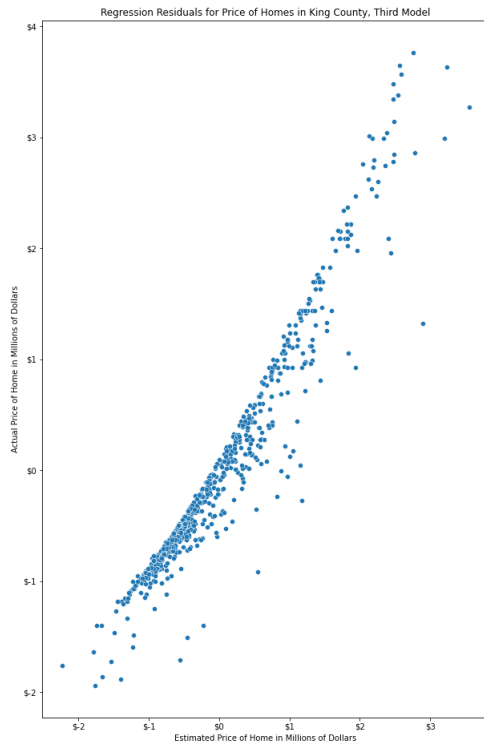
Started by making an initial Dummy Regressor model as a baseline. The R-squared score was close to 0 which is expected, and we could only go up from there!

Second Model



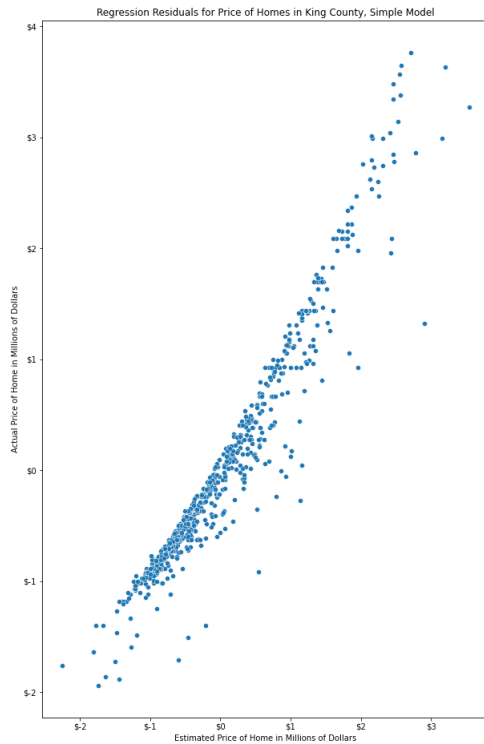
Our second model here focuses on our most correlated feature, which is price_sqft.

Third Model



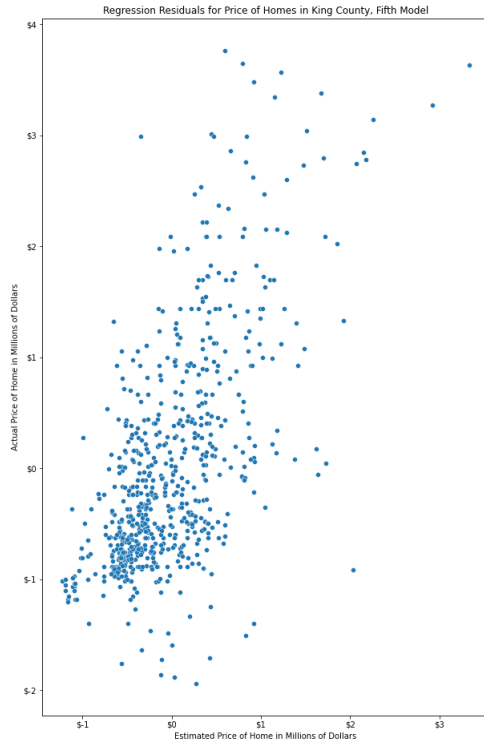
The third model adds onto our most correlated feature by modeling the top three variables. Here we see that our R-squared score is very high at about 0.93.

Fourth Model



Our fourth model includes all of the variables, but we notice that condition number greatly increases even though the R-squared score remains around the same values as our third model. We believe this is due to the multicollinearity introduced by our most correlated value, and as such we want to rerun our model without this variable.

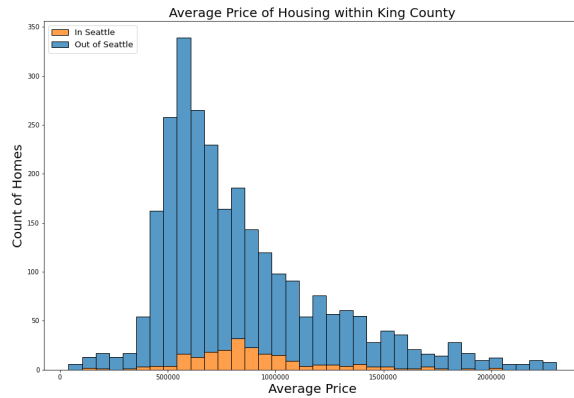
Fifth & Final Model



We arrive at our fifth and final model that looks at the relationship between liveable square feet and building grade. Although our R-squared score has decreased, these two variables have reduced the condition number considerably.

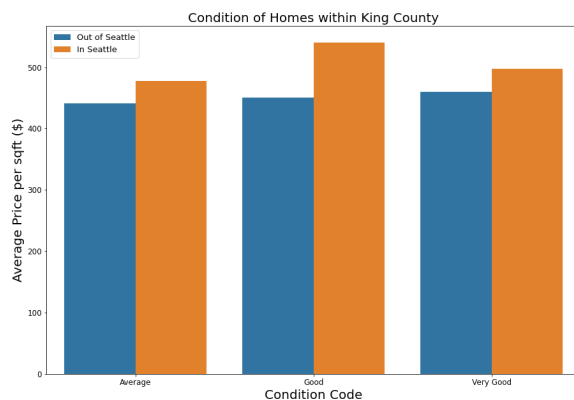
Data Visuals Creation

Average Price of Housing in King County



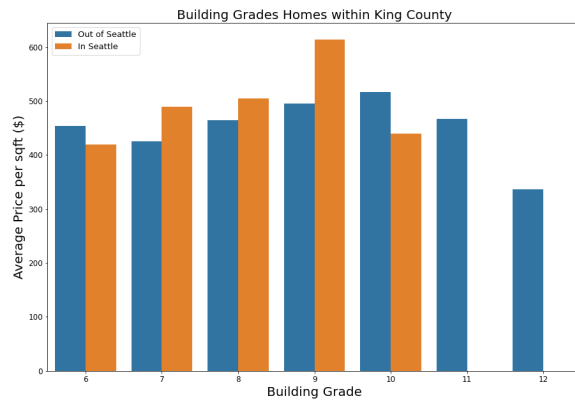
Looking at the average pricing of homes within King County, we see that significantly more families are purchasing homes outside of Seattle compared to purchases within the city. The average price of realty within Seattle is greater than that outside of the city as well which follows the trend of urban realty being more expensive than that of suburb communities.

Condition of Homes Compared to Average Price/SQFT within King County



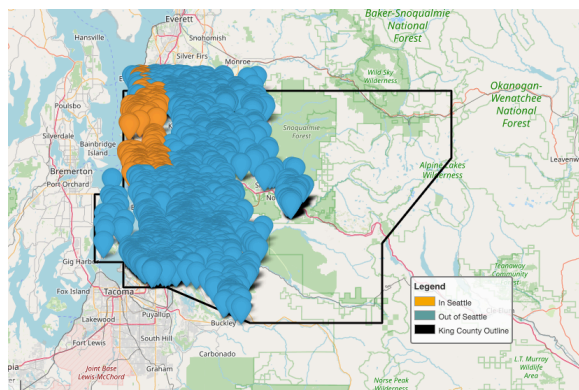
This model focused on homes that were of Average or greater condition to ensure that the families, especially the children are living in suitable conditions and not in homes of disrepair. Homes outside of Seattle have seen a steady incline in price as the conditions increase, whereas those within the city of Seattle have a greater variation in price.

Building Grade of Homes Compared to Average Price/SQFT within King County



This model focused on homes that had a building grade of 6 and above since the grade 6 is the lowest achievable grade that meetings King County building codes. There is an upward trend of average price per square foot as the building grade increases. However, there are only homes outside of Seattle that are of grade 11 and 12 which can be attributed to real estate is cheaper out of the big city or families customizings their homes more in suburban areas whereaas urban areas have more red tape and limitations. In addition, homes of higher grade begin to approach mansion level, and with more property, the bigger the house can be, which may not be possible within the city since there is less space.

Map of Seattle with Markers of Homes within the Dataset



The dataset consisted of 2,731 homes with the numerous fixed

factors and price outliers removed. The homes within the city of Seattle is noted in orange markers and the homes outside of Seattle are noted in blue. The outline of King County can be seen noted in a black line

Conclusion

Our model accounted for approximately 37% of the variability seen in the price of the home after focusing on two main factors, square foot of liveable space and the building grade. As the amount of space in the home and building grade both increase, the average price of the home is expected to increase as well. To be more precise, with one square foot increase in liveable space, you should expect to see an increase in the price of the home on average by about two hundred dollars. With one grade increase in the building grade, you should expect to see an increase in the price of the home on average by about sixty thousand dollars.

Recommendations

- **Plant your roots in a town outside of Seattle**
 - On average the prices of

homes outside of Seattle are lower compared to those within the city of Seattle, which leaves your bank account with more money. Significantly more families purchase homes outside of Seattle so you can surround yourself with other growing families in a family friendly neighborhood as well.

- **Opt for a home with more liveable square feet**
 - Although adding more square footage means more dollar signs on the price tag of the home, a home with a larger space would be advantageous to a growing family. A larger home allows each family member to maintain their own space within a single household.
- **Focus on homes with higher condition code or building grade**
 - As both the condition code and building grade increase, the value of a home increases as well. The higher quality and design of the home, the stronger it

will be to withstand the forces of nature and your children. A quality home has been maintained properly which means less repairs for you and decreased chances of injuries due to the deferred maintenance or home deterioration.

From the start, our focal point was the price of a home. As such, we believe these results would benefit both real estate agencies and families should they be taken into consideration.

Future Insights & Next Steps

- Consider changes in the Housing Market.
- Explore how the changes in demographics impact real-estate trends and the local area.
- Consider changes in interest rates and how this impacts purchaser's ability to obtain a mortgage and real estate demand.