



Learn Git and GitHub without any code!

Using the Hello World guide, you'll start a branch, write comments, and open a pull request.

[Read the guide](#)

[JosephDenney / KingCountyRealEstate](#)

0 stars 0 forks

Star

Unwatch ▾

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

master ▾

...



JosephDenney Update ReadMe.md ...

3 minutes ago

26

[View code](#)

ReadMe.md



King County Housing Analysis



[Slideshow PDF]

[Video Walkthrough](#)

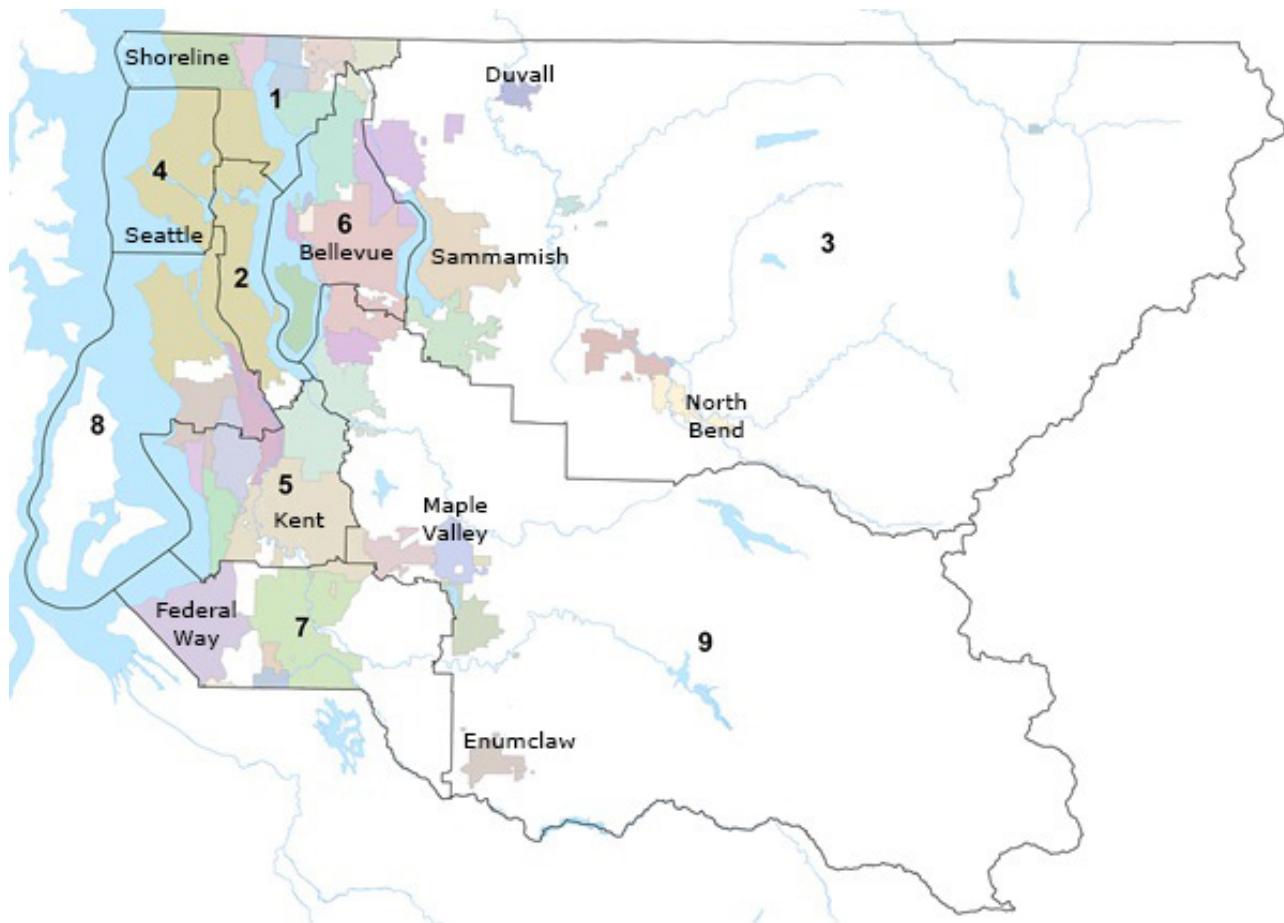
Outline

- Purpose
- King County Map
- Description of Data
- Main Questions
- Summary Conclusions and Insights
- Future Work

Purpose

Create a general understanding of important housing features in King County. Our client is a realtor in the county and has tasked us in helping to explain the most important, or value driving features of homes within the county.

King County General Map



Description of Data

| Variable | Description |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Id | Unique ID for each home sold |
| Date | Date of the home sale |
| Price | Price of each home sold |
| Bedrooms | Number of bedrooms |
| Bathrooms | Number of bathrooms, where .5 accounts for a room with a toilet but no shower |
| Sqft_living | Square footage of the apartments interior living space |
| Sqft_lot | Square footage of the land space |
| Floors | Number of floors |
| Waterfront | A dummy variable for whether the apartment was overlooking the waterfront or not |
| View | An index from 0 to 4 of how good the view of the property was |
| Condition | An index from 1 to 5 on the condition of the apartment, |
| Grade | An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design |
| Sqft_above | The square footage of the interior housing space that is above ground level |
| Sqft_basement | The square footage of the interior housing space that is below ground level |
| Yr_built | The year the house was initially built |
| Yr_renovated | The year of the house's last renovation |
| Zipcode | What zipcode area the house is in |
| Lat | Latitude |
| Long | Longitude |
| Sqft_living15 | The square footage of interior housing living space for the nearest 15 neighbors |
| Sqft_lot15 | The square footage of the land lots of the nearest 15 neighbors |

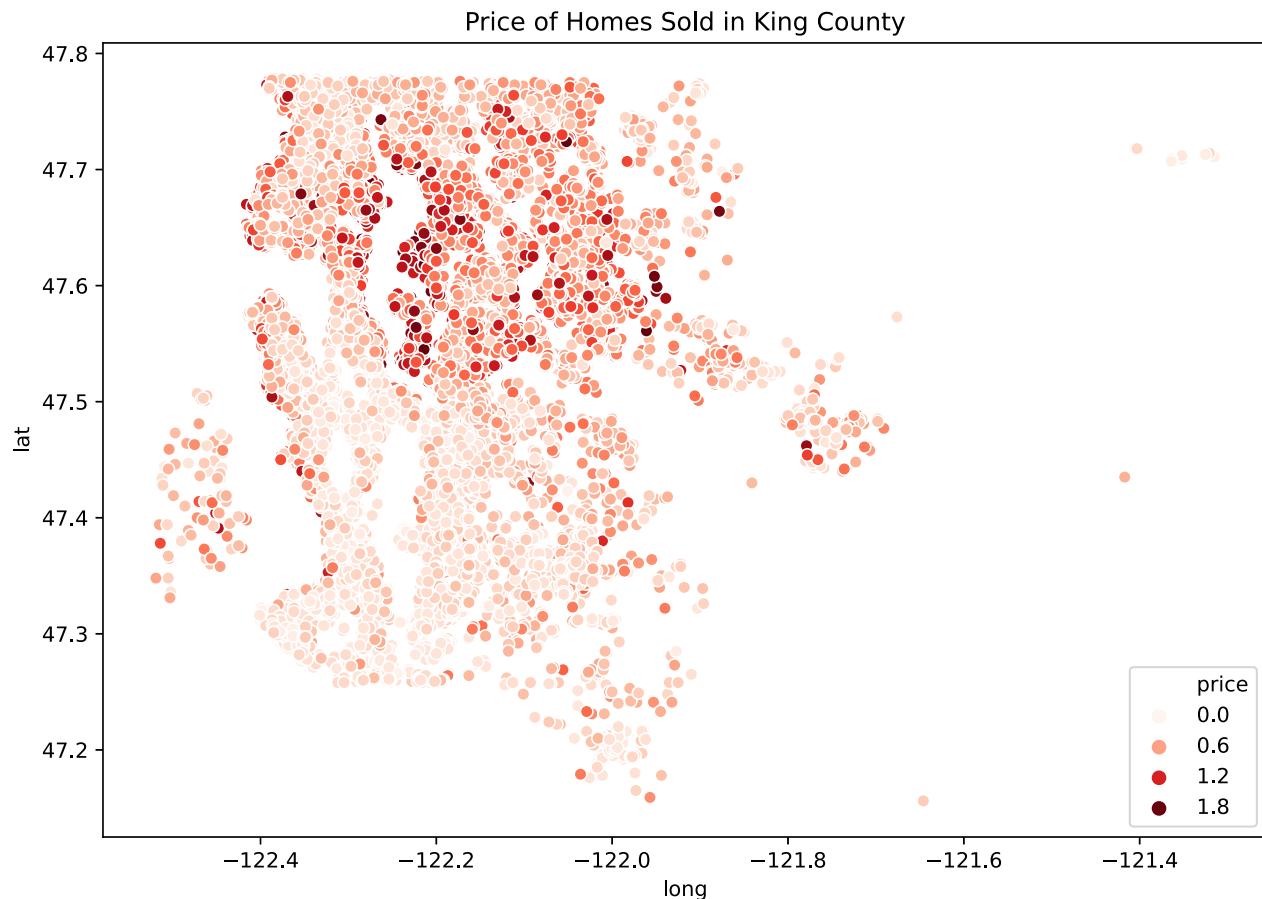
Main Questions

- Question 1: How does longitude and latitude affect home price in King County?
- Question 2: Is temporary condition or overall grade (build quality) affecting home price more significantly?
- Question 3: What parts of King County command the highest values per square foot?
- Question 4: Our realtor client is most interested in Mercer Island. The client would like to know what features of homes on Mercer Island are most desirable. Let's investigate this specific portion of King County to provide them with an actionable recommendation.

Summary Conclusions and Insights

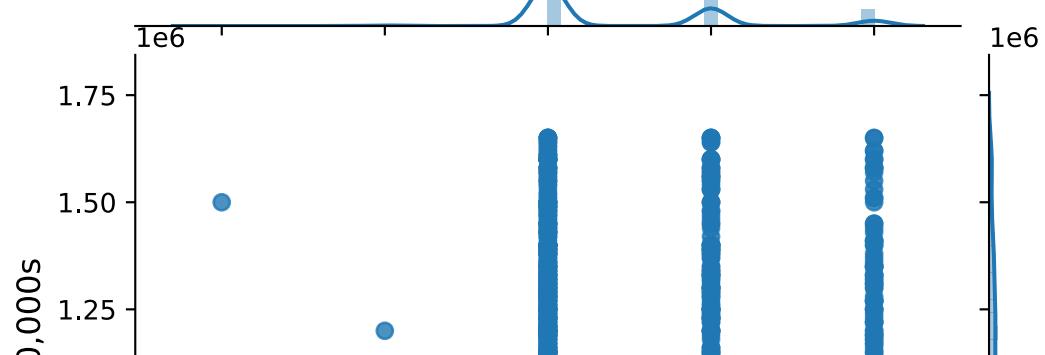
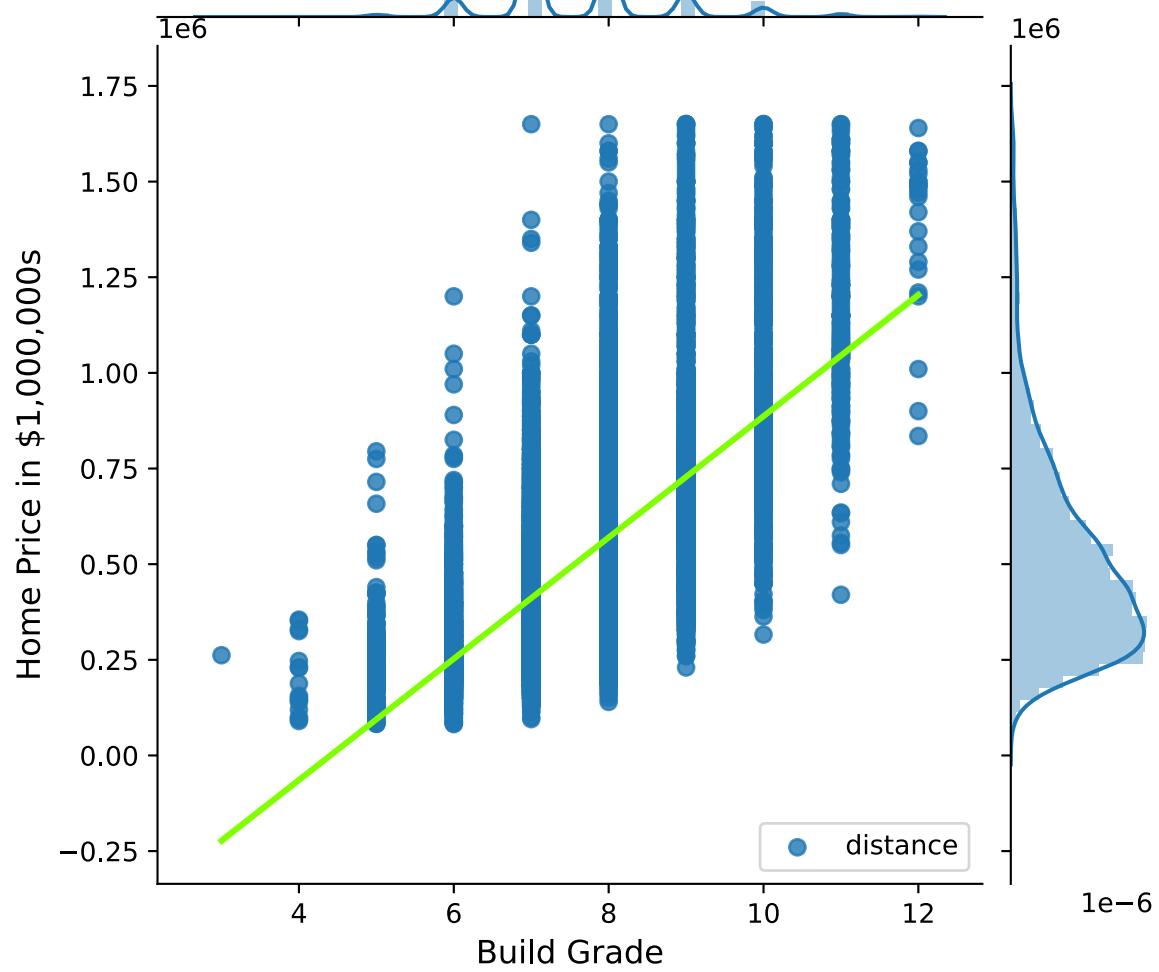
- Question 1 Answer and Further Investigation: When latitude and longitude are included in the regression, their VIF threshold is surpassed by a large margin. A more measured approach might be to look at the data after breaking down the dataset into smaller samples in their specific areas and creating neighborhood specific regressions. This would allow you to truly understand which specific features are affecting home prices, and therefore which to highlight. Location is a huge nonlinear factor on the price of the home.

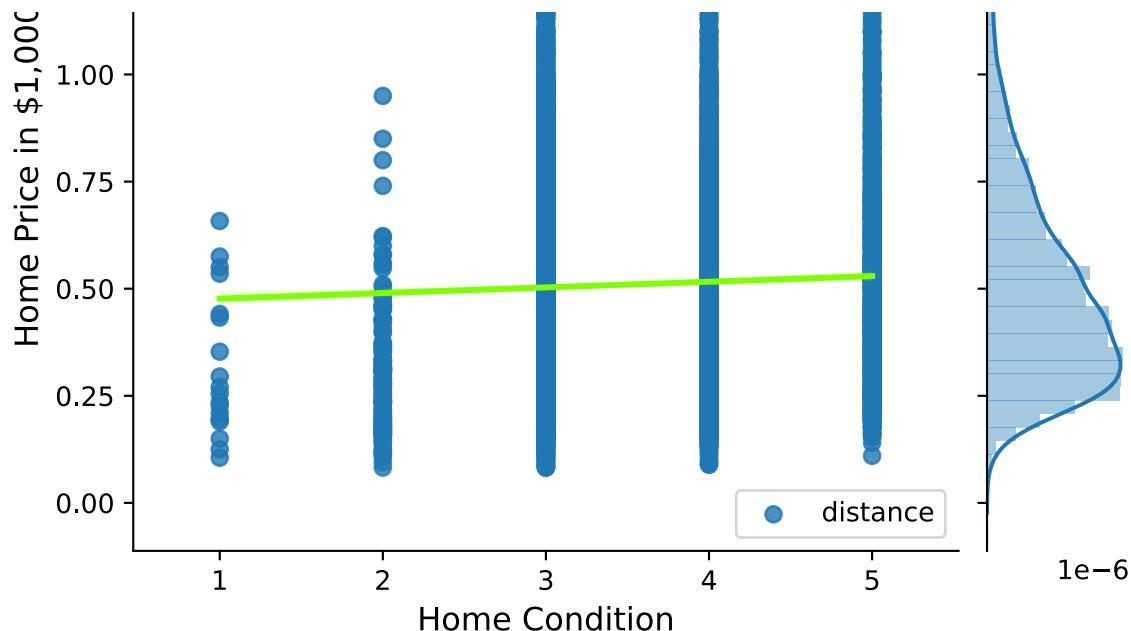
Distance from Seattle, however, does introduce a better feature to the model and is somewhat helpful in the prediction of price. The graph shows that our model is missing some information and is anticipating a less peaky distribution whereas the actual home sales are skewed towards the lower price range. Additionally, the distribution of home prices leans to and peaks to the left of the mean home price, indicating positive skewness in the actual home price sales. This is also why our predictive model also peaks just left of the actual price mean.



- Question 2 Answer and Further Investigation: While neither feature was used in our final regressive predictive model, it is clear that original build grade is strongly positively correlated with home prices. Condition may not affect sale price much. This

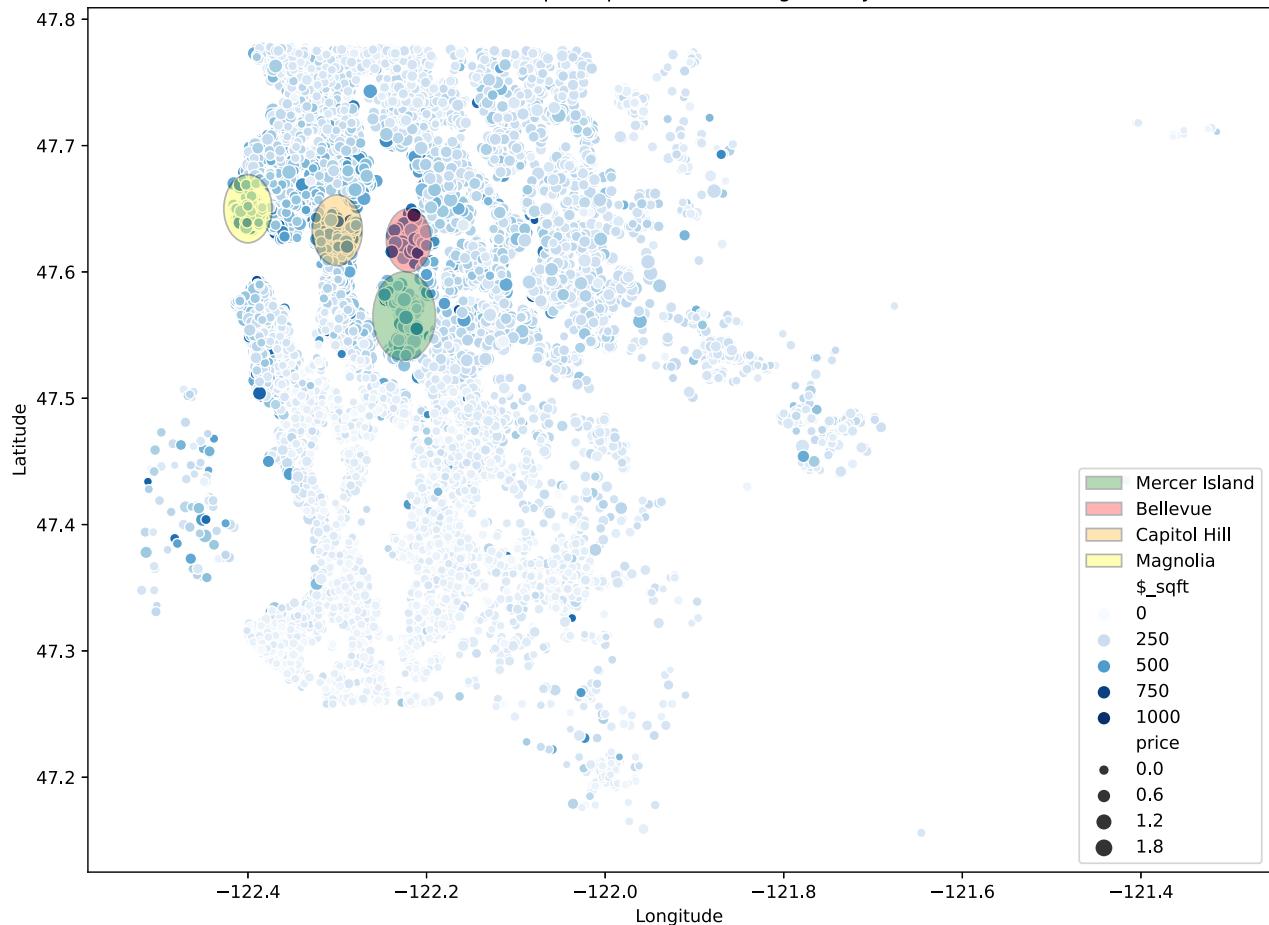
is likely because condition is temporary and can be remedied fairly easily, whereas to change a home's build grade requires substantial renovation from the literal ground up. To further investigate, suggest breaking the data down by geographical location again. The analysis of the entire county is not apples to apples. Real estate in downtown Seattle with a view of the city is not the same as real estate 15 miles south of Seattle.



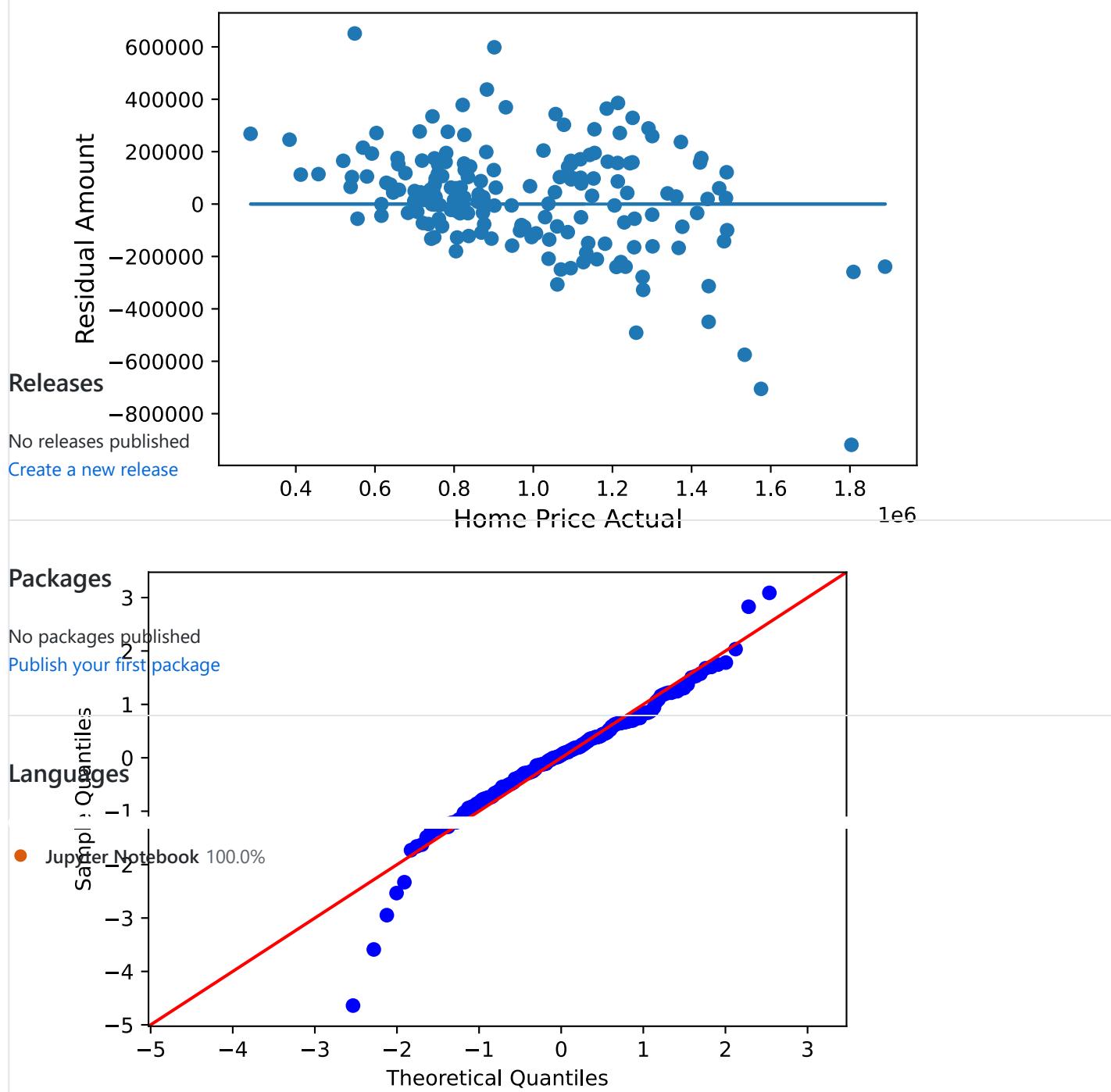


- Question 3 Answer and Further Investigation: Bellevue, Capitol Hill, Magnolia, and Mercer Island have homes that are more expensive per square foot. Mercer Island has large homes and you can get more for your dollar compared with Bellevue. Generally speaking, properties south of Seattle in King County are not as valuable, but we can't pinpoint or recommend why that may be other than location. Isolating these specific locations and neighborhoods in separate regressions analyses will likely be an efficient way to build a better predictive model, the limitation being that the model can't be applied to different neighborhoods.

Price per Square Foot in King County



- Question 4 Answer and Further Investigation: Isolating the home sales on Mercer Island allowed us to isolate a smaller market to understand which features best predict the target. Let's run a quick check model formula for the Mercer Island data. We recommend that the realtor highlight interior square footage, size of lot, and overall proximity to the city. The most important takeaway from our investigations is location, location, location. Can further isolate neighborhoods in further investigation and regressions, but many features that we started with are either trivial or autocorrelated and can't be used in regression.



Future Work

- Analysis of the housing market must be neighborhood specific. Individuals with different income are going to be able to afford what they can afford. Similarly, not every person will put the same importance on specific features that another would. Breaking datasets into neighborhoods is crucial to building a model.
- Isolate which neighborhoods that are most important to the client and create neighborhood specific models.
- Create functions to quickly optimize the model

Thank you!