
Linear Regression of King County Real Estate Data

— By: Blake and Alex —

Purpose

- Use Case Scenario:
 - You are on Zillow looking to buy a house and see a house listed for sale that you like. You don't have any real estate experience and are not familiar with King county at all. How do you know if the listed sales price is a good or bad price?
- Goals:
 - Foolproof UX
 - Simple, but accurate price prediction
 - Use information that is very easily found and verifiable

Thesis of our model

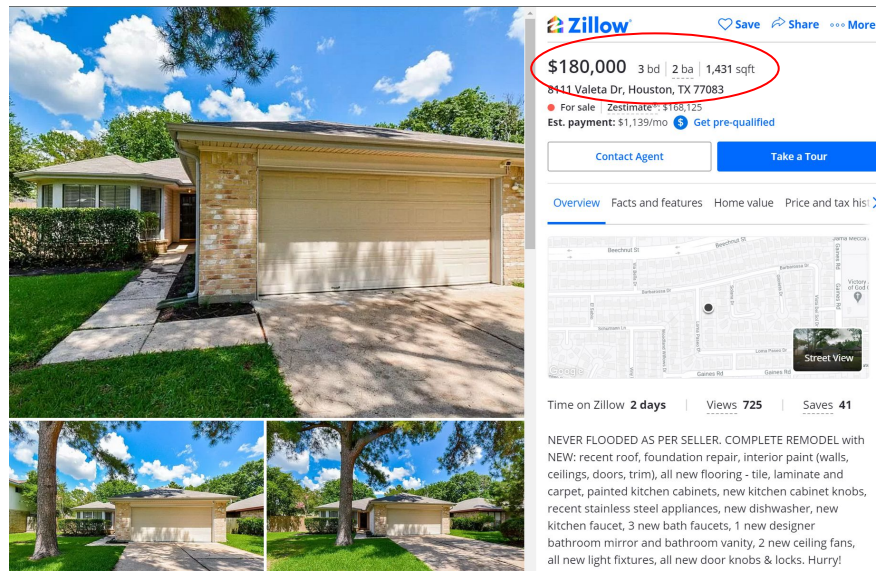
- G.I.G.O.
 - Garbage in = Garbage Out
- Real estate data is often incorrect and/or not vetted
 - For anyone to use this model the data needs to be publicly available
- Most important factors in real estate pricing (Inputs)?
 - Asset type
 - Location
 - How big is the asset

Model Inputs

- Asset Type:
 - Single Family Residences Only
- Only For-Sale Single Family Residences (SFR)
 - Townhomes and Detached homes
 - Our model can be modified to fit other asset models too in future

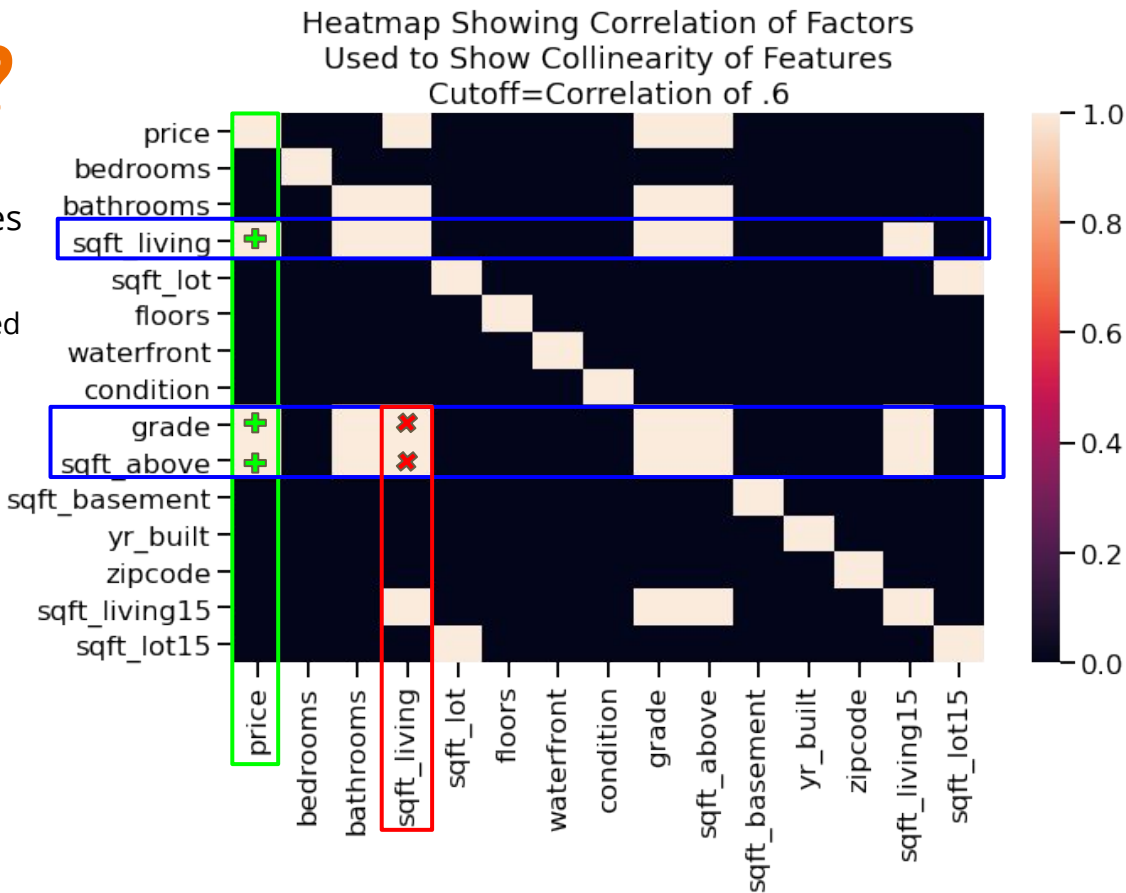
Model Inputs - Location & Size

- Location
 - Zip code
 - Easily verifiable
 - Easily found on internet
 - Goldilocks zone?
 - Not too much data
 - Not too little data
- Size
 - Living SF
 - Easily found on internet
 - Easily verifiable

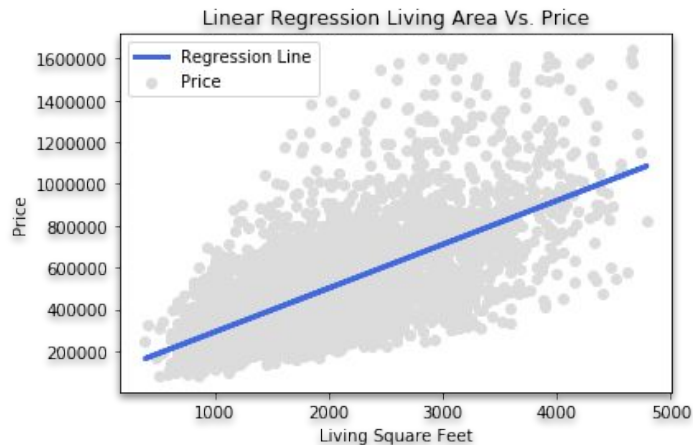


What affects what?

- We want to focus on the white boxes inside the Green Vertical boxes
 - This means these factors are related to Price
- Red Vertical Box
 - Shows sqft_living also related to grade and sqft_above
 - This means these factors are redundant with sqft_living and present no new information
- From this matrix we can determine that our **main focus should be square feet of living area**



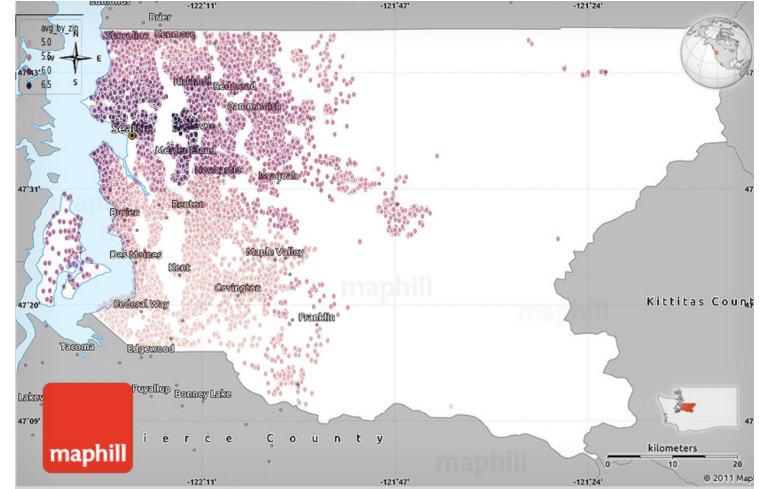
Living area in square feet vs price



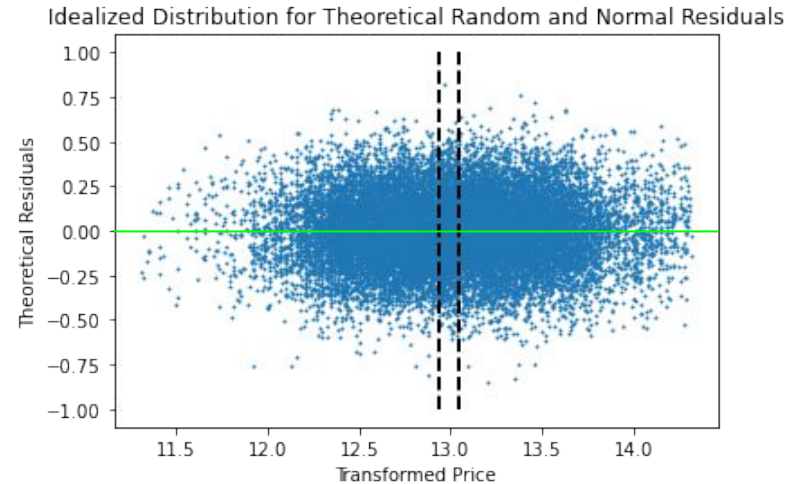
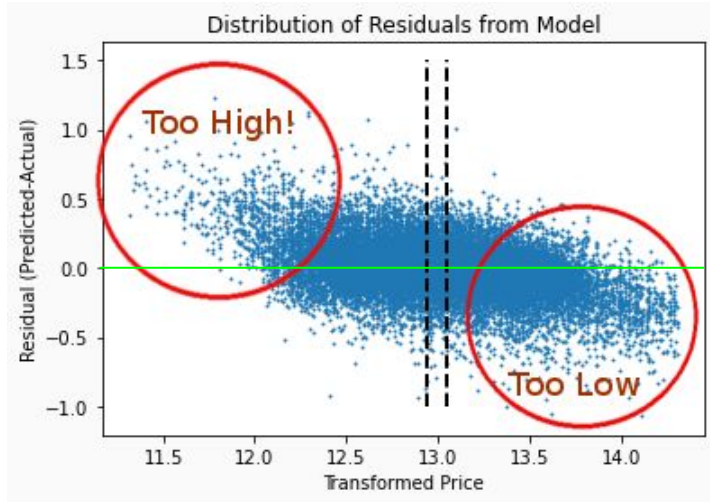
- Living area single strongest predictor for price
- Can see that this accomplishes our goal of helping user not under price or overspend!
 - We can now build on this to make it more accurate

Modeling Zipcode

- Price per square foot living area, darker is greater price
 - This is our modeled data
- We used zip code to create clusters to help predict price
 - Eliminates noise
 - Helps in areas of low data
- Zip code is largely irrefutable and easy data to find



How well does the model work?



- Image on right shows what it should look like in ideal circumstances
 - Majority of points along the 0 value y-axis
- **Shape** of the plot is the relevant information
- Shows that model over-predicts at low prices and under-predicts at high prices

Model Limitations and Characteristics

- Model training data is old
 - Real Estate prices fluctuate over time
 - Training data is from 2014-2015 so not valid today
- Model is trained only for King county
 - Location is very important in real estate
- Model can easily be updated to fit your needs thanks to simple inputs!
 - Can update via extracted data from Zillow, Realtor.com, etc...

Interactive Demo

[Link](#)

Easy as 1,2,3!

1. Run

```
In [ ]: import predict_house_price_interactive
        %run predict_house_price_interactive
```

2. Enter Zip Code

```
In [*]: import predict_house_price_interactive
        %run predict_house_price_interactive
```

What is your zipcode? 98012

What is the total living area of the house?

3. Enter living area in square feet

```
In [1]: import predict_house_price_interactive
        %run predict_house_price_interactive
```

What is your zipcode? 98012

What is the total living area of the house? 2000

Sorry, couldn't find that zip code. Using only living area

Predicted value of house based on living area 2000.0 is 282954.88

Future Work

- Further develop interactive demo to take more factors into account, and be able to produce a range of potential values
- Make a more complex model, as a simple linear model does not hold up well at the high and low ends of the data
- Get data for other counties, and see if the model can be adapted for them
- Get more recent data, to see if the model still holds up

Thank You!

1. Git Hub Project Links:

- a. Alex:
 - i. <https://github.com/anbillinger/dsc-mod-2-project-v2-1-onl01-dtsc-ft-041320>
- b. Blake
 - i. <https://github.com/bsamaha/dsc-mod-2-project-v2-1-onl01-dtsc-ft-041320>

2. Follow our Blogs:

- a. Alex:
 - i. <https://anbillinger.github.io/>
- b. Blake
 - i. <https://medium.com/@blake.samaha16>