Advising Families on Real Estate in King County

JOEL MOTT
FLATIRON SCHOOL

Overview:

- What do families with children consider when purchasing a home?
- How can linear regression connect families with a new home?
- How will it help advise sellers?
- Brief walkthrough of modelling process & results
- Recommendations for buyers and sellers

What do families with school-age children (or younger) consider when purchasing a home?

55% of King County's population are at/below school age or are the age of a parent with such children at home.

The most frequent considerations of families include:

- good schools
- "room to grow", aka square footage
- number of bedrooms
- safe neighborhood

Additional factors may include how families tend to move during the summer.

How will linear regression offer specific insights into connecting families with a new home?

While bar graphs and other simple charts show helpful insight as far as averages and distributions, linear regression can go further.

It uses things like averages and distributions as a springboard into further details, such as:

- forming a line of “best fit” in order to get a better general picture of the data
- showing how far the data is “spread out” from that line
- calculating how well that line accounts for the data
- it can also account for *multiple* factors that influence price

It not only looks behind at existing data, but also ahead by making predictions about future prices

How will this also help advise all sellers?

Calculating how much a seller should consider raising their price if they live in a good school district.

Specific insights into exactly how much square footage and the number of bedrooms can impact listing price.

Some seasonal aspects offered here can help inform sellers.

overview of the linear regression process for this project:

Starts with **understanding the dataset**: what does each data column mean? Which columns are most relevant to your clients' concerns?

Making a **baseline model**: which single factor correlates most with listing price? This will be a benchmark as we refine the linear regression model to your specific needs.

Tailoring the model to families looking for homes: add new columns to the model and see whether it becomes more insightful all while checking for any oversights or pitfalls that come with adding more considerations.

Finalizing the model: balancing between model bias and variance so we can **maximize our ability to make accurate predictions that are also meaningful**.

How did the regression modelling start to work for you?

The **baseline model** showed the correlation between listing price and square footage because that was the highest correlation factor.

It featured a "line of best fit" that was able to explain 49.2% of the variance of how square footage and price relate.

However, the variance of square footage across homes in King County isn't normally distributed. In other words, home square footage across the county doesn't resemble a bell curve, which hinders a regression model's ability to accurately predict.

Subsequently, I looked for ways to (1) consider more factors and (2) make more accurate predictions.

refining your model

Homes were matched to school districts via the King County's tax assessor's website, then cross-referenced the zip codes they encompass from those districts' websites or zipdatamaps.com

I then obtained school district "scores" from part of Niche.com's site , which gives districts "grades" from D-/D/D+ all the way up to A-/A/A+.

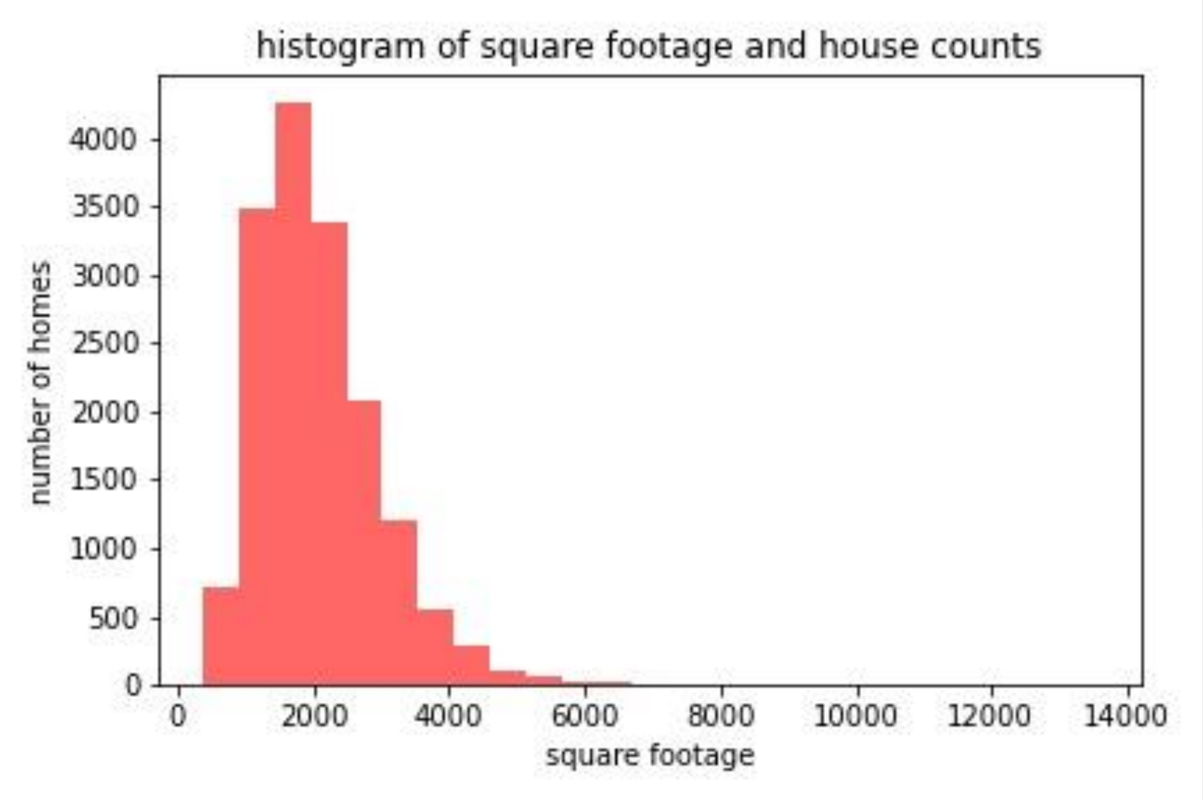
I combined these scores with the preexisting dataset and then added them to a new model that started with just these scores alone and listing prices.

On their own, they didn't have a powerful correlation with price, so it was time to add further factors.

building on school district scores

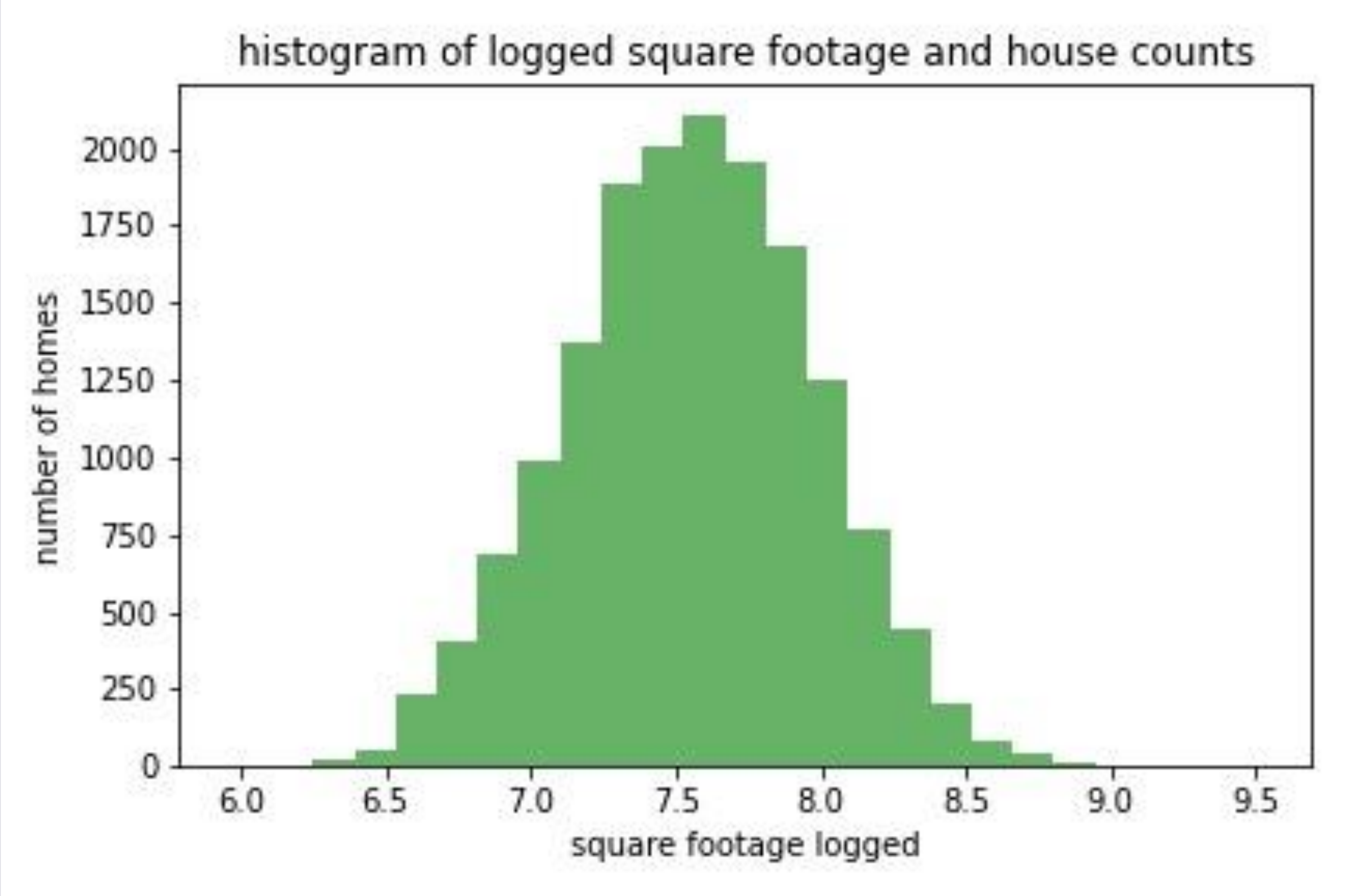
- Since families usually want “room to grow,” I also added square footage to this model. I improved its ability to make predictions by addressing its distribution issue. This image shows how square footage looks in a histogram at first.

- Most homes have between 500 and 4,000 square feet, but enough homes have more (or even much more), so they throw off the distribution.



improving accuracy

- I applied a “logarithmic transformation” to the square footage distribution: it retains all the data but reformats how it’s distributed.
- This resulted in a distribution that looks more like a bell curve.
- Subsequently, square footage was able to make a more accurate prediction of how it correlates with listing price.



further improvements

The model was improving, but it wasn't as good at explaining listing price as the baseline model was.

Subsequently, I also added the number of bedrooms to the factors it considered, which resulted in another small improvement.

Finally, like square footage, the listing prices themselves were not normally distributed, making the model's job harder.

I applied logarithmic transformation to price as well, which **helped the model do an even better job at explaining its data than the baseline could.**

final model results:

price factors	percentage impact on listing price
district score B	-0.048322
district score B+	0.180937
district score A-	0.558151
district score A	0.419581
district score A+	0.604715
# of bedrooms	-0.066777
square footage	0.869222
summer purchase	-0.010940

when compared to a school district score of B-

result explanations

The second column shows the percentage by which the listing price should change based on a one-percent change in the listed factor.

For example, a 1% increase in square footage correlates with a 0.86% increase in price, hence the corresponding “0.869222” value.

All school district scores show a percentage change in comparison to a homes with a school district score of “B-”.

This unseen B- “reference category” helps eliminate potential errors in the district scoring.

result observations

Square footage is still the best predictor of listing price.

District scores from "B+" to "A+" show significant increases, with homes in "A+" districts showing the second-highest correlation of listing price among all these factors.

The number of bedrooms also has a small impact on price, even an overall negative one. This may correspond with residences that have fewer rooms but cost more since they're located in Seattle.

It turns out that purchasing a home during the summer months has almost no impact on price, so family-oriented buyers don't necessarily need to raise their bidding price to secure a home while kids are on summer break.

recommendations & next steps

Any seller in an A-, A, or A+ school district may consider raising their listing price accordingly.

Family-oriented buyers may be advised on what to expect in terms of listing price for more higher-scoring districts.

Neither sellers nor family-oriented buyers would likely need to consider common concerns such as (1) bedrooms or (2) buying a home during the summer months into listing price considerations.

While square footage's dominance as a price factor may be unsurprising, this model provides specific, predictive percentage correlations for clients to consider.

thank you

Joel Mott

Flatiron School

joel.mott8@gmail.com