Pricing Homes in King County

March 2023

Business Problem and Overview

Property values often fluctuate depending on certain features, which can make pricing a home difficult. The goal is to derive **a model that helps the general public determine how a certain property should sell compared to the average home**. The model should act as a **guide for property sellers in King County** so they can pinpoint attributes of a home to emphasize or amend when justifying and setting prices.

The Data

- To create a regression model, public data on homes sold is obtained from King County's government website

- The raw data contains 24 columns, but for the purpose of our regression we will focus on the variables listed to the right

"sqft_living": Square footage of living space in the home

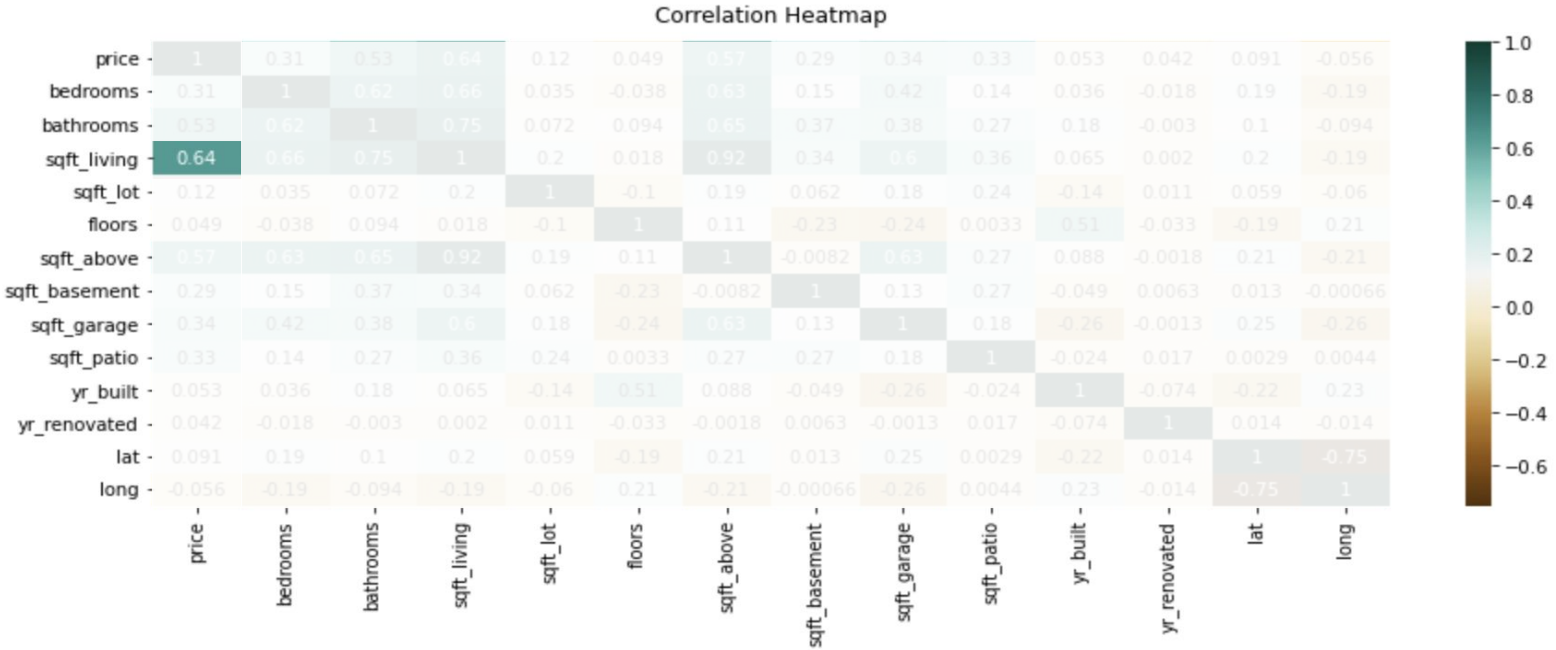
"bathrooms": Number of bathrooms

"grade": Overall grade of the house (construction/design)

"waterfront": Whether the house is on a waterfront

"view": Quality of view from the house

Results: Initial Analysis and Baseline Model



A seaborn heatmap shows that “**sqft_living**” is the **most correlated variable with price**

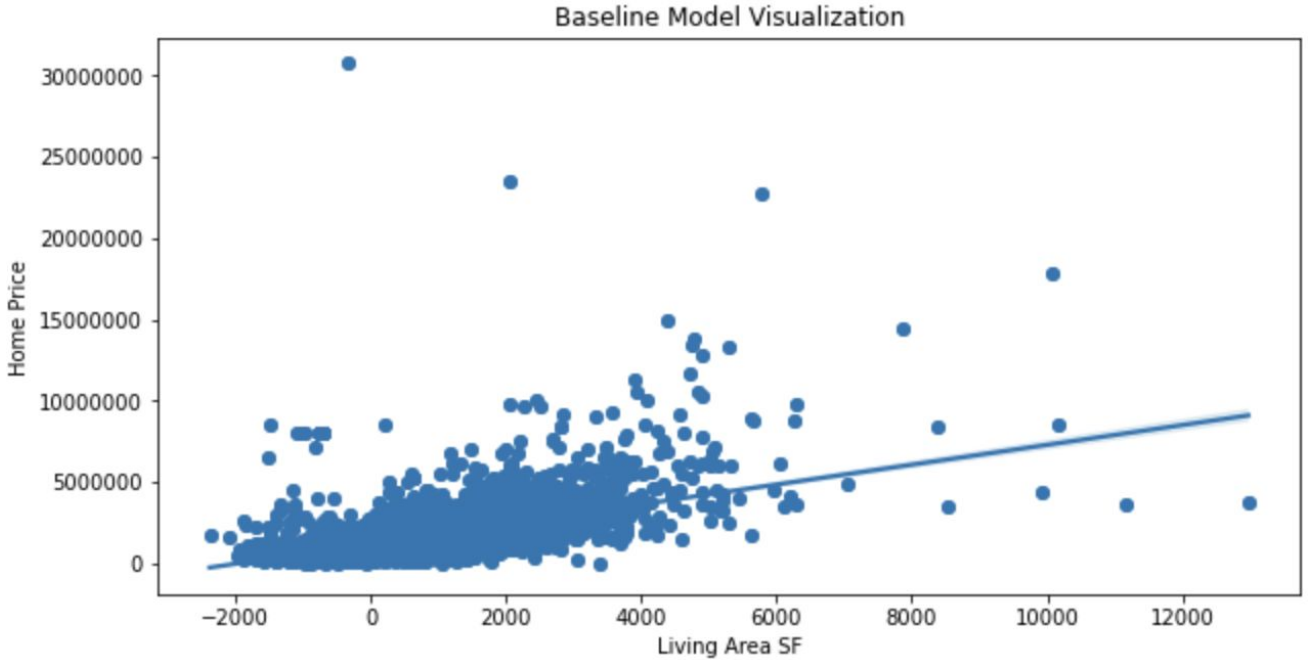
OLS Regression Results

Dep. Variable:	price	R-squared:	0.405
Model:	OLS	Adj. R-squared:	0.405
Method:	Least Squares	F-statistic:	9616.
Date:	Mon, 06 Mar 2023	Prob (F-statistic):	0.00
Time:	21:18:15	Log-Likelihood:	-2.1177e+05
No. Observations:	14126	AIC:	4.235e+05
Df Residuals:	14124	BIC:	4.236e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.223e+06	6600.198	185.237	0.000	1.21e+06	1.24e+06
sqft_living	609.2048	6.213	98.059	0.000	597.027	621.382

Omnibus:	21552.244	Durbin-Watson:	1.806
Prob(Omnibus):	0.000	Jarque-Bera (JB):	31151969.336
Skew:	9.101	Prob(JB):	0.00
Kurtosis:	232.338	Cond. No.	1.06e+03

The baseline model is statistically significant and accounts for **40.5% of price variance**. With a **unit increase in living area**, we can expect price to **rise by $609** and a home of **average area to sell for ~$1.22M**



The baseline model is statistically significant and accounts for **40.5% of price variance**. With a **unit increase in living area**, we can expect price to rise by **$609**.

Results: Overview of Processes

Model 2

- Adding and standardizing the discrete variable (“bathrooms”) to create a multiple regression model

Model 3

- Re-formatting and standardizing “grade” data to transform it from a text into numbers the model can recognize

Model 4

- Incorporating categorical variables (“view” and “waterfront”) to assess the impact that amenities have on average properties

Results: Summary of Models Run

	Model	Independent Variables	R-squared
0	Baseline Model	sqft_living	0.408050
1	Second Model	sqft_living, bathrooms	0.412473
2	Third Model	sqft_living, bathrooms, grade	0.470337
3	Fourth Model	sqft_living, bathrooms, grade, waterfront, view	0.510179

Recall that “independent variables” refers to inputs and “R-squared” tells us the percentage of price variance that the model can account for. **As “R-squared” increases, our models can explain a greater proportion of the price data.**

Results: Final Regression Model

OLS Regression Results

Dep. Variable:	price	R-squared:	0.524
Model:	OLS	Adj. R-squared:	0.524
Method:	Least Squares	F-statistic:	1942.
Date:	Mon, 13 Mar 2023	Prob (F-statistic):	0.00
Time:	07:20:33	Log-Likelihood:	-2.1020e+05
No. Observations:	14126	AIC:	4.204e+05
Df Residuals:	14117	BIC:	4.205e+05
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.162e+06	2.29e+04	50.775	0.000	1.12e+06	1.21e+06
sqft_living	269.3194	10.008	26.910	0.000	249.702	288.937
bathrooms	1.301e+05	1.16e+04	11.255	0.000	1.07e+05	1.53e+05
grade	2.961e+05	7637.279	38.772	0.000	2.81e+05	3.11e+05
waterfront_YES	4.139e+05	7.02e+04	5.893	0.000	2.76e+05	5.52e+05
view_EXCELLENT	8.172e+05	5.94e+04	13.766	0.000	7.01e+05	9.34e+05
view_FAIR	3.001e+05	8.98e+04	3.342	0.001	1.24e+05	4.76e+05
view_NONE	-7.401e+04	2.35e+04	-3.156	0.002	-1.2e+05	-2.8e+04
sqft_living x waterfront_YES	629.9451	31.741	19.846	0.000	567.728	692.162

Omnibus:	21308.665	Durbin-Watson:	1.773
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33610538.388
Skew:	8.842	Prob(JB):	0.00
Kurtosis:	241.309	Cond. No.	1.63e+04

The final model experiments with the relationship between the inputs in the fourth model, and interpretations of the numbers highlighted above are outlined in the following slide

Results: Final Regression Model (Data Interpretation)

- The model accounts for **52.4% of the variance in sale price** and models against a reference home with:
  - Average living area
  - Average number of bathrooms
  - Average grade
  - Average views
  - No waterfront
- The model prices the **typical home** with the aforementioned features at **~$1.2 Million**
- We expect a unit increase in “sqft_living” to raise the value of an average home by $269
- We expect a unit increase in “grade” to raise the value of an average home by ~$300K
- Adjusting variables within a home with "nice-to-have" amenities, such as a waterfront, has a greater impact on value
  - A unit increase in "sqft_living" for an average-sized home with a waterfront adds ~$629 instead of $269

Takeaway and Recommendation

Homesellers interested in maximizing the value of their properties should aim to **include at least one extra "nice-to-have" amenity** (e.g. waterfront, excellent views), as additional features can greatly elevate values of the average property

Future Considerations

- Map geographic distribution of homes with “lat” and “long”
- Leverage other public data from King County’s website such as population and socioeconomic information

Email: kezia.setokusumo@stern.nyu.edu
Github: github.com/keziasetokusumo/p2_project
LinkedIn: www.linkedin.com/in/keziasetokusumo/