Phase 2 Project

Marissa Bush

Outline

Business Problem

Data

Methods

Findings

Conclusions

Business Case

The budding Zillow-type company, J.Hughes Inc., is buying and flipping houses in the Seattle area. They would like to know the types of housing features to prioritize to make the most profit. Knowing which features have the biggest impact on housing prices, J. Hughes Inc. could then make data-driven decisions on which houses to purchase to optimize making a profit on.

Data

Kings County Data Set - Features included:

- id
- date - house was sold
- price
- bedrooms
- bathrooms
- sqft_living
- sqft_lot
- floors
- waterfront
- condition - How good the condition is (overall)
- grade - overall grade given to the housing unit, based on King County grading system
- yr_built
- zip code
- lat
- long

OSEMN Method for Data Analysis

Obtain

Data collected from King County Housing.

Scrub/Explore

Clean data and feature selection

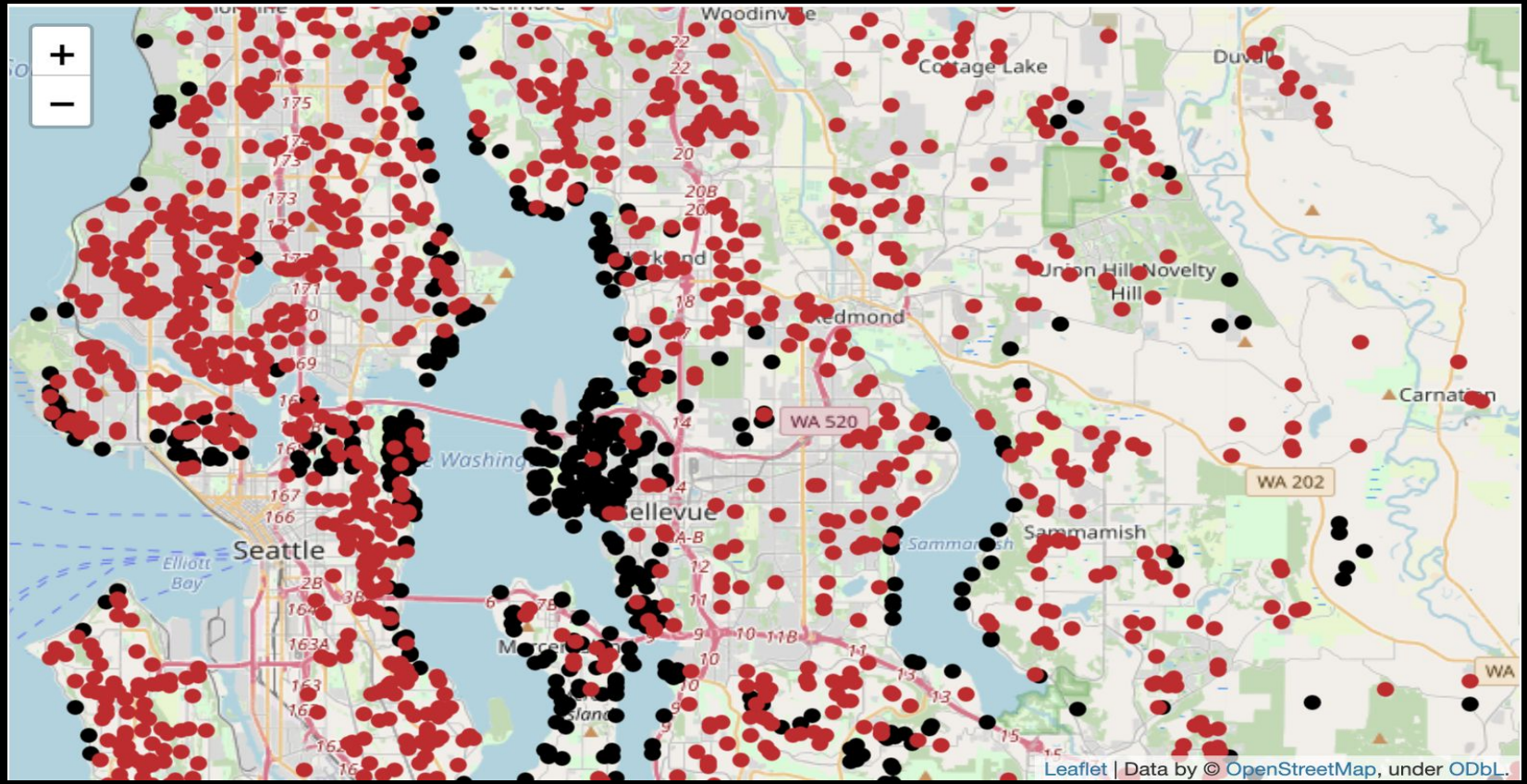
Model

Four Linear Regression models

Interpret

Getting recommendation from model

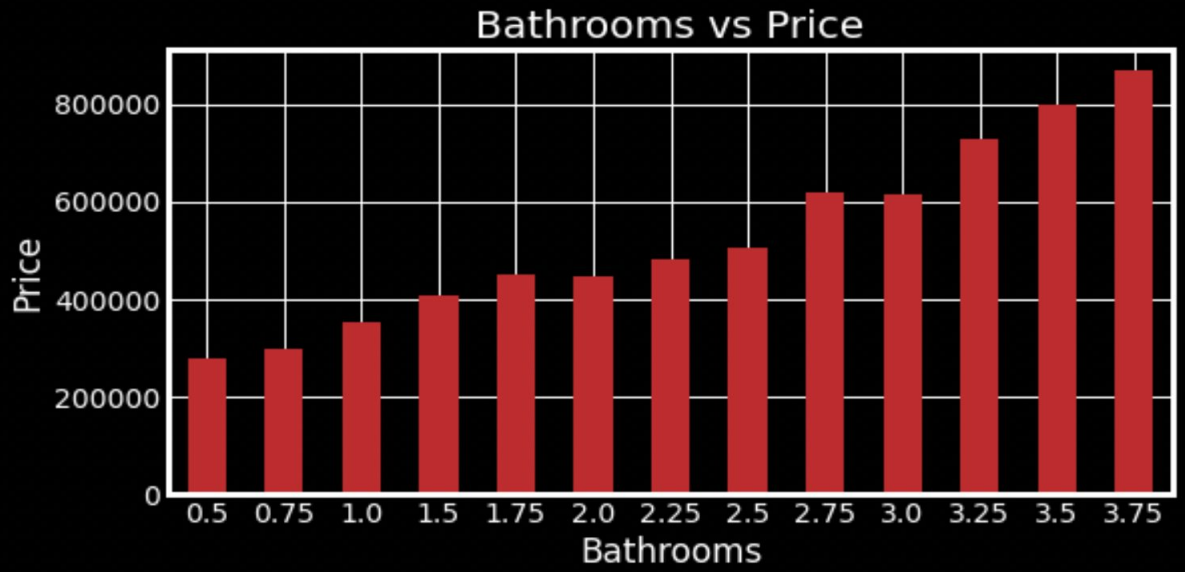
Red: Houses under 1.5 million
Black: Houses above 1.5 million.



Findings

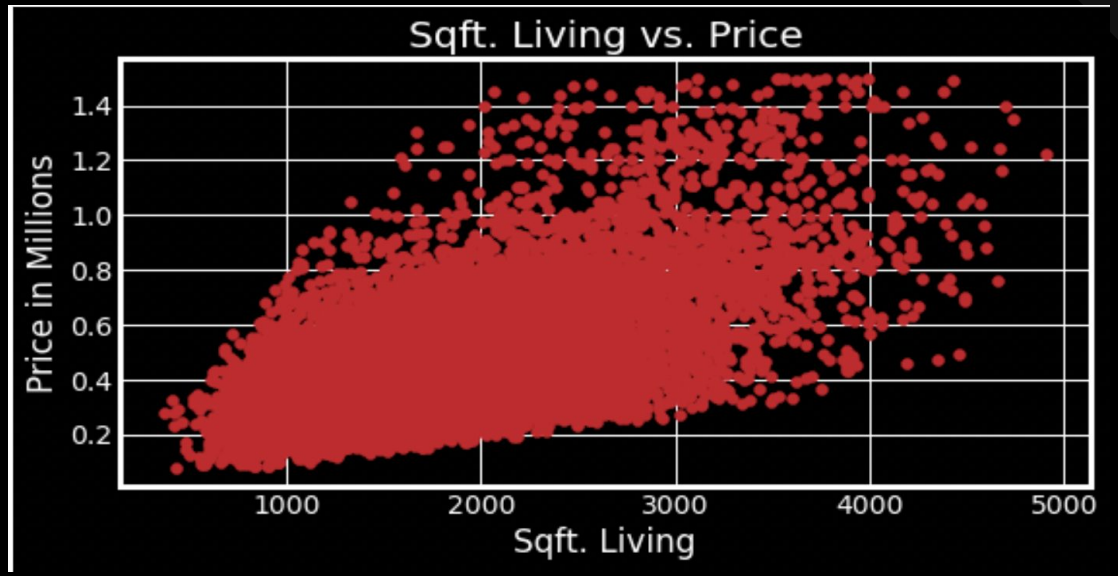
Bathrooms

For every increase in number of bathrooms, $20,940 would be added to the value.



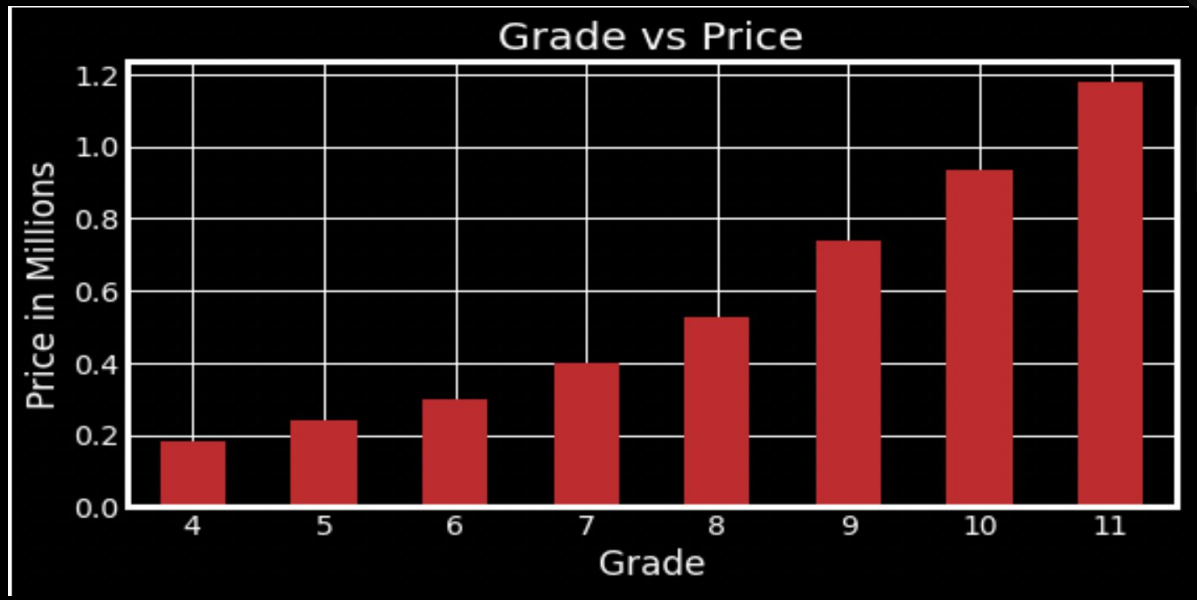
+ Sqft. Living

For every increase in unit of sqft. living, $137 will be added to the value of the home.



+ Grade

A higher “graded” home will add $ 112,200 to the value of the house.



Conclusions + Future Work

Which housing features should be prioritized in order to buy and flip homes for the highest profit?

The following three features/ aspects are recommended:

- **Adding a bathroom to a home increases the value by $20,940.** My recommendation would be to add more bathrooms.
- **My recommendation would be to add sqft.living to a house.** For each unit of additional sq. footage. would add $ 137 to the value.
- **My third recommendation would be to increase the grade of the house.** This will add $112, 200 to the value.

Future work:

- I would like to make additional models with the KC data for houses above 1.5 million.
- Also, possibly look at not just houses, but apartments and other types of housing as well.

Thank you!

Marissabush.02@gmail.com

Github: @Marissa841

Appendix

Model	R^2	RMSE	Number of Features	Best Model
Model 1	0.64	135,717	38	❌
Model 2	0.64	135,922	28	✅
Model 3	0.65	134,932	44	❌
Model 4	0.65	135,442	31	❌

Appendix

Features	Coefficients	
bathrooms	$20,940	
sqft_living	$137	
sqft_lot	-$3	
grade	$112,200	
age_at_sold	$2253	
bedrooms_3	-$29,170	
bedrooms_4	-$49,600	
bedrooms_5	-$60,100	
bedrooms_6	-$77,250	
floors_2_5	$66,090	
floors_3_0	$23,010	
waterfront_1_0	$387,500	🥰
condition_5	$41,240	
zip4_9801	-$35,510	
zip4_9802	-$64,420	
zip4_9803	-$34,220	
zip4_9804	-$51,480	
zip4_9805	-$33,280	
zip4_9806	-$24,870	
zip4_9809	-$180,100	😖
zip4_9810	$62,300	
zip4_9811	$80,720	
zip4_9812	$15,110	
zip4_9815	-$21,830	
zip4_9816	-$80,500	
zip4_9817	-$58,930	
zip4_9818	-$108,900	
zip4_9819	$34,850	