Statistical Inference of House Pricing in King County

By: Killian Kahalley, Chandler O'Neal, and Griffin Riner

Business Problem

Empire Home Remodeling has approached and asked us to provide recommendations and inferences on how to increase the overall returns on their renovations.

Summary

This analysis can be useful to provide Empire Home Remodeling with renovation proposals that would yield the most profitable returns.

In this project we focused on several features of the housing/remodeling market including:

- House Location by Latitude and Longitude
- Grade (quality rating)
- Square footage
- Number of rooms
- Waterfront
- Basement / no basement

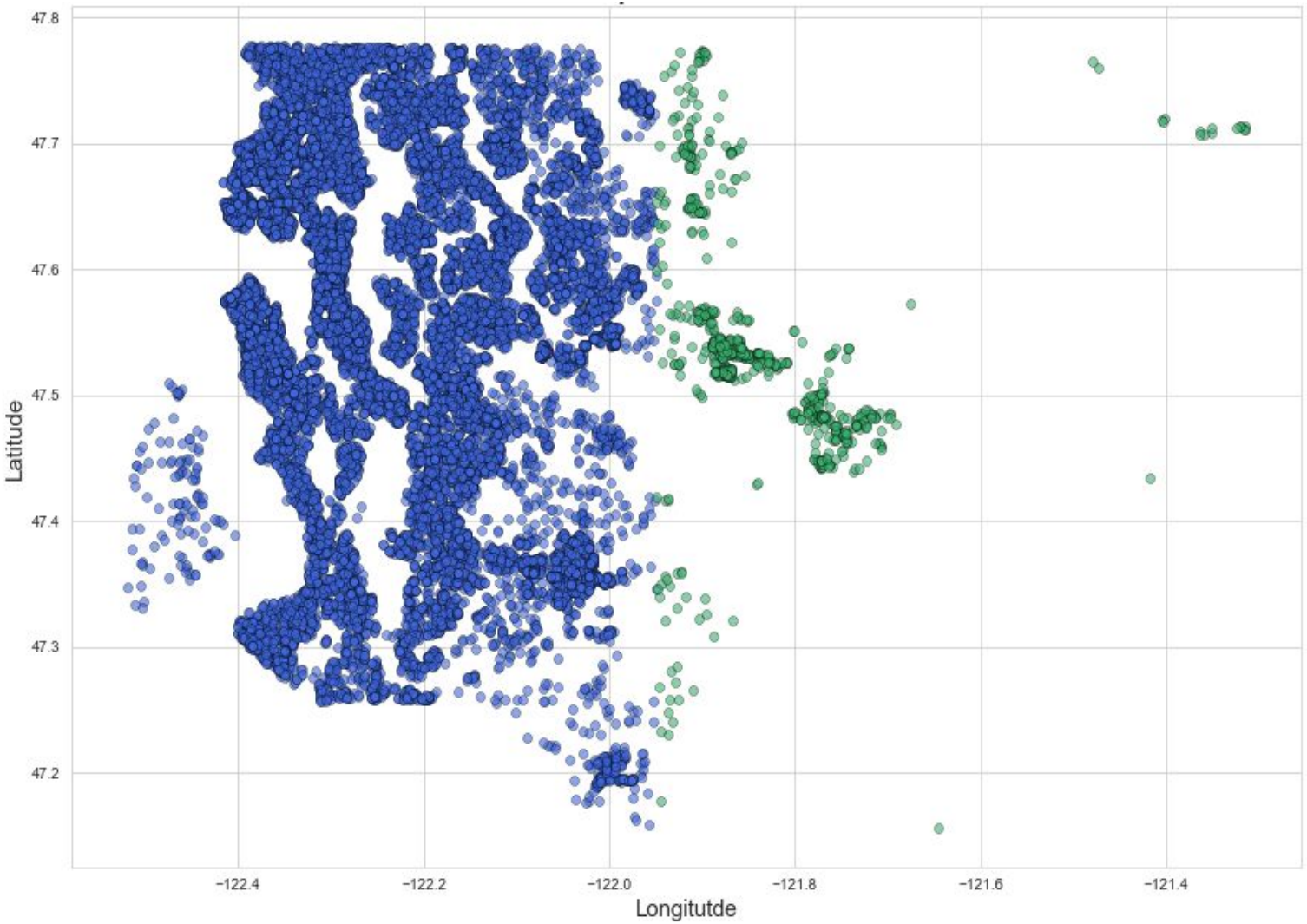
Data & Methods

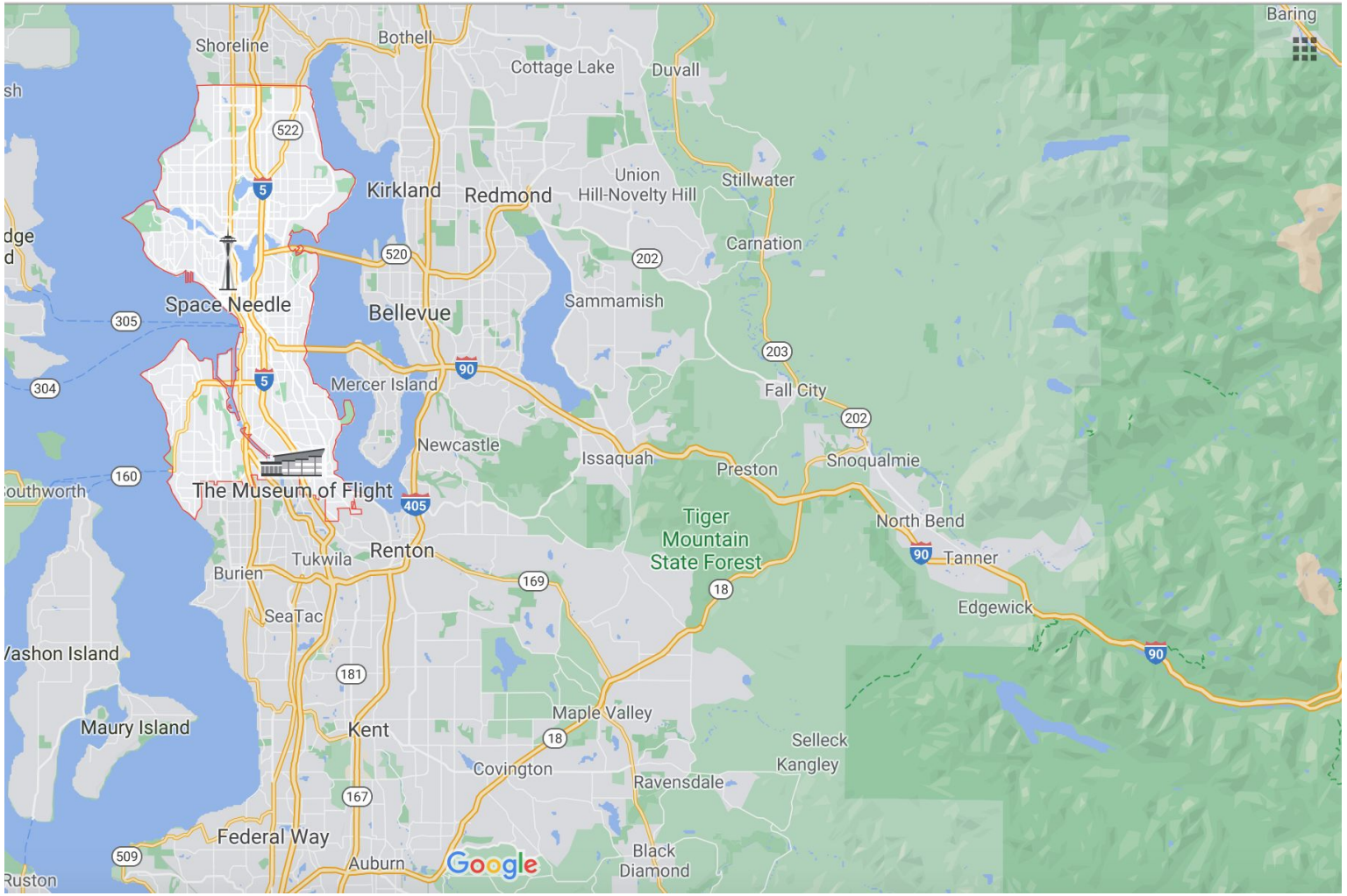
The dataset for King County included housing information contained a collection of **21,597** houses with 21 factors

- We narrowed those factors down to ones we found most useful to renovate.

Using Latitude and longitude, we split the area of King County into two districts, urban and rural. We created a separate model for each district.

This project uses **inferential analysis** to identify the relationship between the data's variables (**house information**) and **price**. To make these analyses, we used an **ordinary least squares linear regression model**.


47.8
47.7
47.6
47.5
47.4
47.3
47.2
Latitude
−122.4
−122.2
−122.0
−121.8
−121.6
−121.4
Longitutde


Shoreline
Bothell
Cottage Lake
Duvall
Baring
522
Kirkland
Redmond
Union
Hill-Novelty Hill
Stillwater
5
520
202
Carnation
Space Needle
Bellevue
Sammamish
305
203
304
Mercer Island
90
Fall City
5
202
Newcastle
Issaquah
Preston
Snoqualmie
160
The Museum of Flight
405
Tiger
Mountain
State Forest
North Bend
Tukwila
Renton
90
Tanner
Burien
169
18
Edgewick
SeaTac
90
181
Maury Island
Kent
Maple Valley
18
Selleck
Kangley
Covington
Ravensdale
167
Federal Way
509
Auburn
Black
Diamond
Google

Our Models (Technical)

Urban Model

R^2= 0.501

	coef	std err	t	P>|t|	[0.025	0.975]
const	11.0645	0.020	552.633	0.000	11.025	11.104
sqft_living	0.0002	5.48e-06	38.914	0.000	0.000	0.000
grade	0.2039	0.003	59.288	0.000	0.197	0.211
waterfront	0.6208	0.033	18.573	0.000	0.555	0.686
has_basement	0.1303	0.005	25.146	0.000	0.120	0.140
bathrooms	-0.0342	0.005	-6.653	0.000	-0.044	-0.024

Rural Model

R^2= 0.750

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.269e+05	2.52e+04	-5.032	0.000	-1.76e+05	-7.74e+04
sqft_living	121.6805	5.014	24.269	0.000	111.837	131.524
grade	4.239e+04	4179.933	10.141	0.000	3.42e+04	5.06e+04
has_basement	-2.609e+04	7351.564	-3.549	0.000	-4.05e+04	-1.17e+04

Assumptions

To check the assumptions of our models we checked for:

- Variance inflation factor
- Normality of the residuals
- Heteroskedasticity and autocorrelation
- Linearity.

All of these assumptions were met for both models.

Results (Urban)

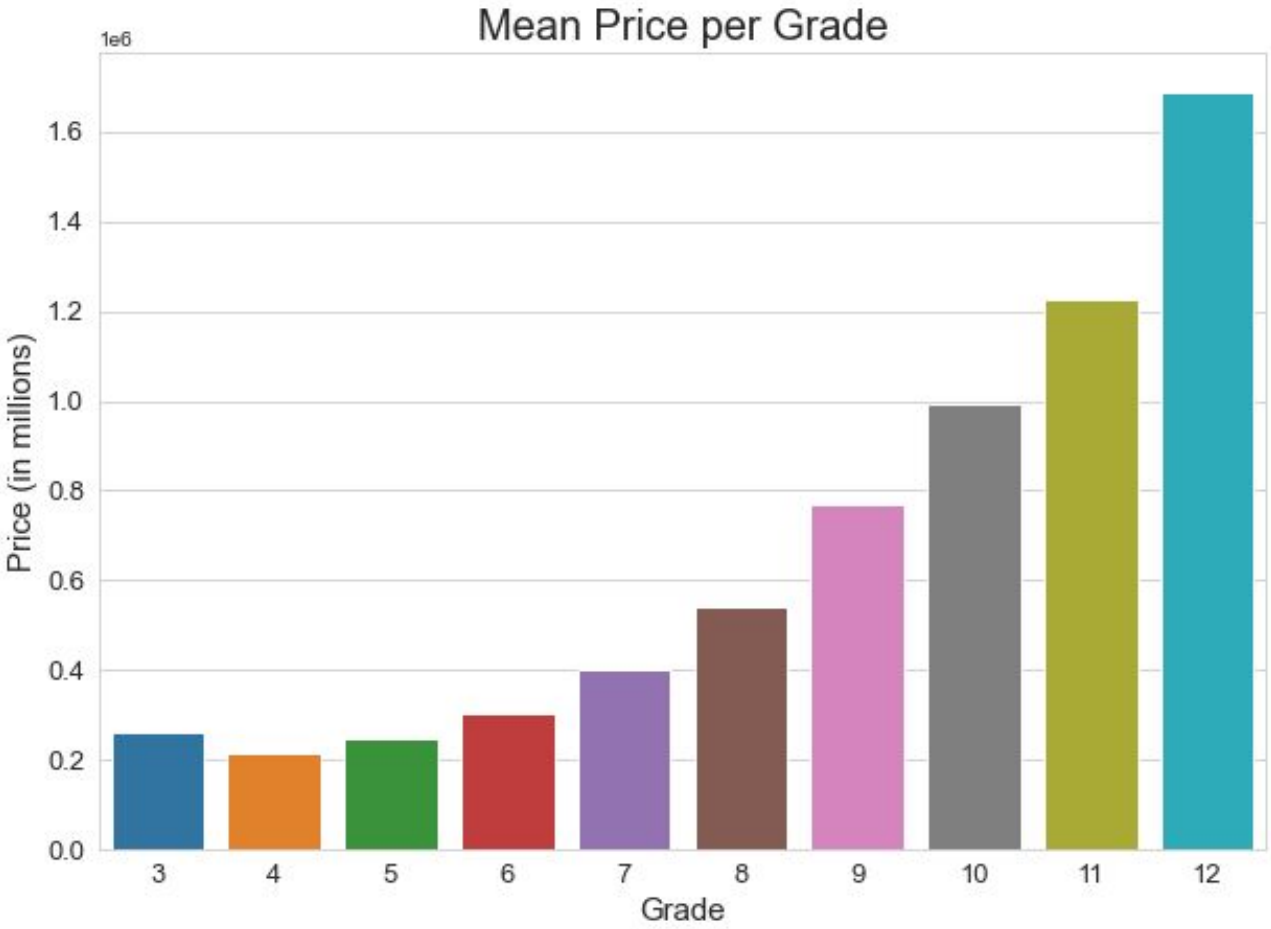
The model's coefficients represent the rate at which the **sales price** changes, given that the **independent variable** increased by 1 unit.

Price was log scaled for the urban model so it would fit our model more cleanly.

- The coefficient for sqft_living is **0.0002**: **(0.02% increase)**
  - Square Footage increases quickly with renovations
- The coefficient for grade is **0.2039**: **(22.62% increase)**
- The coefficient for ratio of bathrooms is **-0.0342**: **(3.36% decrease)**
- The coefficient for has_basement is **0.1303**: **(13.92% increase)**
- The coefficient for waterfront is **0.6208**: **(86.04% increase)**
  - This metric applies to houses with *potential* waterfront views that do not utilize them.

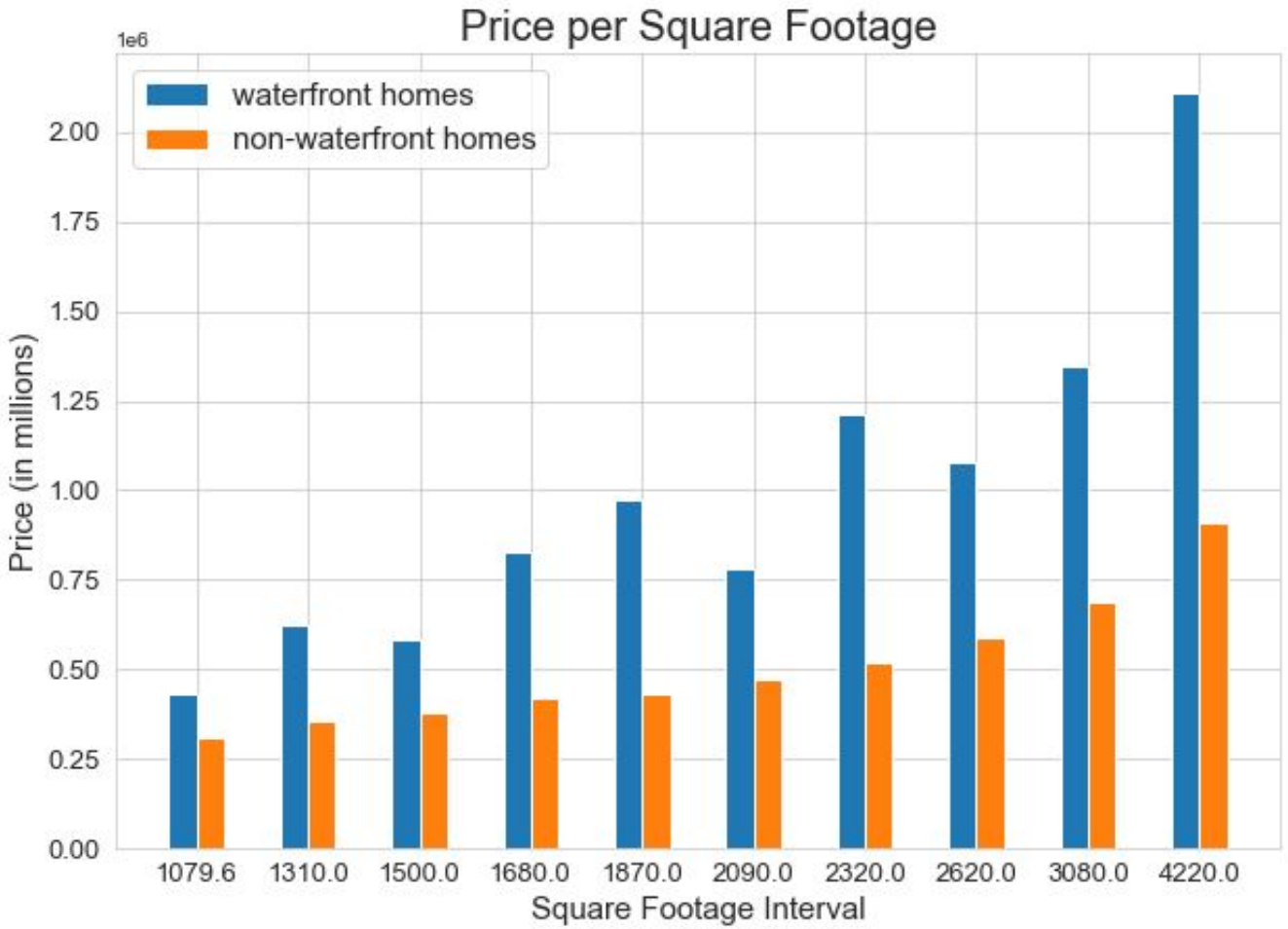
Results (Urban)

Results (Urban)



Results (Urban)

Results (Urban)

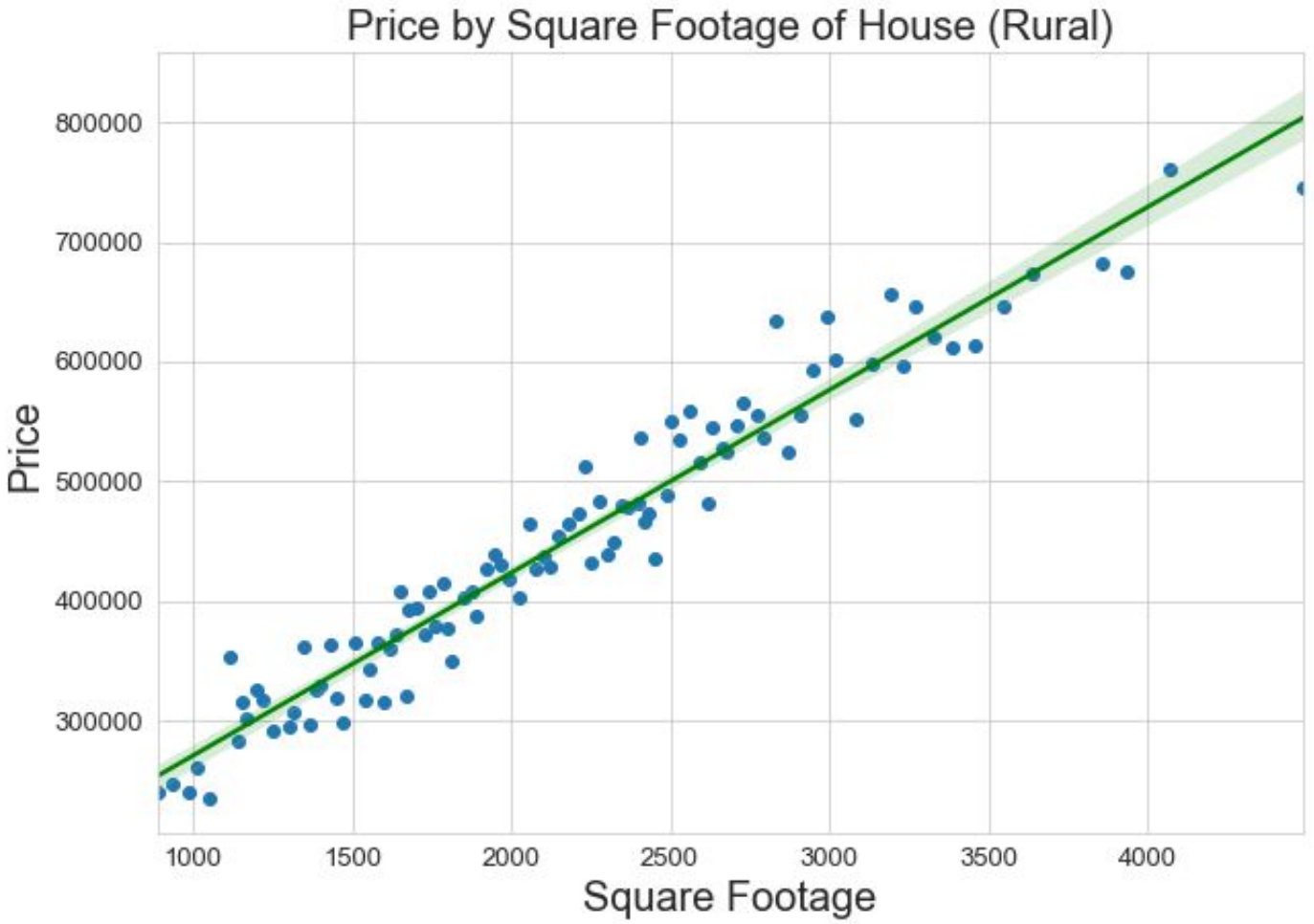


Results (Rural)

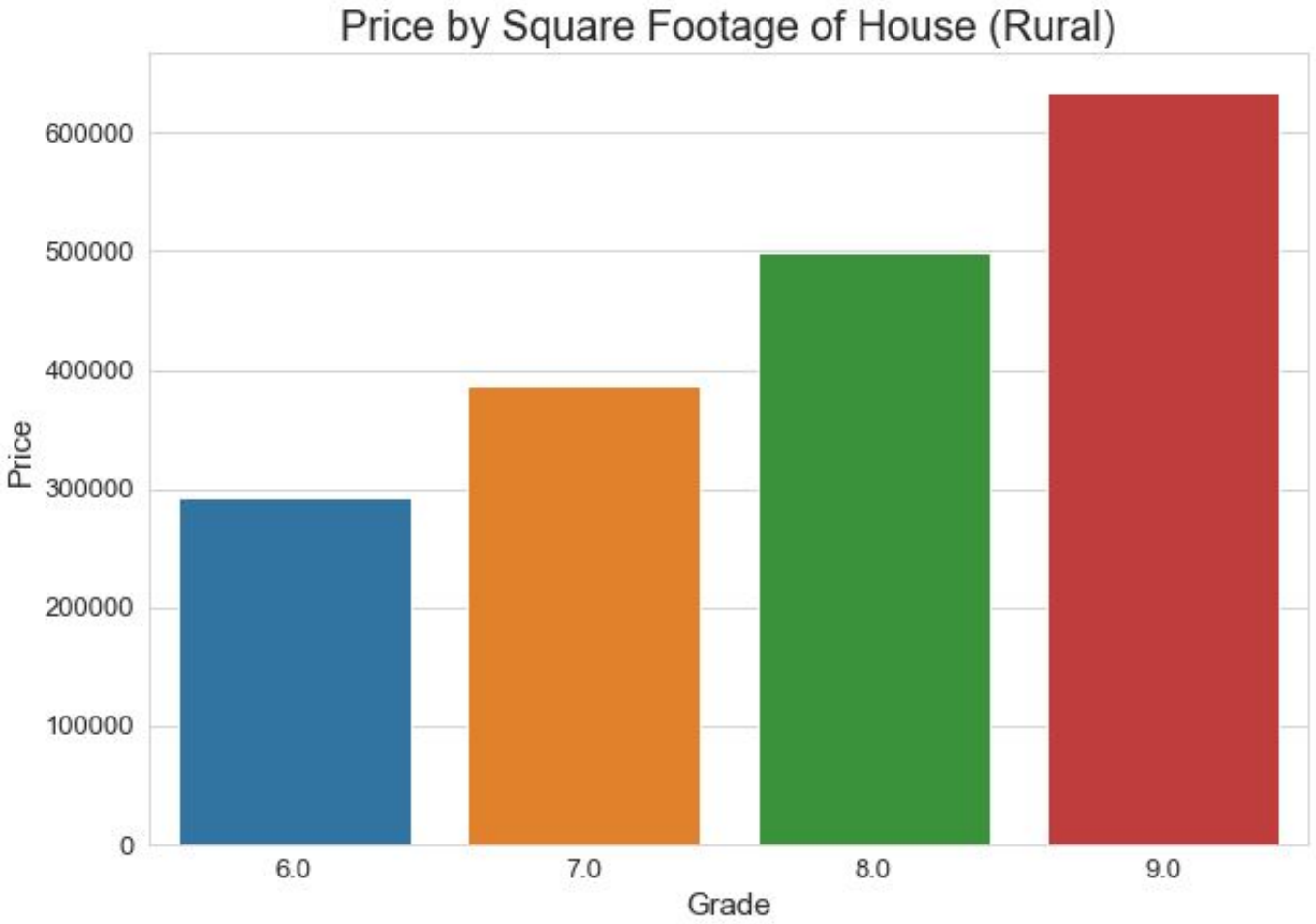
The model's coefficients represent the proportion at which the **sales price** changes, given that the **independent variable** increased by 1 unit.

- The coefficient for grade is **42,390**.
- The coefficient for has_basement is **-26,090.**
    - Are basements a desired feature in modern homes?

Results (Rural)



Results (Rural)



Conclusion in Urban King County:

- Add a basement
- Increase house square footage
- Improve grade
- Reveal waterfront view if view was previously obstructed

Conclusion in Rural King County:

- increase house square footage
- improve grade

Next Steps

- Evaluate interactions between different variables

(How does waterfront/basement effect changing other variables?)

- Further look into bathroom ratio variable

Thank You

Email:

killian.kahalley@gmail.com

jchandleroneal@gmail.com

gnr400800@gmail.com

Github:

@kjkahalley

@jchandleroneal

@GriffinRiner