Modeling King County Housing Prices

Outline

Business Problem

Why did I create these models?

Data

What was used to construct the models?

Methods

How was the data modeled?

Results

What do the models tell us?

Business Problem

Who?

Help Individuals and families interested in moving to King County, WA

What?

What prices should I expect to pay?

What factors affect cost the most?

What If I want to rent a home instead?

How much can I trust the results?

How?

Predictive Modeling

Interpretive Modeling

New Target Modeling

Statistical Analysis

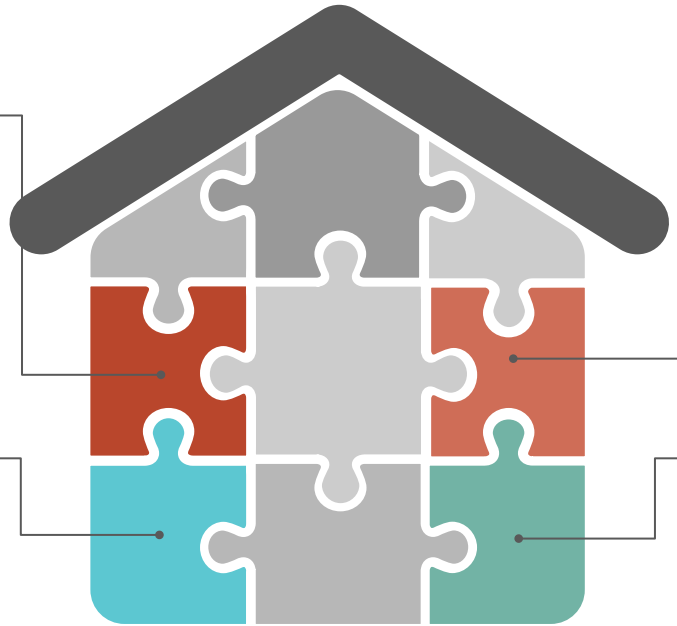
The Data

Roughly 20,000 homes

Ranging from around $75K to $1.1M

Physical Properties

Bedrooms, bathrooms, construction quality, living area, lot area, condition, age, floors, basement, renovations



Geological Information

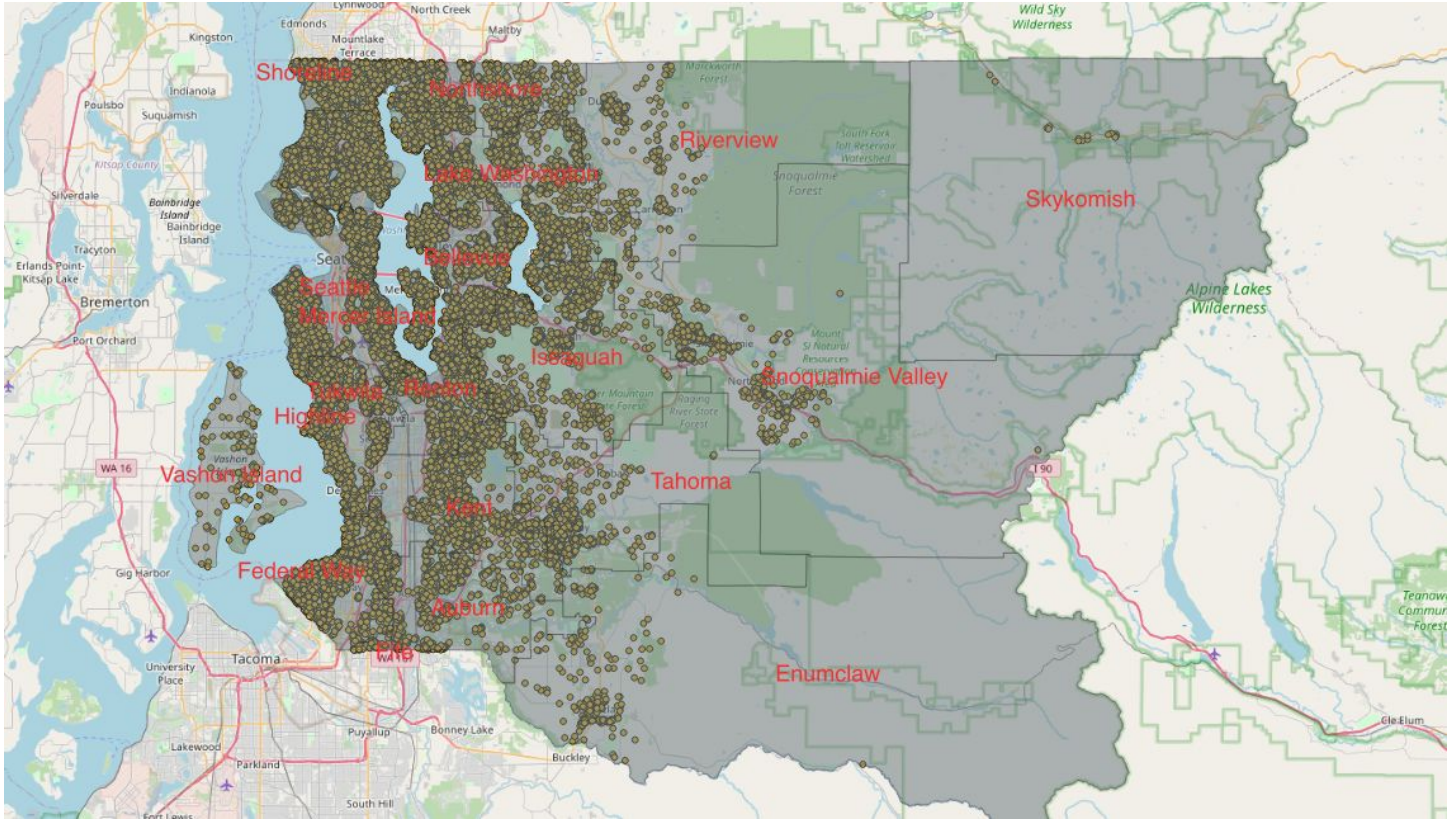
School districts, nearby parks, museums, golf courses and bodies of water, health disparity index*

Income Survey Data

US Census American Community Survey (2014-2018) data of 400 geological tracts: median household income, per capita income, and median contract rent

*Washington State Health Department's tool for analyzing areas of high health risk factors (socioeconomic and environmental)

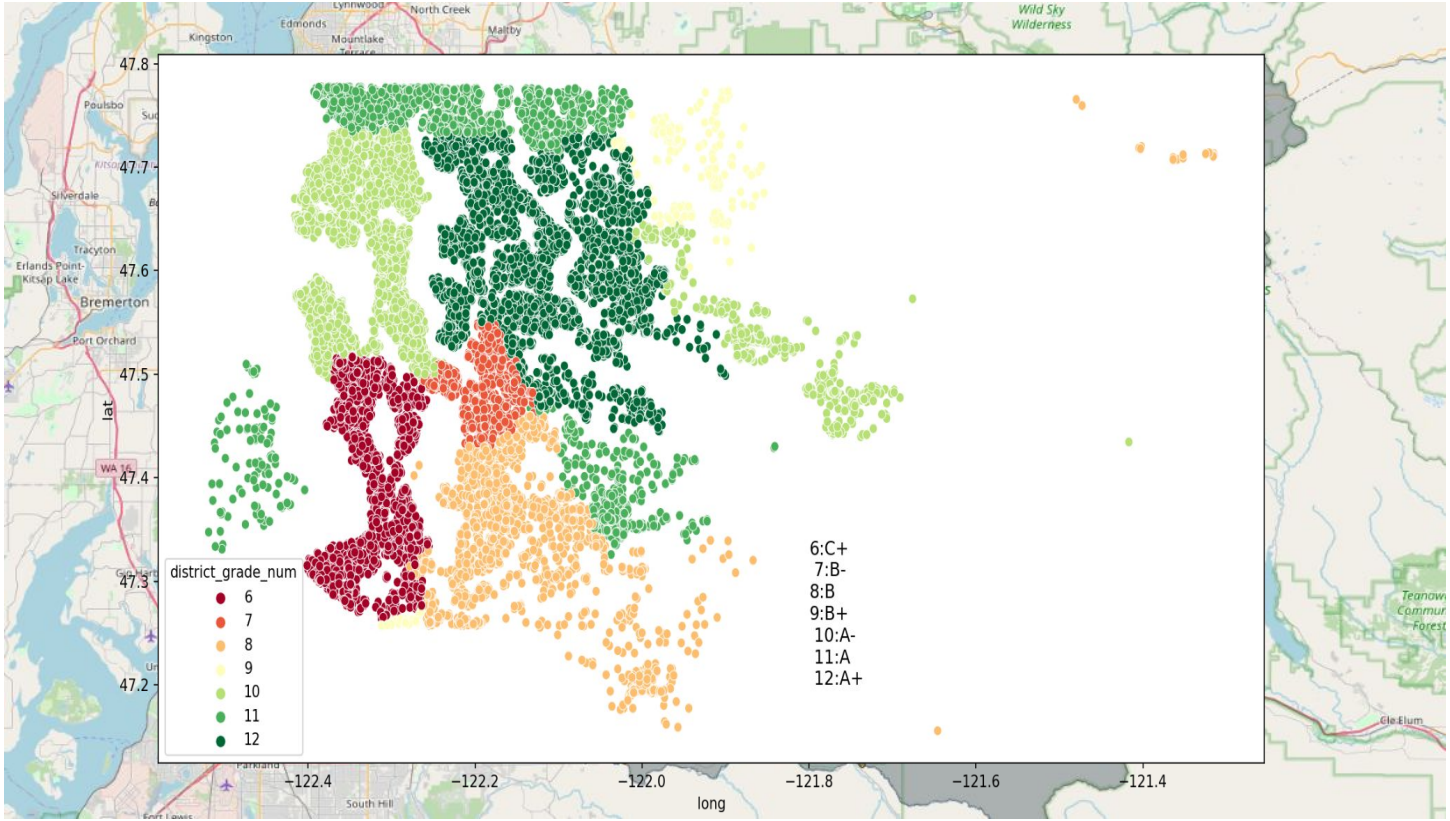
Properties, School Districts, District Ranks



20,000 homes and 20 different school districts

Niche.com rankings were used to transform into numerical data

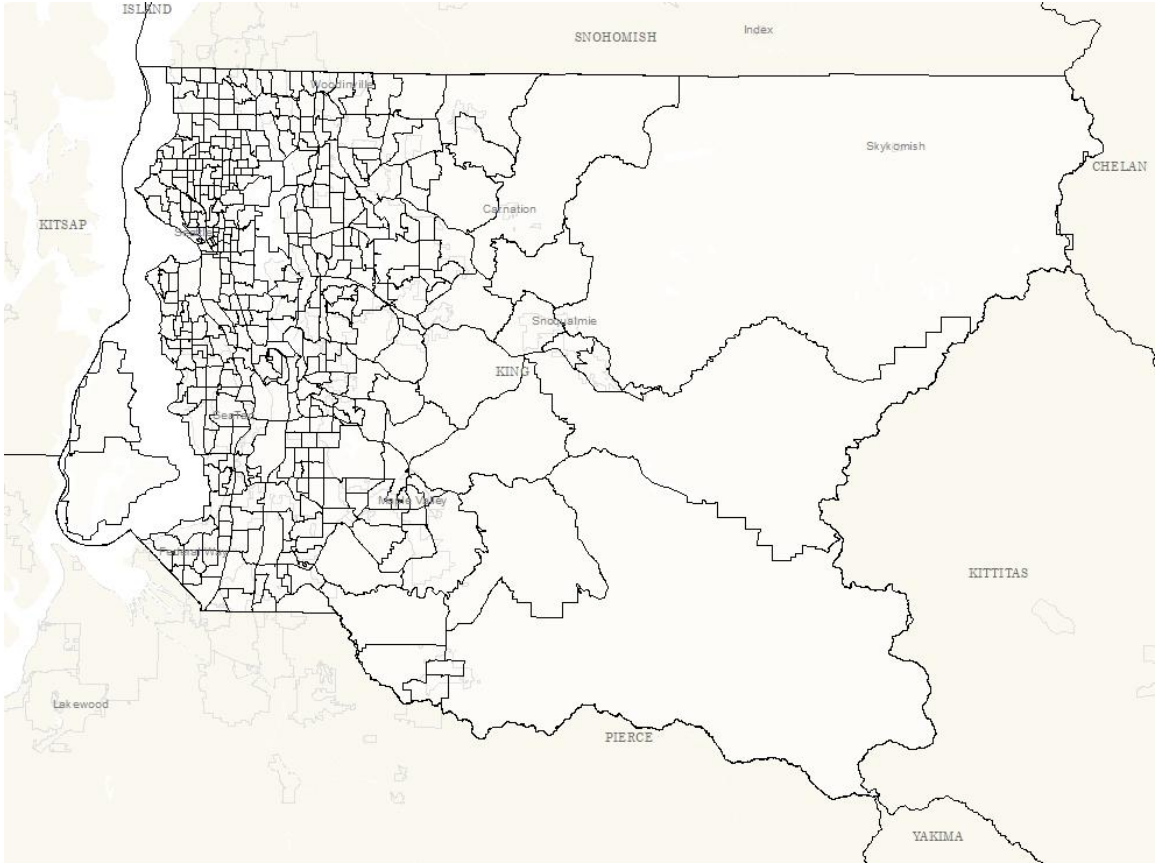
Properties, School Districts, District Ranks



Niche.com uses a ranking scale from D- to A+

Letter grades converted into numerical equivalents

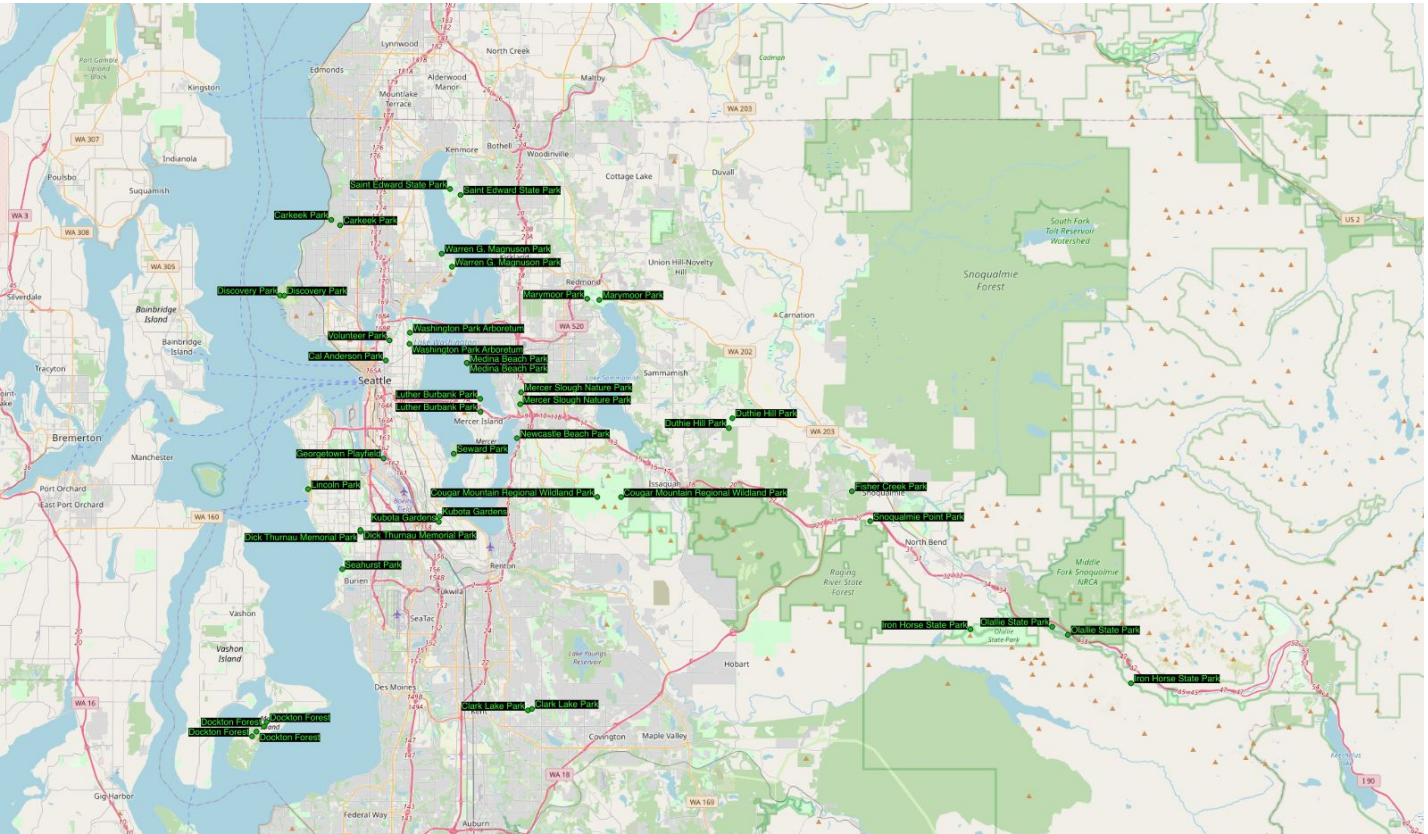
Tract Division for Income Survey



398 census blocks derived from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB)

"Census Blocks: statistical areas bounded on all sides by visible features, such as streets, roads, streams, and railroad tracks, and/or by nonvisible boundaries such as city, town, township, and county limits, and short line-of-sight extensions of streets and roads" -U.S. Census Bureau.

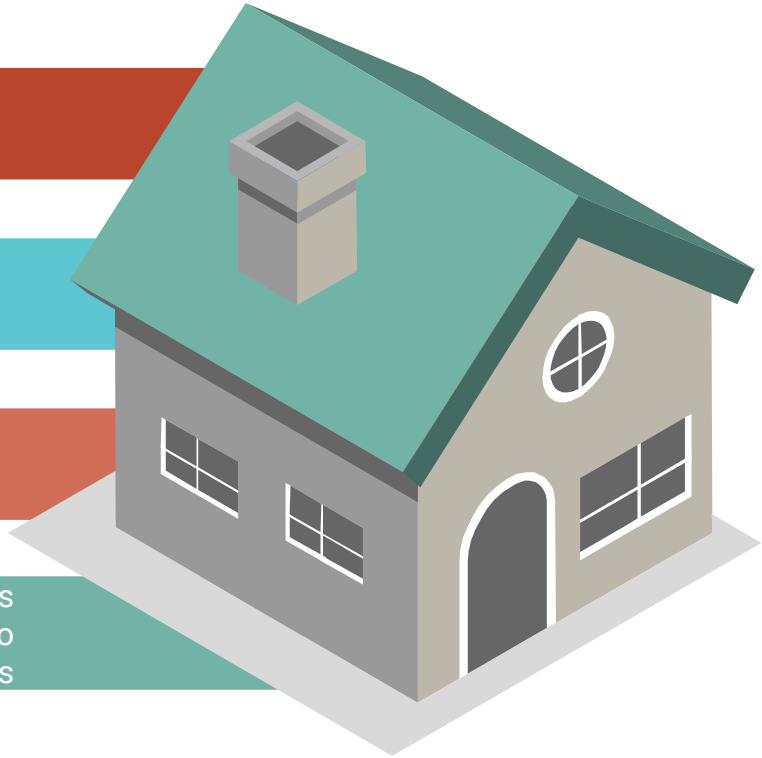
Top Parks



Top park data aggregated from kingcounty.gov, yelp, and various King County nature blogs

Method: OSEMN Data Process

01	**Obtain**	Gather appropriate data (large undertaking for this project)
02	**Scrub and Explore**	Process data and explore the distribution
03	**Model***	Iteratively develop models to solve problem using processed data
04	**Interpret**	Use statistical analysis and domain knowledge to answer questions



*All models in this project are multiple linear regression models

Model 1: Predictive Housing Price Model

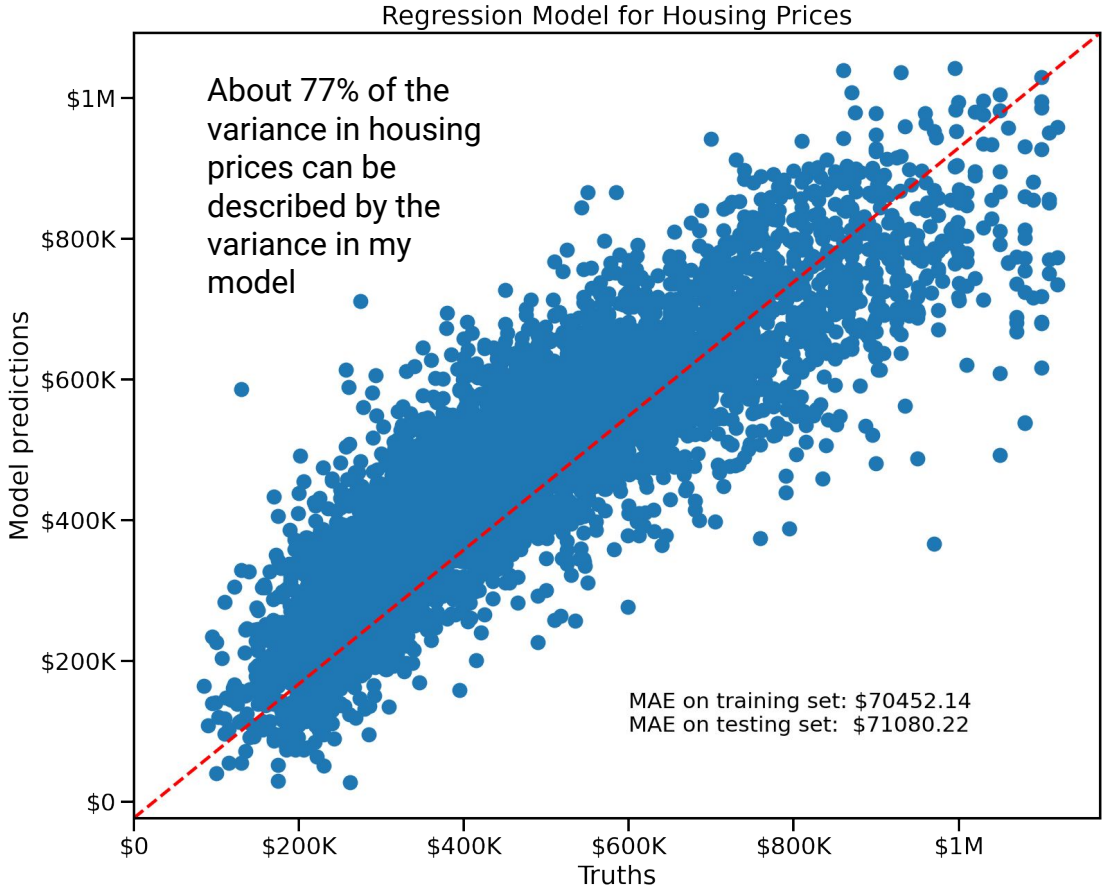
Train Score (R^2):
0.7796727579611459

Train Mean Absolute Error:
70452.14337001787

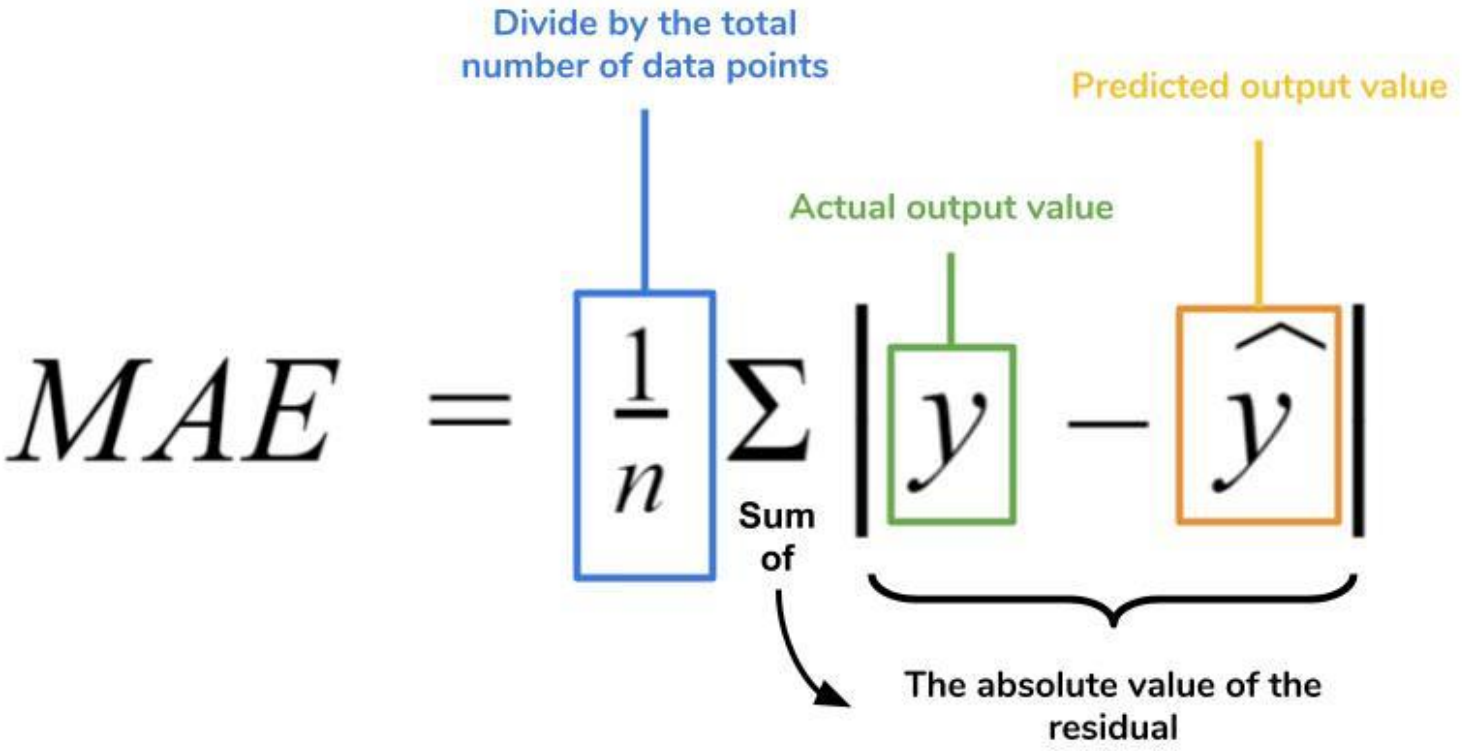
Test Score:(R^2):
0.7722599700234719

Test Mean Absolute Error:
71080.22300422823

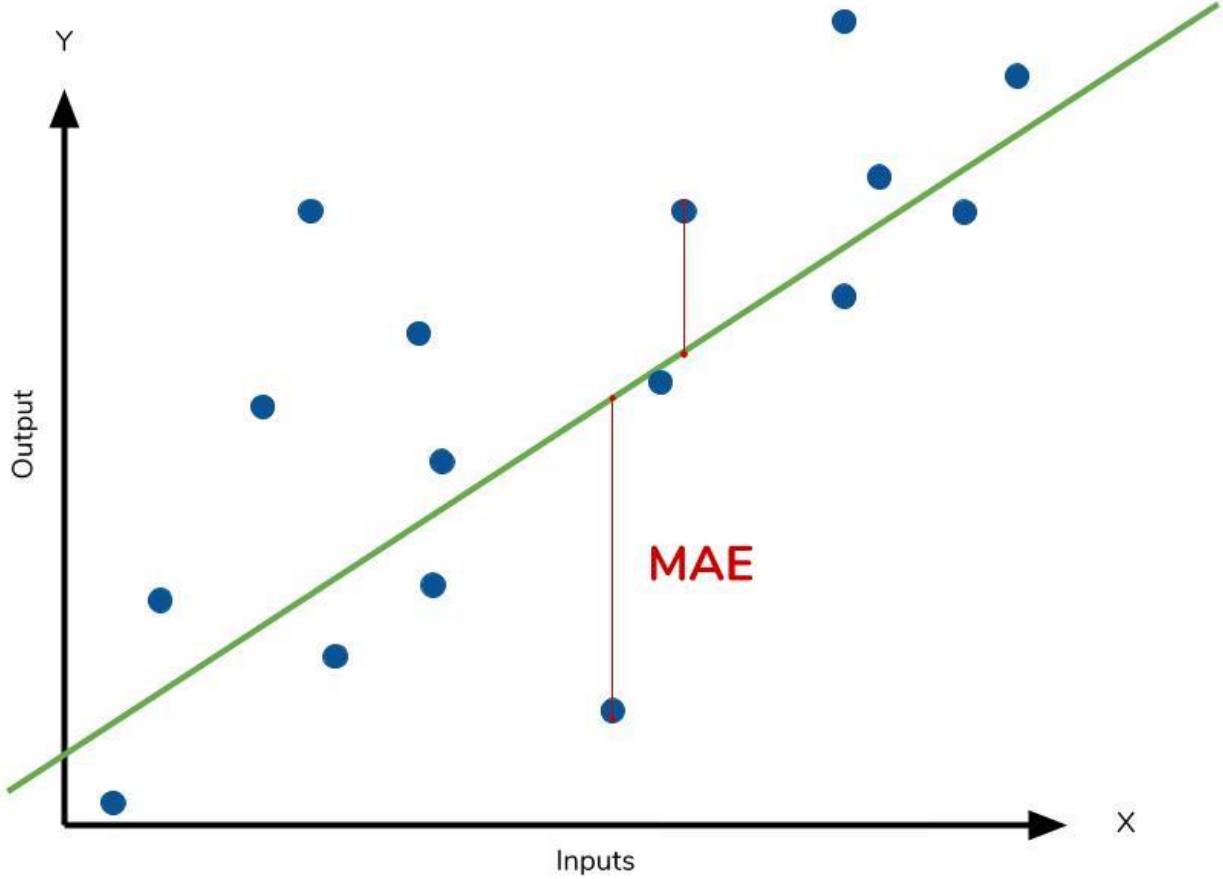
18 total dependent variables that describe the physical, geological, and socioeconomic features of a house



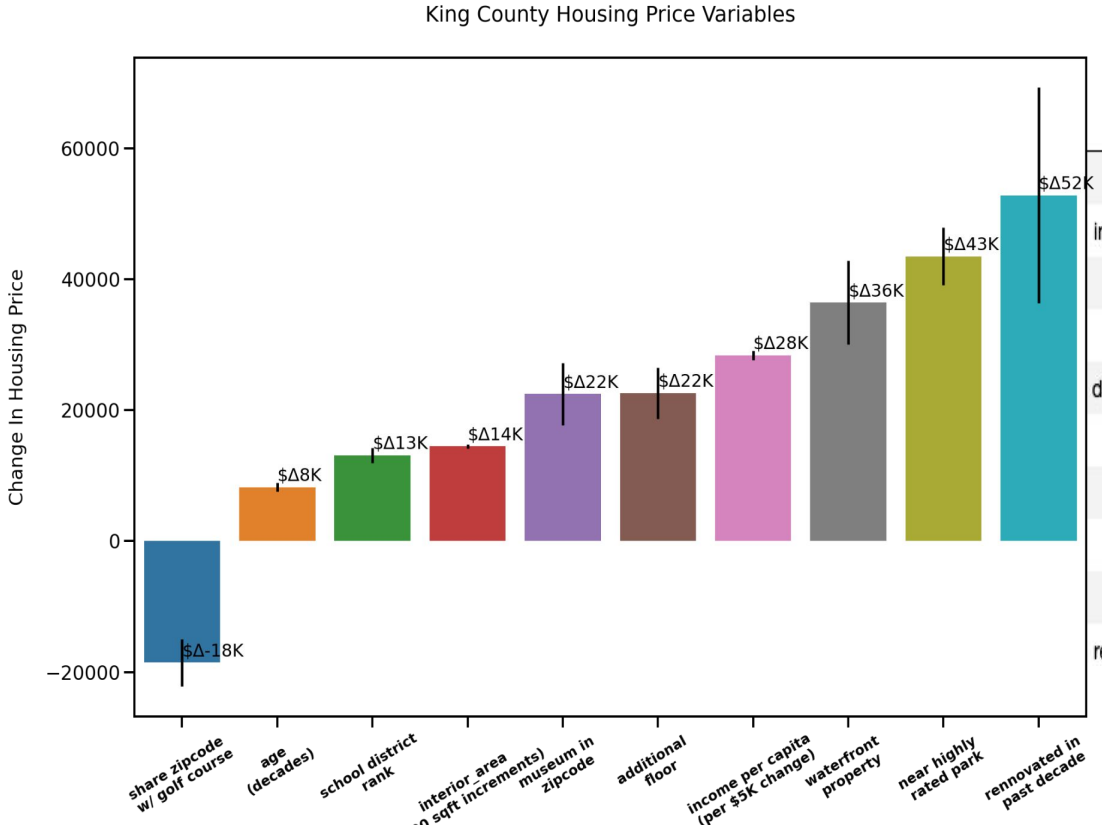
Quick Statistical Review



Quick Statistical Review

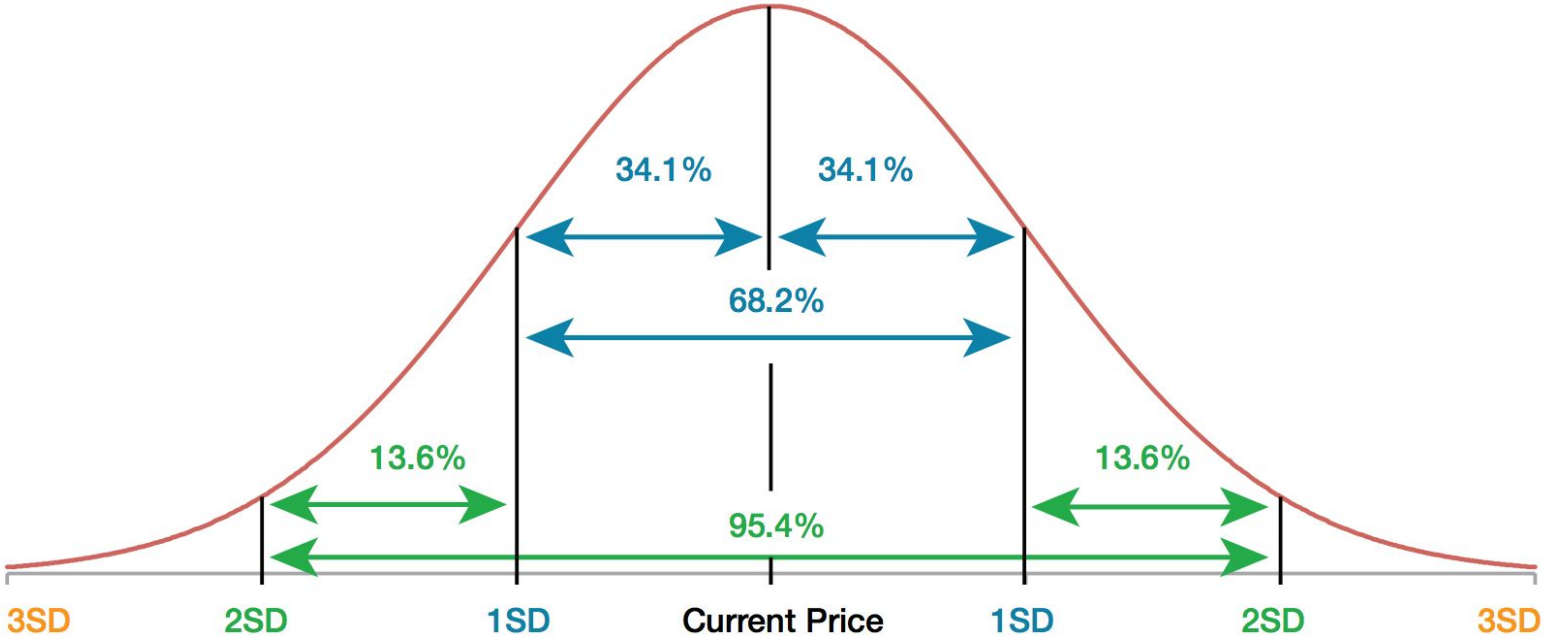


Model 2: Interpretive Housing Price Model



feature	coeff	lower	upper	standard_error	RSE(%)
golf_zipcode	-18575.563897	-22177.291845	-14973.835948	1837.488763	-9.89
income_per_capita	5.675262	5.532927	5.817597	0.072615	1.28
sqft_living	144.689709	141.962402	147.417016	1.391387	0.96
age_at_sale	824.200151	754.793595	893.606707	35.409051	4.30
district_grade_num	13109.617435	11975.863786	14243.371084	578.405593	4.41
museum_zipcode	22461.435461	17708.622908	27214.248015	2424.736067	10.80
multi_story	22591.419374	18648.596335	26534.242413	2011.504793	8.90
new_waterfront	36423.818153	30029.547396	42818.088909	3262.156618	8.96
near_park	43497.294987	39130.200186	47864.389789	2227.954954	5.12
recently_renovated	52803.942494	36335.138841	69272.746147	8401.867684	15.91

Statistical Refresher



Model 3: Predictive Rental Contract Model

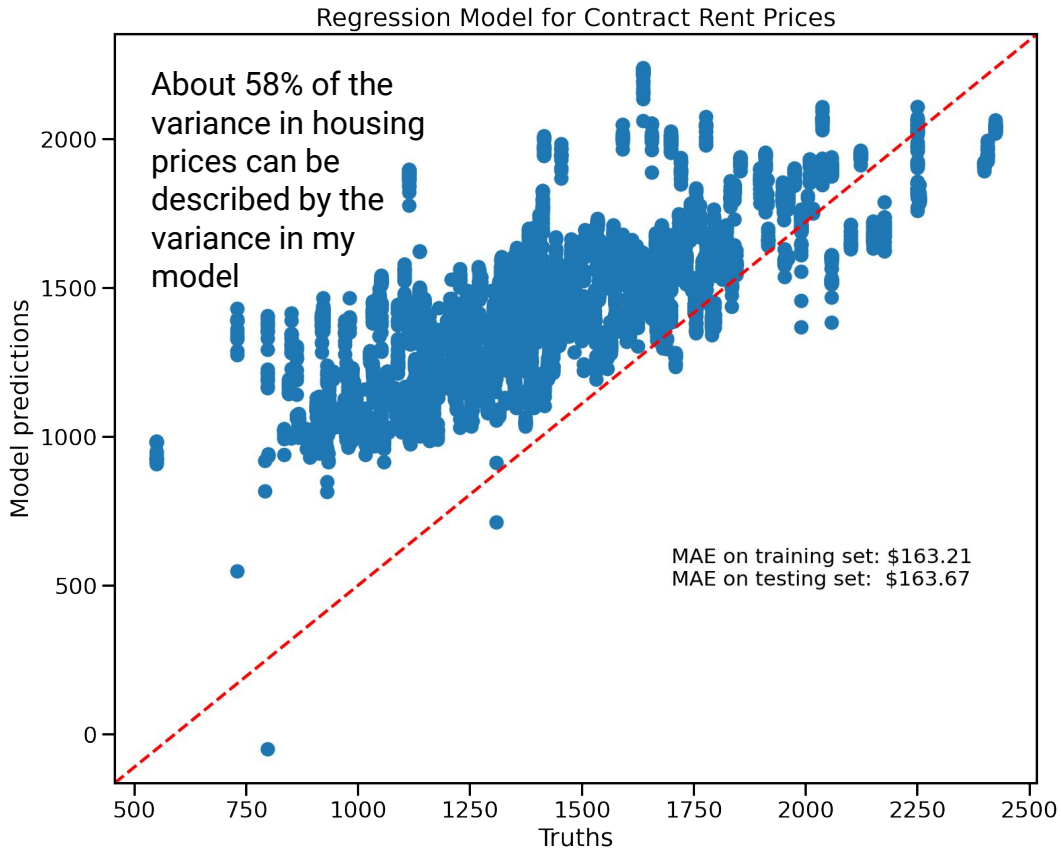
Train Score (R^2):
0.5836960366808019

Train Mean Absolute Error:
63.21287121452102

Test Score:(R^2):
0.5821658335321243

Test Mean Absolute Error:
163.67073938729447

15 dependent variables



Conclusion

Model 1
A predictive linear regression model that can take input from a user and output a range of expected prices for King County houses

Model 2
An interpretive linear regression model that describes individual affects of key features on King County housing prices

Model 3
A predictive model for King County rental contract pricing. Needs improvement.

Future Work
Work on improving the rental predictive model and construct an interpretive rental contract model

Thank you!!!

Author

Dylan Dey

Email

ddey2985@gmail.com

GitHub

@Ddey117

Questions?

Please do not hesitate to email!

Slide Template

By slidesgo.com

Your Time

Was very much appreciated!