Seattle Property Price Prediction

A detailed analysis of the factors affecting property prices in the Seattle area.

Overview

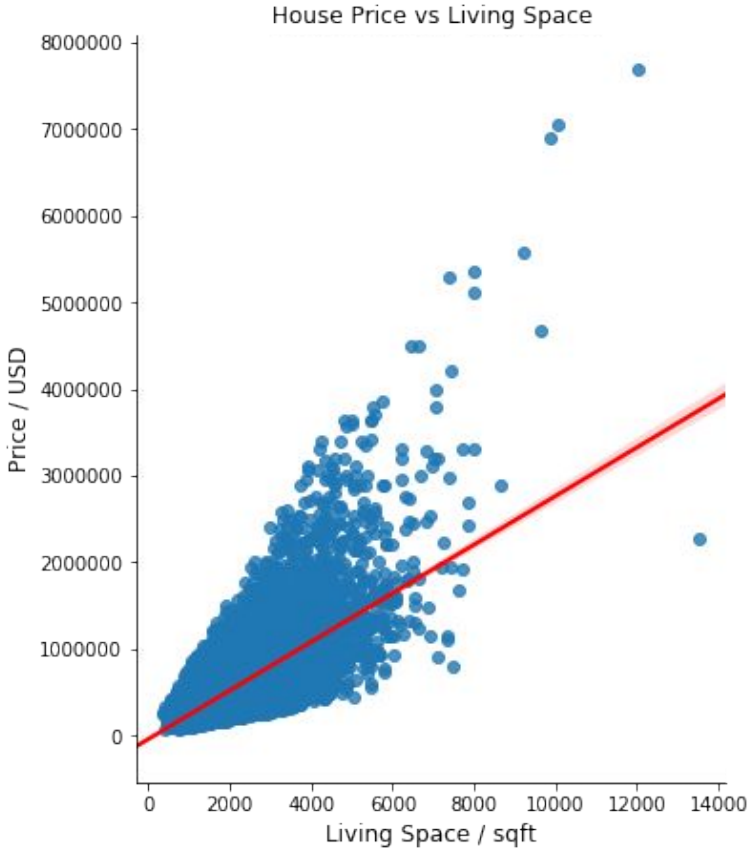
- Our aim is to take the mystery out of house pricing in the Seattle area
- Explain trends and the key factors in prices
- Help customers set reasonable prices for property sales

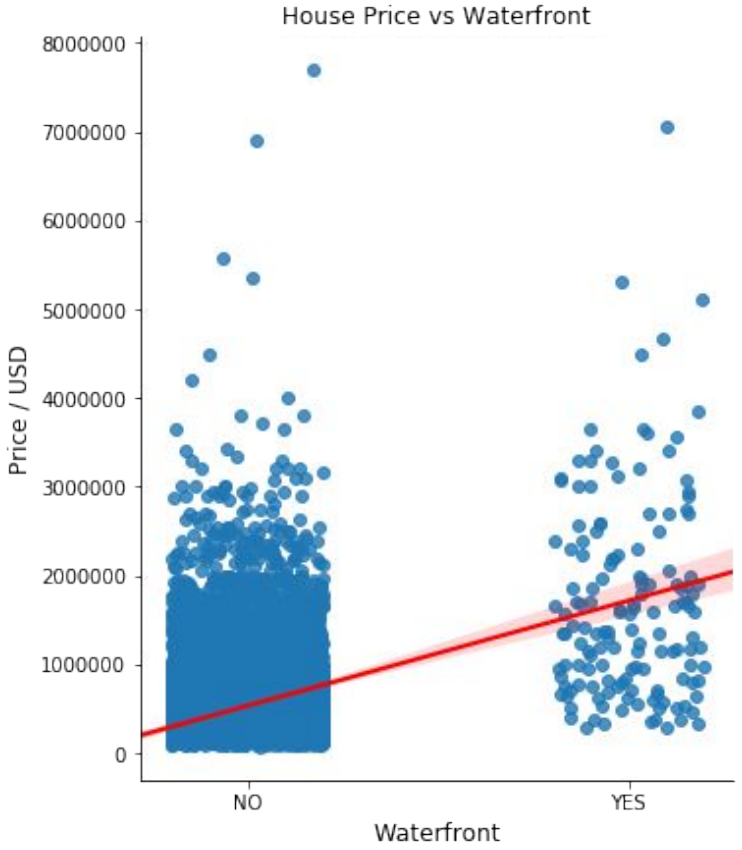
Questions to answer:

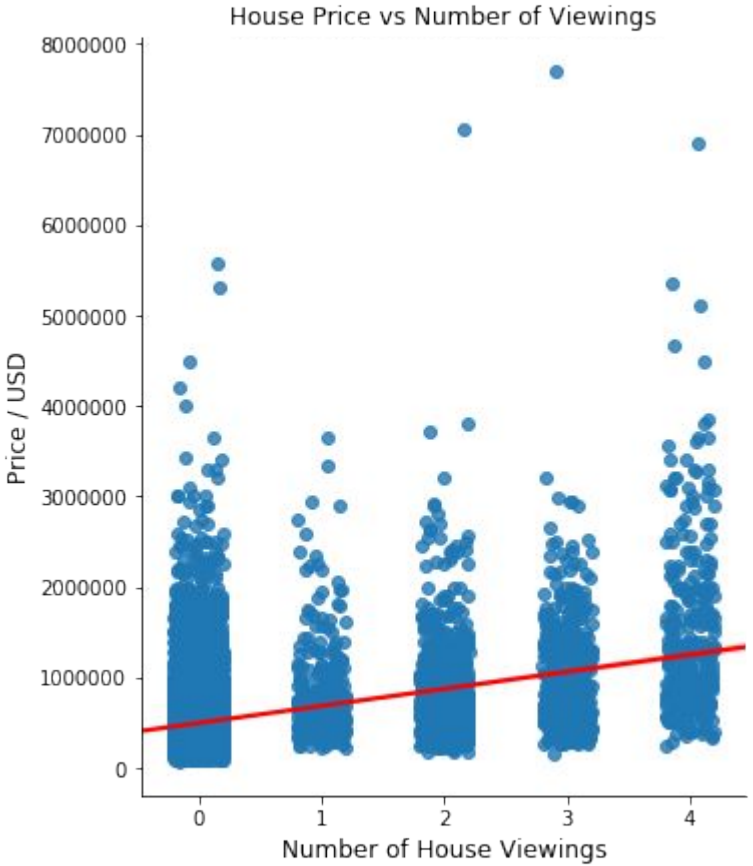
1. What attributes of a property most accurately determine its value?

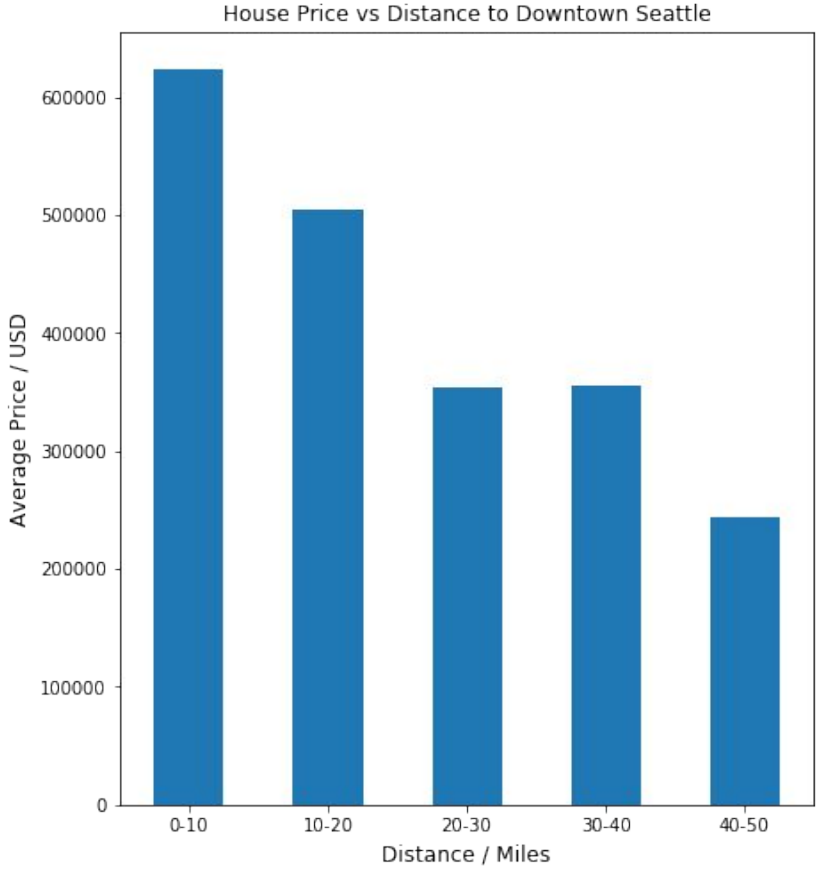
2. What is the relationship between a property's distance from Downtown Seattle and its price?

3. Can predictions be improved by conducting a more local analysis of the data?

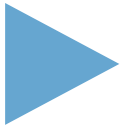

House Price vs Living Space
Price / USD
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
0
2000
4000
6000
8000
10000
12000
14000
Living Space / sqft


House Price vs Waterfront
Price / USD
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
NO
YES
Waterfront


House Price vs Number of Viewings
Price / USD
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
0
0
1
2
3
4
Number of House Viewings


House Price vs Distance to Downtown Seattle
Average Price / USD
600000
500000
400000
300000
200000
100000
0
0-10
10-20
20-30
30-40
40-50
Distance / Miles

To reflect current house prices we need to consider inflation...

2014/15  2015 Dec  2019 Dec

For predicting 2020 sale prices...

$$P = \$235{,}000 + \$420(\text{Sqft L}) + \$1{,}230{,}000(W) - \$27{,}600(D)$$

$$r^2 = 0.643$$

P : Price (USD)

Sqft L : Square Feet of Living Space

W : Is the property is on the waterfront?
Yes : W = 1, No : W = 0

D : Distance from Downtown Seattle

Location, location, location...

Separating the dataset by zipcode improves accuracy

	Zipcode Specific R-squared	Entire Dataset R-squared	Difference in R-squared
Best Performing	0.901	0.643	0.258
Mean	0.700	0.643	0.057
Worst Performing	0.467	0.643	-0.176

= 5.7%

Average increase in R-squared values

Future Work

- Investigate the drivers of house prices at a zipcode level
- Consider additional factors such as proximity to transport links and population density
- Consider more complex relationships between variables

Thank you for listening.

We are happy to take any questions!

Model Output for Entire Dataset

OLS Regression Results

Dep. Variable:	price_dec_19	R-squared:	0.643
Model:	OLS	Adj. R-squared:	0.643
Method:	Least Squares	F-statistic:	1.284e+04
Date:	Wed, 22 Jan 2020	Prob (F-statistic):	0.00
Time:	12:43:40	Log-Likelihood:	-3.0227e+05
No. Observations:	21420	AIC:	6.045e+05
Df Residuals:	21416	BIC:	6.046e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.353e+05	6528.490	36.045	0.000	2.23e+05	2.48e+05
sqft_living	419.9678	2.441	172.077	0.000	415.184	424.752
waterfront	1.23e+06	2.72e+04	45.254	0.000	1.18e+06	1.28e+06
distance to town	-2.76e+04	337.413	-81.808	0.000	-2.83e+04	-2.69e+04

Zipcode Specific Model Output (98039 - Medina)

OLS Regression Results

Dep. Variable:	price	R-squared:	0.901
Model:	OLS	Adj. R-squared:	0.895
Method:	Least Squares	F-statistic:	136.8
Date:	Wed, 22 Jan 2020	Prob (F-statistic):	1.26e-22
Time:	12:53:06	Log-Likelihood:	-696.79
No. Observations:	49	AIC:	1402.
Df Residuals:	45	BIC:	1409.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.693e+06	1.31e+06	1.297	0.201	-9.36e+05	4.32e+06
sqft_living	617.9034	31.673	19.509	0.000	554.111	681.696
waterfront	8.902e+05	3.84e+05	2.318	0.025	1.17e+05	1.66e+06
distance to town	-3.848e+05	2.59e+05	-1.484	0.145	-9.07e+05	1.38e+05