



# House Sales in King County, USA

Predicting house prices,  
using sales data from May  
2014 to May 2015

# TABLE OF CONTENTS



**01**



## **King's County**

Location, Location and some Location

**02**



## **Ben's House's**

Selling a house in King County, buying one with waterfront.

**03**



## **Model Validation**

Test-train-split-validation and the RMSE

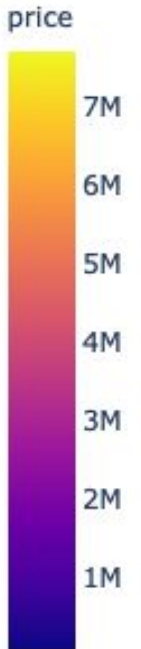
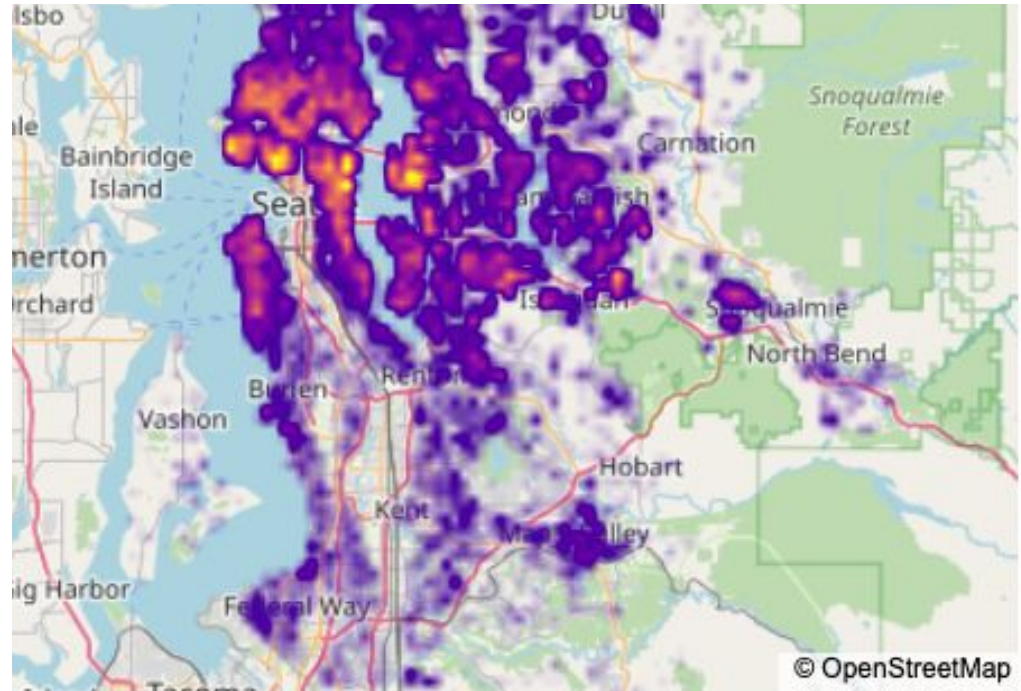
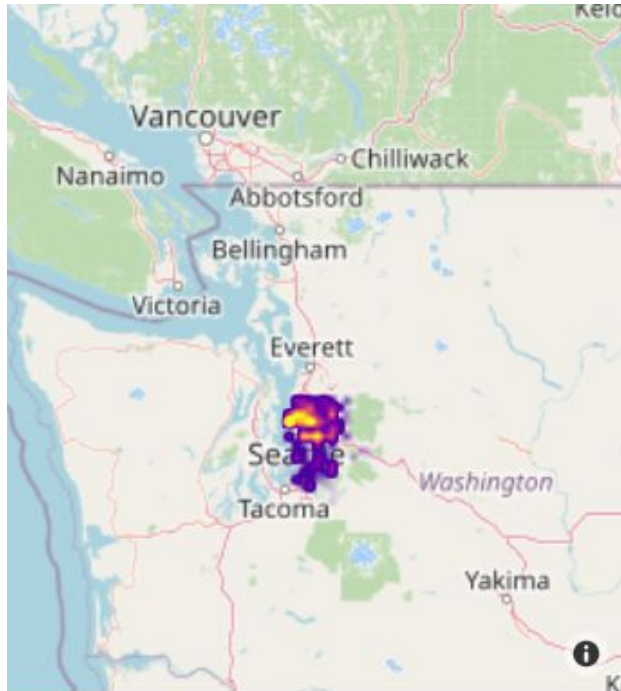
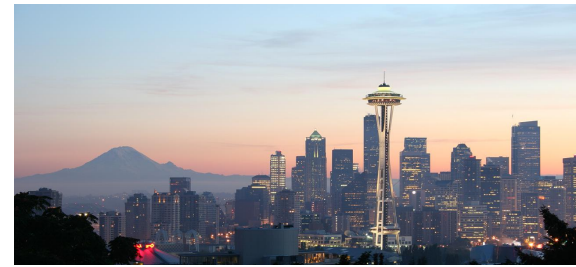
**04**



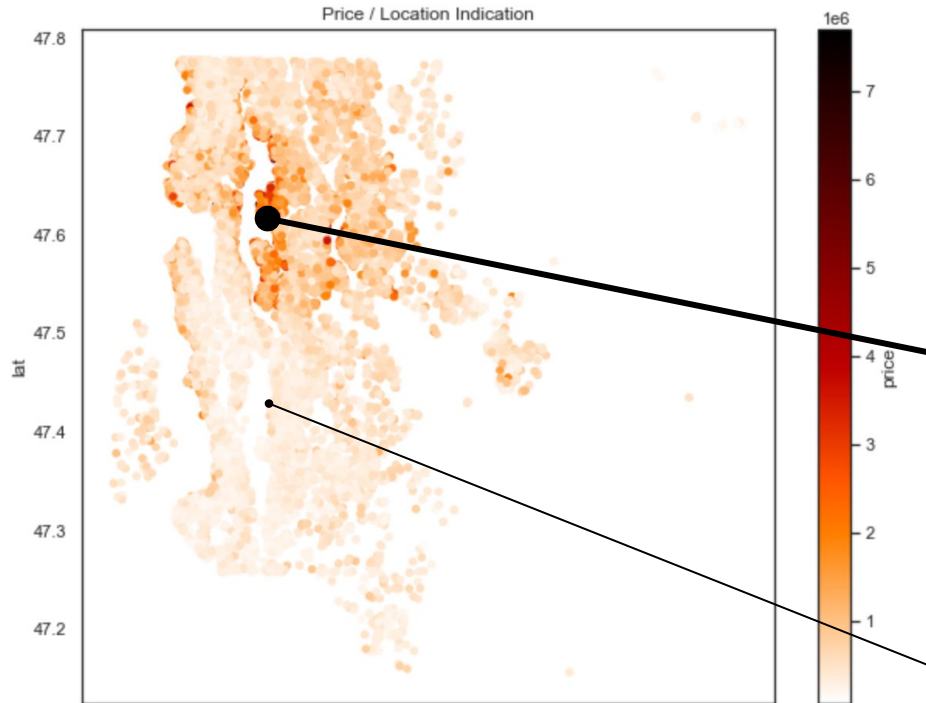
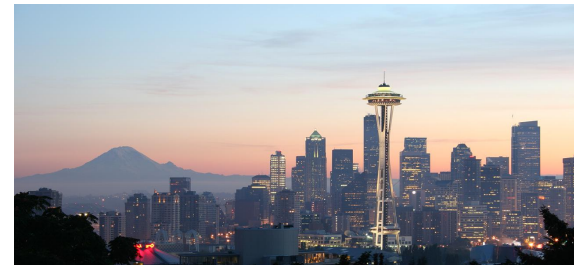
## **Feature Work**

Low and high hanging features.

# King County, USA



# King County, USA

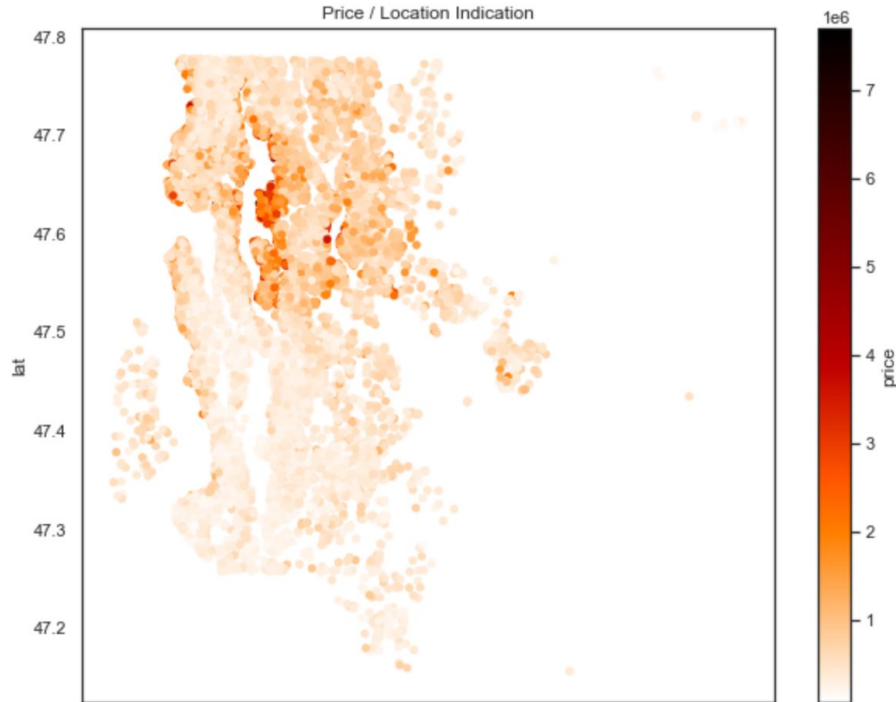


Is this the  
middle of  
Seattle?

Cheap  
waterfront?  
Sadly not.



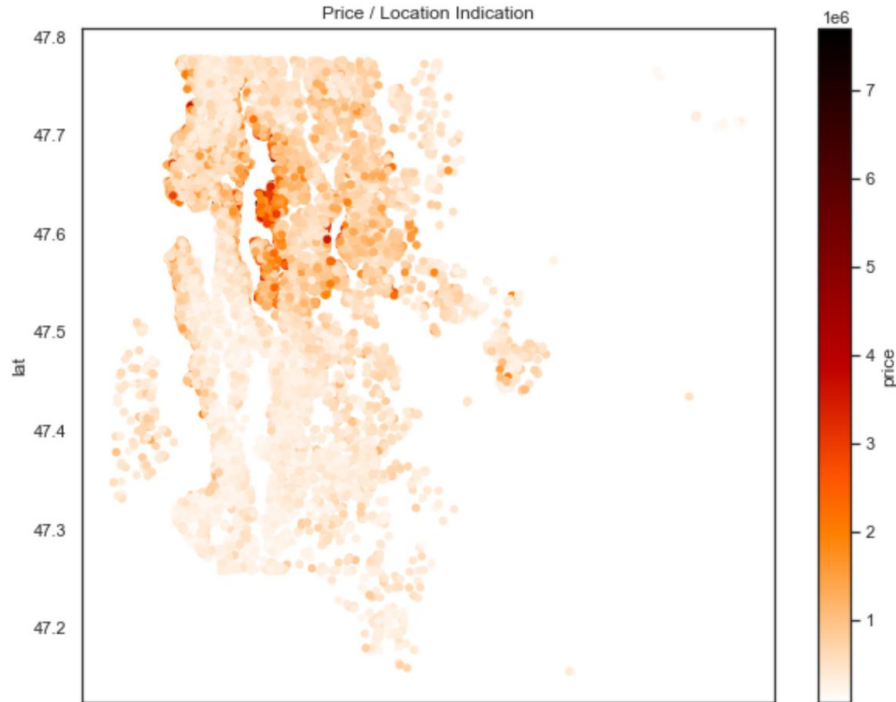
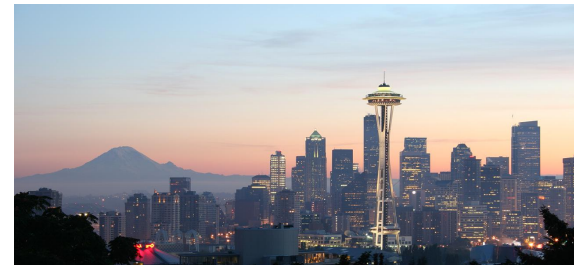
# King County, USA



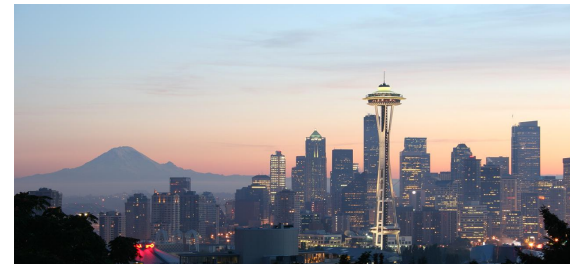
Starbucksiation

[star-buck-si-ation] describes a high density of Starbucks in a given area. Related: 'gentrification', 'capitalism'

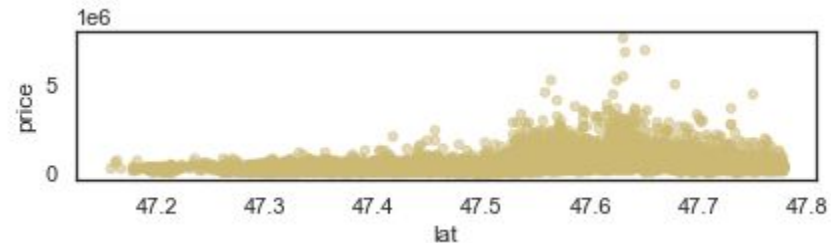
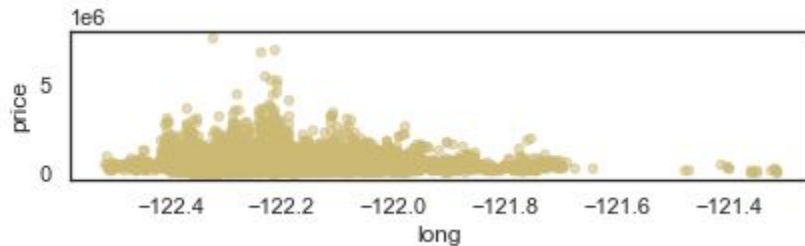
# King County, USA



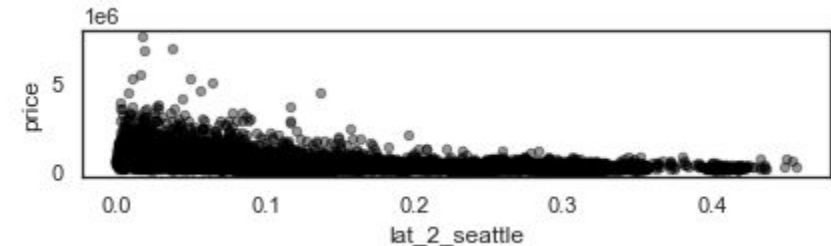
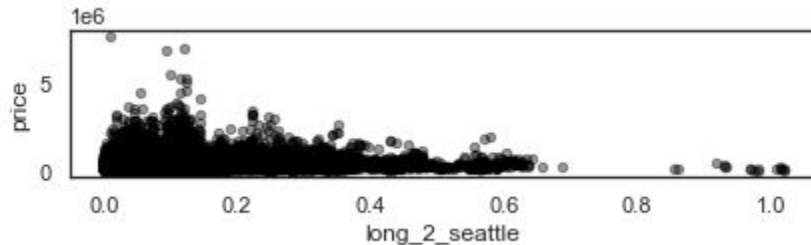
# King County, USA



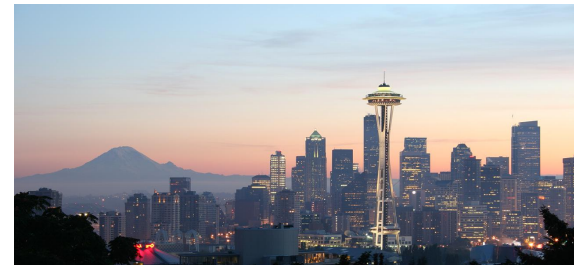
## Absolut Geolocation on price



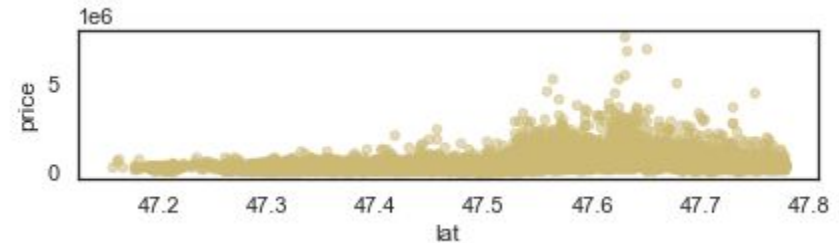
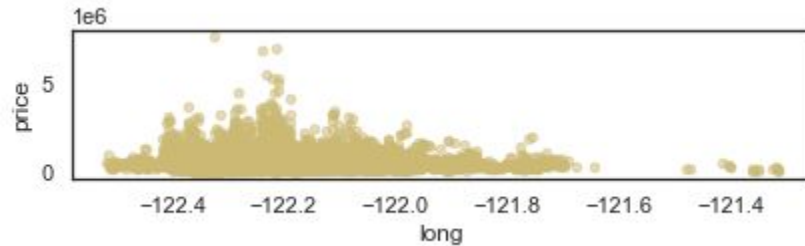
## Geolocation relative to Seattle City Center on price



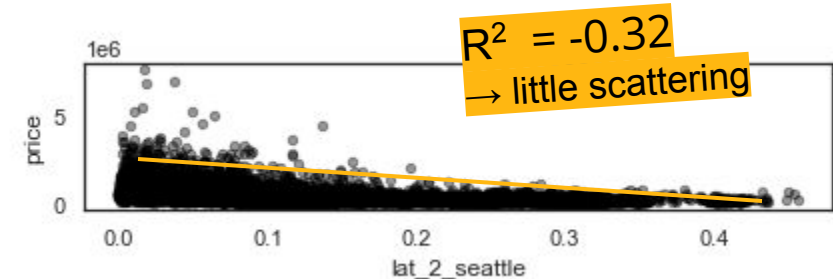
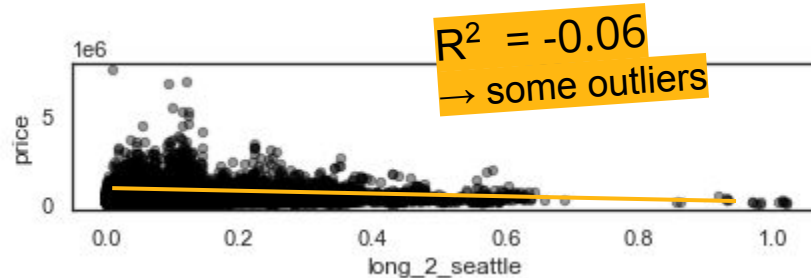
# King County, USA



## Absolut Geolocation on price

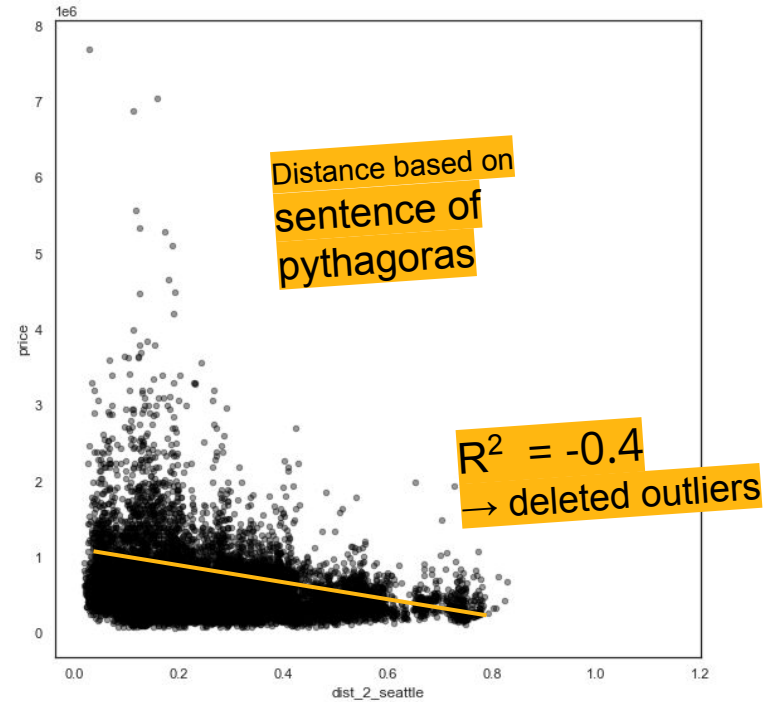
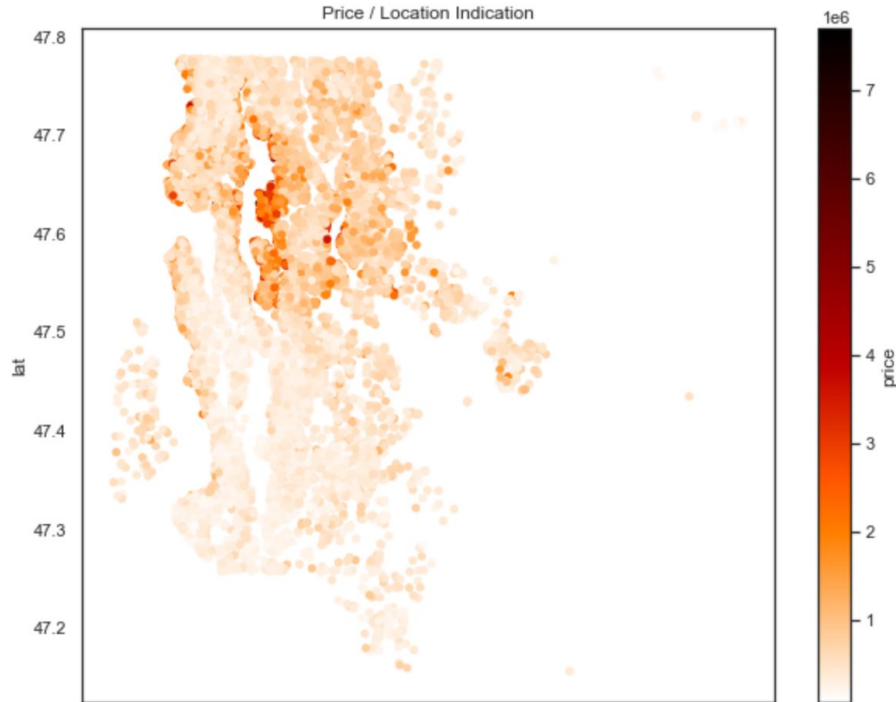


## Geolocation relative to Seattle City Center on price

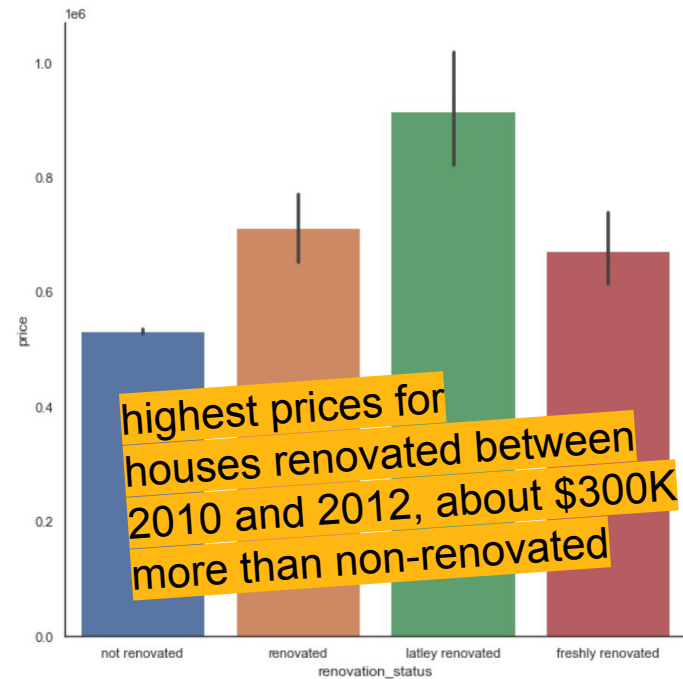
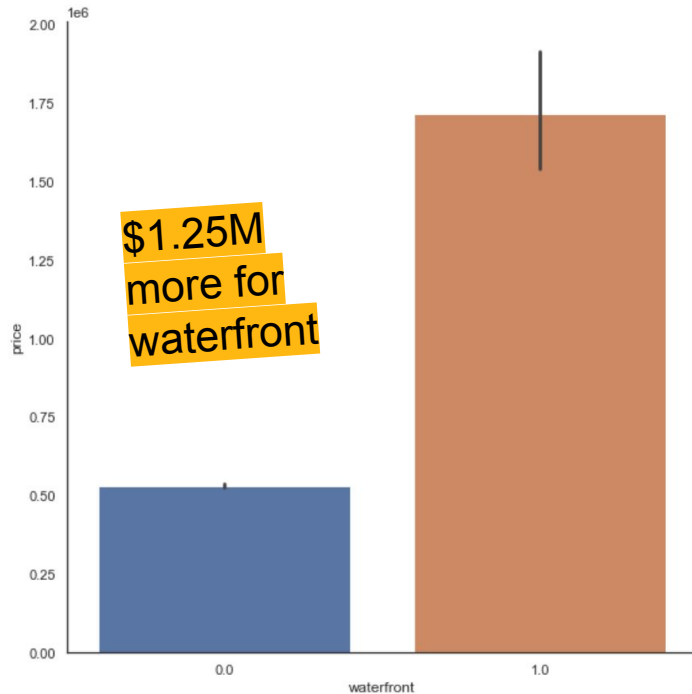




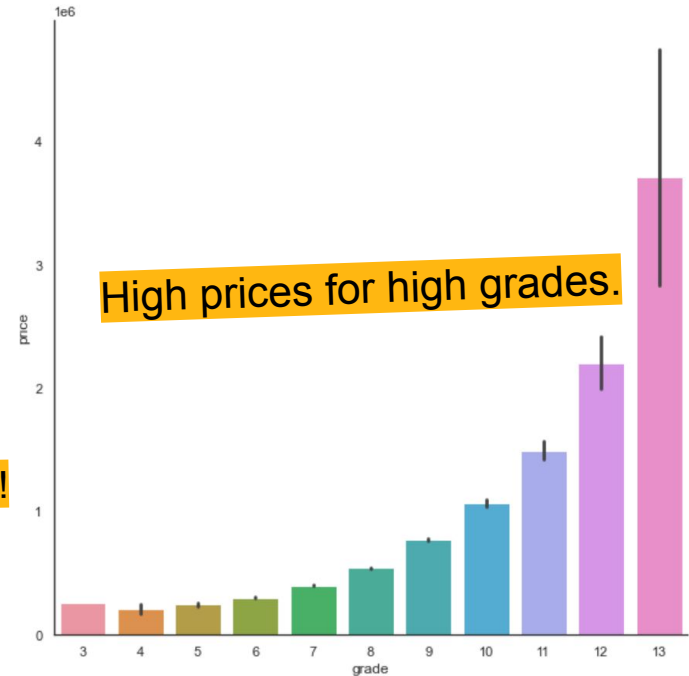
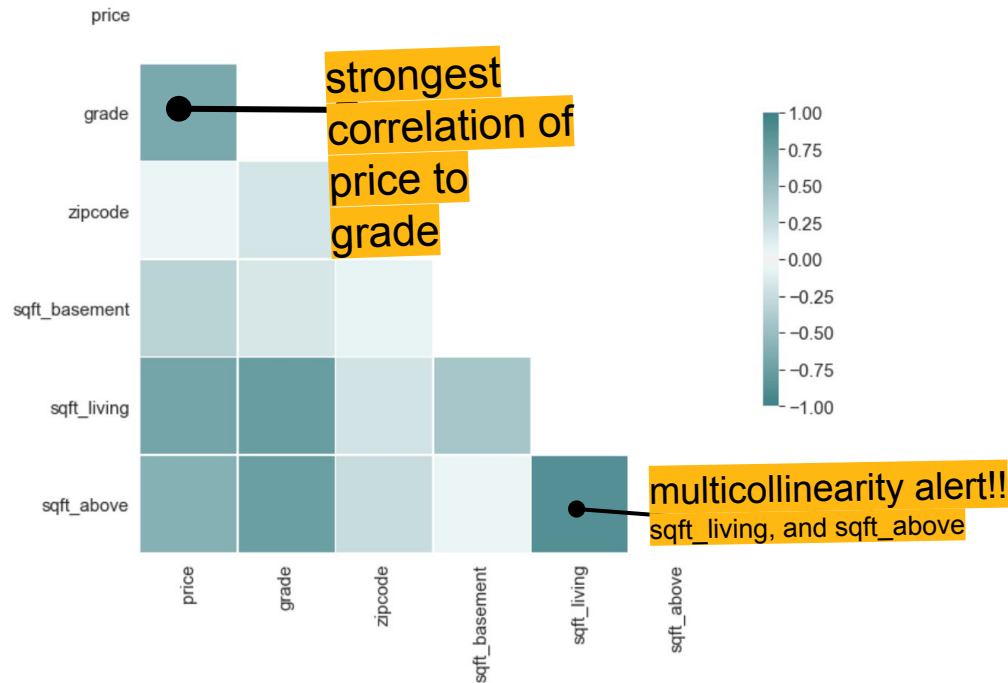
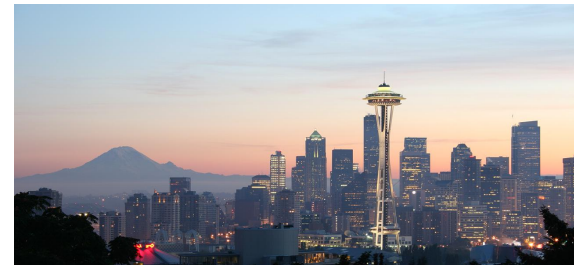
# King County, USA



# King County, USA



# King County, USA



# TABLE OF CONTENTS



**01**

## **King's County**



Quick look at the dataset

**02**



## **Ben's House's**

Selling a house in King County, buying one with waterfront.

**03**



## **Model Validation**

Test-train-split-validation and the RMSE

**04**



## **Feature Work**

Low and high hanging features.



Old friend, I need a new place. This whole home office thing makes me crazy.

I want to live close to the water, what does that cost?

Let's sell my old place. It's run-down at the moment, should I renovate?

Make it cheap, but I want to move in ASAP. Get back to me soon. B.







Macklemore

## Ben's Place:



Bedrooms	3
bathrooms	3
sqft_living	4560
sqft_above	2760
sqft_lot15	13362
waterfront	0
grade	11

**Renovation\_status:** not renovated

→ USD 2,909,336.00

**Renovation\_status:** renovated

→ USD 3,129,036.00

**\$220,000.00**

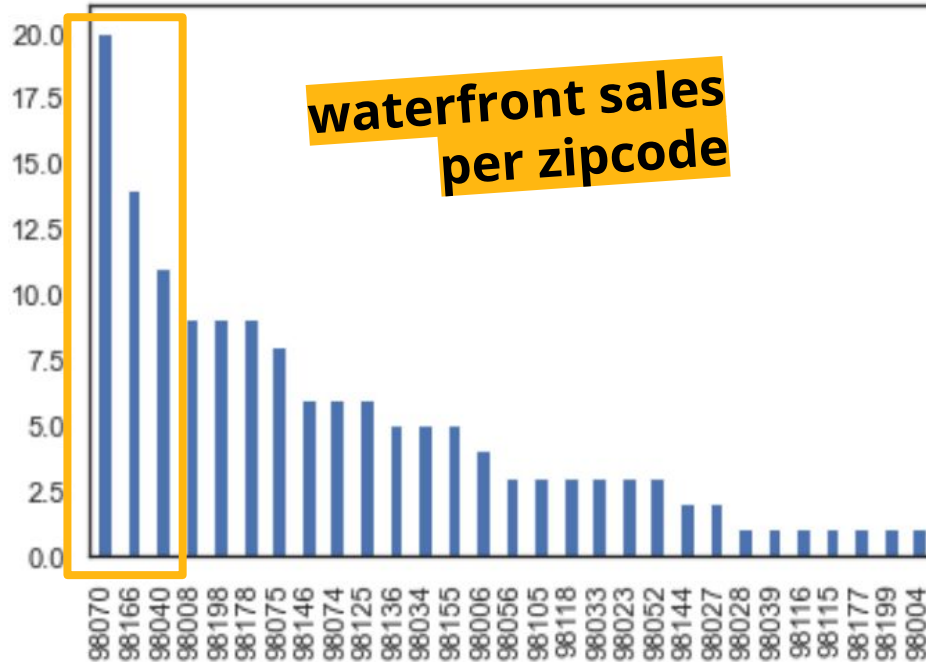
more for a freshly  
renovated house.



# 146 houses ×

At waterfront have been sold last year.  
That's less than 1% of all sales in the region.

# Where to buy Waterfront ✕



**\$ 1,500,000.00**

costs the median house with waterfront. Prices range mostly between \$850K and \$2,300K.

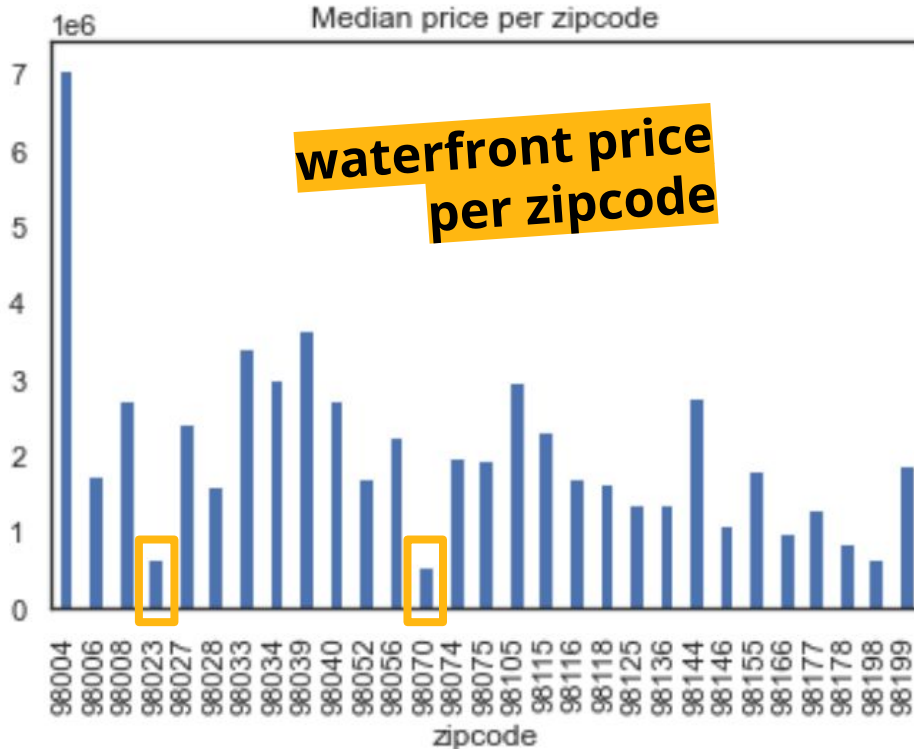
**98070, 98166, 98040**

30% of all houses with waterfront are sold in these zipcode areas. With an median price of only \$1,500K.

**98070, 96023**

zipcode areas with rather cheap waterfront houses.

# Where to buy Waterfront ✕



**\$ 1,500,000.00**

costs the median house with waterfront. Prices range mostly between \$850K and \$2,300K.

**98070, 98166, 98040**

30% of all houses with waterfront are sold in these zipcode areas. With an median price of only \$1,500K.

**98070, 96023**

zipcode areas with rather cheap waterfront houses.



Ben! Good to hear from you. Renovate your place for no more than \$220K. Not worth it. Look for a new place with waterfront in 98070, 98166, and 98040. Will cost you about \$1.5M. Cheaper places in 98070, 96023.

Btw, this analysis is  $48h * \$95 \text{ USD/h} \approx \$5K$  - whenever it suits you. Good talking to you. Hope your family is doing fine.



# TABLE OF CONTENTS



**01**



## King's County

Quick look at the dataset

**02**



## Ben's House's

Buying a house in King County, buying one with waterfront.

**03**



## Model Validation

Test-train-split-validation and the RMSE

**04**



## Feature Work

Low and high hanging features.

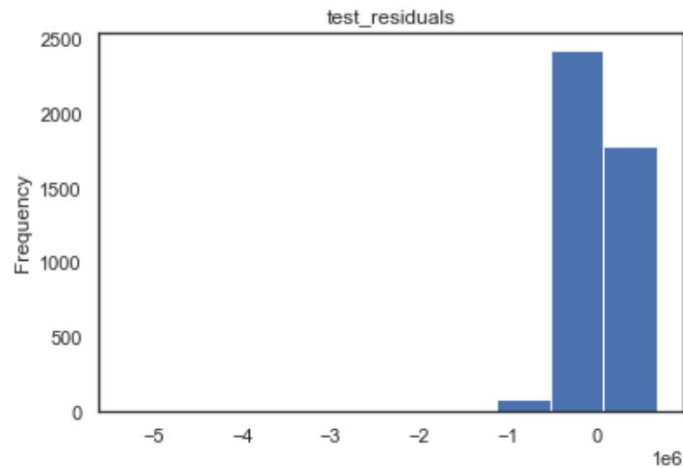
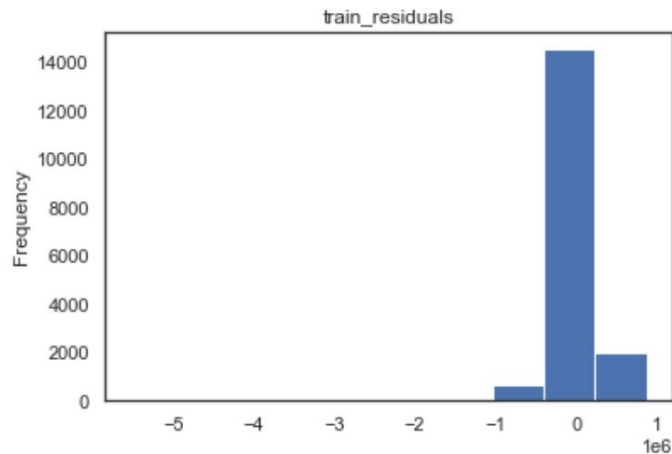
**X\_train: 17,277**

**X\_test: 4,320 ×**

```
sklearn.model_selection.train_test_split(test_size=0.2, random_state=42)
```

**Train RMSE: 239,942**  
**Test RMSE: 236,065<sup>x</sup>**

Root Mean Square Error is an absolute measure of fit,  
we hence talk \$\$\$.



**Train RMSE: 239,942**

**Test RMSE: 236,065** ✕

RMSE is an absolute measure of fit, we hence talk \$\$\$.



<b>Dep. Variable:</b>	price	<b>R-squared:</b>	0.699
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	✗ 0.699
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	3134.
<b>Date:</b>	Thu, 17 Sep 2020	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	09:55:15	<b>Log-Likelihood:</b>	-2.9442e+05
<b>No. Observations:</b>	21597	<b>AIC:</b>	5.889e+05
<b>Df Residuals:</b>	21580	<b>BIC:</b>	5.890e+05
<b>Df Model:</b>	16		
<b>Covariance Type:</b>	nonrobust		

<b>Omnibus:</b>	18683.039	<b>Durbin-Watson:</b>	1.982
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	2003719.883
<b>Skew:</b>	3.651	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	49.619	<b>Cond. No.</b>	3.03e+05

**Here are your  
stats!**

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	4.359e+06	1.15e+05	37.808	0.000	4.13e+06	4.59e+06
<b>has_basement[T.1]</b>	-1.091e+04	3215.445	-3.394	0.001	-1.72e+04	-4612.008
<b>waterfront[T.1.0]</b>	5.437e+05	2.04e+04	26.601	0.000	5.04e+05	5.84e+05
<b>view[T.1.0]</b>	9.548e+04	1.13e+04	8.436	0.000	7.33e+04	1.18e+05
<b>view[T.2.0]</b>	5.384e+04	6856.354	7.852	0.000	4.04e+04	6.73e+04
<b>view[T.3.0]</b>	1.264e+05	9346.662	13.522	0.000	1.08e+05	1.45e+05
<b>view[T.4.0]</b>	2.954e+05	1.42e+04	20.857	0.000	2.68e+05	3.23e+05
<b>renovation_status[T.latley renovated]</b>	9.92e+04	2.28e+04	4.346	0.000	5.45e+04	1.44e+05
<b>renovation_status[T.not renovated]</b>	-2.471e+04	1.87e+04	-1.320	0.187	-6.14e+04	1.2e+04
<b>renovation_status[T.renovated]</b>	-3.044e+04	2.12e+04	-1.435	0.151	-7.2e+04	1.11e+04
<b>bedrooms</b>	-2.721e+04	1854.485	-14.670	0.000	-3.08e+04	-2.36e+04
<b>lat_2_seattle</b>	-1.048e+06	1.8e+04	-58.301	0.000	-1.08e+06	-1.01e+06
<b>yr_built</b>	-2432.4269	61.366	-39.638	0.000	-2552.708	-2312.146
<b>floors</b>	6371.2361	3290.127	1.936	0.053	-77.657	1.28e+04
<b>sqft_living</b>	186.9392	3.106	60.188	0.000	180.851	193.027
<b>grade</b>	1.018e+05	2121.741	47.987	0.000	9.77e+04	1.06e+05
<b>sqft_living15</b>	11.3588	3.332	3.409	0.001	4.828	17.889

# TABLE OF CONTENTS



**01**

## King's County

Quick look at the dataset

**02**

## Ben's House's

Buying a house in King County, buying one with waterfront.

**03**

## Model Validation

Train-split-validation and the RMSE

**04**

## Feature Work

Low and high hanging features.



# Future Work ×

## Low hanging:

- How does the distance (as the crow flies) to a subway station correlates with the price?
- Do schools geo location influence the price?

## High hanging:

- time-to-center by public transportation, and car
- crime data per zipcode
- political data per zipcode

×

# End. Thanks.

×

