



House Sales in King County, Seattle, USA

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

TABLE OF CONTENTS



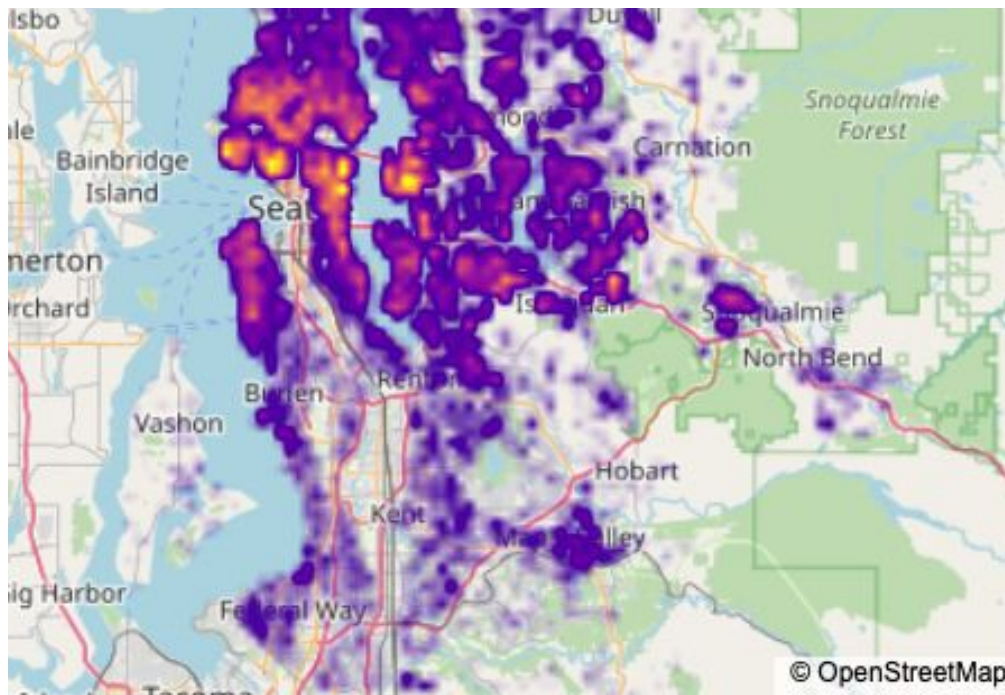
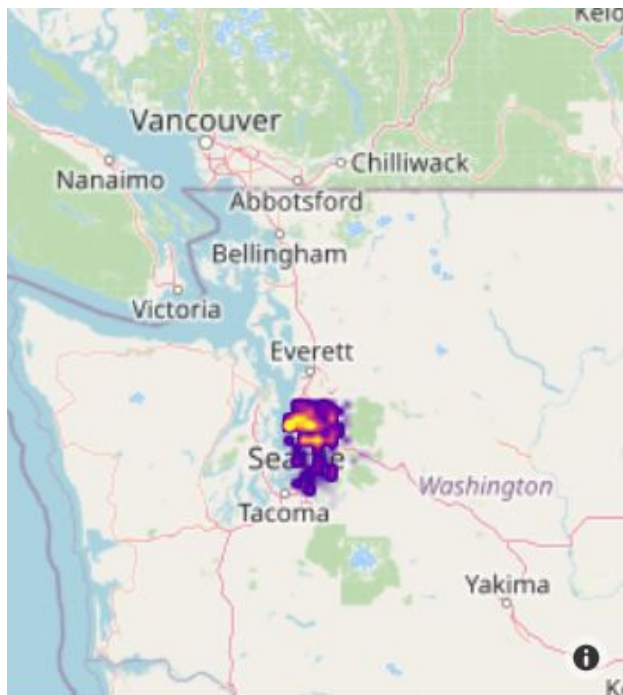
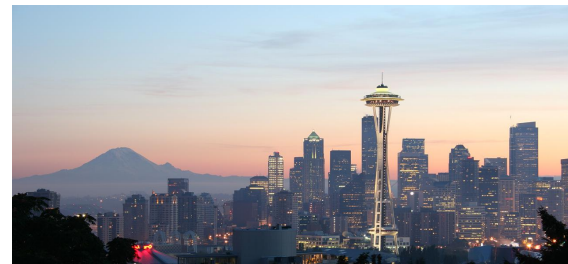
01 × **King's County**
A quick look at the dataset

02 × **Bens's Old House**
Selling a house in King County

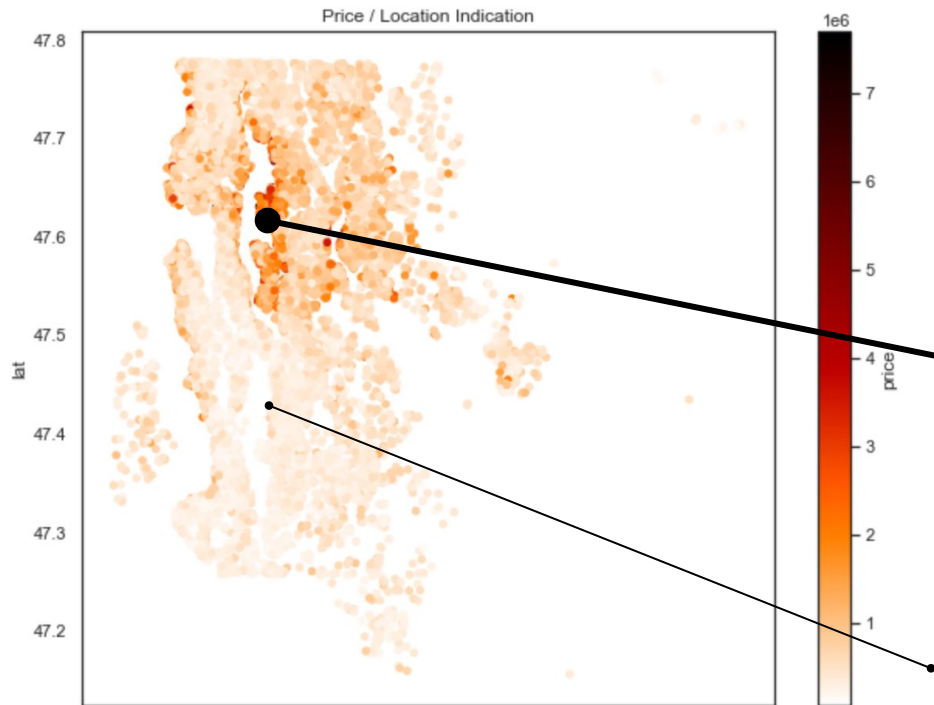
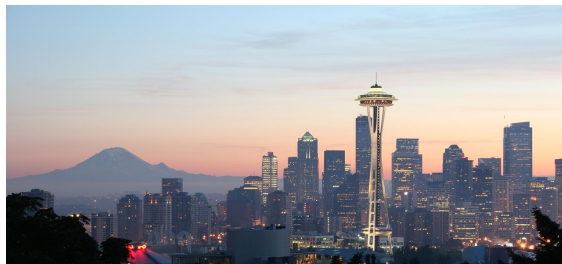
03 × **Karl's New House**
Waterfront, cheap, and available.

04 × **Model Validation**
Test-train-split-validation and the RMSE

King County, USA



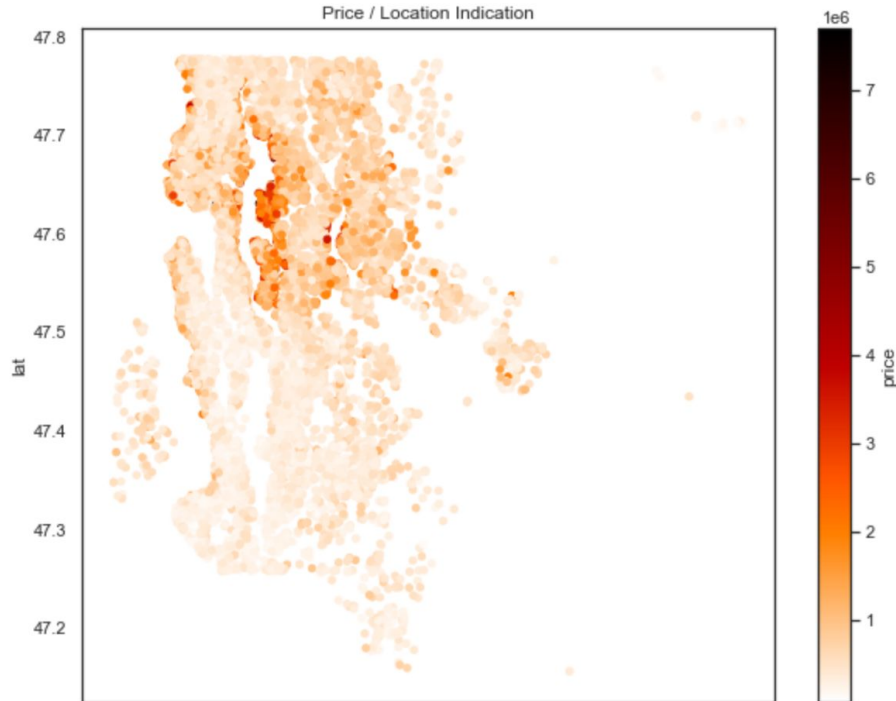
King County, USA



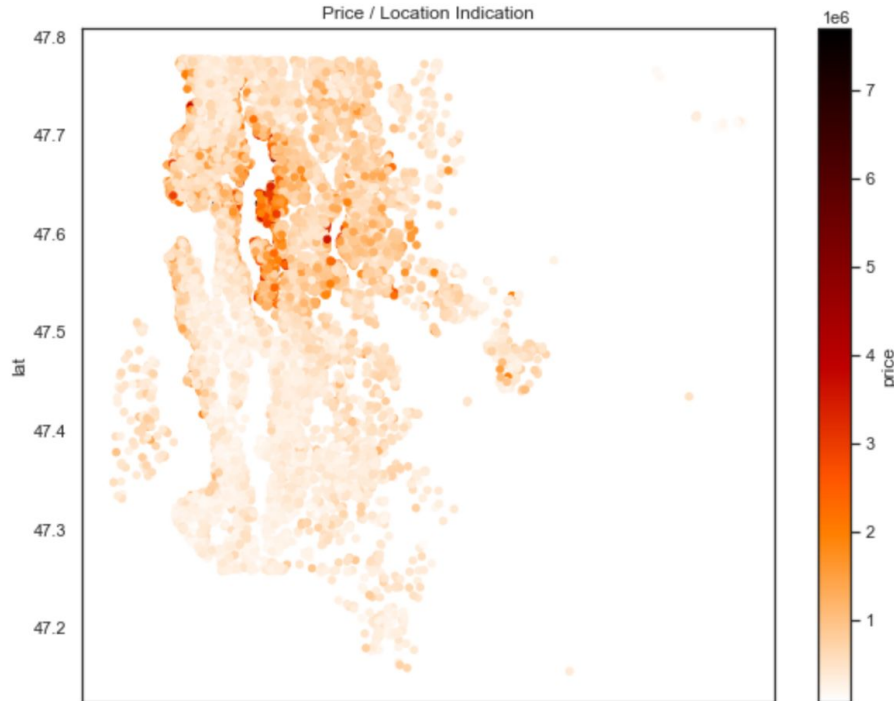
Is this the
middle of
Seattle?

Cheap
waterfront?
Sadly not.

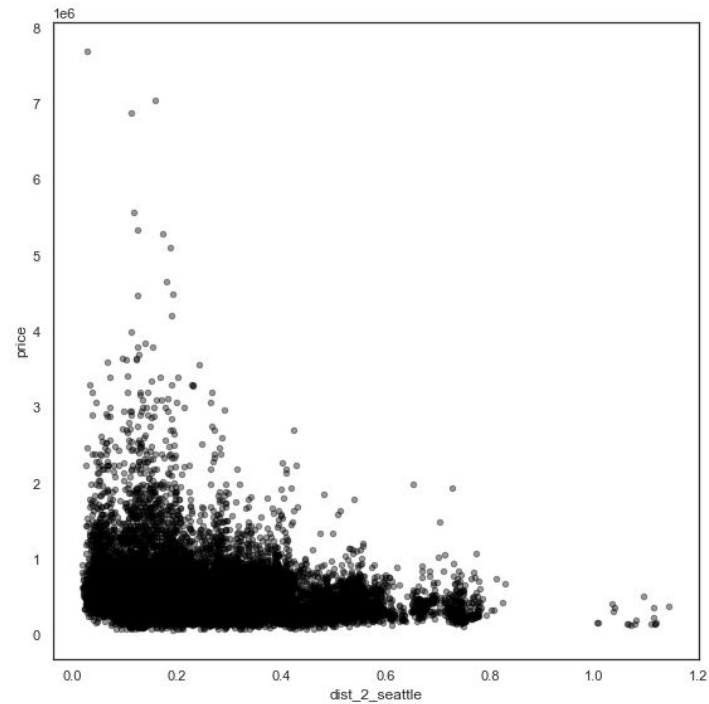
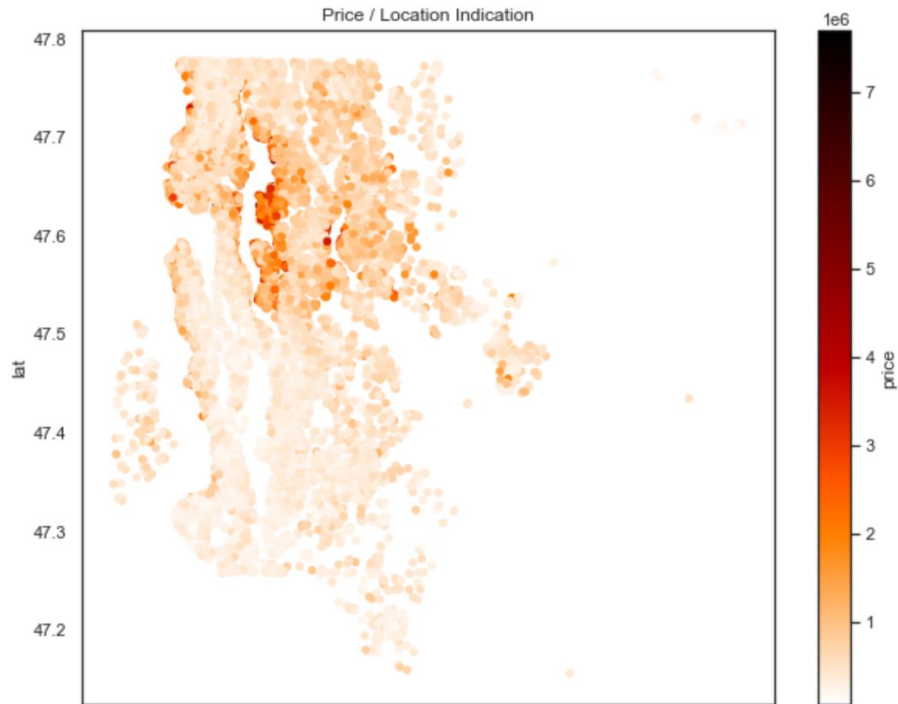
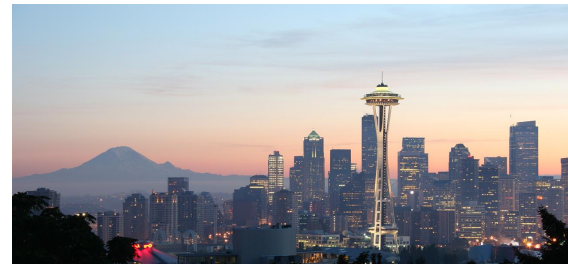
King County, USA



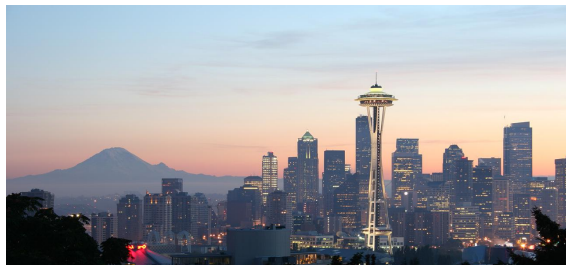
King County, USA



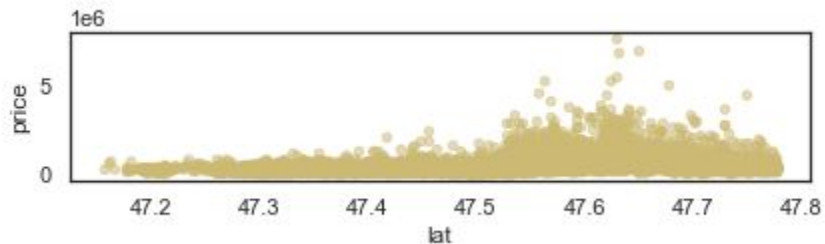
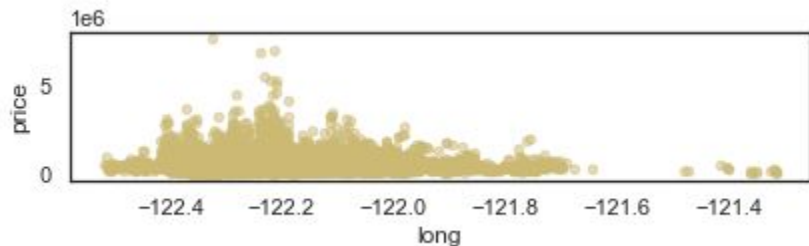
King County, USA



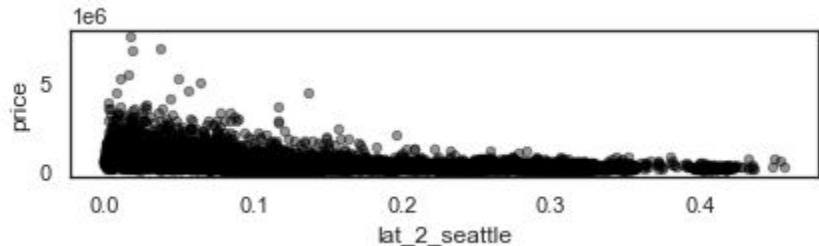
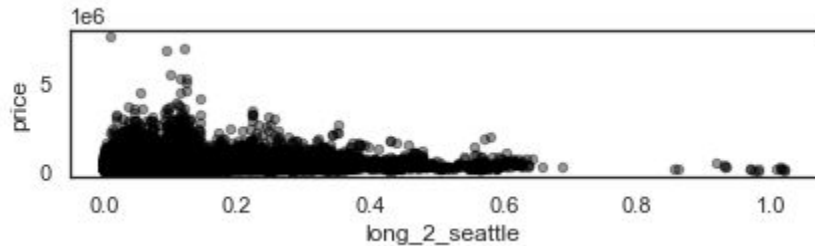
King County, USA



Absolut Geolocation on price



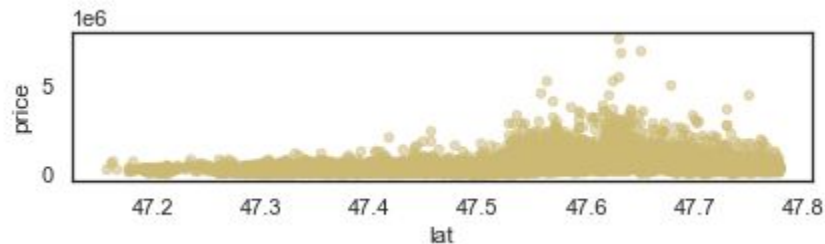
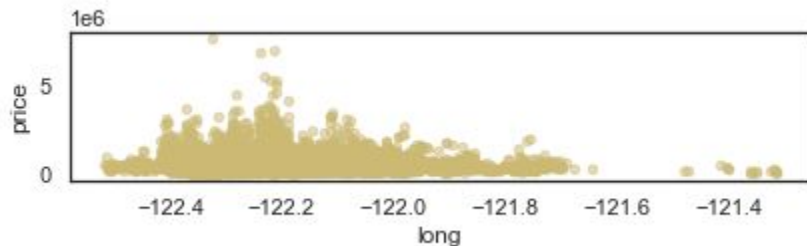
Geolocation relative to Seattle City Center on price



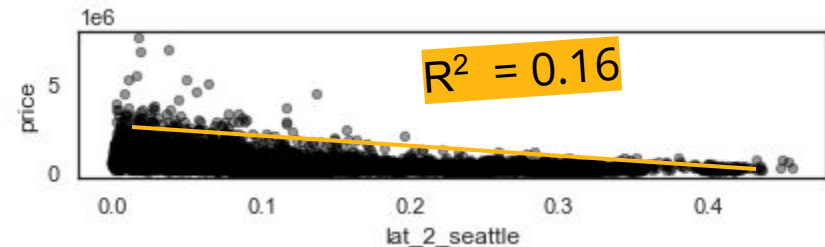
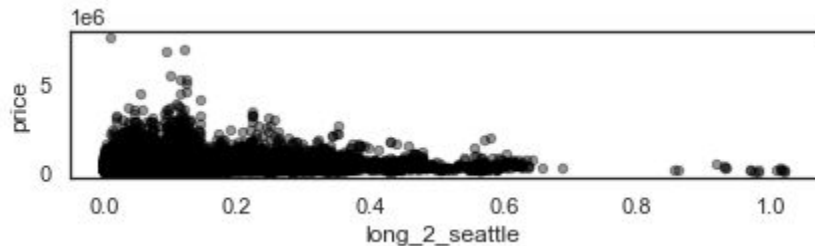
King County, USA



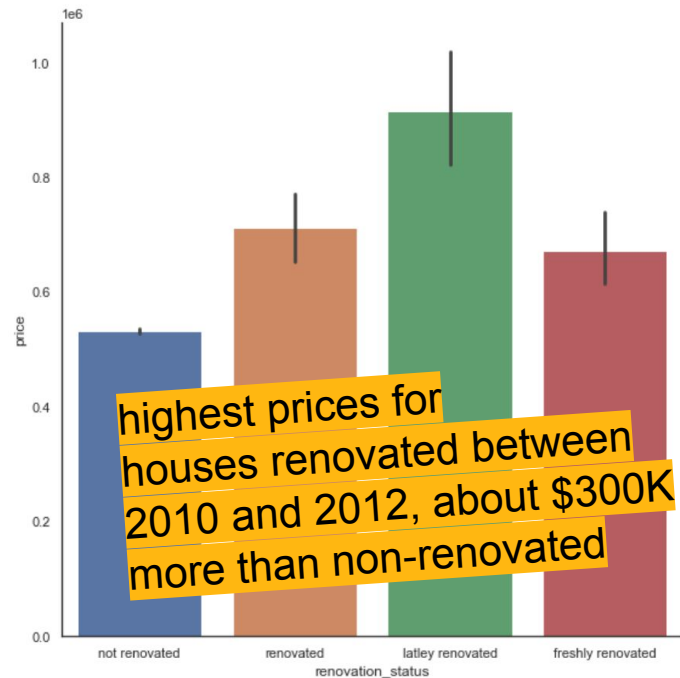
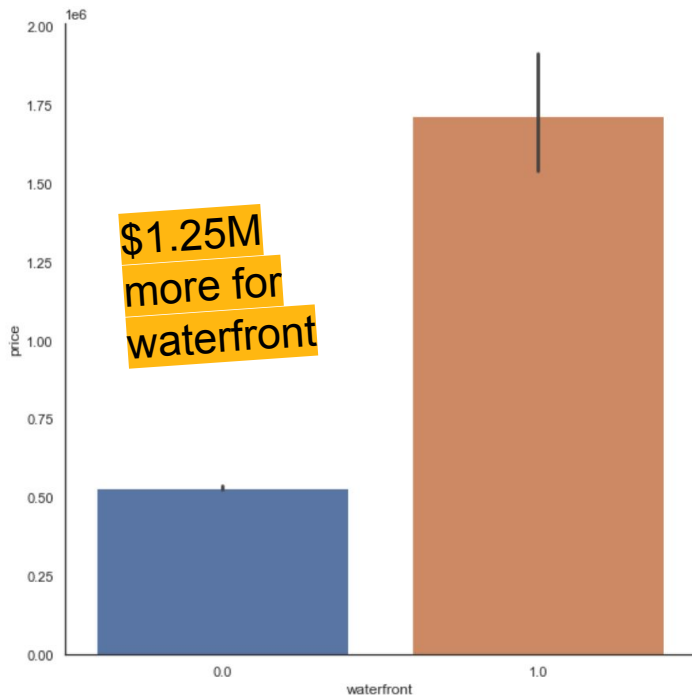
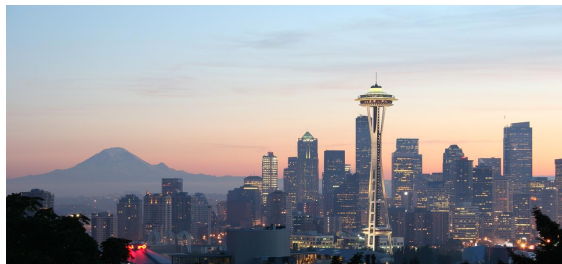
Absolut Geolocation on price



Geolocation relative to Seattle City Center on price



King County, USA



King County, USA

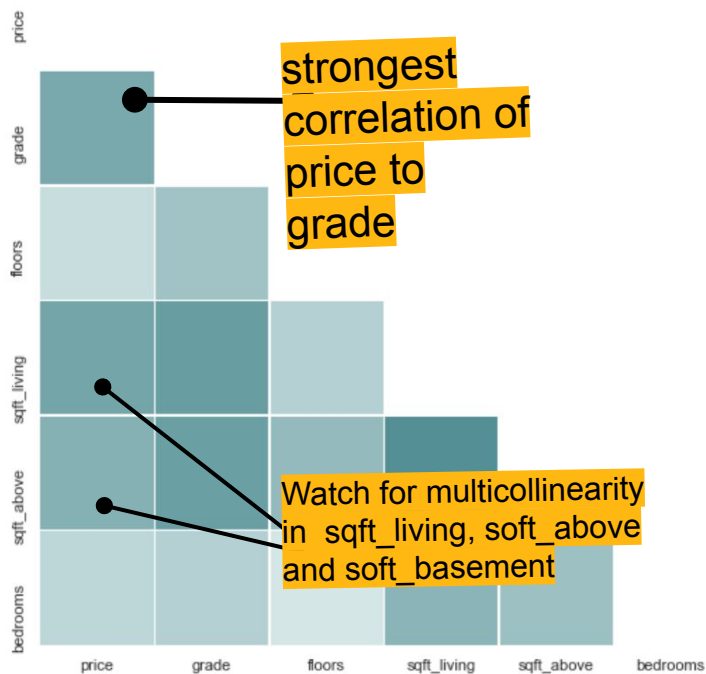
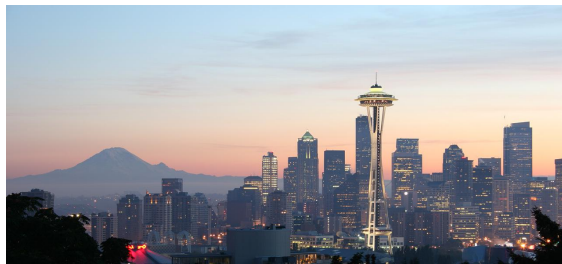


TABLE OF CONTENTS



01

King's County



Quick look at the dataset

02



Ben's Old House

Selling a house in King County

03



Karl's New House

Waterfront, cheap, and available.

04



Model Validation

Test-train-split-validation and the RMSE



Macklemore

Old friend, I need a new place. This whole home office thing makes me crazy. I live in Seattle, but I don't see the water. I live at 98004 right now, you know the place. It's run-down at the moment.

Make it cheap, but I want to move in ASAP.

Get back to me soon. B.

Old friend, I need a new place. This whole

{BS_indicator=True} {\$\$\$\$}

home office thing makes me crazy. I live in Seattle, but I

{hyper sensitive customer}

don't see the water. I live at 98004 right now,

{need: view = 1} {selling price?!}

you know the place. It's run-down at the moment.

{df[Bill's place']}

{optional: renovation}

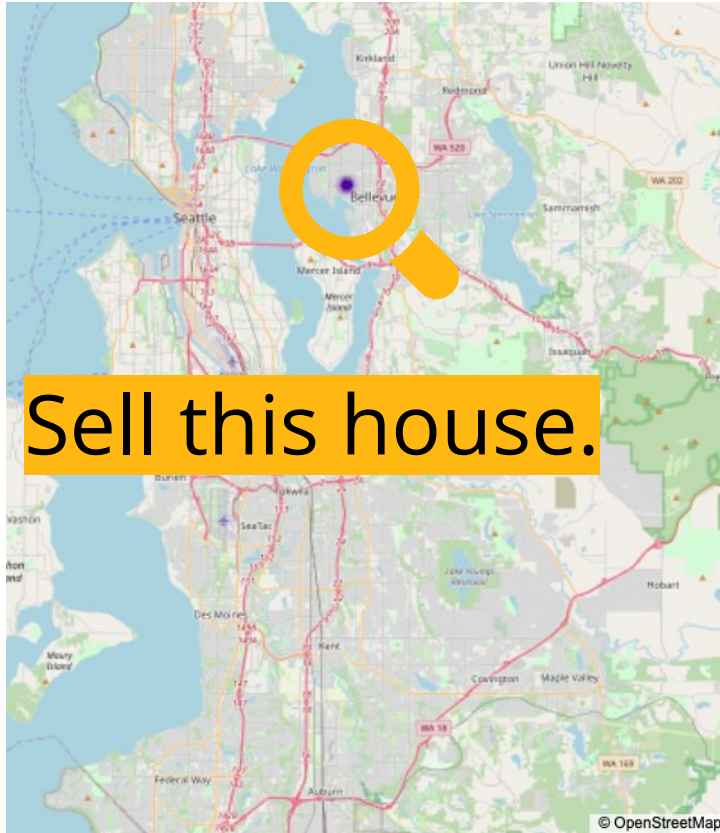
Make it cheap, but I want to move in ASAP.

\$\$\$\$: no

{need: high turnover rate}

Get back to me soon. B.

{deadline: 40h}





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

\$200,000.00

more for a freshly
renovated house.

TABLE OF CONTENTS



01



King's County

Quick look at the dataset

02



Ben's Old House

Selling a house in King County

03



Karl's New House

Waterfront, cheap, and available.

04



Model Validation

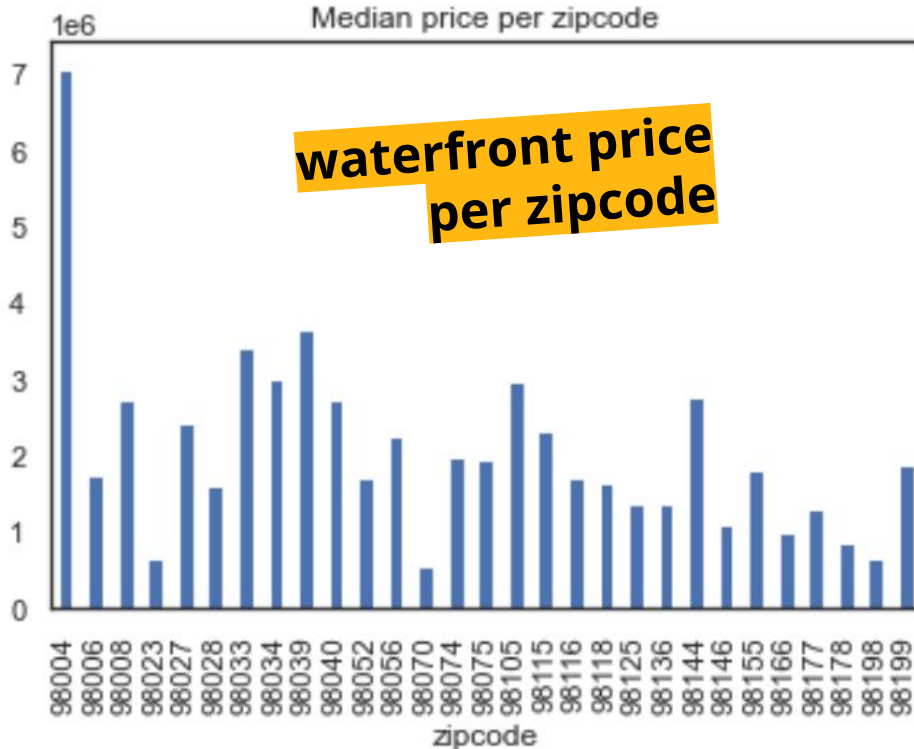
Test-train-split-validation and the RMSE



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.

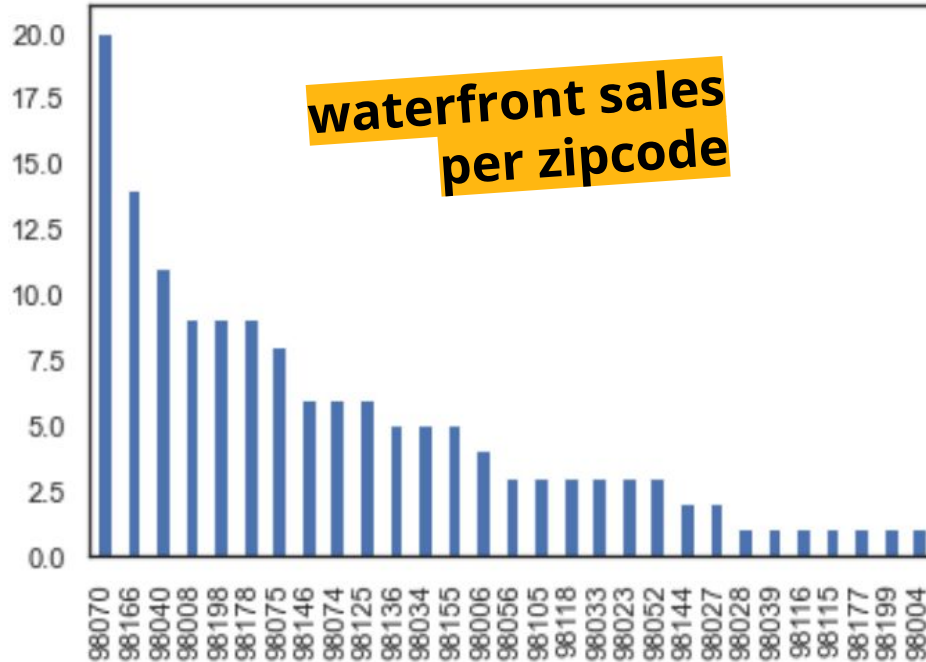
98070, 98166, 98040

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 \$200K. Not worth it. Look for a new place with waterfront in 98070, 98166, and 98040. Will cost you about \$1,500K. Or cheaper in 98200, 98075.

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 Old House



Selling a house in King County

03

Karl's New House



Waterfront, cheap, and available.

04



Model Validation

Test-train-split-validation and the RMSE

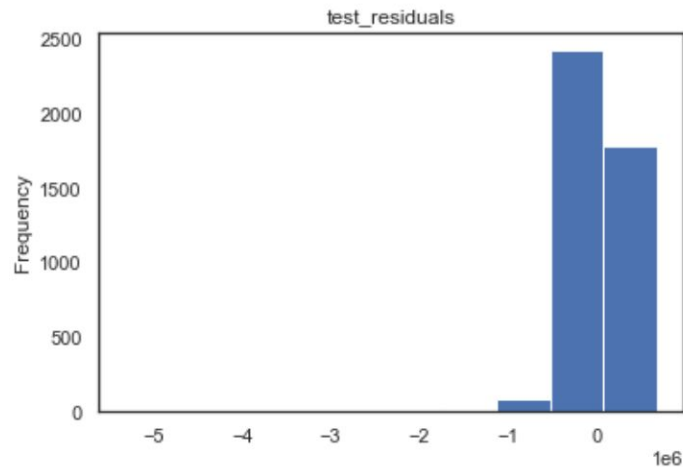
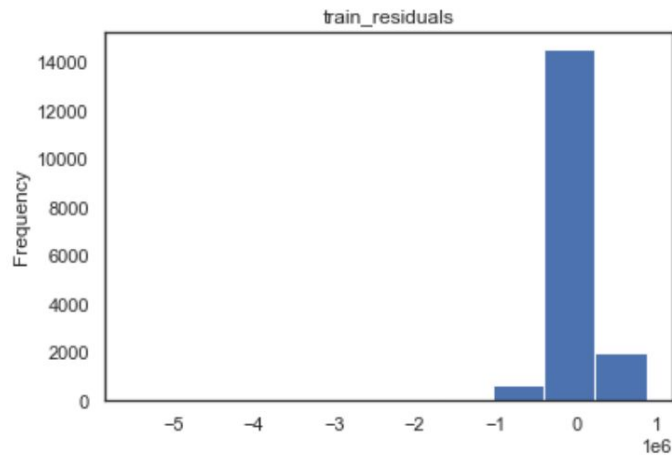
X_train: 17277

X_test: 4320 ✖

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


Train RMSE: 239,942
Test RMSE: 236,065 

RMSE 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 \$\$\$.

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

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

Give me the
stats



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



selling a house in King
County

03

Karl's New House



waterfront, cheap, and
available.

04

Model Validation



cross-train-split-validation
and the RMSE



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' feature
- crime data per zipcode
- Political data per zipcode

×

End. Thanks.

×

