

**HOUSE PRICE PREDICTION PROJECT**

# **DATA SCIENCE FOR HOUSING**

Presented by Seyma Tas



“

**“HOME IS WHERE LOVE  
RESIDES, MEMORIES ARE  
CREATED, FRIENDS ALWAYS  
BELONG, AND LAUGHTER  
NEVER ENDS.”**

# **PROBLEM STATEMENT**

*FAIR HOUSING IS NOT AN OPTION- IT IS THE LAW.*    **CARL HARRIS**

# Questions

- 1) For real estate agents: What are the key factors to predict the price of a house?
- 2) For home buyers: What is the most important factor determining the value of a house?
- 3) For home sellers: How can a house's value be increased?



# Data: King County Housing Dataset

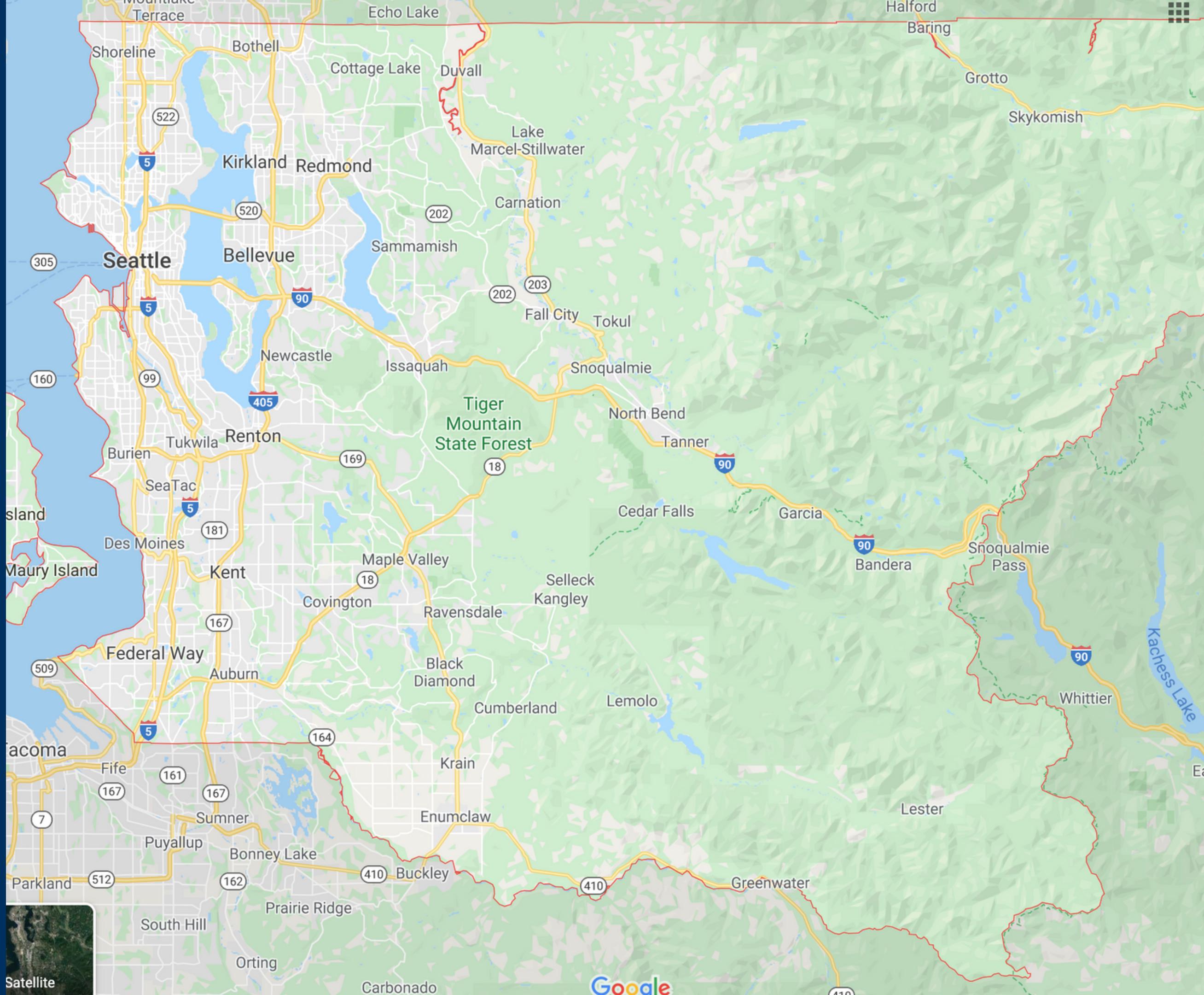


*The data is from the King county, Seattle, Washington.*

*The dataset contains the basic information about 21597 houses sold in 2014 and 2015.*

*Bedrooms, bathrooms, square footage, location, floors, view, condition are some of the given factors to guess the price.*

*I first explored the data then made a model. The model was %89 successful, I need more information to predict the price more efficiently.*



# METHODOLOGY

*TO PREDICT THE PRICE OF A HOUSE WITH THE GIVEN FACTORS*

# METHODOLOGY

## KEY FACTORS

The analysis of key factors determining the price of a house.

## DATA SCIENCE

Data Science techniques to answer business questions.

## ANSWER QUESTIONS

A question of each group; Home buyers, home sellers and real estate agents.

## VISUALIZATIONS

The result is explained by visualisations and maps.

# PARAMETERS IN THE MODEL

LOCATION

SQUARE FOOTAGE OF THE HOUSE  
(BASEMENT, LOT AND LIVING AREA)

NUMBER OF BATHROOMS

NUMBER OF BEDROOMS

GRADE OF THE CONSTRUCTION QUALITY

CONDITION OF RENOVATIONS

THE MONTH HOUSE WAS SOLD

VIEW OF THE HOUSE



# **RECOMMENDATIONS**

*FOR BUYERS, SELLERS AND REAL ESTATE AGENTS*



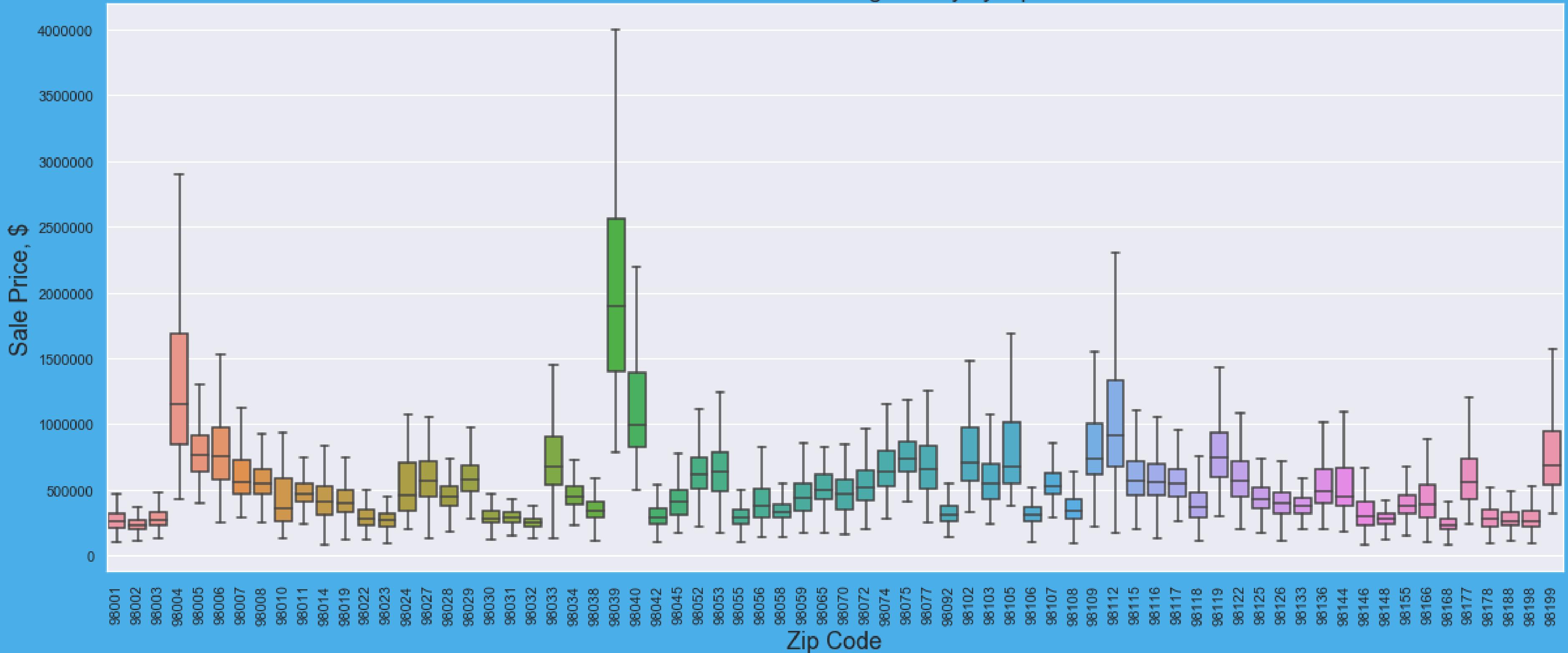
# Recommendations

## 1) For home buyers;

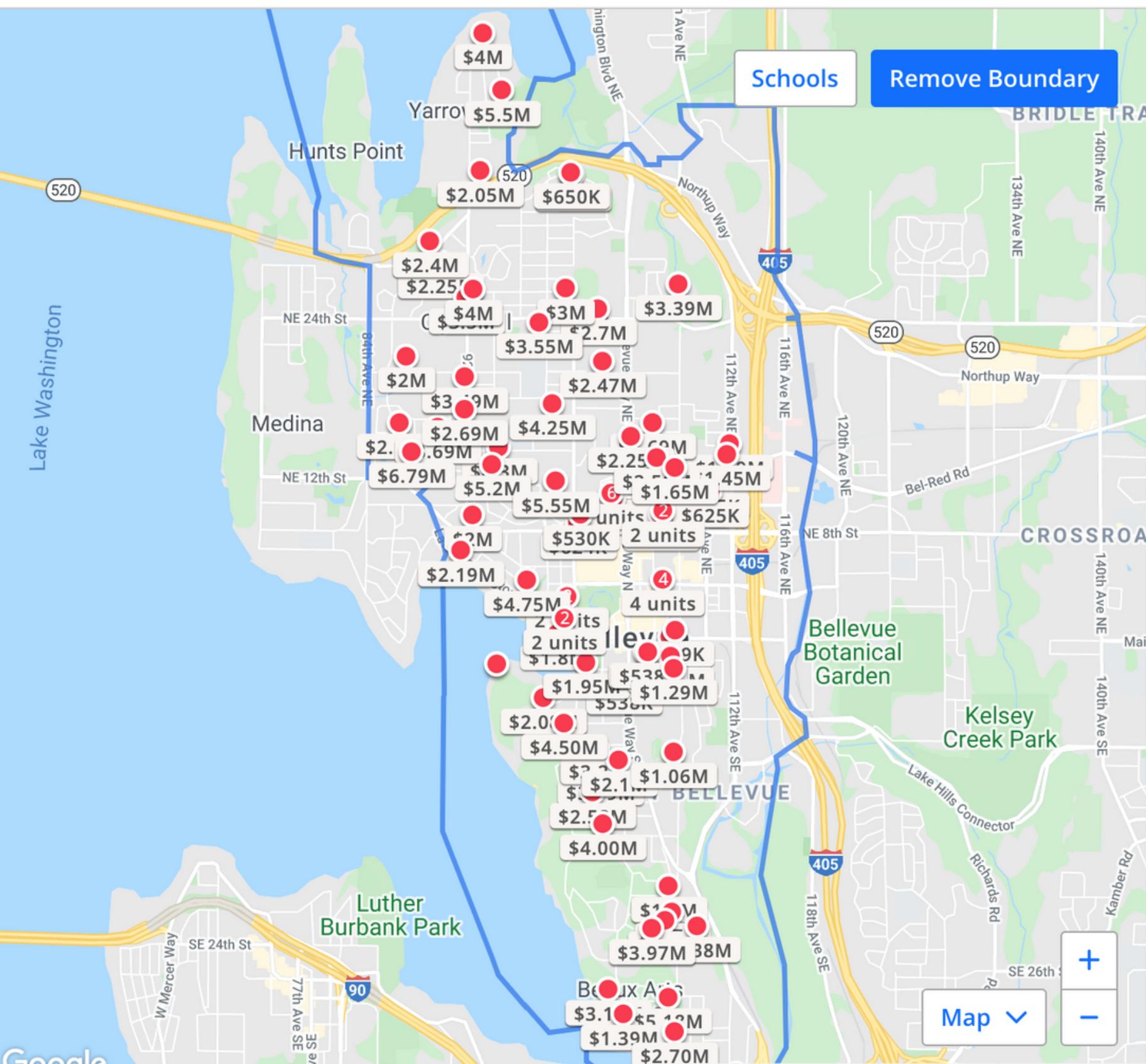
The most important factor determining the value of a house is LOCATION! Because you can do many renovations and upgrades in a house but you can not change the location of it. A good view or waterfront may double the house's price.

Second important factor you can change very difficult is square footage of living area. More rooms mean more value if the location is correct.

## Prices of Houses Sold in King County by Zip Code



 For Sale Price Beds & Baths Home type More Save search



## 98004 Real Estate & Homes For Sale

88 results

Sort by: [Homes for You](#)



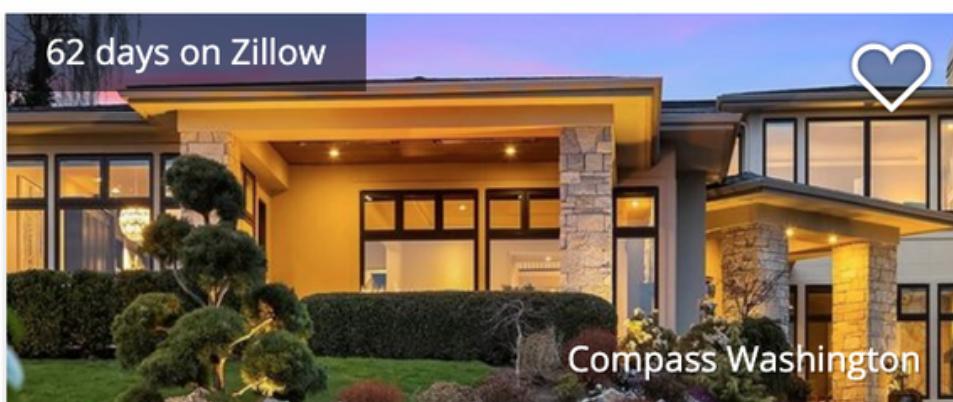
**\$3,970,000** 5 bds | 6 ba | 5,702 sqft  
10460 SE 23rd St, Bellevue, WA 98004  
● House for sale



**\$1,060,000** 3 bds | 2 ba | 1,180 sqft  
814 107th Ave SE, Bellevue, WA 98004  
● For sale by owner



**\$1,650,000** 3 bds | 1.75 ba | 1,510 sqft  
10632 NE 12th Pl, Bellevue, WA 98004  
● House for sale



**\$5,549,000** 5 bds | 5 ba | 5,259 sqft  
9867 Vineyard Crst, Bellevue, WA 98004  
● House for sale



Price cut: \$25,000 (Apr 22)

Mercer Island, WA



For Sale

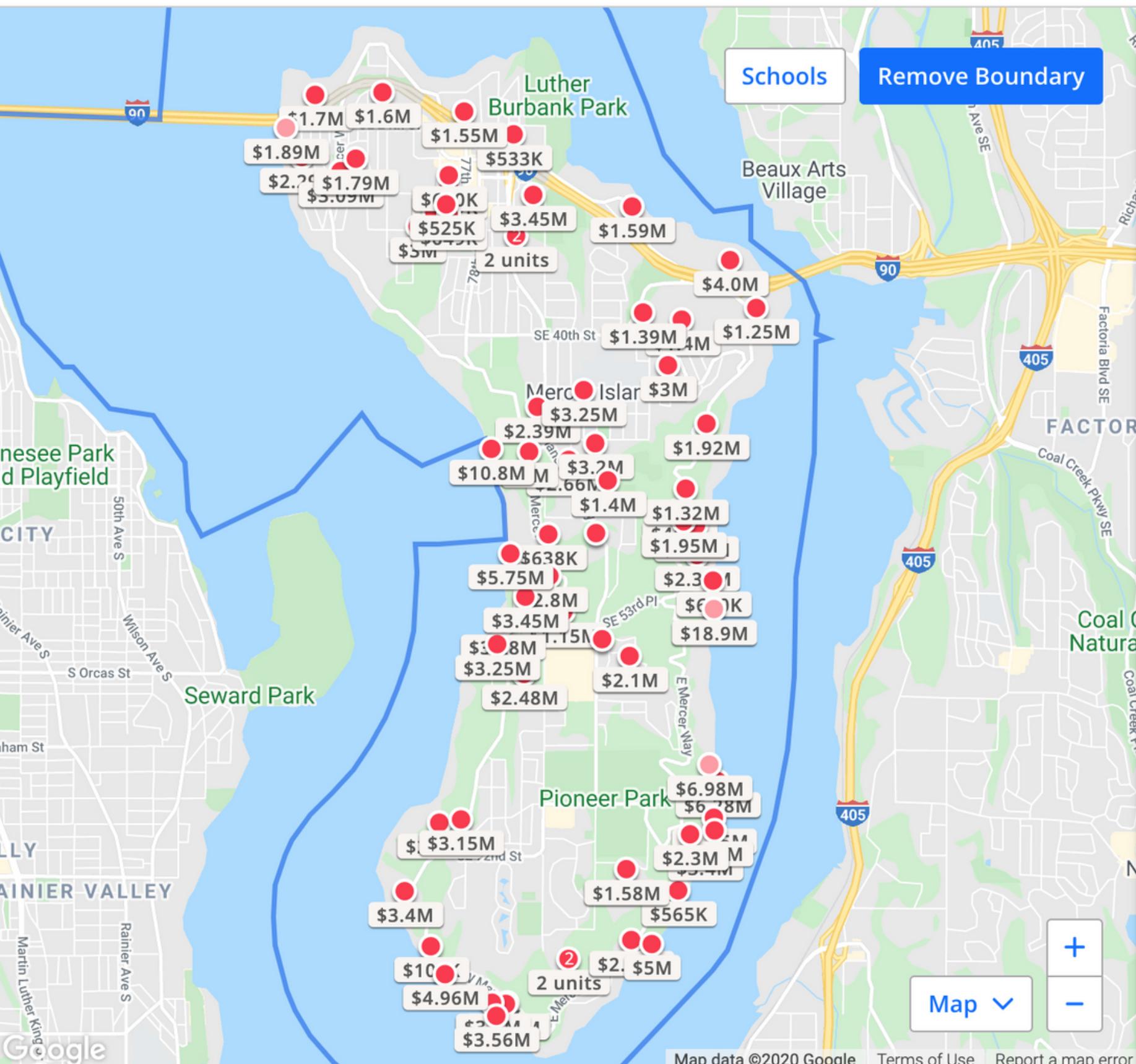
Price

Beds & Baths

Home type

More

Save search



## 98040 Real Estate & Homes For Sale

71 results

Sort by: [Homes for You](#)



**\$18,888,000** 5 bds | 11 ba | 14,940 sqft  
5330 Butterworth Rd, Mercer Island, WA 98040

● House for sale



**\$1,885,000** 4 bds | 3.5 ba | 5,350 sqft  
2403 60th Ave SE, Mercer Island, WA 98040

● House for sale



**\$1,949,950** 4 bds | 5 ba | 3,079 sqft  
4709 E Mercer Way, Mercer Island, WA

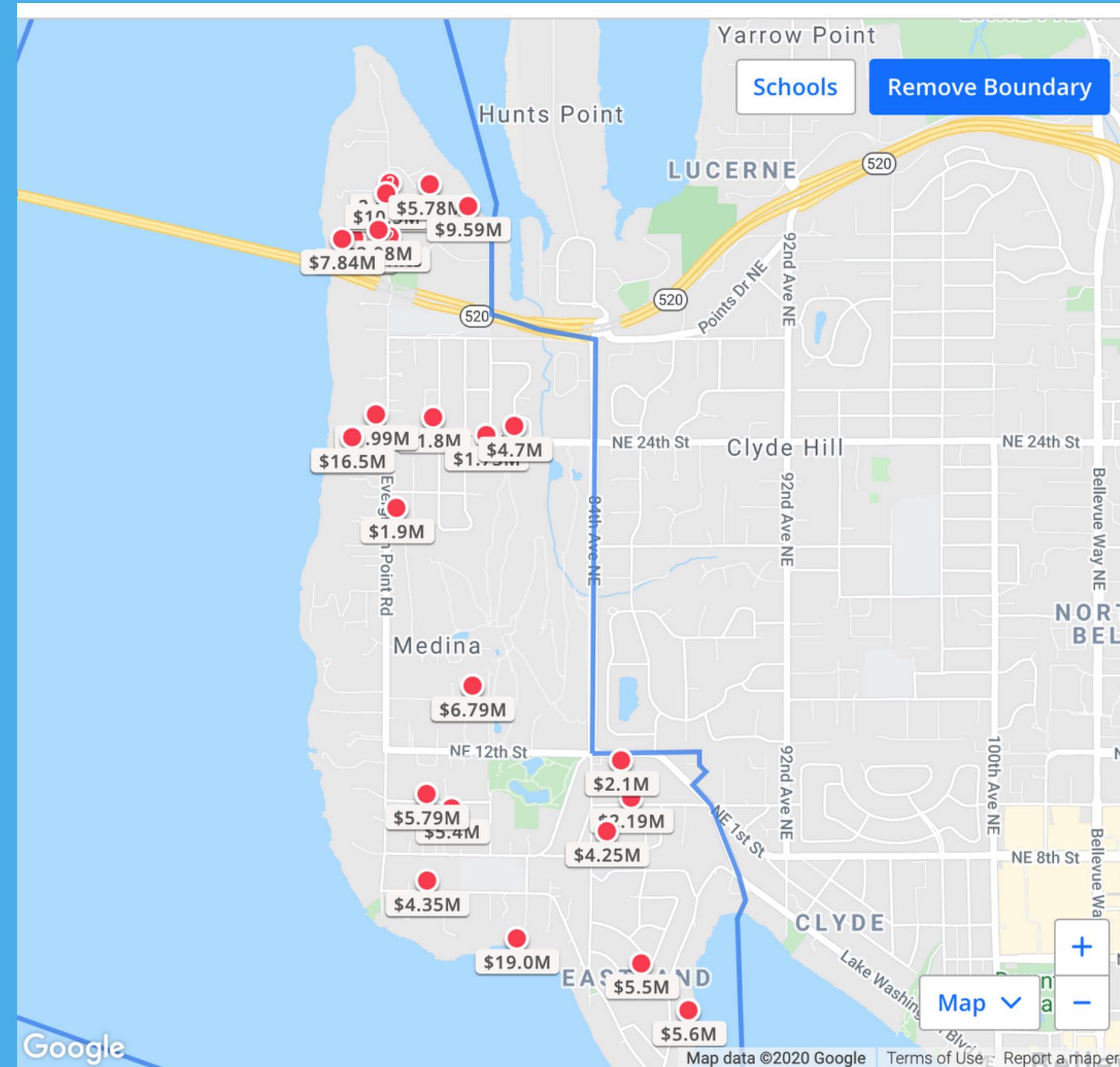
● New construction



**\$1,399,000** 3 bds | 3 ba | 2,710 sqft  
4533 89th Ave SE, Mercer Island, WA 98040

● For sale by owner





98039 Real Estate & Homes For Sale

27 results

Sort by: [Homes for You](#) ▾



**\$5,780,000**      5 bds | 7 ba | 5,854 sqft  
3312 78th Pl NE, Medina, WA 98039

● House for sale



**\$19,000,000** 6 bds | 5 ba | 7,150 sqft  
7935 Overlake Dr W, Medina, WA 98039

● House for sale



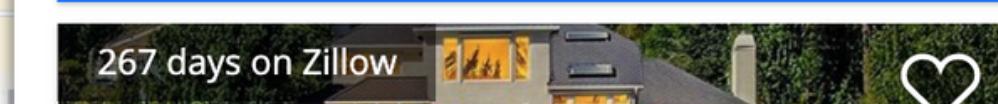
**\$16,500,000**

4 ● Houses for sale



**\$4,700,000**

#### ● New construction



54 days on Zillow

Seattle, WA



For Sale

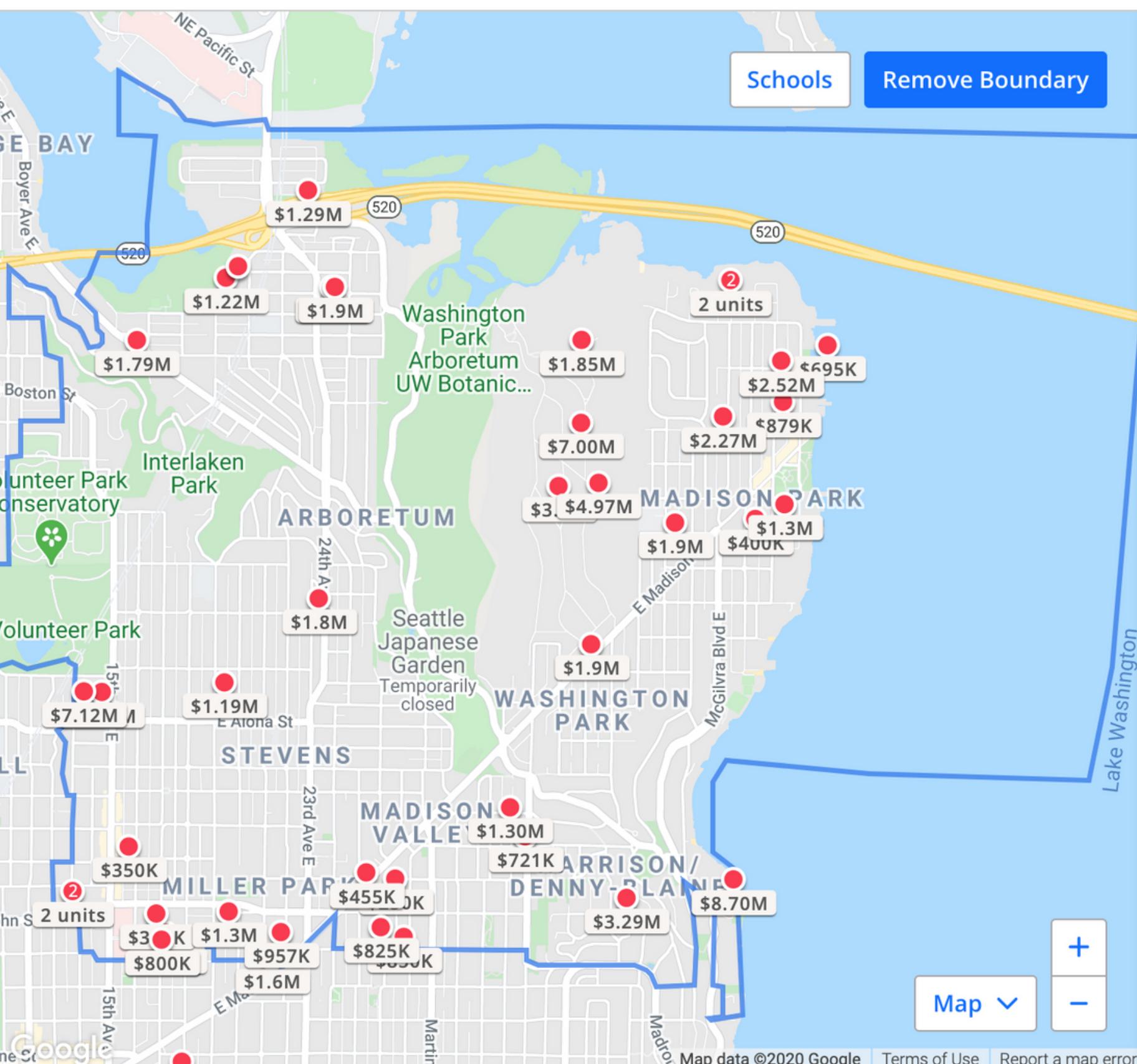
Price

Beds & Baths

Home type

More

Save search



## 98112 Real Estate & Homes For Sale

44 results

Sort by: Homes for You

Price cut: \$100,000 (Feb 11)

**\$1,800,000** 5 bds | 6 ba | 5,456 sqft  
1228 23rd Ave E, Seattle, WA 98112  
● House for sale

Price cut: \$45,000 (Apr 28)

**\$1,849,000** 3 bds | 3 ba | 3,130 sqft  
3356 E Shore Dr, Seattle, WA 98112  
● House for sale

46 days on Zillow

**\$7,000,000** 5 bds | 7 ba | 8,030 sqft  
2001 Broadmoor Dr E, Seattle, WA 98112  
● House for sale

4 days on Zillow

**\$1,300,000** 3 bds | 3 ba | 2,330 sqft  
3024 E Republican St, Seattle, WA 98112  
● House for sale

Kent, WA



For Sale

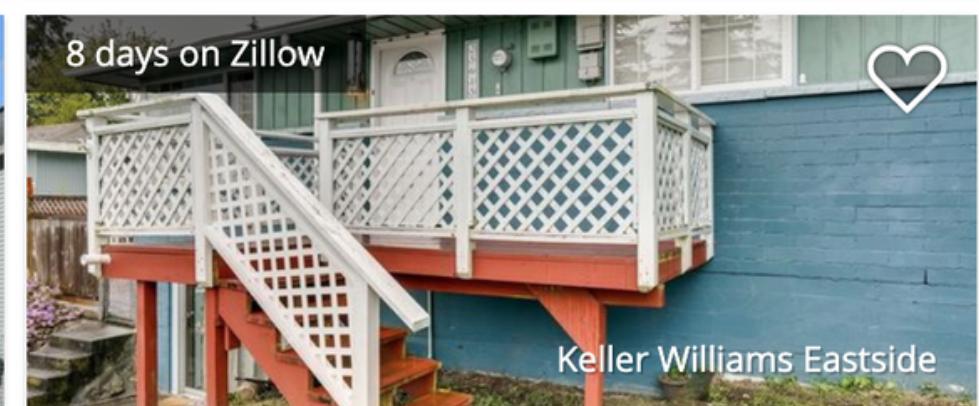
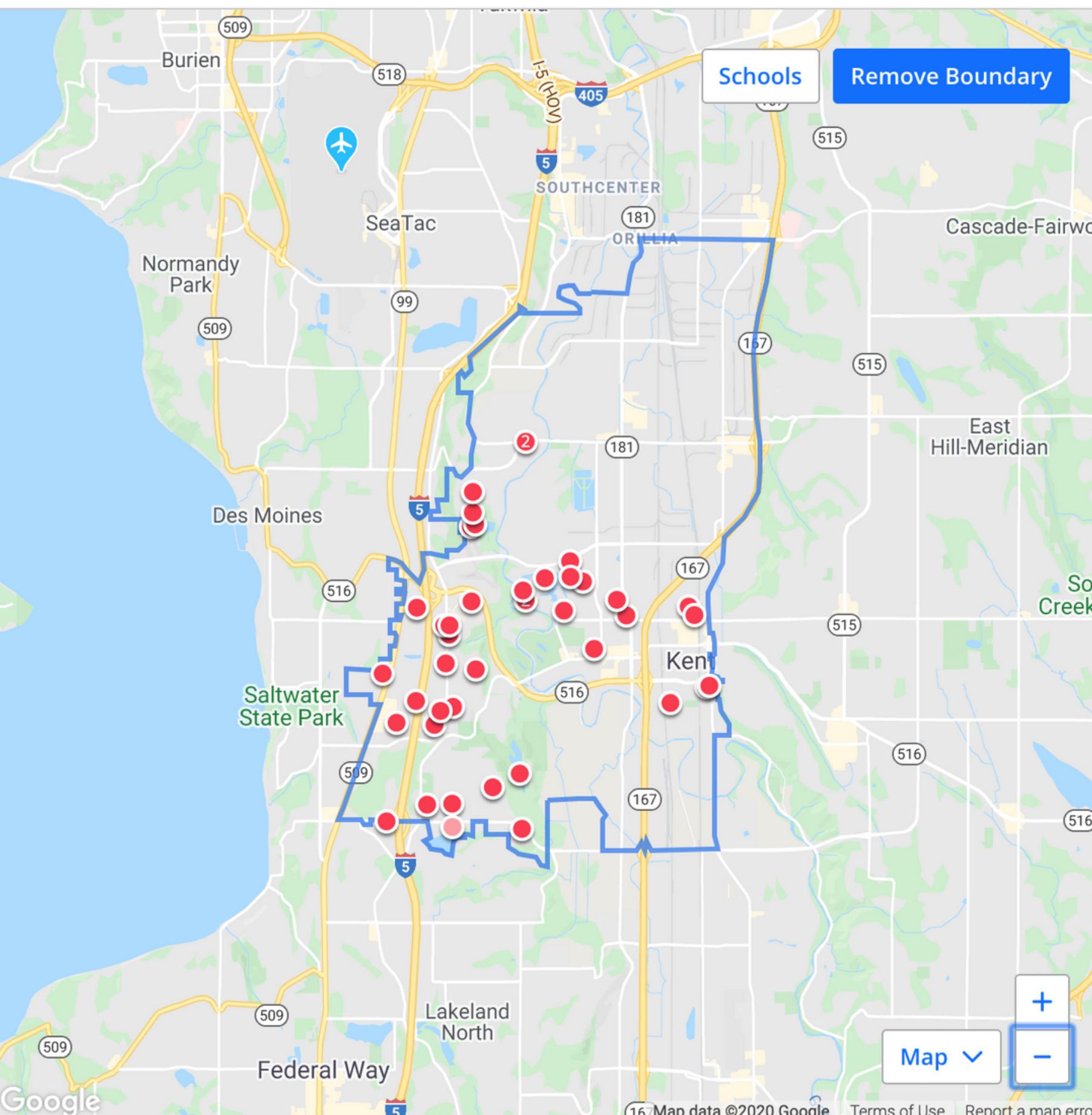
Price

Beds & Baths

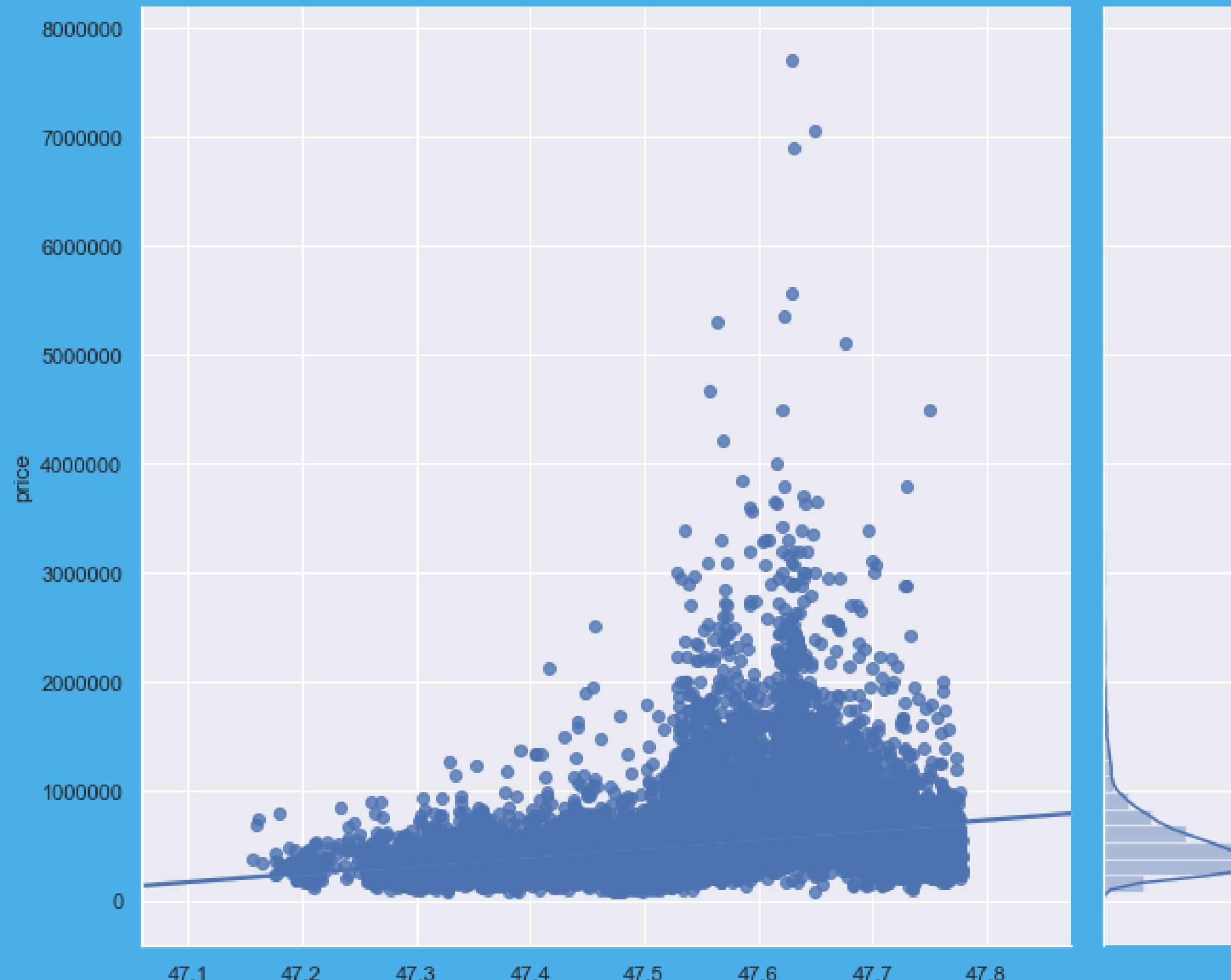
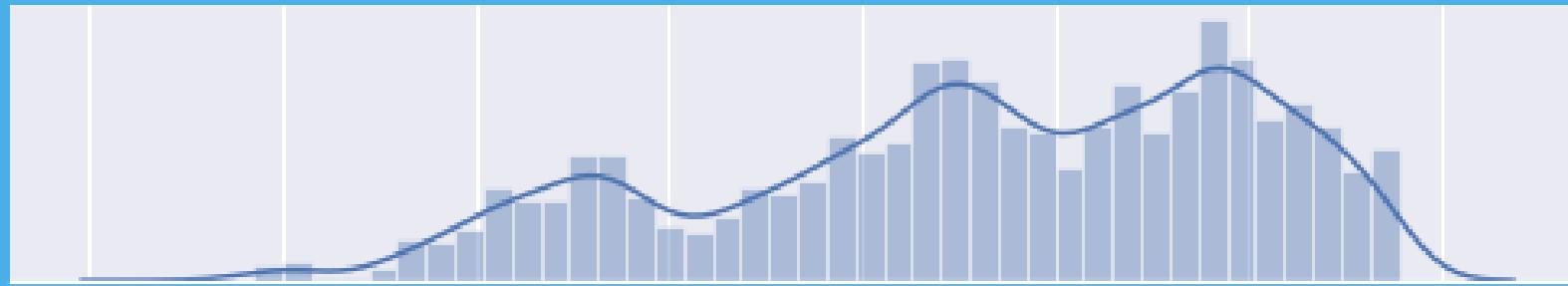
Home type

More

Save search



# PRICE VS. LATITUDE SCATTER REGRESSION PLOT



**We can make two inferences from this graph:**

- i) North of the city is more expensive than the South.
- ii) More houses are sold in the North.

There can be two reasons for this:

- 1) There are more houses in the North and/or
- 2) There is a growing housing market in the North. The city is expanding to the North.

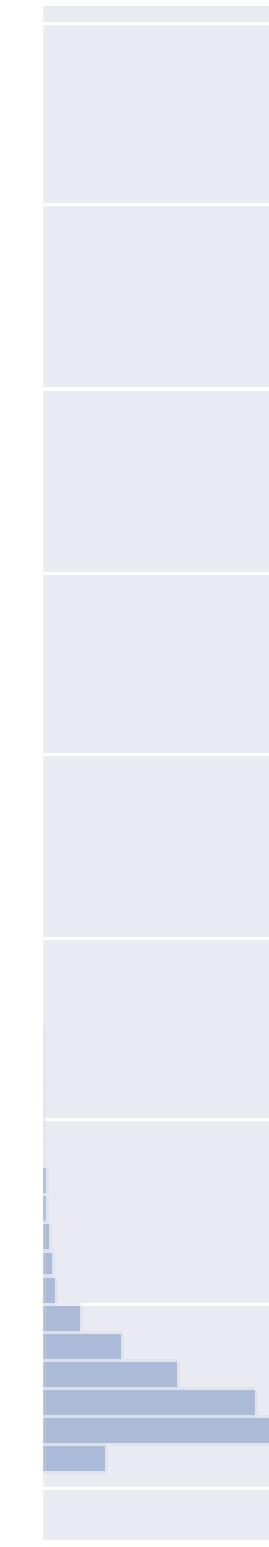
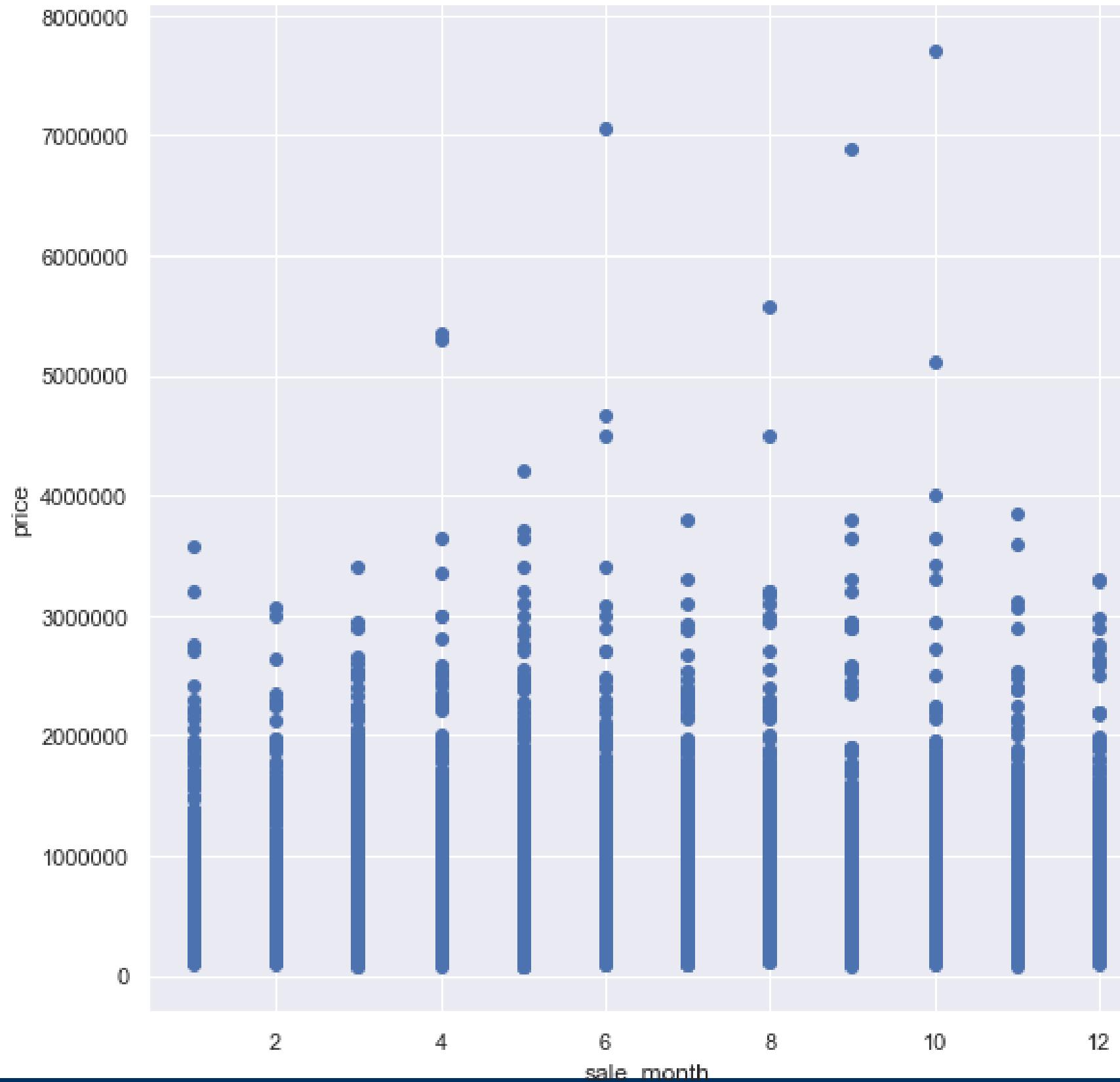
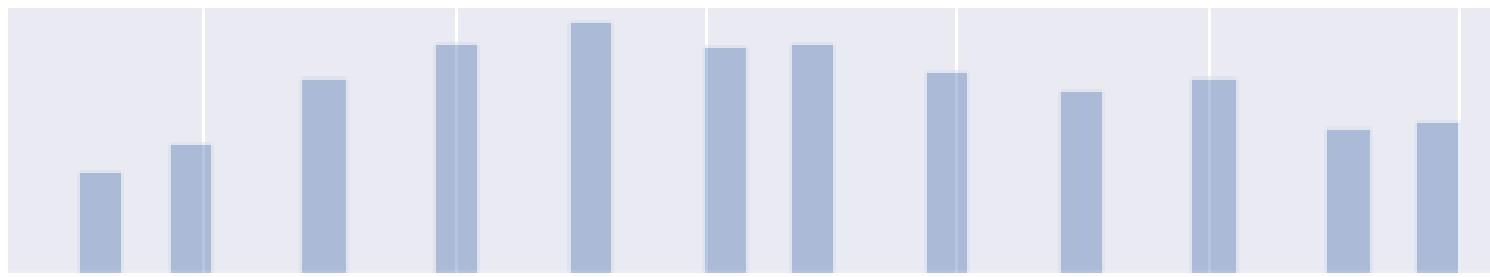
A close-up photograph of an architectural blueprint showing floor plans and elevation details. A claw hammer with a blue handle lies diagonally across the top left corner of the blueprint, symbolizing home improvement or construction.

# Recommendations

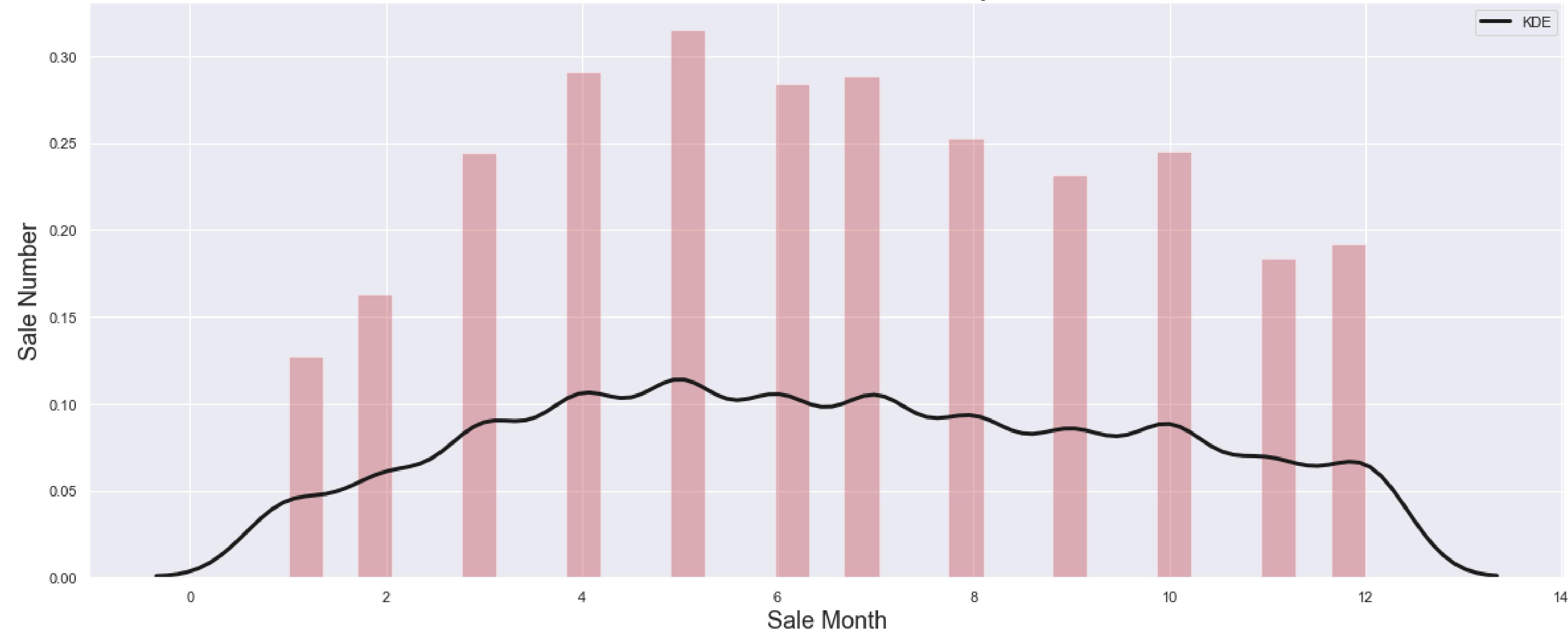
## 2) For home sellers;

Making small updates can increase the price of your house more than you expected.

Sell your house at the right month.



The Distribution of Houses Sold Every Month





# Recommendations

## 3) For real estate agents;

Giving the right price to a house decreases the DOM(days on the market) and makes you work efficiently. My model is 89% predictive. I need more information about houses to make the model efficient.

# FUTURE WORK

*MAKE THE MODEL MORE EFFICIENT*

# Future Work



## Zip code

I didn't use zip code information. I used zip code as a proxy in my model. I had only 70 different zip codes. I could make one hot encoding to these 70 columns. If this model is enlarged, there are going to be thousands of zip codes. We need to use the information included in the zip codes instead of making categorical columns. Zip code contains detailed information about income, house values, rental houses, demographics, education level of the residents, school grades and location. More information can be obtained from the zip code and can be used in the predictions.

# Future Work

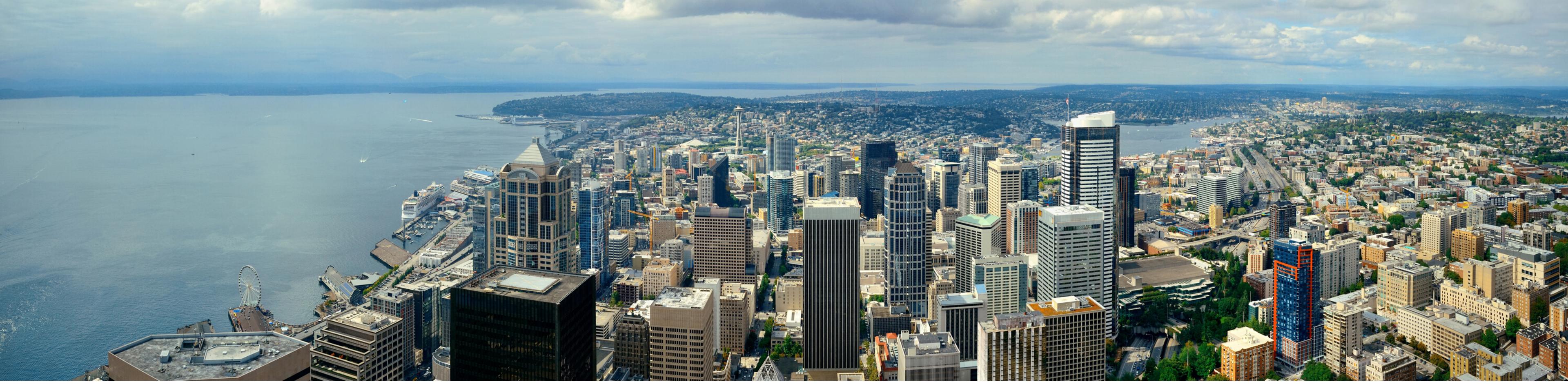


## *More detailed dataset*

I have 20 factors to predict the value of houses. I need approximately 100 predictors to make the house price algorithm work more properly. If I had (at least)garage, garden and pool information I would make more efficient predictions.



# Questions and comments



**Thank you!**