

EDA Project

House Prices in King County, USA

Seda Acar, 16.10.2023

Outline

- Research Questions & Hypotheses
- Data Cleaning
- Client
- Data Exploration
- Results

3 Research Questions & Hypotheses

1. Impact of living space on price?

- Expected: the bigger the sqft area, the higher the price

2. Impact of waterfront on price?

- Expected: if waterfront, then higher price

3. Influence of location on house price?

- Expected:
 - Higher avg house prices for specific neighborhoods/regions
 - if located to water, then higher prices

—> Higher avg prices the more to the west

Client

Jennifer Montgomery

- High budget
- Wants to show off
- Timing within a month
- Waterfront
- Renovated
- High grades
- Resell within 1 year

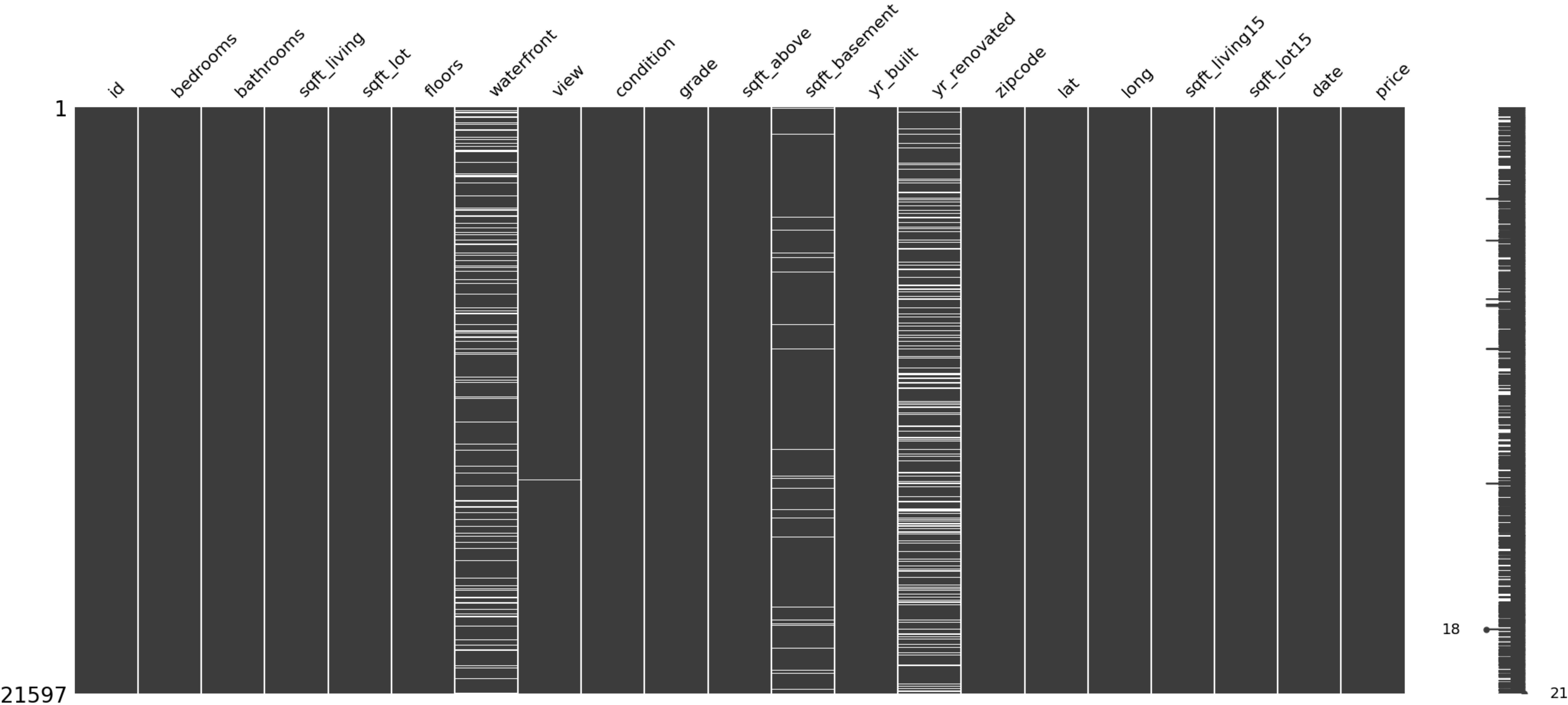
Data Cleaning

Check for ...

- **..Duplicates**
 - `df_kch.duplicated().value_counts()`
- **..Missing values with missingno**
 - `df_kch.isna().sum()`
 - `msno.matrix(df_kch)`

Missing Values

`msno.matrix(df_kch) —> imputed with 0`



Data Adjustments

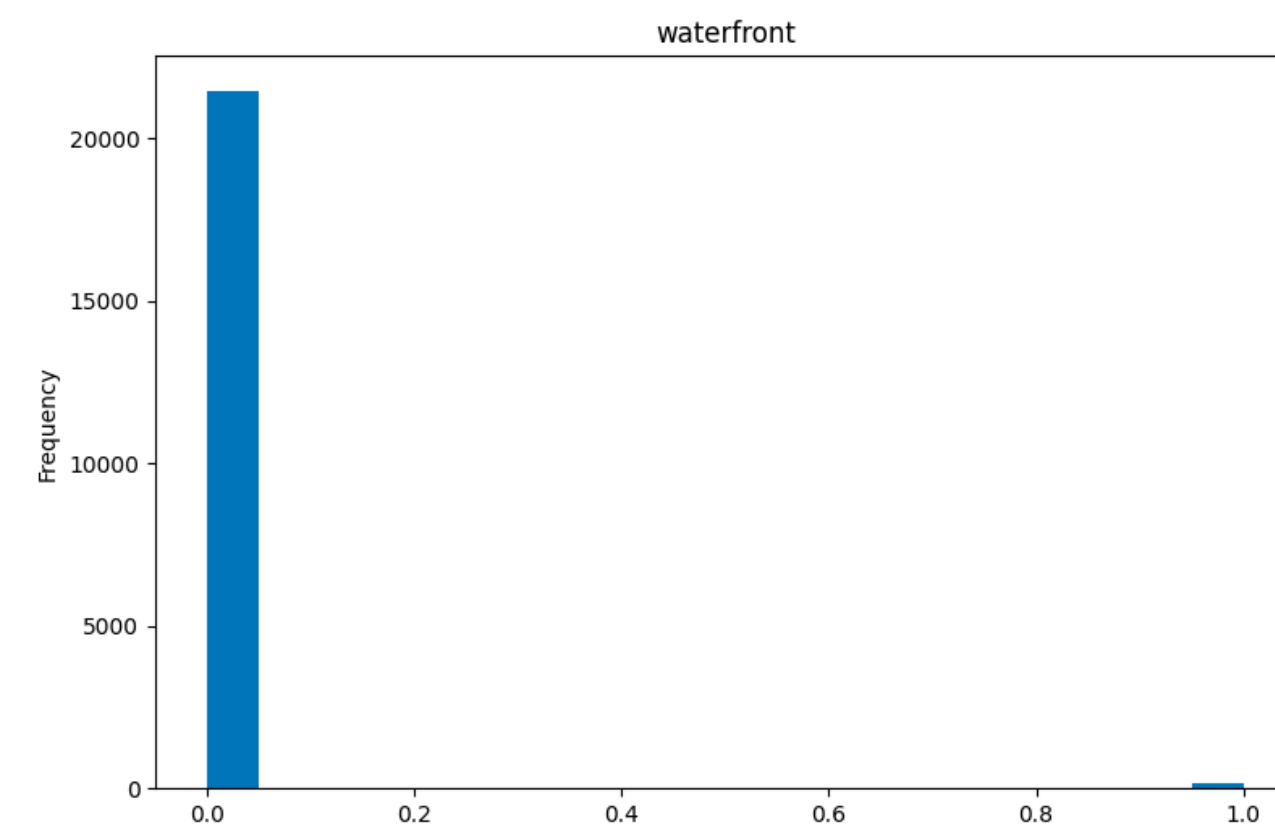
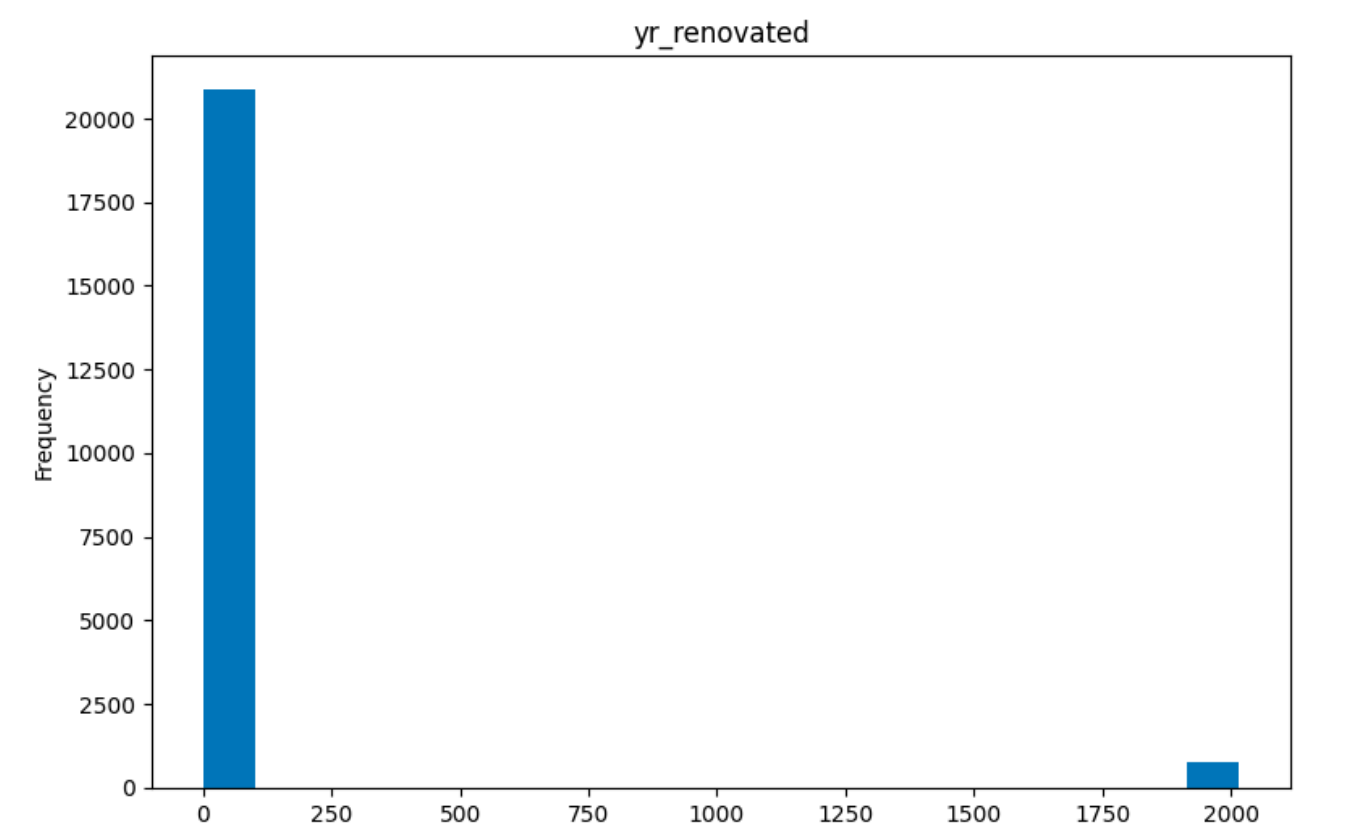
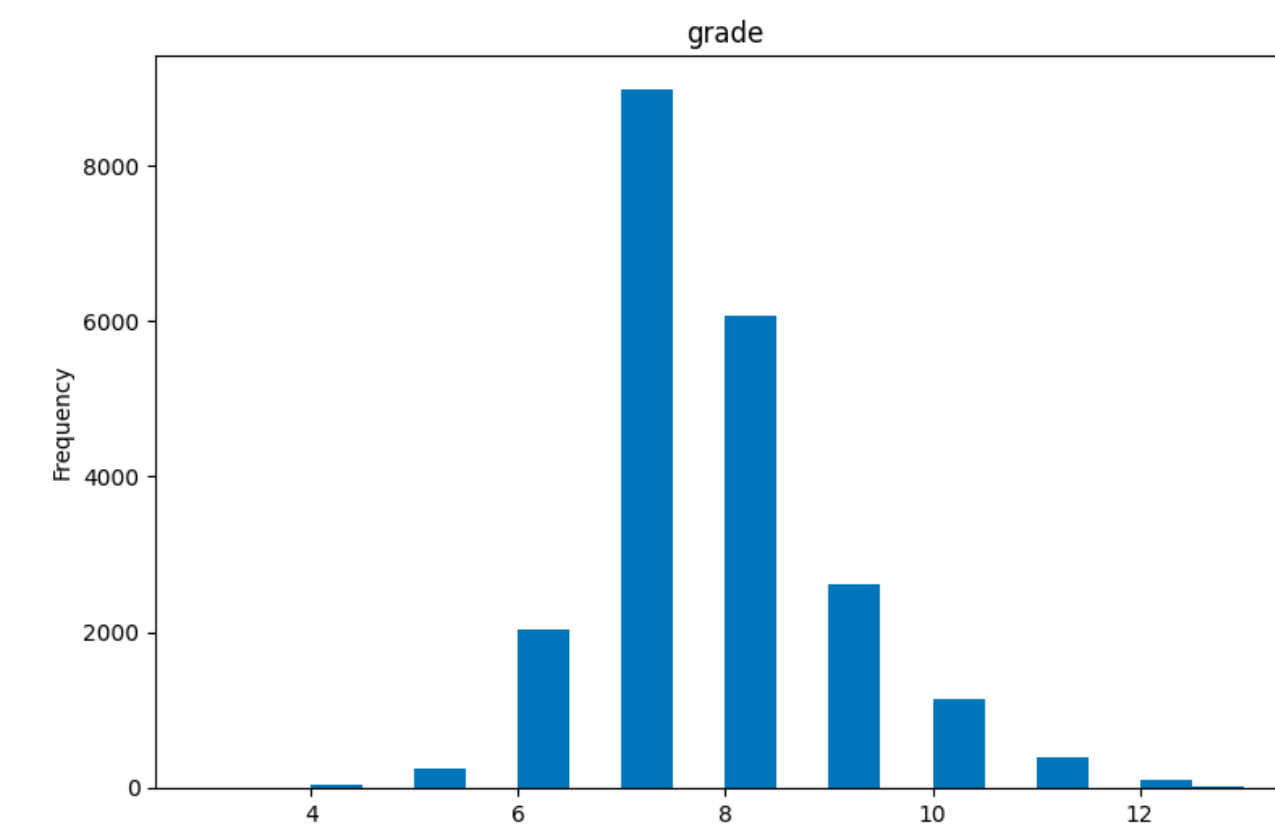
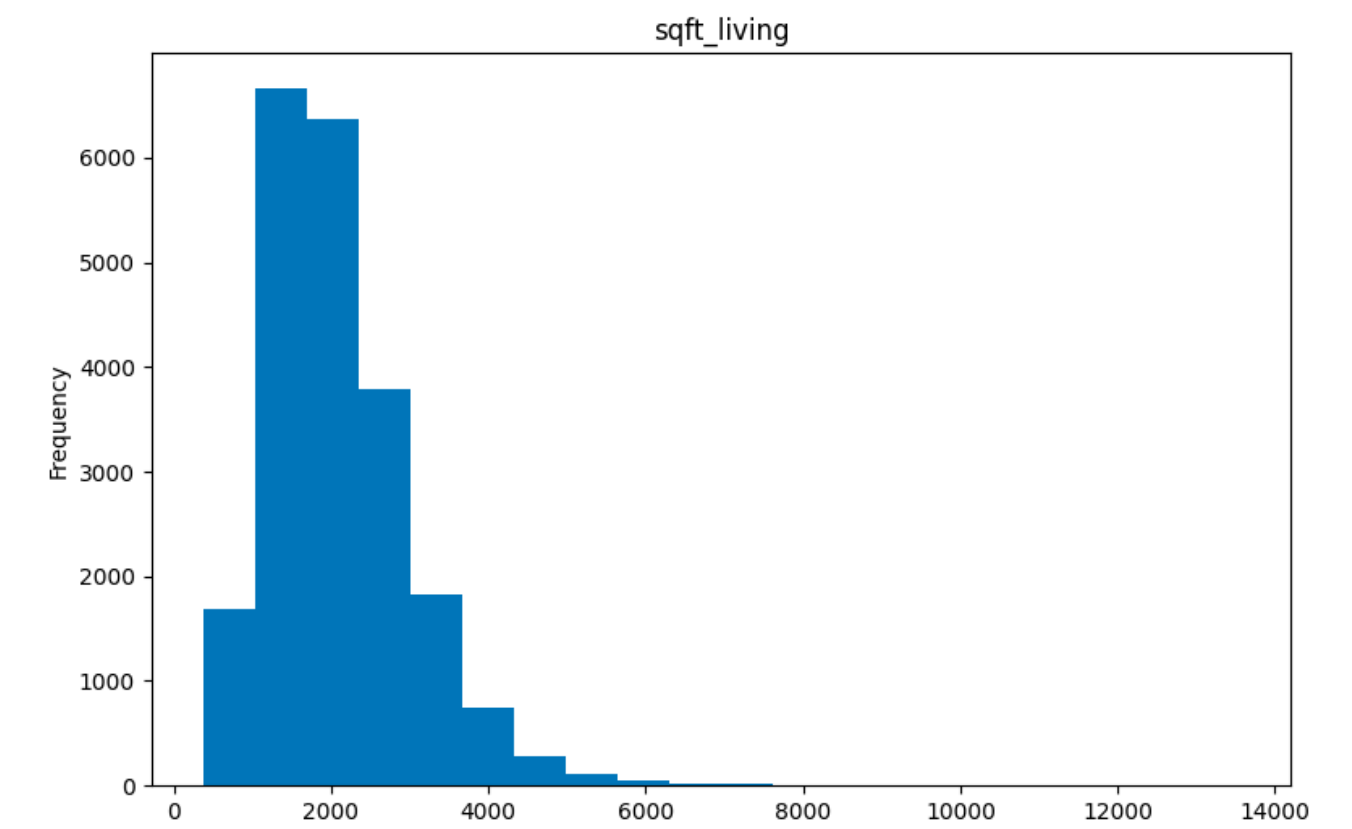
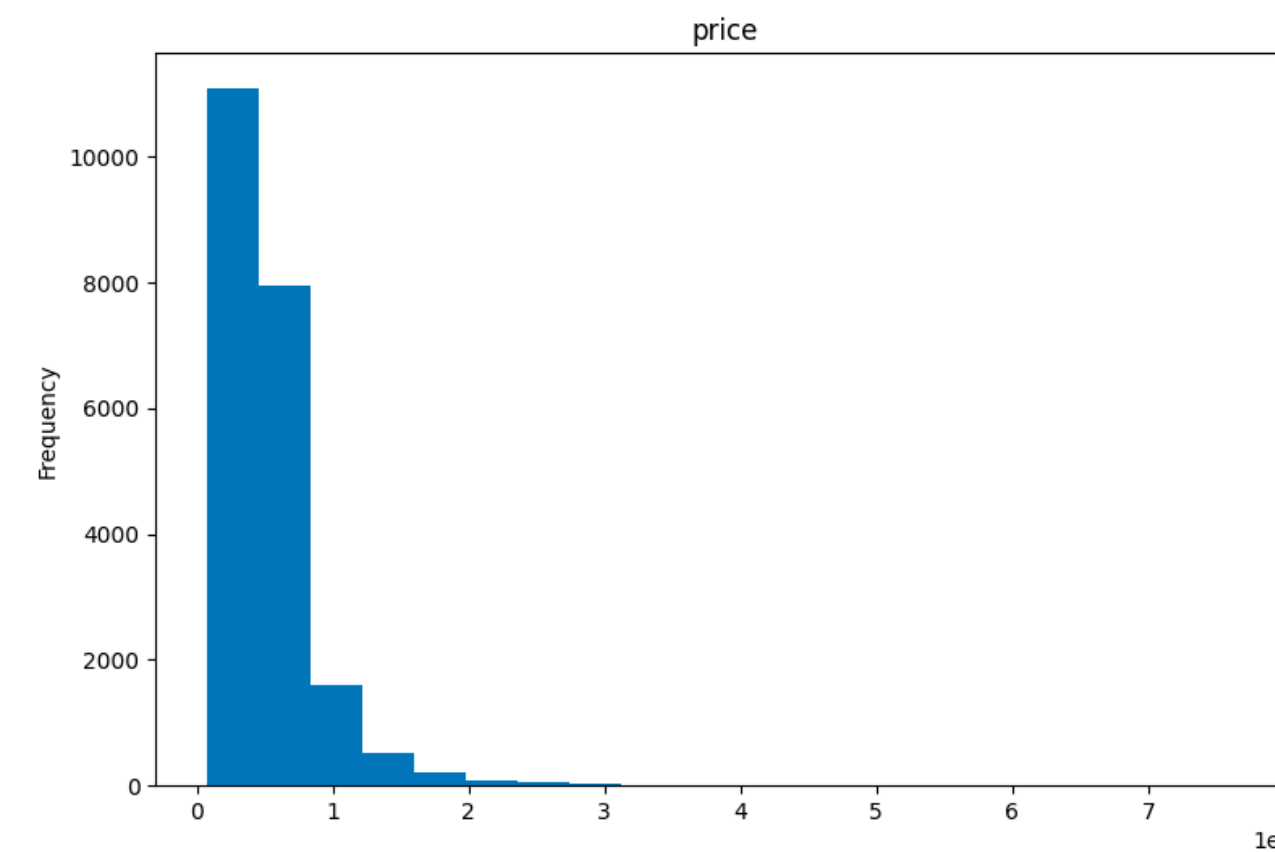
- **Float columns:** converted to int, where no residue
- **Date:** converted to datetime
- **yr_renovated:** values were somehow multiplied by ten
 - Corrected with the following:

```
df_kch['yr_renovated'] = df_kch['yr_renovated'].apply(lambda x:x//10)
```

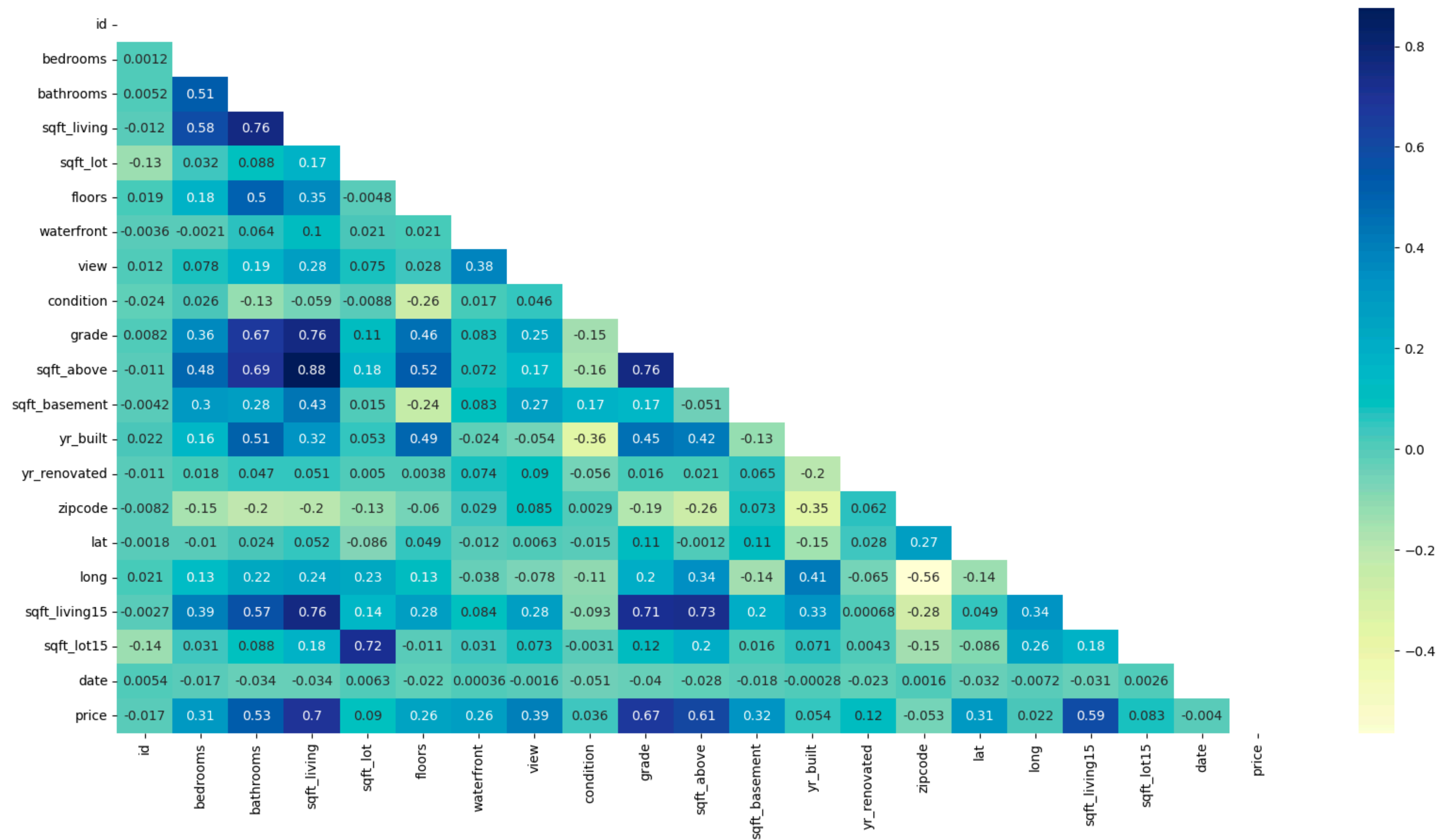

Data Exploration

Histograms

- Further investigation of
 - Price
 - sqft_living
 - Grade
 - yr_renovated
 - Waterfront



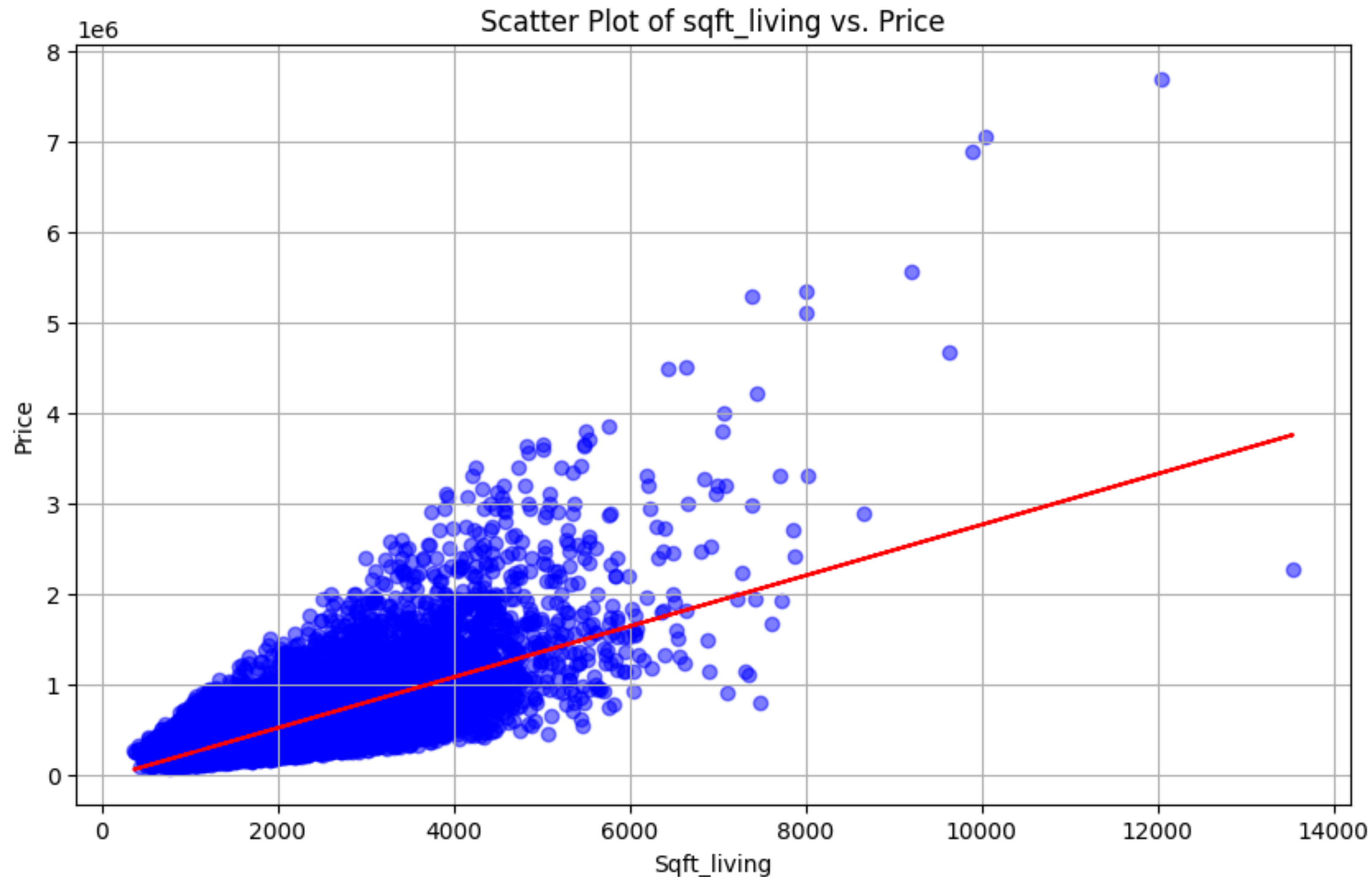
Correlations



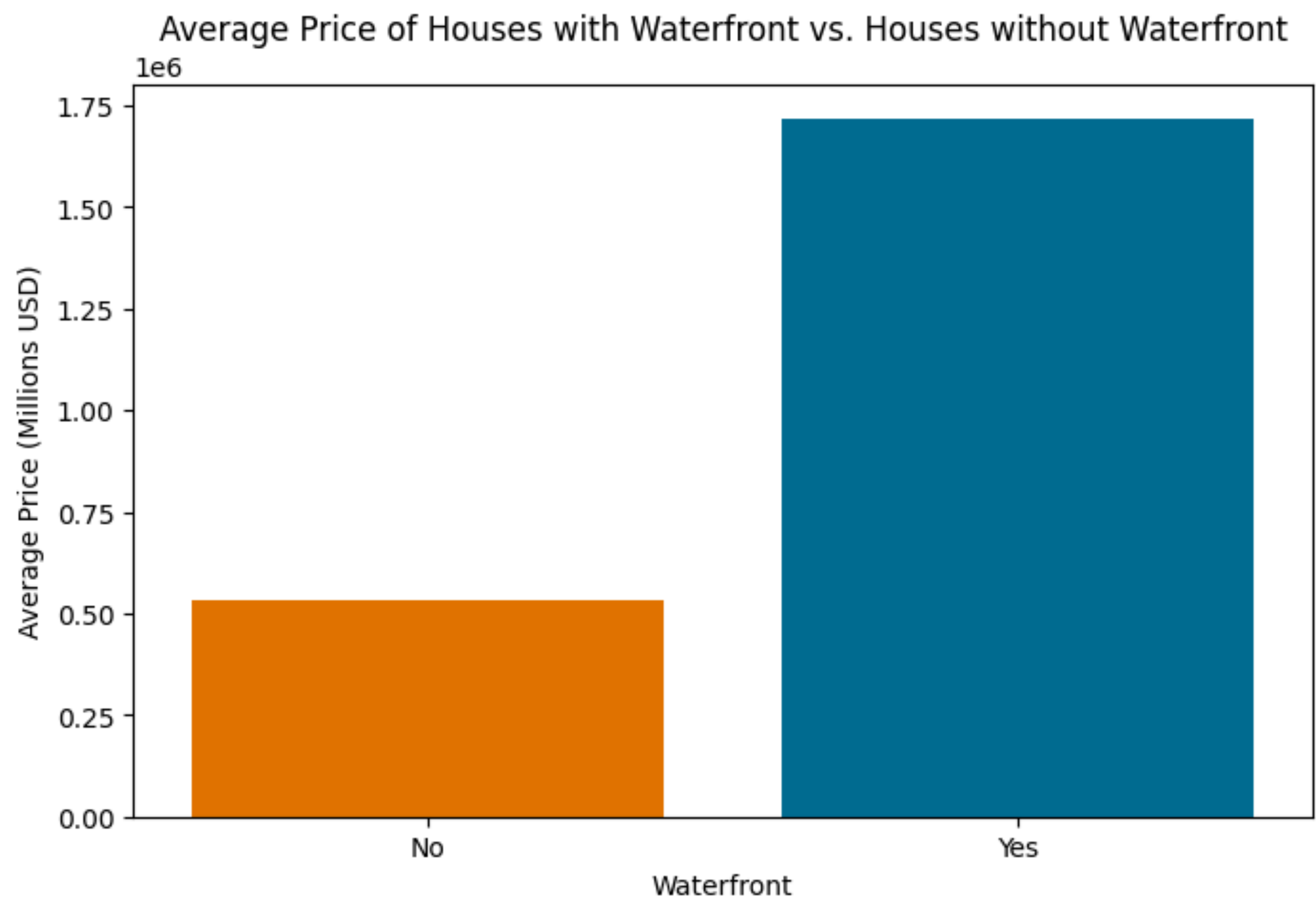
Results

More living space, higher price

Scatter plot with regression line

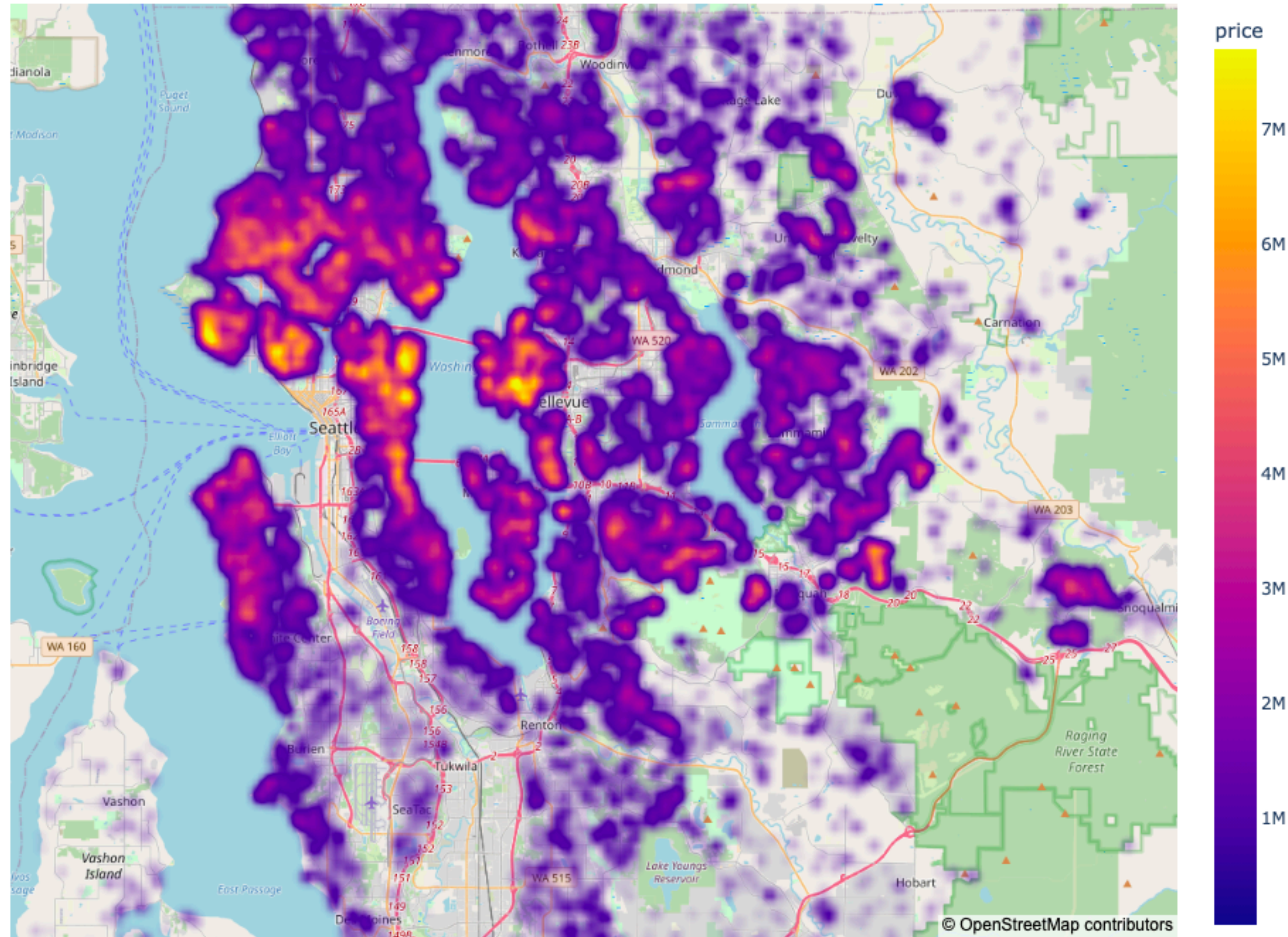


Higher avg price for waterfront properties



If located to water, then higher prices

Density Mapbox

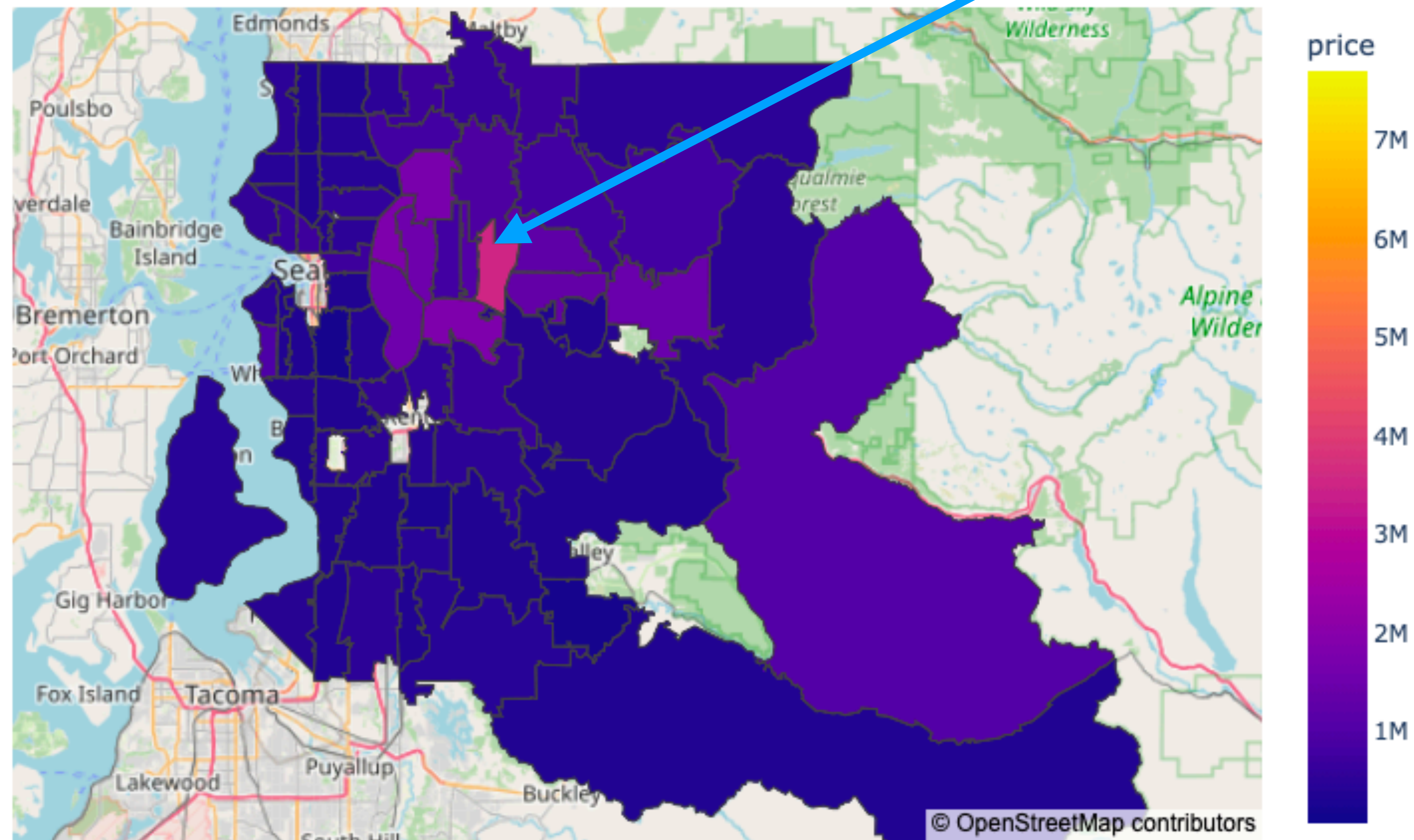


Higher avg house prices for attractive regions

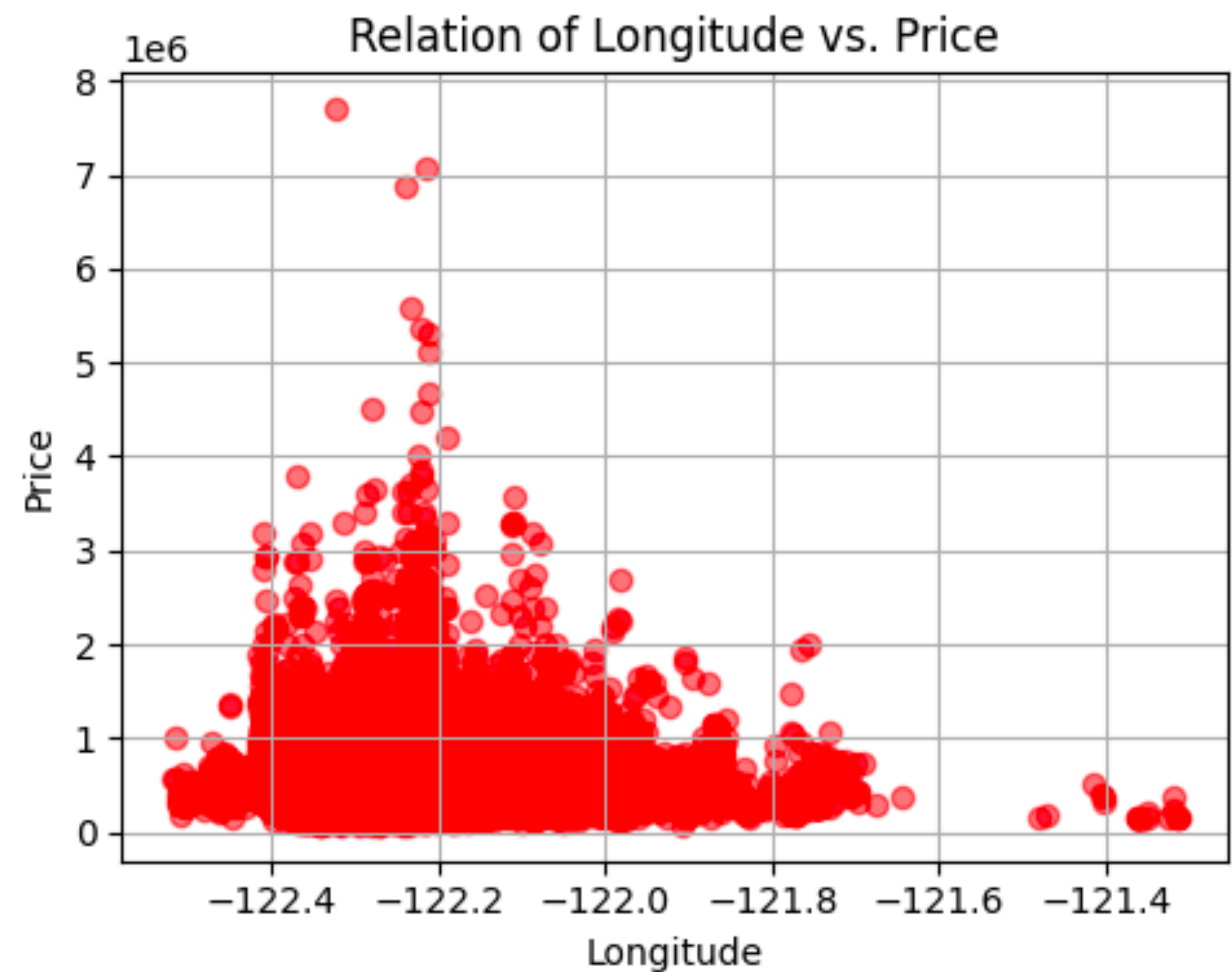
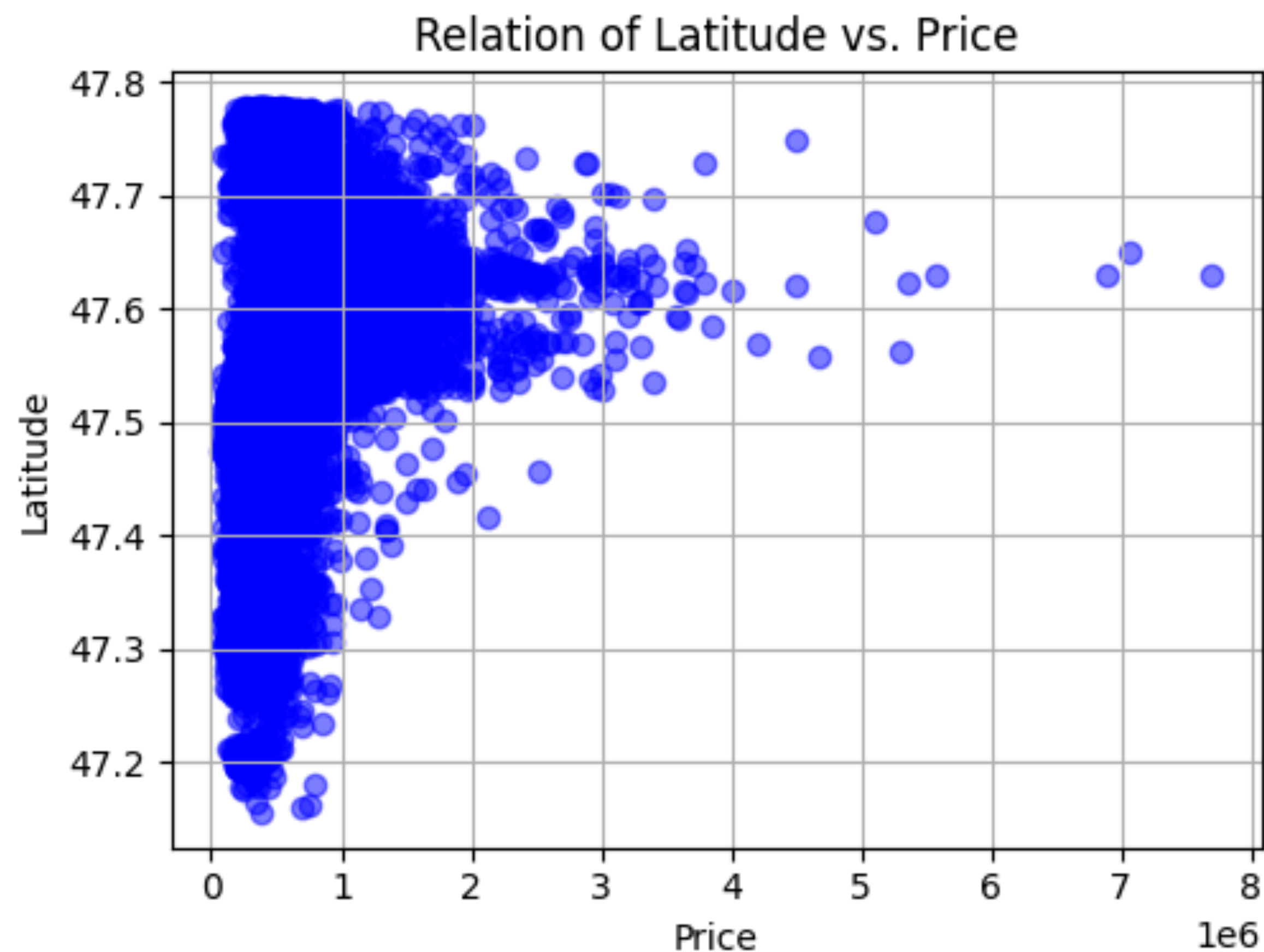
Choropleth Mapbox

Average house prices per zipcode area

Zipcode 98008 -
Bellevue



Relation of Price with Latitude and Longitude

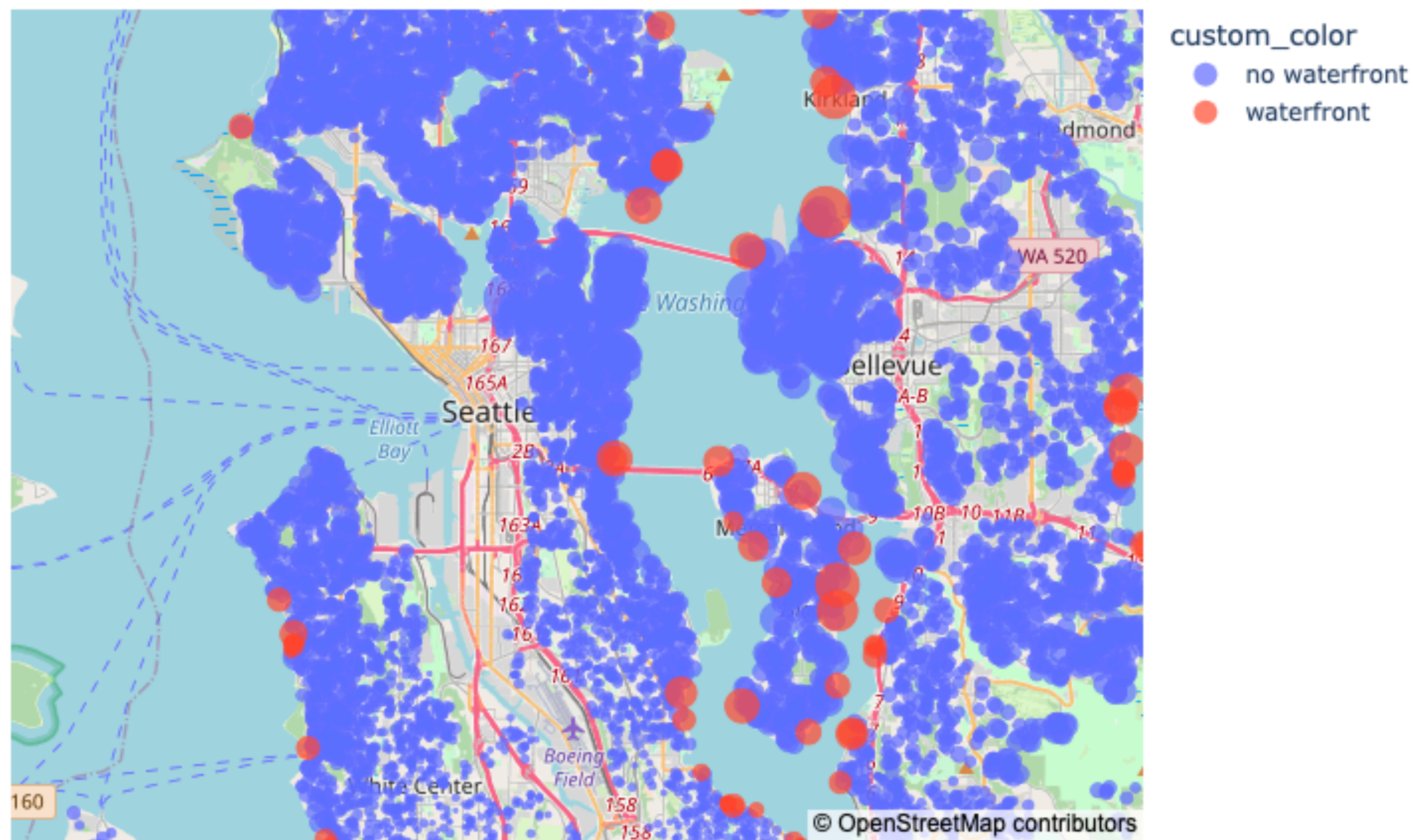


Recommendations for Client

1. Choose a house with a bigger square footage of the home, these tend to be more expensive
2. Waterfront Properties are much more expensive and a perfect choice to show off
3. Search for properties in specific regions, as these seem to be more attractive and higher priced (e.g. zipcode 98008 - Bellevue)

Houses with waterfront

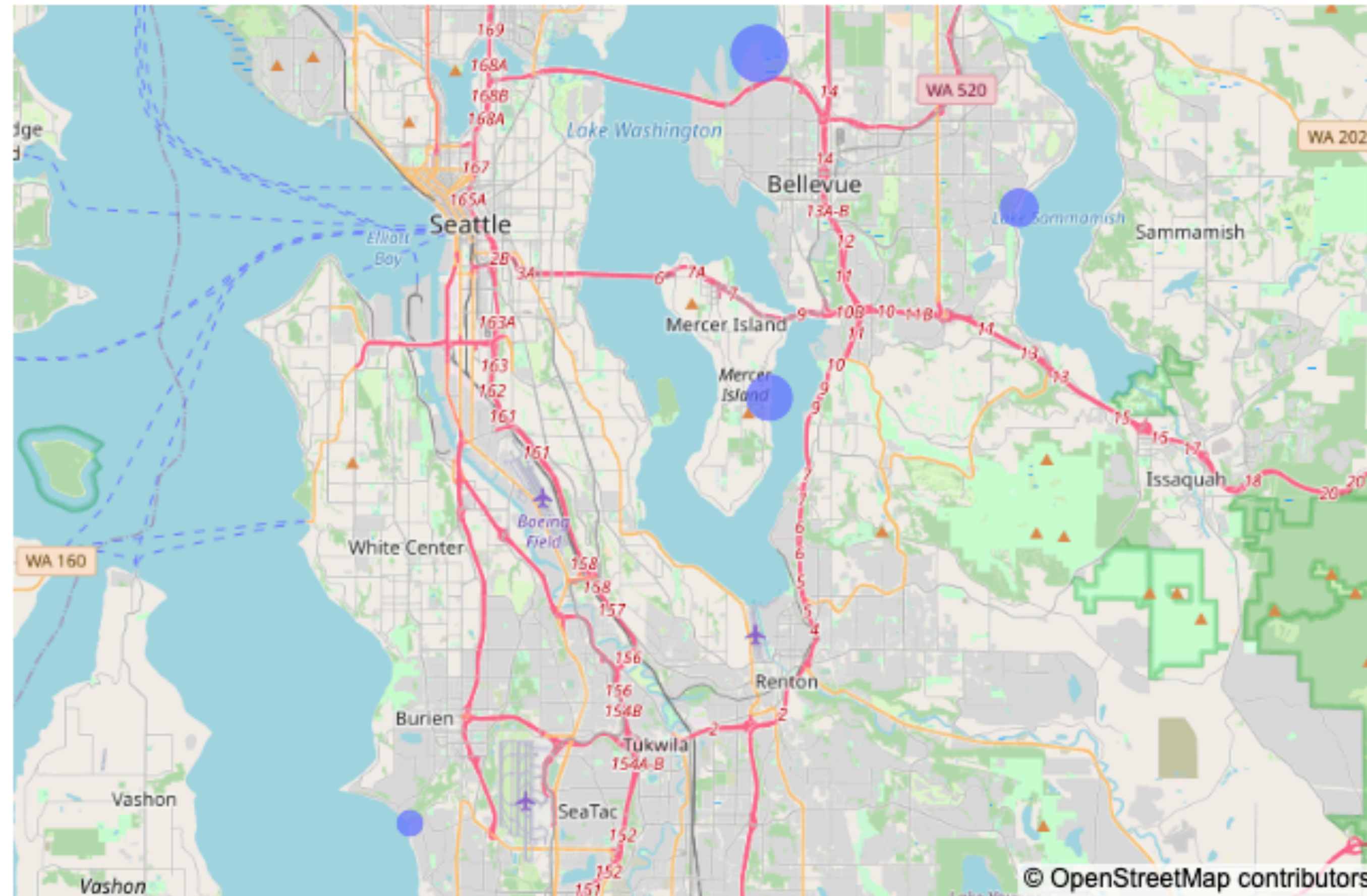
Prices colored according to waterfront



Four Houses meeting Client Requirements

`df_kch.query('waterfront == 1 and yr_renovated != 0 and grade > 10')`

Property candidates for Jennifer Montgomery



Further Research

- More detailed information on Client's wishes for certain amenities nearby
 - Performing distance calculation
- Performing a waterfront check using a water map for missing waterfront values for more accurate search
- Including additional data about city names corresponding to zipcodes to display them on geographical maps

Thank you!