

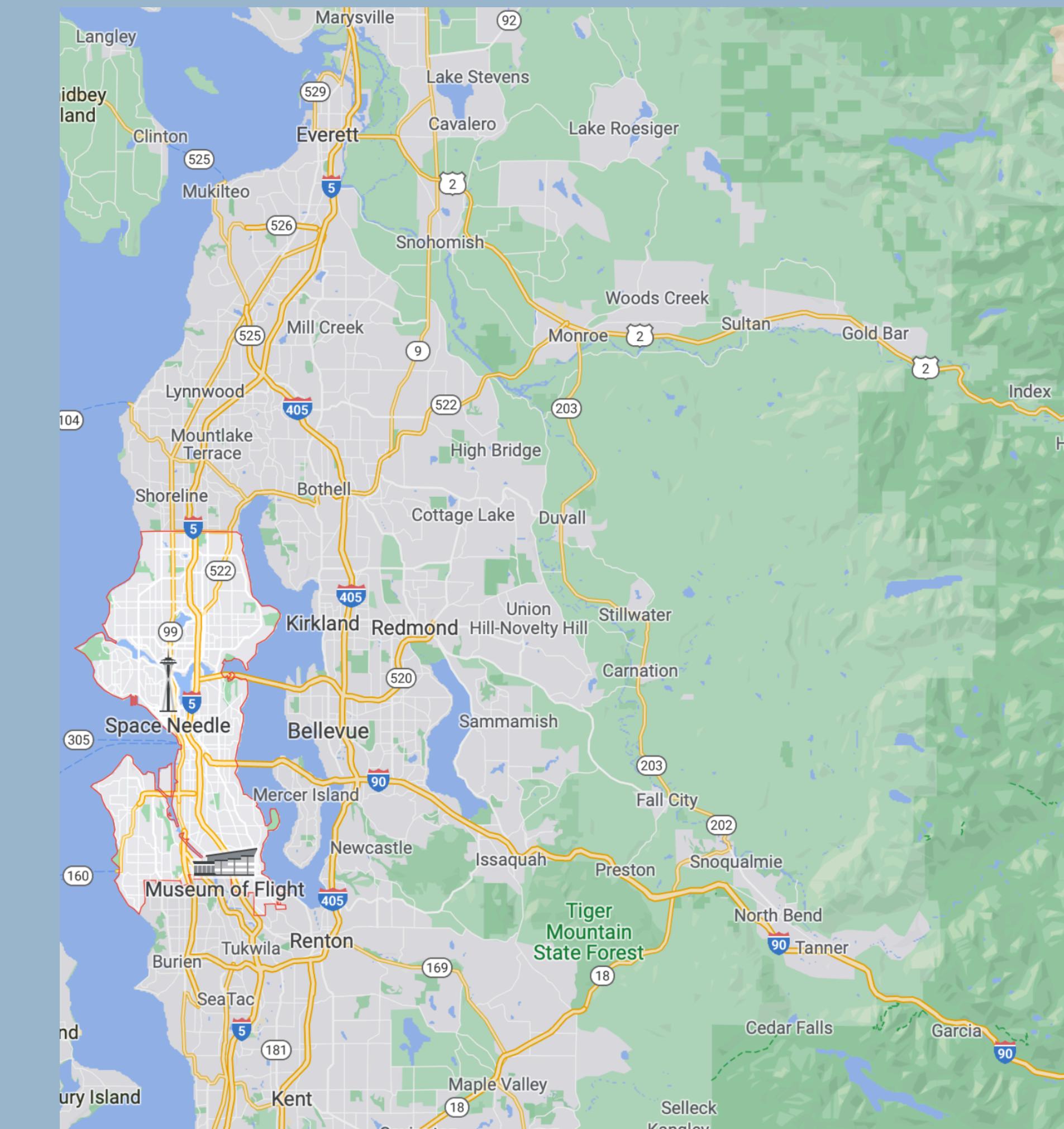
House Sales in King County, USA

EDA-Project in Data Science

Andreas Priebe, 27.01.2022

Understanding the Data and the Task

- King Country (Seattle), USA
- Various data to draw conclusions about house sales from 2014/2015
 - prices, living space, rooms, location...



Checking the data and build some hypothesis

Keep the stakeholder Bonnie Brown in mind

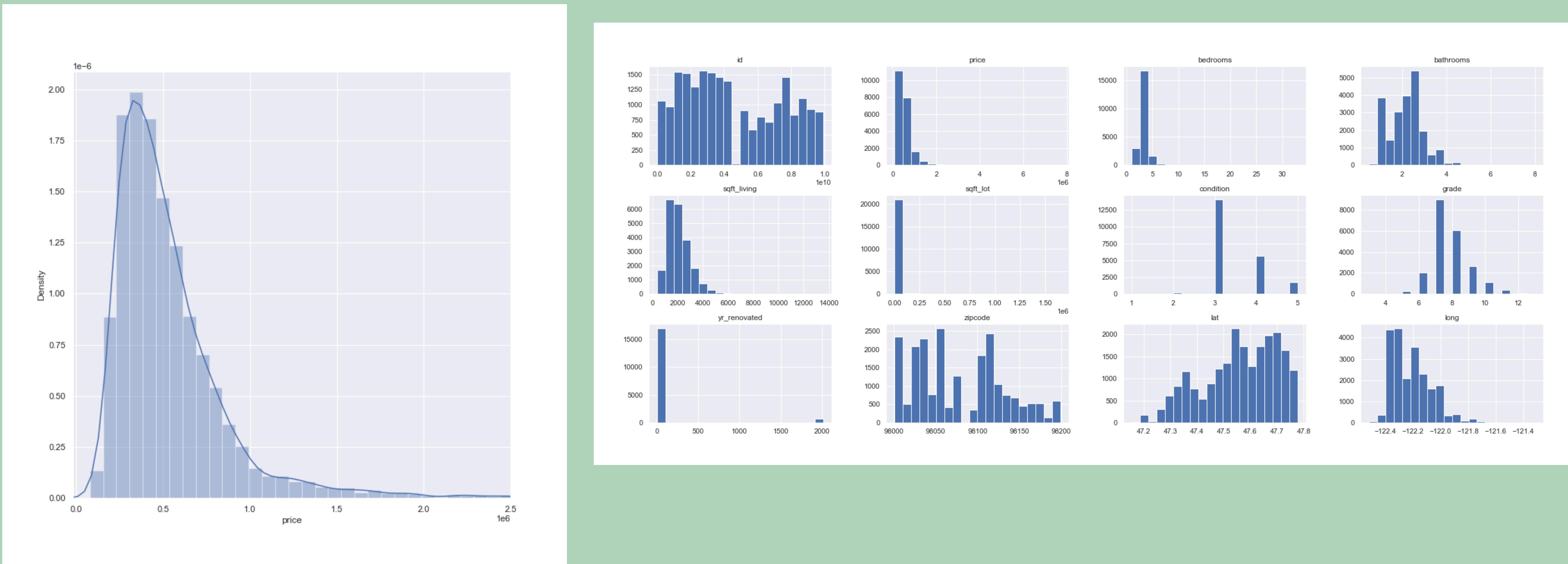
- 1. The number of rooms tend to contribute to the selling price
- 2. The more actual living space the more the price increases
- 3. Depending on the location/the zip-code there is a pattern in selling prices
- 4. Houses with better grades tend to have recently been renovated
- 5. There are times during the year in which selling prices tend to be higher

Who is Bonnie Brown?

- She is a potential seller
- She Has a house and wants to move soon (timing might be important)
- She wants high profit in middle class Neighborhood

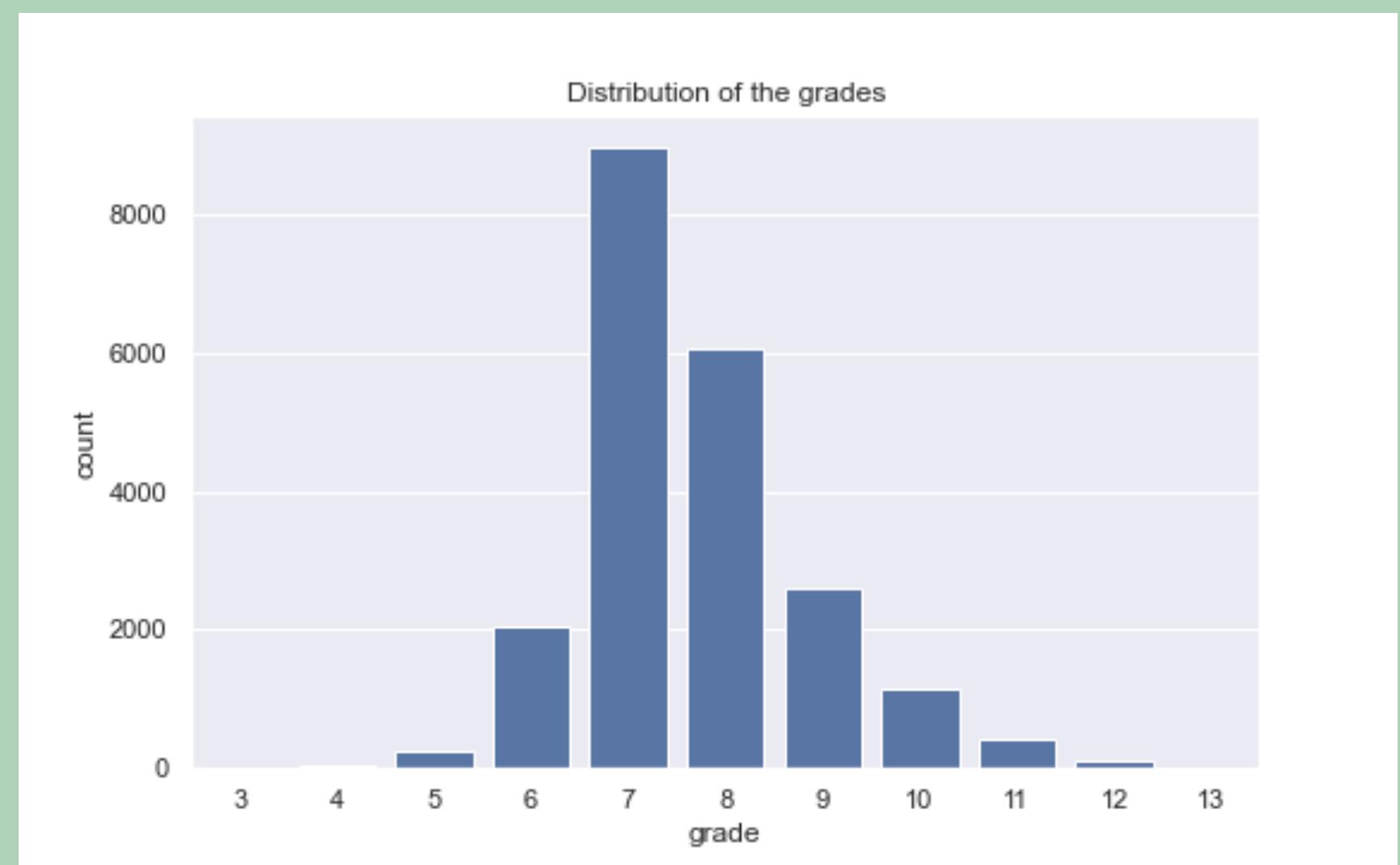
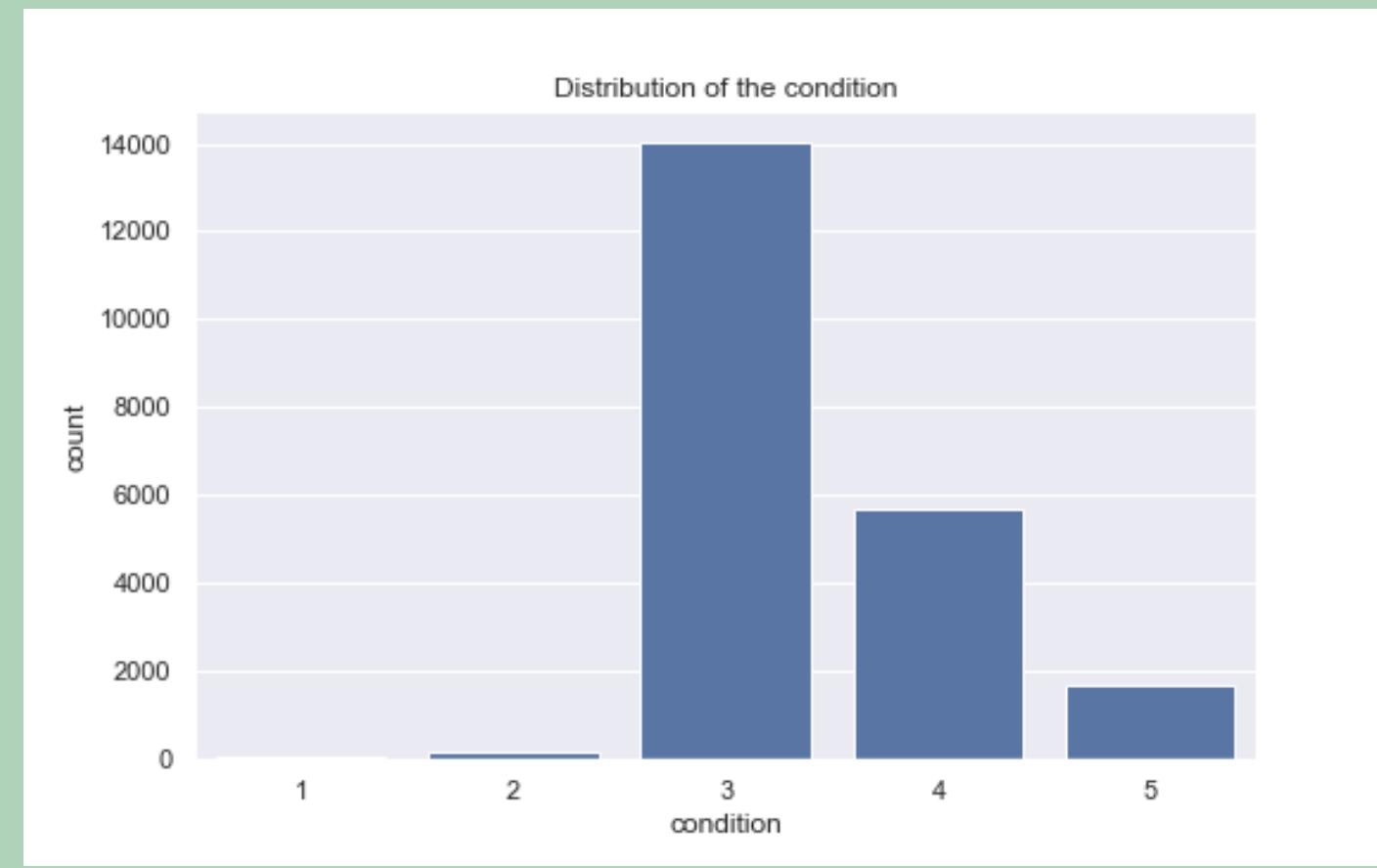
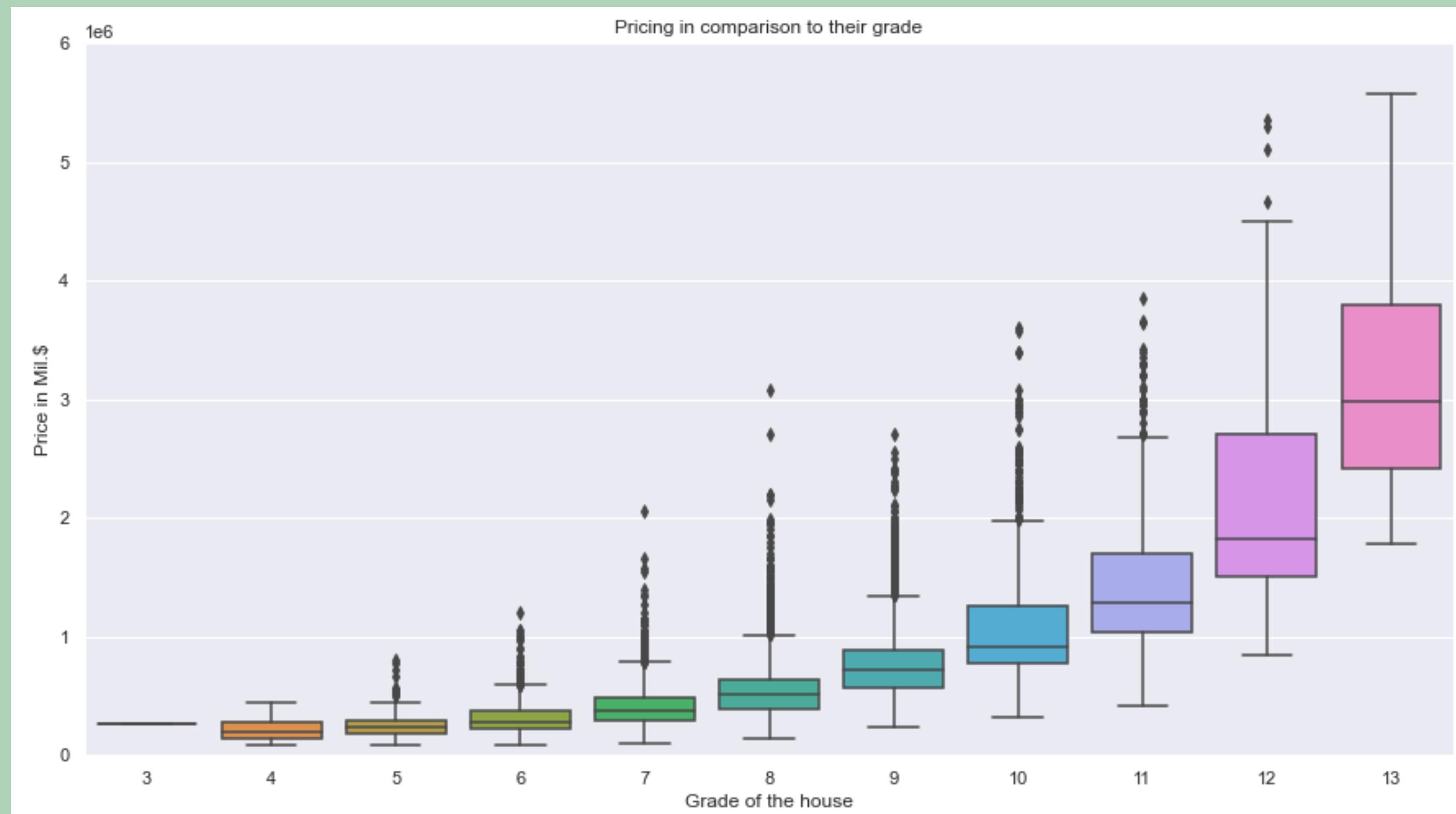
Exploring the data

- Checking the data for connections and distributions
- “middle class neighbourhood” ~ 500.000\$ - 700.000\$



Checking categorial features

- grade and condition tend to contribute to a higher price **but**
 - about 90% awarded grade 7 or 8
 - About 95% awarded with condition 3 or 4



Cleaning the data

- Dropped a few columns (duplicates) to keep only features of interest
- changed the date and checked the the renovation of house
 - Only a few houses have been renovated so I will keep both data

Connections in the data

- Price, Number of bedrooms and bathrooms, living space, lot area, condition, grade, year of renovation

Ist there any correlation?

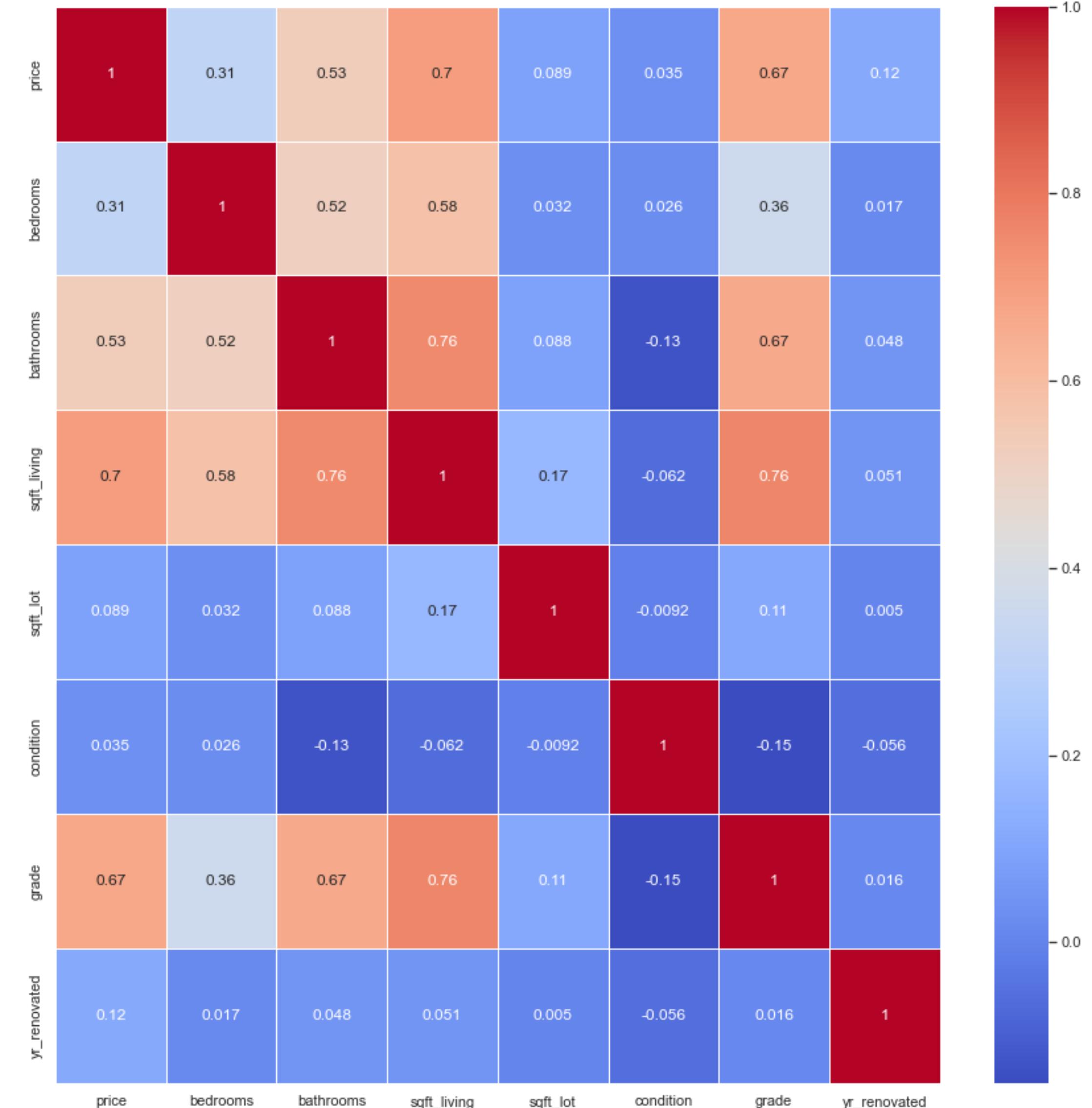
Price and:

Living Space, grade and also bathrooms

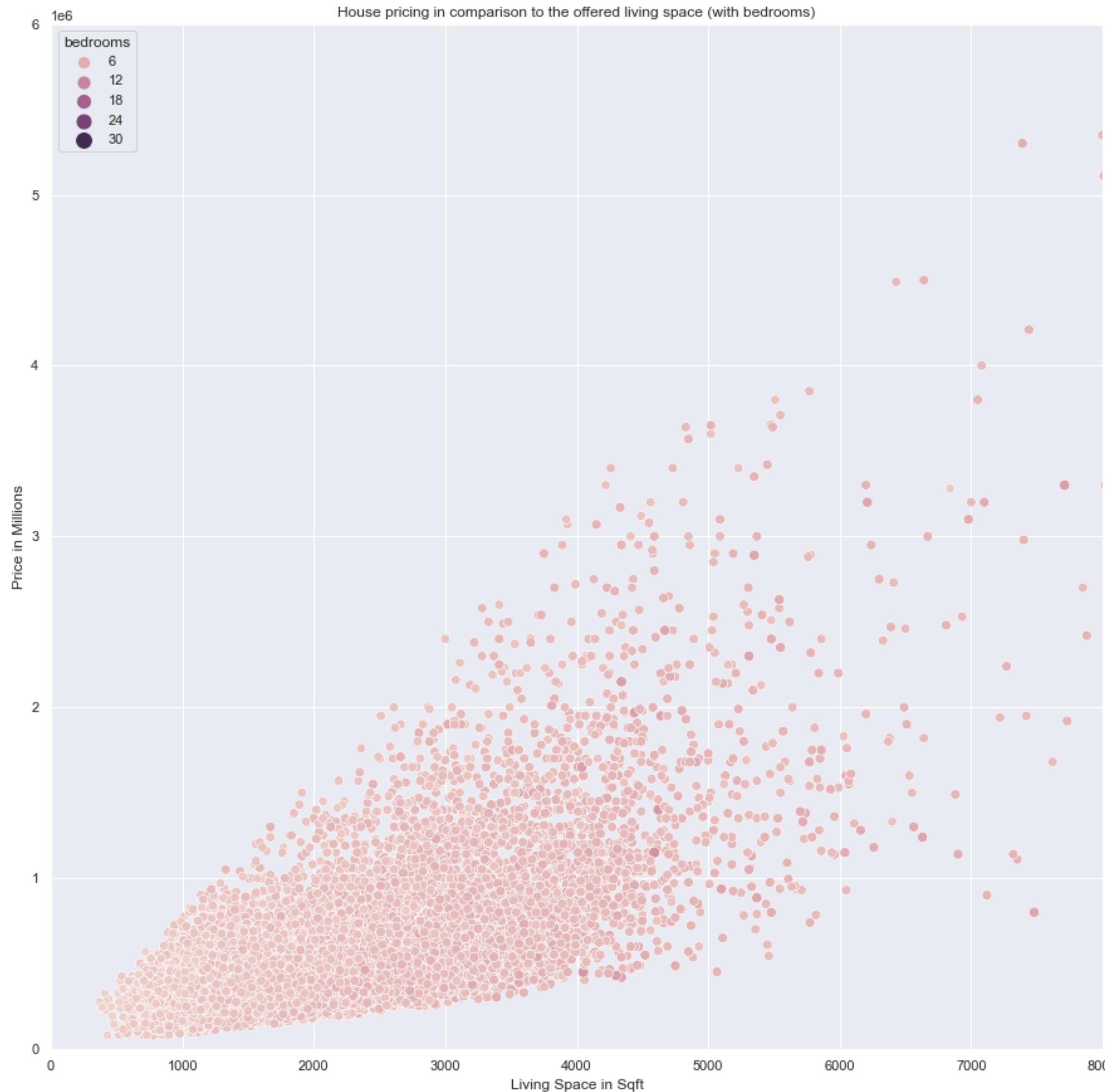
Living space and:

number of bedrooms, bathrooms and the grade of the house (expected)

4. Houses with better grades tend to have recently been renovated



Connections in the data



Connections in the data

1. The number of rooms tend to contribute to the selling price

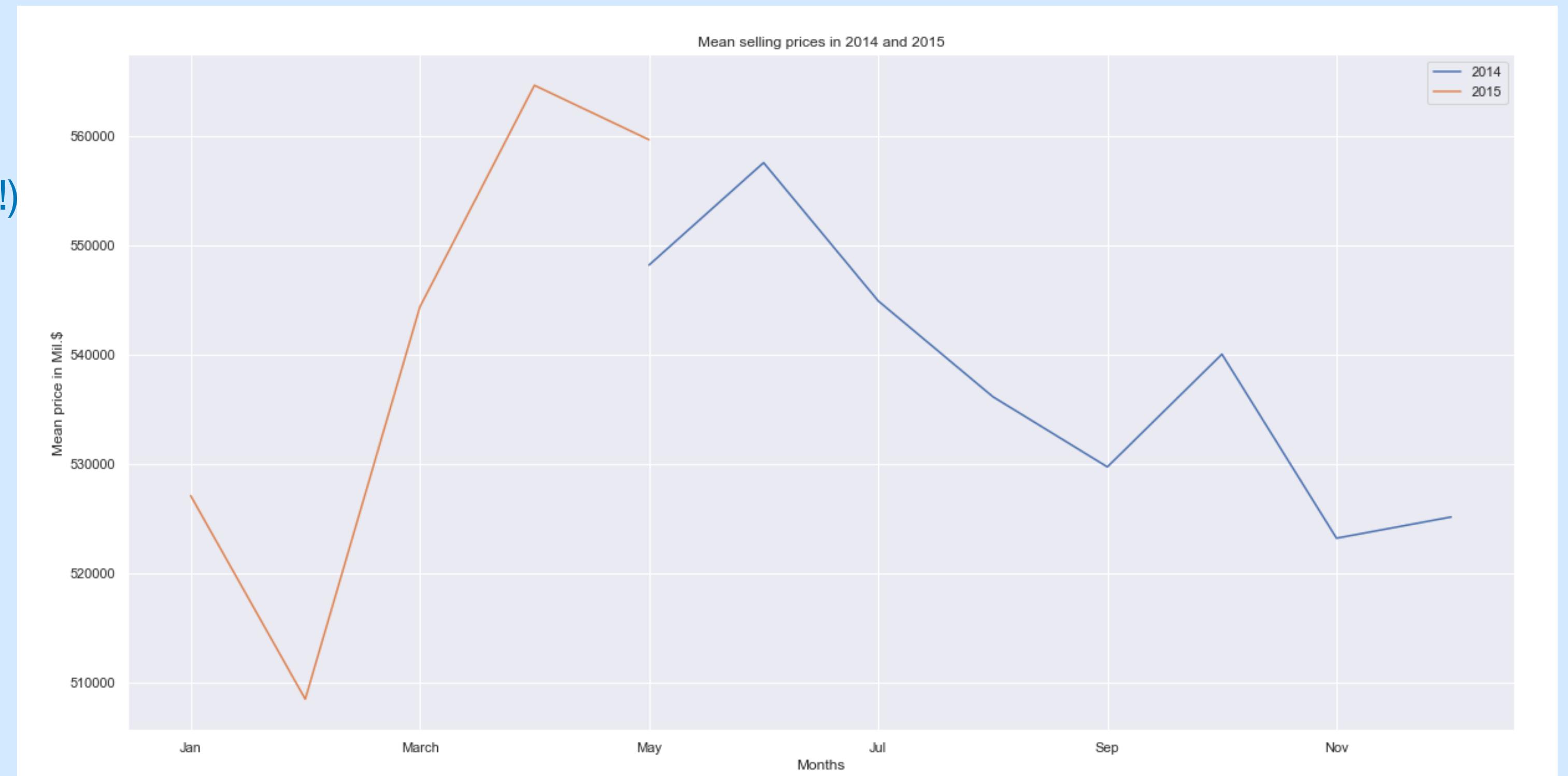
2. The more actual living space the more the price increases

- Yes more actual living space leads to higher selling prices and the houses tend to have more rooms
- Estimated Pearson correlation coefficient ~0.4 to 0.5 depending on the kind of rooms

Connections in the data

5. There are times during the year in which selling prices tend to be higher

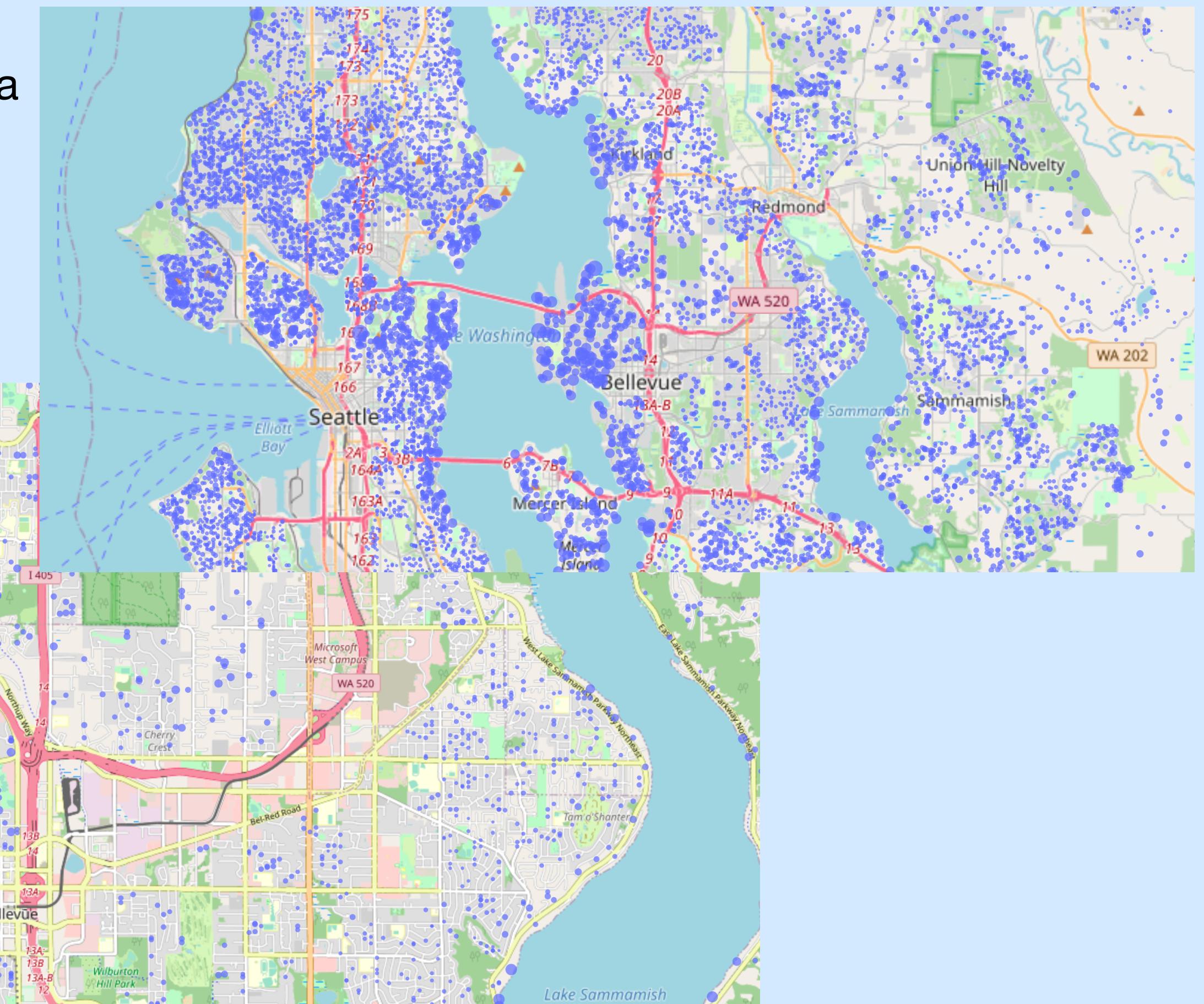
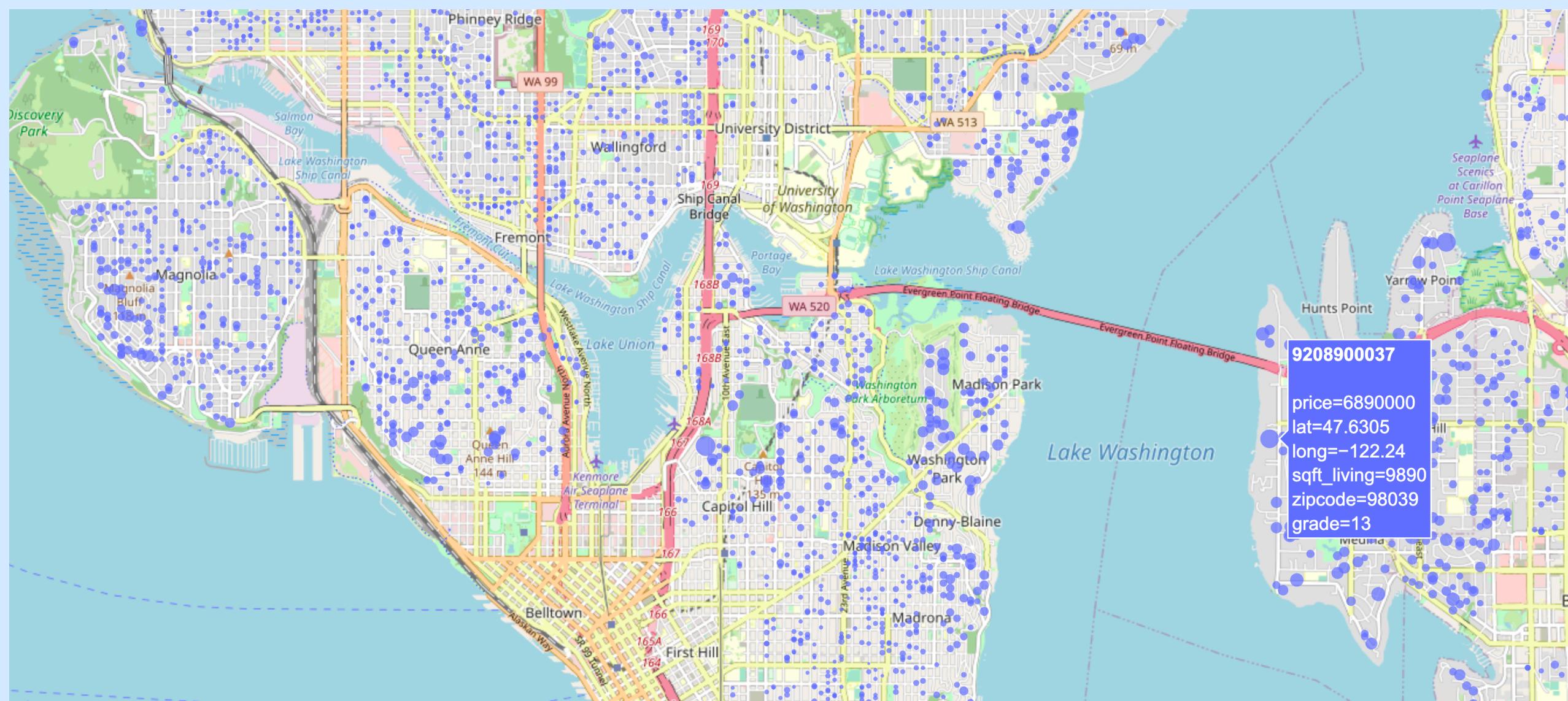
- Yes prices increase in the middle of the year with their peak on April/May
- Spring until summer is a good selling time!
 - ◆ (Keep in mind: Its only a short time period!)



Connections in the data

3. Depending on the location/zip-code there is a pattern in selling prices

- Yes there are many of them located in the outskirts of the city which tend to be less pricier
- inner parts in the city (Bellevue, downtown...) next to the water area have especially high prices (big blue dots)
- Tend to have grades > 10 and an actual living space > 2000 Sqft



Thanks for your attention!

EDA-Project in Data Science

Andreas Priebe, 27.01.2022

So what can we deduce from the data for Bonnie Brown?

- Selling time should be between March-July.
- Price depends on actual living space, number of rooms and the condition/grade of the house. ("middle class" ~500.000 to 700.000\$)
 - Bonnie can't change the house location, but she could do some repairing, painting and small improvements
 - add some additional walls to increase the number of bath or bed rooms
 - If she has not enough living space, she should start with adding some basement or roof space to it
- Speak to the data-collecting company to get more information on what their criteria for the grades are based on ("grade" and "condition" tends to contribute to a higher selling price)
 - No clear difference for changes on middle class houses and the following price increase/decrease.
 - start to calculate for the easiest steps to increase her selling price in a short time span

Summary

1. The number of rooms tend to contribute to the selling price

- **Yes** higher selling prices come along with a higher number of rooms in nearly all cases of higher prices (bathrooms and bedrooms)

2. The more actual living space the more the price increases

- **Yes** also the price correlates to the living space. The estimation of the pearson correlation coefficient is somewhere in the range of 0.4 to 0.5.

3. Depending on the location/the zip-code there is a pattern in selling prices

- **Yes** there is a connection between selling prices and the location. Areas such as Medina and the downtown area which is surrounded by water are particularly noteworthy here.

4. Houses with better grades tend to have recently been renovated

- **Yes and No.** As you can see on the map there are some houses with high prices, which have been renovated more or less recently. However, these houses are to be seen more as historical buildings, which are per se already in the higher segment. Most of the houses of the dataset have not been renovated yet. For this reason, no real connection can be deduced at least from the available data.

5. There are times during the year in which selling prices tend to be higher

- **Yes** - The best time span for selling would be middle of the year, starting with spring. The peak is reached in April/May. Nevertheless we have to keep in mind, that we only have limited data resources for a few months in 2014/15, which makes it very difficult to accurately predict the time ahead.