



EDA on King County Housing Data

Haritha Retnakaran

10.03.22

DATASET:

King County Housing Data: Information about home sales in King County (USA)

STAKEHOLDER INFO:

Jacob Phillips : Buyer

Characteristics :

- unlimited budget
- 4+ bathrooms or smaller house nearby
- big lot (tennis court & pool), golf, historic, no waterfront.

HYPOTHESIS:

1. Are the houses built in later 90's has more features, and hence more expensive? If not, why?
2. How many houses are there with more than 4+ bathrooms?
3. Where are these houses located? near or far from waterfront?

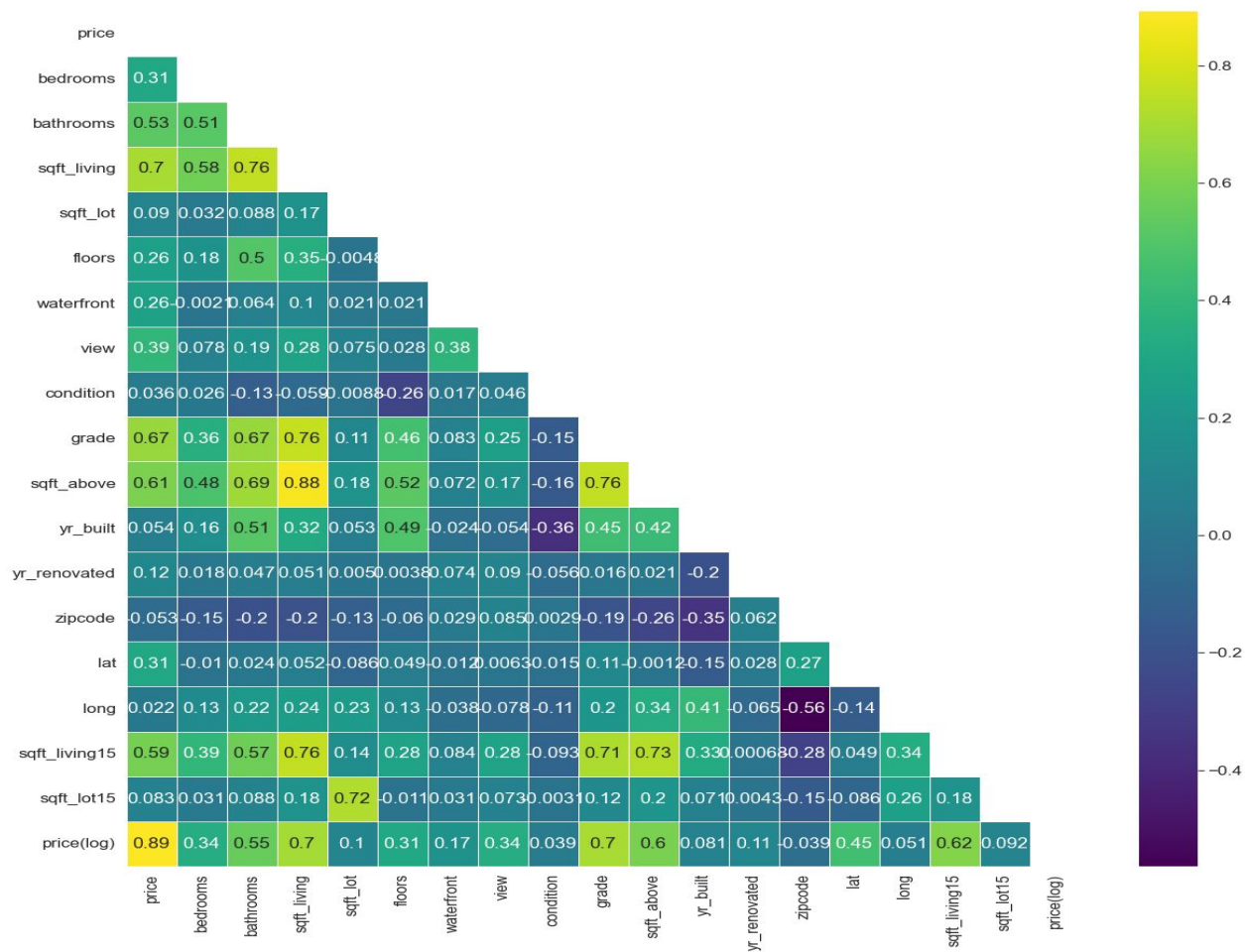
Features of interests:

- Year of built vs features of houses
- Features of houses vs Year of built
- Locations of house vs price

Data cleaning

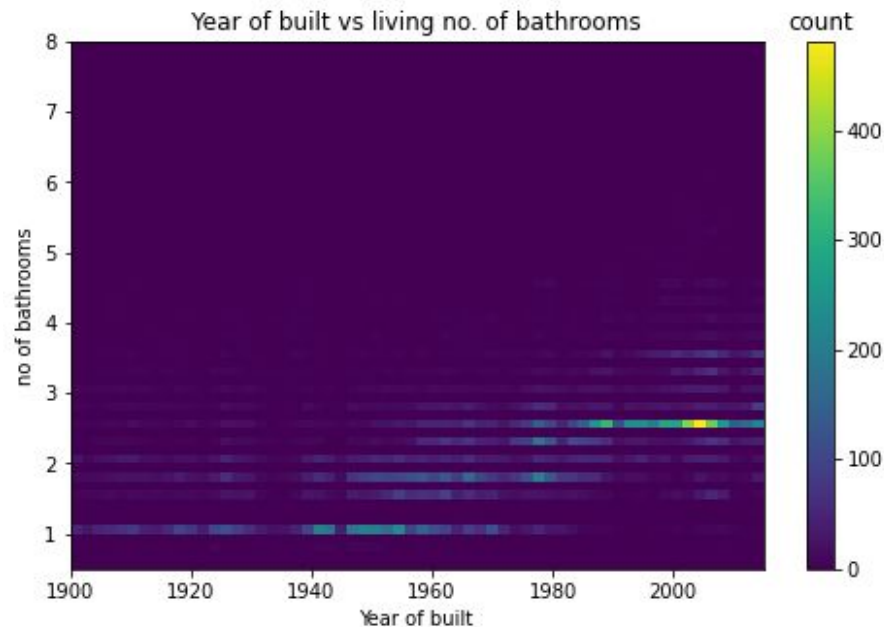
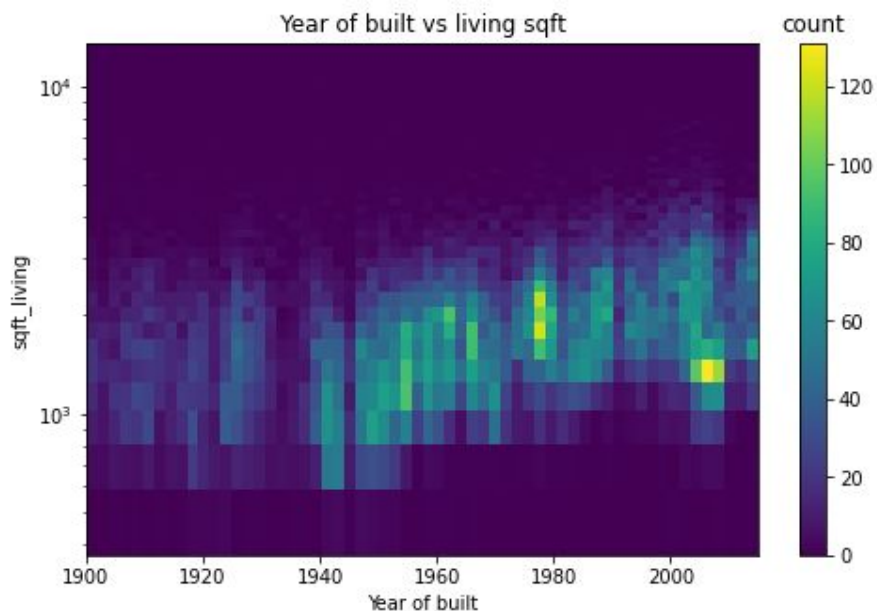
- NaN values → replaced with 0.0
- Date values → Datetime format
- Price values → \log_{10} scale

Correlation matrix



ANALYSIS & RESULTS:

Stakeholders interest:



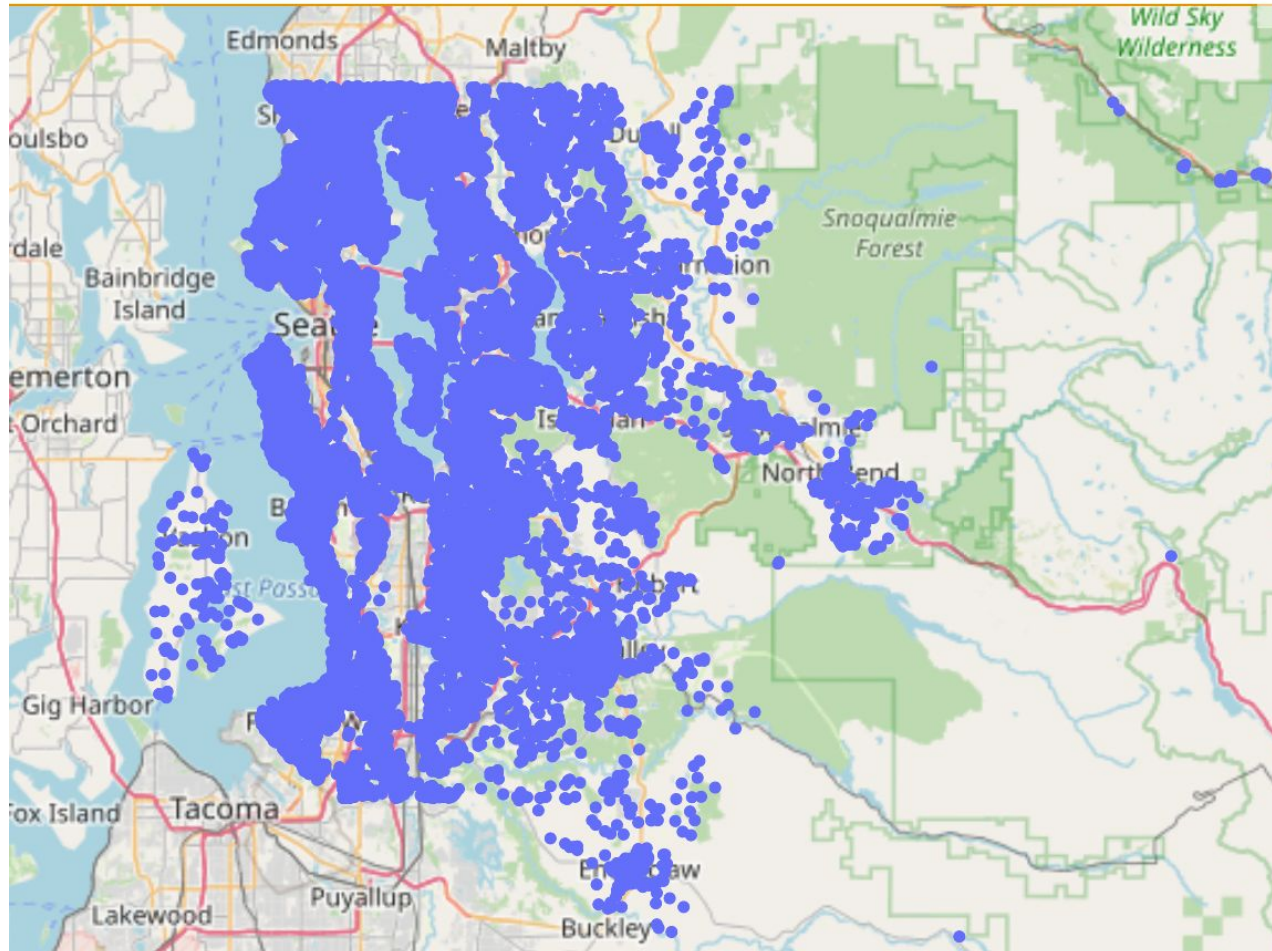
Conclusions

- More no of houses with large size → later years
- More no of houses with 4+ no of bathrooms in later years
- There are fluctuations too!! Why??

Insights

- **Map plot** → price of the houses vs the year built vs location

Location of the houses in County:

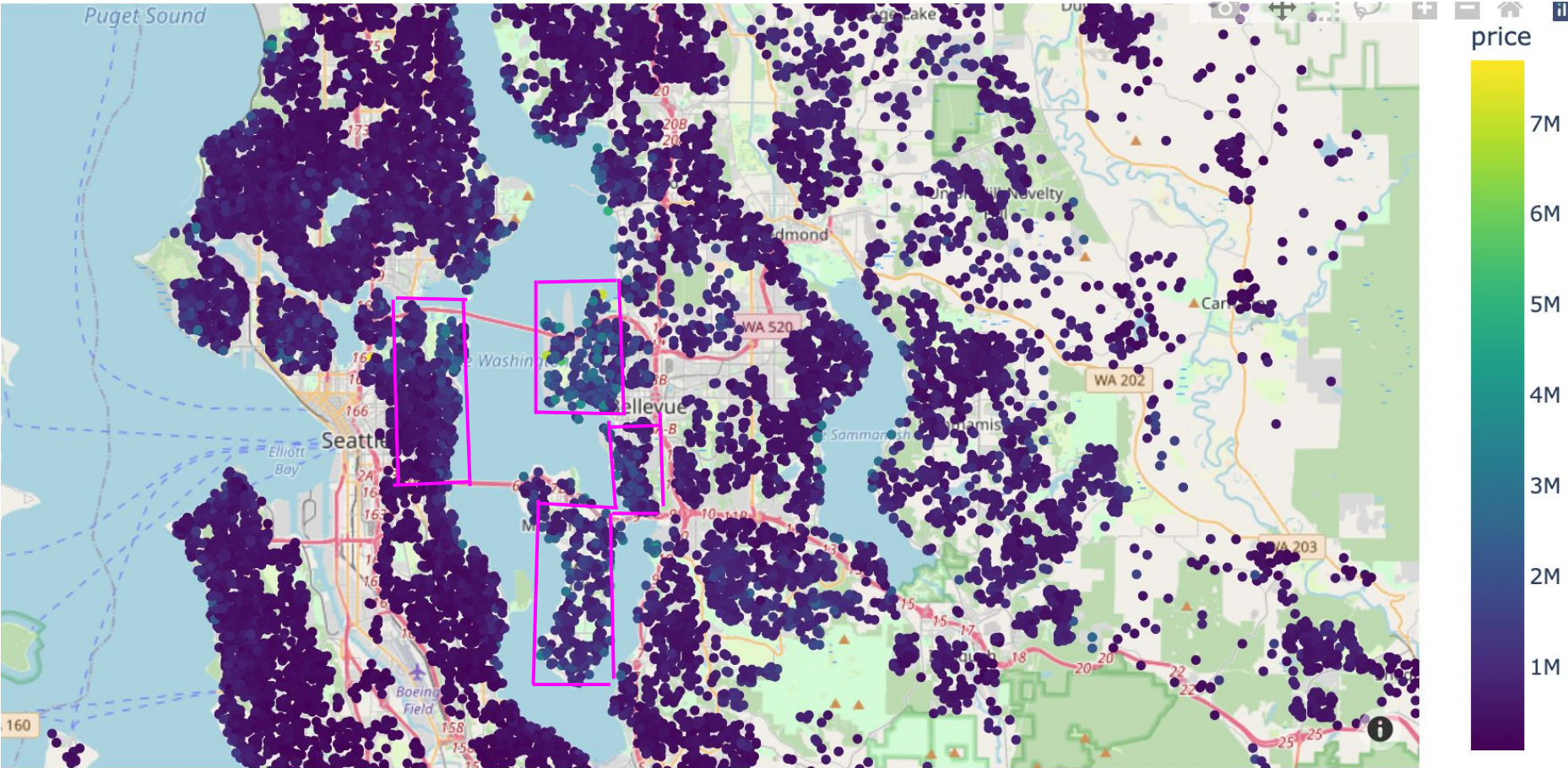


Information on county

Seattle, King County, USA:

- Metropolitan city
- Highly touristic
- Surrounded by sea
- Home to large tech industry: Amazon & microsoft headquarters in Seattle

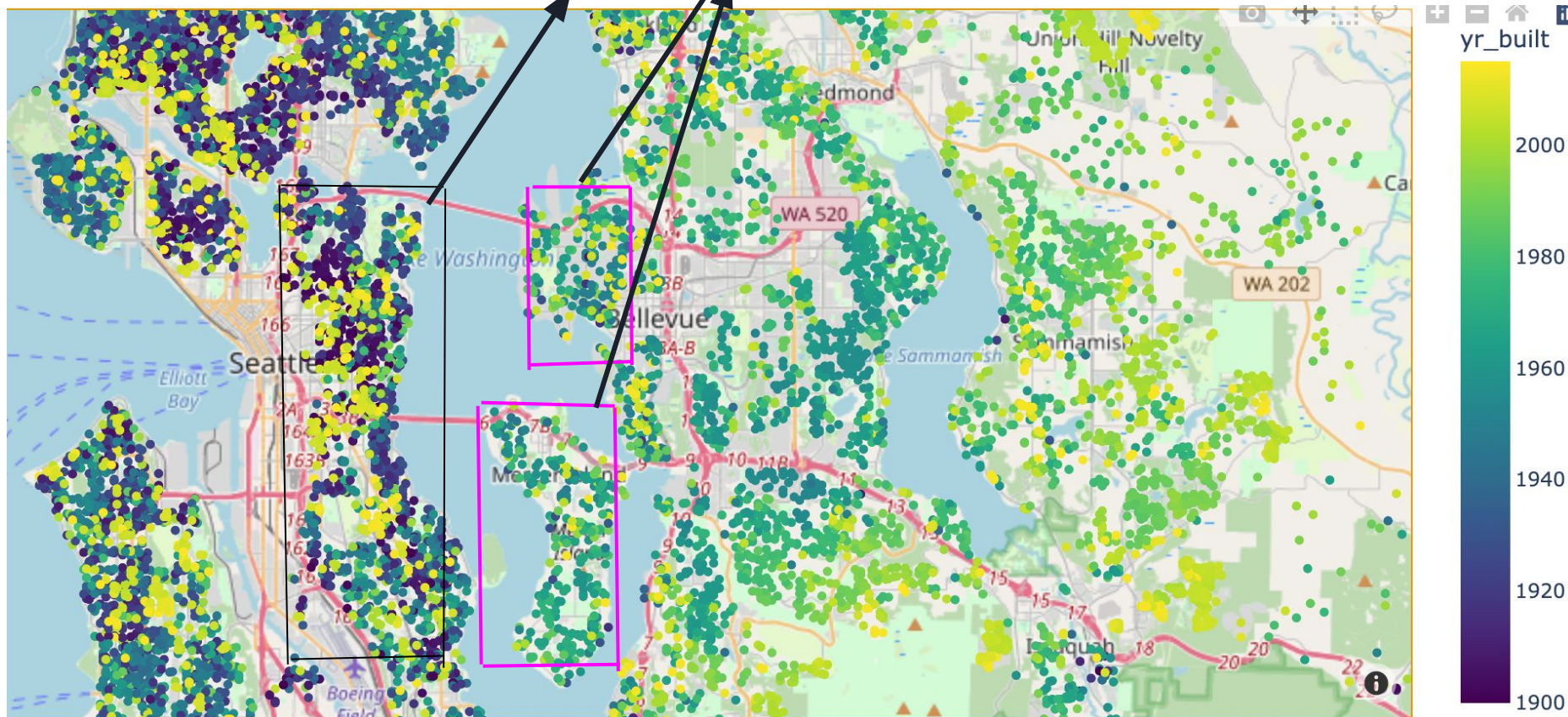
Map with the house price



Map with year of built

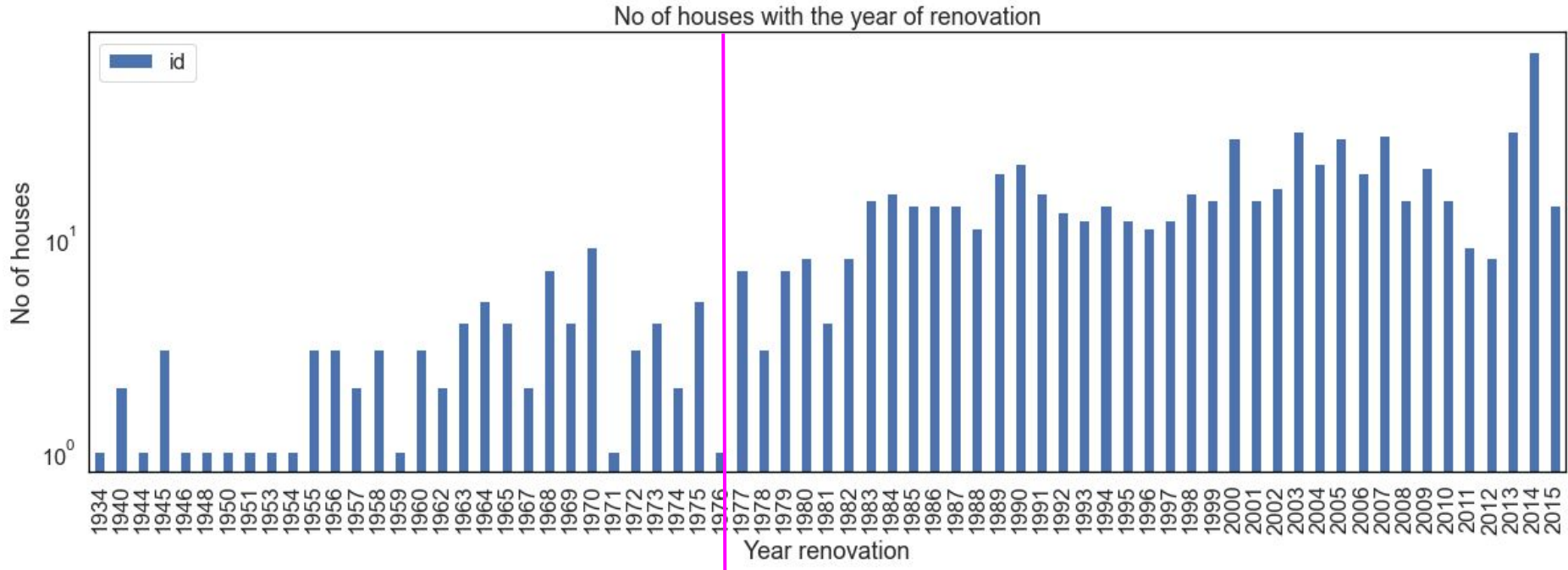
??

Reasonable to the price map



REASONABLE FACTORS:

1) Anything to do with renovation? : around 40% of missing data

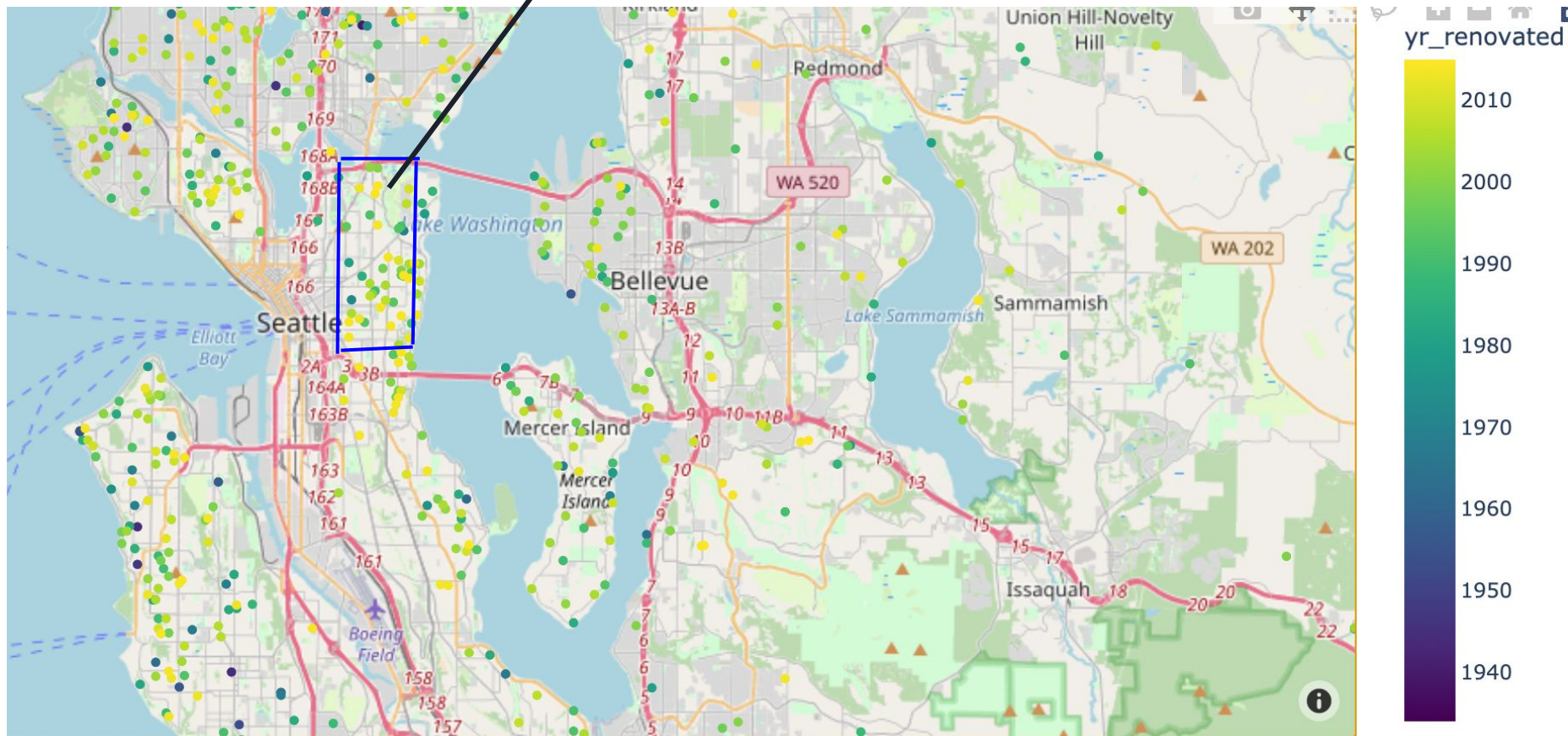


Conclusion:

Explain fluctuations where high price for houses built in early years

Map of renovated houses

Explain fluctuation: high price built in early years



2) Anything to do with recessions in USA?

- 2 main recession period along the years given
 - Late 20th century (1980-82)
 - Early 21st century (2008-09) : **Great recession**

Conclusion:

Explain fluctuations where low price for houses built in later years

Recession vs renovation: why more renovations during recession?

- Travel and recreation budget goes to renovation
- Contractors will be available than before
- Drop in material pricing

3) Other reasons for fluctuations?

- The houses built in early 90's can be larger or have similar size compared to the ones built in 2000.
- May be the houses that were built in early 90's were extensively renovated later and have similar amenities that today's buyers are seeking

RECOMMENDATIONS FOR STAKEHOLDER:

Stakeholder: Buyer

- No waterfront: suburbs or countryside? → mostly built in late 21st century
- Unlimited budget : Most of the houses away from waterfront are less expensive general, maybe he can prioritize his specifications of the house instead of budget

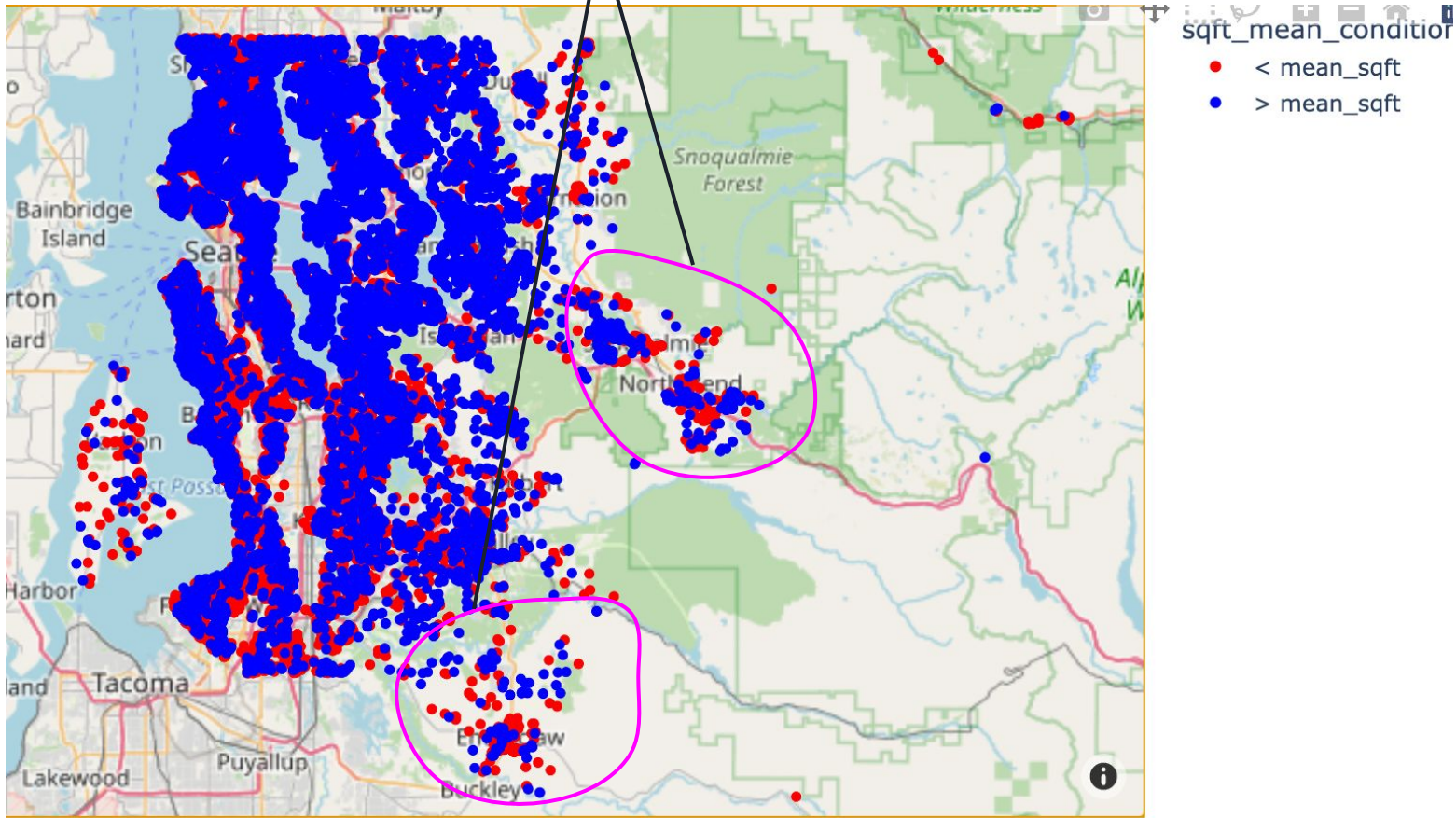
MORE RECOMMENDATIONS:

- What about other characteristics: 4+ bathrooms and smaller houses nearby?

Map of sqft_living of houses

No waterfront, smaller houses nearby

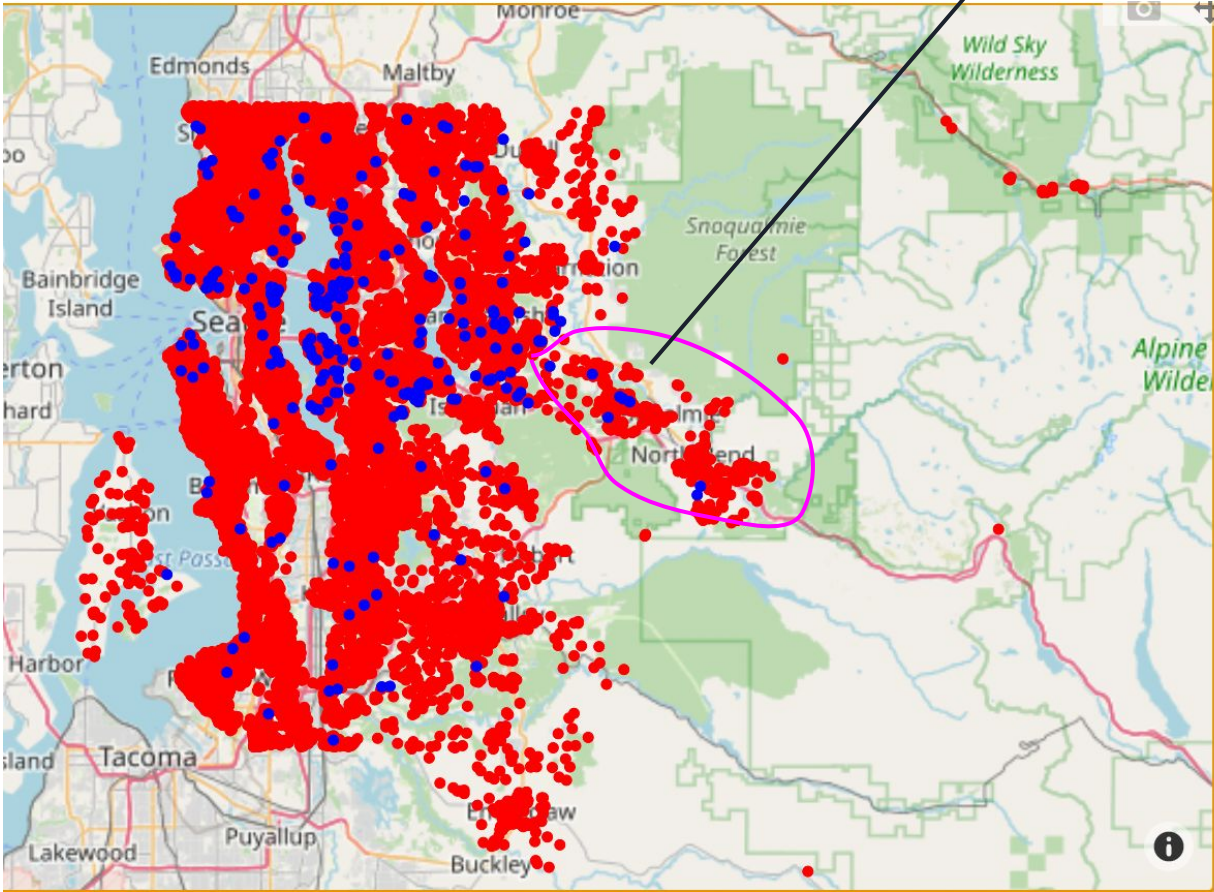
Mean: 2080.322



Map of no. of bathrooms

Interested number: 4+

4+ bathrooms, smaller houses nearby too



FUTURE OUTLOOK

- Look for external data on the seattle housing data which accounts the given data

THANK YOU!!!