King County House Prices Project:

Business case still in development, exploratory results regarding House Price in King County will be shown:


Data Set: containing House Prices for Sales from May 2nd, 2014, to May 27th, 2015.

Besides House Price, the data set contains information about 20 other variables, ranging from information about the selling date, house ID, zipcode area to the various house properties (e.g. living area, year built, grade according to King Counts grading system). Look at 'column_names.md' file for further information.
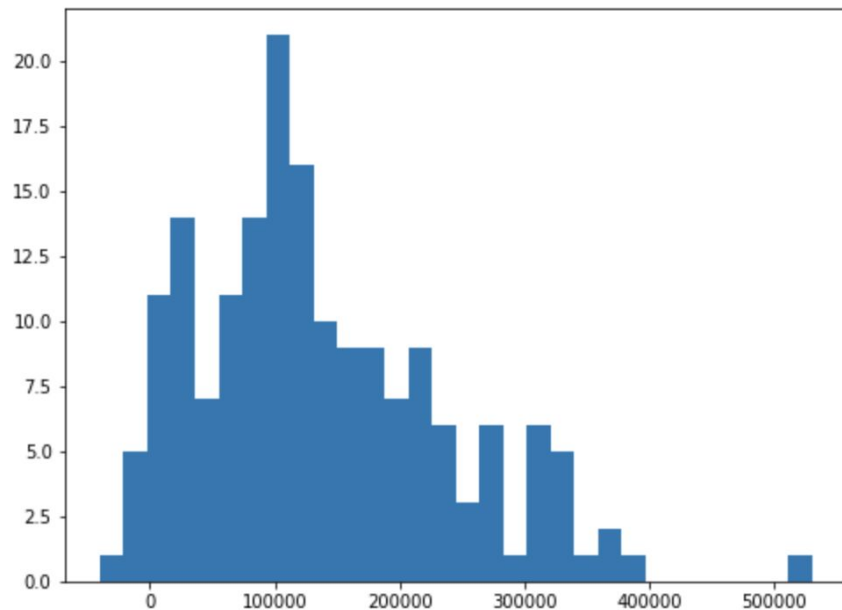
175 houses were sold twice (1 even three times), most of which had significant increases in sales price (mean increase 38%) → area in demand!! (only took most recent price)
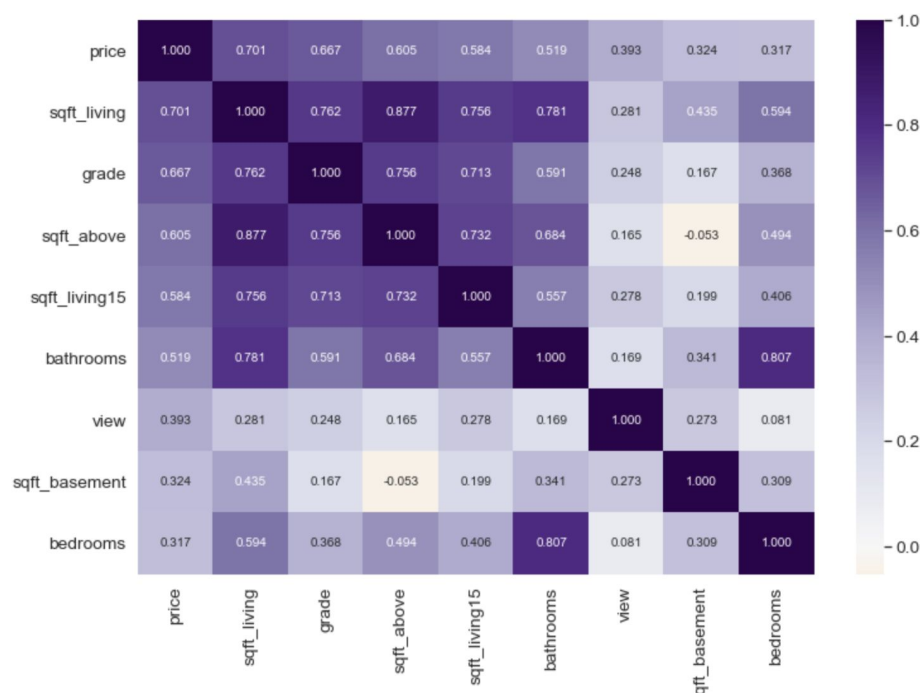
First Results:

House Prices for the King County Data Set roughly follow Log-Normal Distribution:
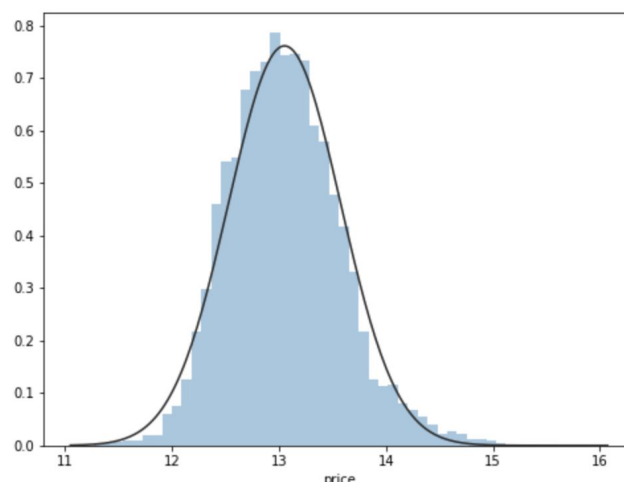
```
Mean Price First Sale:        352263.11
Mean Price Second Sale:       488779.59
Mean Increase in Sale Price: 38.75 %
```


What are the most relevant (non-categorical) variables influencing House Price:


Trying to fit a multiple linear regression model based on most important features, which showed to be the size of the living area, the zipcode of the property and the overall build quality of the house according to the King County grading system:
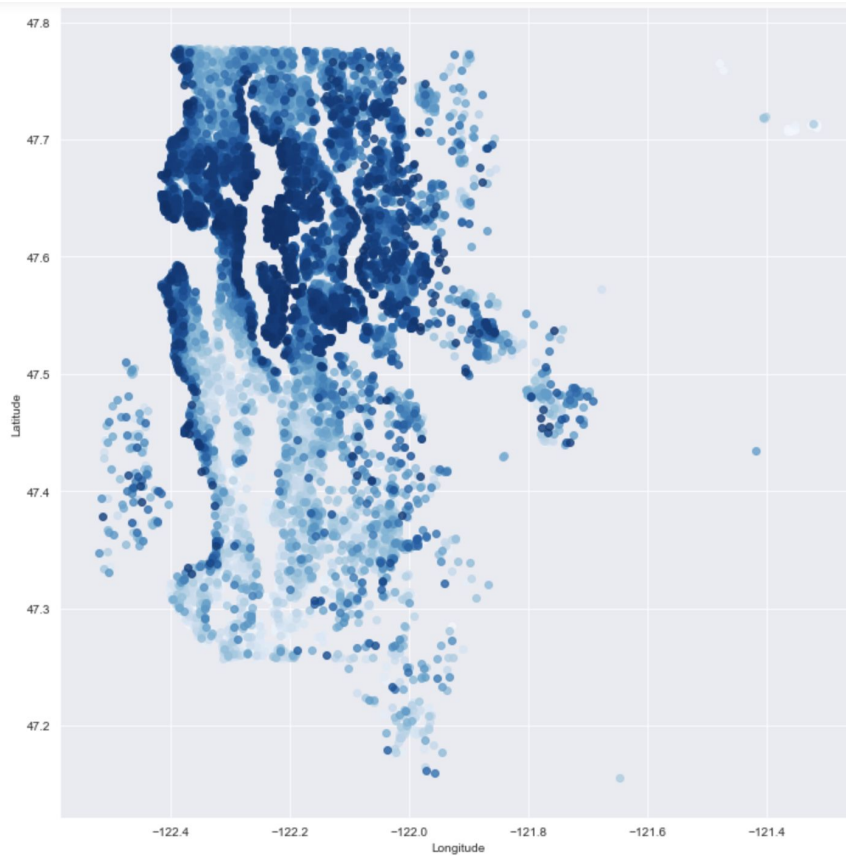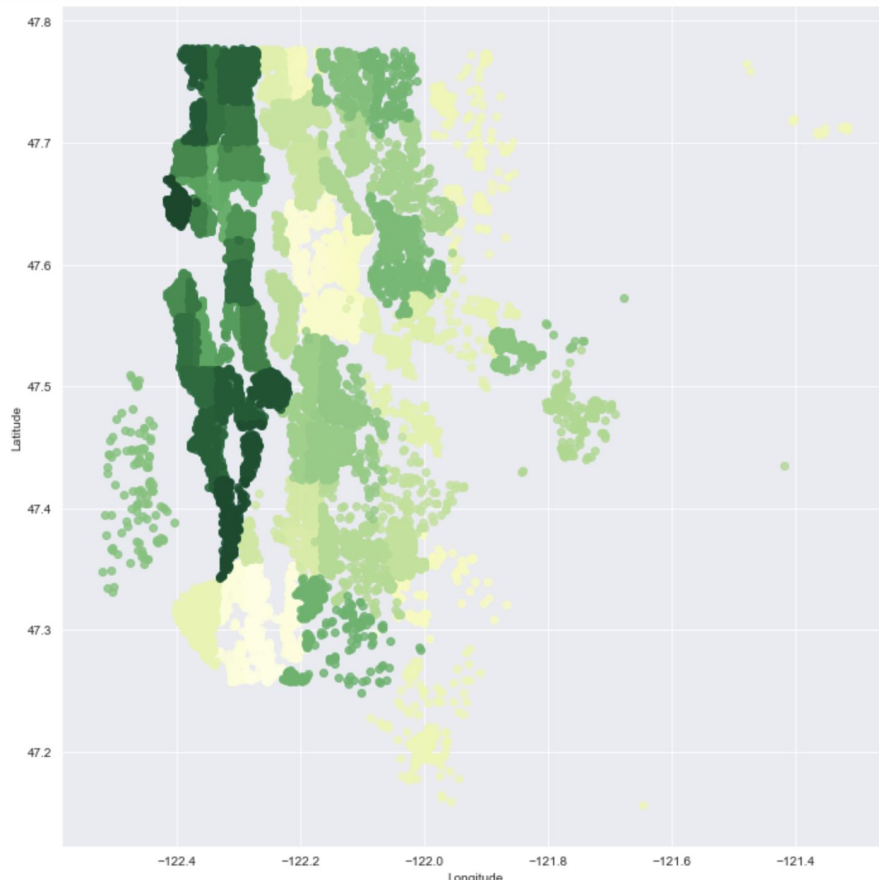
For example, a two dimensional histogram is shown for house price and sqft_living:
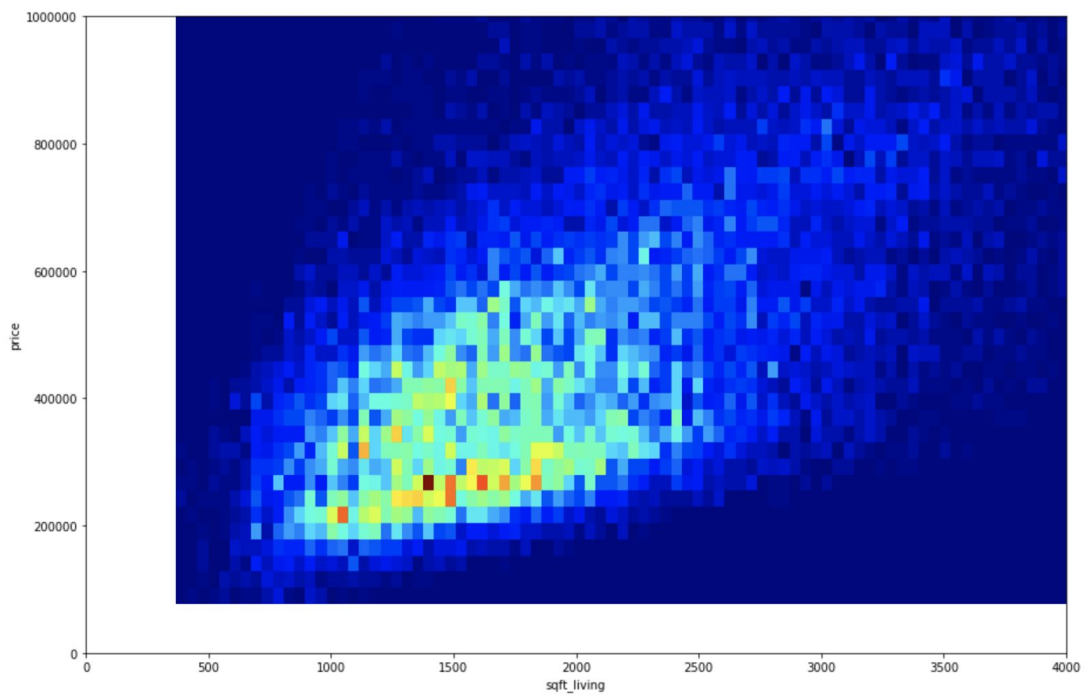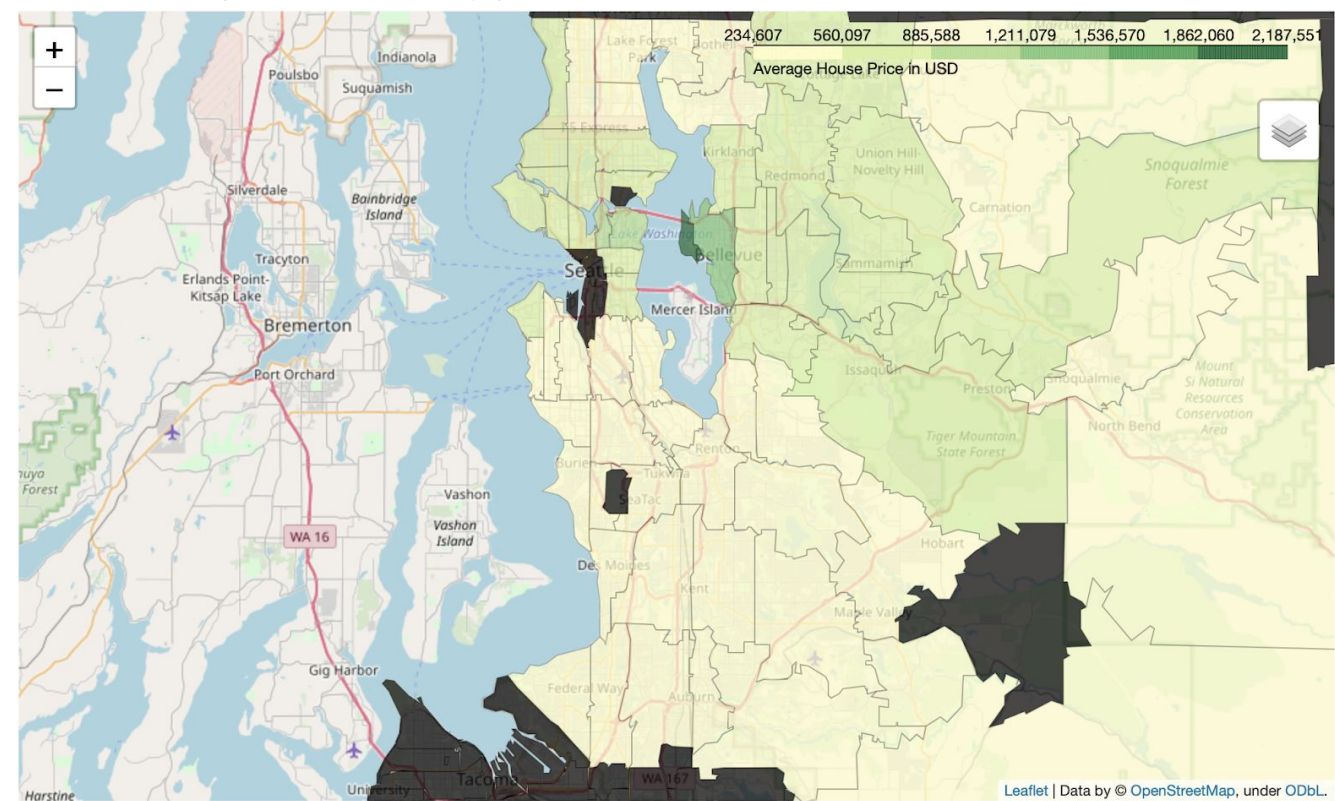

House Prices also increase substantially with the overall quality indicated by grade:
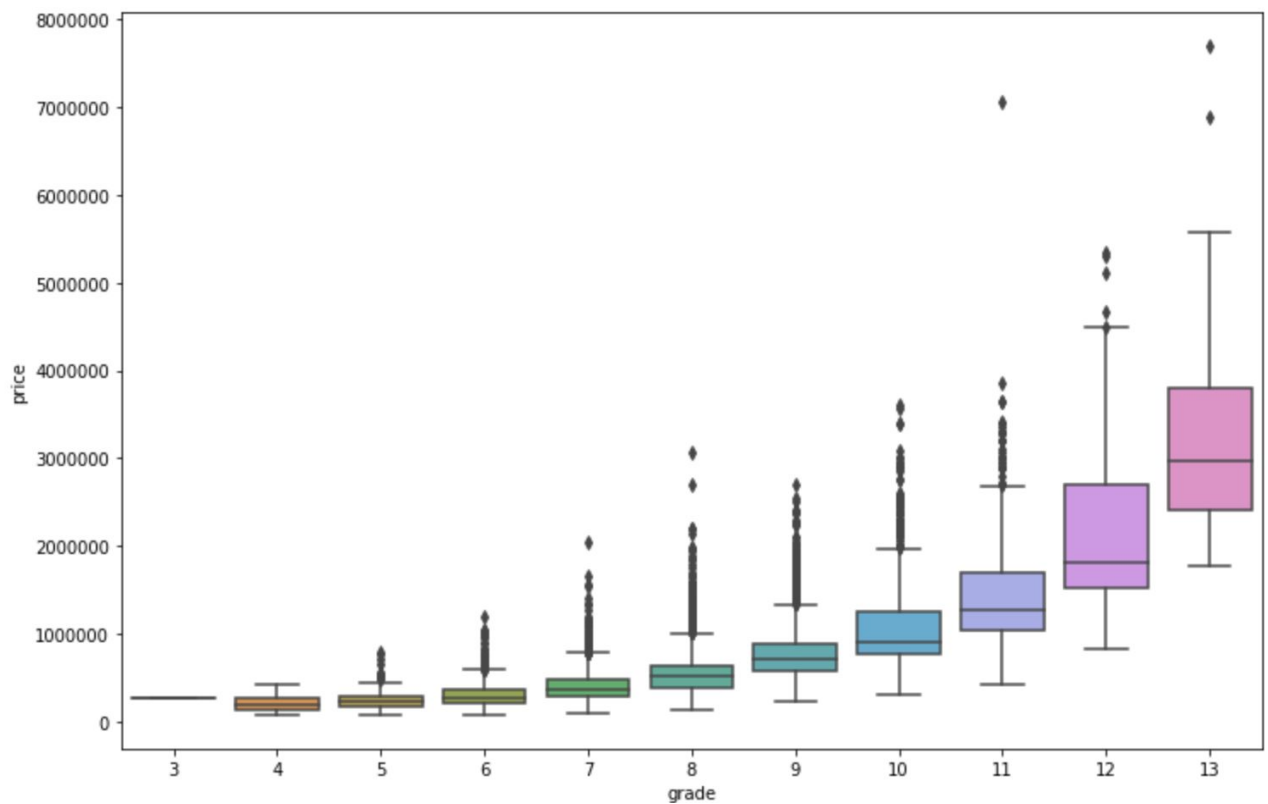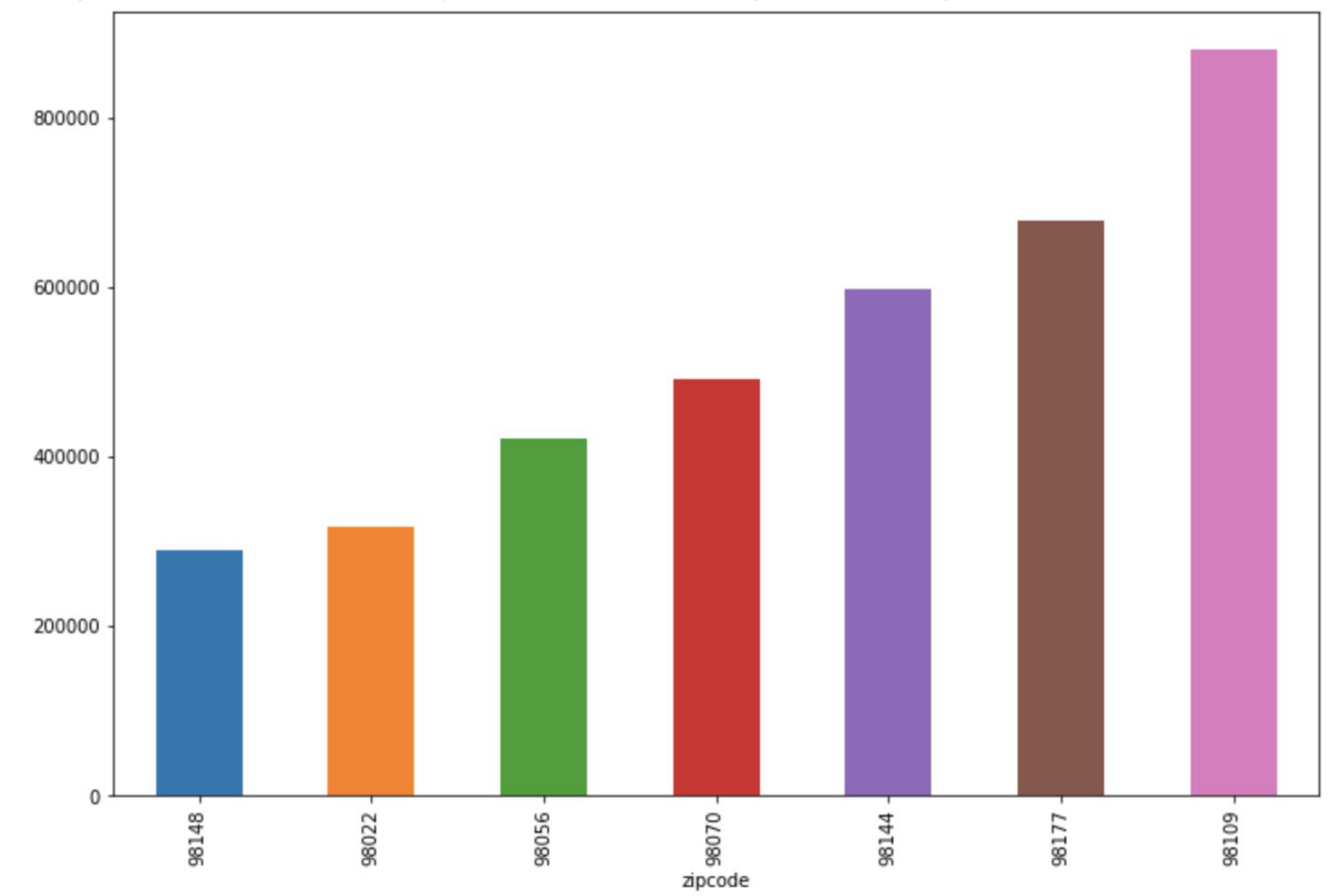
The Zipcode area as a measure of neighborhood had a high impact on house price category.
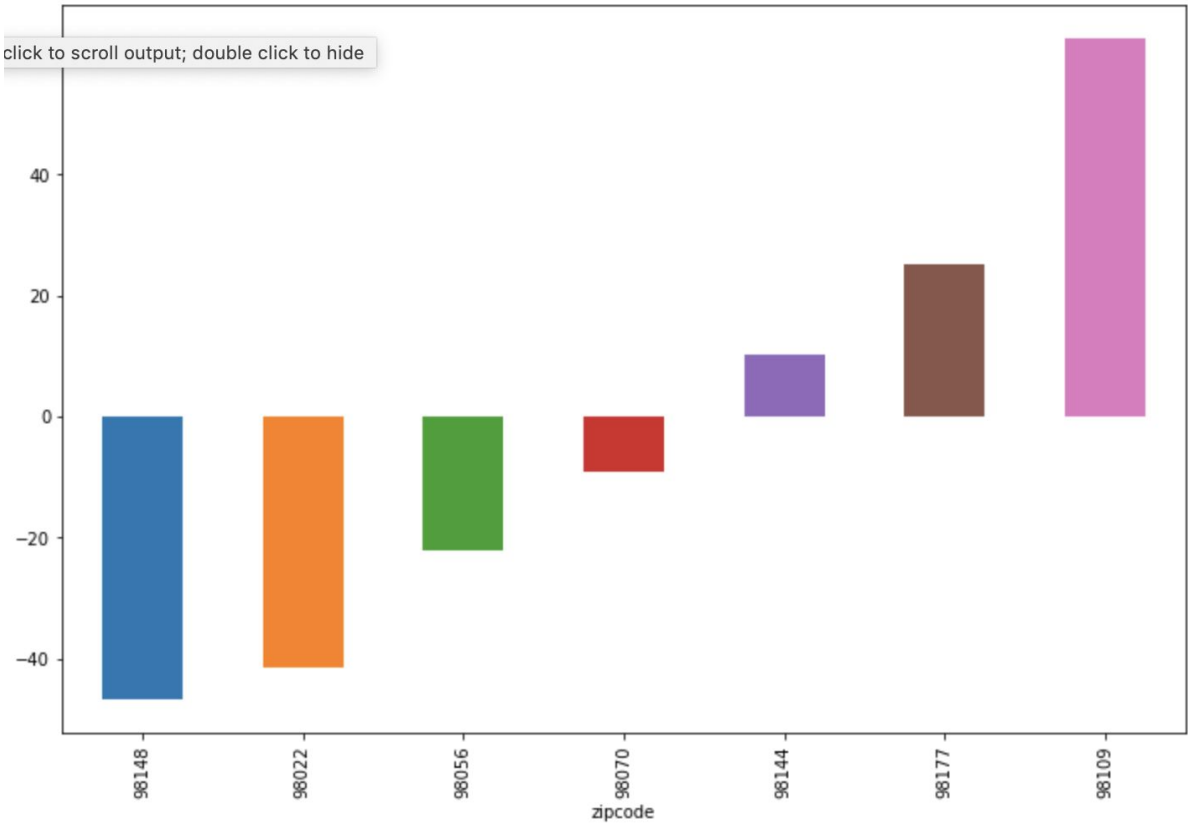Where are the Zipcode areas, and where are the most expansive houses? :

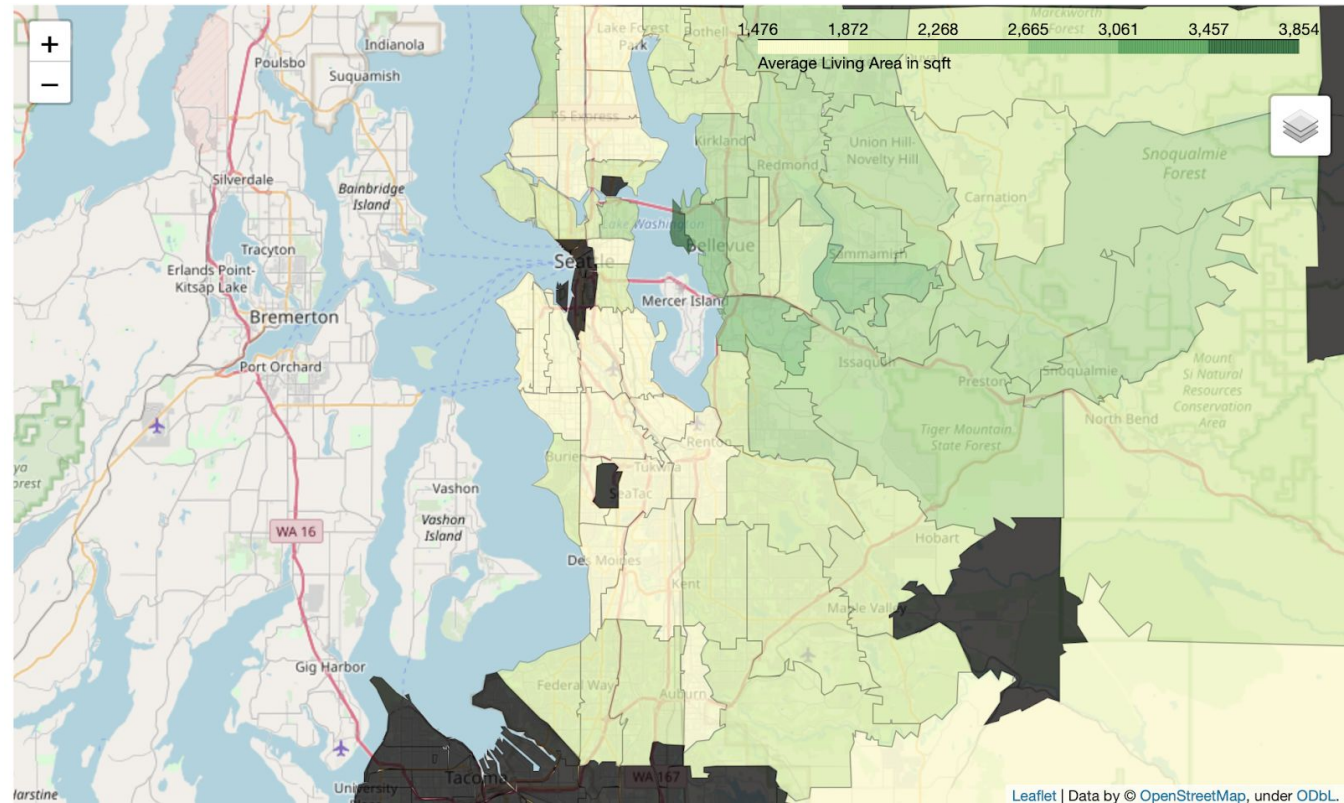What is the average house price in any given Zipcode Area:

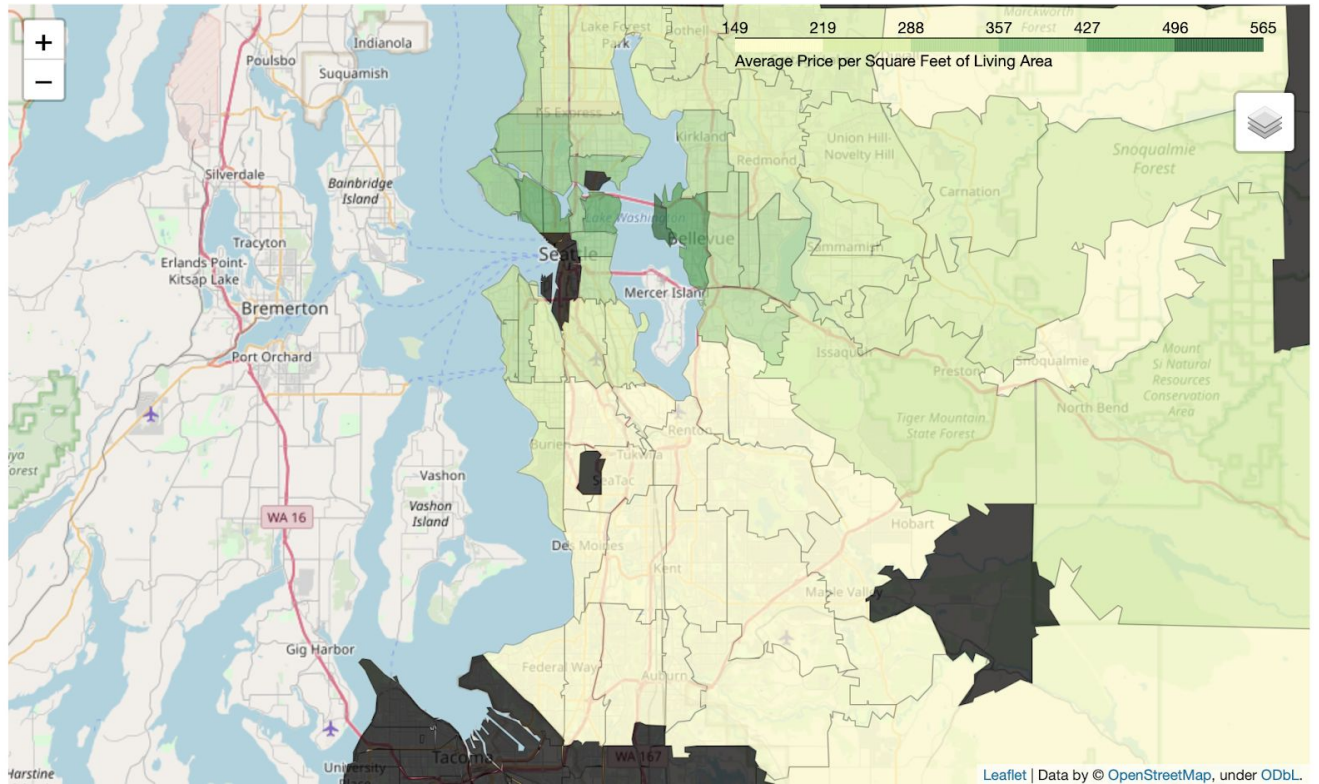Average House Prices for exemplary Zipcode Areas (covering the whole range):

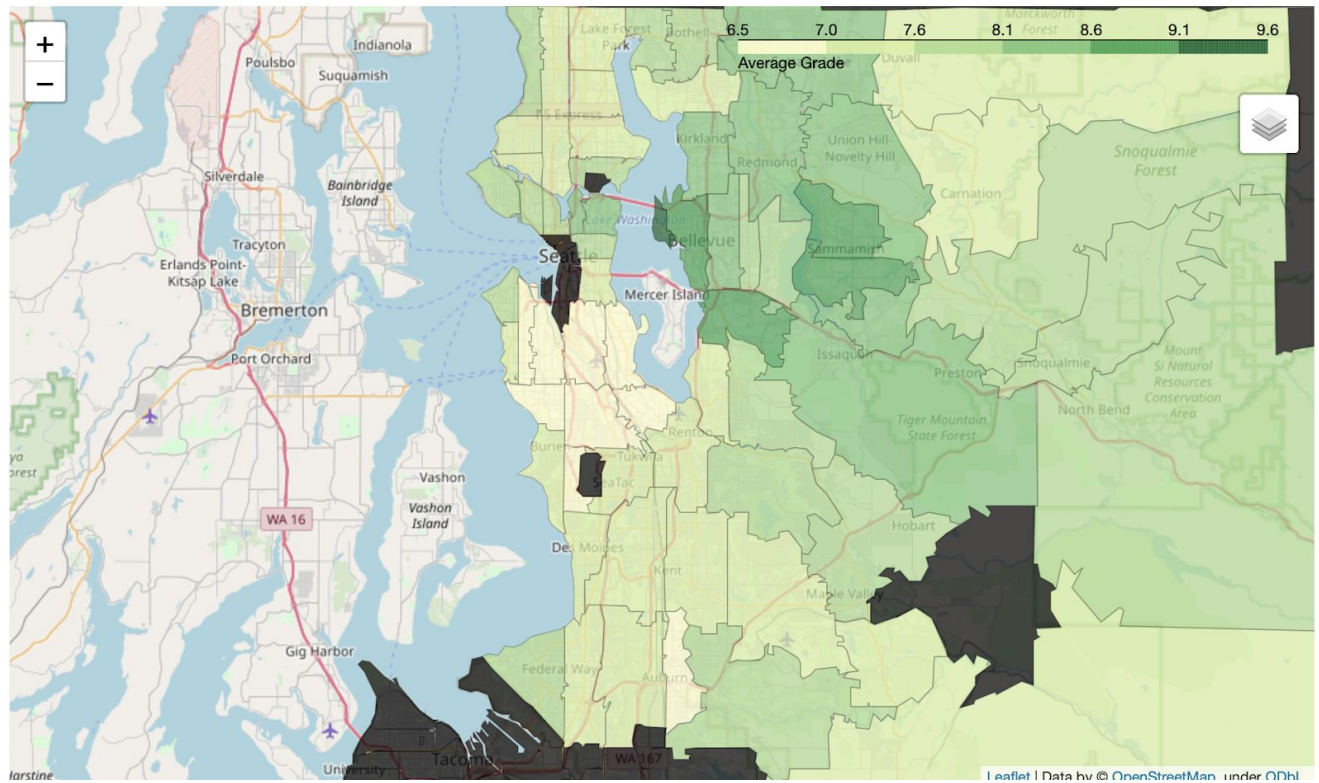Differences to the overall mean house price in percent:

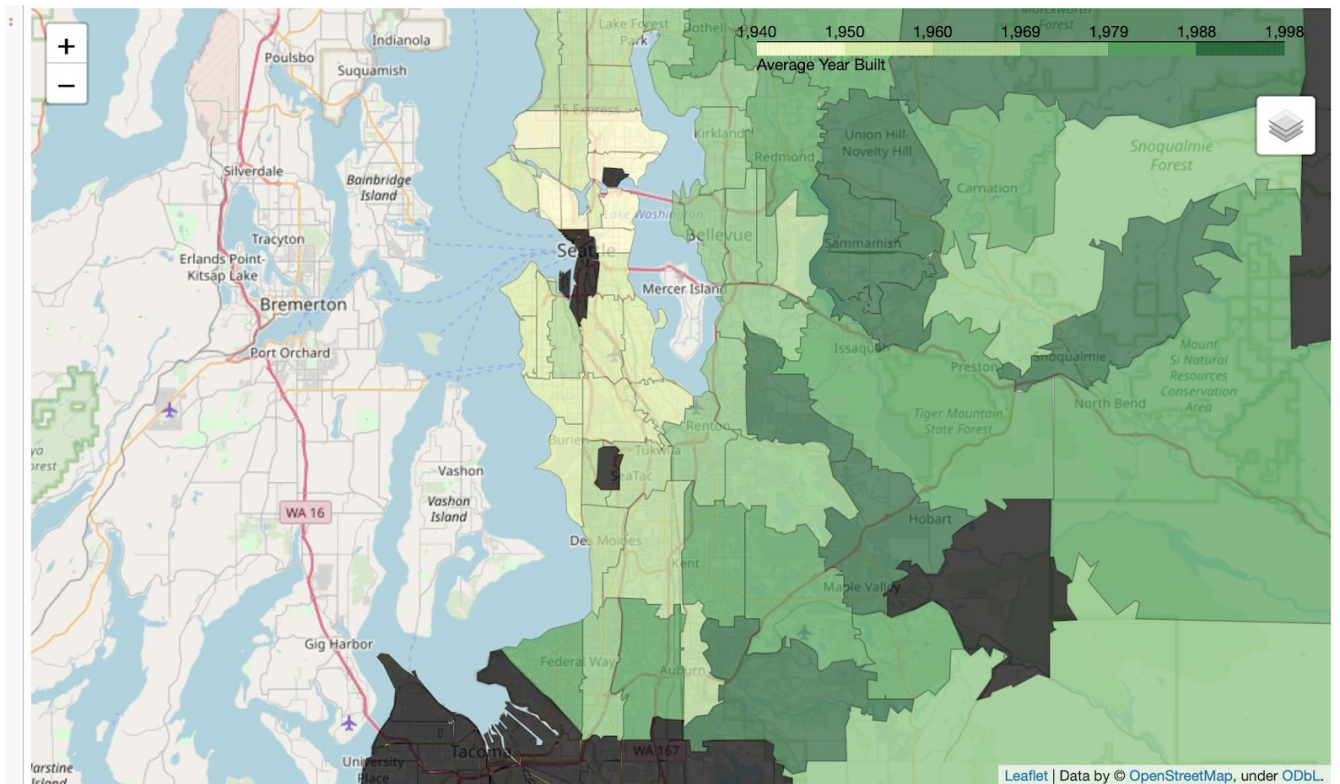What is the corresponding average living area per Zipcode:

What is the average price per square foot in a given Zipcode Area:

What is the average grade of houses in different Zipcode Areas:

What is the average year houses where built in a given Zipcode Area:

Neue Fische Data Science Bootcamp DS2, Daniel Büttner

Is there a "seasonal" difference between House Prices? Shown here are monthly average price differences to the overall average in percent for May 2014 to May 2015. Further exploration is needed. Do all zipcode areas/price categories show this signal? Or is this signal an artefact due to more houses being sold in more expansive categories? ---> Future work. If this proofs to be robust, a recommendation could be derived, for sellers to try to sell in Spring time, and for buyers, to try to buy in wintertime
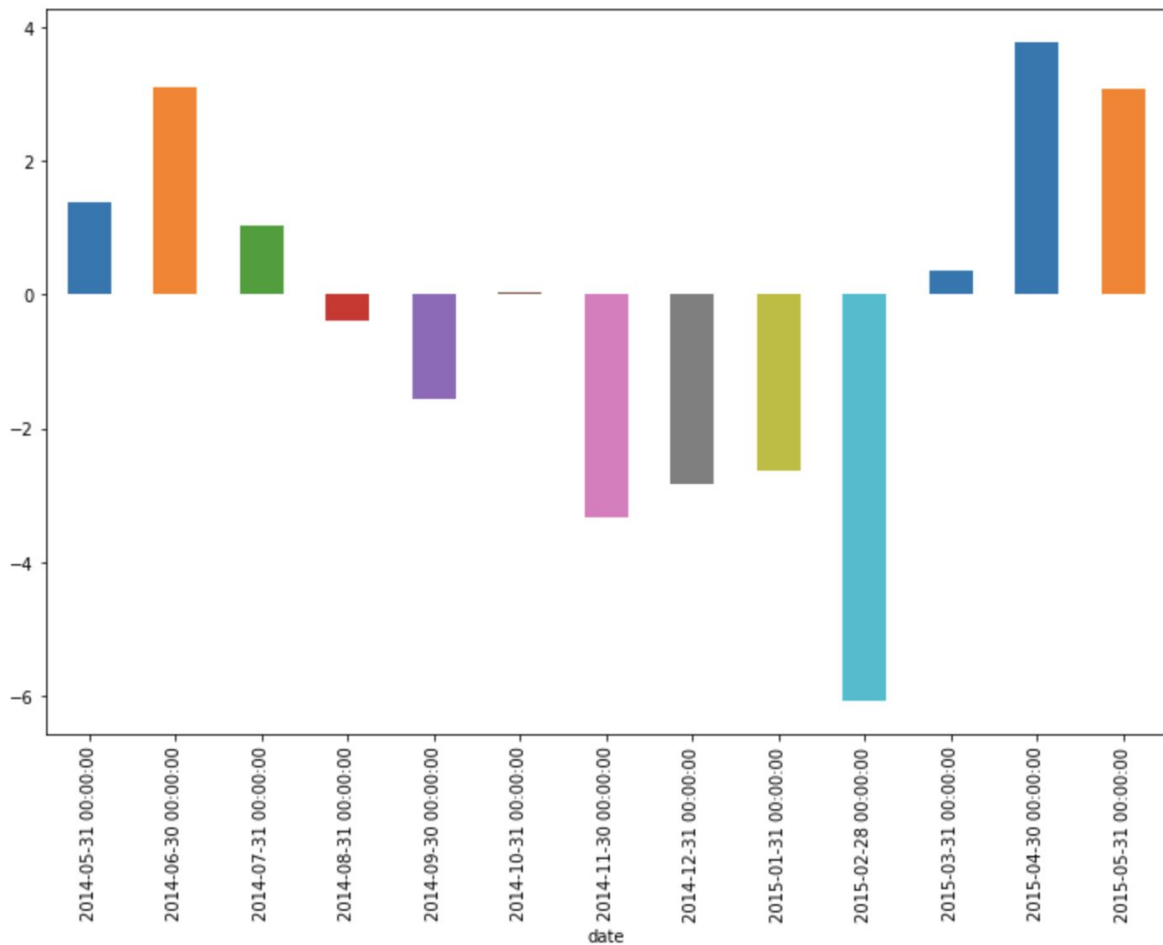
Shown here is the average price per house ordered by the year the house was built in. Year built showed no significant influence during the regression analysis, but it proofs to be the missing ability for a regression model to capture the structural information here: Old houses and new houses are more expansive than mid-century houses.
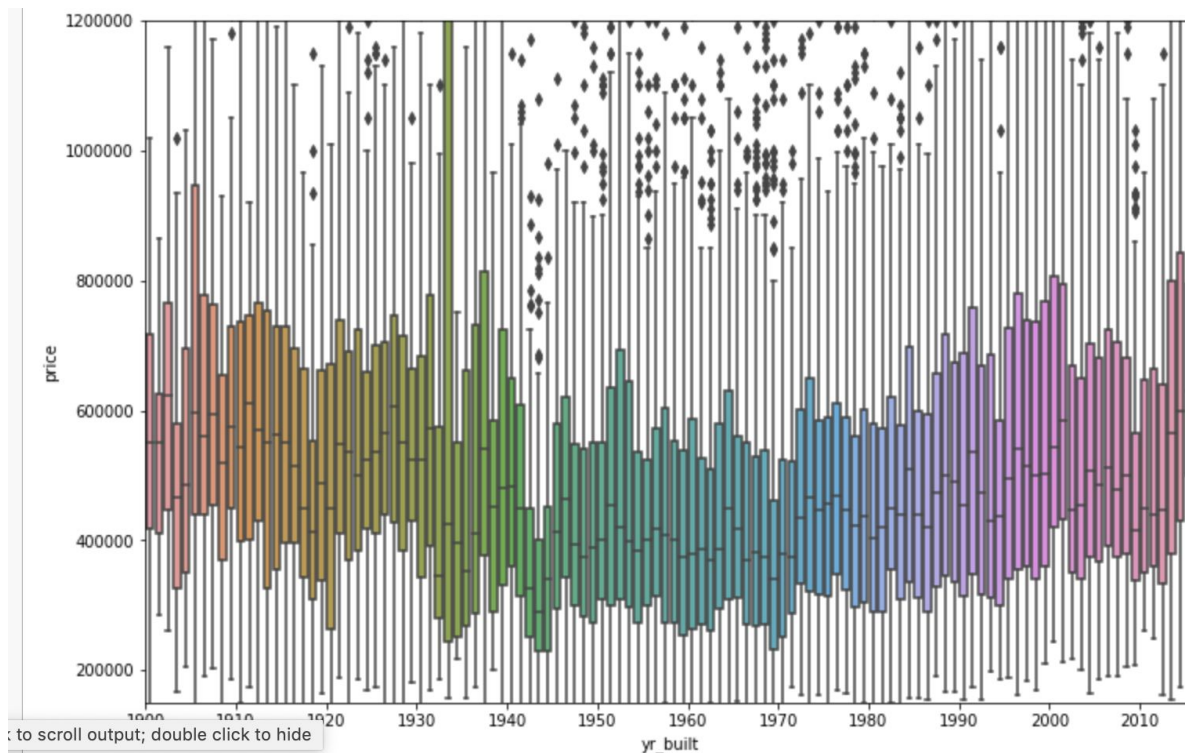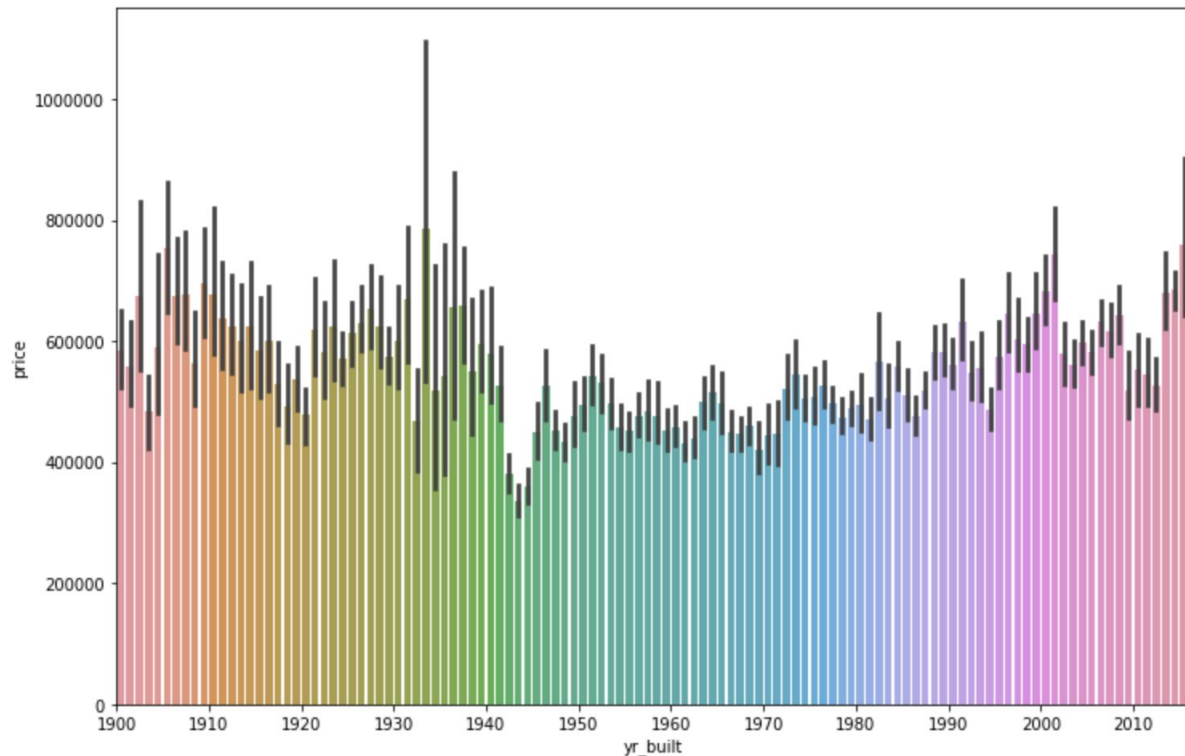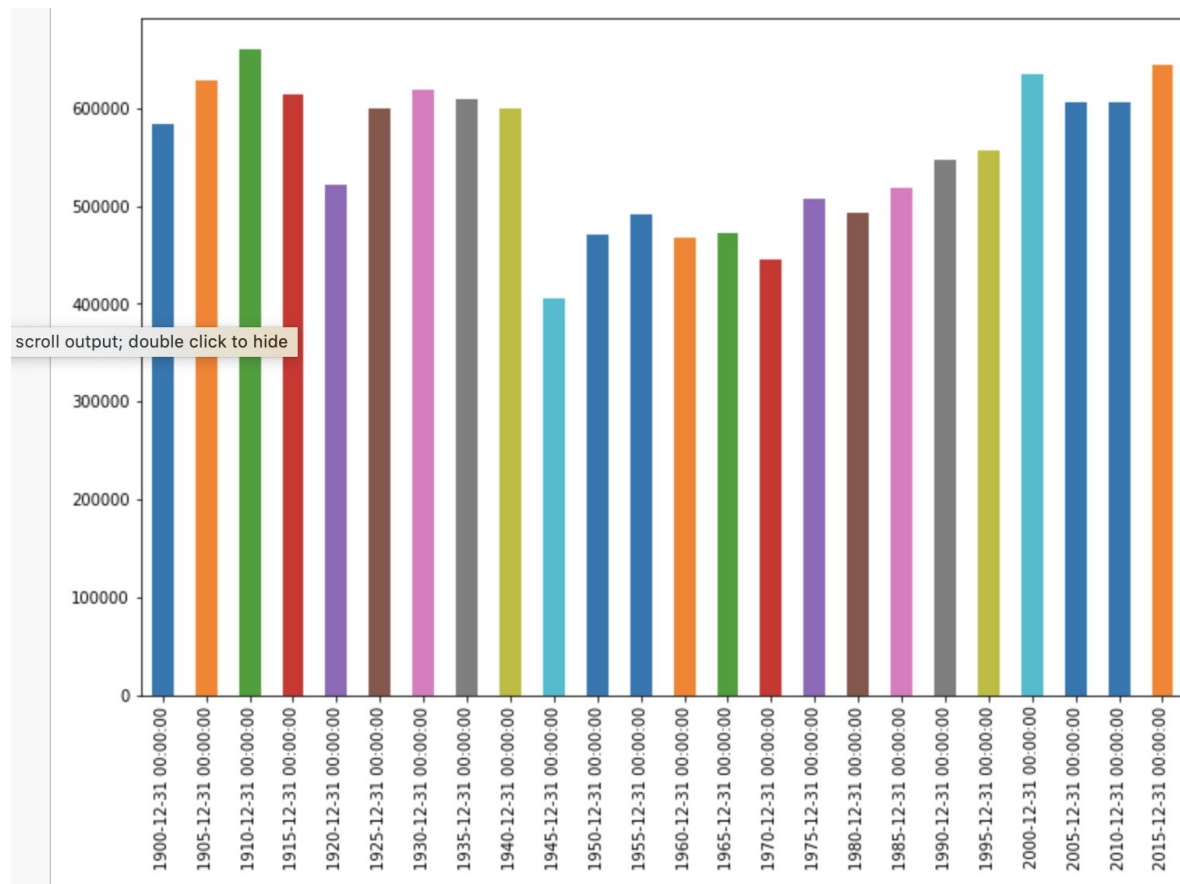
What do temporal differences in the size of the living area look like? Medium sized houses early 20th Century, small houses built during WW2, after that steady increase through 50s/60s, and afterwards. Biggest estates in modern era, but: recession of 2008.
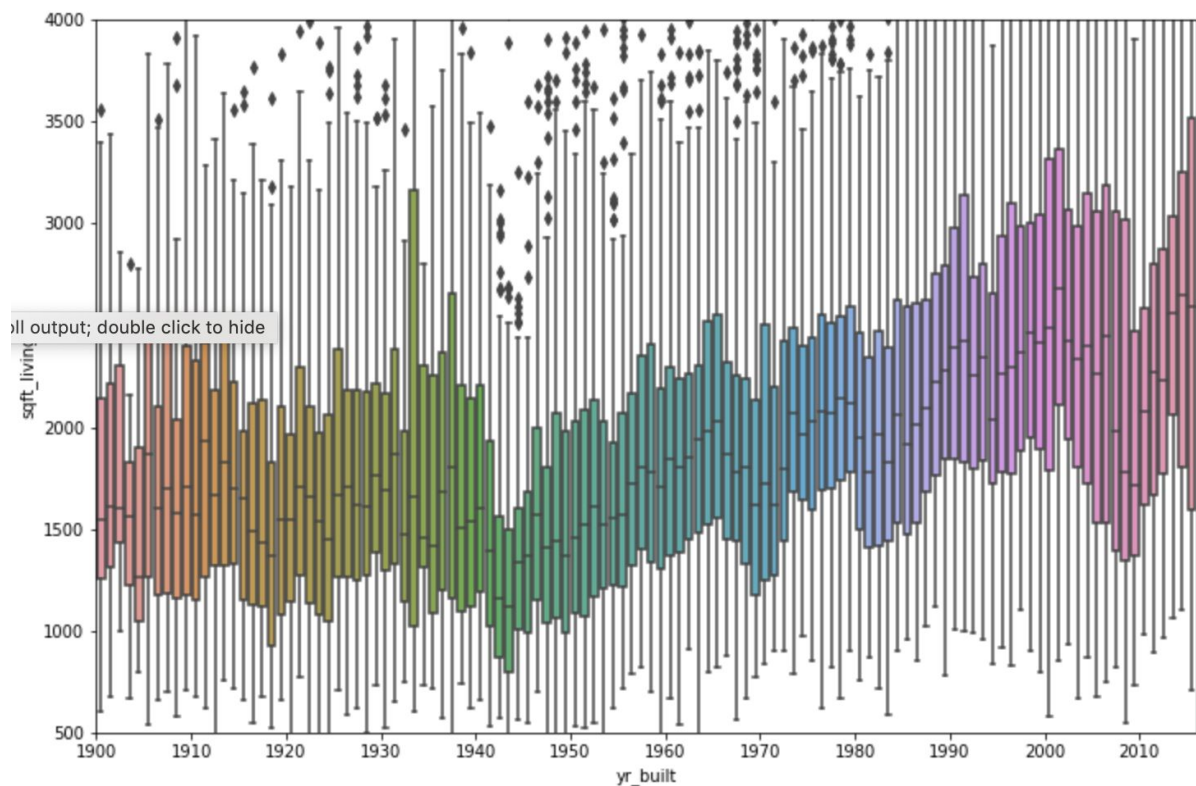
Possible recommendations:

for buyers who are looking for small to medium sized houses for a decent price: look for mid-century properties and/or properties in certain zipcode areas (tbd).

for sellers, or business sellers who are trying to sell for a high price: invest in the overall quality of the property, as long as expected increase of sales price exceeds the investments (tbd). Try to sell during spring time, as prices might be higher then (but this remains to be shown, tbd).

Future work: building a business case on the current findings.

Thanks You for Your Attention!!