

ANALYSIS OF HOUSE PRICES KING LOUIS ESTATE / USA

PRESENTATION 17.09.2020

Britta Röver

AGENDA

- Approach to the Data / Business Understanding
 - Data Cleaning
 - Data Analysis and Exploration
 - Feature Engineering
 - Predictive Modelling
 - Data Visualization
-
- Findings through the Data

DATA

column		type	values	
id	Unique identifier	int		
date	Date house was sold	object		
price	price	float		outliner
bedrooms	Number of bedrooms	int	discret/1-11	outliner
bathrooms	Number of bathrooms	float	discret/0-7	
Sqft_living	Square footage/house	int		
Sqft_lot	Square footage/lot	int		
floors	Number of floors	float	discret/1-4	
waterfront	View to the waterfront	float	1/0	Missing values
view	Indicator for a special view	float	ranking 0-4	Missing values
condition	Indicator for condition	int	discret/1-5	
grade	Indicator for grade / base on a local grading system	int	discret/1-13	

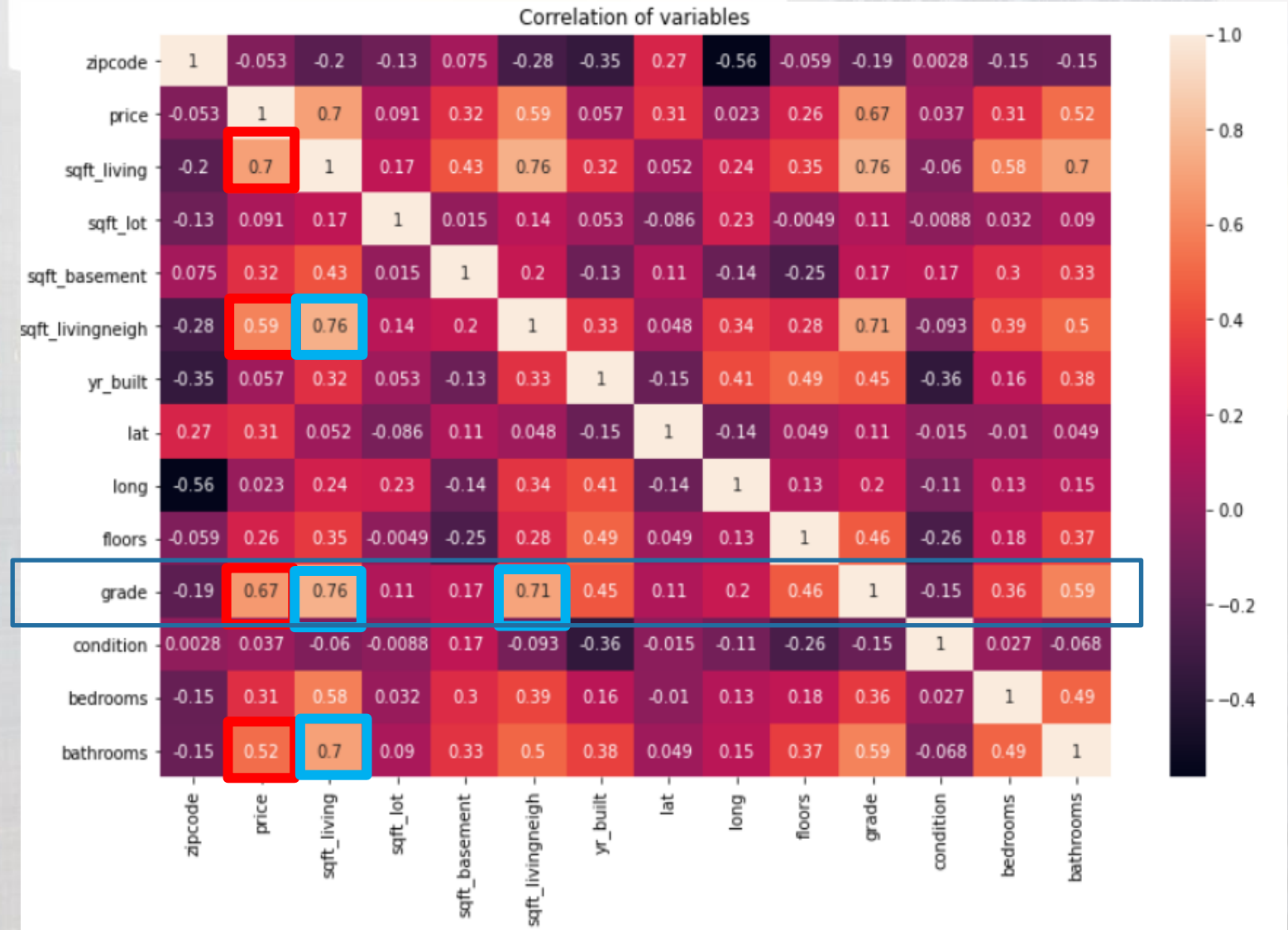
DATA

column		type	values	
Sqft_above	Square footage without basement/house	int		Can be skipped
Sqft_basement	Square footage basement	object		Missing values Wrong data New computation of the value additional feature basement y/n
Yr_built	Built year	float	discret	additional features yr_age and yr_built_cat
Yr_renovated	View to the waterfront	float	discret, 1/0	additional feature renovation
zipcode	Indicator for a special view	float	discret, ranking 0-4	
Lat long	Coordinates (latitude and longitude)	Both float		
Sqft_living Sqft_lot	Square footage /living and lot of the nearest 15 neighbours	int		

GET A BIG PICTURE!

Matrix of correlations

- There seems to be a couple of stronger correlations concerning the price
 - Living space
 - Grade
 - Distance to neighbour
 - bedrooms
- But there are also dependencies between the features, see feature grade



1. SHOT:

Multilinear Regression

- features used
 - Living and lot space
 - grade
 - distance to neighbour
 - Bed- and bathrooms ..
- result: **R2 of 0,704**

OLS Regression Results

Dep. Variable:	price	R-squared:	0.704
Model:	OLS	Adj. R-squared:	0.704
Method:	Least Squares	F-statistic:	3946.
Date:	Thu, 17 Sep 2020	Prob (F-statistic):	0.00
Time:	08:19:18	Log-Likelihood:	-2.9370e+05
No. Observations:	21594	AIC:	5.874e+05
Df Residuals:	21580	BIC:	5.875e+05
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	9.627e+06	2.79e+06	3.445	0.001	4.15e+06	1.51e+07
zipcode	-583.2118	32.000	-18.225	0.000	-645.935	-520.489
waterfront	3.056e+05	8818.216	34.653	0.000	2.88e+05	3.23e+05
yr_built	-2649.0045	64.848	-40.849	0.000	-2776.112	-2521.897
bedrooms	-3.233e+04	1827.460	-17.691	0.000	-3.59e+04	-2.87e+04
bathrooms	4.151e+04	2631.057	15.778	0.000	3.64e+04	4.67e+04
floors	2.654e+04	3048.637	8.705	0.000	2.06e+04	3.25e+04
sqft_living	152.7727	3.185	47.968	0.000	146.530	159.015
sqft_livingneigh	32.7844	3.319	9.878	0.000	26.279	39.290
grade	1.019e+05	2075.170	49.082	0.000	9.78e+04	1.06e+05
condition	2.404e+04	2241.796	10.724	0.000	1.96e+04	2.84e+04
view	5.035e+04	2032.661	24.769	0.000	4.64e+04	5.43e+04
lat	5.879e+05	1.03e+04	56.986	0.000	5.68e+05	6.08e+05
long	-1.972e+05	1.23e+04	-16.017	0.000	-2.21e+05	-1.73e+05
waterfront	3.056e+05	8818.216	34.653	0.000	2.88e+05	3.23e+05

STORY:

A new home for me, my husband and my 3 daughter

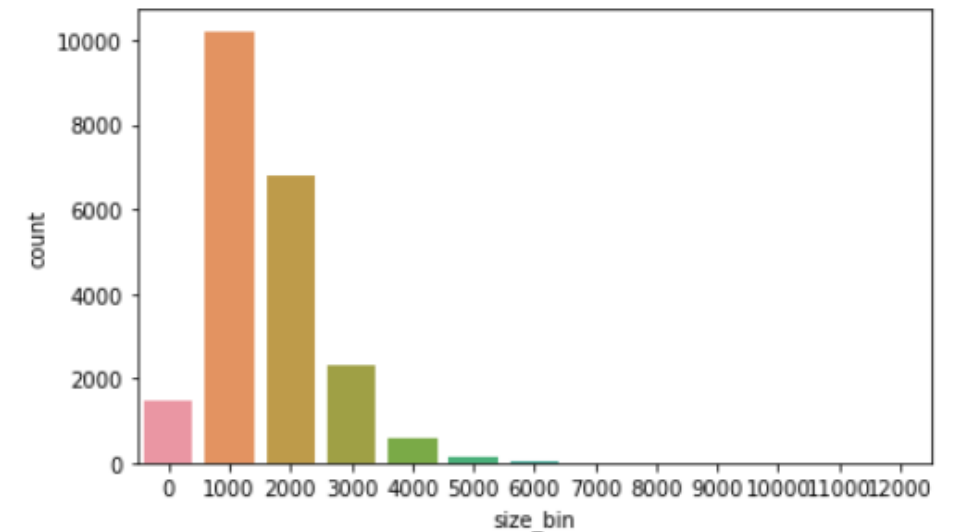
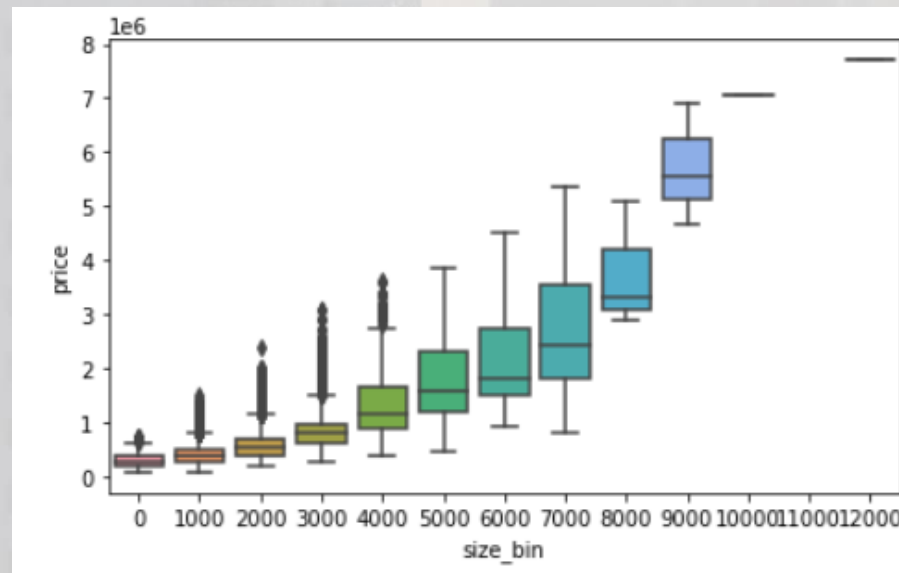
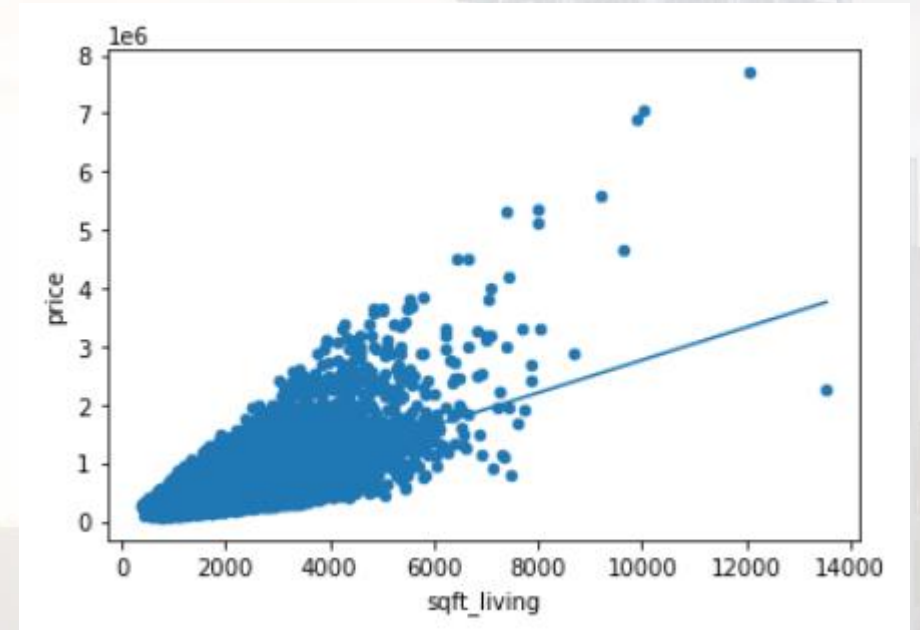
WHAT IS IMPORTANT FOR US?

- The size that means sqft_living and sqft_lot, strong condition for sqft_living ≥ 130 qm²
- The number of bath- and bedrooms:
 - bedrooms at least 4
 - bathrooms at least 2
- The house should have a big basement
- It should have a good condition/grade
- Nice Place! Waterfront!
- Old house?

SIZE:

Finding: SIZE LOT RULES SIZE LIVING!

- There is a strong relationship between the size of the house and the price
- But: 77 % of all houses fulfill the size criteria

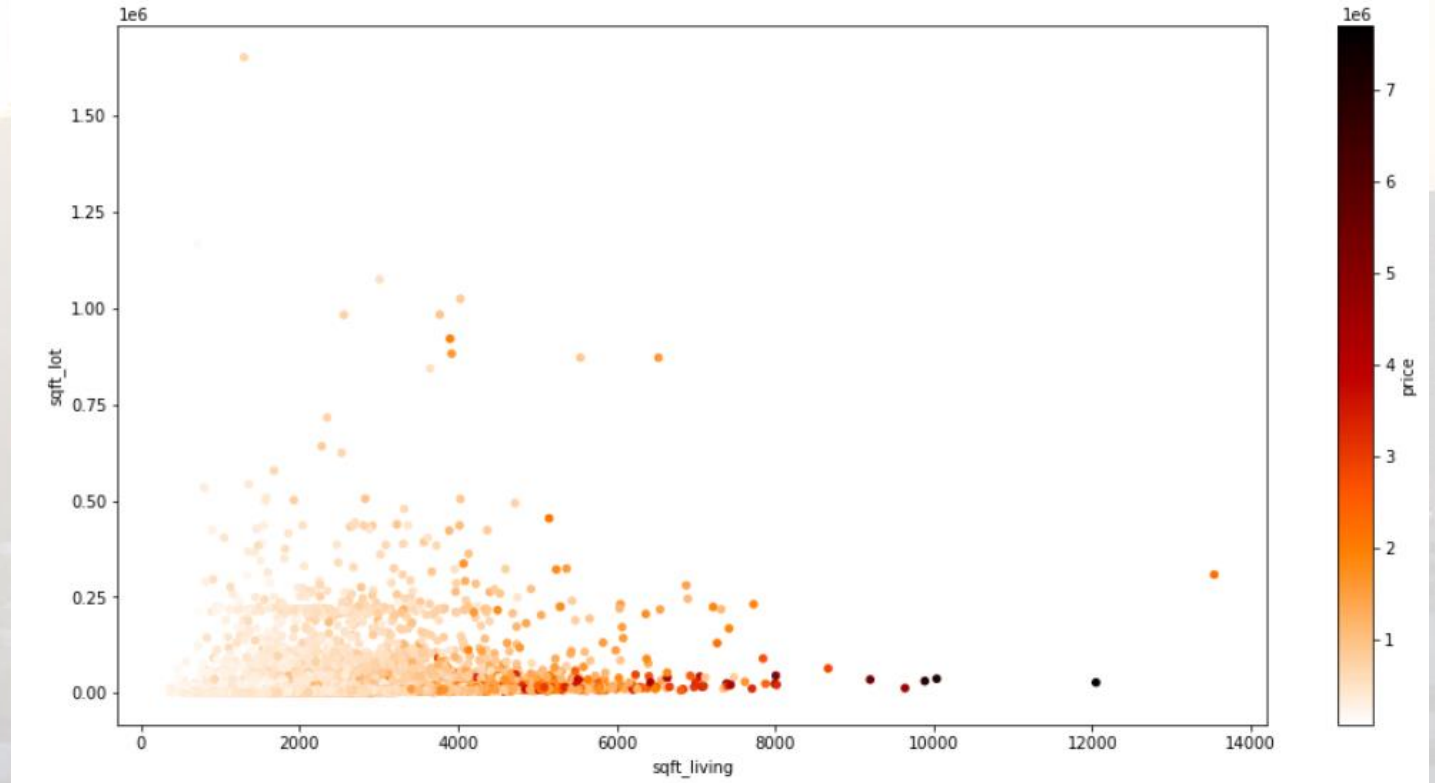
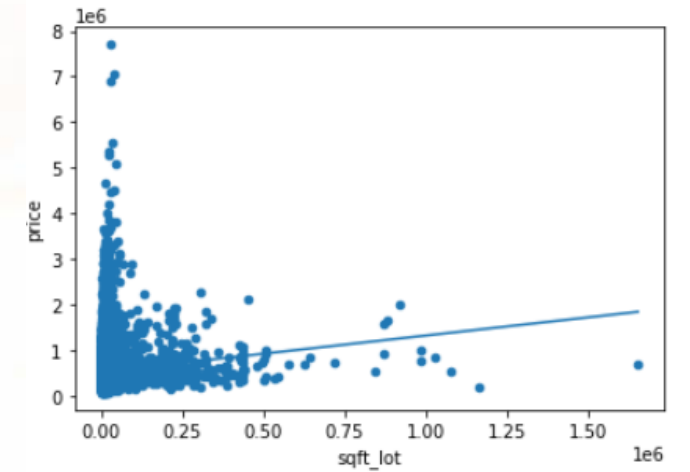


SIZE (2):

Finding: SIZE LOT RULES SIZE LIVING!

- There is no strong relationship between the size of the lot and the price on the one hand and between lot and house size on the other hand

- **BIG LOT and MINIMUM HOUSE SIZE**

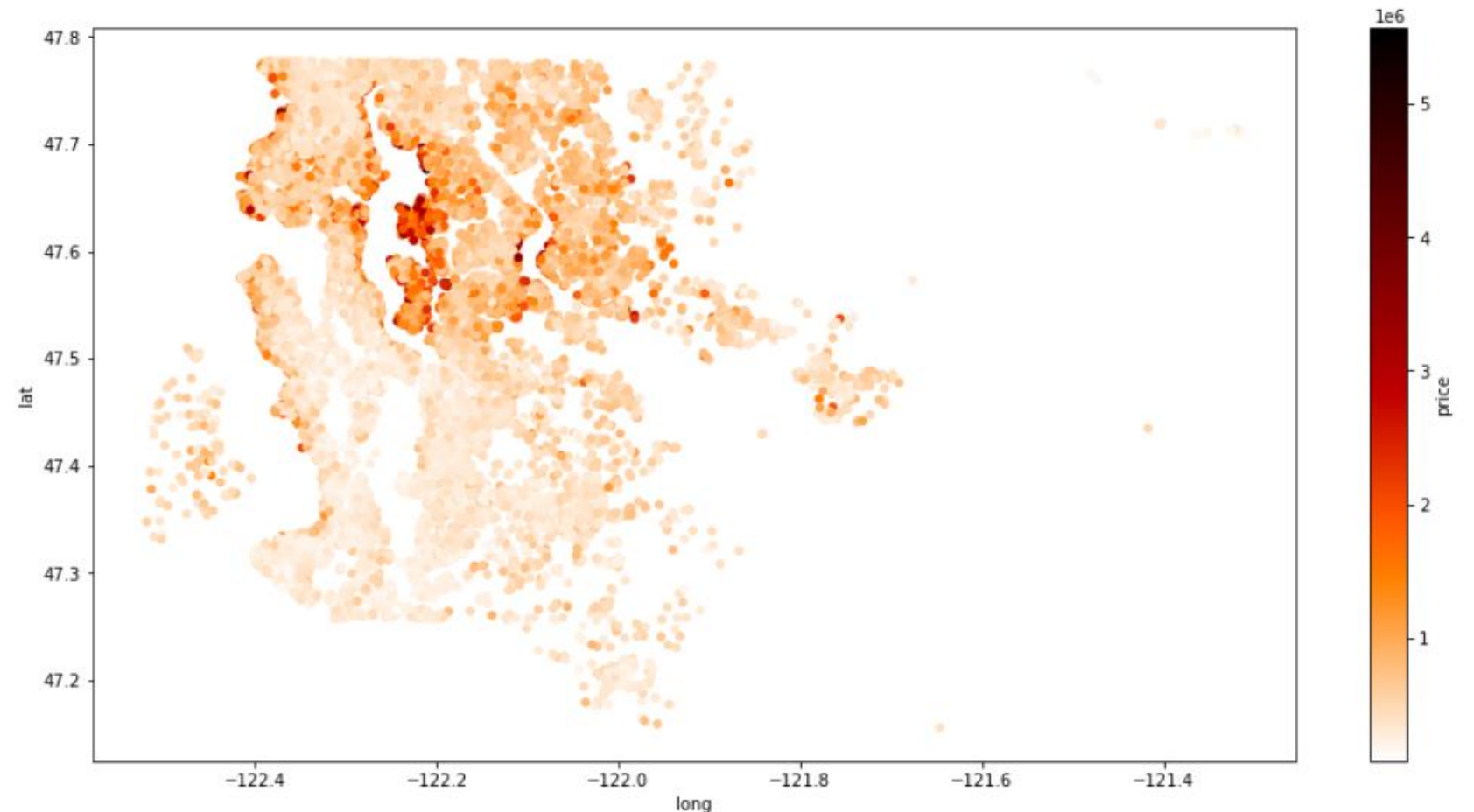


NICE PLACE: Finding : Waterfront is the most expensive option

- But the prices varies
- Maybe the west part of the lake is also nice?
- Have a look!

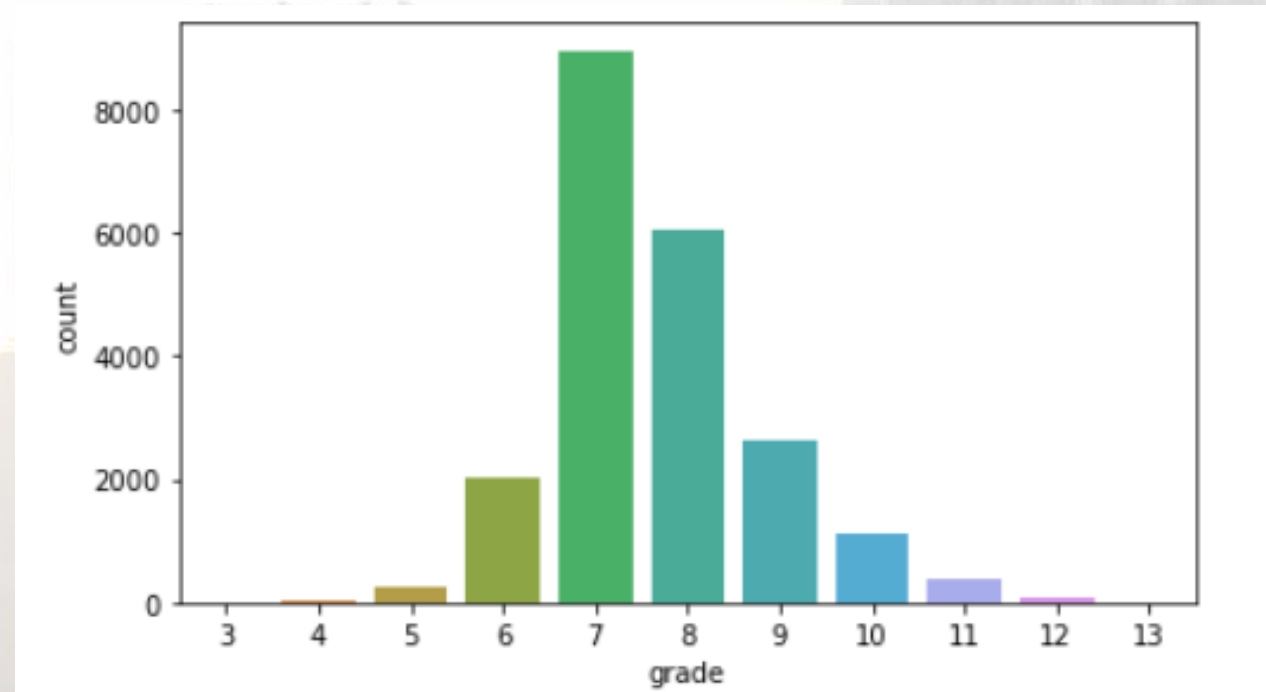
Analysis of columns long, lat

```
immo.plot(kind="scatter", x="long", y="lat", figsize=(16, 8), c="price",  
          cmap="gist_heat_r", colorbar=True, sharex=False);  
plt.show();
```

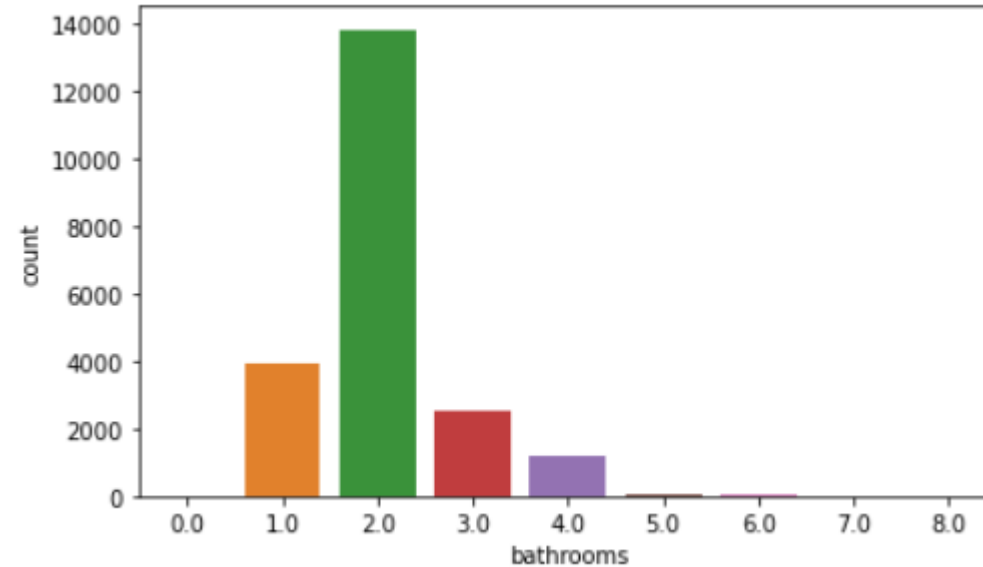
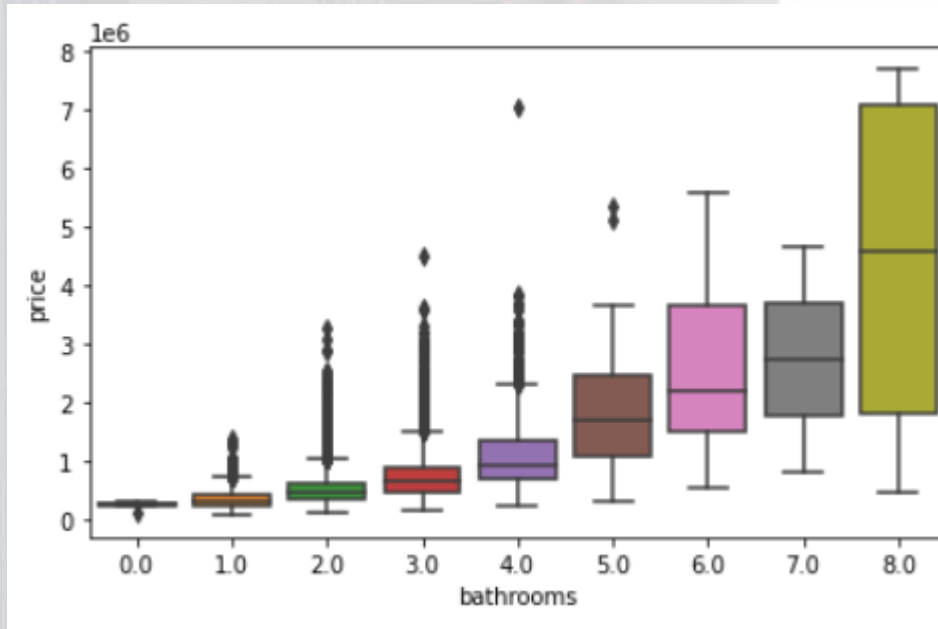


NICE PLACE: Finding 2: the grade is not so important for us!

- Not fully clear by definition
- Strong relationship between grade and bathrooms and bedroomnumbers
- This both is important to us!
- Grade should be at least average (grade==7)
- But at the waterfront only 3 objects have a grade <7

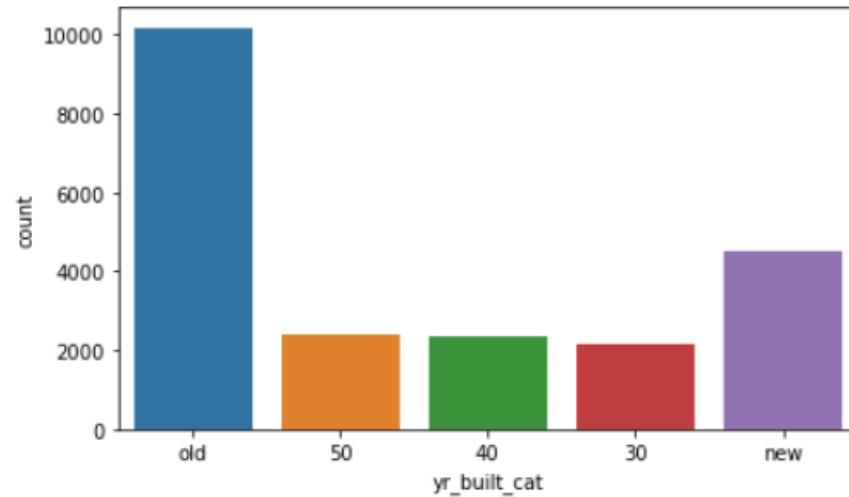
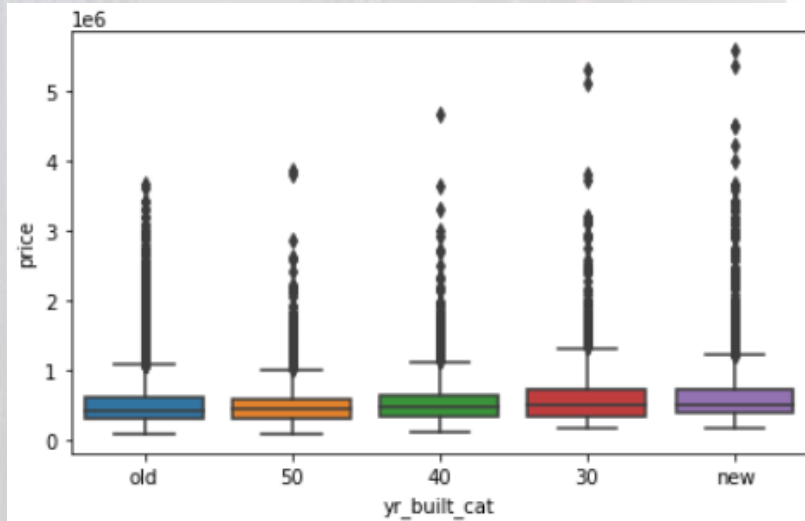


BATHROOMS: Finding 3: 2 Bathrooms are enough!

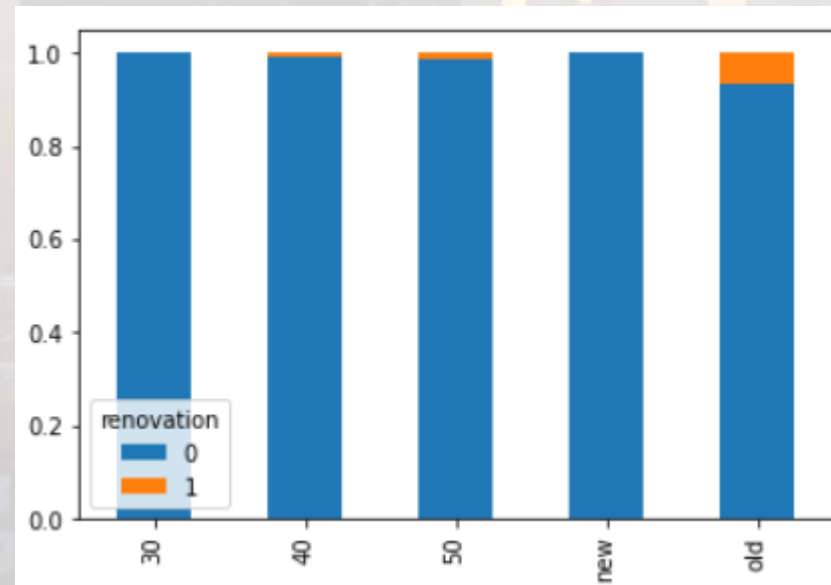


- Bathrooms become more expensive by the size of 2!
- 2 Bathrooms are enough! This also seems to be the standard size

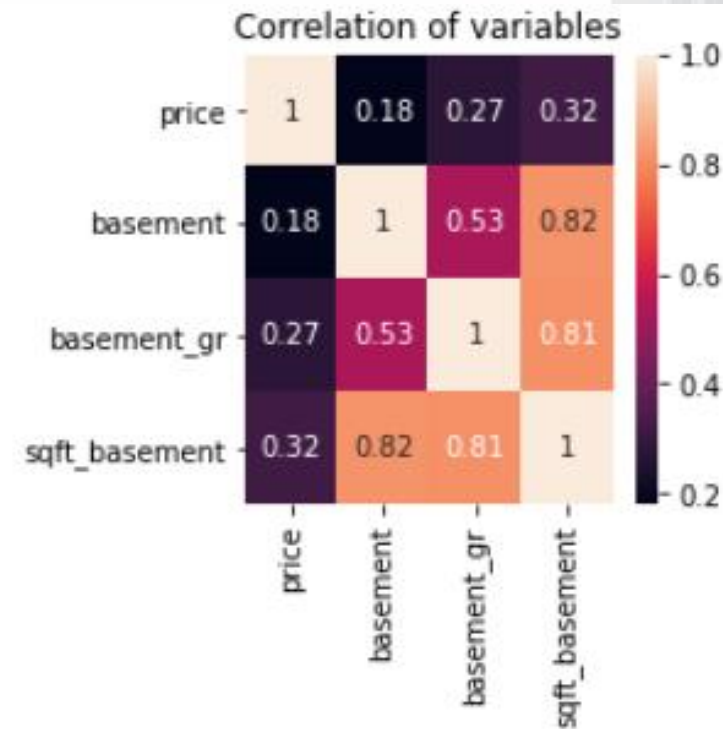
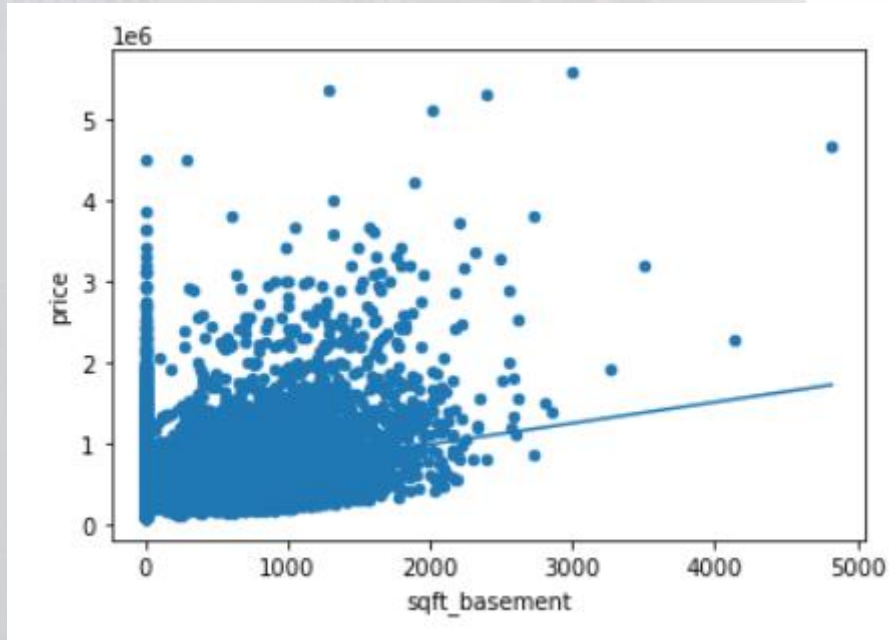
OLD HOUSE: Finding 4: only a small part has been renovated!



- Have a look!



BASEMENT: Finding 4: size is not so expensive!



- The fact basement yes or no has also a important effect to the price!
- The Size is not so expensive

STORY:

A new home for me, my husband and my 3 daughter

WHAT IS IMPORTANT FOR US?

- The size that means sqft_living and sqft_lot, strong condition for sqft_living ≥ 130 qm2 YES
- The number of bath- and bedrooms:
 - bedrooms at least 4
 - bathrooms at least 2
- The house should have a big basement
- It should have a good condition/grade ≤ 7
- Nice Place! Waterfront may be!
- Old house? may be!



THANKS!