House Flippers

Pricing which suits you

Project 1

September 11th, 2019

Columbia | Engineering
The Fu Foundation School of Engineering and Applied Science

DATA ANALYTICS BOOTCAMP - GUILLERMO SAEZ

Agenda

House Flippers

Exploratory Analysis

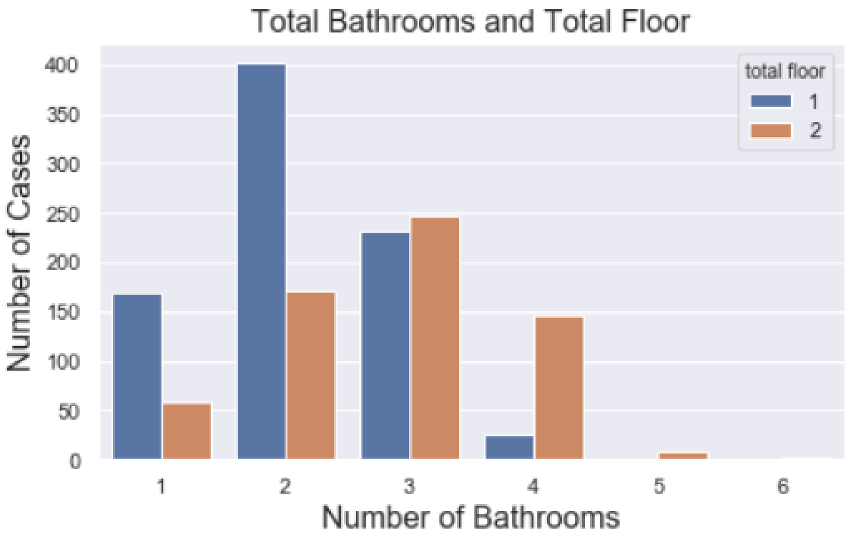
Data Cleaning

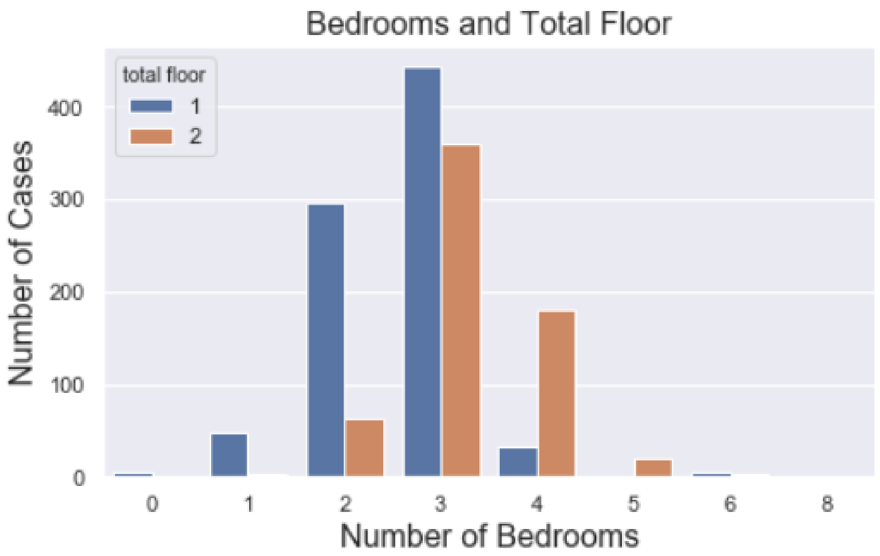
Model

Model Details

Conclusion

Most of the properties have between 2 and 3 bathrooms and the majority has 3 bedrooms

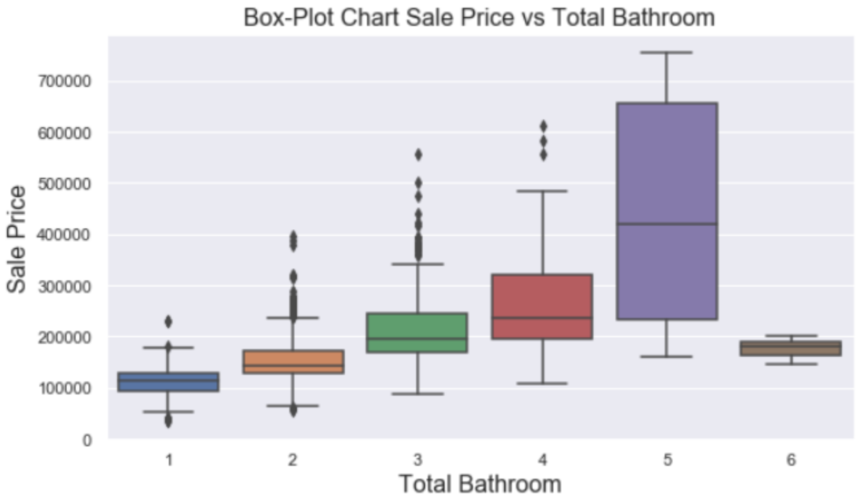


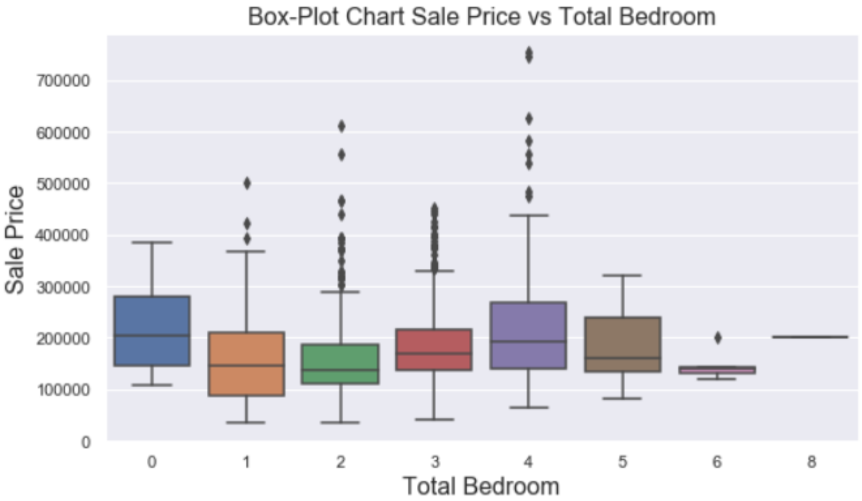


House Flippers

Source: train1 dataset Ames, Iowa, Kaggle

The highest variation of prices occurs when we compare Sale Price with Number of Bathrooms

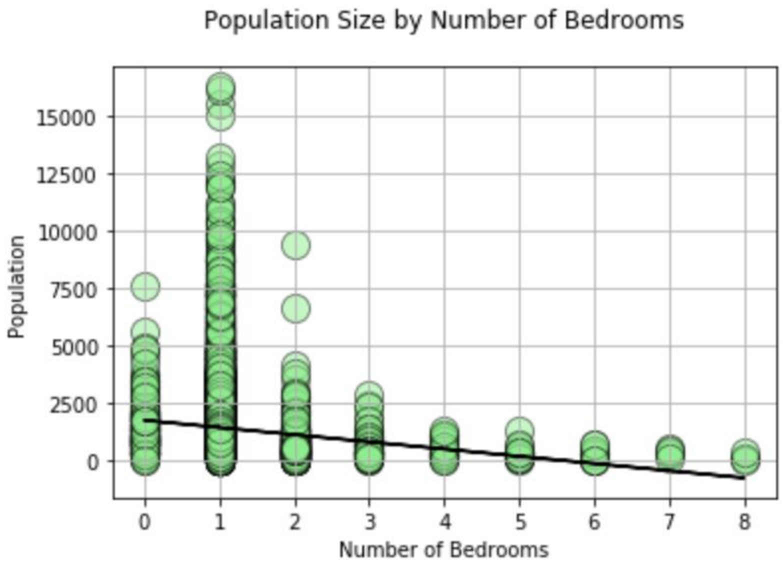




House Flippers

Source: train1 dataset Ames, Iowa, Kaggle

Trends in Sales Price and Population by Number of Bedrooms

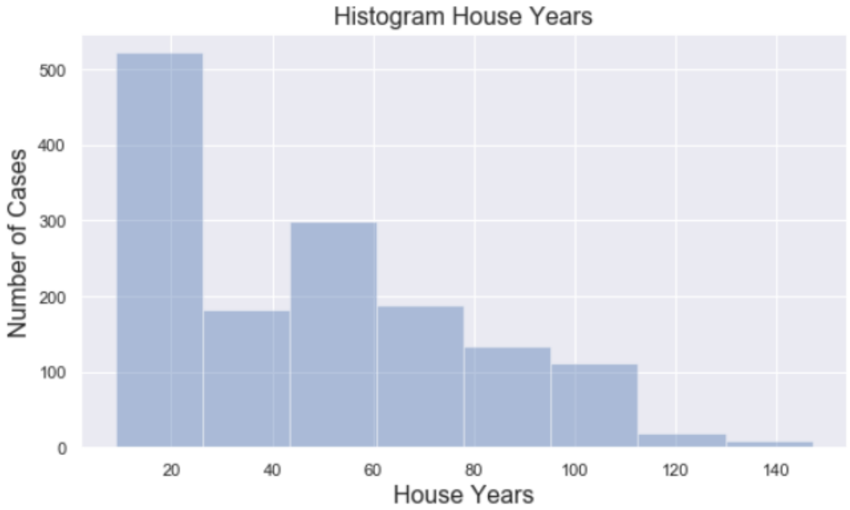


House Flippers

Source: train1 dataset Ames, Iowa, Kaggle

As we expected the older the property, the lower the price will be

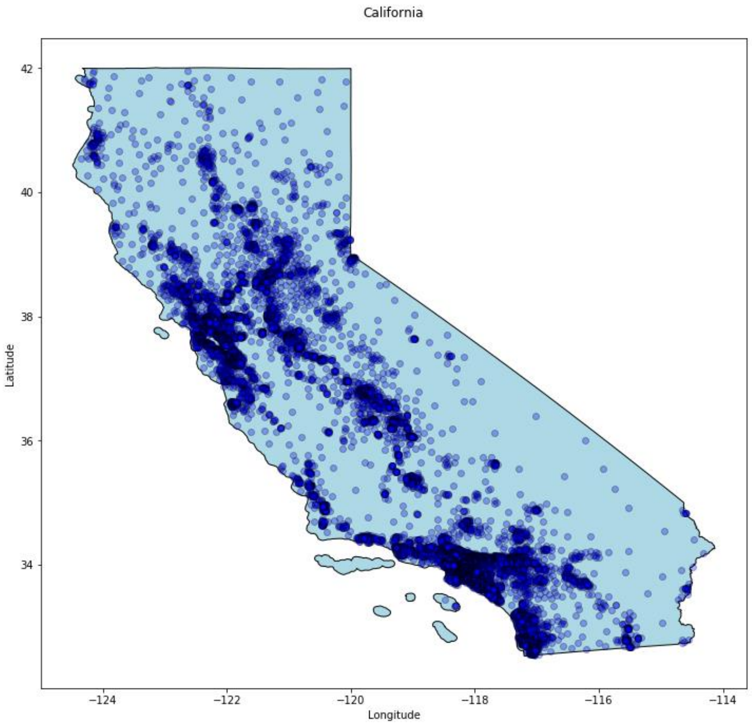
Year Built	Count Sale Price	Avg Sale Price	Std Sale Price
0 to 15	275	248,398	88,616
16 to 30	278	228,168	80,565
31 to 45	166	172,229	48,086
46 to 60	281	148,860	43,246
61 to 75	185	135,350	37,856
76 to 90	93	138,966	62,535
91 to 105	129	127,834	43,988
>105	52	137,238	70,245



House Flippers

Source: train1 dataset Ames, Iowa, Kaggle

California Population Density and Sales Price by Population

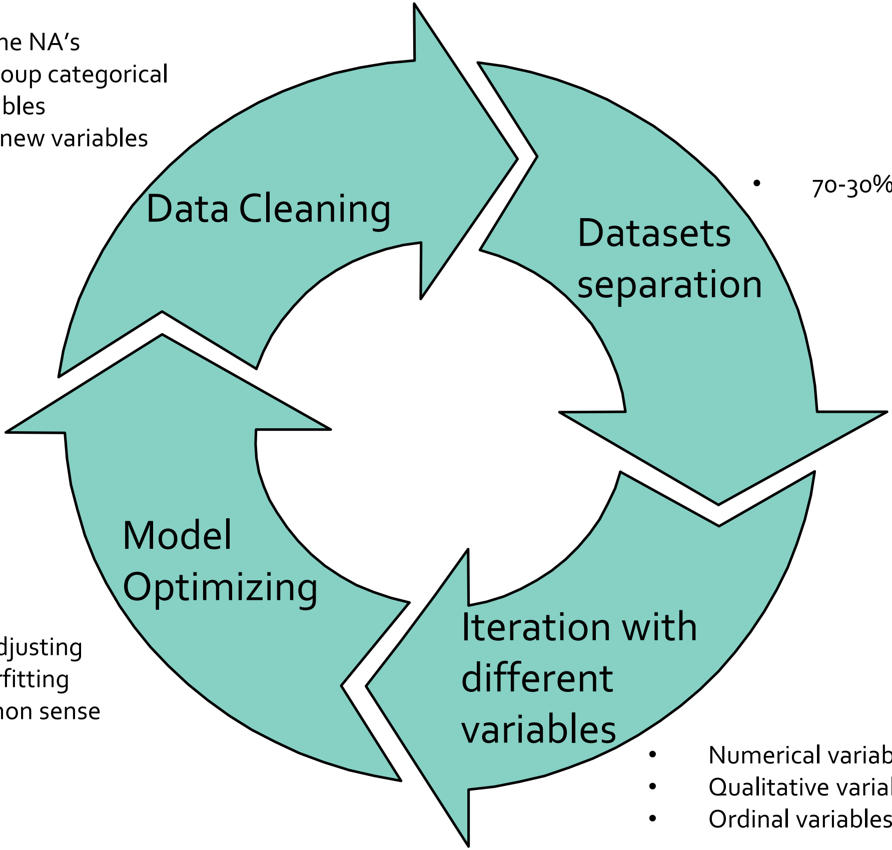


House Flippers

Source: train1 dataset Ames, Iowa, Kaggle

The price prediction model was built through an iterative process applying analytics tools…

- Fill the NA’s
- Regroup categorical variables
- Add new variables



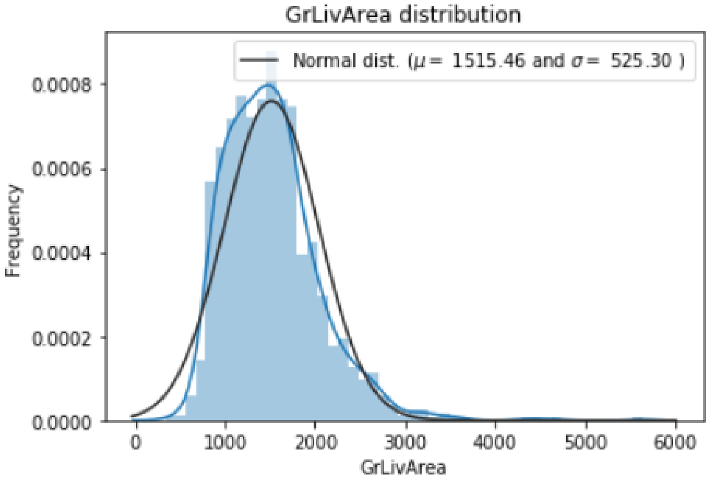
- 70-30%

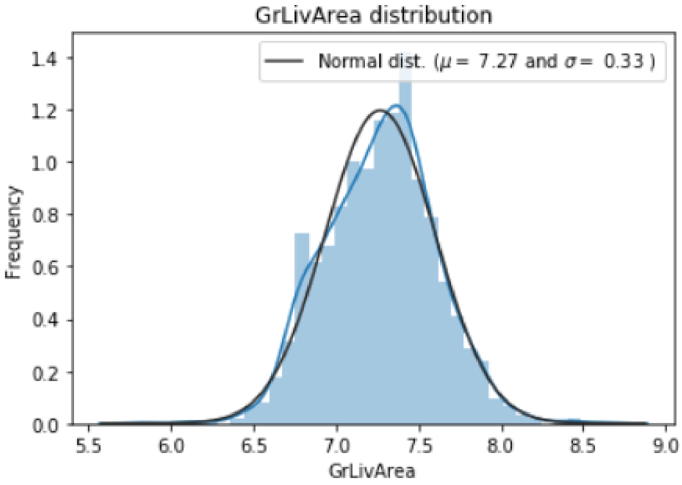
- Numerical variables
- Qualitative variables
- Ordinal variables

- Variable adjusting
- Avoid overfitting
- Add common sense

House Flippers

Validating data and transforming it was the majority part of the work






```python
kitqual = []
for x in train1["KitchenQual"]:
    if x == 'Ex':
        kitqual.append(5)
    elif x == 'Gd':
        kitqual.append(4)
    elif x == 'TA':
        kitqual.append(3)
    elif x == 'Fa':
        kitqual.append(2)
    else:
        kitqual.append(1)
```

	kit_qual
0	4
1	3
2	4
3	4
4	4

Data Cleaning

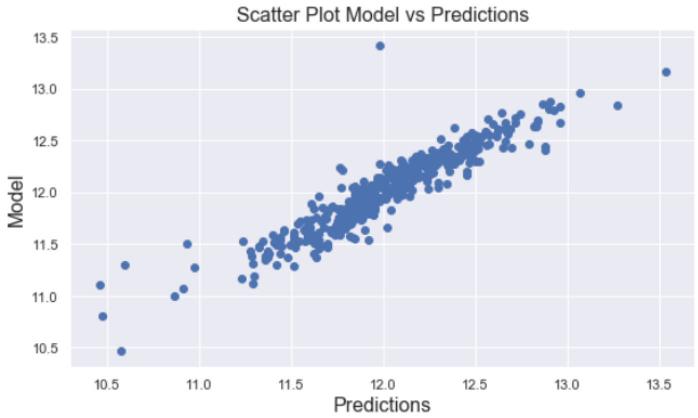
House Flippers

We found 15 significant variables that explained 89% of the prices in the dataset… And 7 variables that can be exploited to increase housing value

- Lot Area
- Building Age
- Basement Area
- Building Age
- Garage Size
- # of Bathrooms
- # of Bedrooms
- # of Floors
- Overall Quality
- Overall Condition
- Type of Property
- Zoning Classification
- Kitchen Quality
- Garage Condition
- Heating Quality

House Flippers

The residuals are normally distributed, and the scatter plot shows a clear straight-line, however we must remove outliers.

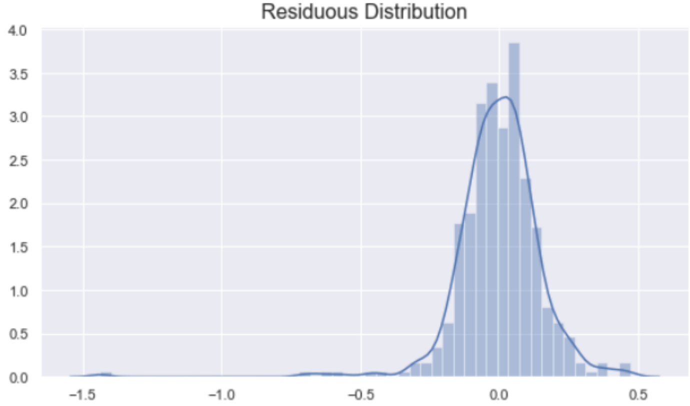


Mean Squared Error (MSE): 0.103
Root Mean Squared Error (RMSE): 0.153

Numerical

Categorical

	Features	Coefs	P[>]
0	Intercept	6.696	0.000
1	LN_LotArea	0.107	0.000
2	LN_GrLivArea	0.491	0.000
3	LN_YearBuilt_Adj	-0.103	0.000
4	LN_Total_Bath	0.103	0.000
5	LN_total floor	-0.152	0.000
6	BedroomAbvGr	-0.024	0.001
7	OverallQual	0.078	0.000
8	OverallCond	0.052	0.000
9	2fmCon	-0.053	0.098
10	Duplex	-0.102	0.000
11	Twnhs	0.013	0.682
12	TwnhsE	0.005	0.816
13	FV	0.143	0.019
14	RH	0.187	0.010
15	RL	0.124	0.028
16	RM	0.062	0.276
17	GarageCars	0.046	0.000
18	kit_qual	0.041	0.000
19	heat_qual	0.011	0.058
20	g_cond	0.012	0.115



House Flippers

Source: train1 dataset Ames, Iowa, Kaggle

Model Details

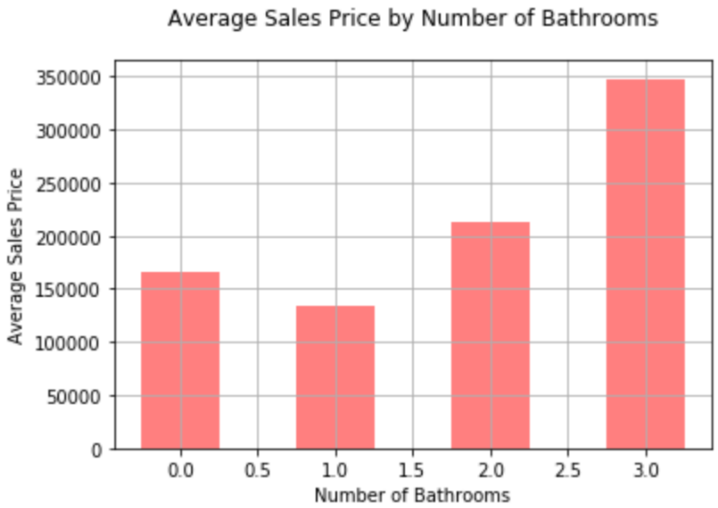
To wrap up: improving this model is possible using more sophisticated techniques and obtaining geographic data

- Designing a model takes time: understanding, organizing, cleaning the data
- Improvement opportunities for this model are related to additional information such as living nearby the ocean, ocean view, economy shocks, new big projects (Amazon), create additional dummy variables, so forth
- Location is very important for the price, however with this data set we cannot validate this assumption
- Removing Outliers will reduce the error of the model
- Machine Learning could be a great option to optimize and improve this model

House Flippers

Appendix

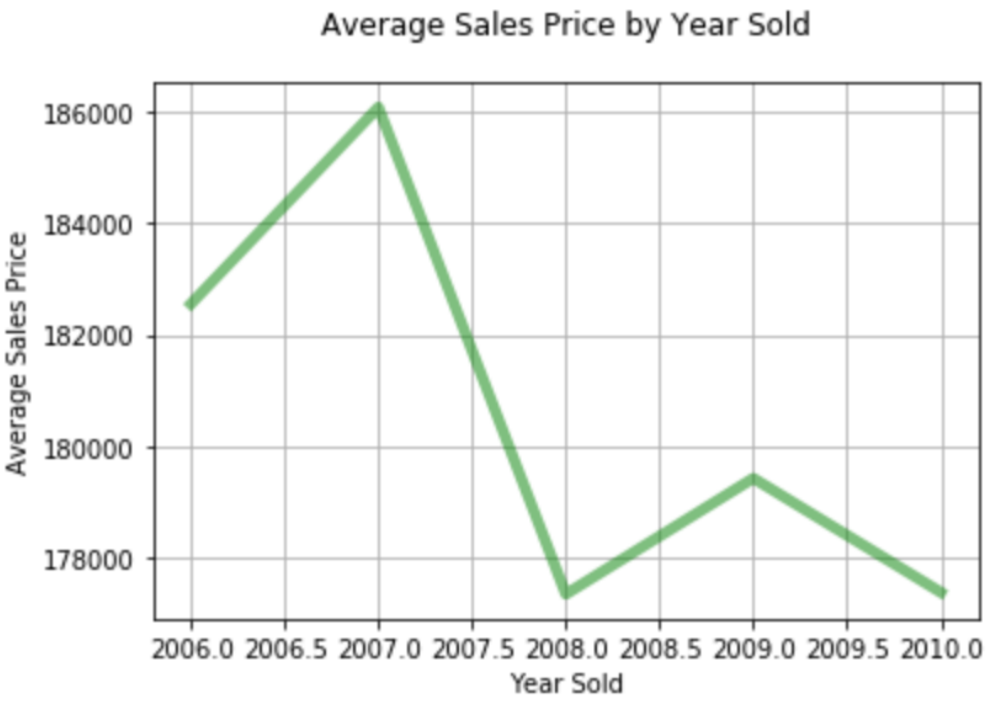
Trends in Sales Price by Number of Bathrooms



FullBath	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
BsmtFullBath	2	2	1	0	2	1	2	1	2
BsmtHalfBath	0	0	0	2	0	0	0	1	0
HalfBath	1	2	1	2	0	0	2	0	1
BedroomAbvGr	0	2	1	2	0	1	0	0	0

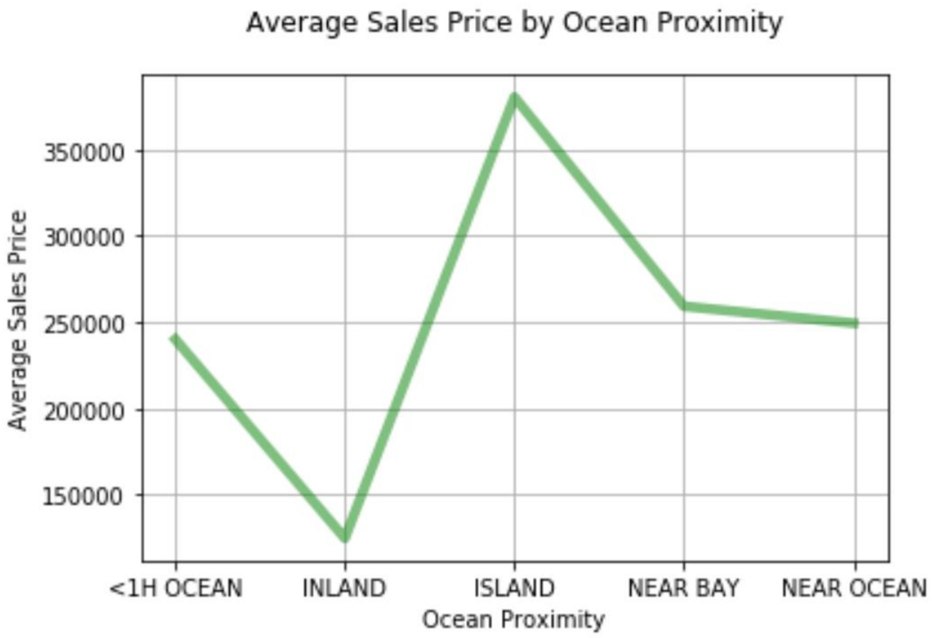
House Flippers

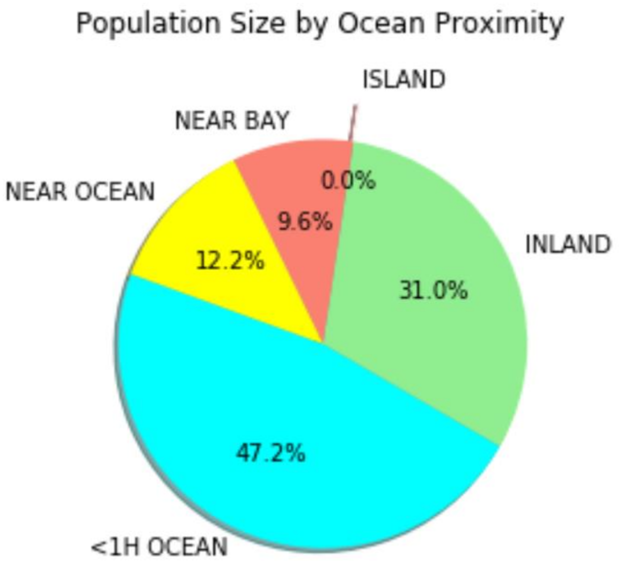
Annual Trends in Sales Price



House Flippers

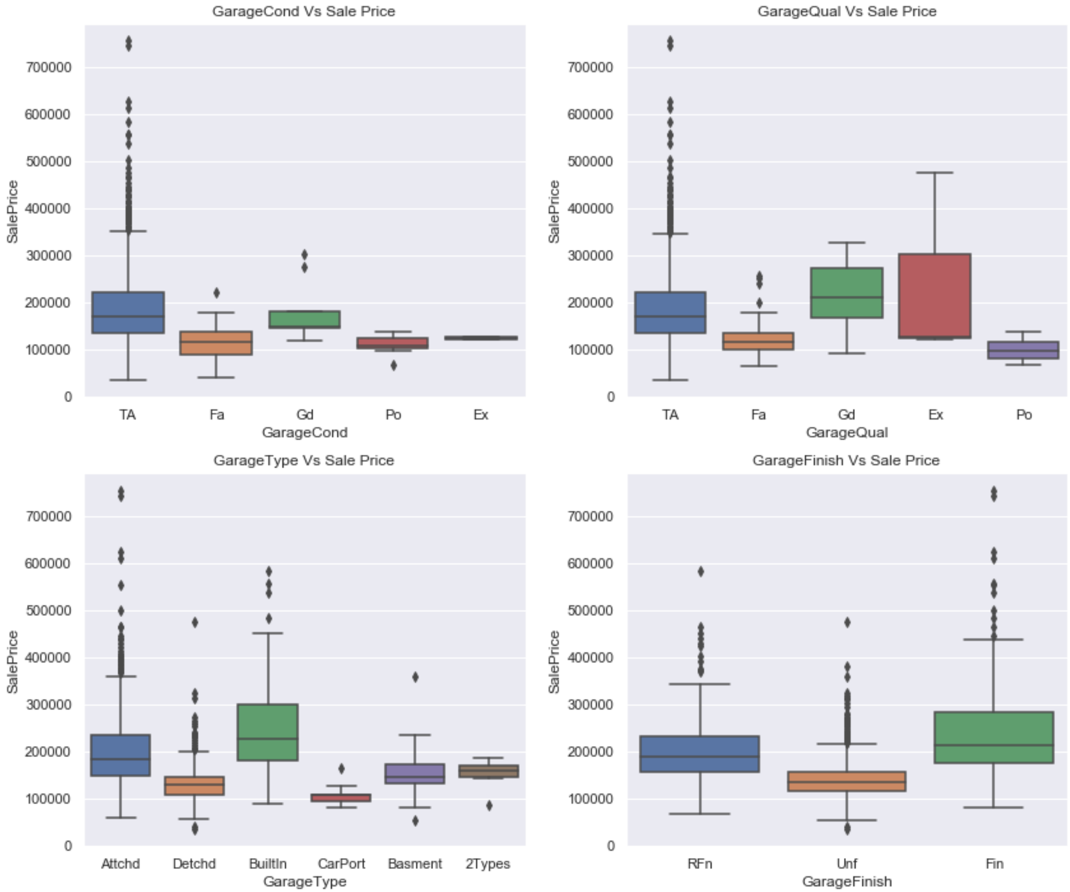
Trends in Sales Price and Population Size by Ocean Proximity





House Flippers

Trends in Sales Price and Population Size by Ocean Proximity

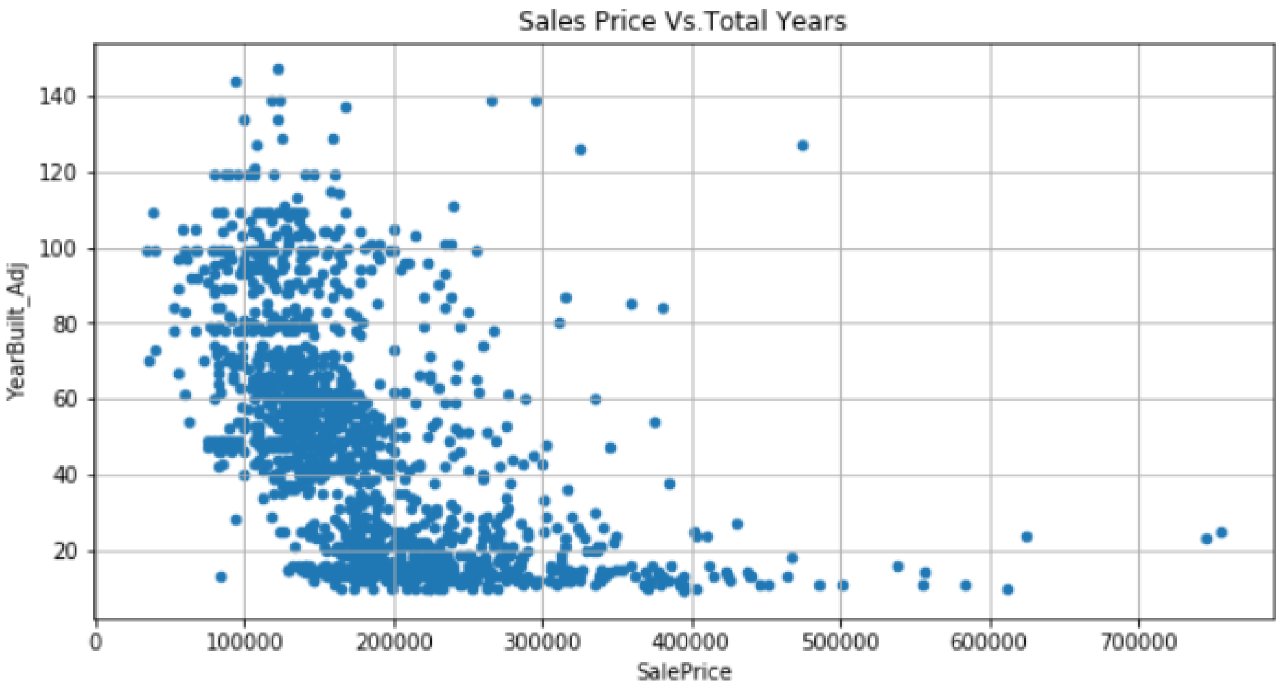


House Flippers

Removing Outliers is a MUST

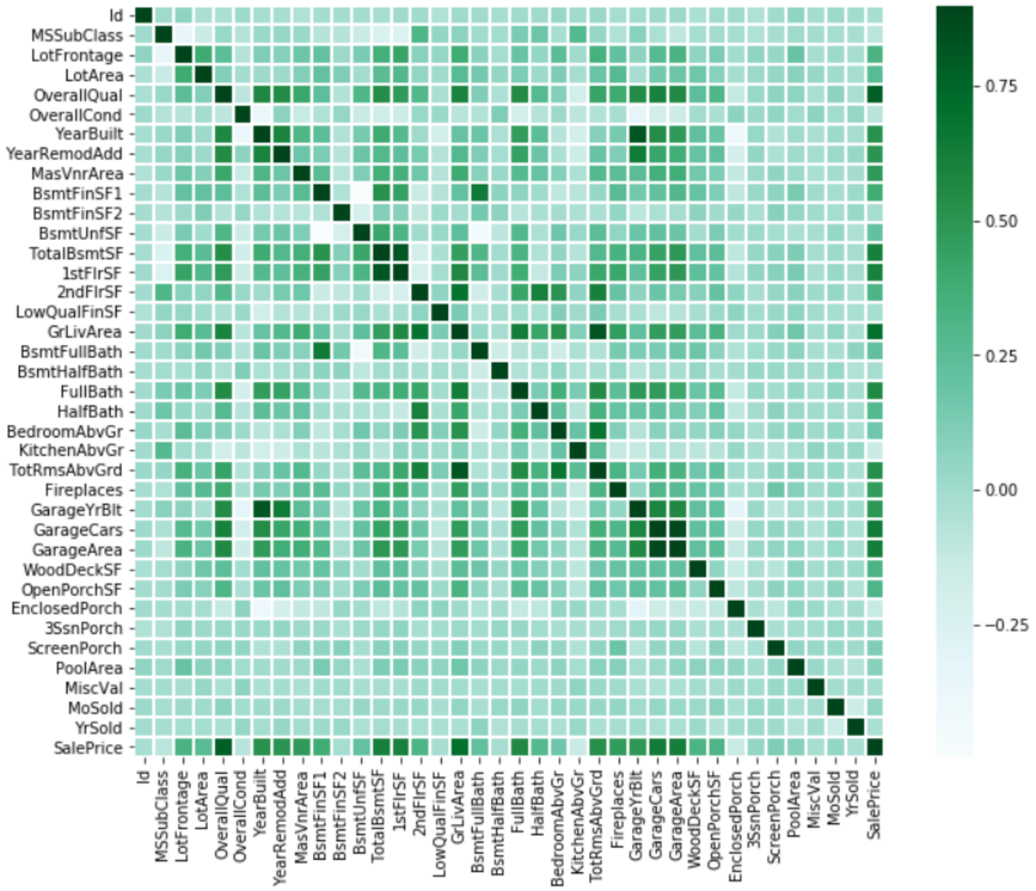
House Flippers

Removing Outliers is a MUST



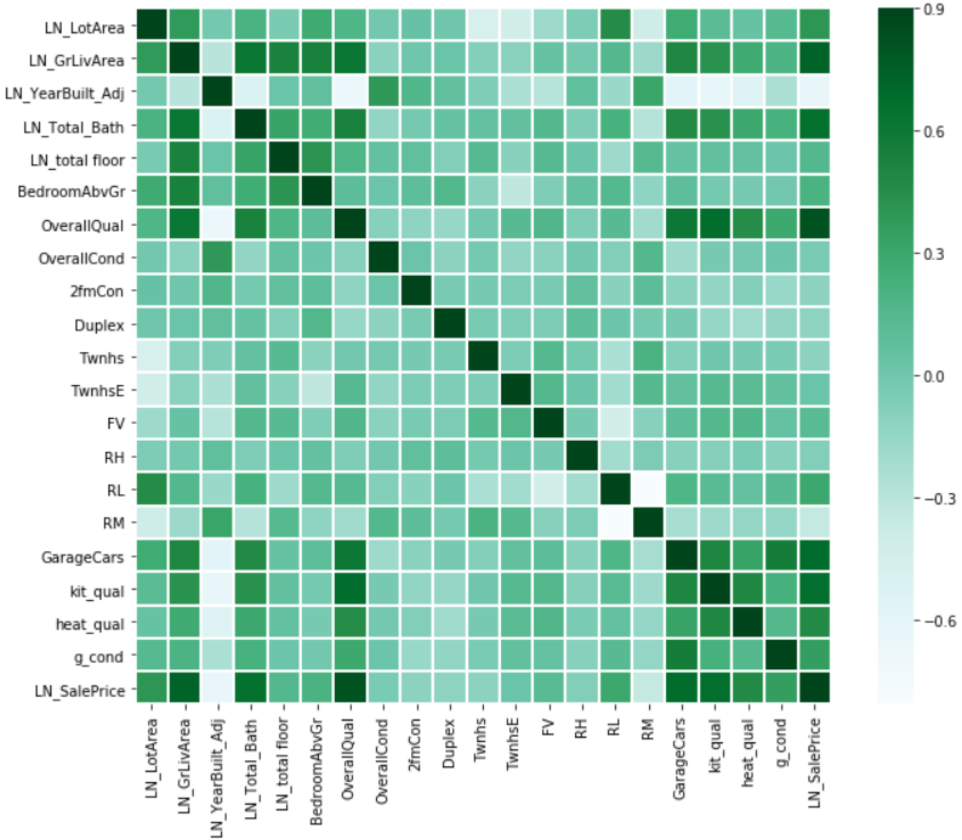
House Flippers

Correlation Matrix



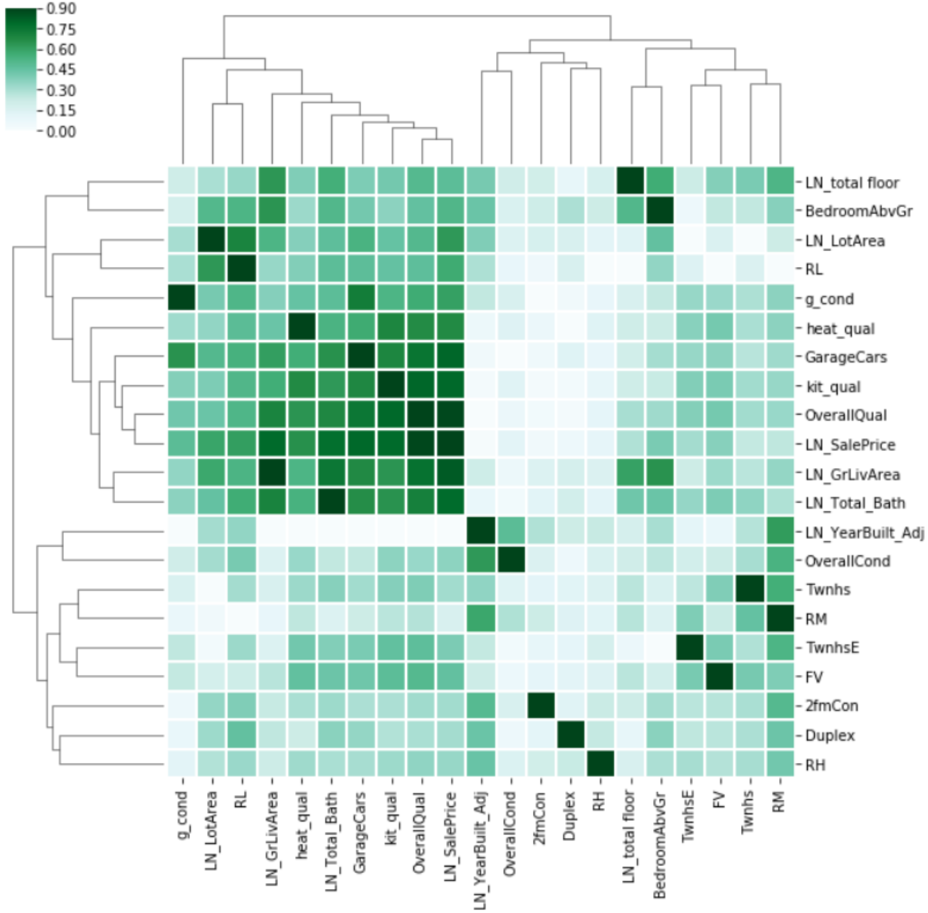
House Flippers

Correlation Matrix

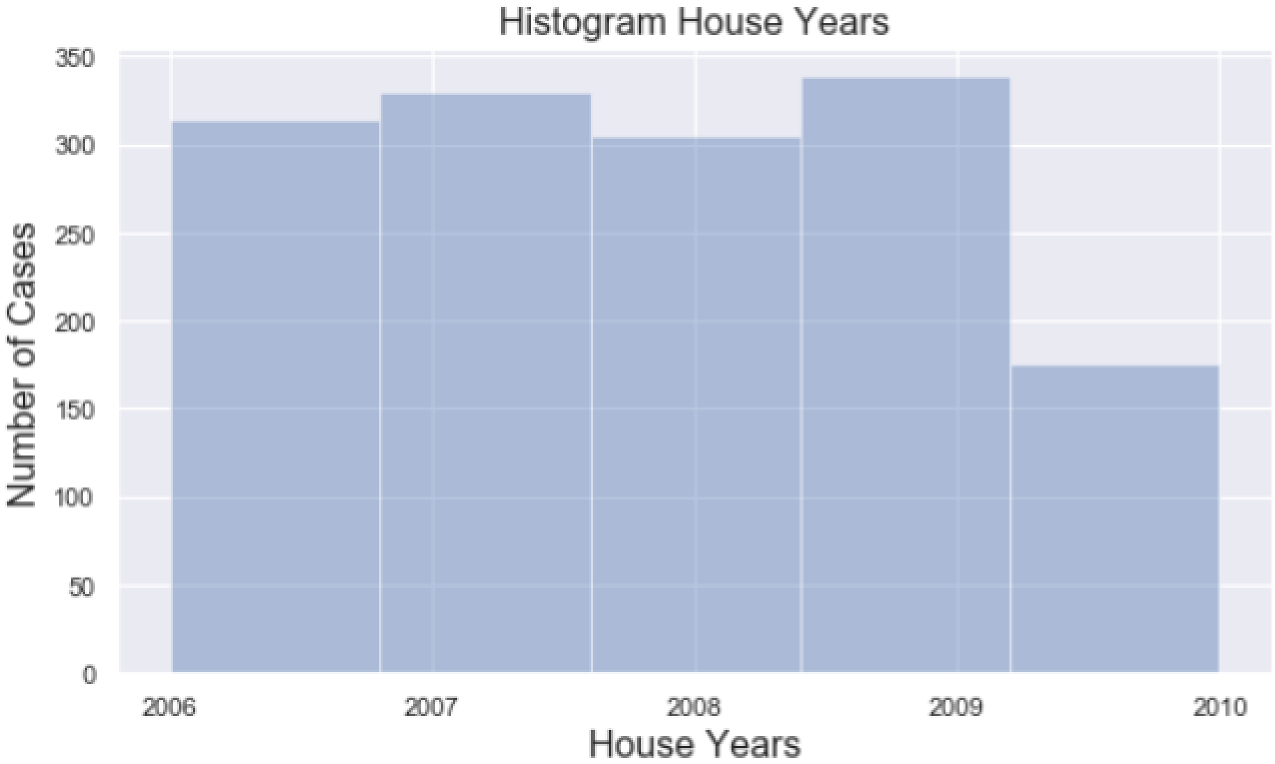


House Flippers

Correlation Matrix



House Flippers



House Flippers