Smart City Redevelopment

Eric Laverdiere
7/10/20

Problem Statement

I was given a data set for the housing values in Ames Iowa and tasked with figuring out a way to predict the potential sale prices of houses in the city.
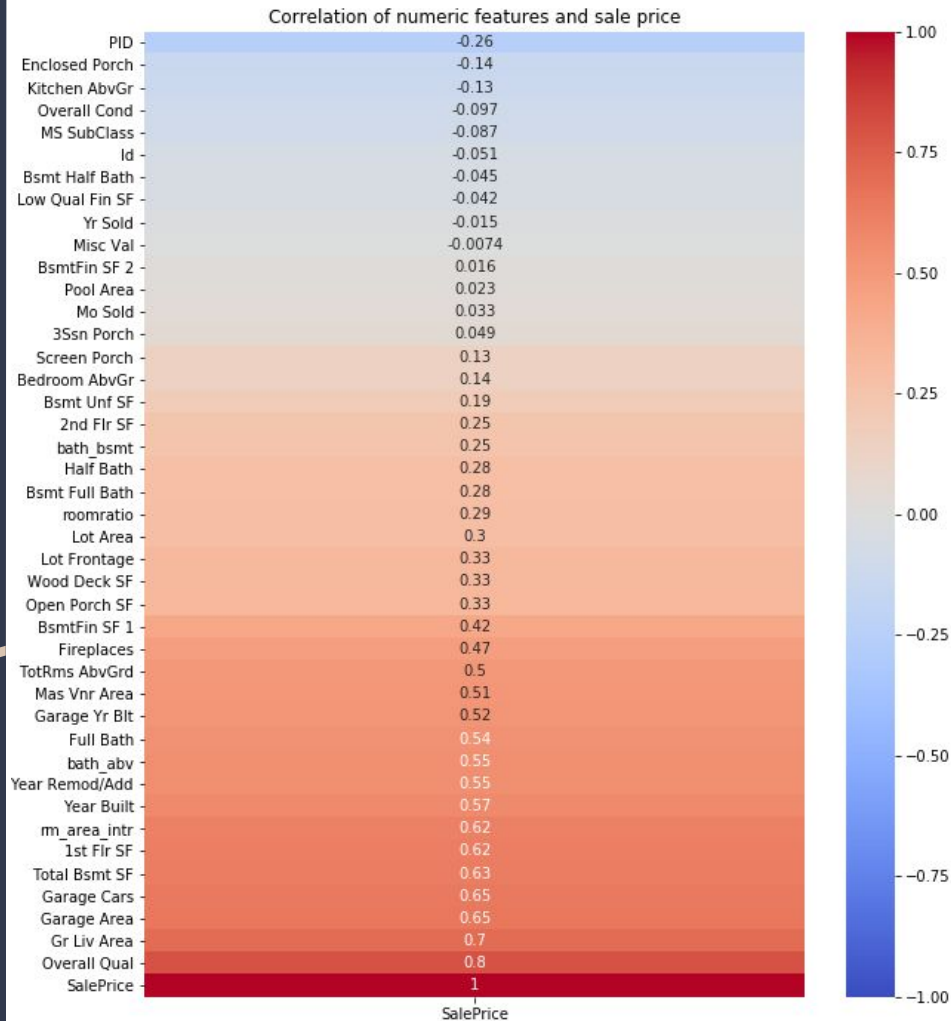
The data will be presented to the Ames city government to try and help them make redevelopment decision for the city while trying to avoid the issues of displacement or gentrification.

What can we figure out based on sale prices and how does that help with redevlopment?

Examining the Data

Using a heatmap to see the correlations of the numerical values with the sale price.

This gives a place to start when looking at important features

Correlation of numeric features and sale price

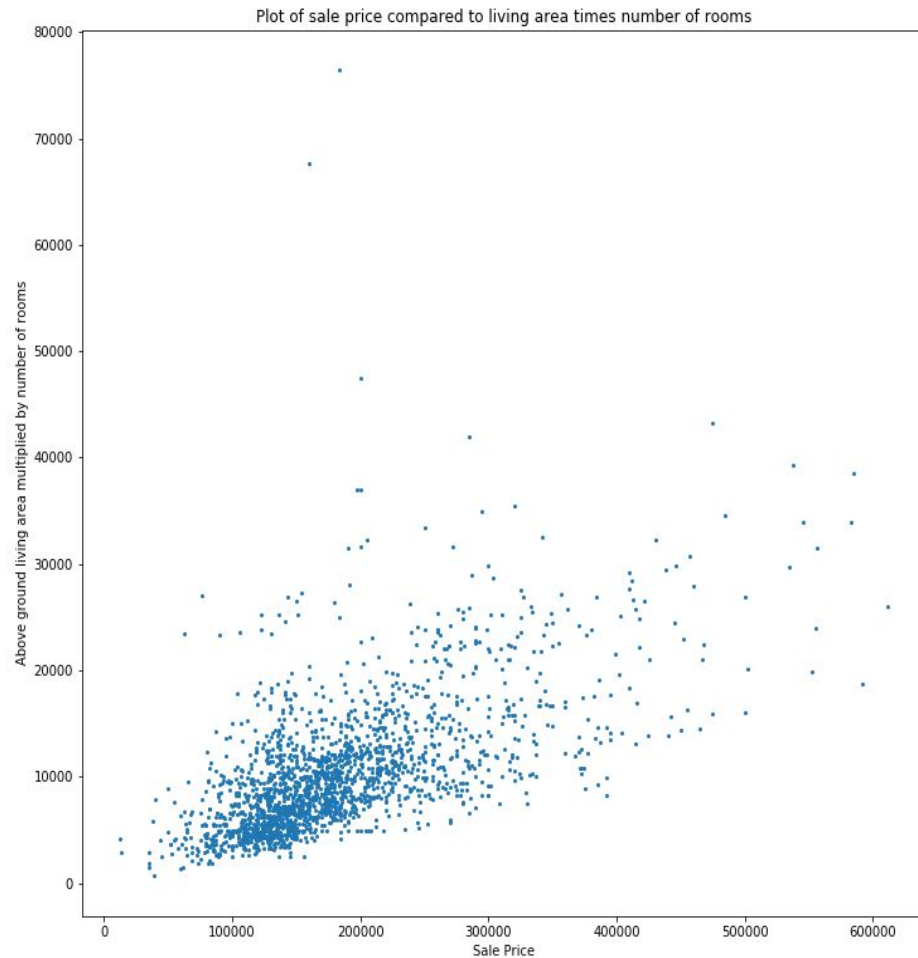|	SalePrice
PID	-0.26
Enclosed Porch	-0.14
Kitchen AbvGr	-0.13
Overall Cond	-0.097
MS SubClass	-0.087
Id	-0.051
Bsmt Half Bath	-0.045
Low Qual Fin SF	-0.042
Yr Sold	-0.015
Misc Val	-0.0074
BsmtFin SF 2	0.016
Pool Area	0.023
Mo Sold	0.033
3Ssn Porch	0.049
Screen Porch	0.13
Bedroom AbvGr	0.14
Bsmt Unf SF	0.19
2nd Flr SF	0.25
bath_bsmt	0.25
Half Bath	0.28
Bsmt Full Bath	0.28
roomratio	0.29
Lot Area	0.3
Lot Frontage	0.33
Wood Deck SF	0.33
Open Porch SF	0.33
BsmtFin SF 1	0.42
Fireplaces	0.47
TotRms AbvGrd	0.5
Mas Vnr Area	0.51
Garage Yr Blt	0.52
Full Bath	0.54
bath_abv	0.55
Year Remod/Add	0.55
Year Built	0.57
rm_area_intr	0.62
1st Flr SF	0.62
Total Bsmt SF	0.63
Garage Cars	0.65
Garage Area	0.65
Gr Liv Area	0.7
Overall Qual	0.8
SalePrice	1

Main features

Some of the main features to take into account:

- Overall quality
- Above ground square footage
- Garage square footage
- Basement square footage
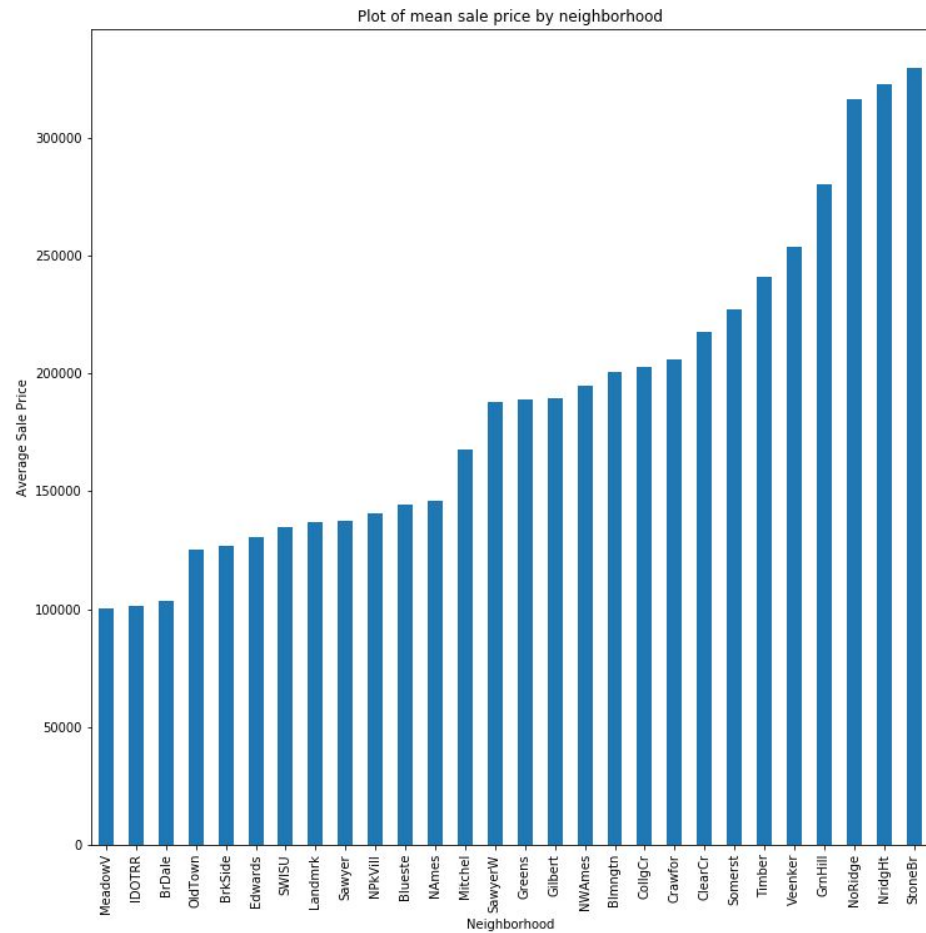- Year Remodeled
- Basement Quality

Additional add features:

- Total number of bathrooms above ground
- Total number of basement bathrooms
- Ratio of bathrooms to rooms above ground

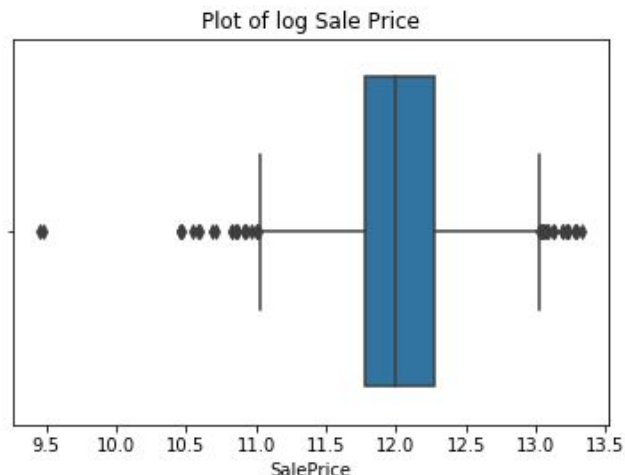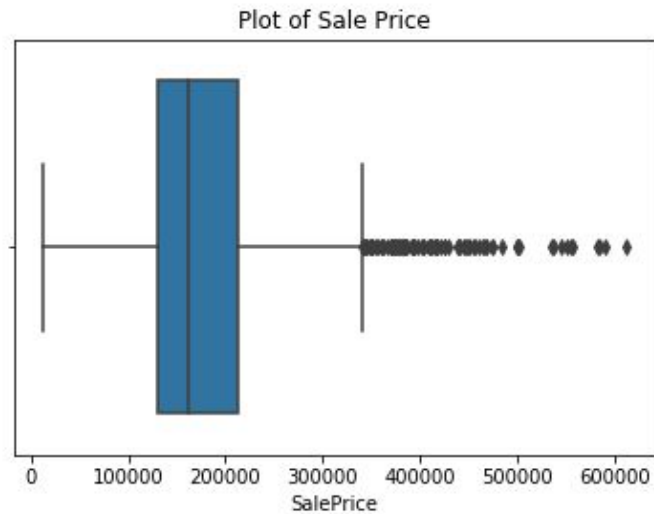# Interaction of above ground rooms and square footage

Plot of sale price compared to living area times number of rooms

Neighborhoods

Plot of mean sale price by neighborhood

Sale Price

Taking the log to normalize and reduce outliers

Plot of Sale Price

Plot of log Sale Price

Regression model

The final r2 score was about .909 so around 91% of the variability in the sale price was explained by the chosen features.
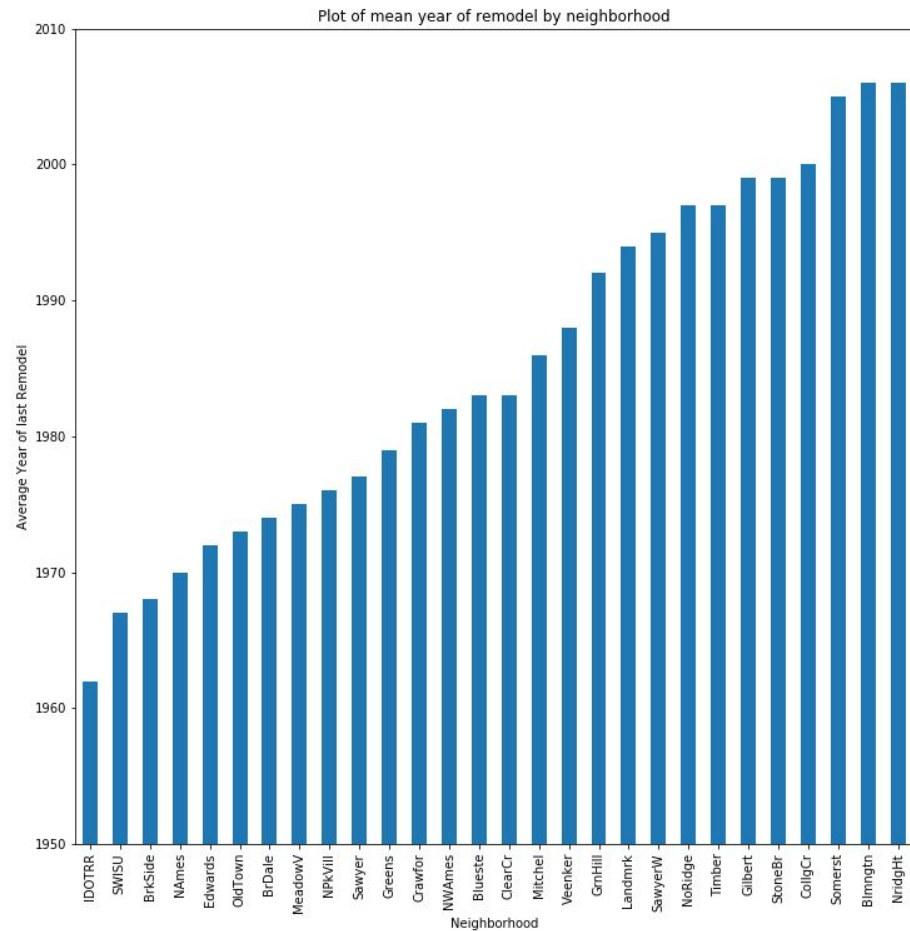

Histogram of the residuals
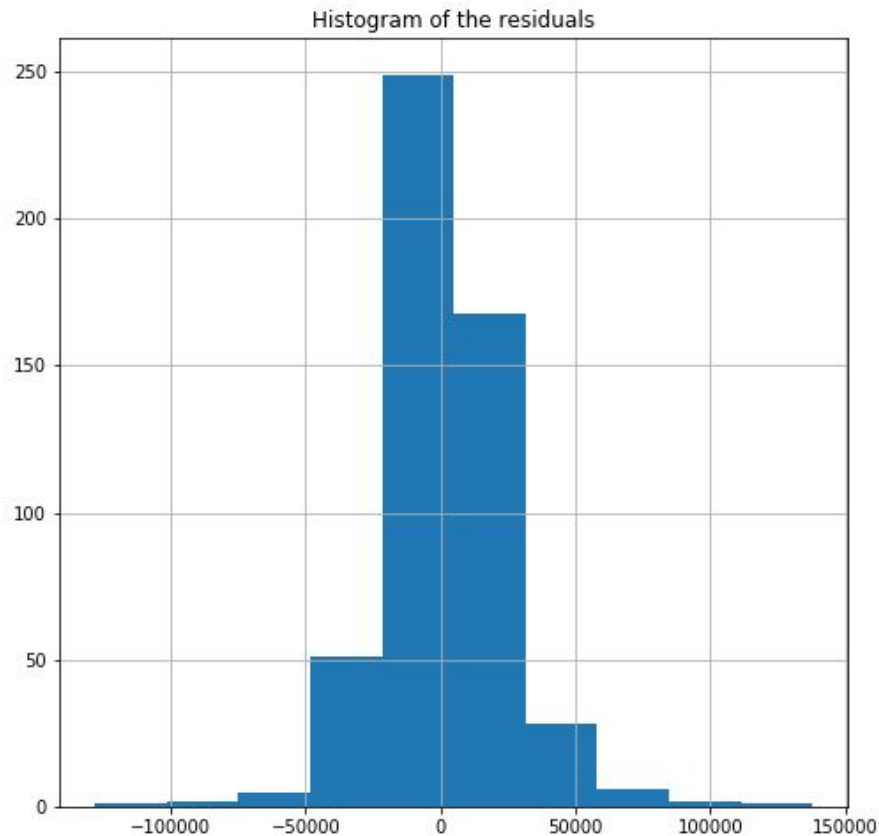
What does this mean for the city?

Model can be used to understand the sale prices in certain neighborhoods and city areas to target investment.

Allow for smart redevelopment

Try to avoid problems sometimes associated with Gentrification(Roos 2020)

Other Factors

Plot of mean year of remodel by neighborhood

Conclusions

Understand what are strong factors in housing prices

Model can help know what places to target for investment

Make sure investment is balanced and does not price out the local residents

What Next?

Further Exploration

- How lot factors affect prices
- Improve the model with regularization techniques
- Look into outside factors that might affect sale price
- Demographics
- Crime
- Quality of Area

References

Roos, Dave (2020)
https://money.howstuffworks.com/gentrification.htm

Grabinsky,Jonathan   Butler,Stuart M. (2015 February 12)
https://www.brookings.edu/blog/social-mobility-memos/2015/02/11/the-anti-poverty-case-for-smart-gentrification-part-2/

Homeguru (2020)
https://www.homeguru.com.au/house-prices