

Predicting home prices with regression:

Client summary

A first capstone project for
Springboard's data science bootcamp

Mike Pierovich
2/26/2020

Problem

What's the home
going to sell for?

And why?



Data

kaggle™

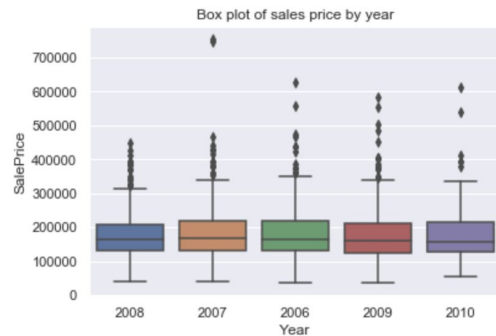
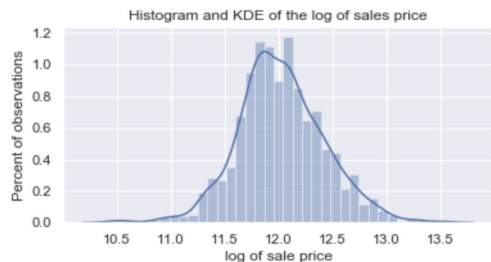
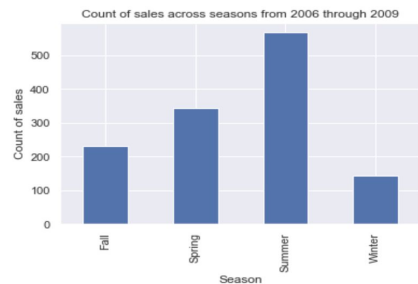
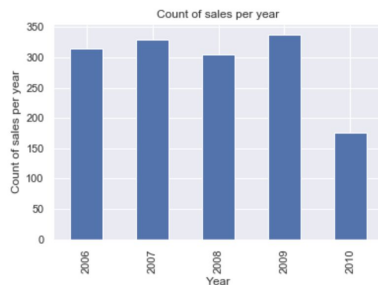
“Ames Housing Data”

- Sales price of 700 homes
- About 80 features describing each home
- From one US city
- 30 miles north of Des Moines



Sales price

- \$160,000 median
- Outliers
- 300 sales/year
- 25 sales/month
- Seasonal
- Prices flat



Home features

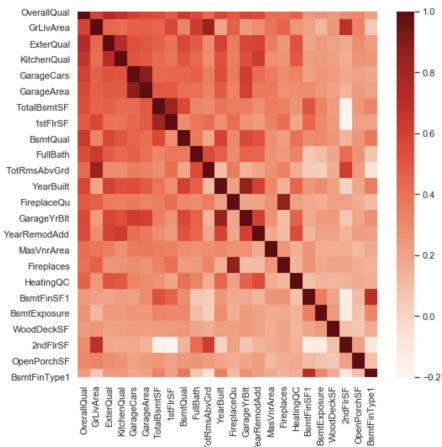
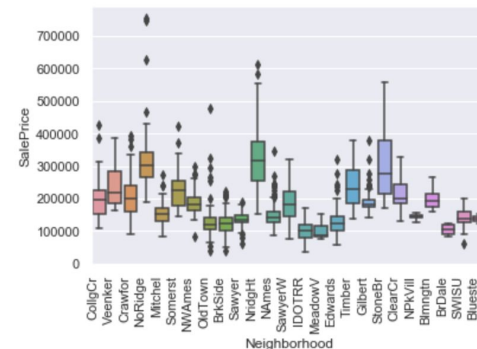
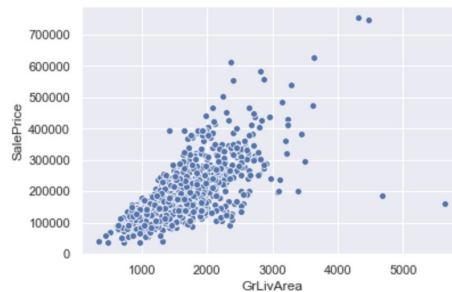
Of 54 non-categorical

- 24 with Pearson's $r > 30\%$
- 10 with cross correlation $< 50\%$

Of 25 categorical

- 2 with Pearson's $r > 30\%$

Selected 12



Models

Build, tune and compare

Linear - baseline		
Linear - baseline	v.	Linear - normalized
Linear - normalized	v.	Lasso
Linear - normalized	v.	Ridge
		Random forest - baseline - all
		Random forest - baseline - all
		Random forest - tuned - all
		Random forest - tuned - high
		Random forest - tuned - all
		Random forest - tuned - high+med
		Random forest - tuned - all
		Random forest - tuned - high+med+low
Linear - normalized	v.	Random forest - tuned - high+med+low

Best performance

Linear - normalized

- RMSE: .15
- MAE: .11
- R2: .86

Recommendations

Launch

- Minimum viable product around
- Predicted price
- Prediction interval
- For a specific property in Ames, Iowa and
- For all properties in Ames, Iowa.

Improve

- Data
 - Location
 - Time
 - Features
- Model
 - Regression
 - Other models