Story County Assessors –Tax Equalization

Ryan McDonald – Data Scientist

General Assembly

https://git.generalassemb.ly/spyder6146/project_2

Problem Statement

Story County requires Property Tax Equalization to occur every 2 years. Ames, IA is the most populated community in the county.

We have been tasked to:

1. Develop opensource tools for County Assessors to utilize for home value assessments for Ames, IA

2. Provide insights on most value-added characteristics to aid potential homeowners when placing offers on Residential homes

Story County
IOWA

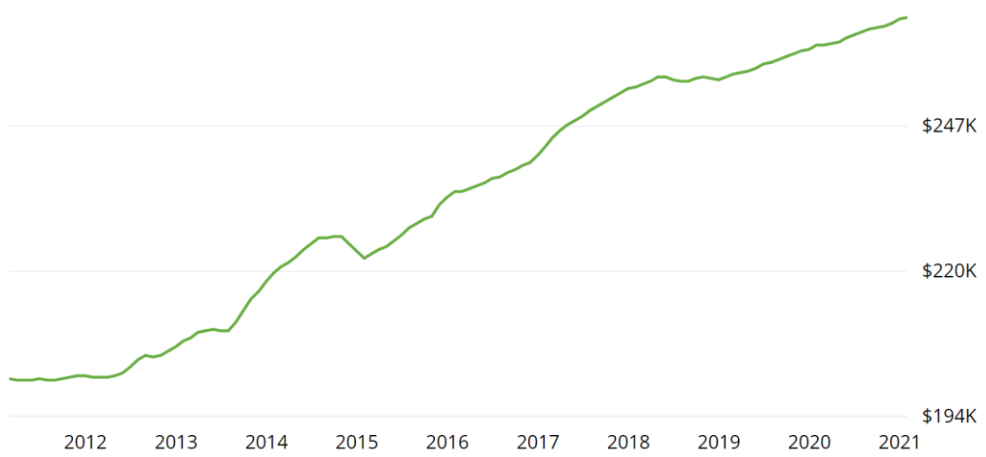
https://www.istockphoto.com/illustrations/ames-ia

Background

Property Tax Assessment Rates have not increased in nearly a decade.

- AMES Residential property tax rate stands at $31.49 per $1000 of assessed value.
- Homes in Ames, IA are taxed on 55.07% of market value.
- Residential property taxes account for 42.99% of all property taxes paid to the County.

Ames, IA Mean Home Price 2011-2021
Ave 3.2% increase / year



https://www. zillow.com/ames-ia/home-values/

Mean Home Value Change 2006-2010			
06-07	07-08	08-09	09-10
3.89%	-0.74%	-1.13%	-4.52%

https://www.storycountyiowa.gov/465/Property-Tax

Background (cont.)

Results are only developed based on data provided in the report.

Data received from County Assessor was collected between 2006 and 2010

- 1,995 home sales were evaluate, each with 81 attributes reviewed

- Not all data was utilized for the analysis. Some categories were merged, new ones were developed, and others dropped entirely

Assumptions:

- Data provided was during the Housing Crisis and property value annual changes between 2006-2010 were sporadic.
    - This may skew models when applied to newer home sales

- Typical historic growth rate for Ames Iowa is ~ + 3.2 / yr.

- Any data manipulation was backed by analysis and deemed appropriate.

Data Analysis - Modeling

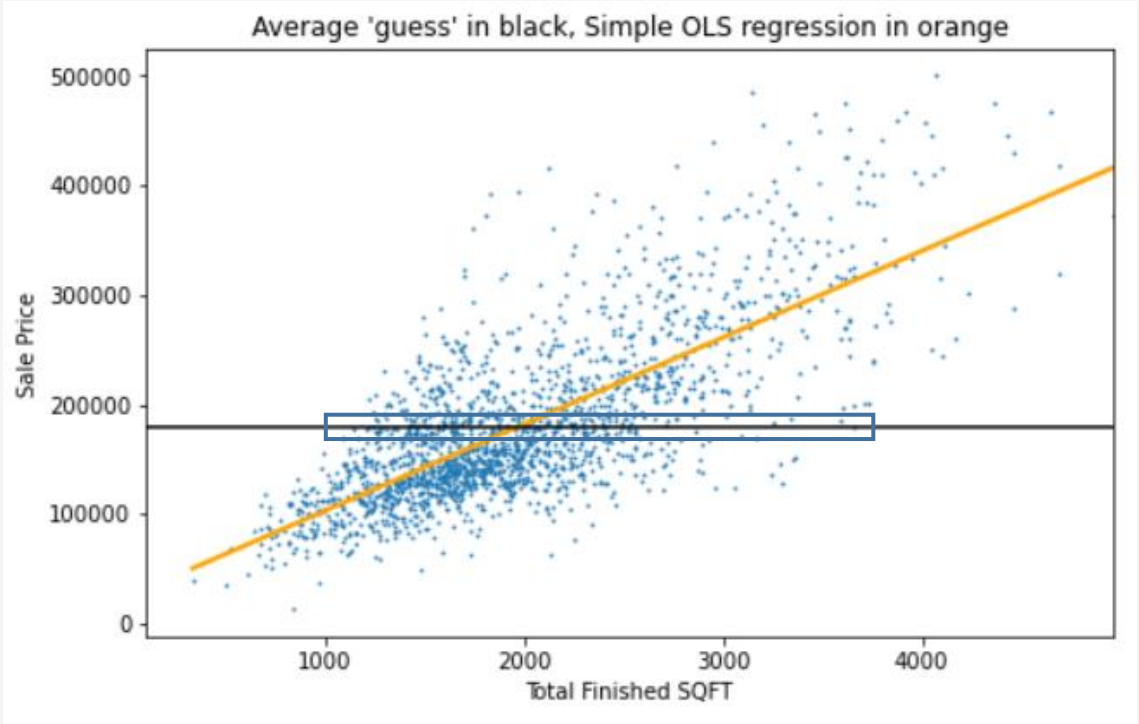
Data was extensively cleaned and interpreted with the problem statement in mind.

Baseline Model was first! Can we guess prices?

Feature	Value
Mean Sale Price	$180, 050
Mean Finished SqFt	1,972
Median Bedrooms	3
Median Baths	2

'Guessing' would only be within 3% of the mean **6.8%** of the time!

Data Analysis – Basic Model



Even a basic linear model will provide much better predictions then a 'guess'!

Modeling on Total Finished SQFT will account for **~56%** of the variability in Sales Price!

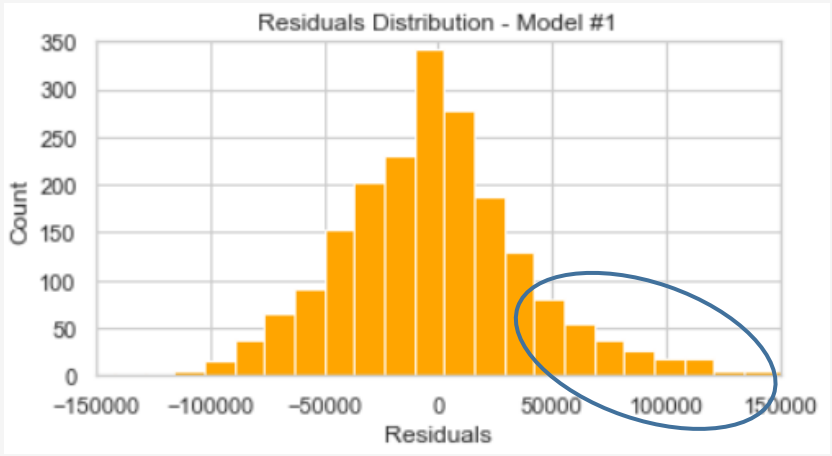
Feature	Correlation to Sale Price
overall_qual	0.807
tot_fin_sq	0.750
gr_liv_area	0.710
garage_area	0.651
garage_cars	0.651
total_bsmt_sf	0.647
1st_flr_sf	0.633
total_baths	0.626
year_built	0.578
year_remod/add	0.554

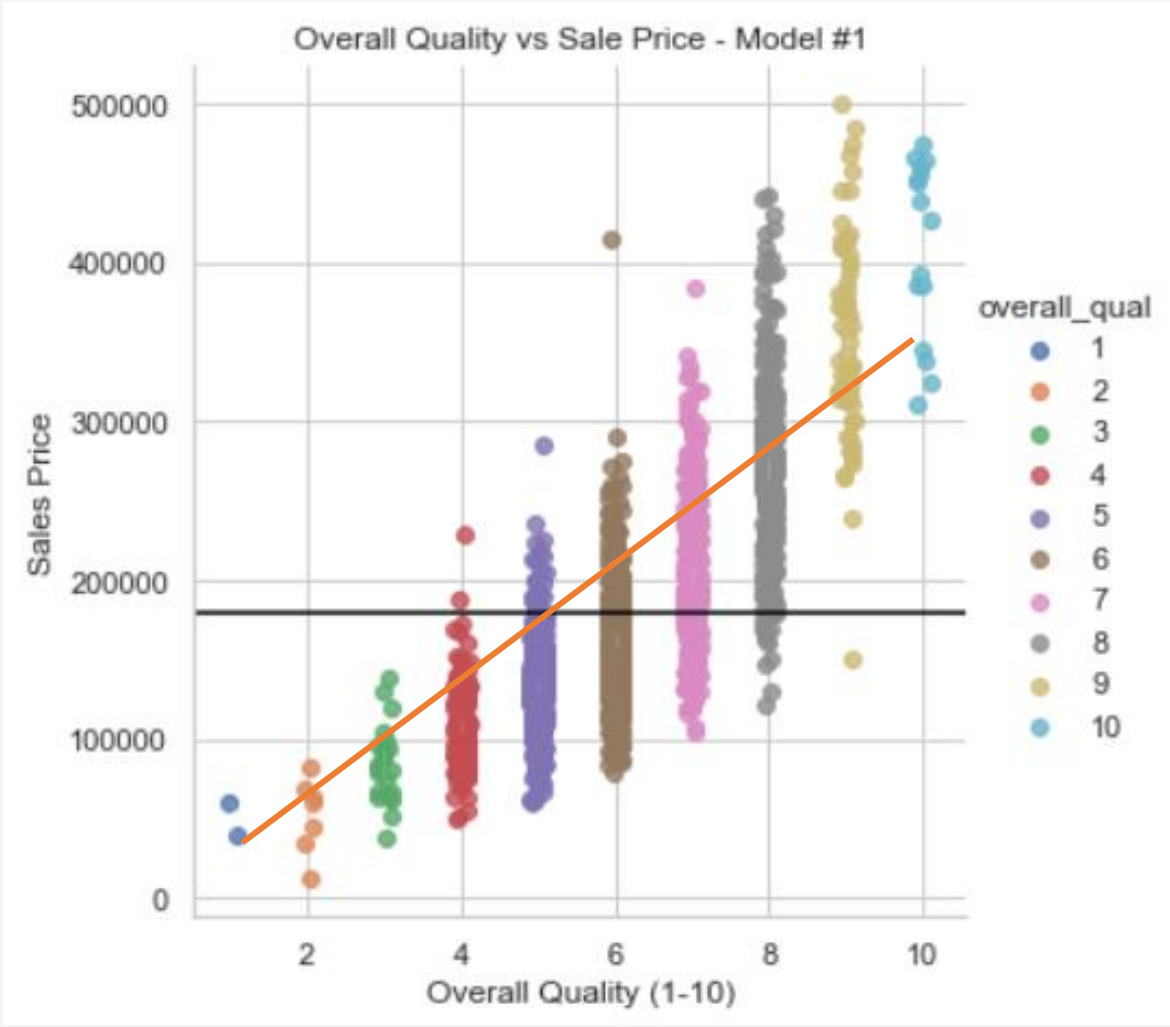
Data Analysis – Basic Model (cont.)

So, how about utilizing the highest correlated feature for predictions?

Modeling on Overall Quality will accurately represent variability in sale price **~66%** of the time

- All things held equal, each unit increase in 'Overall Quality' represents an increase of **$43,140** to the Sale Price!



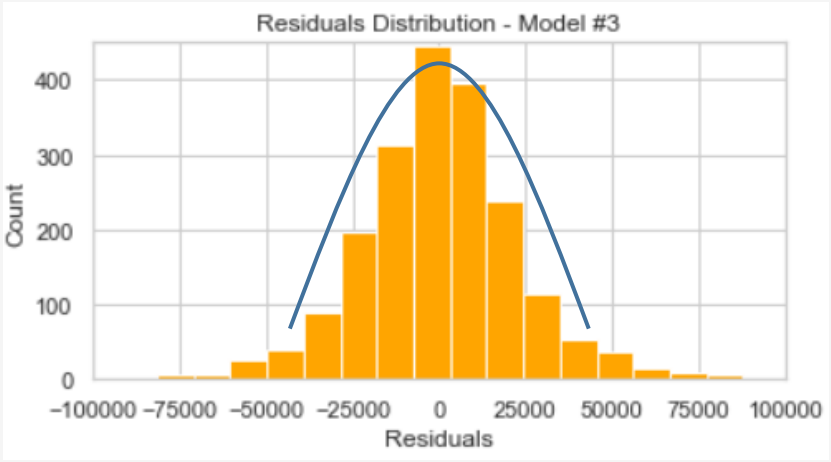


Data Analysis – Multiple Linear Regression

To even better-predict sale price, we've now included the top 40 highest-correlated features!

Modeling these feautres will accurately represent variability in sale price **~87%** of the time

- RSME (root squared mean error) = **$23,794**



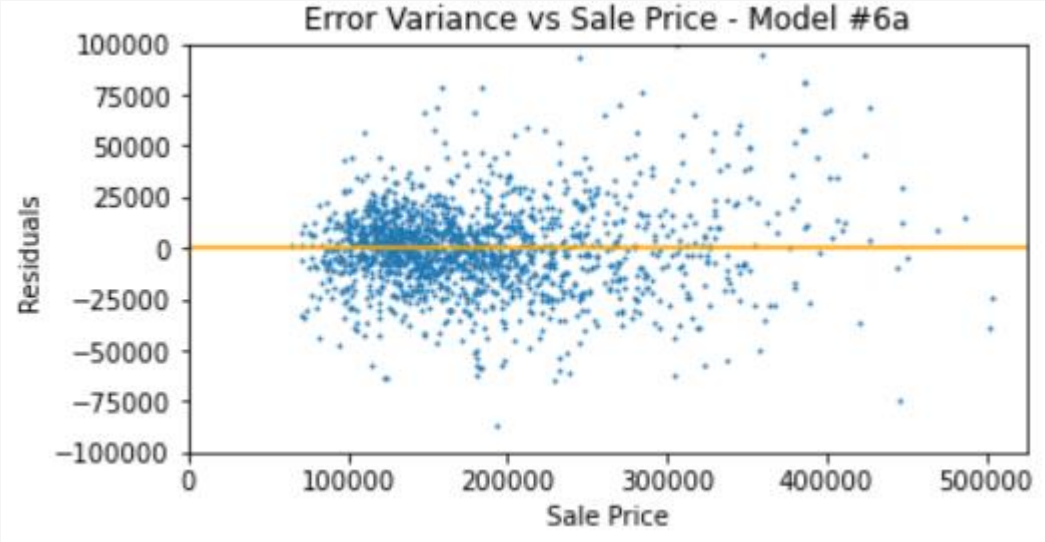
With so many features on varying scales, there is a risk of overfitting!

- **Let introduce some bias to try to improve on the model**

Data Analysis – Final Model Balancing Bias and Variance

How the final model differs from previous:

- Fewer baseline features
- Introduction of engineered features!
  - Polynomial features (creating interaction terms)
  - Dummy Features from object-type variables
  - New features derived from dataset
- Scaling the data in our features
- Utilizing LASSO regularization (different penalty component to Ridge)



Data Analysis – Final Model Stats

Model #6a produced the best results for predicting sale price on unseen test data!

- Ridge regularization was also modeled with only slightly less accurate results.
- Additional accuracies could be gained with more complex regression modeling
- 'Overall Quality' and 'Square Footage' – based features resulted in the highest LASSO Regression Coefficients!

Final Model #6a Stats	
# of Features	18 (5 engineered)
w/ Poly Features	189
Model / fit	Linear Regression
Scalar	StandardScalar
Adj r^2 (train/test)	0.93 / 0.87
RMSE	19,460
LASSO Alpha	66.67
LASSO r^2 (tr/tst)	0.92 / 0.89

Conclusion

Although accuracies may vary based on new test data, Model #6a will be submitted to the Story County Assessors Office for review. This model was able to represent variance in sale price nearly 90% of the time. Further modification to the model can be made in order to utilize in different cities as well.

Homeowners looking to buy or sell a home in Ames, IA should focus efforts on overall quality of features in the home. This will net the greatest increase in value. Total finished square footage and garage size would be the next characteristics to work on. Location will be important as well. The only neighborhoods with mean sales price above $300k are North Ridge Heights and North Ridge.