

Sale Price Determinants of Houses in Ames, IA

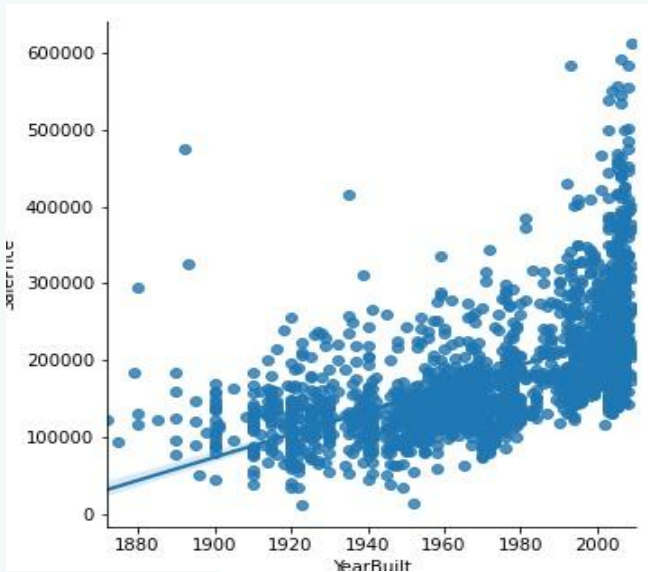
By: Samantha Chu



Problem Statement:

There seems to be an upward trend of the price of houses in Ames, Iowa over the past couple of decades. Young couples and prospective homeowners looking to keep a budget while purchasing a home, would like to be able to estimate how much a home Ames, Iowa will cost them.

- Develop linear regression statistical model to predict price of house
- Suggest/recommend house characteristics that will affect price so home buyers know what price to expect.

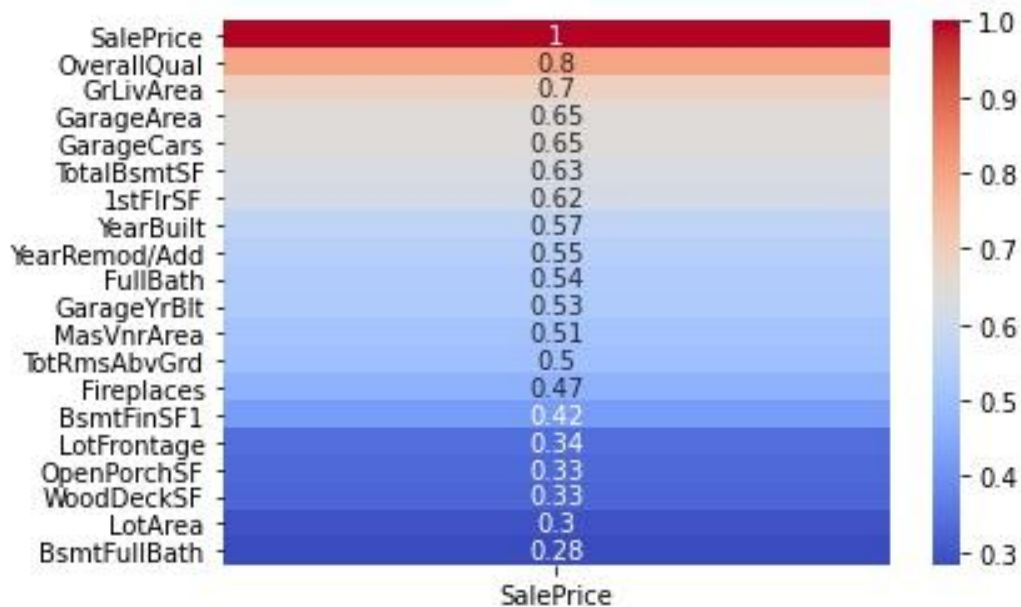


Correlation Coefficients with the Sale Price target variable

Step 1:

Correlation Coefficient with Sale Price:

- Most significant factors that influenced a house's Sale Price
 - Overall Quality
 - Exterior Quality
 - Above Ground Living Area (ft^2) = GrLivArea
 - GarageArea
 - GarageCars
 - Total Square Feet of Basement Area = TotalBsmtSF

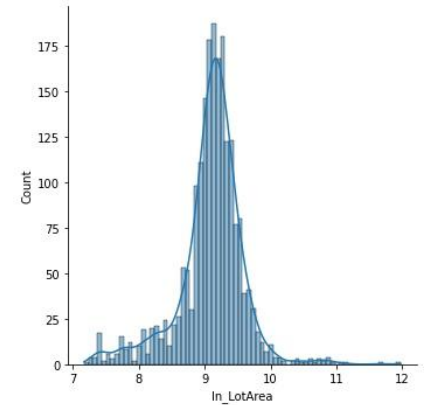
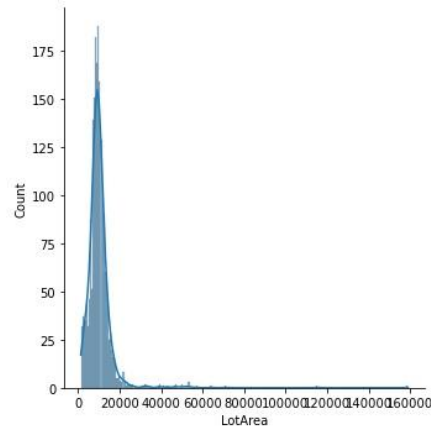
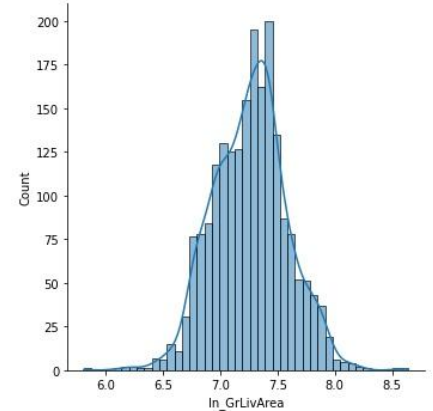
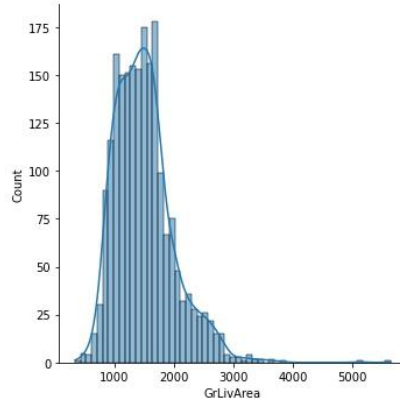


Transformations to match Linear Regression Assumptions

Step 2:

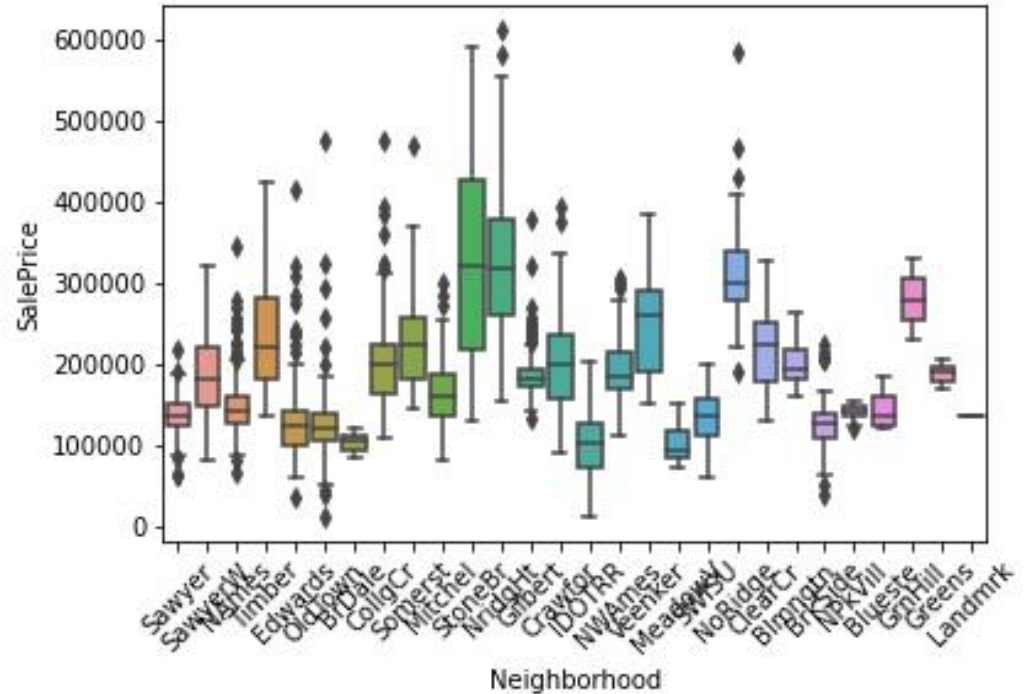
Transformations:

- Above Ground Living Area was right-skewed
- Lot Area was right-skewed as well
- R^2 increased significantly after log transforming



Step 3:

- Neighborhood variable was added
- This feature is of interest to prospective homeowners
- Where/which neighborhood a person lives in matters

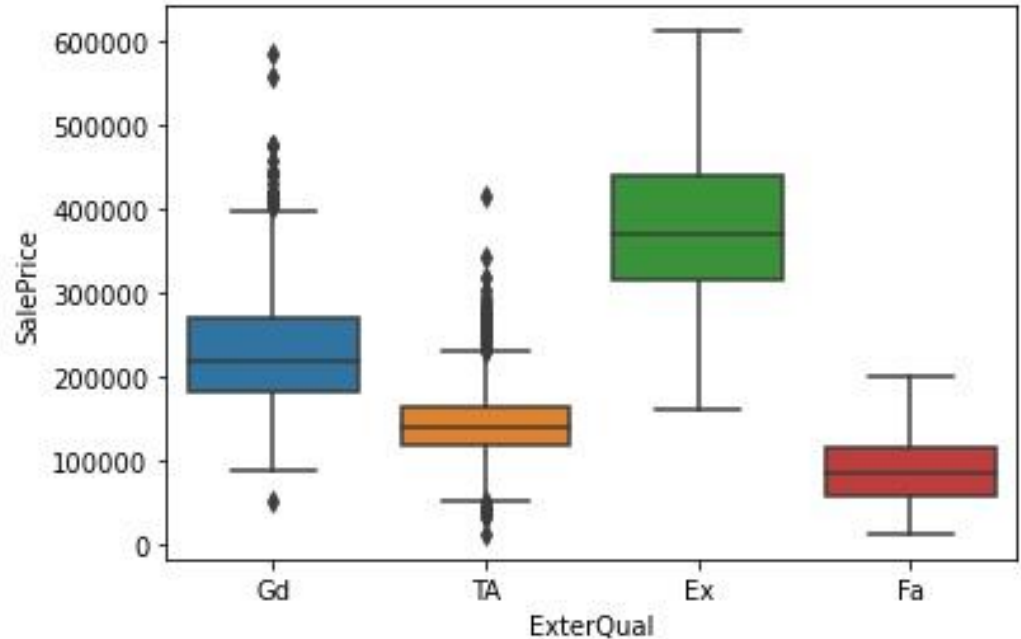


Adding Variables that contributed to variance of Sale Price

Step 3:

Features affecting Sale Price Variance

- Exterior Quality was another variable that I added to influence the variance in Sale Price
- A lot of other variables that I added that were similar to this, and were also ordinal (BsmtQual, HeatingQC)



Looked at Statistical Significance

Step 4:

- Looked at p-values to see significance of the variable

	coef	std err	t	P> t	[0.025	0.975]
const	-1.059e+06	8.06e+04	-13.138	0.000	-1.22e+06	-9.01e+05
In_LotArea	3.027e+04	2341.210	12.927	0.000	2.57e+04	3.49e+04
In_GrLivArea	5.98e+04	3565.636	16.770	0.000	5.28e+04	6.68e+04
GarageArea	22.7350	5.088	4.469	0.000	12.755	32.715
OverallQual	1.295e+04	1028.676	12.592	0.000	1.09e+04	1.5e+04
TotalBsmtSF	7.6179	3.666	2.078	0.038	0.427	14.809
BrDale	1.202e+04	1.14e+04	1.052	0.293	-1.04e+04	3.44e+04
BrkSide	-2529.4090	8983.291	-0.282	0.778	-2.02e+04	1.51e+04
ClearCr	-5092.2805	1.02e+04	-0.498	0.619	-2.52e+04	1.5e+04
CollgCr	-2.008e+04	7876.160	-2.549	0.011	-3.55e+04	-4625.664
Crawfor	4439.4831	8877.653	0.500	0.617	-1.3e+04	2.19e+04
Edwards	-2.212e+04	8380.669	-2.639	0.008	-3.86e+04	-5678.940
Gilbert	-2.151e+04	8379.794	-2.567	0.010	-3.79e+04	-5070.245
IDOTRR	-1.432e+04	9283.570	-1.543	0.123	-3.25e+04	3885.833
MeadowV	1.963e+04	1.06e+04	1.845	0.065	-1236.143	4.05e+04

Looked at Statistical Significance

Step 4:

- Looked at p-values to see significance of the variable

YearBuilt_squ	0.0911	0.019	4.733	0.000	0.053	0.129
YearRemod-Built	197.0255	54.789	3.596	0.000	89.554	304.497
ExterQual	1.725e+04	2210.460	7.804	0.000	1.29e+04	2.16e+04
BsmtQual	2855.9111	1451.494	1.968	0.049	8.734	5703.088
BsmtFinType1	-761.6644	582.895	-1.307	0.192	-1905.041	381.712
BsmtFinSF1	-16.1334	7.907	-2.040	0.041	-31.644	-0.623
BsmtFinType1*BsmtFinSF1	7.9883	1.445	5.527	0.000	5.153	10.823
1stFlrSF	-0.8855	4.048	-0.219	0.827	-8.826	7.055
HeatingQC	2012.2666	1022.668	1.968	0.049	6.253	4018.280
Hip	9647.0630	2107.702	4.577	0.000	5512.700	1.38e+04
Other_roof	-6218.7302	5537.476	-1.123	0.262	-1.71e+04	4643.305



Methodology



#1

**Linear
Regression**

Interpretability
was an important
factor

#2

**Chose and
Transformed
features
carefully**

Looked at p-values
and significance

#3


LASSO

Tried Lasso, but
did not improve
my model that
much with the
features that I
had. ~ Lose
interpretability

#4

Goal

Develop a
predictive model,
that still
maintained some
interpretability



Metrics

R² Score

85-86%

RMSE

Roughly, 29,000-30,000 range

Model Used

Linear Regression

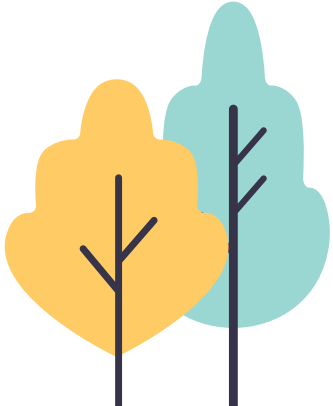




Recommendations

To the prospective homeowner in Ames, Iowa: A few recommendations/interpretations to consider when deciding on a home

1. A house in College Creek neighborhood, then holding everything else constant, on average the sale price of the house will be \$20,080 less than a house bought in Bloomington Heights neighborhood
2. A house bought in Meadow Village neighborhood, on average, has a sale price that is \$19,030 more than a house bought in the Bloomington Heights neighborhood
3. An increase in a house's exterior material quality from, ex: good, to excellent, on average will lead to an increase in the Sale Price by \$17,250.
4. An increase in the Heating Quality and Condition, ex: Fair to average, will lead to an increase in the sale price by \$2,012 dollars.
5. An increase in the ordinal category of the Kitchen Quality ex: Typical to good, will lead to an average increase in the sale price by \$11,330.
6. An increase in the ordinal category of the Fireplace Quality ex: Good to excellent, will lead to an average increase in the sale price by \$1,961.



THANKS

