



Real Estate Valuation Model

Presented by Clarence Alvarez



Background and Task

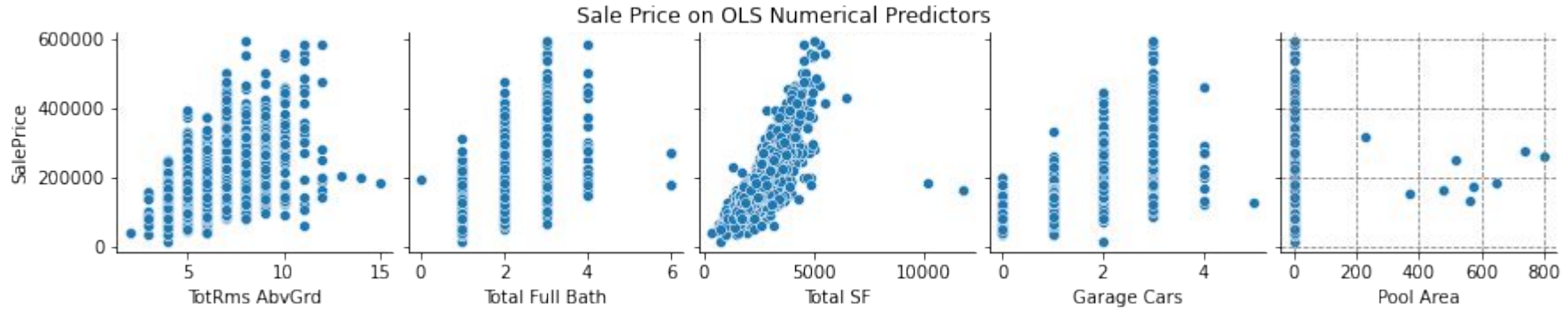
August, 2010

New Data Scientist for new real estate company, Oppenhouser that uses its own valuation model

Management requested for an updated model



Old Model: Ordinary Least Squares Regression



Dummified Variables include: 'Neighborhood' and 'Overall Cond'

- Top 14 largest coefficients were for the dummified variables
- Lowest coefficients for numerical predictors: Total SF, Pool Area

OLS Performance

Cross Val	0.7649
R^2_{train}	0.8019
R^2_{test}	0.8002
RMSE ₂₀₁₀	\$34,311.47
Mean of Residuals ₂₀₀₆₋₂₀₀₉	-\$378.04
Mean of Residuals ₂₀₁₀	-\$1,366.86

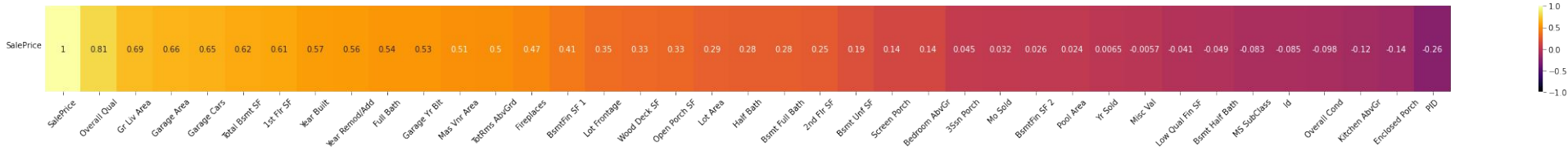
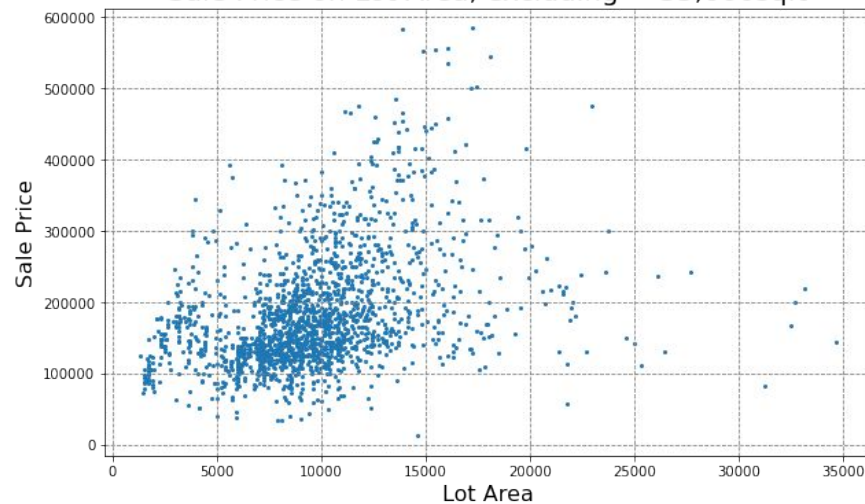


Considerations for Additional Predictors

Mean Sale Price along Timeline



Sale Price on Lot Area, excluding > 35,000sqft



Alternate Model: Ridge and LASSO

New Predictors:

Interaction term between
Overall Quality and Overall Condition

MS SubClass: Identifies the type of dwelling
involved in the sale.



Model Valuations

Metric	Linear	Ridge	Ridge (New Preds)	LASSO	LASSO (New Preds)
Cross Val Score	0.7649	0.7623	0.7897	0.8034	0.8021
R^2_{train}	0.8019	0.9103	0.9261	0.9096	0.9334
R^2_{test}	0.8002	0.8313	0.8795	0.8318	0.8938
RMSE ₂₀₁₀	\$34,311.47	\$28,506.35	\$27,708.28	\$28,513.12	\$27,479.60
Residuals _{'06-'09} *	-\$378.04	-\$305.22	-\$305.22	-\$305.22	-\$305.22
Residuals ₂₀₁₀ *	-\$1,366.86	-\$8,551.63	-\$8,551.63	-\$8,551.63	-\$8,551.63
Optimal α		1.5922	954.5484	28.48	811.1308

Note: * = Mean of Residuals

Recommendation

Adopt the LASSO (w/ New Predictors) model for its low comparatively low RMSE

I encourage underbid the predicted price of future homes to remedy the model's over-estimation.

Re-examine model more frequently due to higher negative mean residuals and decreasing peaks in the mean sale price.