

Willow Real Estate

Using Machine Learning To Predict Housing Prices



A Pilot Program In Ames, IA

Can Willow Provide Better Housing Estimates Than Our Competitors?

- Housing Estimates Are The Most Important Service A Real Estate Database Co. Can Provide.
- How Do We Decide What Home Features Have The Biggest Impact on Sale Prices?
- What Models Provide the Most Accurate Predictions For Sale Prices?

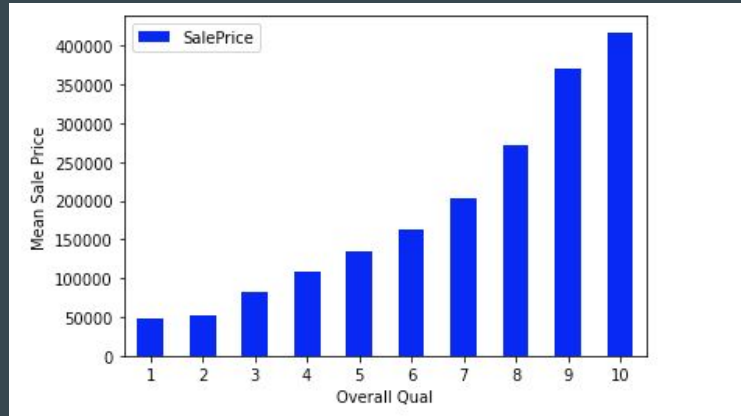


DATA

- Data Was Provided By The Ames, Iowa Assessor's Office
- Over 2,000 Records Of Housing Sales in Ames From 2006 to 2010
- Contains 80+ Categories Of Building and Land Details

Exploratory Data Analysis (EDA)

- Address Missing Values, Identify Any Missing Data & Replace Appropriately
- Identify Any Outliers, Once Identified Decide If The Outliers Should Be Removed
- Visualize Data To Help Identify Strong Correlations To Sales Price, Decide If Further Investigation Is Required.



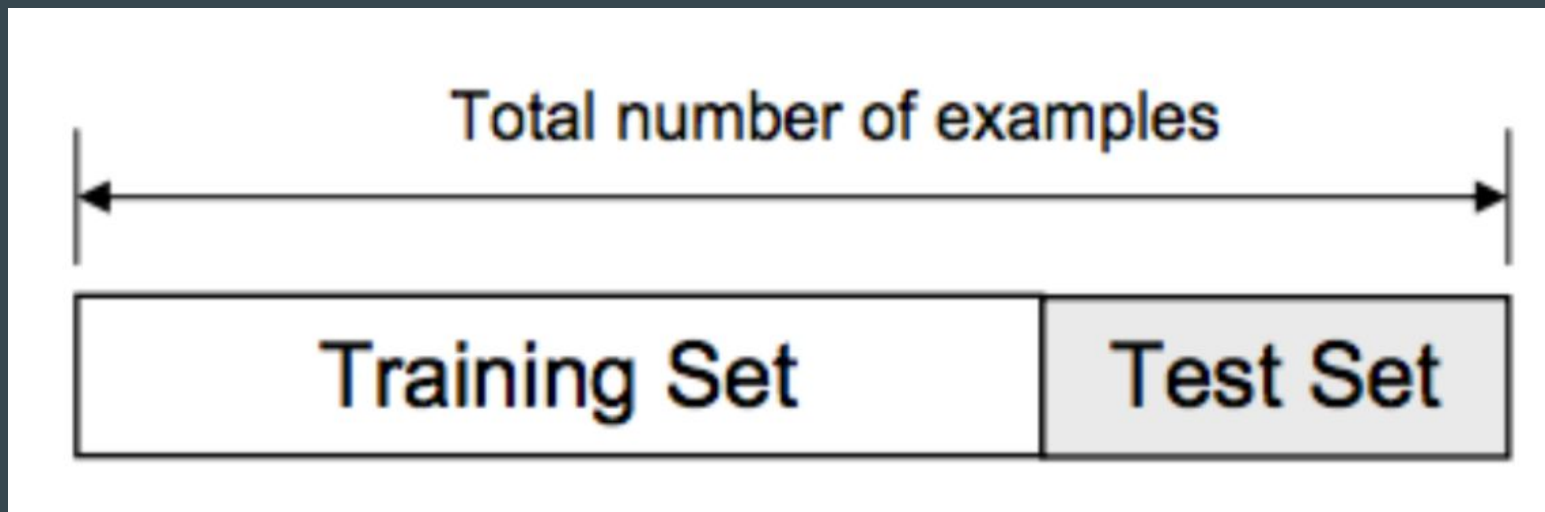
Correlation Heatmap



-Allows Us To Identify The Variables That Have The Highest Correlation To Our Target Variable Sales Price.

Modeling

Train/ Test Split: Data Was Separated Into A Training Set And A Testing Set. Model Is Fitted On The Training Data So We Can Make Predictions On The Testing Data.



Modeling

Baseline Model: Created First To Generate Initial Predictions For Our Dataset. Used To Measure The Performance Of Our Other Models. Derived From the Mean Sales Price.

Train Score (RMSE): 79,526 Test Score (RMSE): 78,375

Linear Regression Model: Multiple Regression Model Where We Will Include Our Selected Features To Predict Our Target Variable Sale Price.

Train Score (RMSE): 37,495 Test Score (RMSE): 31,770

Ridge Model

Train Score (RMSE): 37.574 Test Score (RMSE): 32.101

- Used In An Effort To Reduce Variance In Our Model
- Ridge Model Is Affected By the Largest Coefficients In The Data Set

Lasso Model

Train Score (RMSE): 37.603 Test Score (RMSE): 32.291

- Lasso Model RMSE Score Is Affected By Lack Of Features In The Model

Conclusion

- Based Off Of The Results Of This Model Willow Is Not Ready To Compete With Zillow's Automated Valuation Model (AVM)
- Additional Feature Engineering Is Necessary To Identify More Linear Relationships To Sale Price
- Model Did Not Include Enough Features. More Features Should Be Included And Then Reduced As We See Fit.