Ames Housing Data

Predicting House Prices with Machine Learning

Chiam Xiu Ting

Objective

House price prediction to help sellers to set a more accurate and realistic price to attract potential buyers

Ames Housing Dataset

Dataset Source: Ames, Iowa Assessor's Office

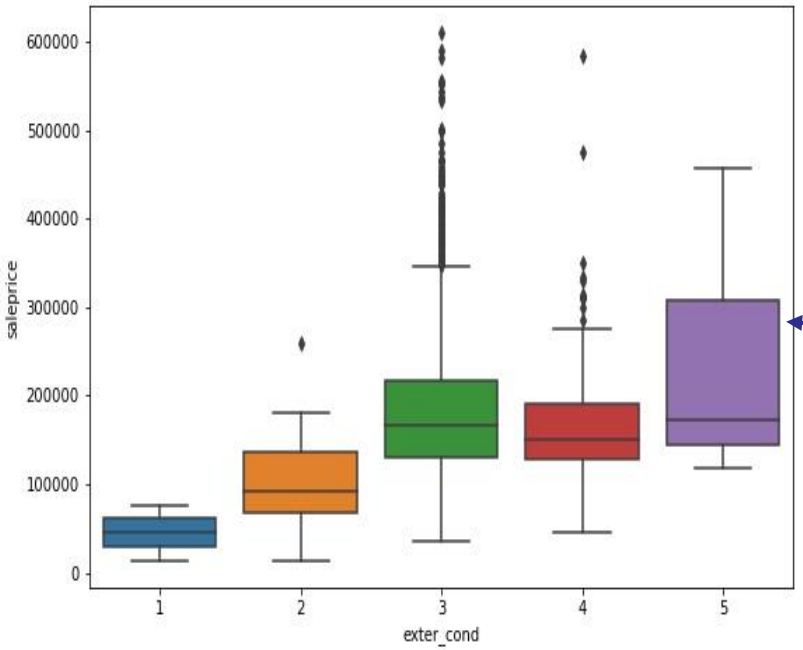
2051 Observations

81 Variables

What do we want to predict?

Sale Price

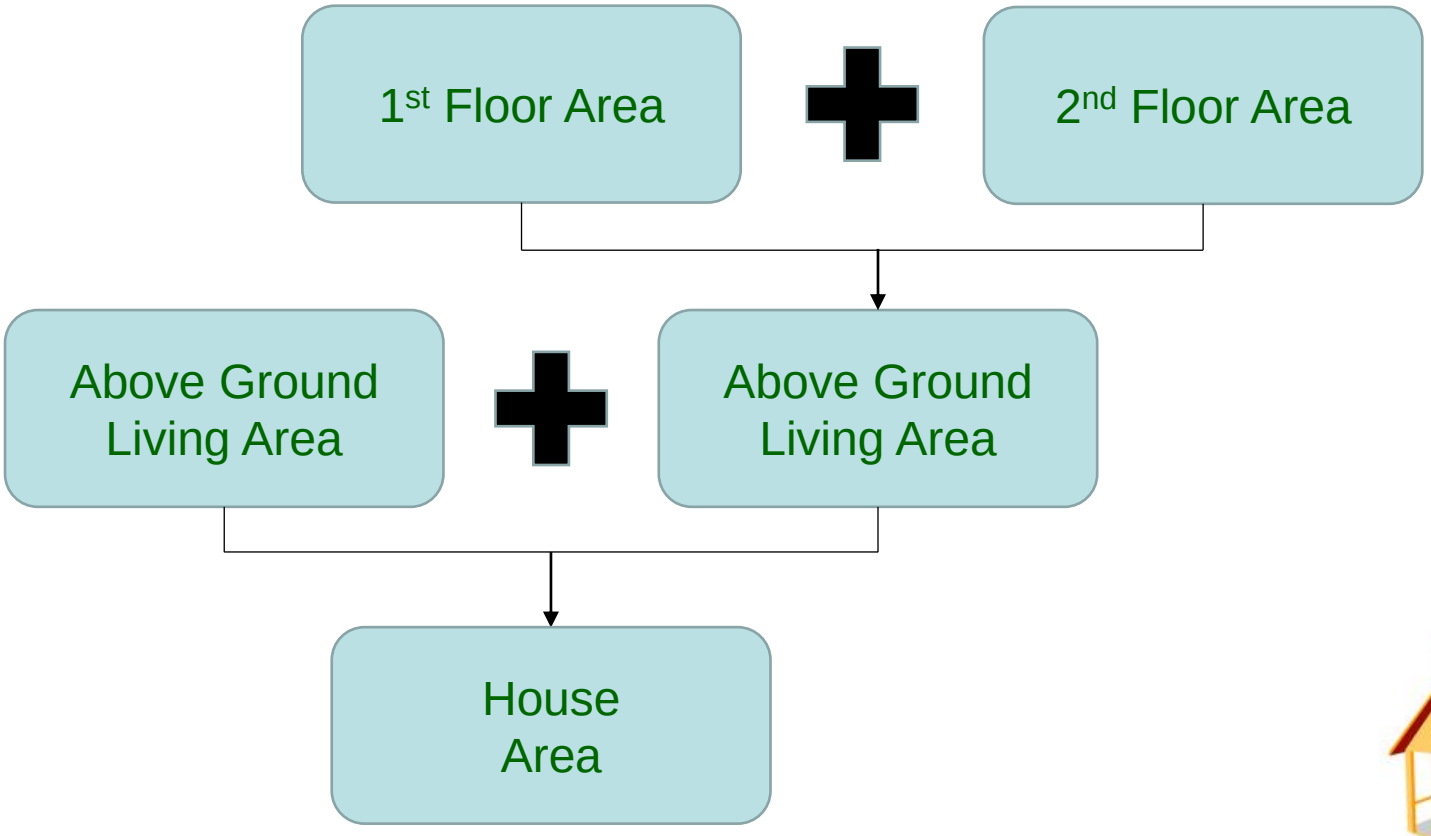
Data Cleaning



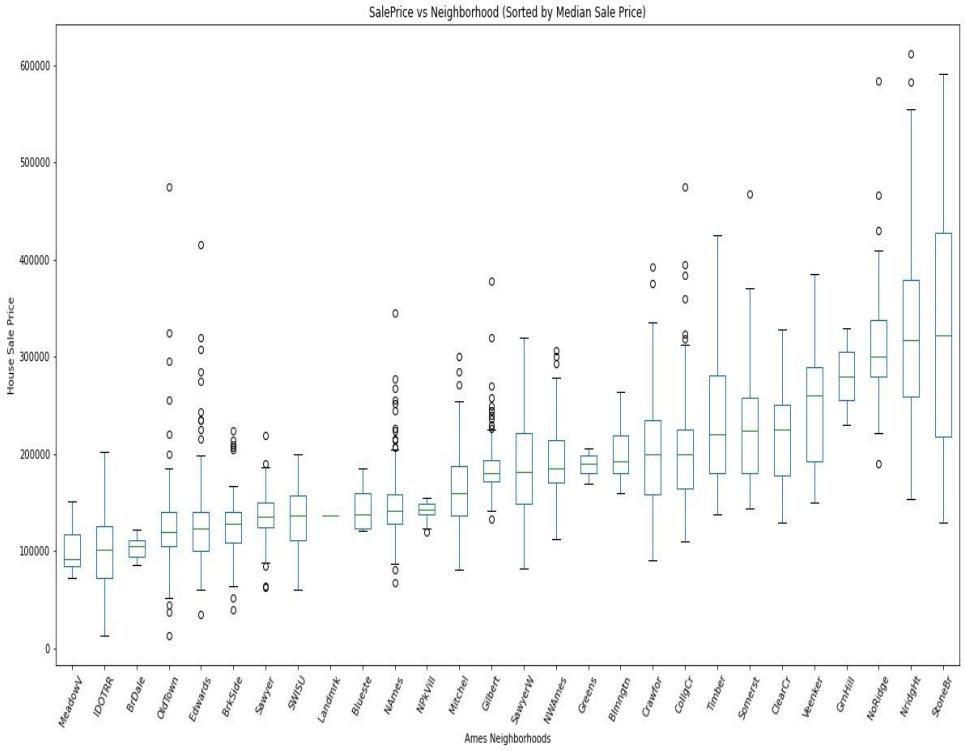
- Understand the dataset and what each variable represents
- Check for NaN values (no data entered)
- Check for wrongly keyed in data
- Fill in null/NaN values
- Removing observations that differ significantly from other observations (outliers)
- Set a numerical rating for variables that can be rated according to their categories. The higher the rating, the better.
- Removed variables with little data and inputs

Feature Engineering

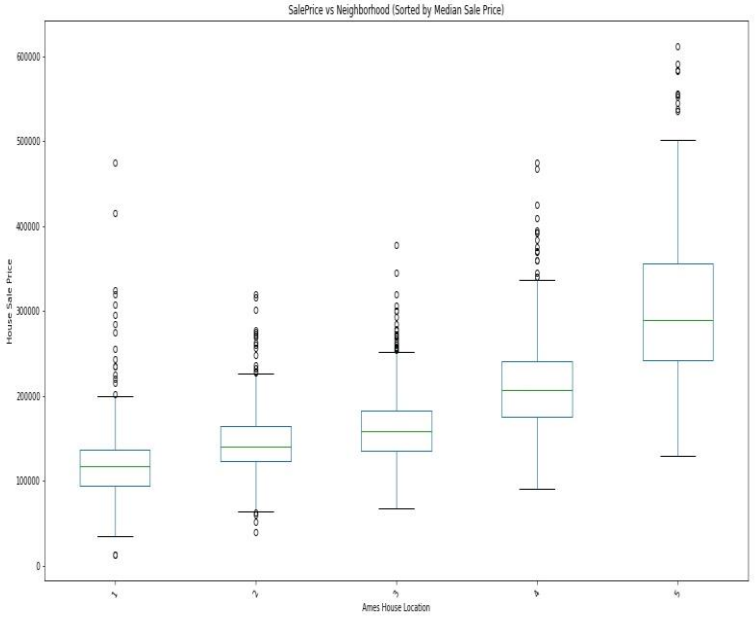
What can be understood from the data?
What is more important?



Feature Engineering



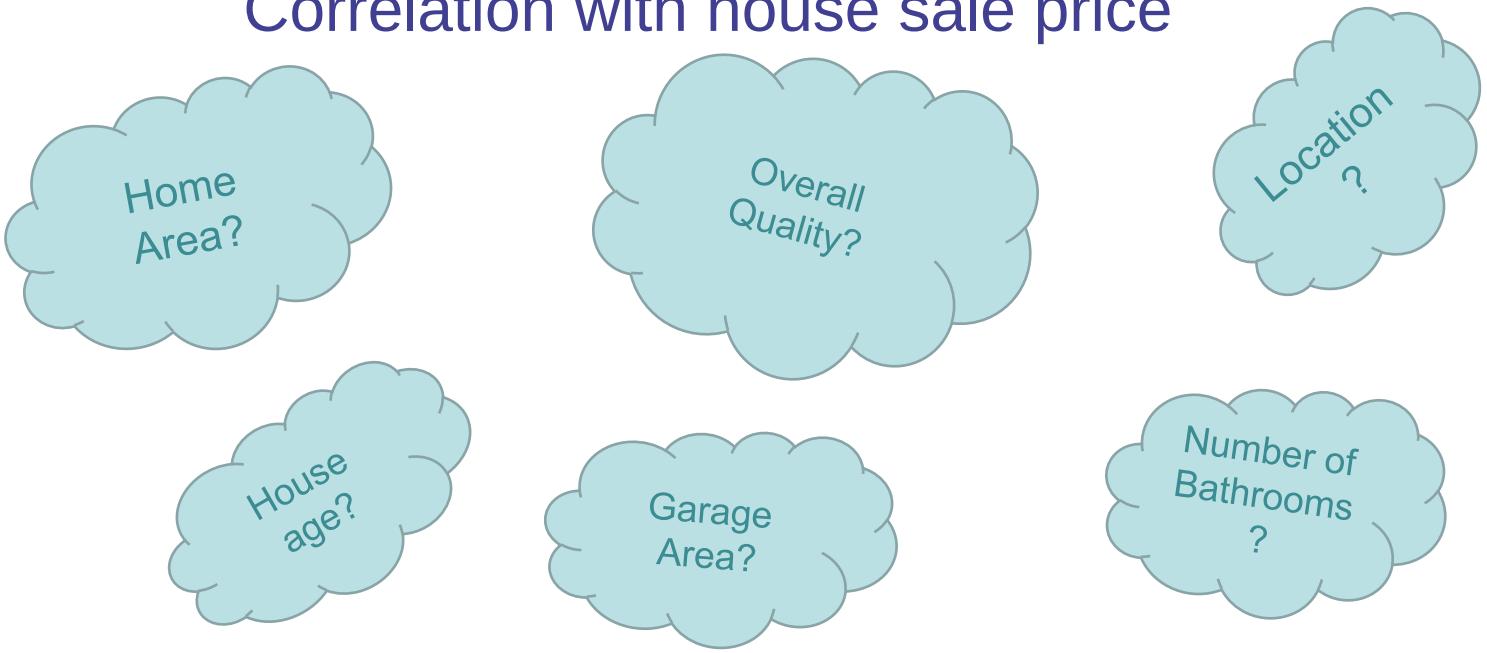
VS



Feature Selection: Methods

What features affect house price the most?

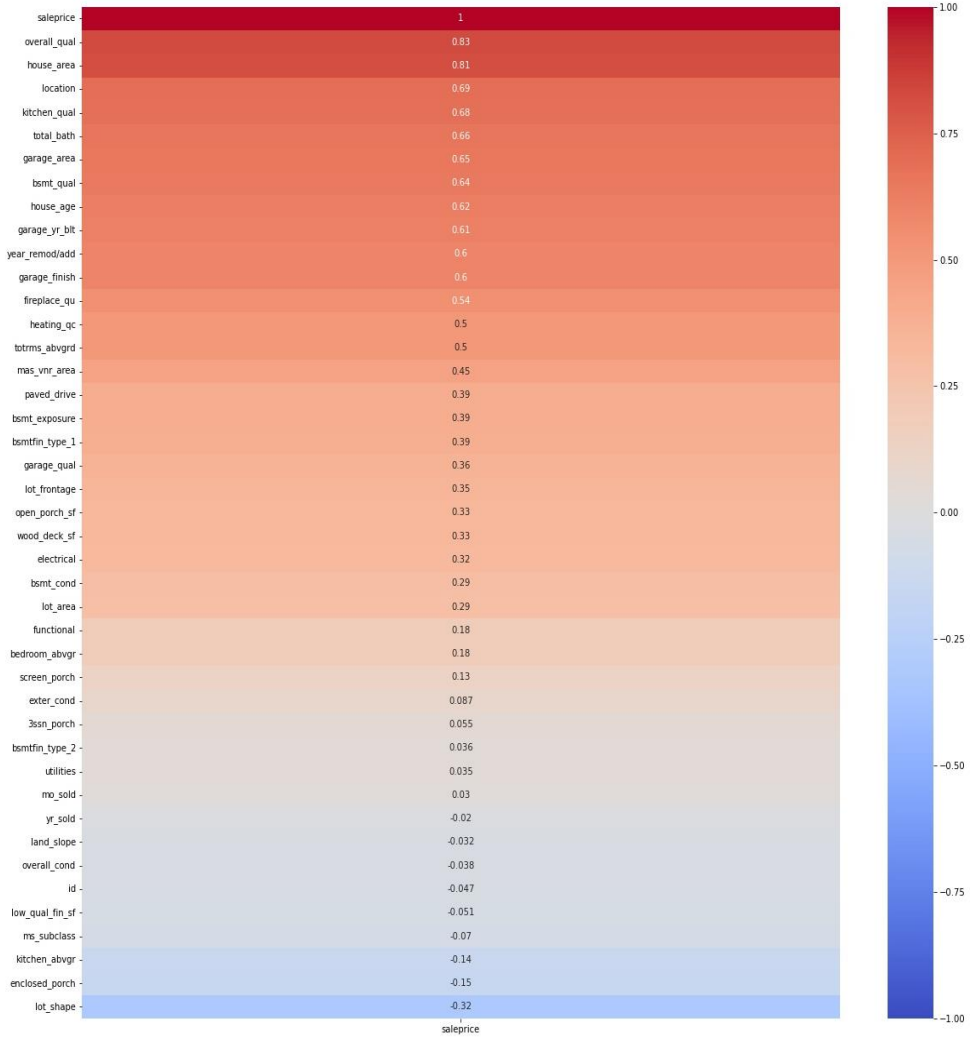
Correlation with house sale price



Can the feature selection be automated?

SelectKBest

Feature Selection: Methods



Features selected using Correlation with price

The chart shows how closely each feature is related to the house sale price

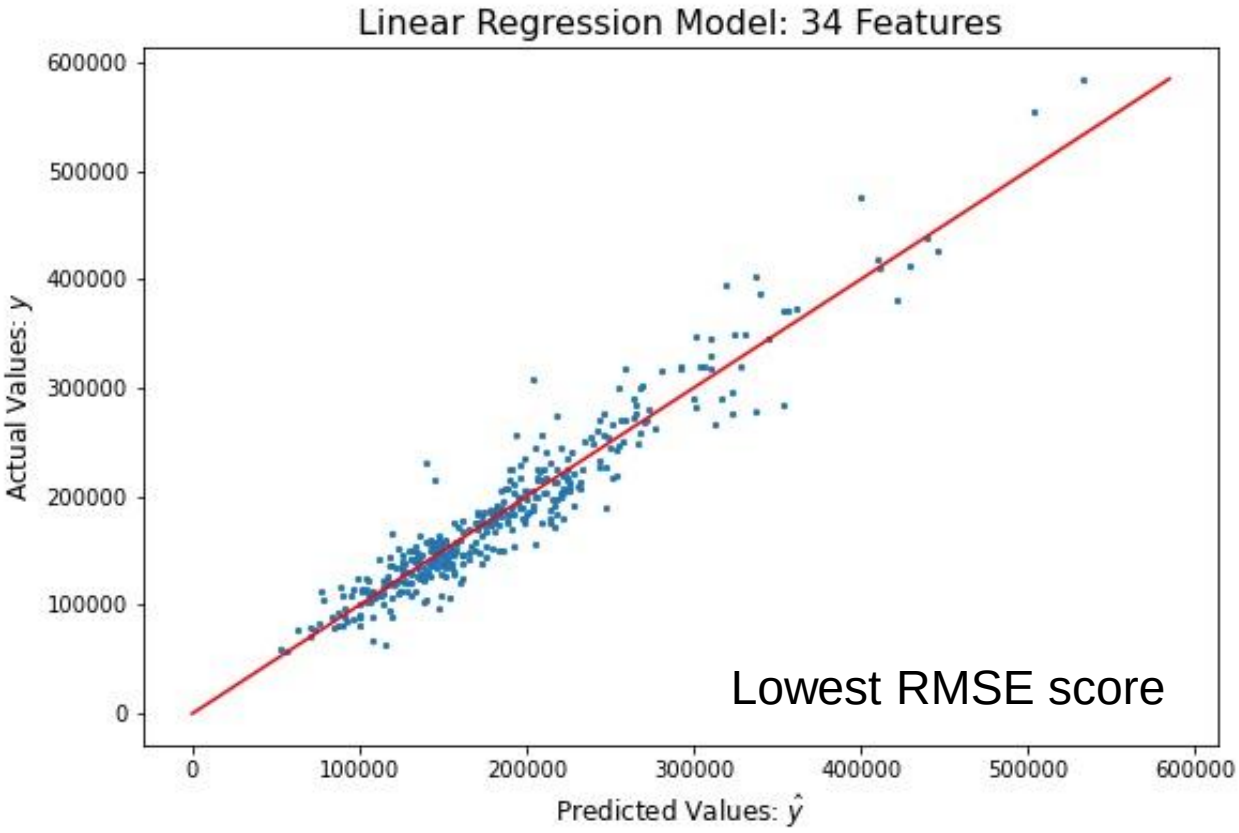
Prediction Models: Score Table

	All = 141 RMSE	All = 141 CVS	k = 32 RMSE	k = 32 CVS	k = 40 RMSE	k = 40 CVS	Pearson RMSE	Pearson CVS
Linear	26341.76569	25772.04854	22262.45114	28666.70869	25143.31035	27530.76764	21861.55688	28669.56642
Ridge	25821.36483	25728.94518	22387.57431	28652.73011	25618.12013	27504.44695	22006.91196	28655.77101
Lasso	25341.01039	25772.02406	22383.26889	28666.70848	26073.23079	27530.76699	22014.29798	28669.56618
ENet	25077.79293	25667.25356	22390.17746	28648.84054	25935.41832	27496.17526	22013.1026	28653.46857

The lower the RMSE Score, the closer is the prediction to the actual price.

Prediction Models

Final Model: Linear Regression Model using 34 features selected using Pearson Correlation Coefficient method



Conclusion / Recommendation

Conclusion:

Useful insights for Home owners (What will affect price?):

- overall quality of the house
- size of the house
- location of the house
- kitchen quality
- total number of bathrooms
- garage area
- age of the house

Recommendation to Company:

Collect information on ‘Condition of Sale’ and do further analysis on whether it will improve the RMSE score of our model.