Real Estate Price Predictor

Springboard — Data Science Career Program

Capstone Project #1

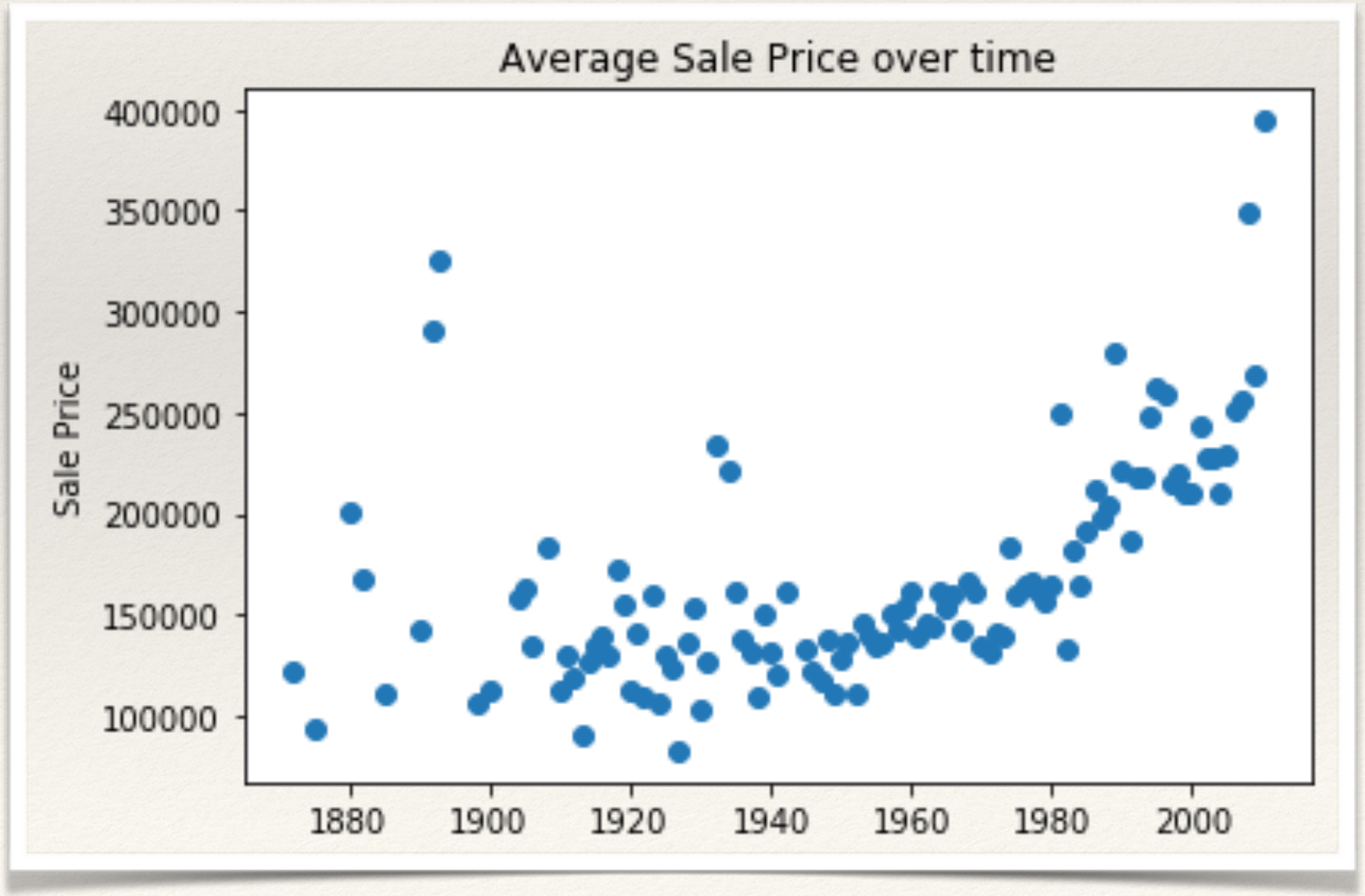
By Kevin Cole

Business Problem

In real estate, finding properties that are undervalued and avoiding properties that are overvalued can help the long term prospects of the investment. Predictive models of housing prices would then be valuable to help investors identify valuable or mispriced properties and sellers to price their properties appropriately. This project seeks to create predictive models for sale price for homes in the Ames, Iowa housing market.

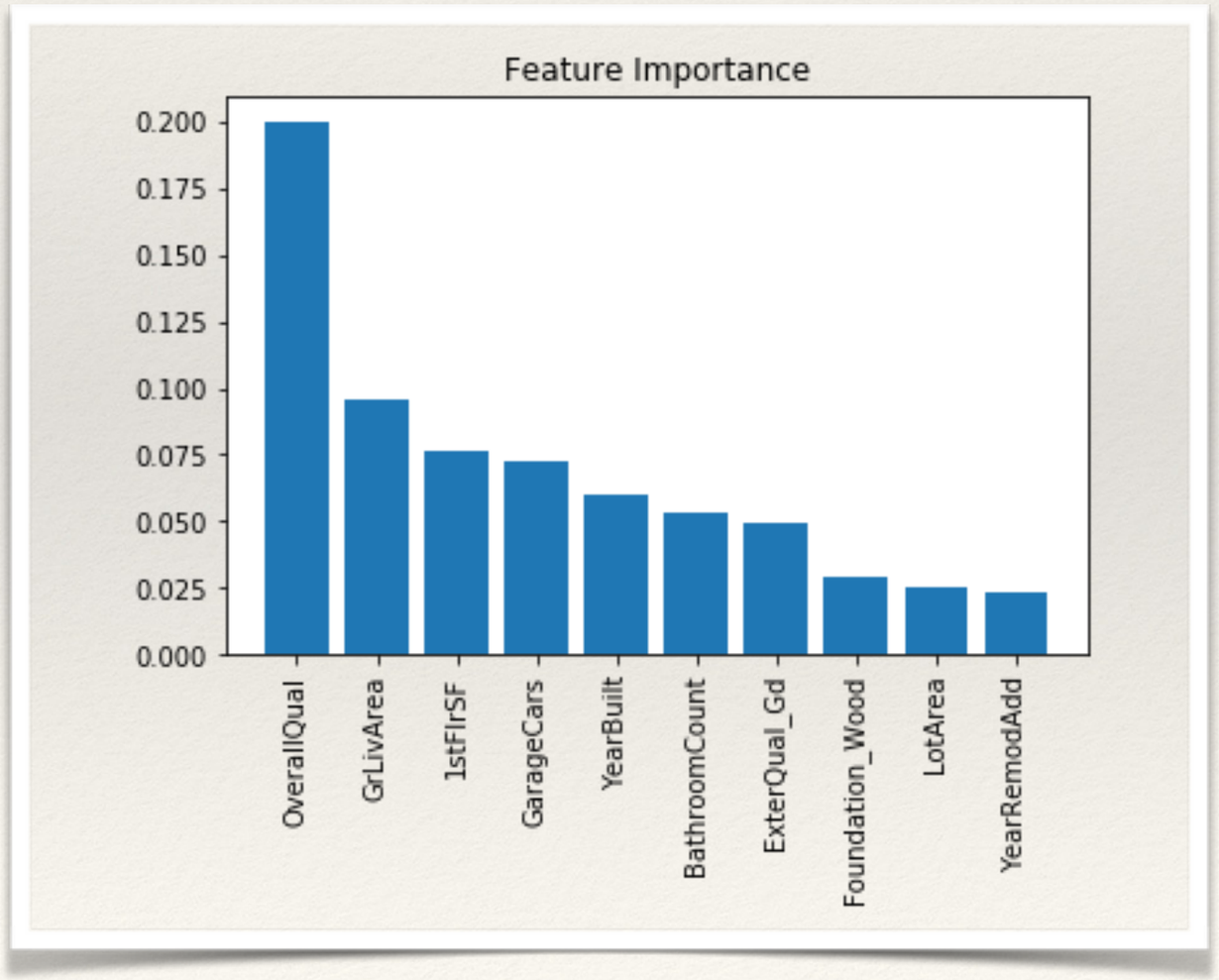
Analysis of local market

- For new constructions, builders and real estate investors should make sure the local market has increasing value.
  - Average sale price in the Ames, Iowa market has been accelerating over the last decades, proving to be a worthwhile location for new construction



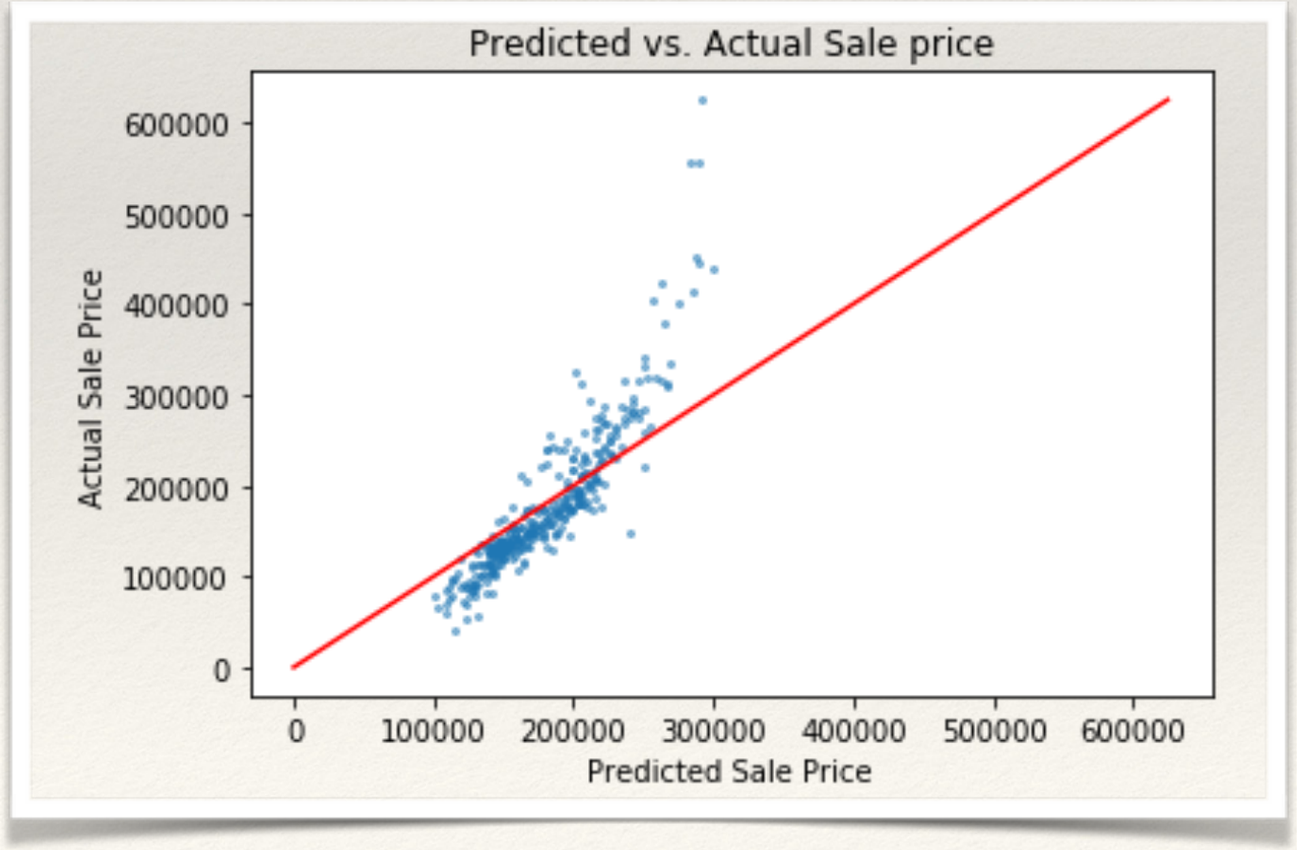
Feature Analysis

- Understanding which features play strong roles in sale price can help investors, builders, and realtors make business decisions.



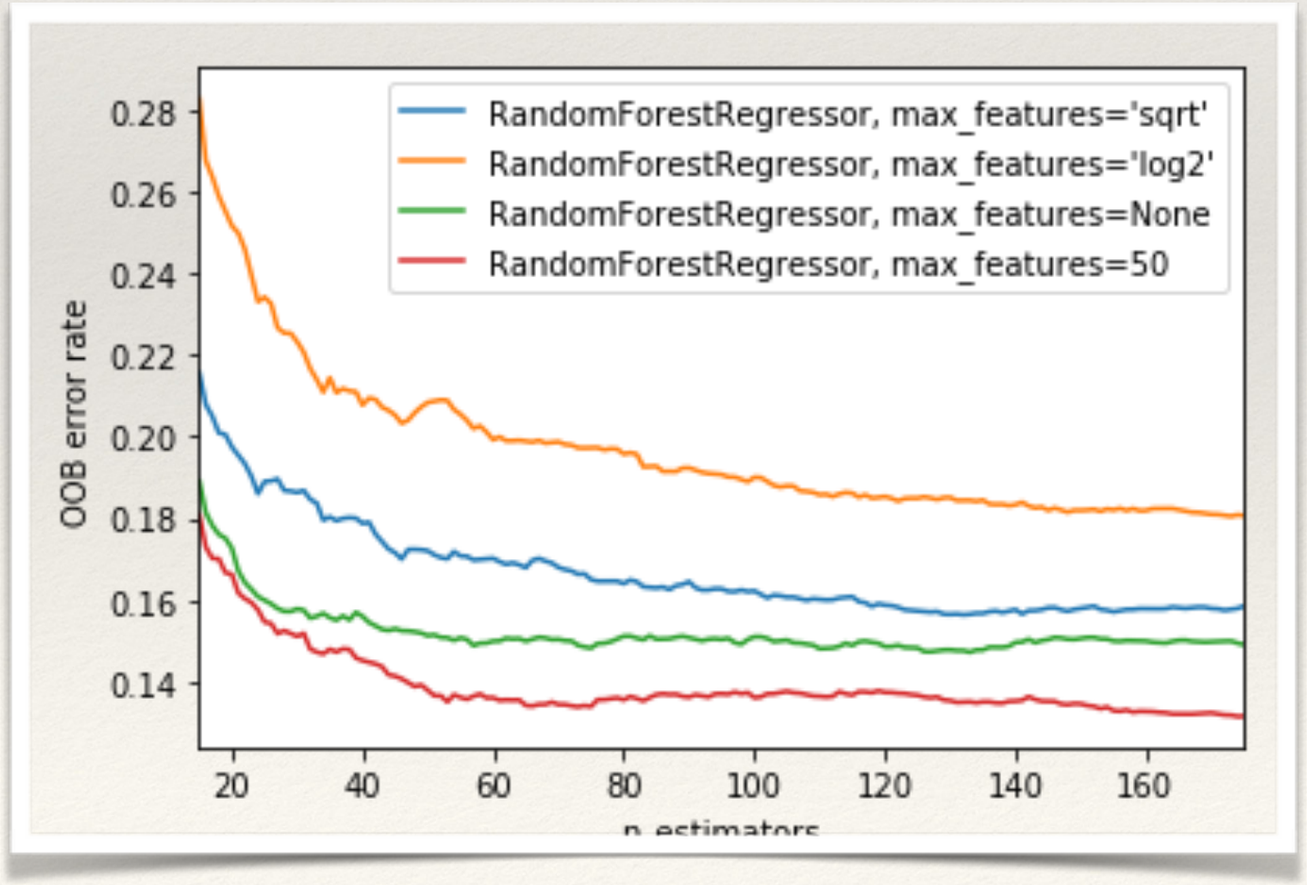
Baseline Model

- Ridge regression was selected as the model.
- Model minimizes overfitting of training data.
- Still has larger errors particularly around more expensive properties.

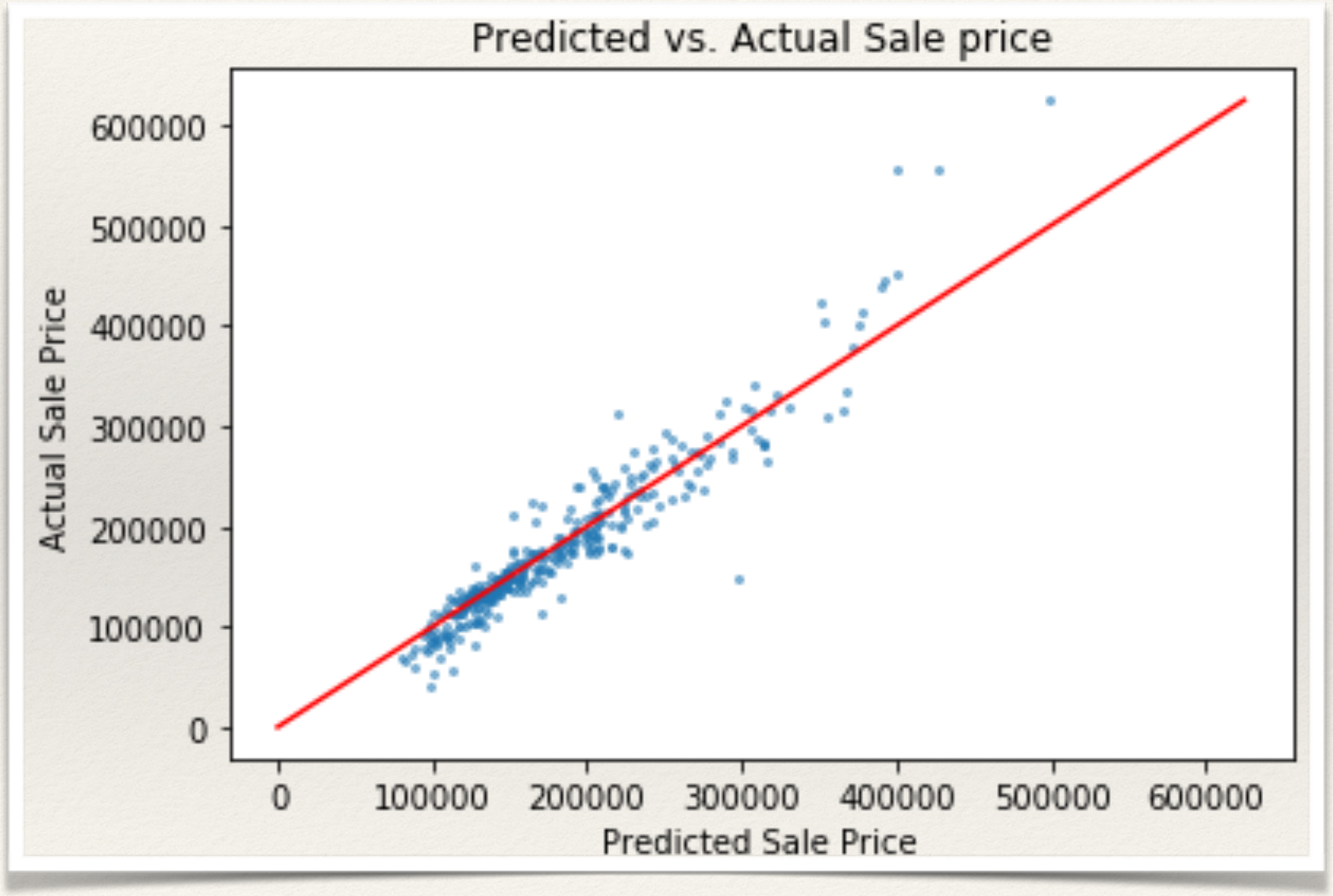


Random Forest Model

- Required tuning of hyperparamaters.
- OOB error measures the prediction error of random forest models
- Error is smallest with 50 features considered per node.
- The error levels out as the number of trees in the forest approaches 100.



Random Forest Model performance



Future Improvements

- Different subsets of the original feature set - given the detections of important features discovered in these models, tuning of the feature set could be strengthened.
- Other models such as Neural Nets could be built and evaluated.
- Expansion of the model to cover nearby markets or different markets all together given data for those markets.

Recommendations

- Undertake actions that can increase a buyer's initial perception of the property. This could include upgrades like paint or flooring prior to listing, or staging efforts at open houses.
- When identifying undervalued properties, the model should be producing a predicted price higher than listed. Due diligence is still required in these instances as there can always be more to the story for an individual property.
- Identifying properties for investment should include not just properties where the model has produced a predicted price higher than listing, but those who score poorly in the high feature importance categories. Identifying categories that are easy to upgrade can lead to fast turn-arounds and remodels.
- When designing new constructions, the coefficients of the linear regression model give you an idea on how the price would grow as a function of features such as number of garages, bathrooms, or square footage. Planning these features into the design can directly lead to a more valuable property.
- Continue to feed and refine the model if it is indeed to be used over time. The model was built on a relatively small data set and more data could prove valuable. The model should not be used in other markets until trained on data from that market as market conditions can vary greatly by region.