

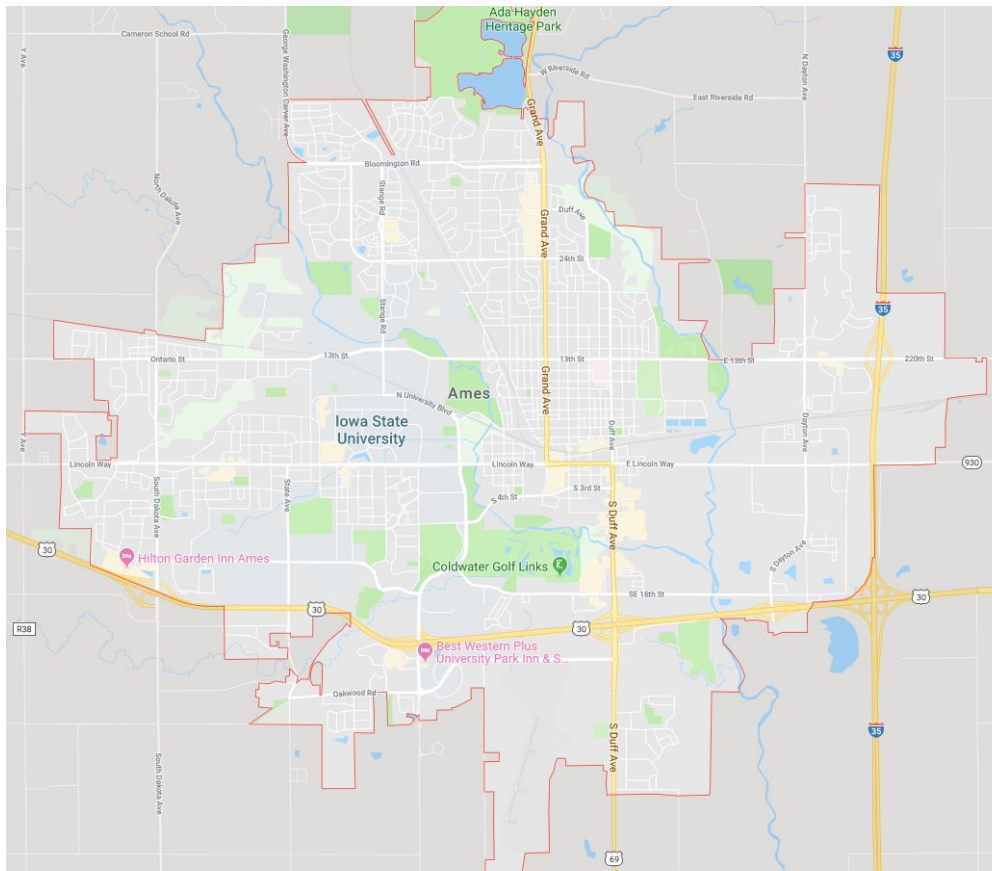
Predicting house prices in the City of Ames

Mai Anh Ly

About the dataset

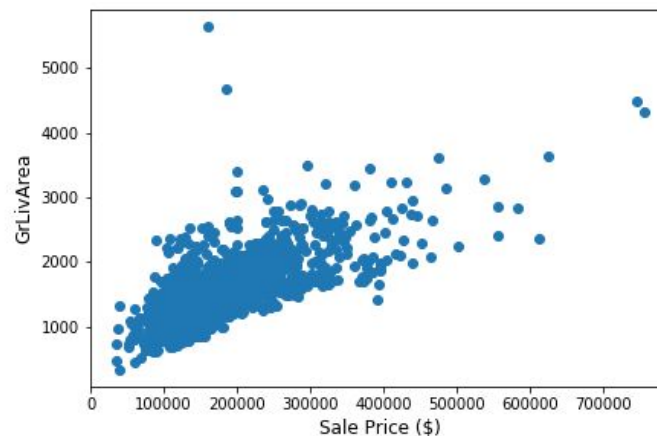
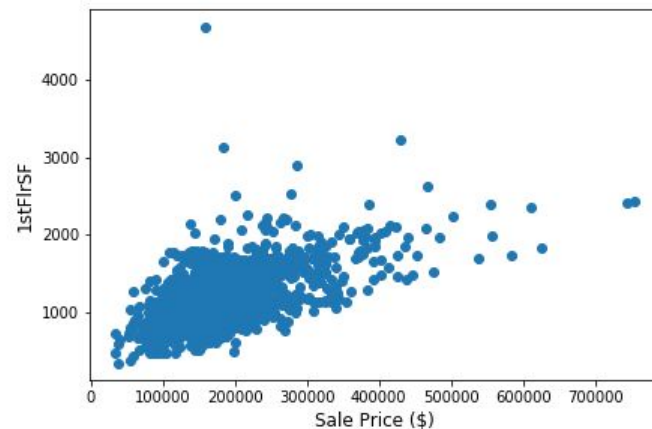
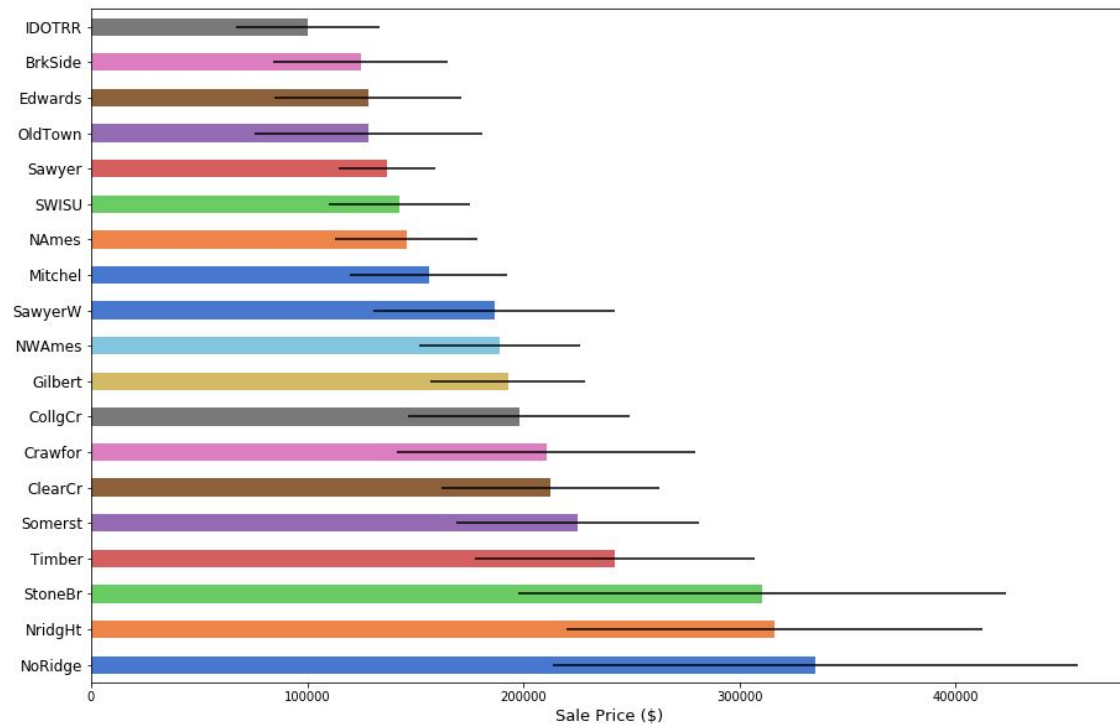
- Data on 1460 houses sold between 2006 to 2010
- Data includes features describing:
 - House/Lot area
 - Quality and condition of house and utilities
 - House neighbourhood and street accessibility

**Predict price of houses
sold in 2010**



Feature selection

- Features in model selected via:
 - Domain knowledge
 - Correlation with house price after EDA

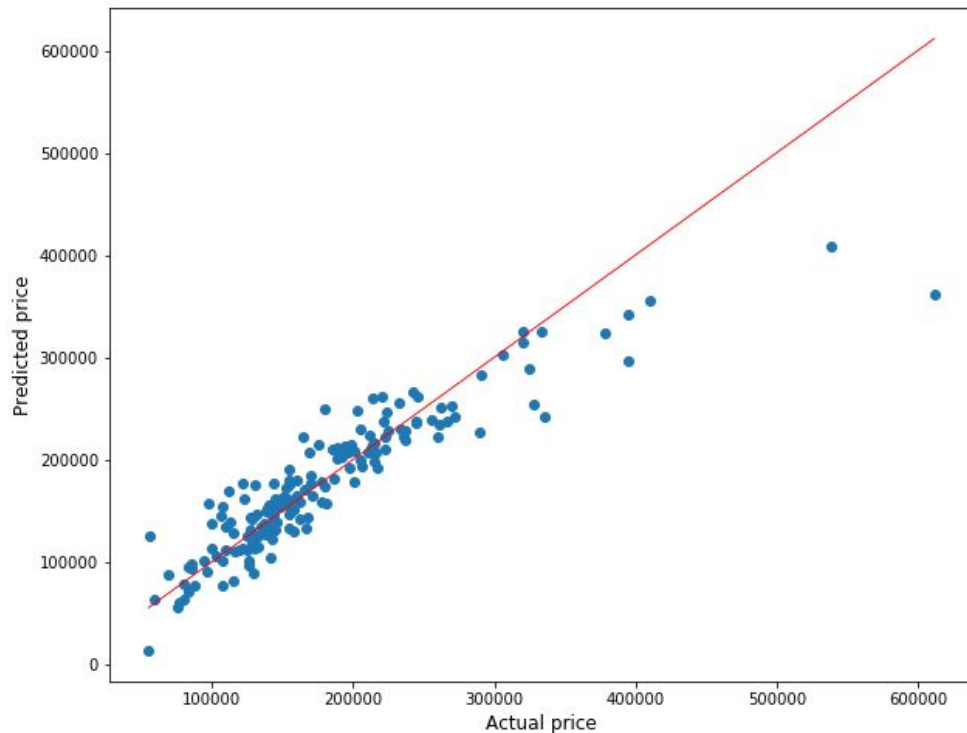


Model evaluation

Selected model can predict price with an upper/lower bound of **\$33 617**
(19% of the mean sale price)

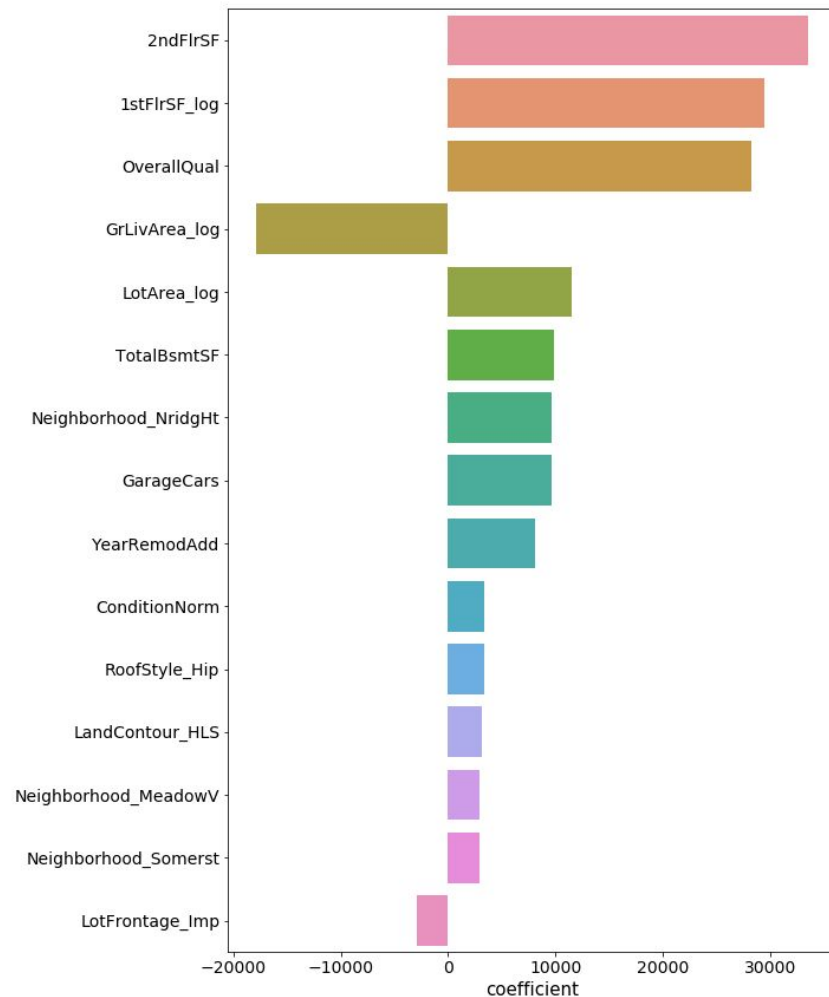
No. features	R2	RMSE (\$)
1	0.61	49858
15	0.82	33617
31	0.81	34561

... but undervalues houses worth
more than \$300 000



Key findings

1. Newer, high quality multi-level houses mean higher prices
2. Location matters!
3. Hip roof style, more car space and being on a hill adds value



Further improvements

- Finetuning the model to improve prediction of high value houses
- Additional data: house addresses, distance to places of interest
- Additional utility: inform clients **when** is the best time to sell