Project 2
Team MLP*

Wandi
Limin
Rohit
Tze Yi

Ames Housing Project

*My Learning Pony

Problem Statement

- Data Scientists within a property agency
- As part of our marketing campaign, we are giving advice to house owners / potential buyers
- (House owner) What type of features would bring the most value to a specific house?
- (Potential Buyers) What features to look out for when purchasing a new house?

Data Science Problem - Predict Ames Housing Prices

Problem Statement

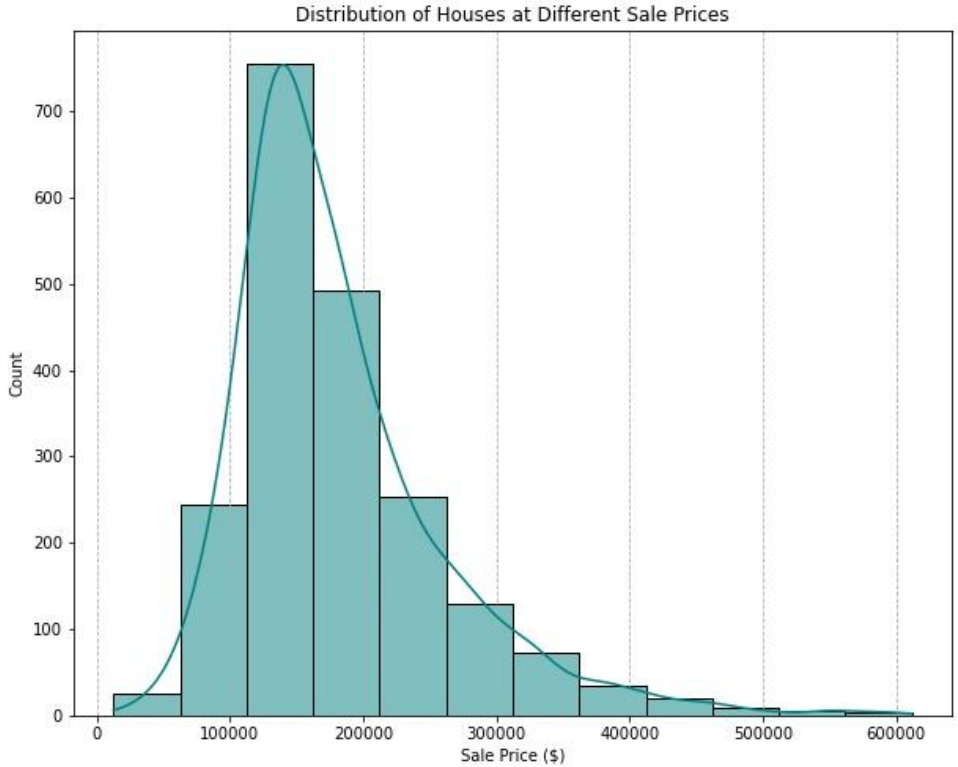
What type of features would best predict the sale of the house?

Concerns to address

- What features add the most value to a home?
- given a set of features, what is the expected sale price of a house?
- given a budget, what kind of house would one be able to afford based on our model?

EDA and Data Cleaning Process

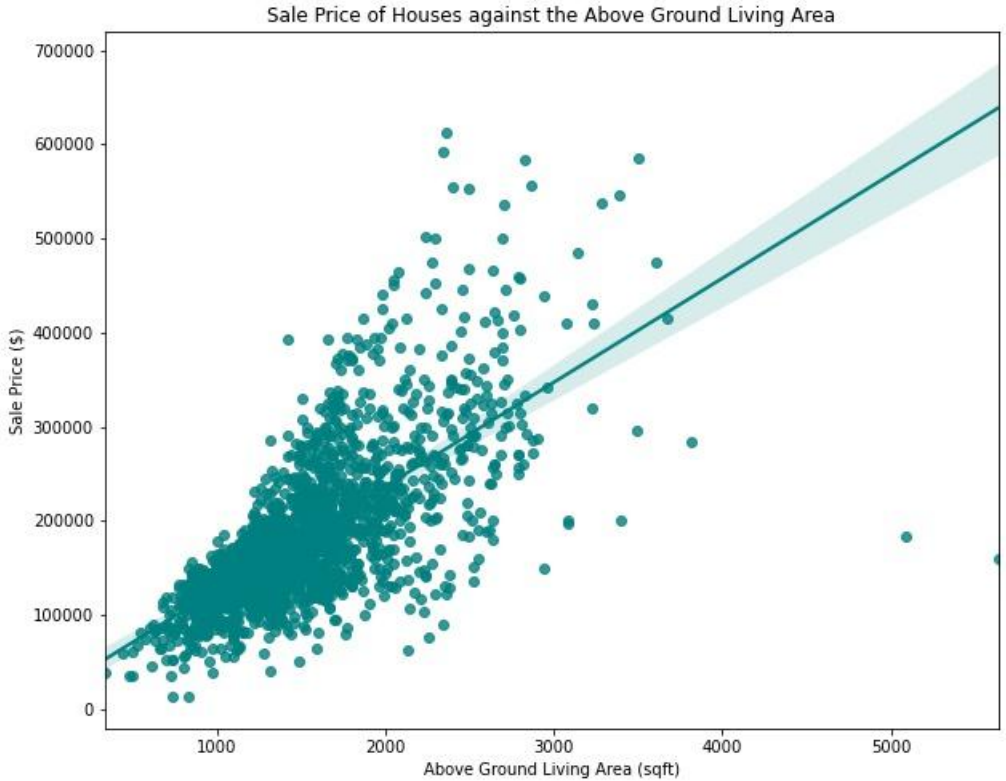
Distribution of House Sales Prices



Based on the data we collect, Ames tends to have houses that are priced around $162,500

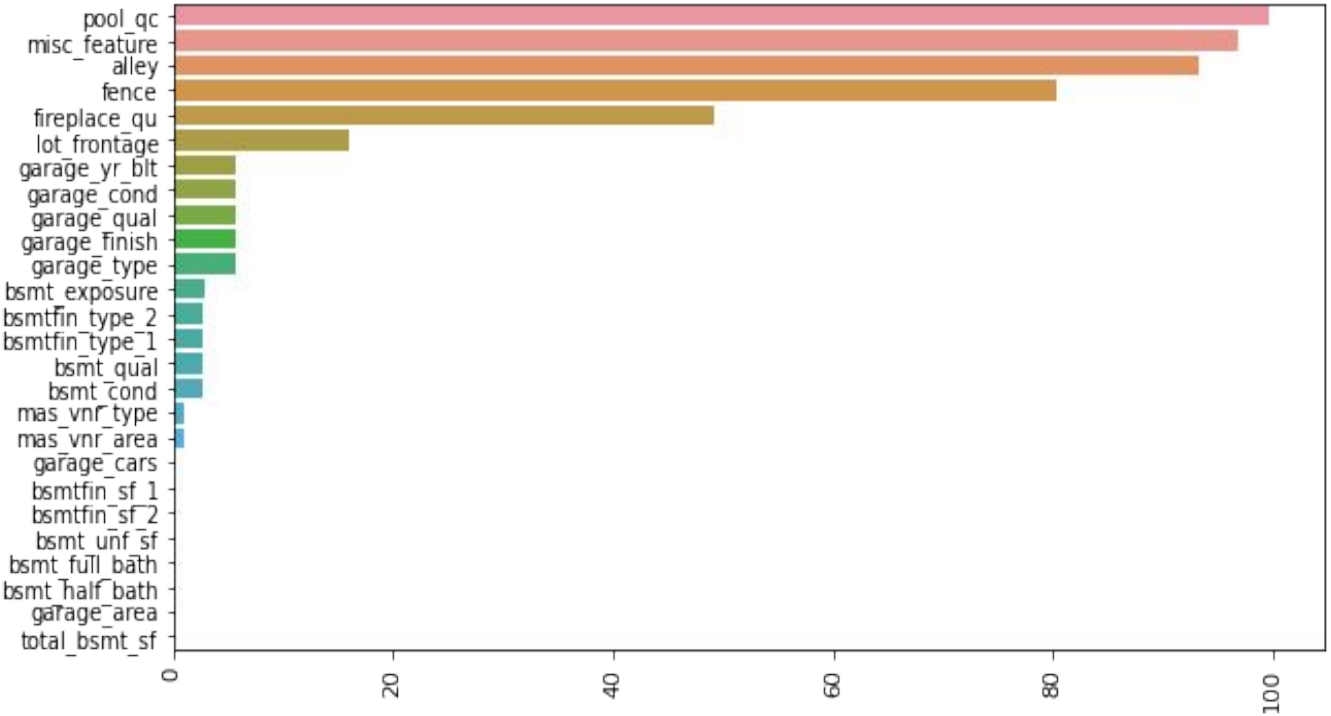
Houses that are priced above this amount tend to be more varied in its value than houses below.

House Size and Sales Price



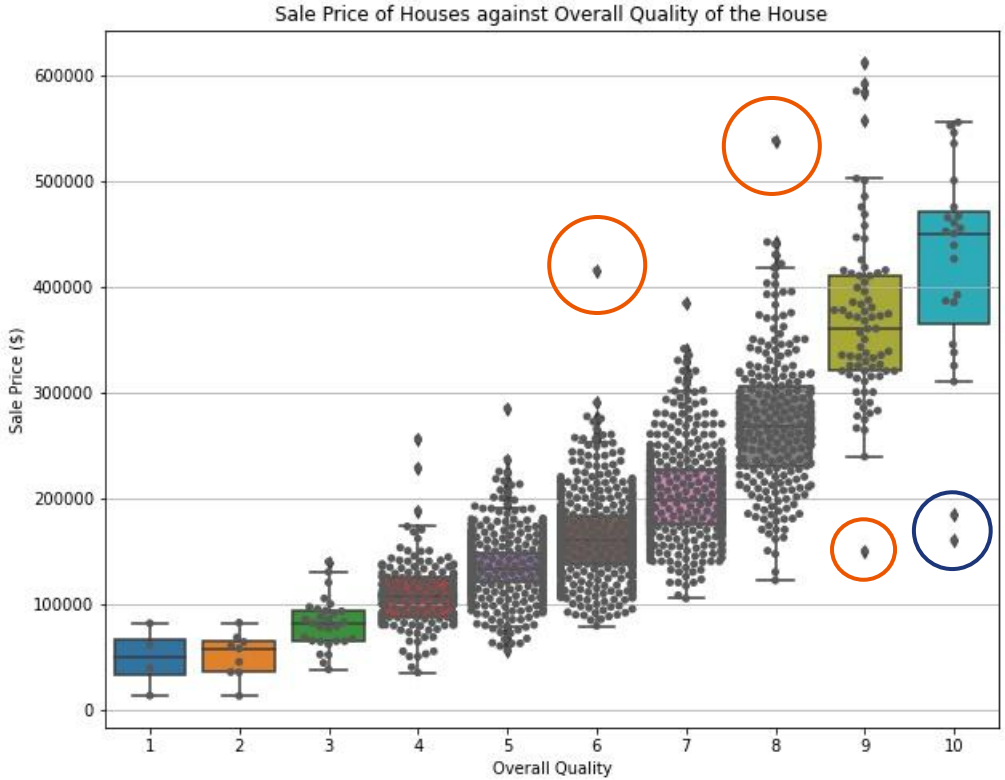
For every unit increase in above ground living area, sale price increases by about $120.

Null Values



In general, most null values represents the absence of a specific feature, hence they are imputed with 0 or ‘None’

Outliers



- 5 houses that need to be removed
  - 2 high quality houses, poorly valued
  - 3 partial sale houses

Pre-processing and Modelling

Procedure/Methodology

Any features combined?

- total_house_sf
- total_porch_sf
- total_bathroom
- total_bsmt bathrooms

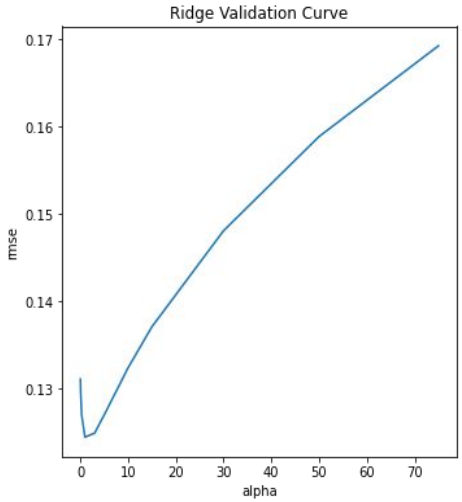
Any interaction terms ? - any investigation done for targets with linear relationships

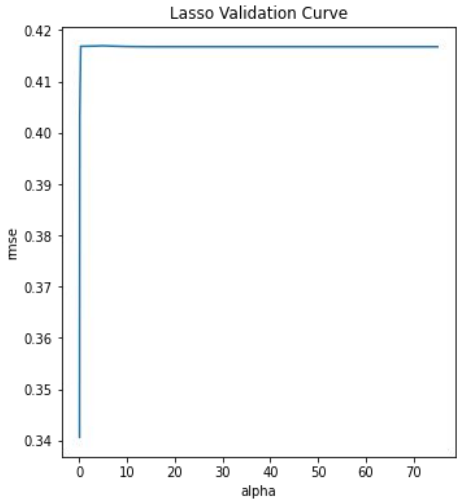
- Neighborhood
- Foundation
- Garage Type
- Ms_subclass
- Masonry Veneer Type

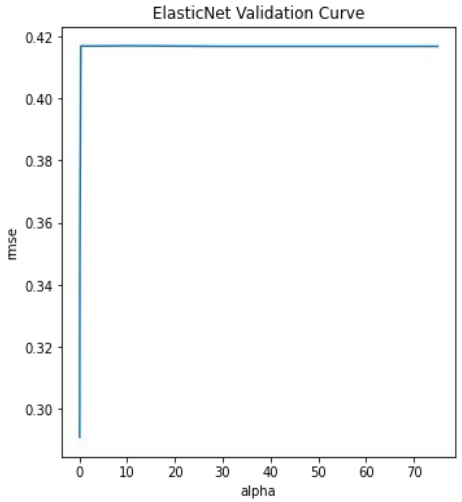
Any collinearity observed and what was dropped because of it?

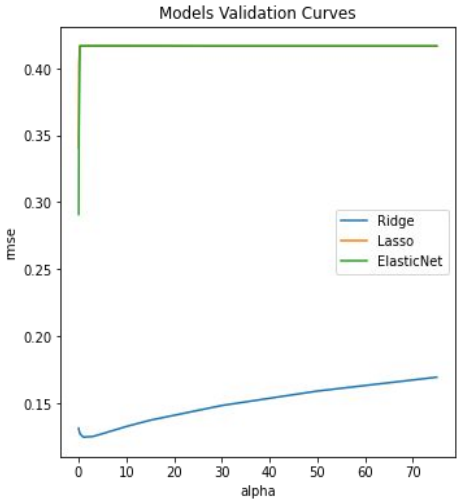
- External condition

Hyper parameter selection









Primary Findings

- Overall, the top 10 features selected by the model to be most predictive mostly applied to the following areas:
  - House size
  - Overall quality and condition ratings
  - Home functionality
  - Neighbourhood

Business Recommendations

Recommendations

Which features appear to add the most value to a home?

- Overall Quality & Above Ground Living Area

Which features hurt the value of a home the most?

- No one features strongly negative impact housing price

	saleprice
saleprice	1.000000
overall_qual	0.801002
gr_liv_area	0.716672
garage_cars	0.652764
total_bsmt_sf	0.649999
garage_area	0.649089
1st_flr_sf	0.632303
year_built	0.574620
year_remod/add	0.554174
garage_yr_blt	0.544069
full_bath	0.542694
mas_vnr_area	0.505241
totrms_abvgrd	0.503861

Recommendations

What are things that homeowners could improve in their homes to increase the value?

- By identifying important features of the house, owners may decide to renovate the house:
  - Garage
  - Basement
  - Kitchens
  - Bathrooms
  - Masonry Veneers

Recommendations

What neighborhoods seem like they might be a good investment?

- Current no. of houses (high population)
- Price of houses at sale (low sale price)

Current Top 3 areas:

➔ Northridge Heights
➔ Stone Brook
➔ Northridge

Recommendations

Edwards

- House Population: 143
- Price: $ 130,493.468531

Sawyer

- House Population: 111
- Price: $ 137,181.909910

Old Town

- House Population: 163
- Price: $ 125,276.300613

Recommendations

Do you feel that this model will generalize to other cities? How could you revise your model to make it more universal OR what date would you need from another city to make a comparable model?

- The model is able to generalize other cities but recommended that they have their own set of housing data for detailed analysis

Conclusion*

House predictions are affected by a myriad of factors that may not be represented in our dataset. Please use our model as a guideline instead of gospel. Thank you.

*Disclaimer
*T&C's apply

Thank You!
Q&A