The Key To Developing Expensive Houses

General Assembly Project 2

By: Liew Chin Xia, Lloyd Lau and Derrick Lim

Introduction

01

Problem statement:

With increasing population growth, there is a strong demand for new homes in Ames. As a data scientist in a private construction company, we are keen to examine the key features that affect sale price. In so doing, our company will be able to maximise profit while serving the demands of the local population.

02

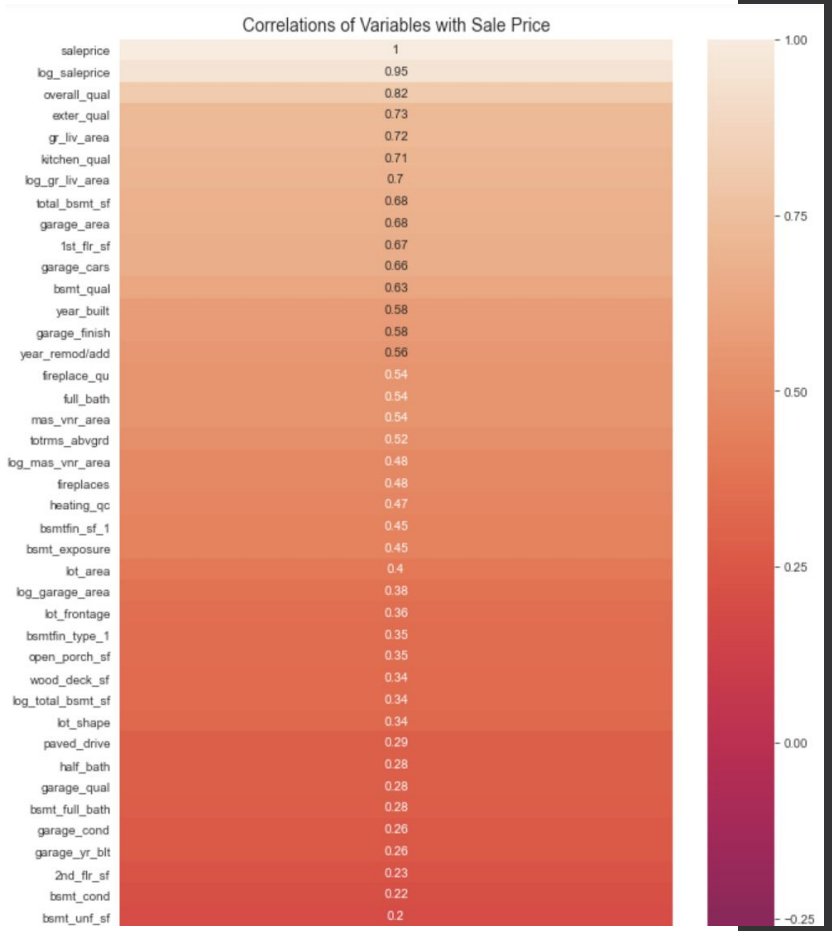
Data analysed:

- Housing-related variables (from Ames Assessor's Office) for residential properties sold in Ames from 2006 to 2010.
- 2051 observations with 81 columns of variables.

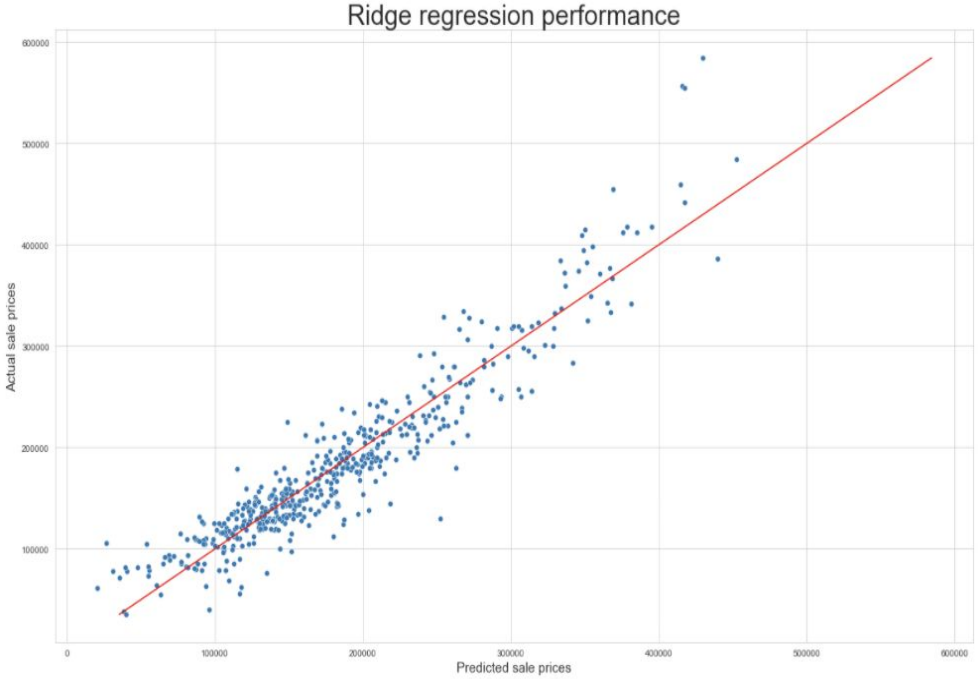
Methodology For Baseline Model

From the correlation heatmap, the following variables have a correlation > 0.5 with sale price

- Overall material and finish of the house (overall_qual)
- Quality of exterior material (exter_qual)
- Quality of kitchen (kitchen_qual)
- Living space above ground (gr_liv_area)
- Size of garage in sq feet (garage_area)
- Size of garage in car capacity (garage_car)
- Total basement area in sq feet (total_bsmt_sf)
- First floor area in sq feet (1st_flr_sf)
- Height of basement (bsmt_qual)
- Original construction date (year_built)
- Interior finish of garage (garage_finish)
- Remodel date (year_remod/add)
- Number of full bathrooms above ground (full_bath)
- Masonry veneer area in square feet (mas_vnr_area)
- Total rooms above ground (totrms_abvgrd)



Findings from Baseline Model



Basic statistics of baseline model:

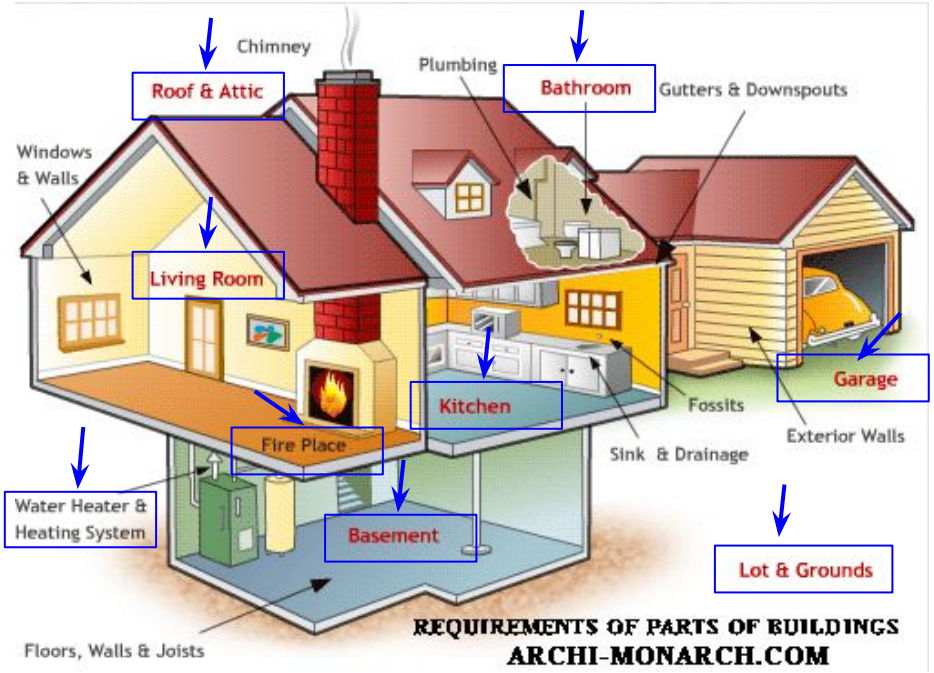
- Ridge train score of 0.8702909684989619
- Ridge test score of 0.8853340503453191
- Cross val score of 0.8609457652058599
- RMSE of 29727 for training set
- RMSE of 27684 for test set

Descriptive Features for Housing

Physical Characteristics
- Type of dwelling, zoning
- Land lot (frontage, area, shape, flatness, configuration, slope)
- Functional
- Neighbourhood and amenities proximity
- Overall house condition and quality
- Construction & remodel date

Material and Foundation
- Housing structure, foundation type
- Roof style, materials, quality and condition
- Masonry veneer type and area

External Facilities
- Paved driveway
- Wood deck area
- Porch
- Pool
- Fence
- Miscellaneous features



Source: https://archi-monarch.com/requirements-of-parts-of-buildings/

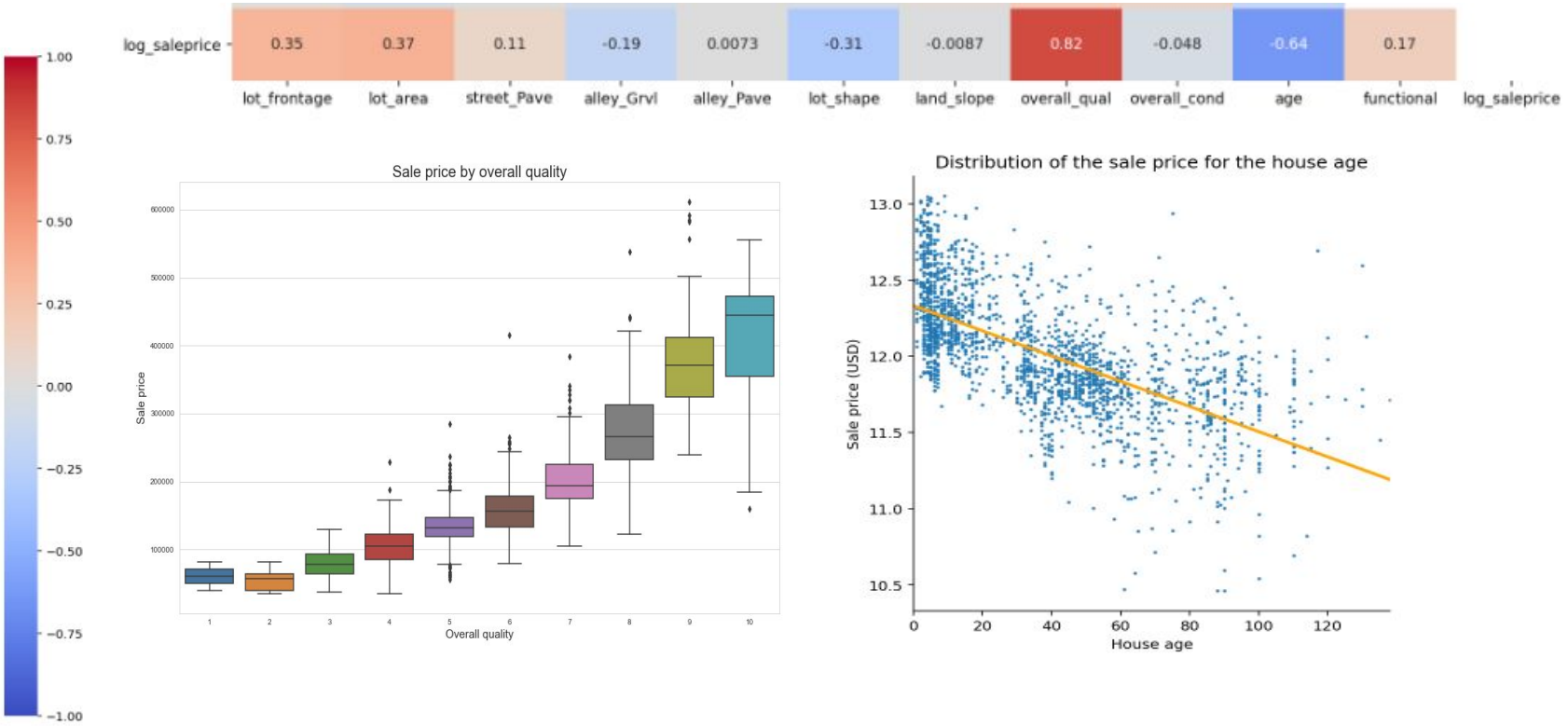
Basement
Quality, Condition, Exposure, Finished/unfinished/total area

Rooms
- First/second/Low quality of finished floors, above grade (ground) living area)
- Full/half bathrooms
- Bedrooms
- Kitchen
- Total rooms

Garage
Location, garage year built, interior finish, total area, car capacity area, quality, condition

Electrical and Heating
- Fireplaces
- Heating
- Central air-conditioning
- Electrical system

Physical Characteristics

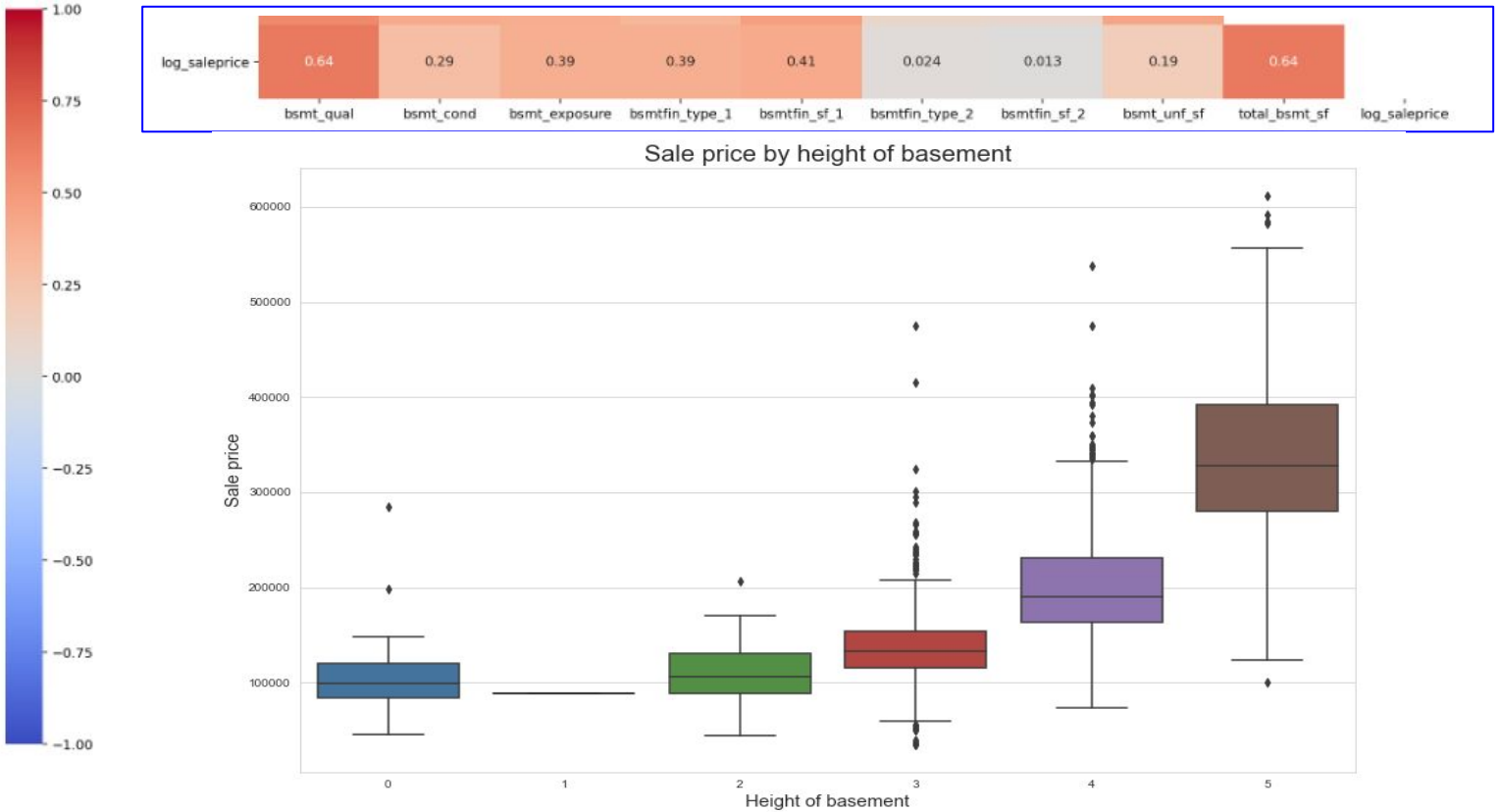


Garage & Material and Foundation

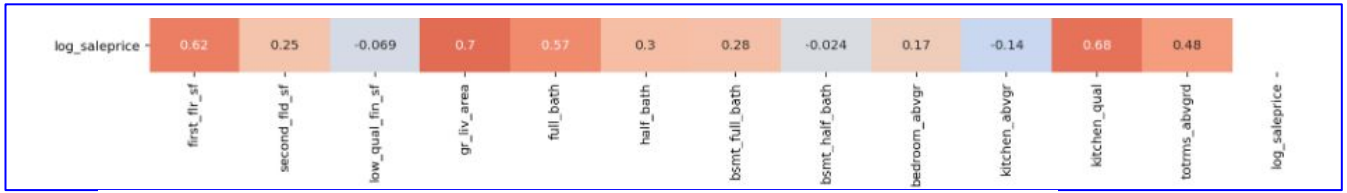
Garage

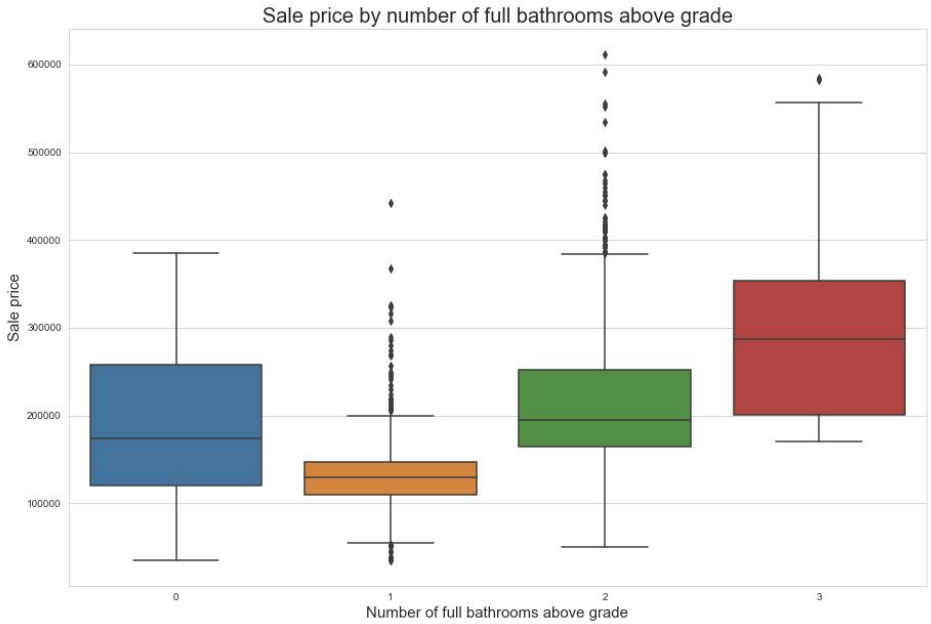
Material and Foundation

Basement

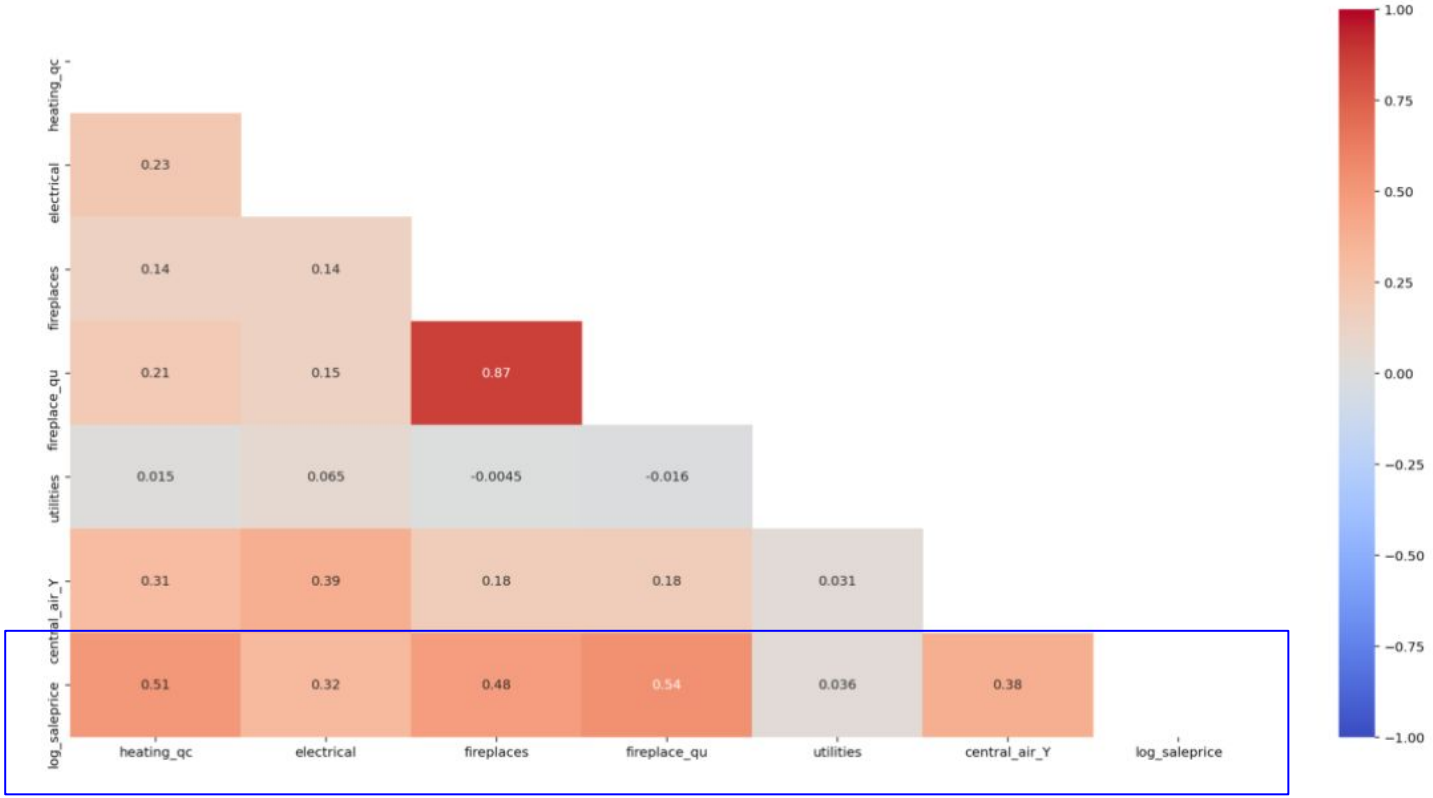

log_saleprice
0.64
0.29
0.39
0.39
0.41
0.024
0.013
0.19
0.64
bsmt_qual
bsmt_cond
bsmt_exposure
bsmtfin_type_1
bsmtfin_sf_1
bsmtfin_type_2
bsmtfin_sf_2
bsmt_unf_sf
total_bsmt_sf
log_saleprice
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Sale price by height of basement
Sale price
600000
500000
400000
300000
200000
100000
0
1
2
3
4
5
Height of basement

House Layout (Rooms and Area)

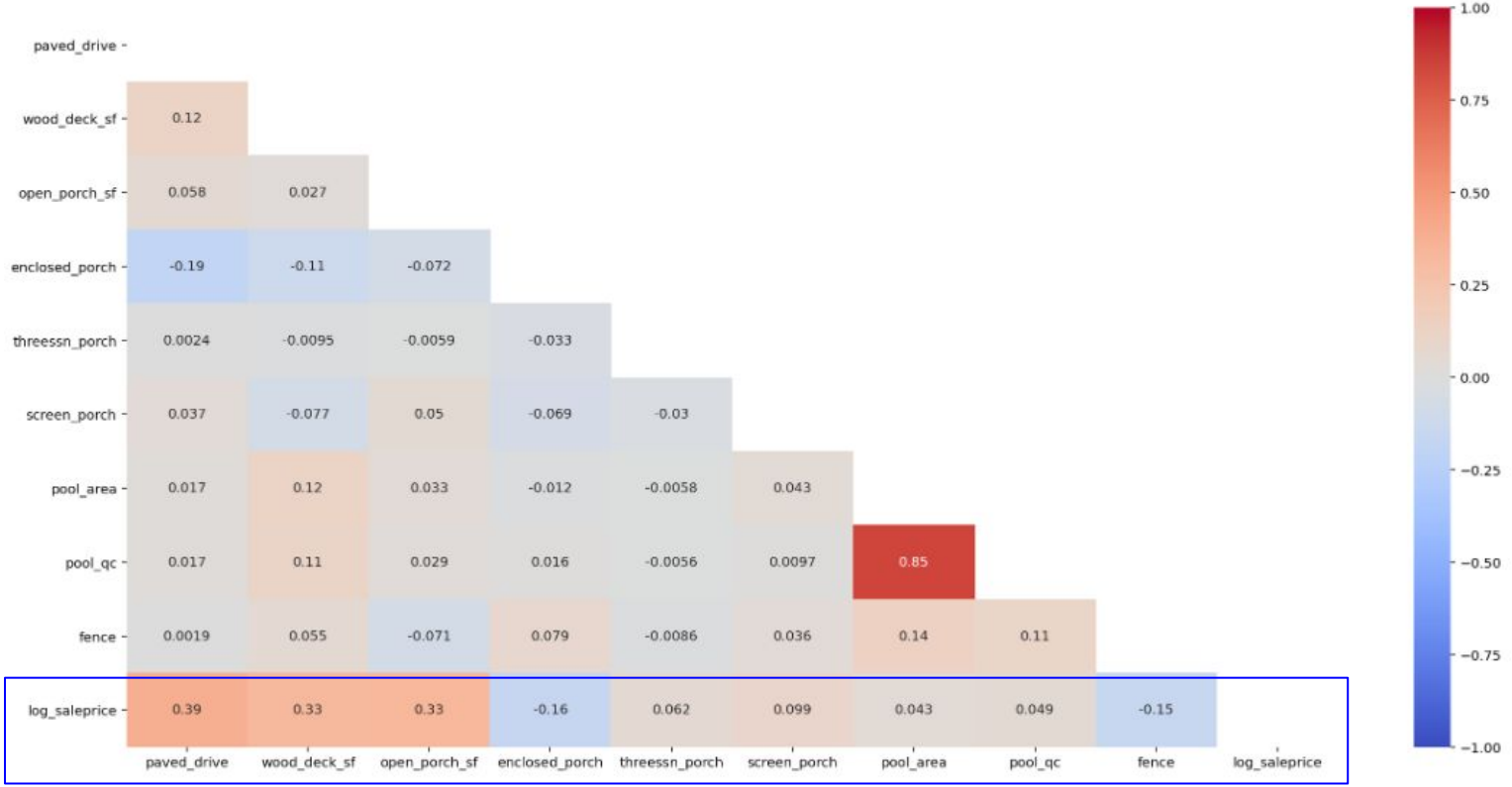




Electrical and Heating



External Facilities



Ridge Regression Modelling

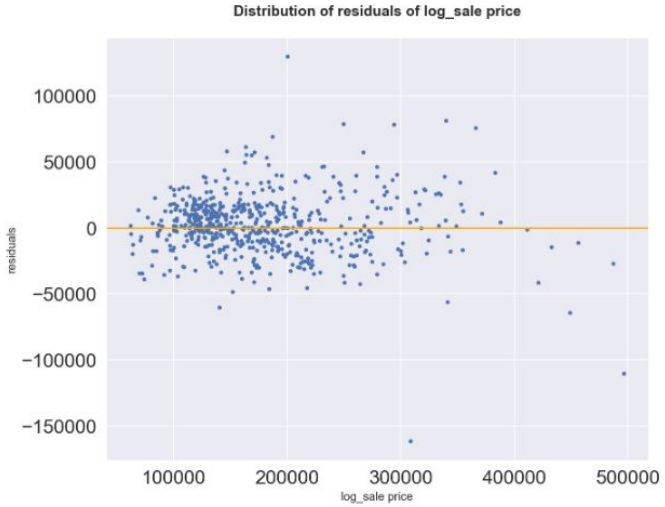
Linear/Lasso Model
The linear regression cross_val_score is 0.8670

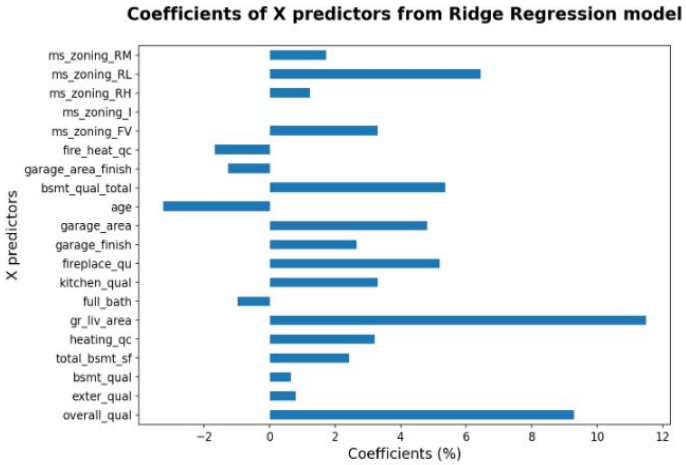
The lasso regression cross_val_score is 0.8664

Ridge Model
The ridge regression cross_val_score is 0.8678

The r2 score of our train set using ridge 0.8746

The r2 score of our test score using ridge 0.8837





Conclusion

- Top features that is correlated to the sales place
- Ridge regression modelling has better performance
- Overall quality has an increasing positive correlation trend with sale price.

Recommendations

- Construction company to consider factors that most affect sale prices when planning for new housing projects so as to maximise profit while serving local population demand
- To explore more detailed analysis such as buyer's behaviour or neighbourhood studies.

Recommendations: Neighbourhoods in Area of Ames

- Northridge Heights
  - Individual styles
  - Northridge Heights Apartme...
  - Downtown Ames
- Iowa State University
  - Individual styles
  - Iowa State University
- Stone Brook Neighborhood
  - Individual styles
  - Stonebrook
  - Stonebrook
  - Stonebrook
  - Stonebrook Crossing