

A dark, grayscale photograph of the New York Stock Exchange building facade, featuring several large, ornate columns and the words "NEW YORK STOCK EXCHANGE" inscribed on the top. The image is used as a background for the slide.

Regression model to predict house prices

GA DSI 22 Project 1: Tan Han Wei (John)

Overview

Part 1

Method

Part 2

Feature
Selection

Part 3

Model choice
and score

A low-angle, black and white photograph of several modern skyscrapers reaching towards a dark sky. The perspective creates a sense of height and scale. The buildings have a grid-like facade of windows.

01

Method

Methodology

- **Feature selection and feature engineering**
- **Create baseline model, evaluate the score**
- **Use regularisation technique, evaluate score**
- **Repeat to improve the model**

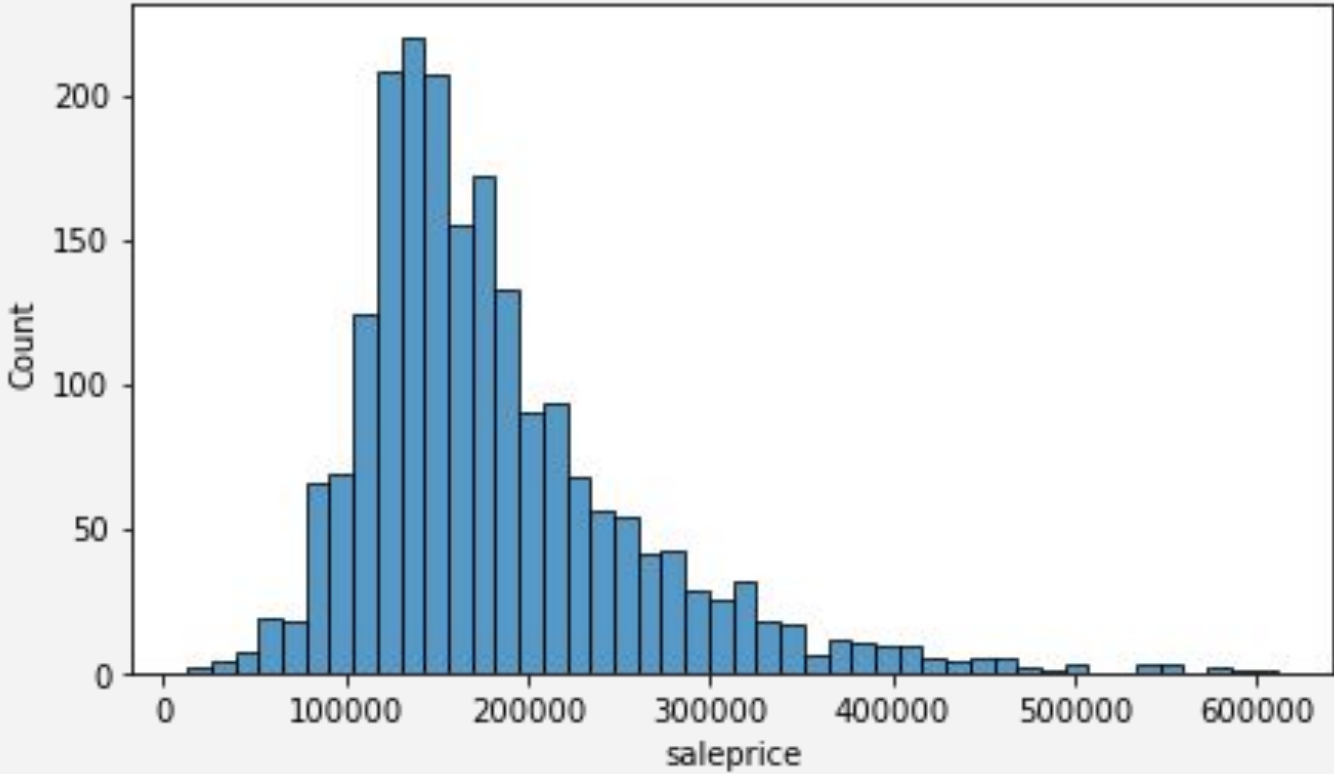
A low-angle, black and white photograph of several skyscrapers reaching towards the sky. The perspective creates a sense of height and scale. The buildings are dark, with some windows reflecting light. The sky is a uniform dark grey.

02

Feature Selection

Dependent variable - Sale Price

Distribution of SalePrice

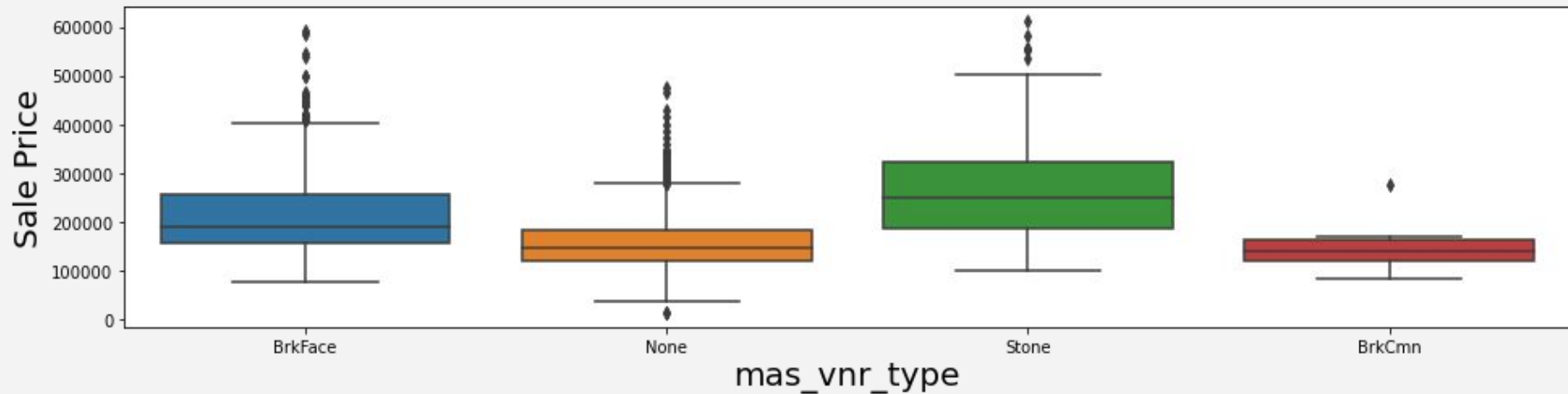
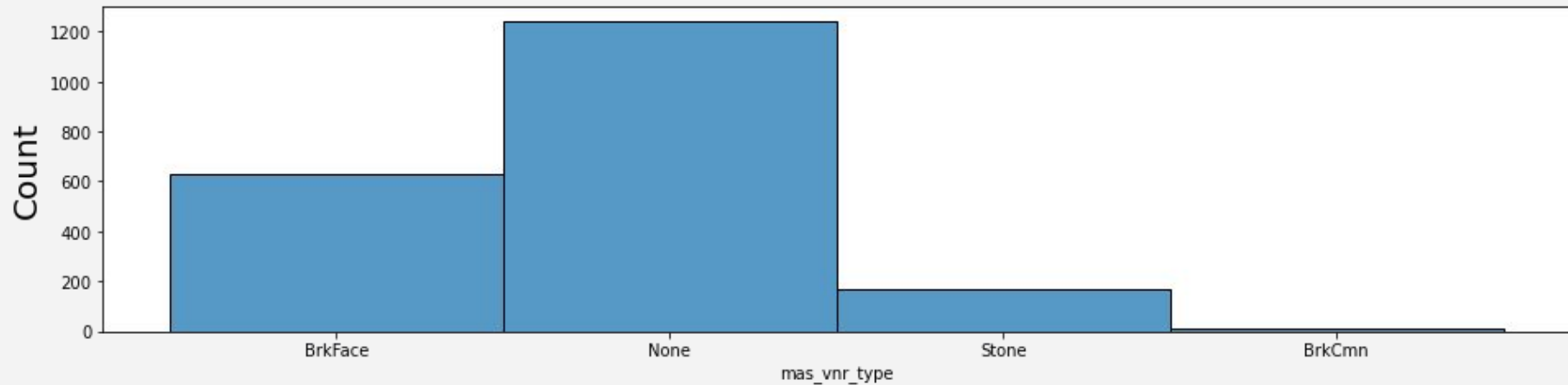


Feature selection

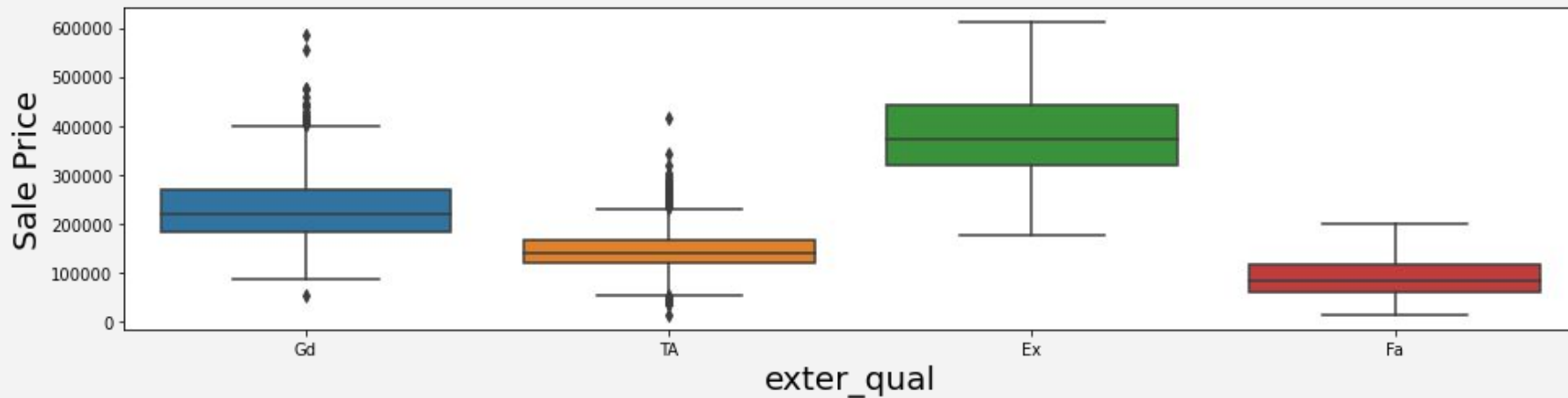
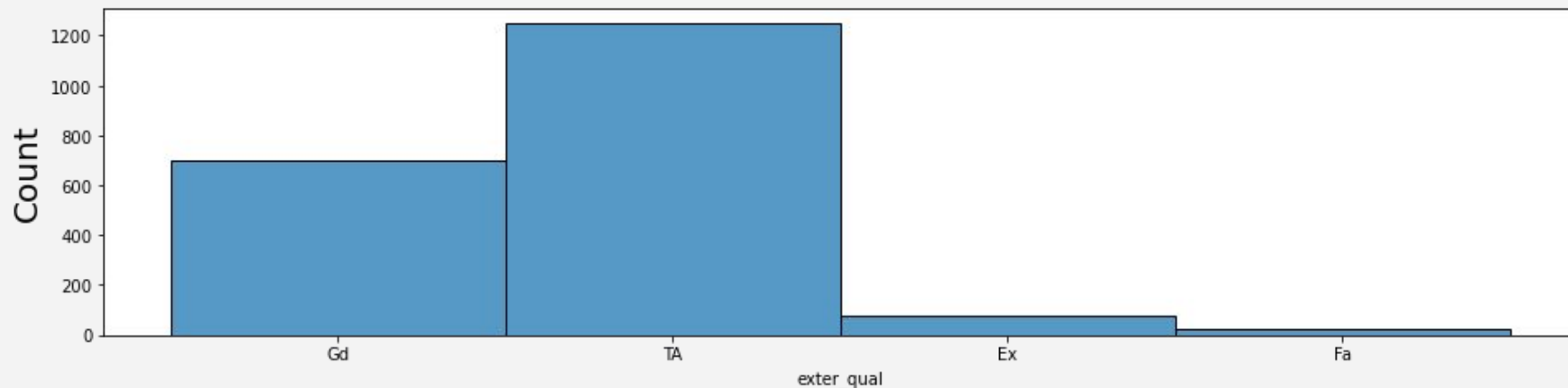
As a rule of thumb, when selecting variables for the model, there are a few factors to take note of:

- 1) Significant variation in the independent variable.**
- 2) For numerical variables, significant correlation in the dependent variable with the independent variable**
- 3) For categorical variables, unique characteristic between the categories**

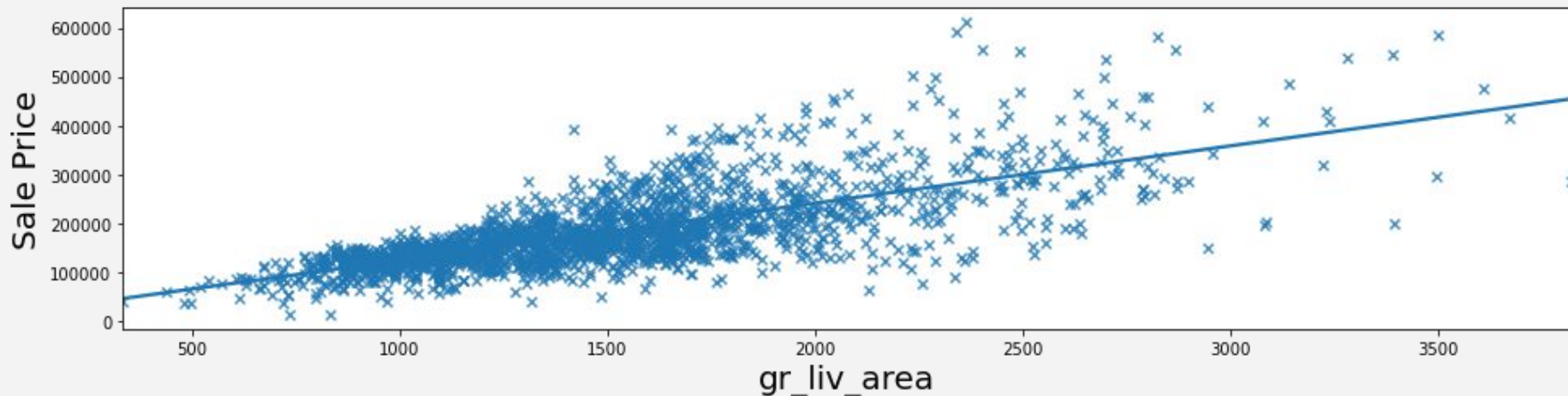
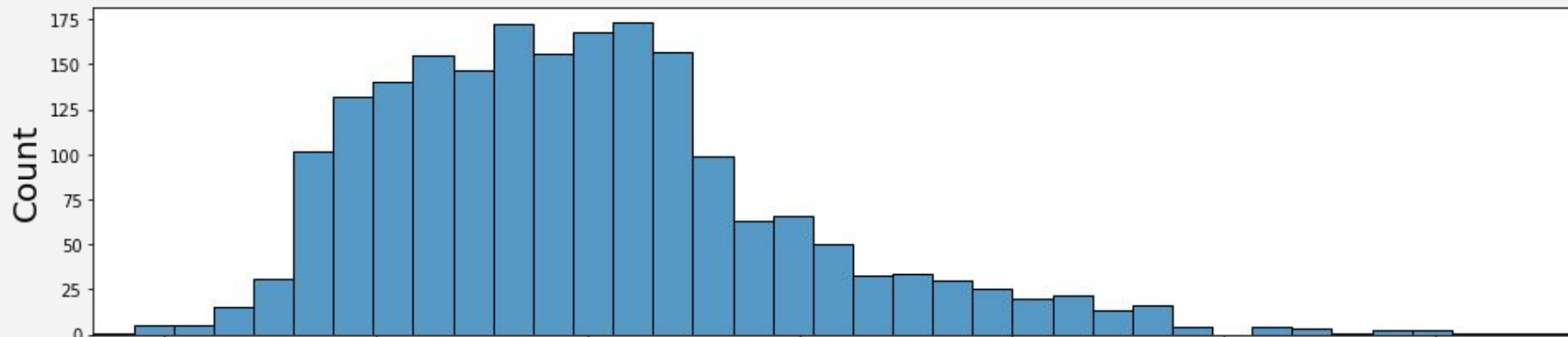
mas_vnr_type histogram & barplot with Sales price



exter_qual histogram & barplot with Sales price



gr_liv_area histogram & scatterplot with Sales price $p = 0.719$



Selected features

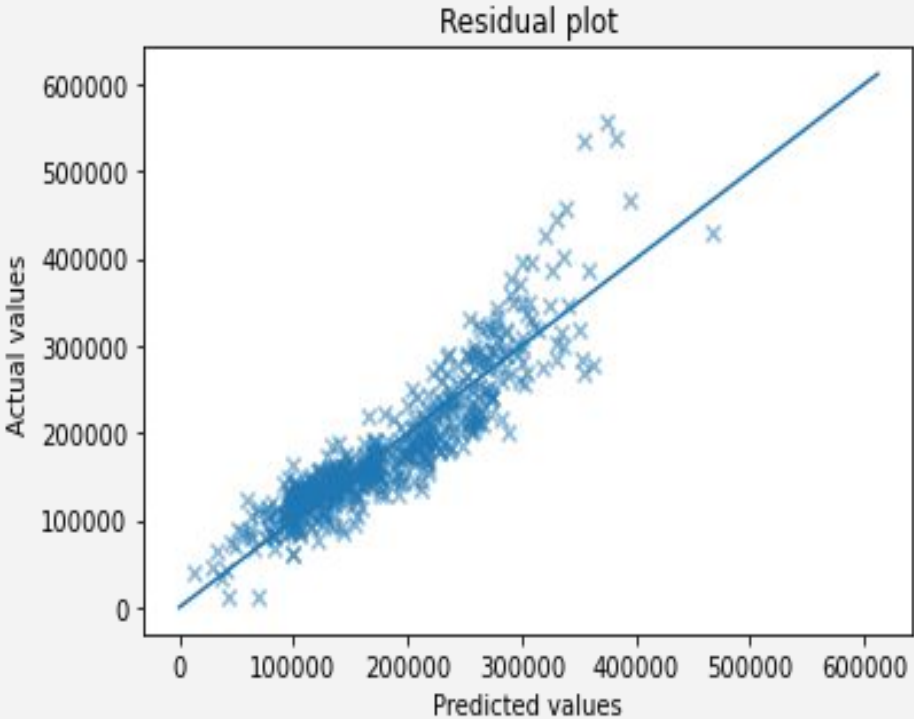
- 1) 'irreg_shape'
- 2) 'inside_lot'
- 3) 'one_floor'
- 4) 'good_qual'
- 5) 'property_age'
- 6) 'have_mas'
- 7) 'good_ext'
- 8) 'total_bsmt_sf'
- 9) '1st_flr_sf'
- 10) '2nd_flr_sf'
- 11) 'gr_liv_area'
- 12) 'total_bath'
- 13) 'good_kitchen'
- 14) 'totrms_abvgrd'

A low-angle, black and white photograph of several modern skyscrapers reaching towards a dark sky. The perspective creates a sense of height and architectural scale. The buildings are composed of many windows, creating a textured pattern of light and dark rectangles.

03

Model Choice

Basic OLS with 14 features



Feature	Coefficients
irreg_shape	6058.88
inside_lot	-656.18
one_floor	-5742.70
good_qual	19247.48
property_age	-325.44
have_mas	8380.92
good_ext	13036.76
total_bsmt_sf	42.64
1st_flr_sf	48.95
2nd_flr_sf	26.81
gr_liv_area	38.84
total_bath	5700.33
good_kitchen	13794.66
totrms_abvgrd	-2541.83

Basic OLS with 14 features

OLS - RMSE 36642

Ridge - RMSE 36505

Lasso - RMSE 36423

For this basic model regularisation did not improve the model, as it does not suffer from overfitting

Overfitted model - PolynomialTransform on 14 variable

Now the regression has 120 variables

OLS - RMSE 77917.89

Ridge - RMSE 31632.67

Lasso - RMSE 31497.68

A grayscale photograph of the New York Stock Exchange building facade. The image shows the ornate classical architecture, including a series of large sculptures (telamones) along the top of the facade. The words "NEW YORK" are visible in large, raised letters on the building's front. A white rectangular box is superimposed over the center of the image, containing the word "Thanks" in a white serif font.

Thanks