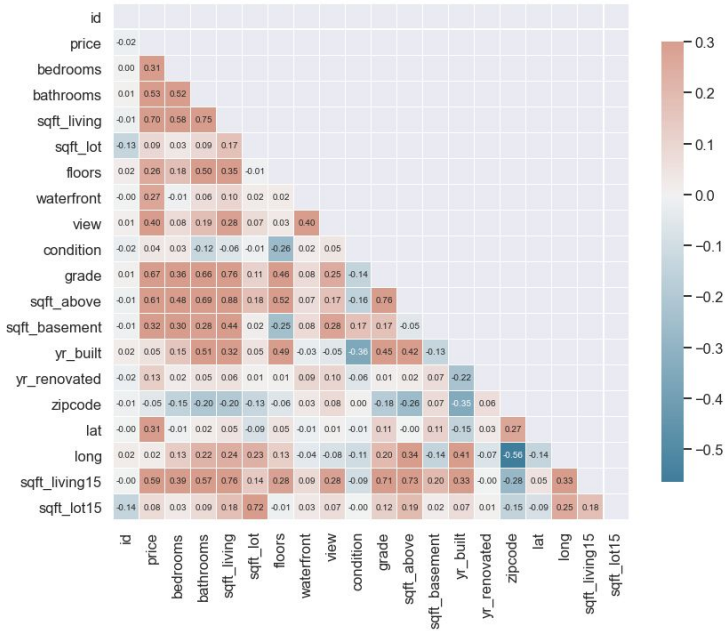
Real Estate Prediction

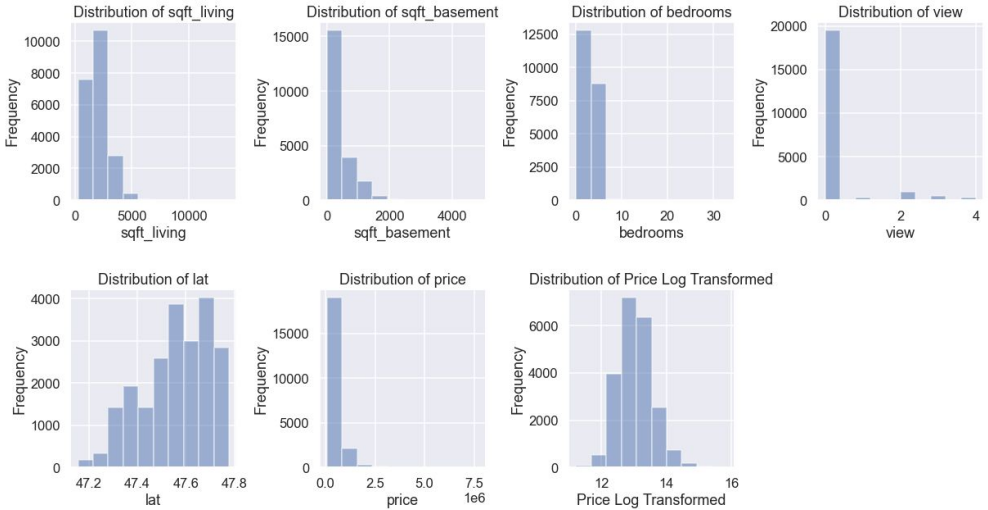
By Kasen Teoh, Chung En Pan, and Jieru Bai

Research Aim and Methodology

- **Aim:** To analyze housing prices to identify which variables influence housing prices the most and to use these variables to predict housing prices
- **Hypothesis:** living space (sqft), bedrooms, and the view would be the most influential and multiple linear regression would be the best model in terms of accuracy and interpretation
- **Models**
  - Multiple linear regression
  - Lasso Regression
  - Neural Network

Identifying Predictors



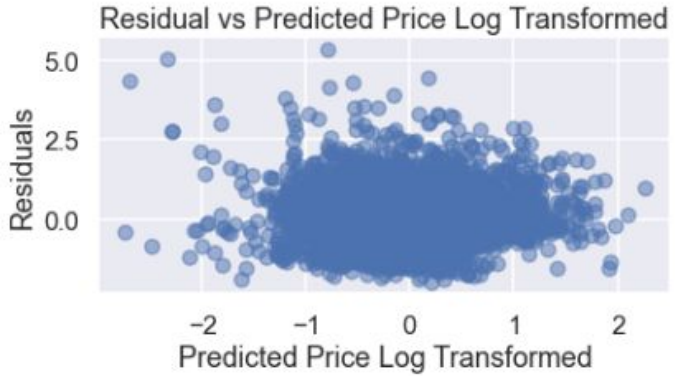


Model	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration	5th Iteration	Mean RMSE
First Model	0.703(0.511)	0.7(0.517)	0.689(0.522)	0.701(0.502)	0.7(0.507)	0.699
Second Model	0.565(0.684)	0.566(0.684)	0.559(0.685)	0.561(0.681)	0.559(0.686)	0.562

Table 1: K-Fold Cross-Validation RMSE and R-squared (R-squared in parentheses)

Multiple Linear Regression Prediction

Adjusted R-Squared: 0.686 RMSE: 0.558



Lasso Regression and Neural Network Prediction

Model	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration	5th Iteration	Mean RMSE
First Model	0.369	0.367	0.362	0.367	0.361	0.3652
Second Model	0.431	0.455	0.45	0.462	0.445	0.4486
Third Model	0.984	0.993	1.013	0.993	1.016	0.9998

Table 3: K-Fold Cross-Validation RMSE for Neural Net and Lasso Reg

Adjusted R-Squared: 0.869 RMSE: 0.362

Results and Discussion

- First part of our hypothesis was half correct, i.e living space and view were among the top 3 most significant; however, bedrooms was not.
- Second part of our hypothesis was incorrect with neural nets outperforming linear regression

  Neural Net RMSE: 0.362 Linear Regression RMSE: 0.558

- Limitations:
  - Timeline of our data was only from 2014-2015
  - Conclusions only valid to Seattle
  - The more data the better for deep learning (neural networks)