Ames Housing Sale price

Breakout Room 2:

Anastasiya, Bryan, Jamie,
Qiwen, Russell, Sheng Jun

Guiding questions

Main

What is the price of the new property?

Additional

- Does an increase of Overall Quality of property affect the sale price?
- Does the price vary with different House Style?
- Does older property cost less than newer ones?

Problem Statement

Who ?

House owners, Landlords, Real estate agents & Clients

Problem ?

Short and long-term financial planning due to the volatility of the housing market

Solution !

Regression models (Linear, Lasso and Ridge)

Data Cleaning & Feature Engineering

Source:
Ames, Iowa, USA
2006 - 2010

Null Value Imputation

- **Mean** for Continuous Variables (Lot Frontage)
- **Existence** for Categorical (Garage and Basement)
- **Removed columns** more than 50% Null (Pool QC)

Convert Data Types

- Categorical Data **into integer** (BsmtExposure:
- Good Exposure 4
- Average Exposure 3
- Minimum Exposure 2
- No Exposure 1
- No Basement 0)

New Columns

- **Age** (Year Sold - Year Remodelled)
- **Square Feet** (Garage + Basement + 1st/2nd Floor SF)

Exploratory Data Analysis
and
Feature selection

Selecting Numerical Variables

Top correlation to Sale Price

Blue - Quality

- Overall Quality **80%** Corr

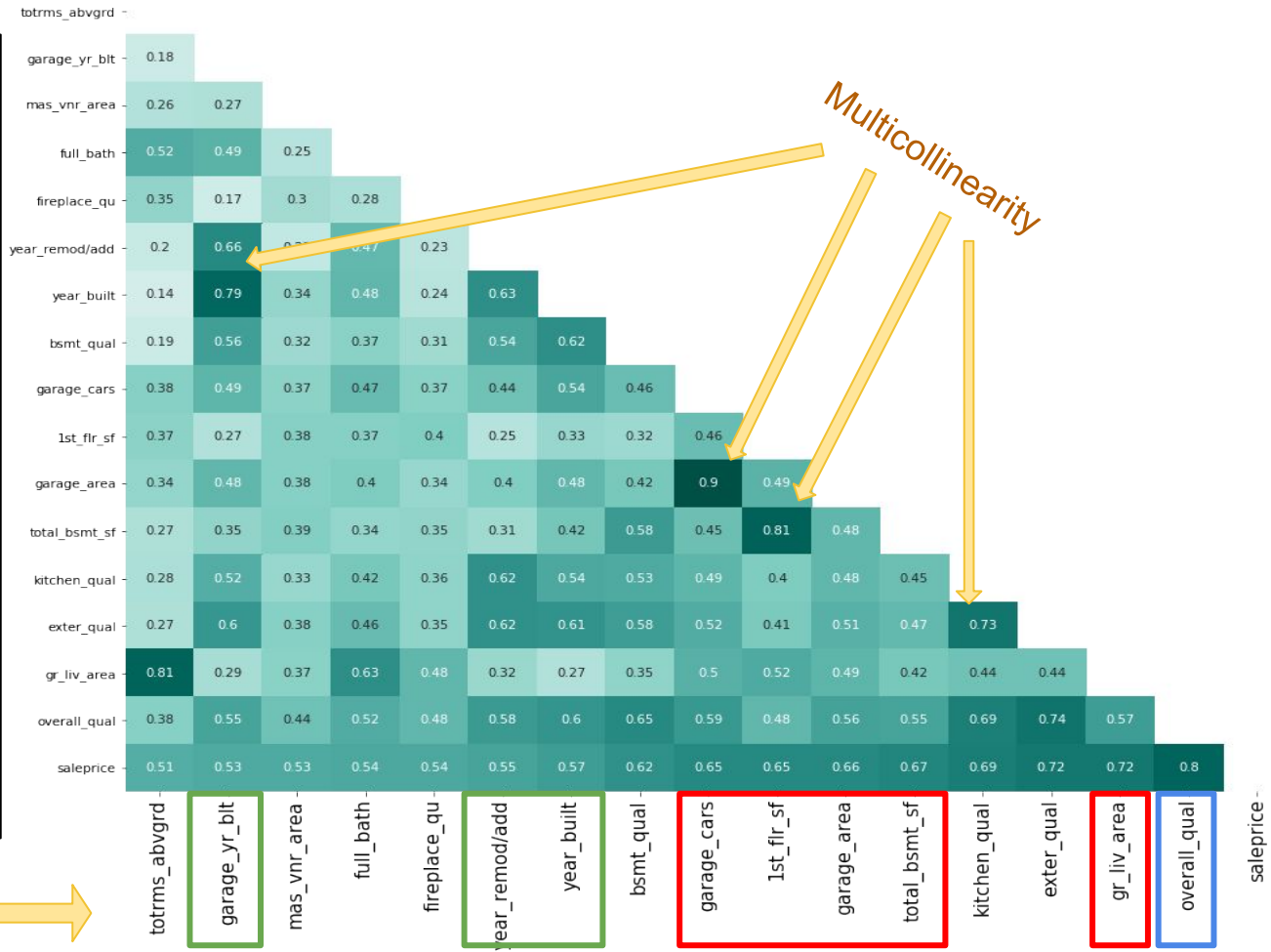
Red - Area

- Garage, Basement, Floor space **above 60%** Corr

Green - Age

- Year Built/Remodelled **above 50%** Corr

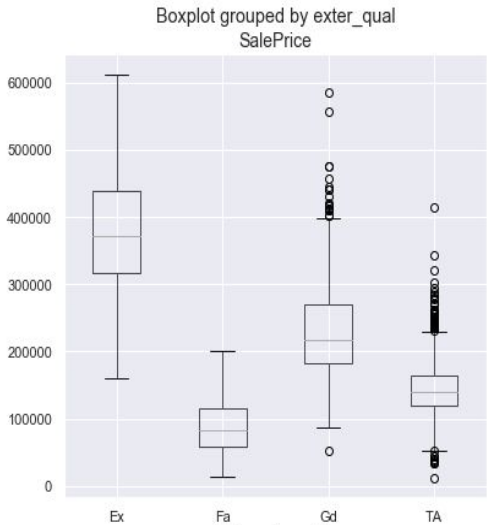
Intercorrelation matrix of features which are represented by numbers with 0.5 and higher correlation coefficient with sale price



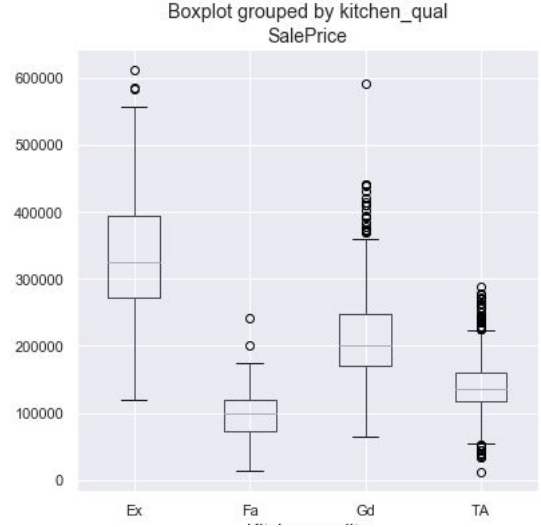
High corr

Low corr

Selecting Categorical Variables



External material quality



Kitchen quality

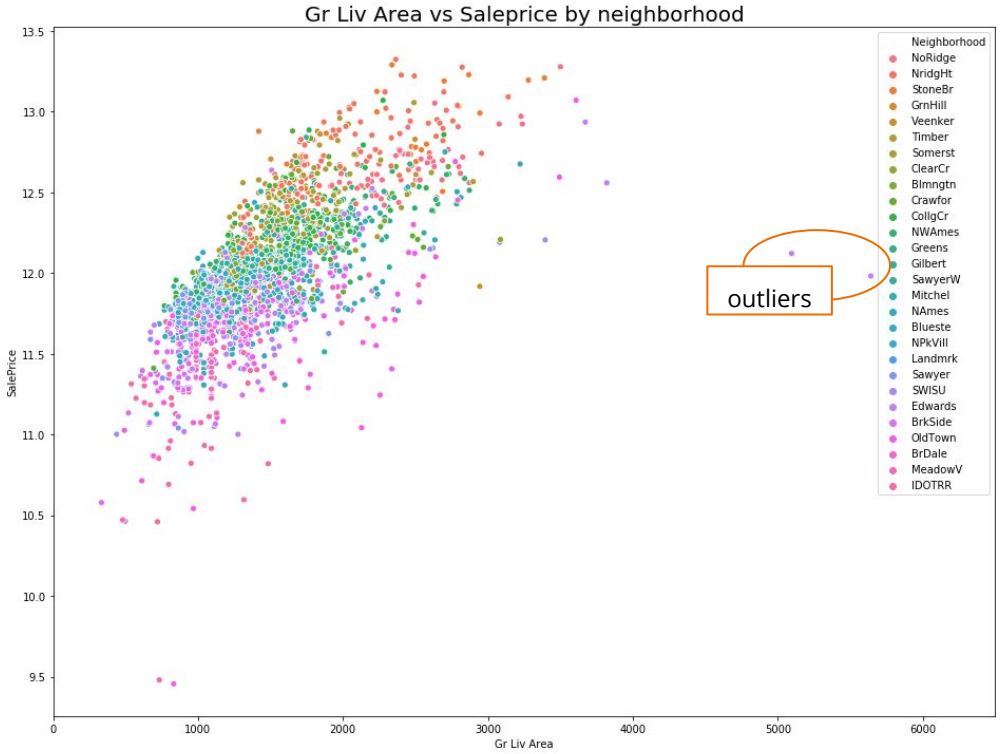
Feature engineering candidates :

- **Meaningful distribution** by classes to sale price

EDA: Linearity of Variables to Sale price

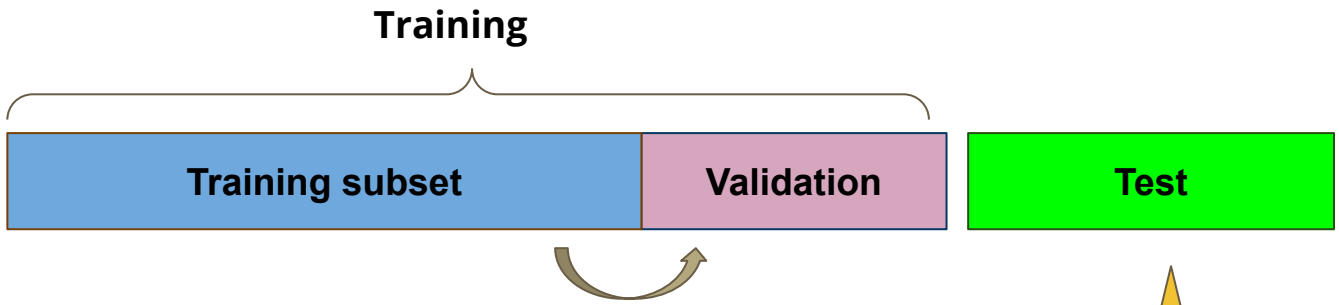
Gr Liv Area vs sale price by neighborhood

- Homoscedastic
- Positive Correlation with Sale Price
- Removed two outliers to prevent overfitting



Model Fitting and Evaluation

Modeling Overview



1. Split into subsets, Cross-validate

2. Evaluate models
   - **Baseline**
   - LinearRegression
   - **Ridge**
   - **Lasso**

Evaluation metrics (against Baseline):
- Root Mean Square Error (RMSE)
- Coef. of Determination (R2)

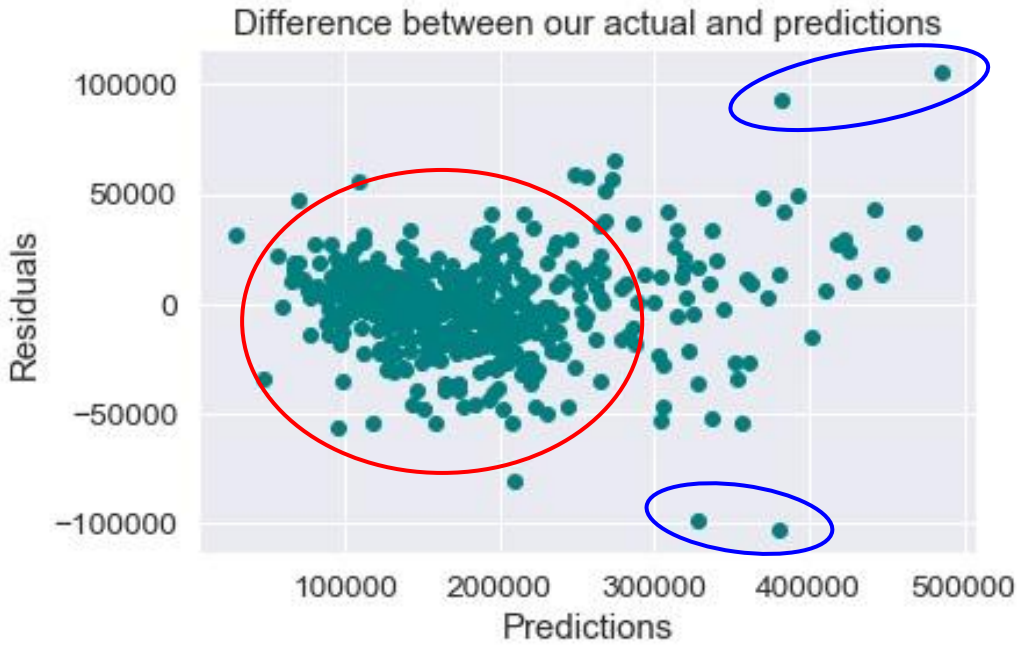
3. Fit features learnt from entire train data to Test set

Model Evaluation

Model Scores	Root Mean Squared Error (lower is better)	R^2 (0 to 1)
Baseline	79,309	0.0000
LR Model	**25,505**	**0.8966**
Lasso	25,556	0.8967
Ridge	25,597	0.8958

Predictions vs. Actuals

- **Low residuals** around highest distributed price range (150K - 200K)
- **High residuals** for outlier prices (**one-off** categorical factors affecting price greatly)



Guiding questions

Main:

What is the price of the new property?

Additional:

- With an increase of Overall Quality of property does price increase?
- Does the price vary with different House Style?
- Does older property cost less than newer ones?

Problem Statement

Who ?

House owners, Landlords, Real estate agents & Clients

Problem ?

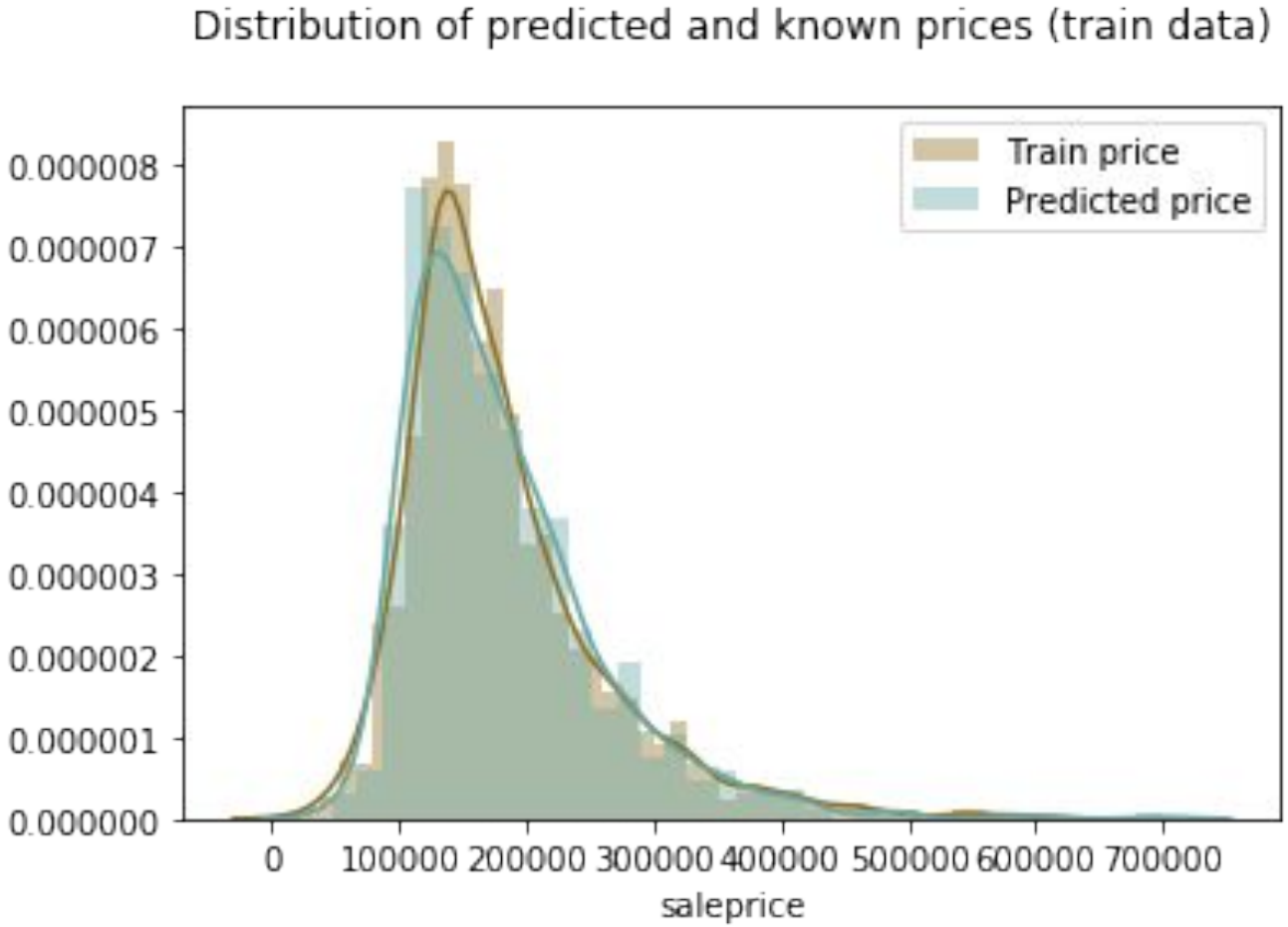
Short and long-term financial planning due to the volatility of the housing market

Solution !

Regression models (Linear, Lasso and Ridge)

Predictions

The distributions of predicted price and known sale price are overlapped. It means that the **range, mean and median are similar**.



Conclusion and Recommendations

Guiding questions

Main:

What is the price of the new property?

Additional:

- With an increase of Overall Quality of property does price increase?
- Does the price vary with different House Style?
- Does older property cost less than newer ones?

Problem Statement

Who ?

House owners, Landlords, Real estate agents & Clients

Problem ?

Short and long-term financial planning due to the volatility of the housing market

Solution !

Regression models (Linear, Lasso and Ridge)

Interpreting Results

- With an increase of Overall Quality of property does price increase?

1. ‘**Overall Quality**’ *(overall_qual corr.: 0.804683)*
    - Overall quality one of the most influencing features. With an increase in 1 grade of quality the price increase on $ 13,986.

Does the price vary with different House Style?

2. '**One storey**'
    - One storey house styles cost more than other styles.

- Does older property cost less than newer ones?

3. **Older property cost less** *(year_built corr.. : 0.573751)*
    - With an increase in one year the property rise in price in $ 5,069.

Stakeholder Recommendations

To House Owners, Landlords, Real Estate Agents:

Features that greatly increase the price of your property:

1. **Basement** and the **overall area** of the property.
2. **Overall material** and **finish quality**.
3. A number of **fireplaces**.
4. '**One storey**' house style.

Stakeholder Recommendations

To House Buyers:

Features to consider when buying a property:

1. Houses which were **built earlier cost less**.
2. **Kitchen quality influence on sale price**. The cost of renovation of the kitchen should be considered. With an increment in kitchen quality by 1, the price of the house rises by $ 6,183.
3. **Less emphasis on the exterior material quality** as an increase in quality results in an increase of only $ 3,879.

Further research suggestions

The current model should be explored on other datasets, as it is possible that the model will have different results based on location.

Based on current model limitations, following research proposed:

- To evaluate the **generalizability** of the model **more data required**.
- Prices **higher than $ 500,000** should be **explored separately**. The model shows worse predictions for such a high price. More data should be collected for high price property separately.
- Features with a negative slope. Additional models should be built to evaluate the influence of the features on price.
- Use a **combination of more complicated models** for prediction.

Thank you

Panel open for Q&A