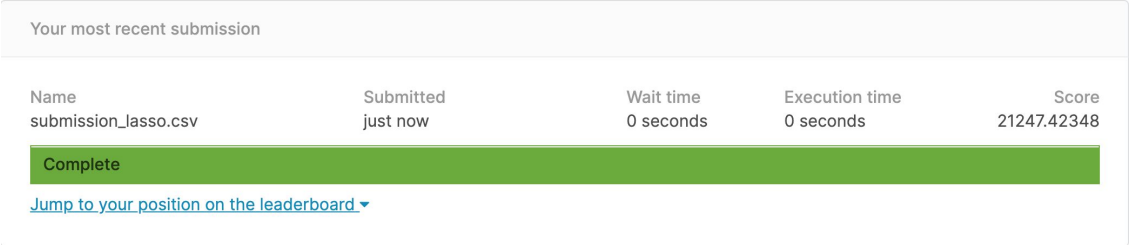
Project 2

AMES, IA HOUSING SALE PRICE PREDICTION

Group 3

Problem Statement

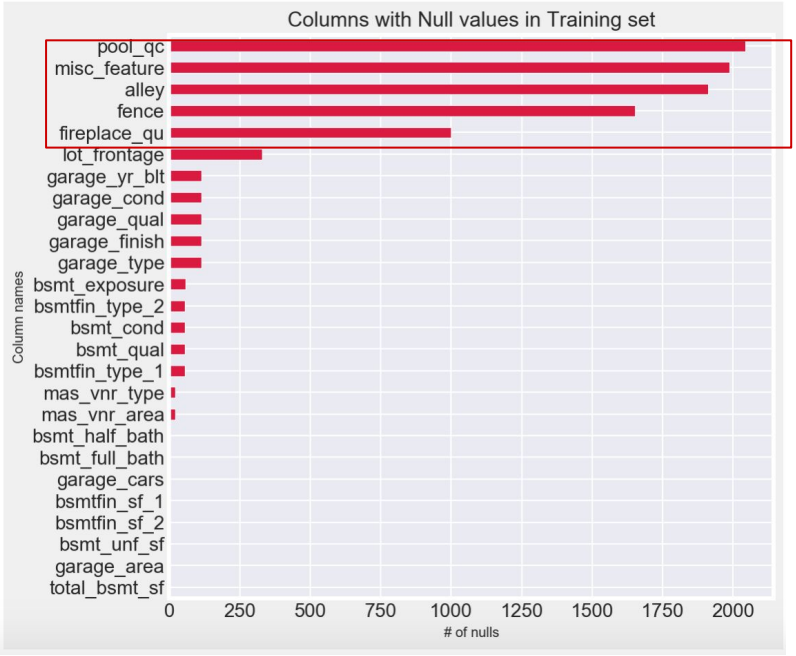
To create a regression model based on the Ames Housing Dataset to predict the most accurate sale price vs other competitors.



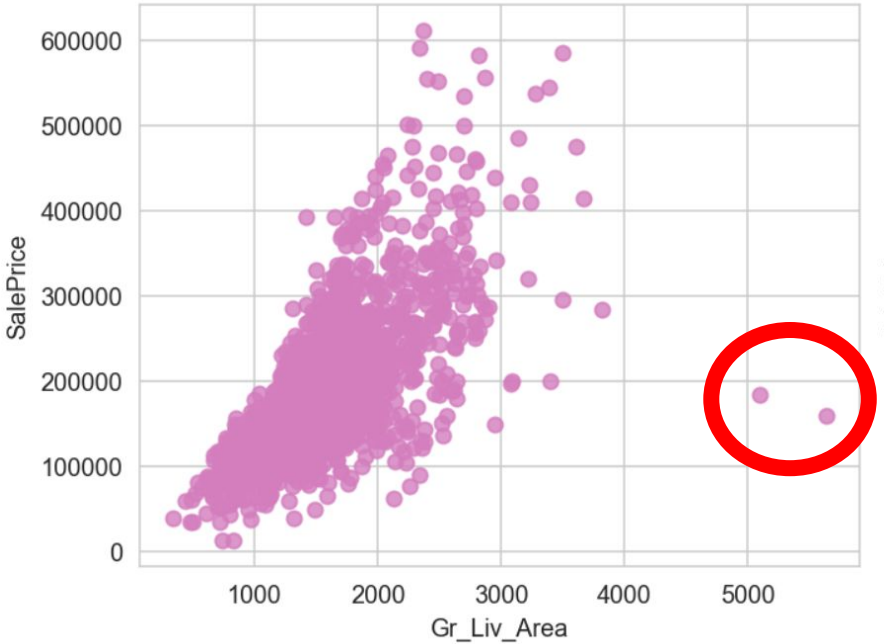
Measure of Success

To analyse the data and make recommendations on development/investment opportunities and how to increase the value of home.

Data Cleaning and EDA

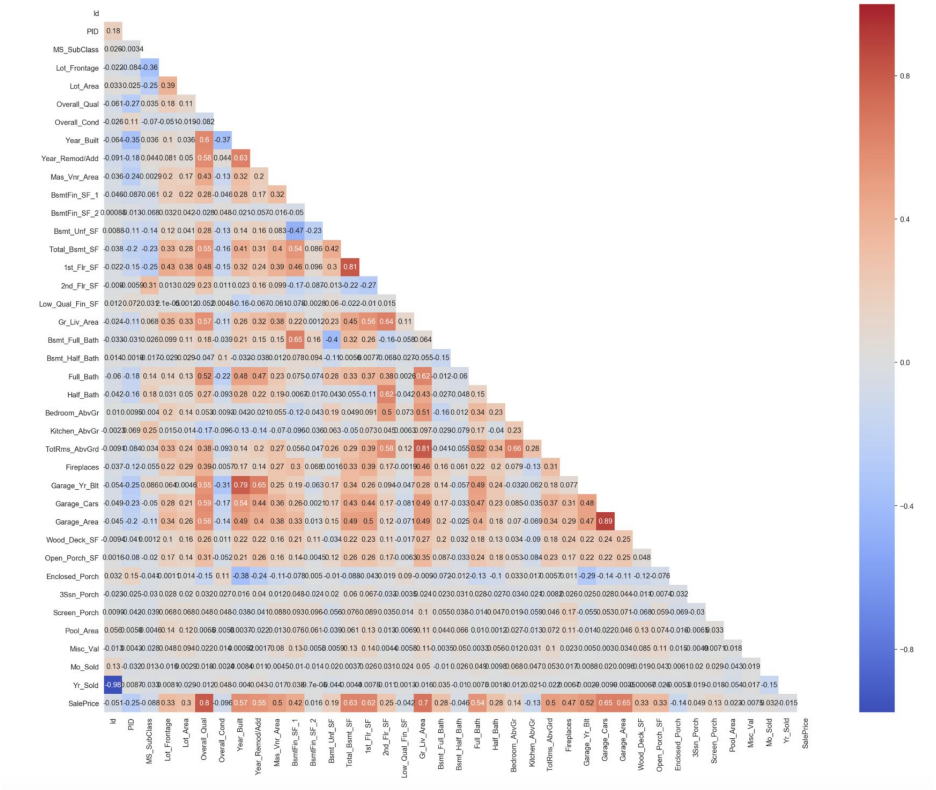


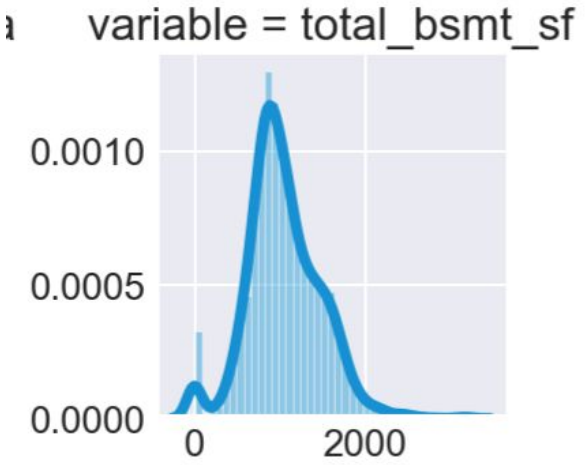
Missing Values

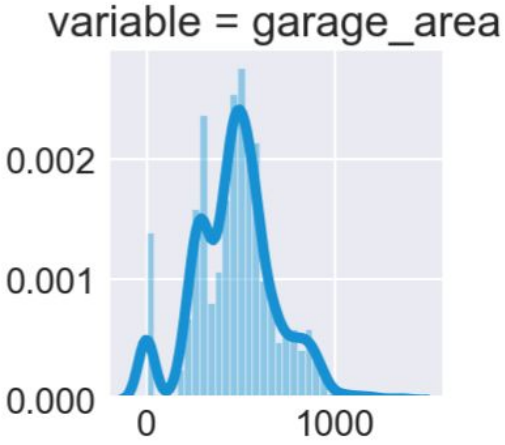


Outliers

Data Cleaning and EDA





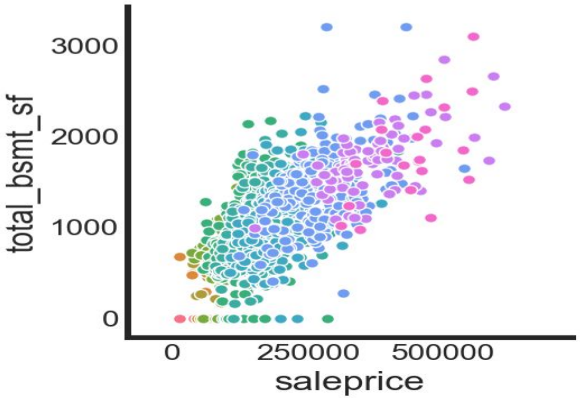


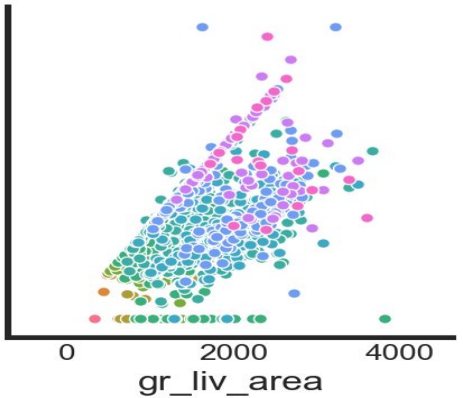
Preprocessing

Change ordinal data to a ranking system

Dummy creation - from 73 to 204 features

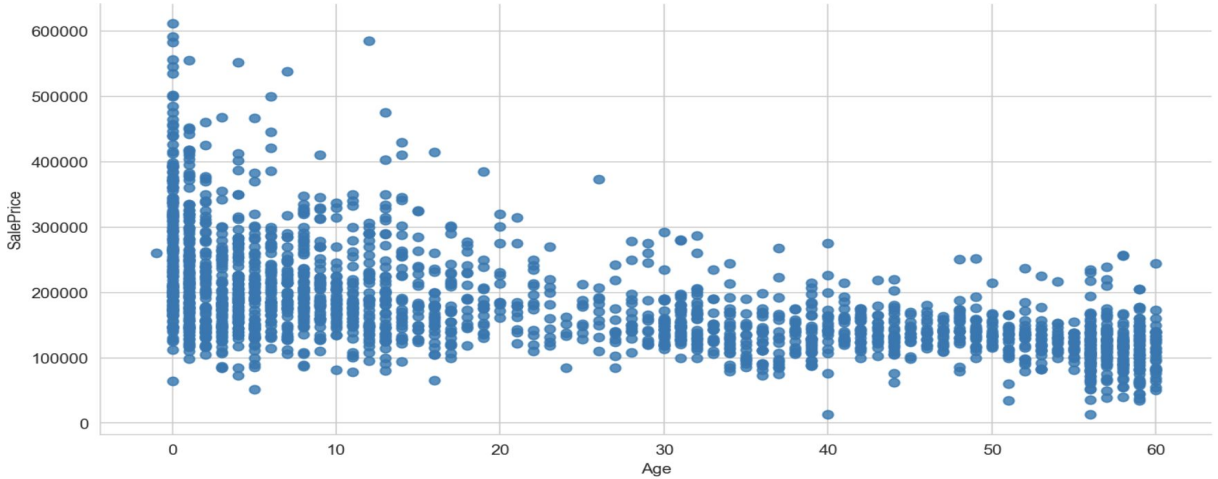
Multicollinearity between variables





Preprocessing

1. Feature engineering (Property age)
2. Sale price normalisation
3. Train-test split (train set: 0.9)
4. Scaling

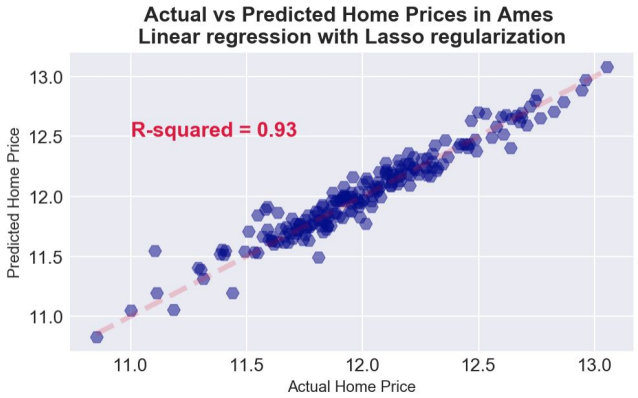


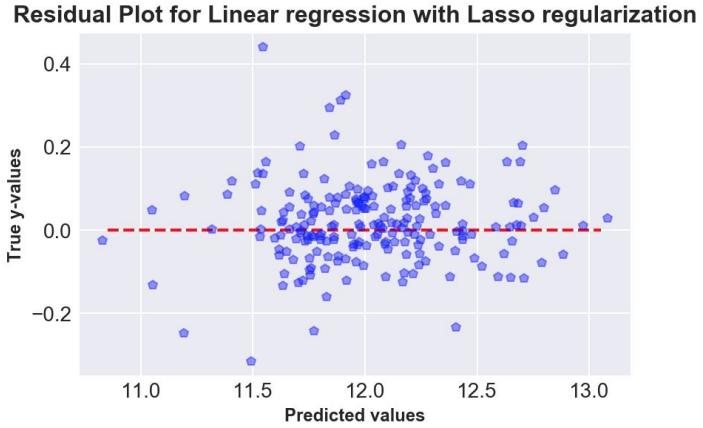
Modeling

	Optimal Value	Scores	
Model	**Hyperparameters**	**R-Squared**	**RMSE**
Vanilla Linear Regression (LR)	-	-845069952923141.5	11076079.2310
LR with Ridge Regularization	Alpha = 76.634	0.92689	0.10301
LR with Lasso Regularization	**Alpha = 0.00316**	**0.93237**	**0.09908**
LR with Elastic Net Regularization	Alpha = 0.00803 Lambda = 0.5	0.93183	0.099475

The selected model for predicting sale price in test dataset is LR with Lasso regularization

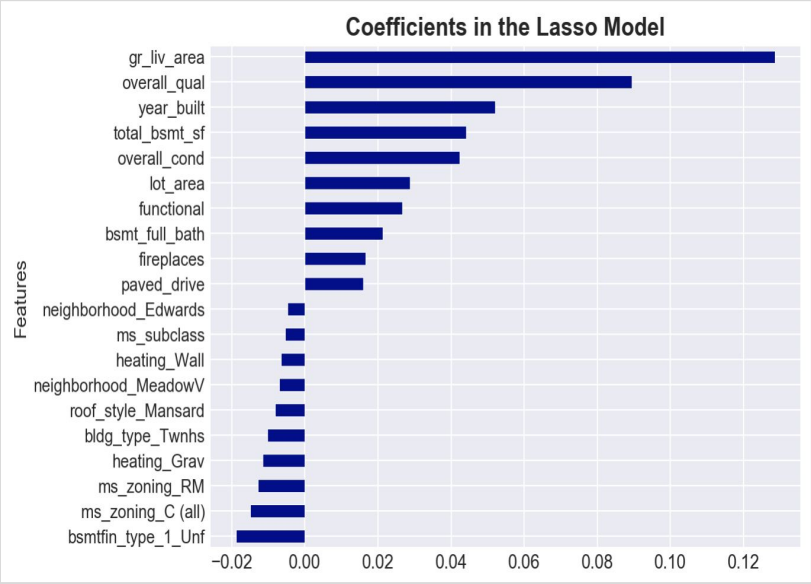
Evaluation





- ➔ Certainly strong model with **R-squared score of 0.93**
- ➔ There are **very few poor predictions** and overall no outliers to be highlighted
- ➔ **No evidence of overfitting** as the R2 score on training set was slightly lower at 0.91
- ➔ The **model generalized well** on unseen data
- ➔ Produced **one of the best RSME score on Kaggle** 21247.42348

Inferential Learning with Lasso



1. Prediction of sale price is not the only output from Lasso Regression model.
2. It helps with **inferential learning** as to the features that positively and negatively impacts sale price of a property.
3. This will be particularly **helpful in making recommendations** to home owners preparing to sell their home.

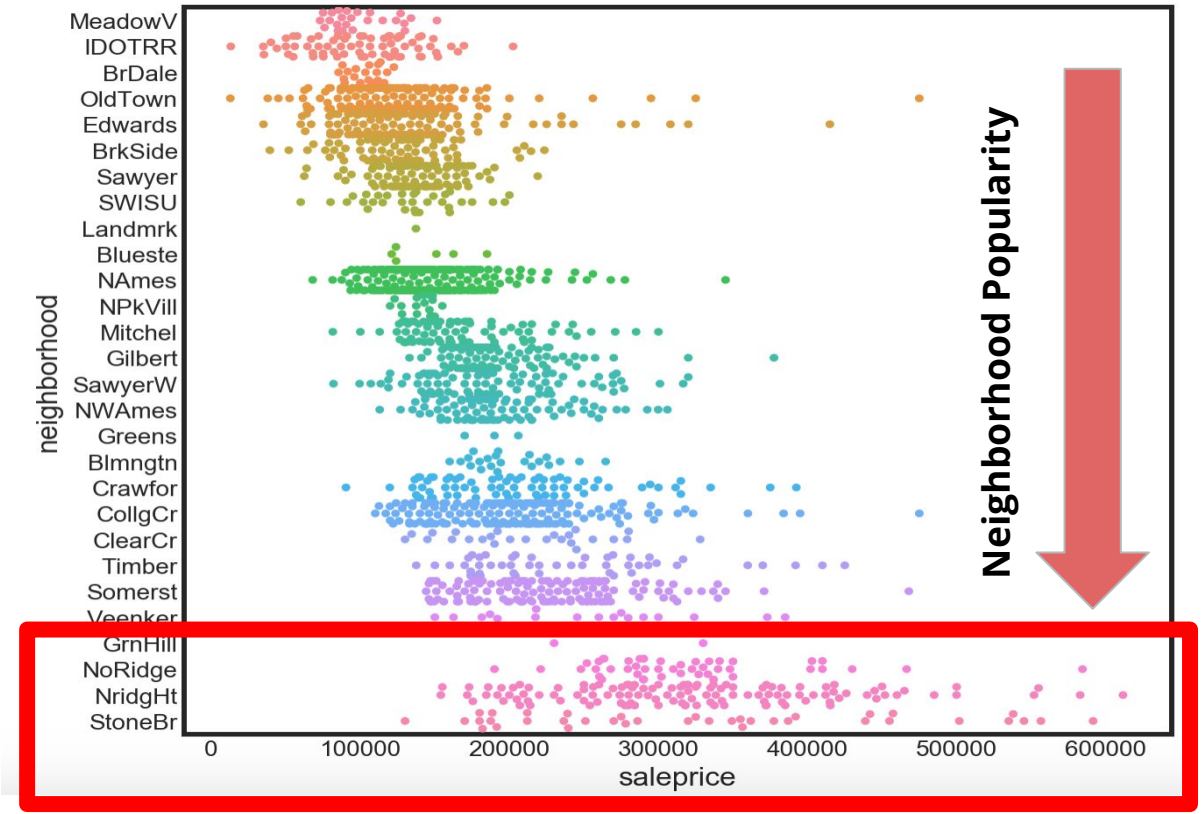
Conclusions and Recommendations (for home owners and real estate co)

- Addressed problem statement by developing a relatively accurate model to predict housing prices based on various features
- Obtained reliable and significant features/Lasso coefficients to guide our recommendations:

Top 3 house features that fetch <u>higher price</u>	**Top 3 house features that <u>hurt price</u>**
1) Bigger houses **2) Well renovated (or good quality) houses** (Priorities: i. Ext qual, ii. kitchen quality, iii. basement qual) 3) Newer houses	**1) Unfinished basement** 2) Houses located in commercial zone 3) Townhouse is not popular

Recommendations for investors and real estate co

- Strong selling point: Good property investment and higher commission
- Richer (with better amenities) neighbourhood



Future Recommendations to increase our user base

- Improve accuracy of our model by collecting more data

- Expand scope of our model to predict housing price for other cities:
  - Supply and demand
  - Interest rates
  - Economic growth
  - Demographics (e.g. profiling of homeowners)

Questions?