
Making Houses price prediction in Ames, Iowa.

By Andrei Muravev 8/9/2021

Introduction

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement, ceiling or the proximity to the railroad. But these factors could affect the price of homes dramatically. How much, how, where and why people buy certain house? To answer these questions, we will look at the Ames House Dataframe to establish trends and predict homes prices. Main goal - find features that affect price of houses the most and what prediction model to use.

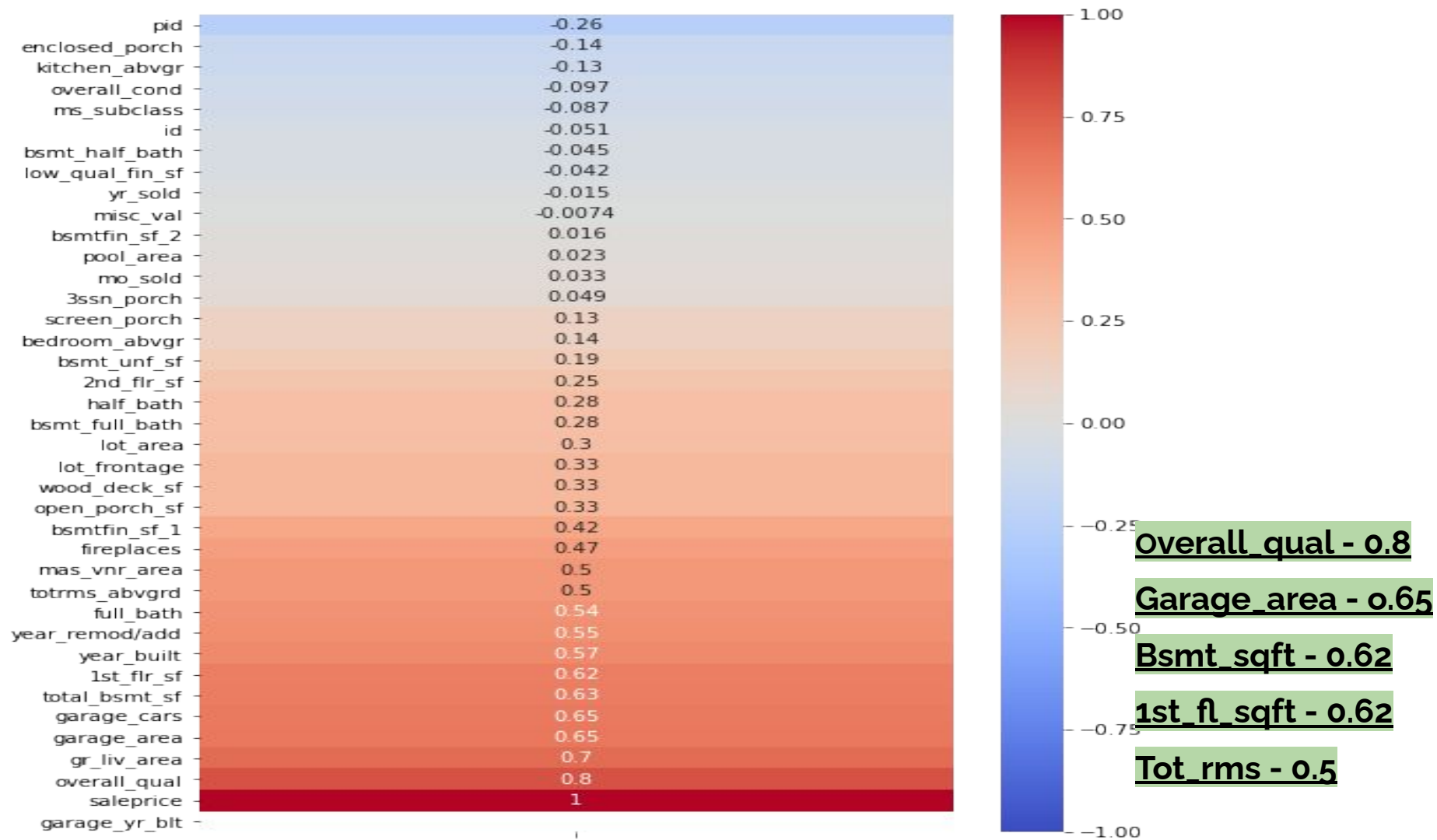


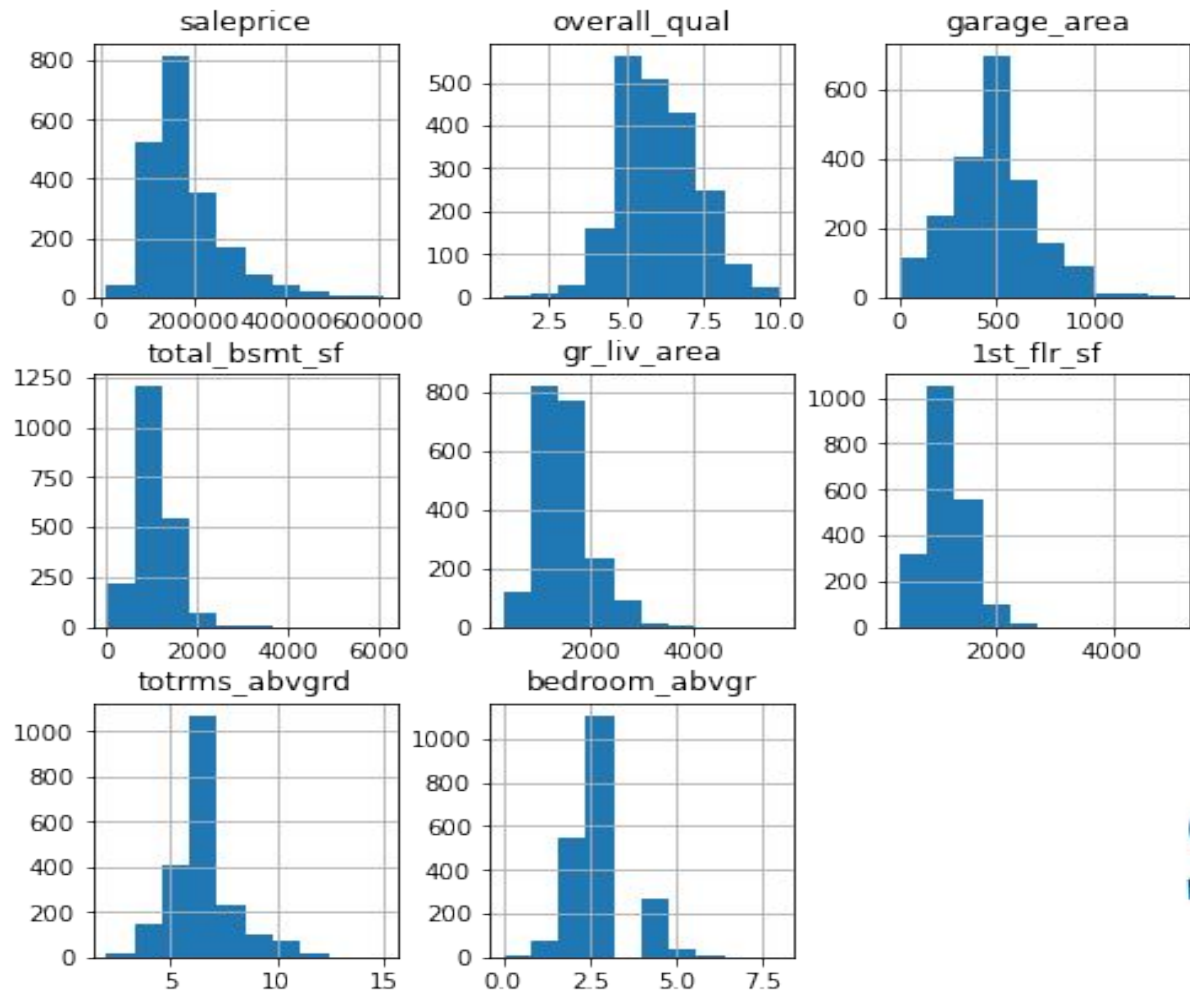


Approach

To predict the price I've used next techniques:

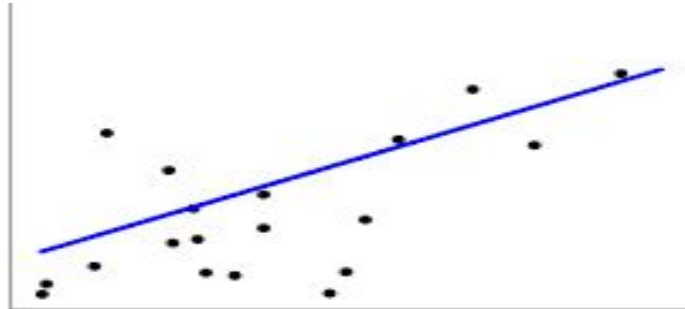
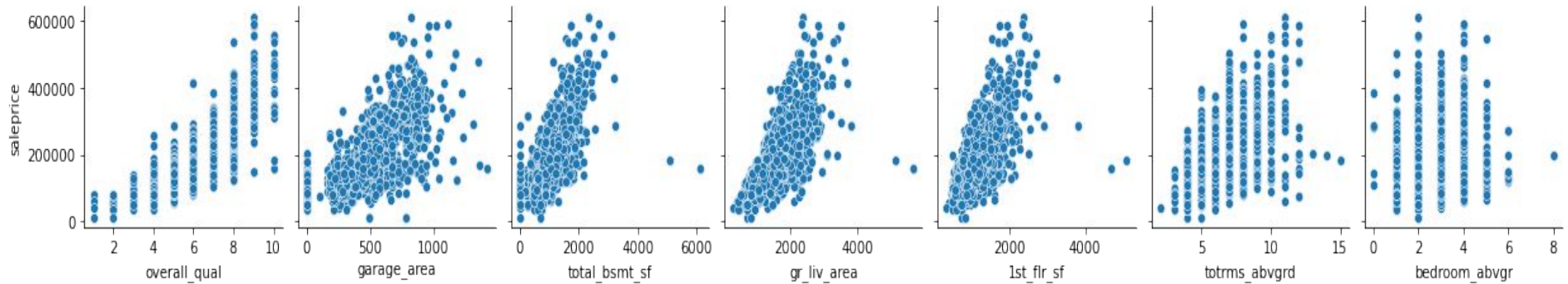
- **Data cleaning**
- **Dummies splits**
- **Check outliers function**
- **Exploratory Data analysis** (graphs, plots, heatmap, correlation tables)
- **Price prediction** (3 models were tested to predict the price: Linear Regression, Lasso and Ridge)



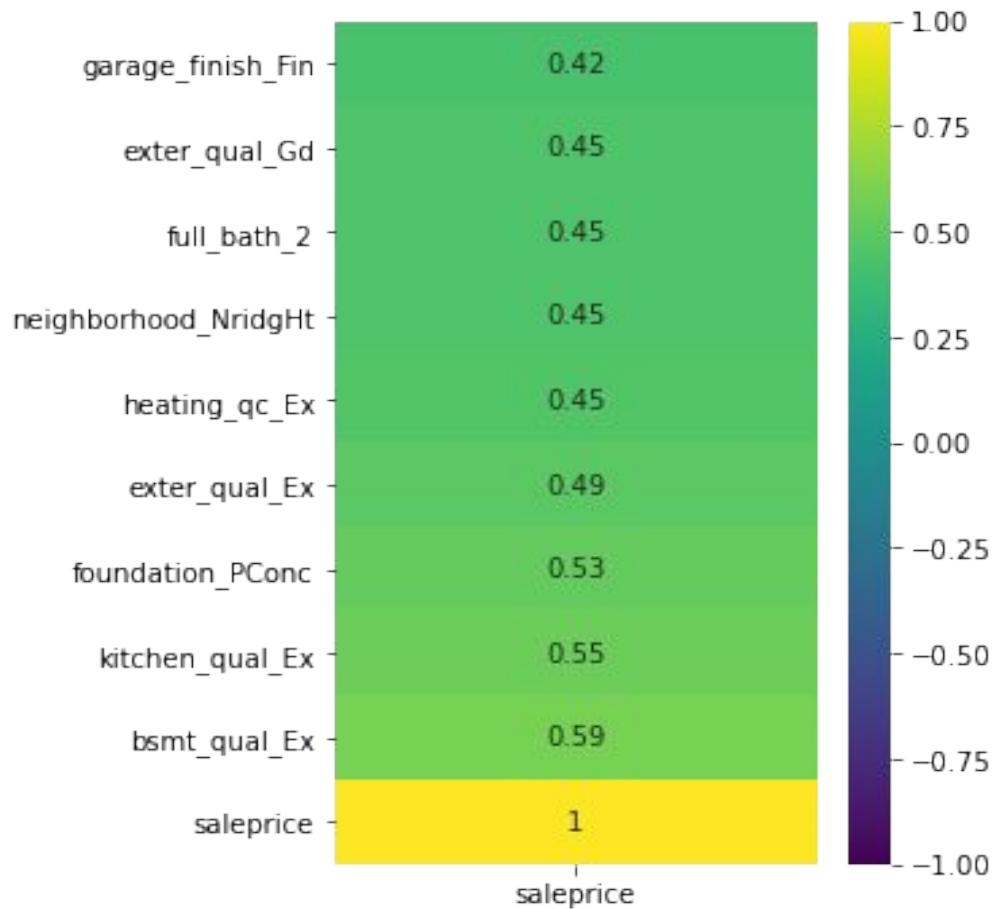


Similarity

Model Preparation



Additional features from not numerical columns



Next features were chosen as best price predictors after model preparation process:

- overall_qual
- garage_area
- total_bsmt_sf
- gr_liv_area
- 1st_flr_sf
- totrms_abvgrd
- bsmt_qual_Ex
- bsmtfin_type_1_GLQ
- bsmt_exposure_Gd
- exter_qual_Ex
- kitchen_qual_Ex
- neighborhood_NridgHt
- central_air_Y
- fireplaces
- full_bath_2
- bedroom_abvgr
- bsmt_qual_Gd



Choosing the best model:

Linear Regression

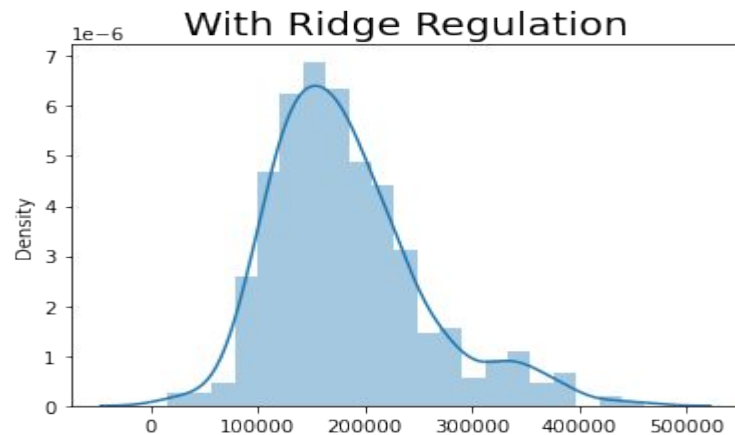
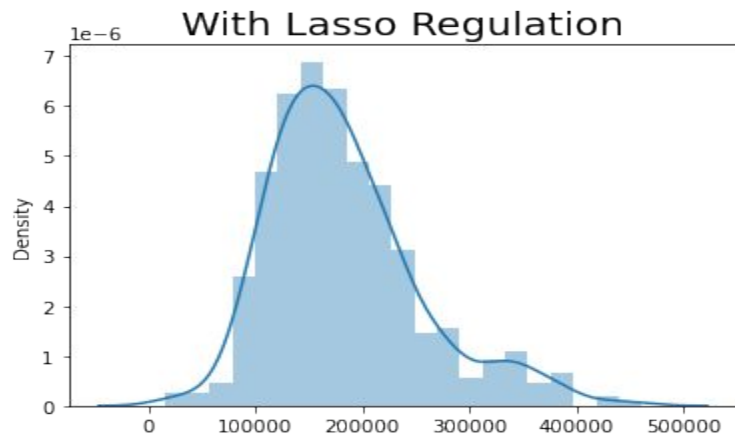
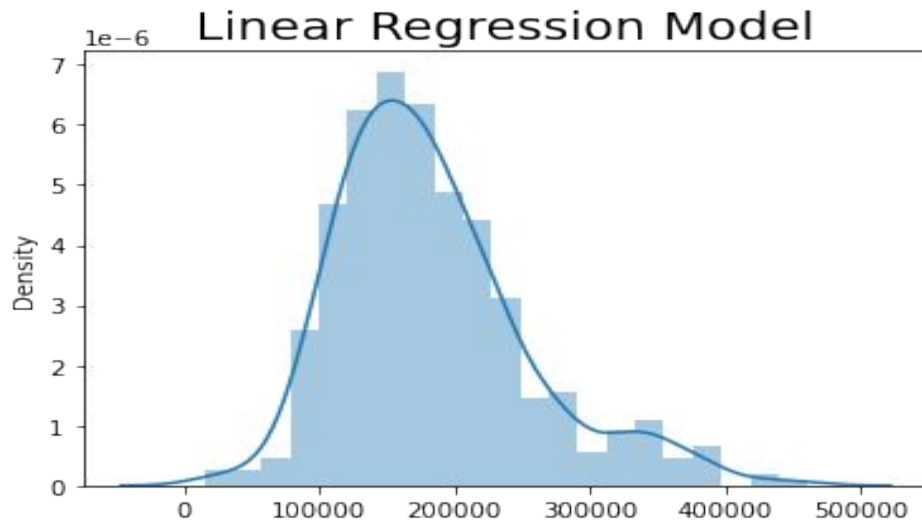
R2 training score 0.8243
R2 testing score 0.8814
Cross_val_score 0.8035
RMSE 26974.20

Lasso

R2 training score 0.8244
R2 testing score 0.8814
cross_val_score 0.8035
RMSE 26974.20

Ridge

R2 training score 0.8244
R2 testing score 0.8814
MSE 727711177 .3209
RMSE 26976.12





Thank you for watching!

For more information and
detailed explanation
please, read my readme
file in my notebook,
especially conclusion part
or contact me at
email@yahoo.com or
phone 777-777-77-77

