# REAL-ESTATE ANALYTICS

*NATHAN MCKINNEY*
*DATA ANALYST AT KENSINGTON ANALYTICS*

# INTRODUCTION

As you enter the housing market, pricing and predictive analytics is key to your launch strategy.

To highlight this, we are looking to predict, given existing data, the price of a house at sale, according to features and relevant industry standards. The model should be able to account for a variety of factors.

# BACKGROUND

- Pricing analytics in U.S. real estate
- Typical features used in comparative market analysis

# WHERE DO WE START?

- Data / Feature Selection

# ABOUT THE DATA

- Ames, Iowa Sale Data
- 2006 to 2010
- Surface area, features, and condition ratings, as well as neighborhood

# FEATURE SELECTION

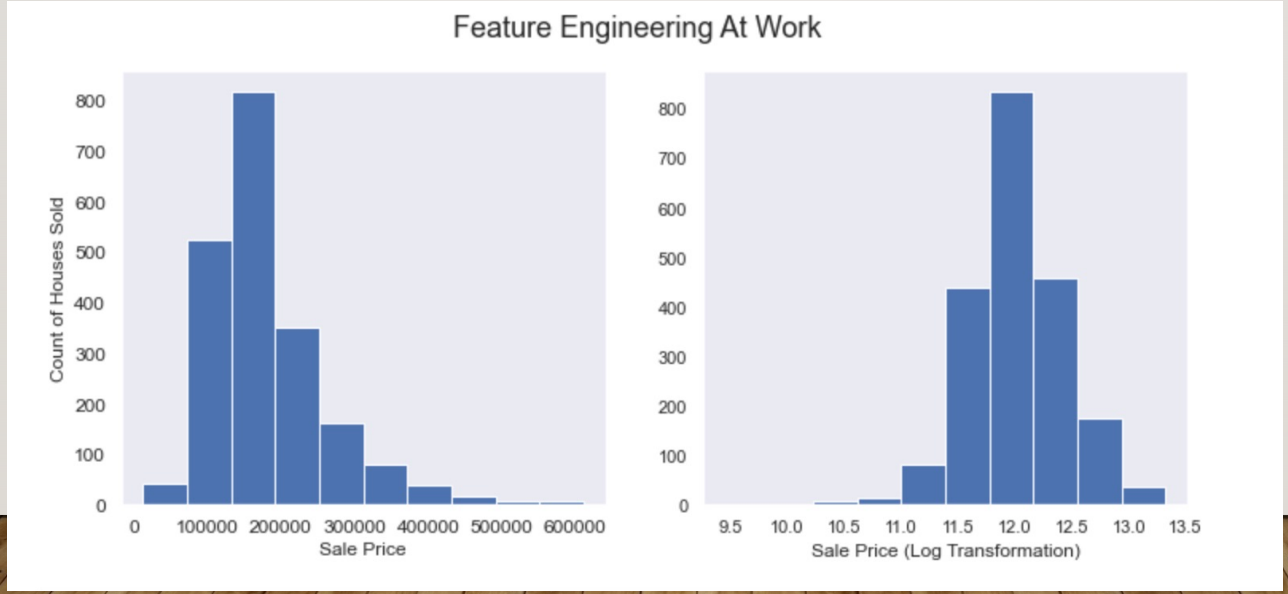
- Prioritized:
  - Surface Area
  - Quality Metrics
  - Certain Features

# WHERE DO WE START?

- Data / Feature Selection
- Feature Engineering

# FEATURE ENGINEERING

- Increase the potential of data with curated features and mathematical adjustments



# WHERE DO WE END?

- Data / Feature Selection
- Feature Engineering
- Our predictive model

# OUR MODEL SPECIFICS

- 16 Features used to predict a sale price
- Null model is mean of housing price
- Lasso, Ridge Regressions with a log transformation

| Some Features |
| --- |
| Overall quality |
| Year Built |
| 1st Floor Surface Area |
| 2nd Floor Surface Area |
| If Kitchen Quality is Excellent |
| If Exterior Quality is Excellent |
| Garage Car Capacity |

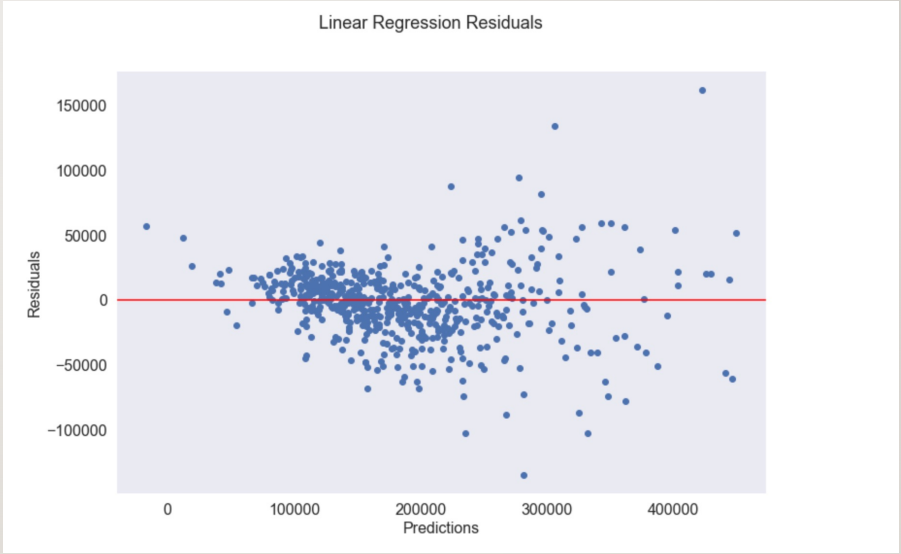
# FINDINGS

# NULL HYPOTHESIS

- Mean housing price used as prediction - $181,250
- Root Mean Squared Error (RMSE)
  - +- 79,200

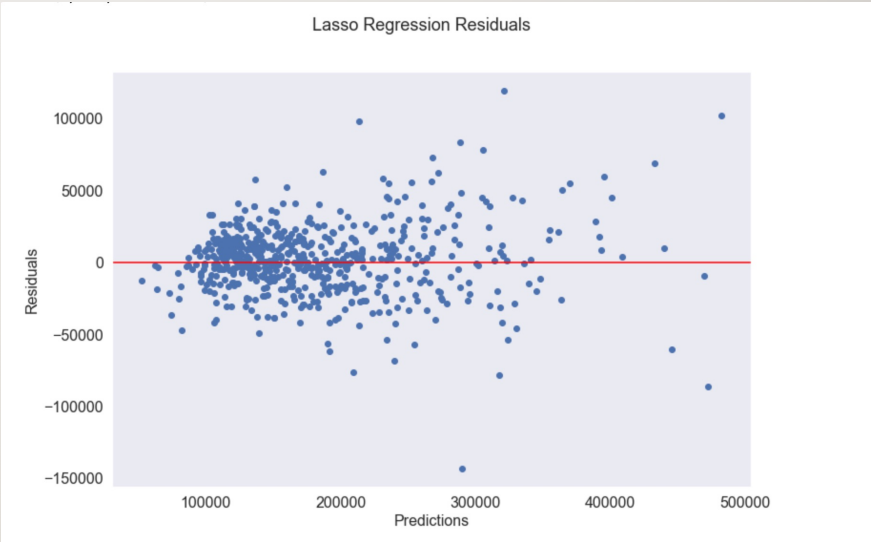
# STANDARD LINEAR REGRESSION (AFTER FEATURE SELECTION)

- RMSE
  - +- 27,800 (Improved 51K)
- R-squared
  - Avg- 88.6%
  - Training- 89.1%
  - Test – 86.6%
- Slightly overfit, but good prediction



# LASSO REGRESSION + LOG TRANSFORMATION

- RMSE
  - +- 24,400 (Improved 4K)
- R-squared
  - Avg- 90.4% (Improved 1.8%)
  - Training- 90.7% (Improved 1.6%)
  - Test – 89.7% (Improved 3.1%)
- Improved fit between test/train, better prediction



# RIDGE REGRESSION + LOG TRANSFORMATION

- RMSE
  - +- 24,200 (up 200)
- R-squared
  - Avg- 90.6%
  - Training- 90.7%
  - Test – 89.7%
- Marginal increase in RMSE

# KEY OBSERVATIONS

- Surface area is a primary predictor in price
- Keeping your kitchens and exteriors in great condition can spur on greater selling prices

| feature | coef |
|---|---|
| 2nd_flr_sf | 19758.0 |
| overall_qual | 18438.0 |
| 1st_flr_sf | 18178.0 |
| total_bsmt_sf | 12525.0 |
| kitchen_qual_Ex | 10063.0 |
| exter_qual_Ex | 9487.0 |
| year_built | 9081.0 |
| garage_area | 8316.0 |
| lot_area | 7735.0 |
| gr_liv_area | 6525.0 |
| year_remod/add | 5872.0 |
| bedroom_abvgr | 5845.0 |
| fireplaces | 3812.0 |
| garage_cars | 2784.0 |
| heating_qc_Ex | 2541.0 |
| full_bath | 2164.0 |

# RECOMMENDATIONS

# AS YOU ENTER THE MARKET

- Prioritize surface area in your strategy
- Periodically review feature importance in pricing to find price multipliers
  - Do consumers prefer marble countertops? Or wood?
  - Would a customer prefer a full upgraded Kitchen?

# POSSIBLE WAYS WE CAN PARTNER WITH YOU

- Develop a specific model for your investment environment
- Dashboard that can evaluate feature importance and changes over a rolling three-month period
- Implement a real-time pricing tool for your use
  - Specify housing metrics and get a prediction for your investment area

# BEFORE YOU GO

- As you approach real estate, data analytics and modeling can drive your competitive advantage.

- We have the domain experience and capabilities needed to provide additional insights and leverage as you enter the market.