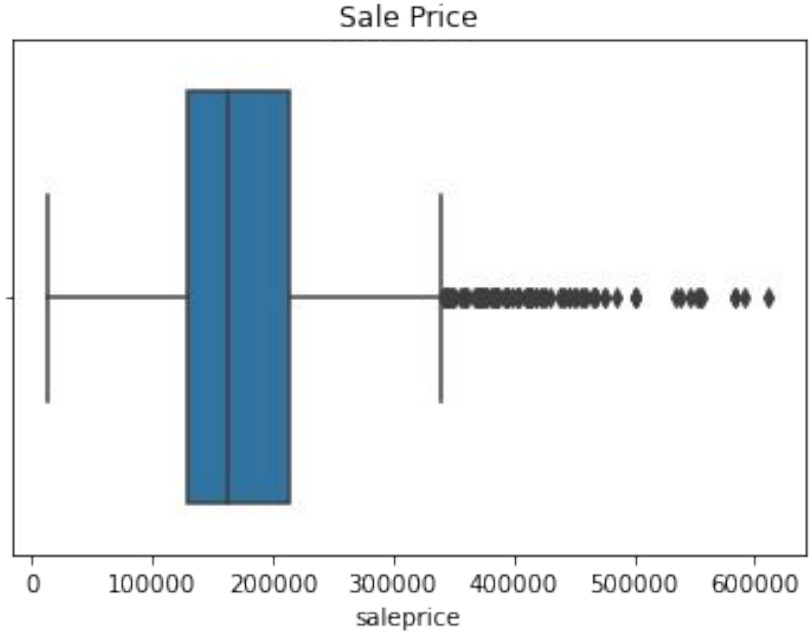
Regression: Predicting House Sale Prices

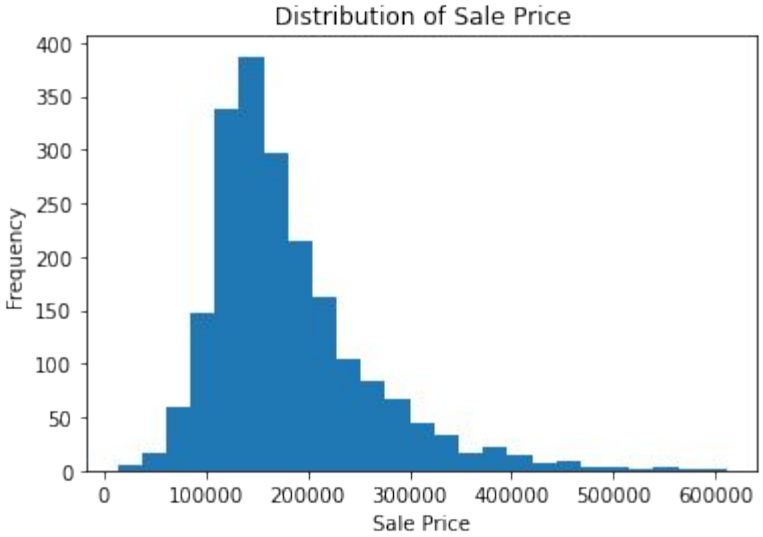
Seung Woo Choi

Problem Statement

- **Context:** Data science consultant for real estate agency, interested in advising homeowners how to maximize the sale price of their houses
- **Goal:** Predict house sale prices within 25,000 USD and determine the most important features for predicting sale price of houses in Ames, Iowa
- **Evaluation metrics:** RMSE & R-squared

Limitation of Outliers



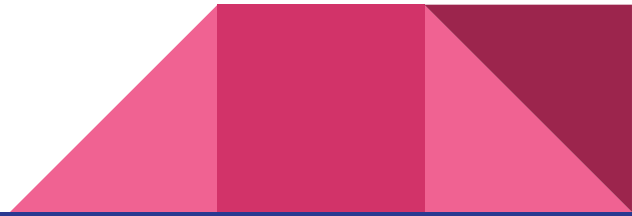


Defining a Successful Model

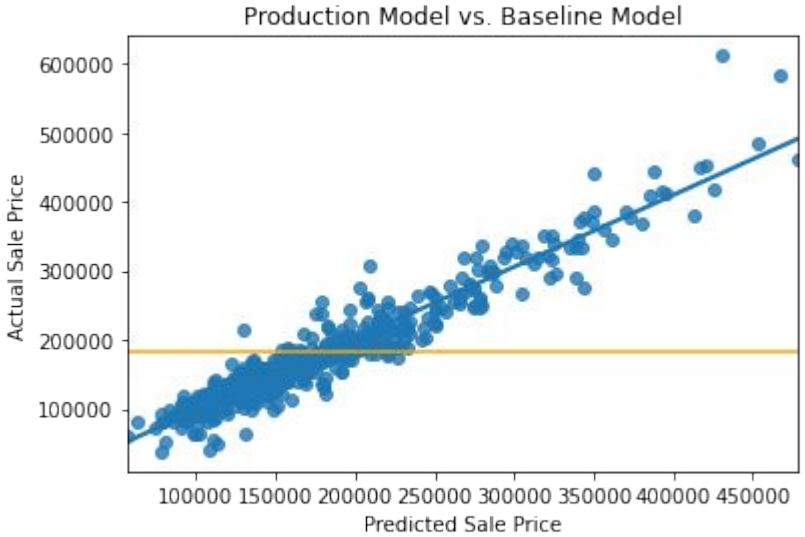
- **Baseline model:**
  - RMSE: $79,262
  - R-squared: 0.0
- **Production model:**
  - Better than the baseline model

Data Source

- Assessor’s Office in Ames, Iowa
- Data collected between 2006 and 2010



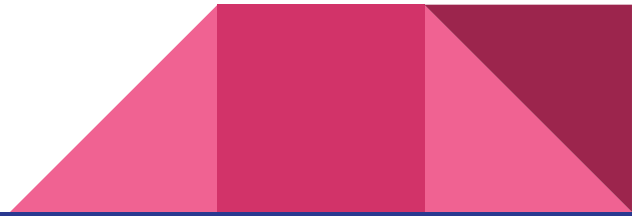
Understanding the Model



- **Model:** Ridge Regression
- **Features:** 14 in total
- **Selection criteria:**
  - Correlation with sale price
  - Domain knowledge

Conclusion

- **Performance of production model:**
  - RMSE: $24,410
  - R-squared: 90.8%
- **Recommendation for future model:**
  - Separate, tiered models based on specific price ranges
- **Feature to improve:**
  - Kitchen quality



Thank You!