Project 2: Ames Price Prediction Model

Image credit: https://www.vecteezy.com/members/khurshidalamek

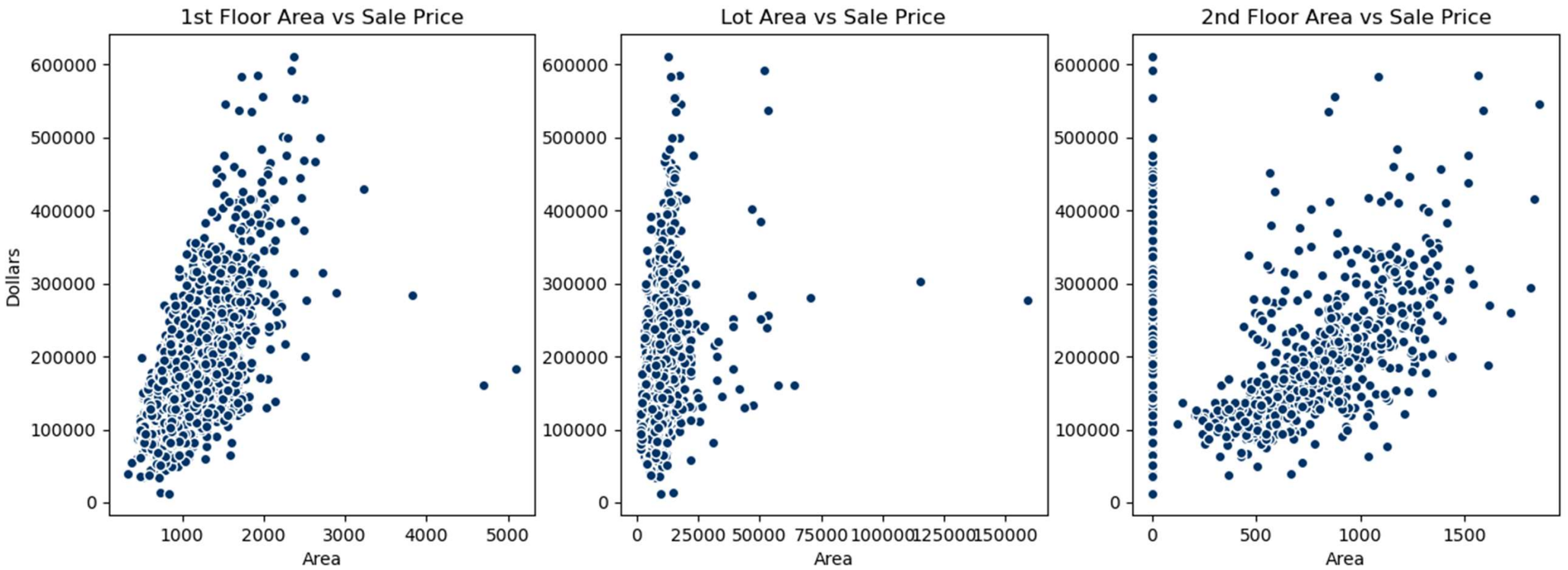
Luis Ulloa

Problem Statement

- **Premise: Homeowners value space, plenty of it**
  - The bigger the house, the bigger the lot, the higher the selling price
  - Aim: use measures of area (space) to build predictive pricing model

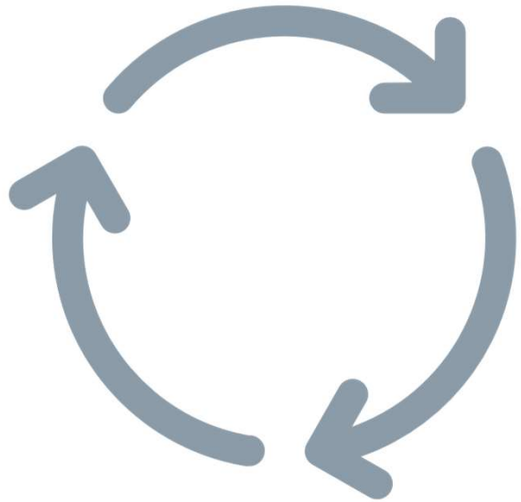
Exploratory Data Analysis

- **Baseline:** 1st floor area, lot area, 2nd floor area



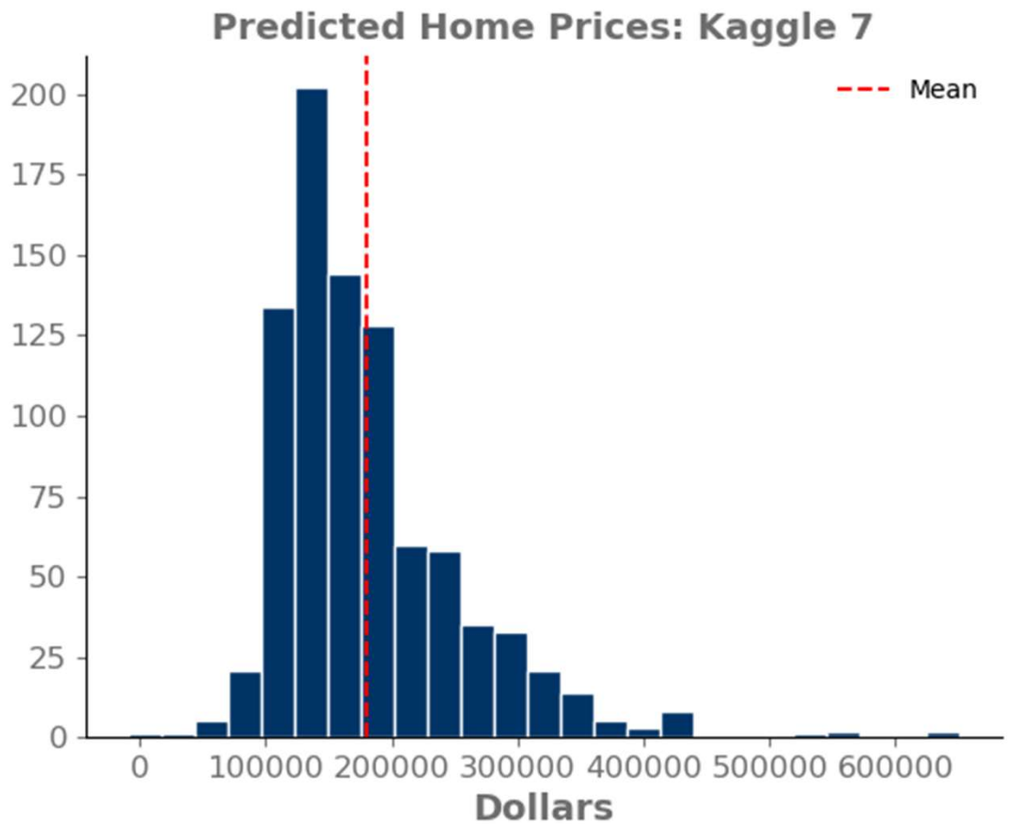
Model Iterations

- **Iteration 2:** combined area columns, automated process
- **Iteration 3:** added time features (year)
- **Iteration 4:** removed some features
- **Iteration 5:** introduced interactive terms
- **Iteration 6:** polynomial features and scaling
- **Iteration 7:** dummified variables



Final Model & Kaggle Preds

- **Features:**
  - Numeric
  - Dummified discrete
  - Interactive terms
  - Polynomial and scaling transforms



Summary

- Engineering features is difficult, but fun work
- Finding the right mix and approach is an art
- Bias and variance are always in conflict

Next Steps

- Further EDA to discover other relationships
- Fine tune model based on data discoveries