

The background image is a landscape photograph. In the foreground, a weathered wooden post supports a strand of barbed wire that runs diagonally across the frame. The ground is covered in dry, golden-brown grass. In the middle ground, there are several bare, leafless trees. The sky is a clear, solid blue. The overall tone is somewhat somber and rural.

AMES HOUSING DATA REGRESSION CHALLENGE

METHODOLOGY

- ▶ Define the Problem
 - ▶ Gather Data
 - ▶ Explore Data
 - ▶ Model the Data
 - ▶ Evaluate Model
- ▶ Answer Problem or Next Steps

GOALS

Create a regression model based on the Ames Housing Dataset

This model will predict the price of a house at sale

AMES HOUSING DATA

THE DETAILS

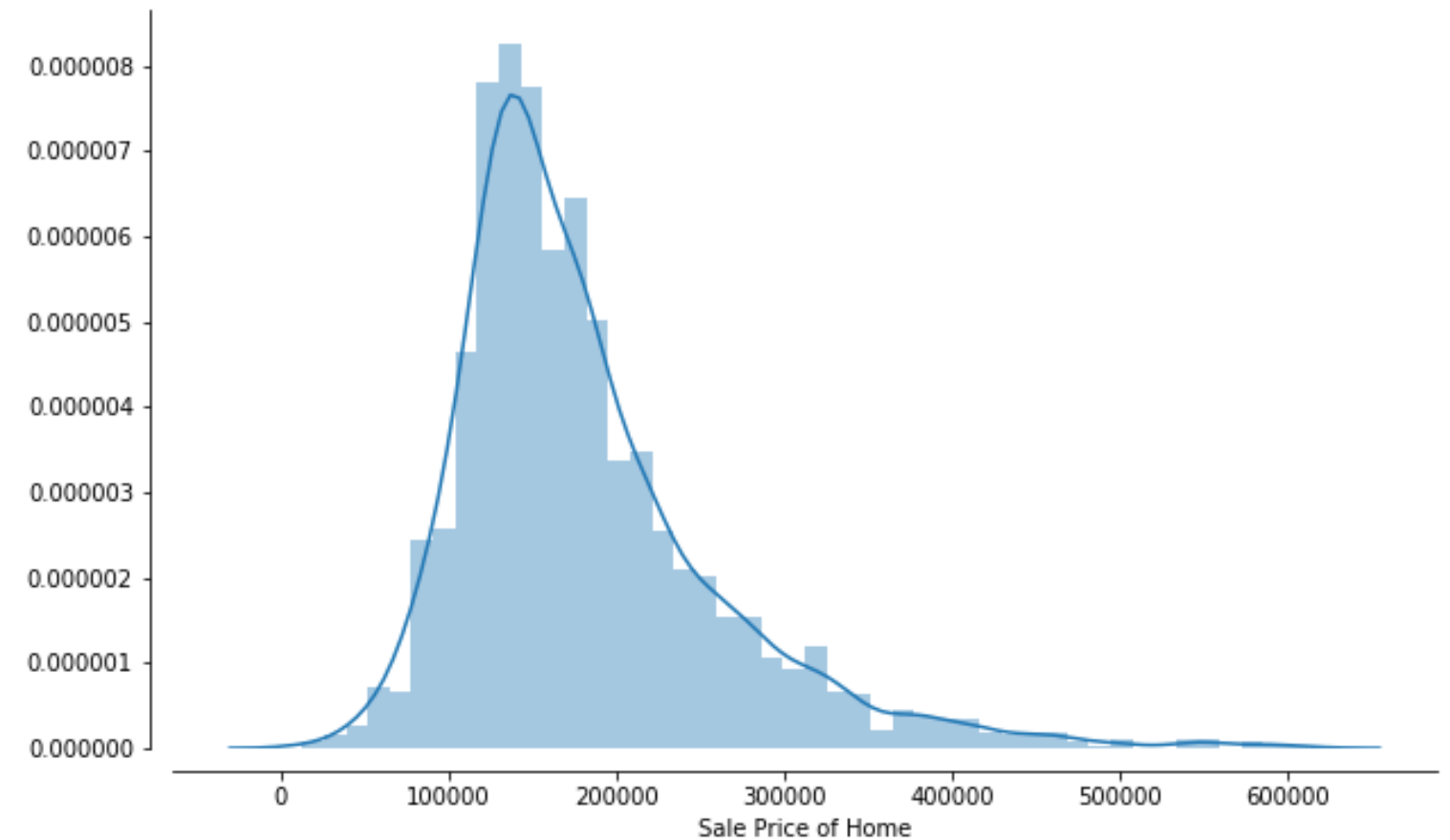
Sale of individual residential property in Ames, Iowa from 2006 to 2010.

- 23 nominal**
- 23 ordinal**
- 14 discrete**
- 20 continuous**

GETTING INTO THE WEEDS

Price Histogram

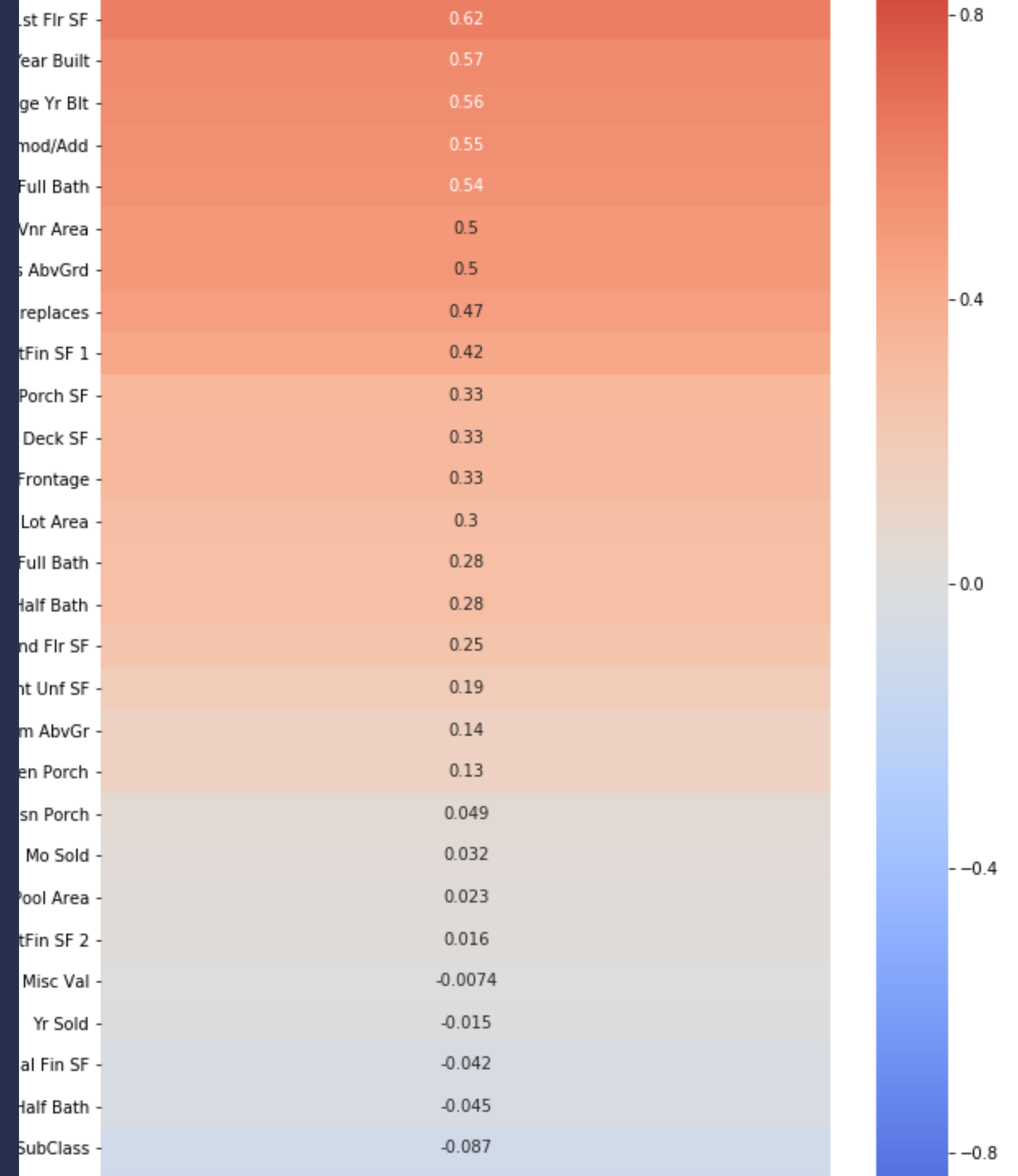
- Skew in price
- Using just a linear regression will not suffice



MORE WEEDS

Correlation Stack

- Several features that show significant correlation to be considered
- Gives reason for adding interaction terms



REGRESSION

ELASTIC NET

- + For case of high dimensionality
- + Multicollinearity among the variables in the data set

RESULTS

Not Great...

First run - 105 out of 106 on Leaderboard

MSE: 105 to 35

NEXT STEPS

Add Interaction Terms

Add dummy columns for categorical features

Different Regression Types

THANK YOU

Questions