



Real Estate: It's Ame-azing!

A REPORT BY EMILY K. SANDERS

Problem Statement

- ▶ Realtors and their clients can gain advantage by correctly estimating how much a house will sell for.
- ▶ In this report, I will demonstrate a model that I believe can aid towards this end.



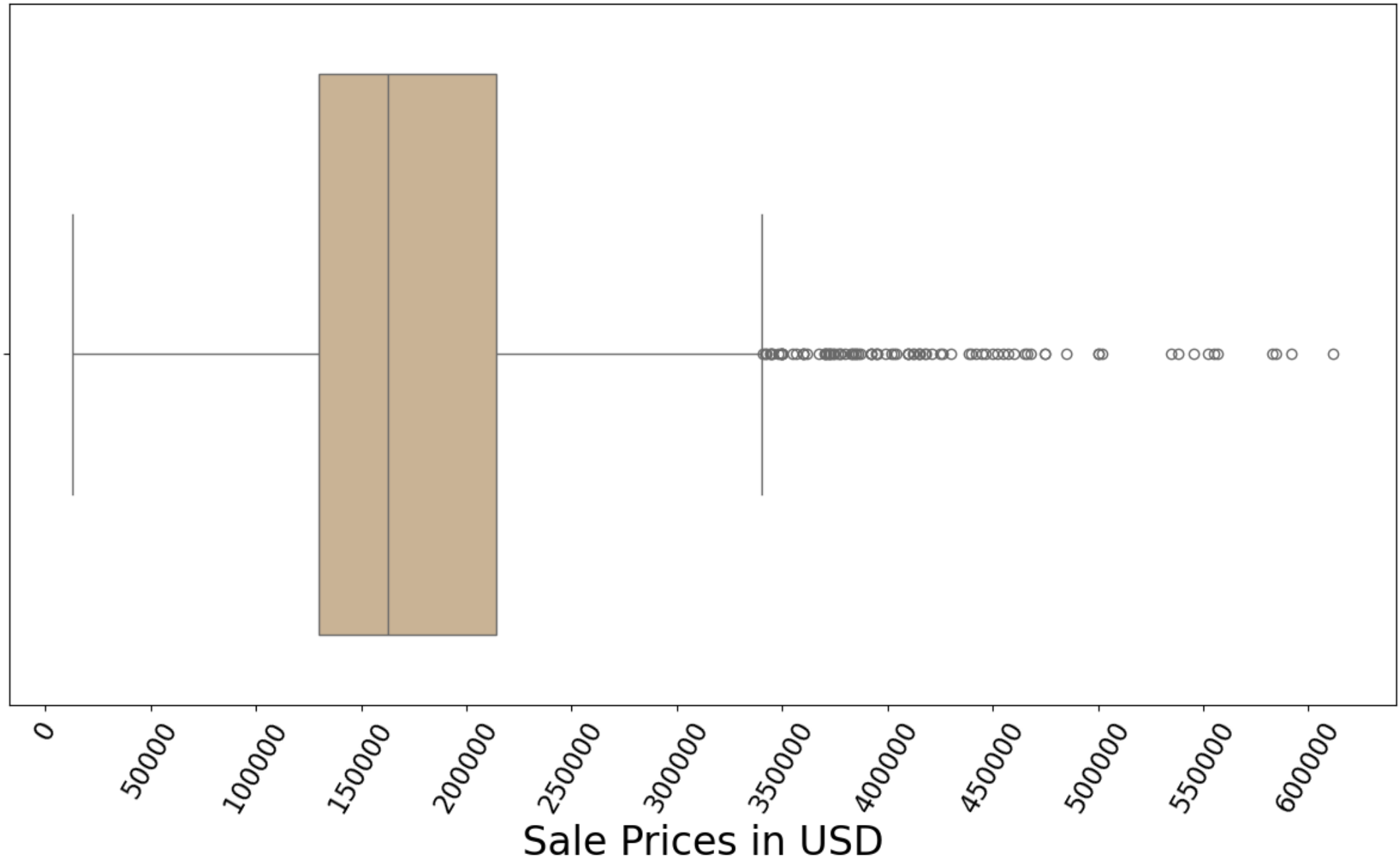
Photo from Zillow



Method

- ▶ I used a dataset originally compiled by the City Assessor.
- ▶ 81 columns of a wide assortment of information on 2051 home sales between 2006 and 2010.
- ▶ Target variable was sale price.
- ▶ A whole lot of preprocessing!

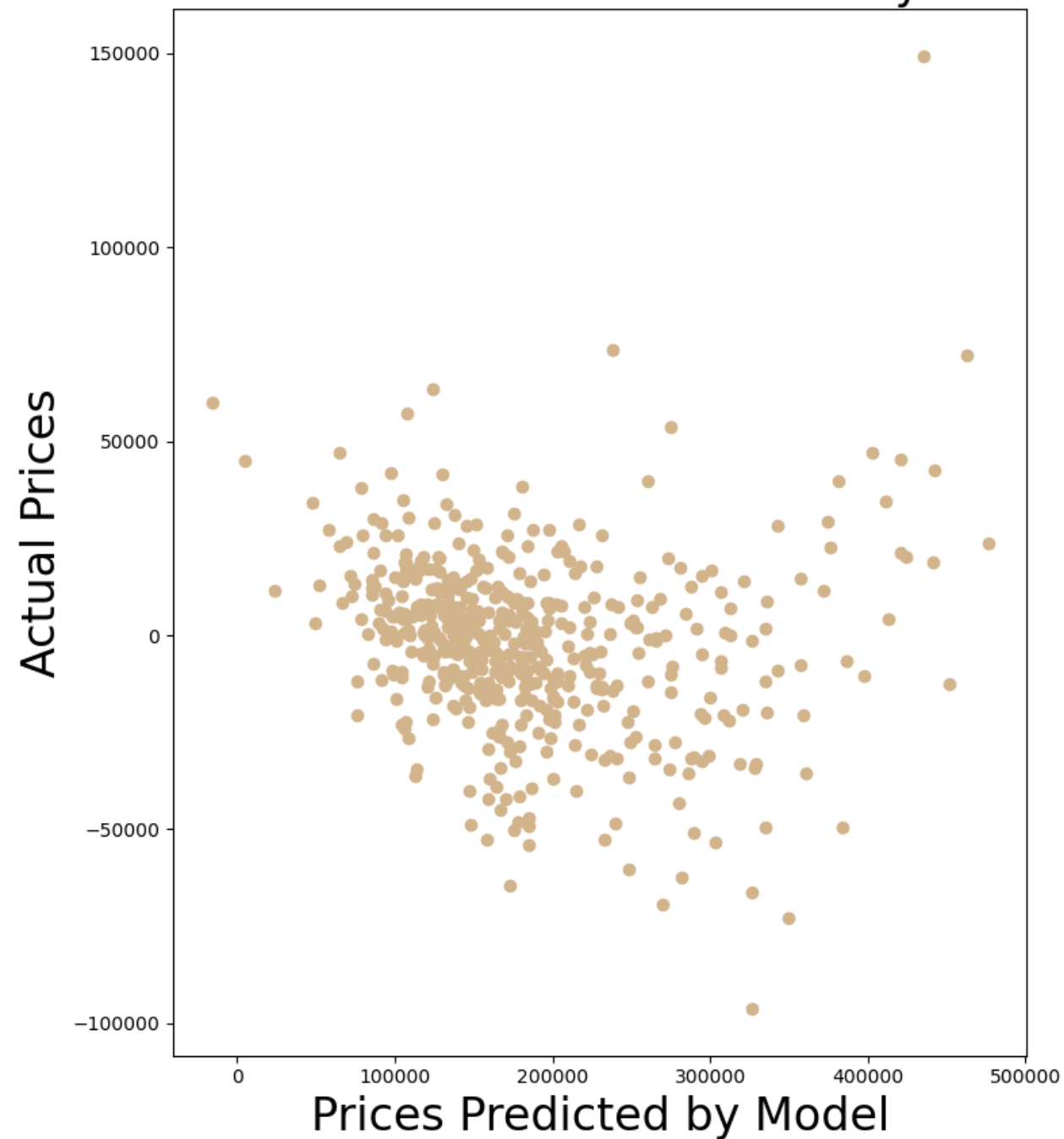
Distribution of Sale Prices



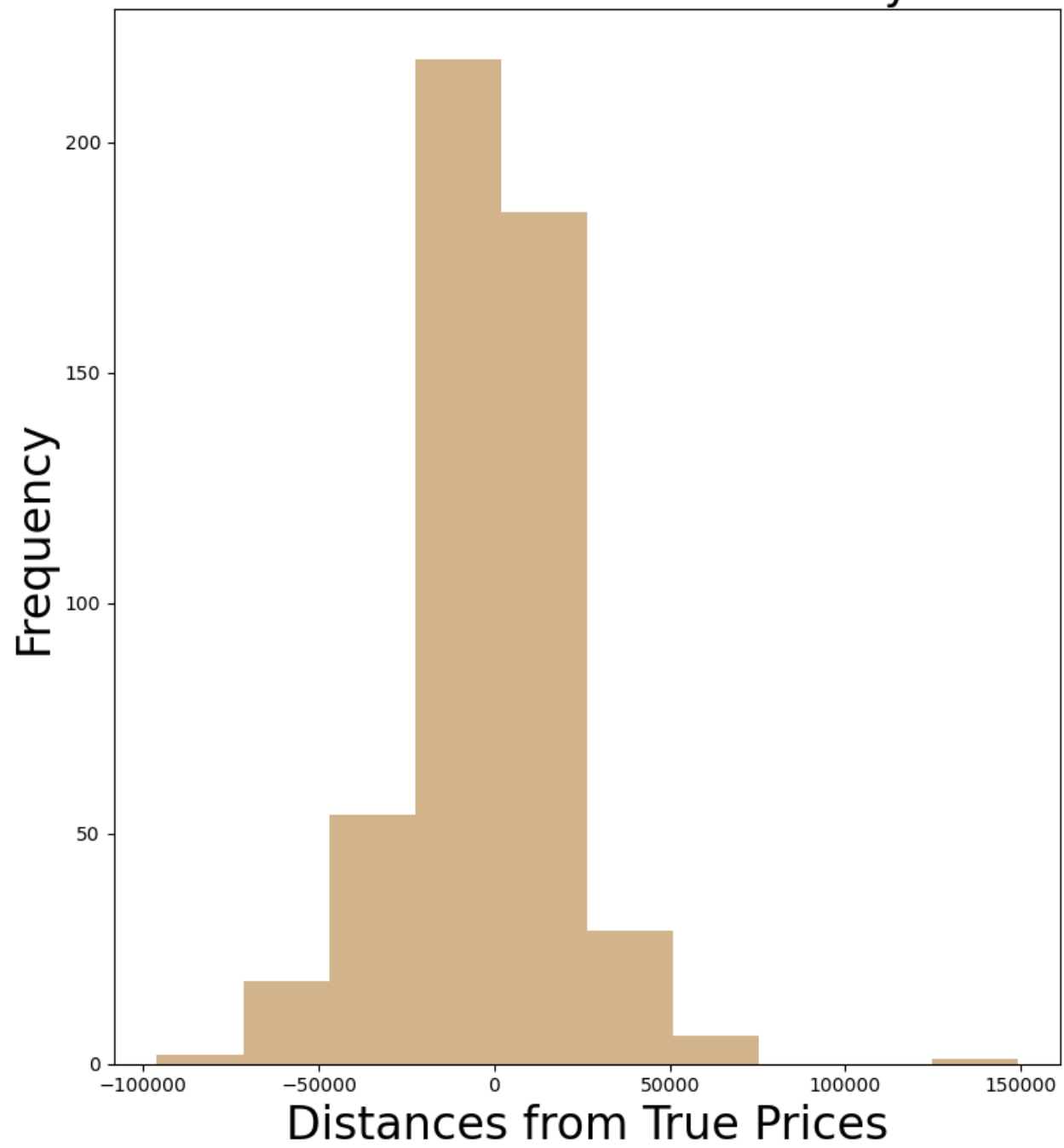
Best Model

- ▶ LASSO regression model.
- ▶ Linear model with all 229 predictors put in, but not all used.
- ▶ Has an average error distance (RMSE) of about \$23,000, and explains about 92% of the variety in home prices.

Residuals - A Test of Efficacy



Residuals - A Test of Efficacy



Recommendations

Future Improvements

- ▶ Collapse minor levels of categorical variables.
- ▶ Keep searching for a human-readable model.
- ▶ Add complex terms.

Immediate Use

- ▶ The model is not perfect, but none ever is!
- ▶ The model can be deployed for immediate use by realtor clients, with some cautions.
- ▶ Maybe add a full bathroom?



Thank you!

- ▶ I will now take questions.