Houses Over List Price Prediction

Salma Katri

January 26, 2022

Background

Motivation

- Housing prices have seen unprecedented y-o-y increases
- Seattle area's bidding wars with houses sold more than $300k over list price
- Redfin available data can help inform homebuyers' decisions

Objective

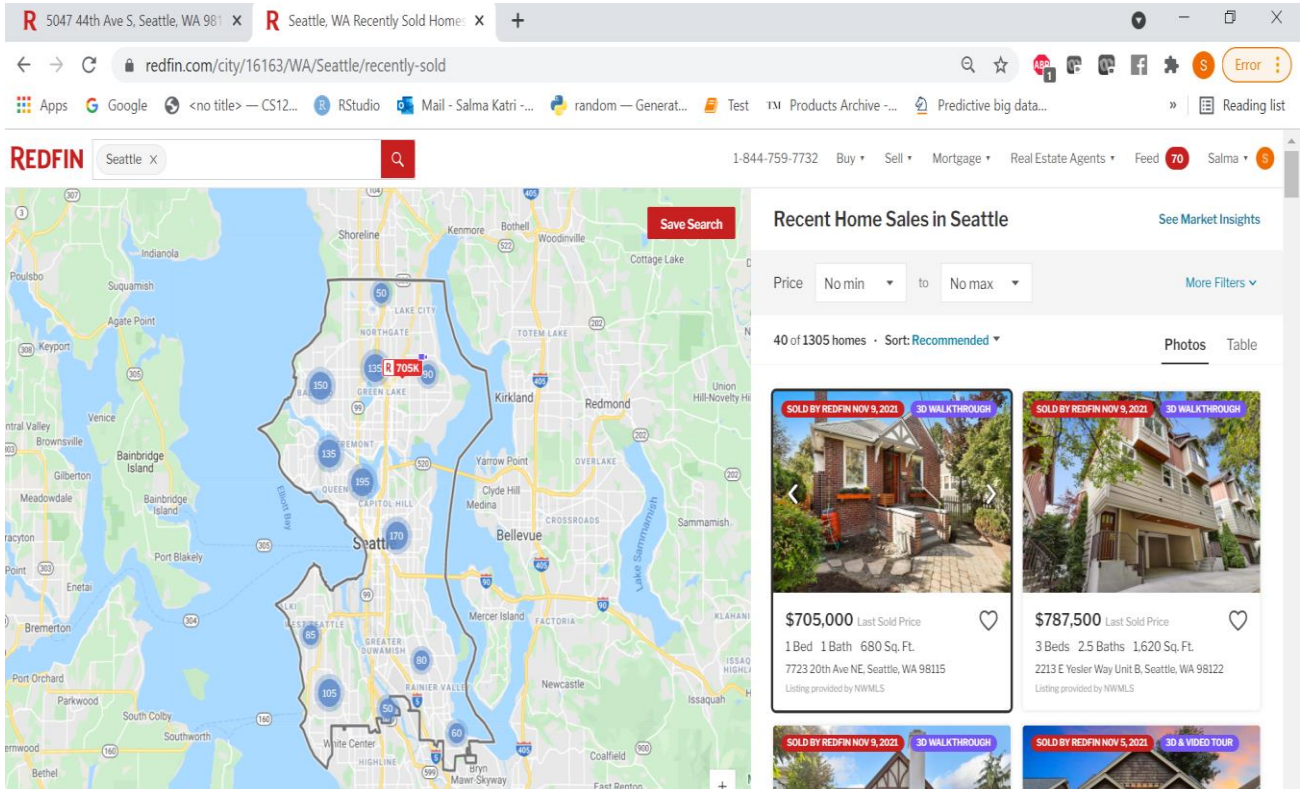
- Predict whether a house will sell for over the list price using Redfin available data

Results

- Random Forest and Gradient Boosting provided better predictions
- Recall score of 94%
- Most significant features include Redfin estimate and days on the market

Data Source

Scraped Redfin sold homes in Seattle Greater Area

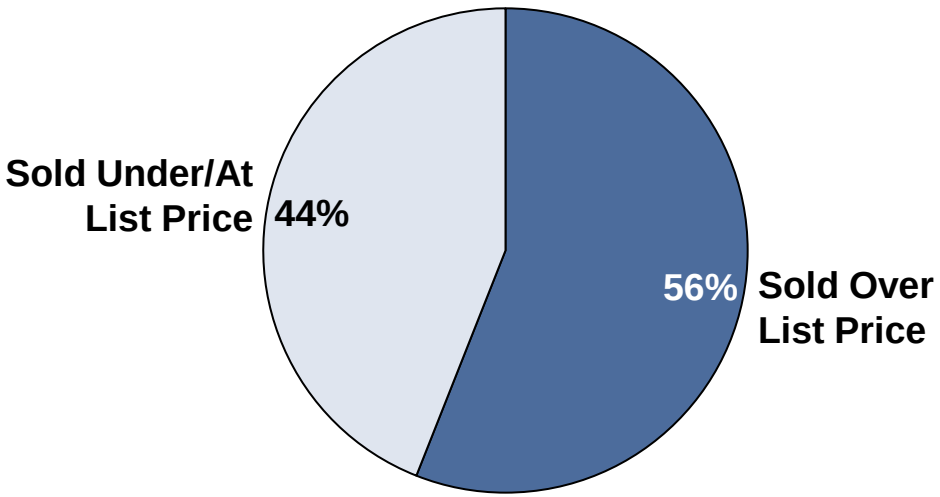


- **2530 rows**
- **24 variables**

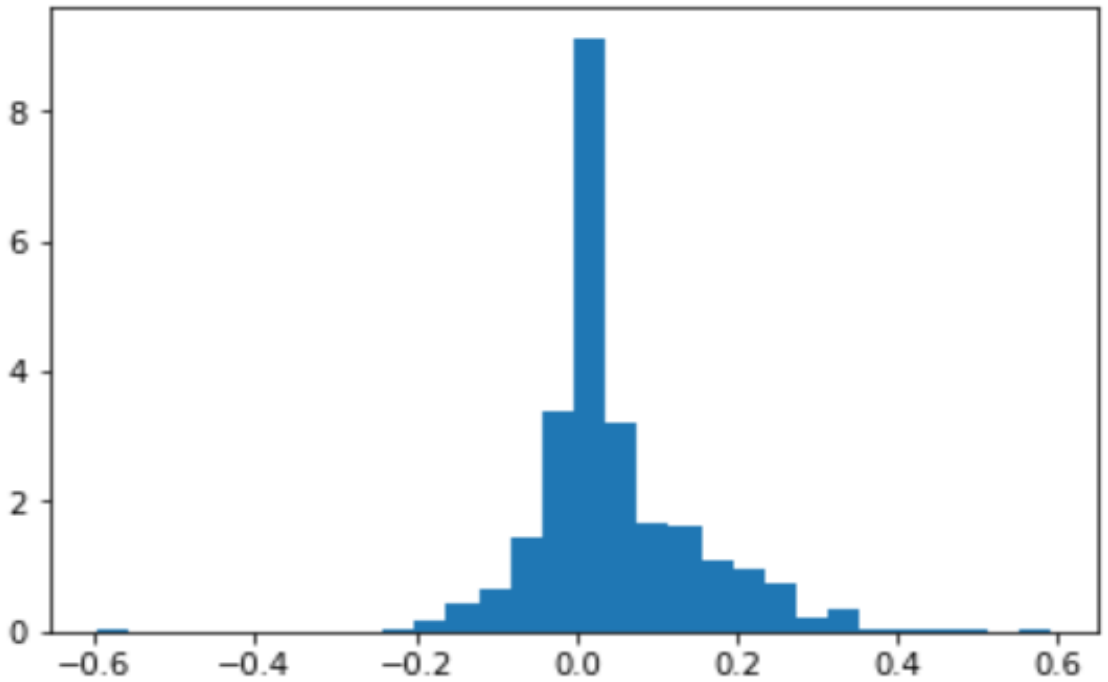
- Dropped variables with many missing values
- Creating new variables: House Age, Over Under Ask (target), Average School Ratings, Difference between Redfin estimate and list
- Binary variables: Renovated and Recently Renovated
- Neighborhood Groups

- **1315 rows**
- **21 variables**

Target Variable



Over/Under Ask Distribution



Models Baseline Comparison

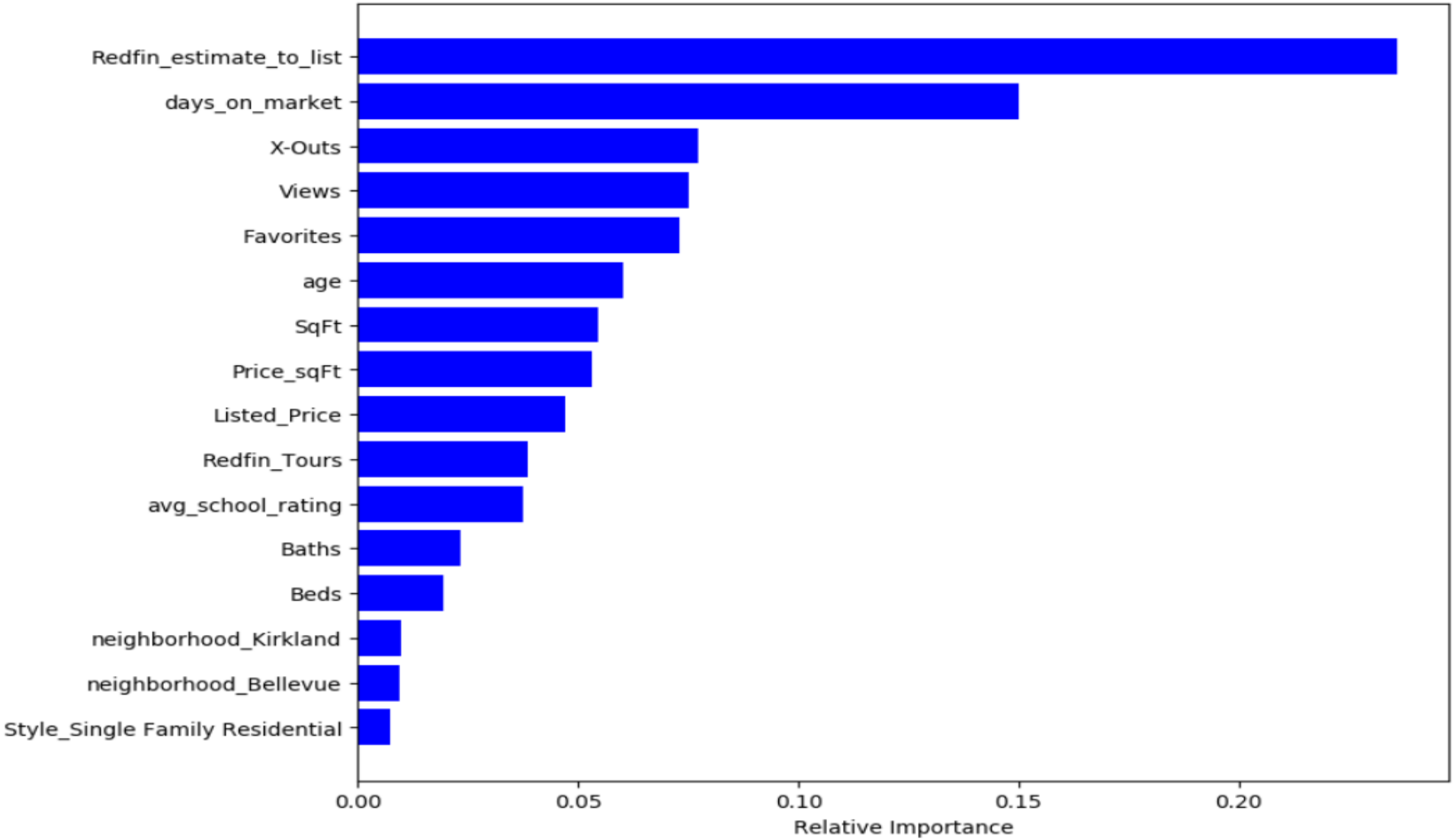
Logistic Regression	**0.78**
KNN	**0.88**
Random Forest	**0.93**
XGBoost	**0.93**

In choosing and tuning the model, Recall was prioritized

Random Forest Tuning

Using Grid Search to tune the parameters, Random Forest model recall score increased from 0.93 to 0.95

Features Importance



Future Work

- Further Tuning
    - More parameters of random forest
    - Tuning XGBoost
- More datapoints and longer period