

PREDICTING HOUSING PRICES IN KING COUNTY, WA

Presented by Matt Schwartz

CONTENTS

03 - Overview and Business Problem

04 - Data

05 - Methods

06 - Results

07 - Conclusions and Further Analysis

OVERVIEW AND BUSINESS PROBLEM

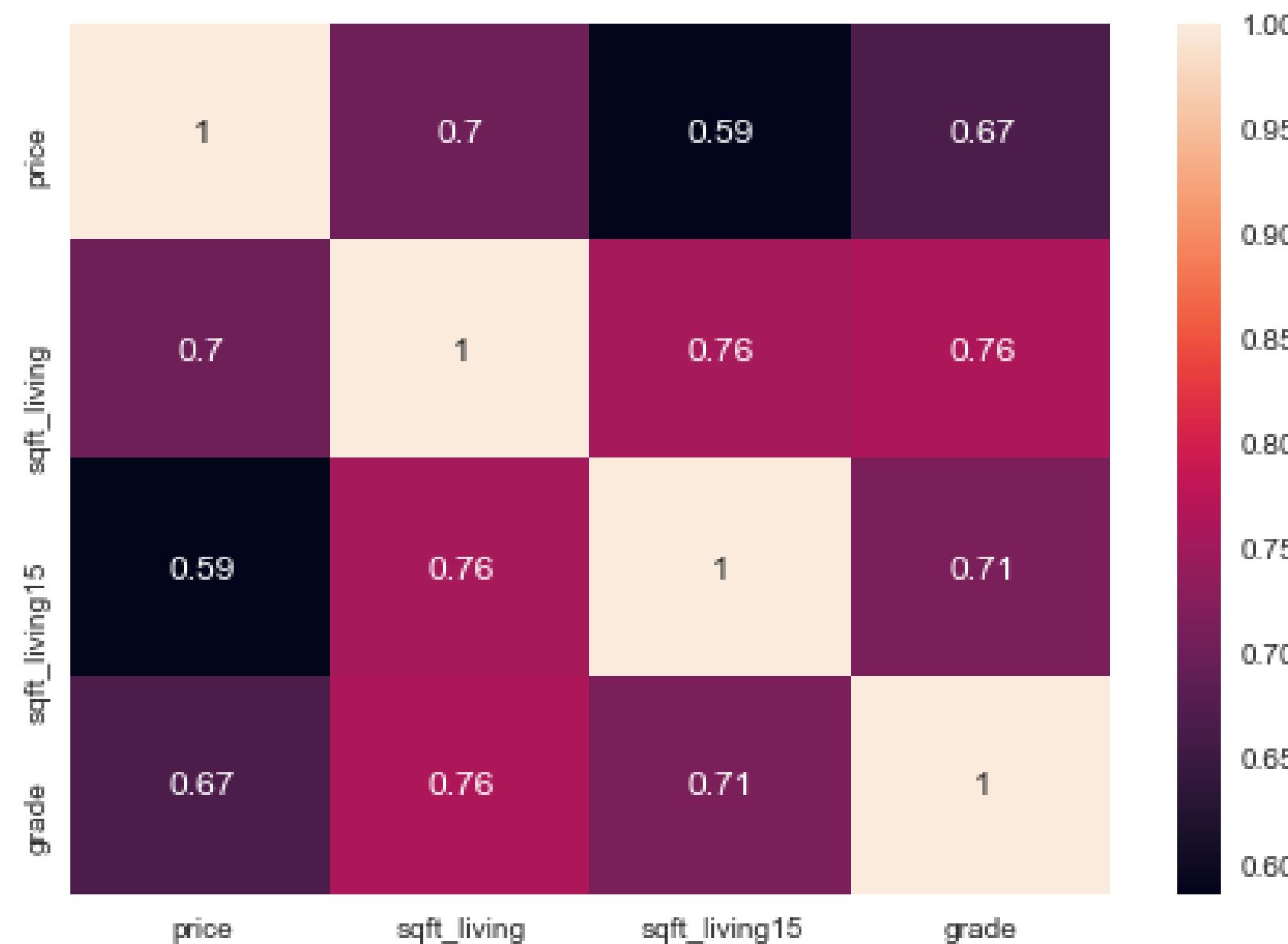
- Home buying is frustrating and tedious
- How do you maintain your leverage and avoid overpaying for a new home?
 - A prediction engine that helps prospective home buyers know if they're getting a good deal
- Models developed here can serve as the backbone for an app or website

DATA



- Over 21K homes in King County, Washington State, and what they sold for
 - County contains Seattle and Bellevue
- Variables include:
 - Square footage
 - Year built/renovated
 - Location
 - Condition and Grade

METHODS



DATA CLEANING

Removing outliers, filling in null values, one hot encoding, addressing collinearity, new variable creation

MULTIPLE LINEAR REGRESSION

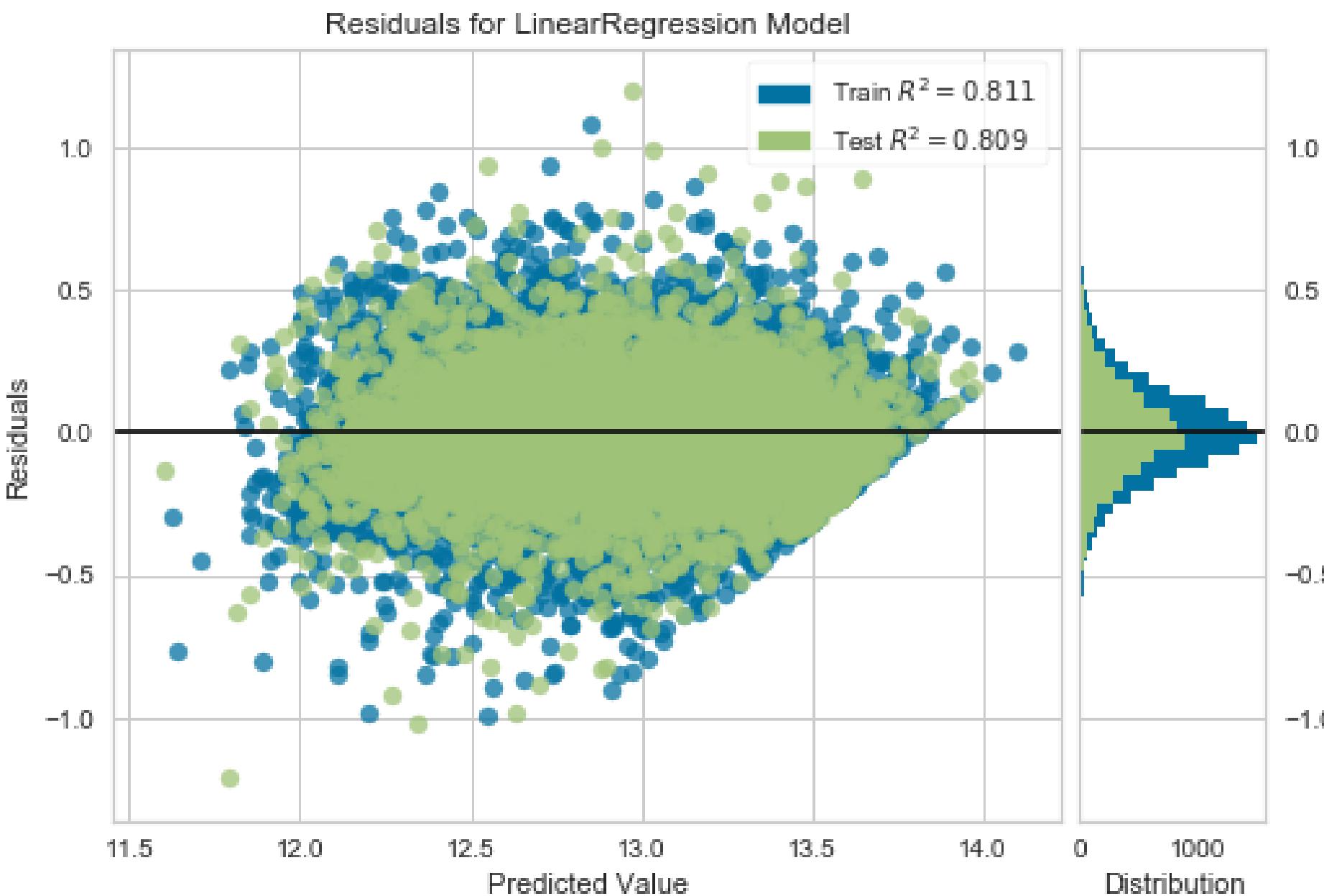
Regressing independent features on the log of price

DEFINING A QUALITY MODEL

R2 and Root Mean Squared Error

RESULTS

HOW'D OUR MODEL DO?



Benchmark Model:

- Train R^2 : **0.65**
- Test R^2 : **0.66**
- Train RMSE: **\$112,216**
- Test RMSE: **\$110,609**

Price Effects:

- Square Footage: **19% increase**
- Distance from downtown: **13% decrease**

Final Model:

- Train R^2 : **0.811**
- Test R^2 : **0.809**
- Train RMSE: **\$85,812**
- Test RMSE: **\$85,490**

CONCLUSIONS AND FURTHER ANALYSIS



ACCURATE PREDICTION ENGINE

- Final model accounts for over 80% of the variation in price
- Slight overfit

LIMITATIONS

- Hidden variable bias
- More data

FUTURE WORK

- Intake more data points and variables to develop reliable model to be used in a housing prediction application

PRESENTED BY, MATT SCHWARTZ

FLATIRON ONLINE PART-TIME DATA SCIENCE COHORT

Phase 2 Project - 1/31/2021

mtschwart@gmail.com



THANK YOU FOR YOUR TIME!
