Linear Regressio

MODELING

BY TRAVIS ROYCE

table of CONTENTS

PROJECT OVERVIEW

BUSINESS OVERVIEW

DATA OVERVIEW

MODELING

REGRESSION RESULTS

NEXT STEPS

THANK YOU

Project Overview

HOUSE PRICE ANALYSIS

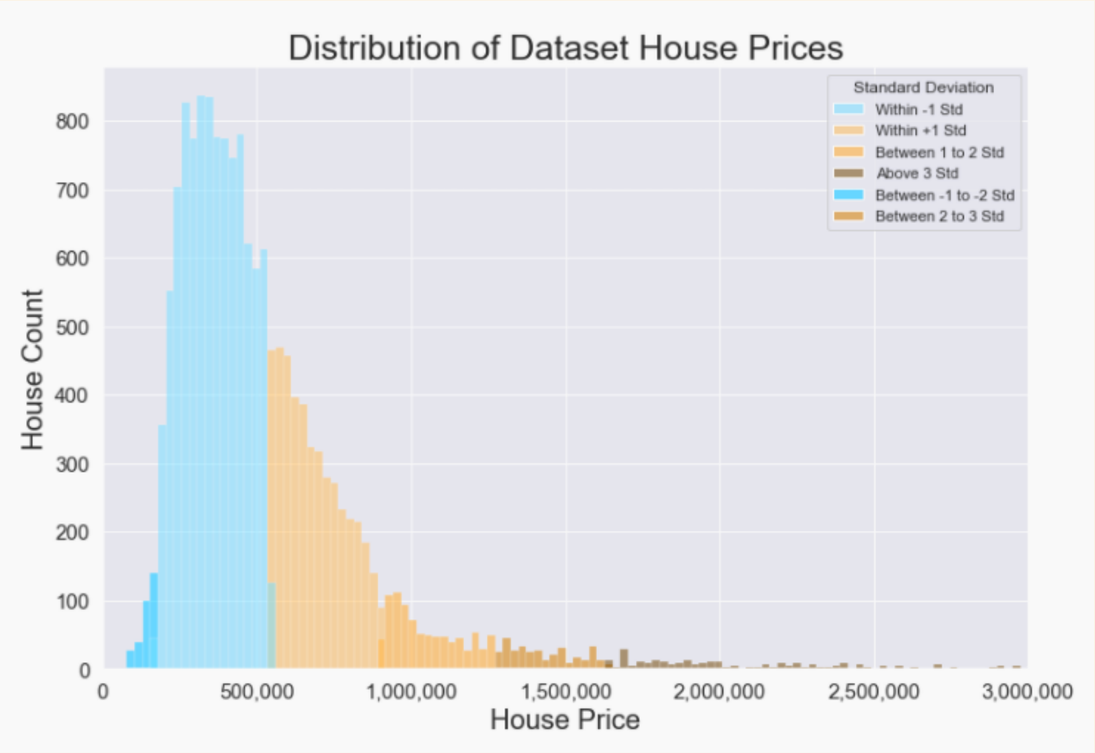
This project utilizes a King County(Washington) house sales data set to model the price of homes (and their features) in the area using multiple linear regression modeling.

Business OVERVIEW

- This project utilizes a King County(Washington) house sales dataset to model the price of homes in the area using multiple linear regression modeling.
- My chosen stakeholder is a hypothetical real estate firm that specializes in "flipping" homes, which involves the purchase, upgrading, and selling of homes.
- The purpose of this project is to provide advice to this real estate firm on what home renovations may increase the values of the homes already owned, and at what prices to purchase and upgrade other homes in King County areas.

Data OVERVIEW

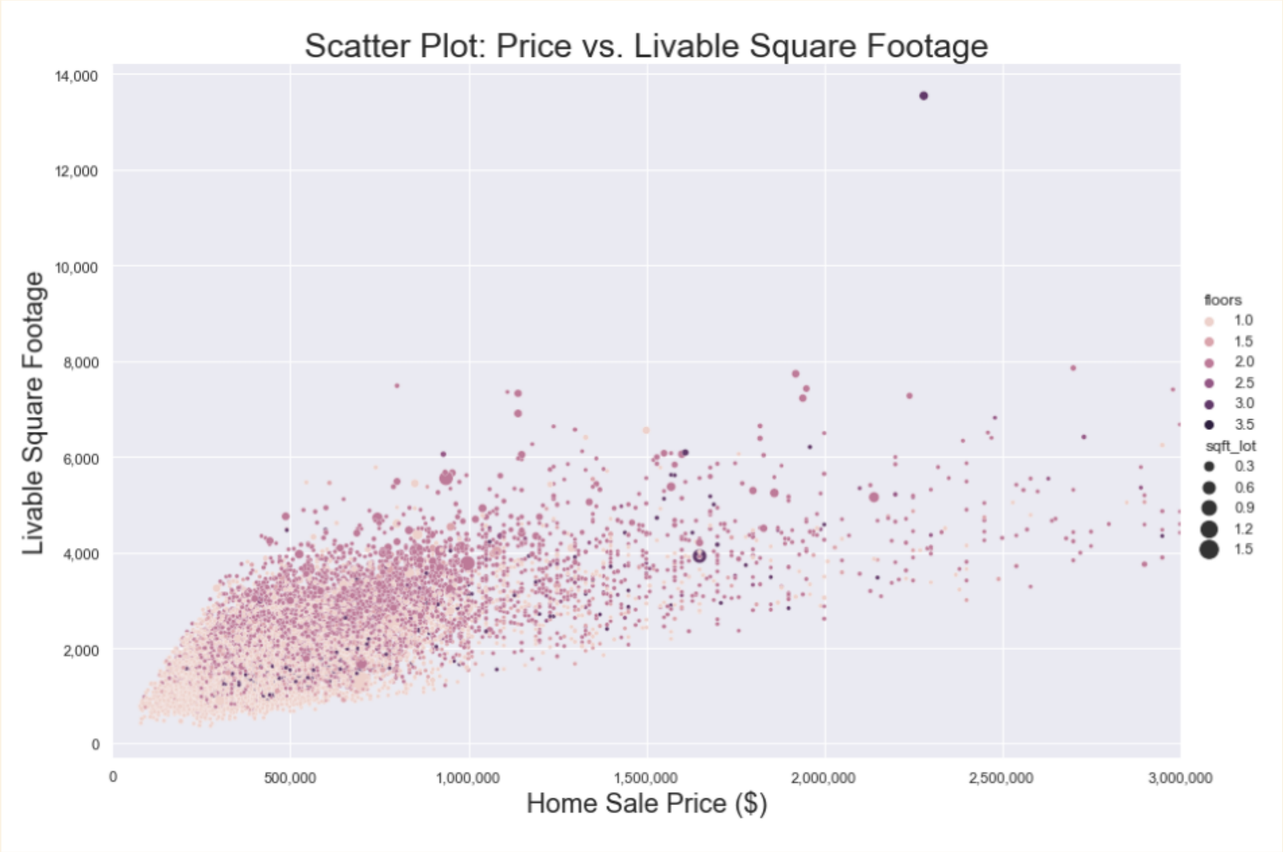
- The data comes from the King County House Sales dataset. The data includes:
  - Home Sale Price
  - # Bedrooms / Bathrooms
  - Sqft of living area / lot
  - Year built / rennovated
  - Sqft of average neighborhood living area / lot
  - Floors
  - Zipcode



Data Modeling

MODEL ONE

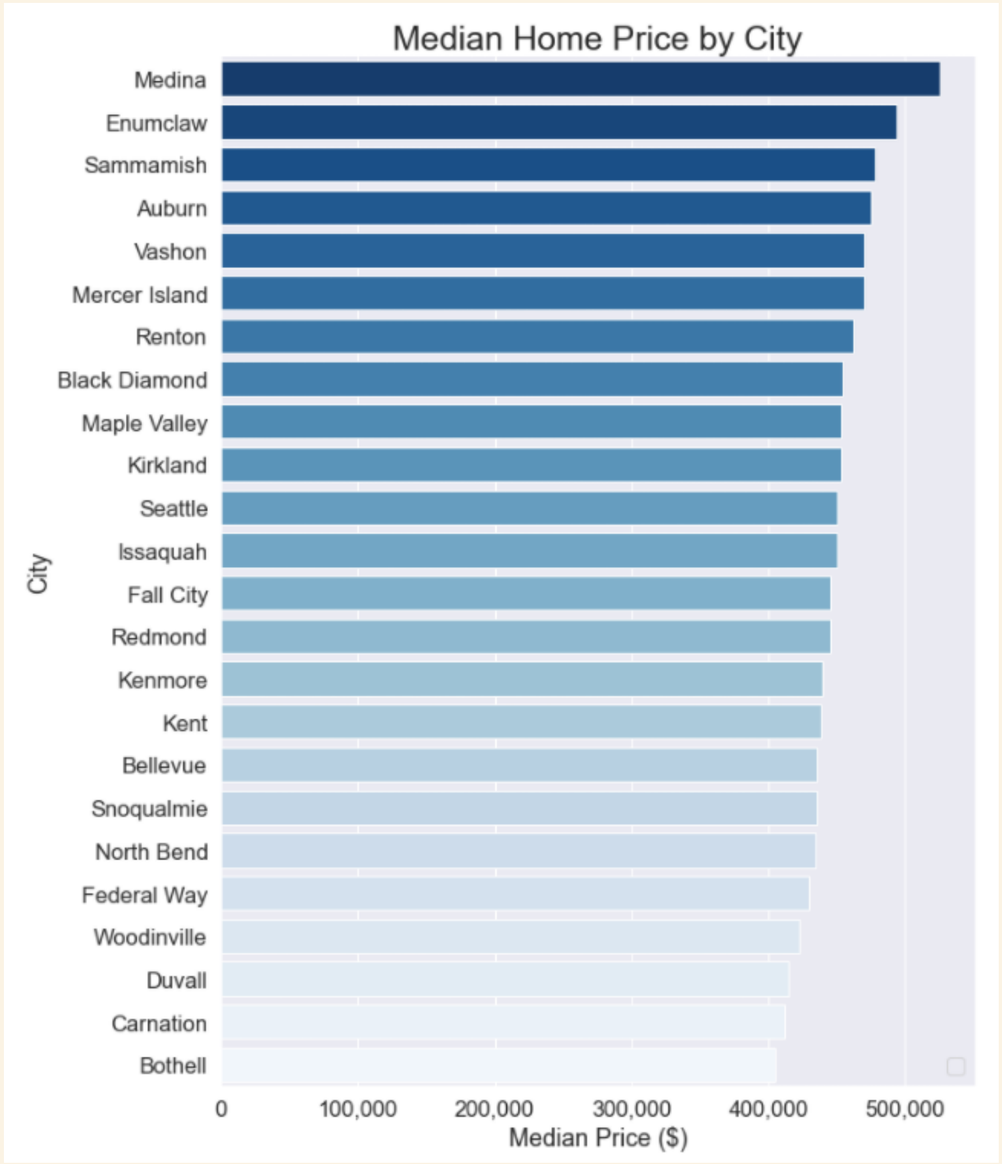
- The first model is based on the variable most correlated to price: the size (in sqft) of the living area.
- This scatter plot includes shading and coloring of the variables: 1) number of floors, and 2) squarefootage of lot.



Data Modeling

MODEL TWO

- The second model built upon the first by:
  - Utilizing the zip code data to add the cities to the dataframe
  - Utilizing most of the variables provided, excluding those with no numerical merit such as "id".



Data Modeling

MODEL THREE

- For the final model (model #3), I applied log transformations to normalize the data, as some variables contained non-normal distributions.

Regression RESULTS

Model 1

- Model 1, based solely off of the livable home square footage, achieved an R-squared of .49, meaning the model explains about 49% of the variance.

Model 2

- Model 2, based on all meaningful variables in the dataset, achieved an R-squared of .71, meaning the model explains about 71% of the variance.

Model 3

- Model 3, based on log-transformed data set variables, achieved an R-squared of .75, meaning the model explains about 75% of the variance.

Regression Results

HOME FEATURES TO ADD

Bathrooms

- According to the final model, adding a bathroom to a home makes that home about $54,800 more valuable.

Size

- Adding more livable space to a home makes that home $132 more valuable for each square foot added.

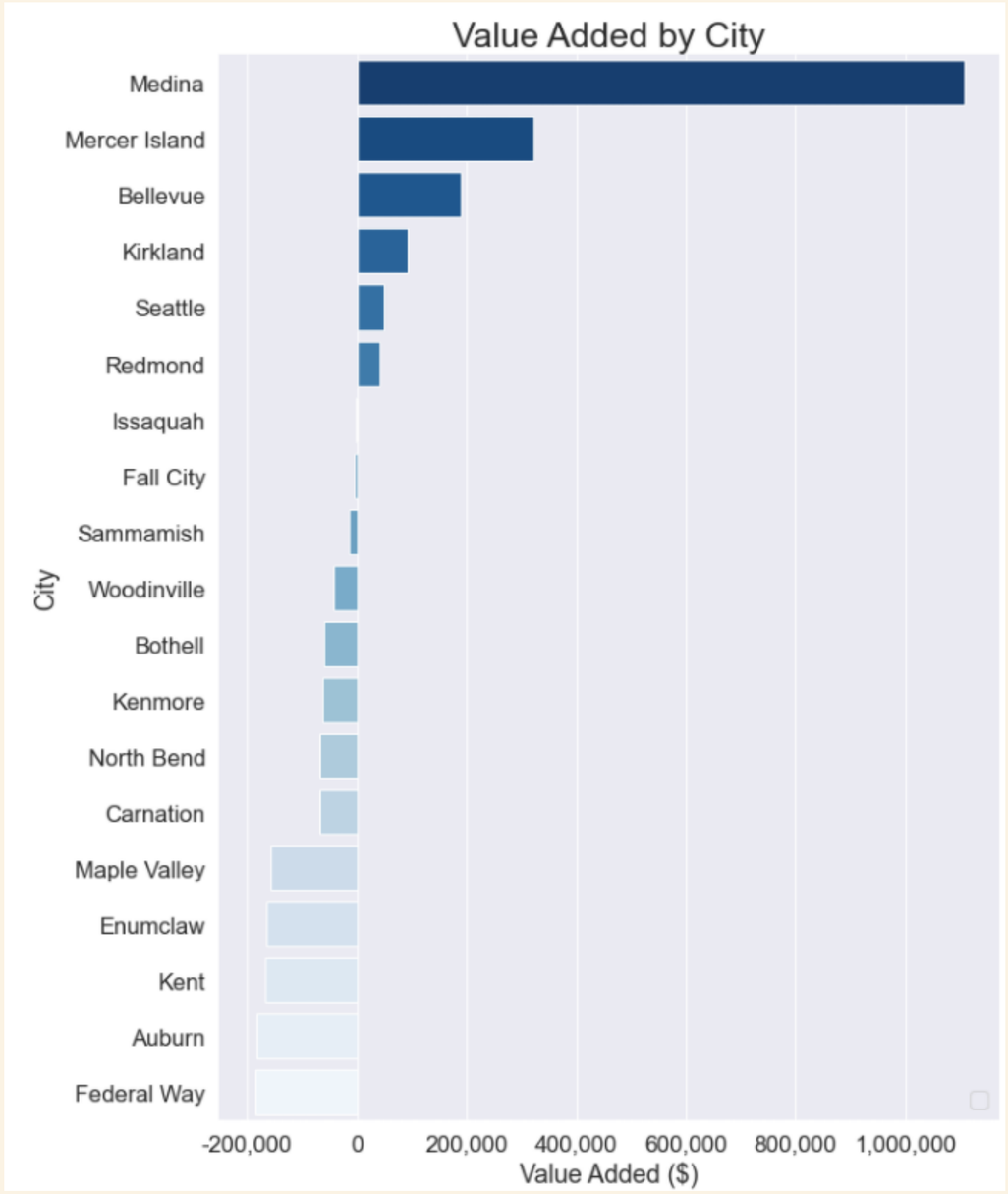
Condition

- Upgrading a home's overall condition (on a scale from 1-10) adds $27,350 to the homes value per increase in scale.

Regression Results

Location

- While changing a home's location is not feasible, location is one of the most important points in determining a home's value.
- This represents how much the city changes the final regression algorithm's value.



Next STEPS

- Accumulate more data! For instance:
  - size and quality of kitchen,
  - size and quality of garden,
  - distance to schools,
  - public transportation,
  - General appliances

- Stratify by year and location
  - It would be interesting to look at how these prices changed through the years, by city.

Thank YOU

Email

TravisCRoyce@gmail.com

GitHub:

@TMCRoyce

LinkedIn

linkedin.com/in/travis-royce