King County Housing Analysis

CeCe Lacey, Warren Umbach, Kyle Vosen

Outline

- Business Problem
- Data
- Methods
- Results

Business Problem

Our goal was to build a model that will help you accurately predict home prices so that you can better serve your clients when helping them buy or sell their home. We would like to investigate how various internal features of a home affect price as well as external features of the larger neighborhood.

Data

We used 3 data sources for this project.

1. Housing data on about 21,000 homes in King County that was given to us

2. Whole Foods location data found on Kaggle

3. Median income by zip code obtained from IPUMS' (Integrated Public Use Microdata Series) National Historic Geographic Information System

Methods

- Cross-Industry Standard Process for Data Mining (CRISP-DM)
    a. Business Understand, Data Understanding, Data Preparation, Modeling, Evaluating, Deployment
- sklearn LinearRegression
- Feature Engineering
- One Hot Encoding
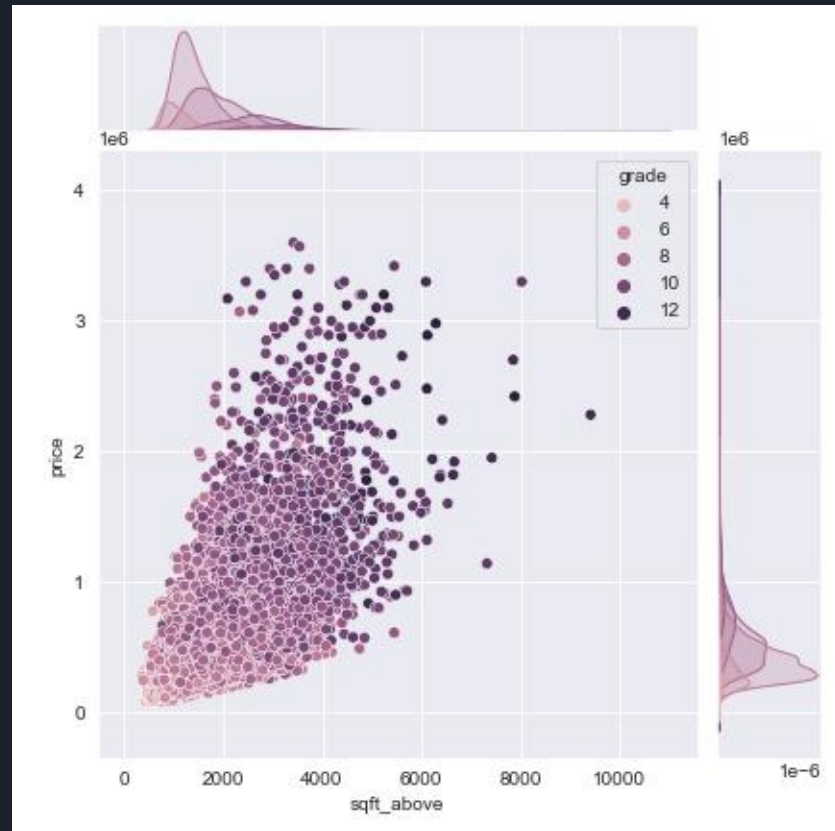- Standard Scaler
- Haversine

Results

Price of all houses in our data set minus outliers.

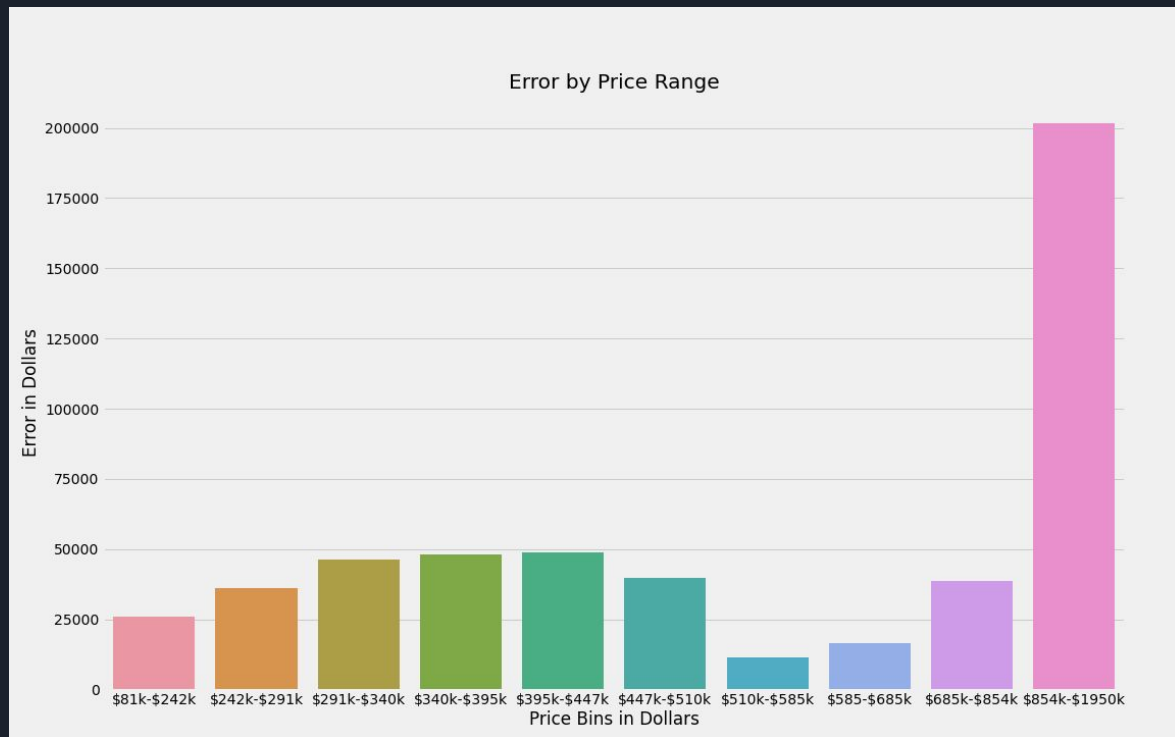Additional- location of Whole Foods

Results

- Comparing price to square footage above ground
- Hue is based on grade scale

Base Model

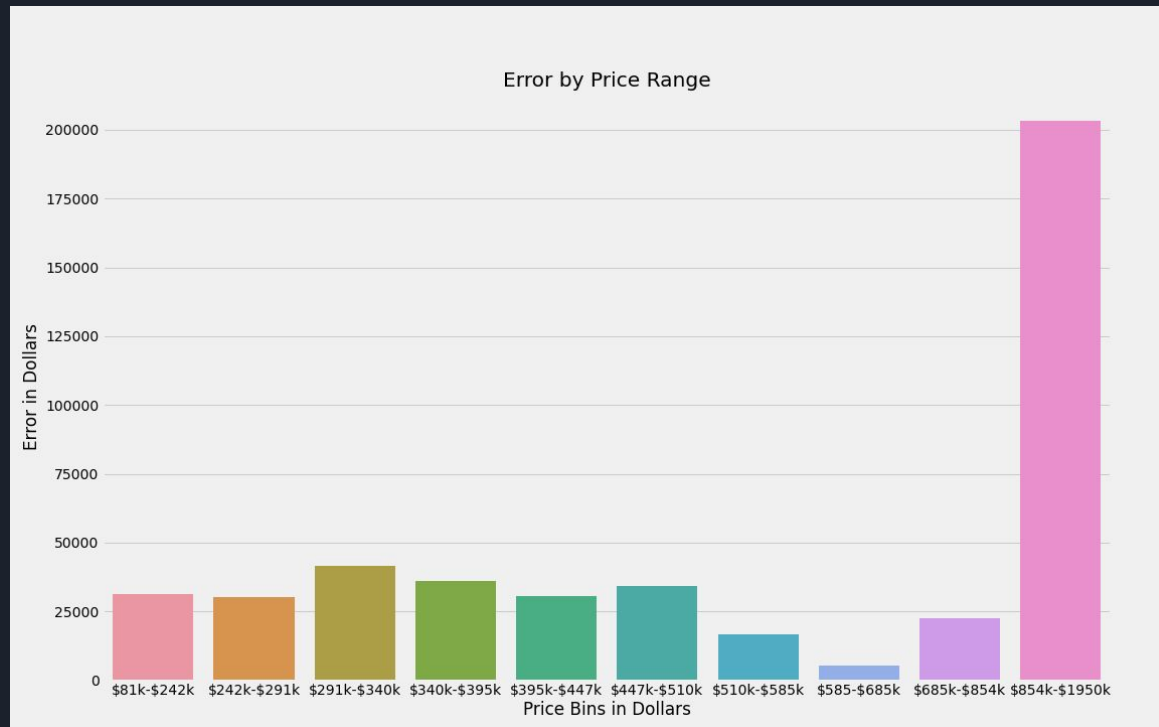## Model Error

- Less than 2.1% at best
- 10.5% on average

Error by Price Range
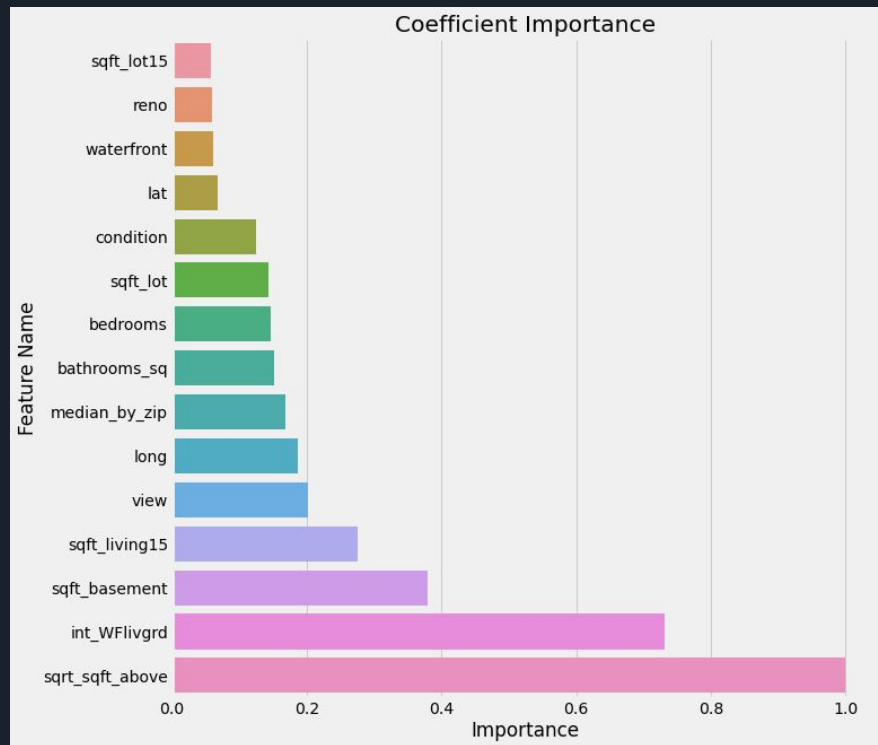
Error in Dollars vs Price Bins in Dollars

Price bins: $81k-$242k, $242k-$291k, $291k-$340k, $340k-$395k, $395k-$447k, $447k-$510k, $510k-$585k, $585-$685k, $685k-$854k, $854k-$1950k

Final Model

Model Error

- Less than 1% at best
- 9% on average

Error by Price Range

Feature Importance

Important features:

- Living space in ft^2
- Distance from Whole Foods
- Housing Grade
- Interactions

Next Steps

Models for individual price ranges

Model with price per square foot as target

More features from outside data like:

- Better census info about:
  - Education
  - income
- Nearest municipal services
  - Nearest bus stop
  - Fire department
  - library
- House features
  - Driveway length
  - fence/ no fence
  - Utility types
  - Internet services

Contact Information

Emails:

kylevosen1999@gmail.com
umbach4@gmail.com
cecelacey@gmail.com

Github:

krvosen
WarrenUm
cclacey