Analyzing King County's Real Estate:
Using Multiple Regression Models to Generate Insights for REIT investors

Tracy Strickel, Jennifer Casias

Silicon Sound Investments
Innovative investments for a sound future

Promising Investments

Identify the most promising real estate investment opportunities.

Provide data-driven insights.

Ensure that our analysis is accurate and reliable.
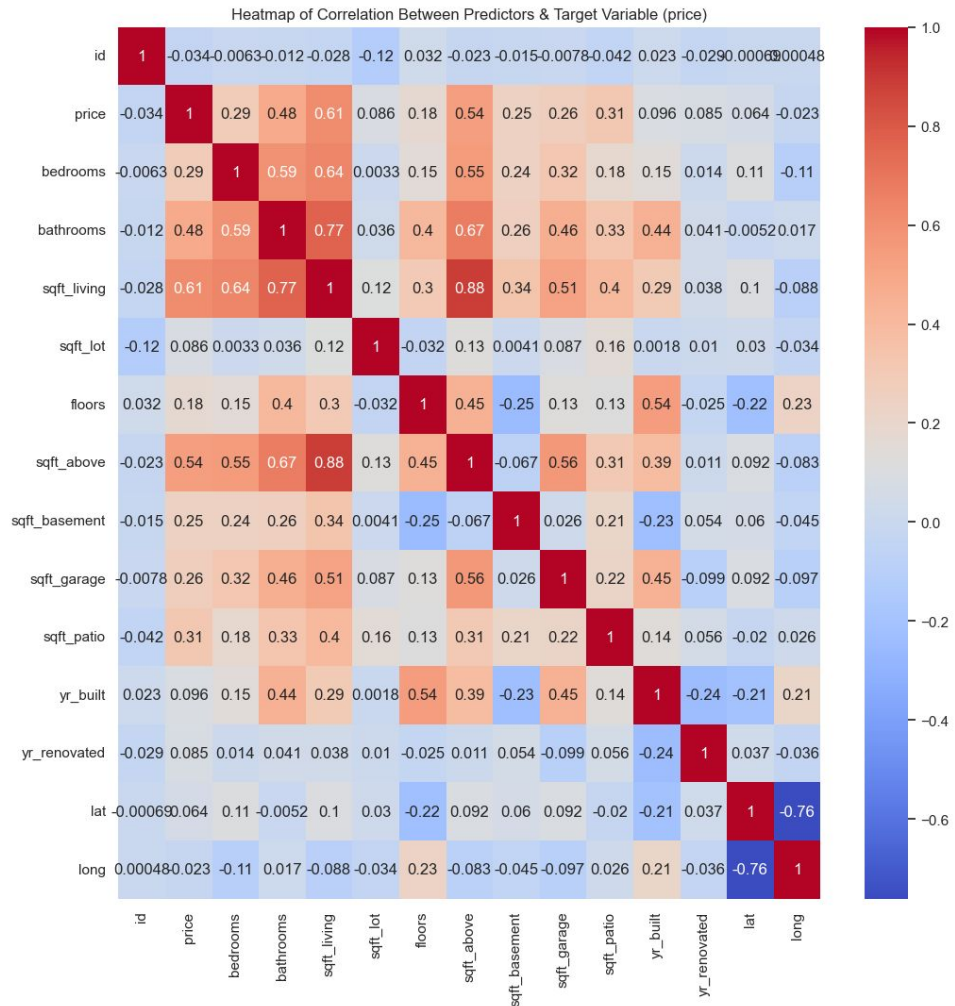
Drive development and growth.

The Data

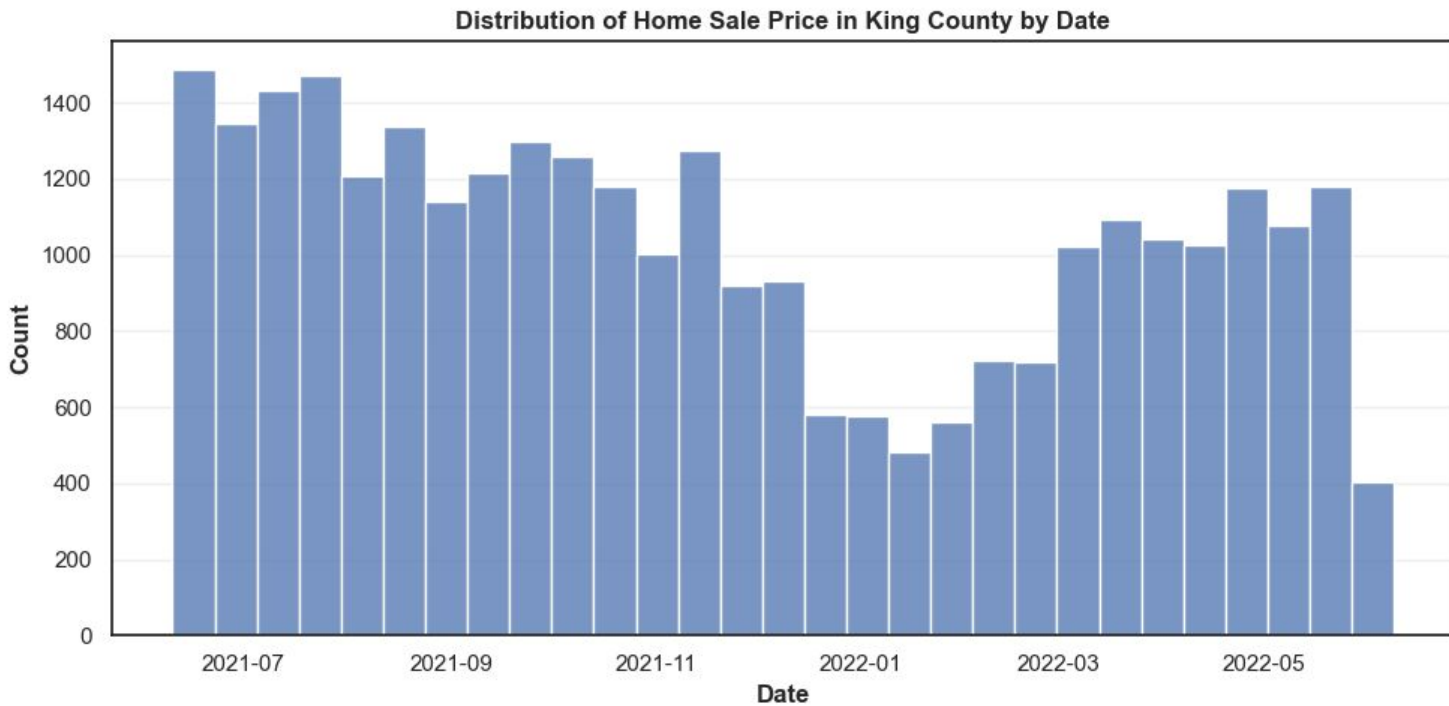**King County's Assessor's Database**

- The dataset contains 30,155 observations with 25 features.
- There are no missing values for the 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'sqft_above', 'sqft_basement', 'sqft_garage', 'sqft_patio', 'yr_built', 'yr_renovated', 'address', 'lat', and 'long' features. However, 'heat_source' and 'sewer_system' features have 32 and 14 missing values respectively.
- The mean sale price of houses in the dataset is around 1.1 million dollars, with a standard deviation of around 900k.
- The mean number of bedrooms is around 3.4, with a minimum of 0 and a maximum of 13.
- The mean number of bathrooms is around 2.3, with a minimum of 0 and a maximum of 10.5.
- The mean square footage of living space in the homes is around 2112, with a standard deviation of around 974.
- The mean lot size is around 16,723, with a standard deviation of around 60,382.
- The median year of sale is 2015.
- The mean house age is 48.1 years, with a standard deviation of 32.1 years.
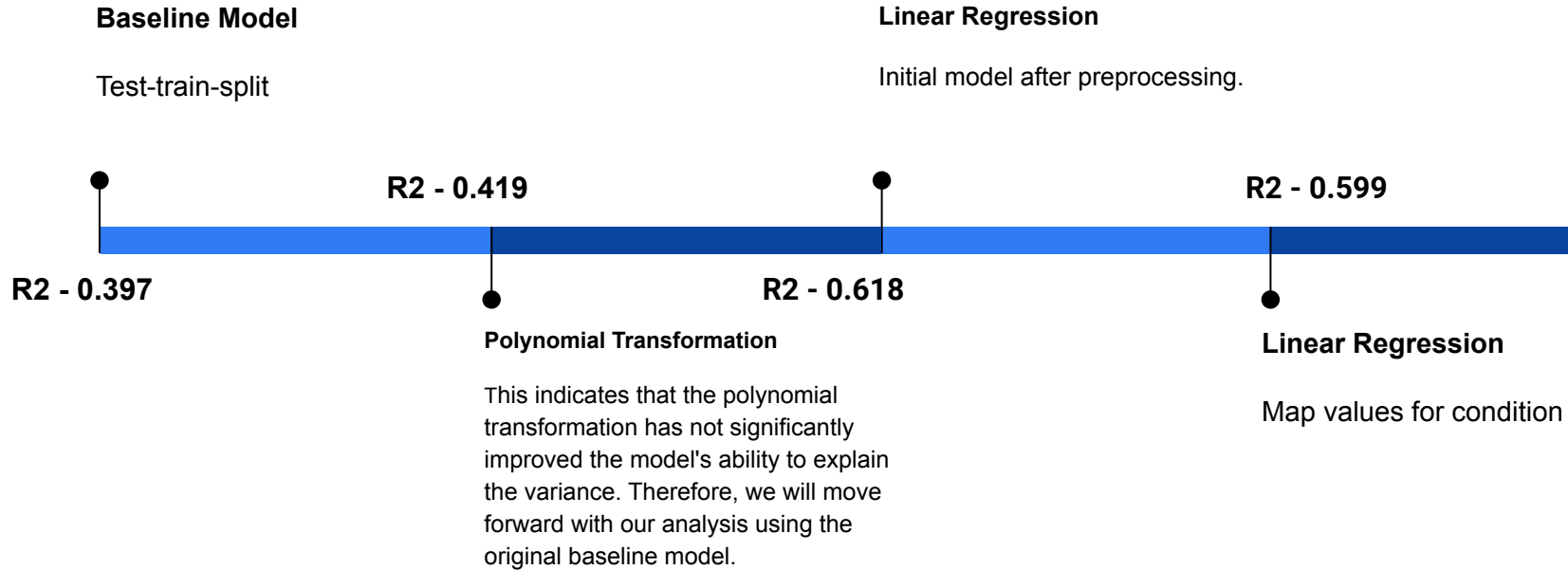
Top Predictors

- Square footage of living space in the home
- Overall grade of the house. Related to the construction and design of the house.
- Square footage of house apart from basement
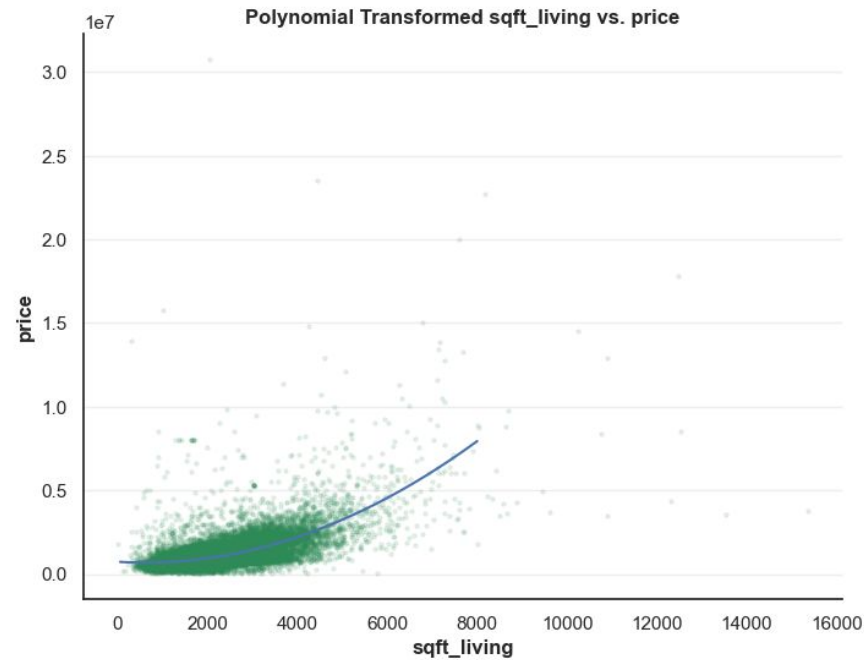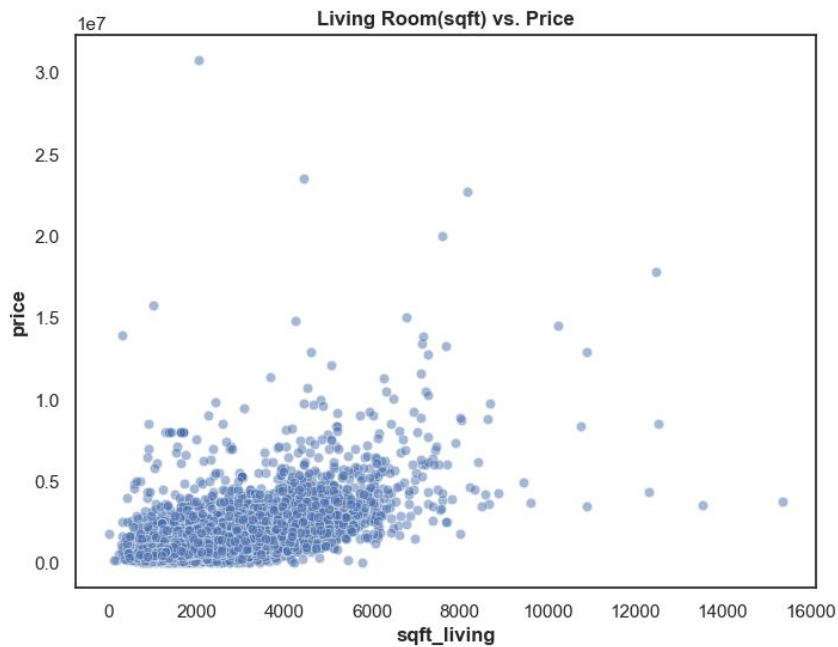- Number of bathrooms
- Quality of view from house

Heatmap of Correlation Between Predictors & Target Variable (price)

Time

Distribution of Home Sale Price in King County by Date

Models

Baseline Model

Test-train-split

Linear Regression

Initial model after preprocessing.

R2 - 0.419

R2 - 0.599

R2 - 0.397

R2 - 0.618

Polynomial Transformation

This indicates that the polynomial transformation has not significantly improved the model's ability to explain the variance. Therefore, we will move forward with our analysis using the original baseline model.

Linear Regression

Map values for condition

Visualization

Models

Linear Regression Model

Map values for bedrooms and bathrooms

R2 - 0.615

R2 - 0.598

Linear Regression Model

Defined columns to scale

Random Forest Model

Broke zip codes apart by dummy variables.

Investments and Potential Returns

R2 - 0.745

Prediction Output

Using the Random Forest Model that can input a set of feature values for a new property, and the model will output a predicted price.

Top Predictors(Random Forest)

Visualize feature scores of the features
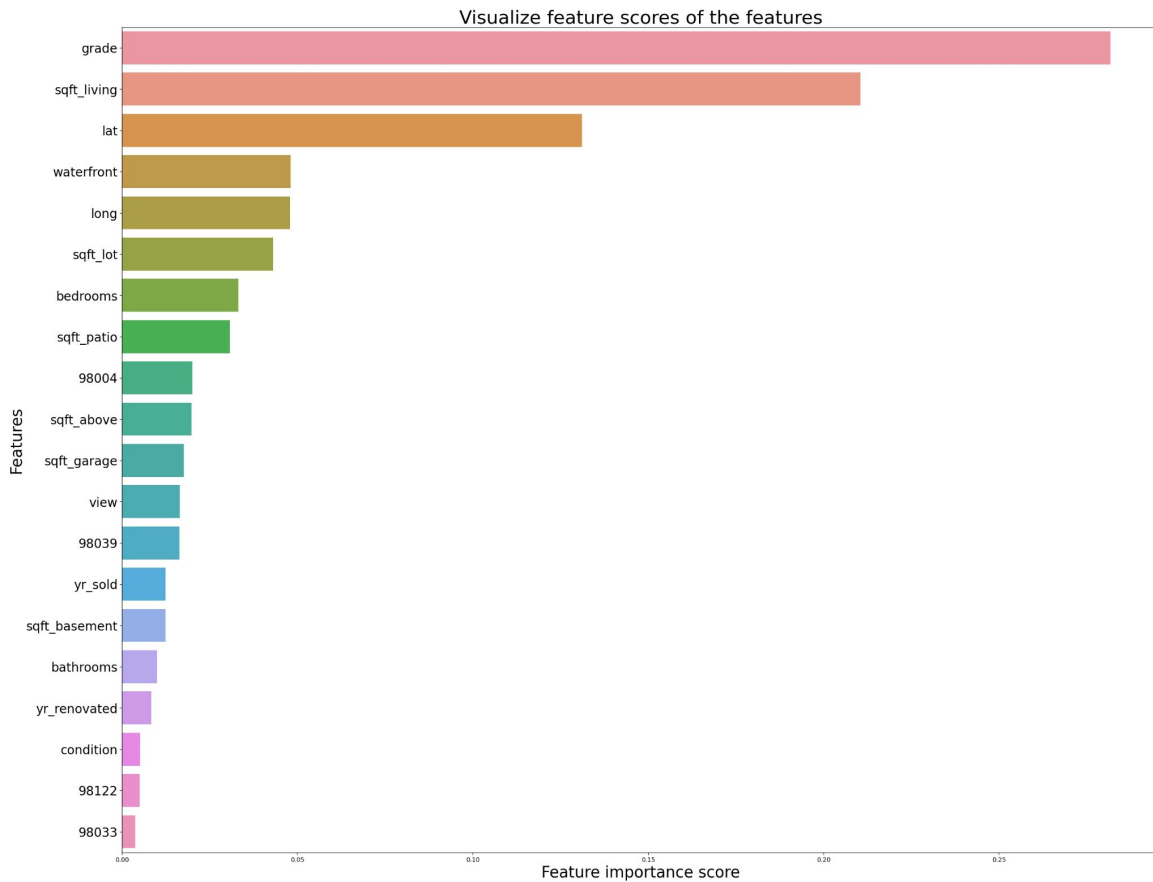
Advance

- Investments and Potential Returns Prediction Output
  - Using Random Forest Model we can input a set of feature values for a new property that will output the predicted price.
- Final step: follow up and iterate
  - Housing market is constantly changing so it is important to periodically update the model with new data
    - Home rentals
    - School districts
    - Workers & home zip code

Reach Out

Tracy Strickel

E-Mail: tracystrickel@gmail.com

LinkedIn: /tracy-strickel-562408139/

GitHub: /hypermoderndragon/

Medium: @tracystrickel

Jennifer Casias

E-Mail: casiasjc@gmail.com

LinkedIn: /jennifercasias/

GitHub: /casiasjc

Medium: @casiasjc

Thank you. Let's stay connected!