Housing Price Per Square Foot Analysis

Using Linear Regression Modeling

By: Nelson Genao

Why Price Per Square Foot (PSF)?

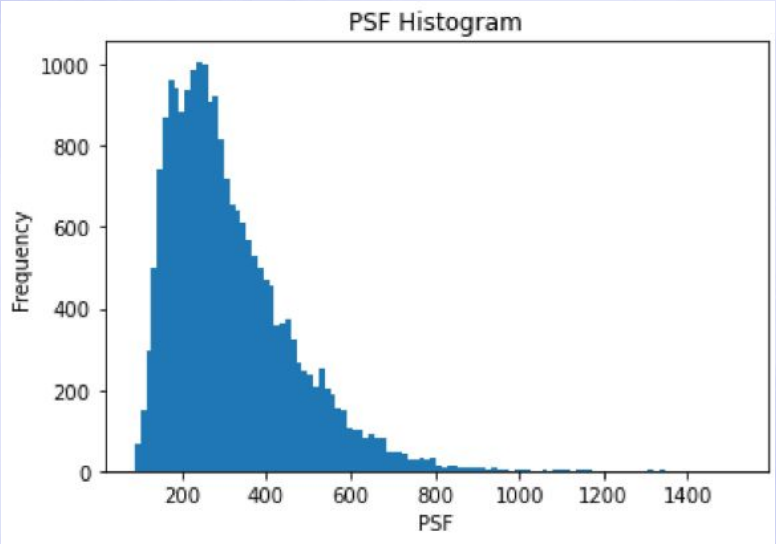
- Popular metric used in the real estate industry
- Useful for quickly comparing the value of similar homes and spaces
- However, should be used in conjunction with other features to make informed decisions

Goal:

- Aiming to find what affects the overall value of your home and how

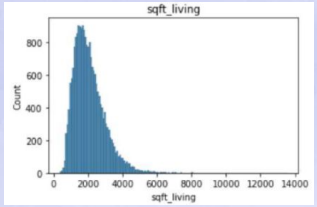
Data

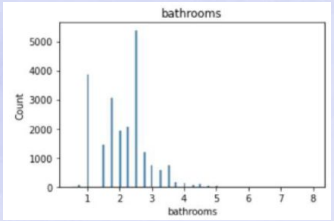
- Used the provided 2015 King's County data set
- Approximately 21,500 homes in final model
- House PSFs range from $80 to $1,500

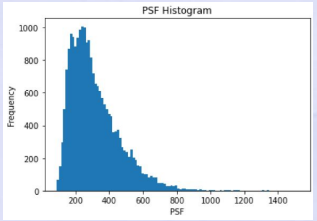


Exploratory Data Analysis

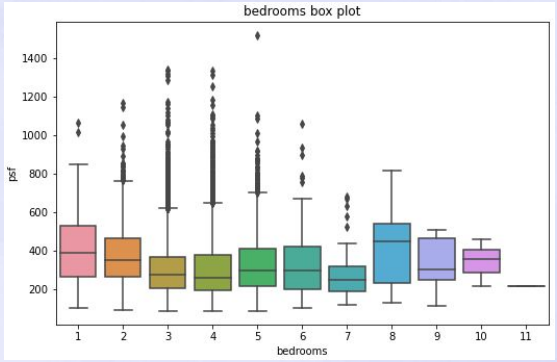
- **Majority of features are categorical such as:**
  - Number of bedrooms, bathrooms or floors
  - Condition and grade of the home
  - Does the home have a basement or by the water?
- **Existence of a right skew**

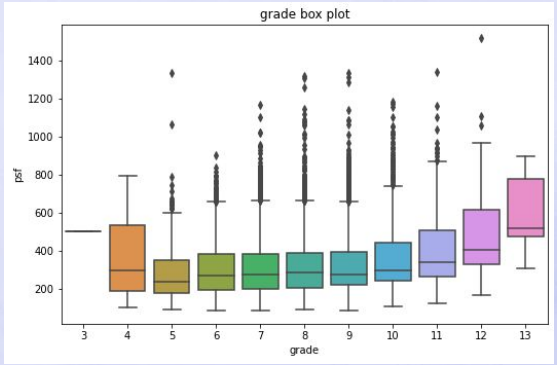


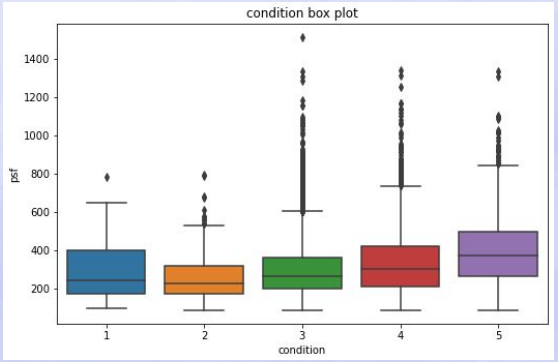




- **Linearity concerns**







Baseline Model

Features:

- Bedrooms, floors, waterfront, condition, grade, year built, bathrooms (dropped)

Modified Features:

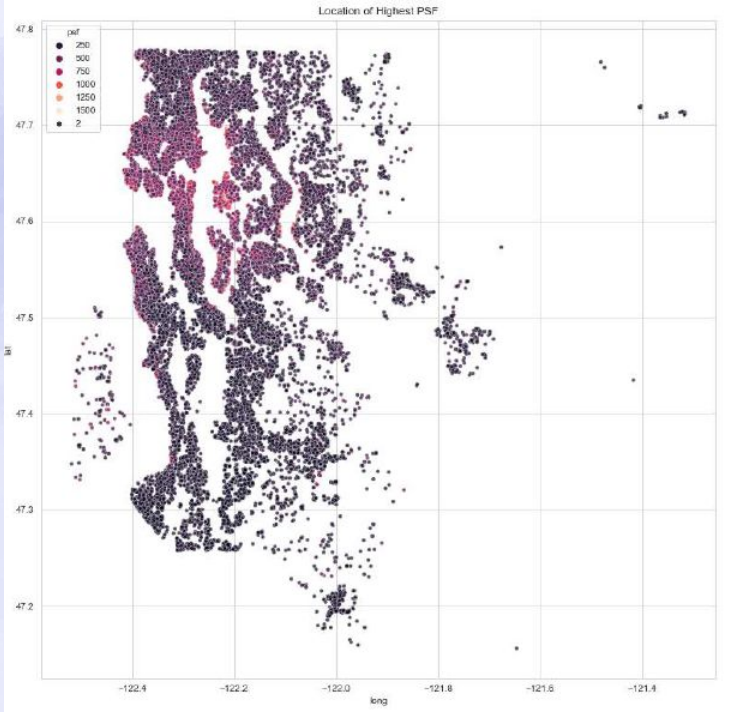
- Living SF to Lot SF ratio
- Having a basement
- Renovations within 27 years

Model Summary:

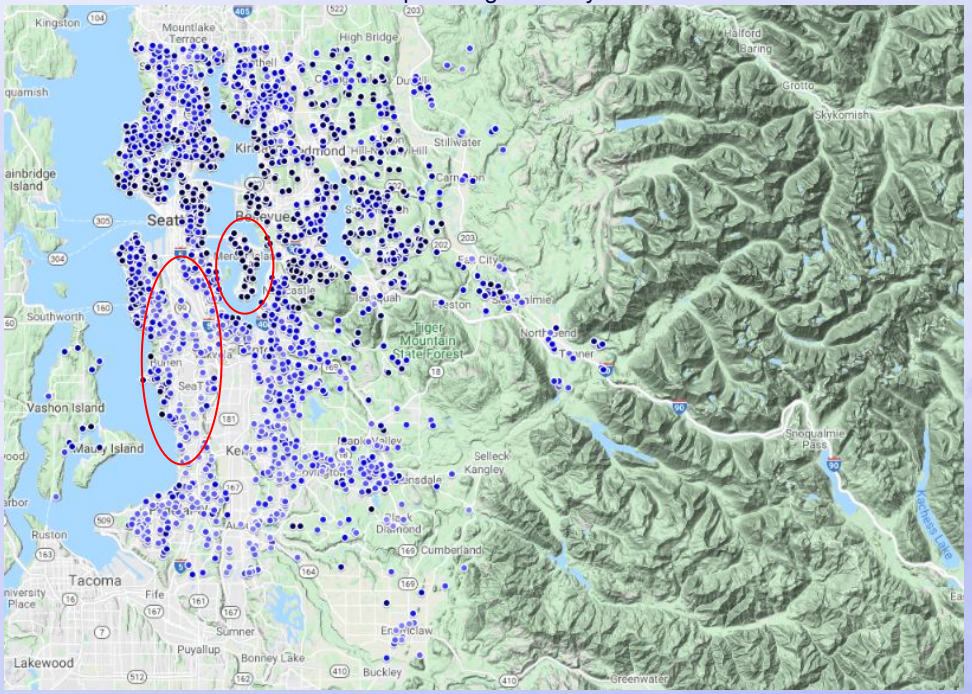
- **R^2** = 0.424
  - Features explained 42% of the variation found in PSF
- **RMSE =** $111.58
  - For comparison, mean PSF is $314.90 and STD PSF is $147.43
- **Most Impactful Features:**
  - Waterfront = $271.61
  - Living to Lot Ratio = $133.46
  - Floors = -$39.47
  - Bedrooms = -$29.94
- **Least Impactful Features:**
  - Year Built = -$1.98

Improvements?

-Exploration of geography and location
-Added back zip codes



Map of King's County



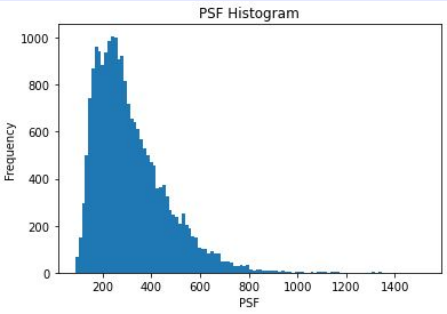
Final Model - Using Zip Codes and Log Transformation

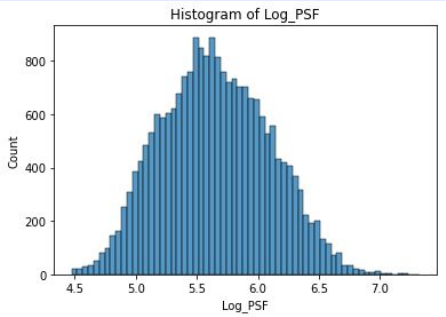
Features:

- Bedrooms, floors, waterfront, condition, grade, zip codes (new), year built (dropped)

Modified Features:

- Living SF to Lot SF ratio
- Having a basement
- Renovations within 27 years



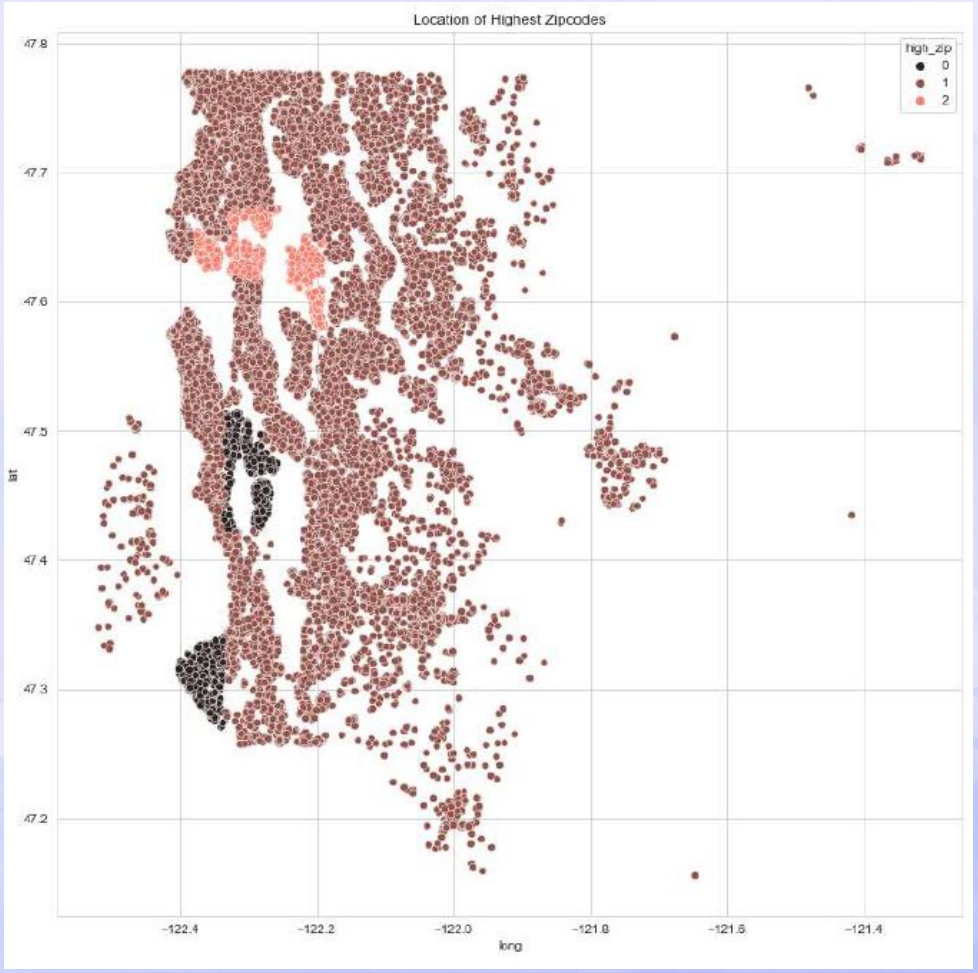


Model Summary:

- **R^2** = 0.76
  - Features explained 76% of the variation found in the log of PSF
- **RMSE** = $77.17
  - For comparison, mean PSF is $314.90 and STD PSF is $147.43
- **Most Impactful Features:**
  - Zip Code = Upwards of 146 to 240%
  - Waterfront = 97%
  - Basement = 27%
- **Least/Worst Impactful Features:**
  - Grade = 4%
  - Bedrooms = -6%
  - Living to Lot Ratio = -10%
  - Floors = -11%

Conclusion

- **LOCATION!**
  - Zip Codes
    - More expensive zip codes include: 98004, 98039, 98109, 98119
    - Cheaper zip codes include: 98148, 98188, 98168, 98023
    - Cheaper zip codes found near airports
  - Waterfront
- More open spaces in their home
- Less floors



Next Steps

- Explore additional location features
  - Proximity of schools, hospitals, parks, monuments, fire department, and police stations
- History of the home
  - Previous fire, flood, or any other damage
  - Presence of termites, or mold
  - Repair or renovation costs
    - Any costs taken into account when negotiating price of the home
- Improve assumptions
  - Features were loosely linear related
  - Residuals could be more normal and less heteroscedastic
    - Model underpredicted high PSF and overpredicted low PSF

Thank you!

github.com/NelGen/NG-Housing-PSF-Modeling-Project

maps.google.com