



Housing Metrics

King's County, WA

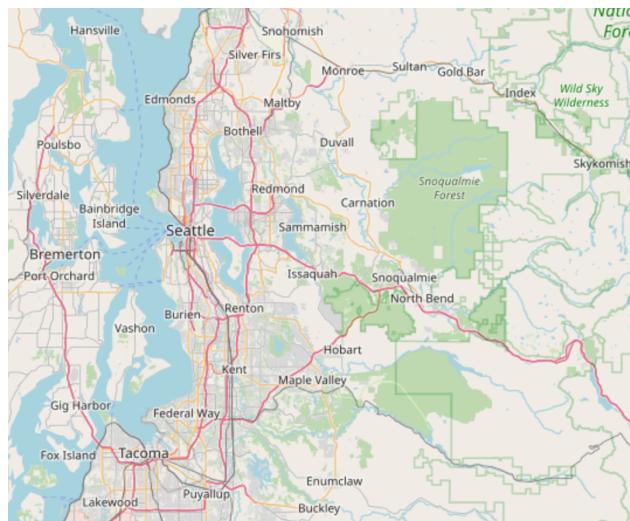
1. Objectives

- Use King's County Housing Data to build a linear model to predict housing price
- Identify the major factors affecting housing price evidenced in the data
- Recommend residential development variables to ensure highest property price on sale

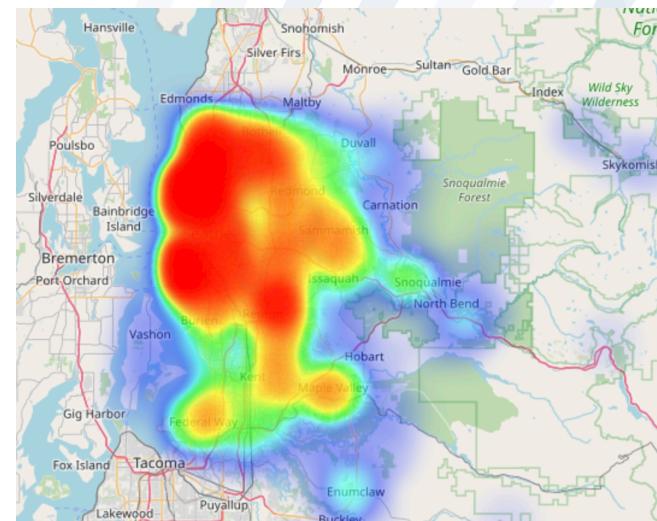
The Data:

- 20,000 sales in the years 2014/2015
- Concentrated around Seattle, WA
- Includes data including:
 - Square footage
 - Zip Code
 - # Bedrooms and Bathrooms
 - Waterfront view, etc.

Seattle



Heatmap of Sales



The OSEMN Process

Data Science OSEMN Model

Obtain

- from other location
- Query from database or API
- Extract from another file
- Generate data (e.g. Sensors)



Scrub

- Filtering lines
- Extracting columns or words
- Replacing values
- Handling missing values
- Converting formats



Explore

- Understanding data
- Deriving statistics
- Creating visualization



Model

- Clustering
- Classification
- Regression
- Dimensionality reduction



INterpret

- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result



Adapted from: [KD Nuggets](#)

2. The Baseline Model

Training Scores:

R2: 0.6971857664620761

Root Mean Squared Error: 198517.72757005875

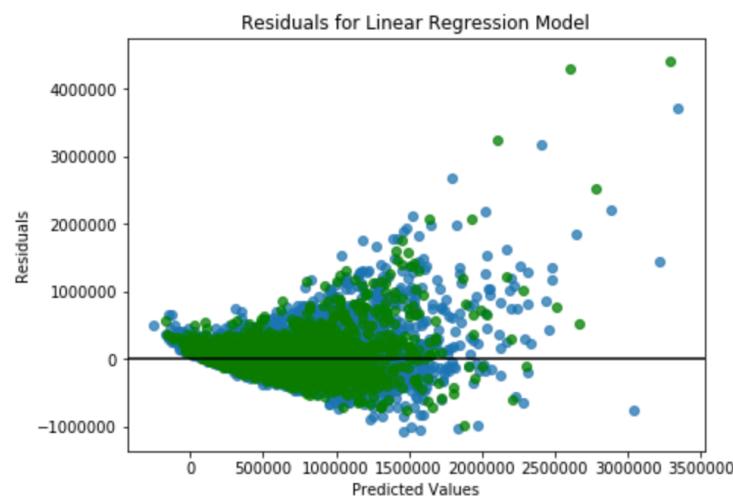
Mean Absolute Error: 127137.95303547624

Testing Scores:

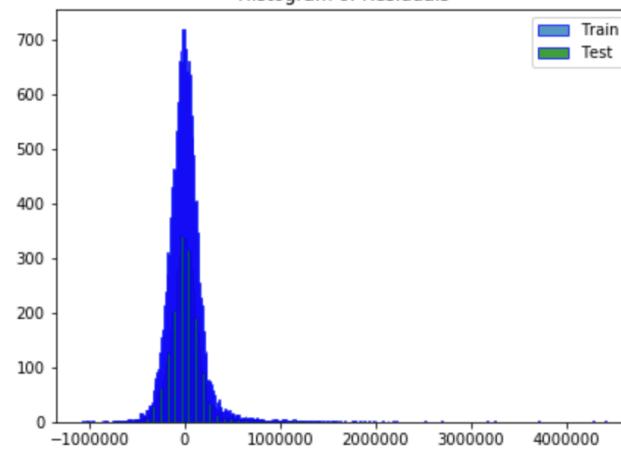
R2: 0.6770320248473616

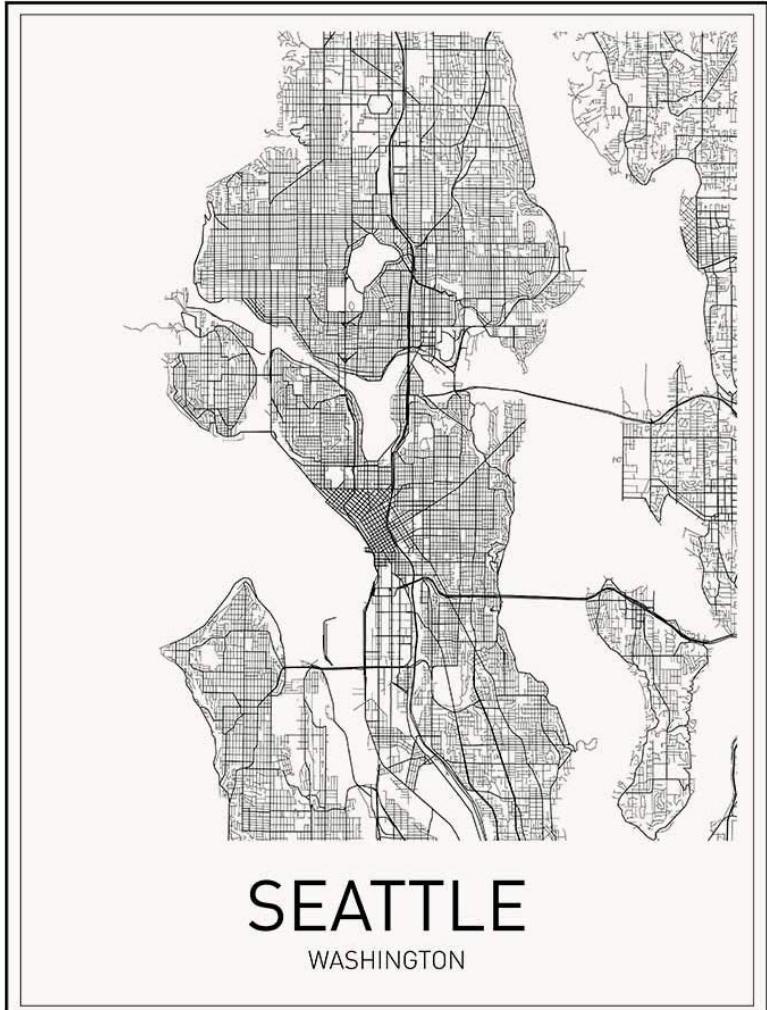
Root Mean Squared Error: 219645.4675533081

Mean Absolute Error: 129729.80964818248



Histogram of Residuals



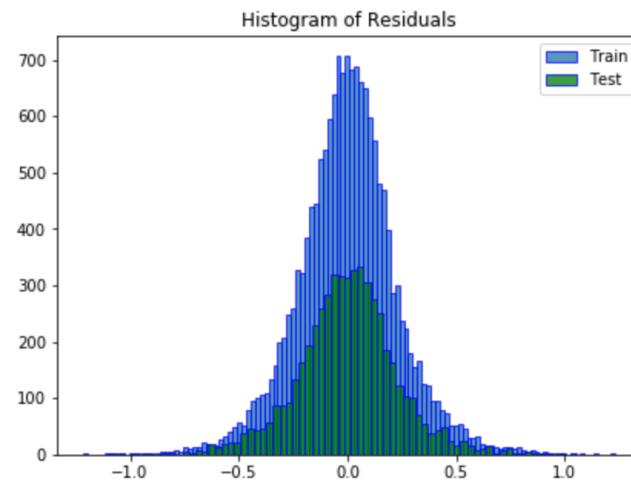
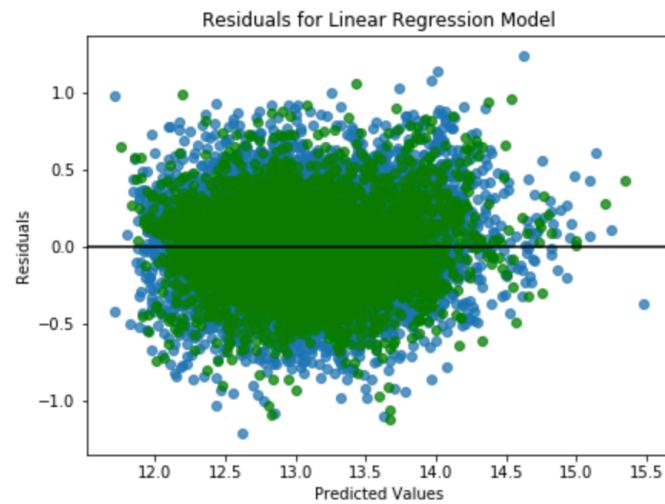


Transformations

- Log transformed price and square footage
- Sorted data by house type (apartment, bungalow or house)
- Organized zip codes into townships.
- Recent construction (built or renovated in last 10 years)

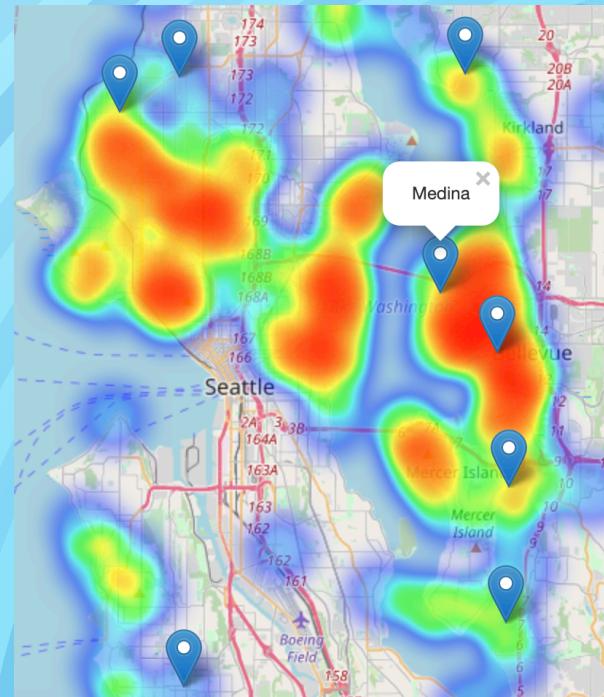
3. Final Model

```
Training Scores:  
R2: 0.8056608461532331  
Root Mean Squared Error: 180610.40224094037  
Mean Absolute Error: 98765.14059214207  
---  
Testing Scores:  
R2: 0.806408576447123  
Root Mean Squared Error: 184639.6900987637  
Mean Absolute Error: 100104.92054029983
```



Major Factors in Price

- Square Footage (0.43)
- If it's in Medina (0.21)
- How far North (0.15)
- Waterfront (0.15)
- If it's in Mercer (0.13)
 - Only 50 Medina sales in data
 - 13 of which are in the top 1% of price by square foot.



Top 1% of cost per sq. ft.

4. Interpretation

To maximize sale price, build homes:

- In Medina or Mercer
- Waterfront View
- As large as possible



Areas of Further Study

- Up and coming neighborhoods
- Factor in quality of neighborhoods
- Housing taxes
- Identify areas of Seattle with affordable land for development

Thank You!

Andrew Mooney

andrew.k.mooney@gmail.com