Module 2 Project:

EVAN HOLDER

Modeling Housing Prices Around Seattle

- Why?

Approach

- Create a model using the King County Housing dataset (Seattle & Seattle metro area)

- Answer these questions:
  - Which housing features drive housing prices?
  - How can I get the best and worst deals?
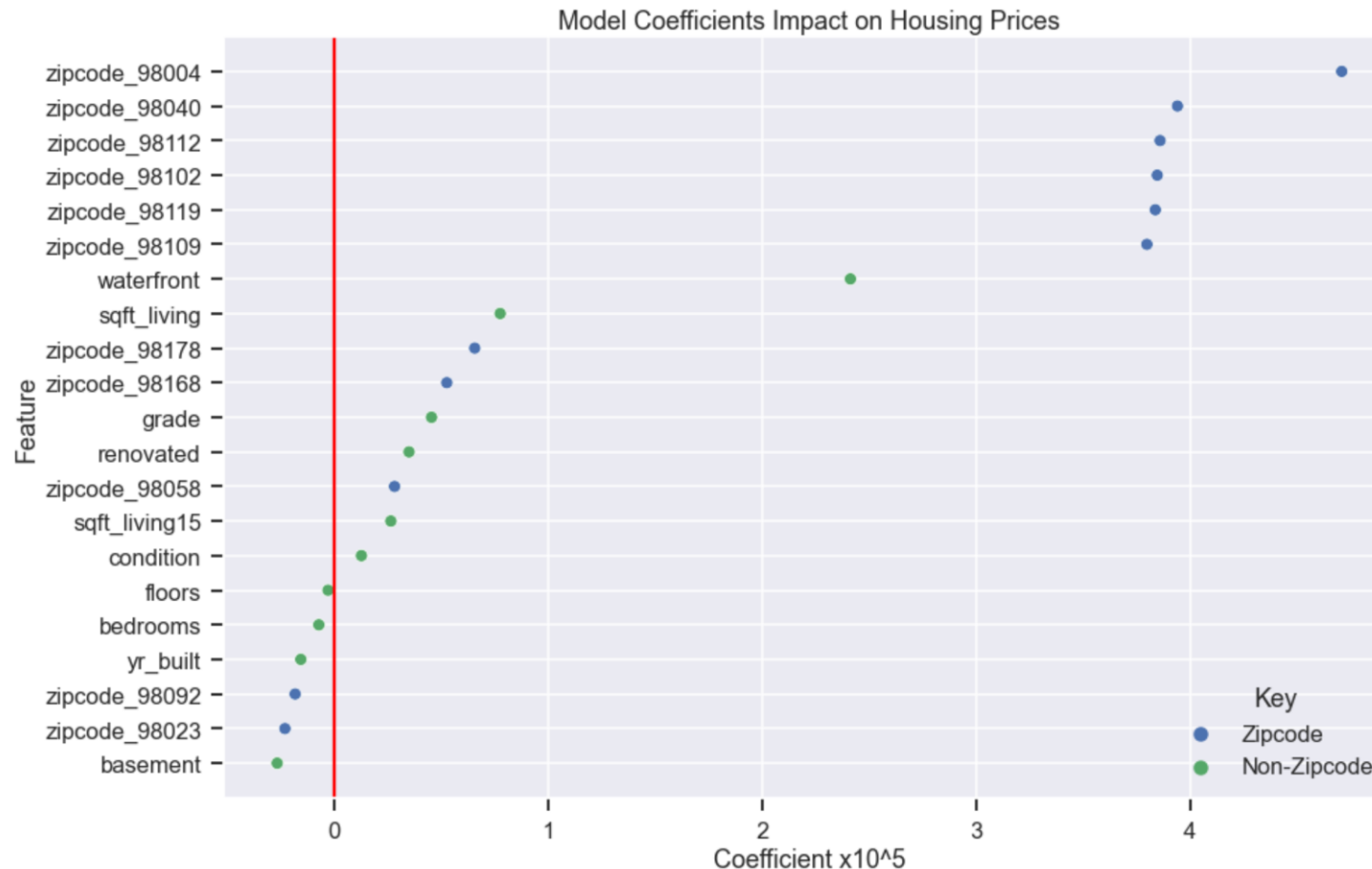
The Model: Features

- <u>Predictors</u>:

  - Bedrooms
  - Bathrooms,
  - Square feet
  - Floors
  - Waterfront

  - Condition
  - Grade
  - Year built
  - Square feet of nearest 15 neighbors

  - Renovated
  - Basement
  - Zipcode

- <u>Outcome Variable</u> : Price

The Model: Driving Features

Model Coefficients Impact on Housing Prices

Rank	Feature	Coefficient
1	zipcode_98004	6.680633
2	waterfront	2.413852
3	sqft_living	0.777310
4	grade	0.456617
5	renovated	0.351373
6	sqft_living15	0.266576
7	basement	-0.264439
8	yr_built	-0.154039
9	condition	0.129015
10	bedrooms	-0.069910
11	floors	-0.026719

The Model: Bedrooms

Model Coefficients Impact on Housing Prices
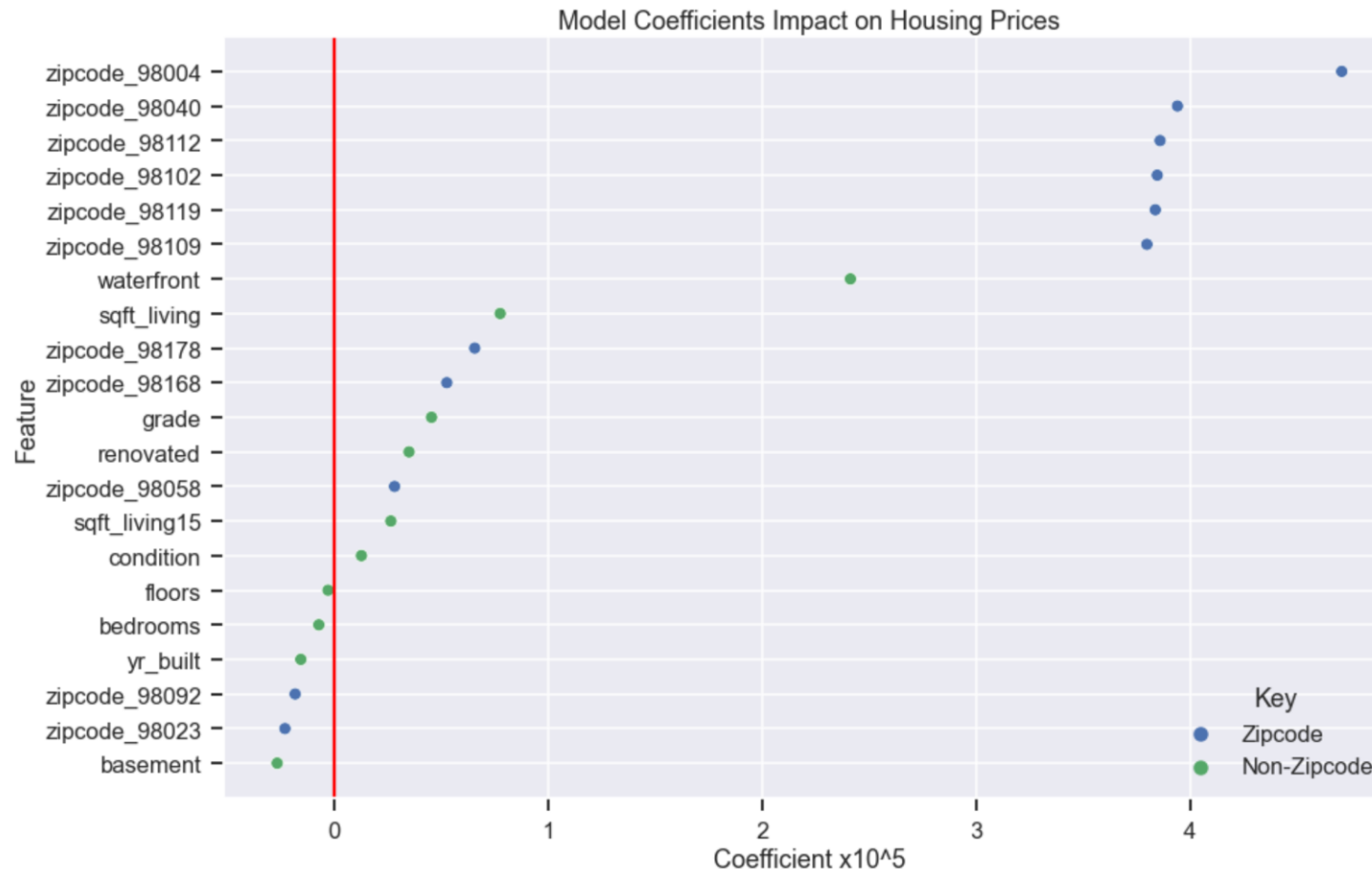
Rank	Feature	Coefficient
1	zipcode_98004	6.680633
2	waterfront	2.413852
3	sqft_living	0.777310
4	grade	0.456617
5	renovated	0.351373
6	sqft_living15	0.266576
7	basement	-0.264439
8	yr_built	-0.154039
9	condition	0.129015
10	bedrooms	-0.069910
11	floors	-0.026719

The Model: Bedrooms and Sqft
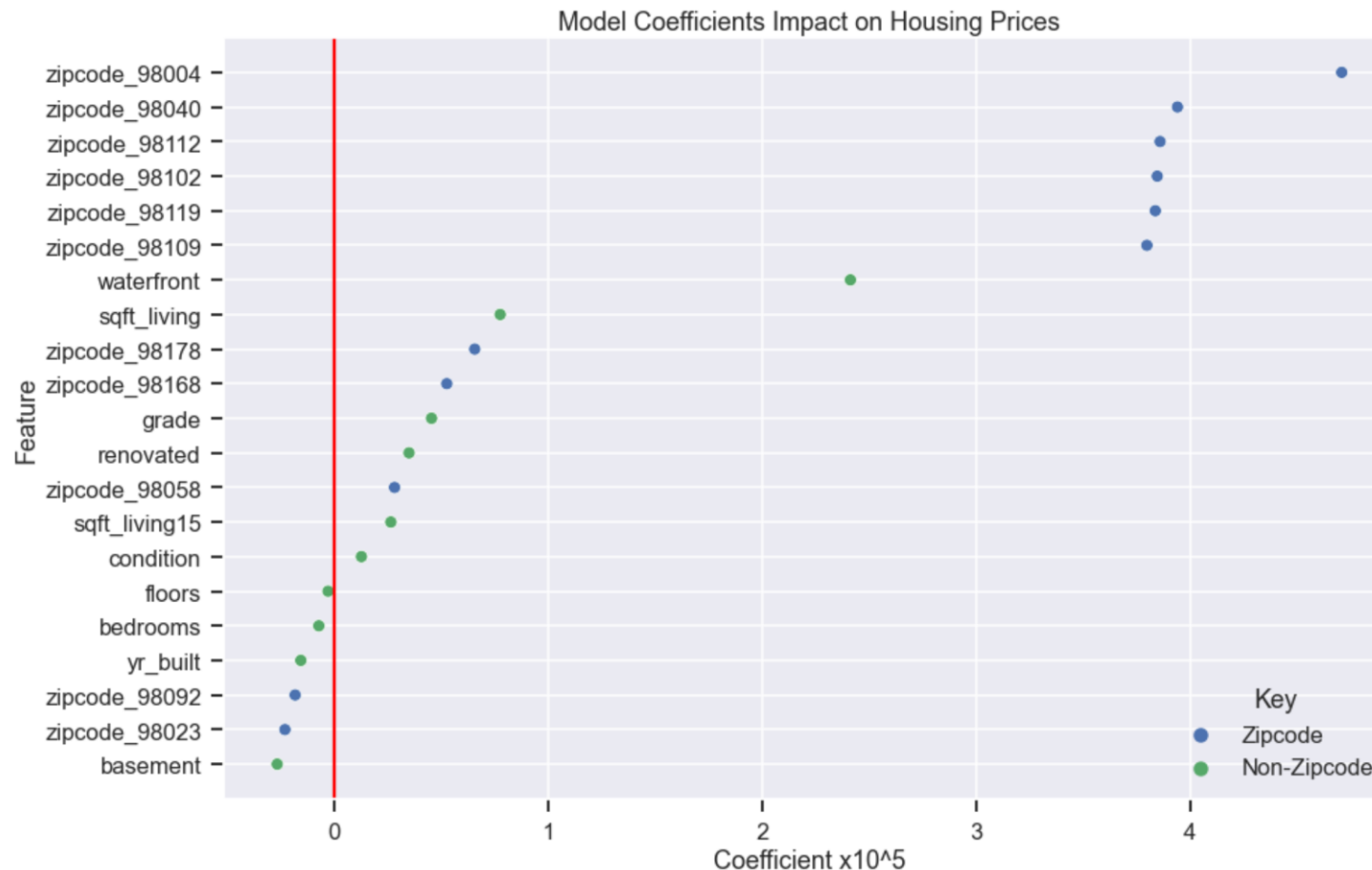
Model Coefficients Impact on Housing Prices

Rank	Feature	Coefficient
1	zipcode_98004	6.680633
2	waterfront	2.413852
3	sqft_living	0.777310
4	grade	0.456617
5	renovated	0.351373
6	sqft_living15	0.266576
7	basement	-0.264439
8	yr_built	-0.154039
9	condition	0.129015
10	bedrooms	-0.069910
11	floors	-0.026719

Model: How well does the it fit the data?

- "Goodness of fit" = percent of errors that we can explain (R-squared)

Price v. Errors

Model: How well does the it fit the data?

- "Goodness of fit" = percent of errors that we can explain (R-squared)

- 81%

Price v. Errors

Model: How well does the it fit the data?

- "Goodness of fit" = percent of errors that we can explain (R-squared)

- 81%

- Model does not perform well where prices are very high and low.

Price v. Errors

The Model: Best/Worst Deals

- Top 100 "best deals" (data points with the most negative errors) v. dataset averages
  - Square Feet, 15 Closest Neighbors Square Feet, Grade, Renovated, Waterfront
  - Zipcodes: 98033, 98115, 98118, 98103, 98006

- Top 100 "worst deals" (data points with the most positive errors) v. dataset averages
  - Renovated, Basements, Waterfront
  - Zipcodes: 98103, 98006, 98033, 98115, 98112, 98177

- Common Attributes
  - Renovated, Waterfront
  - Zipcodes: 98033, 98006, 98115

Summary / Recommendations

- The most important indicator of housing prices is location

- If you are looking for a good price on a house, look for something with a lot of square feet and but yet fewer bedrooms, avoid waterfront properties or any properties that have been newly renovated

Future Work

- Expand the model to include data from larger geographical regions.

The End

- Any questions?