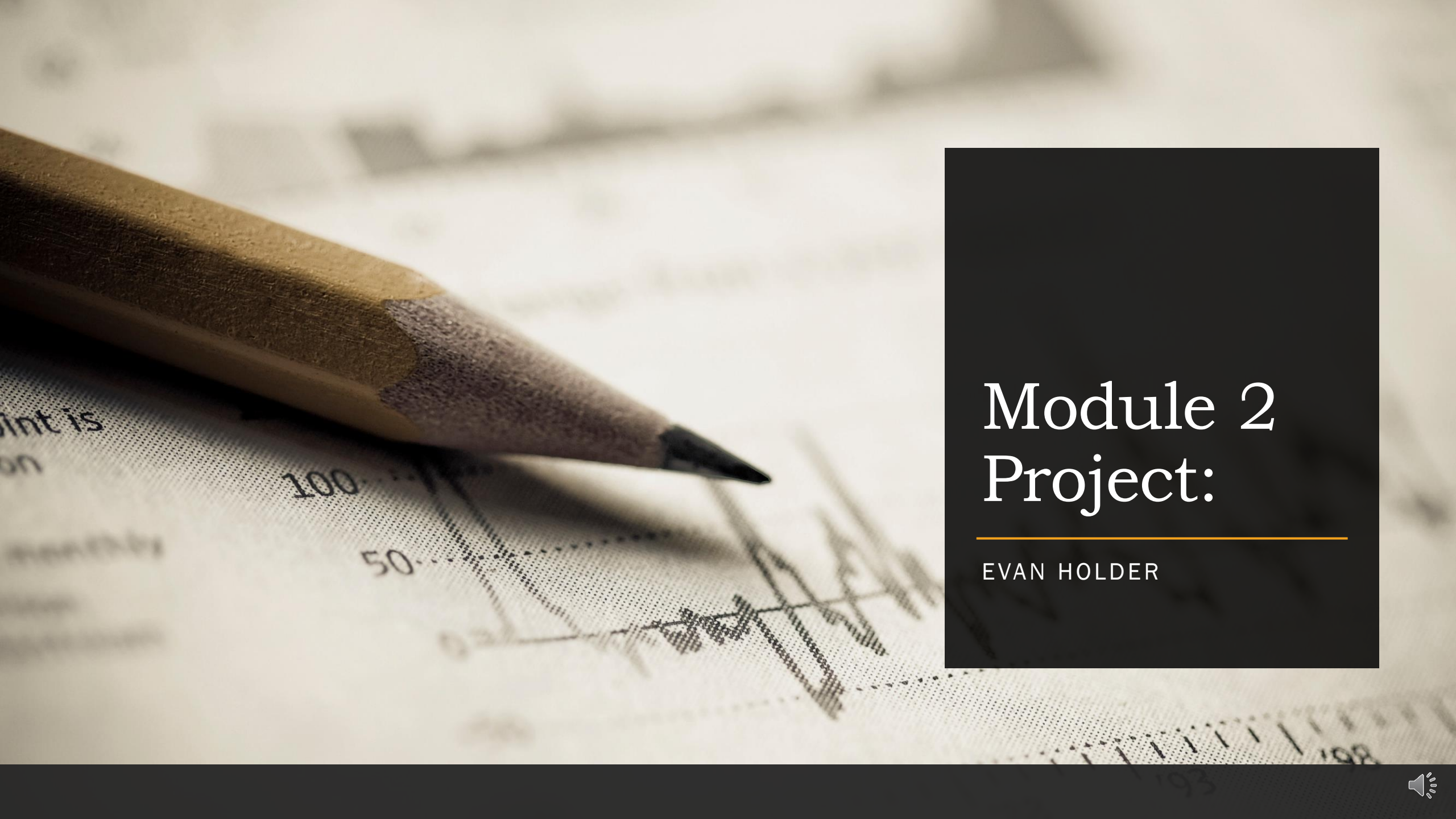


Module 2 Project:

EVAN HOLDER

Modeling Housing Prices Around Seattle

- Why?

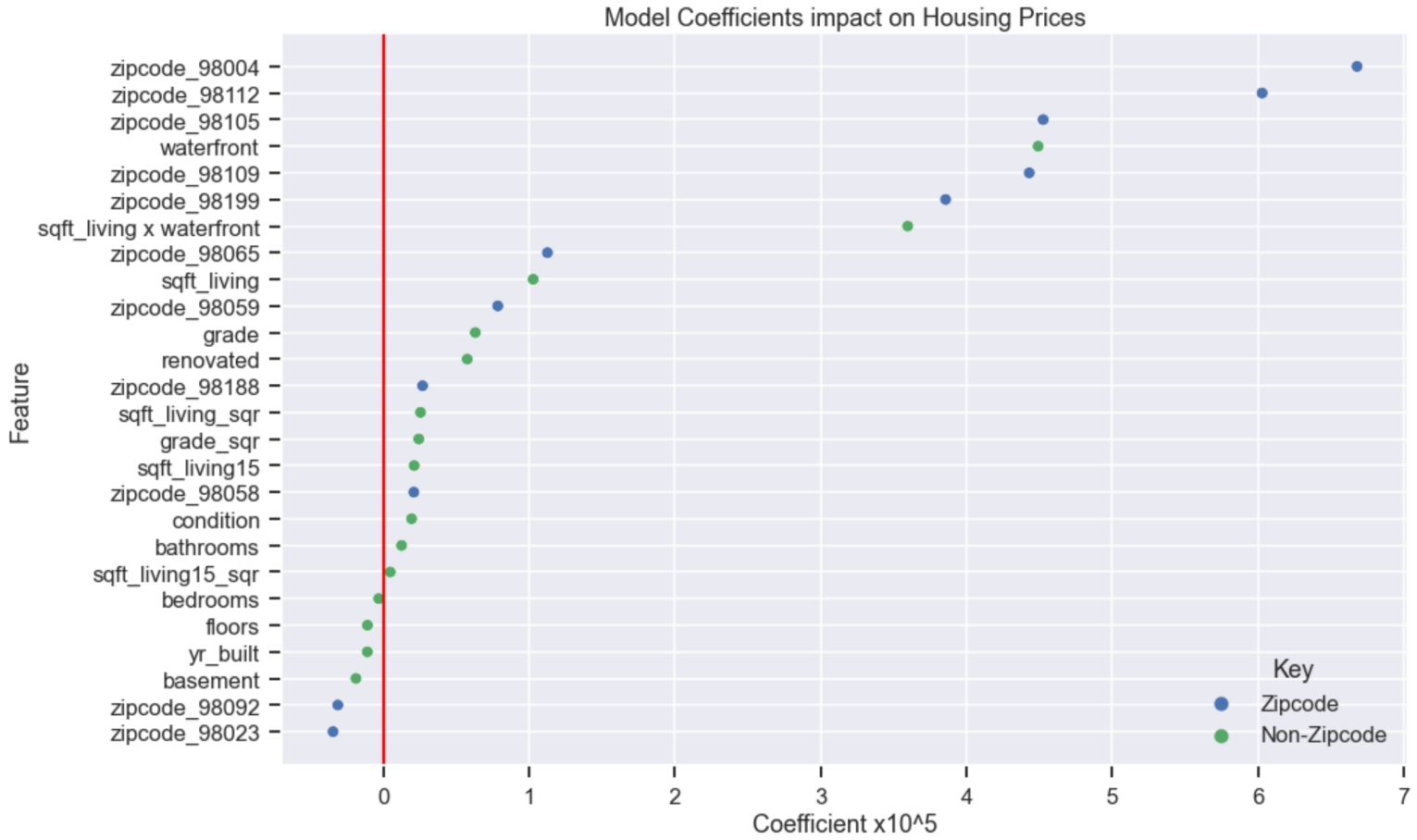
Approach

- 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

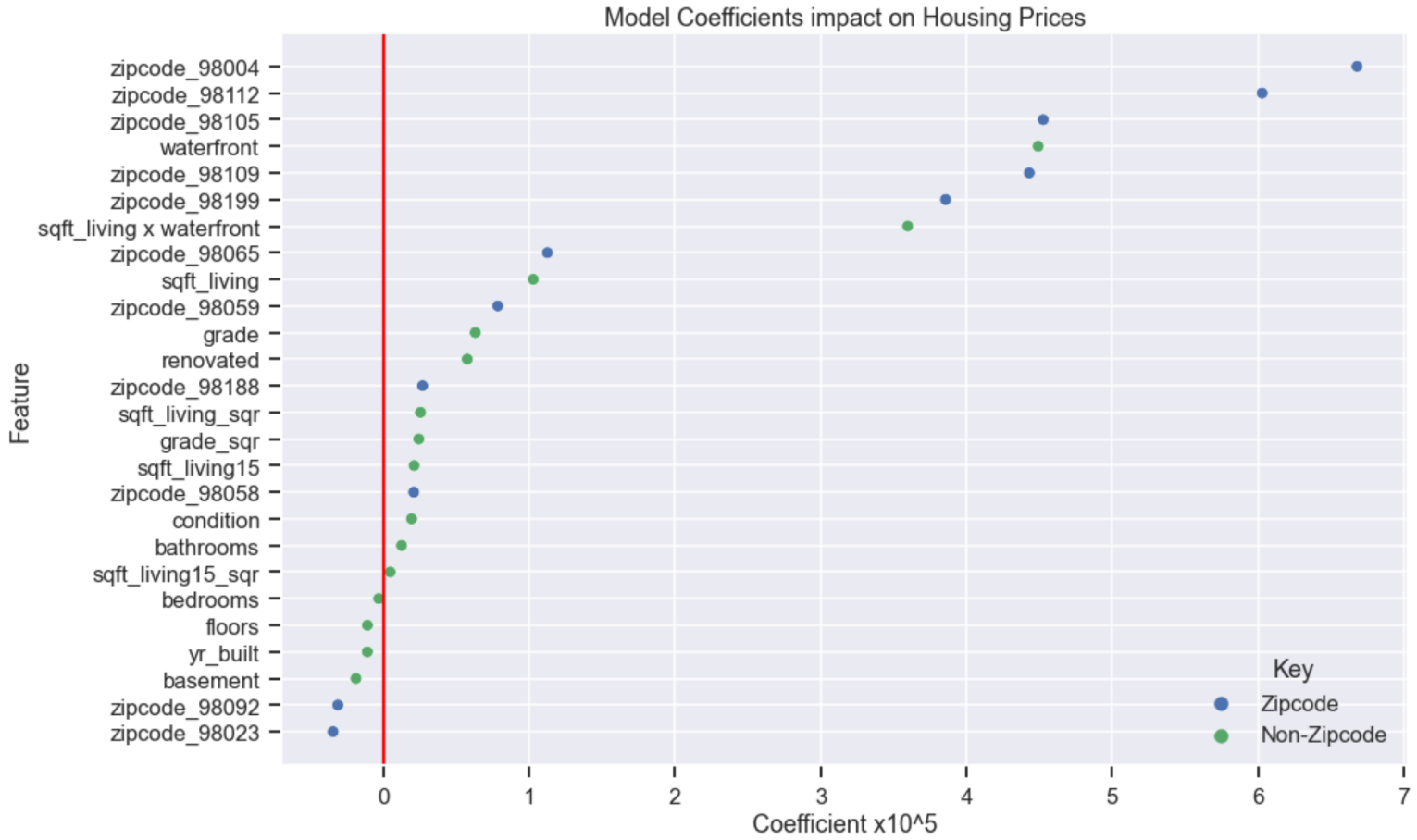
- **Predictors:**
  - Bedrooms
  - Bathrooms,
  - Square feet
  - Floors
  - Waterfront
  - Condition
  - Grade
  - Year built
  - Square feet of nearest 15 neighbors
  - Renovated
  - Basement
  - Zipcode
- **Outcome Variable** : Price

The Model: Driving Features



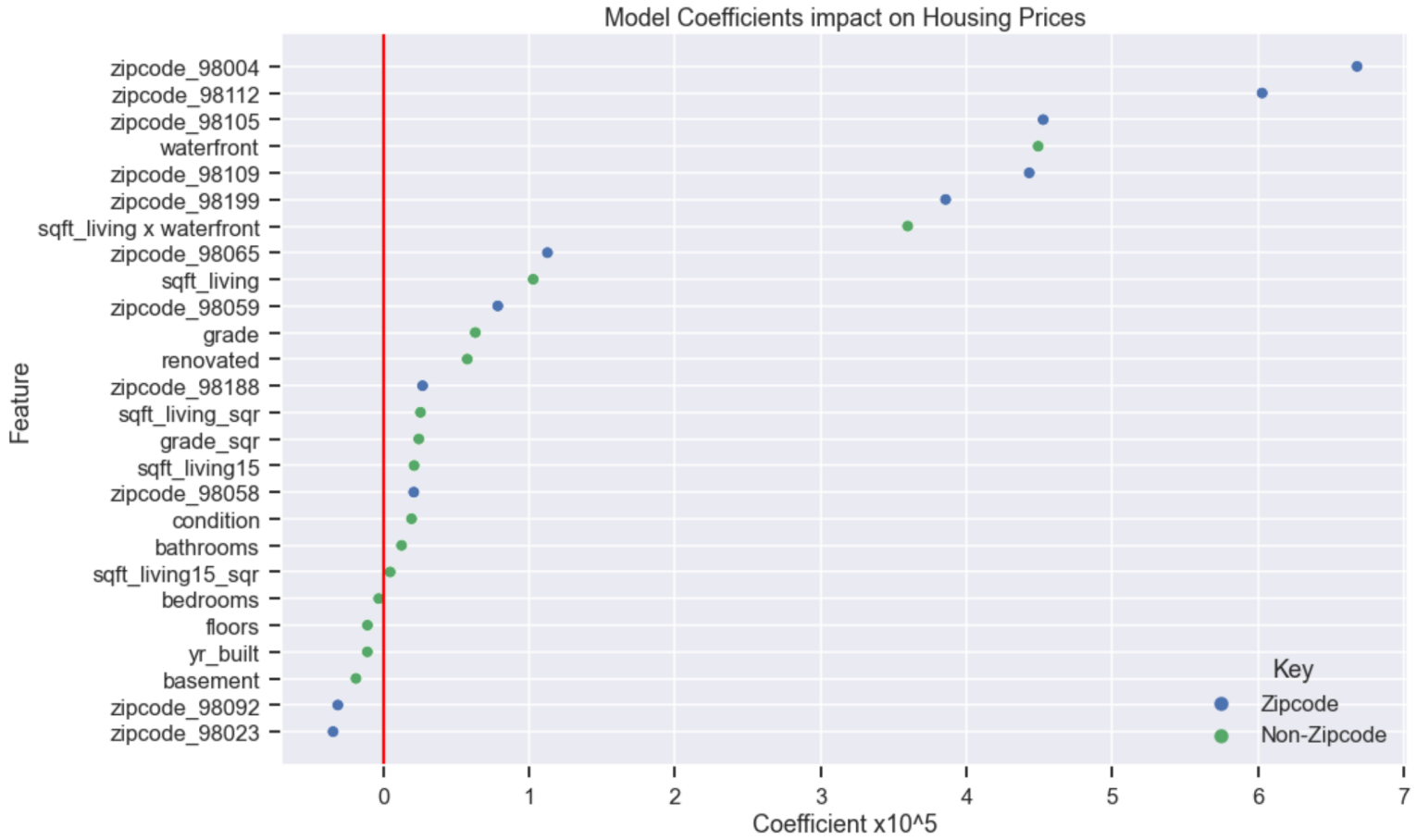
Rank	Feature	Coefficient
1	zipcode_98004	6.680633
2	waterfront	4.493111
3	sqft_living x waterfront	3.598411
4	sqft_living	1.029954
5	grade	0.632588
6	grade_sqr	0.244879
7	sqft_living15	0.212852
8	condition	0.195142
9	basement	-0.186253
10	bathrooms	0.126754
11	yr_built	-0.108136
12	sqft_living15_sqr	0.048538
13	bedrooms	-0.030189

The Model: Bedrooms and Sqft



Rank	Feature	Coefficient
1	zipcode_98004	6.680633
2	waterfront	4.493111
3	sqft_living x waterfront	3.598411
4	sqft_living	1.029954
5	grade	0.632588
6	grade_sqr	0.244879
7	sqft_living15	0.212852
8	condition	0.195142
9	basement	-0.186253
10	bathrooms	0.126754
11	yr_built	-0.108136
12	sqft_living15_sqr	0.048538
13	bedrooms	-0.030189

The Model



Rank	Feature	Coefficient
1	zipcode_98004	6.680633
2	waterfront	4.493111
3	sqft_living x waterfront	3.598411
4	sqft_living	1.029954
5	grade	0.632588
6	grade_sqr	0.244879
7	sqft_living15	0.212852
8	condition	0.195142
9	basement	-0.186253
10	bathrooms	0.126754
11	yr_built	-0.108136
12	sqft_living15_sqr	0.048538
13	bedrooms	-0.030189

Model: How well does the it fit the data?

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

Model: How well does the it fit the data?

- “Goodness of fit” = percent of errors that we can explain (R-squared)
- 81%

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.

The Model: Best/Worst Deals

- Top 10 “best deals” (data points with the most negative errors) v. dataset averages
  - Square Feet
  - Bathrooms
  - Zipcodes: 98006, 98107, 98199

- Top 100 “worst deals” (data points with the most positive errors) v. dataset averages
  - Waterfront
  - Renovated houses
  - Zipcodes: 98112, 98004, 98070

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?