Module 2 Project

Derek VanBriesen

Business Case:

Recommend an ideal geographic location to a home buyer, maximizing cost-effectiveness.

Questions to Consider

- What is the most cost effective location in King County?
- How much money can be saved based solely on location?
- Did any other savings opportunities appear in the data?

New Variables

I introduced a few new measurements to help analyze the data, I will give a brief overview of each one.

ZIP Group

Distance

Seasons

ZIP Group

What is it?

Each ZIP code ranked according to average home price and split into ten equal groups

Why?

Easier to visualize, less cluttered than viewing individual homes or every single ZIP code

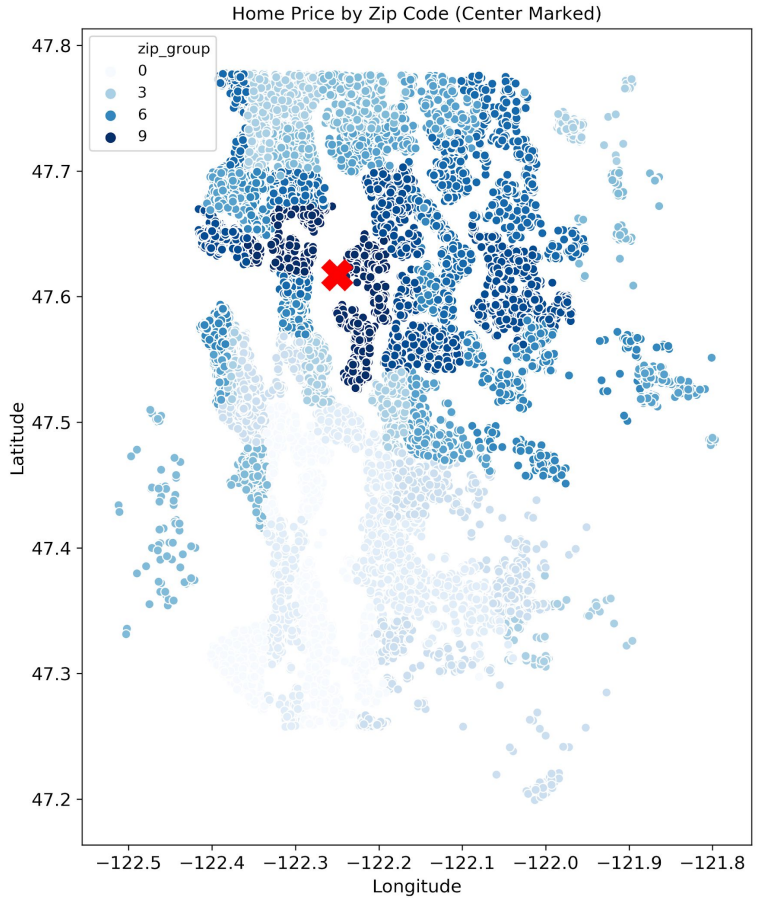
Distance

What is it?

The distance between each home and an estimated “epicenter” where prices are highest

Why?

Can be cross-referenced with ZIP Group in order to help find an optimal location



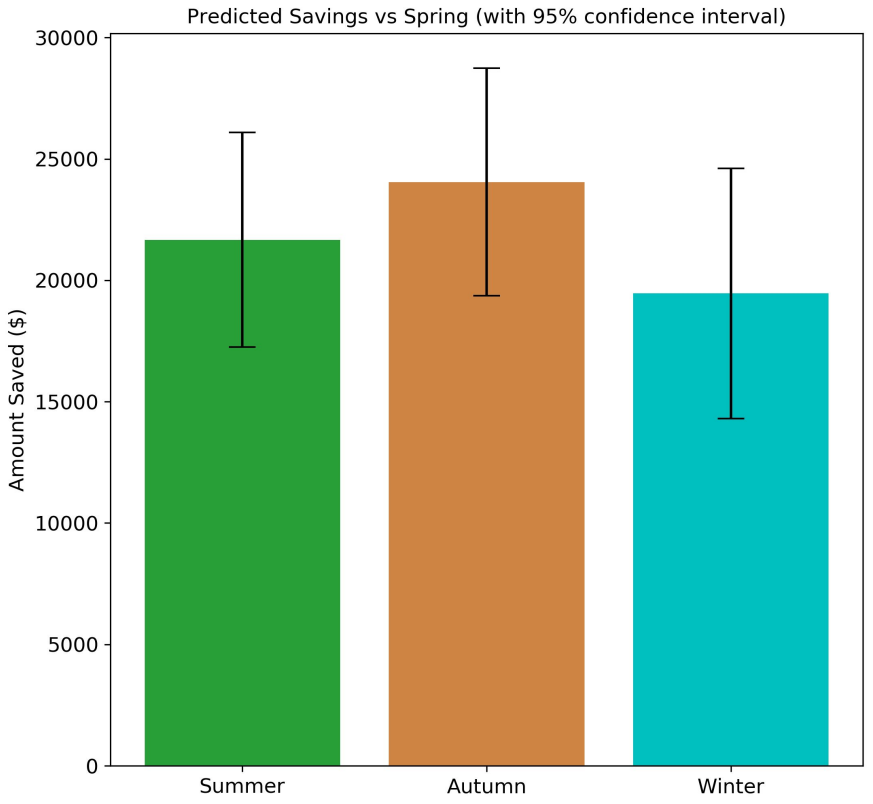
Seasons

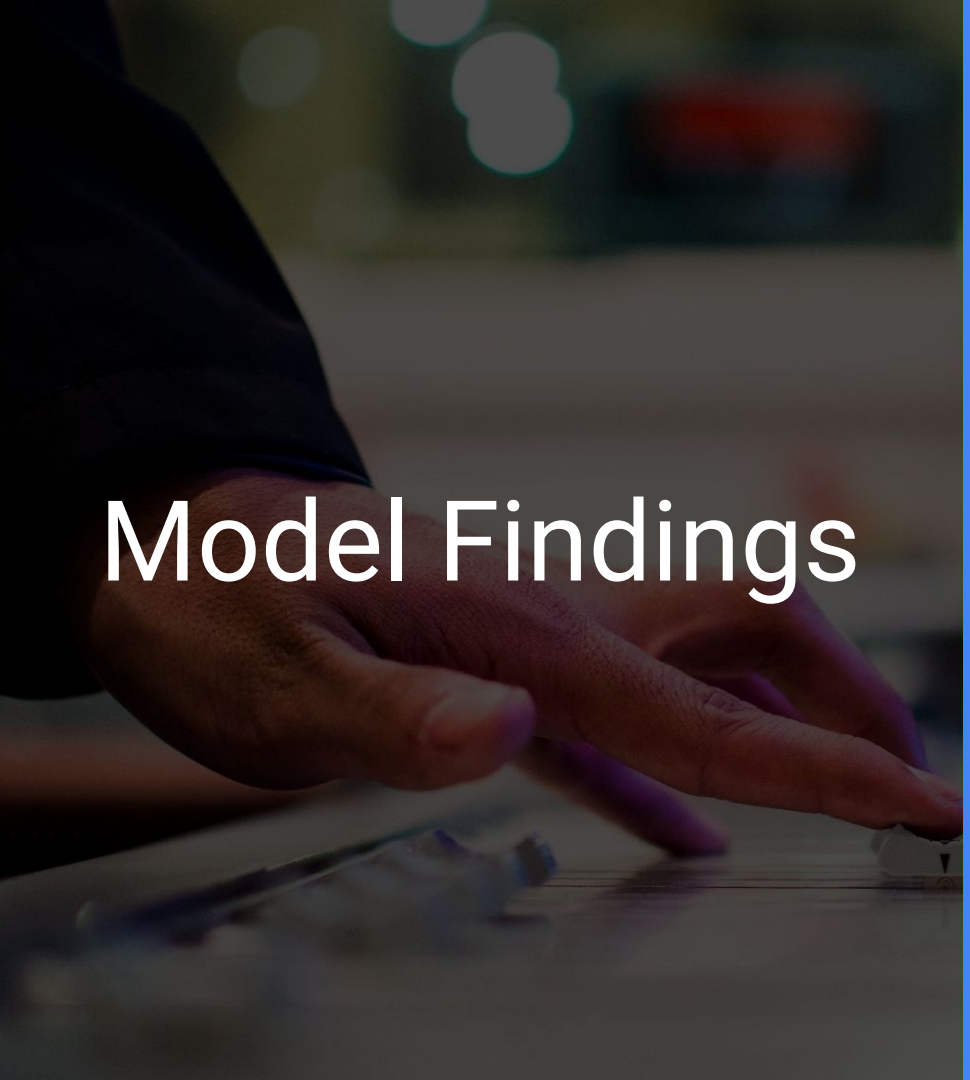
What is it?

Variables representing each of the four seasons (Spring, Summer, Autumn, Winter)

Why?

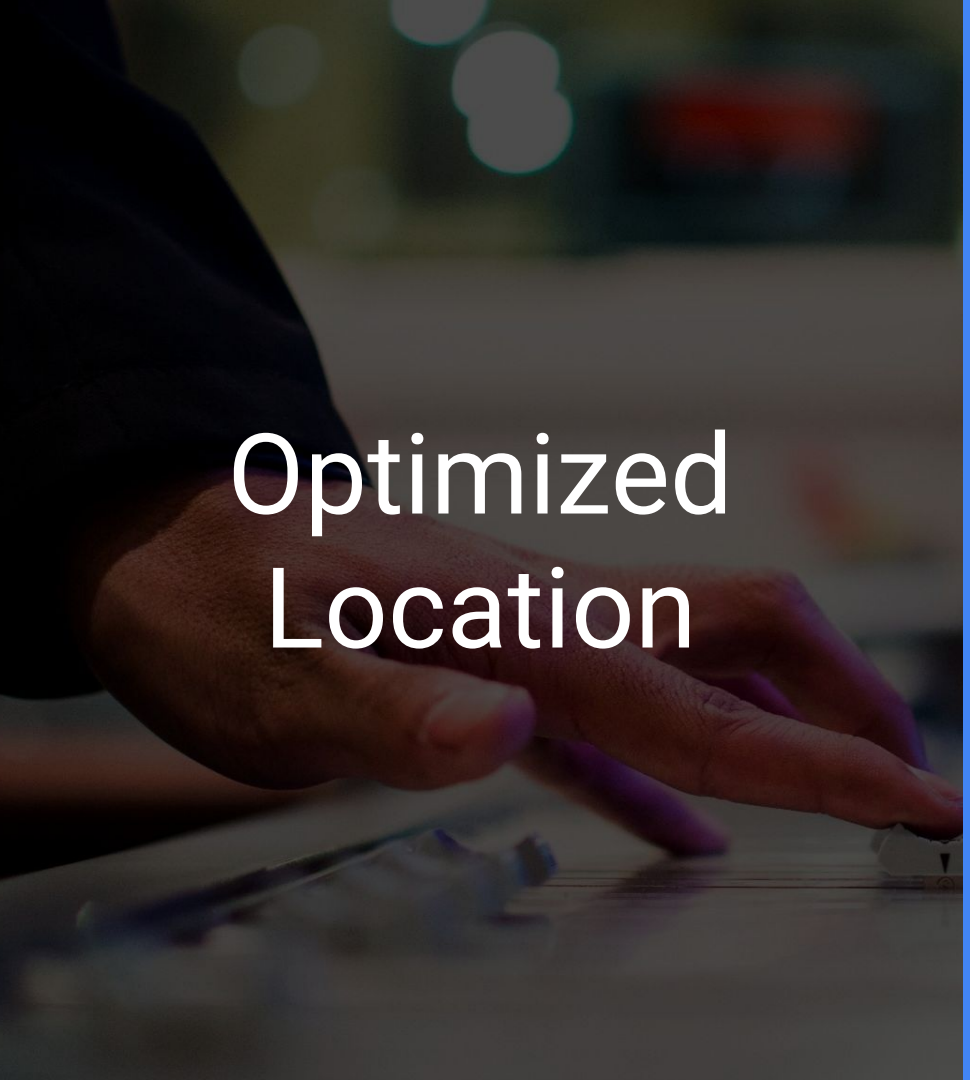
Controllable factor in home buying, it seemed like a simple way to further optimize costs.





Model Findings

- Higher ZIP Group corresponds with higher pricing
- Larger distance value corresponds with lower pricing
- Significant relationship between seasons and price

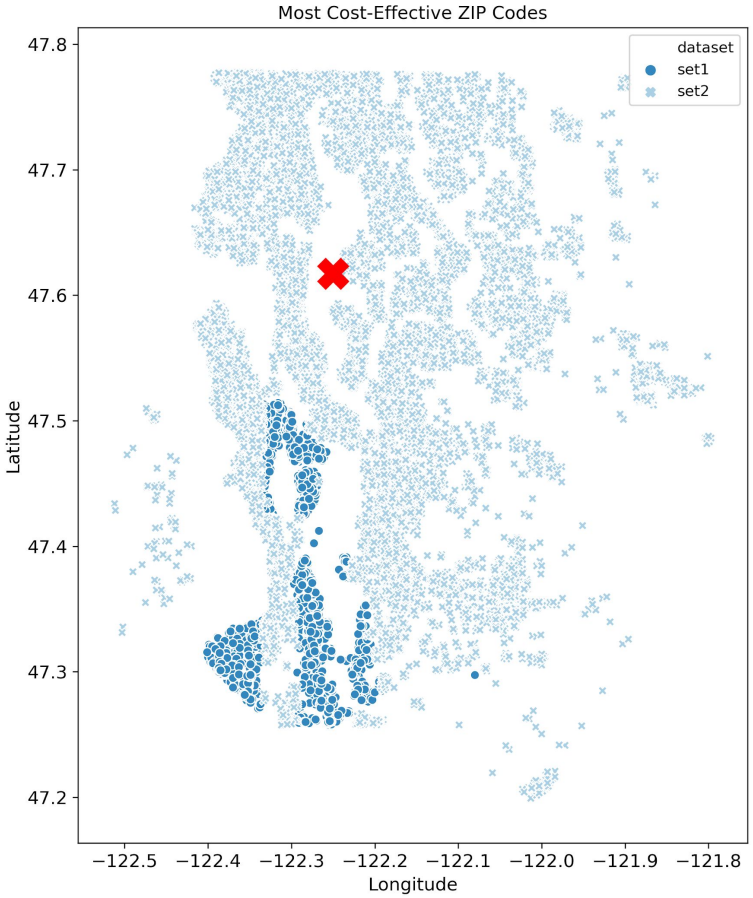


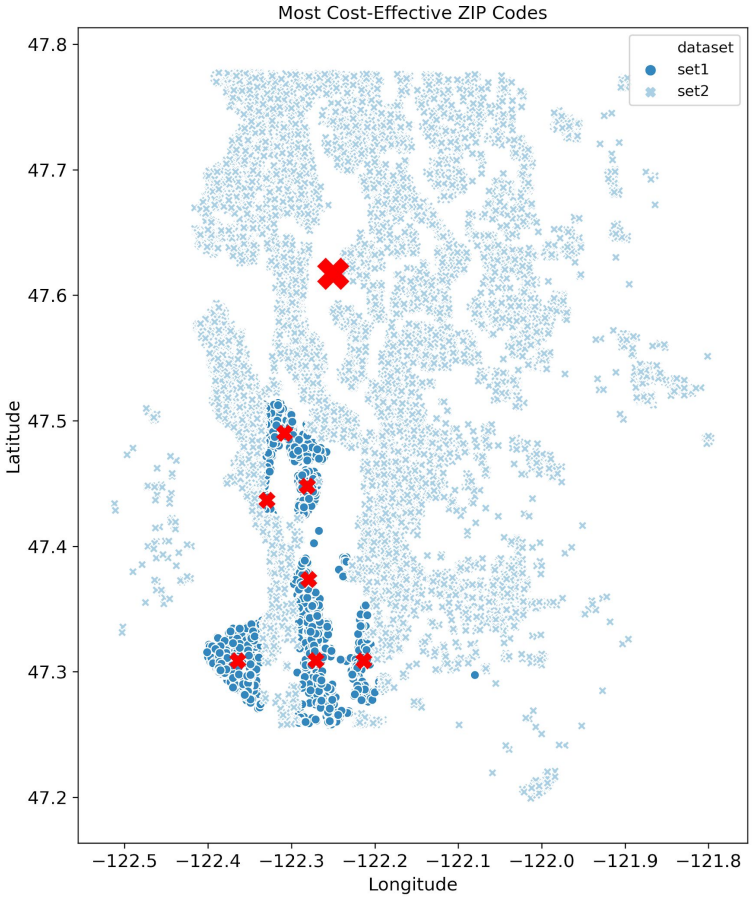
Optimized Location

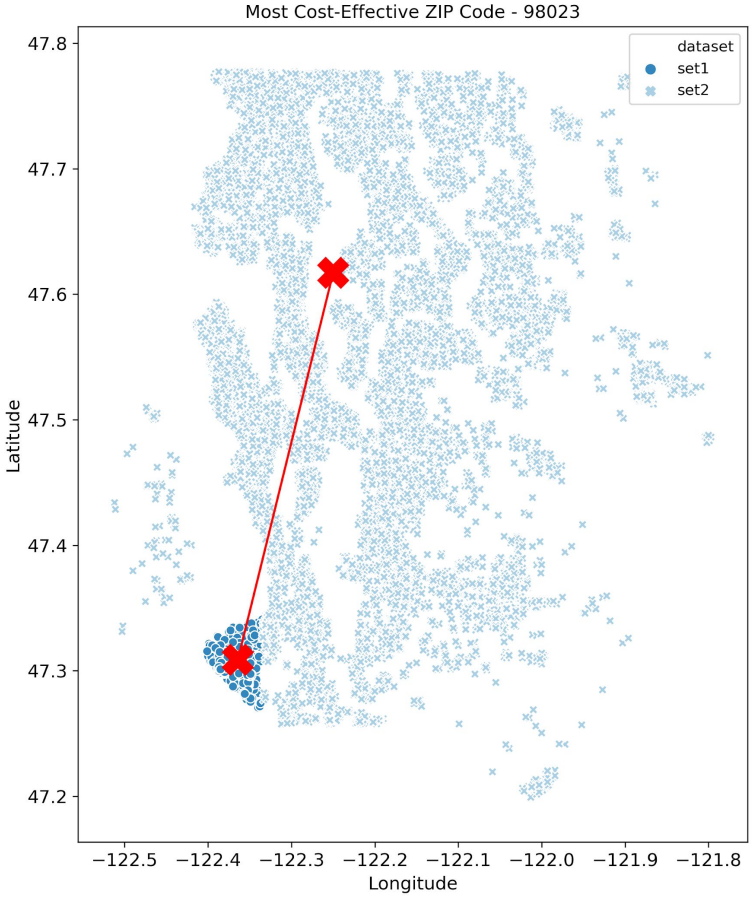
Two main geographic features:

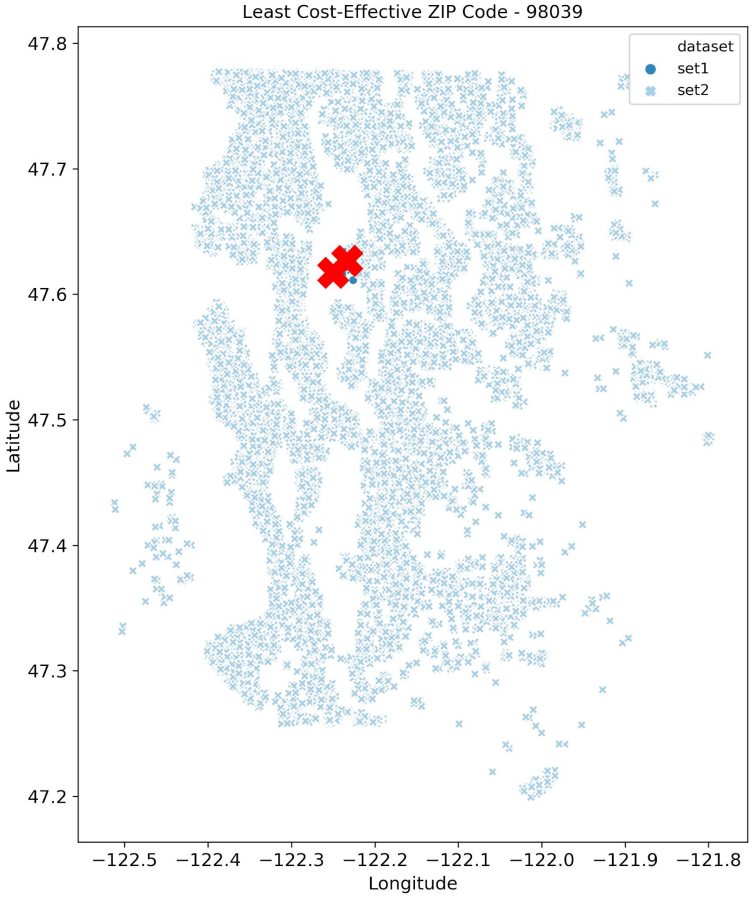
- ZIP Group
- Distance

Goal: Combine the effects of these features to find the lowest geographic cost


Most Cost-Effective ZIP Codes
dataset
set1
set2
47.8
47.7
47.6
47.5
47.4
47.3
47.2
Latitude
−122.5
−122.4
−122.3
−122.2
−122.1
−122.0
−121.9
−121.8
Longitude


Most Cost-Effective ZIP Codes
dataset
set1
set2
47.8
47.7
47.6
47.5
47.4
47.3
47.2
Latitude
−122.5
−122.4
−122.3
−122.2
−122.1
−122.0
−121.9
−121.8
Longitude


Most Cost-Effective ZIP Code - 98023
dataset
set1
set2
47.8
47.7
47.6
47.5
47.4
47.3
47.2
Latitude
−122.5
−122.4
−122.3
−122.2
−122.1
−122.0
−121.9
−121.8
Longitude


Least Cost-Effective ZIP Code - 98039
dataset
set1
set2
47.8
47.7
47.6
47.5
47.4
47.3
47.2
Latitude
−122.5
−122.4
−122.3
−122.2
−122.1
−122.0
−121.9
−121.8
Longitude

How Much Savings?

Most cost-effective ZIP

98023 -- $141,602 less

Least cost-effective ZIP

98039 -- $325,254 more

Predicted Difference

$466,857

Recommendations

- Buy in 98023
  - Most cost-effective ZIP
- Avoid 98039
  - Least cost-effective ZIP
- Buy in autumn, avoid spring

Future Work

- Is there a better way to estimate the epicenter mathematically?
- Different ways of classifying ZIP code areas
  - Coastal
  - Population
- Test other features to better fit model


Thank you for listening!