

Predicting Sales Prices of Projects in King County, WA

An Analysis Using Linear Regression by Joshua Gottlieb

Problem Overview

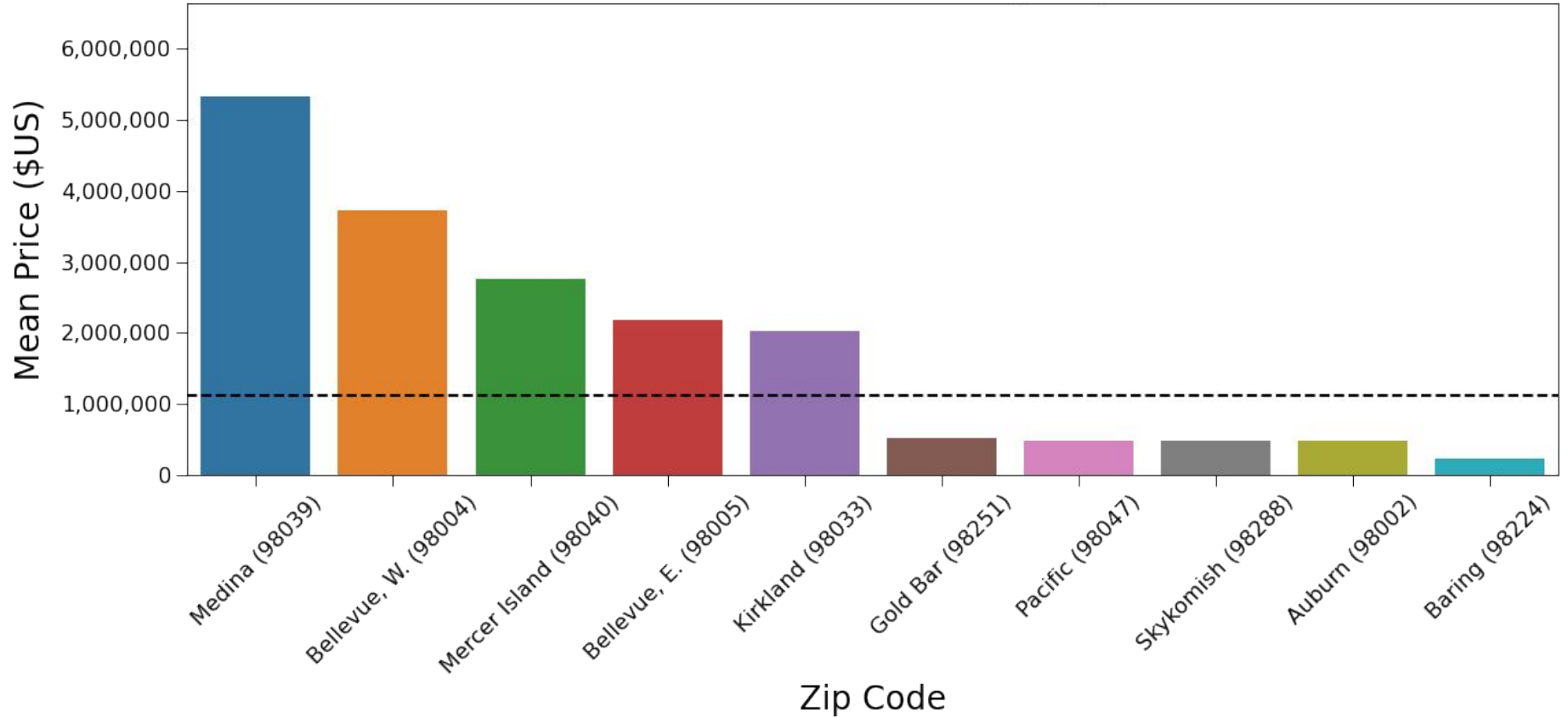
Scenario: A Real Estate Developer has a variety of potential projects that they have outlined, but they do not know which of their options will lead to the greatest value sale.

Business Knowledge

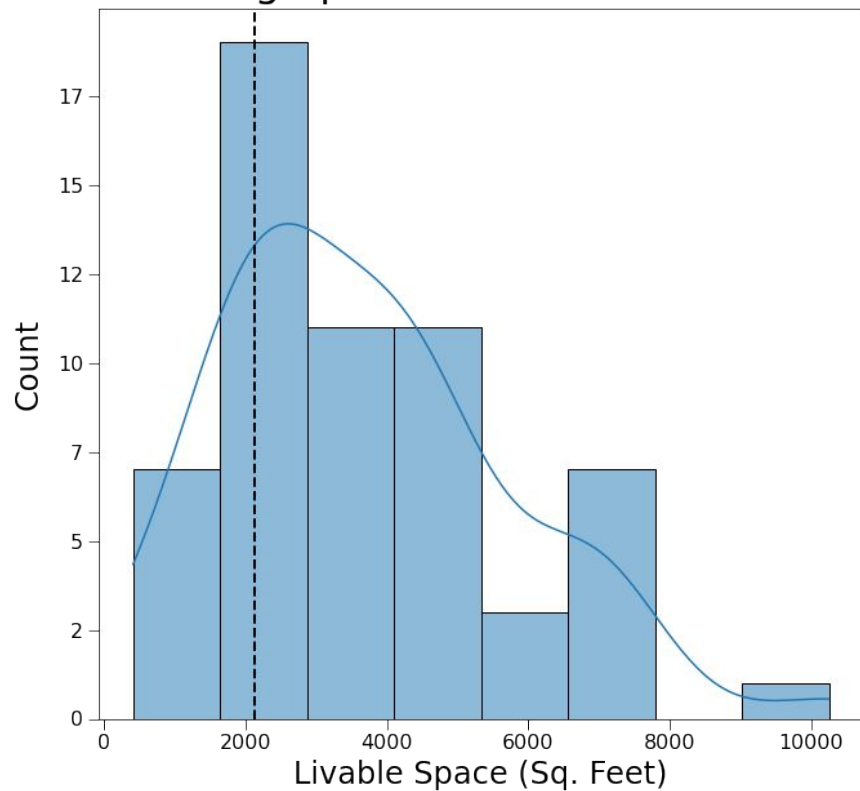
Location, Location, Location

While many factors influence the value of a house, location is closely intertwined with all of these factors, affecting the house's view, privacy, access to urban amenities, and total available land.

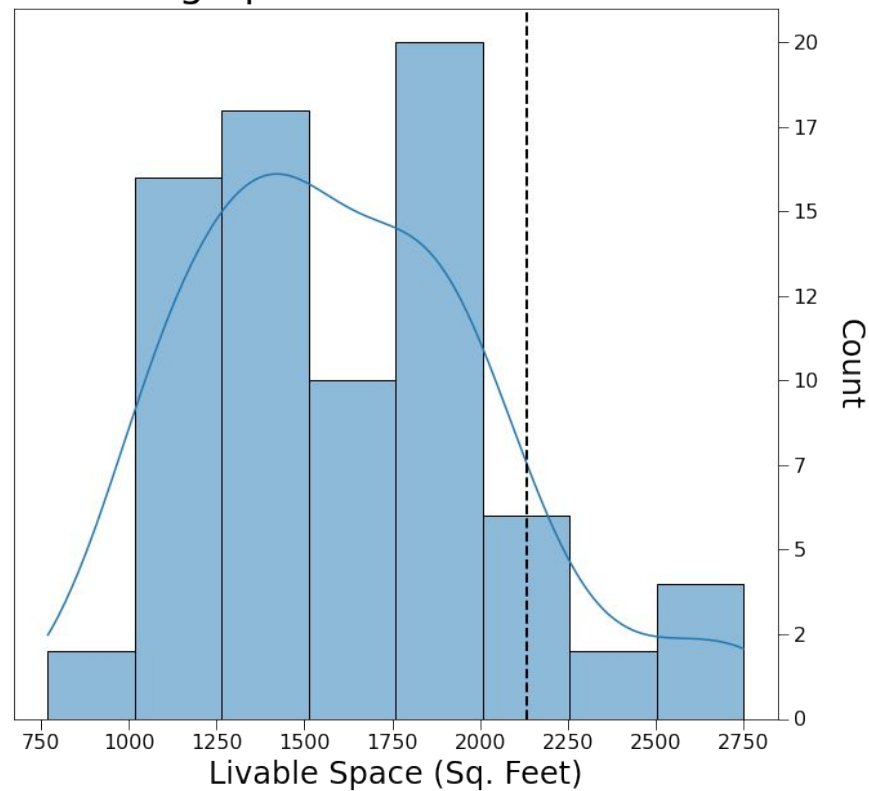
Mean Price of Houses by Zip Code



Living Space in 98039 - Medina



Living Space in 98047 - Pacific



Why Linear Regression?

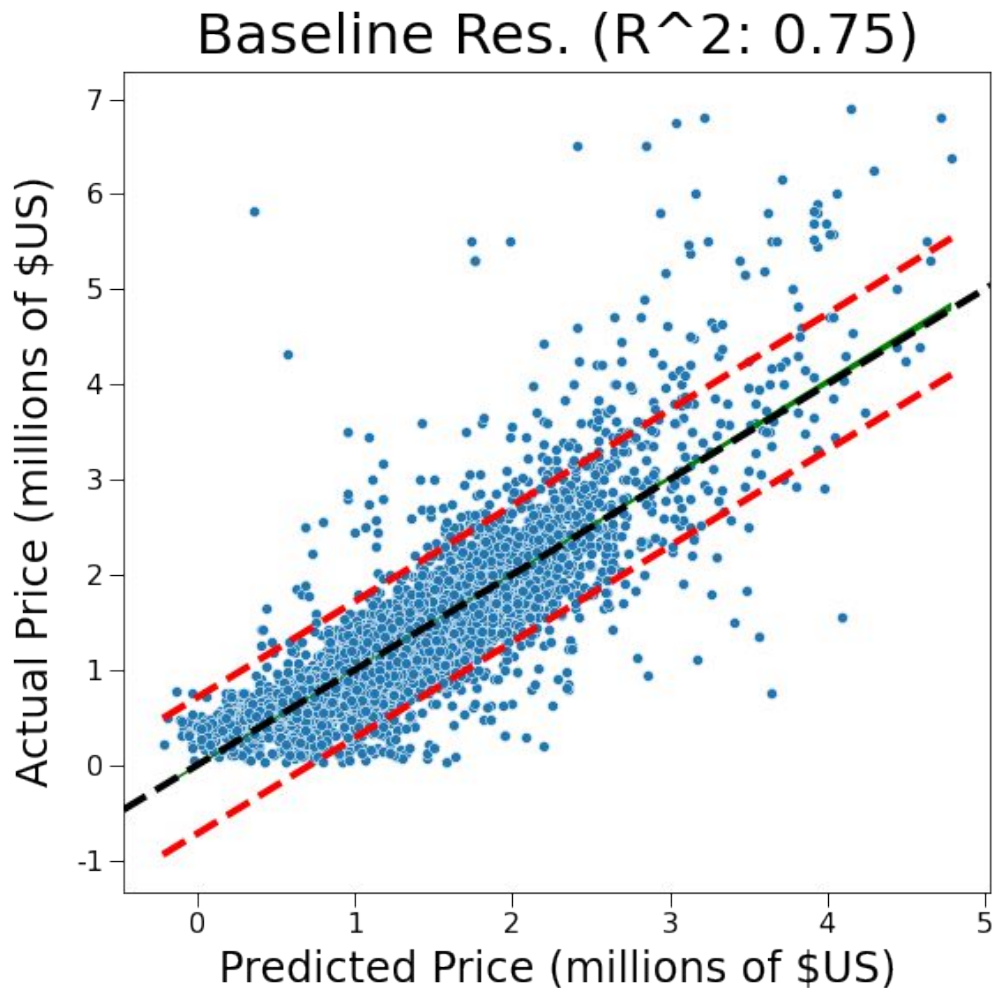
Linear regression allows us to go beyond identifying patterns in our data by creating a model that allows us to predict new values.

Linear regression allows us to estimate the strength of the effects which influence sales price in a concrete, numerical fashion.

Baseline Model

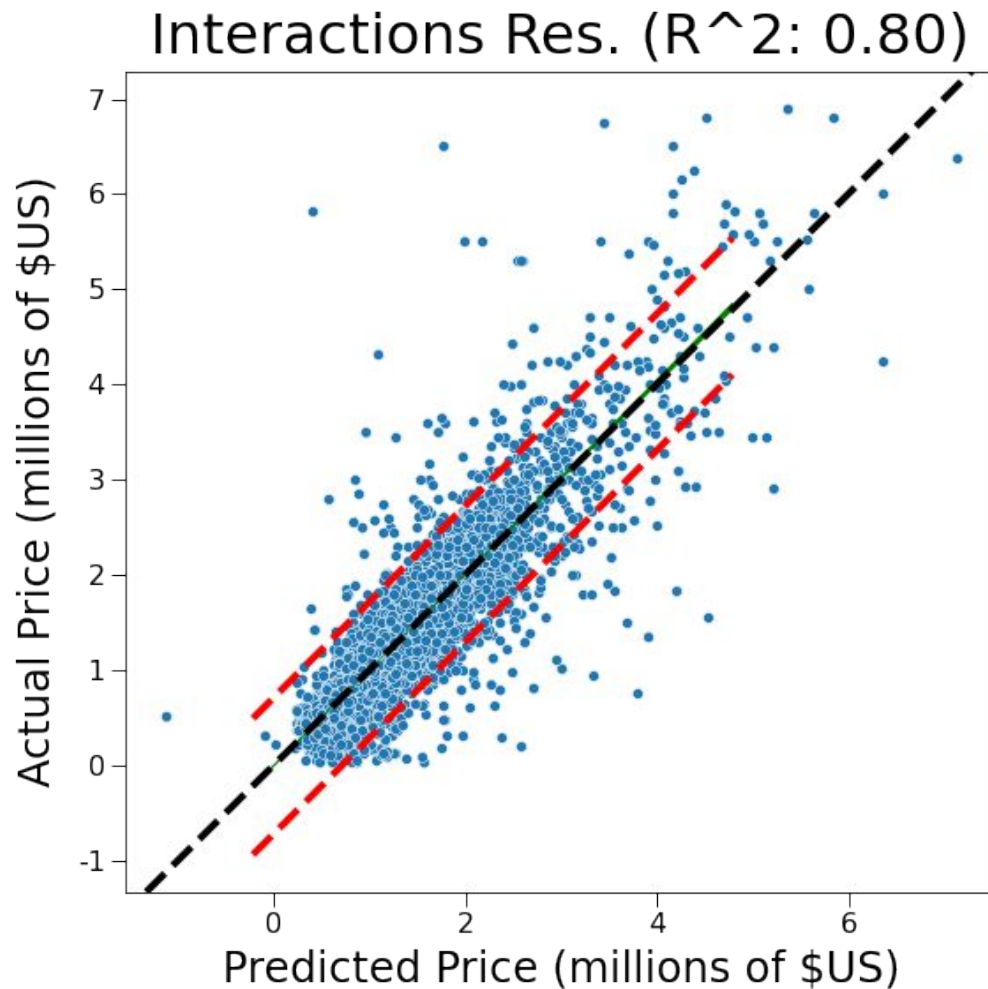
	Coefficient	Confidence Lower	Confidence Upper
Grade	106,165.18	97,775.46	114,554.91
Living Area (Sq. Ft.)	171.89	147.04	196.73
Lot Area (Sq. Ft.)	2.84	1.97	3.71
Waterfront	594,396.48	545,757.68	643,035.29
Zip Code - Bellevue	1,940,324.02	1,883,916.85	1,996,731.18
Zip Code - Medina	2,699,139.58	2,580,999.54	2,817,279.61
Zip Code - Mercer Island	1,205,093.03	1,148,273.87	1,261,912.19

Reference Zip Code:
98042 - Kent

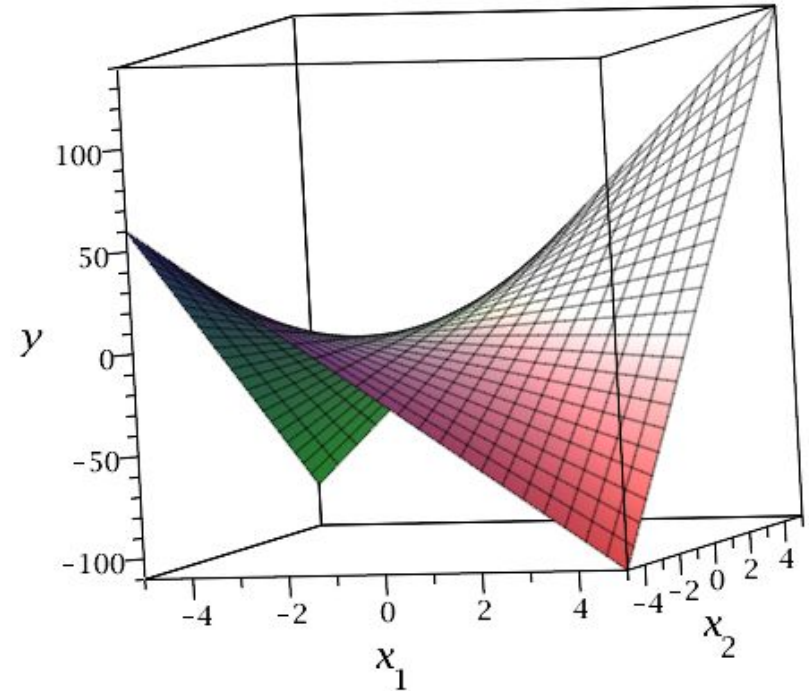
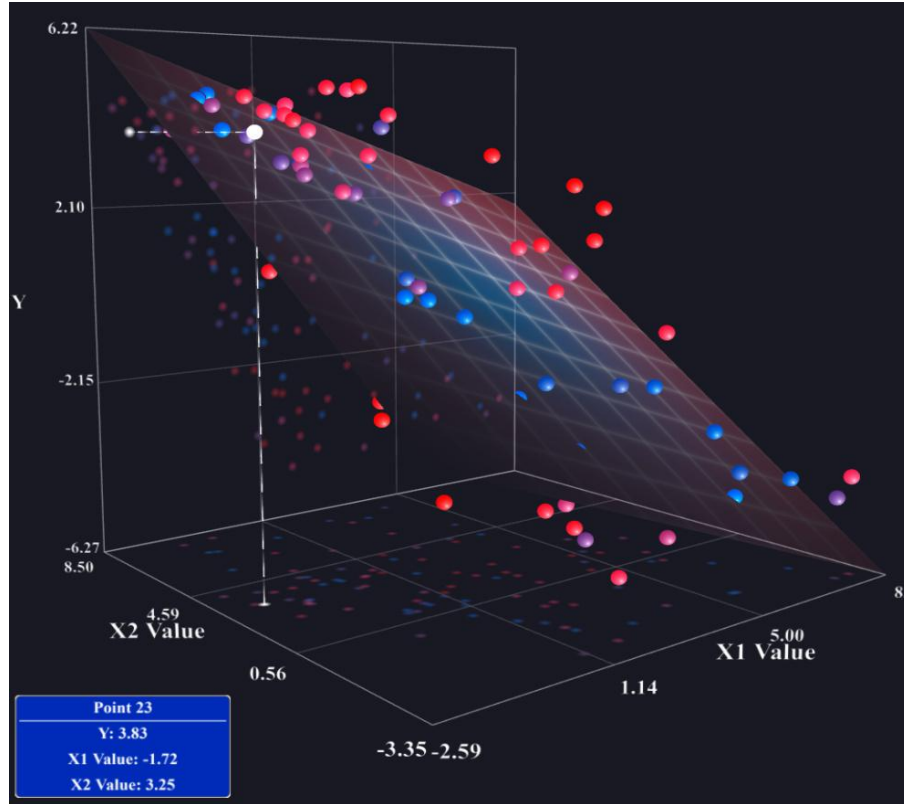


Interactions Model

- 626 Variables
- 528 Interactions
- 98 Stand-Alone
- Built on top of Prior Model



Visualizing Regression



What Can We Do with Interactions?

- Predict sales price of houses with different parameters (zip code, square footage, etc.)
- Estimate the sales price of completed houses, including those on the market to see what they should theoretically sell for and whether current houses are being under- or over-valued

Recommendations and Results

Best Zip Code: Medina, WA adds \$2.7m value

Waterfront: Water adds \$594k in value

Grade: Each point of grade adds \$106k in value

Living Area: Every 1000 sq. ft. adds \$172k in value

For best predictive results, use interactions model

Next Steps

- Gather purchase, building, and renovation data
- Analyze sales history to understand trends over longer periods of time
- Test and model on larger sets of data, especially ultra-expensive properties

Thank you for listening!

Presenter: Joshua Gottlieb

Email: joshuadavidgottlieb@gmail.com

LinkedIn: [linkedin.com/in/joshua-gottlieb](https://www.linkedin.com/in/joshua-gottlieb)

GitHub: github.com/JoshuaGottlieb