

A stylized, monochromatic illustration of the Seattle skyline in shades of orange and red. The Space Needle is the central focus, surrounded by various skyscrapers. The background is a solid orange color.

FOR SALE: King County

Presented by Anna D'Angela, October 19, 2020

An analysis of the King County, Washington housing market.

My client is an emerging real estate brokerage firm in King County. They are looking to maximize their potential commissions so they can expand their business and territory.

Flatiron School, Data Science Program

Phase 2 Project - Linear Regression

Objective

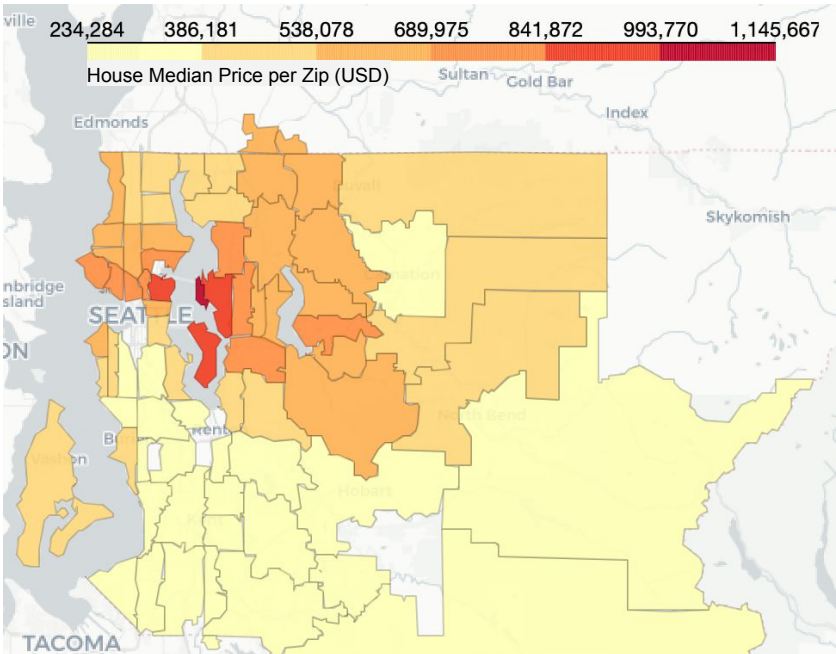
Maximize total sale prices to maximize potential commission for my client.

- 1) Where are the largest and most expensive homes?
- 2) Does age affect price or grade?
- 3) What is the effect of renovation on price and grade?
- 4) Build a model to predict house price.



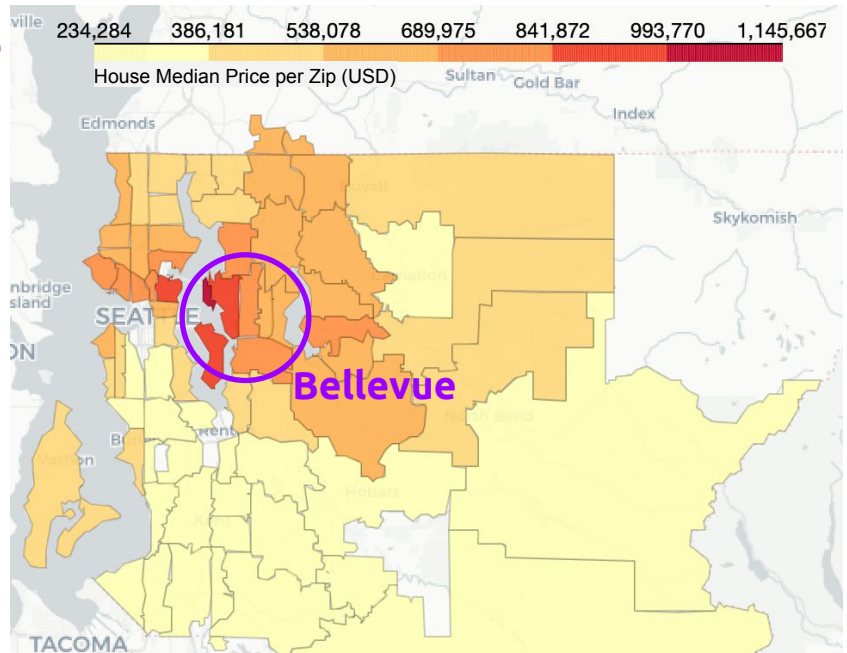
Highest Median Price
by Zip Code:

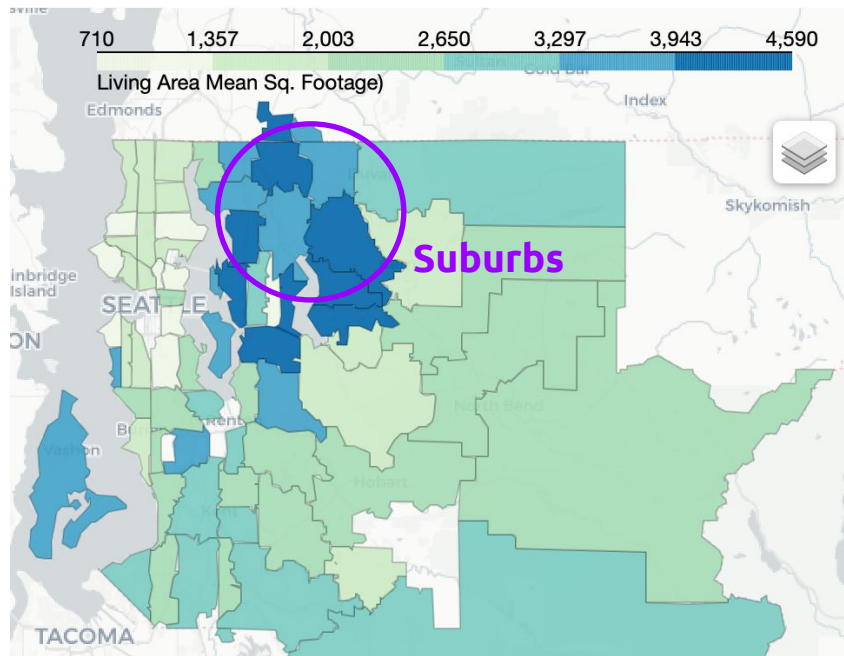
zipcode	mean_price (USD)	pop_density (People / Sq. Mile)
98102	808902.95	15829
98122	634558.14	13594
98105	831464.12	10643
98109	853256.78	10361
98103	584957.35	9905



Highest Median Price by Zip Code:

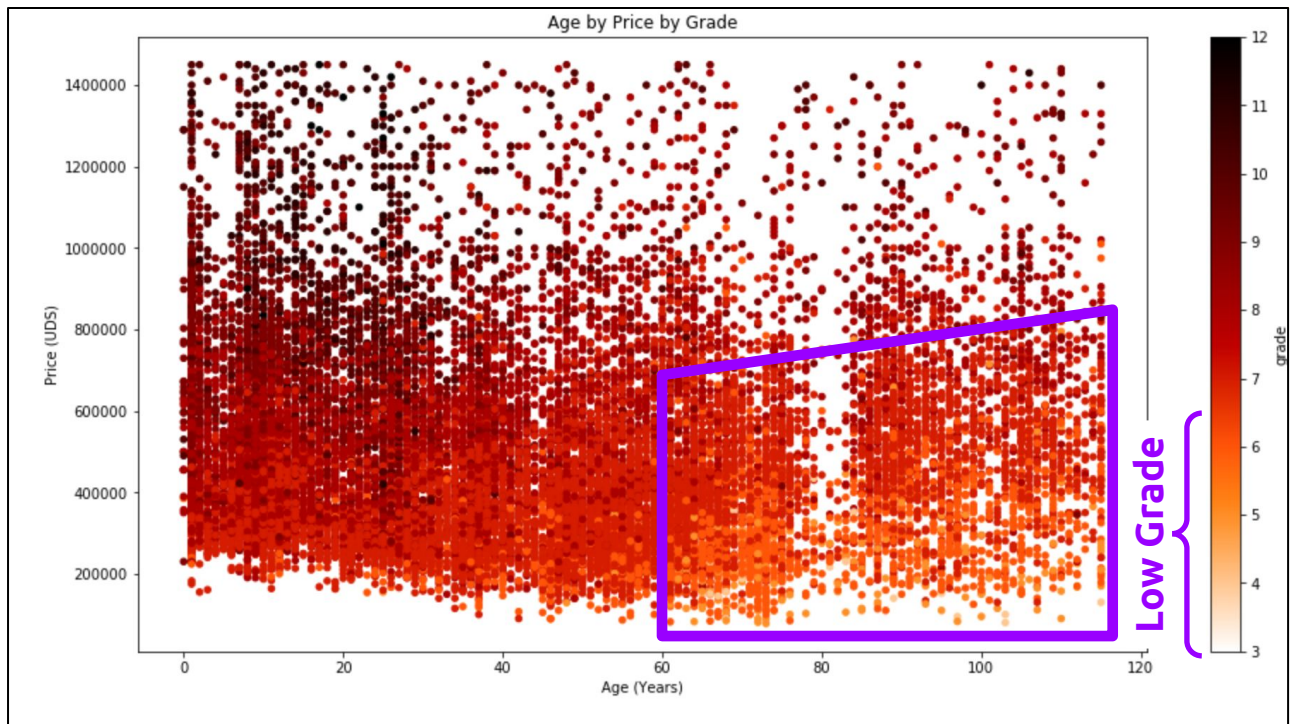
6 Bellevue zip codes 98004 - 98009



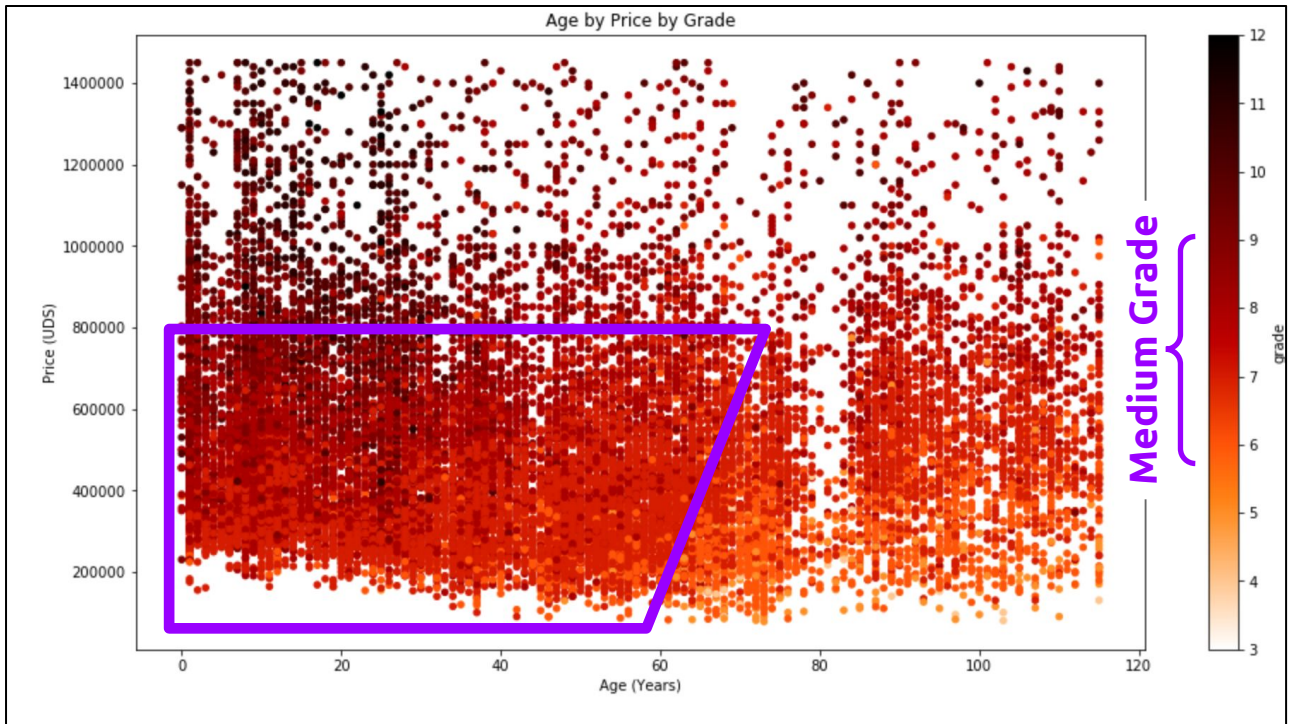


Largest Living Area (Sq. Ft) by Zip Code:

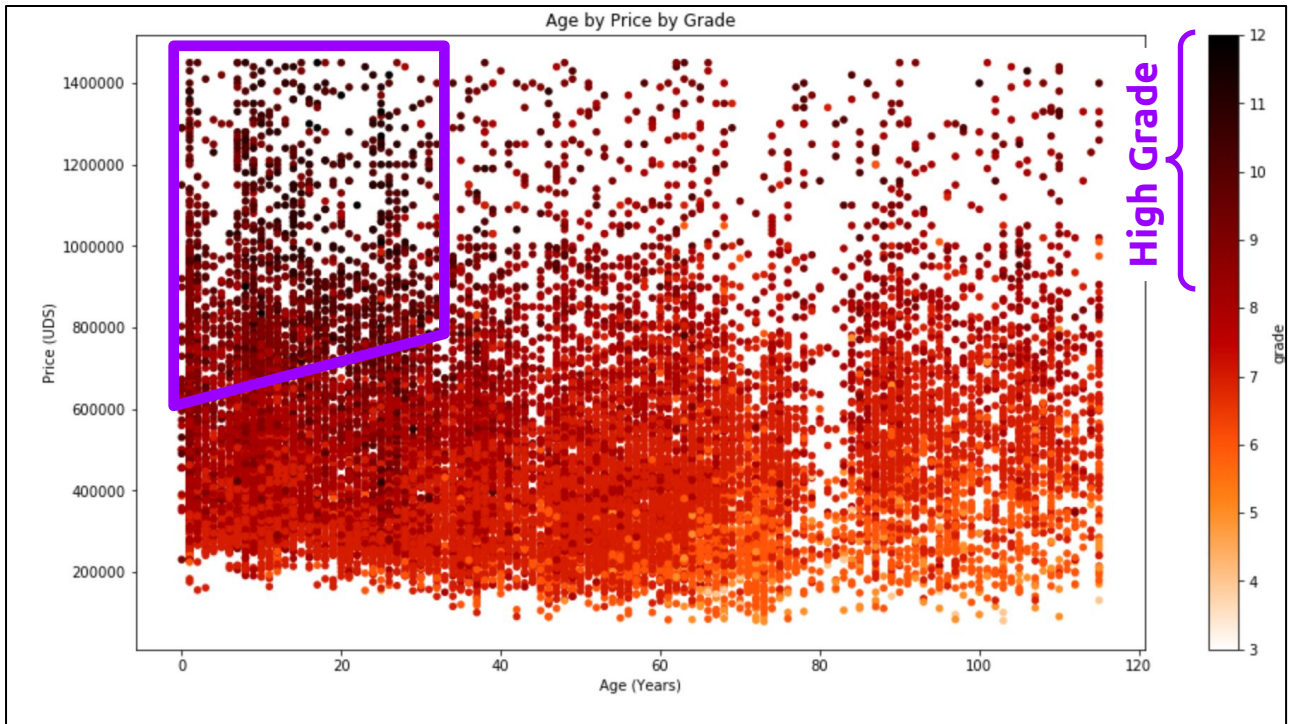
North Creek,
Woodinville,
Redmond,
Kirkland,
and Bellevue



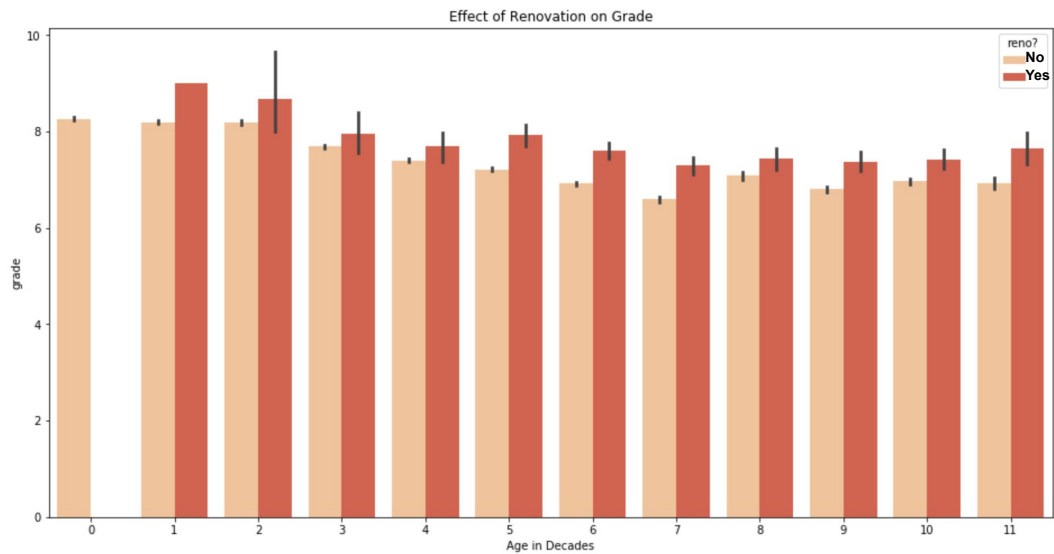
60 to 120 years old, lowest grades (below 7) and lowest priced



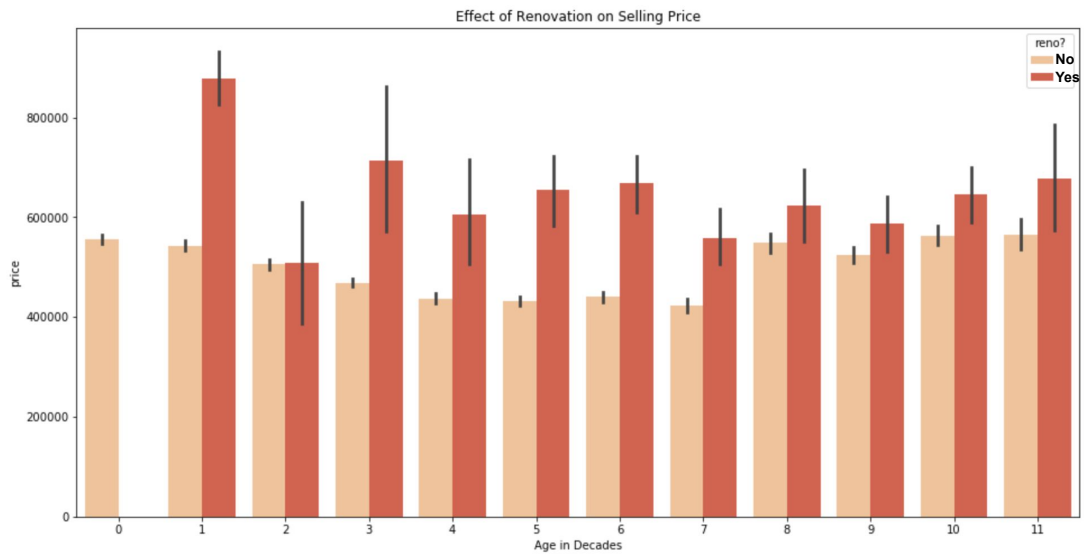
Newer, mid to low grade homes, selling at similar prices to previous slide, how to increase to top tier



Money Makers, Newest (30 years and newer), highest grade, most expensive



Slight ability to improve grade with a renovation, seems most renovations keep homes similar to original grade (most renovations may be focused on improving quality and modernizing homes).



Yet! We see renovation can have a big impact on price. So even if grade does not drastically improve, renovations can increase sale price and are worth considering for flipping homes.

Conclusions

The most expensive homes are on the east side of Lake Washington, which also contains some of the largest homes per population density.

Price is not correlated with age. Grade has a significant effect on price, across all ages.

Renovation can slightly improve grade, and greatly improve selling price.

Predictive Model

Uses 7 Predictors:

- Zip code rank, 1 - 5 (based on median house price)
- Sq footage of living space
- Sq footage of lot
- Population density by zip code
- View rank, 0 - 4
- Yes or no, if property has a functional basement
- Location distance from the highest price, central point

Accuracy

The model can predict with 77 % accuracy based on these features.

Margin of error is 92,000 USD of target home price.

The max price of test data was 1,000,000 USD. Which gives 9.2 % error.

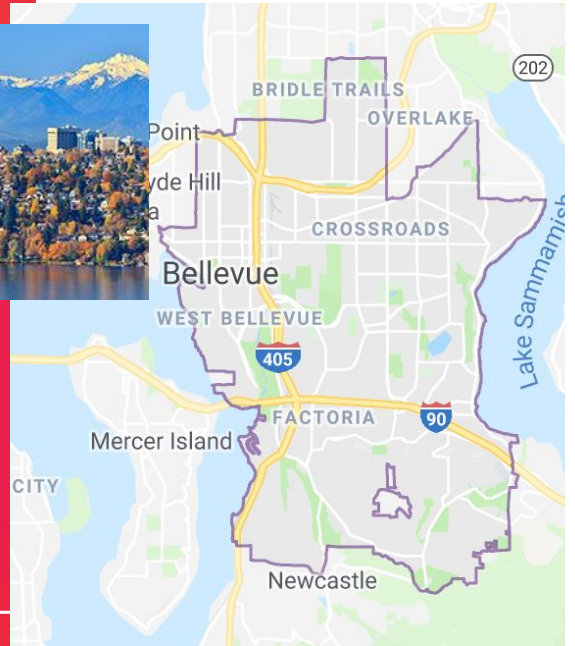




Recommendations

Focus new expansion on the Bellevue area

- High population density (turnover)
- Large, expensive homes
- Consider renovating



Future Work

- Divide top zip codes into neighborhoods, examine desirable neighborhood features/selling points
- Investigate ROI for flipping homes to improve grade
- Look into new development for new, high-grade, custom builds



Thank you!

Anna D'Angela
Detroit, MI

GitHub
@anna-dang

Washington State Motto:
"Alki" - "Bye and Bye"

Limitations - This model used housing data from 05/02/2014 to 05/27/2015 for King County. Some features were engineered on data that changes with time: zip code median price ranking and population density. Its predictive effectiveness is a snapshot of this particular market. As the individual house prices also fluctuate with time, this model will need to be updated as more data becomes available.

	feature	coeficent	top_3_bool
0	log_dist_cent	-4379.626326	False
1	pop_den_by_zip	45272.443174	True
2	view	27856.988744	False
3	zip_rank_price	96771.059288	True
4	log_sqft_lot	7423.453512	False
5	log_sqft_living	100852.594441	True
6	has_basement	-16576.900672	False

OLS Regression Results

	coef	std err	t	P> t	[0.025	0.975]
log_dist_cent	2.636e+04	1066.347	24.717	0.000	2.43e+04	2.84e+04
pop_den_by_zip	3.625e+04	1218.277	29.752	0.000	3.39e+04	3.86e+04
view	2.52e+04	1064.472	23.676	0.000	2.31e+04	2.73e+04
zip_rank_price	1.704e+05	438.524	388.520	0.000	1.7e+05	1.71e+05
log_sqft_lot	1.549e+04	1058.702	14.631	0.000	1.34e+04	1.76e+04
log_sqft_living	7.582e+04	1034.277	73.308	0.000	7.38e+04	7.78e+04
has_basement	4.239e+04	1959.096	21.636	0.000	3.85e+04	4.62e+04

Omnibus:	48.195	Durbin-Watson:	1.864
Prob(Omnibus):	0.000	Jarque-Bera (JB):	54.641
Skew:	-0.078	Prob(JB):	1.36e-12
Kurtosis:	3.220	Cond. No.	5.86

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

