King County Housing Price Analysis

Group Steel:
Brooke Smyth
Ely Lin
Wayne Harrison

Overview

- Business Problem
- Data and feature selection
  - Data cleaning
  - Feature selection
- Model Results
- Conclusion
  - Further analysis

Business Problem

- Major tech transformation project at Bank of Seattle
- Streamline mortgage approval
- Collateral Monitoring (Outstanding Mortgage < 70% Collateral)

Data Understanding

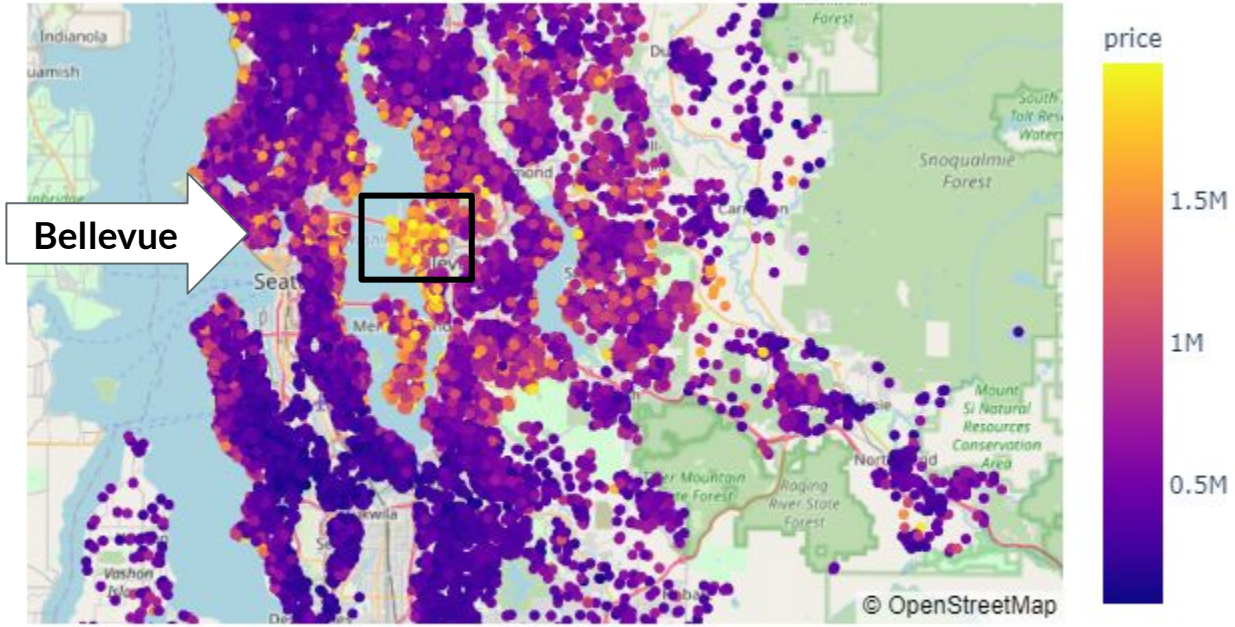
King County Housing Sales

- Informational county website for the public
- Contains over 21,000 records of housing sales within the county
- Records contain over 20 specific characteristics of each home sold

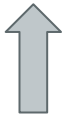
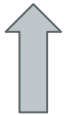
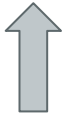
Feature Selection

- Measured the correlation between features and price
- Distance(mi) from Bellevue

General Trend in distance



Model Results

Features			Price	
Distance		1 mile		3.4%
Grade		1 level		16.4 %
Square Footage (living area)		100 sqft		1.5%

Model Accuracy

- explains 76.6% of the variance in price
- Prediction off by $184 thousand on average

Conclusion

- Easy to use for Mortgage Agents
- Faster approval process
- Seamless integration into risk management system
- Easy risk alarm (Outstanding Mortgage < 70% Collateral)

Future Work

- Reduce error
- Explore relationships between features
- Analyze how the time of year affects home prices

Q&A

Appendix

Bedrooms: number of bedrooms

Bathrooms: number of bathrooms

Sqft_living: total square footage of living space in the home

Sqft_lot: total square footage of the lot

Floors: number of floors

View: quality of view from house

Condition: how good the overall condition of the house is. Related to maintenance of house.

Grade: overall grade of the house. Related to the construction and design of the house.

Yr_built: year when house was built

Sqft_living15: the square footage of interior housing living space for the nearest 15 neighbors

Sqft_lot15: the square footage of the land lots of the nearest 15 neighbors

haver_dist_miles: distance between the house and Bellevue using the Haversine formula

OLS Regression Results

Dep. Variable:	price	R-squared:	0.766
Model:	OLS	Adj. R-squared:	0.766
Method:	Least Squares	F-statistic:	4425.
Date:	Thu, 07 Oct 2021	Prob (F-statistic):	0.00
Time:	13:38:44	Log-Likelihood:	-885.91
No. Observations:	16196	AIC:	1798.
Df Residuals:	16183	BIC:	1898.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	17.5750	0.185	95.090	0.000	17.213	17.937
bedrooms	-0.0266	0.003	-9.207	0.000	-0.032	-0.021
bathrooms	0.0678	0.005	14.296	0.000	0.058	0.077
sqft_living	0.0002	5.02e-06	30.509	0.000	0.000	0.000
sqft_lot	5.44e-07	6.6e-08	8.239	0.000	4.15e-07	6.73e-07
floors	0.0755	0.005	16.067	0.000	0.066	0.085
view	0.0731	0.003	25.847	0.000	0.068	0.079
condition	0.0526	0.003	15.616	0.000	0.046	0.059
grade	0.1524	0.003	48.722	0.000	0.146	0.159
yr_built	-0.0032	9.6e-05	-33.001	0.000	-0.003	-0.003
sqft_living15	7.744e-05	4.84e-06	16.016	0.000	6.8e-05	8.69e-05
sqft_lot15	4.528e-07	1.04e-07	4.353	0.000	2.49e-07	6.57e-07
haver_dist_miles	-0.0347	0.000	-88.630	0.000	-0.035	-0.034