

KING COUNTY HOUSE SALE ANALYSIS

Meet the Team

CYNTHIAH MULWO

PATRICK OKORE

DANIEL AKAMA

EDWIN RURUNG'A

DIANA NDUTA

KANE VUNDI

CONTENTS

- Overview
- Business & Data Understanding
- Modeling
- Regression Results
- Conclusion
- Next Steps



KING COUNTY

With nearly 2.2 million residents, King County is the largest county in Washington State. Nationally, it is the 13th largest by population. It is home to Seattle, the state's largest city, and is a hub of innovation, technology, and creative industries. The housing market in King County has been one of the fastest-growing in the country, with strong demand and limited supply leading to a highly competitive market.

BUSINESS UNDERSTANDING

As a result of this strong economy, the population has increased, attracting new employees for burgeoning businesses, and wages for higher-income households have increased. King County has experienced some of the fastest growing housing prices in the nation. The average King County home value is USD 914,300, an 88 percent jump since 2015.

OUR ROLE

At its core, the housing crisis is driven by a supply and demand challenge . Our role as data scientist is to advise new real estate developers who are planning to build new housing facilities on what factors come into play while evaluating homes in King County.

DATA UNDERSTANDING

Date

Date house was sold

Price

Price house was sold

View

Indicator of view from house

An index from 0 to 4 on the quality of view of the house

Bedrooms

Number of bedrooms

Condition

Condition of the house

An index from 1 to 5 on the condition of the house

Lat & Long

Location of house

Coordinates of the region the house is located

Bathrooms

Number of bedrooms

Sqft-Living/Lot/Above

Square footage

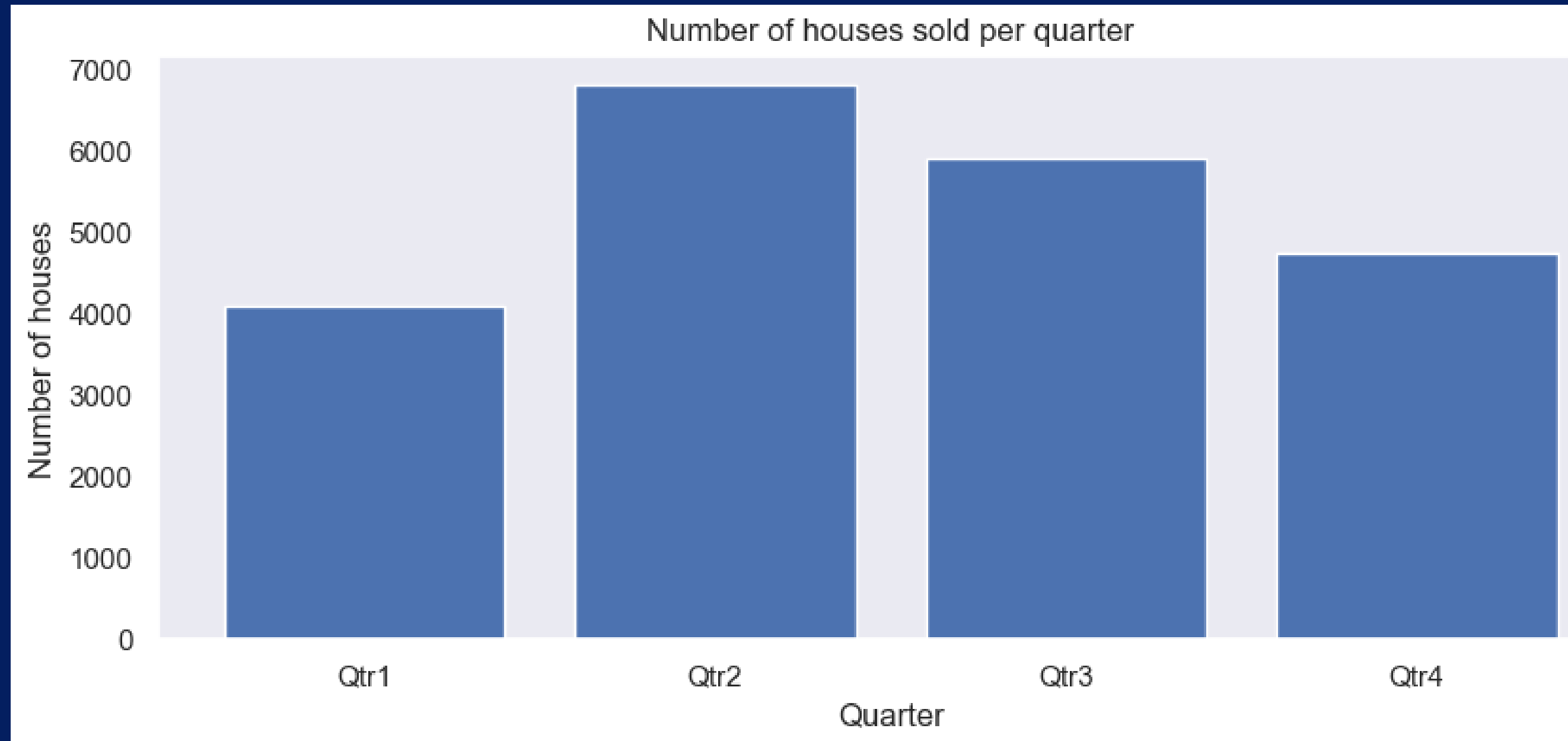
Square footage of particular section of house

DATA ANALYSIS

WE USED
80%
OF THE DATA
COLLECTED



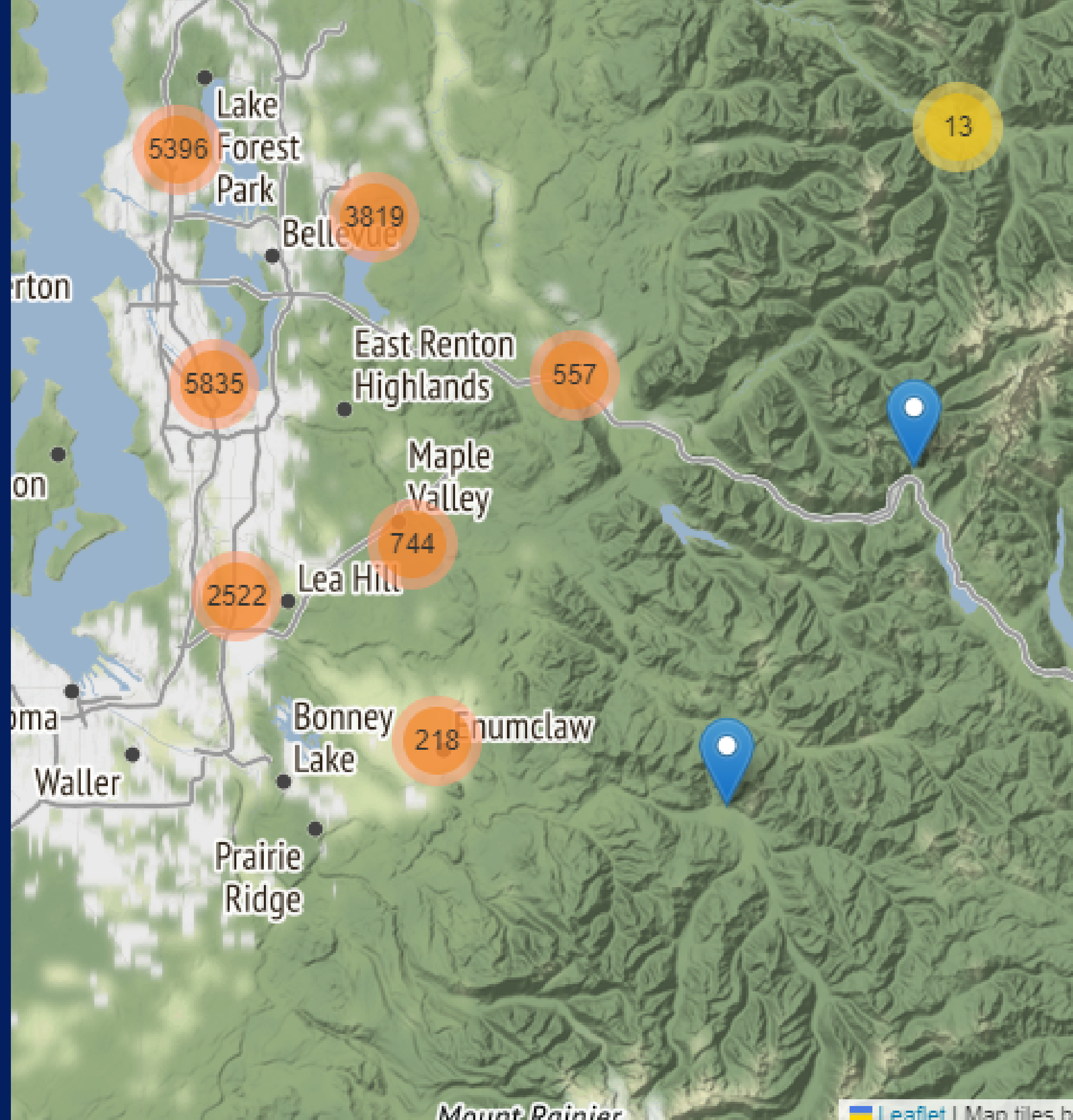
HOUSE SALES PER QUATER



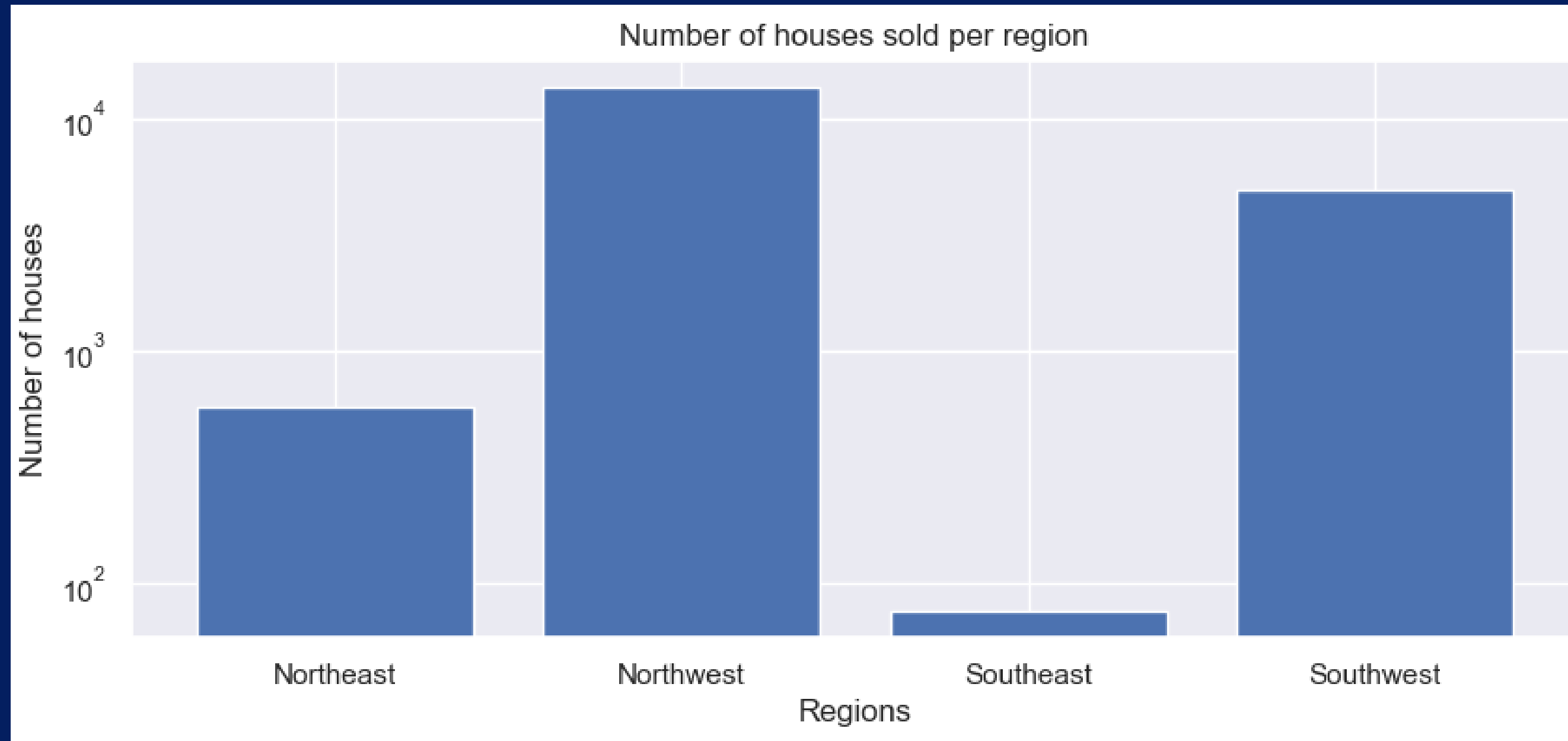
Houses sold in quarter 2 (April, May and June) have the highest price and the trend shows seasonality of house sales.

MAPPING

The King County data is divided into 4 regions, Northwest, Northeast, Southwest and Southeast.



HOUSE SALES PER REGION



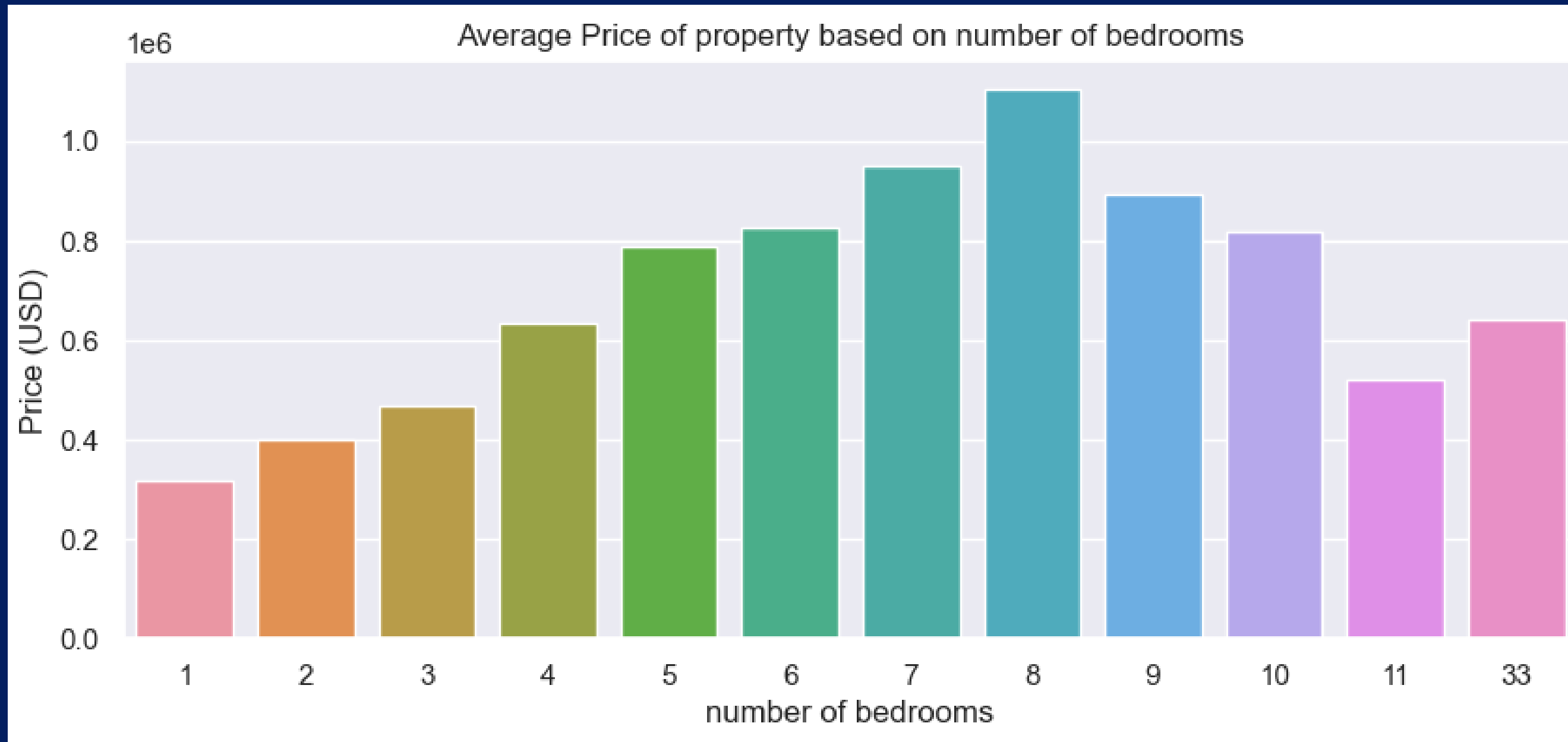
Northwest and Southwest region which is the region surrounding Seattle has the highest number of houses on sale.

DISTRIBUTION OF PRICES



The prices were normally distributed largely because we removed price outliers. There are more low priced houses than highly priced ones.

BEDROOMS VS PRICE



As the bedrooms increase, price increases, upto 8 bedrooms, afterwards the price decreases.

CONDITION VS PRICE



As the condition of the house improves, price increases – houses with a very good condition have higher prices. Thus, keeping property in a good condition will result to a higher valuation

MODELLING



success metrics

MAE

gives the difference between the predicted value (price) and the actual value (mean price). The lower the MAE the better the model.

Rsquared

represents the proportion of the variance for a target that's explained by a predictor in a regression model. The higher the rsquared the better.



BASELINE MODEL

Rsquared

The square footage of the living room only explained 36.3% of variation in price.

MAE

Our final model was off in predicting price by USD 130224.

An increase of 1 sq foot in `sqft_living` will result in an increase in house price by USD 171.29

OLS Regression Results						
Dep. Variable:		price		R-squared:		0.363
Model:		OLS		Adj. R-squared:		0.363
Method:		Least Squares		F-statistic:		1.087e+04
Date:		Thu, 20 Apr 2023		Prob (F-statistic):		0.00
Time:		07:47:41		Log-Likelihood:		-2.5627e+05
No. Observations:		19106		AIC:		5.125e+05
Df Residuals:		19104		BIC:		5.126e+05
Df Model:		1				
Covariance Type:		nonrobust				
		coef	std err	t	P> t	[0.025 0.975]
const		1.386e+05	3401.496	40.745	0.000	1.32e+05 1.45e+05
sqft_living		171.2969	1.643	104.250	0.000	168.076 174.518
Omnibus:		920.239	Durbin-Watson:		1.979	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		1057.149	
Skew:		0.565	Prob(JB):		2.77e-230	
Kurtosis:		3.226	Cond. No.		6.01e+03	

ITERATED MODEL

Rsquared
The final of the living room only explained 60.8% of variation in price.

MAE
Our final model was off in predicting price by USD 95232

An increase of 1 sq foot in `sqft_living` will result in an increase in house price by USD 176

	coef	std err	t	P> t
const	1.744e+05	8624.322	20.227	0.000
bedrooms	-2.059e+04	1475.682	-13.956	0.000
sqft_living	176.2486	1.725	102.154	0.000
floors	2.494e+04	1900.545	13.123	0.000
Qtr2	4936.7353	2686.308	1.838	0.066
Qtr3	-9434.0060	2753.755	-3.426	0.001
Qtr4	-1.374e+04	2889.816	-4.754	0.000
Fair	-4.685e+04	1.07e+04	-4.391	0.000
Good	3.523e+04	2242.338	15.712	0.000
Poor	-8.652e+04	2.72e+04	-3.184	0.001
Very Good	6.301e+04	3616.291	17.424	0.000
Northwest	1.076e+05	5463.186	19.688	0.000
Southeast	-1.481e+04	1.55e+04	-0.953	0.341
Southwest	-1.012e+05	5662.176	-17.877	0.000
EXCELLENT	1.38e+05	1.25e+04	11.082	0.000
FAIR	4321.9548	9195.650	0.470	0.638
GOOD	2.453e+04	8625.322	2.844	0.004
NONE	-8.03e+04	4857.311	-16.532	0.000

CONCLUSION



FEATURE IMPORTANCE:

The number of bedrooms, number of bathrooms, and square footage of the living area are all important features that significantly affect the price of the house.



LOCATION

The location of the house also appears to be important, with different regions having different associated increases or decreases in price. In particular, the north-west region



CONDITION AND VIEW

The condition of the house and the view from the house are also important factors that affect the price

Recommendations

FOCUS ON IMPORTANT FEATURES

Emphasize to realtors the importance of the number of bedrooms, number of bathrooms, and square footage of the living area when determining a property's value.

HIGHLIGHT DESIRABLE LOCATIONS

Encourage realtors to highlight the location of a property, especially in terms of the region it is located in. Realtors should inform clients that different regions have different associated increases or decreases in price.

EMPHASIZE CONDITION AND VIEW

Realtors should also emphasize the importance of the condition of the property and the view from the property.

THANK YOU!

