



Impact of Internal and External Factors of Houses Sold in Kings County WA

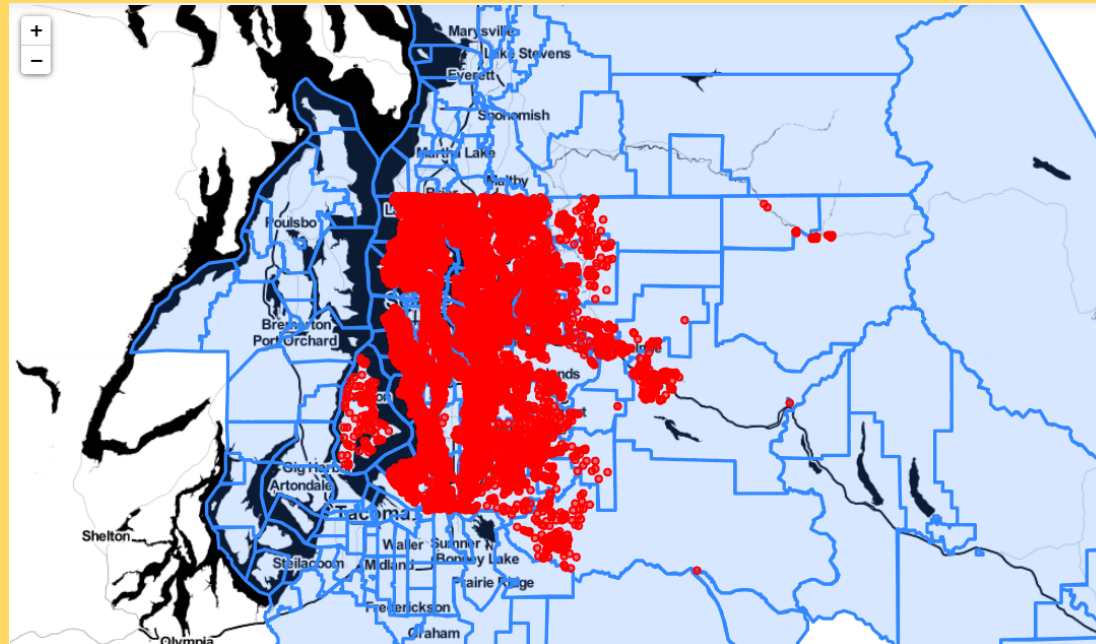
By Raymond Leong

Overview

- We are using Linear Regression modelling in a dataset of Houses sold in Kings County, WA, USA to visualize trends and patterns that can help your company succeed
 - For instance, the if the house is by the water: the price goes up 150% on average
 - On the other hand, if the home inspections is not favorable (on a chart of 1-14), price will not increase until a certain threshold in Grade is met.
- The goal of this project is to examine if external features of the house have more impact on the price compared to internal features

Overview (Continuation)

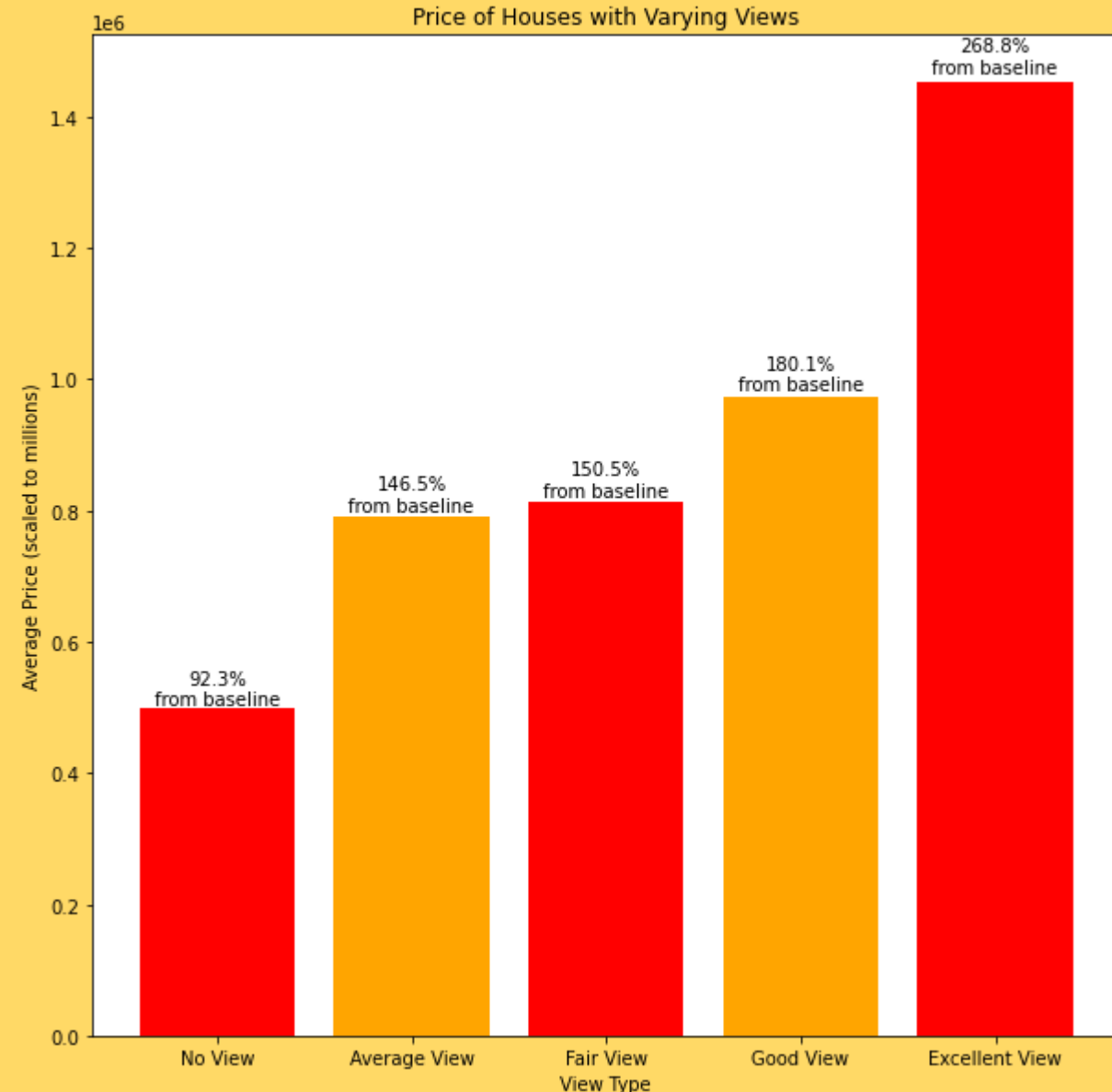
- We are examining the data set from 2014 to 2015. Some implied limitations of this set are: we don't know the buyers demographics, special deals, was the price truly representative of the price listed?



A picture of The houses sold (in red) in Kings County, done with Folium

Business Understanding

- As a Real Estate Company you are familiar with the concept of the hierarchy of prices when you **List** a house for sale, thus it is profitable for you to obtain information about houses that can sell easier.
- You can include the results of this analysis in your seller disclosure and the impact it has, justifying the cost of house based on its features
- The left is an example of % increase from the mean prices of the house based on view type
- Finally, you can use this information if the housing bubble burst



Data Understanding

- The Data consists of 21,597 of houses sold in Kings County in Washington US between 2014 and 2015. There are 20 attributes that may or may not potentially affect the price that we examine.
- External variables relate to the features outside the house. For instance, the view of the house could go from being in a mountain or a lake. Internal features are self explanatory: total living space of the house.
- We used pandas to manipulate the data and create graphs with Seaborn and MatPlot libraries. Folium is a library to generate maps. Sklearn and StatsMod were used to generate our regression. Scikit was used for verification.

OLS Regression Results						
=====						
Dep. Variable:	price_log	R-squared:	0.717			
Model:	OLS	Adj. R-squared:	0.717			
Method:	Least Squares	F-statistic:	2807.			
Date:	Sun, 19 Jun 2022	Prob (F-statistic):	0.00			
Time:	12:11:18	Log-Likelihood:	-2844.6			
No. Observations:	21062	AIC:	5729.			
Df Residuals:	21042	BIC:	5888.			
Df Model:	19					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	12.0122	0.033	367.822	0.000	11.948	12.076
sqft_living	0.0003	3.51e-06	76.345	0.000	0.000	0.000
sqft_living15	0.0002	4.67e-06	34.873	0.000	0.000	0.000
sqft_lot15_new_log	-0.0463	0.003	-16.789	0.000	-0.052	-0.041
condition Fair	-0.1277	0.022	-5.850	0.000	-0.170	-0.085

External Features

VIEW TYPE	WATERFRONT	GRADE	LOCATION
What is it defined as?	What is it defined as?	What is it defined as?	What is it defined as?
Quality of view of the house	If the house is on a waterfront	Relates to the construction and design of the house.	Longitude and latitude
What are some examples?	What are some examples?	What is the scale?	What are some examples?
Includes bays, ship canals, lakes and river/slough waterfronts	See picture below	3-13	As you go north, latitude increases. As you go east, longitude increases

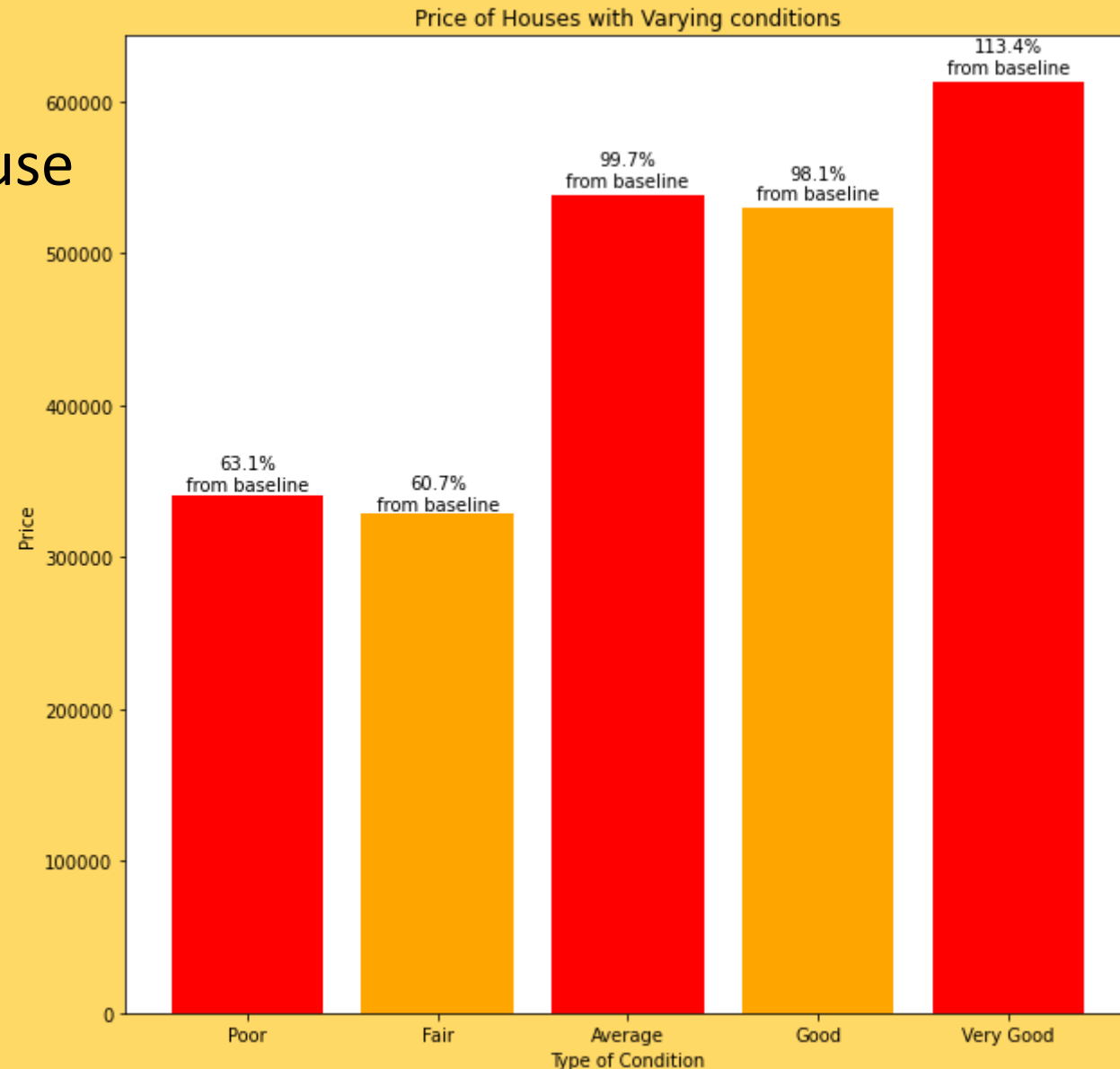


Internal Features

CONDITION	LOT SPACE	OTHER FEATURES
What is it defined as?	What is it defined as?	BASEMENT
How “good” the overall condition of the house is	The amount of space on the lot	Floors
What are some examples?	Why is this important?	Living space
Are the pipes good, does the house need further repairs?	Can indicate possibly the existence of a driveway, pool, etc	Bedrooms/bathrooms

Modeling:

- Examined how the price of the house was correlated to
 - Total Square Footage (TSF)
 - Total Square footage of houses of 15 nearest neighbors (TSF15)
 - Total Square footage of 15 nearest neighbor's lots (TSFL15)
 - Condition
 - View
 - Waterfront
 - Grade
 - Longitude and Latitude



Modelling and Regression Results

- The price of a house is negatively influenced by having no view of the environment
- The price of a house is only positively influenced by having a condition: “Very Good”
- Having a waterfront increases property value by almost 100% (520k vs 1.1m)
- Houses categorized by low grades negatively impact the price until it is “Grade 9: Better”
- 71.7% variation of the prices in this dataset can be explained by our coefficients.

Recommendations & Next Steps

CONDITION		WATERFRONT	ZIPCODE	LOCATION
Houses rated “very good” should be purchased		Having a waterfront triples your property value	There is no significant influence on the difference in price between houses in seattle vs non seattle	Houses more north and west are priced higher
Houses rated “average” and below should have a negative influenced on the price		Not having a waterfront means that the house is likely cheaper		This is because it’s closer to the waterfront properties
VIEW TYPE		GRADE	BASEMENT	
Not having a view drastically decreases the price		At Grade 9, this variable starts having a positive influence on price	Presence of a basement will increase the price	
As long as there is a view, expect the house to cost more			The average increase is around 100k	

Recap Conclusion

Q1

What was the problem?

There is an impending housing bubble burst and we need to evaluate whether external or internal features affect the price of the house more.

Q2

How did we do it?

We've built a Multi Linear Regression model that can create how much weight an attribute can influence a house price based on the coefficients

Q3

What attributes did we look at?

External: View , waterfront, grade, location, Sqft living of neighbors, longitude and latitude

Internal: Condition, lot space, basement, Sqft living,

Thank You!

- Git Hub link: <https://github.com/RaymondLeong94/dsc-phase-2-project-v2-3>