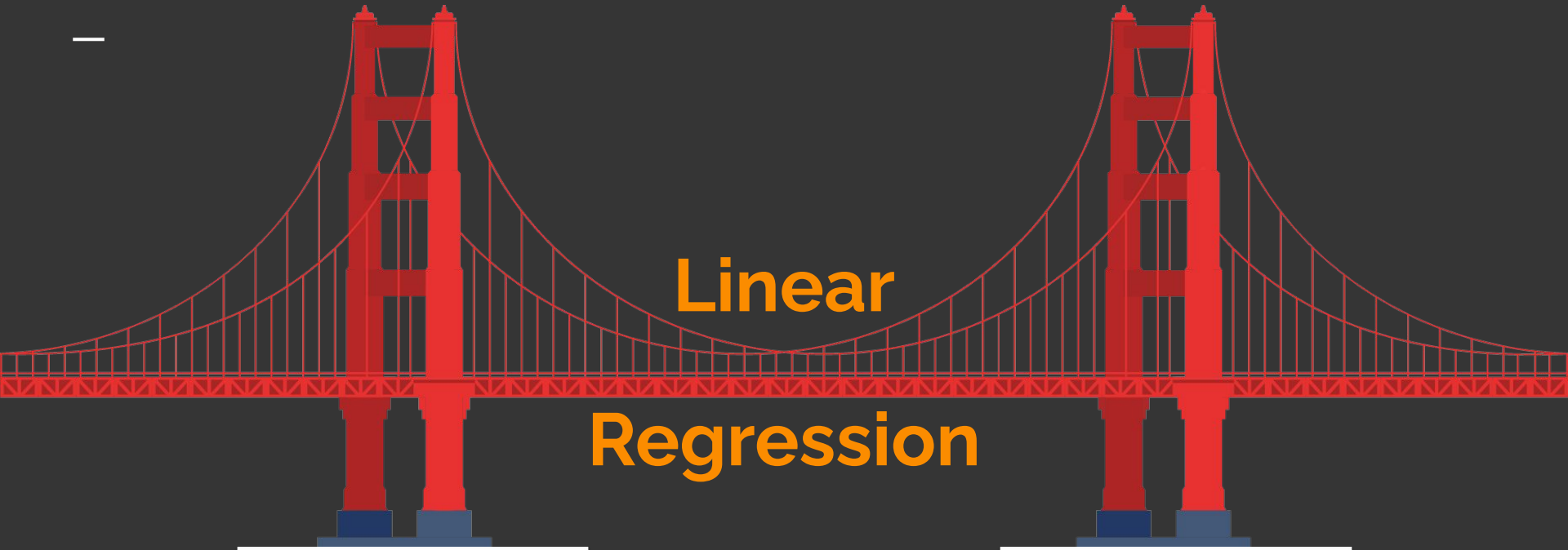
How renovations may increase estimated value of homes.

Linear Regression of King County House-Sales Dataset

Business Problem

The Real Estate Company needs to be able to advise homeowners on how to increase the value of their homes.

They need at least three features for which home renovations can accomplish this, and by how much.



Linear Regression

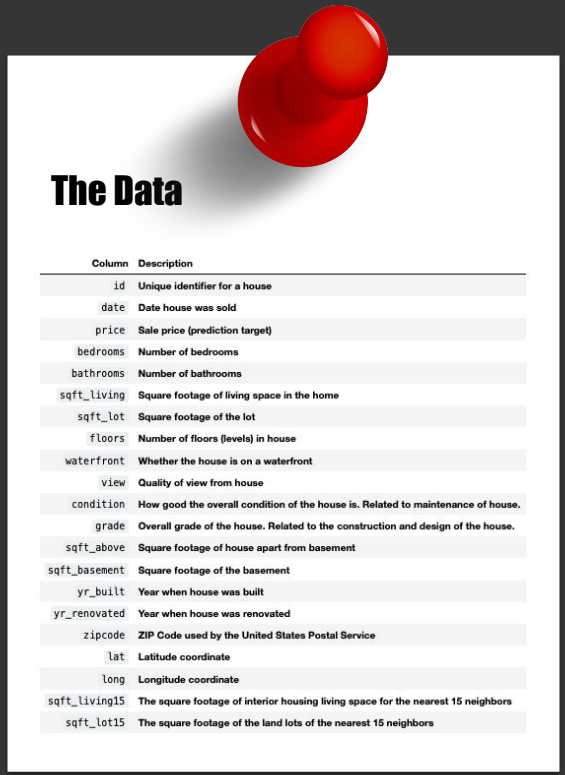
NEED:

Provide Advice to Homeowners

GOAL:

3 Features that Increase the Value of the Home

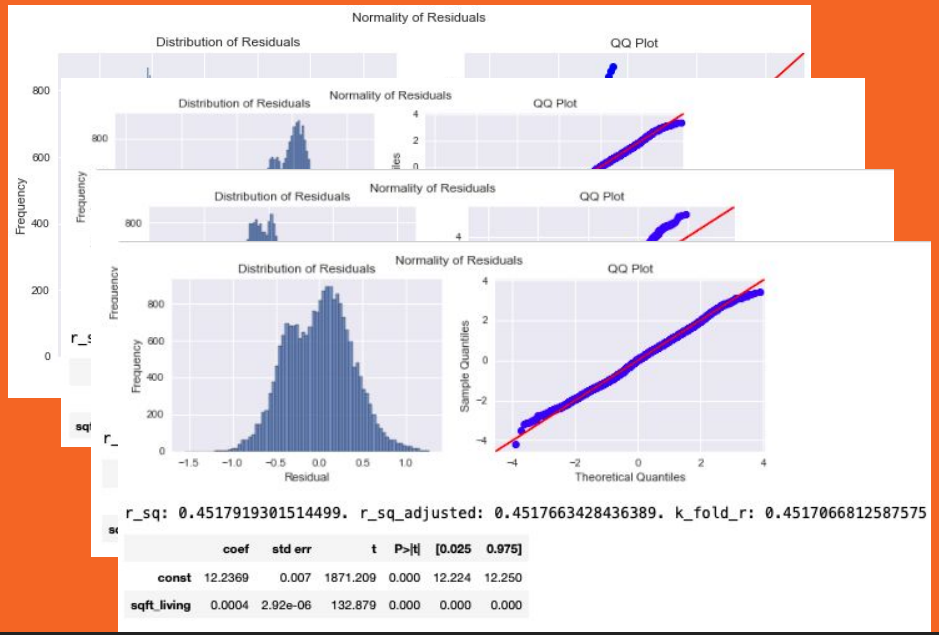
This Regression Analysis uses the King County House-Sales Dataset



The Data

Column	Description
id	Unique identifier for a house
date	Date house was sold
price	Sale price (prediction target)
bedrooms	Number of bedrooms
bathrooms	Number of bathrooms
sqft_living	Square footage of living space in the home
sqft_lot	Square footage of the lot
floors	Number of floors (levels) in house
waterfront	Whether the house is on a waterfront
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.
sqft_above	Square footage of house apart from basement
sqft_basement	Square footage of the basement
yr_built	Year when house was built
yr_renovated	Year when house was renovated
zipcode	ZIP Code used by the United States Postal Service
lat	Latitude coordinate
long	Longitude coordinate
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

Modeling with an iterative approach...



r_sq: 0.4517919301514499. r_sq_adjusted: 0.4517663428436389. k_fold_r: 0.4517066812587575

	coef	std err	t	P>|t|	[0.025	0.975]
const	12.2369	0.007	1871.209	0.000	12.224	12.250
sqft_living	0.0004	2.92e-06	132.879	0.000	0.000	0.000

UNTIL...

Final Model

Almost there

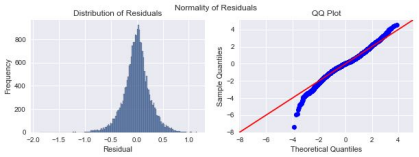
Drop 'sqft_above'

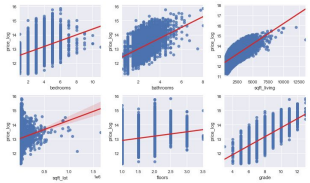
OLS Regression Results

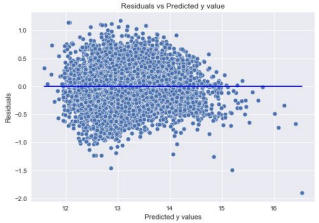
Dep. Variable:	price_log	R-squared:	0.760
Model:	OLS	Adj. R-squared:	0.760
Method:	Least Squares	F-statistic:	2011.
Date:	Tue, 28 Jun 2022	Prob (F-statistic):	0.00
Time:	13:51:02	Log-Likelihood:	-1368.1
No. Observations:	21596	AIC:	2806.
Df Residuals:	21561	BIC:	3086.
Df Model:	34		
Covariance Type:	nonrobust		

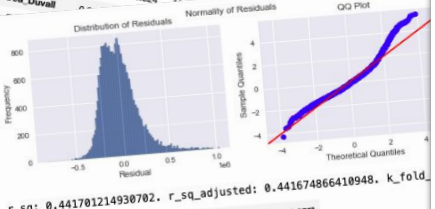
	coef	std err	t	P>|t|	[0.025	0.975]
const	10.9794	0.020	548.982	0.000	10.940	11.019
bedrooms	-0.0083	0.003	-3.281	0.001	-0.013	-0.003
bathrooms	0.0145	0.004	3.623	0.000	0.007	0.022
sqft_living	0.0002	3.98e-06	48.154	0.000	0.000	0.000
sqft_lot	4.861e-07	4.64e-08	10.476	0.000	3.95e-07	5.77e-07
floors	0.0570	0.004	12.891	0.000	0.048	0.066
grade	0.1514	0.003	58.162	0.000	0.146	0.156
view_EXCELLENT	0.1315	0.019	6.880	0.000	0.094	0.169
view_GOOD	0.0397	0.014	2.922	0.003	0.013	0.066
view_NONE	-0.1151	0.008	-14.922	0.000	-0.130	-0.100
condition_Fair	-0.1403	0.020	-6.989	0.000	-0.180	-0.101
condition_Good	0.0834	0.004	19.241	0.000	0.075	0.092
condition_Poor	-0.1483	0.048	-3.081	0.002	-0.243	-0.054
condition_Very Good	0.1803	0.007	26.318	0.000	0.167	0.194
has_basement_YES	0.0355	0.004	8.218	0.000	0.027	0.044
waterfront_YES	0.3303	0.026	12.512	0.000	0.279	0.382
season_sold_spring	0.0466	0.004	12.159	0.000	0.039	0.054
Location/Area_Bellevue Area	0.6932	0.008	83.146	0.000	0.677	0.710
Location/Area_Black Diamond	0.2726	0.027	10.231	0.000	0.220	0.325
Location/Area_Bothell Area	0.4356	0.013	32.518	0.000	0.409	0.462
Location/Area_Carnation	0.3326	0.024	13.661	0.000	0.285	0.380
Location/Area_Duvall	0.3465	0.020	17.458	0.000	0.308	0.385
Location/Area_Enumclaw	0.0749	0.018	4.081	0.000	0.039	0.111
Location/Area_Fall City	0.4629	0.030	15.530	0.000	0.404	0.521
Location/Area_Kent Area	0.0568	0.010	5.822	0.000	0.038	0.076
Location/Area_Maple Valley	0.1812	0.012	14.550	0.000	0.157	0.206
Location/Area_Medina	1.2126	0.037	32.517	0.000	1.139	1.286
Location/Area_Mercer Island	0.8252	0.017	48.783	0.000	0.792	0.858
Location/Area_Newcastle Area	0.2514	0.009	27.856	0.000	0.234	0.269
Location/Area_North Bend	0.3397	0.019	18.270	0.000	0.303	0.376
Location/Area_Sammamish Area	0.5556	0.008	66.889	0.000	0.539	0.572
Location/Area_Seattle Area	0.5544	0.007	78.612	0.000	0.541	0.568
Location/Area_Snoqualmie	0.4142	0.016	25.631	0.000	0.383	0.446
Location/Area_Vashon	0.3660	0.025	14.443	0.000	0.316	0.416
Location/Area_Woodinville	0.4721	0.014	34.726	0.000	0.445	0.499

Omnibus:	769.100	Durbin-Watson:	2.004
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1990.545
Skew:	-0.151	Prob(JB):	0.00
Kurtosis:	4.456	Cond. No.	1.21e+06









r_sq: 0.441701214930702. r_sq_adjusted: 0.441674866410948. k_fold_

	coef	std err	t	P>|t|	[0.025	0.975]
const	8.814e+04	3498.528	25.194	0.000	8.13e+04	9.5e+04
sqft_living	206.0861	1.592	129.475	0.000	202.966	209.206

Findings & Recommendations

CONDITION: Upgrade CONDITION of the home

From average or below to GOOD, will increase home value by 8.3%

To VERY GOOD, it will increase 17.87%

GRADE: Renovate to upgrade the GRADE of the home

For every increase in GRADE rating, the home value will increase by 14.71%

FLOOR: Add a FLOOR

For every additional floor, home values increase 5.89%

Grade:

Related to the construction and design of the house. Classification by construction quality which refers to the types of materials used and the quality of workmanship

Condition:

Related to maintenance of the house

Next Steps:

Remove outliers

Starting with 'sqft_lot' remove outliers from features one by one

Find price per square foot of living

Trying to group homes with similar homes, improving the relevancy of recommendations

Find bathrooms and/or bedrooms per square foot of living

Weeding out anomalies

Building on current model

Exploring new possibilities

Change CONDITION to binary

Add 'condition_above_average' to make it more interpretable

Divide homes by Urban vs Rural

Do Rural home buyers value different features compared to City buyers?

Compare square foot lot to square foot living

Thank You

Email: cassigroesbeck@emailplace.com

GitHub: @AgathaZareth

LinkedIn: linkedin.com/in/cassarra-groesbeck-a64b75229