King County Real Estate Analysis

Our findings, our narrative, our future



Our questions:

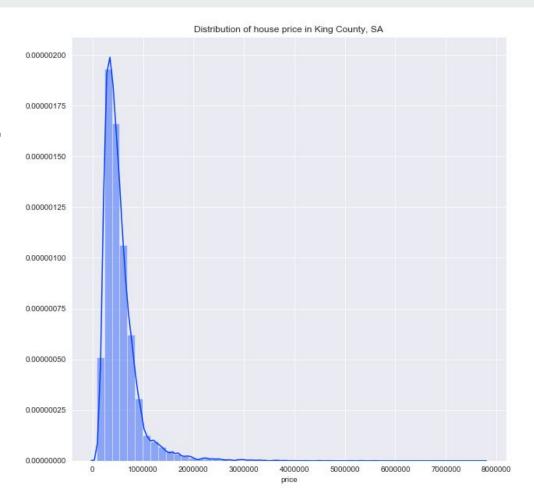
- 1. How accurate a predictor is the amount of square feet of living space?
- 2. Do more recently modified houses have higher prices?
- 3. Are there any clear geographical trends in price?

Stakeholder Overview

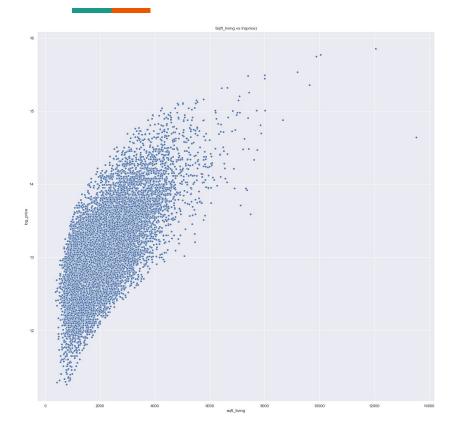
Real Estate Companies savills	Housing Development Firms Taylor Wimpey
Focus on maximising: Sales, Profits	Focussed on maximising sale price after spending on renovations or extensions
Focussed on efficiency in: Costs related to marketing and sales	Minimising cost when investing in house developments

Adjusting the data

- Non-normal distribution
- Data heavily skewed by outliers
- Took the logarithm of our price for final models



Size: An Efficient Price Predictor?



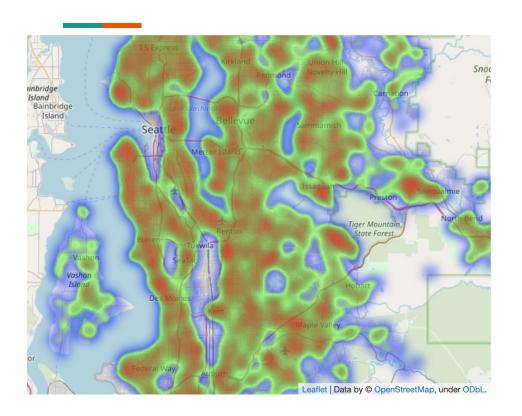
Living Space of a property Vs Price

- Initial data exploration showed a high correlation between living space and price
- Non-normality of price distribution led us to take the log
- Further regression analysis showed it to be the strongest predictor of house price, out of all variables included in our model

Location, location, location ...



- We spotted clusters of high prices around specific locational points
- Locational price data impacts both of our stakeholders:
 - Enabling real estate companies to appropriately price their properties
 - Calculation for a housing development profit margin, based on maximum cost per square foot of building houses or extensions
- This led us into further analysis and mapping

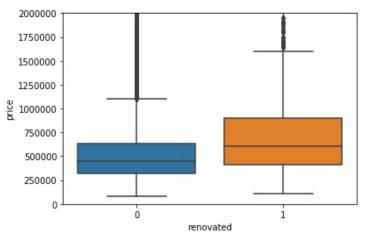


- Our initial visualisation suggested a cluster of high price points around a central area
- Mapping on price per square foot subsets the data, while removing variation based on total property size
- This sets a benchmark ceiling of spending on cost per square foot in various areas, for a housing development firm to profit after a build
- However, more in depth mapping showed there were actually multiple clusters of high priced areas across the dataset
- Key:

Blue: \$231.5 / sqft or lessGreen: \$231.5 - \$270 /sqft

Red::\$405/sqft +

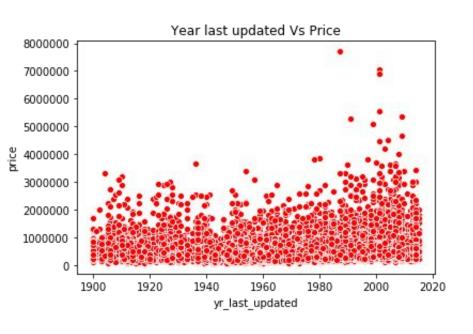
To renovate or not to renovate



Does Renovation affect house price?

- Initial data exploration and visualization showed that the subset of houses that had been renovated in the past were on average a higher price at point of sale
- This led us to include it in our predictive model, which proved less useful
- In conclusion: house renovations on average increase the price, however did not yield useful predictive qualities at a later stage

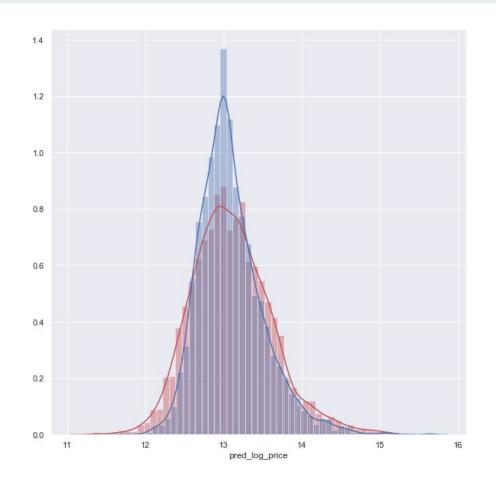
Further exploration around renovations



- Our previous slide led us to further analysis around renovations, and building age
- We created a new variable to test our hypothesis that buildings that had been recently updated (either through initial build or renovation) would be a higher price
- If this was the case, it would be useful to include in our predictive model
- While the scatter plot showed a potential positive correlation, it was very weak

Price Prediction Model

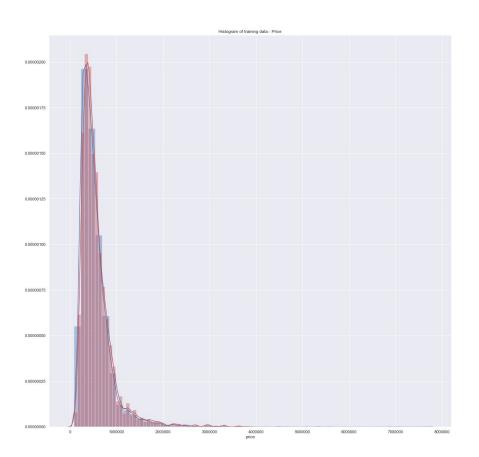
- 1. Both absolute price and cost per square foot are dependent on location
 - a. Average \$264/sqft across whole dataset
 - b. This relationship is not focussed around a central point
- 2. Doubling square footage from 1500 to 3000 results in 82.2% increase in price
- 3. Renovation increases price of 40-year old house by 0.8%; a 90-year old house by 1.8%.
- 4. We are quite confident in our results, however would have achieved similar using purely square footage other variables have a limited impact
- 5. Separate model for **housing development**

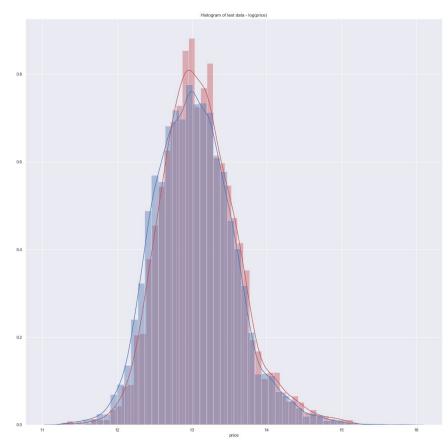


Thank you

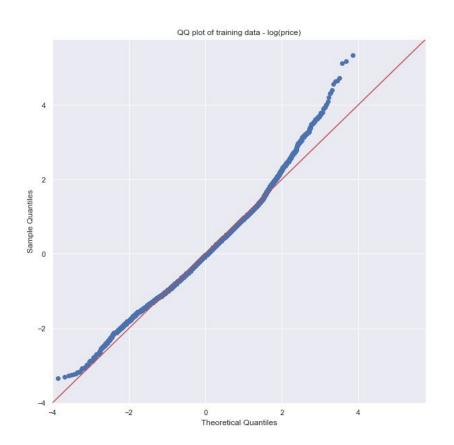
Questions, please

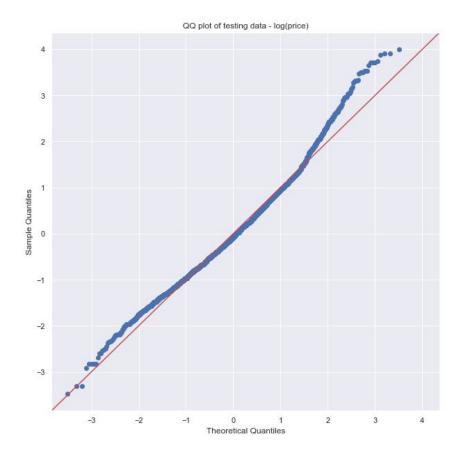
	Heatmap of correlation coefficients for all variables including new ones
а	1 -0.0170.00120.0052-0.012 -0.13 0.019-0.00420.012 -0.0240.0082-0.011 0.022 -0.0120.00820.00180.021-0.0027-0.14 -0.011 0.02 -0.00520.011 0.017 0.014-0.0038
price -	1017 1 031 0.53 0.7 0.09 0.26 0.28 0.4 0.036 0.67 0.61 0.054 0.13 0.053 0.31 0.022 0.59 0.083 0.12 0.098 0.56 0.23 0.2 0.19 0.89
bedrooms 0	0012 031 1 0.51 0.58 0.032 0.18 0.00240.079 0.026 0.36 0.48 0.16 0.018 0.15 0.01 0.13 0.39 0.031 0.018 0.16 0.21 0.071 0.095 0.09 0.34
bathrooms	0052 053 051 1 0.76 0.088 05 0.067 0.19 0.13 0.67 0.69 0.51 0.051 0.2 0.024 0.22 0.57 0.088 0.047 0.53 0.09 0.12 0.14 0.15 0.55
sqft_living -	1.012 0.7 0.58 0.76 1 0.17 0.35 0.11 0.28 0.059 0.76 0.88 0.32 0.056 0.2 0.052 0.24 0.76 0.18 0.051 0.34 0.091 0.12 0.14 0.15 0.7
.sqft_lot	0.13 0.09 0.032 0.088 0.17 1 0.00480023 0.075-0.0088 0.11 0.16 0.053 0.0045 0.13 0.086 0.23 0.14 0.72 0.0051 0.052 0.035 0.26 0.24 0.26 0.1
floors	019 026 018 05 035 0.0048 1 0022 0.028 0.26 0.46 052 0.49 0.0035 0.06 0.049 0.13 028 0.0110.0037 0.5 0.0047 0.037 0.039 0.048 0.31
waterfront -	0042 028 0.00240067 0.11 0.023 0.022 1 0.41 0.018 0.087 0.075 0.026 0.087 0.031 0.013 0.04 0.089 0.032 0.08 0.0073 0.2 0.012 0.0083 0.015 0.18
view	012 0.4 0.079 0.19 0.28 0.075 0.028 0.41 1 0.046 0.25 0.17 0.055 0.1 0.085 0.0061-0.078 0.28 0.073 0.09 0.027 0.22 0.061-0.059 0.07 0.35
andition -	0.024 0.036 0.026 -0.13 -0.0590 0.088 -0.26 0.018 0.046 1 -0.15 -0.16 -0.36 -0.0620 0.029 -0.015 -0.11 -0.0930 0.0340 0.05 -0.39 0.1 -0.09 -0.085 -0.092 0.039
grade 0	0082 067 036 067 076 011 046 0087 025 0.15 1 0.76 0.45 0.017 0.15 1 0.76 0.45 0.017 0.19 0.11 0.2 0.71 0.12 0.015 0.46 0.13 0.057 0.074 0.086 0.7
sqft_above	011 061 048 069 088 018 052 0075 017 016 076 1 042 0022 026-00012 034 073 02 0021 043 0.088 024 026 027 06
yr_built	022 0054 016 051 032 0053 049 0.026 0.055 0.36 045 0.42 1 4.023 0.051 0.41 0.33 0.071 0.2 033 0.29 0.42 0.41 0.43 0.081
у <u>r</u> renovated	0.012 0.13 0.018 0.051 0.056 0.0045 0.0035 0.087 0.1 0.062 0.017 0.022 0.23 1 0.07 0.032 0.0721 0.0026 0.039 1 0.16 0.11 0.085 0.08 0.087 0.12
zipcode -4	0082 0 053 - 0.15 - 0.2 - 0.2 - 0.13 - 0.06 - 0.031 - 0.085 0 0 0 29 - 0.19 - 0.26 - 0.35 - 0.07 - 1 - 0.27 - 0.56 - 0.28 - 0.15 - 0.062 - 0.32 - 0.17 - 0.54 - 0.58 - 0.039
tat -	0018 031 0.01 0.024 0.052 0.086 0.049 0.0130.0061 0.015 0.11 0.0012 0.15 0.032 0.27 1 0.14 0.049 0.086 0.028 0.14 0.47 0.57 0.65 0.51 0.45
long	021 0022 0.13 0.22 0.24 0.23 0.13 0.04 0.078 0.11 0.2 0.34 0.41 0.072 0.56 0.14 1 0.34 0.25 0.065 0.39 0.24 0.78 0.8 0.84 0.051
sqt_living15 -	0027 0.59 0.39 0.57 0.76 0.14 0.28 0.089 0.28 -0.093 0.71 0.73 0.33 0.000250.28 0.049 0.34 1 0.18 0.00062 0.33 0.039 0.19 0.22 0.23 0.62
sqft_lot15	0.14 0.083 0.031 0.088 0.18 0.72 -0.011 0.032 0.073-0.0031 0.12 0.2 0.071 0.0039 0.15 0.086 0.26 0.18 1 0.0044 0.089 0.059 0.29 0.27 0.29 0.092
renovated -	0.011 0.12 0.018 0.047 0.051 0.0051 0.0057 0.08 0.09 0.055 0.015 0.021 0.02 1 0.062 0.028 0.0651 0.0065 0.0044 1 0.15 0.098 0.075 0.071 0.077 0.11
yr_last_updated	0.02 0.098 0.16 0.53 0.34 0.052 0.5 0.00730.027 0.39 0.46 0.43 0.33 0.16 0.32 0.14 0.39 0.33 0.069 0.15 1 0.26 0.39 0.39 0.4 0.120.3
price_per_sqt -4	0052 056 -0.21 -0.09 -0.091 -0.0350 0047 02 022 0.1 0.13 -0.088 -0.29 0.11 0.17 0.47 -0.24 0.039 -0.059 0.098 -0.26 1 -0.51 -0.5 -0.48 0.58
dst_highest_price_prop	011 -0.23 0.071 0.12 0.12 0.26 0.037 -0.012 -0.061 -0.09 0.057 0.24 0.42 -0.085 -0.54 -0.57 0.79 0.19 0.29 -0.075 0.39 -0.51 1 0.97 0.99 -0.27
dist_highest_pricepersqft	017 -0.2 0.095 0.14 0.14 0.24 0.039-0.0083-0.059-0.085 0.074 0.26 0.41 0.08 -0.58 -0.65 0.8 0.22 0.27 0.071 0.39 -0.5 0.97 1 0.97 -0.26
dst_from_needle	014 0.19 0.09 0.15 0.15 0.26 0.048 0.015 0.07 0.092 0.086 0.27 0.43 0.087 0.58 0.51 0.84 0.23 0.29 0.077 0.4 0.48 0.99 0.97 1 0.23
bg_price -{	0038 089 034 055 07 01 031 018 035 0039 07 06 0091 012 0039 045 0051 062 0092 011 012 058 027 026 023 1
	by the second se
	Self No. 1 Self No. 2
	Get High











Multivariate model summary

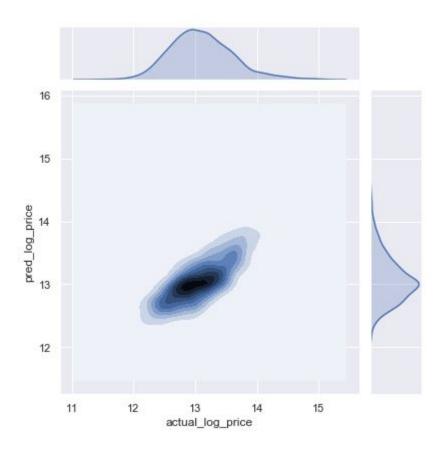
Dep. Variable:	price	R-squared:	0.610
Model:	OLS	Adj. R-squared:	0.610
Method:	Least Squares	F-statistic:	9019.
Date:	Wed, 23 Oct 2019	Prob (F-statistic):	0.00
Time:	14:26:01	Log-Likelihood:	-5372.8
No. Observations:	17276	AIC:	1.075e+04
Df Residuals:	17272	BIC:	1.078e+04
Df Model:	3		

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]			
const	12.0262	0.207	58.151	0.000	11.621	12.432			
sqft_living	0.0004	2.97e-06	144.129	0.000	0.000	0.000			
dist_highest_pricepersqft	-1.3372	0.019	-70.648	0.000	-1.374	-1.300			
yr_last_updated	pdated 0.0002 0.		2.212	0.027	2.69e-05	0.000			
Omnibus:	220.331	Durbi	n-Watson:	1.995					
Prob(Omnibus):	0.000	Jarque-	Bera (JB):	390.197					
Skew:	-0.055		Prob(JB):	1.86e	1.86e-85				
Kurtosis:	3.728		Cond. No.	2.41e+05					

Measures of dispersion for model on test data

	actual_log_price	pred_log_price	error
count	4321.000000	4321.000000	4321.000000
mean	13.104645	13.093593	0.244508
std	0.512552	0.419085	0.185173
min	11.326596	11.690549	0.000230
25%	12.751300	12.814923	0.105758
50%	13.071070	13.031974	0.209871
75%	13.415033	13.310939	0.339736
max	15.150512	15.645168	1.423324



Multivariate model (log(sqft_living) summary

Dep. Variable:	price	R-squared:	0.594		coef	std err	t	P> t	[0.025	0.975]
Model:	OLS	Adj. R-squared:	0.594	const	6.5085	0.046	141.241	0.000	6.418	6.599
Method:	Least Squares	F-statistic:	1.262e+04			0.000				0.000
Date:	Wed, 23 Oct 2019	Prob (F-statistic):	0.00	log_sqft_living dist_highest_pricepersqft		0.006	147.999 -77.939		0.899	0.923
Time:	14:26:02	Log-Likelihood:	-5734.0							
No. Observations:	17276	AIC:	1.147e+04	Omnibus: 48.480	Durbin-Wa	atson:	2.001			
Df Residuals:	17273	BIC:	1.150e+04	Prob(Omnibus) 0.000	Jarque-Bera	a (JB):	49.760			
Df Model:	2			Skew: 0.116	Pro	b(JB):	1.57e-11			
Covariance Type:	nonrobust			Kurtosis: 3.124	Con	ıd. No.	138.			