ROYAL HOMES

KING COUNTY HOUSING LINEAR REGRESSION ANALYSIS

BY NATALYA DORIS

BUSINESS PROBLEM

Real estate company Royal Homes is looking to better understand the King County housing market before they open up shop.

We seek to answer the following questions on behalf of Royal Homes: what types of homes should they be looking to sell to make the most profit? What features lend towards higher sale prices?

THE DATA

This project uses the King County House Sales dataset, which includes the following columns, among others:

- 'price': sale price
- 'waterfront': whether the house is on the waterfront
- 'view': quality of view from house
- 'condition': how good the overall condition of the house is; related to maintenance of house
- 'zipcode': ZIP Code used by the USPS
- 'sqft_living': sq. ft. of living space

Distribution of Kings County House Prices
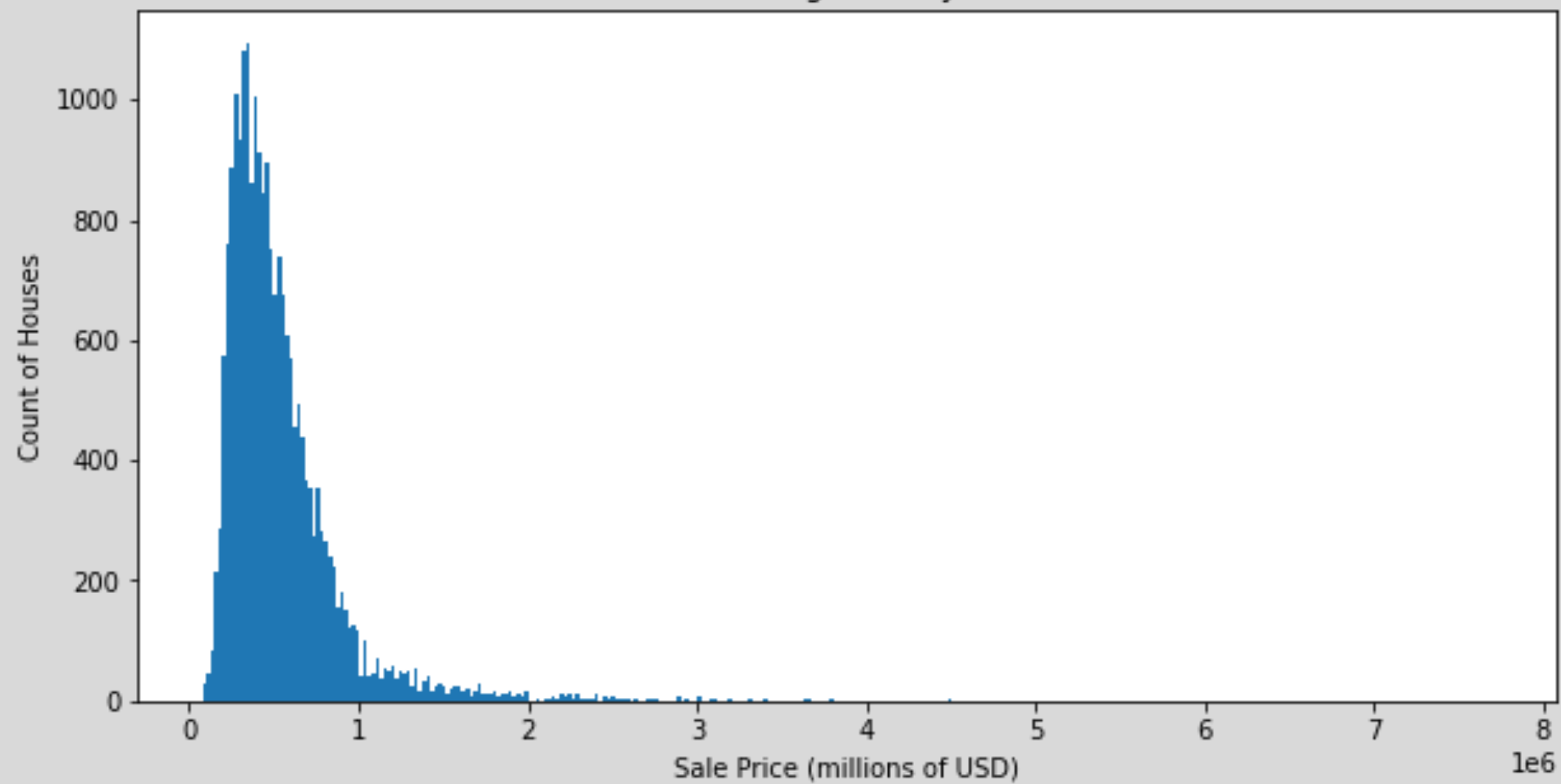
THE MODEL

To better understand how various attributes of a home influence sale price, we perform a multiple linear regression model with the King County data. We determine that **sq. ft. of living space**, **zip code**, **waterfront** status, **view** and **condition** are among the most influential factors in determining sale price.

MODEL SELECTION
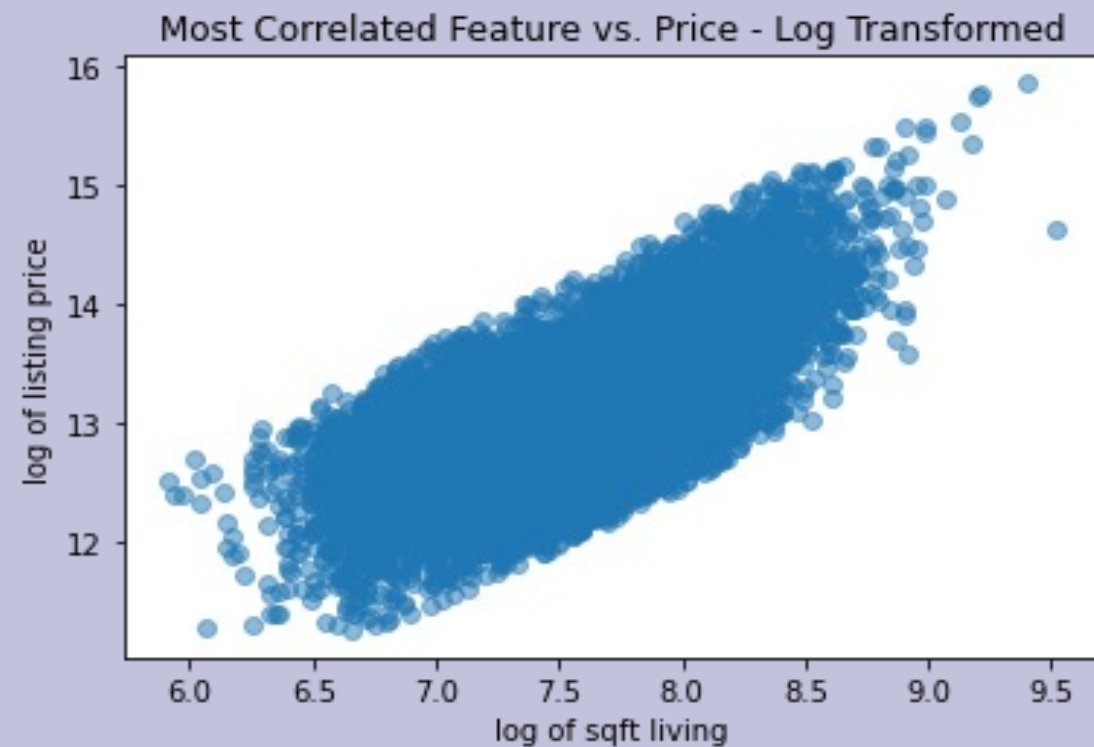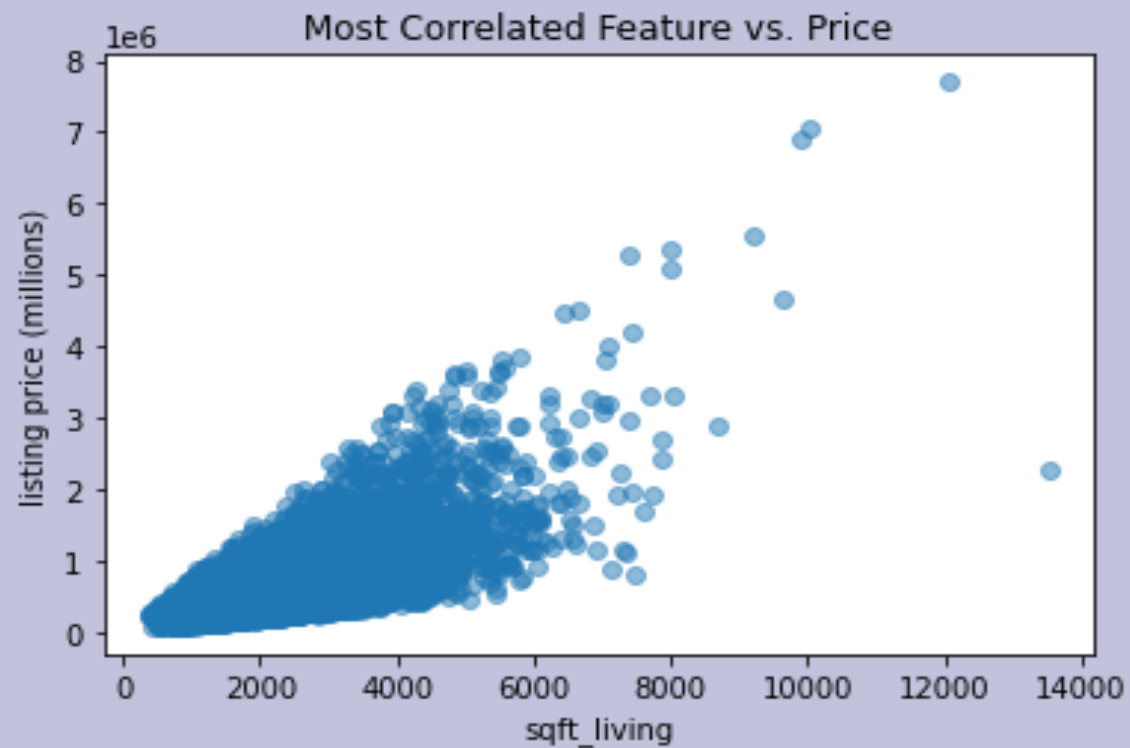
MODEL RESULTS

OLS Regression Results

Dep. Variable:	price	R-squared:	0.845
Model:	OLS	Adj. R-squared:	0.845
Method:	Least Squares	F-statistic:	1547.
Date:	Sun, 31 Jul 2022	Prob (F-statistic):	0.00
Time:	22:48:28	Log-Likelihood:	3357.1
No. Observations:	21597	AIC:	-6560.
Df Residuals:	21520	BIC:	-5946.
Df Model:	76		
Covariance Type:	nonrobust		

Omnibus:	935.360	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3142.856
Skew:	0.053	Prob(JB):	0.00
Kurtosis:	4.866	Cond. No.	281.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	7.5260	0.031	239.220	0.000	7.464	7.588
sqft_living_log	0.6799	0.004	175.090	0.000	0.672	0.687
waterfront_YES	0.4487	0.021	21.034	0.000	0.407	0.491
condition_Fair	-0.1267	0.016	-7.858	0.000	-0.158	-0.095
condition_Good	0.0078	0.003	2.315	0.021	0.001	0.014
condition_Poor	-0.2708	0.039	-6.992	0.000	-0.347	-0.195
condition_VeryGood	0.0564	0.005	10.367	0.000	0.046	0.067
view_EXCELLENT	0.2570	0.015	16.660	0.000	0.227	0.287
view_GOOD	0.1044	0.011	9.547	0.000	0.083	0.126
view_NA	-0.0944	0.027	-3.519	0.000	-0.147	-0.042
view_NONE	-0.1555	0.006	-24.952	0.000	-0.168	-0.143
zip_98003	0.0460	0.014	3.337	0.001	0.019	0.073
zip_98004	1.2450	0.013	94.203	0.000	1.219	1.271
zip_98005	0.8490	0.017	49.425	0.000	0.815	0.883
zip_98006	0.7570	0.011	67.425	0.000	0.735	0.779

RECOMMENDATIONS

- Find properties of larger size in advantageous zip codes
- Waterfront properties and properties with good views sell at higher prices
- Condition matters. A house in poor condition will yield a lower sale price

NEXT STEPS

1. Do a deep dive on top zip codes – what attributes do these houses most commonly feature?

2. Find data that breaks down 'view' into more specific categories – water view vs city view vs greenery, for example

3. Gather data on sales in more recent years – are there any attributes that are more impactful on sale price now?