



# King County Home Price Predictors - An Analysis

By Sameeha Ramadhan - March 2021



# The Data

The King County Housing dataset contains valuable information (such as location, square footage, condition, etc.) about properties in King County, Washington.

---

The final cleaned data contains:

1. 32,732 home sales
2. 22 variables



# Overview

**This analysis will seek to answer a few important questions such as:**

**Which features most strongly affect price?**

**What are some of the features that appear to have a relationship with King County housing sale prices?**

**Are certain claims made by real estate professionals, such as the fact that higher square footage increases home sale price, valid?**





# The Process

1

**Data cleaning and pre-processing:**  
Preparing the data for model inclusion

2

**Exploring and investigating the data:** determining which features should be included.

3

**Building linear models:** testing different features to get the best results

4

**Interpreting the results:** validating claims and making conclusions



## Correlations: For Choosing Features

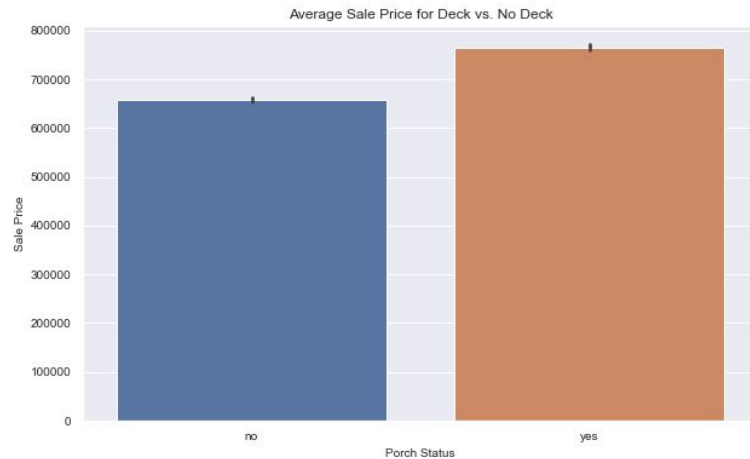
Features highly  
correlated with price  
were considered for  
inclusion.

saleprice	1.000000
bldggrade	0.602441
sqfttotliving	0.582776
township	0.469307
bathfullcount	0.285041
bedrooms	0.269357
sqfttotbasement	0.259356
sqftfinbasement	0.250262
finbasementgrade	0.235648
sqftdeck	0.215990
sqftopenporch	0.189520
bathhalfcount	0.182185
bath3qtrcount	0.175900
heatsystem	0.164649
yrrenovated	0.119026
sqftgarageattached	0.103858
inadequateparking	0.039884
sqftenclosedporch	0.010737
yrbuilt	0.005304
trafficnoise	0.003980
zipcode	-0.002015
condition	-0.002719



# Validating Claims

Home owners appreciate having the ability to enjoy the outdoors while maintaining a slight sense of seclusion, a claim that our model indicated as valid.





## Final Model & Results

Final R-Squared

**R = .685**

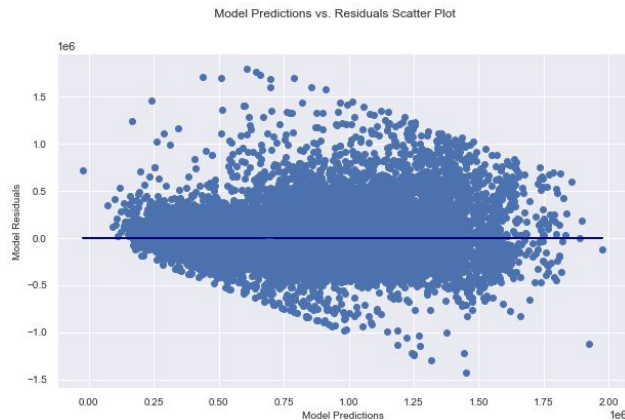
The model can account for about 69% of variability in price.

Final P-Value

**P < 0.05**

The null hypothesis that there is no relationship between price and the predictor values is rejected.

The final model included living space size, bathroom count, deck and enclosed porch sizes, building grade and township (location).





# Recommendations

1

**Consider the location of the home first (most important!)**

2

**The grade of a house has the *greatest* effect on it's value. Choose quality!**

3

**Converting the garage to a bedroom is a good decision. More living space = higher price.**

4

**Adding a deck is a profitable decision.**





# Thank you.

Email: [shramadhan@gmail.com](mailto:shramadhan@gmail.com)

Github: [@samtuleen](https://github.com/samtuleen)

LinkedIn: [linkedin.com/in/sameeha-ramadhan-3a1bba140](https://www.linkedin.com/in/sameeha-ramadhan-3a1bba140)

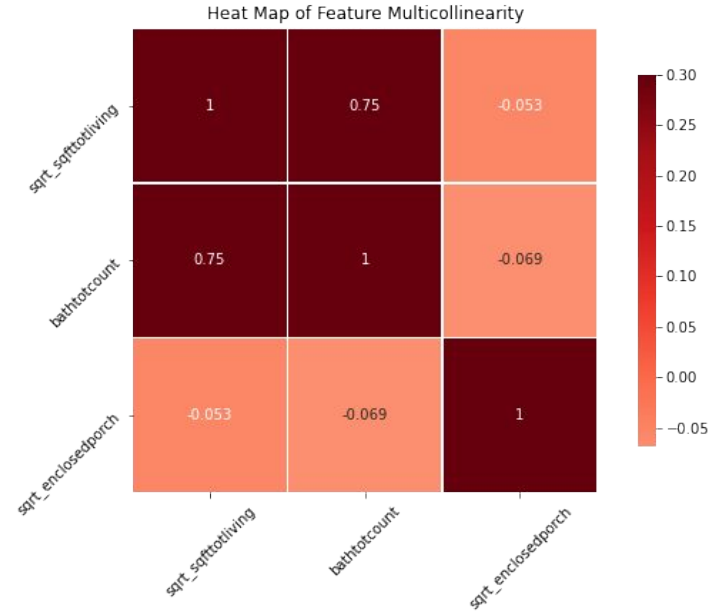




# Appendix



**Two features are highly correlate, however the maximum correlation value is 0.3 on the scale, so they have a 0.3 correlation only, which doesn't mean multicollinearity.**





# Appendix



After running a train test on the model, our final R-squared value equaled 1.

```
In [119]: X_train = sm.add_constant(X_train)
          smmodel = sm.OLS(y_train, X_train).fit()
          smmodel.summary()
```

Out[119]:

OLS Regression Results

Dep. Variable:	saleprice	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	3.320e+31
Date:	Thu, 11 Mar 2021	Prob (F-statistic):	0.00
Time:	12:03:11	Log-Likelihood:	4.6188e+05
No. Observations:	24541	AIC:	-9.237e+05
Df Residuals:	24503	BIC:	-9.234e+05
Df Model:	37		
Covariance Type:	nonrobust		