HOUSE SALE PRICES PREDICTION

Nopasorn Kowathanakul

THE PROJECT

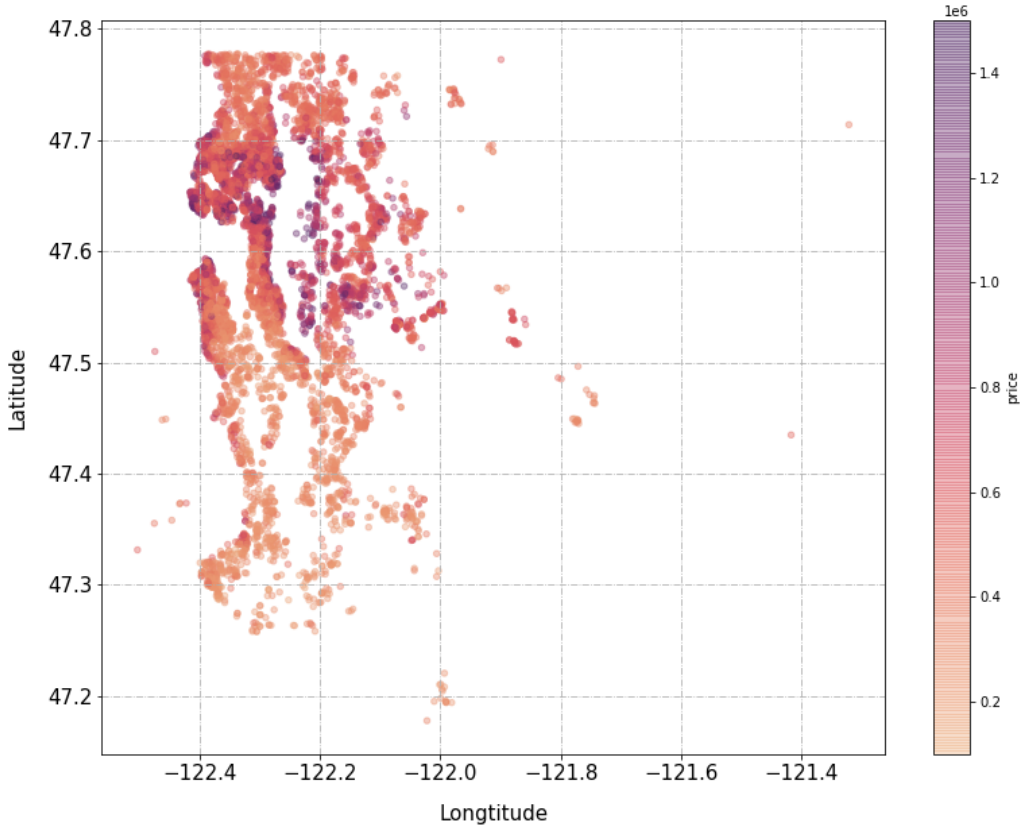
From the sales data in King County, I will do the data analysis and select influence factors for making a prediction model. Then, the following questions will be answered:

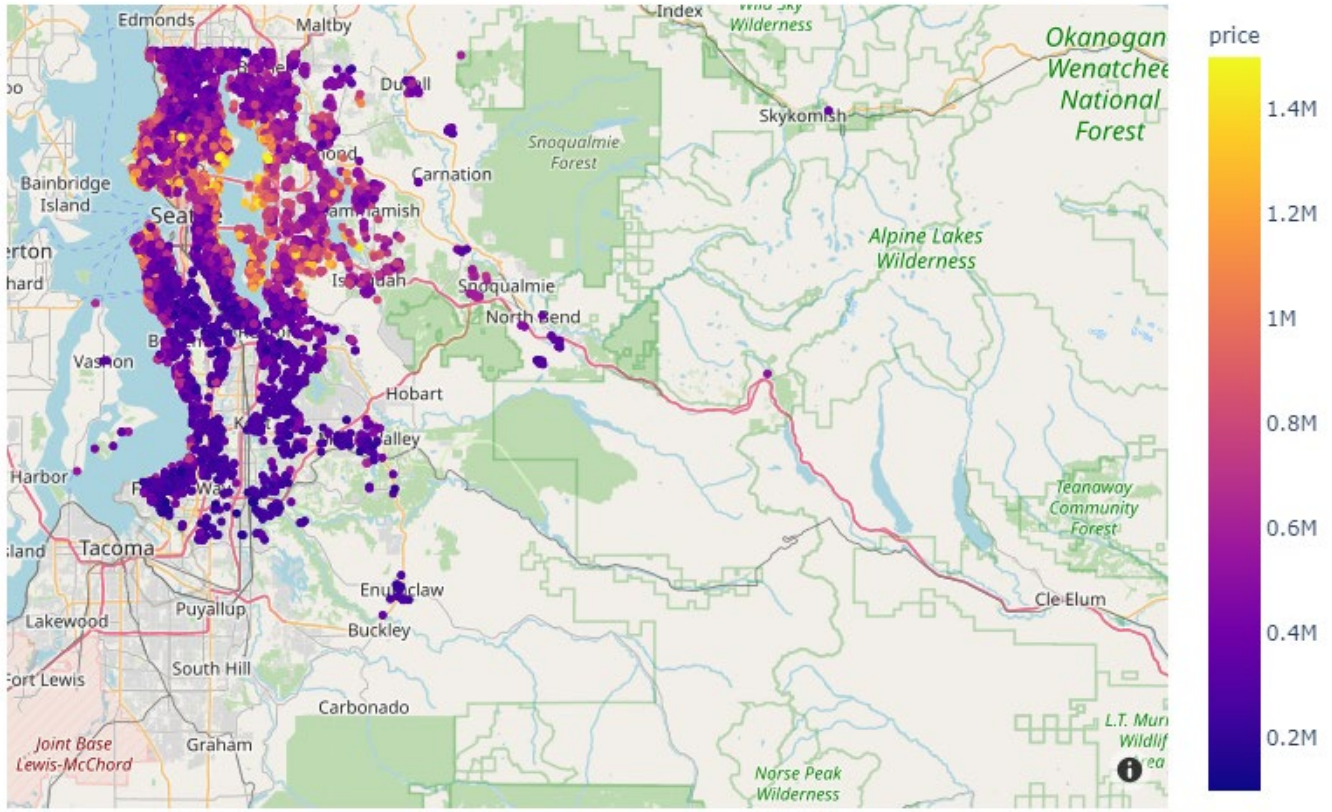
1. How does the location affect the price?
2. Do built and renovated times affect the price?
3. Does selling time affect the price?
4. Can we predict the price from a condition?

The Dataset

KING COUNTY HOUSE SALES DATASET

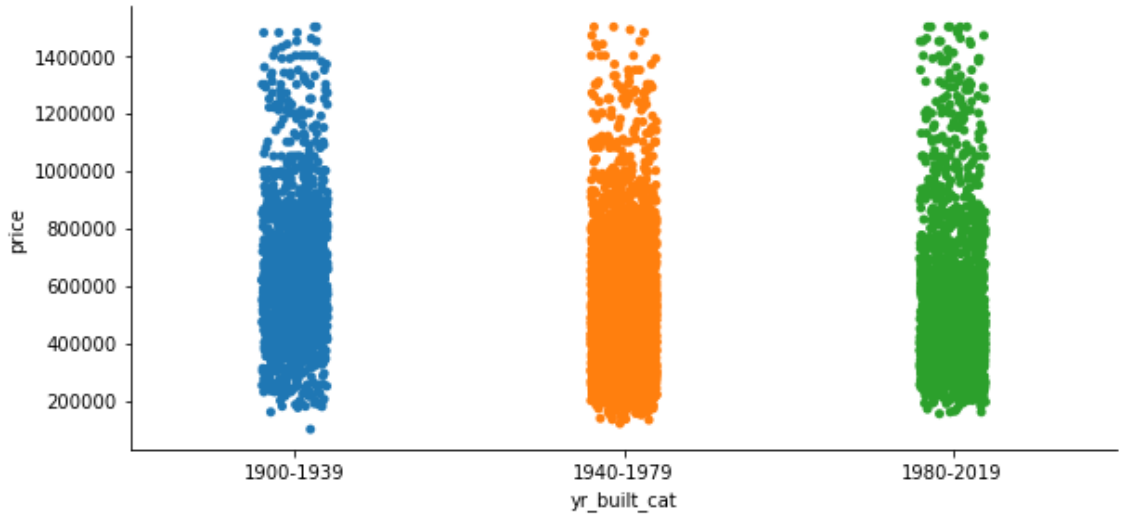
LOCATION EFFECT



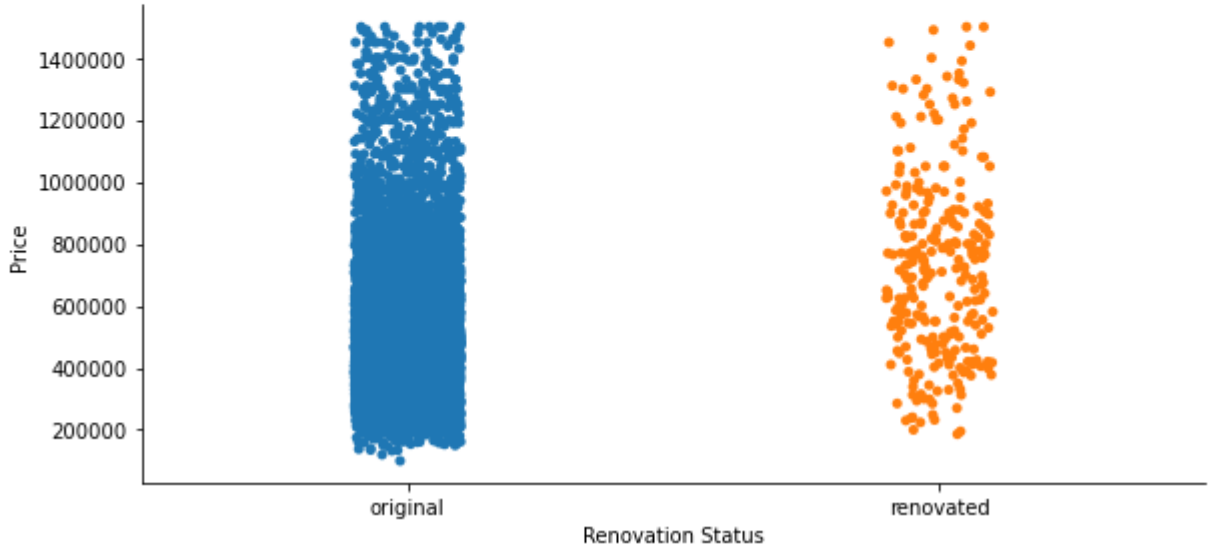


BUILT AND RENOVATED TIMES

```
Mean prices for yr_built_cat
1900-1939    639911.185238
1940-1979    503076.052348
1980-2019    538176.536572
Name: price, dtype: float64

Median prices for yr_built_cat
1900-1939    597000.0
1940-1979    459975.0
1980-2019    475000.0
Name: price, dtype: float64
```

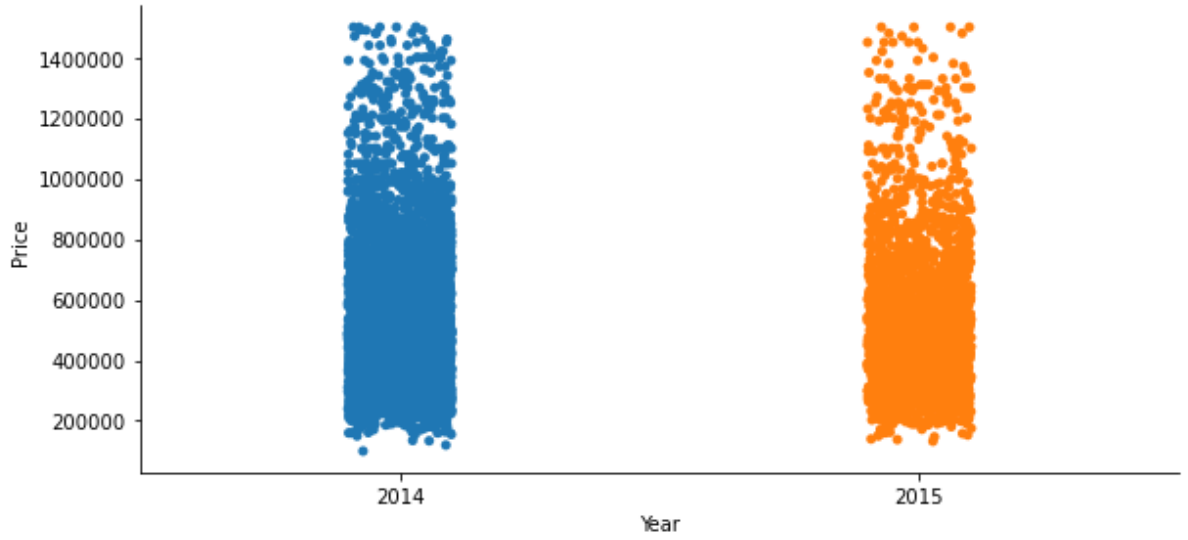



```
Mean prices for renovated
original     532330.073109
renovated    731288.941176
Name: price, dtype: float64

Median prices for renovated
original     484000.0
renovated    721000.0
Name: price, dtype: float64
```

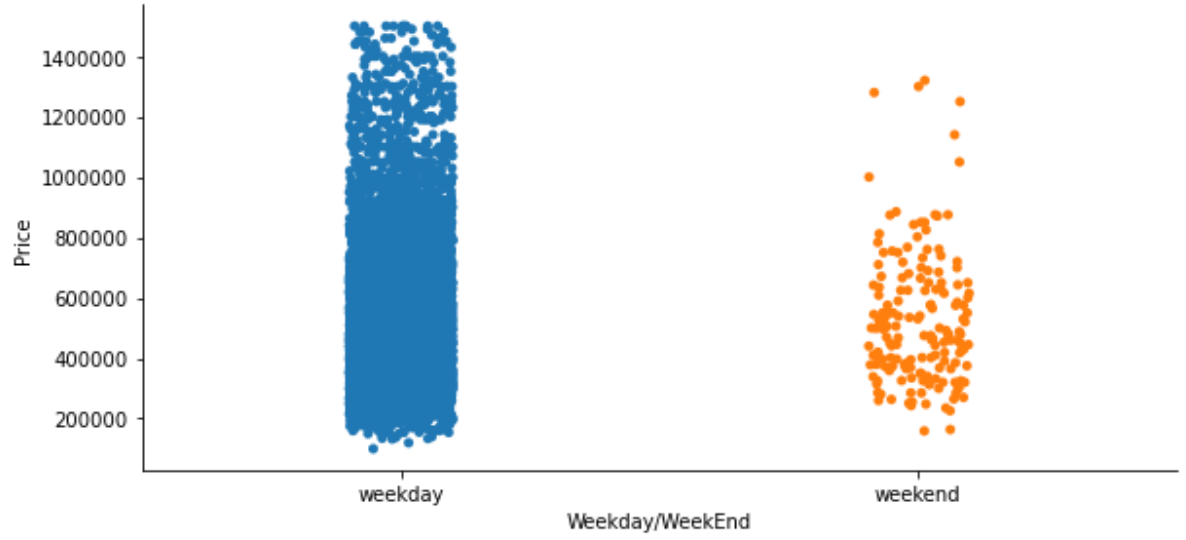


SELLING TIME

```
Mean prices for year
2014    538130.161706
2015    544165.609507
Name: price, dtype: float64

Median prices for year
2014    489000.0
2015    499000.0
Name: price, dtype: float64
```




```
Mean prices for day_cat
weekday    540430.406543
weekend    525963.441860
Name: price, dtype: float64

Median prices for day_cat
weekday    490000.0
weekend    484000.0
Name: price, dtype: float64
```



PRICE PREDICTION

Est.Price = 0.23x sqft_living + 0.12x sqft_above + 0.12x sqft_basement + 0.09x sqft_living15
- 0.05x bedrooms - 0.08x bathrooms + 0.33x grade + 0.22x condition +0.11x floors

OLS Regression Results

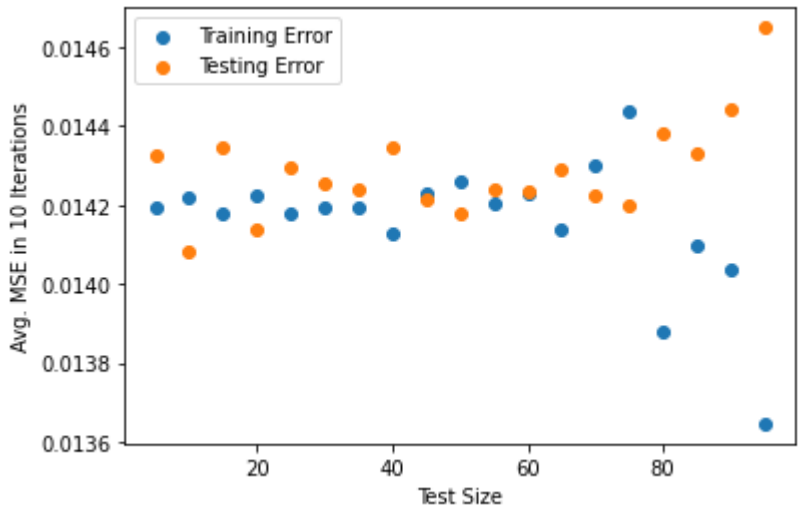
Dep. Variable:	price	R-squared (uncentered):	0.962
Model:	OLS	Adj. R-squared (uncentered):	0.962
Method:	Least Squares	F-statistic:	1.568e+04
Date:	Mon, 24 May 2021	Prob (F-statistic):	0.00
Time:	20:31:38	Log-Likelihood:	3965.2
No. Observations:	5590	AIC:	-7912.
Df Residuals:	5581	BIC:	-7853.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
sqft_living	0.2066	0.031	6.655	0.000	0.146	0.268
sqft_above	0.0447	0.031	1.438	0.151	-0.016	0.106
sqft_basement	0.0574	0.023	2.450	0.014	0.011	0.103
sqft_living15	0.1871	0.016	11.872	0.000	0.156	0.218
bedrooms	-0.0697	0.014	-5.154	0.000	-0.096	-0.043
bathrooms	-0.0634	0.013	-4.727	0.000	-0.090	-0.037
grade	0.3986	0.018	22.281	0.000	0.364	0.434
condition	0.2557	0.013	19.667	0.000	0.230	0.281
floors	0.1125	0.008	14.657	0.000	0.097	0.128

Omnibus:	65.643	Durbin-Watson:	2.009
Prob(Omnibus):	0.000	Jarque-Bera (JB):	62.470
Skew:	-0.227	Prob(JB):	2.72e-14
Kurtosis:	2.750	Cond. No.	52.6

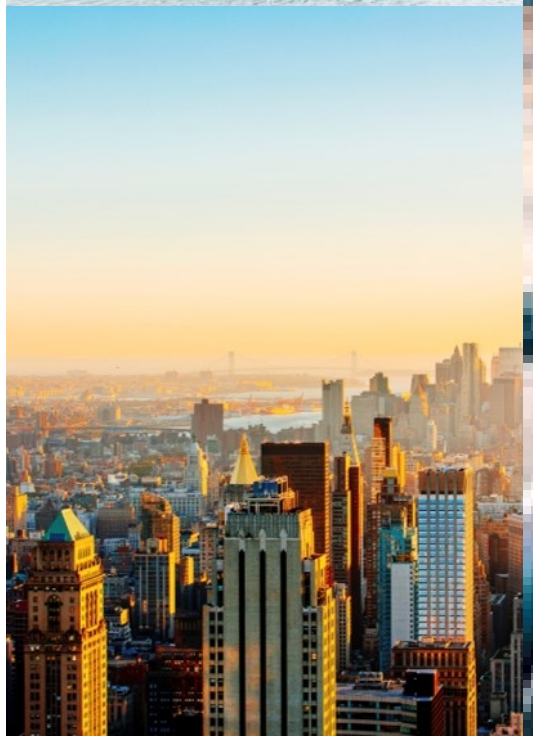
Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

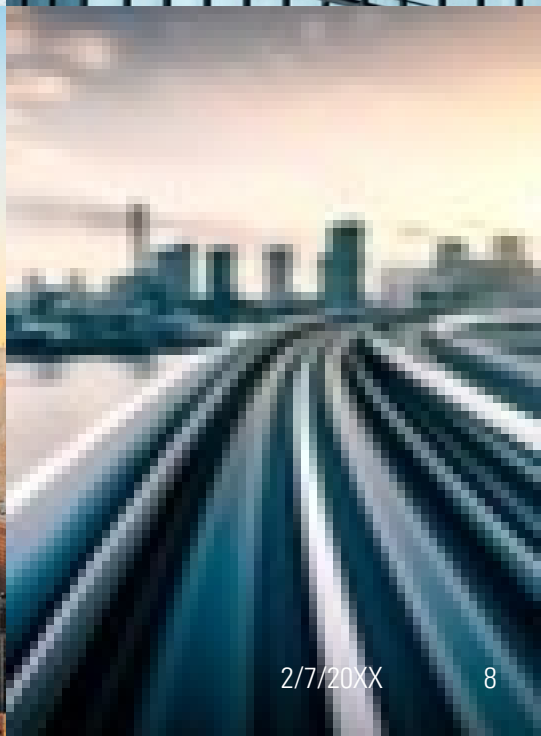
Train Mean Squarred Error: 0.01417106149621563
Test Mean Squarred Error: 0.014334953915793956

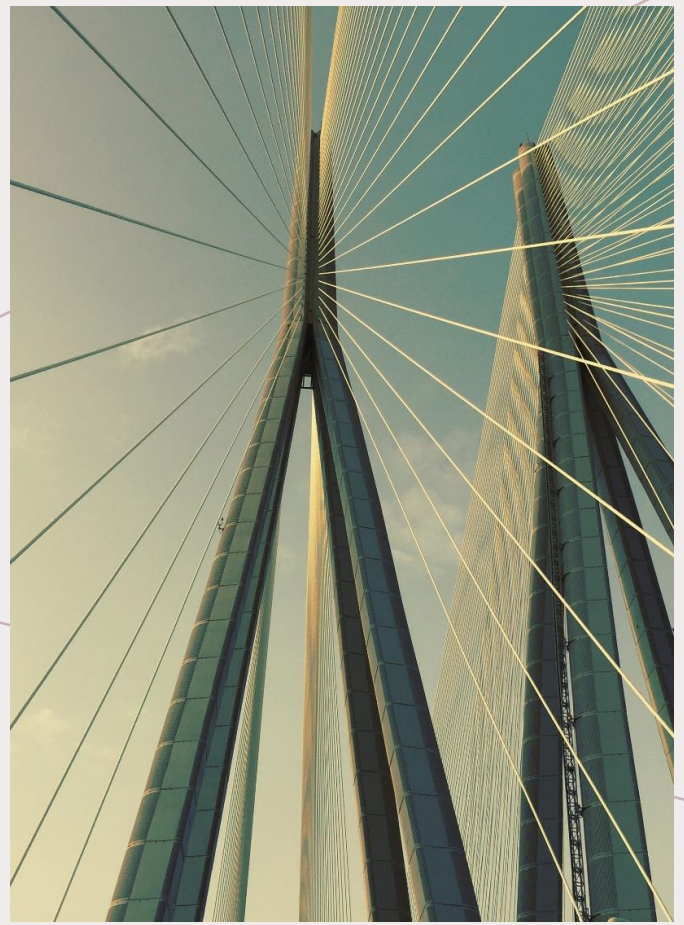


FUTURE PLAN

For the future, I should include more detail about location, such as downtown, shopping mall, supermarket or public transportation, to increase the accuracy of prediction.







THANK YOU