

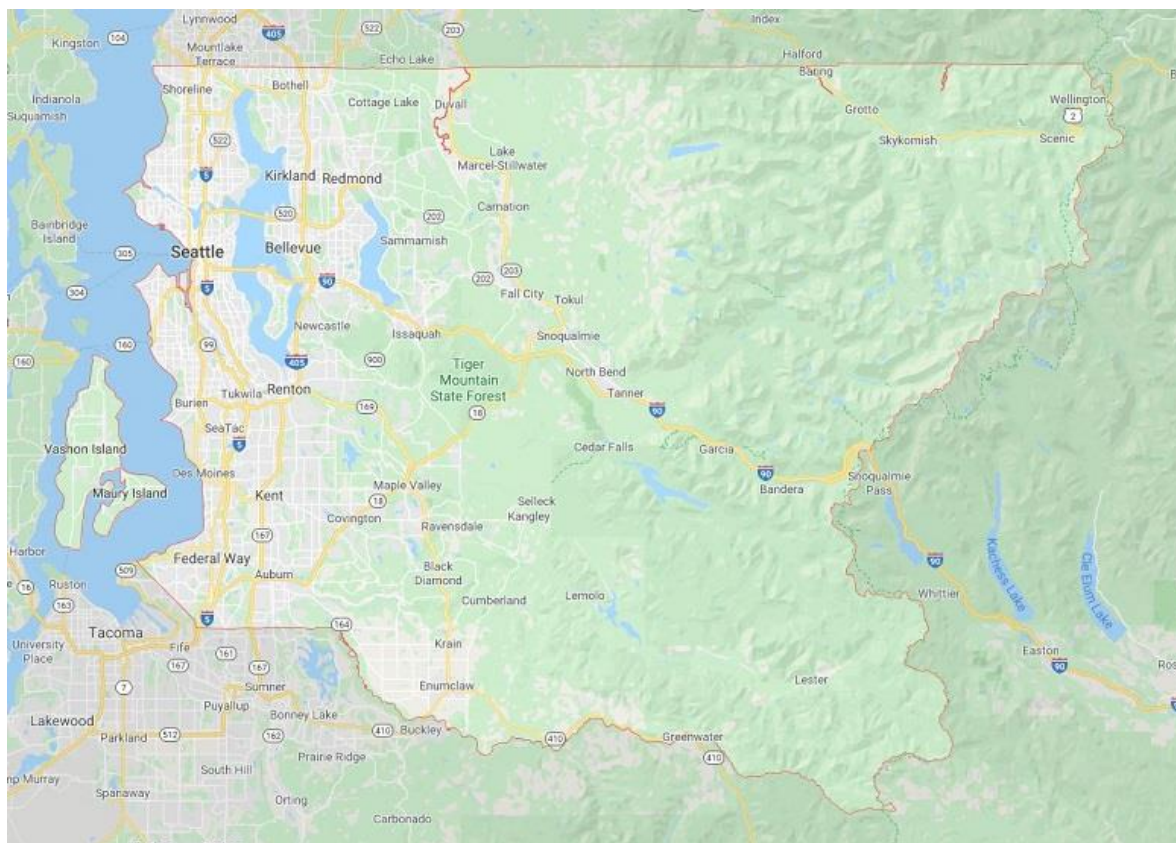
Predicting Real Estate Prices

King County Housing Dataset



The Dataset

More information about the dataset



Column Names and descriptions for King County Data Set

- **id** - unique identified for a house
- **dateDate** - house was sold
- **pricePrice** - is prediction target
- **bedroomsNumber** - of Bedrooms/House
- **bathroomsNumber** - of bathrooms/bedrooms
- **sqft_livingsquare** - footage of the home
- **sqft_lotsquare** - footage of the lot
- **floorsTotal** - floors (levels) in house
- **waterfront** - House which has a view to a waterfront
- **view** - Has been viewed
- **condition** - How good the condition is (Overall)
- **grade** - overall grade given to the housing unit, based on King County grading system
- **sqft_above** - square footage of house apart from basement
- **sqft_basement** - square footage of the basement
- **yr_built** - Built Year
- **yr_renovated** - Year when house was renovated
- **zipcode** - zip
- **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

Dealing with the Data

Is this useable data?

```
df.waterfront.value_counts(normalize=True)
```

```
0.0    0.992342  
1.0    0.007658  
Name: waterfront, dtype: float64
```

```
df.view.value_counts(normalize=True)
```

```
0.0    0.901484  
2.0    0.044763  
3.0    0.023646  
1.0    0.015405  
4.0    0.014702  
Name: view, dtype: float64
```

```
#Check for na values  
df.isna().sum()
```

```
id            0  
date          0  
price         0  
bedrooms      0  
bathrooms     0  
sqft_living    0  
sqft_lot       0  
floors         0  
waterfront    2355  
view          63  
condition      0  
grade          0  
sqft_above     0  
sqft_basement  0  
yr_built       0  
yr_renovated   3810  
zipcode        0  
lat            0  
long           0  
sqft_living15  0  
sqft_lot15     0  
dtype: int64
```

sqft_basement	yr.
0.0	
0.0	
0.0	
0.0	
0.0	
...	
0.0	
?	
700.0	
1040.0	
1040.0	



Exploring the Data

- Question 1: Does renovation have a noticeable effect on price?
- Question 2: Is there a difference in price between a house built in a given time period versus a house renovated in that same time period?
- Question 3: Is there a difference in price based on geographical location in King's County? And if so, why?

Question 1: Does renovation have a noticeable effect on price?



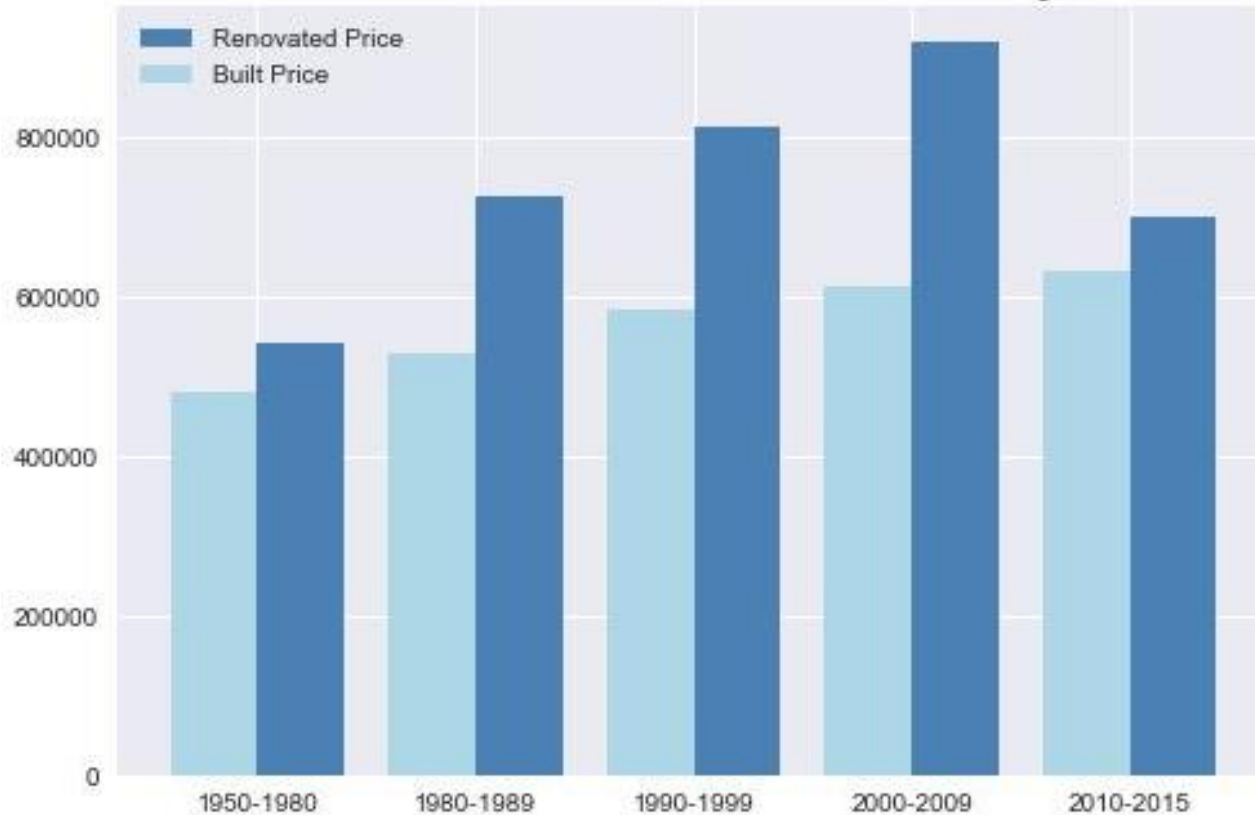
- Renovating a house in King County has a mean price increase of \$237,423 or 144%

Question 1: Does renovation have a noticeable effect on price?

Insight 1: If you are going to sell a house in King County that has not yet been renovated or has not been renovated recently, look into renovating the house and see how much it would cost. If you can spend less than \$230k, you will likely make a profit renovating.

Question 2: Is there a difference in price between a house built in a given time period versus a house renovated in that same time period?

Price of a Renovated House vs Built House by Time Period



- Significantly larger price for renovated houses
- Gap between renovated and built houses seems to be growing
- 2010-2015 might need more data points to observe this trend

Question 2: Is there a difference in price between a house built in a given time period versus a house renovated in that same time period?

Price of a Renovated House vs Built House by Time Period



Insight 2: If you are looking to buy a new home, consider buying a home that has not been renovated. That way, if you renovate, you can greatly improve the value of the home you purchased.

Question 3: Is there a difference in price based on geographical location in King's County?

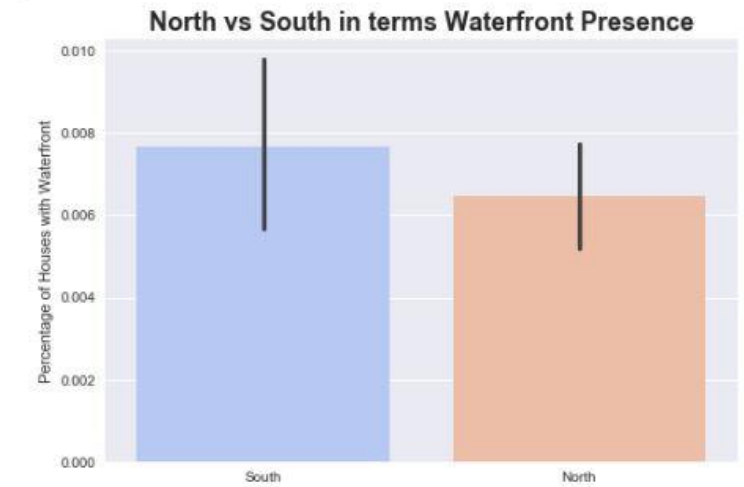
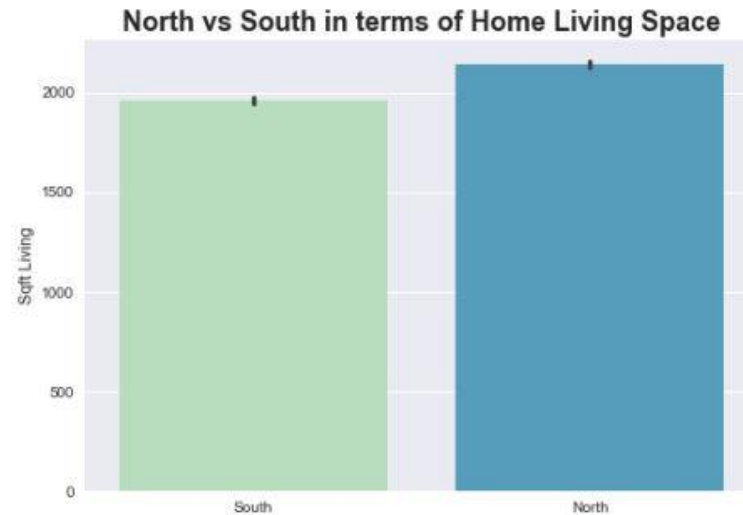
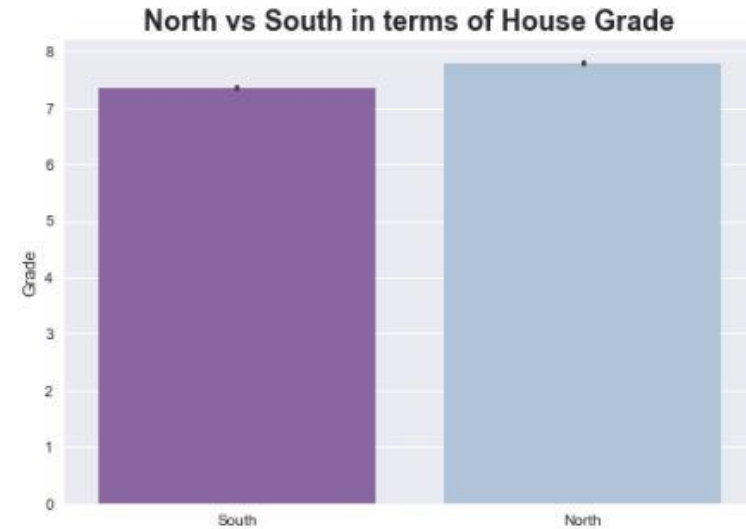


Question 3: Is there a difference in price based on geographical location in King's County?

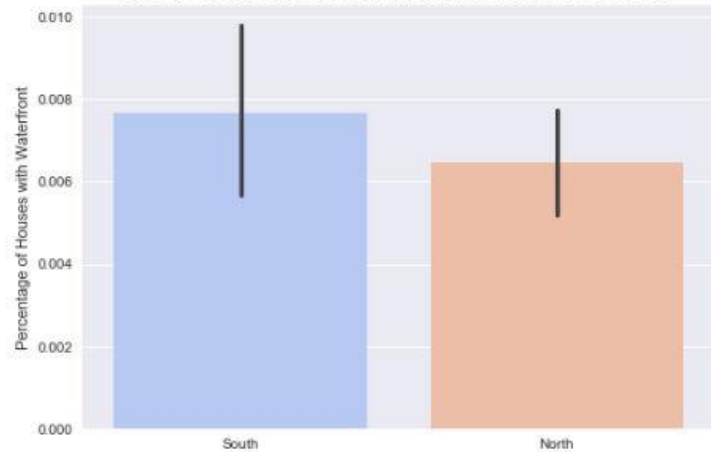


- Houses in the Northern half of Seattle are worth 190% those of in the Southern half - Almost double the price

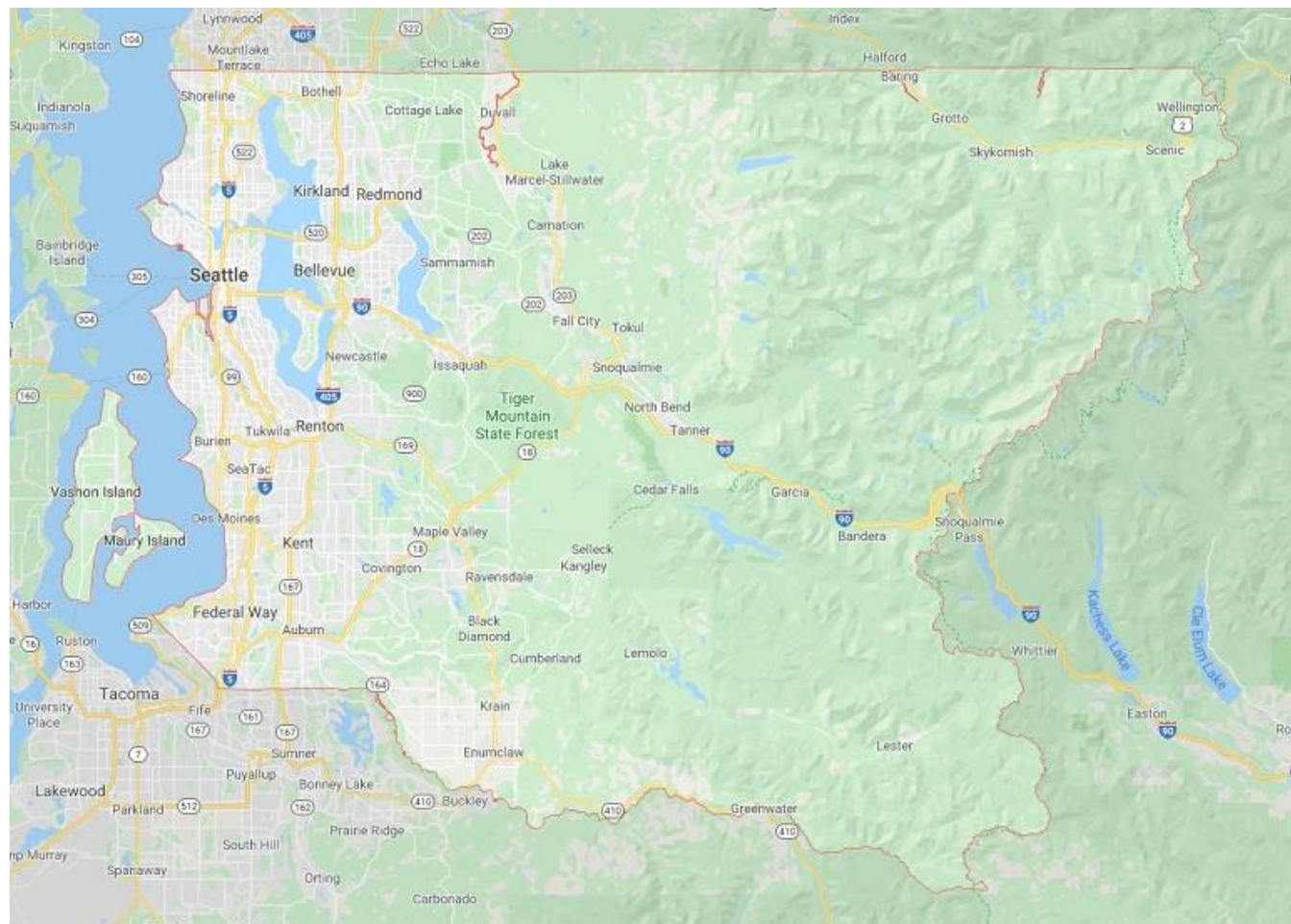
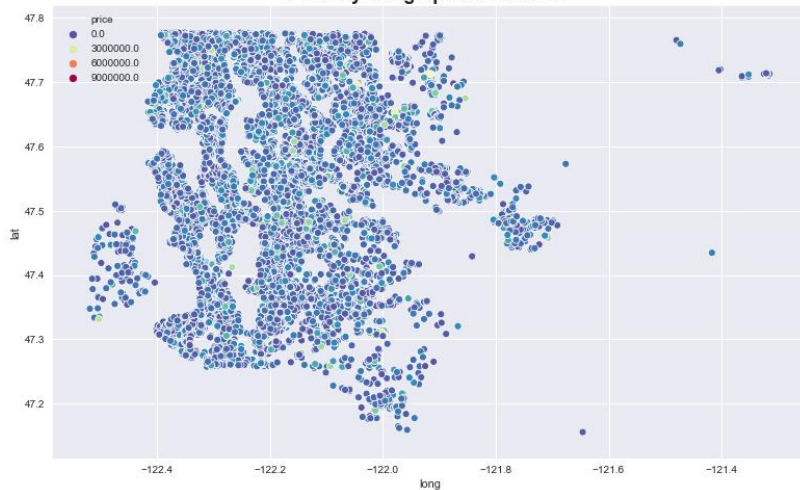
Question 3 (cont.): Why is there a difference in price between Northern and Southern homes in Seattle?



North vs South in terms Waterfront Presence



Price by Geographical Location

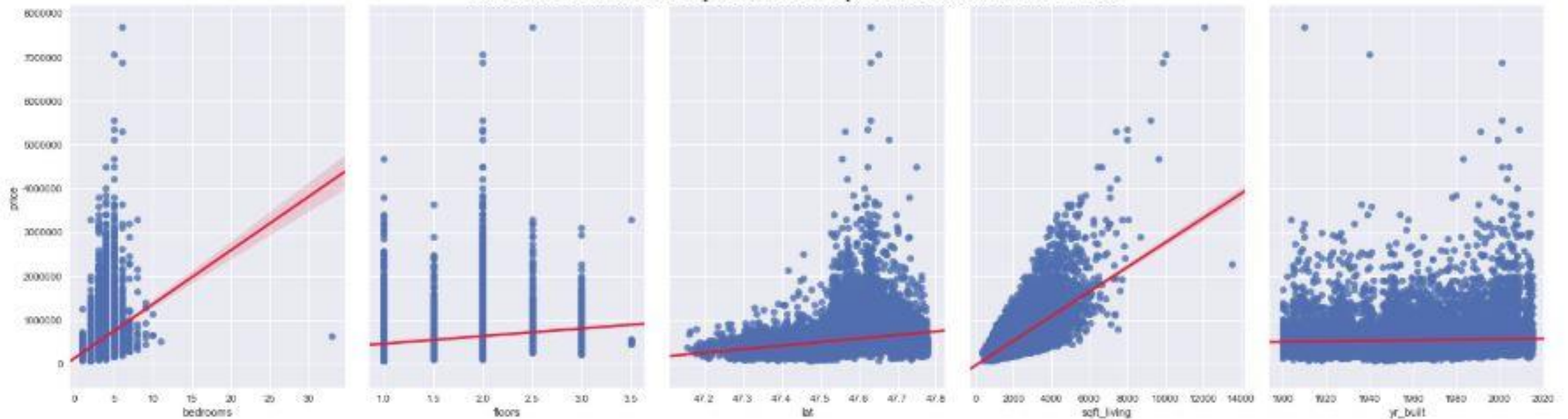


Question 3: Is there a difference in price based on geographical location in King's County?

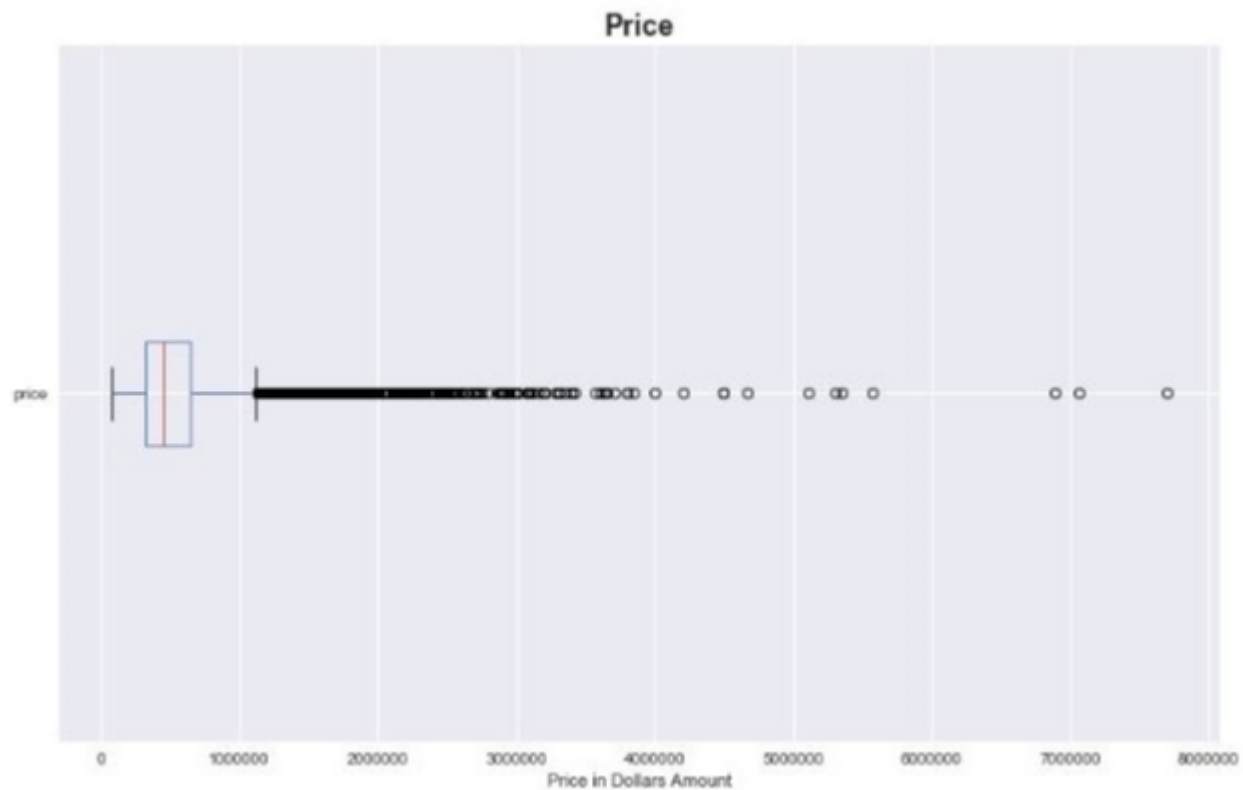
Insight 3: If you are looking to move to King's County, consider buying a home in the Southern part of the city. The homes are almost equivalent in grade and living space but are much cheaper.

Relationship between house features and Price

Pair Plots of Relationships between Dependent Variables and Price



MODEL

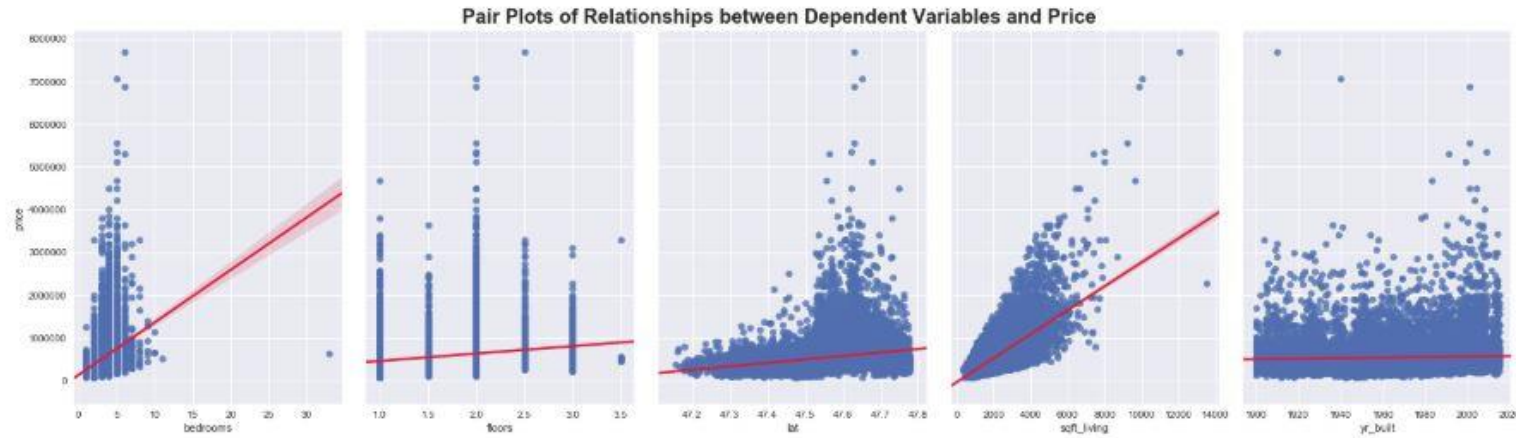


In the linear relationships, there was a skewness due to outliers high in price.

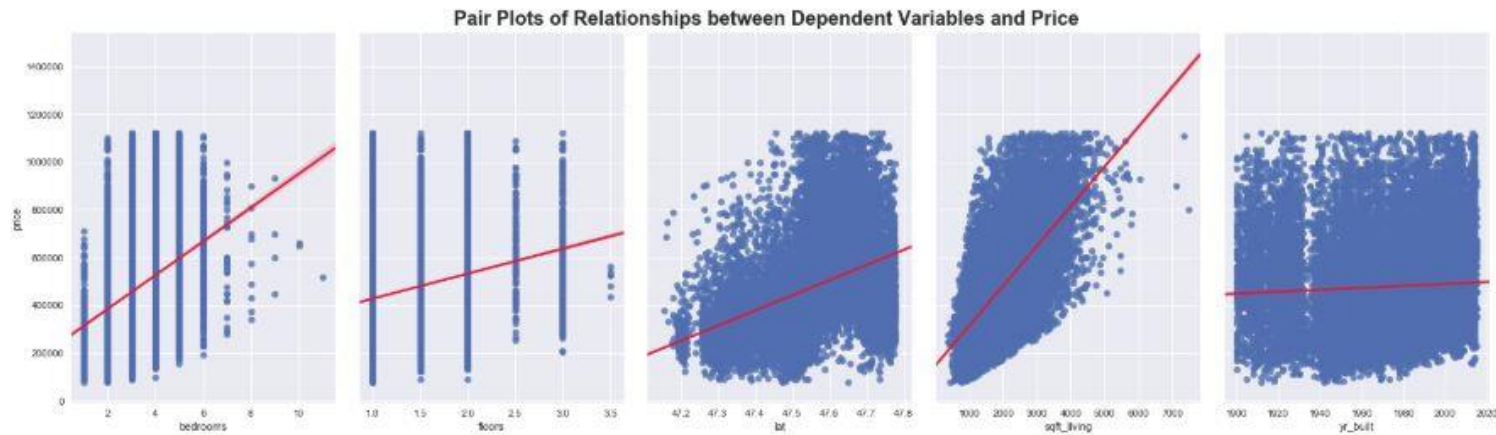
Cut outliers at the upper whisker - \$1,120,000

Linear Relationships before and after removal of Price Outliers

Before



After



Interpreting the OLS Regression Results

OLS Regression Results

Dep. Variable:	price	R-squared (uncentered):	0.903
Model:	OLS	Adj. R-squared (uncentered):	0.903
Method:	Least Squares	F-statistic:	9.423e+04
Date:	Sat, 15 Feb 2020	Prob (F-statistic):	0.00
Time:	21:44:58	Log-Likelihood:	-2.7192e+05
No. Observations:	20267	AIC:	5.438e+05
Df Residuals:	20265	BIC:	5.439e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
lat	3179.6410	65.790	48.330	0.000	3050.687	3308.595
sqft_living	165.4083	1.473	112.293	0.000	162.521	168.296

Omnibus:	924.899	Durbin-Watson:	0.743
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1055.063
Skew:	0.548	Prob(JB):	7.87e-230
Kurtosis:	3.216	Cond. No.	123.



Questions



Predicting Real Estate Prices

King County Housing Dataset

