

# Seattle King County Housing Study



Paranormal Distribution  
Mate Pocs & Naweed Ahmed

 **FLATIRON SCHOOL**

# Table of Contents

- Our Data Set
- Exploring the Data
- Linear Regression
- Recommendations
- Future Work

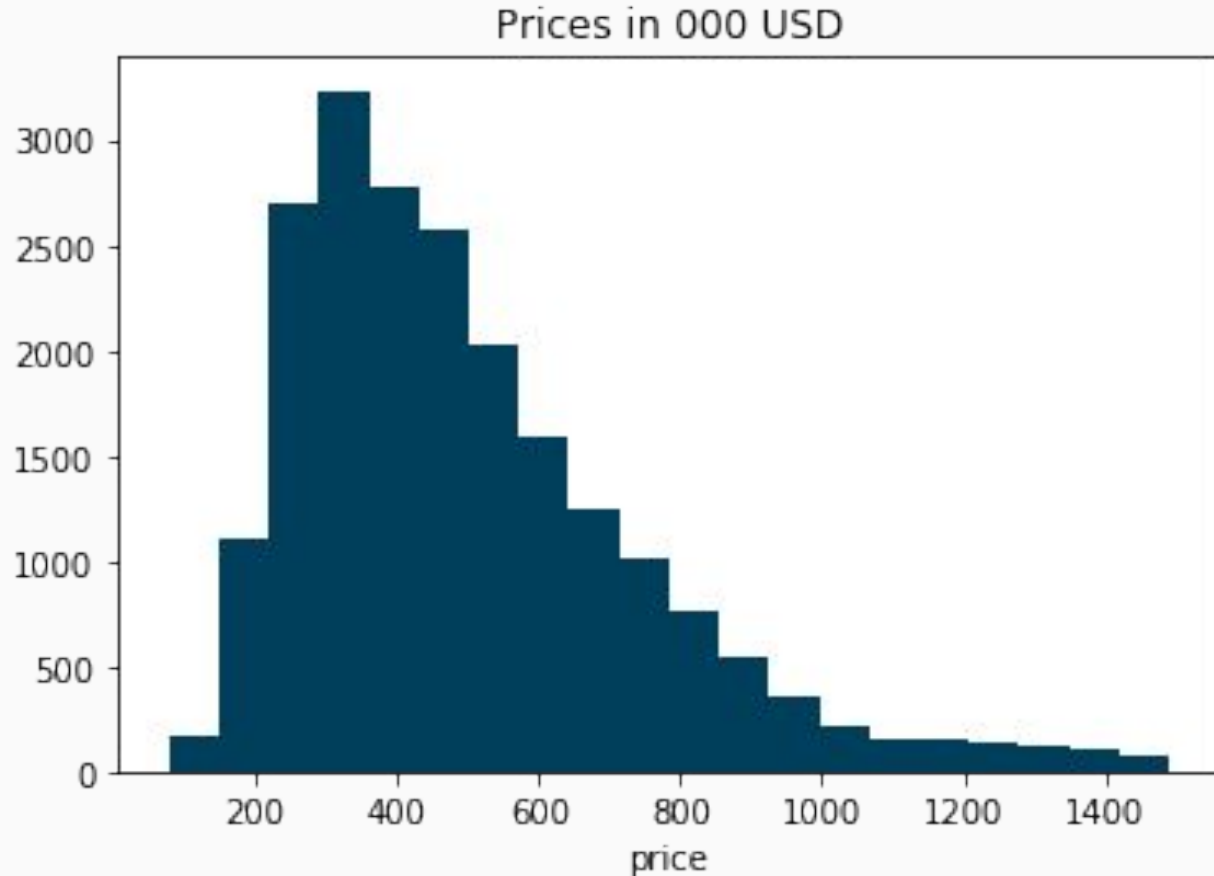
# Our Data Set

The data set shows details of houses sold between May 2014 and May 2015.

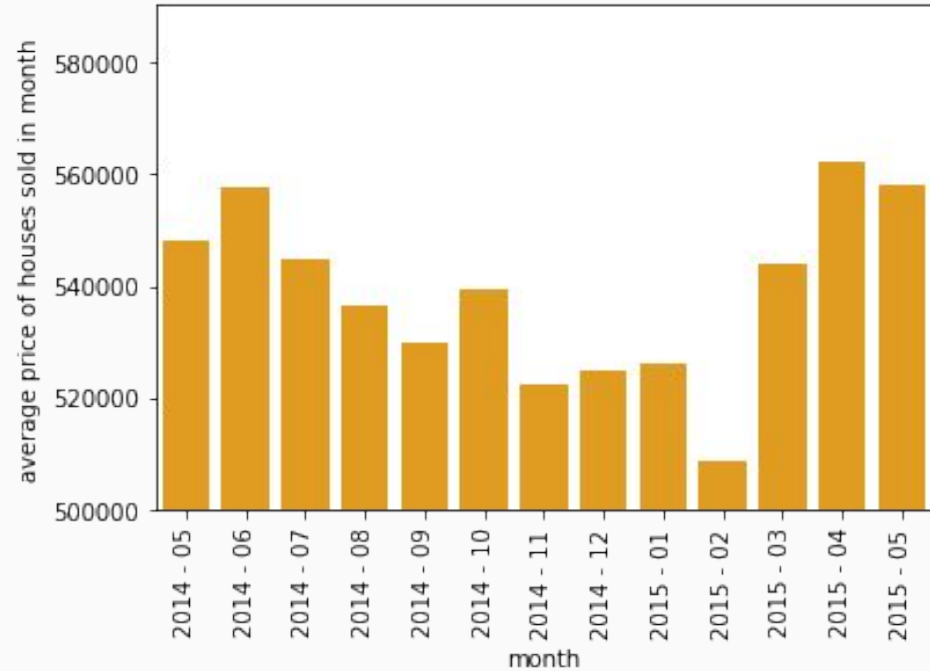
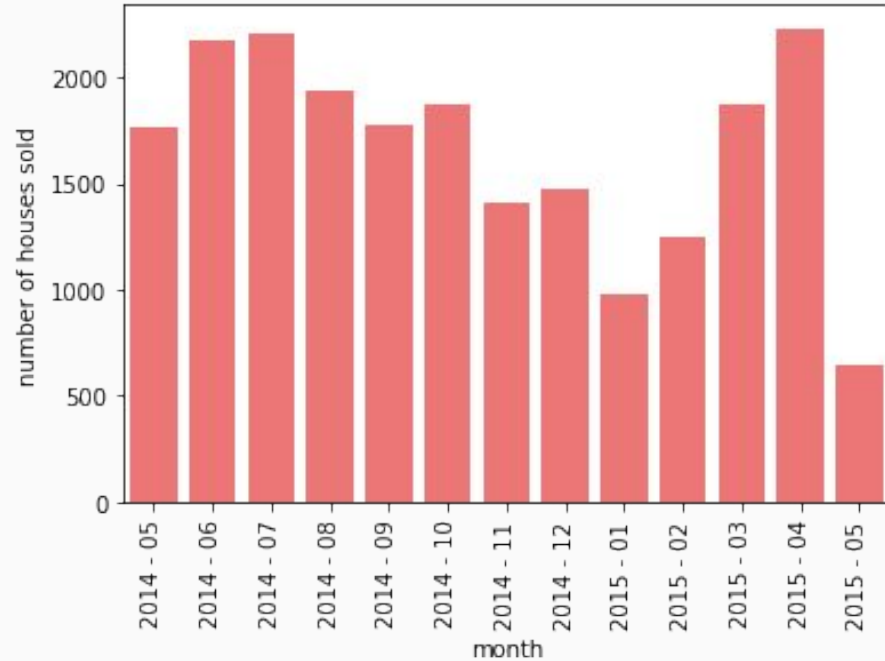
Original data set includes 21,597 rows and 21 columns.

After removing outliers, missing data and columns not deemed appropriate for the model, we were left with 20,031 rows and 19 columns.

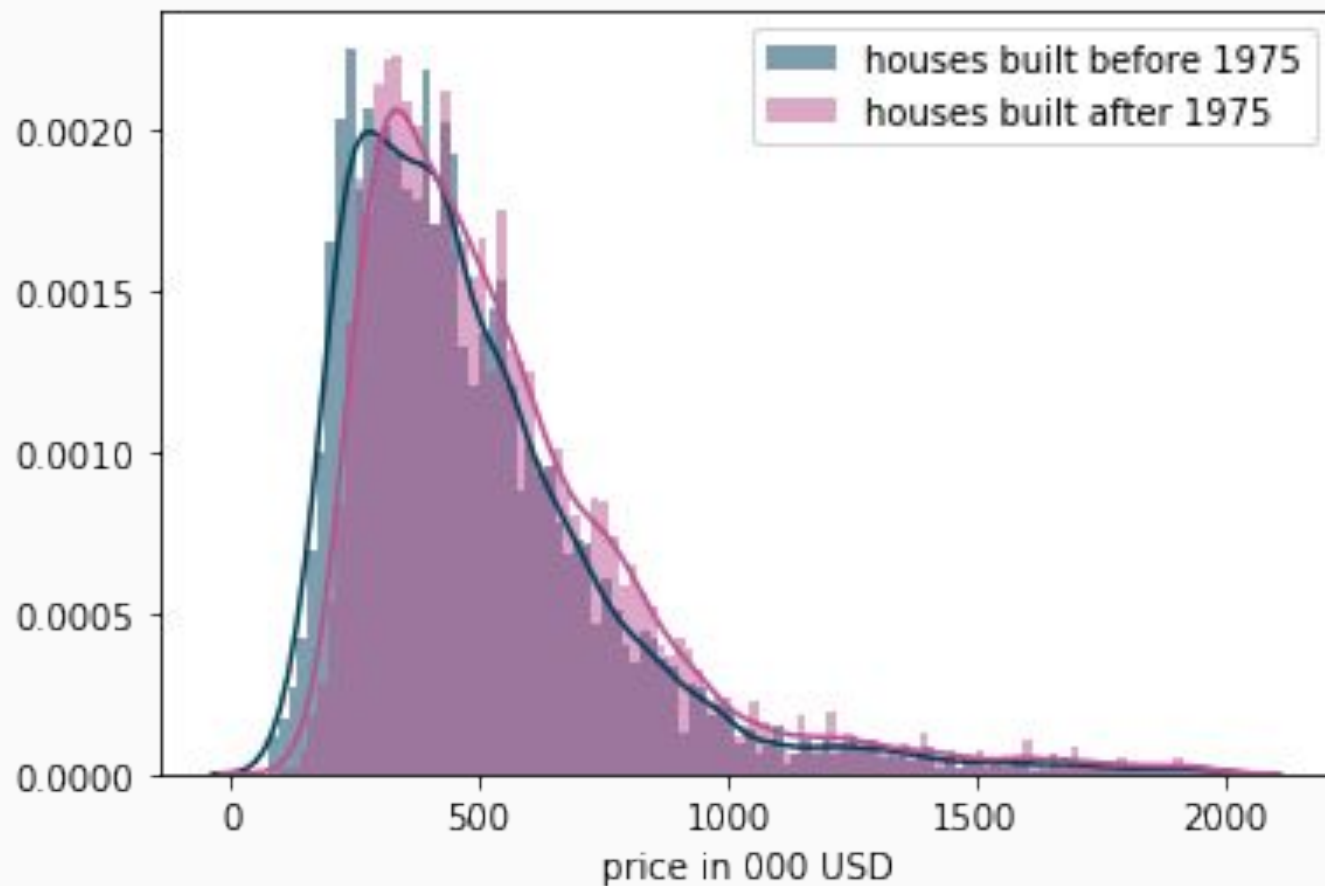
# Exploring the Data - Price histogram



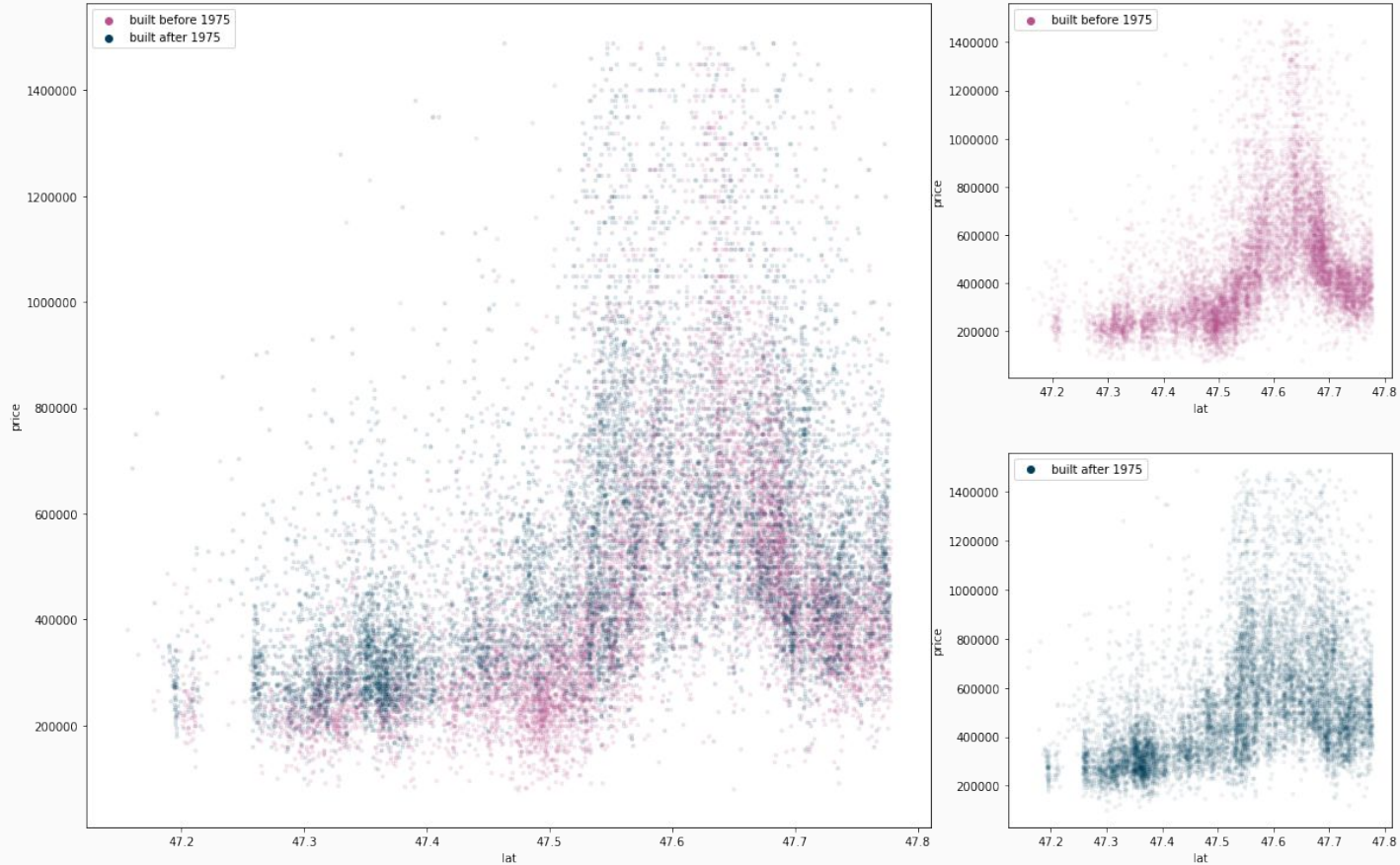
# Exploring the Data - Seasonality



## Exploring the Data - Price comparison of houses built before and after 1975



# Exploring the Data - Latitude vs. prices

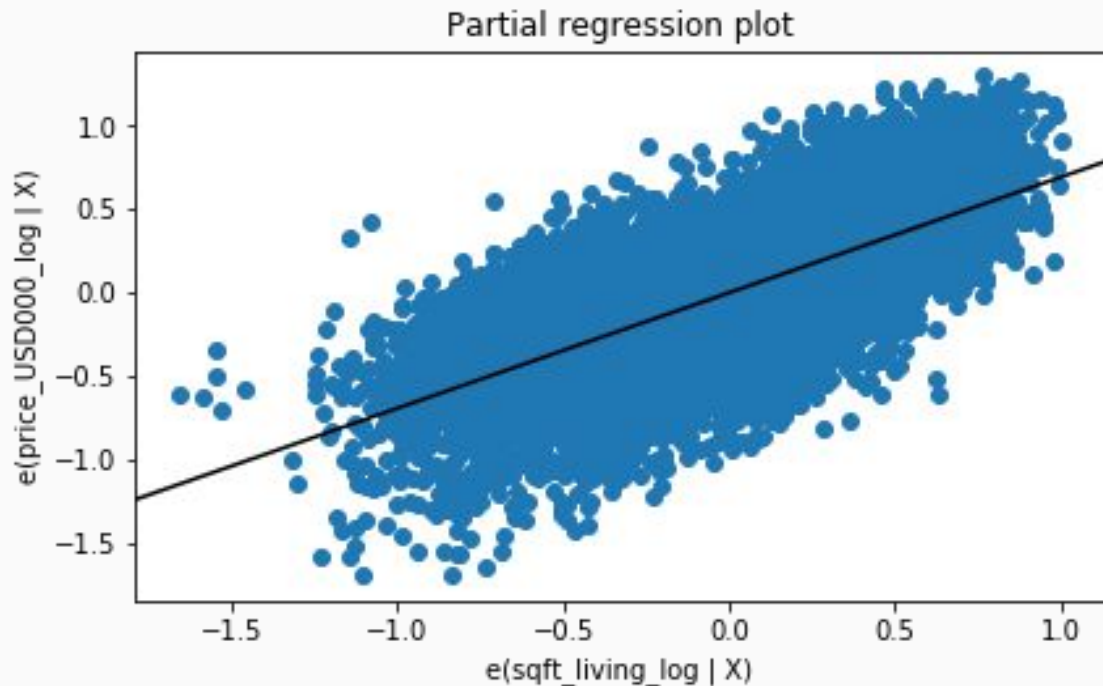




# Linear Regression Analysis

We used 2 independent variables in our regression analysis.

$$\text{price\_USD000\_log} = 0.7246 * \text{sqft\_living\_log} - 2.9815 * \text{lat\_dist\_from\_4762} + 1.0085$$





# Linear Regression Analysis

## OLS Regression Results

```
=====
Dep. Variable:      price_USD000_log      R-squared:                0.703
Model:              OLS                   Adj. R-squared:           0.703
Method:             Least Squares         F-statistic:             2.371e+04
Date:               Tue, 21 Jan 2020       Prob (F-statistic):      0.00
Time:               13:02:48              Log-Likelihood:          -1708.0
No. Observations:   20031                 AIC:                     3422.
Df Residuals:       20028                 BIC:                     3446.
Df Model:           2
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
sqft_living_log	0.7246	0.005	152.681	0.000	0.715	0.734
lat_dist_from_4762	-2.9815	0.020	-146.301	0.000	-3.021	-2.942
constant	1.0085	0.036	28.069	0.000	0.938	1.079

```
=====
Omnibus:            382.472      Durbin-Watson:           1.986
Prob(Omnibus):      0.000       Jarque-Bera (JB):        815.585
Skew:               0.010       Prob(JB):                7.91e-178
Kurtosis:           3.988       Cond. No.                 148.
=====
```

### Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# Recommendations

Based on our findings, we have ascertained that the size of the property and the proximity to a certain latitude has a positive impact on prices.

These two variables explains a large portion of the price variable.

Adjusted R -Squared is 0.703

# Future Work

We have our regression model built which provides a strong model to predict prices. For future work, we can incorporate additional variables into this model and see how it impacts the price.

# Q&A

# Thanks!

Mate Pocs

Naweed Ahmed