

Seattle King County Housing Study



Paranormal Distribution
Mate Pocs & Naweed Ahmed

 **FLATIRON SCHOOL**

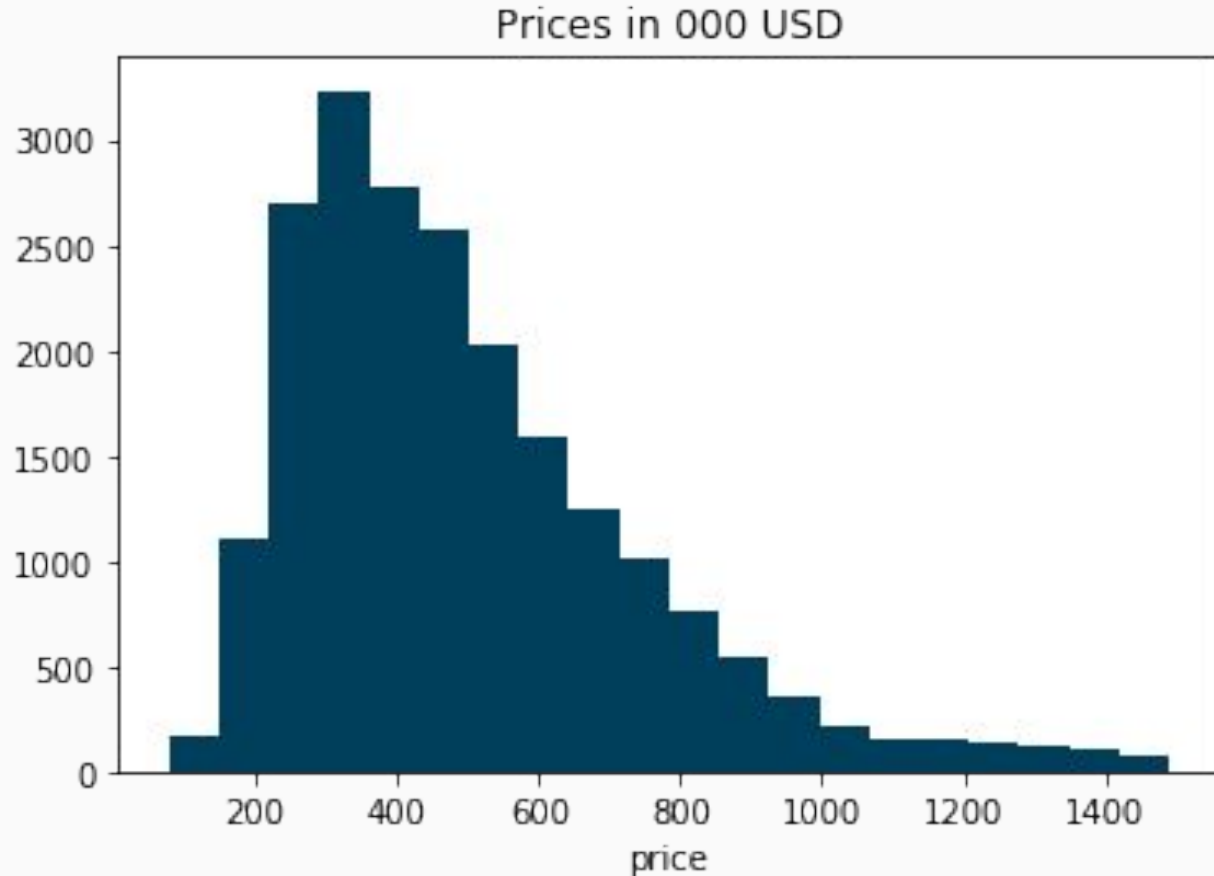
Table of Contents

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

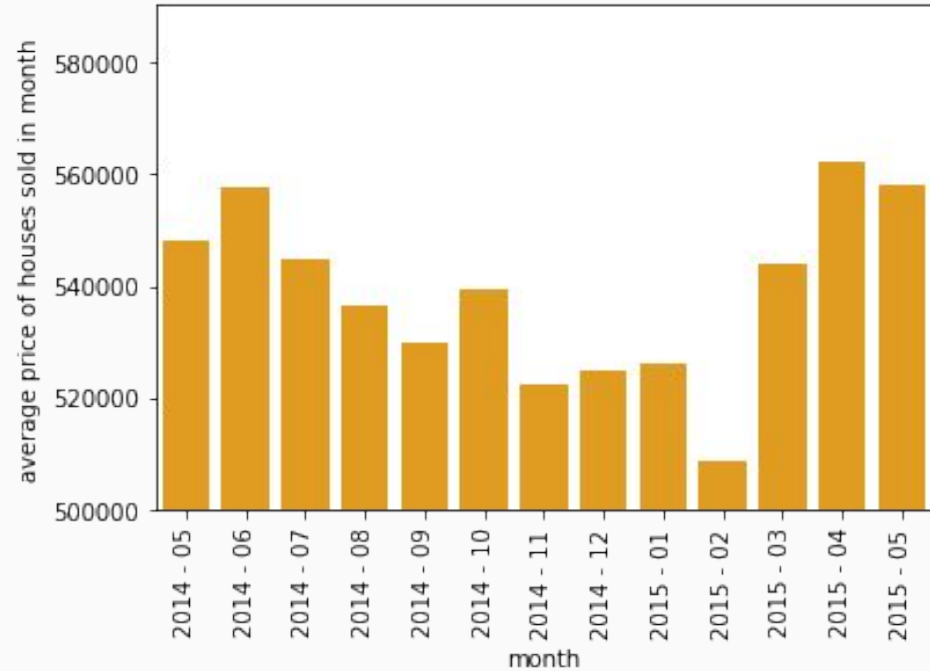
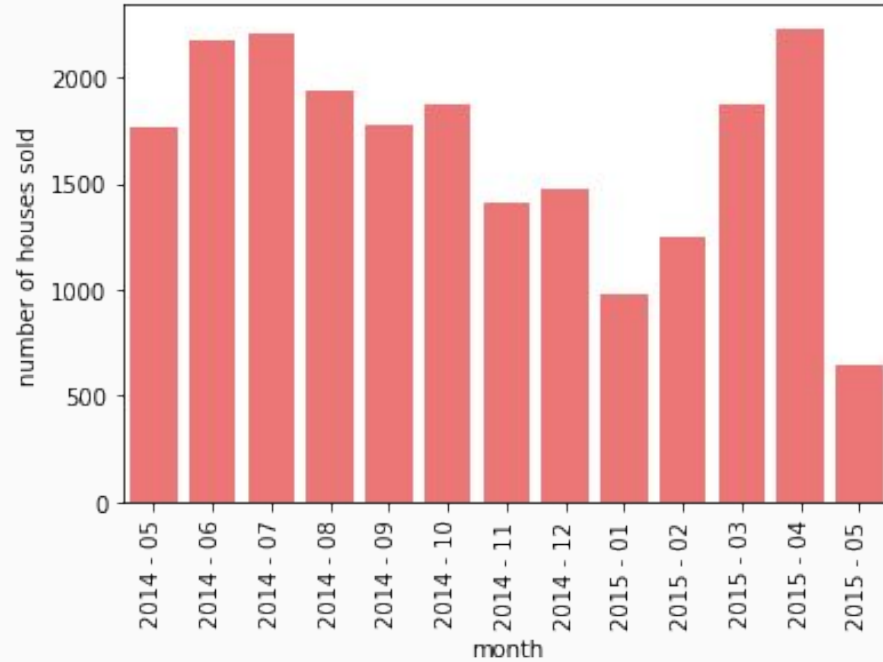
Our Data Set

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

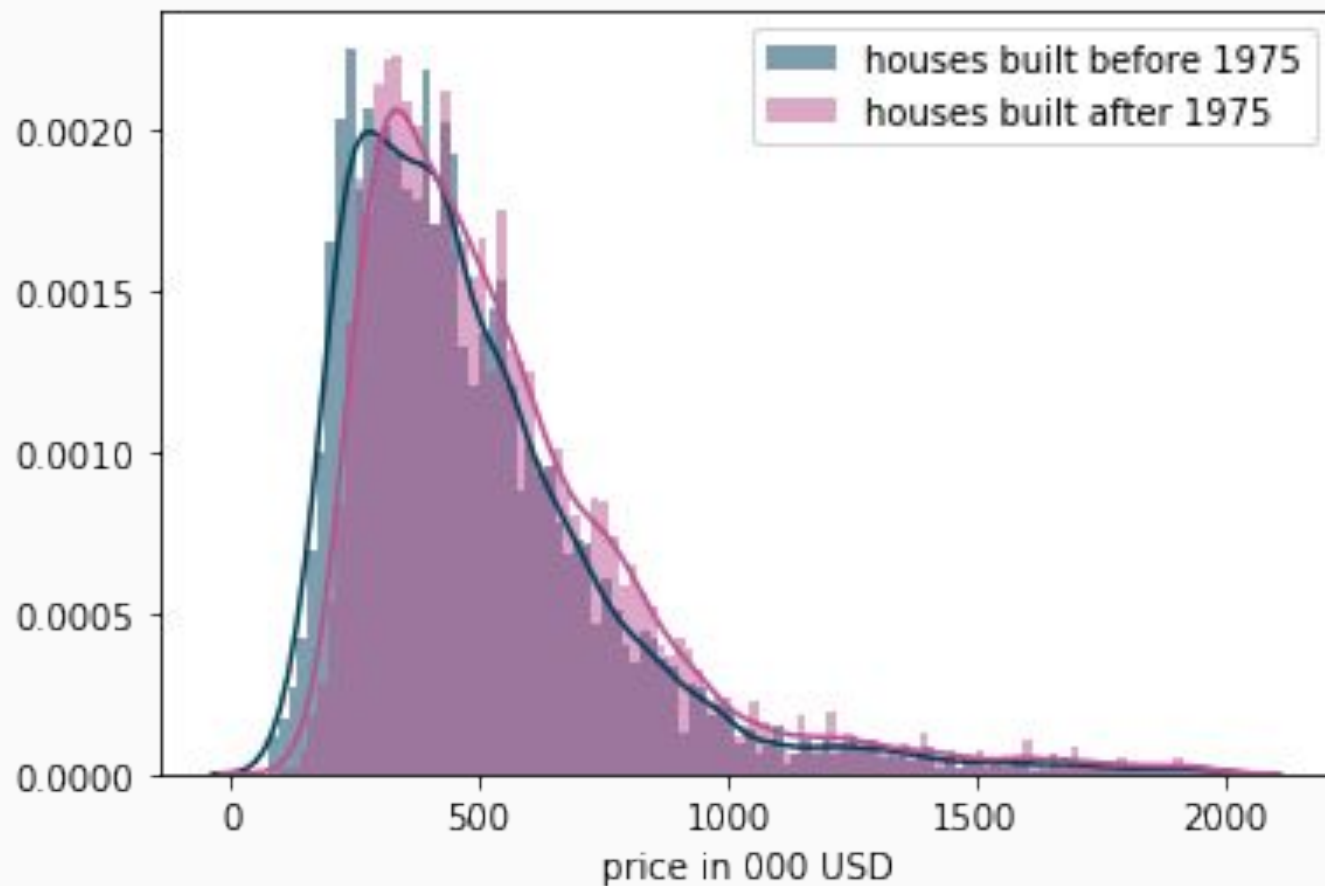
Exploring the Data - Price histogram



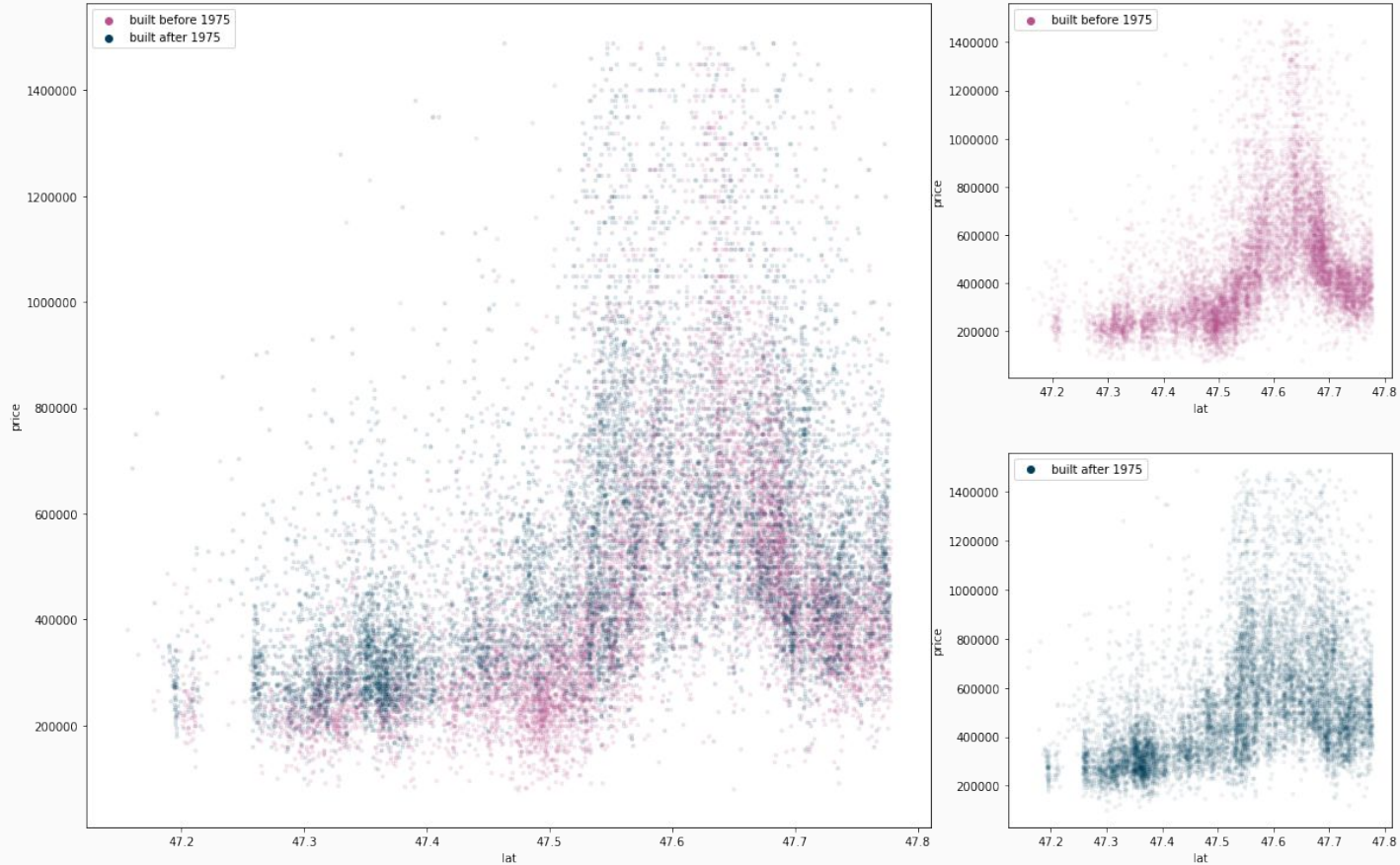
Exploring the Data - Seasonality



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



Exploring the Data - Latitude and prices



Linear Regression Analysis

Variables used : log of sqft_living & lat_dist_from_4762

$$\text{price_USD000_log} = 0.7246 * \text{sqft_living_log} - 2.9815 * \text{lat_dist_from_4762} + 1.0085$$

R- Squared = 0.703

We filtered the data and ran 2 more regression models

Properties built before 1975

$$\text{price_USD000_log} = 0.6858 * \text{sqft_living_log} - 3.4616 * \text{lat_dist_from_4762} + 1.3537$$

R-Squared = 0.662

Properties built after 1975

$$\text{price_USD000_log} = 0.7847 * \text{sqft_living_log} - 2.6159 * \text{lat_dist_from_4762} + 0.4915$$

R2 = 0.770

Linear Regression Analysis

OLS Regression Results

```
=====
Dep. Variable:      price_USD000_log    R-squared:          0.770
Model:              OLS                 Adj. R-squared:      0.770
Method:             Least Squares       F-statistic:         1.617e+04
Date:               Tue, 21 Jan 2020     Prob (F-statistic):  0.00
Time:               14:50:19            Log-Likelihood:      1171.7
No. Observations:   9651                AIC:                 -2337.
Df Residuals:       9648                BIC:                 -2316.
Df Model:           2
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
sqft_living_log	0.7847	0.006	124.273	0.000	0.772	0.797
lat_dist_from_4762	-2.6159	0.022	-118.889	0.000	-2.659	-2.573
constant	0.4915	0.049	10.088	0.000	0.396	0.587

```
=====
Omnibus:            384.008    Durbin-Watson:       1.960
Prob(Omnibus):      0.000     Jarque-Bera (JB):    854.179
Skew:               0.251     Prob(JB):            3.29e-186
Kurtosis:           4.368     Cond. No.            174.
=====
```

Warnings:

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

Conclusions

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

These two variables explain a large portion of the variation in the price variable, adjusted R^2 for houses built after 1975 is 0.770.

Future Work

- Based on our regression model, we suggest to focus on property size and distance from Seattle's city centre in order to predict house prices in the future.
- If we obtain more data for subsequent years we can further analyse seasonal trends in market prices.

Mate Pocs

Naweed Ahmed

Thank You!

Questions?