Final Report By Class group F

Group B

Name of Active Participants: Ahmed Jawad Munir. David Walcher , Barakat OgunJoun

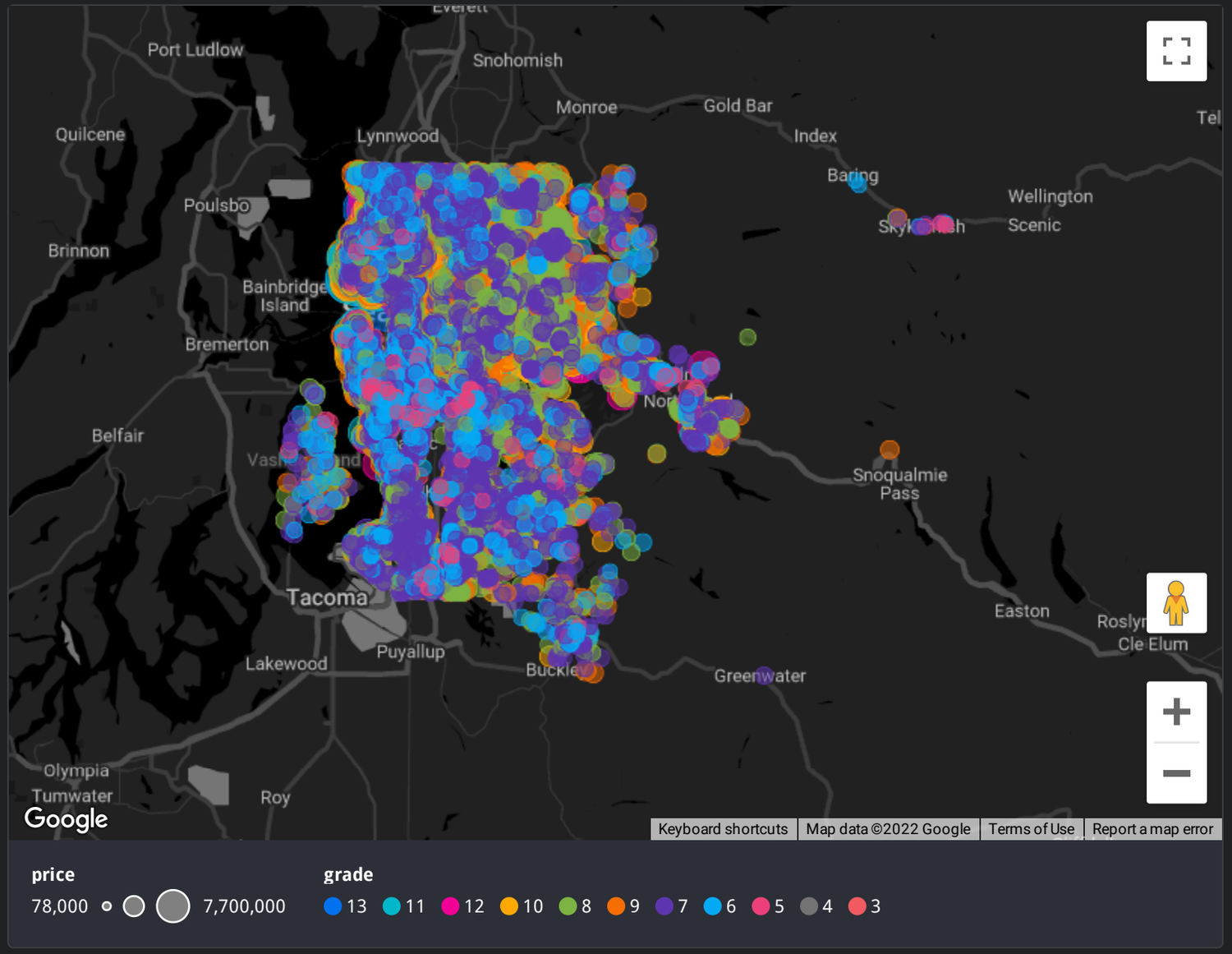
Seattle Real Estate Inc.

HOLDING

INVESTMENTS

Seattle Real Estate Inc.
HOLDING INVESTMENTS

What's the investment?



Year built
1,900 — 2,015

Grade
3 — 13

Price in $
78,000 — 7,700,000

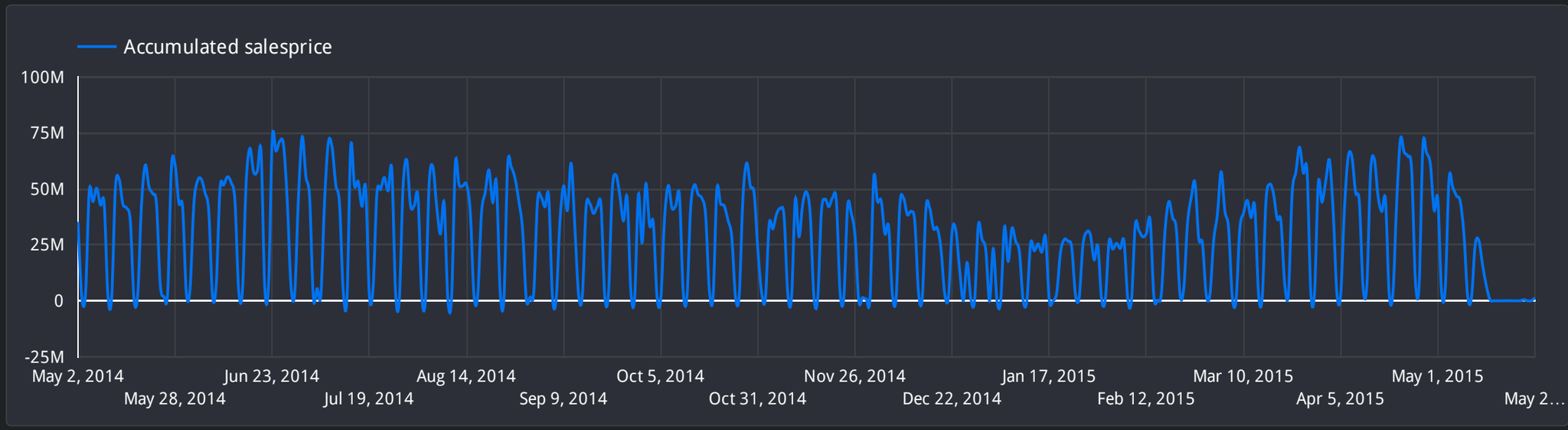
Livingspace in sqare-feet
370 — 13,540

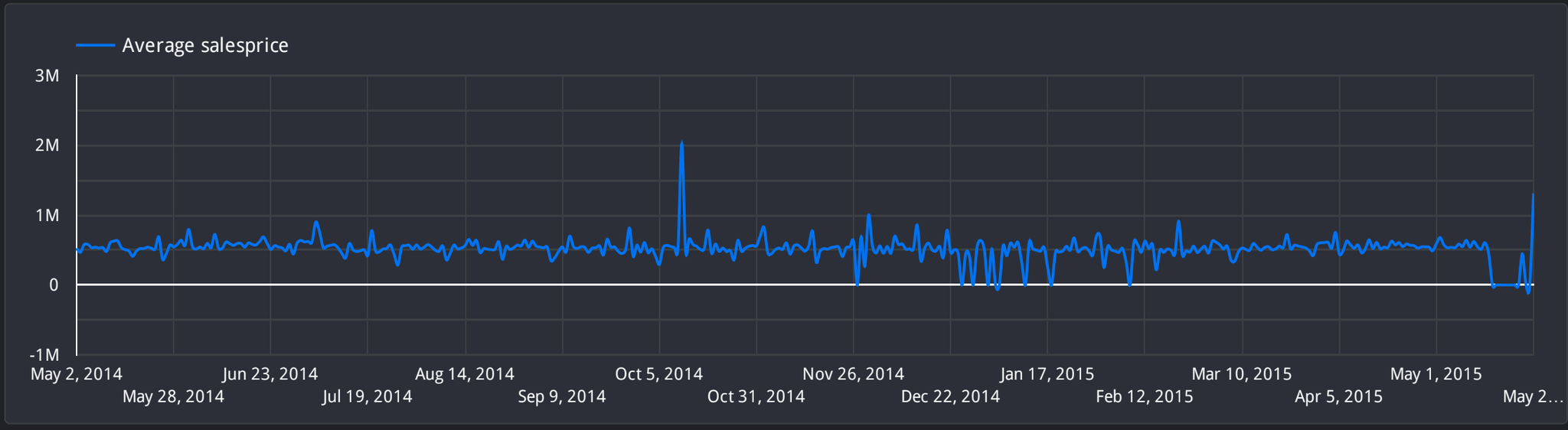
Waterfront
0 — 1

- Real Estate in the Seattle region
- Colour indicates the grade of the property, ranging from 3-13
- Size of the bubbles illustrates the price, ranging from 78,000$ to 7,700,000$
- Map can be adjusted by using the filters on the right

Seattle Real Estate Inc.
HOLDING INVESTMENTS

What's the investment?

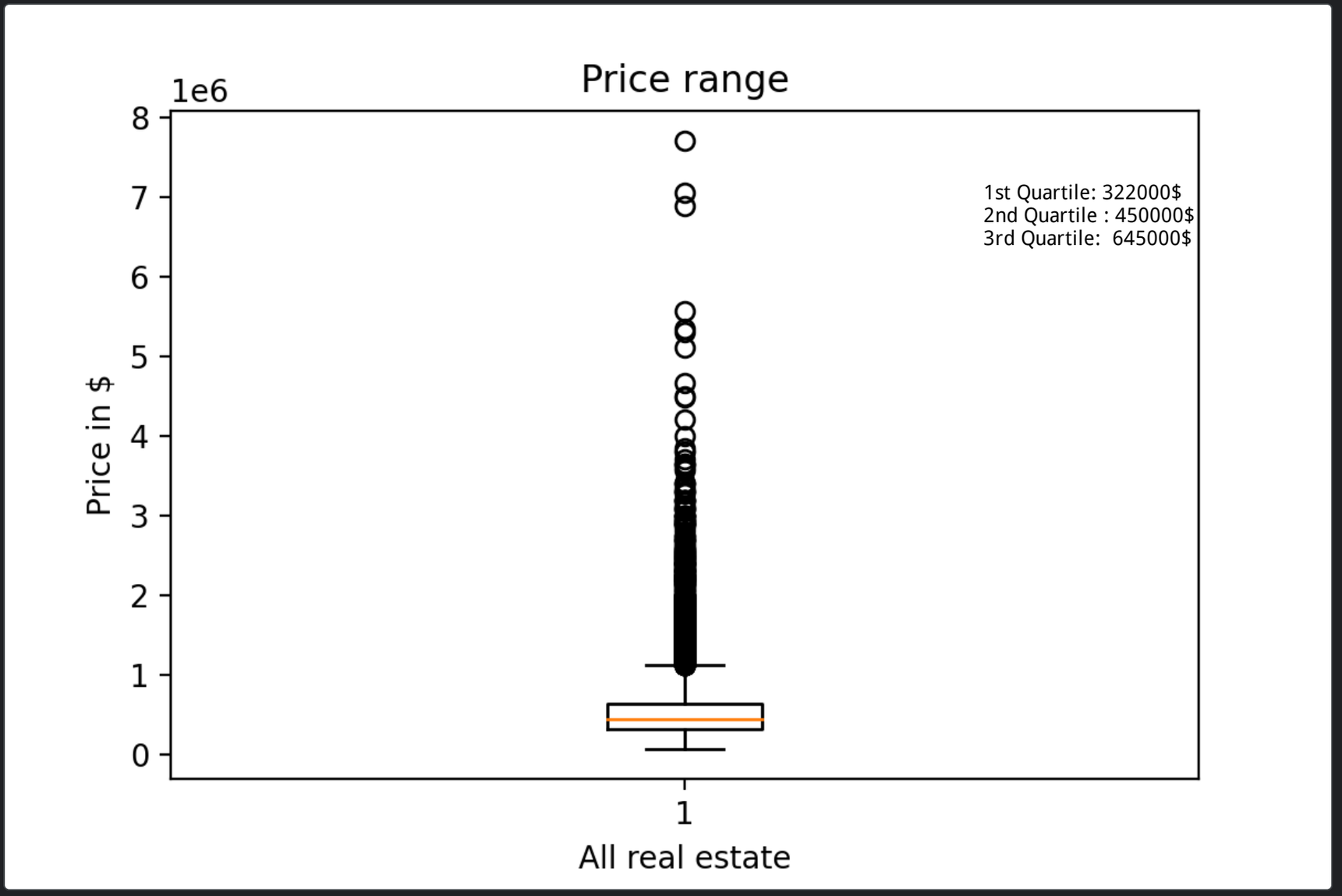




- Less sales in the winter and autumn season
- Average price also decreases in these months
- Possible investment strategy?: Buy in ~January and sell in around May for a 10% profit

Seattle Real Estate Inc.
HOLDING INVESTMENTS

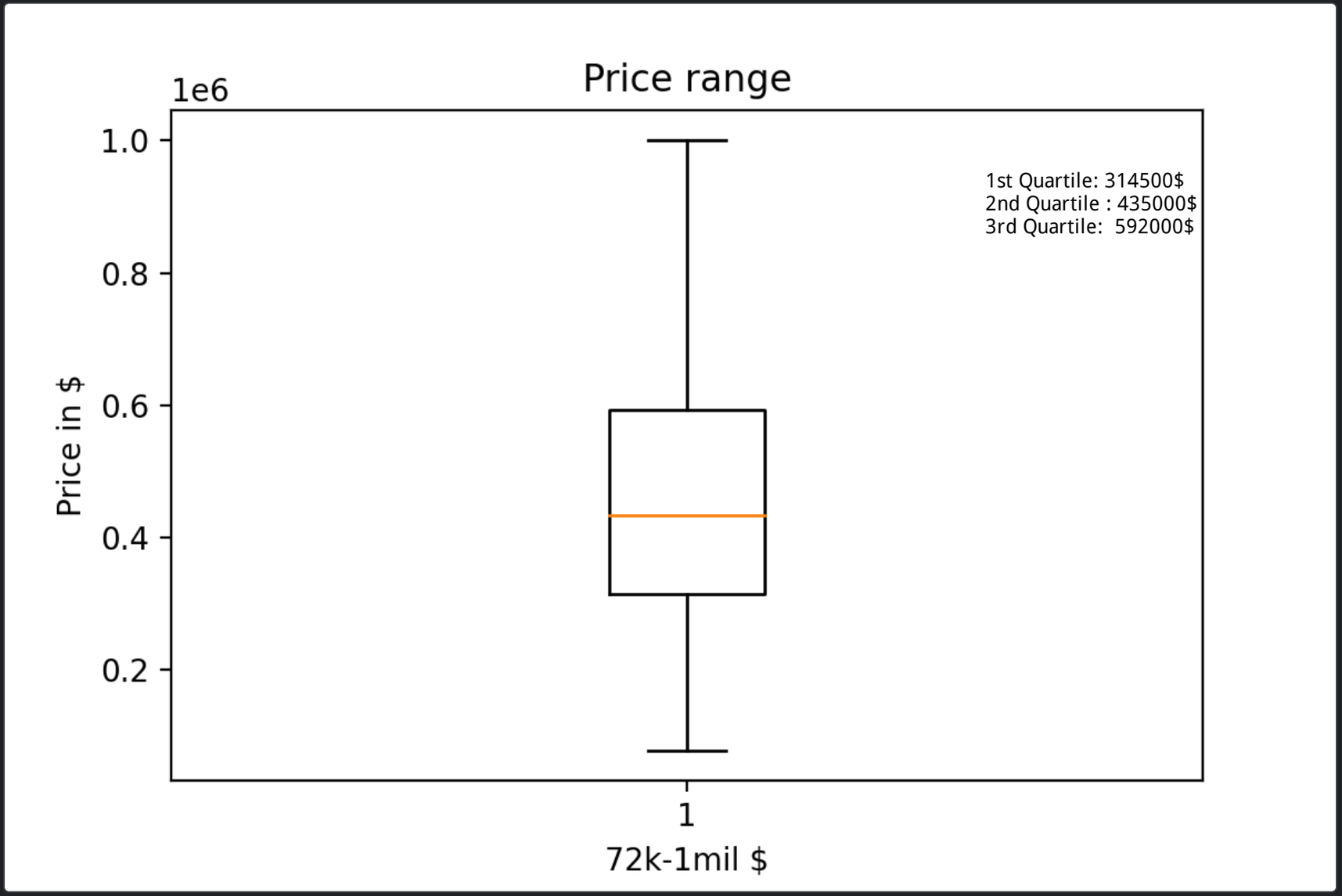
Price ranges



- Most sales from 1.000.000$ onwards can be considered as outliers

Seattle Real Estate Inc.
HOLDING
INVESTMENTS

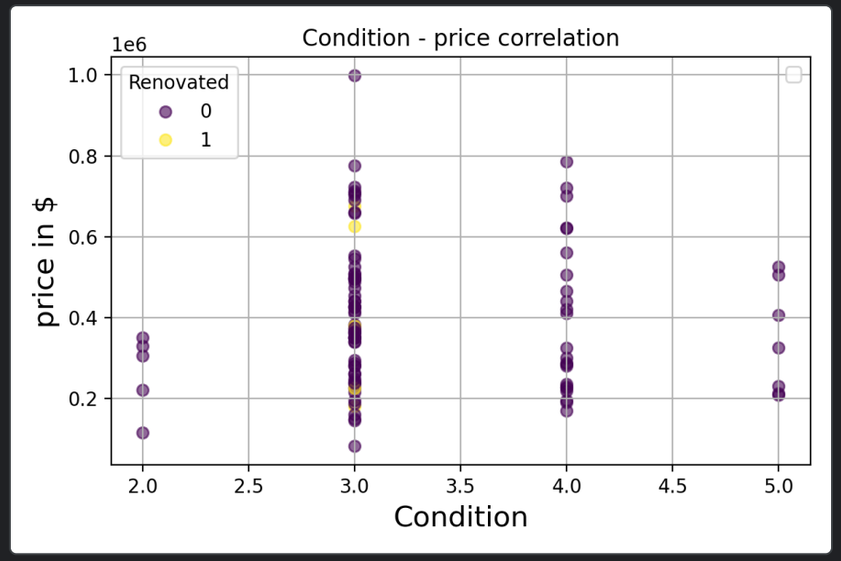
Price ranges

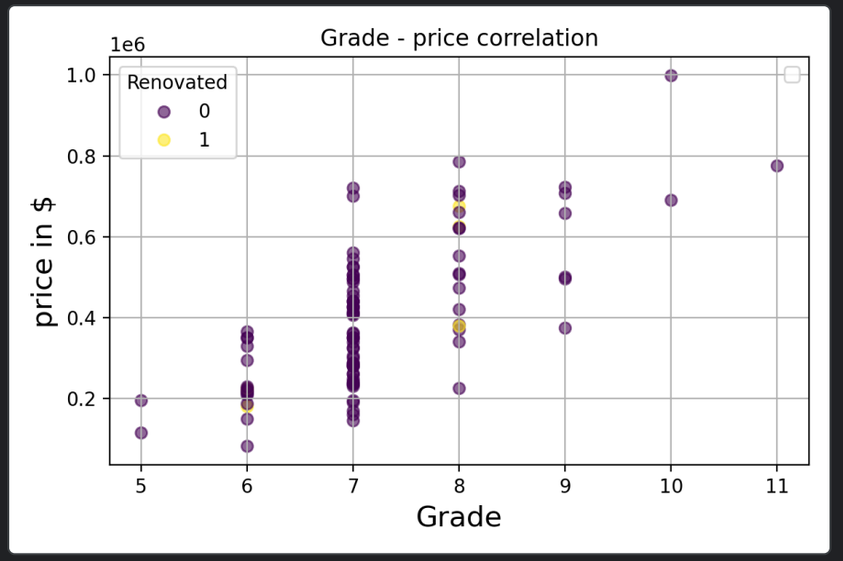


- Slightly skewed price distribution towards the cheaper end of the spectrum

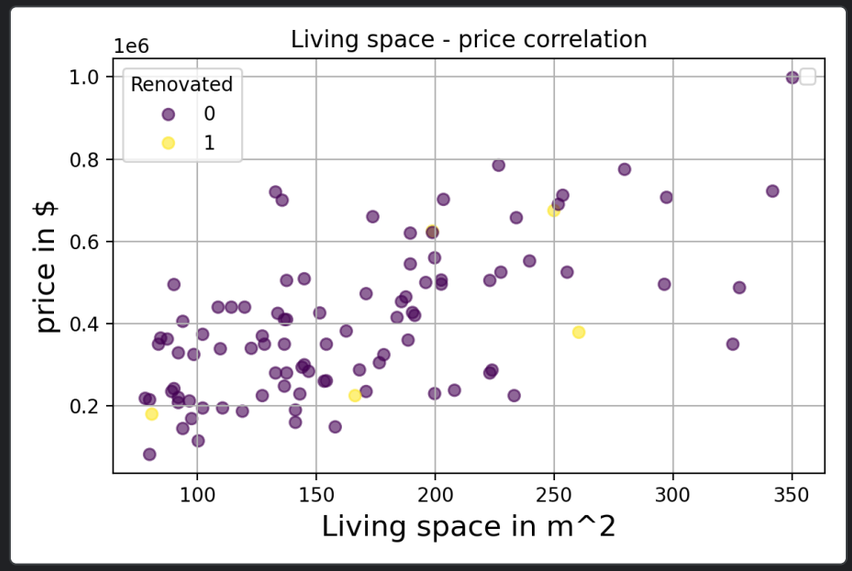
Seattle Real Estate Inc.
HOLDING INVESTMENTS

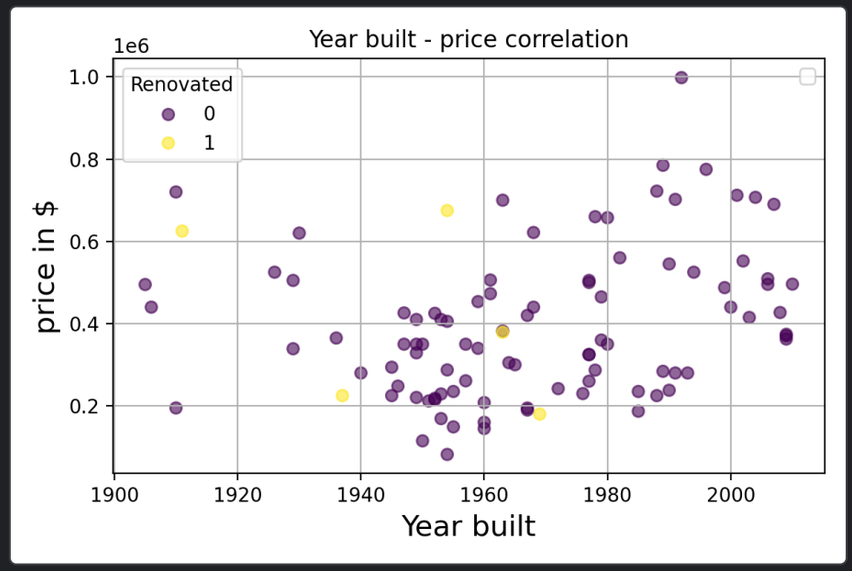
What decides the price?





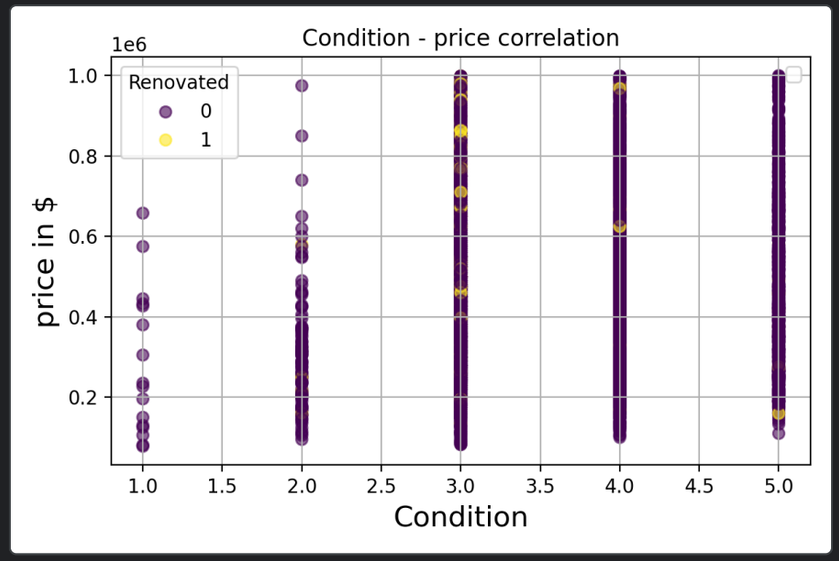
Sample of only 100 random data points - price range {0-1mil$}

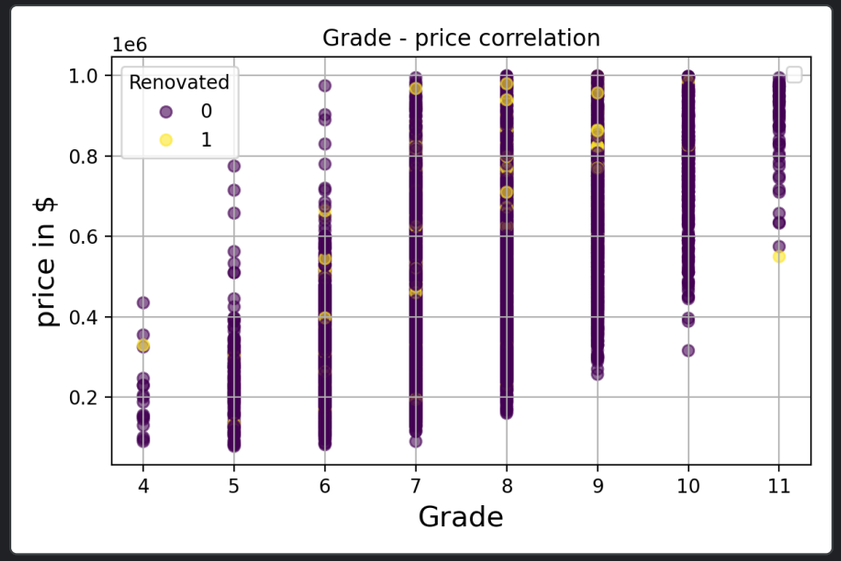




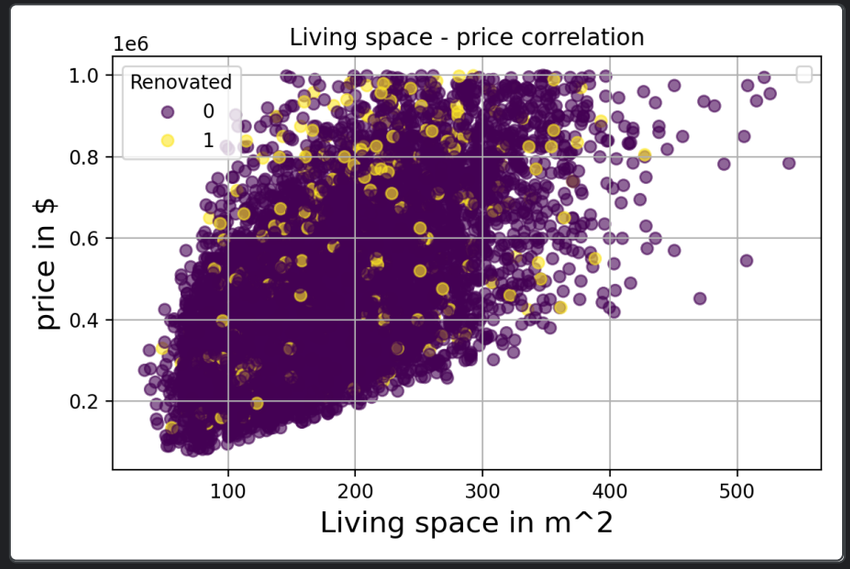
Seattle Real Estate Inc.
HOLDING INVESTMENTS

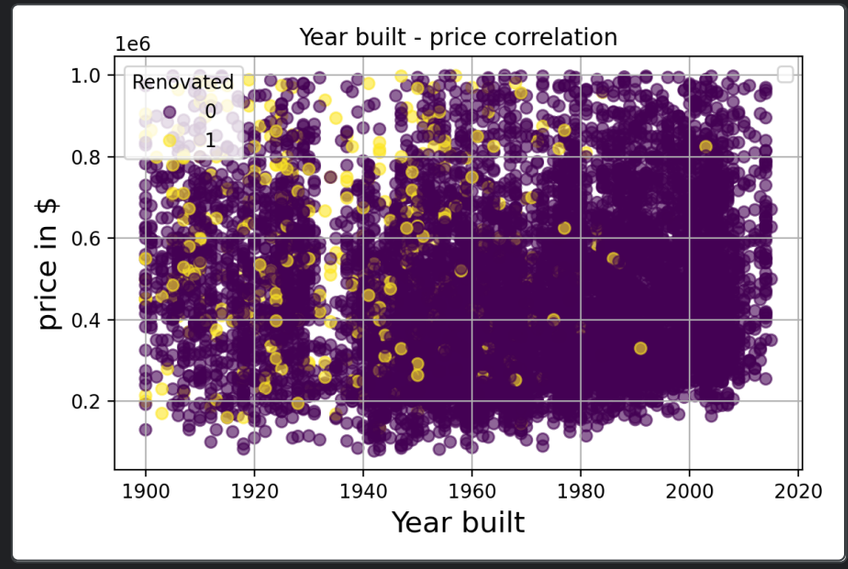
What decides the price?





Sample of only all data points - price range {0-1mil$}

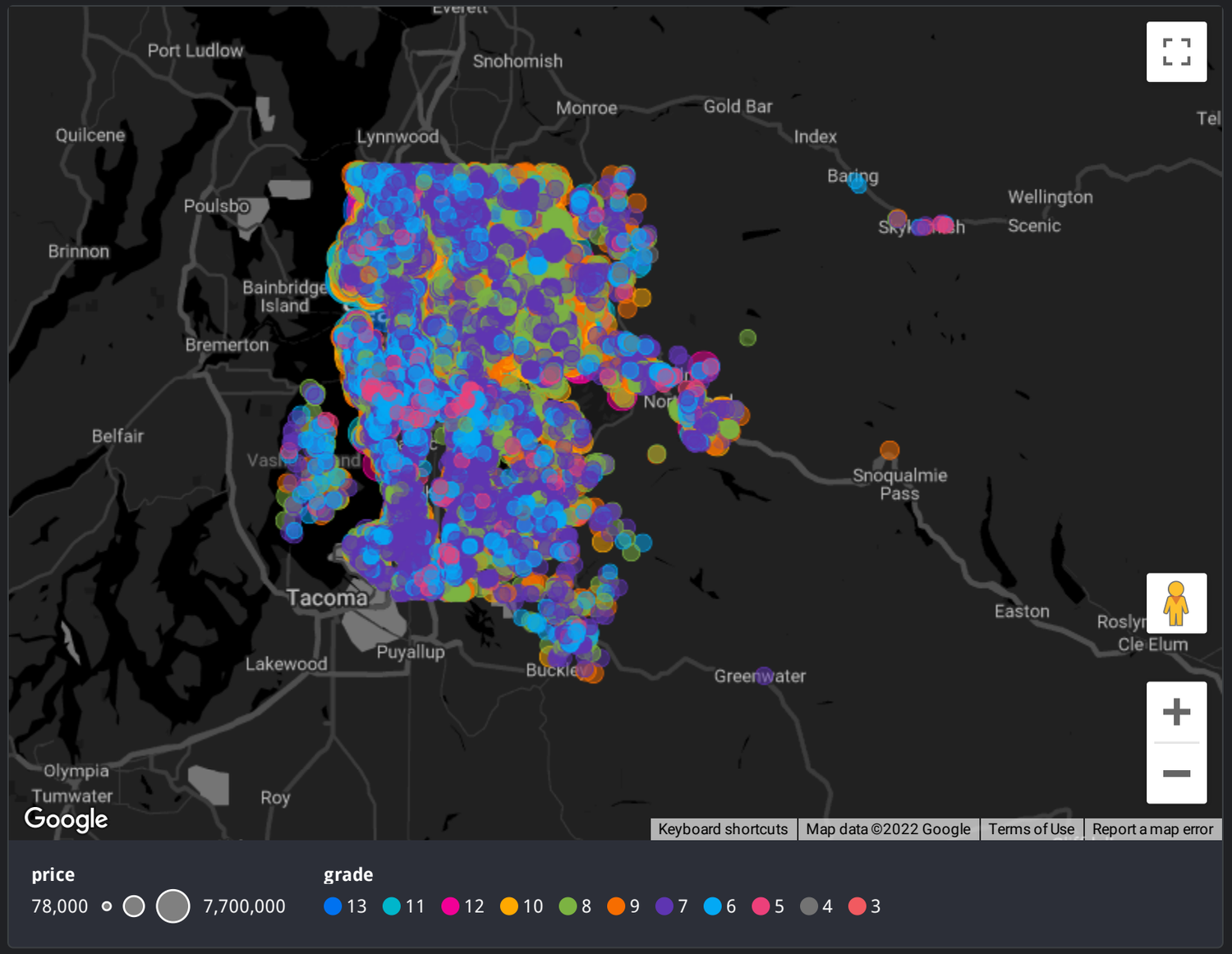




- No clear dependency on the price is visible e.g Very high variance for the same attribute (A house with best condition 5 can have a price which covers the whole spectrum)

Seattle Real Estate Inc.
HOLDING INVESTMENTS

Location decides the price!



Price in $

78,000 — 7,700,000

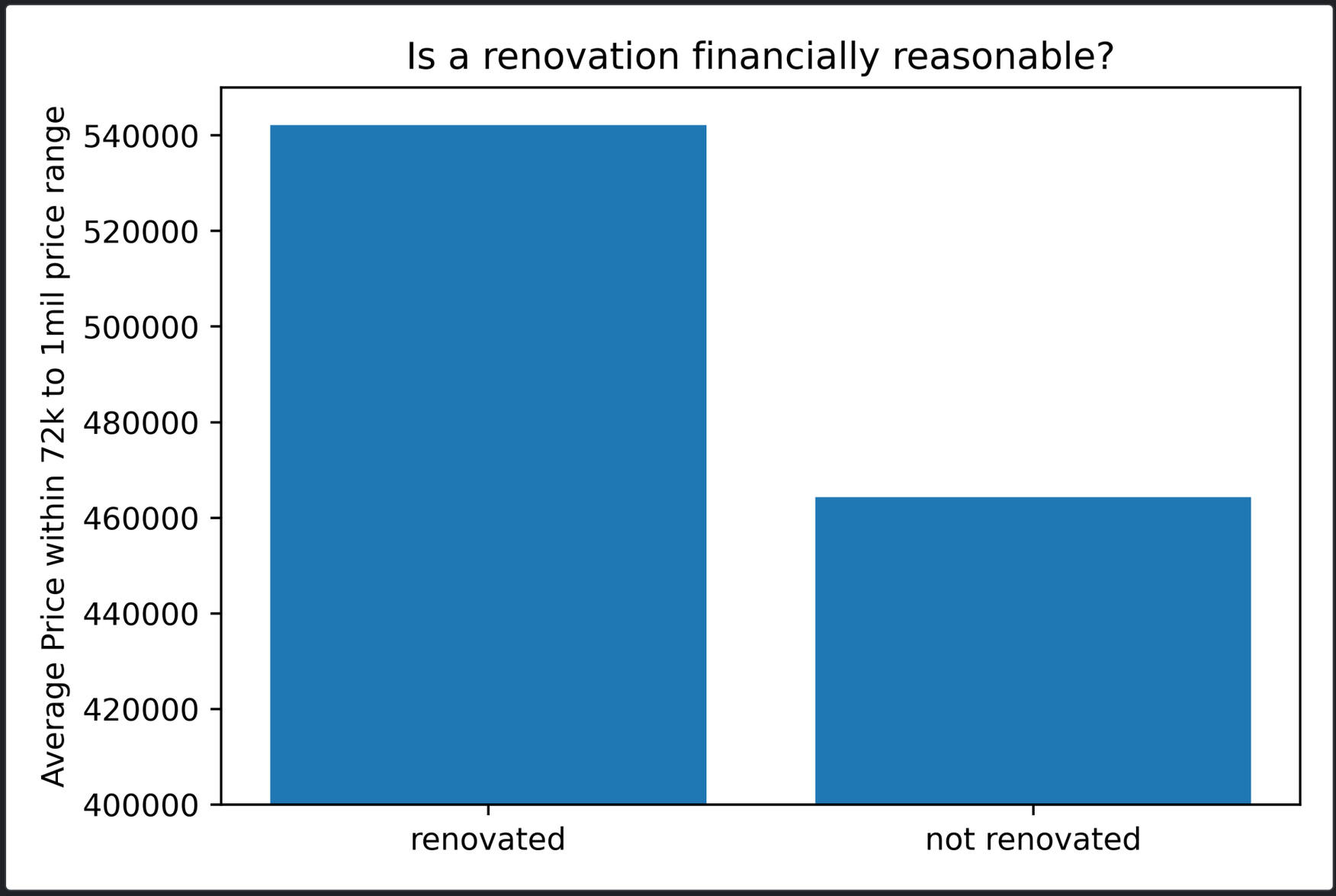
- Set price range from 1.000.000$ to 7.700.000$

- We can clearly see one massive cluster in the Downtown region of Seattle

- City regions are way more expensive than the suburbs or country side

Seattle Real Estate Inc.
HOLDING INVESTMENTS

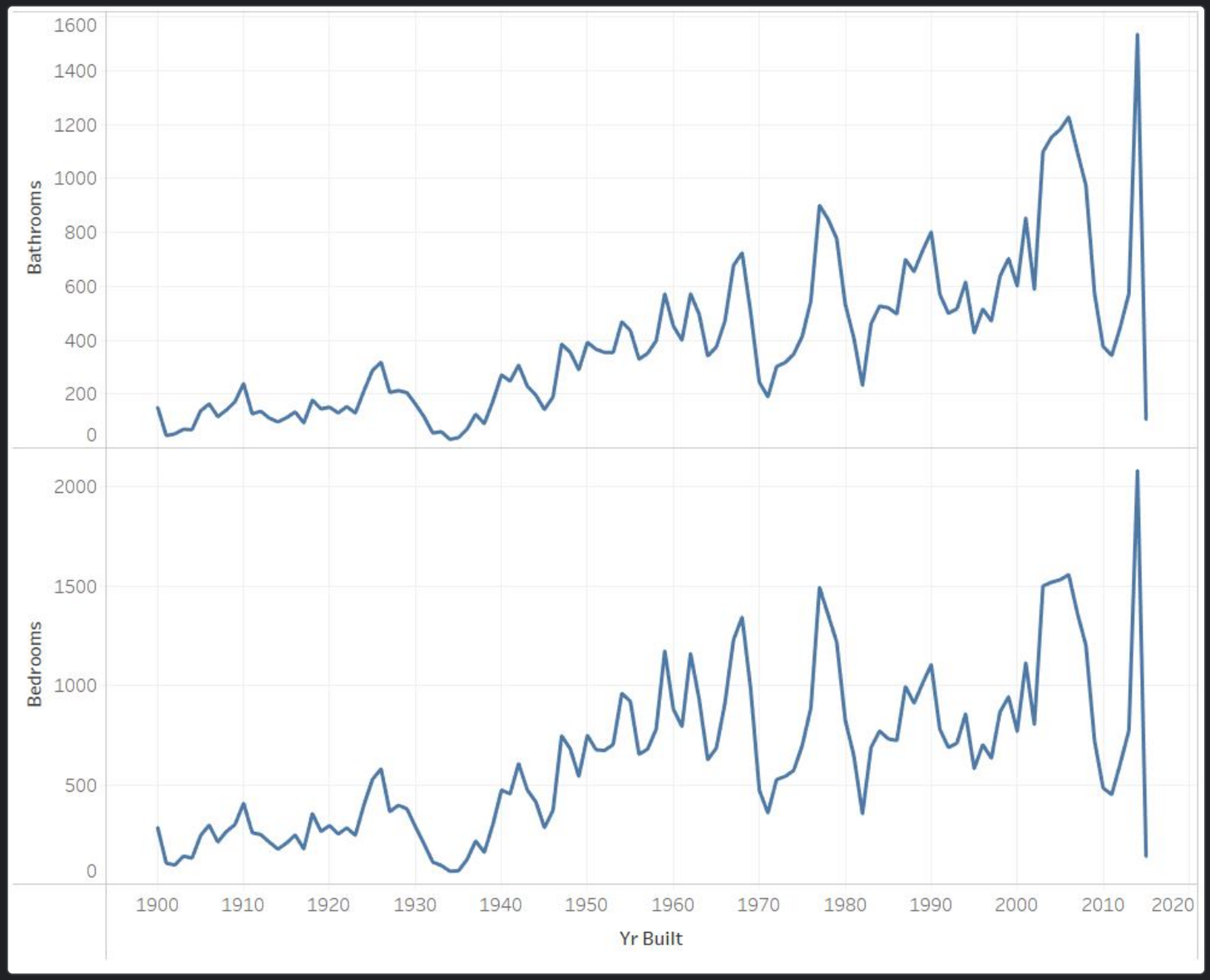
Renovation - reasonable?



- Price difference may be interpreted as a possible renovation budget
- Renovate the house for less than in this case ~70k$ and keep the rest as profit

Seattle Real Estate Inc.
HOLDING INVESTMENTS

Bed- and Bathroom crisis?



- We can clearly see massive dips in the bed- and bathroom amount in the ~1930s, ~1970s, ~1980s and from 2007-2010
- Bed- and Bathroom crisis?
- Probably not, rather: 2nd World War, Oil crisis, 1st Gulf War and the Housing Bubble of 2008

Seattle Real Estate Inc.

HOLDING INVESTMENTS

Thank You

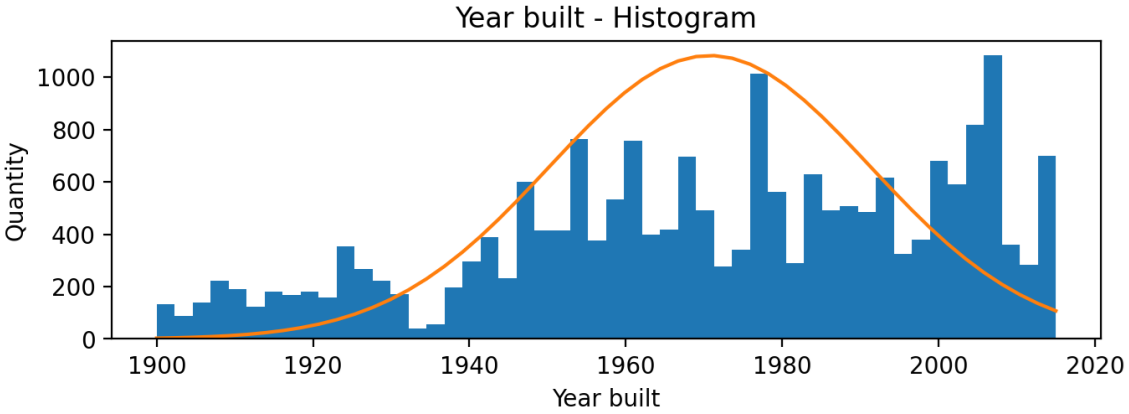
Normality Analysis

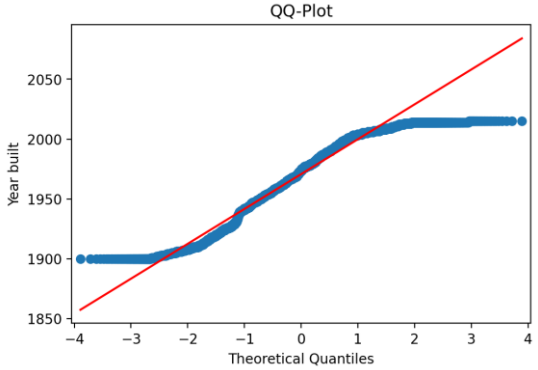
Housing Data - Seattle

Setup

1. Cleaned the data set for outliers (e.g Price (0-1mil$) or Lot space (0-20000 ft^2)
2. Plot Histogram with theoretical gauss curve
3. Calculate mean, standard deviation, skewness and kurtosis
4. Perform Kolmogorov Sminov and Shapiro Wilk's test (Sample size: random 5000)
5. Plot QQ-Plot
6. Transform the data

Year built



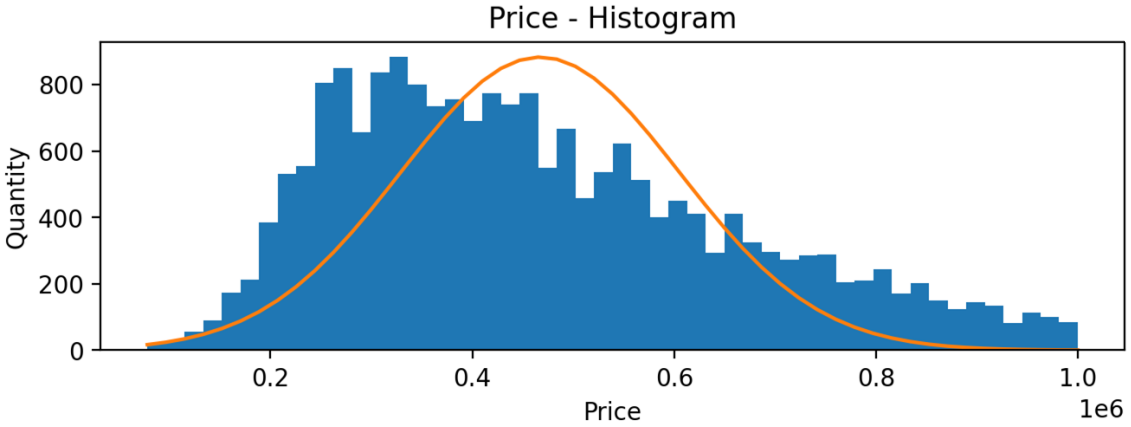


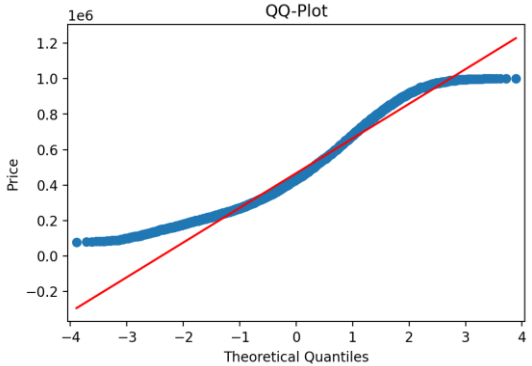
Skewness:-0.46 Kurtosis:-0.64 Mean: 1970.76 Std: 29.1

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- The result of normality test from both Shapiro and Kolmogorov-Sminov arrive at a p-value which is less than 0.05 which shows that the variable **year built** is not normally distributed

Price



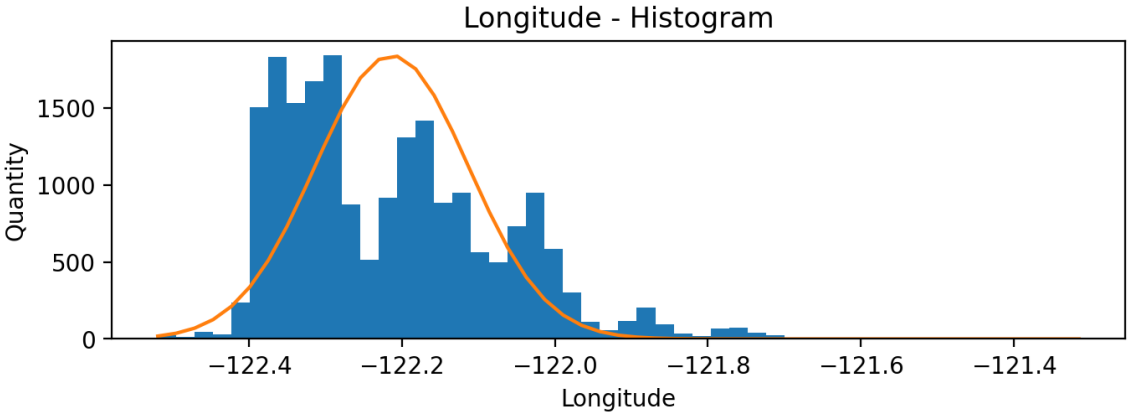


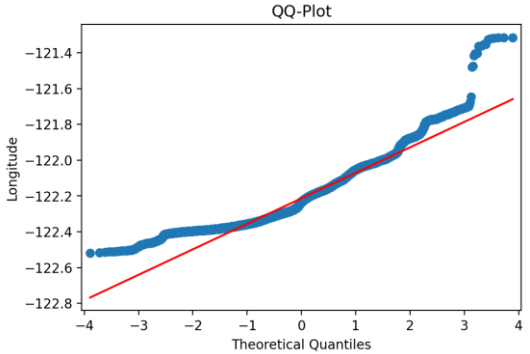
Skewness:0.62 Kurtosis:-0.33 Mean: 467169.75 Std: 195452.45

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- The result of normality test from both Shapiro and Kolmogorov-Sminov arrive at a p-value which is less than 0.05 which shows that the variable **price** is not normally distributed
- Since **price** is continuous and the sample is large, the variable should be transformed before further analysis

Longitude



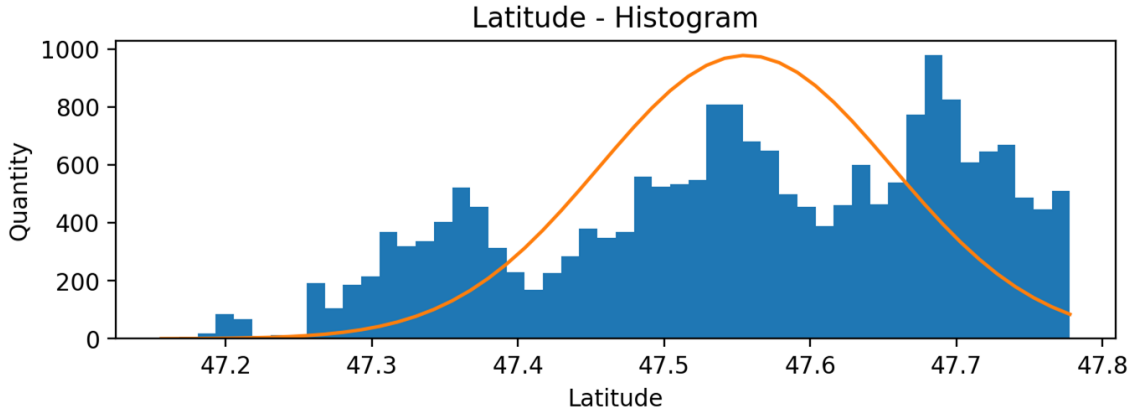


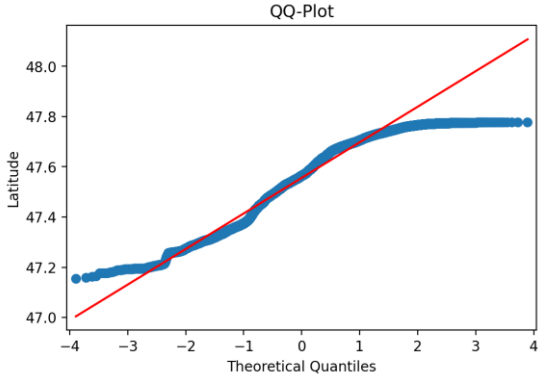
Skewness:0.89 Kurtosis:1.01 Mean: -122.21 Std: 0.14

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- A ‘natural’ variable?
- Not connected to real estate: The dips at -122.3° and -122.1° correspond to two lakes which divide the Seattle city → There obviously aren’t any houses on water

Latitude



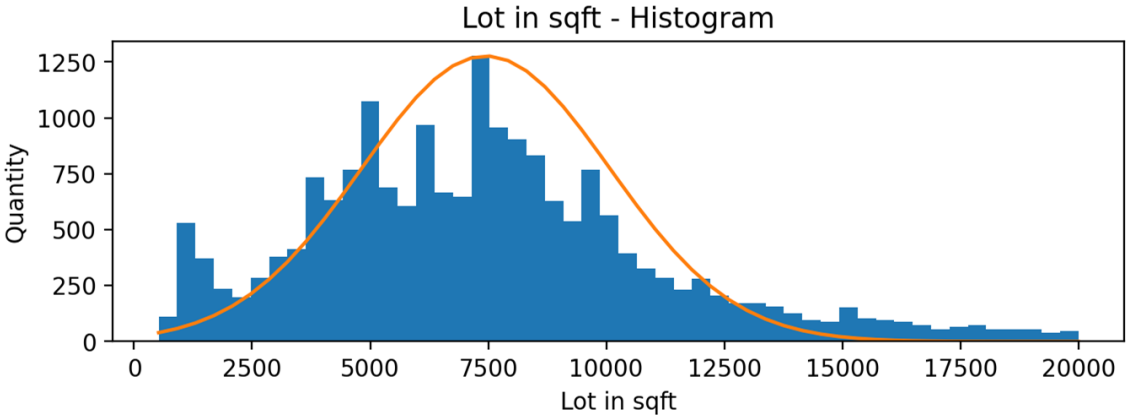


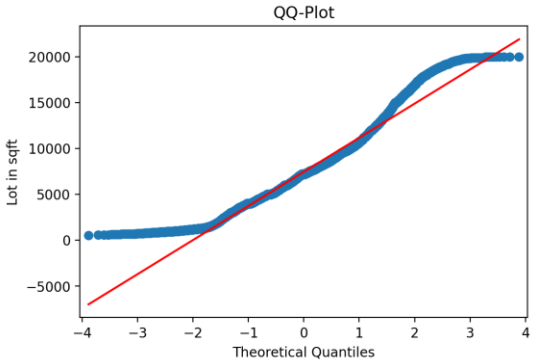
Skewness:-0.42 Kurtosis:-0.8 Mean: 47.56 Std: 0.14

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- Similar conclusion to the slide before (Longitude)
- Both variables can’t be normally distributed

Lotspace



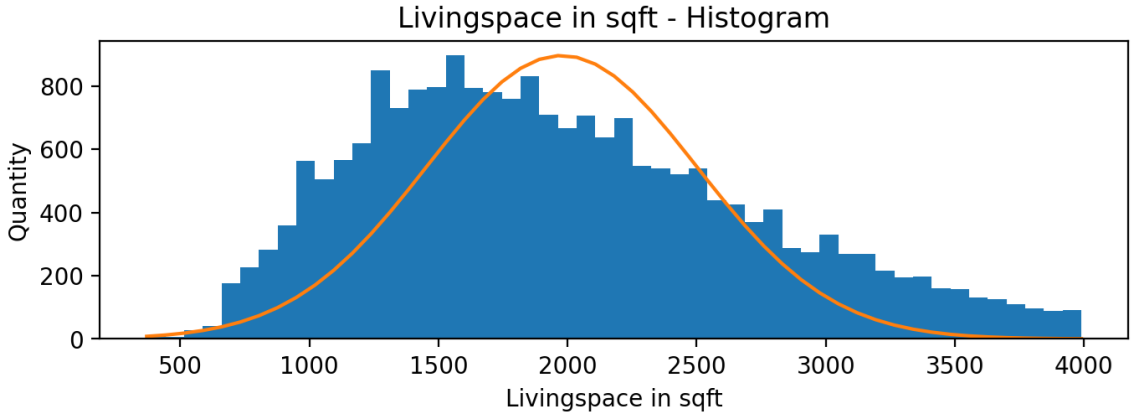


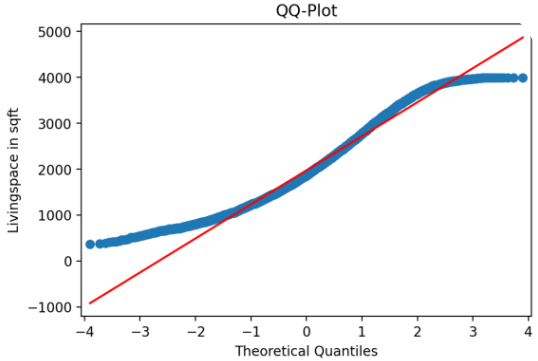
Skewness:0.72 Kurtosis:0.65 Mean: 7448.15 Std: 3717.38

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- The result of normality test from both Shapiro and Kolmogorov-Sminov arrive at a p-value which is less than 0.05 which shows that the variable **lotspace** is not normally distributed
- Since this variable is continuous and the sample is large, it should be transformed before further analysis.

Livingspace



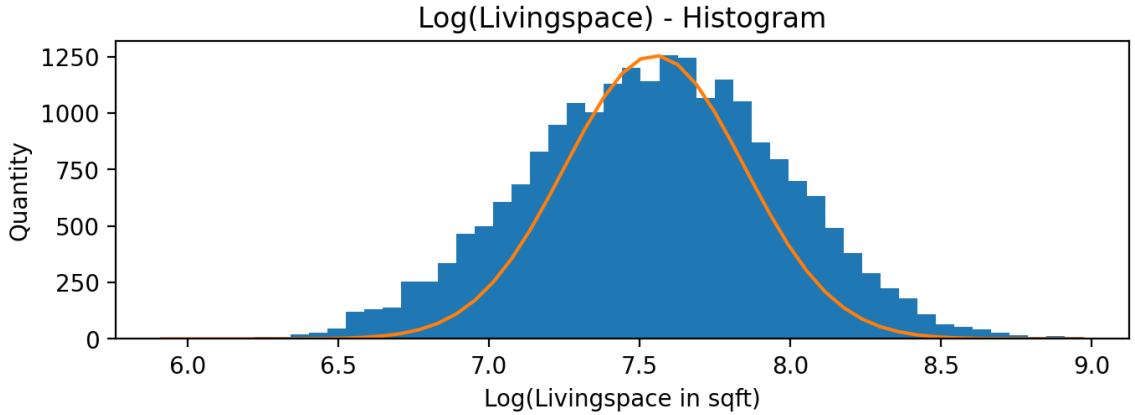


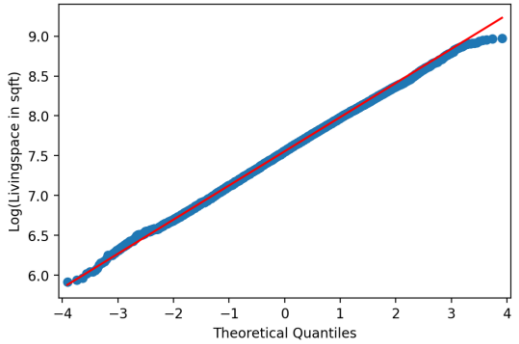
Skewness:0.51 Kurtosis:-0.38 Mean: 1977.21 Std: 741.46

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- **Livingspace** is not normally distributed
- Since **livingspace** is continuous variable and the sample is large, it should be transformed before further analysis

What's next? - Transform Data



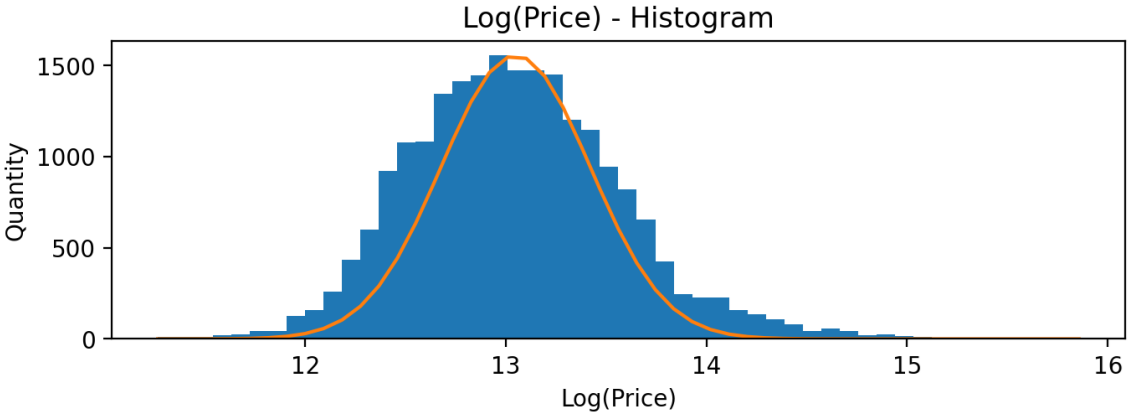


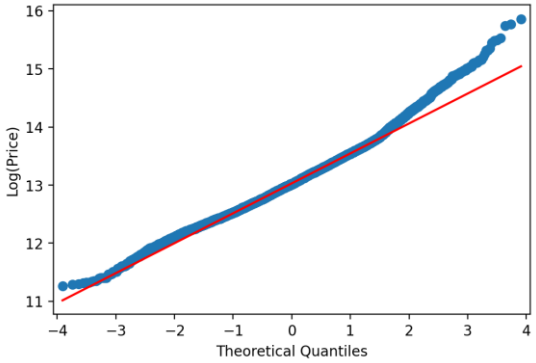
Skewness:-0.05 Kurtosis:-0.15 Mean: 7.55 Std: 0.42

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- Visually the data looks a lot more 'normal' now: QQ-Plot almost perfectly on the 45° degree line
- However the Shapiro and Kolmogorov test still reject the Null-Hypotheses of a normal distribution

Price - Log transformed



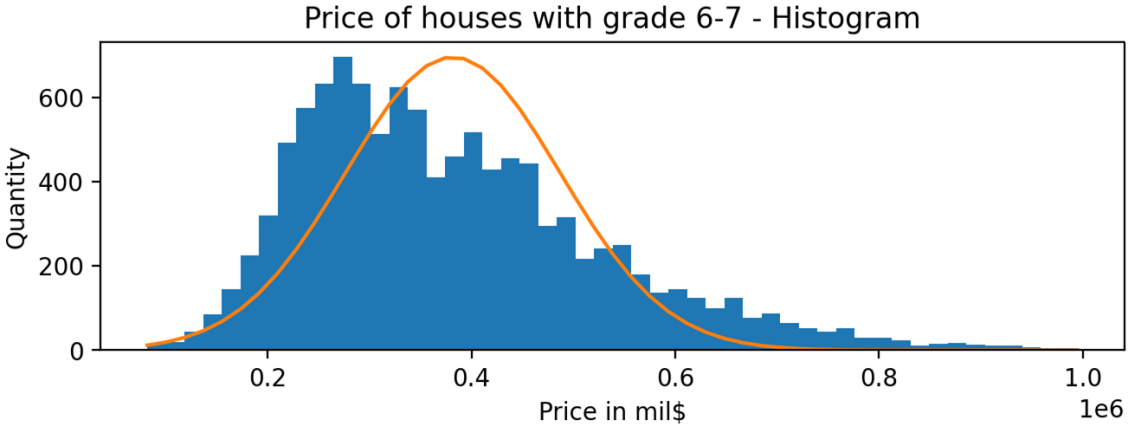


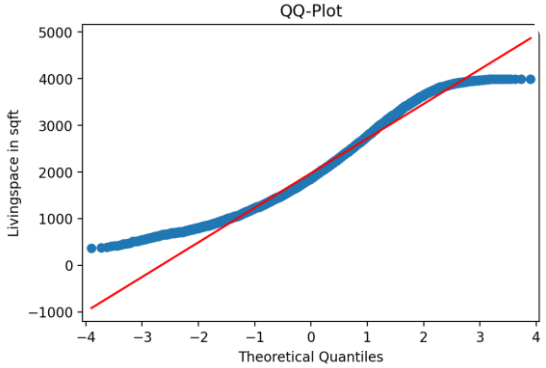
Skewness:0.43 Kurtosis:0.69 Mean: 13.05 Std: 0.53

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- Similarly to the slide before the data looks a lot more ‘normal’ than before
- However the Shapiro and Kolmogorov test also still reject the Null-Hypotheses of a normal distribution

Categorical Analysis





Skewness:0.92 Kurtosis:0.78 Mean: 381761.13 Std: 148666.94

Shapiro p-value: 0.0 Kolmogorov p-value: 0.0

- **Grade 6-7** is the median grade of all houses. More than **50%** of the houses are in this category.
- The result of normality test from both Shapiro and Kolmogorov-Sminov arrive at a p-value which is less than 0.05 which shows that the variable **price for median rated houses** is not normally distributed.

Aim of the analysis

The aim of the analysis is to check the normality of the prices based on the number of bedrooms

Description of variables

The variable that will be used include:

- Price: this contain continuous observations
- Bedrooms: it contain the number of bedrooms. The variable contain categorical observations

Statement of Hypothesis

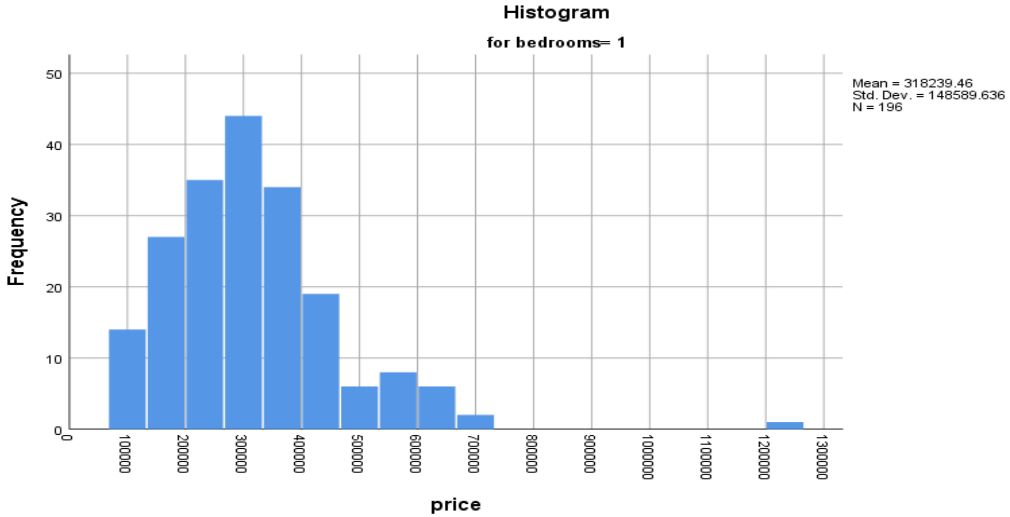
H0 (null hypothesis): The data is normally distributed

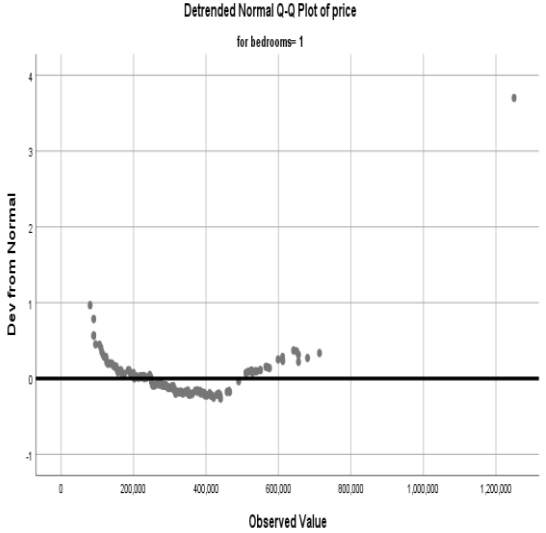
H1: (alternative hypothesis): The data is normally distributed

Decision Rule

Reject null hypothesis (H0) when p-value is less than 0.05

Normality test of Price when bedroom = 1





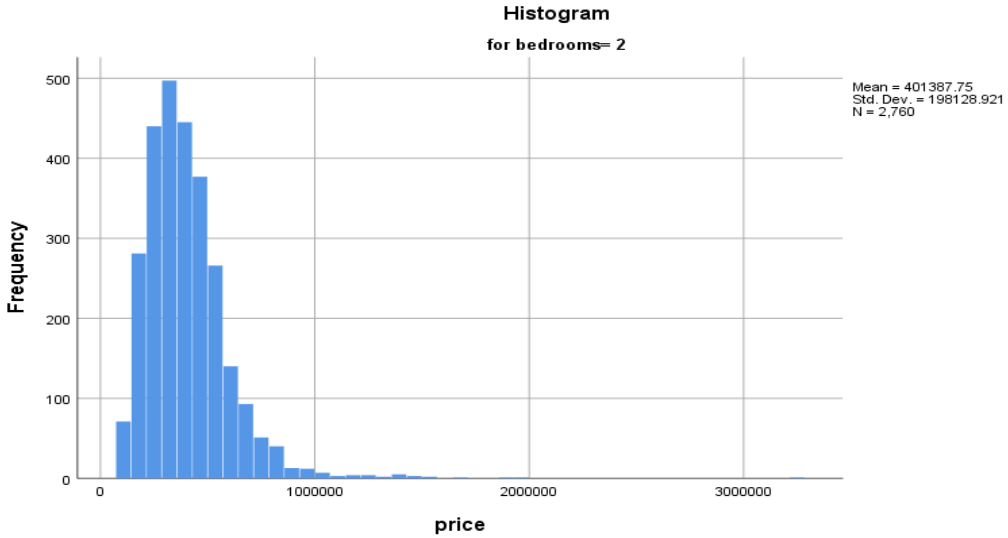
- Kolmogorov Sminov: p-value = 0.01
- Shapiro Wilk: p-value =0.00
- Skewness = 1.686
- Kurtosis = 7.206

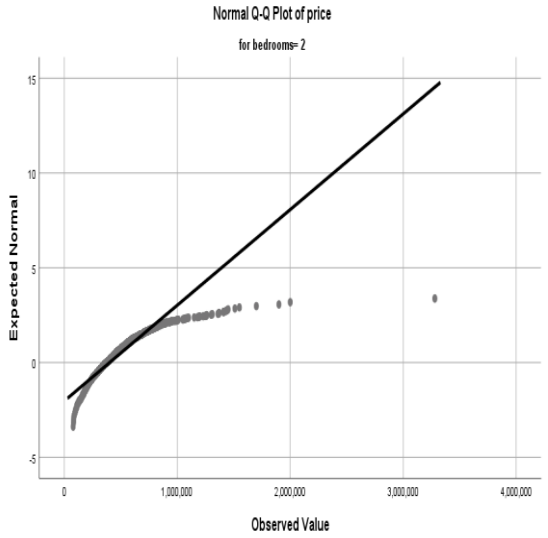
Interpretation of normality test of price when bedroom = 1

- The histogram shows the data is a bit skewed to the right
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 1 is not normally distributed

Normality test of price when bedroom = 2





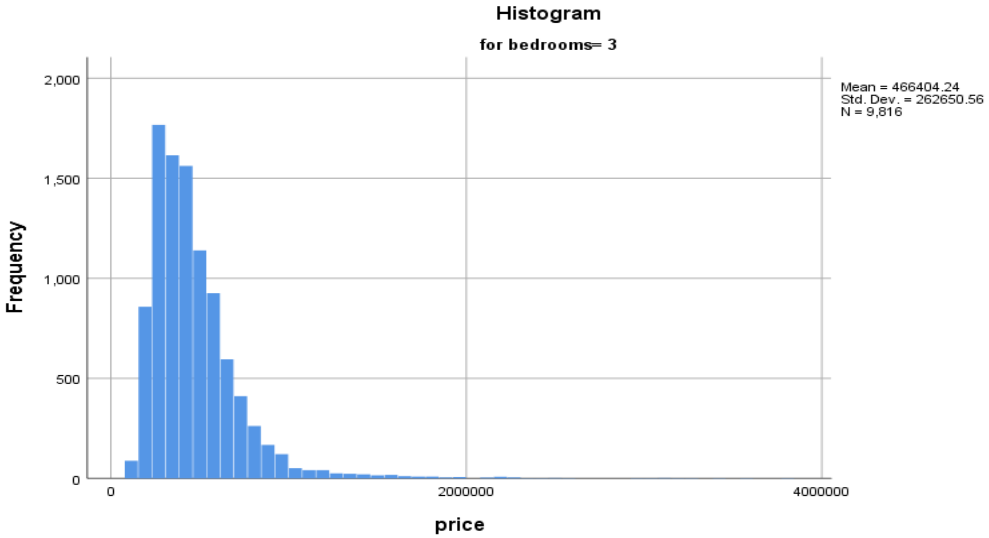
- Kolmogorov Sminov: p-value = 0.00
- Shapiro Wilk: p-value =0.00
- Skewness = 2.761
- Kurtosis = 22.335

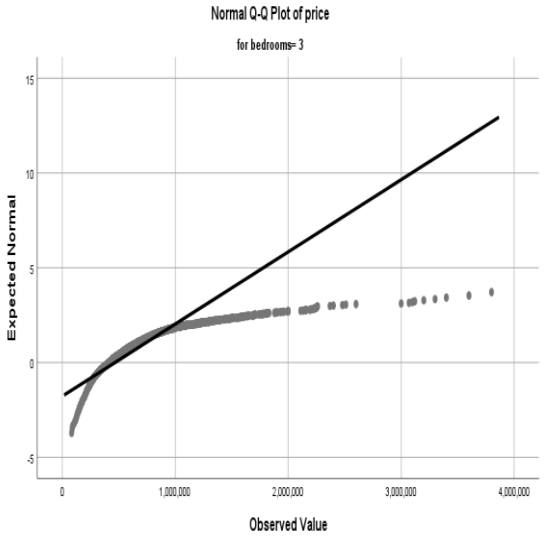
Interpretation of normality test of price when bedroom = 2

- The histogram shows the data is a bit skewed to the right
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 2 is not normally distributed

Normality test of price when bedroom = 3





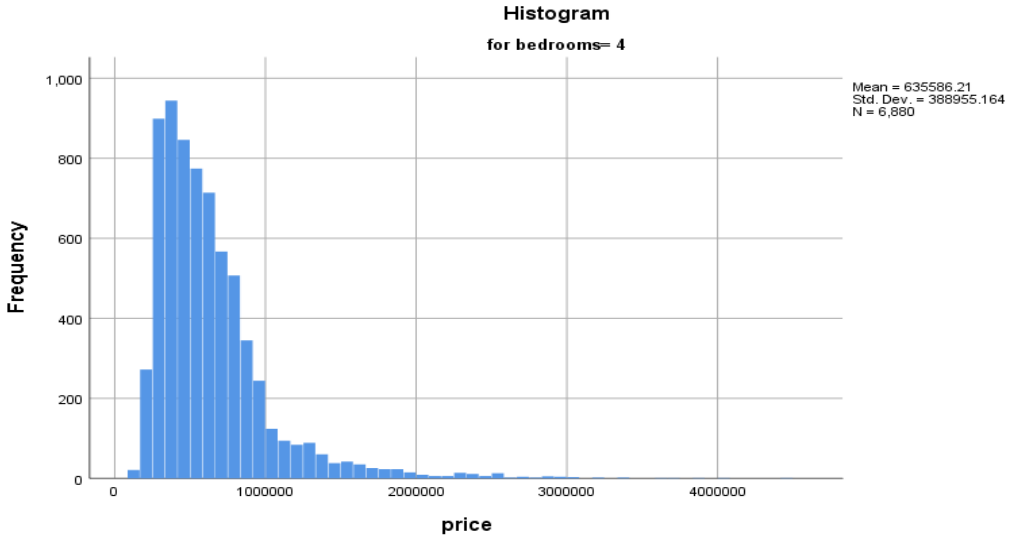
- Kolmogorov Sminov: p-value = 0.00
- Skewness = 3.198
- Kurtosis = 20.596

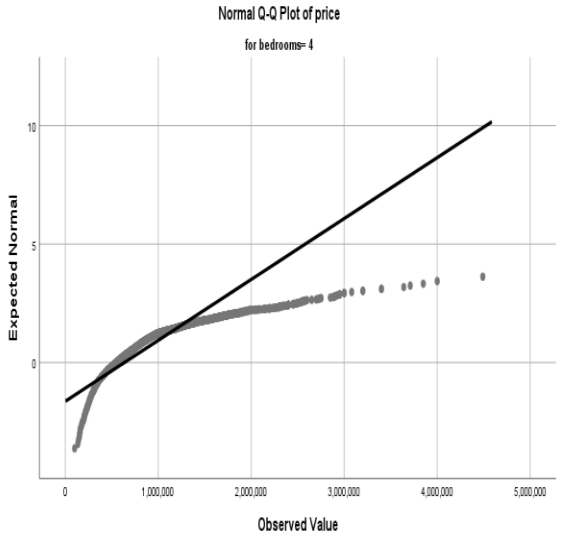
Interpretation of normality test of price when bedroom = 3

- The histogram shows the data is a bit skewed to the right
- Most of the data points on the QQ-plot does not fall on the line
- Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 3 is not normally distributed

Normality test of price when bedroom = 4





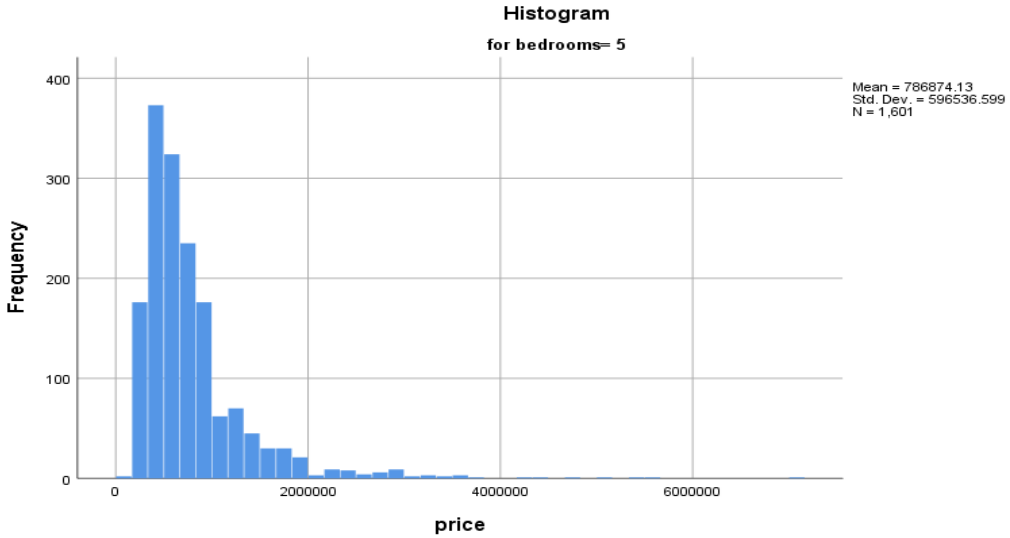
- Kolmogorov Sminov: p-value = 0.00
- Skewness = 2.579
- Kurtosis = 10.971

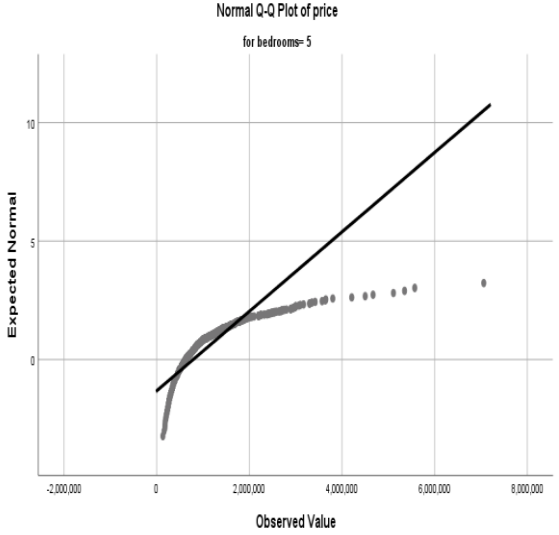
Interpretation of normality test of price when bedroom = 4

- The histogram shows the data is a bit skewed to the right
- Most of the data points on the QQ-plot does not fall on the line
- Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 4 is not normally distributed

Normality test of price when bedroom = 5





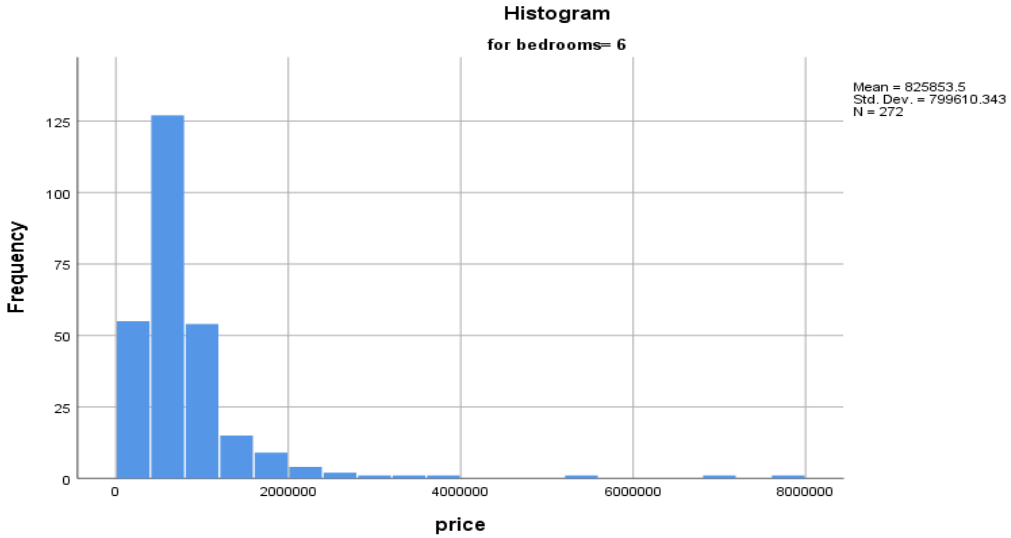
- Kolmogorov Sminov: p-value = 0.00
- Shapiro Wilk: p-value = 0.00
- Skewness = 2.579
- Kurtosis = 10.971

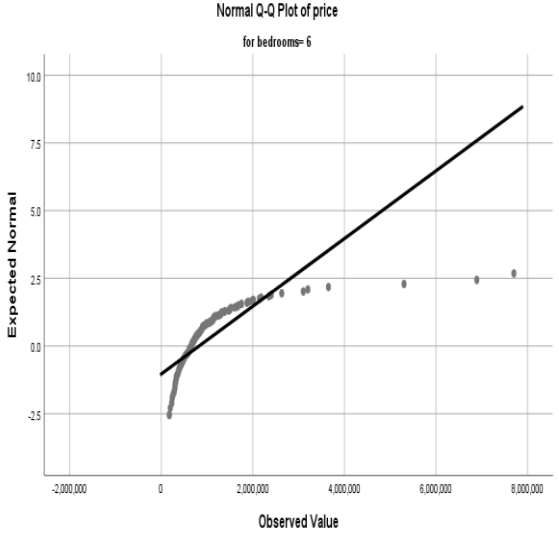
Interpretation of normality test of price when bedroom = 5

- The histogram shows the data is a bit skewed to the right
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 5 is not normally distributed

Normality test of price when bedroom = 6





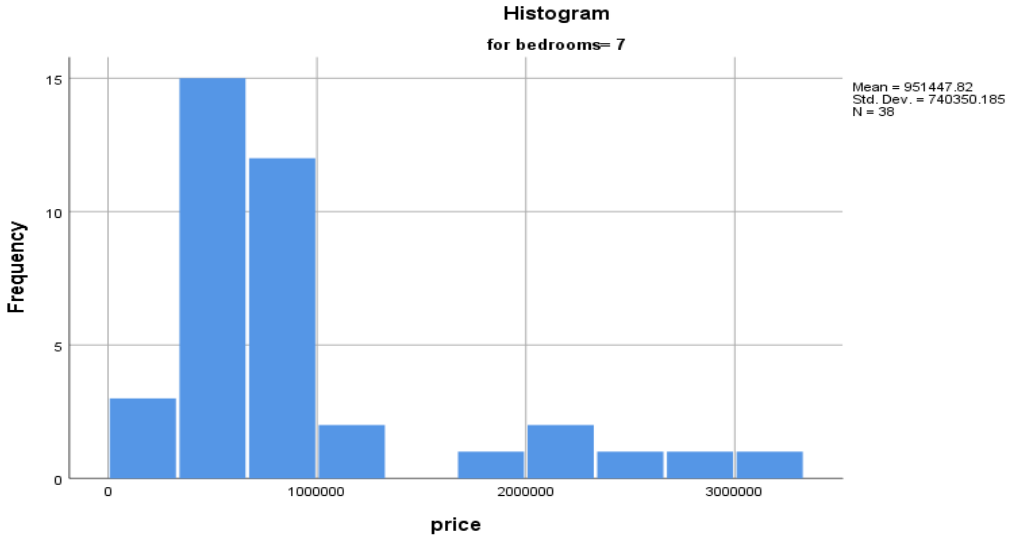
- Kolmogorov Sminov: p-value = 0.00
- Shapiro Wilk: p-value = 0.00
- Skewness = 5.128
- Kurtosis = 35.292

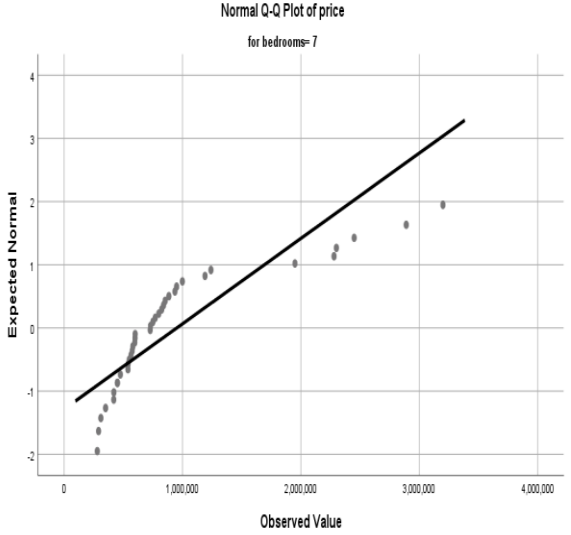
Interpretation of normality test of price when bedroom = 6

- The histogram shows the data is a bit skewed to the right
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 6 is not normally distributed

Normality test of price when bedroom = 7





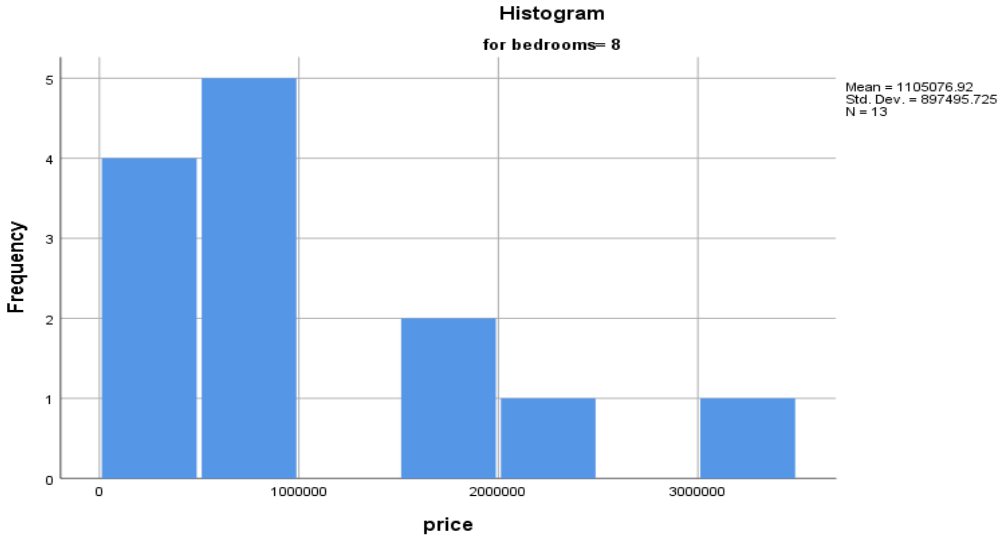
- Kolmogorov Sminov: p-value = 0.00
- Shapiro Wilk: p-value = 0.00
- Skewness = 1.798
- Kurtosis = 2.389

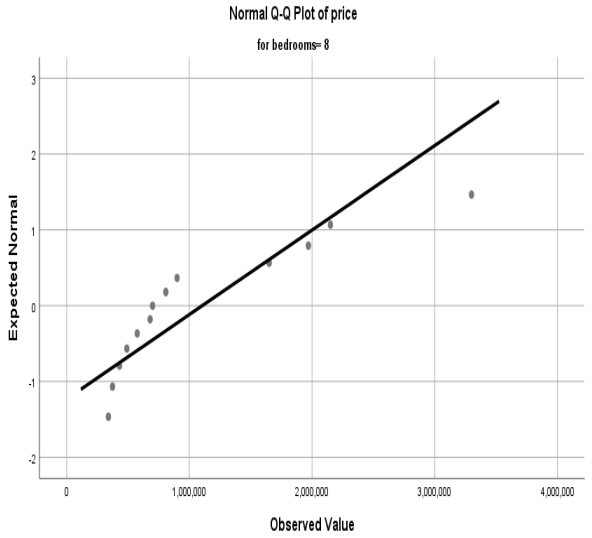
Interpretation of normality test of price when bedroom = 7

- The histogram shows the data is not skewed
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is less than 0.05
- Skewness is greater than 1, this depict that the data is positively skewed
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: prices when bedroom = 7 is not normally distributed

Normality test of price when bedroom = 8





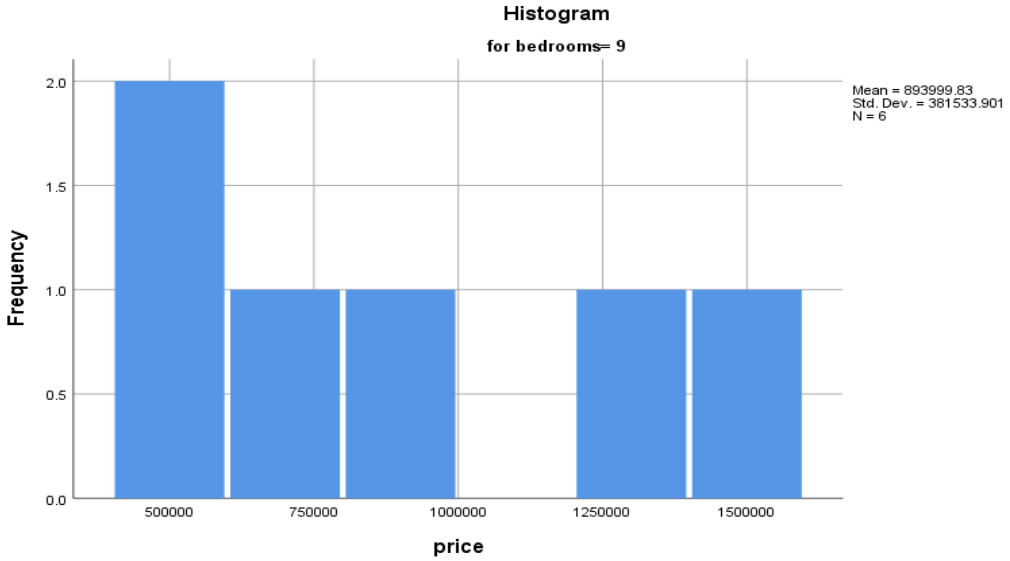
- Kolmogorov Sminov: p-value = 0.06
- Shapiro Wilk: p-value = 0.08
- Skewness = 1.474
- Kurtosis = 1.636

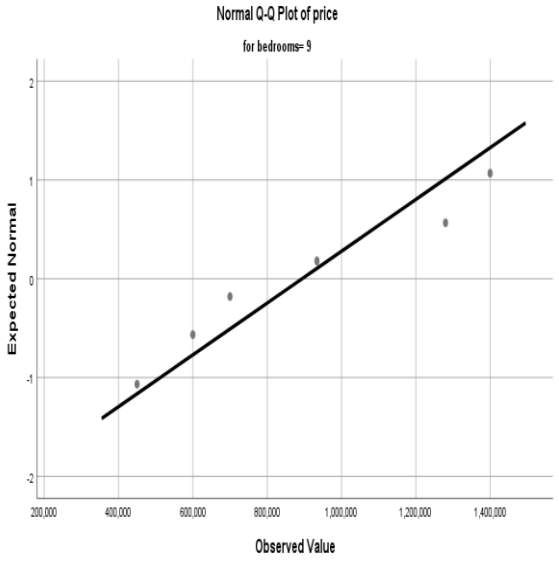
Interpretation of normality test of price when bedroom = 8

- The histogram shows the data is not skewed
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is greater than 0.05
- Skewness is greater than 1
- Kurtosis is greater than 1 which depict the distribution is highly peaked

Conclusion: the normality test for prices when bedroom = 8 is approximately normal. But the visuals show the prices are not normally distributed. Since N= 13 and plot is not the best method to test the normality of data for N < 30, we conclude that the prices of apartments with 8 bedrooms is normally distributed.

Normality test of price when bedroom = 9





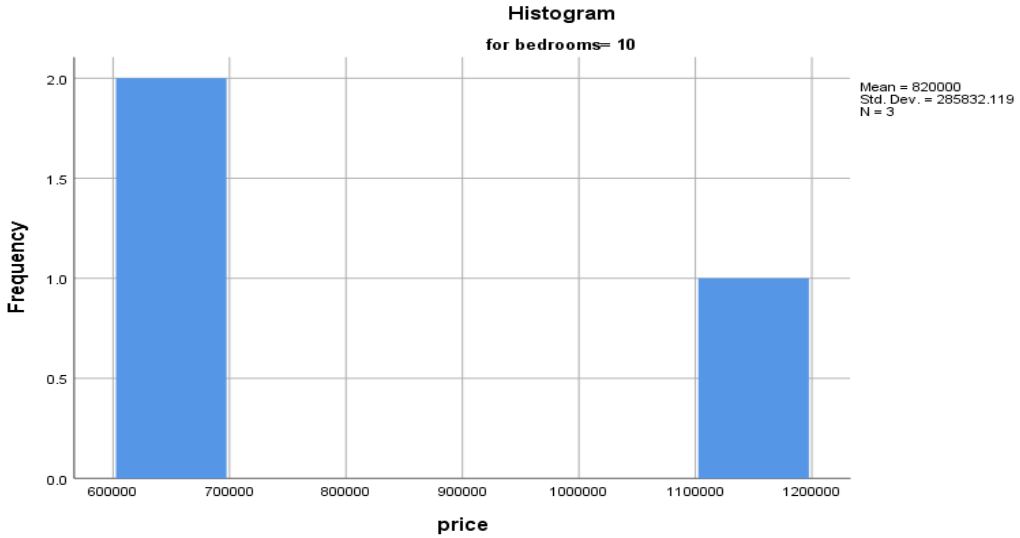
- Kolmogorov Sminov: p-value = 0.200
- Shapiro Wilk: p-value = 0.574
- Skewness = 0.381
- Kurtosis = -1.7774

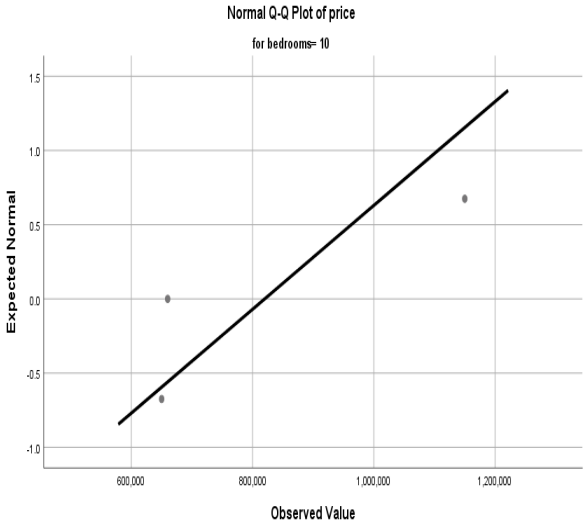
Interpretation of normality test of price when bedroom = 9

- The histogram shows the data is not skewed
- Most of the data points on the QQ-plot does not fall on the line
- Both Shapiro WIlk and Kolmogorov Sminov p-values is greater than 0.05
- Skewness is less than 1, this depict that the data is symmetric
- Kurtosis is greater than one shows that the distribution is too peaked

Conclusion: the normality test for prices when bedroom = 8 is approximately normal. But the visuals show the prices are not normally distributed. Since N = 6 and plots is not the best normality test samples N < 30, we conclude using Shapiro Wilk and Kolmogorov Sminov test that the price of apartments with 8 bedrooms are normally ditributed

Normality test of price when bedroom = 10





- Shapiro Wilk: p-value = 0.033
- Skewness = 1.730

Interpretation of normality test of price when bedroom = 10

- The histogram shows the data is not skewed
- Most of the data points on the QQ-plot does not fall on the line
- Shapiro WIlk p-values is less than 0.05
- Skewness is greater than 1, this depict the data is positively skewed

Conclusion: prices when bedroom = 10 is not normally distributed

Summary of Findings

- Price of all the bedrooms need transformation except for price of 8 and 9 bedrooms
- All prices should be standardized before proceeding to analysis

The Housing Market doesn’t seem to be normally distributed.

Thank You

T-Tests and ANOVA

Seattle - Real Estate Data

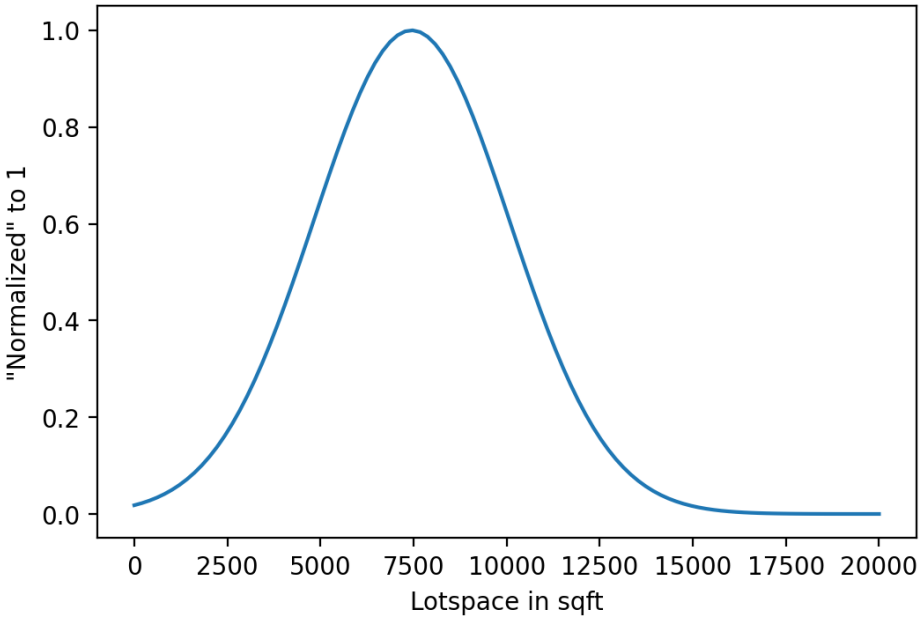
Aim of the Analysis - Methodology

- Perform some T-Tests
- Check weather or which attributes have a distinctive effect on the price of a property

  → Is the price of real estate with different **grades, conditions,** etc. distinct in comparison to each other

- When the sample sizes were massively different from each other, we chose a random subset from the bigger sample with the size of the smaller one and ran the test multiple times

One sample T-Test



- Table 1: One sample T-Test with **H0** = {mean = 7448sqft}
- p-val > 0.05
- Accept **H0**
- Table 2: One sample T-Test with **H0** = {mean = 7600sqft}
- p-val < 0.05
- Reject **H0**

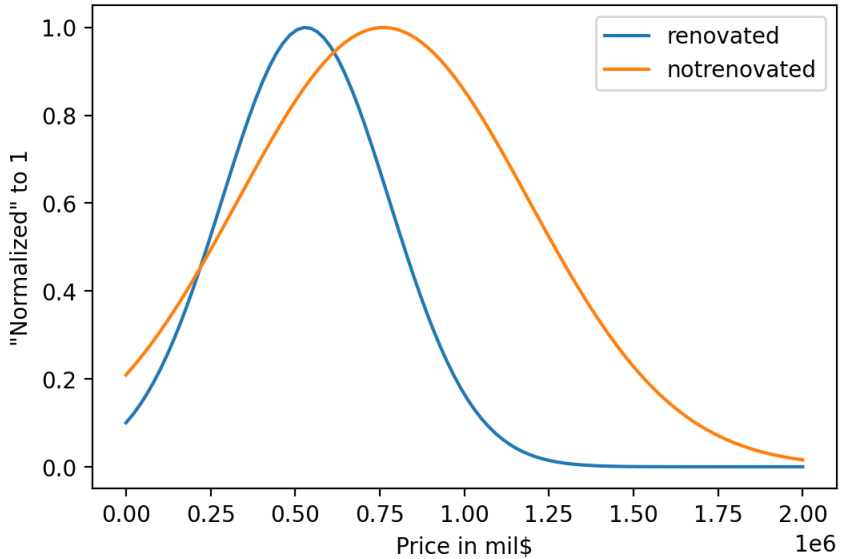
	Hypotheses	statistic	pvalue
T-Test	$7448 ft^2$	0.0055	0.996

Table 1: T-Test: Lotspace

	Hypotheses	statistic	pvalue
T-Test	$7600 ft^2$	-5.67	$1.43 \cdot 10^{-8}$

Table 2: T-Test: Lotspace

Two sample T-Test - Renovation



	statistic	pvalue
T-Test	9.917	1.382e-22

Table 3: T-Test: Renovated vs. Not Renovated

- 'Price-Test' between **Renovated** and **Not Renovated** Houses
- p-val < 0.05
- **Renovated** and **Not renovated** are distributed distinctively

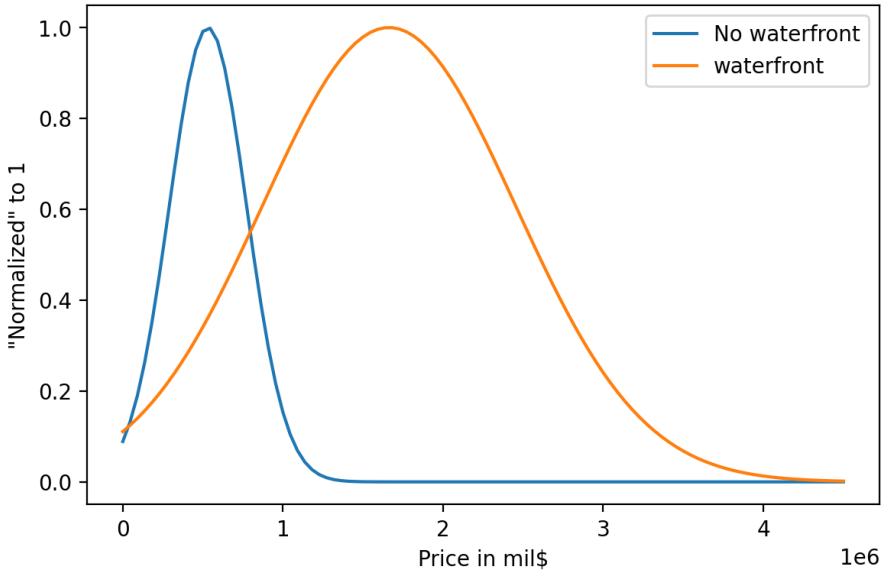
Same T-Test but with more information

	T	dof	alternative	p-val	CI95	cohen-d	BF10	power
T-test	9.91	1477.313	two-sided	1,8e-25	[185463.41 277002.63]	0.464	2.244e+19	1.000

Table 4: T-Test: Renovated vs. Not Renovated

- Arrives at the same conclusion like the test before
- p-val < 0.05
- Reject **H0**
- Power ß = 1
- Probability to correctly reject **H0** is at ~100%

ANOVA - Waterfront

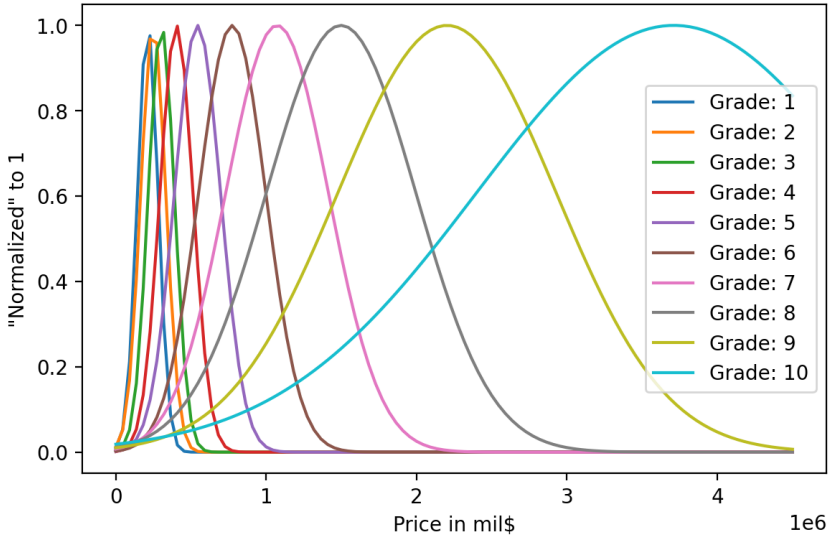


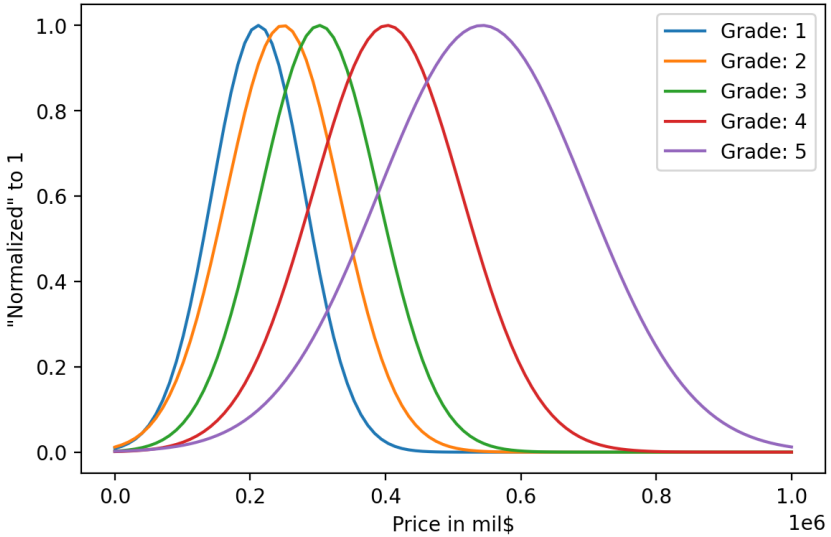
	Source	ddof1	ddof2	F	p-unc	np2
0	waterfront	1	324	176.607134	1,7e-22	0.352786

Table 5: One way ANOVA: Waterfront

- 'Price-ANOVA' between **waterfront** and **No waterfront** Houses
- p-val < 0.05
- **waterfront** and **No waterfront** are distributed distinctively

ANOVA - Grade





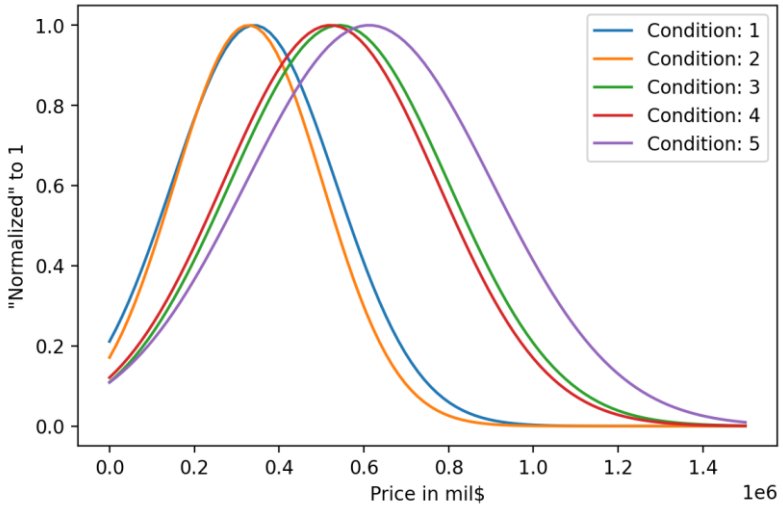
	Source	ddof1	ddof2	F	p-unc	np2
0	grade	10	21585	2337.310688	2,7e-21	0.519886

Table 6: 10-way ANOVA: Grade

Explanation

- ‘Price-ANOVA’ with **grade** as independent variable
- p-val < 0.05
- Different **grades** are distributed distinctively

ANOVA - Condition



	Source	ddof1	ddof2	F	p-unc	np2
0	condition	4	21591	36.851461	7,7e-24	0.006781

Table 7: 5-way ANOVA: Condition

- 'Price-ANOVA' with **condition** as independent variable
- p-val < 0.05
- Different **conditions** are distributed distinctively
- Check 1-2 and 3-4 separately: Means are very close together

Two sample T-Test between similar conditions

- Two sample T-Test between condition 1-2 and 3-4

	statistic	pvalue
T-Test	0.255	0.798

Table 8: T-Test: Condition 1-2

	statistic	pvalue
T-Test	2.811	0.0049

Table 9: T-Test: Condition 3-4

- p-val > 0.05 for condition 1-2: Picked from the same distribution (accept **H0**)
- No significant difference between 1-2 condition regarding price
- p-val < 0.05 for condition 3-4: Not picked from the same distribution (reject **H0**)

→ Reduce condition scale to a 4 step scale (1,2,3,4)?

TWO WAY ANOVA OF THE EFFECT OF YEAR BUILT AND YEAR RENOVATED ON PRICE

H0: Year built has no significant effect on price
H0: Year renovated has no significant effect on price
H0: the interaction between year built and year renovated does not have significant effect on price

Tests of Between-Subjects Effects

Dependent Variable: price

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	4.506E+14[a]	860	5.239E+11	4.408	.000	.155	3790.655	1.000
Intercept	1.741E+14	1	1.741E+14	1464.462	.000	.066	1464.462	1.000
yr_built	8.470E+13	115	7.366E+11	6.197	.000	.033	712.642	1.000
yr_renovated	6.540E+13	69	9.479E+11	7.975	.000	.026	550.251	1.000
yr_built * yr_renovated	2.167E+14	676	3.205E+11	2.697	.000	.081	1822.877	1.000
Error	2.463E+15	20721	1.189E+11					
Total	9.215E+15	21582						
Corrected Total	2.913E+15	21581						

a. R Squared = .155 (Adjusted R Squared = .120)

b. Computed using alpha = .05

Explanation

- P-value < 0.05, the factors have significant effect on price
- Year built (0.033) has more effect on price than year renovated (0.026).
- The interaction between year built and year renovated has the highest effect size (0.081)
- The effect size shows both year built and year renovated has small effect on price (i.e. there other factors that has more effect on price)
- The power analysis shows that 1.000 for all factors. The factors might be overpowered

Most attributes divide the price of a property into distinctive Price ranges.