Dubai Marina Ready Property Analysis

Onkar Singh, August 2019

Executive Summary

This document presents an analysis of data concerning the ready properties in Dubai Marina. The analysis is based on 34887 observations of transactional data, each containing specific characteristics of the property and its price.
After exploring the data by calculating summary and descriptive statistics, and by creating visualizations of the data, several potential relationships between the area's characteristics and price were identified. After exploring the data, a predictive model was created, and finally a regression model to predict an a property's price from its features was created.

Initial Data Exploration

The initial exploration of the data began with some summary and descriptive statistics.

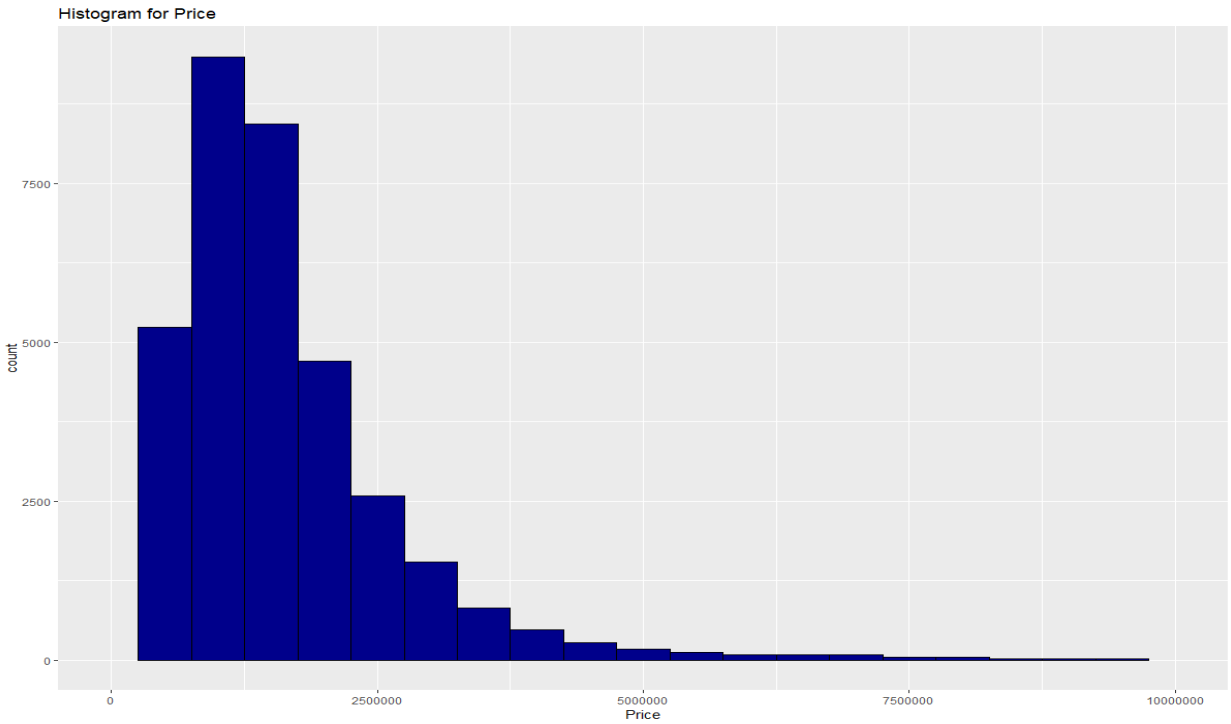
Individual Feature Statistics

Summary statistics for minimum, maximum, mean, median, standard deviation, and distinct count were calculated for numeric columns, and the results taken from 216 observations are shown here:

Column	Min	Max	Mean	Median	Std Dev
Size	32.29	38025.34	1376.31	1198.03	946.0552
Price	26449	38000000	1674817	1377199	1377236
Price/Size	27.98	17865.2	1264.64	1177.13	549.129

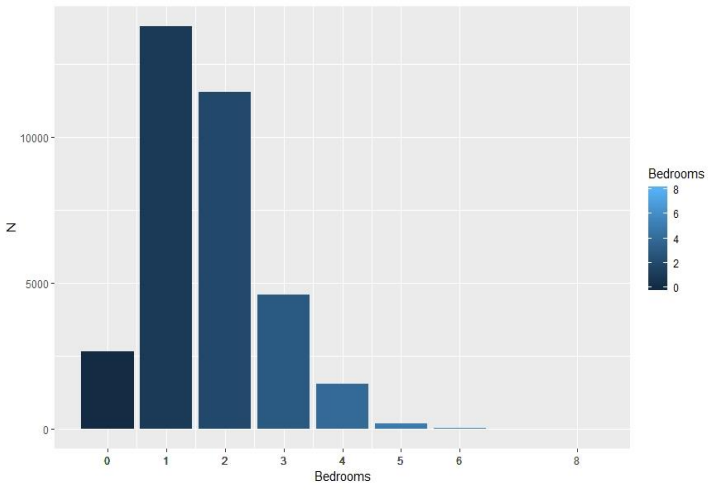
What do we plan on achieving from the exploration and analysis?

Since **Price** is of interest in this analysis, it was noted that the mean and median of this value are significantly different and that the comparatively large standard deviation indicates that there is considerable variance in the prices of the properties. A histogram of the **Price** column shows that the price values are right-skewed – in other words, most properties are priced at the lower end of the price range, as shown here:

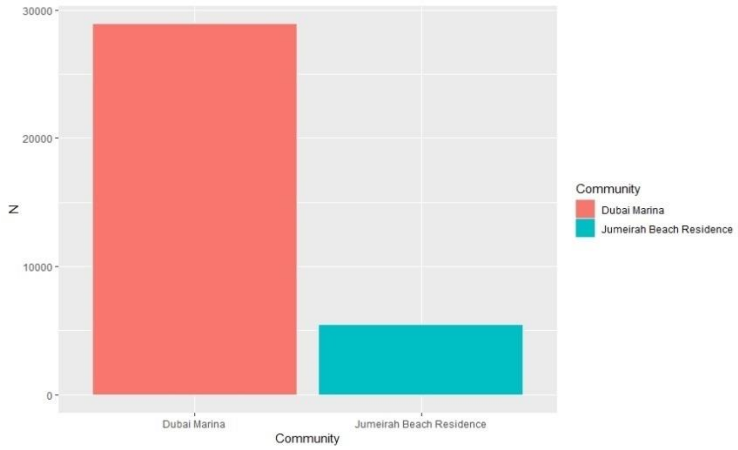


In addition to the numeric values, the observations include categorical features, including:

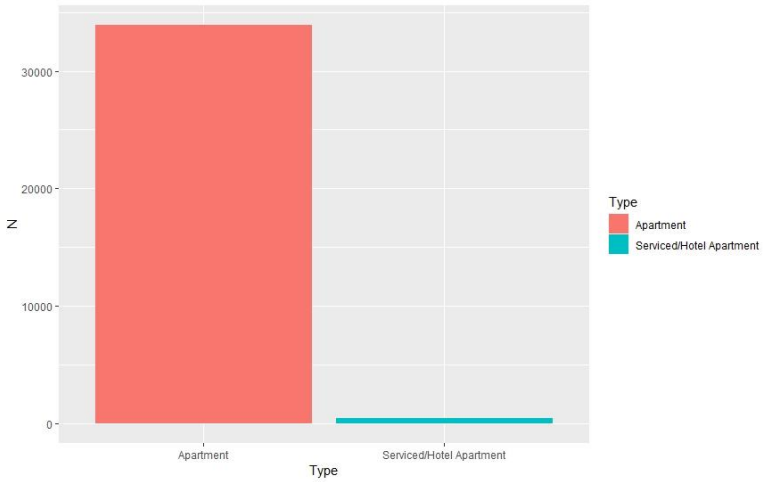
- **DATE** – Date of the Transaction
- **COMMUNITY** – Dubai Marina, Jumeirah Beach Residence
- **PROPERTY** – One of the 160 properties
- **DEVELOPER** – One of the 54 Developers
- **BEDROOMS** – 0(Studio) ,1 ,2 3, 4, 5, 6, 8
- **TYPE** – Apartment, Hotel/Serviced Apartment



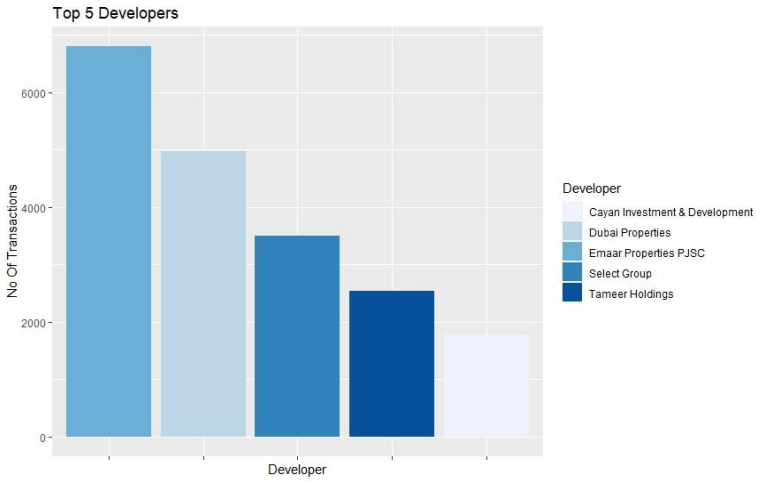
Highest number of transactions have been for one bedroom followed by 2 bedrooms



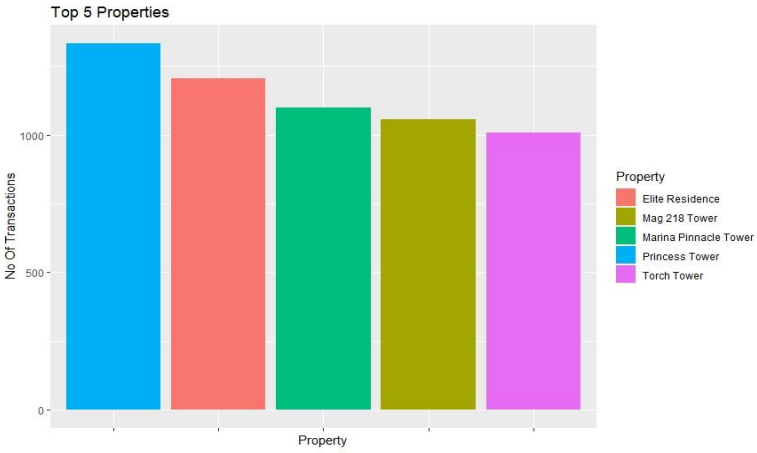
Dubai Marina has seen more number of transactions in comparison to Jumeirah Beach Residences



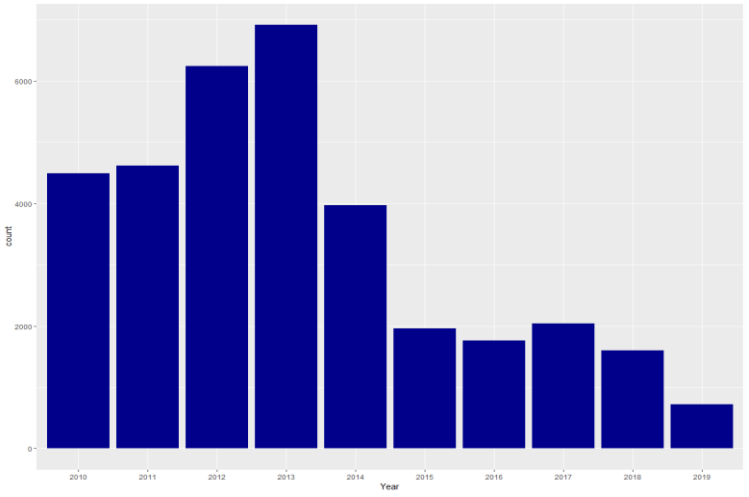
Number of transactions for the type “Apartment” accounts for more than 90% of the transactions.



Top 5 Developers in number of transactions for the area.



Top 5 properties with the highest number of transactions



The area has seen highest transactions in 2013 followed by 2012.

Correlation and Apparent Relationships

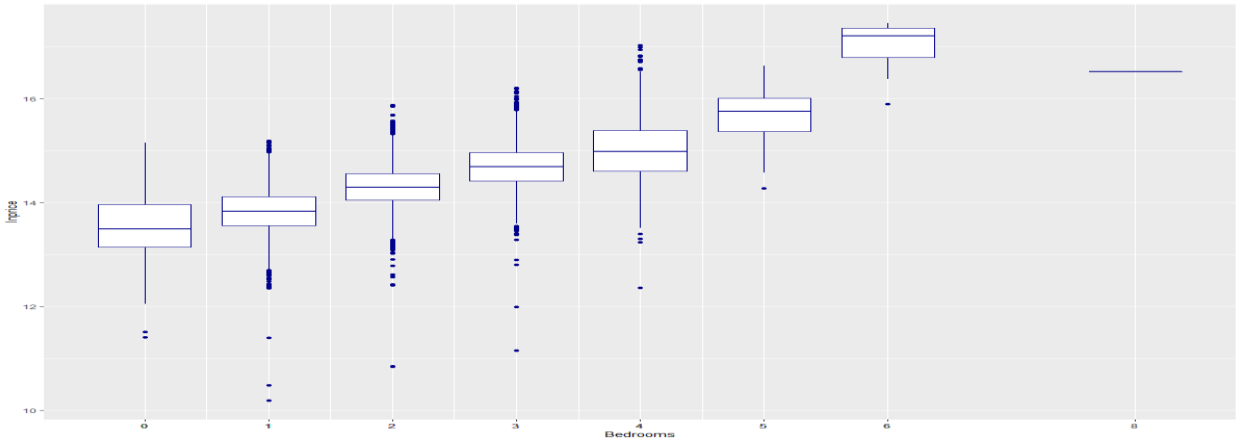
After exploring the individual features, an attempt was made to identify relationships between features in the data – in particular, between **Price** and the other features.

Categorical Relationships

An attempt was made to discern any apparent relationship between categorical feature values and price. The following box-plots show the categorical columns that seem to exhibit a relationship with the log of price:

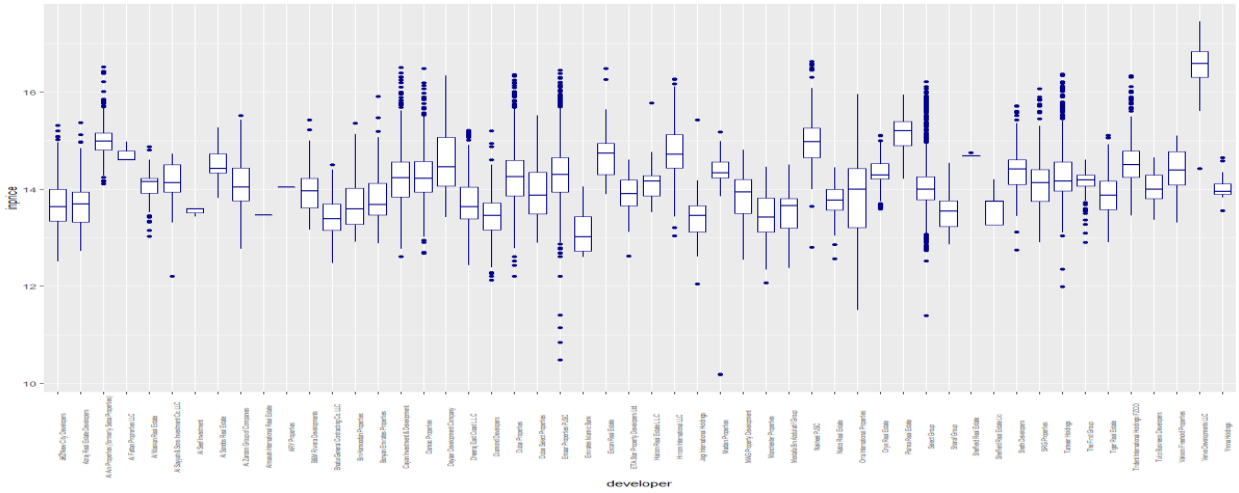
Bedrooms

The price index moves up as the number of bedrooms increase.



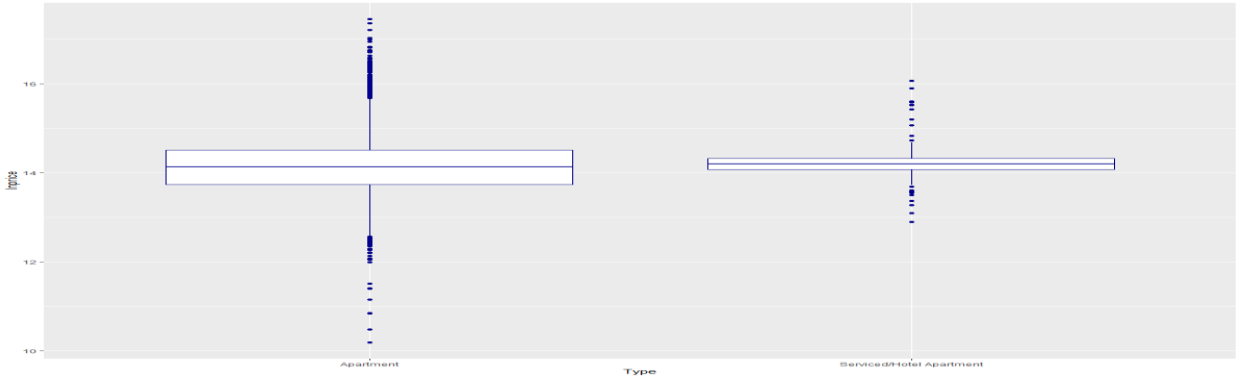
Developers

The price index varies when compared to Developers.



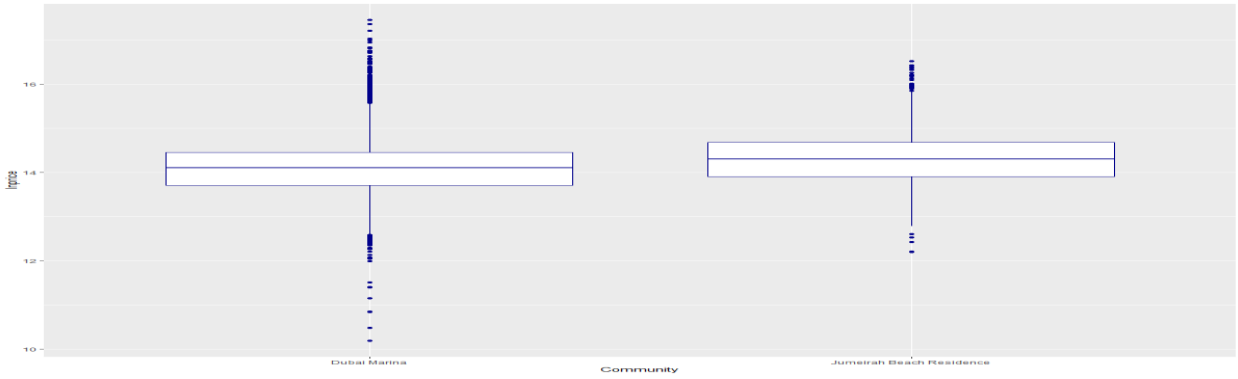
Type

Both the box plots' show no skew are normally distributed. However, the hotel apartment box plot shows a lot of outliers.



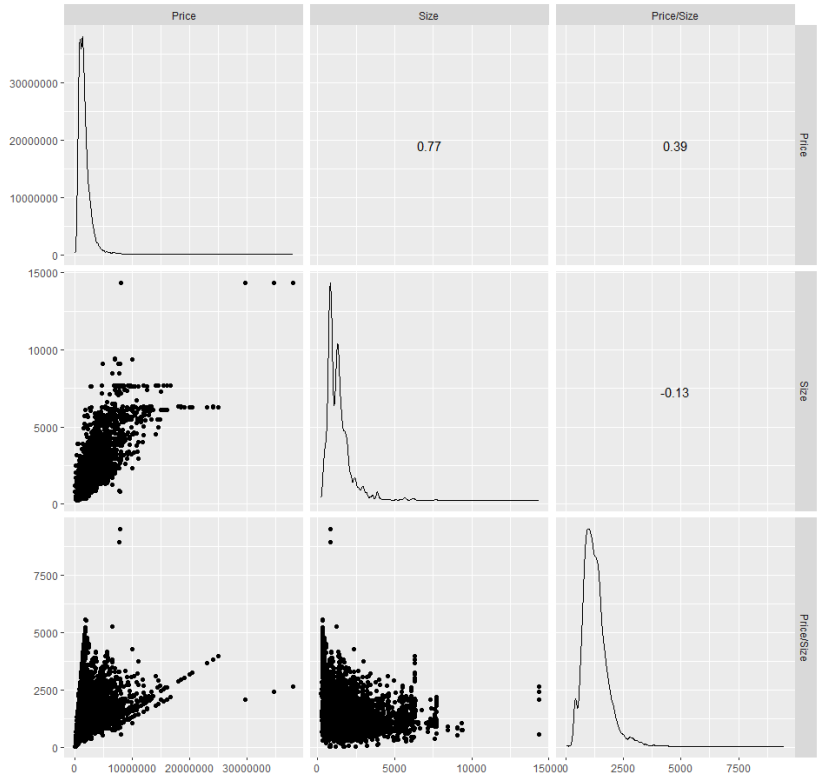
Community

Normally distributed box plots with extended outliers in Dubai Marina

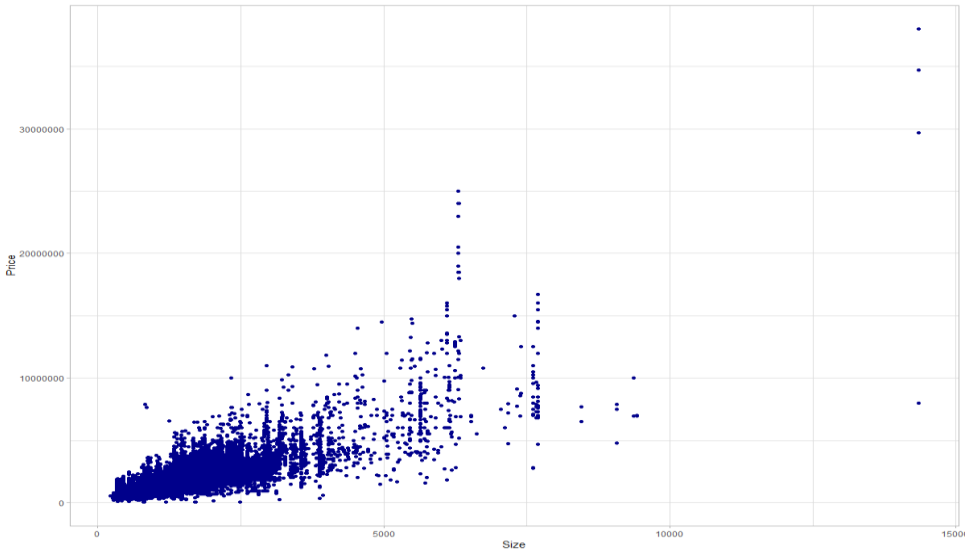


Numeric Relationships

The correlation between the numeric columns was then calculated with the following results:



We see the correlation between Size and Price is *r* = .77. Correlations range between zero (no association between variables) and 1.0 (a one-to-one association). Below is a visualization of the relationship using GGPLOT2



Significance Test

Here we conduct a T test to test the correlation for significance

P-value	< 0.00000000000000022
T	220.67
95 percent confidence interval	0.7615046, 0.7702526
Cor	0.765914

- P Value less than .05, we can say that this result would happen very rarely by chance.
- T value is 220.67 which states that that the correlation is 220.67 times larger than would typically be expected by chance. Thus, chance seems an unlikely explanation for the result.
- We see that significance test has given us a 95% confidence interval of [0.76, 0.77], meaning that we are 95% confident that the population value (ρ) is in that range.

We reject the null hypothesis and conclude that this correlation is not due to chance.

Linear Regression

After proving a strong correlation between size and price, we shall now create a linear model between Size and Price.

Price= 62719+1174(Size)

Residuals:				
Min -8907908	1Q -376929	Median -59678	3Q 305198	Max 21092092

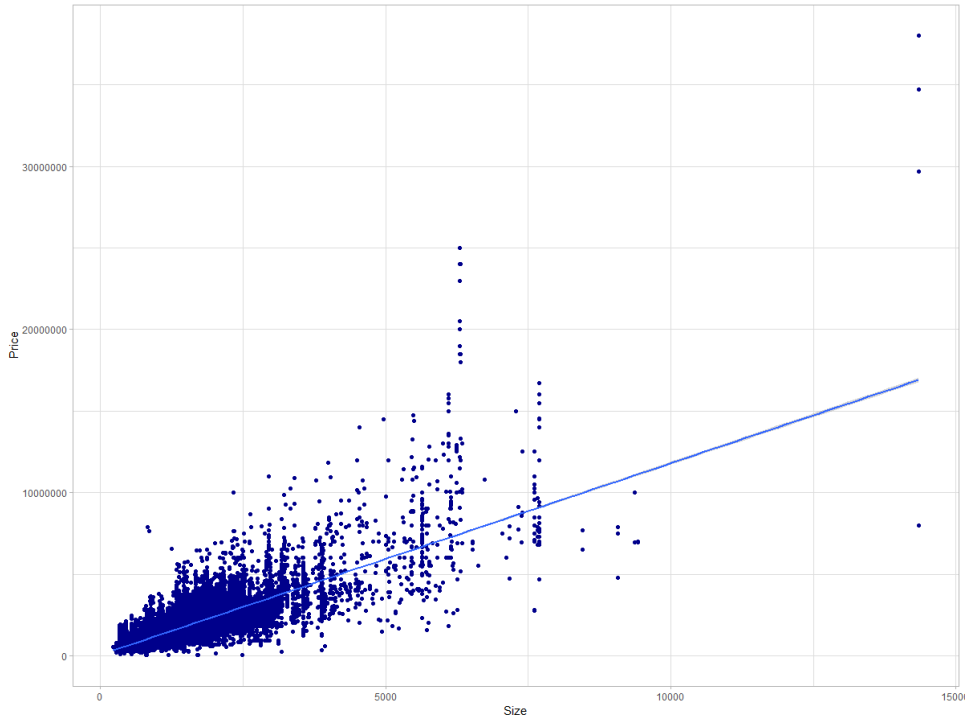
Coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	62718.72	8583.36	7.307	0.000000000000279
Size ---	1173.95	5.32	220.670	< 0.0000000000000002

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 851100 on 34314 degrees of freedom
Multiple R-squared: 0.5866, **Adjusted R-squared**: 0.5866
F-statistic: 4.87e+04 on 1 and 34314 DF, **p-value**: < 0.00000000000000022

Conclusions and predictions-

- An increase of 1 sqft would increase the price by AED 1174
- 58.66% of change in Price has been explained by Size.
- A big F statistic number concludes that there is a strong relationship between the variables.



Predictions-

Size(Sqft)	Price(AED)
1000	1,236,667
2000	2,410,616
3000	3,584,565