

Final Project

Predicting Property Prices in Dubai
by Mercy Ogbede

#Background

Dubai's real estate market is characterized by rapid growth and diverse property offerings, making it essential to understand how size influence property prices. Specifically, property size (in square feet) and bedroom sizes are significant indicators of value. Research has shown that larger properties typically command higher prices, and properties with more bedrooms are often more desirable for families (Haffner et al., 2010; Abraha & Moges 2024). However, there maybe instances where more rooms may not significantly translate to higher prices.

Motivation This research is timely, given the dynamic nature of housing market, driven by urbanization and changing demographics. Insights from this study will benefit potential buyers, real estate professionals, and policymakers by clarifying how property characteristics impact market prices.

Research Question Does size and location affect property price in Dubai?

Hypothesis Property size and location are positively correlated with property prices in Dubai

Explanation: This hypothesis posits that larger properties and those with more bedroom sizes are perceived as more desirable, potentially leading to higher prices. The location of the property also influences price. While intuitive, this study will empirically test the relationship within the context of Dubai's real estate market, aiming to understand how these factors influence property prices.

Data description

I will be working on Dubai Properties- Apartment Dataset <https://www.kaggle.com/datasets/dataregress/dubai-properties-dataset>. The dataset is part of a personal project on apartment pricing downloaded from Kaggle <https://www.kaggle.com/datasets/dataregress/dubai-properties-dataset>. The author stated that the data was scraped from the real estate portal and it is anonymized, consisting of more than 1800+ properties containing 38 features. The import variables for my research are the; prices, no_of_bedrooms and size_in_sqft.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()     masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
# loading the data
```

```
properties_data <- read.csv("properties_data.csv")
```

```
head(properties_data)
```

	id	neighborhood	latitude	longitude	price	size_in_sqft
1	5528049	Palm Jumeirah	25.11321	55.13893	2700000	1079
2	6008529	Palm Jumeirah	25.10681	55.15120	2850000	1582
3	6034542	Jumeirah Lake Towers	25.06330	55.13773	1150000	1951
4	6326063	Culture Village	25.22730	55.34176	2850000	2020
5	6356778	Palm Jumeirah	25.11427	55.13976	1729200	507
6	6356784	Palm Jumeirah	25.11427	55.13976	3119900	1015

	price_per_sqft	no_of_bedrooms	no_of_bathrooms	quality	maid_room	unfurnished
1	2502.32	1	2	Medium	False	False
2	1801.52	2	2	Medium	False	False
3	589.44	3	5	Medium	True	True
4	1410.89	2	3	Low	False	True
5	3410.65	0	1	Medium	False	False
6	3073.79	1	2	Medium	False	False

	balcony	barbecue_area	built_in_wardrobes	central_ac	childrens_play_area
1	True	True	False	True	True
2	True	False	True	True	True
3	True	False	True	False	False
4	True	False	False	False	False
5	False	False	True	True	False
6	False	False	True	True	False

	childrens_pool	concierge	covered_parking	kitchen_appliances	lobby_in_building	
1	False	True	False		True	False
2	False	False	False		False	False

3	False	False	True	False	False		
4	False	True	True	False	False		
5	False	False	True	True	False		
6	False	False	True	True	False		
maid_service networked pets_allowed private_garden private_gym							
1	False	False	True	False	False		
2	False	False	False	False	False		
3	False	False	False	False	False		
4	False	False	True	False	False		
5	False	True	False	False	False		
6	False	True	False	False	False		
private_jacuzzi private_pool security shared_gym shared_pool shared_spa study							
1	False	False	False	True	False	False	False
2	False	False	False	True	True	False	False
3	True	False	True	True	True	False	False
4	False	False	False	False	False	False	False
5	False	False	True	True	True	True	False
6	False	False	True	True	True	True	False
vastu_compliant view_of_landmark view_of_water walk_in_closet							
1	False	False	True	False			
2	False	False	True	False			
3	False	True	True	True			
4	False	False	False	False			
5	False	True	True	False			
6	False	True	True	False			

```
# Displaying the data structure
summary(properties_data)
```

id	neighborhood	latitude	longitude
Min. :5528049	Length:1905	Min. :24.87	Min. :55.07
1st Qu.:7560167	Class :character	1st Qu.:25.07	1st Qu.:55.15
Median :7631829	Mode :character	Median :25.10	Median :55.21
Mean :7573308		Mean :25.12	Mean :55.21
3rd Qu.:7670328		3rd Qu.:25.19	3rd Qu.:55.27
Max. :7706643		Max. :25.27	Max. :55.44
price	size_in_sqft	price_per_sqft	no_of_bedrooms
Min. : 220000	Min. : 294	Min. : 361.9	Min. :0.000
1st Qu.: 890000	1st Qu.: 840	1st Qu.: 870.9	1st Qu.:1.000
Median : 1400000	Median :1271	Median :1169.6	Median :2.000
Mean : 2085830	Mean :1417	Mean :1327.2	Mean :1.793
3rd Qu.: 2200000	3rd Qu.:1703	3rd Qu.:1622.5	3rd Qu.:2.000

Max. :35000000	Max. :9576	Max. :4805.9	Max. :5.000
no_of_bathrooms	quality	maid_room	unfurnished
Min. :1.000	Length:1905	Length:1905	Length:1905
1st Qu.:2.000	Class :character	Class :character	Class :character
Median :2.000	Mode :character	Mode :character	Mode :character
Mean :2.513			
3rd Qu.:3.000			
Max. :6.000			
balcony	barbecue_area	built_in_wardrobes	central_ac
Length:1905	Length:1905	Length:1905	Length:1905
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
childrens_play_area	childrens_pool	concierge	covered_parking
Length:1905	Length:1905	Length:1905	Length:1905
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
kitchen_appliances	lobby_in_building	maid_service	networked
Length:1905	Length:1905	Length:1905	Length:1905
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
pets_allowed	private_garden	private_gym	private_jacuzzi
Length:1905	Length:1905	Length:1905	Length:1905
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
private_pool	security	shared_gym	shared_pool
Length:1905	Length:1905	Length:1905	Length:1905
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

shared_spa	study	vastu_compliant	view_of_landmark
Length:1905	Length:1905	Length:1905	Length:1905
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

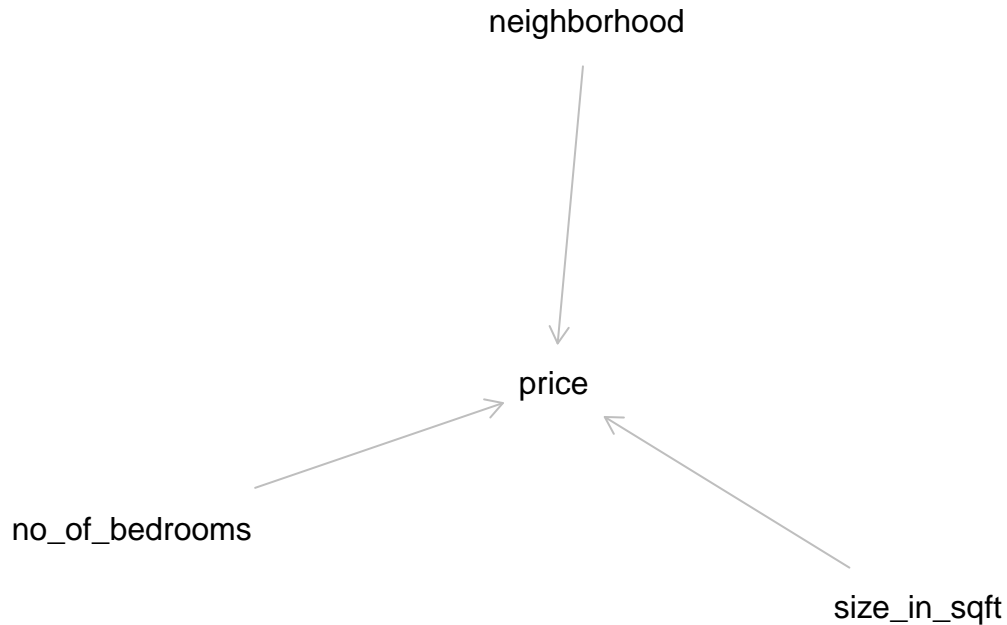
view_of_water	walk_in_closet
Length:1905	Length:1905
Class :character	Class :character
Mode :character	Mode :character

A summary of the data shows that the minimum price of a house is 220000 AED, the median price is 1400000 AED and the maximum is 35000000 AED. The minimum property size in square feet is 294, median value is 1271 and the maximum is 9576. Median value for number of bedrooms is 2 with the maximum value at 5.

*** Creating a dag to visualize the data

```
library(dagitty)
dag <- dagitty("dag {
  size_in_sqft -> price
  no_of_bedrooms -> price
  neighborhood -> price
}")
plot(dag)
```

Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to



#Applying a regression model to visualize relationship between variables

```
# Fitting a linear regression model to understand the relationship between price and property
price_model_1 <- lm(price ~ size_in_sqft, data = properties_data)
summary(price_model_1)
```

Call:

```
lm(formula = price ~ size_in_sqft, data = properties_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-10246746	-712540	74260	696250	17247970

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.658e+06	7.378e+04	-22.48	<2e-16 ***
size_in_sqft	2.642e+03	4.407e+01	59.95	<2e-16 ***

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

Residual standard error: 1714000 on 1903 degrees of freedom

Multiple R-squared: 0.6538, Adjusted R-squared: 0.6536

F-statistic: 3594 on 1 and 1903 DF, p-value: < 2.2e-16

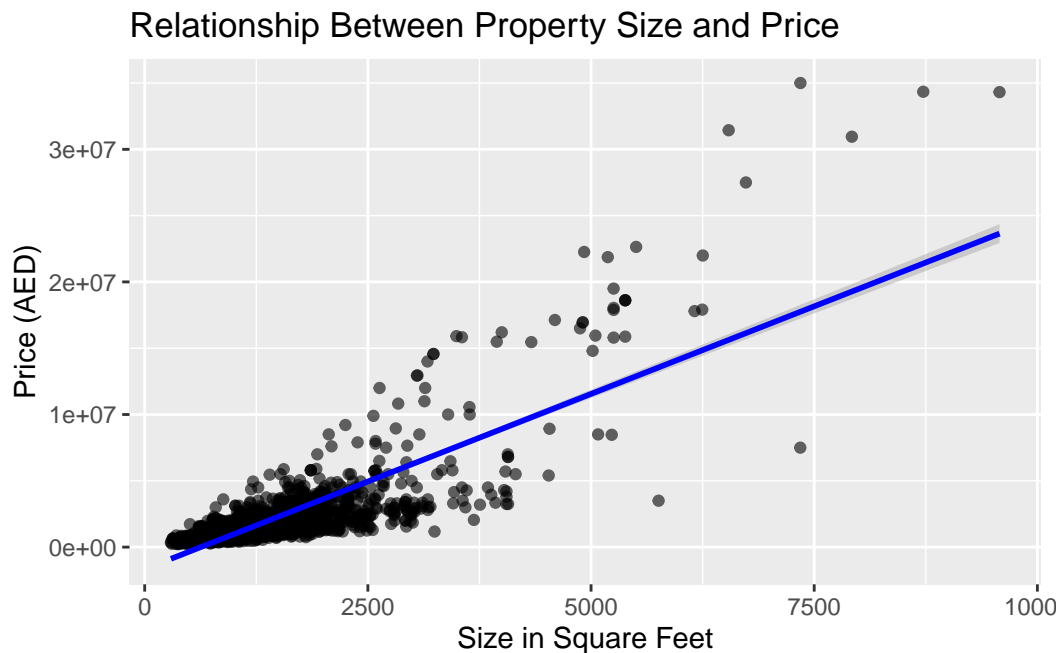
The model shows a strong positive relationship between property size and price, with prices increasing by 2,642 AED for every additional square foot. The relationship is significant (p-value < 2e-16). With an R-squared at 0.6538, 65.4% of the variation in property prices is explained by size_in_sqft. This indicates a strong relationship, meaning property size is a key driver of property prices. Additionally, the large F-statistic and extremely small p-value (< 2.2e-16) confirm that the overall model is significant.

#Below is the model visualization for the relationship between price and size regardless of the bedroom number

```
library(ggplot2)

# Visualization for Model 1: price ~ size_in_sqft
ggplot(properties_data, aes(x = size_in_sqft, y = price)) +
  geom_point(alpha = 0.6, color = "black") +
  geom_smooth(method = "lm", col = "blue") +
  labs(title = "Relationship Between Property Size and Price",
       x = "Size in Square Feet", y = "Price (AED)")
```

`geom_smooth()` using formula = 'y ~ x'



#Next, finding the relationship between price and approximate bedroom size

```
# Creating a new column for size_per_bedroom
properties_data$size_per_bedroom <- ifelse(properties_data$no_of_bedrooms > 0,
                                           properties_data$size_in_sqft /properties_data$no_of_bedrooms,NA)

# Selecting only relevant columns: price, size_in_sqft, no_of_bedrooms, and size_per_bedroom
new_properties_data <- properties_data[, c("price", "size_in_sqft", "no_of_bedrooms", "size_per_bedroom")]

# Viewing the first few rows of the updated dataset
head(new_properties_data)
```

	price	size_in_sqft	no_of_bedrooms	size_per_bedroom
1	2700000	1079	1	1079.0000
2	2850000	1582	2	791.0000
3	1150000	1951	3	650.3333
4	2850000	2020	2	1010.0000
5	1729200	507	0	NA
6	3119900	1015	1	1015.0000

```
#Next, running the model
price_model_2 <- lm(price ~ size_per_bedroom, data = new_properties_data)
summary(price_model_2)
```

Call:

```
lm(formula = price ~ size_per_bedroom, data = new_properties_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-6937984	-1237256	-272734	605439	26969472

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2159612.5	219751.3	-9.828	<2e-16 ***
size_per_bedroom	5548.7	268.2	20.688	<2e-16 ***

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

Residual standard error: 2679000 on 1779 degrees of freedom
(124 observations deleted due to missingness)

Multiple R-squared: 0.1939, Adjusted R-squared: 0.1935
F-statistic: 428 on 1 and 1779 DF, p-value: < 2.2e-16

From the model, for every 1 square foot increase in average room size, the property price increases by approximately 5,549 AED. This relationship is statistically significant (p-value < 2e-16). With an R-squared at 0.1939, only 19.4% of the variation in property prices is explained by size_per_bedroom. This indicates that size_per_bedroom has a moderate influence but may not fully explain price variability. Furthermore, with a residual standard error at 2,679,000, the average prediction error remains large, suggesting high variability in property prices that may not be accounted for solely by size_per_bedroom. However, with an F-statistic at 428 the model is significant (p-value < 2.2e-16), which still proves that size_per_bedroom is a valid predictor of property price.

Next, visualizing the model to draw the regression line

```
library(ggplot2)

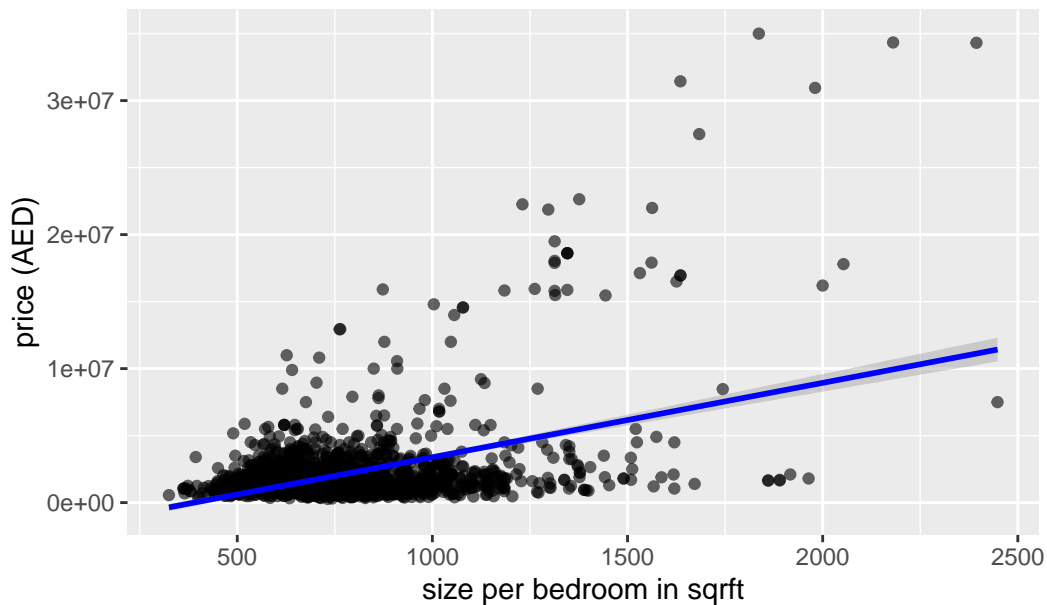
# Visualization for model 3: price ~ size_per_bedroom
ggplot(new_properties_data, aes(x = size_per_bedroom, y = price)) +
  geom_point(alpha = 0.6, color = "black") +
  geom_smooth(method = "lm", col = "blue") +
  labs(title = "relationship between approximate room size and price",
       x = "size per bedroom in sqrft", y = "price (AED)")
```

``geom_smooth()`` using formula = 'y ~ x'

Warning: Removed 124 rows containing non-finite outside the scale range
(``stat_smooth()``).

Warning: Removed 124 rows containing missing values or values outside the scale range
(``geom_point()``).

relationship between approximate room size and price



#Next, getting a multiple regression using size in sqft, number of bedrooms and neighborhood/location of property to predict price.

```
price_model_mult <- lm(price ~ size_in_sqft + no_of_bedrooms + neighborhood, data = properties_data)
summary(price_model_mult)
```

Call:

```
lm(formula = price ~ size_in_sqft + no_of_bedrooms + neighborhood,
    data = properties_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-13992111	-593165	49315	559713	14287713

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-4.373e+06	4.942e+05	-8.849
size_in_sqft	3.479e+03	6.817e+01	51.035
no_of_bedrooms	-9.862e+05	6.273e+04	-15.721
neighborhoodA1 Barsha	3.558e+06	9.047e+05	3.932
neighborhoodA1 Furjan	3.130e+06	5.839e+05	5.361
neighborhoodA1 Kifaf	3.084e+06	7.532e+05	4.094

neighborhoodAl Quoz	2.645e+06	1.598e+06	1.655
neighborhoodAl Sufouh	3.181e+06	7.272e+05	4.375
neighborhoodArjan	3.266e+06	6.706e+05	4.870
neighborhoodBarsha Heights (Tecom)	2.874e+06	9.041e+05	3.179
neighborhoodBluewaters	3.994e+06	7.254e+05	5.507
neighborhoodBusiness Bay	3.275e+06	5.097e+05	6.425
neighborhoodCity Walk	2.899e+06	6.329e+05	4.580
neighborhoodCulture Village	2.434e+06	6.661e+05	3.655
neighborhoodDAMAC Hills	3.132e+06	5.901e+05	5.308
neighborhoodDIFC	2.625e+06	5.565e+05	4.717
neighborhoodDiscovery Gardens	2.674e+06	7.261e+05	3.683
neighborhoodDowntown Dubai	4.088e+06	4.954e+05	8.252
neighborhoodDubai Creek Harbour (The Lagoons)	3.797e+06	5.493e+05	6.913
neighborhoodDubai Festival City	1.107e+05	8.356e+05	0.132
neighborhoodDubai Harbour	4.335e+06	5.630e+05	7.699
neighborhoodDubai Healthcare City	4.005e+06	1.183e+06	3.386
neighborhoodDubai Hills Estate	3.428e+06	5.335e+05	6.425
neighborhoodDubai Land	2.763e+06	6.712e+05	4.117
neighborhoodDubai Marina	2.977e+06	4.960e+05	6.002
neighborhoodDubai Production City (IMPZ)	3.097e+06	7.273e+05	4.258
neighborhoodDubai Residence Complex	3.042e+06	1.599e+06	1.903
neighborhoodDubai Silicon Oasis	2.633e+06	6.563e+05	4.011
neighborhoodDubai South (Dubai World Central)	1.928e+06	8.360e+05	2.307
neighborhoodDubai Sports City	2.985e+06	5.758e+05	5.185
neighborhoodFalcon City of Wonders	3.864e+06	1.006e+06	3.841
neighborhoodGreen Community	1.547e+04	1.180e+06	0.013
neighborhoodGreens	2.859e+06	5.627e+05	5.080
neighborhoodInternational City	3.159e+06	7.908e+05	3.994
neighborhoodJebel Ali	3.409e+06	1.182e+06	2.883
neighborhoodJumeirah	4.865e+06	5.436e+05	8.950
neighborhoodJumeirah Beach Residence	3.433e+06	5.078e+05	6.760
neighborhoodJumeirah Golf Estates	3.007e+06	1.182e+06	2.544
neighborhoodJumeirah Lake Towers	2.559e+06	5.198e+05	4.924
neighborhoodJumeirah Village Circle	3.026e+06	5.005e+05	6.045
neighborhoodJumeirah Village Triangle	3.358e+06	1.005e+06	3.340
neighborhoodMeydan	2.741e+06	6.118e+05	4.480
neighborhoodMina Rashid	6.448e+05	1.596e+06	0.404
neighborhoodMirdif	3.108e+06	7.925e+05	3.922
neighborhoodMohammed Bin Rashid City	4.035e+06	5.613e+05	7.190
neighborhoodMotor City	1.851e+06	5.673e+05	3.263
neighborhoodMudon	3.583e+06	9.068e+05	3.951
neighborhoodOld Town	3.297e+06	6.112e+05	5.395
neighborhoodPalm Jumeirah	3.474e+06	4.961e+05	7.002

neighborhoodRemraam	2.728e+06	7.898e+05	3.454
neighborhoodThe Hills	3.024e+06	6.696e+05	4.516
neighborhoodThe Views	2.976e+06	5.366e+05	5.547
neighborhoodTown Square	3.409e+06	5.732e+05	5.948
neighborhoodUmm Suqeim	3.704e+06	6.456e+05	5.737
neighborhoodwasl gate	4.496e+06	1.601e+06	2.808
neighborhoodWorld Trade Center	7.921e+05	7.237e+05	1.094
	Pr(> t)		
(Intercept)	< 2e-16	***	
size_in_sqft	< 2e-16	***	
no_of_bedrooms	< 2e-16	***	
neighborhoodAl Barsha	8.72e-05	***	
neighborhoodAl Furjan	9.32e-08	***	
neighborhoodAl Kifaf	4.42e-05	***	
neighborhoodAl Quoz	0.098059	.	
neighborhoodAl Sufouh	1.28e-05	***	
neighborhoodArjan	1.21e-06	***	
neighborhoodBarsha Heights (Tecom)	0.001500	**	
neighborhoodBluewaters	4.17e-08	***	
neighborhoodBusiness Bay	1.67e-10	***	
neighborhoodCity Walk	4.96e-06	***	
neighborhoodCulture Village	0.000265	***	
neighborhoodDAMAC Hills	1.24e-07	***	
neighborhoodDIFC	2.57e-06	***	
neighborhoodDiscovery Gardens	0.000237	***	
neighborhoodDowntown Dubai	2.92e-16	***	
neighborhoodDubai Creek Harbour (The Lagoons)	6.52e-12	***	
neighborhoodDubai Festival City	0.894606		
neighborhoodDubai Harbour	2.22e-14	***	
neighborhoodDubai Healthcare City	0.000724	***	
neighborhoodDubai Hills Estate	1.67e-10	***	
neighborhoodDubai Land	4.01e-05	***	
neighborhoodDubai Marina	2.33e-09	***	
neighborhoodDubai Production City (IMPZ)	2.17e-05	***	
neighborhoodDubai Residence Complex	0.057163	.	
neighborhoodDubai Silicon Oasis	6.28e-05	***	
neighborhoodDubai South (Dubai World Central)	0.021183	*	
neighborhoodDubai Sports City	2.39e-07	***	
neighborhoodFalcon City of Wonders	0.000127	***	
neighborhoodGreen Community	0.989542		
neighborhoodGreens	4.15e-07	***	
neighborhoodInternational City	6.74e-05	***	
neighborhoodJebel Ali	0.003982	**	

neighborhoodJumeirah	< 2e-16 ***
neighborhoodJumeirah Beach Residence	1.84e-11 ***
neighborhoodJumeirah Golf Estates	0.011028 *
neighborhoodJumeirah Lake Towers	9.22e-07 ***
neighborhoodJumeirah Village Circle	1.80e-09 ***
neighborhoodJumeirah Village Triangle	0.000854 ***
neighborhoodMeydan	7.92e-06 ***
neighborhoodMina Rashid	0.686332
neighborhoodMirdif	9.10e-05 ***
neighborhoodMohammed Bin Rashid City	9.37e-13 ***
neighborhoodMotor City	0.001121 **
neighborhoodMudon	8.07e-05 ***
neighborhoodOld Town	7.75e-08 ***
neighborhoodPalm Jumeirah	3.51e-12 ***
neighborhoodRemraam	0.000566 ***
neighborhoodThe Hills	6.70e-06 ***
neighborhoodThe Views	3.33e-08 ***
neighborhoodTown Square	3.24e-09 ***
neighborhoodUmm Suqeim	1.12e-08 ***
neighborhoodwasl gate	0.005041 **
neighborhoodWorld Trade Center	0.273896

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

Residual standard error: 1522000 on 1849 degrees of freedom
Multiple R-squared: 0.735, Adjusted R-squared: 0.7272
F-statistic: 93.26 on 55 and 1849 DF, p-value: < 2.2e-16

Key Insights From the Model Size and Bedrooms:

size_in_sqft remains a strong positive predictor of price (Estimate = 3,479, p-value < 2e-16).

No_of_bedrooms unexpectedly has a negative coefficient (-986,200 AED), suggesting that properties with more bedrooms may have lower prices when size is held constant. This negative coefficient for no_of_bedrooms in the regression model still makes sense when considering that property size is held constant. For instance, if two properties have the same total size but differ in the number of bedrooms, the property with more bedrooms will likely have smaller individual rooms. Smaller rooms can reduce the perceived value of the property because buyers may prioritize spacious open layouts over a higher count of cramped rooms. Imagine property A is 1,200 sqft with 2 bedrooms (here bedrooms are larger and more spacious), and property B is 1,200 sqft with 4 bedrooms (here bedrooms are smaller, potentially making the property feel less luxurious or functional). This trade-off in bedroom size can lead to lower

prices for properties with more bedrooms but the same total size, and shows buyers' preference for quality of space over quantity.

Neighborhoods:

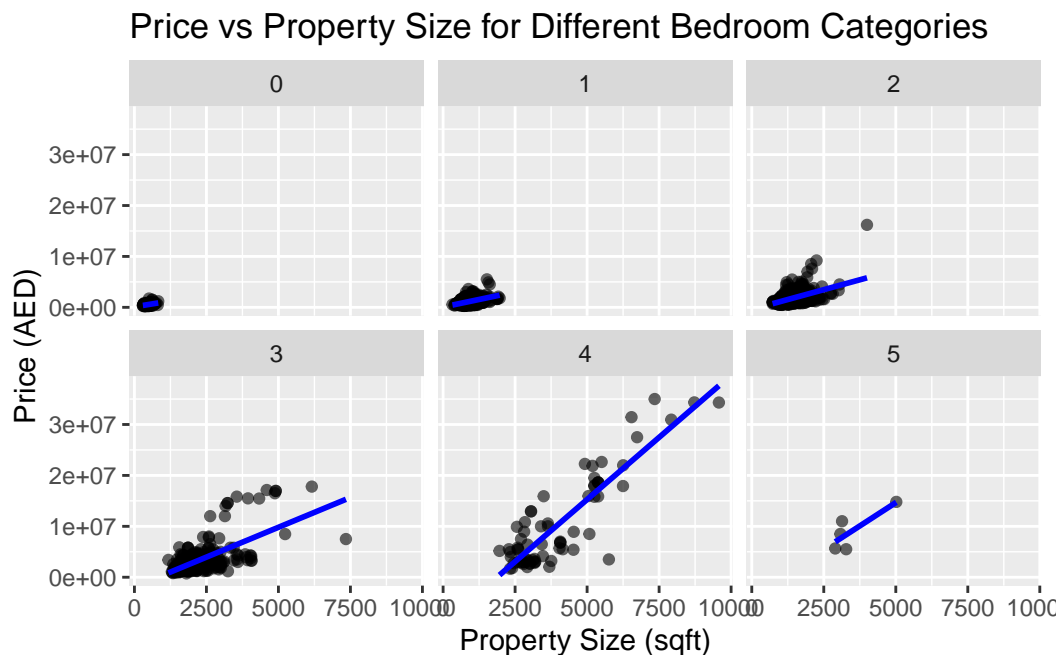
Many neighborhoods (e.g., Downtown Dubai, Palm Jumeirah, Jumeirah) have significantly positive coefficients (denoted by ***), confirming their higher price influence. Some neighborhoods such as Green Community and Mina Rashid show non-significant effects.

Model Fit: The Adjusted R-squared is 0.727, meaning the model explains ~72.7% of the variability in property prices. The F-statistic is highly significant ($p < 2.2e-16$), confirming the overall model's validity.

#Next, assuming the size is not held constant, and each number of bedroom is held constant, we can run a model to understand the relationship between price and property size for every bedroom number.

```
ggplot(properties_data, aes(x = size_in_sqft, y = price)) +  
  geom_point(alpha = 0.6) +  
  geom_smooth(method = "lm", se = FALSE, color = "blue") +  
  facet_wrap(~ no_of_bedrooms) +  
  labs(title = "Price vs Property Size for Different Bedroom Categories",  
       x = "Property Size (sqft)", y = "Price (AED)")
```

`geom_smooth()` using formula = 'y ~ x'



“Price vs property size for each bedroom category”

The plot reveals how the relationship between property size and price shifts across bedroom categories.

For properties with 0 bedrooms, the trend is flat and prices sit firmly in the lower range. This likely reflects small studios or unconventional spaces where size doesn't strongly dictate value.

Moving up to 1- and 2-bedroom properties, one could see a slight upward trend, but the data remains tightly clustered. Here, price seems less sensitive to property size, suggesting that buyers prioritize other features, like location or amenities, over square footage.

Things start to change with 3-bedroom properties. A clear upward trend emerges, where increasing property size drives prices higher. At the same time, the spread of data points widens, hinting at more variability—perhaps due to differences in location, quality, or luxury features.

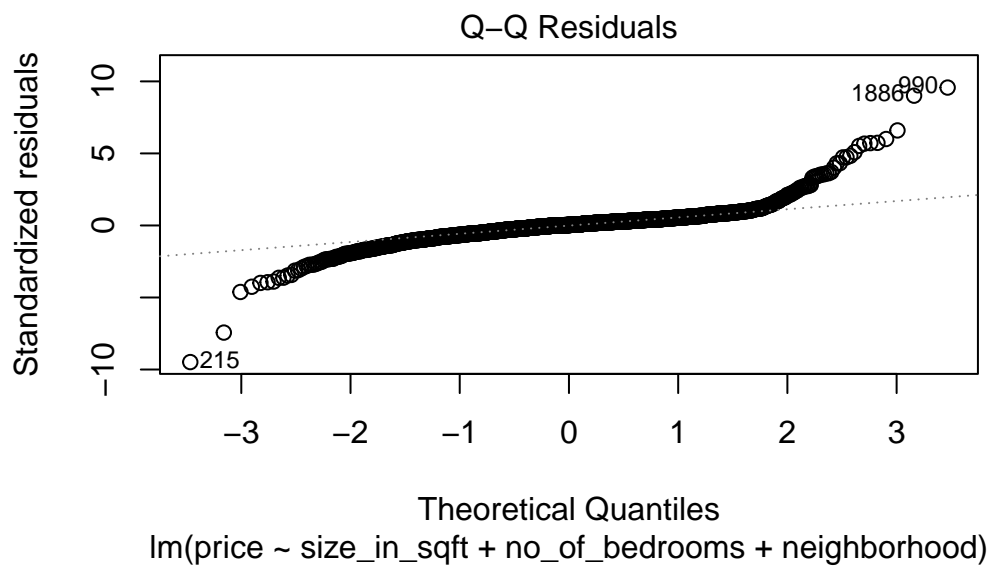
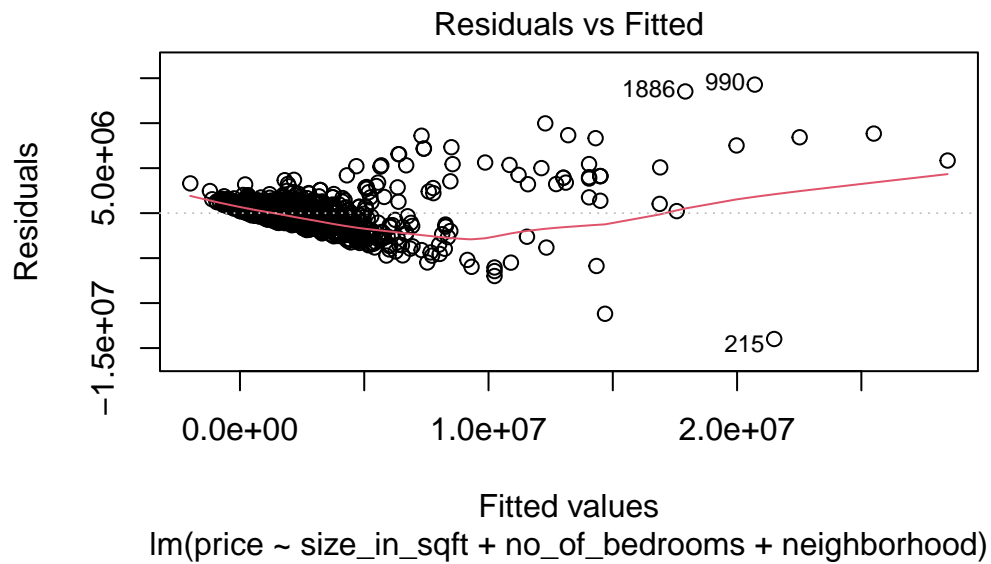
The most striking relationship is seen for 4-bedroom properties, where the regression line takes a sharp climb. Larger homes with four bedrooms clearly command higher prices, reinforcing their appeal to families or buyers seeking spacious, premium living.

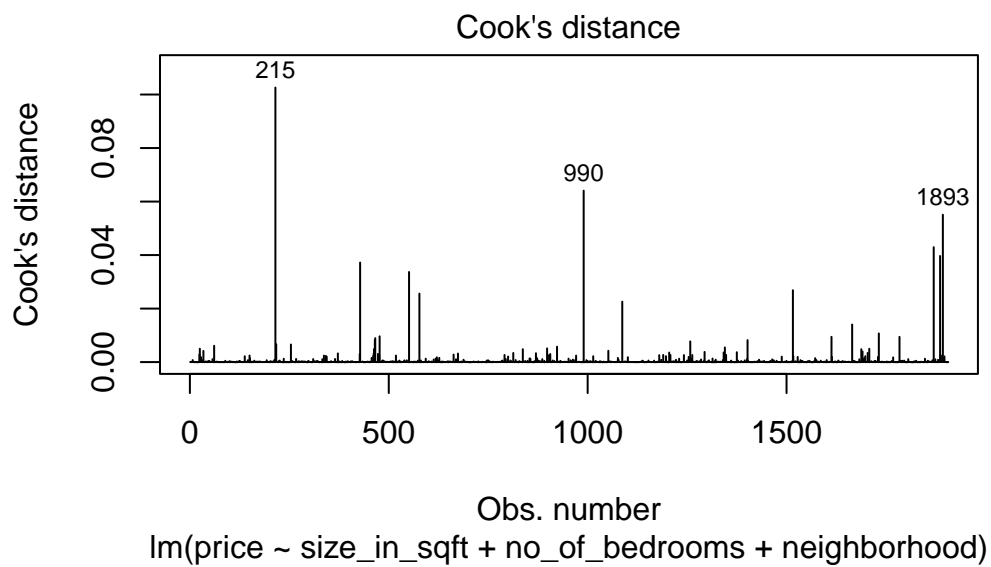
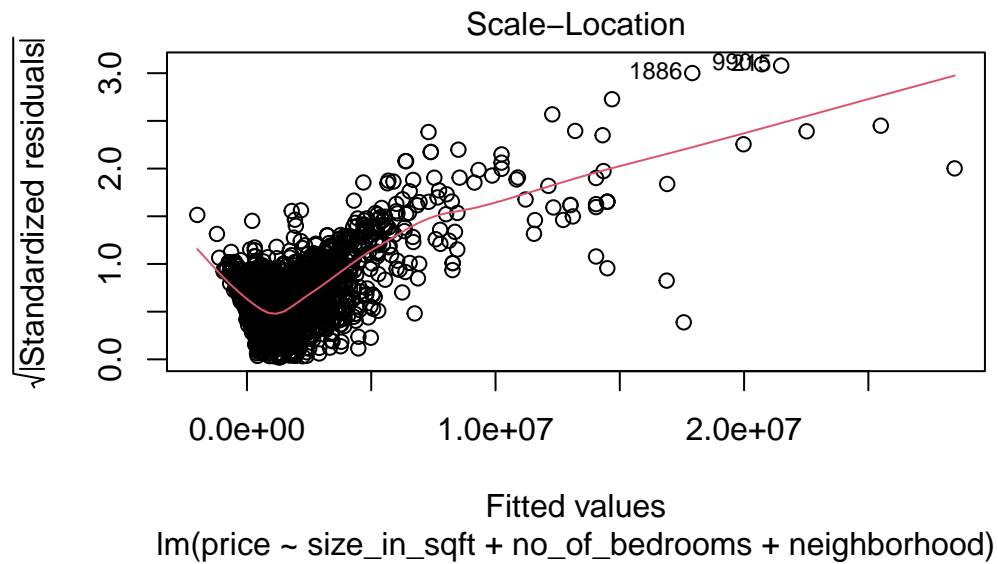
For 5-bedroom properties, while the positive trend continues, the data is sparse. It's clear that size matters, but with limited observations, it's hard to draw confident conclusions about this segment.

I will have to check for linearity and homoscedasticity and normality of residuals in the data.

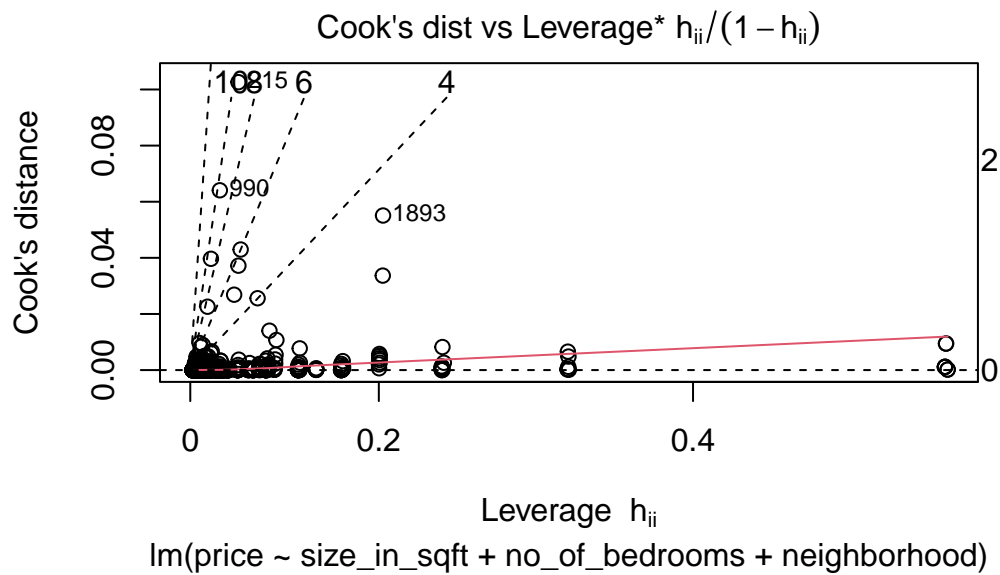
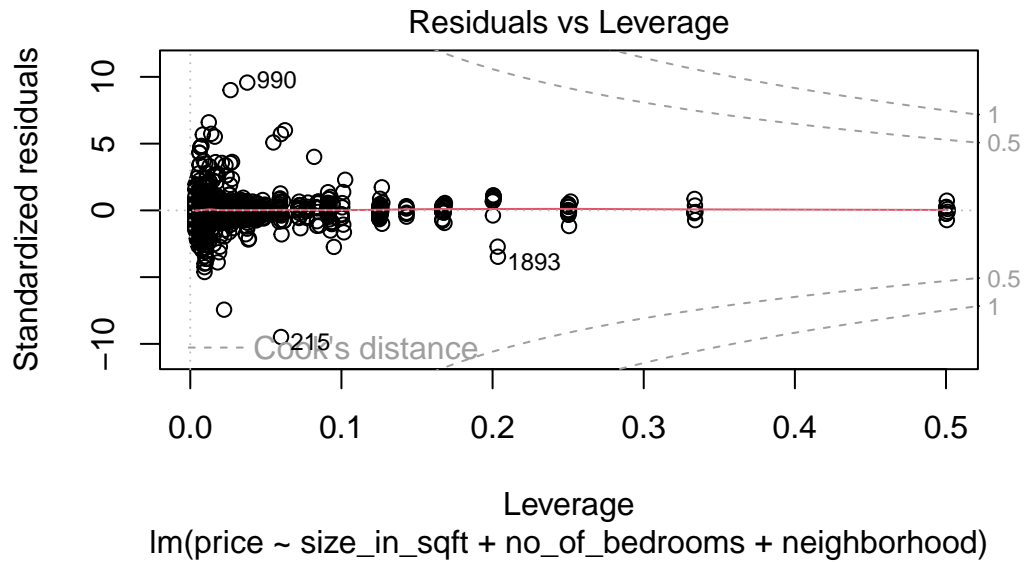
```
plot(price_model_mult, which = 1:6)
```

```
Warning: not plotting observations with leverage one:  
338, 1733, 1849, 1898
```





Warning: not plotting observations with leverage one:
338, 1733, 1849, 1898



Influential Observations: Certain data points, particularly 215, 990, and 1893, stand out as influential. These points have a combination of: High leverage (unusual predictor values). Large Cook's distance, meaning they significantly impact the regression results. This doesn't imply errors, but these points could represent extreme properties, such as luxury homes or

anomalies.

Uneven Residual Variance (Heteroscedasticity) The Scale-Location plot shows an upward spread of residuals as the predicted prices increase. This indicates that price variability grows for higher property values, common in real estate data.

Non-Normal Residuals The Q-Q plot highlights deviations from normality, especially at the extremes, where points like 215, 990, and 1886 stand out. This suggests that extreme values in price are impacting the residual distribution.

Non-Linearity The Residuals vs Fitted plot reveals a curved pattern, suggesting that the relationship between size_in_sqft and price is not perfectly linear. This could be due to diminishing returns as property size increases.

I will identify as well as remove these influential points in the data.

```
# Identifying influential points
influential_points <- cooks.distance(lm(price ~ size_in_sqft + no_of_bedrooms + neighborhood
which(influential_points > (4 / nrow(properties_data)))
```

```
24 25 34 61 138 150 215 217 254 337 342 372 427 428 461 462
24 25 34 61 138 150 215 217 254 337 342 372 427 428 461 462
463 464 465 466 473 476 477 518 551 577 663 674 791 813 837 870
463 464 465 466 473 476 477 518 551 577 663 674 791 813 837 870
898 903 906 923 971 990 1014 1052 1087 1180 1190 1197 1205 1208 1242 1258
898 903 906 923 971 990 1014 1052 1087 1180 1190 1197 1205 1208 1242 1258
1263 1294 1342 1345 1348 1375 1402 1516 1517 1613 1665 1688 1691 1698 1704 1708
1263 1294 1342 1345 1348 1375 1402 1516 1517 1613 1665 1688 1691 1698 1704 1708
1732 1784 1870 1886 1888 1893 1897
1732 1784 1870 1886 1888 1893 1897
```

```
# Removing influential points and refitting the model
properties_clean <- properties_data[-c(215, 990, 1893), ]
model_clean <- lm(price ~ size_in_sqft + no_of_bedrooms + neighborhood, data = properties_clean)
summary(model_clean)
```

Call:

```
lm(formula = price ~ size_in_sqft + no_of_bedrooms + neighborhood,
    data = properties_clean)
```

Residuals:

Min	1Q	Median	3Q	Max
-11307369	-602591	45005	544354	13477902

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-4.427e+06	4.699e+05	-9.423
size_in_sqft	3.519e+03	6.744e+01	52.177
no_of_bedrooms	-1.012e+06	6.058e+04	-16.706
neighborhoodAl Barsha	3.606e+06	8.589e+05	4.199
neighborhoodAl Furjan	3.186e+06	5.548e+05	5.742
neighborhoodAl Kifaf	3.125e+06	7.151e+05	4.370
neighborhoodAl Quoz	2.676e+06	1.517e+06	1.764
neighborhoodAl Sufouh	3.233e+06	6.905e+05	4.682
neighborhoodArjan	3.317e+06	6.370e+05	5.207
neighborhoodBarsha Heights (Tecom)	2.923e+06	8.583e+05	3.406
neighborhoodBluewaters	4.037e+06	6.887e+05	5.862
neighborhoodBusiness Bay	3.461e+06	4.848e+05	7.140
neighborhoodCity Walk	2.933e+06	6.008e+05	4.882
neighborhoodCulture Village	2.456e+06	6.322e+05	3.885
neighborhoodDAMAC Hills	3.175e+06	5.605e+05	5.665
neighborhoodDIFC	2.660e+06	5.284e+05	5.034
neighborhoodDiscovery Gardens	2.715e+06	6.894e+05	3.939
neighborhoodDowntown Dubai	4.134e+06	4.707e+05	8.784
neighborhoodDubai Creek Harbour (The Lagoons)	3.853e+06	5.220e+05	7.382
neighborhoodDubai Festival City	1.317e+05	7.931e+05	0.166
neighborhoodDubai Harbour	4.388e+06	5.350e+05	8.203
neighborhoodDubai Healthcare City	4.057e+06	1.123e+06	3.613
neighborhoodDubai Hills Estate	3.484e+06	5.070e+05	6.872
neighborhoodDubai Land	2.818e+06	6.375e+05	4.420
neighborhoodDubai Marina	3.024e+06	4.712e+05	6.417
neighborhoodDubai Production City (IMPZ)	3.149e+06	6.907e+05	4.559
neighborhoodDubai Residence Complex	3.079e+06	1.517e+06	2.029
neighborhoodDubai Silicon Oasis	2.678e+06	6.233e+05	4.297
neighborhoodDubai South (Dubai World Central)	3.150e+06	8.585e+05	3.670
neighborhoodDubai Sports City	3.035e+06	5.470e+05	5.549
neighborhoodFalcon City of Wonders	3.910e+06	9.551e+05	4.094
neighborhoodGreen Community	4.165e+04	1.120e+06	0.037
neighborhoodGreens	2.911e+06	5.346e+05	5.445
neighborhoodInternational City	3.209e+06	7.509e+05	4.274
neighborhoodJebel Ali	3.461e+06	1.122e+06	3.084
neighborhoodJumeirah	4.905e+06	5.163e+05	9.501

neighborhoodJumeirah Beach Residence	3.478e+06	4.824e+05	7.210
neighborhoodJumeirah Golf Estates	3.058e+06	1.122e+06	2.726
neighborhoodJumeirah Lake Towers	2.603e+06	4.937e+05	5.272
neighborhoodJumeirah Village Circle	3.076e+06	4.756e+05	6.467
neighborhoodJumeirah Village Triangle	3.407e+06	9.544e+05	3.570
neighborhoodMeydan	2.789e+06	5.811e+05	4.799
neighborhoodMina Rashid	6.684e+05	1.515e+06	0.441
neighborhoodMirdif	3.166e+06	7.526e+05	4.206
neighborhoodMohammed Bin Rashid City	4.090e+06	5.333e+05	7.670
neighborhoodMotor City	1.891e+06	5.387e+05	3.510
neighborhoodMudon	3.645e+06	8.611e+05	4.233
neighborhoodOld Town	3.342e+06	5.804e+05	5.758
neighborhoodPalm Jumeirah	3.417e+06	4.712e+05	7.253
neighborhoodRemraam	2.770e+06	7.499e+05	3.695
neighborhoodThe Hills	3.069e+06	6.358e+05	4.826
neighborhoodThe Views	3.025e+06	5.097e+05	5.935
neighborhoodTown Square	3.472e+06	5.448e+05	6.372
neighborhoodUmm Suqeim	3.754e+06	6.132e+05	6.122
neighborhoodwasl gate	4.574e+06	1.520e+06	3.008
neighborhoodWorld Trade Center	8.134e+05	6.869e+05	1.184

	Pr(> t)
(Intercept)	< 2e-16 ***
size_in_sqft	< 2e-16 ***
no_of_bedrooms	< 2e-16 ***
neighborhoodAl Barsha	2.81e-05 ***
neighborhoodAl Furjan	1.09e-08 ***
neighborhoodAl Kifaf	1.31e-05 ***
neighborhoodAl Quoz	0.077856 .
neighborhoodAl Sufouh	3.05e-06 ***
neighborhoodArjan	2.13e-07 ***
neighborhoodBarsha Heights (Tecom)	0.000674 ***
neighborhoodBluewaters	5.41e-09 ***
neighborhoodBusiness Bay	1.34e-12 ***
neighborhoodCity Walk	1.14e-06 ***
neighborhoodCulture Village	0.000106 ***
neighborhoodDAMAC Hills	1.70e-08 ***
neighborhoodDIFC	5.28e-07 ***
neighborhoodDiscovery Gardens	8.50e-05 ***
neighborhoodDowntown Dubai	< 2e-16 ***
neighborhoodDubai Creek Harbour (The Lagoons)	2.34e-13 ***
neighborhoodDubai Festival City	0.868075
neighborhoodDubai Harbour	4.34e-16 ***
neighborhoodDubai Healthcare City	0.000311 ***

neighborhoodDubai Hills Estate	8.62e-12 ***
neighborhoodDubai Land	1.05e-05 ***
neighborhoodDubai Marina	1.76e-10 ***
neighborhoodDubai Production City (IMPZ)	5.47e-06 ***
neighborhoodDubai Residence Complex	0.042574 *
neighborhoodDubai Silicon Oasis	1.83e-05 ***
neighborhoodDubai South (Dubai World Central)	0.000250 ***
neighborhoodDubai Sports City	3.30e-08 ***
neighborhoodFalcon City of Wonders	4.43e-05 ***
neighborhoodGreen Community	0.970333
neighborhoodGreens	5.88e-08 ***
neighborhoodInternational City	2.02e-05 ***
neighborhoodJebel Ali	0.002072 **
neighborhoodJumeirah	< 2e-16 ***
neighborhoodJumeirah Beach Residence	8.11e-13 ***
neighborhoodJumeirah Golf Estates	0.006481 **
neighborhoodJumeirah Lake Towers	1.51e-07 ***
neighborhoodJumeirah Village Circle	1.27e-10 ***
neighborhoodJumeirah Village Triangle	0.000367 ***
neighborhoodMeydan	1.72e-06 ***
neighborhoodMina Rashid	0.659164
neighborhoodMirdif	2.72e-05 ***
neighborhoodMohammed Bin Rashid City	2.77e-14 ***
neighborhoodMotor City	0.000459 ***
neighborhoodMudon	2.41e-05 ***
neighborhoodOld Town	9.96e-09 ***
neighborhoodPalm Jumeirah	5.97e-13 ***
neighborhoodRemraam	0.000227 ***
neighborhoodThe Hills	1.51e-06 ***
neighborhoodThe Views	3.50e-09 ***
neighborhoodTown Square	2.35e-10 ***
neighborhoodUmm Suqeim	1.13e-09 ***
neighborhoodwasl gate	0.002661 **
neighborhoodWorld Trade Center	0.236495

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

Residual standard error: 1444000 on 1846 degrees of freedom
Multiple R-squared: 0.7441, Adjusted R-squared: 0.7365
F-statistic: 97.6 on 55 and 1846 DF, p-value: < 2.2e-16

Summary of results after addressing influential points:

Model Improvements:

The removal of influential observations has resulted in a cleaner model fit. Residual standard error is now 1,444,000 AED (initially 1522000 AED), indicating some improvement. Adjusted R-squared has increased to 0.7365, so the model now explains ~73.7% of the variability in property prices instead of ~72.7%.

Key Findings:

size_in_sqft: Remains a strong positive predictor (Estimate = 3,519 AED per sqft, $p < 2e-16$). **no_of_bedrooms:** Retains its negative effect (Estimate = -1,012,000 AED, $p < 2e-16$), reaffirming that properties with more bedrooms but the same size have smaller room spaces, reducing perceived value. **Neighborhood Effects:** Several neighborhoods, like Jumeirah, Palm Jumeirah, and Downtown Dubai, continue to show strong positive associations with price. **Non-Significant Neighborhoods:** Dubai Festival City, Mina Rashid, and World Trade Center remain statistically insignificant, suggesting weak or no association with price.

Next, I will apply a log-transformation to address heteroscedasticity and non-normality.

```
# Log-transforming the price variable
properties_data$log_price <- log(properties_data$price)

# Refitting the model with log-transformed price
model_log <- lm(log_price ~ size_in_sqft + no_of_bedrooms + neighborhood, data = properties_data)
summary(model_log)
```

Call:

```
lm(formula = log_price ~ size_in_sqft + no_of_bedrooms + neighborhood,
    data = properties_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.29619	-0.16195	-0.01119	0.14611	1.33602

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	13.3694224	0.0927655	144.121
size_in_sqft	0.0003931	0.0000128	30.724
no_of_bedrooms	0.2030914	0.0117765	17.245

neighborhoodAl Barsha	-0.4039732	0.1698339	-2.379
neighborhoodAl Furjan	-0.5972849	0.1096163	-5.449
neighborhoodAl Kifaf	0.1756063	0.1413891	1.242
neighborhoodAl Quoz	-0.8114541	0.3000354	-2.705
neighborhoodAl Sufouh	-0.2143442	0.1365049	-1.570
neighborhoodArjan	-0.4967831	0.1258965	-3.946
neighborhoodBarsha Heights (Tecom)	-0.4872953	0.1697129	-2.871
neighborhoodBluewaters	0.5570892	0.1361694	4.091
neighborhoodBusiness Bay	-0.0837117	0.0956804	-0.875
neighborhoodCity Walk	0.2843385	0.1188039	2.393
neighborhoodCulture Village	0.1300803	0.1250378	1.040
neighborhoodDAMAC Hills	-0.3641690	0.1107803	-3.287
neighborhoodDIFC	0.1204513	0.1044772	1.153
neighborhoodDiscovery Gardens	-0.9257892	0.1363076	-6.792
neighborhoodDowntown Dubai	0.2846638	0.0930022	3.061
neighborhoodDubai Creek Harbour (The Lagoons)	0.0497372	0.1031195	0.482
neighborhoodDubai Festival City	-0.3912382	0.1568619	-2.494
neighborhoodDubai Harbour	0.3304869	0.1056953	3.127
neighborhoodDubai Healthcare City	-0.4507934	0.2220333	-2.030
neighborhoodDubai Hills Estate	-0.1683562	0.1001513	-1.681
neighborhoodDubai Land	-1.0661301	0.1259989	-8.461
neighborhoodDubai Marina	-0.0809805	0.0931078	-0.870
neighborhoodDubai Production City (IMPZ)	-0.9437001	0.1365393	-6.912
neighborhoodDubai Residence Complex	-1.0440950	0.3000891	-3.479
neighborhoodDubai Silicon Oasis	-0.8828478	0.1232051	-7.166
neighborhoodDubai South (Dubai World Central)	-0.8981719	0.1569354	-5.723
neighborhoodDubai Sports City	-0.7961703	0.1080832	-7.366
neighborhoodFalcon City of Wonders	-0.1955178	0.1888617	-1.035
neighborhoodGreen Community	-0.7356774	0.2214924	-3.321
neighborhoodGreens	-0.3982767	0.1056358	-3.770
neighborhoodInternational City	-1.0165219	0.1484545	-6.847
neighborhoodJebel Ali	-0.7648783	0.2219443	-3.446
neighborhoodJumeirah	0.4342860	0.1020538	4.255
neighborhoodJumeirah Beach Residence	0.1165386	0.0953218	1.223
neighborhoodJumeirah Golf Estates	-0.4529966	0.2218526	-2.042
neighborhoodJumeirah Lake Towers	-0.4768816	0.0975706	-4.888
neighborhoodJumeirah Village Circle	-0.5506178	0.0939547	-5.860
neighborhoodJumeirah Village Triangle	-0.5716734	0.1887353	-3.029
neighborhoodMeydan	-0.3062621	0.1148544	-2.667
neighborhoodMina Rashid	-0.3678619	0.2996962	-1.227
neighborhoodMirdif	-0.2165181	0.1487763	-1.455
neighborhoodMohammed Bin Rashid City	0.0423290	0.1053634	0.402
neighborhoodMotor City	-0.6561382	0.1064899	-6.162

neighborhoodMudon	-0.2020260	0.1702293	-1.187
neighborhoodOld Town	0.1325989	0.1147321	1.156
neighborhoodPalm Jumeirah	0.1986635	0.0931366	2.133
neighborhoodRemraam	-0.9089962	0.1482677	-6.131
neighborhoodThe Hills	0.0893208	0.1256983	0.711
neighborhoodThe Views	-0.1550221	0.1007232	-1.539
neighborhoodTown Square	-0.6318095	0.1076081	-5.871
neighborhoodUmm Suqeim	0.2777940	0.1212006	2.292
neighborhoodwasl gate	-0.6170337	0.3006221	-2.053
neighborhoodWorld Trade Center	-0.2331111	0.1358565	-1.716
	Pr(> t)		
(Intercept)	< 2e-16	***	
size_in_sqft	< 2e-16	***	
no_of_bedrooms	< 2e-16	***	
neighborhoodAl Barsha	0.017478	*	
neighborhoodAl Furjan	5.75e-08	***	
neighborhoodAl Kifaf	0.214391		
neighborhoodAl Quoz	0.006903	**	
neighborhoodAl Sufouh	0.116533		
neighborhoodArjan	8.25e-05	***	
neighborhoodBarsha Heights (Tecom)	0.004135	**	
neighborhoodBluewaters	4.48e-05	***	
neighborhoodBusiness Bay	0.381737		
neighborhoodCity Walk	0.016795	*	
neighborhoodCulture Village	0.298324		
neighborhoodDAMAC Hills	0.001030	**	
neighborhoodDIFC	0.249102		
neighborhoodDiscovery Gardens	1.49e-11	***	
neighborhoodDowntown Dubai	0.002239	**	
neighborhoodDubai Creek Harbour (The Lagoons)	0.629632		
neighborhoodDubai Festival City	0.012712	*	
neighborhoodDubai Harbour	0.001795	**	
neighborhoodDubai Healthcare City	0.042469	*	
neighborhoodDubai Hills Estate	0.092928	.	
neighborhoodDubai Land	< 2e-16	***	
neighborhoodDubai Marina	0.384550		
neighborhoodDubai Production City (IMPZ)	6.57e-12	***	
neighborhoodDubai Residence Complex	0.000514	***	
neighborhoodDubai Silicon Oasis	1.11e-12	***	
neighborhoodDubai South (Dubai World Central)	1.22e-08	***	
neighborhoodDubai Sports City	2.63e-13	***	
neighborhoodFalcon City of Wonders	0.300691		
neighborhoodGreen Community	0.000913	***	

neighborhoodGreens	0.000168 ***
neighborhoodInternational City	1.02e-11 ***
neighborhoodJebel Ali	0.000581 ***
neighborhoodJumeirah	2.19e-05 ***
neighborhoodJumeirah Beach Residence	0.221644
neighborhoodJumeirah Golf Estates	0.041305 *
neighborhoodJumeirah Lake Towers	1.11e-06 ***
neighborhoodJumeirah Village Circle	5.45e-09 ***
neighborhoodJumeirah Village Triangle	0.002488 **
neighborhoodMeydan	0.007731 **
neighborhoodMina Rashid	0.219810
neighborhoodMirdif	0.145749
neighborhoodMohammed Bin Rashid City	0.687920
neighborhoodMotor City	8.82e-10 ***
neighborhoodMudon	0.235464
neighborhoodOld Town	0.247942
neighborhoodPalm Jumeirah	0.033053 *
neighborhoodRemraam	1.07e-09 ***
neighborhoodThe Hills	0.477424
neighborhoodThe Views	0.123953
neighborhoodTown Square	5.11e-09 ***
neighborhoodUmm Suqeim	0.022017 *
neighborhoodwasl gate	0.040259 *
neighborhoodWorld Trade Center	0.086355 .

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

Residual standard error: 0.2857 on 1849 degrees of freedom

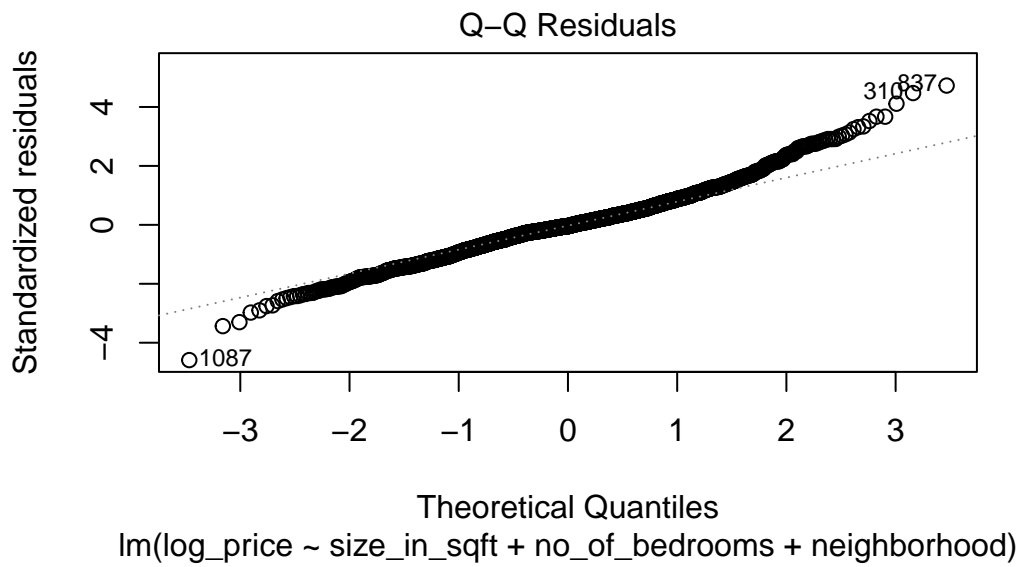
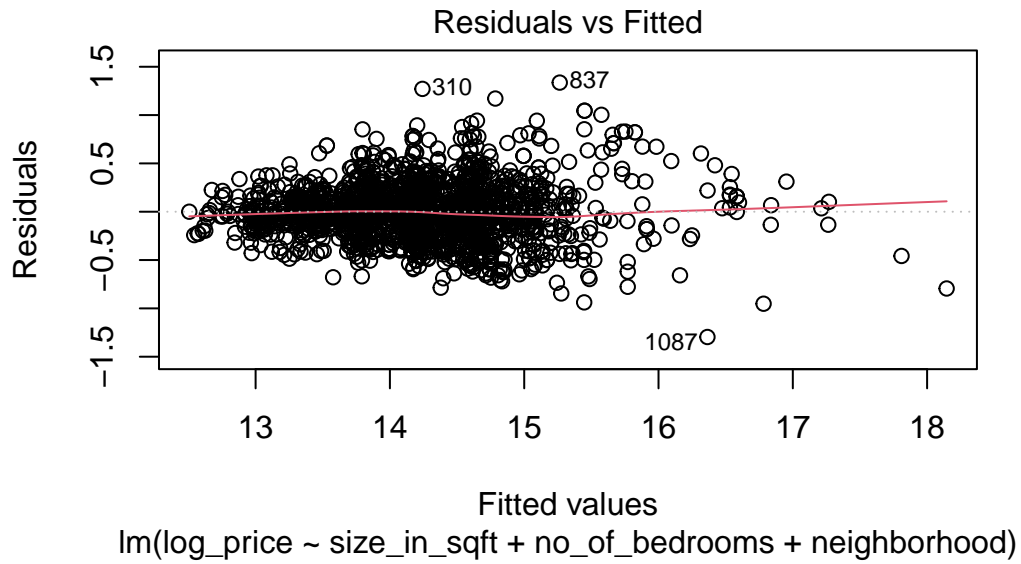
Multiple R-squared: 0.8581, Adjusted R-squared: 0.8539

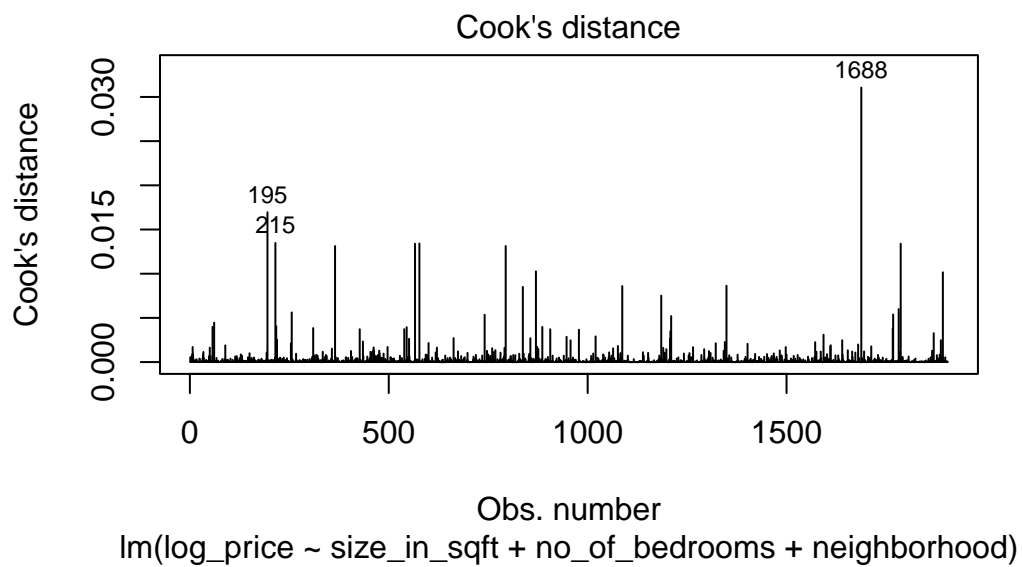
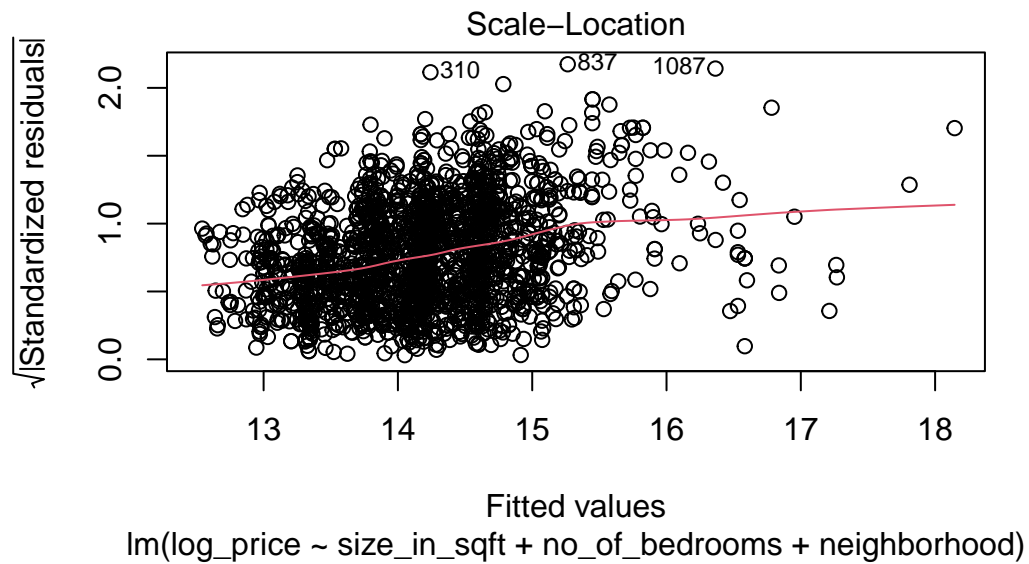
F-statistic: 203.3 on 55 and 1849 DF, p-value: < 2.2e-16

```
# Diagnostic plots for the new model
#par(mfrow = c(2, 2))
plot(model_log, which = 1:6)
```

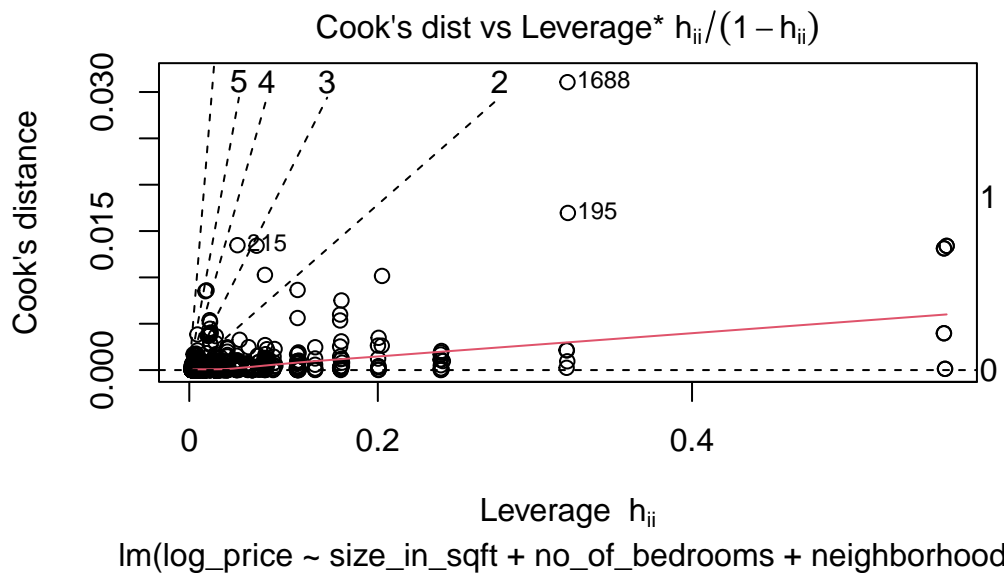
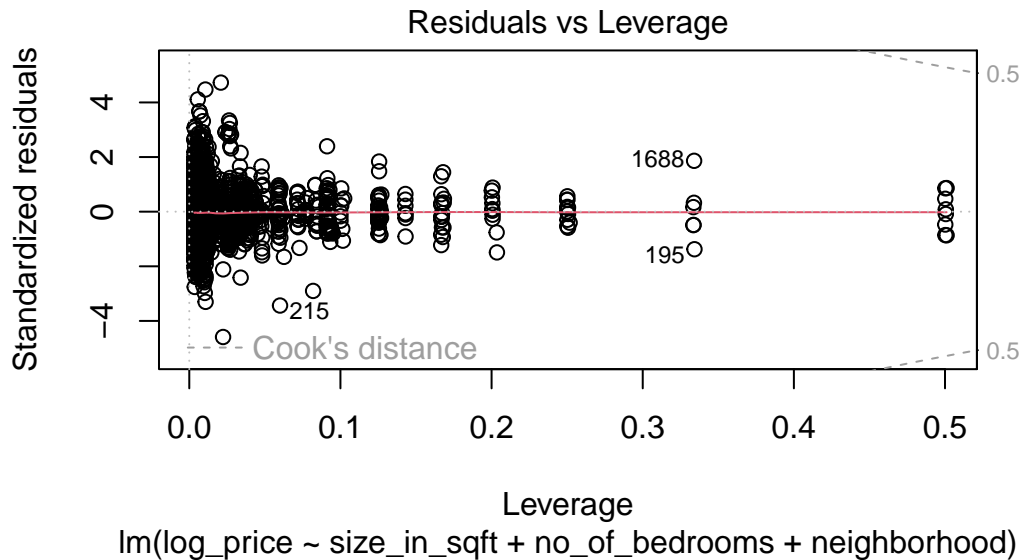
Warning: not plotting observations with leverage one:

338, 1733, 1849, 1898





Warning: not plotting observations with leverage one:
338, 1733, 1849, 1898



Diagnostic Analysis of the Model: The diagnostics provide valuable insights into the behavior of the model: $\text{lm}(\log_price \sim \text{size_in_sqft} + \text{no_of_bedrooms} + \text{neighborhood})$.

Influence of Observations: The Cook's Distance vs Leverage plot highlights observations 195, 215, and 1688 as points with moderate leverage and influence. While these points slightly

impact the model, their Cook's distance values remain low, suggesting they are not extreme enough to destabilize the regression. The Cook's Distance plot confirms this – while these points stand out, their influence is manageable.

Residual Behavior The Residuals vs Leverage plot shows no concerning patterns, with most points clustered around zero. High-leverage points like 1688 and 195 do not appear to distort the model disproportionately.

The Residuals vs Fitted plot paints a clear picture: residuals are largely scattered without obvious patterns, confirming that the relationship between predictors and the log-transformed price is well captured. Residuals for extreme fitted values still show slight spread, suggesting minor variability in the higher range.

Homoscedasticity and Variance The Scale-Location plot demonstrates improved residual variance after the log transformation. The slight upward trend in standardized residuals at higher fitted values indicates that some heteroscedasticity persists but is considerably reduced.

Normality of Residuals The Q-Q plot shows residuals closely following the theoretical quantiles, confirming approximate normality. Minor deviations at the tails, particularly around points 310, 837, and 1087, suggest mild non-normality for extreme observations.

The log transformation of price has significantly improved the model's stability and adherence to assumptions:

Linearity: Residuals show no systematic patterns. **Homoscedasticity:** Variance is largely stable. **Normality:** Residuals are approximately normal, with only minor tail deviations. **Influence:** Points 195, 215, and 1688 merit attention but do not destabilize the model. Overall, the model now performs well, providing a reliable foundation for understanding price dynamics in Dubai's real estate market.

#Conclusion The analysis presented here has revealed a clear and significant relationship between property price, size, number of bedrooms and neighborhood in Dubai's real estate market. While it is generally known that property prices increase with size, the strength and magnitude of this relationship are often unclear. By quantifying this relationship, the results show that for every additional square foot of a property, property prices increase by approximately 2,642 AED. This finding provides a data-driven understanding of the price-size dynamic, which goes beyond assumptions and anecdotal evidence. With an R-squared value of 65.4%, the model demonstrates that property size explains a substantial portion of the variability in prices, making it one of the key drivers in determining value. Unexpectedly, when holding property size constant, having more bedrooms had a negative association with price. This suggests a preference for larger, more spacious rooms rather than a higher number of smaller bedrooms. Also, neighborhoods like Jumeirah, Palm Jumeirah, and Downtown Dubai showed strong positive associations with property prices, while others like Mina Rashid and World Trade Center were insignificant.

These insights are particularly valuable for potential buyers and sellers in Dubai, as they can make more informed decisions about pricing and valuation. For real estate professionals, this

quantitative relationship can serve as a foundation for property appraisals, price predictions, and market analysis. Policymakers and urban planners may also find the results useful in understanding market dynamics and improving housing strategies. However, it is important to note that while size is a major determinant of price, additional factors—such as location, amenities, property age, and market demand—likely influence prices as well. Future analyses incorporating these variables can provide an even more comprehensive understanding of Dubai’s real estate market.

References:

- Haffner, M. E. A., & Boumeester, H. J. F. M. (2010). The Affordability of Housing in the Netherlands: An Increasing Income Gap Between Renting and Owning? *Housing Studies*, 25(6), 799–820. <https://doi.org/10.1080/02673037.2010.511472>
- Abraha, G., & Moges, K. (2024). Journal of Urban Development Studies. *Journal of Urban Development Studies*, 4(1).