

# Excel Games: Data Analysis Stages

~MUHAMMAD FADHIL ABIDIN~



# Table of Contents

The following is a business problem regarding the data analysis process that will be carried out using Ms. Excel:

<u>Step 1</u>	Clean the Initial Data, do it in "Data Cleaning" sheet without removing original data. Just add more columns for your cleaning process.
<u>Step 2</u>	Check for outliers, do the process in "Outliers Checking" sheet and Remove outliers and input clean data in "Clean Data" sheet.
<u>Step 3</u>	Create statistical measures to know the data distribution and provide insights from the result, in "Stat Measures" sheet.
<u>Step 4</u>	"EDA", Top 3 locations with the most properties? From the Top 3, please provide insights on the characteristics of the properties.
<u>Step 5</u>	What aspect makes high impact on the pricing in Mont Kiara? (Correlation) and Please give price recommendation for 3 rooms, 2 bathrooms, 2 car parks, with the size of 1000 sq. ft. in Mont Kiara. (Regression).

# Procedures for processing data analysis using Excel

01

## Analysis ToolPak

Activate on File >> Option >>  
Add-ins >> Excel  
Add-ins>>Analysis ToolPak.

02

## Usage

Analysis ToolPak help you in statistics and data analysis.

03

## Documentation

Write every Process in "Steps Documentation" sheet.

04

## Level up challenge

Finish before 90 minutes!



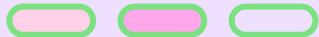
# Initial Data

1	Data Context: The dataset is about list of properties in some locations in Kuala Lumpur. Properties characteristics also included, such as price, rooms, bathrooms, car parks, property type, size, and furnishing status.							
2								
3	Location	Price	Rooms	Bathrooms	Car Parks	Property Type	Size	Furnishing
4	KLCC, Kuala Lumpur	RM 1,250,000	2+1	3	2	Serviced Residence	Built-up : 1,335 sq. ft.	Fully Furnished
5	Damansara Heights, Kuala Lumpur	RM 6,800,000		6	7	Bungalow	Land area : 6900 sq. ft.	Partly Furnished
6	Dutamas, Kuala Lumpur	RM 1,030,000		3	4	2 Condominium (Corner)	Built-up : 1,875 sq. ft.	Partly Furnished
7	Cheras, Kuala Lumpur							
8	Bukit Jalil, Kuala Lumpur	RM 900,000	4+1		3	2 Condominium (Corner)	Built-up : 1,513 sq. ft.	Partly Furnished
9	Taman Tun Dr Ismail, Kuala Lumpur	RM 5,350,000	4+2		5	4 Bungalow	Land area : 7200 sq. ft.	Partly Furnished
10	Seputeh, Kuala Lumpur							
11	Taman Tun Dr Ismail, Kuala Lumpur	RM 2,600,000		5	4	4 Semi-detached House	Land area : 3600 sq. ft.	Partly Furnished
12	Taman Tun Dr Ismail, Kuala Lumpur	RM 1,950,000	4+1		4	3 2-sty Terrace/Link House (EndLot)	Land area : 25x75 sq. ft.	Partly Furnished
13	Sri Petaling, Kuala Lumpur	RM 385,000		3	2	1 Apartment (Intermediate)	Built-up : 904 sq. ft.	Partly Furnished
14	Ampang Hilir, Kuala Lumpur							
15	Taman Tun Dr Ismail, Kuala Lumpur	RM 1,680,000		4	3	2-sty Terrace/Link House (Intermediate)	Land area : 22 x 80 sq. ft.	Partly Furnished
16	Taman Tun Dr Ismail, Kuala Lumpur	RM 1,700,000	3+1		3	2-sty Terrace/Link House (Intermediate)	Land area : 1900 sq. ft.	Partly Furnished
17	Taman Tun Dr Ismail, Kuala Lumpur	RM 4,580,000	6+1		5	5 Bungalow (Intermediate)	Land area : 6000 sq. ft.	Partly Furnished
18	Bukit Jalil, Kuala Lumpur							
19	Taman Tun Dr Ismail, Kuala Lumpur	RM 3,100,000	4+1		3	Semi-detached House (Intermediate)	Land area : 3600 sq. ft.	Partly Furnished
20	Bukit Tunku (Kenny Hills), Kuala Lumpur	RM 9,000,000	6+1		7	4 Bungalow (Corner)	Land area : 8500 sq. ft.	Partly Furnished
21	Damansara Heights, Kuala Lumpur	RM 4,500,000	5+1		7	Bungalow (Corner)	Built-up : 4,842 sq. ft.	Partly Furnished
22	Mont Kiara, Kuala Lumpur	RM 1,780,000	4+1		4	2 Condominium (Corner)	Built-up : 1,830 sq. ft.	Partly Furnished
23	Mont Kiara, Kuala Lumpur	RM 3,450,000	4+1		6	3 Condominium (Corner)	Built-up : 3,720 sq. ft.	Fully Furnished
24	Desa ParkCity, Kuala Lumpur	RM 1,500,000	3+2		4	2 Condominium (Corner)	Built-up : 1,798 sq. ft.	Partly Furnished
25	Damansara Heights, Kuala Lumpur	RM 1,550,000		1	1	1 Serviced Residence (Intermediate)	Built-up : 904 sq. ft.	Fully Furnished
26	Mont Kiara, Kuala Lumpur	RM 1,500,000	3+1		4	3 Condominium	Built-up : 2,163 sq. ft.	Fully Furnished

01

# Clean the initial data without removing original data.

Add more columns for cleaning process!



# Clean “Price”

As we can see the data in the "Price" column is irregular where it contains numbers and strings so we need to format it into numbers so that it can be operated using VBA (alt+F11) >> Insert >> Module then paste the code to create a VBA ExtractMoney function to take the value from the given cell, then remove the text "RM" and the comma using the Replace function. After that, the Trim function is used to remove extra spaces. Finally, the cleaned text is converted into numbers using the Val function, and the result is returned as a numeric value. Then u can use =ExtractMoney(B2).



	=ExtractMoney(B2)
K	L
<b>Price Clean</b>	<b>Car Parks Clean</b>
1250000	
6800000	
1030000	
0	
900000	
5350000	
0	
2600000	
1950000	
385000	
0	
1680000	
1700000	
4580000	
0	
3100000	
9000000	

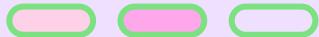
## ExtractMoney

```
Function ExtractMoney(rng As Range) As Double
    Dim text As String
    ' Get the cell value
    text = rng.Value

    ' Remove "RM" and commas from the text
    text = Replace(text, "RM", "")
    text = Replace(text, ",", "")

    ' Trim any extra spaces
    text = Trim(text)

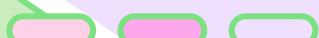
    ' Convert the cleaned text to a number
    ExtractMoney = Val(text)
End Function
```



# Clean “Rooms”

As we can see the data in the "Rooms" column is irregular where it contains numbers, strings and mathematical operations so we need to evaluate this column that it can be operated using VBA (alt+F11) >> Insert >> Module then paste the code to create a VBA EvaluateFormula function that accepts a string containing an Excel cell reference or expression as an argument, then uses the Evaluate method to calculate the result of the expression. The evaluation result is returned as the output of the EvaluateFormula function. Because there is a "Studio" value in this column, we have to use IF to change the value to 1. So the complete formula is:

```
=IF(C2="Studio";1;IF(ISNUMBER(EvaluateFormula(C2));EvaluateFormula(C2);0))
```





Rooms Clean	Bathrooms Clean
3	
6	
3	
0	
5	
6	
0	
5	
5	
3	
0	
4	
4	
7	
0	
5	
7	

## EvaluateFormula

```
Function EvaluateFormula(cell As String) As Variant  
    EvaluateFormula = Evaluate(cell)  
End Function
```

# Clean “Bathrooms”

In the “bathrooms” column it is not that difficult because we only need to overcome blank values by using IF only, if the condition is blank then it will be filled with the value 0 as a representation of the missing value but if not then it returns the value in the column.

So the complete formula will be as follows:

```
=IF(D2="";0;D2).
```

# Clean “Car Parks”

In the “Car Parks” column it is not that difficult because we only need to overcome blank values by using IF only, if the condition is blank then it will be filled with the value 0 as a representation of the missing value but if not then it returns the value in the column.

So the complete formula will be as follows:

```
=IF(E2=""";0;E2).
```

## Bathrooms

	=IF(D2=""";0;D2)
	J
	Bathrooms Clean
3	3
6	7
3	4
0	0
5	3
6	5
0	0
5	4
5	4
3	2

## Car Parks

	=IF(E2=""";0;E2)
	L
	Car Parks Clean
0	2
0	0
0	2
0	0
0	2
0	4
0	0
0	4
0	3
0	1

# Clean “Size”

In the "Size" column there are many variations of values that exist because of this we will only take the numbers after the colon character to make it easier using VBA (alt + F11) >> Insert >> Module then paste the code to create a VBA ExtractNumber takes the value from the cell, removes unnecessary text such as "sq. ft.", "Land area :", "Built-up :", and commas, then cleans up spaces using Trim. If the text contains an "x" (for example, 25x75), the function separates it, multiplies the two parts, and returns the result. If there is no "x", the function directly converts the text to a number using Val and returns it as a numeric value. Then u can use: =ExtractNumber(G2).

=ExtractNumber(G2)

M	N	O
Size Clean		
1335		
6900		
1875		
0		
1513		
7200		
0		
3600		
1875		
904		
0		
1760		
1900		
6000		
0		
3600		
8500		

# ExtractNumber



Function ExtractNumber(rng As Range) As Double

```
Dim text As String
Dim result As Double
Dim parts() As String
Dim i As Integer
```

```
' Get the value from the cell
text = rng.Value
```

```
' Remove unwanted text and trim spaces
text = Replace(text, "sq. ft.", "")
text = Replace(text, "Land area :", "")
text = Replace(text, "Built-up :", "")
text = Replace(text, ",", "")
text = Trim(text)
```



```
' Check if the text contains an "x" (e.g., 25x75)
```

```
If InStr(text, "x") > 0 Then
    ' Split the text by "x" and calculate the product
    parts = Split(text, "x")
    result = 1
    For i = LBound(parts) To UBound(parts)
        result = result * Val(Trim(parts(i)))
    Next i
    ExtractNumber = result
Else
    ' If no "x", return the numeric value
    ExtractNumber = Val(text)
End If
End Function
```

02

# Outliers Checking



Do the process in "Outliers Checking" sheet!

At this stage I put the data starting from the header in the 5th row because above it I will create a place to calculate the average and standard deviation for each numeric column which will be used to check for outliers.

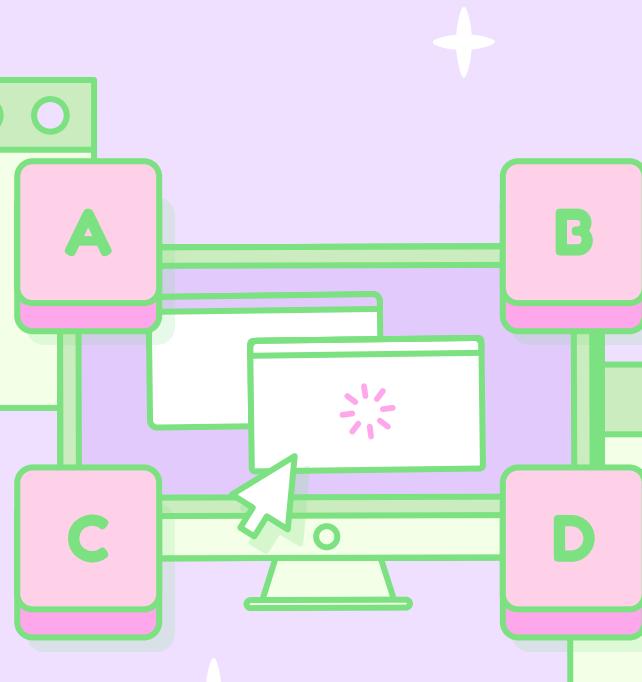
# Updated Data

	Location	Price Clean (RM)	Rooms Clean	Bathrooms Clean	Car Parks Clean	Property Type	Size Clean (Sq Ft)	Furnishing
6	KLCC, Kuala Lumpur	1250000	3	3	2	Serviced Residence	1335	Fully Furnished
7	Dutamas, Kuala Lumpur	1030000	3	4	2	Condominium (Corner)	1875	Partly Furnished
8	Bukit Jalil, Kuala Lumpur	900000	5	3	2	Condominium (Corner)	1513	Partly Furnished
9	Taman Tun Dr Ismail, Kuala Lumpur	5350000	6	5	4	Bungalow	7200	Partly Furnished
10	Taman Tun Dr Ismail, Kuala Lumpur	2600000	5	4	4	Semi-detached House	3600	Partly Furnished
11	Taman Tun Dr Ismail, Kuala Lumpur	1950000	5	4	3	2-sty Terrace/Link	1875	Partly Furnished
12	Sri Petaling, Kuala Lumpur	385000	3	2	1	Apartment (Interme)	904	Partly Furnished
13	Taman Tun Dr Ismail, Kuala Lumpur	4580000	7	5	5	Bungalow (Interme)	6000	Partly Furnished
14	Bukit Tunku (Kenny Hills), Kuala Lumpur	9000000	7	7	4	Bungalow (Corner)	8500	Partly Furnished
15	Mont Kiara, Kuala Lumpur	1780000	5	4	2	Condominium (Corner)	1830	Partly Furnished
16	Mont Kiara, Kuala Lumpur	3450000	5	6	3	Condominium (Corner)	3720	Fully Furnished
17	Desa ParkCity, Kuala Lumpur	1500000	5	4	2	Condominium (Corner)	1798	Partly Furnished
18	Damansara Heights, Kuala Lumpur	1550000	1	1	1	Serviced Residence	904	Fully Furnished
19	Mont Kiara, Kuala Lumpur	1500000	4	4	3	Condominium	2163	Fully Furnished
20	Bangsar South, Kuala Lumpur	490000	1	1	1	Serviced Residence	520	Fully Furnished
21	Bukit Jalil, Kuala Lumpur	610000	4	2	2	Condominium (Intermediate)	1236	Partly Furnished
22	Dutamas, Kuala Lumpur	1035880	4	4	1	Condominium (Enclosed)	1876	Partly Furnished
23	Ampang Hilir, Kuala Lumpur	3300000	5	4	4	Condominium	3536	Unfurnished
24	Ampang Hilir, Kuala Lumpur	460000	1	1	1	Serviced Residence	613	Fully Furnished
25	KLCC, Kuala Lumpur	2400000	2	2	1	Serviced Residence	1006	Fully Furnished
26	Jalan Klang Lama (Old Klang Road), Kuala Lumpur	1438000	5	4	3	Condominium (Corner)	2378	Partly Furnished
27	Mont Kiara, Kuala Lumpur	1720000	5	6	3	Condominium (Corner)	2020	Partly Furnished
28	KLCC, Kuala Lumpur	4280000	4	4	2	Serviced Residence	2195	Fully Furnished
29	KLCC, Kuala Lumpur	2300000	1	2	1	Serviced Residence	1023	Partly Furnished
30	Damansara Heights, Kuala Lumpur	4200000	6	5	5	Bungalow	5274	Partly Furnished
31	Sunqai Besi, Kuala Lumpur	900000	6	5	3	2-sty Terrace/Link	20	Fully Furnished

# Excel Formula

## Average Formula:

=AVERAGE(x1:x999)



## Standard Deviation Formula

=STDEV.S(x1:x999)



# Calculation

Perform the average and standard deviation calculations for all numeric columns until the results are at least as shown in the following image:

F3	A	B	C	D	E	F	G
1		Price Clean (RM)	Rooms Clean	Bathrooms Clean	Car Parks Clean	Size Clean (Sq Ft)	
2	Average:	1875458,15	3,67	3,19	2,19	34164,42	
3	Standard Deviation:	2393877,06	1,50	1,63	1,38	834825,78	



# Z-Score Calculation

## Formula:

(value - Mean)/Std deviation

## Explanation

Calculate the z-score of each value from all columns using the previously calculated mean and standard deviation, for example: =(B6-\$B\$2)/\$B\$3, where B6 is the first value from the "Price" column and B2 and B3 are the previously calculated mean and standard deviation.

# Outliers or Not

	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
4		Z-Score													RESULT
5	Price	Clean	Rooms	Bathrooms	Clean	Car Parks	Clean	Size	Clean	R1	R2	R3	R4		
6	-0.281274131	-0.450497264	-0.116893239	-0.139391293	-0.039325	Good	Good	Good	Good	Good	Good	Good	Good	Good	
7	-0.353175258	-0.450497264	0.497851423	-0.139391293	-0.038678	Good	Good	Good	Good	Good	Good	Good	Good	Good	
8	-0.40749047	0.886822927	-0.116893239	-0.139391293	-0.039112	Good	Good	Good	Good	Good	Good	Good	Good	Good	
9	1.45142869	1.555483022	1.112596084	1.306307874	-0.032299	Good	Good	Good	Good	Good	Good	Good	Good	Good	
10	0.302664603	0.886822927	0.497851423	1.306307874	-0.036612	Good	Good	Good	Good	Good	Good	Good	Good	Good	
11	0.031138545	0.886822927	0.497851423	0.583458291	-0.038678	Good	Good	Good	Good	Good	Good	Good	Good	Good	
12	-0.622812653	-0.450497264	-0.7316379	-0.862240876	-0.039841	Good	Good	Good	Good	Good	Good	Good	Good	Good	
13	1.129774746	2.224143118	1.112596084	2.029157458	-0.033737	Good	Good	Good	Good	Good	Good	Good	Good	Good	
14	2.976151933	2.224143118	2.342085406	1.306307874	-0.030742	Good	Good	Good	Good	Good	Good	Good	Good	Good	
15	-0.039875962	0.886822927	0.497851423	-0.139391293	-0.038732	Good	Good	Good	Good	Good	Good	Good	Good	Good	
16	0.857737139	0.886822927	1.727340745	0.583458291	-0.038468	Good	Good	Good	Good	Good	Good	Good	Good	Good	
17	-0.156841032	0.886822927	0.497851423	-0.139391293	-0.03877	Good	Good	Good	Good	Good	Good	Good	Good	Good	
18	-0.135954413	-1.787817456	-1.346382561	-0.862240876	-0.039841	Good	Good	Good	Good	Good	Good	Good	Good	Good	
19	-0.156841032	0.218162831	0.497851423	0.583458291	-0.038333	Good	Good	Good	Good	Good	Good	Good	Good	Good	
20	-0.578750752	-1.787817456	-1.346382561	-0.862240876	-0.040301	Good	Good	Good	Good	Good	Good	Good	Good	Good	
21	-0.528622864	0.218162831	-0.7316379	-0.139391293	-0.039443	Good	Good	Good	Good	Good	Good	Good	Good	Good	
22	-0.350718992	0.218162831	0.497851423	-0.862240876	-0.038677	Good	Good	Good	Good	Good	Good	Good	Good	Good	
23	0.595077279	0.886822927	0.497851423	1.306307874	-0.036688	Good	Good	Good	Good	Good	Good	Good	Good	Good	
24	-0.591282724	-1.787817456	-1.346382561	-0.862240876	-0.040419	Good	Good	Good	Good	Good	Good	Good	Good	Good	
25	0.219118123	-1.11915736	-0.7316379	-0.862240876	-0.039719	Good	Good	Good	Good	Good	Good	Good	Good	Good	
26	-0.182740441	0.886822927	0.497851423	0.583458291	-0.038076	Good	Good	Good	Good	Good	Good	Good	Good	Good	
27	-0.064939905	0.886822927	1.727340745	0.583458291	-0.038504	Good	Good	Good	Good	Good	Good	Good	Good	Good	
28	1.00455027	0.218162831	0.497851423	-0.139391293	-0.038295	Good	Good	Good	Good	Good	Good	Good	Good	Good	
29	0.177344884	-1.787817456	-0.7316379	-0.862240876	-0.039699	Good	Good	Good	Good	Good	Good	Good	Good	Good	
30	0.971038435	1.555483022	1.112596084	2.029157458	-0.034607	Good	Good	Good	Good	Good	Good	Good	Good	Good	

## Outliers Checking

Next we will determine whether a field is an outlier or not by using the Formula =IF(OR(I6>3;I6<-3); "Outlier"; "Good") checking the value in cell I6. If the value is greater than 3 or less than -3, then the result is "Outlier". If it does not meet both conditions, the result is "Good". This formula is used to identify whether the value in I6 is an outlier based on the specified limits.



**Cmd**

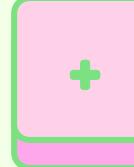
**A**

After checking all the Z-Score results and determining whether they are outliers or not, copy all the results and delete the outliers using the filter on the next sheet and don't forget to delete the blank data (0) in each numeric column that has been processed previously.

**Z**

**C**

**Ctrl**



# Updated Data

1	Location	Price Clean (RM)	Rooms Clean	Bathrooms Clean	Car Parks Clean	Property Type	Size Clean (Sq Ft)	Furnishing
2	KLCC, Kuala Lumpur	1250000	3	3	2	Serviced Residence	1335	Fully Furnished
3	Dutamas, Kuala Lumpur	1030000	3	4	2	Condominium (Corn)	1875	Partly Furnished
4	Bukit Jalil, Kuala Lumpur	900000	5	3	2	Condominium (Corn)	1513	Partly Furnished
5	Taman Tun Dr Ismail, Kuala Lumpur	5350000	6	5	4	Bungalow	7200	Partly Furnished
6	Taman Tun Dr Ismail, Kuala Lumpur	2600000	5	4	4	Semi-detached Hous	3600	Partly Furnished
7	Taman Tun Dr Ismail, Kuala Lumpur	1950000	5	4	3	2-sty Terrace/Link H	1875	Partly Furnished
8	Sri Petaling, Kuala Lumpur	385000	3	2	1	Apartment (Intermed)	904	Partly Furnished
9	Taman Tun Dr Ismail, Kuala Lumpur	4580000	7	5	5	Bungalow (Intermedi	6000	Partly Furnished
10	Bukit Tunku (Kenny Hills), Kuala Lumpur	9000000	7	7	4	Bungalow (Corner)	8500	Partly Furnished
11	Mont Kiara, Kuala Lumpur	1780000	5	4	2	Condominium (Corn)	1830	Partly Furnished
12	Mont Kiara, Kuala Lumpur	3450000	5	6	3	Condominium (Corn)	3720	Fully Furnished
13	Desa ParkCity, Kuala Lumpur	1500000	5	4	2	Condominium (Corn)	1798	Partly Furnished
14	Damansara Heights, Kuala Lumpur	1550000	1	1	1	Serviced Residence	904	Fully Furnished
15	Mont Kiara, Kuala Lumpur	1500000	4	4	3	Condominium	2163	Fully Furnished
16	Bangsar South, Kuala Lumpur	490000	1	1	1	Serviced Residence	520	Fully Furnished
17	Bukit Jalil, Kuala Lumpur	610000	4	2	2	Condominium (Interr	1236	Partly Furnished
18	Dutamas, Kuala Lumpur	1035880	4	4	1	Condominium (EndL	1876	Partly Furnished
19	Ampang Hilir, Kuala Lumpur	3300000	5	4	4	Condominium	3536	Unfurnished
20	Ampang Hilir, Kuala Lumpur	460000	1	1	1	Serviced Residence	613	Fully Furnished
21	KLCC, Kuala Lumpur	2400000	2	2	1	Serviced Residence	1006	Fully Furnished
22	Jalan Klang Lama (Old Klang Road), Kua	1438000	5	4	3	Condominium (Corn)	2378	Partly Furnished
23	Mont Kiara, Kuala Lumpur	1720000	5	6	3	Condominium (Corn)	2020	Partly Furnished
24	KLCC, Kuala Lumpur	4280000	4	4	2	Serviced Residence	2195	Fully Furnished
25	KLCC, Kuala Lumpur	2300000	1	2	1	Serviced Residence	1023	Partly Furnished
	mansara Heights, Kuala Lumpur	4200000	6	5	5	Bungalow	5274	Partly Furnished
	gai Besi, Kuala Lumpur	900000	6	5	3	2-sty Terrace/Link H	20	Fully Furnished

03

# Statistical Measures To Know The Data Distribution

Do the process in "Stat Measures" sheet!

# Descriptive Statistics



Descriptive Statistics

Input

Input Range:

Grouped By:  Columns  Rows

Labels in first row

Output options

Output Range:

New Worksheet Ply:  
 New Workbook

Summary statistics

Confidence Level for Mean:  %

Kth Largest:

Kth Smallest:

By using the toolpak analysis on the Data >> Data Analysis >> Descriptive Statistics tab. We can do descriptive statistics faster by entering the input range as in the picture where I will check the entire "Price" column and mark that the first column is a label then mark the output will be displayed starting from column J1, and ask the program to display the summary output!



# Result

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Price Clean (RM)	Rooms Clean	Bathrooms Clean	Car Parks Clean	Property Type	Size Clean (Sq Ft)	Furnishing		Price Clean (RM)		Rooms Clean		Bathrooms Clean			
2	1250000	3	3	2	Serviced Residence	1335 Fully Furnished										
3	1030000	3	4	2	Condominium (Corner)	1875 Partly Furnished		Mean	1603980,29	Mean	3,576438	Mean	3,072900			
4	900000	5	3	2	Condominium (Corner)	1513 Partly Furnished		Standard Error	25546,8414	Standard Error	0,0241844	Standard Error	0,0256062			
5	1100000	6	5	4	Bungalow	7200 Partly Furnished		Median	1139000	Median	4	Median	3			
6	1100000	5	4	4	Semi-detached House	3600 Partly Furnished		Mode	1000000	Mode	3	Mode	2			
7	1100000	5	4	3	2-sty Terrace/Link House	1875 Partly Furnished		Standard Deviation	1456618,08	Standard Deviation	1,3789355	Standard Deviation	1,4600024			
8	1100000	3	2	1	Apartment (Intermed)	904 Partly Furnished		Sample Variance	2,1217E+12	Sample Variance	1,9014631	Sample Variance	2,1316069			
9	1100000	7	5	5	Bungalow (Intermedi)	6000 Partly Furnished		Kurtosis	5,88585247	Kurtosis	-0,0724253	Kurtosis	-0,258027			
10	1100000	7	7	4	Bungalow (Corner)	8500 Partly Furnished		Skewness	2,2471235	Skewness	0,2433666	Skewness	0,7323345			
11	1100000	5	4	2	Condominium (Corner)	1830 Partly Furnished		Range	8998850	Range	7	Range	7			
12	1100000	5	6	3	Condominium (Corner)	3720 Fully Furnished		Minimum	1150	Minimum	1	Minimum	1			
13	1100000	5	4	2	Condominium (Corner)	1798 Partly Furnished		Maximum	9000000	Maximum	8	Maximum	8			
14	1100000	1	1	1	Serviced Residence	904 Fully Furnished		Sum	5214539915	Sum	11627	Sum	9990			
15	1100000	4	4	3	Condominium	2163 Fully Furnished		Count	3251	Count	3251	Count	3251			
16	1100000	1	1	1	Serviced Residence	520 Fully Furnished										
17	1100000	4	2	2	Condominium (Inter)	1236 Partly Furnished										
18	1100000	4	4	1	Condominium (EndL)	1876 Partly Furnished										
19	1100000	5	4	4	Condominium	3536 Unfurnished										
20	1100000	1	1	1	Serviced Residence	613 Fully Furnished		Mean	2,07105506	Mean	2155,7804					
21	1100000	2	2	1	Serviced Residence	1006 Fully Furnished		Standard Error	0,01859812	Standard Error	243,93103					
22	1100000	5	4	3	Condominium (Corner)	2378 Partly Furnished		Median	2	Median	1440					
23	1100000	5	6	3	Condominium (Corner)	2020 Partly Furnished		Mode	2	Mode	1650					
24	1100000	4	4	2	Serviced Residence	2195 Fully Furnished		Standard Deviation	1,06041883	Standard Deviation	13908,348					
25	1100000	1	2	1	Serviced Residence	1023 Partly Furnished		Sample Variance	1,12448809	Sample Variance	193442132					
26	1100000	6	5	5	Bungalow	5274 Partly Furnished		Kurtosis	2,11959115	Kurtosis	3170,7865					
27	1100000	6	5	3	2-sty Terrace/Link H	20 Fully Furnished		Skewness	1,33766912	Skewness	55,963944					
28	1100000	4	4	3	Serviced Residence	3095 Fully Furnished		Range	5	Range	789997					
29	1100000	2	2	1	Serviced Residence	896 Fully Furnished		Minimum	1	Minimum	3					
30	1100000	3	2	2	Condominium (Corner)	1200 Partly Furnished		Maximum	6	Maximum	790000					
31	1100000	5	4	4	Condominium (EndL)	1772 Partly Furnished		Sum	6733	Sum	7008442					
32	1100000	1	1	1	Serviced Residence	5214539915		Count	3251	Count	3251					

# Question and answer

## Question:

Provide insight from the statistical results!

## Answer:

Overall, the average price of renting a hotel in Kuala Lumpur with four rooms, three bathrooms, two car parks, with a size of 2155 Sq Ft is RM 1,603,980.

04

# Exploratory Data Analysis & Find Three Location With The Most Properties

Do the process in "EDA" sheet!

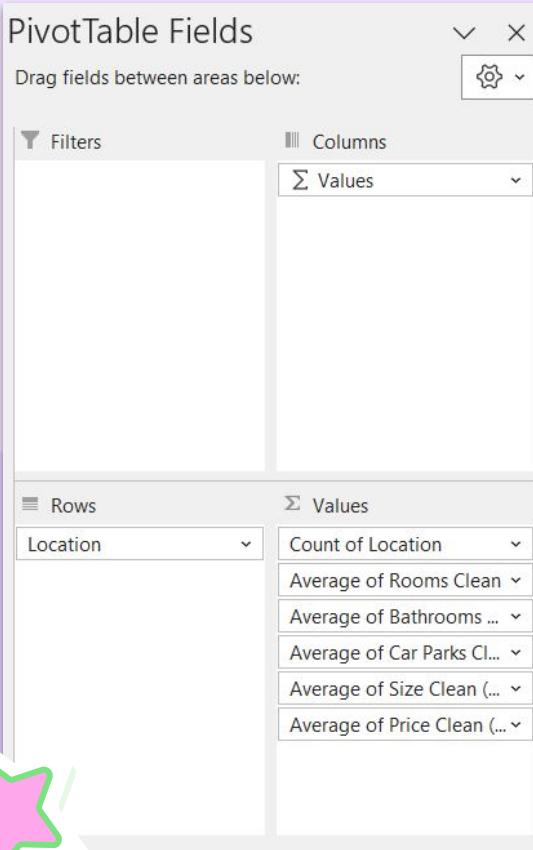
# Exploratory Data Analysis

	B	C	D	E	F	G	H						
1	Price	Clean (RM)	Rooms	Clean	Bathrooms	Clean	Car Parks	Clean	Property	Type	Size	Clean (Sq Ft)	Furnishing
2	Fit	You can double-click a value to see which detailed values make up the summarized total.											
3													
4	(Tell me more)												
5													
6	Taman Tun Dr Ismail, Kuala Lumpur	1250000	3	2	2	2	2	1325	Serviced Residence		1325	Fully Furnished	
7		1030000	3	4	2	2	2	1030	Condominium (Com)		1030	Partly Furnished	
8		900000	5	3	2	2	2	900	Condominium (Com)		900	Partly Furnished	
9		5350000	6	5	4	4	4	5350	Bungalow		5350	Partly Furnished	
10		2600000	5	4	3	3	3	2600	Semi-detached House		2600	Partly Furnished	
11		1950000	5	4	3	3	3	1950	3-2-sty Terrace/Link H		1950	Partly Furnished	
12		385000	3	2	2	2	2	385	Apartments (Intermed)		385	Partly Furnished	
13		458000	7	5	4	5	5	4580	Bungalow (Intermedi)		4580	6000 Partly Furnished	
14		900000	7	7	6	6	6	900	Bungalow (Corner)		900	8500 Partly Furnished	
15		1780000	5	4	3	3	3	1780	Condominium (Com)		1780	1630 Partly Furnished	
16		3450000	5	6	5	5	5	3450	Condominium (Com)		3450	3720 Fully Furnished	
17		1500000	5	4	3	3	3	1500	Condominium (Com)		1500	1790 Partly Furnished	
18		1550000	1	1	1	1	1	1550	Serviced Residence		1550	904 Fully Furnished	
19		1500000	4	4	3	3	3	1500	Condominium (End)		1500	2163 Fully Furnished	
20		490000	1	1	1	1	1	490	Serviced Residence		490	520 Fully Furnished	
21		610000	4	2	2	2	2	610	Condominium (Inter)		610	1236 Partly Furnished	
22		1035880	4	4	3	3	3	1035	Condominium (End)		1035	1876 Partly Furnished	
23		3300000	5	4	4	4	4	3300	Condominium		3300	3536 Unfurnished	
24		460000	1	1	1	1	1	460	Serviced Residence		460	613 Fully Furnished	
25		2400000	2	2	2	2	2	2400	Serviced Residence		2400	1006 Fully Furnished	
26		1438000	5	4	3	3	3	1438	Condominium (Com)		1438	2378 Partly Furnished	
27		1720000	5	6	3	3	3	1720	Condominium (Com)		1720	2020 Partly Furnished	
28		4280000	4	4	2	2	2	4280	Serviced Residence		4280	2195 Fully Furnished	
29		2300000	1	2	1	1	1	2300	Serviced Residence		2300	1023 Partly Furnished	
30		4200000	6	5	5	5	5	4200	Bungalow		4200	5274 Fully Furnished	
31		900000	6	5	3	3	3	900	3-3-sty Terrace/Link H		900	20 Fully Furnished	

## Pivot Table

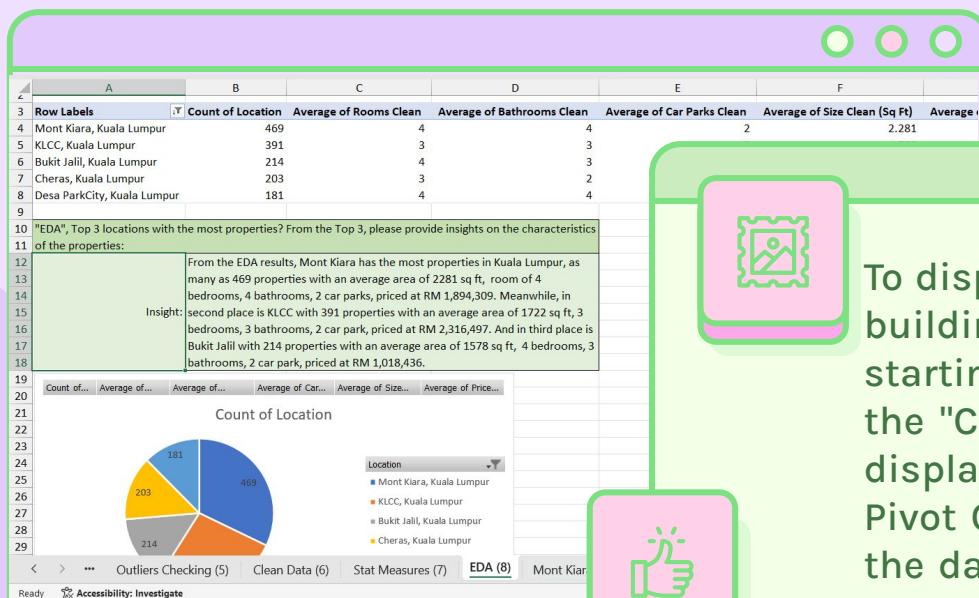
We will use pivot table to create EDA based on our business question. For that we have to select all data then go to insert tab >> pivot table, create pivot table on new sheet.

# Pivot Table Setup



Here is a breakdown of the pivot table setup where we also calculate the location column as a parameter to find out which hotel in Kuala Lumpur has the most buildings, then the other columns calculate the average of each numeric column as insight for our analysis.

# Pivot Table Result



To display the hotels with the most buildings, we can filter the data starting from the largest value in the "Count of Location" column. To display the Pie Chart, click Insert >> Pivot Chart >> Pie Chart, then filter the data value above 165 to narrow the search and make it easier to get insights.



05

# Correlation and Regression

Do the process in "Mont Kiara" sheet!

# Question and answer

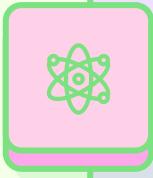
## Question:

What aspect makes high impact on the pricing in Mont Kiara? (Correlation) and Please give price recommendation for 3 rooms, 2 bathrooms, 2 car parks, with the size of 1000 sq. ft. in Mont Kiara. (Regression)



## Answer:

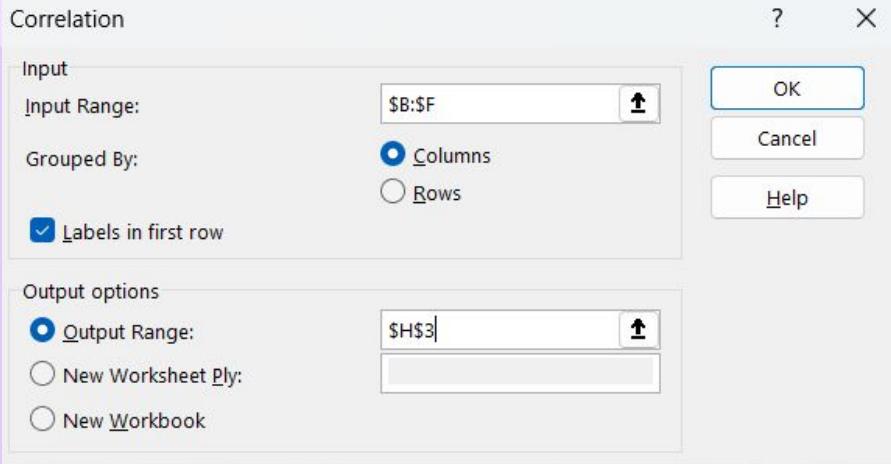
To answer the business problem we need to filter the data only for Mont Kiara! Therefore use the filter on the "Location" column and exclude Mont Kiara then delete all existing fields and deactivate the filter again so that now we only have Mont Kiara as our main location.



# Updated Data

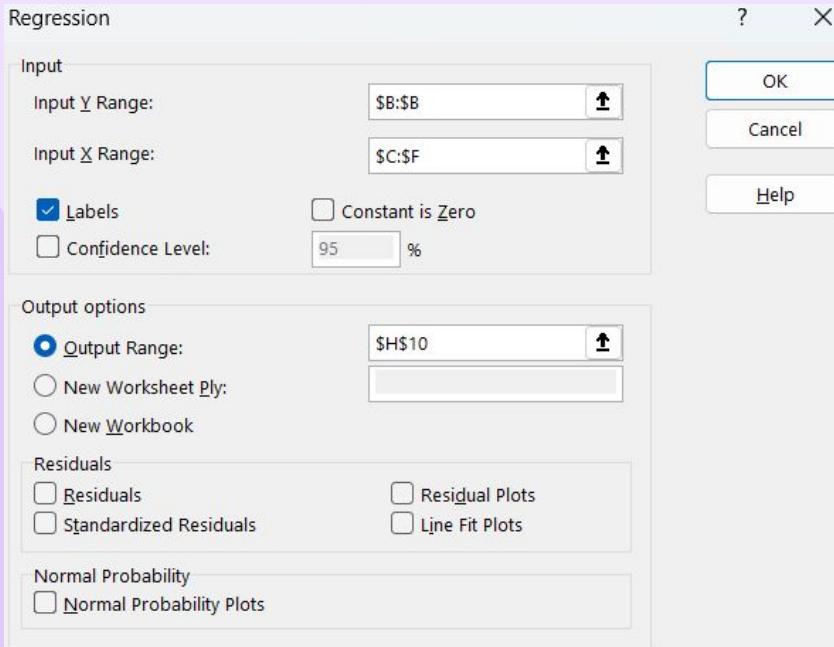
	A	B	C	D	E	F
1	Location	Price Clean (RM)	Rooms Clean	Bathrooms Clean	Car Parks Clean	Size Clean (Sq Ft)
2	Mont Kiara, Kuala Lumpur	1780000	5	4	2	1830
3	Mont Kiara, Kuala Lumpur	3450000	5	6	3	3720
4	Mont Kiara, Kuala Lumpur	1500000	4	4	3	2163
5	Mont Kiara, Kuala Lumpur	1720000	5	6	3	2020
6	Mont Kiara, Kuala Lumpur	1080000	2	2	1	896
7	Mont Kiara, Kuala Lumpur	1150000	3	2	2	1200
8	Mont Kiara, Kuala Lumpur	2300000	5	5	2	2702
9	Mont Kiara, Kuala Lumpur	1830000	4	4	2	1830
10	Mont Kiara, Kuala Lumpur	2180000	5	4	3	3000
11	Mont Kiara, Kuala Lumpur	3450000	5	6	3	3720
12	Mont Kiara, Kuala Lumpur	2400000	5	5	3	2702
13	Mont Kiara, Kuala Lumpur	1850000	5	4	2	1830
14	Mont Kiara, Kuala Lumpur	3100000	5	6	3	3400
15	Mont Kiara, Kuala Lumpur	2600000	5	4	3	3041
16	Mont Kiara, Kuala Lumpur	1630000	5	4	2	2065
17	Mont Kiara, Kuala Lumpur	2480000	5	5	3	3041
18	Mont Kiara, Kuala Lumpur	2480000	5	5	3	3041
	Mont Kiara, Kuala Lumpur	2200000	4	4	2	2243
	Mont Kiara, Kuala Lumpur	1550000	4	4	2	1496

# Correlation



Use the toolpak analysis to perform correlation on the Data >> Data Analysis >> Correlation. In the input range column, enter all the location of numeric columns we have, then mark the first row as labels and use the output range as a determination of where the correlation output will be displayed.

# Regression



For regression select the "price" column as the independent variable and the other columns as dependent variables on the X and Y inputs. Check labels and select the output range to determine where the regression output will be displayed.

Cmd

A

# Clue

To answer the regression question,  
we need to add the intercept  
coefficient with all the products of  
the desired dependent variable with  
the output of the dependent  
variable regression coefficient.

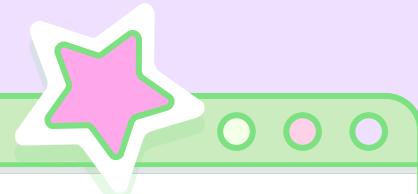
Z

C

Ctrl



# Final Result





# THANKS

Feel free to ask me anything on the  
comment below.