Property Listing Company Analysis

RevoU FSDA Week 2-3 Intermediate Assignment

Spreadsheet & Statistics

by **Riadhi Nur Fajrina**

TABLE OF CONTENTS

BACKGROUND OVERVIEW

Business Background

Core Business Problem

Data Explanation

BUSINESS BACKGROUND

ABC Company is a property listing company in Malaysia. Their mission is to provide various available choices of property to their users around Kuala Lumpur and other big cities in Malaysia.
The company want to **maximize the profit** through joint-profit sharing of 20%.

The head of Data tasked me to perform **end-to-end analysis** and **provide insights** from the data set provided to help the company achieve their goals.

CORE BUSINESS PROBLEM

1. Maximize the profit through 20% joint-sharing profit.
2. The more property sold, the more revenue the company get.
3. The highest-priced property brings the highest percentage of revenue, but hard to sell.
4. Provide insights and make recommendations that can satisfy the company and potential buyers.

DATASET OVERVIEW

This dataset is mainly consist with 5000 dataset of Poperty Listing in Kuala Lumpur.
The information is about :

- Location
- Price
  Rooms, Bathrooms, Carparks
- Property Type
- Size
- Furnishing

To gather the insights, we can perform *descriptive analytics* for the Price, Rooms, Bathrooms, Carparks and Size columns as they all are numerical data type. And we can use *pivot table* for the Location, Property Type and Furnishing as they all categorical data type.

DATA CLEANING

DATA CLEANING PROCESS

Removing Duplicates

- Remove 68 duplicate rows from the dataset

Removing Irrelevant Values

- Remove rows with missing value that don't give any information. 23 rows removed.
- Remove *Bungalow Land* and *Residential Land* because they barely provide any information. 47 rows removed.
- Remove anomaly data (missing value and overly low) in Price and Size column. 141 rows removed.
- Remove outliers that exceed *Batas Atas* in Price and Size after performing Descriptive Analystic. 454 rows removed.

Standarize Values

- Separate text using *split text to column* (remove "RM" in Price , text after " **,** " in Location and text after "" **(** " in Property Type).
- Convert non-numerical values into numerical values in Price, Rooms and Size.
- Standarize Size units into sqft (from sq.m, acres to sqft).

Handling Missing Value

- Perform *XLOOKUP* function to fill missing value in Rooms, Bathrooms and Carparks from the Desa Park City property

From 5000 rows to 4267 rows after data cleaning

DATA CLEANING DESC STATS

Outliers must be removed (using *Batas Atas Outlier* as the limit) because they can affect the average of Price and Size, causing issues when recommending our properties to clients.

Price	
Count	4721
Minimum	105000
Maximum	50000000
Mean	2045950.729
Median	1300000
Mode	1,200,000
Sum	9658933392
Range	49895000
Sample Variance	5683375516525
Standard Deviation	2383983.12
Standard Error	34696.5406
Skewness	4.721146871
Kurtosis	48.63037036
Coefficient of Variation	1.165220201
Confidence Level(95%)	66910.05927
Quartile 1	718000
Quartile 3	2450000
IQR	1732000
Batas Bawah Outlier	-1880000
Batas Atas Outlier	5048000

Size	
Count	4721
Minimum	304
Maximum	790000
Mean	2594.959314
Median	1626
Mode	1,650
Sum	12250802.92
Range	789696
Sample Variance	142556199.4
Standard Deviation	11939.69009
Standard Error	173.7705013
Skewness	61.17216895
Kurtosis	4038.657145
Coefficient of Variation	4.601108782
Confidence Level(95%)	338.252199
Quartile 1	1087
Quartile 3	2767
IQR	1680
Batas Bawah Outlier	-1433
Batas Atas Outlier	5287

DESCRIPTIVE STATISTICS

(after Data Cleaning Process)

PRICE

- The Price column has a wide range. The property prices range from RM 105,000 to RM 5,070,000.
- In representing the character of Property Price, we can use Median Score (1,200,000) as the central tendency.
- Price skewness is positive (Mean > Median).
- Outliers will not be removed because this is the Descriptive Analystics result after removing outliers in the Data Cleaning Process.

Price	
Count	4267
Minimum	105,000
Maximum	5,070,000
Mean	1,479,683
Median	1,200,000
Mode	1,200,000
Sum	6,313,806,418
Range	4,965,000
Sample Variance	1050909184795
Standard Deviation	1025138.617
Standard Error	15693.55278
Skewness	1.187886881
Kurtosis	48.63037036
Coefficient of Variation	0.6928097237
Confidence Level(95%)	66910.05927
Quartile 1	676775
Quartile 2	1200000
Quartile 3	1990000
Quartile 4	5070000
IQR	1313225
Batas Bawah Outlier	-1293062.5
Batas Atas Outlier	3959837.5

SIZE

- The Size column has a wide range. Property sizes range from 304 to 5,274 square feet.
- In representing the character of Property Size, we can use Median Score (1,507) as the central tendency.
- Size skewness is positive (Mean > Median).
- Outliers will not be removed because this is the Descriptive Analystics result after removing outliers in the Data Cleaning Process.

Size	
Count	4267
Minimum	304
Maximum	5,274
Mean	1,803
Median	1,507
Mode	1,650
Sum	7,692,011
Range	4,970
Sample Variance	1018014.847
Standard Deviation	1008.967218
Standard Error	15.44598947
Skewness	1.144320574
Kurtosis	4038.657145
Coefficient of Variation	0.5597057932
Confidence Level(95%)	338.252199
Quartile 1	1050
Quartile 2	1507
Quartile 3	2250
Quartile 4	5274
IQR	1200
Batas Bawah Outlier	-750
Batas Atas Outlier	4050

ROOMS

- The Rooms column has a wide range. The number of rooms on the property ranges from 1 to 16.
- We can identify 4 as the central tendency in representing the character of Room Property. Because the Mean, Median, and Mode are all in close proximity (4),
- Room skewness is **slightly positive** (Mean > Median).
- Outliers will not be removed because this is the Descriptive Analystics result after removing outliers in the Data Cleaning Process.

Rooms	
Count	4267
Minimum	1
Maximum	16
Mean	3.6
Median	4
Mode	4
Sum	15436
Range	15
Sample Variance	1.865871722
Standard Deviation	1.365969151
Standard Error	0.02091122957
Skewness	0.3549859954
Kurtosis	9.585853829
Coefficient of Variation	0.3775971993
Confidence Level(95%)	0.044263018
Quartile 1	3
Quartile 2	4
Quartile 3	4
Quartile 4	16
IQR	1
Batas Bawah Outlier	1.5
Batas Atas Outlier	5.5

BATHROOMS

- The Bathrooms column has a wide range. The number of rooms on the property ranges from 1 to 9.
- In representing the character of Property Bathrooms, we can use Median Score (3) as the central tendency.
- Bathroom skewness is **slightly positive** (Mean > Median).
- Outliers will not be removed because this is the Descriptive Analystics result after removing outliers in the Data Cleaning Process.

Bathrooms	
Count	4267
Minimum	1
Maximum	9
Mean	3.1
Median	3
Mode	2
Sum	13256
Range	8
Sample Variance	2.055903025
Standard Deviation	1.43384205
Standard Error	0.0219502763
Skewness	0.6364706566
Kurtosis	9.259923971
Coefficient of Variation	0.4615422472
Confidence Level(95%)	0.049501796
Quartile 1	2
Quartile 2	3
Quartile 3	4
Quartile 4	9
IQR	2
Batas Bawah Outlier	-1
Batas Atas Outlier	7

CARPARKS

- The Carparks column has a wide range. The number of rooms on the property ranges from 1 to 28.
- We can identify 2 as the central tendency in representing the character of Carparks Property. Because the Mean, Median, and Mode are all in close proximity (2),
- Carparks skewness is **symmetrical** (Mean = Median).
- Outliers will not be removed because this is the Descriptive Analytics result after removing outliers in the Data Cleaning Process.

Car Parks	
Count	4267
Minimum	1
Maximum	28
Mean	2.1
Median	2
Mode	2
Sum	9078
Range	27
Sample Variance	1.415059873
Standard Deviation	1.189562892
Standard Error	0.01821067679
Skewness	5.192307037
Kurtosis	62.2412155
Coefficient of Variation	0.559139112
Confidence Level(95%)	0.033685652
Quartile 1	2
Quartile 2	2
Quartile 3	2
Quartile 4	28
IQR	0
Batas Bawah Outlier	2
Batas Atas Outlier	2

EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

Given that prices below Q1 are too low, we can concentrate more on directing potential buyers to properties with price ranges between Q1-Q2 and Q3-Q4 in order to turn a profit after doing Descriptive Statistics.

Properties between Q1 and Q2 will be referred to as Affordable Properties, while properties between Q3 and Q4 will be referred to as Luxury Properties.

Quartile 1	676775
Quartile 2	1200000
Quartile 3	1990000
Quartile 4	5070000

Affordable

Location	Count
Mont Kiara	106
Bukit Jalil	100
KLCC	97
Cheras	89
Dutamas	75
Sentul	67
Kepong	67
Desa Park City	57
Jalan Klang Lama (Old Klang Road)	52

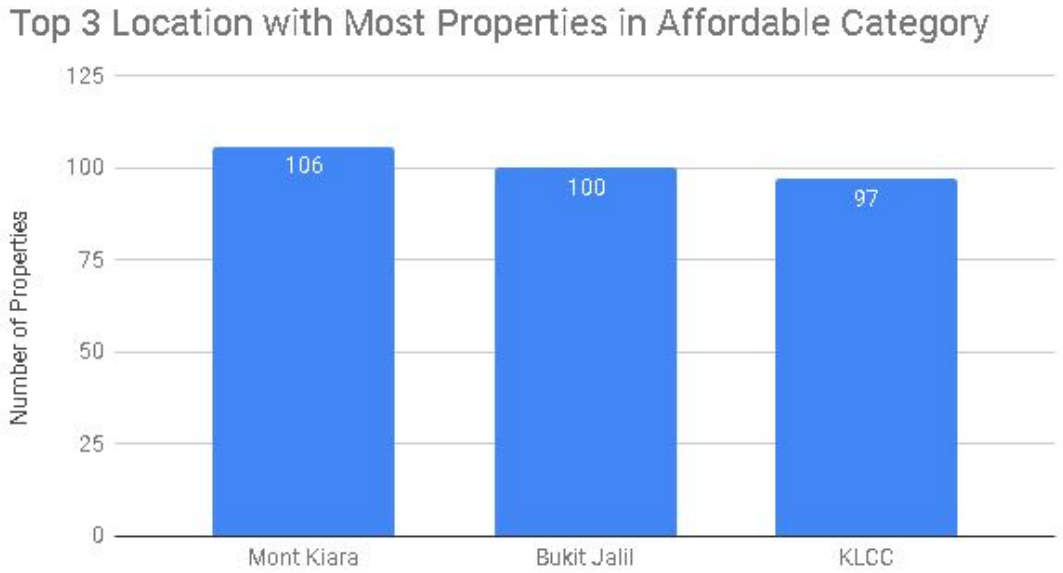
Luxury

Location	Count
Mont Kiara	279
KLCC	248
Desa Park City	136
Sri Hartamas	70
Damansara Heights	47
Bangsar	31
Ampang Hilir	25
Bukit Jalil	25
City Centre	25

CHARACTERISTIC OF AFFORDABLE PROPERTIES

Price between RM 1,990,000 to RM 5,070,000 (Q3-Q4)

Selling the **Top 3 Locations** with the **Most Affordably Properties** can be our first priority.



CHARACTERISTIC OF AFFORDABLE PROPERTIES

Location	Avg. Price	Avg. Rooms	Avg. Bathrooms	Avg. Car Parks	Avg. Size	Most Property Type	Most Furnishing
Mont Kiara	938,622	3	2	2	1,235	Condominium	Fully Furnished
Bukit Jalil	927,590	4	3	2	1,462	Condominium	Partly Furnished
KLCC	998,718	2	2	2	874	Serviced Residences	Fully Furnished

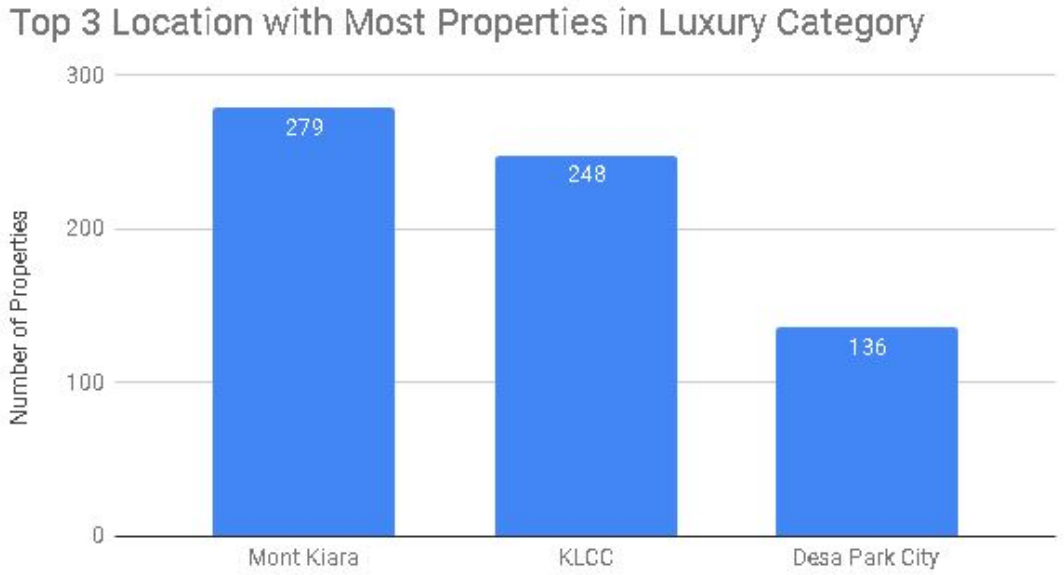
Insights :

- **Mount Kiara** has the **most properties available** in the Affordable Properties category, while **KLCC** has the **fewest**.
- **KLCC** has the **highest average price** of any property in the Affordable Properties category, while **Bukit Jalil** has the **lowest**.
- This category of **rooms** has a **wide range of variations**.
- Most of the properties has **2 Bathrooms and 2 Car Parks**.
- Except in **KLCC**, the majority of properties in this category are **Condominiums**.
- Except in **Bukti Jalil**, the **majority** of the properties in this category are **Fully Furnished.**

CHARACTERISTIC OF LUXURY PROPERTIES

Price between RM 676,775 to RM 1,200,000 (Q1-Q2)

Selling the **Top 3 Locations** with the **Most Luxury Properties** can be our priority.



CHARACTERISTIC OF LUXURY PROPERTIES

Location	Avg. Price	Avg. Rooms	Avg. Bathrooms	Avg. Car Parks	Avg. Size	Most Property Type	Most Furnishing
Mont Kiara	2,881,037	5	5	2	3,306	Condominium	Partly Furnished
KLCC	2,984,591	3	4	2	2,305	Serviced Residence	Partly Furnished
Desa Park City	2,865,963	5	5	2	2,716	3-sty Terrace/Link House	Partly Furnished

Insights :

- **Mount Kiara** has the **most properties available** in the Luxury Properties category, while **Desa Park City** has the **fewest**.
- **KLCC** has the **highest average price** of any property in the Luxury Properties category, while **Mount Kiara** has the **lowest**.
- Except in **KLCC,** the majority of the properties has **5 Rooms, 5 Bathrooms and 2 Car Parks**.
- This category contains **various Property Types**.
- In this category, the **majority** of the properties are **Partially Furnished**.

RECOMMENDATION

- If buyers want to live in a **small house** for a **lower price**, they can choose the **Affordable Property**.
- **Bukit Jalil** is an option for buyers who want to live in a l**ower-priced property**. However, the majority of the property is **Partially Furnished**.
- If buyers want to live in a **more expensive, fully furnished property**, they can choose **Mount Kiara**.

- If buyers want to live in a **large house**, they can choose **Luxury Properties**, which are **more expensive** than Affordable Properties.
- Buyers can choose what **Type of Property** they want or where the **Location** they want to live because there is **so much variety in Luxury Properties**.

STATISTICAL MEASUREMENT

(using Desa Park City dataset)

CORRELATION

	Price	*Rooms*	*Bathrooms*	*Car Parks*	*Size*
Price	1				
Rooms	0.72	1			
Bathrooms	0.77	0.82	1		
Car Parks	0.63	0.63	0.65	1	
Size	**0.80**	0.67	0.71	0.53	1

- **Size** has the strongest positive correlation with **Price**.
- **Car Parks** has the lowest positive correlation with **Price**.
- This implies that the **larger the Size** of the Property, the **higher the Price** of the Property.

REGRESSION : MODEL 1

	Price	Rooms	Bathrooms	Car Parks	Size
Price	1				
Rooms	0.72	1			
Bathrooms	0.77	0.82	1		
Car Parks	0.63	0.63	0.65	1	
Size	0.80	0.67	0.71	0.53	1

Significance F
0

	Coefficients	P-value
Intercept	0	#N/A
Rooms	36314.99144	**0.4092907952**
Bathrooms	169673.9105	0.00017473328
Car Parks	121598.7239	0.01700308031
Size	456.7218771	0

Regression Statistics	
Multiple R	0.9772869799
R Square	0.955089841
Adjusted R Square	**0.954315528**
Standard Error	438406.1022
Observations	177

- The **correlation** between **Bathrooms** and **Rooms** is **high**, indicating **multicollinearity**.
- **Significance F** value is **less than the alpha threshold value (5%)**, indicating that it can represent all existing data.
- **Size, Bathrooms, and Car Parks** all have **significant coefficients (P-value 5%)**, whereas **Rooms do not (P-value > 5%)**. We must run the Regression Model once more.
- The **adjusted R Square result of 95%** indicates that we have a good Regression Model. The dependent can be described by the independent variables.

REGRESSION : MODEL 2

	Price	Bathrooms	Car Parks	Size
Price	1			
Bathrooms	0.77	1		
Car Parks	0.63	0.65	1	
Size	0.80	0.71	0.53	1

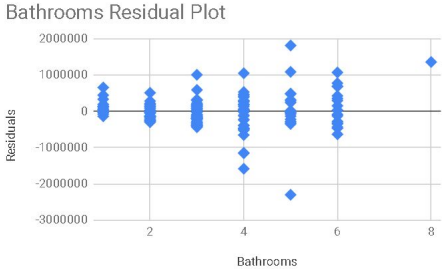
Significance F
0

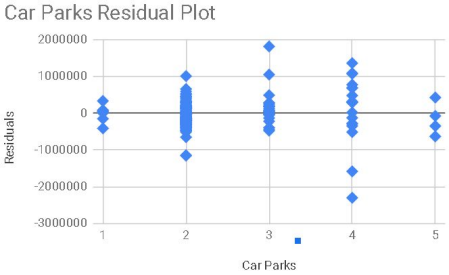
	Coefficients	P-value
Intercept	0	#N/A
Bathrooms	190018.9719	0.00000062515
Car Parks	139184.7749	0.00269311940
Size	469.5544291	**0**

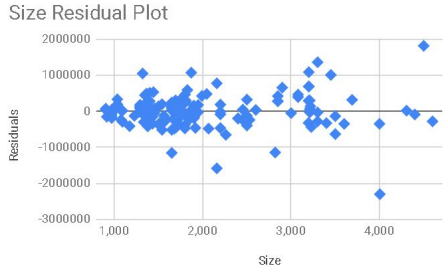
Regression Statistics	
Multiple R	0.9772869799
R Square	0.955089841
Adjusted R Square	**0.954315528**
Standard Error	438406.1022
Observations	177

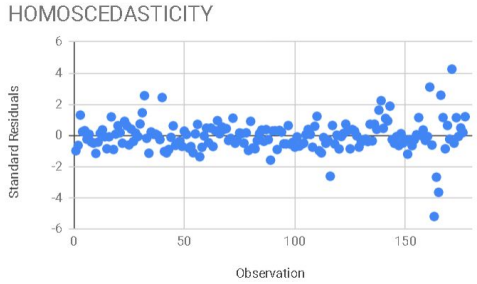
- The **correlation** between **Bathrooms** and **Rooms** is **high**, indicating **multicollinearity**.
- **Significance F** value is **less than the alpha threshold value (5%)**, indicating that it can represent all existing data.
- **All independent variables** have **significant coefficients (P-value 5%)**. We can use all variables to predict prices.
- The **adjusted R Square result of 95%** indicates that we have a good Regression Model. The dependent can be described by the independent variables.

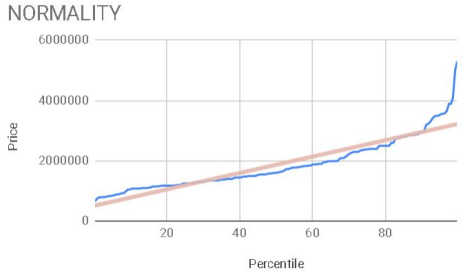
HOMOSCEDASTICITY & NORMALITY











- The Residual Plot appears to be dispersed and does not approach a consistent pattern for each observation point.
- Because we are using cross section data rather than time series data, the data will be non-autocorrelational.
- The Normal Probability Plot demonstrates that the residuals (observed) distribution fits the Z-Score (theoretical) distribution relatively well.

PRICE PREDICTION

	Coefficients
Intercept	0
Bathrooms	**190018.9719**
Car Parks	**139184.7749**
Size	**469.5544291**

We can predict property prices in Desa Park City using the results of the regression model we created with this formula:

Price (RM) = 0(# of Rooms) + 190018.9719(# of Bathrooms) + 139184.7749(# of Car Parks) + 469.5544291(Size)

Assume there is one loyal buyer who is looking for a property in Desa Park CIty with 3 rooms, 4 bathrooms, 3 car parks and 2200 sqft. The price recommendation is as follows:

Price (RM) = (0*4) + (190018.9719*4) + (139184.7749*3) + (469.5544291*2200)
0 + 760075.885 + 417554.325 + 1033019.74
RM 2,210,590.95

THANKS!

Feel free to comment below, DMs or reach me at riadhigodjay@gmail.com for any feedbacks or suggenstions