

PROPERTY LISTING

in Kuala Lumpur

Statistics and Spreadsheet

Study case

Awal Rais S

Week 2 Intermediate Assignment

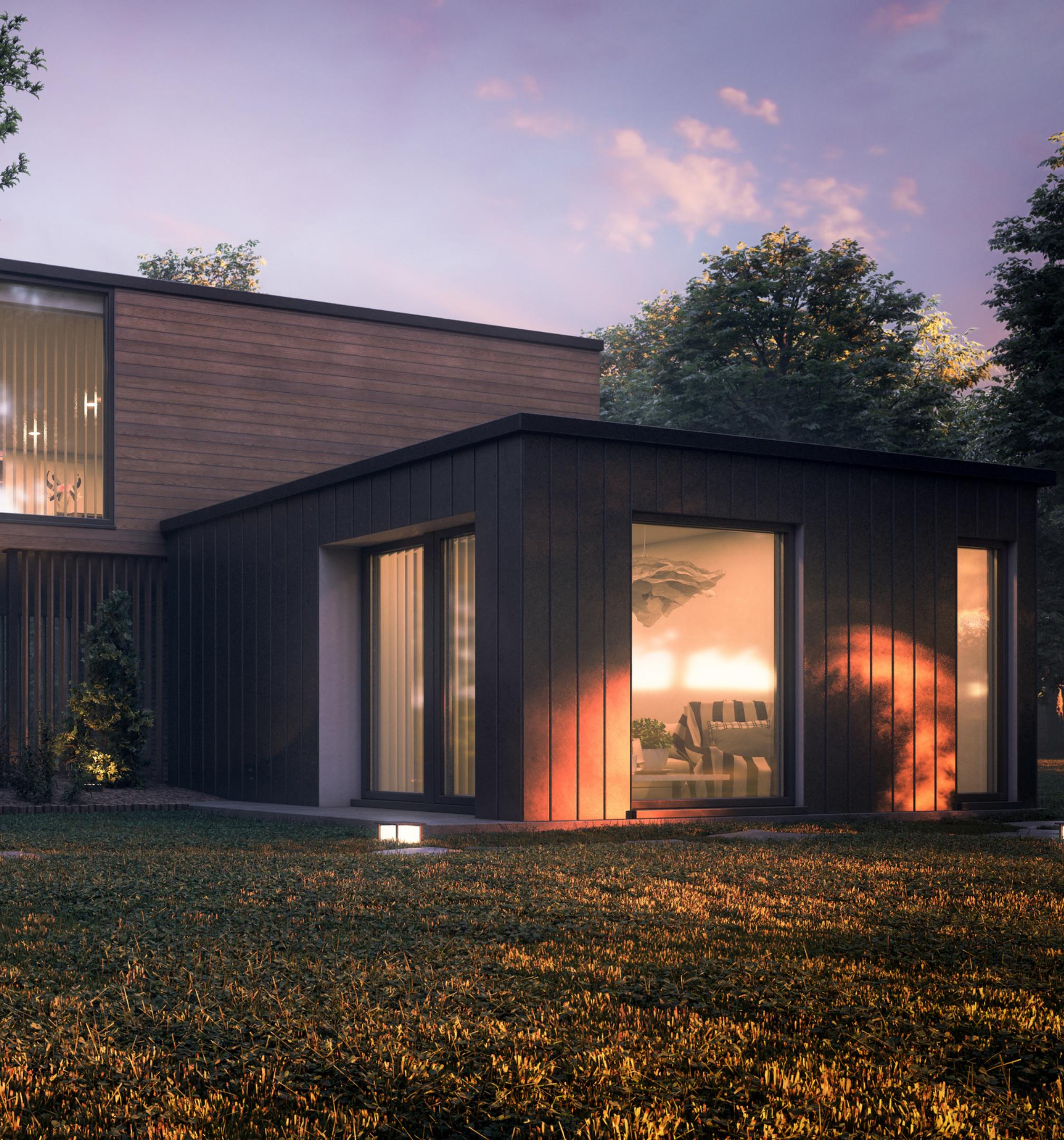




TABLE OF CONTENT

BACKGROUND BUSINESS

DATASET OVERVIEW

DATA & STATISTICAL PROCESS

INSIGHT AND RECOMMENDATIONS

BACKGROUND BUSINESS



Sofie Residence is a property listing company in Malaysia. The company mission is to provide various available choices of property to their users which providing property listing around Kuala Lumpur and other big cities in Malaysia. This company want to maximize the profit through joint-profit sharing of 20%. It means that the highest-priced property brings the highest percentage of revenue to the company. As a data analyst I need to deliver insights for the company which can help achieve that goals.



DATASET OVERVIEW

This dataset is mainly consist with **5000 dataset of Property Listing in Kuala Lumpur.**

The information is mainly about :

- Location
- Price
- Rooms
- Bathrooms
- Car Parks
- Property types
- Property character
- Size
- Furnishing

Dataset link : [dataset and spreadsheet link](#)



DATA & STATISTICAL PROCESS



✓ Data cleaning

Data cleaning process including removal outliers

✓ Descriptive Analysis

Explaining about the result of descriptive statistics

✓ Exploratory Data Analysis

Explaining about EDA and characteristic for each type property

✓ Correlation & Regression

Explaining about correlation between dependent and independent variables, price prediction for suggested property

DATA CLEANING



Data
cleaning

: This step includes following step :
1. Removal irrelevant values
2. Removal of Duplicates
3. Handling missing Value
4. Converting Data type
5. Removal Outliers

***removal tolerance : 15%**

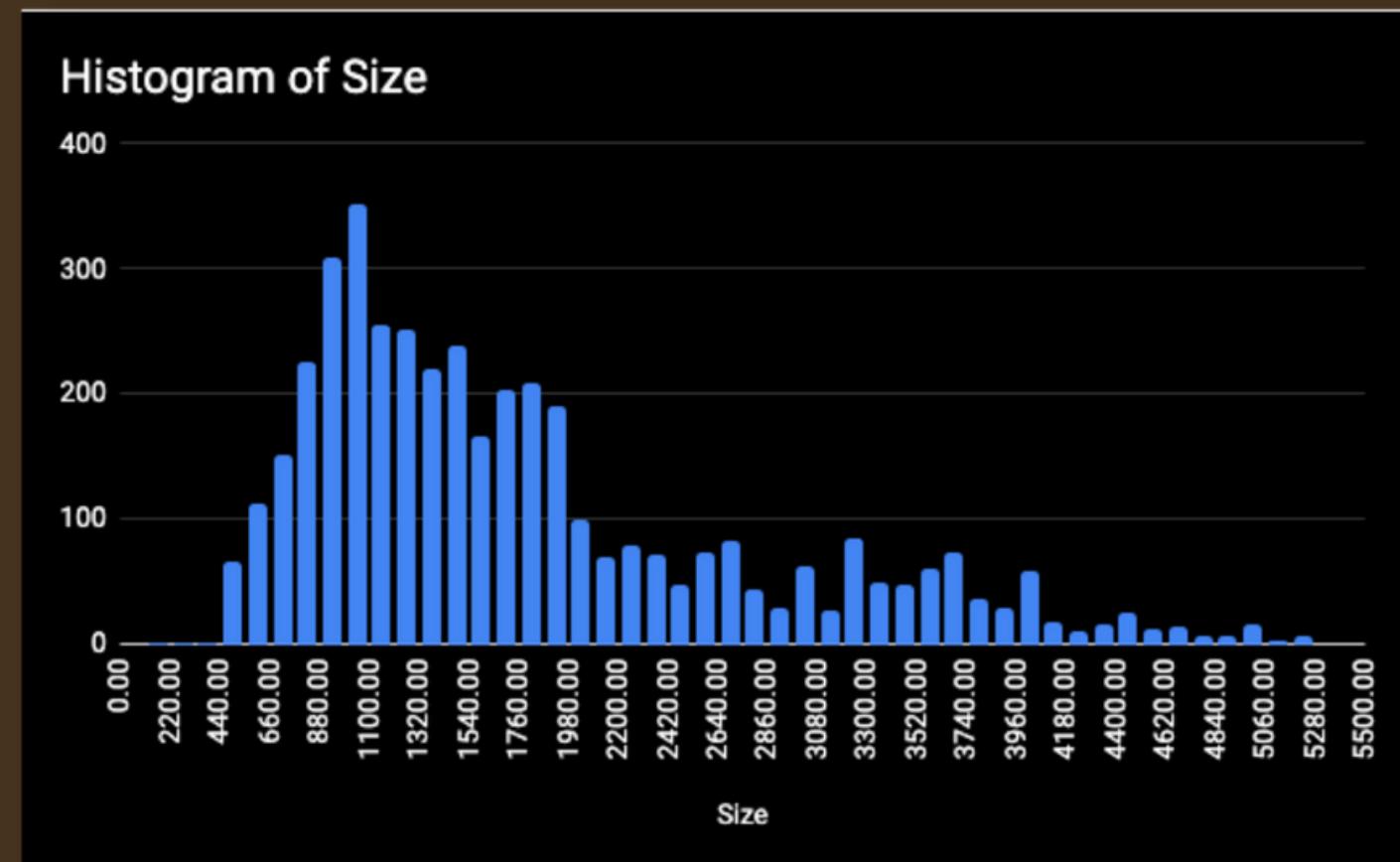
Before data cleaning : 5000 rows
After data cleaning. : 4230 rows



**Property
Listing**

TOTAL : 770 data removed including outliers

DESCRIPTIVE ANALYSIS



Mean : 2044849.6

Median : 1300000

Mode : 1200000

Price has **positive skewness** (Mean > Median > Mode)

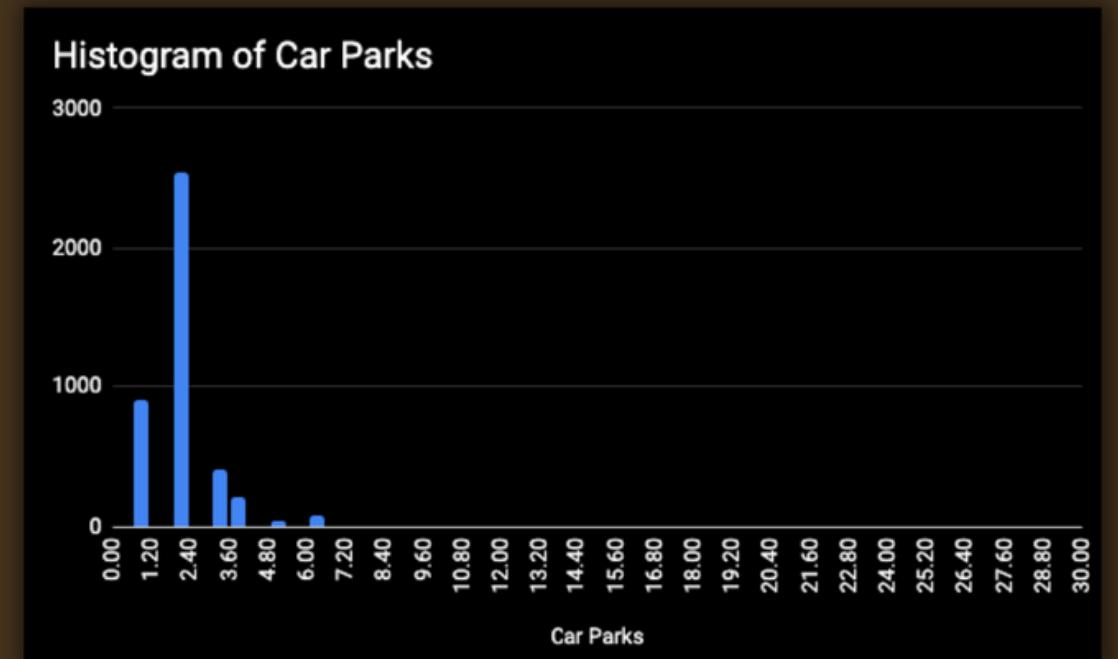
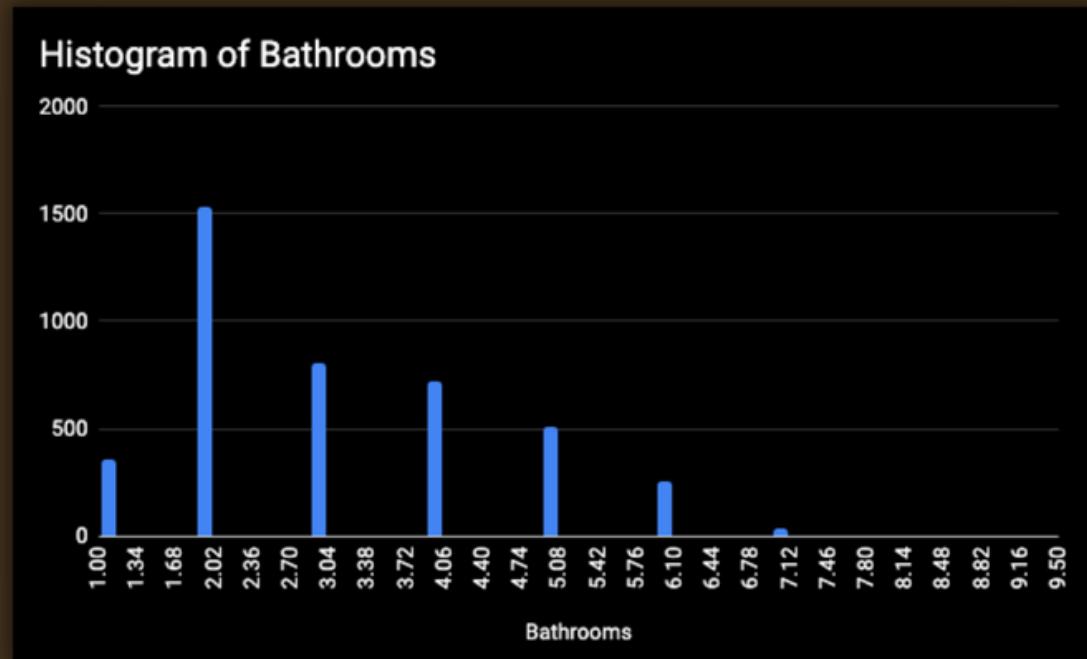
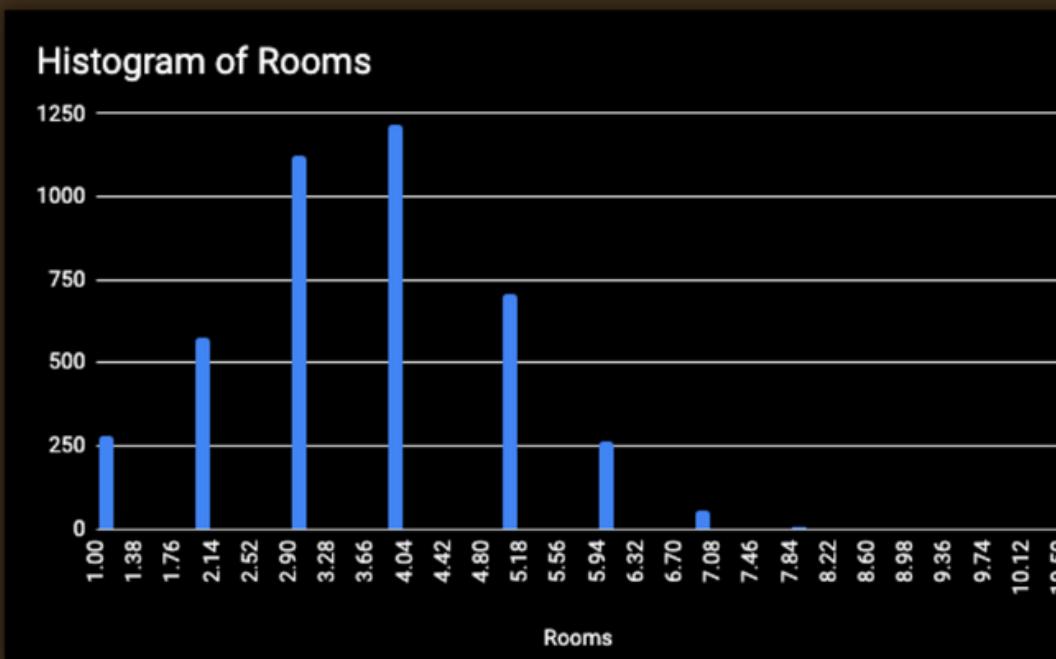
Mean : 2416.4

Median : 1626

Mode : 1650

Size has **positive skewness** (Mean > Median > Mode)

DESCRIPTIVE ANALYSIS



Mean : 4.00

Median : 4.00

Mode : 4.00

Rooms has **symmetrical distribution**
(Mean = Median = Mode)

Mean : 3.38

Median : 3.00

Mode : 2.00

Bathroom has **positive skewness** (Mean > Median > Mode)

Mean : 2.15

Median : 2.00

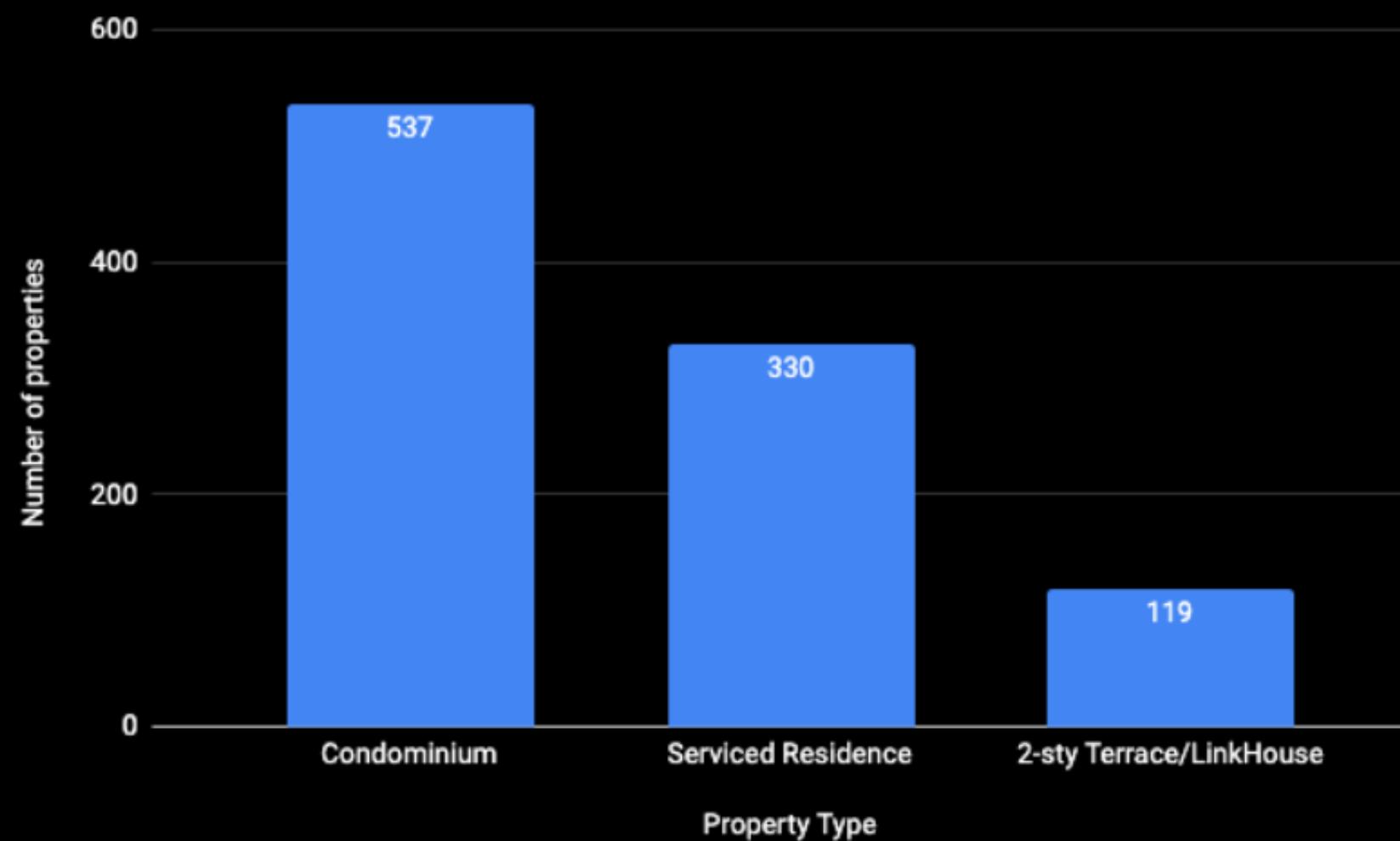
Mode : 2.00

Car park has slightly **positive skewness**
(Mean > Median > Mode)

EXPLORATORY DATA ANALYSIS



Top 3 Affordable based on Property type



PROPERTIES WITH AFFORDABLE PRICE

Based on property type, **Condominium** has the highest number of properties, meanwhile **2-sty Terrace/Link House** has the lowest numbers of affordable properties.



Property
Listing



**HOW IS THE MAIN
CHARACTERISTIC OF
TOP 3 WITH
AFFORDABLE PRICE ?**

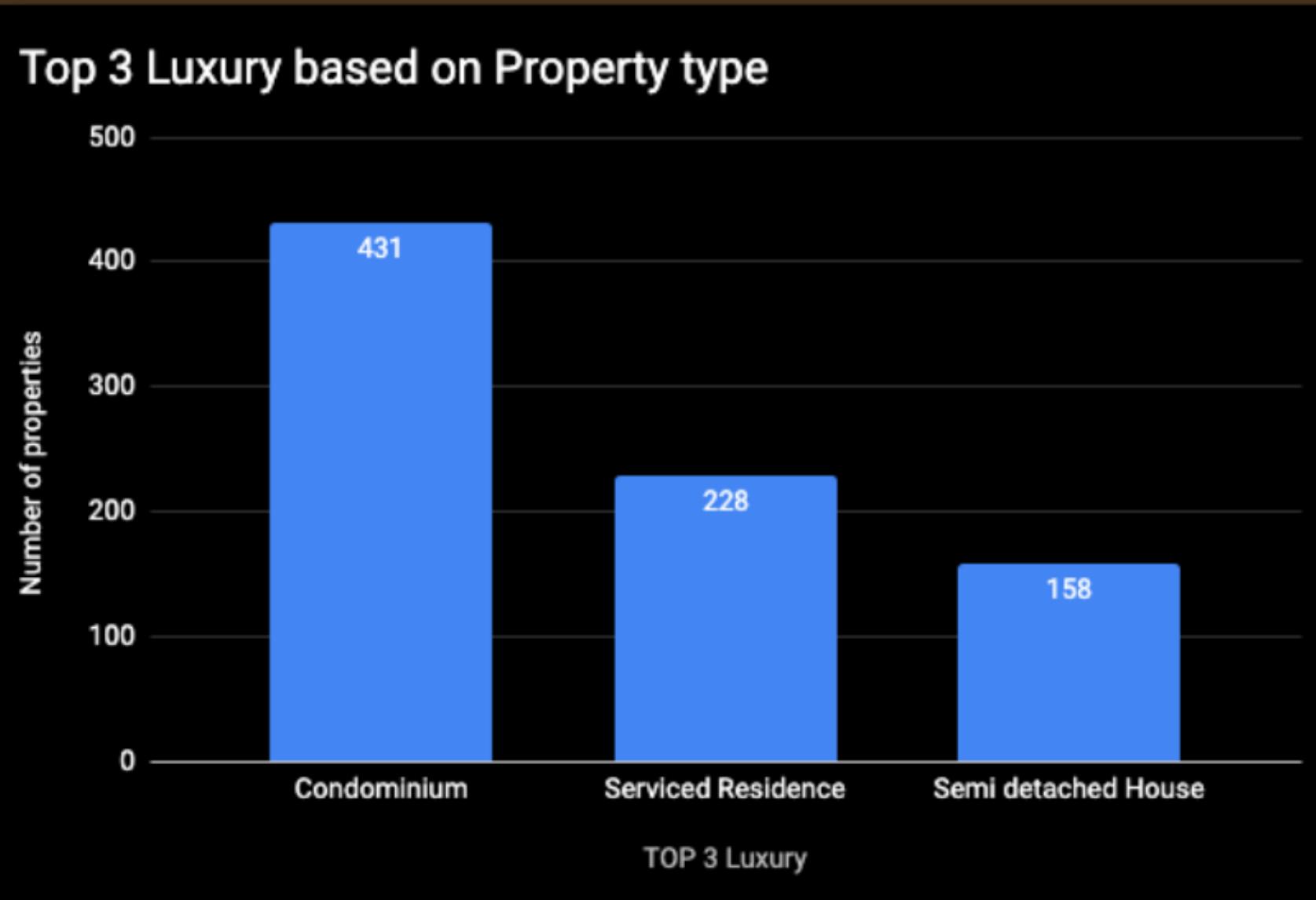
INSIGHT & CHARACTERISTICS



Property type	Avg Room	Avg Car Parks	Avg Bathrooms	Avg Size	Avg Price	Most Furnishing
Condominium	3	2	3	1399	916,876	Partly furnished
Serviced Residence	3	2	3	1,001	900,024	Furnished
2-sty Terrace/LinkHouse	3	2	3	1,756	867,809	Partly furnished

- Condominium has the **highest average price** among other properties. Meanwhile 2-sty Terrace/LinkHouse has **the lowest average price**.
- Most of properties has **3 rooms, 2 Car Parks, and 3 Bathrooms** in this category.
- Most of properties are **partly furnished**, except serviced residence.
- **2-sty Terrace / Link House** has **biggest average size** with **lowest price** compared to others.

EXPLORATORY DATA ANALYSIS



PROPERTIES WITH LUXURY PRICE

Based on property type, **Condominium** has the highest number of properties, meanwhile **Semi detached house** has the lowest numbers of Luxury properties.



**HOW IS THE MAIN
CHARACTERISTIC
OF TOP 3 WITH
LUXURY PRICE ?**

INSIGHT & CHARACTERISTICS



Property type	Room	Car Parks	Bathrooms	Avg Size	Avg Price	Most Furnishing
Condominium	5	3	2	3,006	2,961,519	Partly Furnished
Serviced Residence	5	2	5	3,012	2,968,858	Partly Furnished
Semi detached House	5	2	5	3,017	2,973,453	Partly Furnished

- Semi detached House has the **highest average price** among other properties. Meanwhile Condominium has **the lowest average price**.
- Most of properties has **5 rooms, 2 Car Parks, and 5 Bathrooms** in this category.
- All of properties are **partly furnished** and has slightly different in **average size**.
- With a slightly different of average size, **Condominium** is more worth to get rather than others, with more **bigger car parks** and **lowest average price**.

SO, WHICH ONE YOU SHOULD GET IT NOW?

Affordable or take it to another level
with a luxury one?

Here's Our Recommendations !



- With a simple minimalistic size, **Serviced Residence** would be the best options for you with Furnished building already. Also with affordable price, **Serviced Residence** with Fully Furnished **is our best seller** among other in affordable category
- Bigger, more space and Luxury. You can do everything in one place. No worry! we have the best choice for you. **Condominium** would be fit to you, if you want a bigger workplace, a bigger residence with a bunch of luxury features. Spent more to get more.

Check assumption test

- Non - Multicollinearity
- Homoscedasity
- Non - autocorrelation

Correlation and Regression

	Price	Room	Bathroom	Car Park	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 highest correlation to price** (strong positive correlation)
- **Car Park has the lowest correlation to price** (low positive correlation)

Summary Output of 1st Model Regression

Regression Statistics	
Multiple R	0.9773774868
R Square	0.9552667518
Adjusted R Square	0.954232457
Standard Error	438804.5118
Observations	177

- **Adjusted R-Square** is 95% means that we have a good model regression

	df	SS	MS	F	Significance F
Regression	4	711348629867558	177837157466890	923.5923759	0
Residual	173	33311046132442	192549399609		
Total	177	744659676000000			

- **Significance F** is less than 5%, it means our regression model represent all variables.

Summary Output of 1st Model Regression

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0	#N/A	#N/A	#N/A
Room	36314.991	43903.69057	0.8271512252	0.4092907952
Bathroom	169673.91	44227.07676	3.836426075	0.0001747332814
Car Park	121598.72	50456.95675	2.409949624	0.01700308031
Size	456.72187	58.03331114	7.869995148	0

- **p-value** on **Room** variables has more than 5%. Thus we need to **exclude Room** variable

Summary Output of 2nd Model Regression

Regression Statistics	
Multiple R	0.9772697201
R Square	0.9550561058
Adjusted R Square	0.9542767319
Standard Error	439632.9517
Observations	176



- **Adjusted R-Square** is 95% means that we still have a good model regression

	df	SS	MS	F	Significance F
Regression	3	710533832119621	236844610706540	1225.414553	0
Residual	173	33436943880379	193277132257		
Total	176	743970776000000			



- **Significance F** is less than 5%, it means our regression model represent all variables.

Summary Output of 1st Model Regression

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0	#N/A	#N/A	#N/A
Bathroom	190018.9719	36723.18234	5.174360167	0.000000625154703
Car Park	139184.7749	45717.33181	3.044464088	0.002693119403
Size	469.5544291	55.87038742	8.40435248	0

- **p-value** on **all variables** are **below 5%**. therefore we could use **all variables** for prediction

Price Prediction

	Coefficients	Standard Error	t Stat	P-value
Intercept	0	#N/A	#N/A	#N/A
Bathroom	190018.9719	36723.18234	5.174360167	0.000006251547035
Car Park	139184.7749	45717.33181	3.044464088	0.002693119403
Size	469.5544291	55.87038742	8.40435248	0

Based on the result of regression mode, our regression equation will be :

$$\text{Price} = 190018.9719 * \text{rooms} + 139184.7749 * \text{Car Park} + 469.5544291 * \text{Size}$$

Price Prediction

	Coefficients	Standard Error	t Stat	P-value
Intercept	0	#N/A	#N/A	#N/A
Bathroom	190018.9719	36723.18234	5.174360167	0.0000006251547035
Car Park	139184.7749	45717.33181	3.044464088	0.002693119403
Size	469.5544291	55.87038742	8.40435248	0

so, the price prediction for 3 Rooms, 4 Bathrooms, 3 Car Parks and 2200 sqft will be :

$$\text{Price} = (0 * 3) + (4 * 190018.9719) + (3 * 139184.7749) + (2200 * 469.5544291)$$

Price = RM 2,210,649.96

THANK YOU!

Let's connect !

 awalraiss@gmail.com

 /awalraiss

 awalraiss