

RENT REALITIES: WHAT AFFECTS PRICES, FUTURE TRENDS, AND SMART CHOICES

Group members:

Zhen Xue, Gue (33521352)

Chun Kei, Chow (33520771)

Sunaina, Rayaprol (33944091)

Hui Shawn, Lim (34017259)

Lecturer:

Dr. Ashish Dutt




OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

References

OVERVIEW

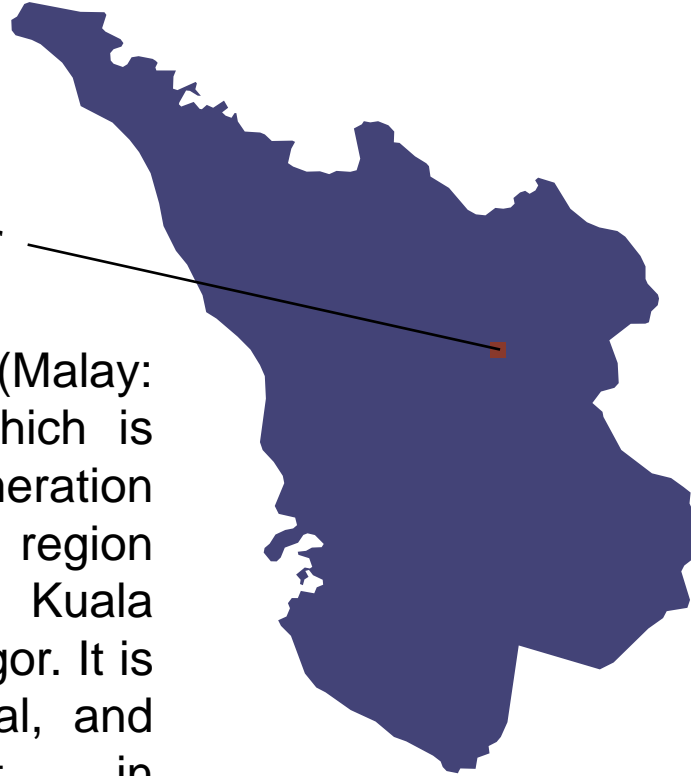
- 
- 1 Background**
 - 2 Research questions**
 - 3 Data pre-processing**
 - 4 Exploratory Data Analysis**
 - 5 Evaluation & Modelling**
 - 6 Discussion**
 - 7 Conclusion**

SECTION 1: **BACKGROUND**

SECTION 1: BACKGROUND

Kuala Lumpur

Klang Valley (Malay: *Lembah Klang*), which is an urban conglomeration in Malaysia, is a region that encompasses Kuala Lumpur, and Selangor. It is the cultural, political, and economic heart in Malaysia.



Where is Klang Valley?

Population and Situation

Benefit of get to know with the trend



Action

SECTION 1: BACKGROUND

2.311 M

people living in Klang Valley by 2024
(Worldometer, 2023)

Due to high population, significant urbanisation and high inflation rates in Klang Valley, housing prices in the cities are unpredictable. This necessitates that people from all walks of life closely monitor the housing rental market and its trends (Zainon et al., 2017).

Where is Klang Valley?

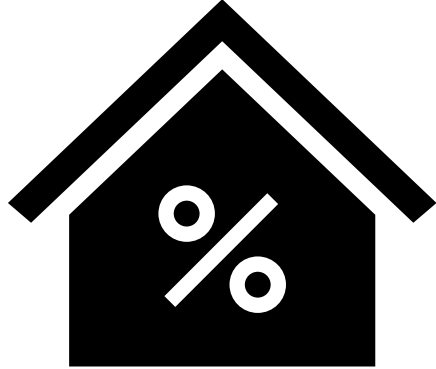
Population and Situation

Benefit of get to know with the trend

Action



SECTION 1: BACKGROUND



For renters, it will greatly help in budgeting;

For landlords, it will help in determining house rental price to attract and retain tenants.

Population and
Situation

**Benefit of get to know
with the trend**

Action



SECTION 1: BACKGROUND




Moreover, in this project, we analysed the data and making several models to predict the future trend of the rental price.

Benefit of get to know
with the trend

Action



OVERVIEW

- 
- 1 Background**
 - 2 Research questions**
 - 3 Data pre-processing**
 - 4 Exploratory Data Analysis**
 - 5 Evaluation & Modelling**
 - 6 Discussion**
 - 7 Conclusion**

SECTION 1: **BACKGROUND**

OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

SECTION 2: RESEARCH QUESTIONS



SECTION 2: RESEARCH QUESTIONS

01

What are the significant factors influences the rent for the property?

02

How to predict the movement of rental in future years based on previous trends?

03

How to find the most suitable rental area (or property name, type etc.) to be invested, or rented?

04

For property developers, where to potentially build a new housing area?

OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

SECTION 2: RESEARCH QUESTIONS

OVERVIEW

- 1 Background
- 2 Research questions
- 3 Data pre-processing
- 4 Exploratory Data Analysis
- 5 Evaluation & Modelling
- 6 Discussion
- 7 Conclusion

SECTION 3:

DATA

PREPROCESSING

SECTION 3: DATA PREPROCESSING



OVERVIEW OF OUR DATASET

The first 3 data of our dataset in `malaysia_house_rental_price_data.csv` is as follows:

	prop_name	completion_year	monthly_rent	property_type	rooms	parking	bathroom	size	furnished	facilities	additional_facilities	city	area
0	The Hipster @ Taman Desa	2022.0	4200.0	Condominium	5.0	2.0	6.0	1842.0	Fully Furnished	Minimart, Gymnasium, Security, Playground, Swi...	Air-Cond, Cooking Allowed, Washing Machine	Kuala Lumpur	Taman Desa
1	Segar Courts	NaN	2300.0	Condominium	3.0	1.0	2.0	1170.0	Partially Furnished	Playground, Parking, Barbeque area, Security, ...	Air-Cond, Cooking Allowed, Near KTM/LRT	Kuala Lumpur	Cheras
2	Pangsapuri Teratak Muhibbah 2	NaN	1000.0	Apartment	3.0	NaN	2.0	650.0	Fully Furnished	Minimart, Jogging Track, Lift, Swimming Pool	NaN	Kuala Lumpur	Taman Desa

Features:	Property name, Completion year, Property type, Rooms, Bathrooms, Parking slots, Size of property, Furnished condition, Facilities and Additional Facilities available, City, Area
Target feature:	Monthly rent

ISSUES IN OUR DATASET

Repetitive Information

Overlapping data in facilities and additional_facilities because they have been cumulatively entered as strings

Blank Cells

Numerous instances of missing values throughout the dataset.

Inconsistent City Names

Kuala Lumpur and Selangor have unwanted spaces in the entry names.

Not-referable monthly-rent and size Information

Some data seems unreliable. i.e., “rent: RM240000 in Sunway Vivaldi” and “size: 1 sqft in Kelana Sentral”.

Inconsistent data types

Some columns have suboptimal datatype. i.e., year (float64).

POST DATA CLEANING

Objectives

- No outliers
- No missing data
- Fix structural errors.
- Remove duplicate data



1

Merged facilities &
additional_facilities

2

Cleared
observations with
missing values

3

Corrected format of
city names

4

Converted all
variables to desired
datatypes

5

Took 95%
confidence interval
of required data

POST DATA CLEANING

Thus, we end up with:

	prop_name	completion_year	monthly_rent	property_type	rooms	parking	bathroom	size	furnished	facilities	city	area
3	Sentul Point Suite Apartment	2020	1700	Apartment	2	1	2	743	Partially Furnished	[Parking, Playground, Swimming Pool, Squash Co...	Kuala Lumpur	Sentul
7	Arte Plus Jalan Ampang	2018	1550	Service Residence	1	1	1	700	Fully Furnished	[Parking, Gymnasium, Playground, Security, Lif...	Kuala Lumpur	Ampang
8	Nova I	2014	1400	Apartment	2	1	1	750	Fully Furnished	[Playground, Security, Parking, Lift, Swimming...	Kuala Lumpur	Segambut

OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

SECTION 3:

DATA

PREPROCESSING

OVERVIEW

- 
- 1 Background
 - 2 Research questions
 - 3 Data pre-processing
 - 4 **Exploratory Data Analysis**
 - 5 Evaluation & Modelling
 - 6 Discussion
 - 7 Conclusion

SECTION 4:

EXPLORATORY DATA ANALYSIS

SECTION 4: EXPLORATORY DATA ANALYSIS



INVESTIGABLE FACTORS

01

**COMPLETION
YEAR**

02

FACILITIES

03

CITY & AREA

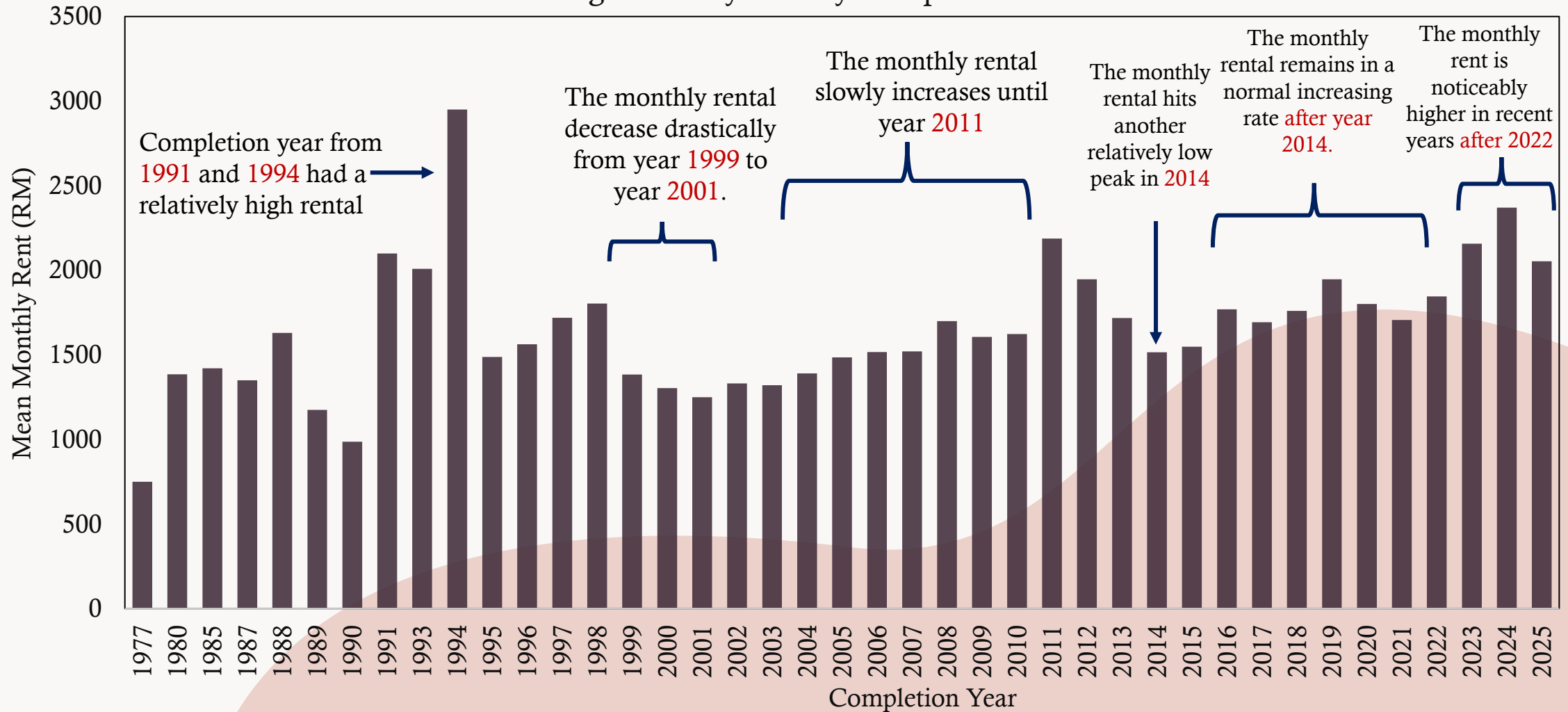
INVESTIGABLE FACTORS

01

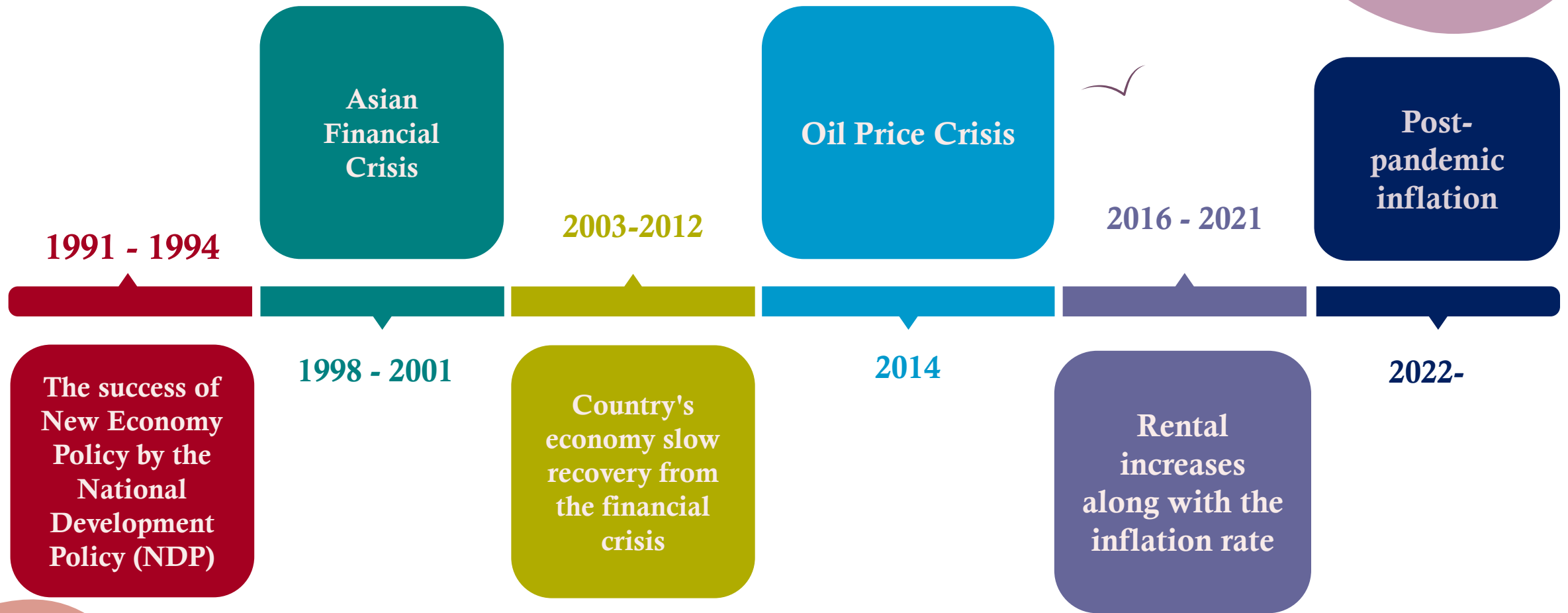
**COMPLETION
YEAR**

COMPLETION YEAR

Average Monthly Rent by Completion Year



COMPLETION YEAR



(Chan et al., 2018)

INVESTIGABLE FACTORS

02

FACILITIES

FACILITIES

Observations

General situation of monthly rent for rentals with:

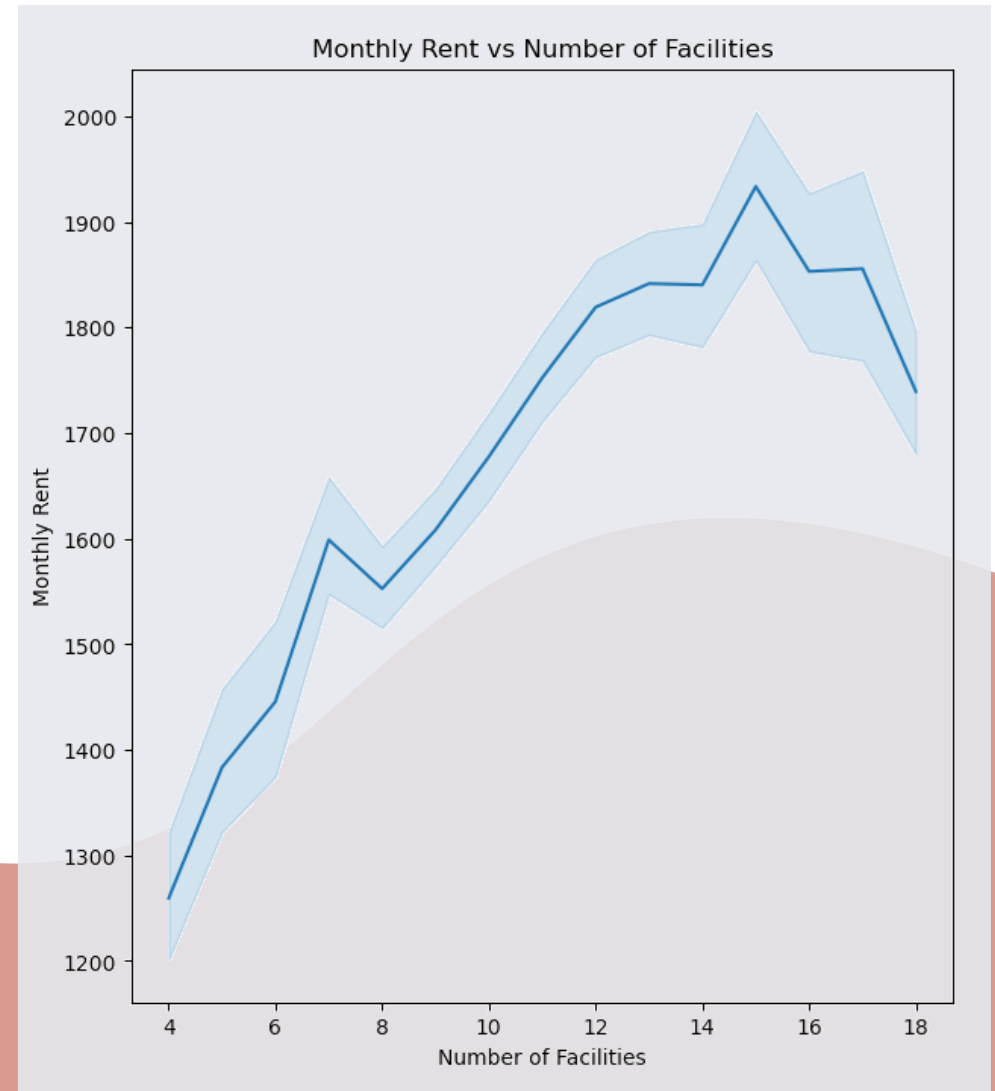
- Up to 15 facilities : Generally increases
- More than 15 facilities : Generally decreases

Why?

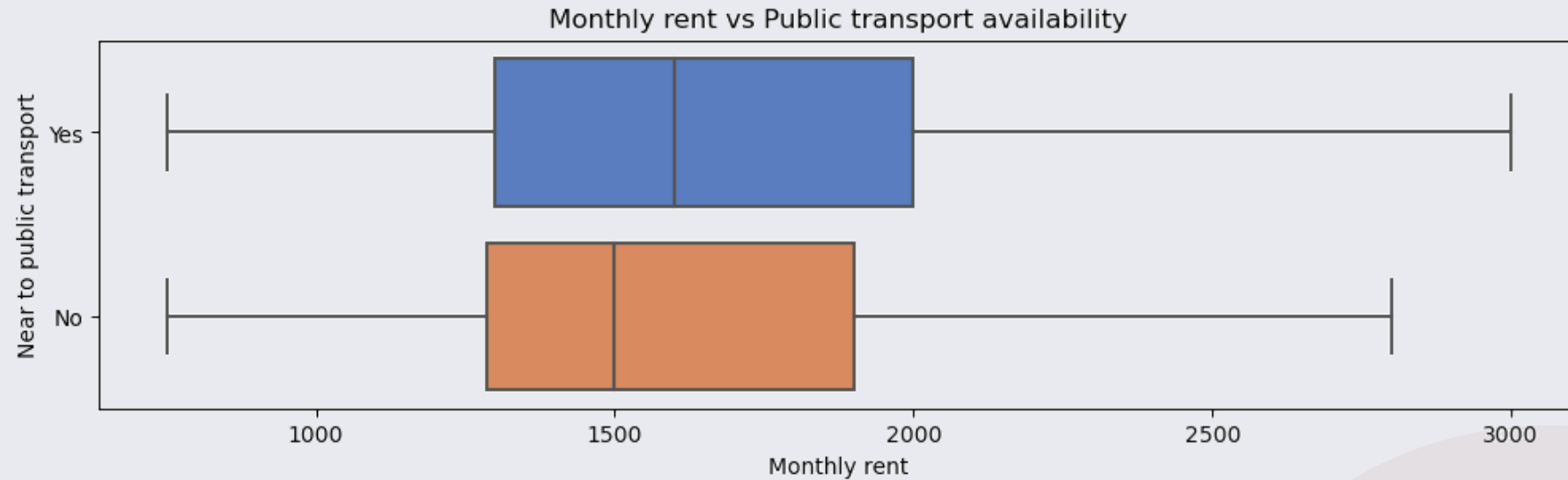
Similarities of rentals despite having >15 facilities:

- At far ends of public transport lines (Cyberjaya, Gombak)
- Located far away from high-paying job opportunities
- Located far away from most educational institutes, recreational outlets, and malls

So, the demand for housing is low in those areas.



RENT VS PUBLIC TRANSPORT AVAILABILITY



Observations

The overall monthly rent is higher for units that have proximity to public transport (KTM/LRT).

Why?

The demand is higher, because of:

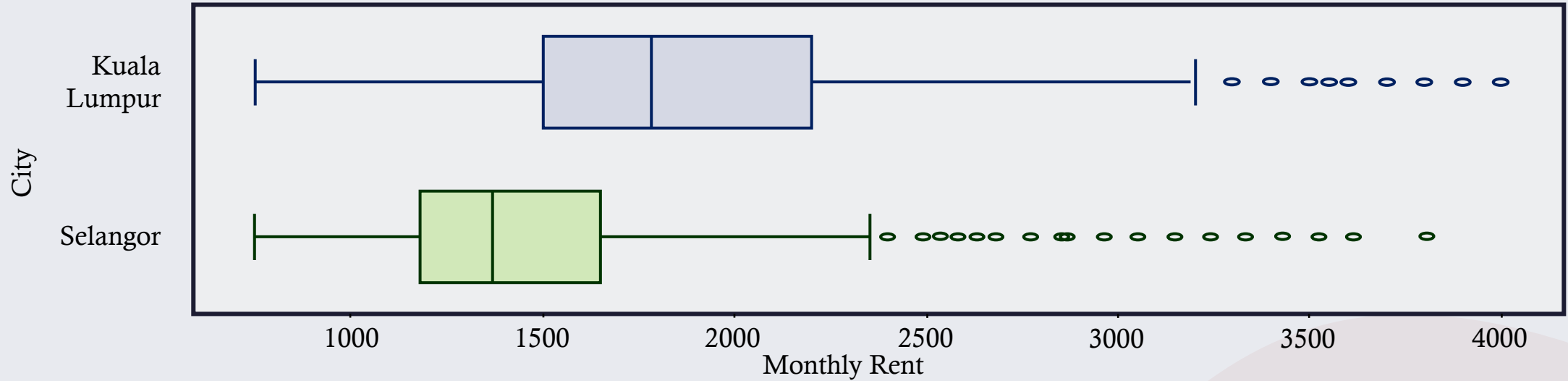
- Convenience
- Better accessibility
- Renters will not need to spend on personal vehicles
- Reduce carbon footprint (eco-friendly)

INVESTIGABLE FACTORS

03

CITY & AREA

CITY & AREA



What we see

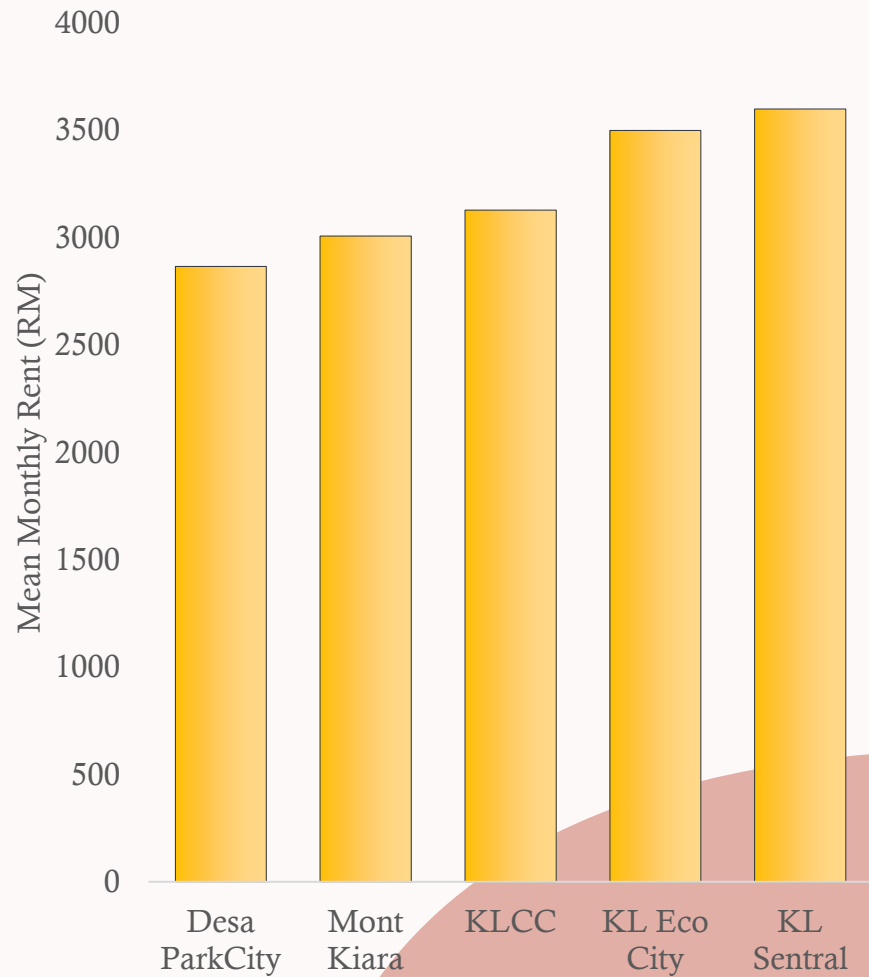
In overall, the monthly rent is higher in Kuala Lumpur than Selangor.

Why is it so?

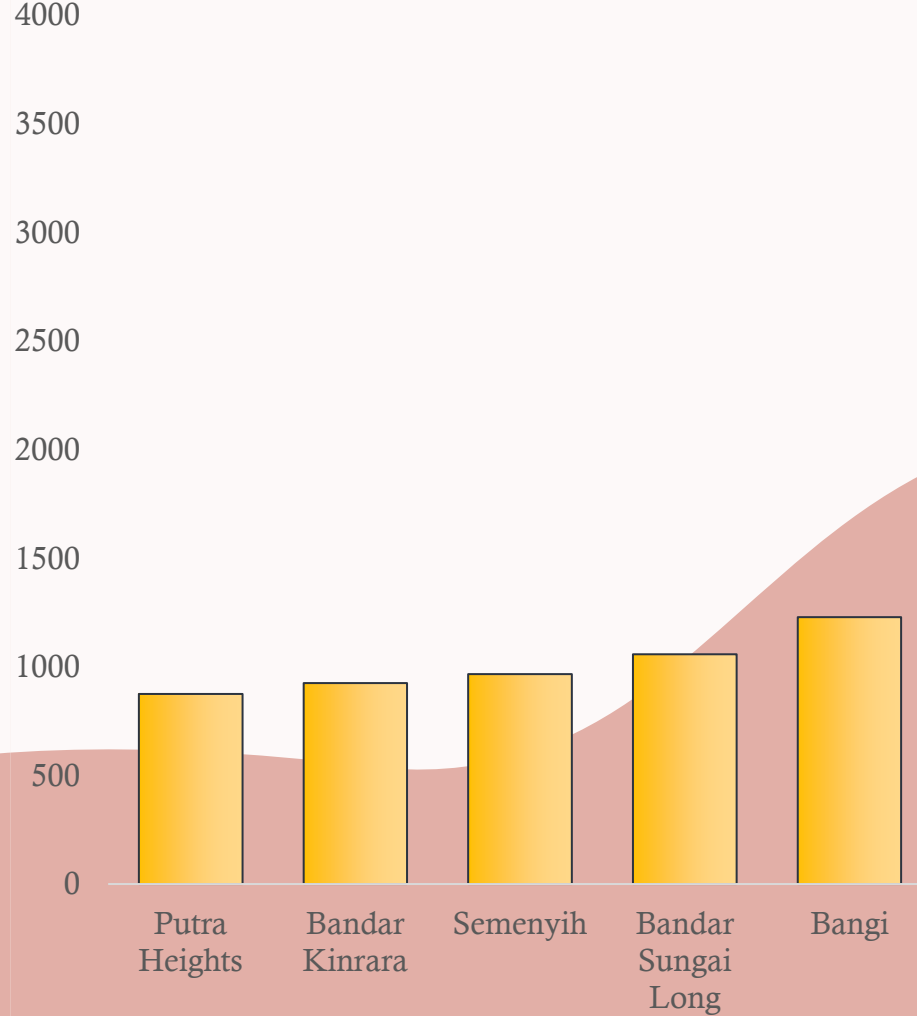
Kuala Lumpur is the economic centre of Malaysia, which consist of numerous high consumption area.

CITY & AREA

Mean Monthly Rent: Top 10 Areas



Mean Monthly Rent: Bottom 10 Areas



Observation

All top 10 areas are mostly from **Kuala Lumpur**.

All bottom 10 areas are mostly from **Selangor**.

Further Approach

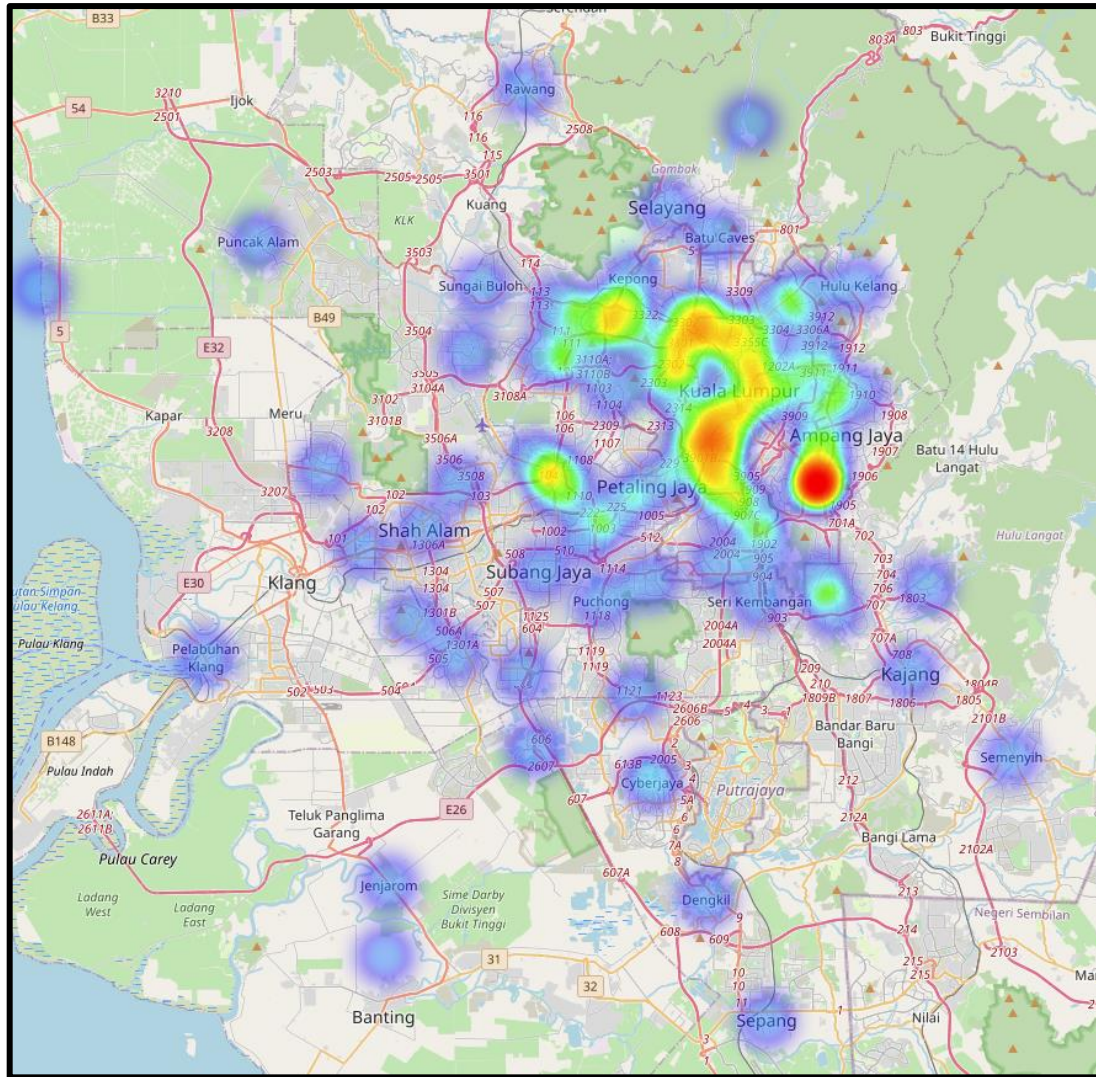
We extend our investigation by using service from `Nominatim`, which is a geolocator module from `geopy`.

What are we collecting?

- Longitude
- Latitude

Lastly, we build a heatmap based on the monthly rent.

CITY & AREA



What we see

We see that there is not just one cluster, but numerous ones, showing yellow or red (indicating high rental prices).

Why is it so?

This shows that the existence of a city (urbanisation) can bring an increase in rental prices and affect the areas near them, based on the distance.

For clearer pic:



<https://sites.google.com/student.monash.edu/preview-of-heatmap-hrp-area/home?pli=1&authuser=2>

Low (RM700)

High (RM4000)

CITY & AREA: GAUSSIAN MIXTURE MODEL

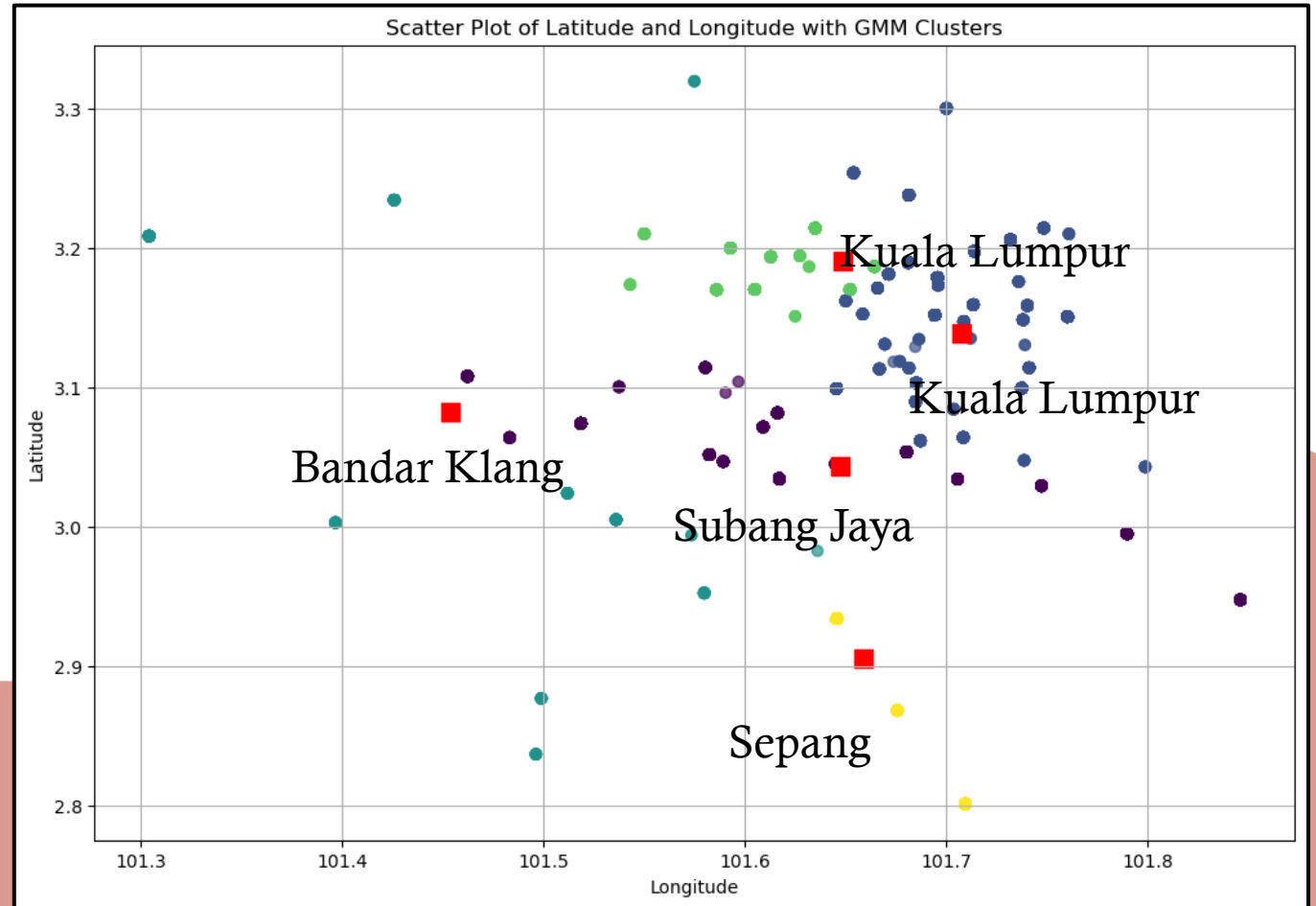
We made a Gaussian Mixture Model since clusters exist.

Why Gaussian Mixture Model?

- GMM can handle complex data distributions.
- GMM allows soft clustering, which a data point can belong to multiple clusters.

Approach

We set clusters of 5 and find each of the centres. Then, we calculated distance of every point to each centres in clusters.



OVERVIEW

- 
- 1 **Background**
 - 2 **Research questions**
 - 3 **Data pre-processing**
 - 4 **Exploratory Data Analysis**
 - 5 **Evaluation & Modelling**
 - 6 **Discussion**
 - 7 **Conclusion**

SECTION 4:

EXPLORATORY DATA ANALYSIS

OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

SECTION 5: **EVALUATING & MODELLING**

SECTION 5: EVALUATING & MODELLING



DATA CORRELATION AT A GLANCE



What we see

Evaluating the **Pearson Correlation** in heatmap, we see that each factor are weakly related as there are multiple factors that brings effect to the monthly rent.

Approach

We require at least 2 models consist of:

- Multiple Linear Regression Model
- Polynomial Regression Model

MODELLING

Approach for every model

We separated current data into:

70% training set

AND

30% testing set

We have built a Multiple Linear Regression Model.

Multiple Linear Regression Model: Best fit hyperplane

$$E(X) = 20.91x_1 - 170x_2 - 14.63x_3 + 122.71x_4 + 1.24x_5 + 21.40x_6 - 19.57x_7 - 41431.6$$

where:

x_1 : completion year

x_2 : rooms

x_3 : parking

x_4 : bathroom

x_5 : size

x_6 : facilities num

x_7 : distance to centre

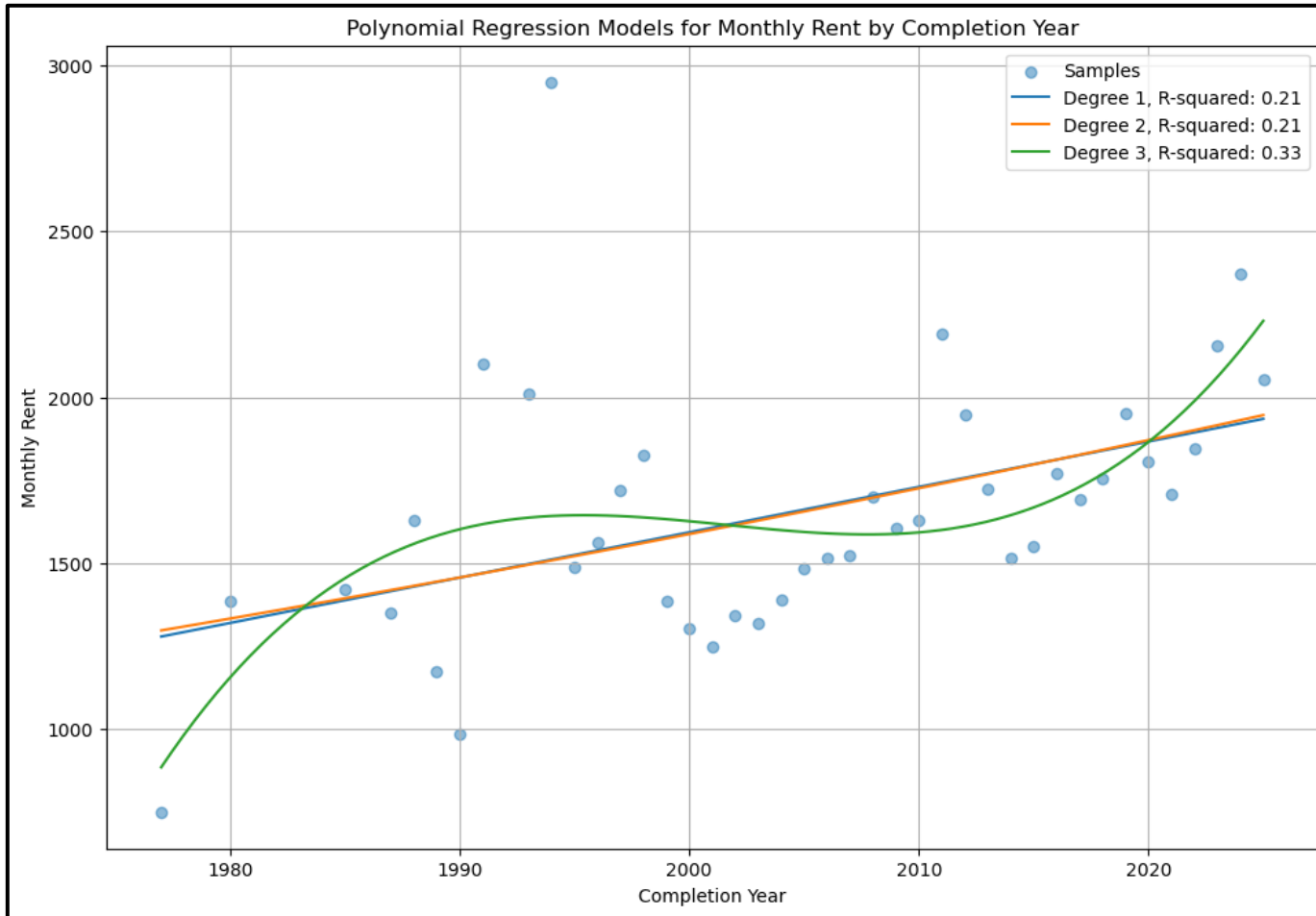
$E(X)$: expected monthly rent for that specific type of room

Testing score: 0.266

Testing score: 0.263

POLYNOMIAL REGRESSION MODEL OF MONTHLY RENT AND COMPLETION YEAR

To predict the future trend, we built a Polynomial Regression Model by taking mean for each year.



Degree 1 equation:

$$13.91x - 2.631 \times 10^4$$

Degree 2 equation:

$$0.5528x^2 - 2209x + 2.207 \times 10^6$$

Degree 3 equation:

$$0.0607x^3 - 364.58x^2 + 7.3 \times 10^5x - 4.9 \times 10^8$$

R-squared value

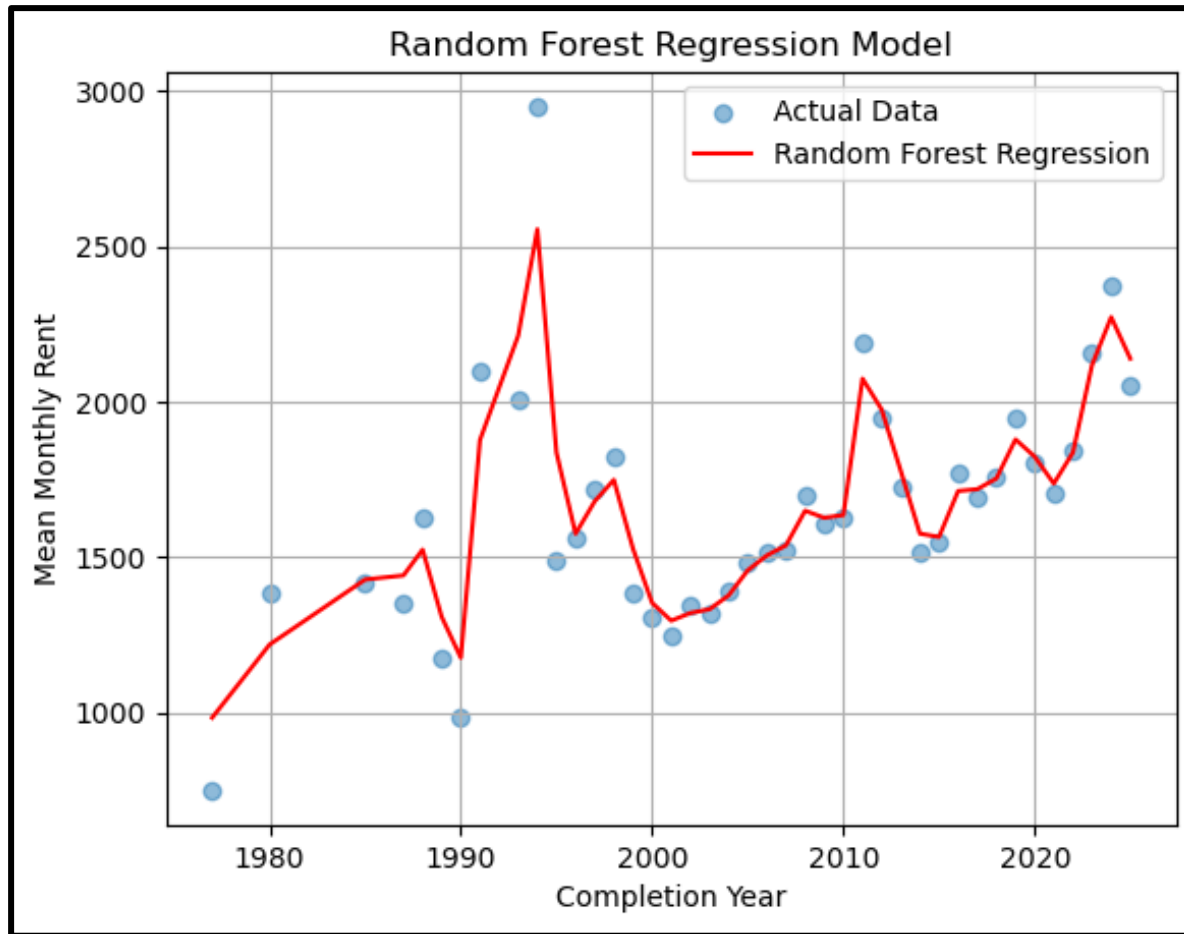
Degree 1: 0.21

Degree 2: 0.21

Degree 3: 0.33

RANDOM FOREST REGRESSOR MODEL OF MONTHLY RENT AND COMPLETION YEAR

This predictive model uses a combination of mean and decisions from trees or previous data (years) to predict the future data, where the final prediction is the average of all other previous predictions.



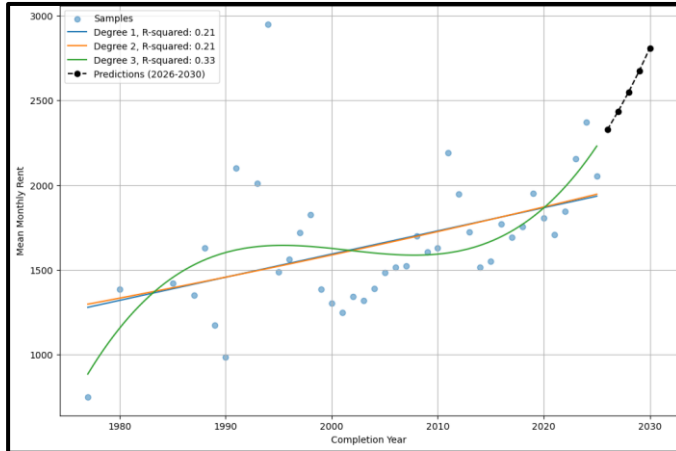
R-squared value: 0.8978

This model **overfitted**.

- This is because a random forest regressor model uses **average method** to improve the predictive model's accuracy
- This is coherent with the grouping of monthly rent by completion year using the mean.

RANDOM FOREST REGRESSOR MODEL VS POLYNOMIAL REGRESSION MODEL

Which model to use? This depends on the prediction of future monthly rents.

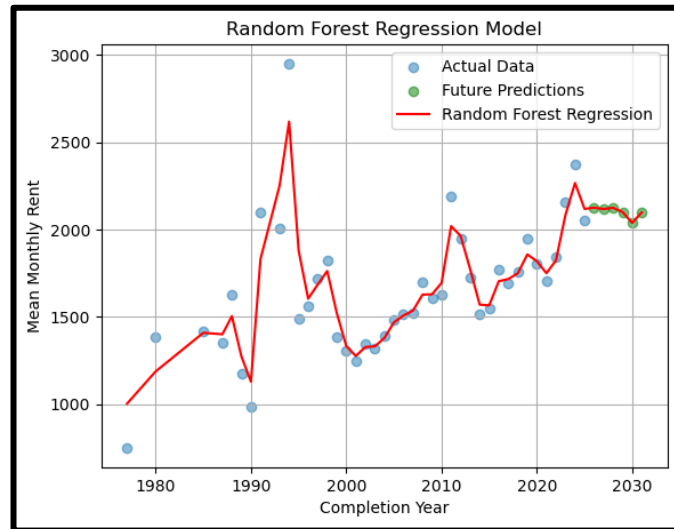


Polynomial Regression Model (power 3)

2026:\$2327.11
2027:\$2433.12
2028:\$2548.39
2029:\$2673.28

We only see increasing trend in this model

Polynomial regression model always shows an upward trend



Random Forest Regressor Model

2026:\$2192.27
2027:\$2142.86
2028:\$2256.07
2029:\$2144.28

We see a decreasing trend following by increasing trend in this model

Random forest regressor model makes decision based on previous trends

OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

SECTION 5: **EVALUATING & MODELLING**

OVERVIEW

- 1 **Background**
- 2 **Research questions**
- 3 **Data pre-processing**
- 4 **Exploratory Data Analysis**
- 5 **Evaluation & Modelling**
- 6 **Discussion**
- 7 **Conclusion**

SECTION 6: **DISCUSSION**

SECTION 6: DISCUSSION

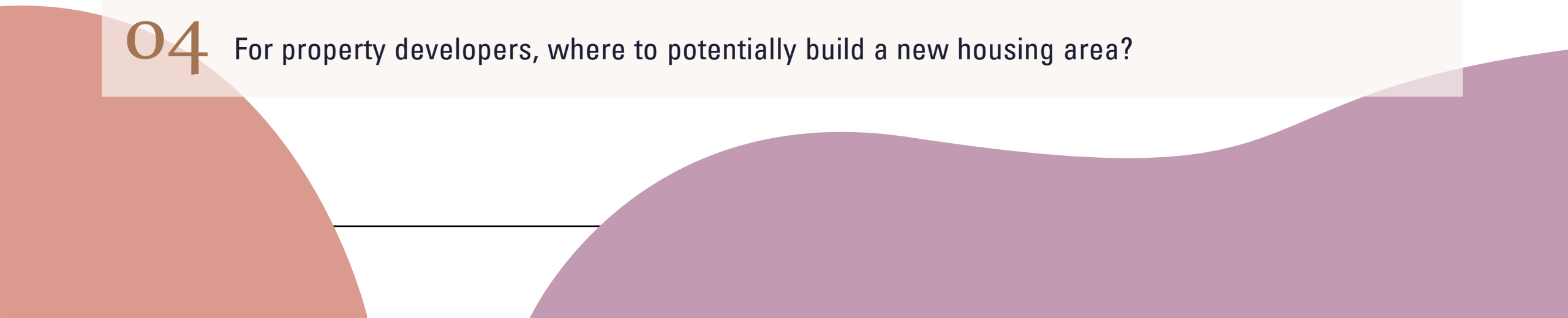
Recalling back questions from **Section 2**:

01 What are the significant factors influences the rent for the property?

02 How to predict the movement of rental in future years based on previous trends?

03 How to find the most suitable rental area (or property name, type etc.) to be invested, or rented?

04 For property developers, where to potentially build a new housing area?

Decorative wavy shapes in shades of red and purple at the bottom of the page.

SECTION 6: DISCUSSION

O1 What are the significant factors influences the rent for the property?

In this project, based on the result of f-test from multiple linear regression model, we found out the significant factors:

1. Completion_year
2. Rooms
3. Parking
4. Bathroom
5. Size
6. Facilities_num (Number of facilities)
7. Distance_to_centre (Distance of property to the centre of each urban)

SECTION 6: DISCUSSION

02 How to predict the movement of rental in future years based on previous trends?

We could put some input to the regression equation, that could give the overall evaluation on how would the rental fee goes in the future.

The regression equation:

$$\hat{y} = 0.0607x^3 - 364.58x^2 + 7.3 * 10^5x - 4.9 * 10^8$$

For example:

In the year of 2026, the estimated mean room price would be:

$$\hat{y} = \text{RM } 2327.11$$

compare to mean room price when 2021 which is RM 1705.44

SECTION 6: DISCUSSION

03 How to find the most suitable rental area (or property name, type etc.) to be invested, or rented?

For rental, the specific requirement could be inserted to the equation, and it could present a tenant either the locus, that is available, or the rental price (if specified the location)

The equation of hyperplane based on all the 7 significant factors that affect the rent:

$$E(X) = 20.91x_1 - 170x_2 - 14.63x_3 + 122.71x_4 + 1.24x_5 + 21.40x_6 - 19.57x_7 - 41431.6$$

For example:

Unit that is built in 2021 with 1 room, 1 parking, 1 bathroom, with a size of 400 sqft, 3 facilities, with distance to the centre of 3km brings:

$$E(X) = \text{RM } 1267.08$$

SECTION 6: DISCUSSION

04 For property developers, where to potentially build a new housing area?

Based on a heatmap created when analysing *Rent vs City*, property developers is suggested to build a new property at the yellow and green areas because these are the areas that have potential to give back a high return of investment (ROI).



OVERVIEW

- 
- 1 **Background**
 - 2 **Research questions**
 - 3 **Data pre-processing**
 - 4 **Exploratory Data Analysis**
 - 5 **Evaluation & Modelling**
 - 6 Discussion**
 - 7 **Conclusion**

SECTION 6: **DISCUSSION**

OVERVIEW

- 
- 1 Background**
 - 2 Research questions**
 - 3 Data pre-processing**
 - 4 Exploratory Data Analysis**
 - 5 Evaluation & Modelling**
 - 6 Discussion**
 - 7 Conclusion**

SECTION 7: **CONCLUSION**

SECTION 7: CONCLUSION

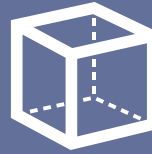
In this study, we had done...

Analysis



There are mainly 7 crucial factors that impact rental prices which provide insights into market dynamics.

Modelling



We made 3 models namely multiple linear regression model, polynomial regression model and random forest regression model for prediction.

Visualisation



Visualised all results and model in the form including heatmap, boxplot, histogram and so on.

SECTION 7: CONCLUSION

Our suggestions:

Tenants

Tenants should always consider the **various factors** that affects the rental to find a most suitable rent that suits themselves the best.

Investors

Investors should keep an eye on the **events** that will make rental price fluctuates before investing a property.

Developers

Property developers should focus on **potential area** to build a property by monitoring different areas based on the different aspects.



This has proved the importance of predictive models in understanding and forecasting rental market dynamics in the real estate sector.



DANKE !



References

- Chan, K. H., Ho, G. Y., Lim, F. Y., Lim, M. Y., & Yap, Y. L. (2018). *Determinants of housing price index in Malaysia* (Doctoral dissertation, UTAR).
- Worldometer. (2023). *Malaysia Population*. <https://www.worldometers.info/world-population/malaysia-population/>
- Zainon, N., Mohd-Rahim, F. A., Sulaiman, S., Abd-Karim, S. B., & Hamzah, A. (2017). Factors affecting the demand of affordable housing among the middle-income groups in Klang Valley Malaysia. *Journal of Design and Built Environment*, 1-10. <https://doi.org/10.22452/jdbe.sp2017no1.1>