Rental Pricing in KL

Ken Law

2023-02-27

Rental Prices in Kuala Lumpur (02/12/2023)

Description

This dataset is provided by ARIEWIJAYA on Kaggle. It is scraped from local rental advertising listing websites in Kuala Lumpur. It contains a few listings from the surrounding outskirt areas of KL. There are 9991 listings total with 13 variables. Primary key is set as ‘ads_id’ by the dataset owner.

Variables

1. Advertising ID: Unique ID assigned to each *primary key*
2. Property name: (Character)
3. Completion year: (Date/Integer)
4. Monthly rent: Currency in Ringgit Malaysia (Currency/Numeric)
5. Location: ‘City - Region’ (Location/Character)
6. Property Type: Categorical e.g. Apartment, Condo (Factor)
7. Rooms: (Integer)
8. Parking: (Integer)
9. Bathroom: (Integer)
10. Size: Calculated in sq. ft. (Integer)
11. Furnished: Fully, Partially, or Not Furnished (Factor)
12. Facilities: Listed amenities and facilities provided (Factor)
13. Additional Facilities: Air-con, Public Transport, etc (Factor)

Business Task

Market Analysis:
Analyze the location, property type, size, and completion year columns to understand the current market trends and demand for specific types of properties in certain areas.

Pricing Optimization:
Use the completion year, location, property type, size, furnished, and facilities columns to determine the optimal price for a property.

Customer Segmentation:
Analyze the facilities and additional facilities columns to identify patterns in customer preferences and segment the market accordingly.

Property Selection:
Use the completion year, location, property type, size, furnished, facilities, parking, bathroom, and rooms columns to identify properties that meet specific customer needs and preferences.

Advertising and Marketing:
Use the advertising ID, property name, completion year, location, property type, size, furnished, and facilities columns to create targeted advertising campaigns and marketing strategies.

Inventory Management:
Use the furnished column to track which properties are fully, partially, or not furnished and manage inventory accordingly.

Maintenance and Improvement:
Use the completion year and facilities columns to identify areas where improvements or maintenance may be needed to keep the property competitive in the market.

Revenue Forecasting:
Use the monthly rent and size columns to forecast potential revenue and evaluate the profitability of a rental property.

Performance Tracking:
Use the monthly rent, size, parking, bathroom, and rooms columns to track the performance of a property over time and identify areas for improvement.

Risk Assessment:
Use the location, completion year, and property type columns to assess the risk of investing in a specific property.

Data Processing

Loading the packages:

```
library(here)
```

```
## here() starts at C:/Users/lawke/Desktop/portfolio_2023/rental-pricing-in-kuala-lumpur-malaysia
```

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.0     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.1     ✔ tibble    3.1.8
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors
```

```
library(dplyr)
library(readr)
library(stringr)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(skimr)
```

Open CSV file:

```
mudah_apartment_clean <- read_csv("C:/Users/lawke/Desktop/portfolio_2023/rental-pricing-in-kuala-lumpur-malaysia/
mudah-apartment-clean.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 9991 Columns: 13
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (8): prop_name, monthly_rent, location, property_type, size, furnished, ...
## dbl (5): ads_id, completion_year, rooms, parking, bathroom
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Quick look:

```
skim_without_charts(mudah_apartment_clean)
```

Data summary

Name	mudah_apartment_clean
Number of rows	9991
Number of columns	13
_____	
Column type frequency:	
character	8
numeric	5

_____

Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
prop_name	508	0.95	3	82	0	1292	0
monthly_rent	0	1.00	15	22	0	273	0
location	0	1.00	18	35	0	60	0
property_type	0	1.00	4	50	0	10	0
size	0	1.00	8	15	0	931	0
furnished	1	1.00	13	19	0	3	0
facilities	895	0.91	4	167	0	5981	0
additional_facilities	2831	0.72	8	68	0	145	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
ads_id	0	1.00	99364754.80	4005482.88	16525511	99753474	99932996	100147960	100323185
completion_year	4373	0.56	2014.81	7.42	1980	2011	2017	2020	2025
rooms	4	1.00	2.76	0.78	1	2	3	3	9
parking	2630	0.74	1.35	0.53	1	1	1	2	10
bathroom	2	1.00	1.95	0.58	1	2	2	2	7

```
glimpse(mudah_apartment_clean)
```

```
## Rows: 9,991
## Columns: 13
## $ ads_id                <dbl> 100323185, 100203973, 100323128, 100191767, 9702…
## $ prop_name             <chr> "The Hipster @ Taman Desa", "Segar Courts", "Pan…
## $ completion_year       <dbl> 2022, NA, NA, 2020, NA, NA, NA, 2018, 2014, NA, …
## $ monthly_rent          <chr> "RM 4 200 per month", "RM 2 300 per month", "RM …
## $ location              <chr> "Kuala Lumpur - Taman Desa", "Kuala Lumpur - Che…
## $ property_type         <chr> "Condominium", "Condominium", "Apartment", "Apar…
## $ rooms                 <dbl> 5, 3, 3, 2, 1, 3, 3, 1, 2, 3, 2, 4, 1, 3, 5, 3, …
## $ parking               <dbl> 2, 1, NA, 1, 1, 1, 2, 1, 1, 1, NA, 2, 1, 1, NA, …
## $ bathroom              <dbl> 6, 2, 2, 2, 1, 2, 2, 1, 1, 2, 2, 2, 1, 2, 2, 2, …
## $ size                  <chr> "1842 sq.ft.", "1170 sq.ft.", "650 sq.ft.", "743…
## $ furnished             <chr> "Fully Furnished", "Partially Furnished", "Fully…
## $ facilities            <chr> "Minimart, Gymnasium, Security, Playground, Swim…
## $ additional_facilities <chr> "Air-Cond, Cooking Allowed, Washing Machine", "A…
```

```
head(mudah_apartment_clean)
```

```
## # A tibble: 6 × 13
##      ads_id prop_n…¹ compl…² month…³ locat…⁴ prope…⁵ rooms parking bathr…⁶ size
##       <dbl> <chr>      <dbl> <chr>   <chr>   <chr>   <dbl>   <dbl>   <dbl> <chr>
## 1 100323185 The Hip…    2022 RM 4 2… Kuala … Condom…     5       2       6 1842…
## 2 100203973 Segar C…      NA RM 2 3… Kuala … Condom…     3       1       2 1170…
## 3 100323128 Pangsap…      NA RM 1 0… Kuala … Apartm…     3      NA       2 650 …
## 4 100191767 Sentul …    2020 RM 1 7… Kuala … Apartm…     2       1       2 743 …
## 5  97022692 Arte Mo…      NA RM 1 2… Kuala … Servic…     1       1       1 494 …
## 6 100322897 Residen…      NA RM 1 5… Kuala … Apartm…     3       1       2 884 …
## # … with 3 more variables: furnished <chr>, facilities <chr>,
## #   additional_facilities <chr>, and abbreviated variable names ¹prop_name,
## #   ²completion_year, ³monthly_rent, ⁴location, ⁵property_type, ⁶bathroom
```

```
str(mudah_apartment_clean)
```

```
## spc_tbl_ [9,991 × 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ads_id               : num [1:9991] 1.0e+08 1.0e+08 1.0e+08 1.0e+08 9.7e+07 ...
##  $ prop_name            : chr [1:9991] "The Hipster @ Taman Desa" "Segar Courts" "Pangsapuri Teratak Muhibbah
2" "Sentul Point Suite Apartment" ...
##  $ completion_year      : num [1:9991] 2022 NA NA 2020 NA ...
##  $ monthly_rent         : chr [1:9991] "RM 4 200 per month" "RM 2 300 per month" "RM 1 000 per month" "RM 1 70
0 per month" ...
##  $ location             : chr [1:9991] "Kuala Lumpur - Taman Desa" "Kuala Lumpur - Cheras" "Kuala Lumpur - Tam
an Desa" "Kuala Lumpur - Sentul" ...
##  $ property_type        : chr [1:9991] "Condominium" "Condominium" "Apartment" "Apartment" ...
##  $ rooms                : num [1:9991] 5 3 3 2 1 3 3 1 2 3 ...
##  $ parking              : num [1:9991] 2 1 NA 1 1 1 2 1 1 1 ...
##  $ bathroom             : num [1:9991] 6 2 2 2 1 2 2 1 1 2 ...
##  $ size                 : chr [1:9991] "1842 sq.ft." "1170 sq.ft." "650 sq.ft." "743 sq.ft." ...
##  $ furnished            : chr [1:9991] "Fully Furnished" "Partially Furnished" "Fully Furnished" "Partially Fu
rnished" ...
##  $ facilities           : chr [1:9991] "Minimart, Gymnasium, Security, Playground, Swimming Pool, Parking, Lif
t, Barbeque area, Multipurpose hall, Jogging Track" "Playground, Parking, Barbeque area, Security, Jogging Track,
Swimming Pool, Gymnasium, Lift, Sauna" "Minimart, Jogging Track, Lift, Swimming Pool" "Parking, Playground, Swimm
ing Pool, Squash Court, Security, Minimart, Gymnasium, Lift" ...
##  $ additional_facilities: chr [1:9991] "Air-Cond, Cooking Allowed, Washing Machine" "Air-Cond, Cooking Allowed
, Near KTM/LRT" NA "Cooking Allowed, Near KTM/LRT, Washing Machine" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ads_id = col_double(),
##   ..   prop_name = col_character(),
##   ..   completion_year = col_double(),
##   ..   monthly_rent = col_character(),
##   ..   location = col_character(),
##   ..   property_type = col_character(),
##   ..   rooms = col_double(),
##   ..   parking = col_double(),
##   ..   bathroom = col_double(),
##   ..   size = col_character(),
##   ..   furnished = col_character(),
##   ..   facilities = col_character(),
##   ..   additional_facilities = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Split location into city and region:

```
mudah_apartment_clean <- mudah_apartment_clean %>%
  mutate(city = str_split(location, "-", simplify = TRUE)[, 1],
         region = str_split(location, "-", simplify = TRUE)[, 2]) %>% select(-location)
```

Data Validation

```
class(mudah_apartment_clean$prop_name)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$completion_year)
```

```
## [1] "numeric"
```

```
class(mudah_apartment_clean$monthly_rent)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$city)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$region)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$rooms)
```

```
## [1] "numeric"
```

```
class(mudah_apartment_clean$parking)
```

```
## [1] "numeric"
```

```
class(mudah_apartment_clean$bathroom)
```

```
## [1] "numeric"
```

```
class(mudah_apartment_clean$size)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$furnished)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$facilities)
```

```
## [1] "character"
```

```
class(mudah_apartment_clean$additional_facilities)
```

```
## [1] "character"
```

Data Cleaning and Transformation

Convert monthly_rent (character to numeric):

```
mudah_apartment_clean <- mudah_apartment_clean %>%
  mutate(monthly_rent = as.integer(str_replace_all(mudah_apartment_clean$monthly_rent, c("RM" = "", "per" = "", "
month" = "", " " = ""))))


class(mudah_apartment_clean$monthly_rent)
```

```
## [1] "integer"
```

Convert size (character to integer):

```
mudah_apartment_clean <- mudah_apartment_clean %>%
  mutate(size = as.integer((str_remove(size, "sq.ft."))))
class(mudah_apartment_clean$size)
```

```
## [1] "integer"
```

Convert furnished (character to factor):

```
mudah_apartment_clean <- mudah_apartment_clean %>%
  mutate(furnished = as.factor(furnished))
class(mudah_apartment_clean$furnished)
```

```
## [1] "factor"
```

Remove Properties with no names (NA):

```
mudah_apartment_clean <- subset(mudah_apartment_clean, !is.na(mudah_apartment_clean$prop_name))
```

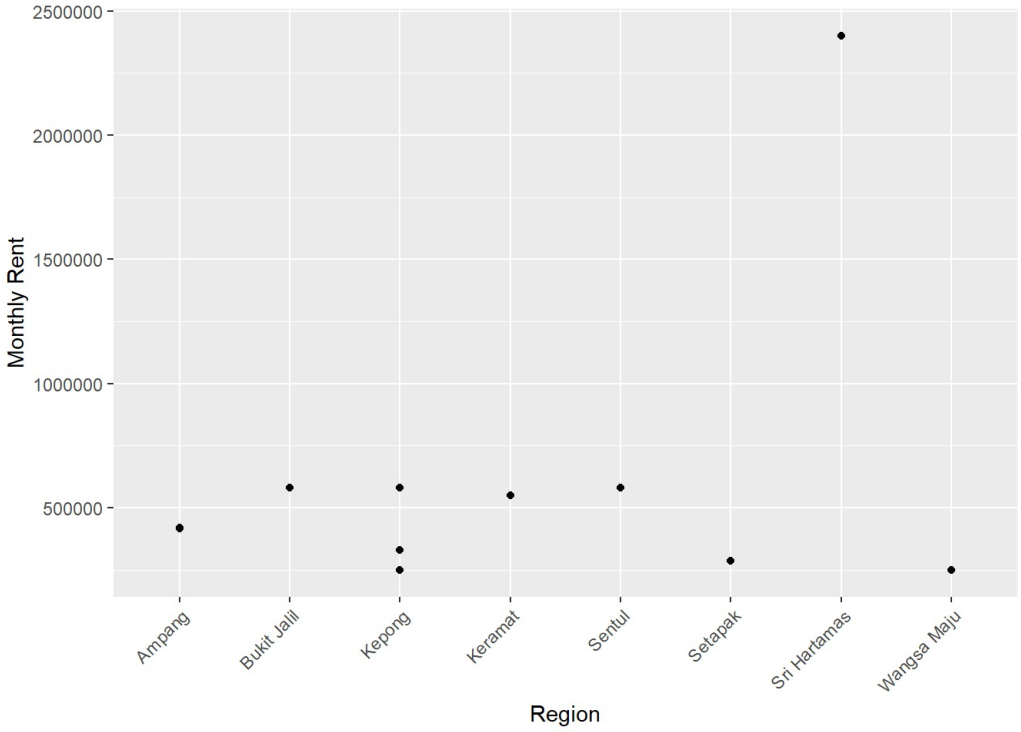
Analysis

Which regions in KL are the most expensive?

```
expensive_region <-  subset(mudah_apartment_clean, monthly_rent > 200000)
ggplot(data = expensive_region, aes(x = expensive_region$region , y = expensive_region$monthly_rent)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(x = "Region", y = "Monthly Rent")
```

```
## Warning: Use of `expensive_region$region` is discouraged.
## i Use `region` instead.
```

```
## Warning: Use of `expensive_region$monthly_rent` is discouraged.
## i Use `monthly_rent` instead.
```



Which rental properties have the highest rent?

```
mudah_apartment_clean <- mudah_apartment_clean %>%  arrange(desc(monthly_rent))
head(mudah_apartment_clean)
```

```
## # A tibble: 6 × 14
##      ads_id prop_n…¹ compl…² month…³ prope…⁴ rooms parking bathr…⁵  size furni…⁶
##       <dbl> <chr>      <dbl>   <int> <chr>   <dbl>   <dbl>   <dbl> <int> <fct>
## 1 100058432 Sunway …    2011 2400000 Condom…     4       3       4  2573 Fully …
## 2 100223741 The Tam…    2005  580000 Condom…     4       1       3  1345 Fully …
## 3  99990155 United …    2019  580000 Servic…     3       2       2   958 Fully …
## 4  99812277 Parkhil…    2019  580000 Condom…     3       2       2  1100 Fully …
## 5  99812339 The Hau…    2019  550000 Condom…     3       2       2  1061 Partia…
## 6  99888010 The Ele…    2015  419000 Servic…     1       1       1   705 Partia…
## # … with 4 more variables: facilities <chr>, additional_facilities <chr>,
## #   city <chr>, region <chr>, and abbreviated variable names ¹prop_name,
## #   ²completion_year, ³monthly_rent, ⁴property_type, ⁵bathroom, ⁶furnished
```

1. Sunway Vivaldi = RM 2,400,000
2. The Tamarind = RM 580,000
3. United Point Residence @ North Kiara = RM 580,000
4. Parkhill Residence Bukit Jalil = RM 580,000
5. The Haute = RM 550,000

- Monthly rent

Which rental properties have the lowest rent?

```
tail(mudah_apartment_clean)
```

```
## # A tibble: 6 × 14
##      ads_id prop_n…¹ compl…² month…³ prope…⁴ rooms parking bathr…⁵  size furni…⁶
##       <dbl> <chr>      <dbl>   <int> <chr>   <dbl>   <dbl>   <dbl> <int> <fct>
## 1 100183866 Melur A…    2006     100 Apartm…     1      NA       1   100 Not Fu…
## 2 100163744 Platinu…      NA     100 Condom…     1       1       1     1 Not Fu…
## 3 100129853 Platinu…    2013     100 Condom…     1      NA       1   100 Not Fu…
## 4  99872270 Platinu…      NA     100 Condom…     1      NA       1   500 Not Fu…
## 5  99974016 Idaman …      NA      90 Condom…     1       1       1   100 Not Fu…
## 6  99410613 Setapak…    1994      80 Condom…     1      NA       1     5 Not Fu…
## # … with 4 more variables: facilities <chr>, additional_facilities <chr>,
## #   city <chr>, region <chr>, and abbreviated variable names ¹prop_name,
## #   ²completion_year, ³monthly_rent, ⁴property_type, ⁵bathroom, ⁶furnished
```

1. Setapak Ria Condo = RM 80
2. Idaman Sutera = RM 90
3. Platinum Lake PV 15 = RM 100
4. Platinum Hill PV 2 = RM 100
5. Platinum Lake PV 8 = RM 100

- Monthly rent