Data Analysis of Rental Pricing in Klang Valley

Rental Pricing Trend within Klang Valley

Dataset was taken from Kaggle, webscraped from Mudah.com by a public user. This project aims to provide visual insights on Klang Valley's rental trend with given category (provided in Mudah website) such as location, size, number of bedrooms, parking or furnished. Chosen this project as I am also actively looking for a new unit to move in Klang Valley, having this insights will help me compare my options, evaluating them within the actual market trend.

Installing data packages and library

```
options(repos = "https://cloud.r-project.org")
```

```
install.packages("stringr")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'stringr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 4.3.1
```

```
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```r
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.1

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
install.packages("dplyr")
```

```
## Warning: package 'dplyr' is in use and will not be installed
```

```r
library(dplyr)
library(readr)
install.packages("wordcloud")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'wordcloud' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'wordcloud'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\1\AppData\Local\R\win-library\4.3\00LOCK\wordcloud\libs\x64\wordcloud.dll
## to C:\Users\1\AppData\Local\R\win-library\4.3\wordcloud\libs\x64\wordcloud.dll:
## Permission denied

## Warning: restored 'wordcloud'

##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```r
library(wordcloud)
```

```
## Warning: package 'wordcloud' was built under R version 4.3.1
```

```
## Loading required package: RColorBrewer
```

```r
install.packages("broom")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'broom' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```r
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.3.1
```

```r
rental_df = read.csv("Rental Pricing.csv")
```

Data Exploration

To describe the whole dataset (rows and columns), also checking on how many empty cells are within the dataset

```r
colnames(rental_df)
```

```
##  [1] "ads_id"                "property_type"         "prop_name"
##  [4] "completion_year"       "monthly_rent"          "district"
##  [7] "parking"               "bathroom"              "size"
## [10] "furnished"             "facilities"            "additional_facilities"
## [13] "region"
```

```r
str(rental_df)
```

```
## 'data.frame':    19991 obs. of  13 variables:
##  $ ads_id               : int  100323185 100203973 100323128 100191767 97022692 100322897 100322962 1
##  $ property_type        : chr  "Condominium" "Condominium" "Apartment" "Apartment" ...
##  $ prop_name            : chr  "The Hipster @ Taman Desa" "Segar Courts" "Pangsapuri Teratak Muhibbal
##  $ completion_year      : int  2022 NA NA 2020 NA NA NA 2018 2014 NA ...
##  $ monthly_rent         : num  4200 2300 1000 1700 1299 ...
##  $ district             : chr  " Taman Desa" " Cheras" " Taman Desa" " Sentul" ...
##  $ parking              : int  2 1 NA 1 1 1 2 1 1 1 ...
##  $ bathroom             : int  6 2 2 2 1 2 2 1 1 2 ...
##  $ size                 : int  1842 1170 650 743 494 884 982 700 750 862 ...
##  $ furnished            : chr  "Fully Furnished" "Partially Furnished" "Fully Furnished" "Partially F
##  $ facilities           : chr  "Minimart, Gymnasium, Security, Playground, Swimming Pool, Parking, Li
##  $ additional_facilities: chr  "Air-Cond, Cooking Allowed, Washing Machine" "Air-Cond, Cooking Allowe
##  $ region               : chr  "Kuala Lumpur" "Kuala Lumpur" "Kuala Lumpur" "Kuala Lumpur" ...
```

```
head(rental_df)
```

```
##      ads_id     property_type                    prop_name completion_year
## 1 100323185       Condominium      The Hipster @ Taman Desa            2022
## 2 100203973       Condominium                  Segar Courts              NA
## 3 100323128         Apartment Pangsapuri Teratak Muhibbah 2              NA
## 4 100191767         Apartment  Sentul Point Suite Apartment            2020
## 5  97022692 Service Residence               Arte Mont Kiara              NA
## 6 100322897         Apartment      Residensi Vista Wirajaya              NA
##   monthly_rent    district parking bathroom size           furnished
## 1         4200  Taman Desa       2        6 1842     Fully Furnished
## 2         2300      Cheras       1        2 1170 Partially Furnished
## 3         1000  Taman Desa      NA        2  650     Fully Furnished
## 4         1700      Sentul       1        2  743 Partially Furnished
## 5         1299  Mont Kiara       1        1  494       Not Furnished
## 6         1500     Setapak       1        2  884 Partially Furnished
##
## 1            Minimart, Gymnasium, Security, Playground, Swimming Pool, Parking, Lift, Barbeque area
## 2                                 Playground, Parking, Barbeque area, Security, Jogging Track, Swi
## 3                                                                                          Minimart,
## 4                                              Parking, Playground, Swimming Pool, Squash Court, S
## 5 Parking, Security, Lift, Swimming Pool, Playground, Gymnasium, Barbeque area, Minimart, Multipurpo
## 6                                                        Parking, Security, Lift, Swimming P
##                           additional_facilities       region
## 1     Air-Cond, Cooking Allowed, Washing Machine Kuala Lumpur
## 2        Air-Cond, Cooking Allowed, Near KTM/LRT Kuala Lumpur
## 3                                               Kuala Lumpur
## 4 Cooking Allowed, Near KTM/LRT, Washing Machine Kuala Lumpur
## 5                                     Air-Cond Kuala Lumpur
## 6                  Cooking Allowed, Near KTM/LRT Kuala Lumpur
```

```
emptycell = colSums(is.na(rental_df))
print(emptycell)
```

```
##                ads_id         property_type             prop_name
##                     0                     0                     0
##       completion_year          monthly_rent              district
##                  9185                     2                     0
##               parking              bathroom                  size
##                  5702                     6                     0
##             furnished            facilities additional_facilities
##                     0                     0                     0
##                region
##                     0
```

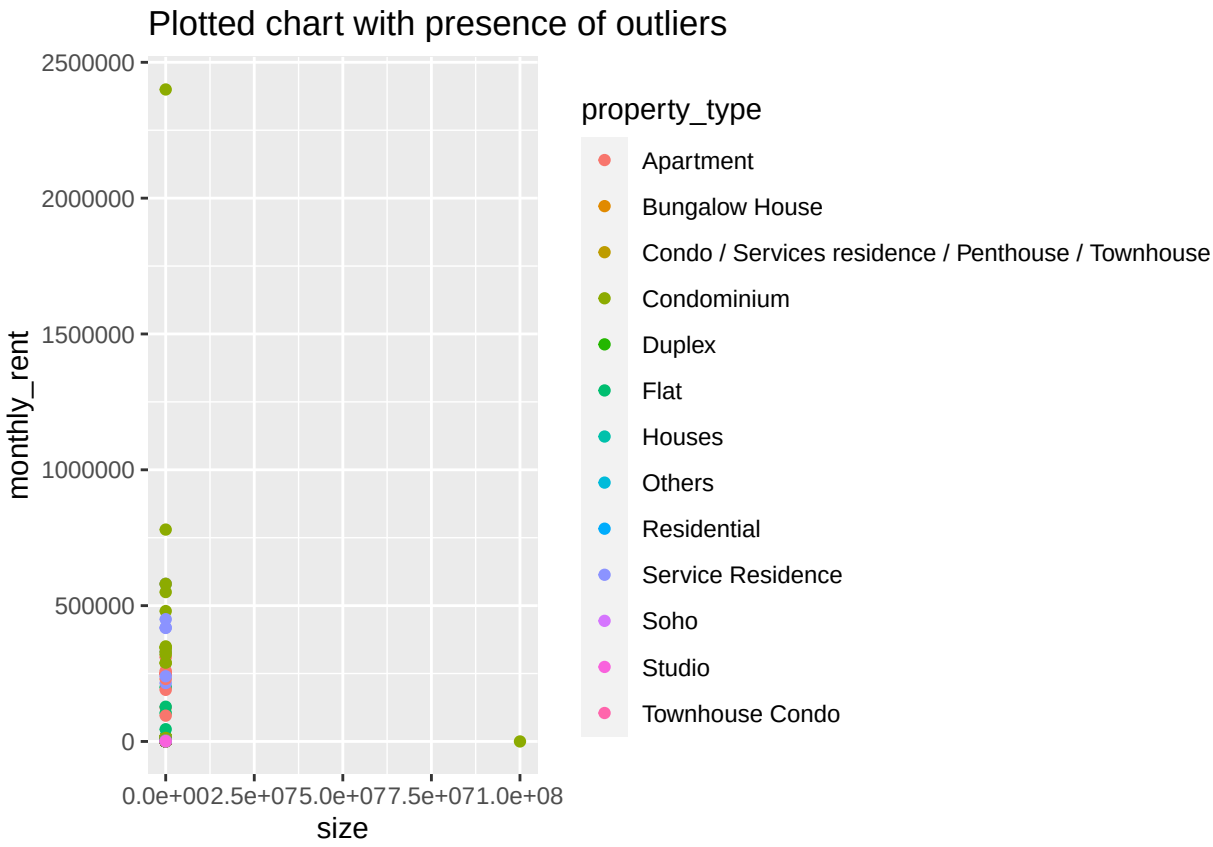
Dataset is already good as it is. Empty cells in column 'parking' and 'completion_year' will be ignored as it will not be used as much in the upcoming analysis.

Data Preparation

Checking for numerical outliers in the dataset ### Checking for outliers and important variable column of interest; size and monthly rent. Plotting graph with outliers will make our plots skewed.

```
ggplot(data = rental_df) +
  geom_point(mapping = aes(x = size, y = monthly_rent, color = property_type)) +
  labs(title = "Plotted chart with presence of outliers")
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



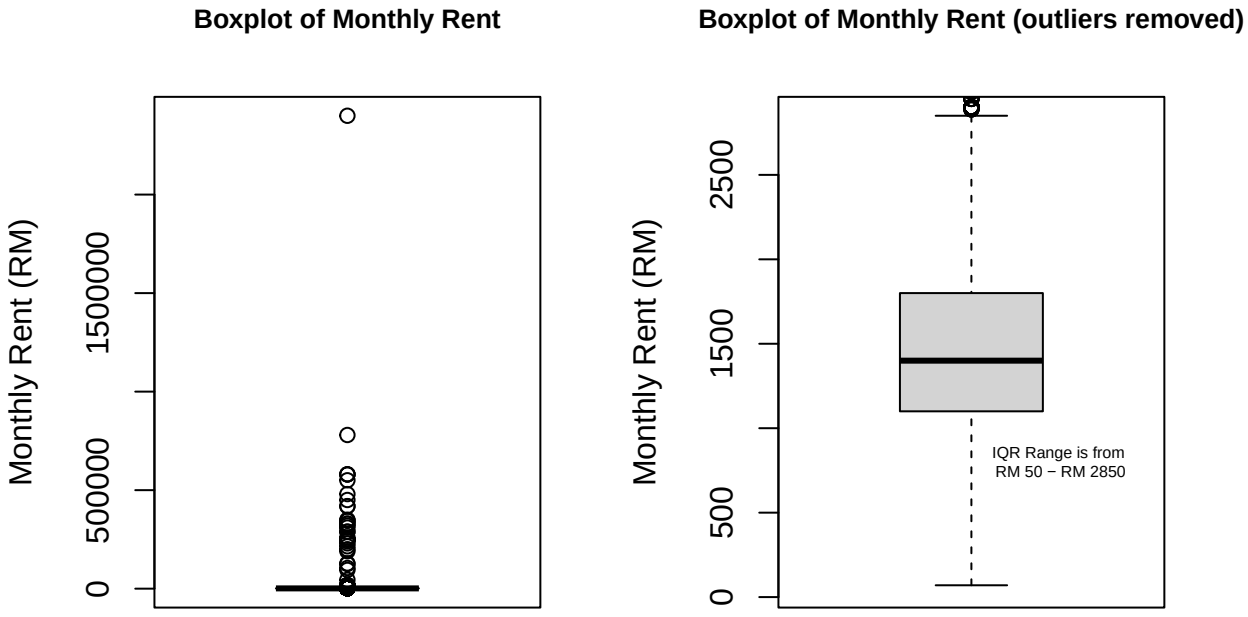

```
rentsummary = summary(rental_df$monthly_rent)
q1 = rentsummary["1st Qu."]
median = rentsummary['Median']
q3 = rentsummary["3rd Qu."]

lower_limit = q1 - 1.5 * IQR(rental_df$monthly_rent, na.rm = TRUE)
upper_limit = q3 + 1.5 * IQR(rental_df$monthly_rent, na.rm = TRUE)


print(lower_limit)
```

```
## 1st Qu.
##      50
```

```
print(upper_limit)
```

```
## 3rd Qu.
##    2850
```

```r
par(mfrow = c(1, 2))
boxplot(rental_df$monthly_rent, main = "Boxplot of Monthly Rent", ylab = "Monthly Rent (RM)", cex.main =
boxplot(rental_df$monthly_rent, main = "Boxplot of Monthly Rent (outliers removed)", ylab = "Monthly Re
text(x = 1.25, y = 800, labels = "IQR Range is from \nRM 50 - RM 2850", cex = 0.5)
```



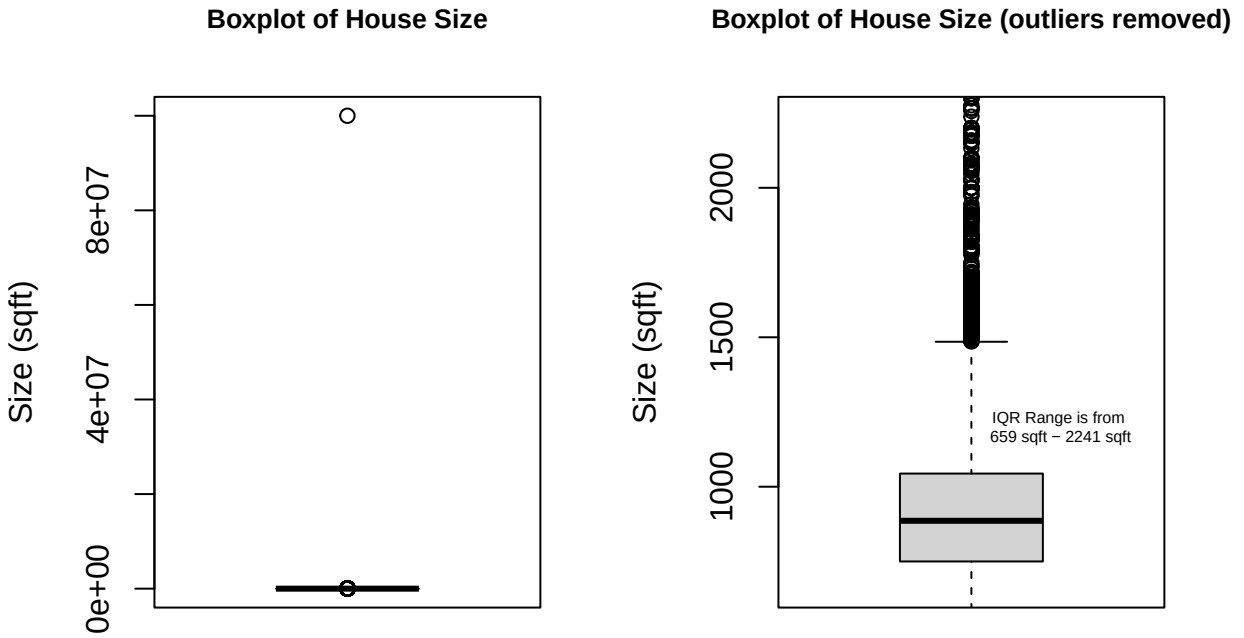

```r
sizesummary = summary(rental_df$size)
q1_size = sizesummary["1st Qu."]
median_size = sizesummary['Median']
q3_size = sizesummary["3rd Qu."]

size_lowerlimit = q1 - 1.5 * IQR(rental_df$size, na.rm = TRUE)
size_upperlimit = q3 + 1.5 * IQR(rental_df$size, na.rm = TRUE)


print(size_lowerlimit)
```

```
## 1st Qu.
##     659
```

```r
print(size_upperlimit)
```

```
## 3rd Qu.
##    2241
```

```r
par(mfrow = c(1, 2))
boxplot(rental_df$size, main = "Boxplot of House Size", ylab = "Size (sqft)", cex.main = 0.8)
boxplot(rental_df$size, main = "Boxplot of House Size (outliers removed)", ylab = "Size (sqft)", ylim =
text(x = 1.25, y = 1200, labels = "IQR Range is from \n659 sqft - 2241 sqft", cex = 0.5)
```



Creating new dataframe after removing outliers in our numerical column: monthly rent and size.

```r
filteredrental = rental_df %>%
  filter(monthly_rent <= 2850 & size <= 2241 & monthly_rent > 100 & size > 100)

dim(filteredrental)
```

```
## [1] 18626    13
```

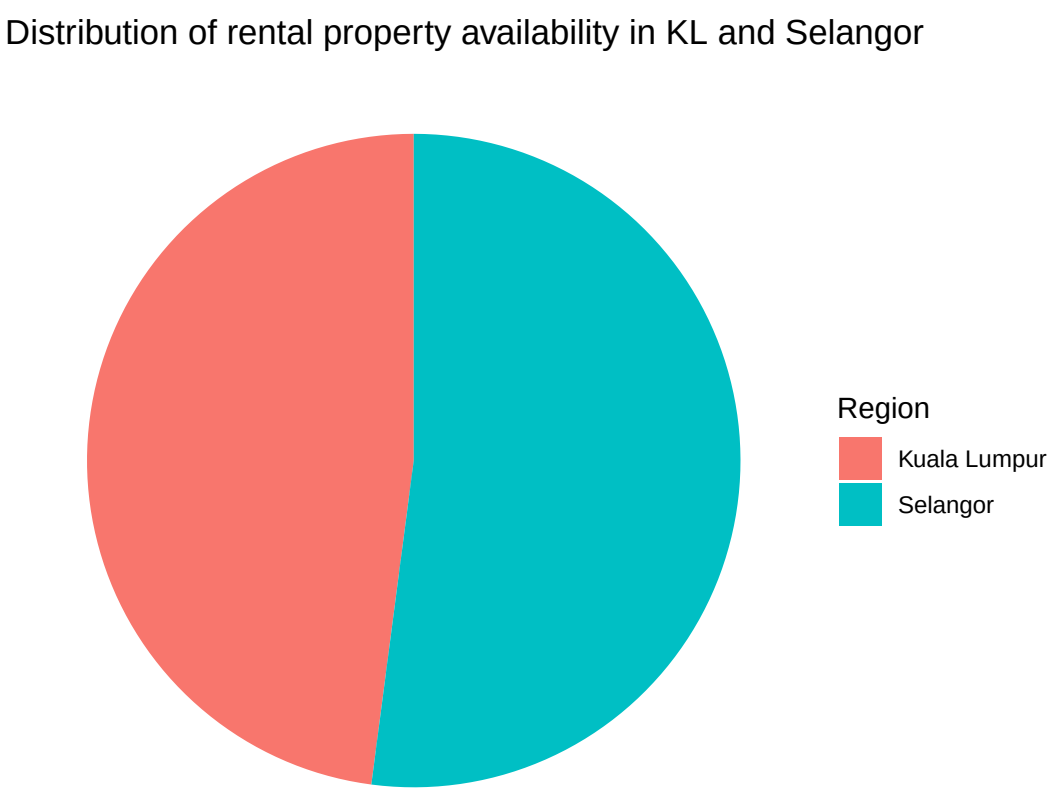
```r
summary(filteredrental)
```

```
##      ads_id          property_type       prop_name         completion_year
##  Min.   : 16525511   Length:18626       Length:18626       Min.   :1977
##  1st Qu.: 99850470   Class :character   Class :character   1st Qu.:2013
##  Median :100228333   Mode  :character   Mode  :character   Median :2017
##  Mean   : 99723326                                         Mean   :2015
##  3rd Qu.:100627338                                         3rd Qu.:2020
##  Max.   :100854617                                         Max.   :2025
##                                                            NA's   :8693
```

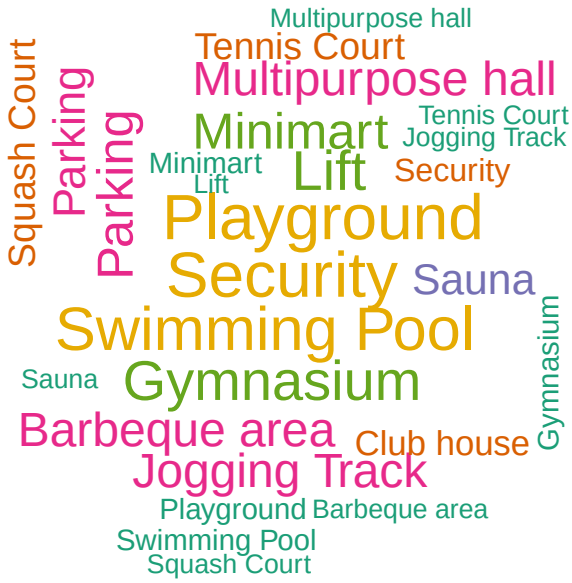
```
##   monthly_rent    district            parking          bathroom
##  Min.   : 120   Length:18626       Min.   : 1.000   Min.   :1.000
##  1st Qu.:1100   Class :character   1st Qu.: 1.000   1st Qu.:2.000
##  Median :1400   Mode  :character   Median : 1.000   Median :2.000
##  Mean   :1447                      Mean   : 1.391   Mean   :1.849
##  3rd Qu.:1750                      3rd Qu.: 2.000   3rd Qu.:2.000
##  Max.   :2850                      Max.   :10.000   Max.   :8.000
##                                    NA's   :5356     NA's   :2
##       size          furnished          facilities        additional_facilities
##  Min.   : 104.0   Length:18626       Length:18626       Length:18626
##  1st Qu.: 737.0   Class :character   Class :character   Class :character
##  Median : 867.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 884.6
##  3rd Qu.:1012.0
##  Max.   :2238.0
##
##     region
##  Length:18626
##  Class :character
##  Mode  :character
##
##
##
##
```

Data Visualization **To easily illustrate trends in our dataset.**

```
ggplot(data = filteredrental, aes(x ="", fill = region)) +
  geom_bar(width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  labs(title = "Distribution of rental property availability in KL and Selangor", fill = "Region")
```



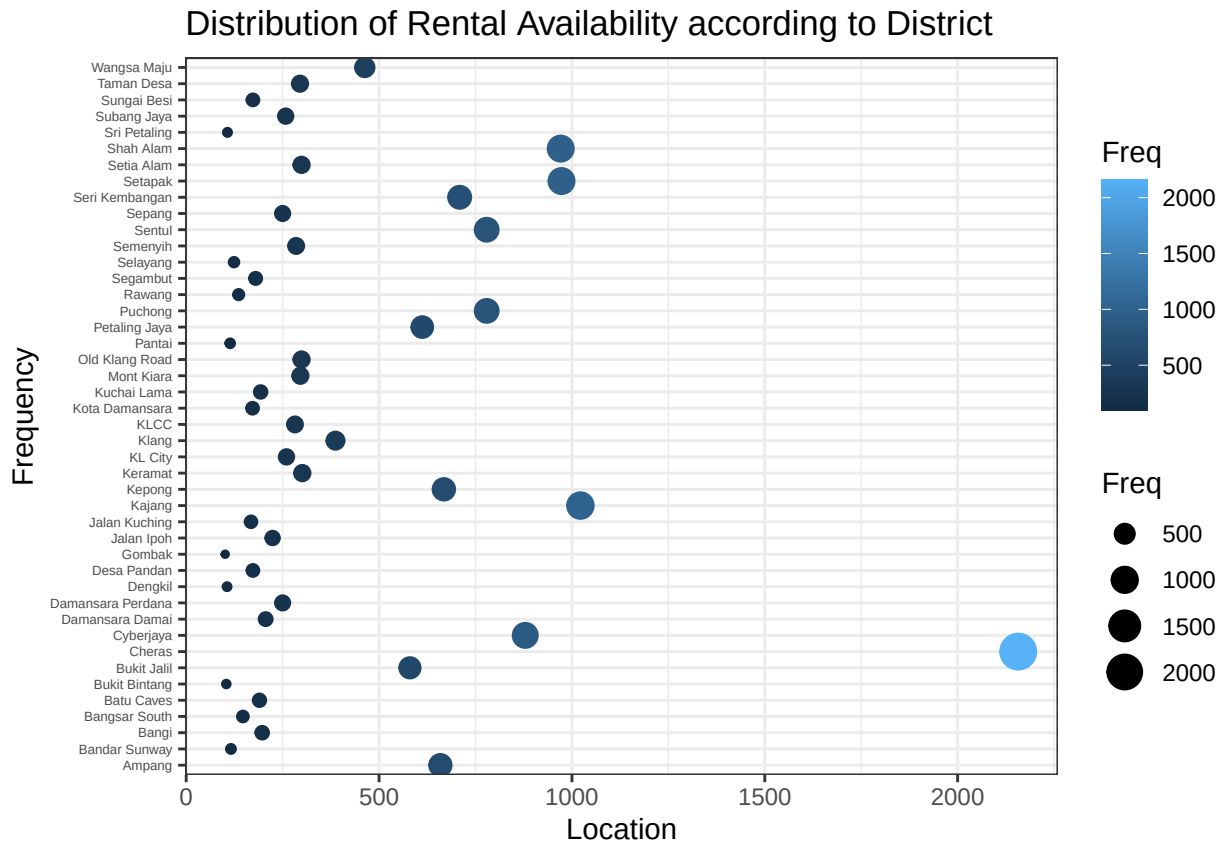

```
facility_lists = strsplit(filteredrental$facilities, ",")

# Create a list to store the frequency counts
facility_counts = vector("list", length = length(facility_lists))

# Count the frequency of each variable in the 'facilities' column
for (i in seq_along(facility_lists)) {
  facility_counts[[i]] = table(facility_lists[[i]])
}

facility_characters = unlist(facility_lists)

# Count the frequency of each character
character_frequency = table(facility_characters)

facility_df = as.data.frame(character_frequency) %>%
  filter(!(facility_characters %in% c(11,10,6))) %>%
  arrange(desc(Freq))

wordcloud(words = facility_df$facility_characters, freq = facility_df$Freq,
          scale = c(2, 0.8), min.freq = 100, random.order = FALSE,
          rot.per = 0.2, colors = brewer.pal(6, "Dark2"),
          main = "Distribution of facilities mentioned")
```



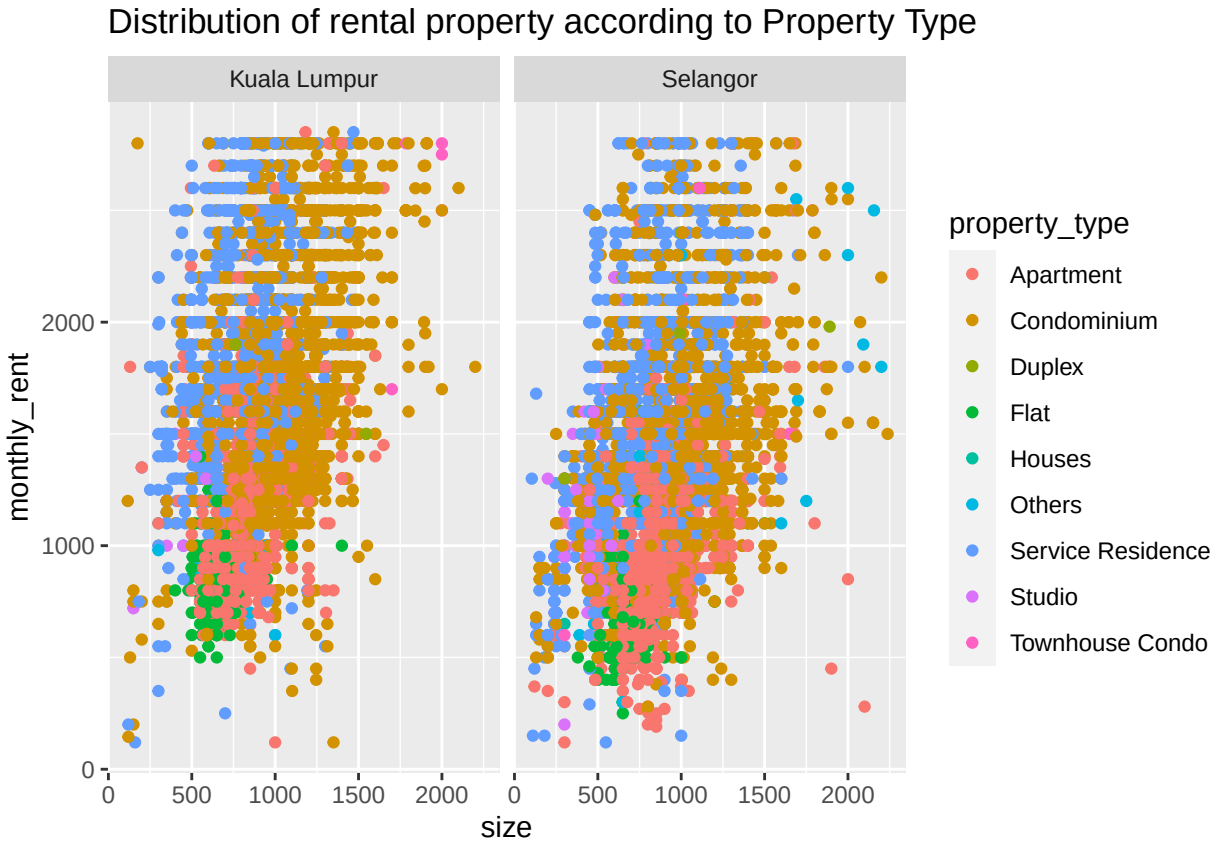
Most listed rentals in Klang Valley are frequently associated with these facilities; security, playground, swimming pool and so on.

```
location_count = (table(rental_df$district)) %>%
  as.data.frame() %>%
  arrange(desc(Freq))
```

```
location_count %>%
  filter(Freq >= 100) %>%
  ggplot(aes(x = Freq, y = Var1, size = Freq, color = Freq)) +
  geom_point() +
  labs(title = "Distribution of Rental Availability according to District", x = "Location", y = "Freque
  theme_bw() +
  theme(axis.text.y = element_text(size = 5))
```




```r
ggplot (data = filteredrental) +
  geom_point(mapping = aes(x = size, y = monthly_rent, color = property_type)) +
  facet_wrap(~region) +
  labs(title = "Distribution of rental property according to Property Type")
```

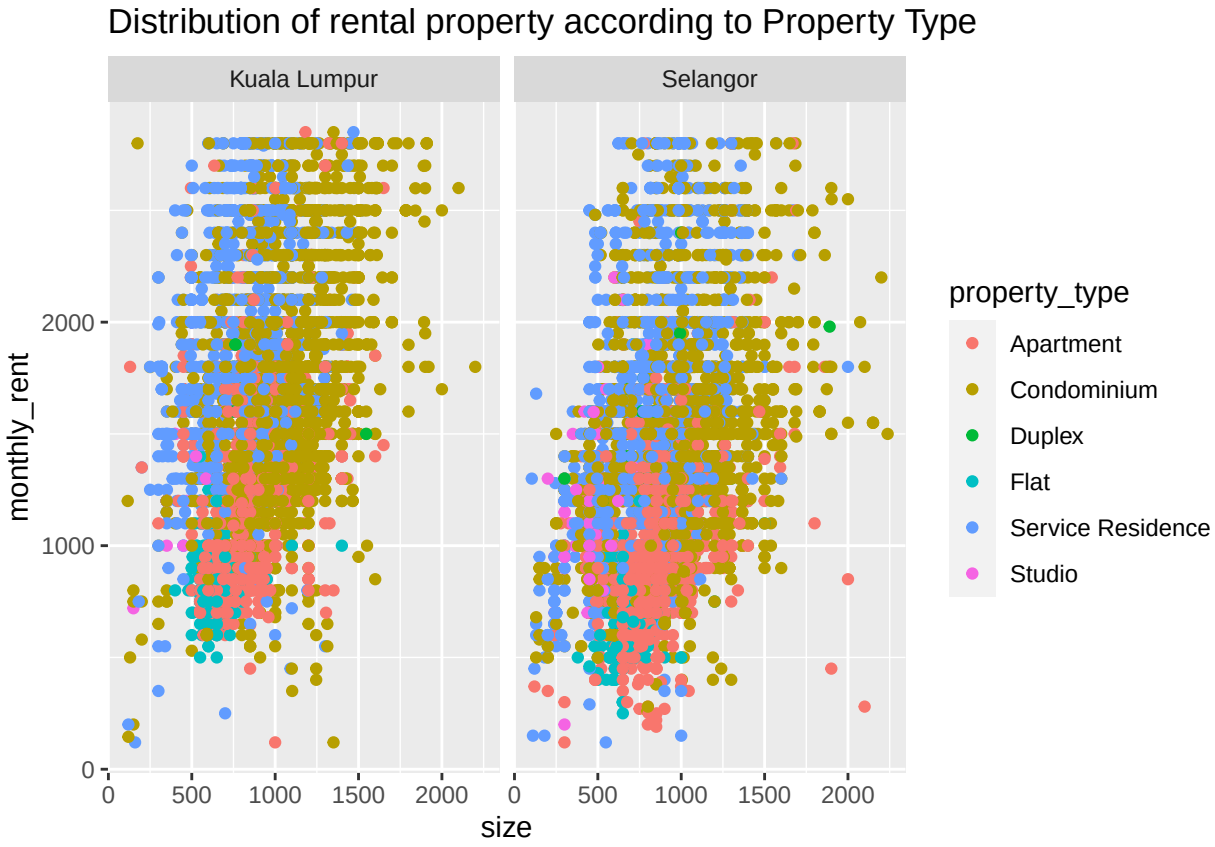


Removing property type listed with less than 50, to visually improving the plot and analysis.

```
table(filteredrental$property_type)
```

```
##
##         Apartment       Condominium            Duplex              Flat
##              5194              7681                69               577
##            Houses            Others Service Residence            Studio
##                 1                78              4808               187
##   Townhouse Condo
##                31
```

```
apt_rental = filteredrental %>%
  filter(property_type %in%  c("Apartment", "Condominium", "Service Residence", "Flat", "Studio", "Duplo

ggplot (data = apt_rental) +
  geom_point(mapping = aes(x = size, y = monthly_rent, color = property_type)) +
  facet_wrap(~region) +
  labs(title = "Distribution of rental property according to Property Type")
```

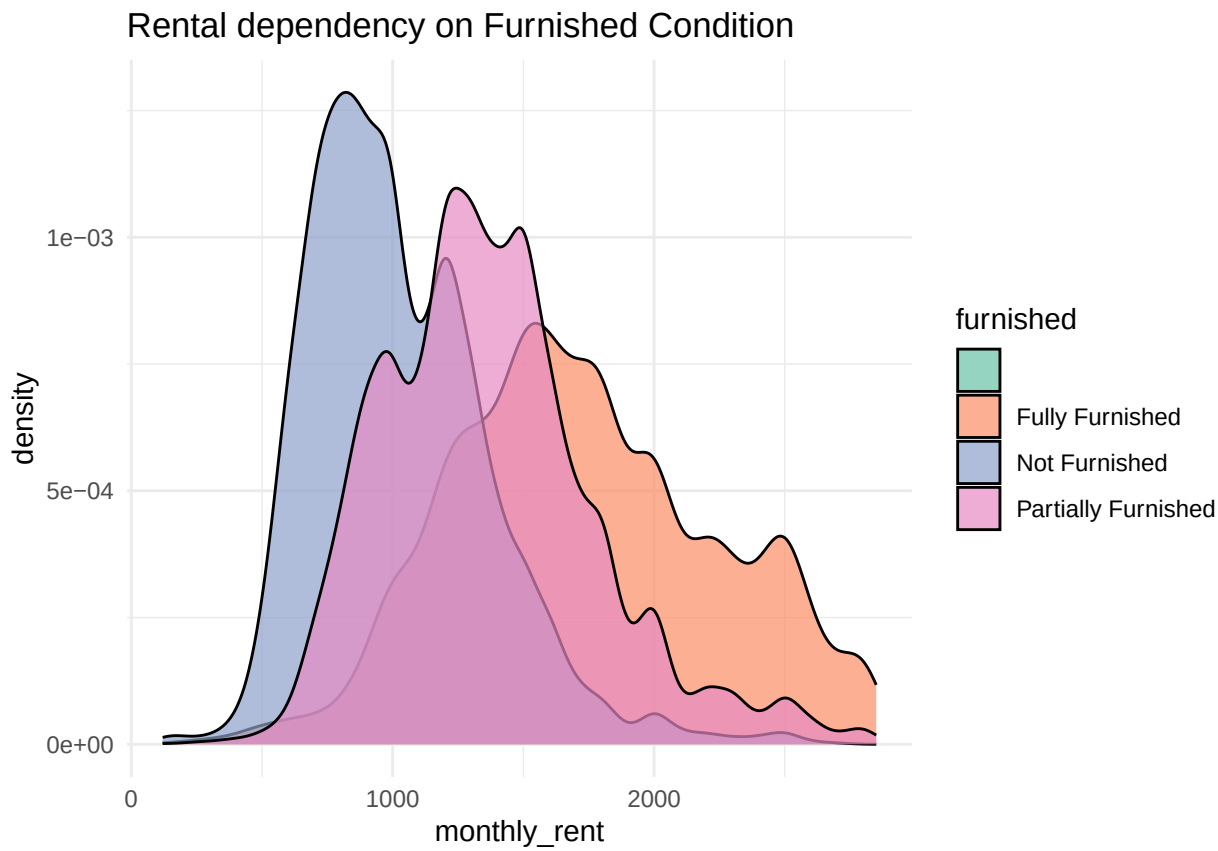


Analysing rental dependency on variable of interest.

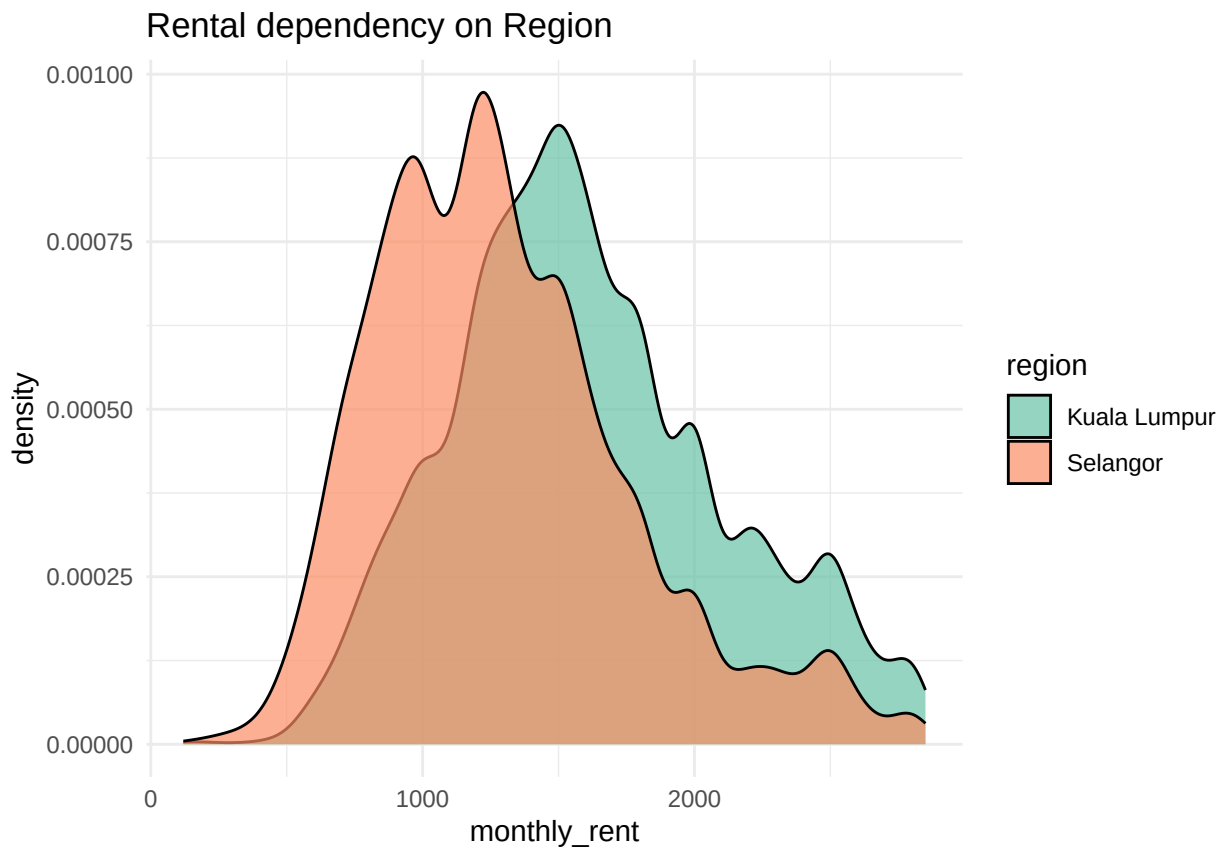
```
ggplot(filteredrental, aes(x = monthly_rent, fill = furnished)) +
  geom_density(alpha = 0.7) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Rental dependency on Furnished Condition") +
  theme_minimal()
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

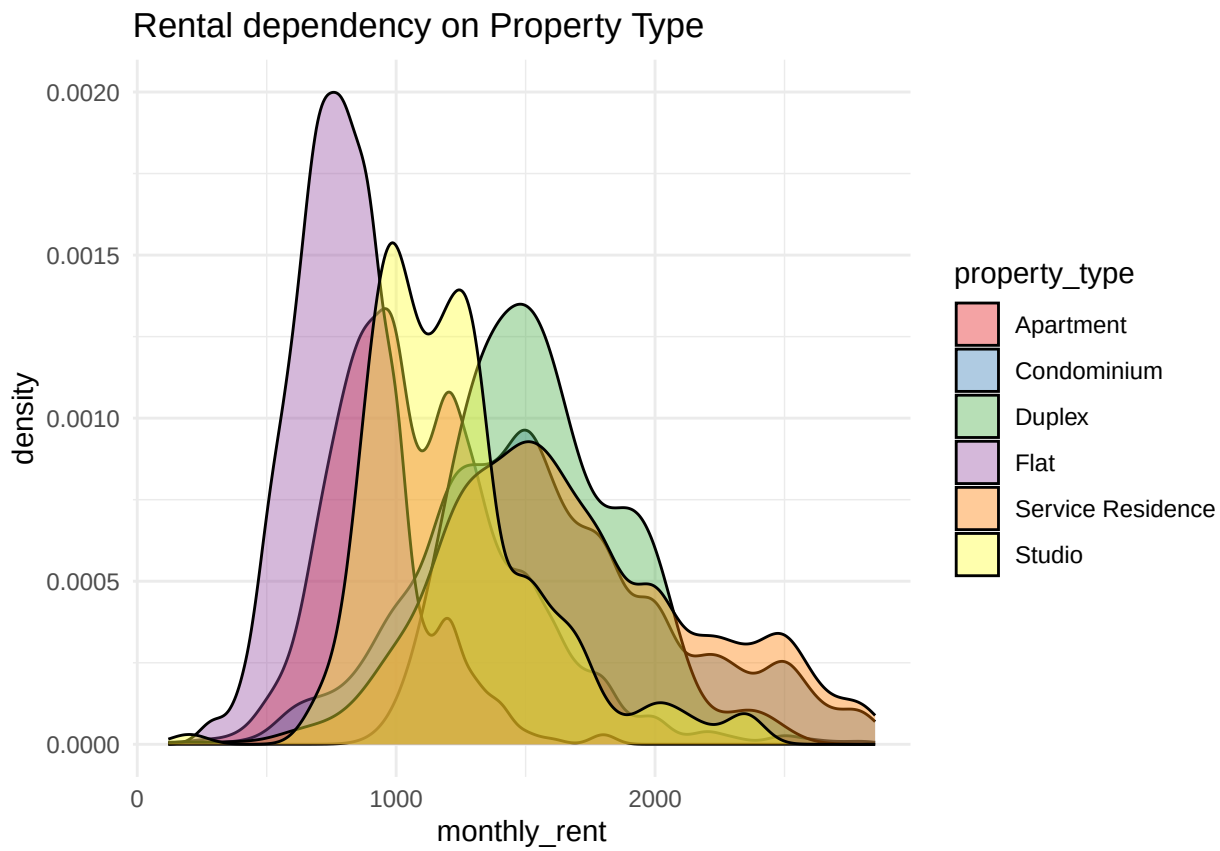
```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```




```
ggplot(filteredrental, aes(x = monthly_rent, fill = region)) +
  geom_density(alpha = 0.7) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Rental dependency on Region") +
  theme_minimal()
```



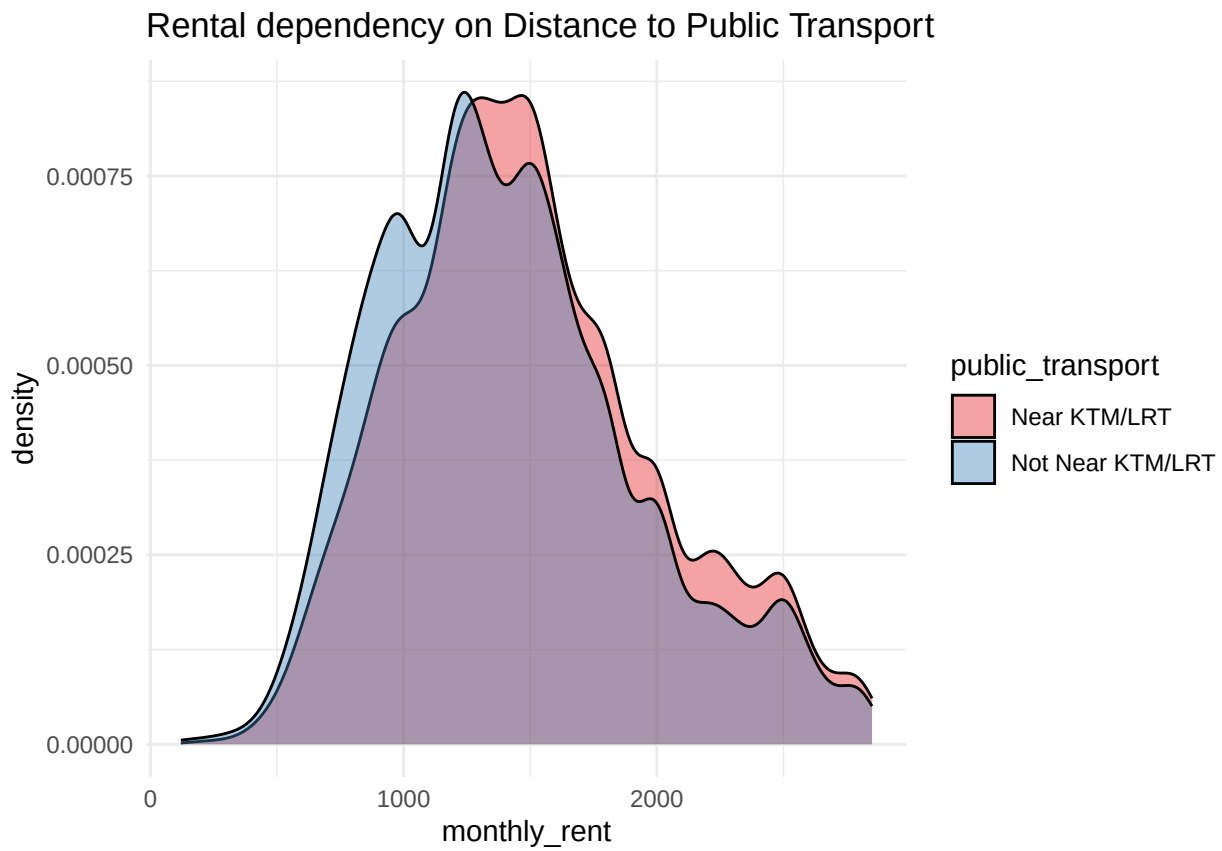

```
ggplot(apt_rental, aes(x = monthly_rent, fill = property_type)) +
  geom_density(alpha = 0.4) +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Rental dependency on Property Type") +
  theme_minimal()
```



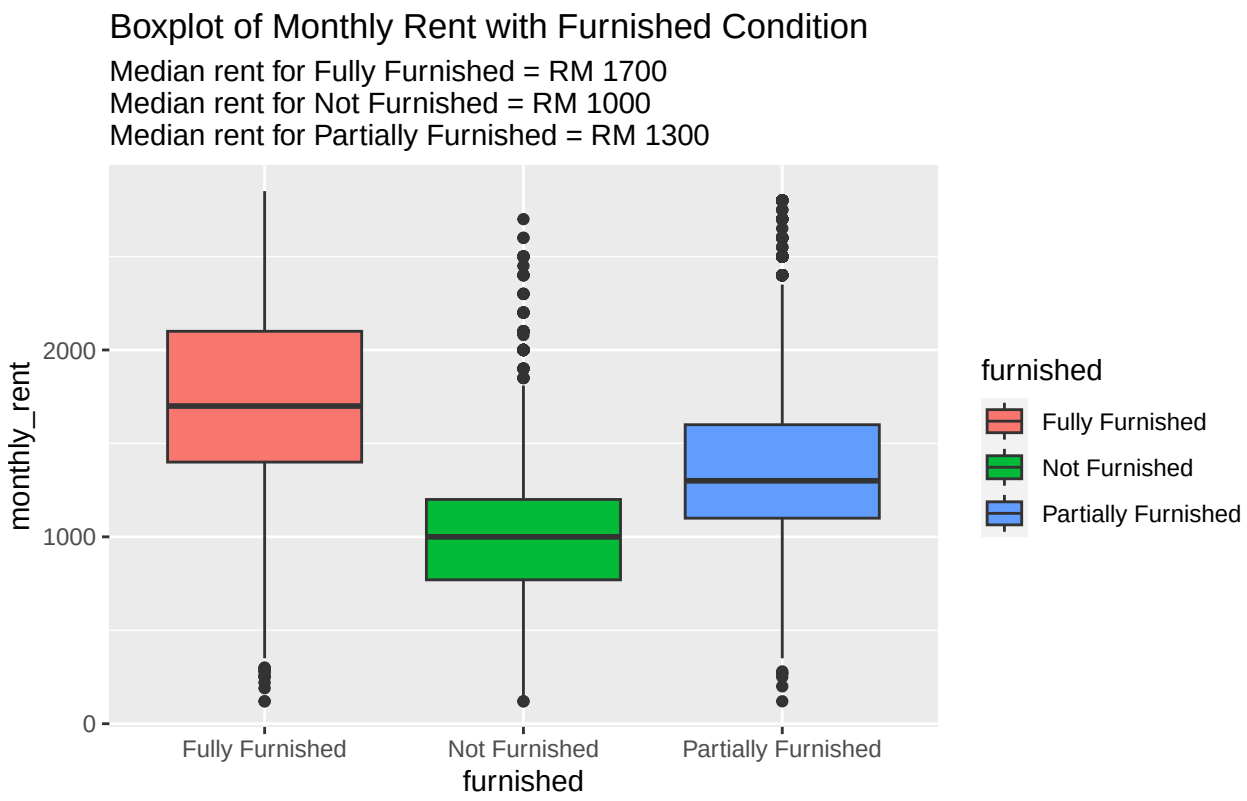
Extracting "Near KTM/LRT" information into a new column, added onto dataframe.

```
rental_publictransport = filteredrental %>%
  mutate(
    public_transport = str_extract(additional_facilities, "Near KTM/LRT"),
    public_transport = replace(public_transport, is.na(public_transport), "Not Near KTM/LRT"))
```

```
ggplot(rental_publictransport, aes(x = monthly_rent, fill = public_transport)) +
  geom_density(alpha = 0.4) +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Rental dependency on Distance to Public Transport") +
  theme_minimal()
```



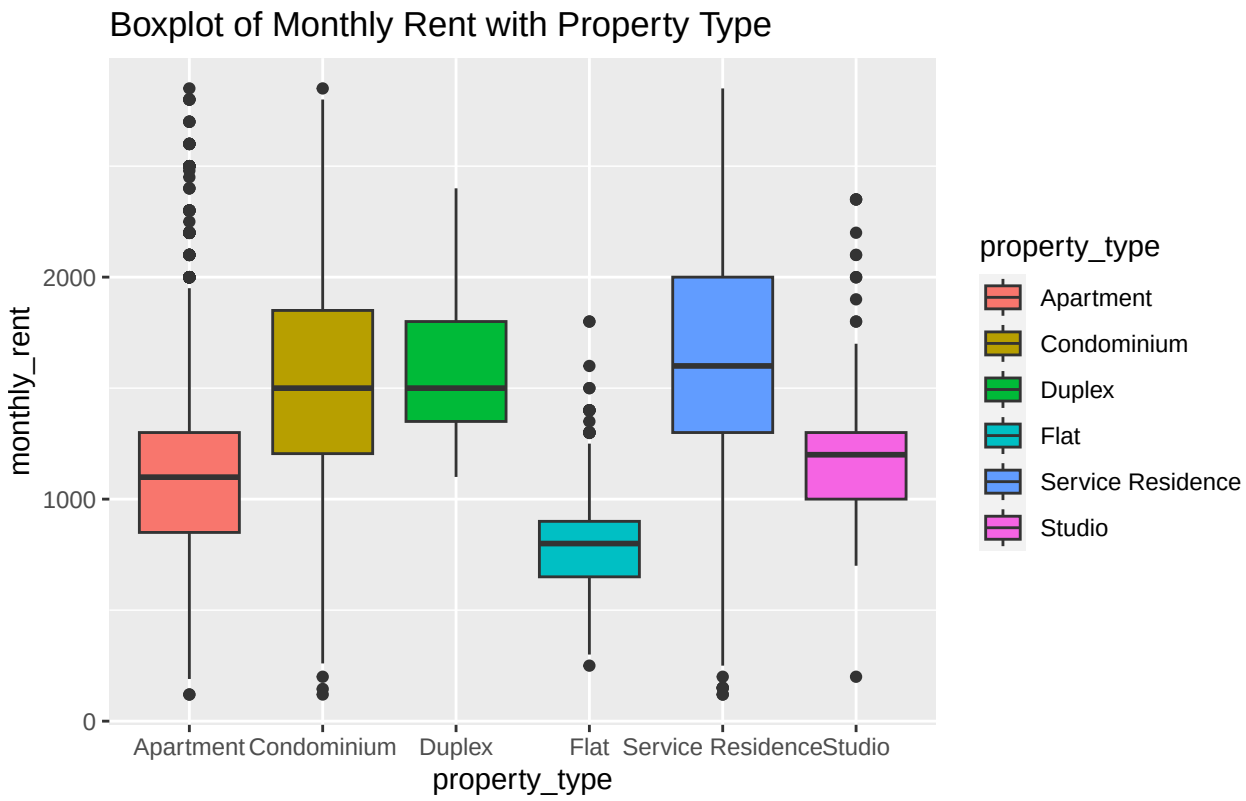

```
filteredrental %>%
  filter(!is.na(furnished) & furnished != "") %>%
  select(furnished, monthly_rent) %>%
  ggplot(aes(x = furnished, y = monthly_rent, fill = furnished)) +
  geom_boxplot() +
  labs(title = "Boxplot of Monthly Rent with Furnished Condition",
       subtitle = "Median rent for Fully Furnished = RM 1700 \nMedian rent for Not Furnished = RM 1000 \
  theme(plot.margin = margin(b=30))
```




```
median_data = filteredrental %>%
  filter(!is.na(furnished) & furnished != "") %>%
  group_by(furnished) %>%
  summarize(median_rent = median(monthly_rent))
print(median_data)
```

```
## # A tibble: 3 x 2
##   furnished           median_rent
##   <chr>                     <dbl>
## 1 Fully Furnished            1700
## 2 Not Furnished              1000
## 3 Partially Furnished        1300
```

```
apt_rental %>%
  select(property_type, monthly_rent) %>%
  ggplot(aes(x = property_type, y = monthly_rent, fill = property_type)) +
  geom_boxplot() +
  labs(title = "Boxplot of Monthly Rent with Property Type") +
  theme(plot.margin = margin(b=30))
```




```
propertymedian_data = apt_rental %>%
  group_by(property_type) %>%
  summarize(propertymedian_rent = median(monthly_rent))
print(propertymedian_data)
```

```
## # A tibble: 6 x 2
##   property_type     propertymedian_rent
##   <chr>                           <dbl>
## 1 Apartment                        1099
## 2 Condominium                      1500
## 3 Duplex                           1500
## 4 Flat                              800
## 5 Service Residence                1600
## 6 Studio                           1200
```

Insights gained from all the visuals generated As a potential tenant, I'd always consider an affordable rent of below RM 1.5K as my budget, before deciding on a house. From this analysis, RM1.5K per month offers a lot of potential unit such as: **1)** Apartment, flat, and condominium shows a median of monthly rent close to RM 1.5k. However, service residence can be considered out of the budget. **2)**Not furnished and partially furnished rental are still within RM 1.5k budget, however a fully furnished unit median is around RM1.7k **3)** Monthlyl rental for units in Kl are slighly higher that units in Selangor. **4)** Vicinity of public transport, whether it is near KTM/LRT or not, does not significantly affect monthly rental as much. **5)** Will proceed on using a machine learning model to explore more on variables correlation and feature importance.

Model prediction using Random Forest

```
ohe_df = filteredrental %>%
  filter(property_type %in%  c("Apartment", "Condominium", "Service Residence", "Flat", "Studio", "Duple
  mutate(
    public_transport = str_extract(additional_facilities, "Near KTM/LRT"),
    public_transport = replace(public_transport, is.na(public_transport), "Not Near KTM/LRT")) %>%
  filter(complete.cases(.))

install.packages("fastDummies")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'fastDummies' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```
library(fastDummies)
```

```
## Warning: package 'fastDummies' was built under R version 4.3.1

## Thank you for using fastDummies!

## To acknowledge our work, please cite the package:

## Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from
```

```
ohe_df = dummy_columns(ohe_df, select_column = c("property_type", "furnished", "region", "public_transp
  select(-'ads_id')
```

Removing non-numeric column

```
ohe_numerical = list()
for (col in names(ohe_df)) {
  if (is.numeric(ohe_df[[col]]))
    ohe_numerical[[col]] = ohe_df[[col]]
}
ohe_num_df = data.frame(ohe_numerical)
```

```
cor(ohe_num_df)
```

```
##                                completion_year monthly_rent      parking
## completion_year                    1.000000000  0.156476519  0.302906721
## monthly_rent                       0.156476519  1.000000000  0.236038936
## parking                            0.302906721  0.236038936  1.000000000
## bathroom                          -0.055847337  0.243541825  0.308305121
## size                              -0.212728460  0.336251870  0.326509749
```

```
## property_type_Apartment              -0.077835651 -0.228439517 -0.090615866
## property_type_Condominium            -0.213141725  0.090509125  0.064336949
## property_type_Duplex                 -0.013414094 -0.007075511 -0.024944866
## property_type_Flat                   -0.037364657 -0.053604933 -0.008434526
## property_type_Service.Residence       0.285258181  0.091137882  0.016522652
## property_type_Studio                  0.003068963 -0.073679369 -0.070723932
## furnished_Fully.Furnished            -0.030513273  0.399213816 -0.037552978
## furnished_Not.Furnished              -0.019914078 -0.301919267 -0.053978944
## furnished_Partially.Furnished         0.043506532 -0.212670940  0.072224913
## region_Kuala.Lumpur                   0.026772980  0.355657928 -0.117887470
## region_Selangor                      -0.026772980 -0.355657928  0.117887470
## public_transport_Near.KTM.LRT         0.061270312  0.045134140 -0.012382742
## public_transport_Not.Near.KTM.LRT    -0.061270312 -0.045134140  0.012382742
##                                       bathroom        size
## completion_year                   -5.584734e-02 -0.21272846
## monthly_rent                       2.435418e-01  0.33625187
## parking                            3.083051e-01  0.32650975
## bathroom                           1.000000e+00  0.68705259
## size                               6.870526e-01  1.00000000
## property_type_Apartment            5.399443e-02 -0.05873774
## property_type_Condominium          1.830820e-01  0.32423788
## property_type_Duplex              -6.235472e-05 -0.02073657
## property_type_Flat                -1.981827e-02 -0.02961620
## property_type_Service.Residence   -1.958776e-01 -0.26250571
## property_type_Studio              -1.732475e-01 -0.15690403
## furnished_Fully.Furnished         -1.665083e-01 -0.11432207
## furnished_Not.Furnished            2.501572e-02 -0.02912032
## furnished_Partially.Furnished      1.526699e-01  0.13416720
## region_Kuala.Lumpur                1.723227e-01  0.13232738
## region_Selangor                   -1.723227e-01 -0.13232738
## public_transport_Near.KTM.LRT      5.954890e-03 -0.03088843
## public_transport_Not.Near.KTM.LRT -5.954890e-03  0.03088843
##                                   property_type_Apartment
## completion_year                                -0.07783565
## monthly_rent                                   -0.22843952
## parking                                        -0.09061587
## bathroom                                        0.05399443
## size                                           -0.05873774
## property_type_Apartment                         1.00000000
## property_type_Condominium                      -0.40806530
## property_type_Duplex                           -0.02324636
## property_type_Flat                             -0.01228613
## property_type_Service.Residence                -0.28973753
## property_type_Studio                           -0.03957417
## furnished_Fully.Furnished                      -0.14041567
## furnished_Not.Furnished                         0.16249386
## furnished_Partially.Furnished                   0.03911515
## region_Kuala.Lumpur                            -0.03012189
## region_Selangor                                 0.03012189
## public_transport_Near.KTM.LRT                  -0.01910161
## public_transport_Not.Near.KTM.LRT               0.01910161
##                                   property_type_Condominium
## completion_year                                 -0.21314172
## monthly_rent                                     0.09050912
```

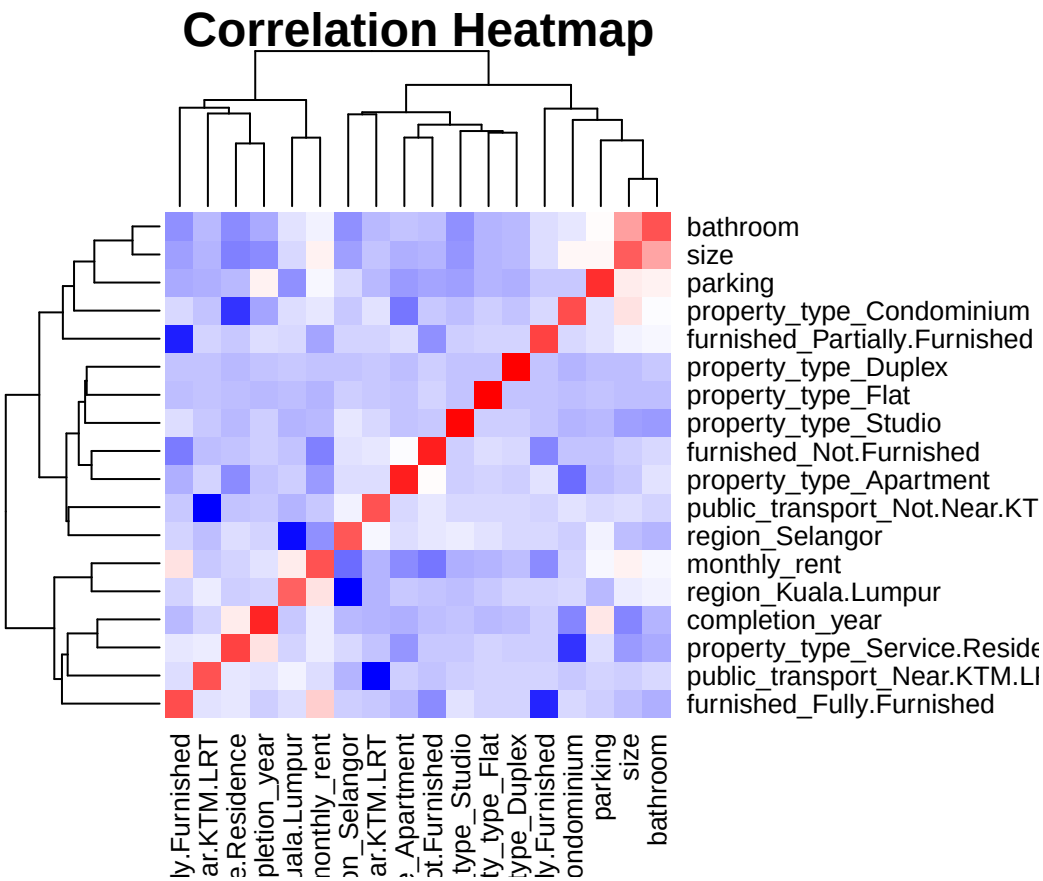
```
## parking                                          0.06433695
## bathroom                                         0.18308203
## size                                             0.32423788
## property_type_Apartment                         -0.40806530
## property_type_Condominium                        1.00000000
## property_type_Duplex                            -0.05832651
## property_type_Flat                              -0.03082664
## property_type_Service.Residence                 -0.72696903
## property_type_Studio                            -0.09929399
## furnished_Fully.Furnished                        0.01750938
## furnished_Not.Furnished                         -0.05456261
## furnished_Partially.Furnished                    0.01686493
## region_Kuala.Lumpur                              0.05064233
## region_Selangor                                 -0.05064233
## public_transport_Near.KTM.LRT                   -0.06424446
## public_transport_Not.Near.KTM.LRT                0.06424446
##                                  property_type_Duplex property_type_Flat
## completion_year                        -1.341409e-02       -0.037364657
## monthly_rent                           -7.075511e-03       -0.053604933
## parking                                -2.494487e-02       -0.008434526
## bathroom                               -6.235472e-05       -0.019818267
## size                                   -2.073657e-02       -0.029616200
## property_type_Apartment                -2.324636e-02       -0.012286127
## property_type_Condominium              -5.832651e-02       -0.030826638
## property_type_Duplex                    1.000000e+00       -0.001756109
## property_type_Flat                     -1.756109e-03        1.000000000
## property_type_Service.Residence        -4.141342e-02       -0.021887757
## property_type_Studio                   -5.656505e-03       -0.002989567
## furnished_Fully.Furnished              -8.326850e-03       -0.020093195
## furnished_Not.Furnished                 1.662659e-02        0.044979128
## furnished_Partially.Furnished          -2.111596e-03       -0.008174934
## region_Kuala.Lumpur                     9.632494e-03       -0.031487338
## region_Selangor                        -9.632494e-03        0.031487338
## public_transport_Near.KTM.LRT          -9.750312e-03       -0.007282756
## public_transport_Not.Near.KTM.LRT       9.750312e-03        0.007282756
##                                  property_type_Service.Residence
## completion_year                                       0.28525818
## monthly_rent                                          0.09113788
## parking                                               0.01652265
## bathroom                                             -0.19587765
## size                                                 -0.26250571
## property_type_Apartment                              -0.28973753
## property_type_Condominium                            -0.72696903
## property_type_Duplex                                 -0.04141342
## property_type_Flat                                   -0.02188776
## property_type_Service.Residence                       1.00000000
## property_type_Studio                                 -0.07050145
## furnished_Fully.Furnished                             0.07652797
## furnished_Not.Furnished                              -0.06189621
## furnished_Partially.Furnished                        -0.03822053
## region_Kuala.Lumpur                                  -0.01252240
## region_Selangor                                       0.01252240
## public_transport_Near.KTM.LRT                         0.08798382
## public_transport_Not.Near.KTM.LRT                    -0.08798382
```

```
##                                   property_type_Studio
## completion_year                          0.003068963
## monthly_rent                            -0.073679369
## parking                                 -0.070723932
## bathroom                                -0.173247469
## size                                    -0.156904033
## property_type_Apartment                 -0.039574170
## property_type_Condominium               -0.099293992
## property_type_Duplex                    -0.005656505
## property_type_Flat                      -0.002989567
## property_type_Service.Residence         -0.070501451
## property_type_Studio                     1.000000000
## furnished_Fully.Furnished                0.049162540
## furnished_Not.Furnished                 -0.021389019
## furnished_Partially.Furnished           -0.036200358
## region_Kuala.Lumpur                     -0.087785523
## region_Selangor                          0.087785523
## public_transport_Near.KTM.LRT           -0.022267228
## public_transport_Not.Near.KTM.LRT        0.022267228
##                                   furnished_Fully.Furnished
## completion_year                                -0.03051327
## monthly_rent                                    0.39921382
## parking                                        -0.03755298
## bathroom                                       -0.16650829
## size                                           -0.11432207
## property_type_Apartment                        -0.14041567
## property_type_Condominium                       0.01750938
## property_type_Duplex                           -0.00832685
## property_type_Flat                             -0.02009319
## property_type_Service.Residence                 0.07652797
## property_type_Studio                            0.04916254
## furnished_Fully.Furnished                       1.00000000
## furnished_Not.Furnished                        -0.33215999
## furnished_Partially.Furnished                  -0.80157208
## region_Kuala.Lumpur                             0.02560502
## region_Selangor                                -0.02560502
## public_transport_Near.KTM.LRT                   0.05292138
## public_transport_Not.Near.KTM.LRT              -0.05292138
##                                   furnished_Not.Furnished
## completion_year                             -0.01991408
## monthly_rent                                -0.30191927
## parking                                     -0.05397894
## bathroom                                     0.02501572
## size                                        -0.02912032
## property_type_Apartment                      0.16249386
## property_type_Condominium                   -0.05456261
## property_type_Duplex                         0.01662659
## property_type_Flat                           0.04497913
## property_type_Service.Residence             -0.06189621
## property_type_Studio                        -0.02138902
## furnished_Fully.Furnished                   -0.33215999
## furnished_Not.Furnished                      1.00000000
## furnished_Partially.Furnished               -0.29770116
## region_Kuala.Lumpur                         -0.06034122
```

```
## region_Selangor                                0.06034122
## public_transport_Near.KTM.LRT                 -0.08211101
## public_transport_Not.Near.KTM.LRT              0.08211101
##                                 furnished_Partially.Furnished
## completion_year                                   0.043506532
## monthly_rent                                     -0.212670940
## parking                                           0.072224913
## bathroom                                          0.152669936
## size                                              0.134167196
## property_type_Apartment                           0.039115154
## property_type_Condominium                         0.016864935
## property_type_Duplex                             -0.002111596
## property_type_Flat                               -0.008174934
## property_type_Service.Residence                  -0.038220534
## property_type_Studio                             -0.036200358
## furnished_Fully.Furnished                        -0.801572078
## furnished_Not.Furnished                          -0.297701158
## furnished_Partially.Furnished                     1.000000000
## region_Kuala.Lumpur                               0.012334133
## region_Selangor                                  -0.012334133
## public_transport_Near.KTM.LRT                    -0.001513808
## public_transport_Not.Near.KTM.LRT                 0.001513808
##                                 region_Kuala.Lumpur region_Selangor
## completion_year                        0.026772980    -0.026772980
## monthly_rent                           0.355657928    -0.355657928
## parking                               -0.117887470     0.117887470
## bathroom                               0.172322725    -0.172322725
## size                                   0.132327378    -0.132327378
## property_type_Apartment               -0.030121892     0.030121892
## property_type_Condominium              0.050642331    -0.050642331
## property_type_Duplex                   0.009632494    -0.009632494
## property_type_Flat                    -0.031487338     0.031487338
## property_type_Service.Residence       -0.012522397     0.012522397
## property_type_Studio                  -0.087785523     0.087785523
## furnished_Fully.Furnished              0.025605022    -0.025605022
## furnished_Not.Furnished               -0.060341223     0.060341223
## furnished_Partially.Furnished          0.012334133    -0.012334133
## region_Kuala.Lumpur                    1.000000000    -1.000000000
## region_Selangor                       -1.000000000     1.000000000
## public_transport_Near.KTM.LRT          0.140421382    -0.140421382
## public_transport_Not.Near.KTM.LRT     -0.140421382     0.140421382
##                                 public_transport_Near.KTM.LRT
## completion_year                                   0.061270312
## monthly_rent                                      0.045134140
## parking                                          -0.012382742
## bathroom                                          0.005954890
## size                                             -0.030888429
## property_type_Apartment                          -0.019101610
## property_type_Condominium                        -0.064244459
## property_type_Duplex                             -0.009750312
## property_type_Flat                               -0.007282756
## property_type_Service.Residence                   0.087983822
## property_type_Studio                             -0.022267228
## furnished_Fully.Furnished                         0.052921381
```

```
## furnished_Not.Furnished                              -0.082111007
## furnished_Partially.Furnished                        -0.001513808
## region_Kuala.Lumpur                                   0.140421382
## region_Selangor                                      -0.140421382
## public_transport_Near.KTM.LRT                         1.000000000
## public_transport_Not.Near.KTM.LRT                    -1.000000000
##                                    public_transport_Not.Near.KTM.LRT
## completion_year                                          -0.061270312
## monthly_rent                                             -0.045134140
## parking                                                   0.012382742
## bathroom                                                 -0.005954890
## size                                                      0.030888429
## property_type_Apartment                                   0.019101610
## property_type_Condominium                                 0.064244459
## property_type_Duplex                                      0.009750312
## property_type_Flat                                        0.007282756
## property_type_Service.Residence                          -0.087983822
## property_type_Studio                                      0.022267228
## furnished_Fully.Furnished                                -0.052921381
## furnished_Not.Furnished                                   0.082111007
## furnished_Partially.Furnished                             0.001513808
## region_Kuala.Lumpur                                      -0.140421382
## region_Selangor                                           0.140421382
## public_transport_Near.KTM.LRT                            -1.000000000
## public_transport_Not.Near.KTM.LRT                         1.000000000
```

```
correlation_table = cor(ohe_num_df)
cor_df = data.frame(variable = colnames(ohe_num_df), correlation = correlation_table)
heatmap(correlation_table,
        col = colorRampPalette(c("blue", "white", "red"))(100),
        main = "Correlation Heatmap",
        xlab = "",
        ylab = "")
```



Train model

```
install.packages("randomForest")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'randomForest' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'randomForest'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\1\AppData\Local\R\win-library\4.3\00LOCK\randomForest\libs\x64\randomForest.dll
## to
## C:\Users\1\AppData\Local\R\win-library\4.3\randomForest\libs\x64\randomForest.dll:
## Permission denied

## Warning: restored 'randomForest'

##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpAlY9F6\downloaded_packages
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.3.1
```

```
## randomForest 4.7-1.1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

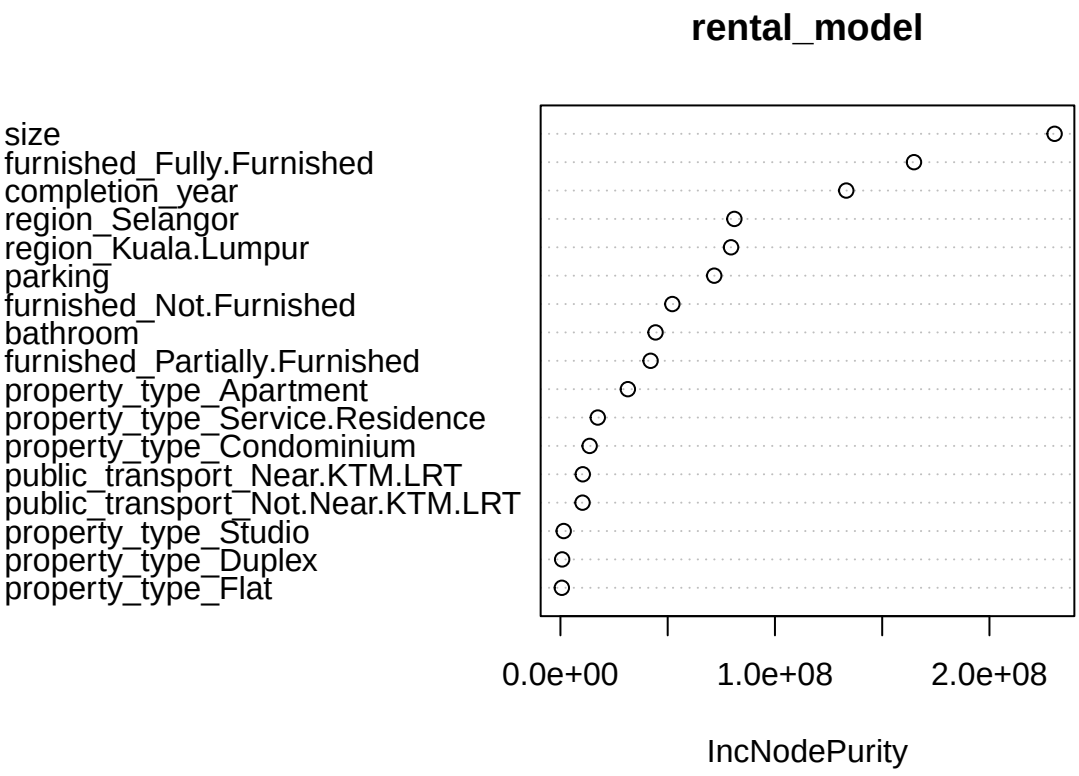
```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
set.seed(123)
train_index = sample(1:nrow(ohe_num_df), 0.8 * nrow(ohe_num_df))
train_data = ohe_num_df[train_index, ]
test_data = ohe_num_df[-train_index, ]

# Define the target variable and predictors
target_variable = "monthly_rent"
predictors = setdiff(names(ohe_num_df), target_variable)

# Train a Random Forest model
rental_model = randomForest(formula(paste(target_variable, "~", paste(predictors, collapse = "+"))),
                      data = train_data)

# Print the model summary
print(rental_model)
```

```
##
## Call:
##  randomForest(formula = formula(paste(target_variable, "~", paste(predictors,      collapse = "+"))
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
##           Mean of squared residuals: 84714.21
##                     % Var explained: 63.45
```

Evaluating feature importance on targeted variable of 'monthly rent'

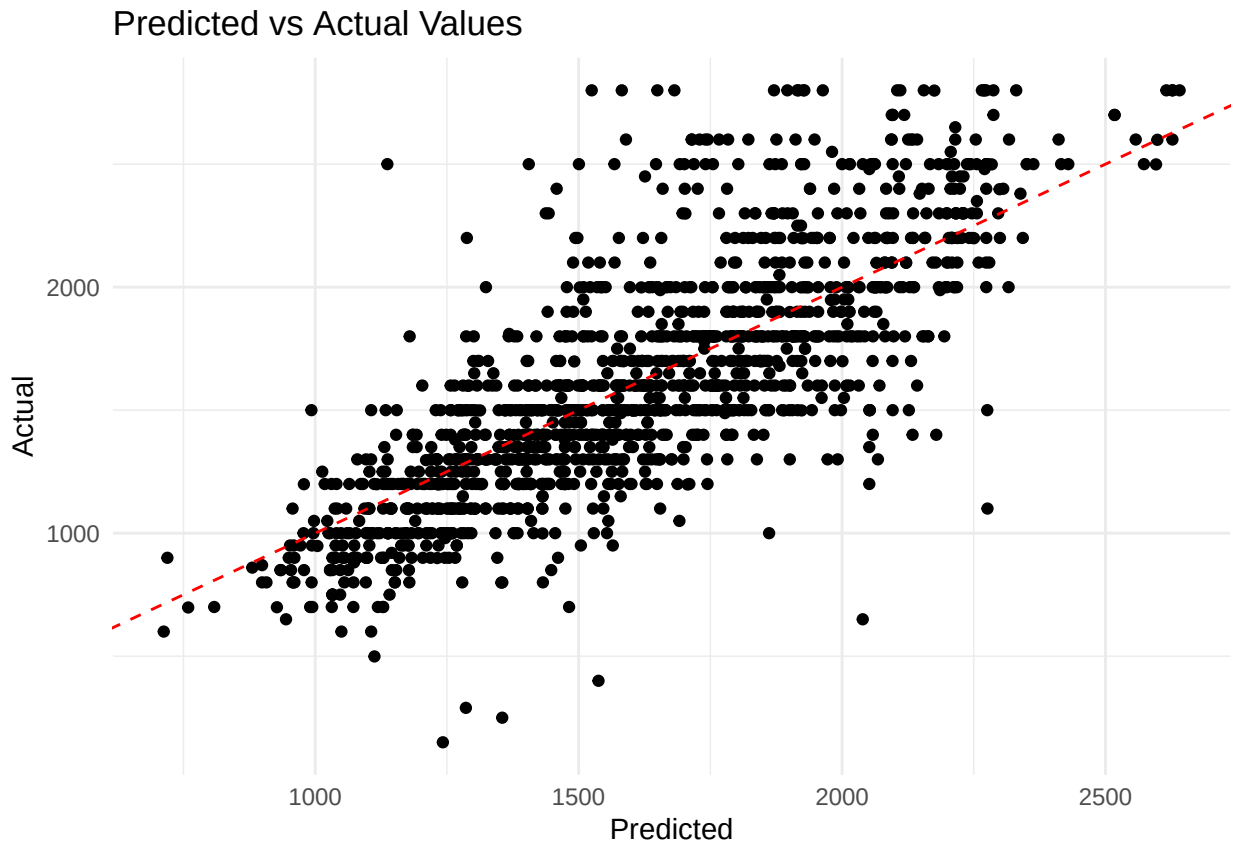
```
varImpPlot(rental_model)
```




```
prediction = predict(rental_model, newdata = test_data)

actual = test_data$monthly_rent
ss_total = sum((actual - mean(actual))^2)
ss_residual = sum((actual - prediction)^2)
r_squared = 1 - (ss_residual / ss_total)

print(r_squared)
```

```
## [1] 0.6225412
```

```
# Create a data frame for plotting
prediction_df = data.frame(Actual = actual, Predicted = prediction)

# Create Scatter Plot
ggplot(prediction_df, aes(x = Predicted, y = Actual)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +  # Add diagonal line
  labs(title = "Predicted vs Actual Values", x = "Predicted", y = "Actual") +
  theme_minimal()
```




```
head(prediction_df)
```

```
##    Actual Predicted
## 4    2000  2031.451
## 10   1199  1473.708
## 11   1100  1295.426
## 12   2500  2166.986
## 14   1400  1726.554
## 15   1800  1474.362
```

Conclusion

Random Forest model r-squared value is at 0.6225412, which only represent 60% of accuracy instead of 100%. Model may be needed to be revised, improved or undergo hyperparameter tuning (of which I'm still exploring). As for now, this model has illustrated that top 3 variables that highly affected monthly rent in Klang Valley is the unit's size, it's furnished condition and the year it's build. Personally, I always thought that public transport facilities (whether its within/outside vicinity) will highly affect rental price, but my analysis shows the contrast. This also shows that Klang Valley is definitely a city of cars, and most its potential residents are car owner, hence why it doesnt affect rental price.