property analysis

Zhong

2025-06-30

#Step 1 clean data

# 1. Load required packages  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.4.3

## Warning: package 'ggplot2' was built under R version 4.4.2

## Warning: package 'tibble' was built under R version 4.4.2

## Warning: package 'tidyr' was built under R version 4.4.2

## Warning: package 'readr' was built under R version 4.4.2

## Warning: package 'purrr' was built under R version 4.4.2

## Warning: package 'dplyr' was built under R version 4.4.2

## Warning: package 'stringr' was built under R version 4.4.2

## Warning: package 'forcats' was built under R version 4.4.2

## Warning: package 'lubridate' was built under R version 4.4.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(stringr)  
  
# 2. Read raw data  
df <- read\_csv("condo\_prices.csv", show\_col\_types = FALSE)

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

# 3. Clean price field  
df <- df %>%  
 mutate(price = as.numeric(str\_replace\_all(as.character(price), "[^0-9.]", "")))  
  
# 4. Extract state from address  
df <- df %>%  
 mutate(state = str\_trim(word(Address, -1, sep = fixed(","))),  
 state = replace\_na(state, "Unknown"))  
  
# 5. Convert size, bedroom, bathroom, and completion year to numeric  
df <- df %>%  
 mutate(  
 size\_sqft = as.numeric(str\_extract(str\_remove\_all(`Property Size`, ","), "\\d+")),  
 Bedroom = as.numeric(str\_replace\_all(as.character(Bedroom), "[^0-9.]", "")),  
 Bathroom = as.numeric(str\_replace\_all(as.character(Bathroom), "[^0-9.]", "")),  
 `Completion Year` = as.numeric(str\_replace\_all(as.character(`Completion Year`), "[^0-9]", ""))  
 )  
  
# 6. Add missing indicators and impute by group median  
num\_cols <- c("Bedroom", "Bathroom", "size\_sqft", "Completion Year")  
for (col in num\_cols) {  
 df[[paste0(col, "\_missing")]] <- as.integer(is.na(df[[col]]))  
}  
df <- df %>%  
 group\_by(state, `Property Type`) %>%  
 mutate(across(all\_of(num\_cols),  
 ~ ifelse(is.na(.x), median(.x, na.rm = TRUE), .x))) %>%  
 ungroup()  
  
# 7. Create age variable (capped between 0 and 100)  
df <- df %>%  
 mutate(age = 2025 - `Completion Year`,  
 age = ifelse(age < 0 | age > 100, NA, age))  
  
# 8. Process parking lot data  
df <- df %>%  
 mutate(  
 parking\_missing = as.integer(is.na(`Parking Lot`)),  
 parking\_lot = as.integer(replace\_na(as.numeric(`Parking Lot`), 0))  
 )

## Warning: There was 1 warning in `mutate()`.  
## ℹ In argument: `parking\_lot = as.integer(replace\_na(as.numeric(`Parking Lot`),  
## 0))`.  
## Caused by warning in `replace\_na()`:  
## ! NAs introduced by coercion

# 9. Generate amenity count variable  
split\_len <- function(x) ifelse(  
 is.na(x) | x == "",  
 0L,  
 lengths(str\_split(x, "[,\\n;]+"))  
)  
df <- df %>%  
 mutate(  
 amenity\_count = split\_len(Facilities),  
 facilities\_missing = as.integer(is.na(Facilities))  
 )  
  
# 10. Generate transport variables  
df <- df %>%  
 mutate(  
 bus\_stop\_count = split\_len(`Bus Stop`),  
 railway\_count = split\_len(str\_c(`Nearby Railway Station`, `Railway Station`, sep = ",")),  
 transport\_index = bus\_stop\_count + railway\_count,  
 transport\_missing =  
 as.integer(is.na(`Bus Stop`) &  
 is.na(`Nearby Railway Station`) &  
 is.na(`Railway Station`))  
 )  
  
# 11. Handle missing categorical variables  
cat\_cols <- c("Tenure Type", "Property Type", "state")  
for (col in cat\_cols) {  
 miss\_col <- paste0(str\_replace\_all(col, " ", "\_"), "\_missing")  
 df[[miss\_col]] <- as.integer(is.na(df[[col]]))  
 df[[col]] <- replace\_na(df[[col]], "Unknown")  
}  
  
# 12. Encode floor level range  
df <- df %>%  
 mutate(  
 floor\_level\_encoded = case\_when(  
 str\_detect(str\_to\_lower(`Floor Range`), "low") ~ 0,  
 str\_detect(str\_to\_lower(`Floor Range`), "middle") ~ 1,  
 str\_detect(str\_to\_lower(`Floor Range`), "high") ~ 2,  
 TRUE ~ NA\_real\_  
 )  
 )  
  
# 13. Clean number of floors and total units  
df <- df %>%  
 mutate(  
 `# of Floors` = as.numeric(na\_if(`# of Floors`, "-")),  
 `Total Units` = as.numeric(na\_if(`Total Units`, "-"))  
 )  
  
# 14. Remove redundant columns  
drop\_cols <- c(  
 "description", "Facilities", "Bus Stop", "Mall", "Nearby School", "Nearby Mall",  
 "Hospital", "School", "Park", "Railway Station", "Nearby Railway Station",  
 "Firm Type", "Firm Number", "REN Number", "Ad List", "Parking Lot",  
 "Property Size", "Address", "Building Name", "Developer"  
)  
df <- df %>% select(-any\_of(drop\_cols))  
  
# 15. Drop rows with missing price  
df <- df %>% filter(!is.na(price))

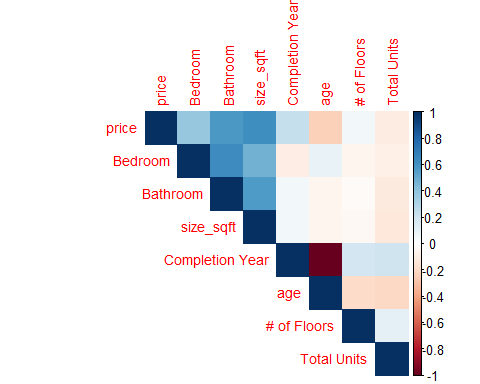
#step 2 EDA

# 1. Load required packages and read cleaned data  
library(tidyverse)  
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.2

## corrplot 0.95 loaded

df <- read\_csv("condo\_prices\_cleaned.csv", show\_col\_types = FALSE)  
  
# 2. Plot price distribution (histogram + density)  
p1 <- ggplot(df, aes(x = price)) +   
 geom\_histogram(binwidth = 50000, fill = "steelblue") +   
 labs(title = "Price Distribution", x = "Price (RM)", y = "Count") +  
 theme\_minimal(base\_size = 13)  
  
p2 <- ggplot(df, aes(x = price)) +   
 geom\_density(fill = "skyblue", alpha = 0.6) +   
 labs(title = "Price Density Plot", x = "Price (RM)", y = "Density") +  
 theme\_minimal(base\_size = 13)  
  
# 3. Scatter plot of size vs. price (with regression line)  
p3 <- ggplot(df, aes(x = size\_sqft, y = price)) +   
 geom\_point(alpha = 0.4, color = "darkorange") +  
 geom\_smooth(method = "lm", se = FALSE, color = "black") +  
 labs(title = "Price vs Size", x = "Size (sqft)", y = "Price (RM)") +  
 theme\_minimal(base\_size = 13)  
  
# 4. Boxplot of top 10 states by median price  
top10\_states <- df %>%  
 group\_by(state) %>%  
 summarise(median\_price = median(price, na.rm = TRUE)) %>%  
 arrange(desc(median\_price)) %>%  
 slice(1:10) %>%  
 pull(state)  
  
p4 <- ggplot(df %>% filter(state %in% top10\_states),  
 aes(x = reorder(state, price, median), y = price)) +  
 geom\_boxplot(fill = "lightblue", outlier.size = 0.8) +  
 coord\_flip() +  
 labs(title = "Price Distribution",  
 x = "State", y = "Price (RM)") +  
 theme\_minimal(base\_size = 13)  
  
# 5. Correlation heatmap of numeric variables  
cor\_matrix <- cor(df %>%  
 select(price, Bedroom, Bathroom, size\_sqft, `Completion Year`, age, `# of Floors`, `Total Units`),  
 use = "complete.obs")  
p5 <- corrplot(cor\_matrix, method = "color", type = "upper", tl.cex = 0.9)



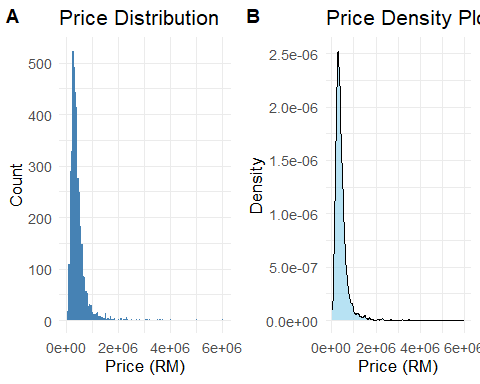
# 6. Top 10 states by average price per sqft  
df <- df %>%  
 mutate(price\_per\_sqft = price / size\_sqft)  
  
top\_pps <- df %>%  
 group\_by(state) %>%  
 summarise(mean\_pps = mean(price\_per\_sqft, na.rm = TRUE)) %>%  
 filter(mean\_pps < quantile(mean\_pps, 0.99)) %>%  
 arrange(desc(mean\_pps)) %>%  
 slice(1:10)  
  
p6 <- ggplot(top\_pps, aes(x = reorder(state, mean\_pps), y = mean\_pps)) +  
 geom\_col(fill = "royalblue") +  
 coord\_flip() +  
 labs(title = "Top 10 States by Price per Sqft", x = "State", y = "Price per Sqft (RM)") +  
 theme\_minimal(base\_size = 13)  
#combine plots  
library(cowplot)

## Warning: package 'cowplot' was built under R version 4.4.2

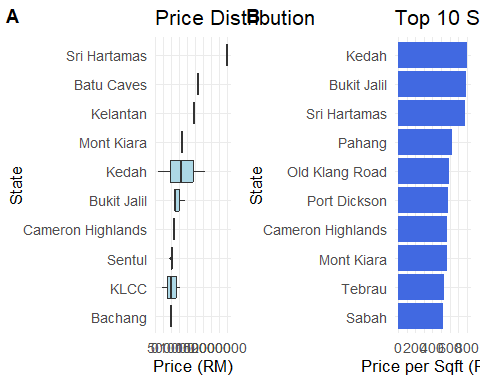
##   
## Attaching package: 'cowplot'

## The following object is masked from 'package:lubridate':  
##   
## stamp

plot\_grid(p1, p2, labels = c("A", "B"), ncol = 2)



plot\_grid(p4, p6, labels = c("A", "B"), ncol = 2)

 #XGBoost

# 1. Load modeling packages  
library(data.table)

## Warning: package 'data.table' was built under R version 4.4.2

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:lubridate':  
##   
## hour, isoweek, mday, minute, month, quarter, second, wday, week,  
## yday, year

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

library(caret)

## Warning: package 'caret' was built under R version 4.4.3

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(xgboost)

## Warning: package 'xgboost' was built under R version 4.4.3

##   
## Attaching package: 'xgboost'

## The following object is masked from 'package:dplyr':  
##   
## slice

library(Matrix)

## Warning: package 'Matrix' was built under R version 4.4.3

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

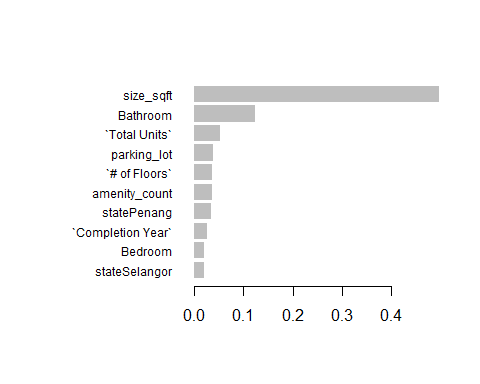
# 2. Read data and split into training/testing sets  
df <- fread("condo\_prices\_cleaned.csv")  
set.seed(2025)  
train\_idx <- createDataPartition(df$price, p = 0.8, list = FALSE)  
train\_df <- df[train\_idx]  
test\_df <- df[-train\_idx]  
  
# 3. Remove constant columns  
all\_feats <- setdiff(names(df), "price")  
is\_const <- sapply(train\_df[, all\_feats, with = FALSE], function(x) length(unique(x)) <= 1)  
feat\_final <- setdiff(all\_feats, names(is\_const)[is\_const])  
  
# 4. One-hot encoding  
dv <- dummyVars(~ ., data = train\_df[, feat\_final, with = FALSE])  
train\_m <- predict(dv, newdata = train\_df[, feat\_final, with = FALSE])  
test\_m <- predict(dv, newdata = test\_df[, feat\_final, with = FALSE])  
  
# 5. Align test dummy columns with training  
miss <- setdiff(colnames(train\_m), colnames(test\_m))  
if (length(miss) > 0) {  
 zero\_mat <- matrix(0, nrow(test\_m), length(miss), dimnames = list(NULL, miss))  
 test\_m <- cbind(test\_m, zero\_mat)  
}  
test\_m <- test\_m[, colnames(train\_m), drop = FALSE]  
  
# 6. Create DMatrix objects  
dtrain <- xgb.DMatrix(data = train\_m, label = train\_df$price)  
dtest <- xgb.DMatrix(data = test\_m, label = test\_df$price)  
  
# 7. Set parameters and apply cross-validation with early stopping  
params <- list(  
 objective = "reg:squarederror",  
 eval\_metric = "rmse",  
 eta = 0.1,  
 max\_depth = 6,  
 subsample = 0.8,  
 colsample\_bytree = 0.8  
)  
cv\_res <- xgb.cv(params, dtrain, nrounds = 1000, nfold = 5,  
 early\_stopping\_rounds = 20, verbose = 1)

## [1] train-rmse:500345.178117+7514.826708 test-rmse:501198.617621+33570.171647   
## Multiple eval metrics are present. Will use test\_rmse for early stopping.  
## Will train until test\_rmse hasn't improved in 20 rounds.  
##   
## [2] train-rmse:458046.918219+7166.732802 test-rmse:460262.958467+32454.901357   
## [3] train-rmse:419840.218757+7165.655902 test-rmse:425016.581853+29556.771353   
## [4] train-rmse:386035.147006+6710.929630 test-rmse:394580.262240+28734.510414   
## [5] train-rmse:354903.692882+6308.523343 test-rmse:365890.786273+27684.130445   
## [6] train-rmse:326846.544469+5915.548169 test-rmse:340797.335556+26477.896439   
## [7] train-rmse:302876.106612+6668.503035 test-rmse:320143.207823+25730.833117   
## [8] train-rmse:280454.300787+6492.229695 test-rmse:300706.711437+24716.916579   
## [9] train-rmse:260954.554821+5672.238689 test-rmse:284485.077511+26402.214478   
## [10] train-rmse:243705.531771+5986.329964 test-rmse:270033.065840+25557.797780   
## [11] train-rmse:227932.312718+6019.212056 test-rmse:256475.513959+25679.928628   
## [12] train-rmse:213117.734809+5859.887688 test-rmse:246474.754199+24567.807867   
## [13] train-rmse:200014.713002+5756.268073 test-rmse:237831.517326+23197.879226   
## [14] train-rmse:188697.212925+5802.765649 test-rmse:230304.998286+22772.455318   
## [15] train-rmse:178349.858189+5566.175578 test-rmse:223952.217968+20605.493491   
## [16] train-rmse:169404.998199+5229.465168 test-rmse:218457.115793+19943.591061   
## [17] train-rmse:160932.419774+4807.118793 test-rmse:213359.270535+19924.507835   
## [18] train-rmse:153776.597806+4354.067827 test-rmse:209300.264967+19040.413890   
## [19] train-rmse:147462.873795+4322.096216 test-rmse:206418.813037+18663.922235   
## [20] train-rmse:141433.843628+4051.386192 test-rmse:203754.830265+18274.600059   
## [21] train-rmse:136233.035035+3777.930970 test-rmse:200890.324189+18684.092593   
## [22] train-rmse:131944.506853+3579.992047 test-rmse:199045.493263+18402.414085   
## [23] train-rmse:127415.362515+3403.477279 test-rmse:197562.567062+17807.587092   
## [24] train-rmse:123398.916214+3539.383188 test-rmse:195700.511017+17259.998439   
## [25] train-rmse:119736.926917+3362.778884 test-rmse:194726.700168+17068.761479   
## [26] train-rmse:116598.791044+3212.199293 test-rmse:194318.798911+17202.376796   
## [27] train-rmse:113712.322442+3115.652665 test-rmse:194033.498872+17230.842398   
## [28] train-rmse:111076.655676+3173.630515 test-rmse:193285.196211+16724.836113   
## [29] train-rmse:108867.982285+2916.117983 test-rmse:192288.834904+16514.655811   
## [30] train-rmse:106750.034675+2767.197683 test-rmse:191452.696384+16468.274565   
## [31] train-rmse:104814.147722+2668.019457 test-rmse:190476.541957+15918.813204   
## [32] train-rmse:103030.816174+2583.288740 test-rmse:189947.180917+15947.626524   
## [33] train-rmse:101450.148982+2797.932556 test-rmse:189320.728096+15774.212664   
## [34] train-rmse:100039.404816+2635.029782 test-rmse:189178.114136+15970.149831   
## [35] train-rmse:98495.348245+2557.819577 test-rmse:189104.212611+16389.608182   
## [36] train-rmse:97099.283859+2593.858140 test-rmse:189283.999269+16741.820992   
## [37] train-rmse:95747.790333+2370.321392 test-rmse:189662.068260+16968.458149   
## [38] train-rmse:94400.274888+2400.806096 test-rmse:189445.920051+16746.562235   
## [39] train-rmse:93155.709307+2269.297153 test-rmse:189332.063097+16345.559589   
## [40] train-rmse:92194.585658+2435.012437 test-rmse:188844.283810+16200.182714   
## [41] train-rmse:91362.879816+2299.053032 test-rmse:189143.604525+16027.519698   
## [42] train-rmse:90520.306008+2117.999888 test-rmse:188692.871310+15844.773752   
## [43] train-rmse:89687.626806+2174.686955 test-rmse:188577.884338+15934.426765   
## [44] train-rmse:88711.587217+2069.371612 test-rmse:188304.413393+15554.282588   
## [45] train-rmse:87964.056436+2281.013368 test-rmse:188251.484957+15474.131646   
## [46] train-rmse:87367.417310+2324.724088 test-rmse:188502.968227+15923.890165   
## [47] train-rmse:86651.930285+2275.291756 test-rmse:188138.970774+15591.653866   
## [48] train-rmse:86052.942286+2297.476194 test-rmse:187994.723436+15389.527390   
## [49] train-rmse:85387.486937+2222.895316 test-rmse:187978.301808+15640.063350   
## [50] train-rmse:84760.110659+2312.833287 test-rmse:187800.497388+15611.368397   
## [51] train-rmse:84172.515699+2498.755381 test-rmse:187848.719652+15668.541805   
## [52] train-rmse:83732.093475+2553.075989 test-rmse:188213.835883+16016.611358   
## [53] train-rmse:83215.172355+2614.944489 test-rmse:187805.844776+15770.801531   
## [54] train-rmse:82794.262958+2579.430972 test-rmse:187834.679101+16232.531110   
## [55] train-rmse:82340.597825+2461.783173 test-rmse:187874.619678+16316.794679   
## [56] train-rmse:81984.517841+2423.007933 test-rmse:187644.459757+16046.716047   
## [57] train-rmse:81448.953554+2351.683551 test-rmse:187452.299123+15925.643491   
## [58] train-rmse:81015.972936+2269.105905 test-rmse:187386.395843+15924.770951   
## [59] train-rmse:80637.052623+2230.542542 test-rmse:187227.998178+15673.144219   
## [60] train-rmse:80164.647292+2368.552301 test-rmse:186856.026372+15424.741756   
## [61] train-rmse:79782.586143+2290.900536 test-rmse:186711.845821+15252.859561   
## [62] train-rmse:79343.980188+2299.018302 test-rmse:186708.960324+15265.419066   
## [63] train-rmse:78765.397821+2160.747778 test-rmse:186745.498077+15280.783933   
## [64] train-rmse:78409.956740+2158.428872 test-rmse:186711.044219+15485.964567   
## [65] train-rmse:78055.058050+2110.585910 test-rmse:186697.873303+15304.812692   
## [66] train-rmse:77615.331455+1962.661308 test-rmse:186335.159799+15356.708053   
## [67] train-rmse:77101.023632+2024.167399 test-rmse:186261.070374+15349.547279   
## [68] train-rmse:76761.349791+2025.951519 test-rmse:186199.027963+15072.356265   
## [69] train-rmse:76338.241903+2046.153550 test-rmse:186263.852543+15217.388601   
## [70] train-rmse:75999.946511+1996.643053 test-rmse:186272.151333+15200.558994   
## [71] train-rmse:75513.365608+1997.248171 test-rmse:186232.940583+15256.123874   
## [72] train-rmse:75142.269507+2036.960626 test-rmse:186292.389404+15210.155436   
## [73] train-rmse:74733.378823+2136.878032 test-rmse:186274.710064+15126.146205   
## [74] train-rmse:74420.239524+2154.165904 test-rmse:186236.971723+15064.509145   
## [75] train-rmse:73968.655216+2064.157620 test-rmse:186022.457701+14931.917392   
## [76] train-rmse:73591.090656+2083.732959 test-rmse:185959.514187+14953.024253   
## [77] train-rmse:73287.984525+2137.023095 test-rmse:185953.170083+14966.552075   
## [78] train-rmse:72981.869095+2126.003336 test-rmse:185891.013278+14885.145683   
## [79] train-rmse:72675.929187+2097.948404 test-rmse:185765.445337+14742.252918   
## [80] train-rmse:72284.594735+2178.777146 test-rmse:185545.615084+14615.079619   
## [81] train-rmse:72060.690175+2217.649448 test-rmse:185426.359645+14528.325141   
## [82] train-rmse:71779.017304+2115.028280 test-rmse:185303.396892+14643.091657   
## [83] train-rmse:71413.379827+2142.434895 test-rmse:185180.850996+14602.997426   
## [84] train-rmse:71062.118878+2105.113741 test-rmse:185139.241312+14600.095992   
## [85] train-rmse:70831.632178+2132.094063 test-rmse:185201.374282+14671.970349   
## [86] train-rmse:70589.847659+2104.702558 test-rmse:185317.642219+14742.788060   
## [87] train-rmse:70368.307444+2051.023973 test-rmse:185310.906101+14810.979161   
## [88] train-rmse:70093.752017+2173.211807 test-rmse:185243.360982+14763.343316   
## [89] train-rmse:69868.275956+2249.436832 test-rmse:185217.056848+14740.806811   
## [90] train-rmse:69540.371189+2255.165700 test-rmse:185034.537147+14679.630614   
## [91] train-rmse:69199.090589+2351.506321 test-rmse:185022.319674+14633.539313   
## [92] train-rmse:68932.709658+2384.823826 test-rmse:185031.272188+14657.662228   
## [93] train-rmse:68668.666397+2428.088549 test-rmse:184992.997643+14478.483162   
## [94] train-rmse:68393.296763+2320.785744 test-rmse:184957.585051+14491.164580   
## [95] train-rmse:68116.499270+2248.581307 test-rmse:184862.698185+14643.699396   
## [96] train-rmse:67731.957243+2301.547507 test-rmse:184774.993372+14656.203326   
## [97] train-rmse:67397.792192+2266.008576 test-rmse:184814.630879+14630.158210   
## [98] train-rmse:67144.001344+2308.201448 test-rmse:184756.246539+14664.583638   
## [99] train-rmse:66729.801397+2226.021693 test-rmse:184633.251714+14739.317187   
## [100] train-rmse:66462.529102+2220.016269 test-rmse:184597.122680+14702.698717   
## [101] train-rmse:66259.901410+2118.435483 test-rmse:184600.352714+14658.717494   
## [102] train-rmse:66116.546714+2127.215319 test-rmse:184542.403663+14569.378677   
## [103] train-rmse:65930.923981+2116.405547 test-rmse:184512.724176+14606.499206   
## [104] train-rmse:65743.984087+2033.795528 test-rmse:184508.893963+14600.977934   
## [105] train-rmse:65319.878056+2085.927878 test-rmse:184378.073092+14511.324144   
## [106] train-rmse:64981.943612+2135.249598 test-rmse:184231.423674+14454.508286   
## [107] train-rmse:64749.909383+2098.038887 test-rmse:184190.166135+14363.534259   
## [108] train-rmse:64366.122424+2037.349936 test-rmse:184175.564085+14412.792418   
## [109] train-rmse:64223.121646+1997.739433 test-rmse:184102.871268+14364.900328   
## [110] train-rmse:63934.078757+1875.455720 test-rmse:184053.459201+14444.921968   
## [111] train-rmse:63708.835301+1936.213275 test-rmse:183968.518155+14411.007922   
## [112] train-rmse:63403.356954+1894.666769 test-rmse:183962.361154+14401.003544   
## [113] train-rmse:63161.071115+2074.930633 test-rmse:183977.758728+14422.060202   
## [114] train-rmse:62871.369090+2048.427455 test-rmse:183950.478773+14347.567900   
## [115] train-rmse:62652.377124+1993.886287 test-rmse:183945.876375+14348.344462   
## [116] train-rmse:62511.985213+1979.943329 test-rmse:183964.935022+14408.069546   
## [117] train-rmse:62318.345060+1902.907708 test-rmse:183943.755246+14389.949831   
## [118] train-rmse:62103.585847+1922.163246 test-rmse:183878.405969+14337.769786   
## [119] train-rmse:61885.115180+1892.573155 test-rmse:183901.708813+14427.654610   
## [120] train-rmse:61709.748016+1930.561106 test-rmse:183897.494950+14390.536122   
## [121] train-rmse:61505.651921+1981.643776 test-rmse:183817.757101+14388.819066   
## [122] train-rmse:61295.459651+1983.298499 test-rmse:183773.782943+14346.894442   
## [123] train-rmse:61191.744094+1968.790477 test-rmse:183753.395485+14345.702104   
## [124] train-rmse:60941.466341+2041.534301 test-rmse:183653.947504+14272.241571   
## [125] train-rmse:60833.771472+2014.411808 test-rmse:183595.964455+14314.221795   
## [126] train-rmse:60539.519712+1924.025278 test-rmse:183490.923361+14335.299926   
## [127] train-rmse:60308.864147+1853.388654 test-rmse:183423.027300+14316.033503   
## [128] train-rmse:60076.179437+1897.024267 test-rmse:183350.353058+14259.572936   
## [129] train-rmse:59778.420382+1773.811375 test-rmse:183316.173187+14238.274034   
## [130] train-rmse:59544.082429+1825.351150 test-rmse:183244.421799+14191.405732   
## [131] train-rmse:59391.571091+1820.158813 test-rmse:183223.095014+14209.664407   
## [132] train-rmse:59057.036394+1792.159803 test-rmse:183196.174526+14184.351704   
## [133] train-rmse:58960.392768+1821.152718 test-rmse:183191.738490+14172.010914   
## [134] train-rmse:58836.485872+1884.641983 test-rmse:183180.536424+14128.539254   
## [135] train-rmse:58703.425410+1910.301193 test-rmse:183170.418918+14075.985218   
## [136] train-rmse:58429.197668+1845.248195 test-rmse:183147.341688+14087.042160   
## [137] train-rmse:58185.005980+1837.075479 test-rmse:183110.643719+14110.300881   
## [138] train-rmse:58006.446284+1772.731968 test-rmse:183098.337854+14167.973569   
## [139] train-rmse:57771.819091+1816.718743 test-rmse:183048.356135+14229.541330   
## [140] train-rmse:57637.434627+1785.773222 test-rmse:183007.263544+14269.174662   
## [141] train-rmse:57335.029919+1756.854945 test-rmse:182983.612013+14247.801477   
## [142] train-rmse:57126.553188+1839.319612 test-rmse:182931.061628+14178.250199   
## [143] train-rmse:56848.673261+1720.894655 test-rmse:182881.509132+14217.975167   
## [144] train-rmse:56646.120006+1656.394660 test-rmse:182860.793473+14224.884623   
## [145] train-rmse:56431.572743+1737.867484 test-rmse:182821.028144+14162.406940   
## [146] train-rmse:56185.512890+1734.404940 test-rmse:182759.085182+14127.546355   
## [147] train-rmse:55941.052266+1725.172998 test-rmse:182742.694059+14114.868306   
## [148] train-rmse:55863.191422+1756.780115 test-rmse:182730.389671+14114.962691   
## [149] train-rmse:55751.688618+1754.715452 test-rmse:182736.274881+14025.147937   
## [150] train-rmse:55619.451524+1736.671937 test-rmse:182718.812971+14062.001392   
## [151] train-rmse:55459.290200+1773.940031 test-rmse:182715.933862+13923.611411   
## [152] train-rmse:55280.131585+1754.253398 test-rmse:182665.643163+13895.325928   
## [153] train-rmse:55116.385210+1774.678486 test-rmse:182656.036031+13906.203942   
## [154] train-rmse:54949.066233+1760.608591 test-rmse:182627.824651+13899.610544   
## [155] train-rmse:54820.809267+1738.595065 test-rmse:182610.063651+13889.888242   
## [156] train-rmse:54692.953417+1692.727158 test-rmse:182578.115457+13872.753924   
## [157] train-rmse:54437.723871+1716.433164 test-rmse:182522.347358+13834.253628   
## [158] train-rmse:54343.760454+1740.082519 test-rmse:182455.965028+13944.868312   
## [159] train-rmse:54186.908442+1692.304704 test-rmse:182513.261106+13809.499225   
## [160] train-rmse:53949.784565+1627.399359 test-rmse:182505.169584+13855.067069   
## [161] train-rmse:53813.002343+1618.415780 test-rmse:182488.769259+13817.803720   
## [162] train-rmse:53571.151833+1687.240905 test-rmse:182480.009519+13850.216041   
## [163] train-rmse:53414.665045+1662.393882 test-rmse:182461.016338+13899.390978   
## [164] train-rmse:53217.446244+1636.015859 test-rmse:182483.719353+13944.711069   
## [165] train-rmse:52933.047881+1614.327648 test-rmse:182366.431201+13963.177291   
## [166] train-rmse:52674.784705+1616.664357 test-rmse:182347.981777+13951.496947   
## [167] train-rmse:52436.268008+1611.776070 test-rmse:182302.373779+13980.287391   
## [168] train-rmse:52268.271269+1638.941794 test-rmse:182261.761971+13980.827666   
## [169] train-rmse:52124.798800+1626.222687 test-rmse:182238.863904+13990.403957   
## [170] train-rmse:51925.539522+1639.314607 test-rmse:182256.572737+13936.509870   
## [171] train-rmse:51755.522672+1673.145546 test-rmse:182207.330917+13926.417353   
## [172] train-rmse:51616.089225+1714.094547 test-rmse:182208.529763+13926.618058   
## [173] train-rmse:51530.901797+1743.784772 test-rmse:182212.563123+13939.129003   
## [174] train-rmse:51352.095737+1725.482417 test-rmse:182187.200676+13925.627181   
## [175] train-rmse:51194.187918+1812.074372 test-rmse:182065.103709+13928.710436   
## [176] train-rmse:51040.645606+1875.804970 test-rmse:182050.374604+13921.073709   
## [177] train-rmse:50946.595732+1840.952090 test-rmse:182075.033001+13918.481022   
## [178] train-rmse:50778.216833+1924.521747 test-rmse:182039.208659+13927.084672   
## [179] train-rmse:50505.097677+1902.048822 test-rmse:182022.496428+13975.544365   
## [180] train-rmse:50340.932648+1901.618101 test-rmse:182006.932944+13985.352652   
## [181] train-rmse:50158.185904+1926.052636 test-rmse:181984.477774+14024.644798   
## [182] train-rmse:50054.385233+1916.383428 test-rmse:181990.377456+13993.870348   
## [183] train-rmse:49858.304352+1900.700547 test-rmse:181979.227719+14030.264545   
## [184] train-rmse:49680.666576+1888.540269 test-rmse:181938.879853+14024.533073   
## [185] train-rmse:49554.826221+1870.147855 test-rmse:181934.805998+14008.581121   
## [186] train-rmse:49500.634063+1881.203921 test-rmse:181944.549016+14008.294592   
## [187] train-rmse:49312.764903+1829.243297 test-rmse:181934.010224+13944.694103   
## [188] train-rmse:49179.125939+1790.852347 test-rmse:181948.584487+13930.443124   
## [189] train-rmse:49098.943572+1818.226729 test-rmse:181962.250568+13934.985660   
## [190] train-rmse:49022.705764+1792.872609 test-rmse:181955.044780+13967.161323   
## [191] train-rmse:48792.451349+1835.102023 test-rmse:181895.137188+13995.677372   
## [192] train-rmse:48668.112402+1865.935064 test-rmse:181894.368647+14000.438127   
## [193] train-rmse:48534.037123+1853.332238 test-rmse:181884.531493+14025.997258   
## [194] train-rmse:48403.416932+1811.187537 test-rmse:181885.020860+13999.553701   
## [195] train-rmse:48232.988266+1747.444225 test-rmse:181882.493664+13987.068297   
## [196] train-rmse:48073.987373+1705.297323 test-rmse:181860.227567+13978.709077   
## [197] train-rmse:47836.440954+1669.585492 test-rmse:181867.893438+13966.482144   
## [198] train-rmse:47687.661371+1637.772262 test-rmse:181903.265679+14000.893287   
## [199] train-rmse:47557.593420+1646.993538 test-rmse:181899.168319+14002.422075   
## [200] train-rmse:47428.592831+1692.991128 test-rmse:181845.832685+13972.178674   
## [201] train-rmse:47269.843086+1719.060917 test-rmse:181844.487335+13999.967436   
## [202] train-rmse:47139.939753+1749.591078 test-rmse:181873.853875+14031.082240   
## [203] train-rmse:47025.639611+1713.679755 test-rmse:181952.368389+14015.173503   
## [204] train-rmse:46842.058311+1686.962112 test-rmse:181926.700357+14015.507665   
## [205] train-rmse:46661.135446+1635.209735 test-rmse:181885.540027+14077.563599   
## [206] train-rmse:46559.260555+1679.922979 test-rmse:181873.661203+14090.806251   
## [207] train-rmse:46406.140303+1657.474793 test-rmse:181774.023418+14183.245911   
## [208] train-rmse:46331.268758+1628.973657 test-rmse:181772.410002+14165.099163   
## [209] train-rmse:46209.792801+1712.306698 test-rmse:181779.854364+14188.909490   
## [210] train-rmse:46031.821710+1657.245966 test-rmse:181776.311881+14238.992494   
## [211] train-rmse:45882.653181+1719.005580 test-rmse:181795.570737+14285.991398   
## [212] train-rmse:45774.706773+1769.445592 test-rmse:181784.462150+14256.112199   
## [213] train-rmse:45654.146735+1767.992101 test-rmse:181783.222963+14239.230659   
## [214] train-rmse:45477.339383+1662.544846 test-rmse:181760.996283+14226.616631   
## [215] train-rmse:45383.946797+1638.955323 test-rmse:181775.229107+14181.219648   
## [216] train-rmse:45235.712563+1664.380839 test-rmse:181755.903076+14169.346918   
## [217] train-rmse:45115.824208+1617.672323 test-rmse:181737.416167+14173.648840   
## [218] train-rmse:44974.752856+1635.275423 test-rmse:181728.755007+14196.370669   
## [219] train-rmse:44841.931622+1606.242357 test-rmse:181711.421045+14194.516445   
## [220] train-rmse:44715.154397+1623.435034 test-rmse:181700.707739+14182.883108   
## [221] train-rmse:44599.256230+1641.560537 test-rmse:181660.851238+14158.414897   
## [222] train-rmse:44430.146945+1558.428703 test-rmse:181651.750225+14178.568603   
## [223] train-rmse:44326.120657+1538.431956 test-rmse:181657.703305+14208.945023   
## [224] train-rmse:44199.498985+1514.498304 test-rmse:181632.746332+14150.519740   
## [225] train-rmse:44119.686619+1529.930958 test-rmse:181632.258973+14157.997757   
## [226] train-rmse:44009.374356+1527.021788 test-rmse:181625.087957+14147.485682   
## [227] train-rmse:43857.552404+1506.248559 test-rmse:181647.538202+14161.539134   
## [228] train-rmse:43737.986727+1483.537079 test-rmse:181652.735077+14153.078296   
## [229] train-rmse:43583.720393+1473.797725 test-rmse:181648.636302+14178.788534   
## [230] train-rmse:43458.809842+1442.636277 test-rmse:181640.659676+14170.444637   
## [231] train-rmse:43361.549495+1422.330467 test-rmse:181631.235365+14173.484520   
## [232] train-rmse:43260.451446+1383.456652 test-rmse:181589.464895+14202.854889   
## [233] train-rmse:43157.919228+1374.113757 test-rmse:181593.122772+14201.437016   
## [234] train-rmse:43005.195589+1418.865578 test-rmse:181570.395045+14217.064469   
## [235] train-rmse:42839.893319+1359.890195 test-rmse:181554.948133+14207.677368   
## [236] train-rmse:42687.239444+1453.207558 test-rmse:181546.156628+14221.563027   
## [237] train-rmse:42587.633568+1514.379069 test-rmse:181544.869112+14218.768075   
## [238] train-rmse:42439.030802+1507.172419 test-rmse:181526.034756+14208.917687   
## [239] train-rmse:42310.683074+1586.346942 test-rmse:181537.305629+14172.840782   
## [240] train-rmse:42150.285970+1597.151639 test-rmse:181508.087300+14170.196377   
## [241] train-rmse:42003.404425+1639.997367 test-rmse:181502.713284+14197.460546   
## [242] train-rmse:41870.062329+1643.356126 test-rmse:181496.302660+14185.976423   
## [243] train-rmse:41757.689623+1620.034334 test-rmse:181470.530838+14189.918109   
## [244] train-rmse:41603.857262+1622.313587 test-rmse:181473.612465+14166.574366   
## [245] train-rmse:41479.843156+1572.185314 test-rmse:181469.240743+14159.082892   
## [246] train-rmse:41311.141991+1585.923868 test-rmse:181435.586654+14124.004249   
## [247] train-rmse:41194.701246+1585.369918 test-rmse:181416.348761+14115.604353   
## [248] train-rmse:41064.167596+1644.995783 test-rmse:181414.393828+14109.291259   
## [249] train-rmse:40978.154577+1664.943588 test-rmse:181380.484969+14110.746118   
## [250] train-rmse:40899.078064+1661.851942 test-rmse:181370.647512+14125.143071   
## [251] train-rmse:40771.083847+1675.879822 test-rmse:181357.980550+14112.149295   
## [252] train-rmse:40652.266883+1622.410846 test-rmse:181369.714138+14120.202396   
## [253] train-rmse:40514.595909+1625.867640 test-rmse:181330.336594+14126.894440   
## [254] train-rmse:40404.468119+1647.322952 test-rmse:181328.348120+14150.929867   
## [255] train-rmse:40355.346699+1633.772078 test-rmse:181350.557274+14131.244749   
## [256] train-rmse:40257.902261+1637.021888 test-rmse:181312.819956+14127.656348   
## [257] train-rmse:40144.750411+1628.151960 test-rmse:181259.280383+14169.616719   
## [258] train-rmse:39981.270475+1587.550987 test-rmse:181261.777762+14175.085707   
## [259] train-rmse:39843.786524+1587.015499 test-rmse:181252.237441+14181.575285   
## [260] train-rmse:39729.324407+1513.091956 test-rmse:181210.466335+14190.674011   
## [261] train-rmse:39619.423005+1488.205604 test-rmse:181184.026659+14203.435968   
## [262] train-rmse:39557.536146+1502.162409 test-rmse:181166.380796+14207.731683   
## [263] train-rmse:39484.311315+1474.714166 test-rmse:181152.462535+14211.708225   
## [264] train-rmse:39376.840212+1503.797433 test-rmse:181137.414397+14222.898674   
## [265] train-rmse:39235.014649+1538.931318 test-rmse:181104.701838+14219.352479   
## [266] train-rmse:39131.324203+1506.888177 test-rmse:181081.510522+14209.080019   
## [267] train-rmse:39008.328530+1514.757593 test-rmse:181087.780368+14209.283692   
## [268] train-rmse:38867.851148+1536.508576 test-rmse:181061.513845+14199.376123   
## [269] train-rmse:38754.775863+1497.056195 test-rmse:181054.640941+14203.189062   
## [270] train-rmse:38657.025467+1488.094844 test-rmse:181042.559161+14209.640987   
## [271] train-rmse:38579.706933+1491.567680 test-rmse:181020.007344+14239.804720   
## [272] train-rmse:38412.592203+1420.259242 test-rmse:181019.945863+14232.686267   
## [273] train-rmse:38308.202603+1406.875803 test-rmse:180993.419852+14241.693915   
## [274] train-rmse:38204.494175+1430.268861 test-rmse:180982.381430+14248.905676   
## [275] train-rmse:38088.355100+1393.215863 test-rmse:181013.120571+14254.143539   
## [276] train-rmse:37998.311266+1394.243601 test-rmse:181002.318951+14265.833289   
## [277] train-rmse:37935.059230+1387.496355 test-rmse:180992.883117+14289.964944   
## [278] train-rmse:37795.987745+1431.343645 test-rmse:180984.802319+14267.609766   
## [279] train-rmse:37702.004590+1399.815157 test-rmse:180987.171335+14261.667856   
## [280] train-rmse:37598.647611+1390.488587 test-rmse:180962.979858+14293.460387   
## [281] train-rmse:37513.477141+1362.808217 test-rmse:180982.580507+14296.649384   
## [282] train-rmse:37370.088044+1327.455207 test-rmse:180958.537145+14312.392536   
## [283] train-rmse:37313.096825+1340.182842 test-rmse:180955.101969+14290.464640   
## [284] train-rmse:37220.973802+1315.490819 test-rmse:180948.027669+14294.683967   
## [285] train-rmse:37122.494714+1319.563592 test-rmse:180931.611552+14281.782552   
## [286] train-rmse:37043.242053+1326.251811 test-rmse:180984.625060+14253.025314   
## [287] train-rmse:36953.650504+1326.997848 test-rmse:180989.307495+14245.787131   
## [288] train-rmse:36882.094271+1331.146845 test-rmse:181001.308851+14249.456519   
## [289] train-rmse:36779.875993+1333.190150 test-rmse:181012.892652+14256.504469   
## [290] train-rmse:36666.521925+1311.972113 test-rmse:180991.245106+14245.694476   
## [291] train-rmse:36575.931993+1290.057830 test-rmse:180979.571174+14247.309680   
## [292] train-rmse:36527.140782+1278.084880 test-rmse:180989.186183+14232.094838   
## [293] train-rmse:36430.983548+1252.860294 test-rmse:180980.089737+14239.021029   
## [294] train-rmse:36372.384796+1245.951251 test-rmse:180962.240513+14238.410568   
## [295] train-rmse:36278.430666+1227.141142 test-rmse:180948.001691+14239.786874   
## [296] train-rmse:36180.132617+1235.626007 test-rmse:180932.833785+14233.089206   
## [297] train-rmse:36051.259049+1202.862108 test-rmse:180911.561236+14235.028574   
## [298] train-rmse:35936.262524+1205.336398 test-rmse:180917.425898+14261.756175   
## [299] train-rmse:35857.052158+1245.226970 test-rmse:180909.941325+14256.525486   
## [300] train-rmse:35724.110543+1258.070915 test-rmse:180891.932864+14295.517862   
## [301] train-rmse:35590.238771+1246.510021 test-rmse:180859.453220+14319.075103   
## [302] train-rmse:35513.820407+1267.748533 test-rmse:180850.195715+14315.954573   
## [303] train-rmse:35472.888370+1261.521896 test-rmse:180861.709283+14295.382902   
## [304] train-rmse:35358.975063+1212.413533 test-rmse:180876.793468+14292.736627   
## [305] train-rmse:35273.024940+1253.822436 test-rmse:180855.383053+14299.699459   
## [306] train-rmse:35182.488633+1265.424596 test-rmse:180834.275791+14303.533693   
## [307] train-rmse:35087.507390+1223.162195 test-rmse:180830.180197+14317.940530   
## [308] train-rmse:34991.788321+1216.735529 test-rmse:180849.816588+14306.931248   
## [309] train-rmse:34908.827904+1233.072156 test-rmse:180868.112262+14325.721855   
## [310] train-rmse:34850.329157+1249.312152 test-rmse:180886.619882+14317.140068   
## [311] train-rmse:34740.749068+1241.730301 test-rmse:180874.699767+14329.634547   
## [312] train-rmse:34637.982644+1228.126432 test-rmse:180864.371825+14326.090916   
## [313] train-rmse:34569.628826+1260.541943 test-rmse:180856.699524+14320.261451   
## [314] train-rmse:34494.481016+1255.541688 test-rmse:180839.762705+14341.437367   
## [315] train-rmse:34390.215425+1263.407543 test-rmse:180787.551359+14377.655739   
## [316] train-rmse:34271.947623+1256.066299 test-rmse:180774.884954+14386.296885   
## [317] train-rmse:34226.281833+1231.900003 test-rmse:180775.797621+14369.862302   
## [318] train-rmse:34149.404911+1208.935822 test-rmse:180785.025533+14369.862924   
## [319] train-rmse:34062.601502+1205.186245 test-rmse:180777.643681+14381.479171   
## [320] train-rmse:33998.303601+1214.740631 test-rmse:180784.808696+14394.508412   
## [321] train-rmse:33891.133424+1221.080918 test-rmse:180764.041487+14441.757445   
## [322] train-rmse:33806.678531+1204.359631 test-rmse:180718.119680+14498.125694   
## [323] train-rmse:33718.489214+1171.496484 test-rmse:180709.324149+14505.830333   
## [324] train-rmse:33618.027386+1154.360612 test-rmse:180714.114001+14486.313420   
## [325] train-rmse:33509.070835+1148.010394 test-rmse:180708.787838+14479.178447   
## [326] train-rmse:33366.612555+1170.412656 test-rmse:180725.535449+14493.596625   
## [327] train-rmse:33275.056338+1165.892712 test-rmse:180722.541369+14502.652289   
## [328] train-rmse:33224.629532+1154.899796 test-rmse:180719.586437+14510.001134   
## [329] train-rmse:33101.529674+1131.877168 test-rmse:180724.273796+14553.627989   
## [330] train-rmse:33001.426375+1138.771817 test-rmse:180708.005913+14547.286053   
## [331] train-rmse:32916.002598+1116.071749 test-rmse:180676.586584+14542.381457   
## [332] train-rmse:32851.087683+1145.249312 test-rmse:180706.848626+14555.226558   
## [333] train-rmse:32793.203125+1149.220198 test-rmse:180699.550826+14558.956573   
## [334] train-rmse:32722.589098+1177.928946 test-rmse:180702.436123+14552.837499   
## [335] train-rmse:32665.265919+1188.073648 test-rmse:180689.490646+14567.704787   
## [336] train-rmse:32578.575906+1149.464638 test-rmse:180695.334175+14559.830315   
## [337] train-rmse:32527.903414+1171.835219 test-rmse:180702.697986+14540.545711   
## [338] train-rmse:32440.166532+1174.207104 test-rmse:180695.118114+14536.284939   
## [339] train-rmse:32354.061452+1165.652507 test-rmse:180693.206303+14514.340514   
## [340] train-rmse:32267.942043+1194.596032 test-rmse:180684.162306+14521.638409   
## [341] train-rmse:32175.806761+1143.693001 test-rmse:180687.856390+14548.956080   
## [342] train-rmse:32104.175329+1150.479321 test-rmse:180671.531157+14532.524816   
## [343] train-rmse:32021.808363+1148.298871 test-rmse:180681.395597+14519.333066   
## [344] train-rmse:31927.750228+1149.147104 test-rmse:180673.253485+14522.260310   
## [345] train-rmse:31848.030081+1137.942277 test-rmse:180673.034356+14524.885736   
## [346] train-rmse:31757.844117+1115.874589 test-rmse:180687.307873+14493.349455   
## [347] train-rmse:31632.209167+1098.567814 test-rmse:180696.791718+14514.741687   
## [348] train-rmse:31564.932578+1099.076608 test-rmse:180683.172662+14530.327270   
## [349] train-rmse:31510.694503+1104.205498 test-rmse:180683.563172+14535.898827   
## [350] train-rmse:31455.655892+1081.952710 test-rmse:180657.400048+14538.325174   
## [351] train-rmse:31347.060807+1073.279303 test-rmse:180663.715800+14538.303935   
## [352] train-rmse:31277.774815+1061.431741 test-rmse:180676.560978+14516.641538   
## [353] train-rmse:31175.658377+1075.310501 test-rmse:180677.543001+14501.827492   
## [354] train-rmse:31099.074336+1102.573948 test-rmse:180665.896224+14493.915329   
## [355] train-rmse:31041.226762+1092.111601 test-rmse:180653.807680+14501.407311   
## [356] train-rmse:30934.028486+1069.327951 test-rmse:180640.076036+14493.928644   
## [357] train-rmse:30850.539984+1068.540679 test-rmse:180632.963322+14479.511090   
## [358] train-rmse:30770.602328+1057.463830 test-rmse:180615.388675+14465.855087   
## [359] train-rmse:30669.499731+1055.804054 test-rmse:180662.075904+14423.068810   
## [360] train-rmse:30594.975767+1042.792806 test-rmse:180674.359041+14433.995173   
## [361] train-rmse:30541.482006+1039.320675 test-rmse:180675.978884+14429.483502   
## [362] train-rmse:30475.583571+1064.528796 test-rmse:180655.257516+14438.704589   
## [363] train-rmse:30416.710877+1067.311536 test-rmse:180645.329346+14453.592359   
## [364] train-rmse:30358.400851+1051.641964 test-rmse:180635.653189+14463.328835   
## [365] train-rmse:30267.689602+1051.070095 test-rmse:180645.529811+14457.681132   
## [366] train-rmse:30196.457532+1076.760176 test-rmse:180648.778727+14490.438649   
## [367] train-rmse:30123.575786+1119.345218 test-rmse:180628.092228+14489.831233   
## [368] train-rmse:30078.273400+1121.736255 test-rmse:180631.262017+14476.081565   
## [369] train-rmse:29996.063560+1096.553045 test-rmse:180645.725542+14463.202186   
## [370] train-rmse:29923.667536+1083.283250 test-rmse:180616.318541+14503.051168   
## [371] train-rmse:29812.745299+1108.777020 test-rmse:180628.672805+14512.165215   
## [372] train-rmse:29720.275284+1101.353988 test-rmse:180595.145191+14526.324824   
## [373] train-rmse:29641.429403+1138.499843 test-rmse:180595.253747+14522.742885   
## [374] train-rmse:29538.743887+1132.079822 test-rmse:180612.300174+14533.523943   
## [375] train-rmse:29435.257923+1122.195594 test-rmse:180597.840345+14538.168682   
## [376] train-rmse:29355.826341+1141.439517 test-rmse:180577.759060+14532.984871   
## [377] train-rmse:29300.926215+1130.707251 test-rmse:180577.738776+14540.691227   
## [378] train-rmse:29204.832373+1128.744138 test-rmse:180579.057117+14538.654493   
## [379] train-rmse:29122.960952+1115.515154 test-rmse:180576.119431+14527.159092   
## [380] train-rmse:29045.435058+1152.307372 test-rmse:180566.092545+14533.210021   
## [381] train-rmse:28975.806879+1178.377407 test-rmse:180561.087404+14531.402820   
## [382] train-rmse:28917.856867+1187.778853 test-rmse:180561.808718+14534.672981   
## [383] train-rmse:28849.827685+1174.196119 test-rmse:180575.580657+14533.065507   
## [384] train-rmse:28775.591350+1162.447314 test-rmse:180568.140978+14508.299188   
## [385] train-rmse:28724.586417+1159.439145 test-rmse:180546.523156+14476.971159   
## [386] train-rmse:28628.324394+1154.787209 test-rmse:180524.576672+14462.207361   
## [387] train-rmse:28543.437524+1176.428689 test-rmse:180514.216582+14449.457788   
## [388] train-rmse:28464.114220+1198.314165 test-rmse:180475.652280+14450.490873   
## [389] train-rmse:28381.222198+1209.003541 test-rmse:180489.570935+14449.398663   
## [390] train-rmse:28286.159953+1210.440210 test-rmse:180489.677024+14453.559145   
## [391] train-rmse:28212.676996+1193.510459 test-rmse:180518.261469+14473.216503   
## [392] train-rmse:28149.549389+1199.518191 test-rmse:180495.639418+14467.868567   
## [393] train-rmse:28080.797520+1178.429023 test-rmse:180497.845001+14478.957058   
## [394] train-rmse:27967.664865+1157.236505 test-rmse:180491.488889+14429.941688   
## [395] train-rmse:27916.904575+1184.149633 test-rmse:180496.631989+14425.515083   
## [396] train-rmse:27811.935638+1148.139743 test-rmse:180475.474453+14423.611810   
## [397] train-rmse:27758.846889+1139.968114 test-rmse:180461.344057+14428.671383   
## [398] train-rmse:27723.362033+1144.574594 test-rmse:180461.135991+14439.866126   
## [399] train-rmse:27683.466069+1132.298528 test-rmse:180469.962029+14443.547590   
## [400] train-rmse:27629.628282+1112.938309 test-rmse:180481.277153+14443.385000   
## [401] train-rmse:27572.996107+1119.218449 test-rmse:180460.010451+14418.584322   
## [402] train-rmse:27497.556851+1096.517287 test-rmse:180469.158326+14427.268585   
## [403] train-rmse:27421.727230+1076.553012 test-rmse:180468.475549+14417.653871   
## [404] train-rmse:27363.773960+1082.731559 test-rmse:180453.121713+14426.700979   
## [405] train-rmse:27281.165483+1082.324337 test-rmse:180453.907665+14438.229303   
## [406] train-rmse:27194.123135+1069.428598 test-rmse:180452.406299+14425.006321   
## [407] train-rmse:27121.395311+1062.506379 test-rmse:180471.600142+14425.514040   
## [408] train-rmse:27058.590407+1028.507129 test-rmse:180487.067200+14419.610814   
## [409] train-rmse:27000.240938+1039.218817 test-rmse:180470.185671+14423.085702   
## [410] train-rmse:26924.972062+1047.627461 test-rmse:180475.294877+14416.945803   
## [411] train-rmse:26840.824235+1044.099861 test-rmse:180470.811615+14403.268606   
## [412] train-rmse:26812.890504+1033.828312 test-rmse:180474.536366+14402.655534   
## [413] train-rmse:26746.708628+1011.774783 test-rmse:180459.502893+14396.041940   
## [414] train-rmse:26682.195782+1020.961761 test-rmse:180463.103956+14385.643350   
## [415] train-rmse:26607.677533+996.891527 test-rmse:180474.581386+14389.732609   
## [416] train-rmse:26540.248995+994.326872 test-rmse:180466.860732+14386.124843   
## [417] train-rmse:26484.776829+978.467621 test-rmse:180466.719090+14386.475613   
## [418] train-rmse:26435.981055+959.627327 test-rmse:180458.899507+14389.050754   
## [419] train-rmse:26378.068964+963.838676 test-rmse:180463.705561+14378.462266   
## [420] train-rmse:26335.993550+982.263305 test-rmse:180464.454255+14372.789515   
## [421] train-rmse:26279.145670+981.355179 test-rmse:180466.692028+14375.477888   
## [422] train-rmse:26236.241039+964.898116 test-rmse:180468.968605+14375.564829   
## [423] train-rmse:26181.289685+947.505470 test-rmse:180455.778553+14380.357936   
## [424] train-rmse:26106.750538+951.549865 test-rmse:180454.439604+14366.696842   
## [425] train-rmse:26048.873892+959.964440 test-rmse:180467.140967+14380.037594   
## [426] train-rmse:25985.135942+976.526199 test-rmse:180472.952749+14379.813754   
## Stopping. Best iteration:  
## [406] train-rmse:27194.123135+1069.428598 test-rmse:180452.406299+14425.006321

best\_n <- cv\_res$best\_iteration  
  
# 8. Train XGBoost model  
model <- xgb.train(params, dtrain, nrounds = best\_n)  
  
# 9. Evaluate model on test set  
preds <- predict(model, dtest)  
rmse <- sqrt(mean((preds - test\_df$price)^2))  
cat("XGBoost Test RMSE =", round(rmse, 2), "\n")

## XGBoost Test RMSE = 126224.3

# 10. Extract and plot top 10 features  
imp <- xgb.importance(feature\_names = colnames(train\_m), model = model)  
xgb.plot.importance(imp[1:10, ])

 #OLS

# 11. Fit OLS model using top 8 features from XGBoost  
top\_feats <- imp$Feature[1:8]  
train\_sub <- as.data.frame(train\_m)[, top\_feats]  
train\_sub$price <- train\_df$price  
  
test\_sub <- as.data.frame(test\_m)[, top\_feats]  
test\_sub$price <- test\_df$price  
  
lm\_mod <- lm(price ~ ., data = train\_sub)  
summary(lm\_mod)

##   
## Call:  
## lm(formula = price ~ ., data = train\_sub)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2627759 -101133 -27732 50317 2196252   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.882e+07 1.975e+06 -9.526 < 2e-16 \*\*\*  
## size\_sqft 2.172e+02 1.213e+01 17.899 < 2e-16 \*\*\*  
## Bathroom 1.712e+05 1.375e+04 12.447 < 2e-16 \*\*\*  
## `\\`Total Units\\`` -2.197e+01 1.142e+01 -1.923 0.05464 .   
## parking\_lot 4.337e+04 9.312e+03 4.658 3.52e-06 \*\*\*  
## `\\`# of Floors\\`` 2.497e+02 2.182e+02 1.144 0.25278   
## amenity\_count 5.821e+03 2.059e+03 2.827 0.00477 \*\*   
## statePenang 1.506e+05 1.504e+04 10.017 < 2e-16 \*\*\*  
## `\\`Completion Year\\`` 9.236e+03 9.850e+02 9.377 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 247800 on 1323 degrees of freedom  
## (1870 observations deleted due to missingness)  
## Multiple R-squared: 0.5543, Adjusted R-squared: 0.5516   
## F-statistic: 205.7 on 8 and 1323 DF, p-value: < 2.2e-16

# 12. OLS model evaluation  
pred <- predict(lm\_mod, newdata = test\_sub)  
cat("OLS Test RMSE =", round(sqrt(mean((pred - test\_sub$price)^2)), 2), "\n")

## OLS Test RMSE = NA

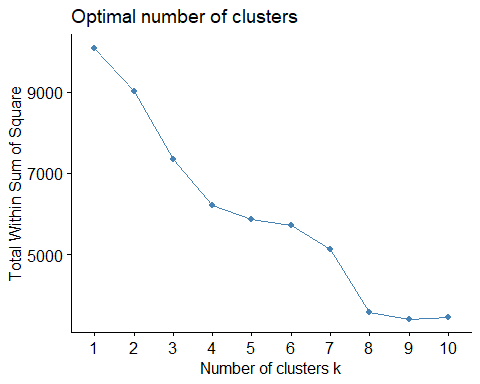
#Cluster

# 1. Load clustering packages  
library(dplyr)  
library(cluster)  
library(factoextra)

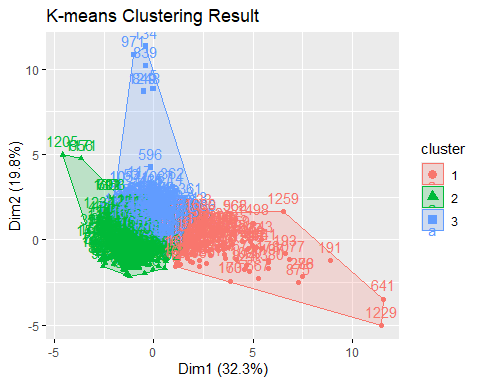
## Warning: package 'factoextra' was built under R version 4.4.3

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

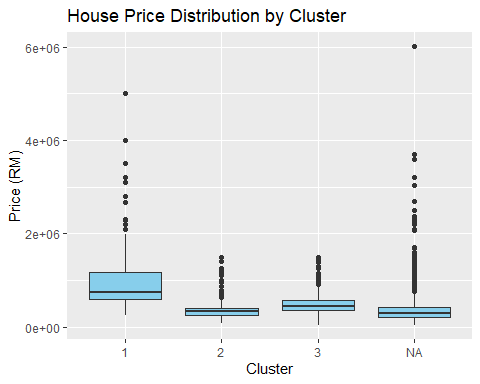
# 2. Select variables for clustering  
cluster\_vars <- c("size\_sqft", "Bathroom", "Total Units", "parking\_lot", "# of Floors", "amenity\_count")  
  
# 3. Prepare data and drop missing  
cluster\_data <- df %>%  
 mutate(row\_id = row\_number()) %>%  
 select(row\_id, all\_of(cluster\_vars)) %>%  
 na.omit()  
  
# 4. Standardize variables  
cluster\_scaled <- scale(cluster\_data[, -1])  
  
# 5. Apply K-means clustering (k = 3)  
set.seed(2025)  
km\_res <- kmeans(cluster\_scaled, centers = 3, nstart = 25)  
  
# 6. Assign cluster labels back to original data  
df$cluster <- NA  
df$cluster[cluster\_data$row\_id] <- km\_res$cluster  
df$cluster <- factor(df$cluster)  
  
# 7. (Optional) Elbow method plot  
fviz\_nbclust(cluster\_scaled, kmeans, method = "wss")



# 8. Visualize clusters (2D projection)  
fviz\_cluster(km\_res, data = cluster\_scaled, main = "K-means Clustering Result")



# 9. Price by cluster boxplot  
ggplot(df, aes(x = cluster, y = price)) +  
 geom\_boxplot(fill = "skyblue") +  
 labs(title = "House Price Distribution by Cluster", x = "Cluster", y = "Price (RM)")



# 10. Summarize mean values by cluster  
cluster\_summary <- df %>%  
 filter(!is.na(cluster)) %>%  
 group\_by(cluster) %>%  
 summarise(across(all\_of(cluster\_vars), mean, na.rm = TRUE))

## Warning: There was 1 warning in `summarise()`.  
## ℹ In argument: `across(all\_of(cluster\_vars), mean, na.rm = TRUE)`.  
## ℹ In group 1: `cluster = 1`.  
## Caused by warning:  
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.  
## Supply arguments directly to `.fns` through an anonymous function instead.  
##   
## # Previously  
## across(a:b, mean, na.rm = TRUE)  
##   
## # Now  
## across(a:b, \(x) mean(x, na.rm = TRUE))

write.csv(cluster\_summary, "cluster\_summary.csv", row.names = FALSE)