Realestate_Noone

2023-06-11

데이터 로드 및 변환

```
library(dplyr)
```

```
##
## 다음의 패키지를 부착합니다: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# finalretaildata 불러오기

#setwd("C:₩₩Rtest₩₩realestate")
data_whole <- read.csv("RealEstateData/FinalRetailData_1차 수정.csv", h = T, fileEncoding = "cp
949")
data_whole <- subset(data_whole, Rejion == "노원구")  # 다른 구 분석하려면 이 부분 변경경
summary(data_whole)
```

```
##      index       transaction_id     apartment_id       city
##  Min.   :25704   Min.   :1121715   Min.   :  794   Length:9160
##  1st Qu.:28022   1st Qu.:1124044   1st Qu.: 5865   Class :character
##  Median :30341   Median :1126369   Median : 5964   Mode  :character
##  Mean   :30337   Mean   :1126376   Mean   : 7374
##  3rd Qu.:32648   3rd Qu.:1128694   3rd Qu.: 9896
##  Max.   :34976   Max.   :1131136   Max.   :12621
##     dong              jibun              apt              addr_kr
##  Length:9160        Length:9160        Length:9160        Length:9160
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  exclusive_use_area year_of_completion transaction_year_month
##  Min.   : 12.42     Min.   :1976       Length:9160
##  1st Qu.: 49.50     1st Qu.:1988       Class :character
##  Median : 59.34     Median :1992       Mode  :character
##  Mean   : 64.42     Mean   :1993
##  3rd Qu.: 84.65     3rd Qu.:1998
##  Max.   :180.34     Max.   :2016
##  transaction_date       floor        transaction_real_price      year
##  Length:9160        Min.   : 1.000   Min.   : 9000          Min.   :2017
##  Class :character   1st Qu.: 4.000   1st Qu.:25900          1st Qu.:2017
##  Mode  :character   Median : 8.000   Median :31600          Median :2017
##                     Mean   : 8.035   Mean   :33917          Mean   :2017
##                     3rd Qu.:12.000   3rd Qu.:40000          3rd Qu.:2017
##                     Max.   :30.000   Max.   :82600          Max.   :2017
##    Latitude        Hardness        Rejion           bigMarket05
##  Min.   :127.0   Min.   :37.61   Length:9160        Min.   :0.000
##  1st Qu.:127.1   1st Qu.:37.63   Class :character   1st Qu.:1.000
##  Median :127.1   Median :37.65   Mode  :character   Median :1.000
##  Mean   :127.1   Mean   :37.65                      Mean   :1.569
##  3rd Qu.:127.1   3rd Qu.:37.66                      3rd Qu.:2.000
##  Max.   :127.1   Max.   :37.69                      Max.   :5.000
##   bigMarket10     bigMarket15       school05         school10
##  Min.   :0.000   Min.   : 1.00   Min.   : 0.000   Min.   : 2.00
##  1st Qu.:3.000   1st Qu.: 7.00   1st Qu.: 3.000   1st Qu.:12.00
##  Median :5.000   Median : 9.00   Median : 4.000   Median :15.00
##  Mean   :4.449   Mean   : 8.77   Mean   : 4.232   Mean   :15.31
##  3rd Qu.:6.000   3rd Qu.:11.00   3rd Qu.: 6.000   3rd Qu.:20.00
##  Max.   :9.000   Max.   :15.00   Max.   :10.000   Max.   :26.00
##    school15        subway05         subway10        subway15
##  Min.   : 7.00   Min.   :0.0000   Min.   :0.000   Min.   : 0.000
##  1st Qu.:25.00   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.: 4.000
##  Median :28.00   Median :1.0000   Median :2.000   Median : 5.000
##  Mean   :29.54   Mean   :0.6774   Mean   :2.384   Mean   : 5.081
##  3rd Qu.:35.00   3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.: 6.000
##  Max.   :46.00   Max.   :3.0000   Max.   :7.000   Max.   :10.000
##   hospital05       hospital10      hospital15       movie05
##  Min.   :  1.00   Min.   :  9.0   Min.   : 32.0   Min.   :0.000
##  1st Qu.: 16.00   1st Qu.: 67.0   1st Qu.:144.0   1st Qu.:0.000
##  Median : 27.00   Median : 85.0   Median :198.0   Median :1.000
##  Mean   : 35.75   Mean   :109.9   Mean   :229.3   Mean   :1.358
##  3rd Qu.: 42.00   3rd Qu.:118.0   3rd Qu.:332.0   3rd Qu.:2.000
```

```
##  Max.   :187.00   Max.   :309.0   Max.   :414.0   Max.   :6.000
##     movie10          movie15          kid05           kid10
##  Min.   : 0.00   Min.   : 1.00   Min.   : 4.00   Min.   :13.00
##  1st Qu.: 3.00   1st Qu.: 8.00   1st Qu.:16.00   1st Qu.:61.00
##  Median : 5.00   Median :11.00   Median :21.00   Median :69.00
##  Mean   : 4.81   Mean   :10.74   Mean   :21.26   Mean   :69.26
##  3rd Qu.: 7.00   3rd Qu.:13.00   3rd Qu.:25.00   3rd Qu.:81.00
##  Max.   :10.00   Max.   :18.00   Max.   :47.00   Max.   :99.00
##      kid15          office05        office10         office15
##  Min.   : 30.0   Min.   :0.000   Min.   : 2.000   Min.   : 5.00
##  1st Qu.:122.0   1st Qu.:1.000   1st Qu.: 7.000   1st Qu.:15.00
##  Median :139.0   Median :2.000   Median : 8.000   Median :17.00
##  Mean   :136.6   Mean   :2.022   Mean   : 8.164   Mean   :16.77
##  3rd Qu.:155.0   3rd Qu.:3.000   3rd Qu.:10.000   3rd Qu.:19.00
##  Max.   :195.0   Max.   :7.000   Max.   :14.000   Max.   :23.00
```

```
str(data_whole)
```

```
## 'data.frame':    9160 obs. of  39 variables:
##  $ index                : int  25704 25705 25706 25707 25708 25709 25710 25711 25712 25713
...
##  $ transaction_id       : int  1121715 1121716 1121717 1121718 1121719 1121720 1121721 1121
722 1121723 1121724 ...
##  $ apartment_id         : int  8588 9807 12149 12149 3443 5445 11664 11722 4591 4576 ...
##  $ city                 : chr  "서울특별시" "서울특별시" "서울특별시" "서울특별시" ...
##  $ dong                 : chr  "월계동" "월계동" "월계동" "월계동" ...
##  $ jibun                : chr  "946" "556" "929" "929" ...
##  $ apt                  : chr  "우남푸르미아" "주공2" "현대" "현대" ...
##  $ addr_kr              : chr  "월계동 946 우남푸르미아" "월계동 556 주공2" "월계동 929 현
대" "월계동 929 현대" ...
##  $ exclusive_use_area   : num  59.9 84.8 60 85 84.6 ...
##  $ year_of_completion   : int  2006 1992 2000 2000 2005 2006 2000 2002 1986 1986 ...
##  $ transaction_year_month: chr  "2017-01-01" "2017-01-01" "2017-01-01" "2017-01-01" ...
##  $ transaction_date     : chr  "1~10" "1~10" "1~10" "1~10" ...
##  $ floor                : int  9 3 22 11 4 4 4 17 5 6 ...
##  $ transaction_real_price: int  29000 32000 30500 41000 35700 34000 42700 42300 28800 29100
...
##  $ year                 : int  2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...
##  $ Latitude             : num  127 127 127 127 127 ...
##  $ Hardness             : num  37.6 37.6 37.6 37.6 37.6 ...
##  $ Rejion               : chr  "노원구" "노원구" "노원구" "노원구" ...
##  $ bigMarket05          : int  1 1 0 0 0 2 0 1 2 2 ...
##  $ bigMarket10          : int  4 4 5 5 4 3 4 3 5 5 ...
##  $ bigMarket15          : int  9 8 5 5 9 8 6 8 9 9 ...
##  $ school05             : int  5 3 6 6 3 0 2 1 0 0 ...
##  $ school10             : int  15 15 9 9 15 13 8 7 12 12 ...
##  $ school15             : int  28 28 28 28 28 30 28 19 25 25 ...
##  $ subway05             : int  1 0 0 0 1 2 2 2 2 2 ...
##  $ subway10             : int  1 3 5 5 2 2 4 7 5 5 ...
##  $ subway15             : int  5 5 7 7 7 8 9 10 9 9 ...
##  $ hospital05           : int  13 12 8 8 10 11 8 24 14 14 ...
##  $ hospital10           : int  38 40 86 86 54 43 61 100 101 101 ...
##  $ hospital15           : int  113 122 183 183 115 175 193 240 194 194 ...
##  $ movie05              : int  0 0 2 2 0 1 2 2 0 0 ...
##  $ movie10              : int  1 0 4 4 1 5 4 6 5 5 ...
##  $ movie15              : int  7 5 7 7 8 6 8 9 9 9 ...
##  $ kid05                : int  12 14 18 18 15 20 21 20 24 24 ...
##  $ kid10                : int  53 65 58 58 59 75 71 70 72 72 ...
##  $ kid15                : int  139 137 140 140 142 148 152 148 154 154 ...
##  $ office05             : int  3 2 5 5 2 1 2 3 2 2 ...
##  $ office10             : int  5 7 10 10 6 9 10 11 10 10 ...
##  $ office15             : int  15 18 21 21 15 21 20 21 19 19 ...
```

```
data_whole %>% colnames()
```

```
##  [1] "index"                "transaction_id"         "apartment_id"
##  [4] "city"                 "dong"                   "jibun"
##  [7] "apt"                  "addr_kr"                "exclusive_use_area"
## [10] "year_of_completion"   "transaction_year_month" "transaction_date"
## [13] "floor"                "transaction_real_price" "year"
## [16] "Latitude"             "Hardness"               "Rejion"
## [19] "bigMarket05"          "bigMarket10"            "bigMarket15"
## [22] "school05"             "school10"               "school15"
## [25] "subway05"             "subway10"               "subway15"
## [28] "hospital05"           "hospital10"             "hospital15"
## [31] "movie05"              "movie10"                "movie15"
## [34] "kid05"                "kid10"                  "kid15"
## [37] "office05"             "office10"               "office15"
```

```
# # Was:
# data %>% select(filterCol)
#
# # Now:
# data %>% select(all_of(filterCol))
filterCol<-c("index", "transaction_id", "apartment_id", "city", "jibun", "apt", "addr_kr", "Lat
itude", "Hardness", "year", "Rejion")
data_whole<-data_whole %>% select(-all_of(filterCol))
str(data_whole)
```

```
## 'data.frame':    9160 obs. of  28 variables:
##  $ dong                  : chr  "월계동" "월계동" "월계동" "월계동" ...
##  $ exclusive_use_area    : num  59.9 84.8 60 85 84.6 ...
##  $ year_of_completion    : int  2006 1992 2000 2000 2005 2006 2000 2002 1986 1986 ...
##  $ transaction_year_month: chr  "2017-01-01" "2017-01-01" "2017-01-01" "2017-01-01" ...
##  $ transaction_date      : chr  "1~10" "1~10" "1~10" "1~10" ...
##  $ floor                 : int  9 3 22 11 4 4 4 17 5 6 ...
##  $ transaction_real_price: int  29000 32000 30500 41000 35700 34000 42700 42300 28800 29100
...
##  $ bigMarket05           : int  1 1 0 0 0 2 0 1 2 2 ...
##  $ bigMarket10           : int  4 4 5 5 4 3 4 3 5 5 ...
##  $ bigMarket15           : int  9 8 5 5 9 8 6 8 9 9 ...
##  $ school05              : int  5 3 6 6 3 0 2 1 0 0 ...
##  $ school10              : int  15 15 9 9 15 13 8 7 12 12 ...
##  $ school15              : int  28 28 28 28 28 30 28 19 25 25 ...
##  $ subway05              : int  1 0 0 0 1 2 2 2 2 2 ...
##  $ subway10              : int  1 3 5 5 2 2 4 7 5 5 ...
##  $ subway15              : int  5 5 7 7 7 8 9 10 9 9 ...
##  $ hospital05            : int  13 12 8 8 10 11 8 24 14 14 ...
##  $ hospital10            : int  38 40 86 86 54 43 61 100 101 101 ...
##  $ hospital15            : int  113 122 183 183 115 175 193 240 194 194 ...
##  $ movie05               : int  0 0 2 2 0 1 2 2 0 0 ...
##  $ movie10               : int  1 0 4 4 1 5 4 6 5 5 ...
##  $ movie15               : int  7 5 7 7 8 6 8 9 9 9 ...
##  $ kid05                 : int  12 14 18 18 15 20 21 20 24 24 ...
##  $ kid10                 : int  53 65 58 58 59 75 71 70 72 72 ...
##  $ kid15                 : int  139 137 140 140 142 148 152 148 154 154 ...
##  $ office05              : int  3 2 5 5 2 1 2 3 2 2 ...
##  $ office10              : int  5 7 10 10 6 9 10 11 10 10 ...
##  $ office15              : int  15 18 21 21 15 21 20 21 19 19 ...
```

```
# 면적당 가격 변수 추가 및 real_price 변수 제거
data_whole$transaction_real_price <- as.numeric(data_whole$transaction_real_price)
data_whole$unit_price <- data_whole$transaction_real_price / data_whole$exclusive_use_area
data_whole$transaction_real_price <- NULL
str(data_whole)
```

```
## 'data.frame':    9160 obs. of  28 variables:
##  $ dong                  : chr  "월계동" "월계동" "월계동" "월계동" ...
##  $ exclusive_use_area    : num  59.9 84.8 60 85 84.6 ...
##  $ year_of_completion    : int  2006 1992 2000 2000 2005 2006 2000 2002 1986 1986 ...
##  $ transaction_year_month: chr  "2017-01-01" "2017-01-01" "2017-01-01" "2017-01-01" ...
##  $ transaction_date      : chr  "1~10" "1~10" "1~10" "1~10" ...
##  $ floor                 : int  9 3 22 11 4 4 4 17 5 6 ...
##  $ bigMarket05           : int  1 1 0 0 0 2 0 1 2 2 ...
##  $ bigMarket10           : int  4 4 5 5 4 3 4 3 5 5 ...
##  $ bigMarket15           : int  9 8 5 5 9 8 6 8 9 9 ...
##  $ school05              : int  5 3 6 6 3 0 2 1 0 0 ...
##  $ school10              : int  15 15 9 9 15 13 8 7 12 12 ...
##  $ school15              : int  28 28 28 28 28 30 28 19 25 25 ...
##  $ subway05              : int  1 0 0 0 1 2 2 2 2 2 ...
##  $ subway10              : int  1 3 5 5 2 2 4 7 5 5 ...
##  $ subway15              : int  5 5 7 7 7 8 9 10 9 9 ...
##  $ hospital05            : int  13 12 8 8 10 11 8 24 14 14 ...
##  $ hospital10            : int  38 40 86 86 54 43 61 100 101 101 ...
##  $ hospital15            : int  113 122 183 183 115 175 193 240 194 194 ...
##  $ movie05               : int  0 0 2 2 0 1 2 2 0 0 ...
##  $ movie10               : int  1 0 4 4 1 5 4 6 5 5 ...
##  $ movie15               : int  7 5 7 7 8 6 8 9 9 9 ...
##  $ kid05                 : int  12 14 18 18 15 20 21 20 24 24 ...
##  $ kid10                 : int  53 65 58 58 59 75 71 70 72 72 ...
##  $ kid15                 : int  139 137 140 140 142 148 152 148 154 154 ...
##  $ office05              : int  3 2 5 5 2 1 2 3 2 2 ...
##  $ office10              : int  5 7 10 10 6 9 10 11 10 10 ...
##  $ office15              : int  15 18 21 21 15 21 20 21 19 19 ...
##  $ unit_price            : num  484 377 509 482 422 ...
```

```
# transaction_month 변수 추가 및 transaction_year_month, transaction_date, apt 변수 제거
data_whole$transaction_month <- substr(data_whole$transaction_year_month, 6, 7)
data_whole$transaction_year_month <- NULL
data_whole$transaction_date <- NULL
data_whole$apt <- NULL
str(data_whole)
```

```
## 'data.frame':    9160 obs. of  27 variables:
##  $ dong             : chr  "월계동" "월계동" "월계동" "월계동" ...
##  $ exclusive_use_area: num  59.9 84.8 60 85 84.6 ...
##  $ year_of_completion: int  2006 1992 2000 2000 2005 2006 2000 2002 1986 1986 ...
##  $ floor            : int  9 3 22 11 4 4 4 17 5 6 ...
##  $ bigMarket05      : int  1 1 0 0 0 2 0 1 2 2 ...
##  $ bigMarket10      : int  4 4 5 5 4 3 4 3 5 5 ...
##  $ bigMarket15      : int  9 8 5 5 9 8 6 8 9 9 ...
##  $ school05         : int  5 3 6 6 3 0 2 1 0 0 ...
##  $ school10         : int  15 15 9 9 15 13 8 7 12 12 ...
##  $ school15         : int  28 28 28 28 28 30 28 19 25 25 ...
##  $ subway05         : int  1 0 0 0 1 2 2 2 2 2 ...
##  $ subway10         : int  1 3 5 5 2 2 4 7 5 5 ...
##  $ subway15         : int  5 5 7 7 7 8 9 10 9 9 ...
##  $ hospital05       : int  13 12 8 8 10 11 8 24 14 14 ...
##  $ hospital10       : int  38 40 86 86 54 43 61 100 101 101 ...
##  $ hospital15       : int  113 122 183 183 115 175 193 240 194 194 ...
##  $ movie05          : int  0 0 2 2 0 1 2 2 0 0 ...
##  $ movie10          : int  1 0 4 4 1 5 4 6 5 5 ...
##  $ movie15          : int  7 5 7 7 8 6 8 9 9 9 ...
##  $ kid05            : int  12 14 18 18 15 20 21 20 24 24 ...
##  $ kid10            : int  53 65 58 58 59 75 71 70 72 72 ...
##  $ kid15            : int  139 137 140 140 142 148 152 148 154 154 ...
##  $ office05         : int  3 2 5 5 2 1 2 3 2 2 ...
##  $ office10         : int  5 7 10 10 6 9 10 11 10 10 ...
##  $ office15         : int  15 18 21 21 15 21 20 21 19 19 ...
##  $ unit_price       : num  484 377 509 482 422 ...
##  $ transaction_month : chr  "01" "01" "01" "01" ...
```

```
# factor 형으로 변환
data_whole$year <- as.factor(data_whole$year)
data_whole$dong <- as.factor(data_whole$dong)
data_whole$transaction_month <- as.factor(data_whole$transaction_month) # 거래월에 따른 가격 변화 확인

# 변환 결과 확인
str(data_whole)
```

```
## 'data.frame':    9160 obs. of  28 variables:
##  $ dong              : Factor w/ 5 levels "공릉동","상계동",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ exclusive_use_area: num  59.9 84.8 60 85 84.6 ...
##  $ year_of_completion: int  2006 1992 2000 2000 2005 2006 2000 2002 1986 1986 ...
##  $ floor             : int  9 3 22 11 4 4 4 17 5 6 ...
##  $ bigMarket05       : int  1 1 0 0 0 2 0 1 2 2 ...
##  $ bigMarket10       : int  4 4 5 5 4 3 4 3 5 5 ...
##  $ bigMarket15       : int  9 8 5 5 9 8 6 8 9 9 ...
##  $ school05          : int  5 3 6 6 3 0 2 1 0 0 ...
##  $ school10          : int  15 15 9 9 15 13 8 7 12 12 ...
##  $ school15          : int  28 28 28 28 28 30 28 19 25 25 ...
##  $ subway05          : int  1 0 0 0 1 2 2 2 2 2 ...
##  $ subway10          : int  1 3 5 5 2 2 4 7 5 5 ...
##  $ subway15          : int  5 5 7 7 7 8 9 10 9 9 ...
##  $ hospital05        : int  13 12 8 8 10 11 8 24 14 14 ...
##  $ hospital10        : int  38 40 86 86 54 43 61 100 101 101 ...
##  $ hospital15        : int  113 122 183 183 115 175 193 240 194 194 ...
##  $ movie05           : int  0 0 2 2 0 1 2 2 0 0 ...
##  $ movie10           : int  1 0 4 4 1 5 4 6 5 5 ...
##  $ movie15           : int  7 5 7 7 8 6 8 9 9 9 ...
##  $ kid05             : int  12 14 18 18 15 20 21 20 24 24 ...
##  $ kid10             : int  53 65 58 58 59 75 71 70 72 72 ...
##  $ kid15             : int  139 137 140 140 142 148 152 148 154 154 ...
##  $ office05          : int  3 2 5 5 2 1 2 3 2 2 ...
##  $ office10          : int  5 7 10 10 6 9 10 11 10 10 ...
##  $ office15          : int  15 18 21 21 15 21 20 21 19 19 ...
##  $ unit_price        : num  484 377 509 482 422 ...
##  $ transaction_month : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year              : Factor w/ 32 levels "1976","1983",..: 25 11 19 19 24 25 19 21 5 5 ...
```

컬럼 값 Exploration 및 데이터 변환

```
library(ggplot2)

# year of completion -- 준공년도
summary(data_whole$year_of_completion)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1976    1988    1992    1993    1998    2016
```

```
data_whole$year_of_completion_f <- cut(data_whole$year_of_completion, breaks = c(0, 1997, 2001,
2007, Inf), labels = c("1st", "2nd", "3rd", "4th"))
data_whole$year_of_completion <- NULL
summary(data_whole$year_of_completion_f)
```

```
##  1st  2nd  3rd  4th
## 6513 1775  712  160
```

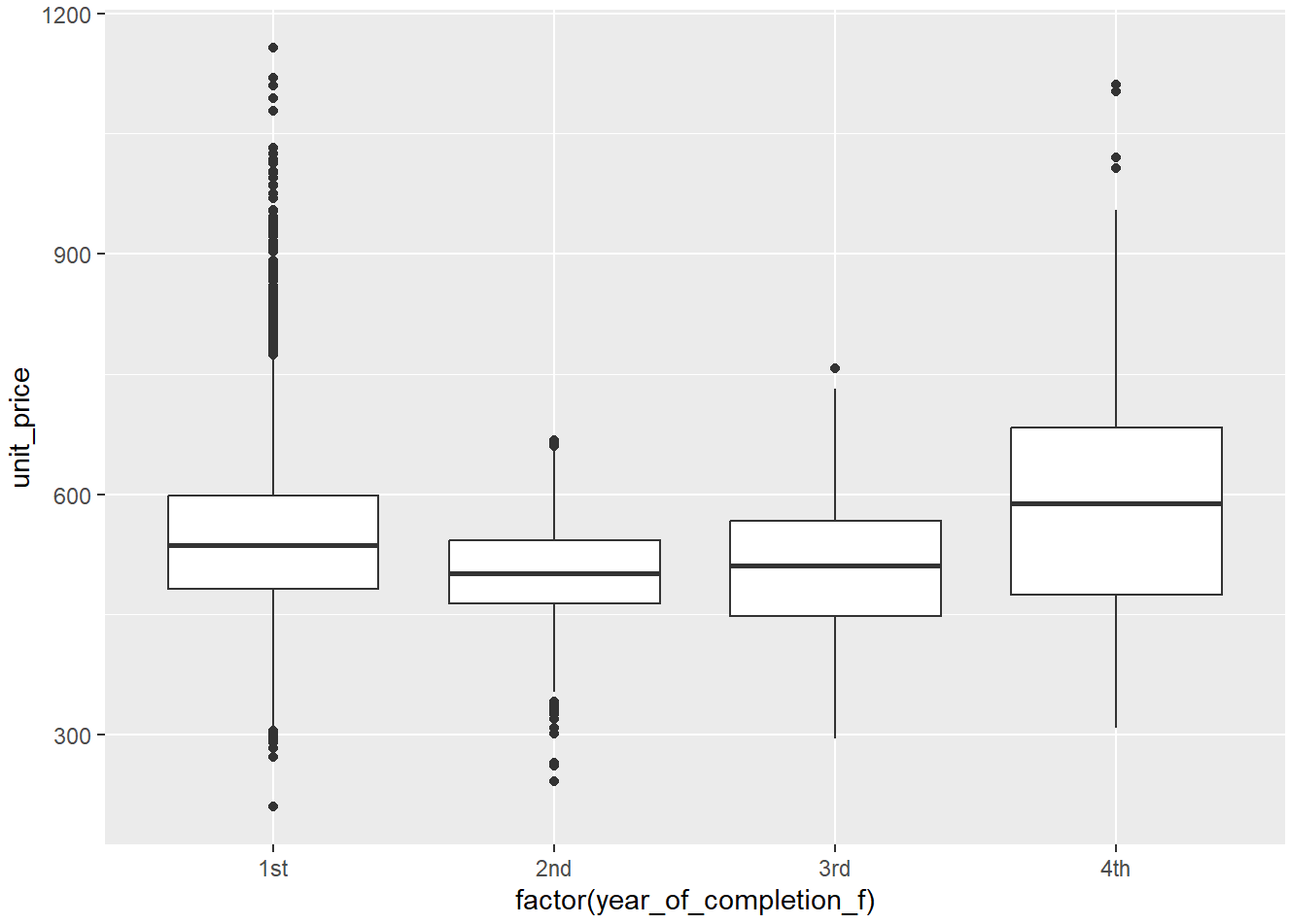
```
# 전체 가격 분포
data_whole %>% summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price))
```

```
##   count avg_price std_price
## 1  9160  538.6799  100.0453
```

```
# 준공년도 factor별 가격 분포
data_whole %>% group_by(year_of_completion_f) %>%
  summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price))
```

```
## # A tibble: 4 × 4
##   year_of_completion_f count avg_price std_price
##   <fct>                <int>     <dbl>     <dbl>
## 1 1st                   6513      550.      105.
## 2 2nd                   1775      502.       60.4
## 3 3rd                    712      514.       86.3
## 4 4th                    160      614.      161.
```

```
ggplot(data = data_whole, aes(x = factor(year_of_completion_f), y = unit_price)) + geom_boxplot
()
```




```
# 동별 가격 분포
summary(data_whole$dong)
```

```
## 공릉동 상계동 월계동 중계동 하계동
##   1096   3549   1487   2214    814
```

```
data_whole %>% group_by(dong) %>%
  summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price)) # dong별 평균 및 표준편차
```

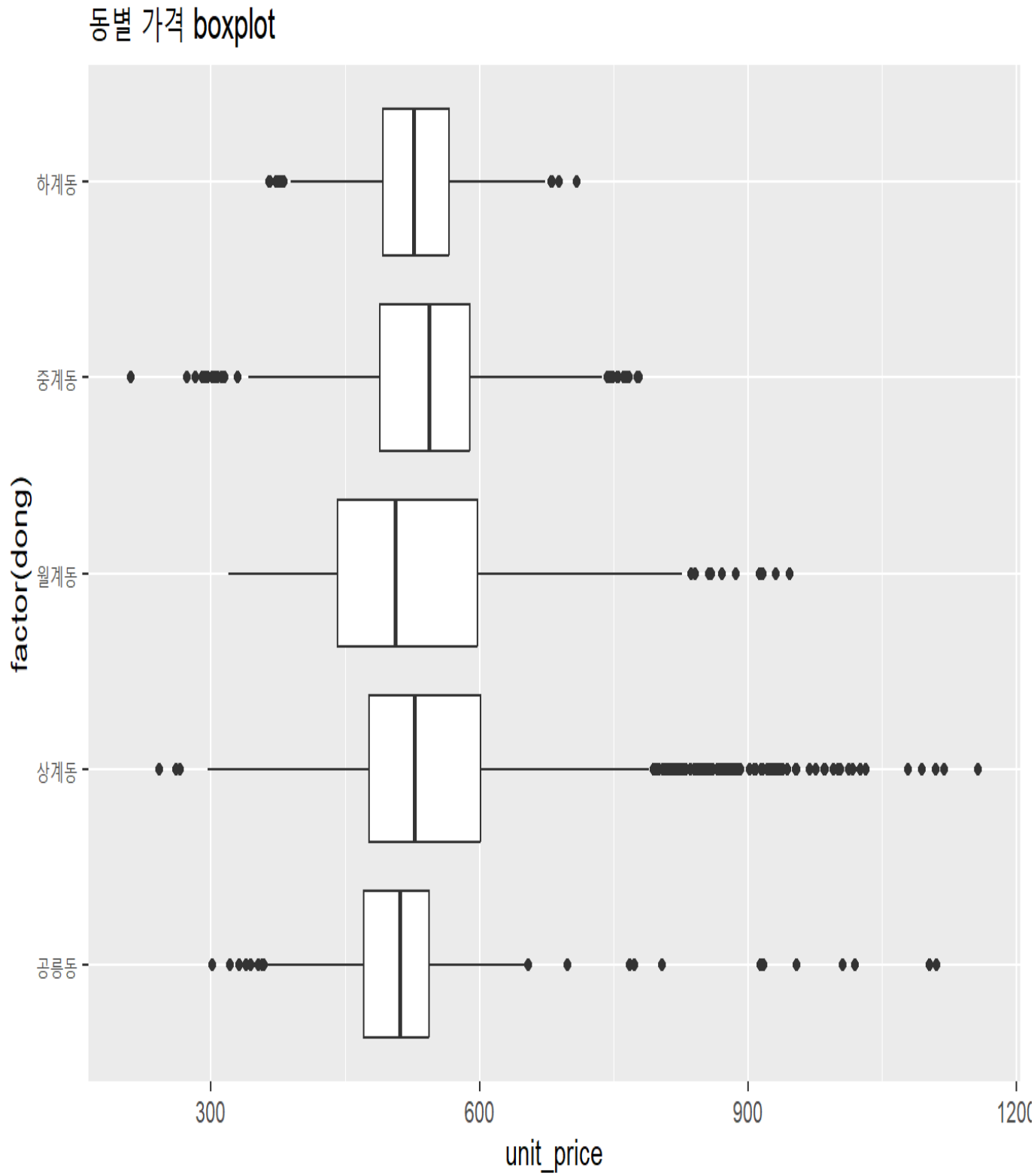
```
## # A tibble: 5 × 4
##   dong   count avg_price std_price
##   <fct>  <int>     <dbl>     <dbl>
## 1 공릉동  1096      508.      71.0
## 2 상계동  3549      553.     119.
## 3 월계동  1487      525.     108.
## 4 중계동  2214      544.      79.1
## 5 하계동   814      530.      58.7
```

```
View(data_whole %>% group_by(dong) %>%
  summarize(count = n(), avg_price = mean(unit_price), std_price = sd(unit_price))) # dong별 평균 및 표준편차
```

```
# unit price
summary(data_whole$unit_price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   210.0   474.0   526.1   538.7   585.7  1157.0
```

```
ggplot(data = data_whole, aes(x = factor(dong), y = unit_price)) + geom_boxplot() + coord_flip() +ggtitle("동별 가격 boxplot")
```



노원구 단위당 가격 분석

트레이닝 데이터와 테스트 데이터로 split

```
# Data transformation for Tree & Regression Model
data_whole1 <- data_whole

install.packages('caTools', repos ="http://cran.us.r-project.org")
```

```
## 패키지 'caTools'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUISW₩AppData₩Local₩Temp₩RtmpEZswNf₩downloaded_packages
```

```
library(caTools)

set.seed(123)
sample = sample.split(data_whole1$unit_price, SplitRatio = .7)
data_train1 = subset(data_whole1, sample == TRUE)
data_test1  = subset(data_whole1, sample == FALSE)

str(data_train1); mean(data_train1$unit_price)
```

```
## 'data.frame':    6412 obs. of  28 variables:
##  $ dong                : Factor w/ 5 levels "공릉동","상계동",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ exclusive_use_area  : num  59.9 60 84.8 114.8 50.1 ...
##  $ floor               : int  9 22 4 4 5 6 10 2 2 6 ...
##  $ bigMarket05         : int  1 0 2 0 2 2 1 1 1 2 ...
##  $ bigMarket10         : int  4 5 3 4 5 5 4 4 3 5 ...
##  $ bigMarket15         : int  9 5 8 6 9 9 8 8 6 9 ...
##  $ school05            : int  5 6 0 2 0 0 3 3 9 0 ...
##  $ school10            : int  15 9 13 8 12 12 15 15 14 12 ...
##  $ school15            : int  28 28 30 28 25 25 28 28 25 25 ...
##  $ subway05            : int  1 0 2 2 2 2 0 0 0 2 ...
##  $ subway10            : int  1 5 2 4 5 5 3 3 1 5 ...
##  $ subway15            : int  5 7 8 9 9 9 5 5 4 9 ...
##  $ hospital05          : int  13 8 11 8 14 14 12 12 7 14 ...
##  $ hospital10          : int  38 86 43 61 101 101 40 40 38 101 ...
##  $ hospital15          : int  113 183 175 193 194 194 122 122 99 194 ...
##  $ movie05             : int  0 2 1 2 0 0 0 0 0 0 ...
##  $ movie10             : int  1 4 5 4 5 5 0 0 0 5 ...
##  $ movie15             : int  7 7 6 8 9 9 5 5 6 9 ...
##  $ kid05               : int  12 18 20 21 24 24 14 14 11 24 ...
##  $ kid10               : int  53 58 75 71 72 72 65 65 43 72 ...
##  $ kid15               : int  139 140 148 152 154 154 137 137 111 154 ...
##  $ office05            : int  3 5 1 2 2 2 2 2 1 2 ...
##  $ office10            : int  5 10 9 10 10 10 7 7 8 10 ...
##  $ office15            : int  15 21 21 20 19 19 18 18 13 19 ...
##  $ unit_price          : num  484 509 401 372 574 ...
##  $ transaction_month   : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year                : Factor w/ 32 levels "1976","1983",..: 25 19 25 19 5 5 11 11 17 5
...
##  $ year_of_completion_f: Factor w/ 4 levels "1st","2nd","3rd",..: 3 2 3 2 1 1 1 1 2 1 ...
```

```
## [1] 539.699
```

```
str(data_test1);mean(data_test1$unit_price)
```

```
## 'data.frame':    2748 obs. of  28 variables:
##  $ dong               : Factor w/ 5 levels "공릉동","상계동",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ exclusive_use_area  : num  84.8 85 84.6 85 59.9 ...
##  $ floor              : int  3 11 4 17 5 18 1 4 10 7 ...
##  $ bigMarket05        : int  1 0 0 1 0 1 2 1 0 1 ...
##  $ bigMarket10        : int  4 5 4 3 2 3 5 4 4 4 ...
##  $ bigMarket15        : int  8 5 9 8 4 8 9 8 9 8 ...
##  $ school05           : int  3 6 3 1 3 1 0 3 3 4 ...
##  $ school10           : int  15 9 15 7 13 7 12 15 12 8 ...
##  $ school15           : int  28 28 28 19 22 19 25 28 29 29 ...
##  $ subway05           : int  0 0 1 2 0 2 2 0 1 2 ...
##  $ subway10           : int  3 5 2 7 0 7 5 3 4 4 ...
##  $ subway15           : int  5 7 7 10 2 10 9 5 6 9 ...
##  $ hospital05         : int  12 8 10 24 3 24 14 12 10 10 ...
##  $ hospital10         : int  40 86 54 100 57 100 101 40 55 60 ...
##  $ hospital15         : int  122 183 115 240 80 240 194 122 126 198 ...
##  $ movie05            : int  0 2 0 2 0 2 0 0 0 2 ...
##  $ movie10            : int  0 4 1 6 0 6 5 0 2 4 ...
##  $ movie15            : int  5 7 8 9 5 9 9 5 8 7 ...
##  $ kid05              : int  14 18 15 20 11 20 24 14 17 23 ...
##  $ kid10              : int  65 58 59 70 49 70 72 65 67 70 ...
##  $ kid15              : int  137 140 142 148 106 148 154 137 145 151 ...
##  $ office05           : int  2 5 2 3 1 3 2 2 1 2 ...
##  $ office10           : int  7 10 6 11 7 11 10 7 6 9 ...
##  $ office15           : int  18 21 15 21 15 21 19 18 18 20 ...
##  $ unit_price         : num  377 482 422 498 406 ...
##  $ transaction_month  : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year               : Factor w/ 32 levels "1976","1983",..: 11 19 24 21 17 21 5 11 14 18
...
##  $ year_of_completion_f: Factor w/ 4 levels "1st","2nd","3rd",..: 1 2 3 3 2 3 1 1 1 2 ...
```

```
## [1] 536.3021
```

Decision Tree

```
install.packages("rpart", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'rpart'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
```

```
## Warning: 패키지 'rpart'의 이전설치를 삭제할 수 없습니다
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE):
## C:WUsersWLUISWWorkSpaceWRWR-4.3.0WlibraryW00L0CKWrpartWlibsWx64Wrpart.dll를
## C:WUsersWLUISWWorkSpaceWRWR-4.3.0WlibraryWrpartWlibsWx64Wrpart.dll로 복사하는데
## 문제가 발생했습니다: Permission denied
```

```
## Warning: 'rpart'를 복구하였습니다
```

```
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:WUsersWLUISWAppDataWLocalWTempWRtmpEZswNfWdownloaded_packages
```

```
library(rpart)
tree1 <- rpart(unit_price~.-year,
               data=data_train1,
               method = "anova",
               control = rpart.control(minsplit = 50, maxdepth = 5))

# tree 결과
print(tree1)
```

```
## n= 6412
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 6412 66477240.0 539.6990
##    2) exclusive_use_area>=38.58 6054 41797700.0 527.3488
##      4) hospital15< 314.5 4254 21862740.0 503.4056
##        8) movie10< 4.5 2153  8991443.0 478.7103
##         16) exclusive_use_area>=59.9965 869  3547706.0 452.4992 *
##         17) exclusive_use_area< 59.9965 1284  4442657.0 496.4497 *
##        9) movie10>=4.5 2101 10212760.0 528.7120
##         18) dong=공릉동,중계동,하계동 1706  5400641.0 515.8183
##           36) school10< 14.5 825  2460518.0 490.7893 *
##           37) school10>=14.5 881  1939332.0 539.2563 *
##         19) dong=월계동 395  3303542.0 584.4001
##           38) exclusive_use_area>=59.95 145   606002.5 486.9981 *
##           39) exclusive_use_area< 59.95 250   524034.6 640.8932 *
##      5) hospital15>=314.5 1800 11732720.0 583.9346
##       10) exclusive_use_area>=84.935 196   952535.5 502.2512 *
##       11) exclusive_use_area< 84.935 1604  9312639.0 593.9159
##         22) transaction_month=01,02,03,04,05,06 1044  4675721.0 571.3356
##           44) kid15< 187.5 980  3672269.0 564.1653 *
##           45) kid15>=187.5 64   181562.1 681.1298 *
##         23) transaction_month=07,08,09,10,11 560  3112245.0 636.0120 *
##    3) exclusive_use_area< 38.58 358  8140876.0 748.5483
##      6) subway05< 1.5 220  2800634.0 669.1959
##       12) kid15< 159.5 139   711886.4 602.7048 *
##       13) kid15>=159.5 81   419655.5 783.2980 *
##      7) subway05>=1.5 138  1746503.0 875.0521 *
```

```
summary(tree1)
```

```
## Call:
## rpart(formula = unit_price ~ . - year, data = data_train1, method = "anova",
##     control = rpart.control(minsplit = 50, maxdepth = 5))
##   n= 6412
##
##            CP nsplit rel error    xerror        xstd
## 1  0.24878679      0 1.0000000 1.0002288 0.029060246
## 2  0.12338423      1 0.7512132 0.7522252 0.015014154
## 3  0.05405969      2 0.6278290 0.6289149 0.013360266
## 4  0.03999167      3 0.5737693 0.5823195 0.012248926
## 5  0.02769431      4 0.5337776 0.5424704 0.011808631
## 6  0.02510772      6 0.4783890 0.4872602 0.011267948
## 7  0.02250559      7 0.4532813 0.4681798 0.010943372
## 8  0.01505899      9 0.4082701 0.4248675 0.009989956
## 9  0.01505464     10 0.3932111 0.4047923 0.008985099
## 10 0.01236347     11 0.3781565 0.3852021 0.008791926
## 11 0.01000000     12 0.3657930 0.3708029 0.008292214
##
## Variable importance
## exclusive_use_area           school10         hospital15           school15
##                 20                  9                  9                  9
##              kid15         hospital10        bigMarket15           subway05
##                  8                  6                  5                  5
##            movie05               dong              kid10              kid05
##                  4                  4                  3                  3
##            movie10            movie15         hospital05           office15
##                  2                  2                  2                  2
##        bigMarket10   transaction_month           school05           subway15
##                  2                  1                  1                  1
##        bigMarket05
##                  1
##
## Node number 1: 6412 observations,    complexity param=0.2487868
##   mean=539.699, MSE=10367.63
##   left son=2 (6054 obs) right son=3 (358 obs)
##   Primary splits:
##       exclusive_use_area < 38.58   to the right, improve=0.2487868, (0 missing)
##       kid15              < 175     to the left,  improve=0.2242459, (0 missing)
##       hospital15         < 314.5   to the left,  improve=0.2091796, (0 missing)
##       bigMarket15        < 10.5    to the left,  improve=0.2011589, (0 missing)
##       hospital10         < 228.5   to the left,  improve=0.1972842, (0 missing)
##
## Node number 2: 6054 observations,    complexity param=0.1233842
##   mean=527.3488, MSE=6904.146
##   left son=4 (4254 obs) right son=5 (1800 obs)
##   Primary splits:
##       hospital15  < 314.5   to the left,  improve=0.1962367, (0 missing)
##       school10    < 16.5    to the left,  improve=0.1755251, (0 missing)
##       bigMarket15 < 10.5    to the left,  improve=0.1741203, (0 missing)
##       school15    < 33.5    to the left,  improve=0.1552185, (0 missing)
##       kid15       < 175.5   to the left,  improve=0.1271009, (0 missing)
##   Surrogate splits:
##       school10    < 17.5    to the left,  agree=0.911, adj=0.702, (0 split)
##       hospital10  < 118.5   to the left,  agree=0.886, adj=0.616, (0 split)
##       school15    < 33.5    to the left,  agree=0.877, adj=0.585, (0 split)
```

```
##       kid15       < 158.5   to the left,  agree=0.838, adj=0.454, (0 split)
##       bigMarket15 < 11.5    to the left,  agree=0.836, adj=0.448, (0 split)
##
## Node number 3: 358 observations,    complexity param=0.05405969
##   mean=748.5483, MSE=22739.88
##   left son=6 (220 obs) right son=7 (138 obs)
##   Primary splits:
##       subway05   < 1.5     to the left,  improve=0.4414438, (0 missing)
##       kid10      < 81.5    to the left,  improve=0.4190922, (0 missing)
##       hospital10 < 228.5   to the left,  improve=0.4127252, (0 missing)
##       kid15      < 159.5   to the left,  improve=0.4104417, (0 missing)
##       school10   < 23      to the left,  improve=0.3814535, (0 missing)
##   Surrogate splits:
##       movie05    < 5.5     to the left,  agree=0.888, adj=0.710, (0 split)
##       school15   < 41      to the left,  agree=0.885, adj=0.703, (0 split)
##       office15   < 18.5    to the left,  agree=0.883, adj=0.696, (0 split)
##       school10   < 20.5    to the left,  agree=0.877, adj=0.681, (0 split)
##       hospital05 < 140     to the left,  agree=0.860, adj=0.638, (0 split)
##
## Node number 4: 4254 observations,    complexity param=0.03999167
##   mean=503.4056, MSE=5139.337
##   left son=8 (2153 obs) right son=9 (2101 obs)
##   Primary splits:
##       movie10  < 4.5     to the left,  improve=0.12160120, (0 missing)
##       subway05 < 1.5     to the left,  improve=0.09424765, (0 missing)
##       subway15 < 8.5     to the left,  improve=0.08187867, (0 missing)
##       school05 < 0.5     to the right, improve=0.07831775, (0 missing)
##       kid05    < 23.5    to the left,  improve=0.06754998, (0 missing)
##   Surrogate splits:
##       dong        splits as  RLLRR,       agree=0.824, adj=0.644, (0 split)
##       movie15     < 8.5     to the left,  agree=0.787, adj=0.568, (0 split)
##       bigMarket15 < 7.5     to the left,  agree=0.688, adj=0.368, (0 split)
##       movie05     < 0.5     to the left,  agree=0.670, adj=0.332, (0 split)
##       kid15       < 138.5   to the right, agree=0.632, adj=0.254, (0 split)
##
## Node number 5: 1800 observations,    complexity param=0.02250559
##   mean=583.9346, MSE=6518.178
##   left son=10 (196 obs) right son=11 (1604 obs)
##   Primary splits:
##       exclusive_use_area < 84.935  to the right, improve=0.12508150, (0 missing)
##       transaction_month  splits as  LLLLLLRRRRR, improve=0.11328940, (0 missing)
##       kid15              < 185.5   to the left,  improve=0.09235544, (0 missing)
##       school10           < 23.5    to the left,  improve=0.09196631, (0 missing)
##       hospital05         < 53      to the left,  improve=0.08220254, (0 missing)
##   Surrogate splits:
##       office10    < 12.5    to the right, agree=0.916, adj=0.230, (0 split)
##       bigMarket15 < 12.5    to the right, agree=0.913, adj=0.204, (0 split)
##       bigMarket10 < 8.5     to the right, agree=0.908, adj=0.153, (0 split)
##       school05    < 9.5     to the right, agree=0.908, adj=0.153, (0 split)
##       hospital05  < 7.5     to the left,  agree=0.908, adj=0.153, (0 split)
##
## Node number 6: 220 observations,    complexity param=0.02510772
##   mean=669.1959, MSE=12730.15
##   left son=12 (139 obs) right son=13 (81 obs)
##   Primary splits:
##       kid15      < 159.5   to the left,  improve=0.5959693, (0 missing)
```

```
##       kid10      < 81.5    to the left,  improve=0.5570633, (0 missing)
##       kid05      < 29.5    to the left,  improve=0.5265200, (0 missing)
##       hospital10 < 236     to the left,  improve=0.5265200, (0 missing)
##       school15   < 37.5    to the left,  improve=0.4744166, (0 missing)
##   Surrogate splits:
##       hospital10  < 227.5   to the left,  agree=0.955, adj=0.877, (0 split)
##       kid10       < 81.5    to the left,  agree=0.955, adj=0.877, (0 split)
##       kid05       < 29.5    to the left,  agree=0.945, adj=0.852, (0 split)
##       school15    < 37.5    to the left,  agree=0.905, adj=0.741, (0 split)
##       bigMarket05 < 1.5     to the right, agree=0.882, adj=0.679, (0 split)
##
## Node number 7: 138 observations
##   mean=875.0521, MSE=12655.82
##
## Node number 8: 2153 observations,    complexity param=0.01505899
##   mean=478.7103, MSE=4176.239
##   left son=16 (869 obs) right son=17 (1284 obs)
##   Primary splits:
##       exclusive_use_area < 59.9965 to the right, improve=0.11133690, (0 missing)
##       kid05              < 14.5    to the left,  improve=0.07040976, (0 missing)
##       kid10              < 50.5    to the left,  improve=0.05769652, (0 missing)
##       movie15            < 4.5     to the left,  improve=0.05567656, (0 missing)
##       office15           < 17.5    to the left,  improve=0.05298594, (0 missing)
##   Surrogate splits:
##       year_of_completion_f splits as  RRLL,        agree=0.676, adj=0.198, (0 split)
##       movie15              < 4.5     to the left,  agree=0.655, adj=0.146, (0 split)
##       school10             < 8.5     to the left,  agree=0.649, adj=0.131, (0 split)
##       movie05              < 0.5     to the right, agree=0.647, adj=0.124, (0 split)
##       kid05                < 14.5    to the left,  agree=0.636, adj=0.098, (0 split)
##
## Node number 9: 2101 observations,    complexity param=0.02769431
##   mean=528.712, MSE=4860.905
##   left son=18 (1706 obs) right son=19 (395 obs)
##   Primary splits:
##       dong     splits as  L-RLL,       improve=0.1477150, (0 missing)
##       subway15 < 8.5     to the left,  improve=0.1312419, (0 missing)
##       kid15    < 147     to the left,  improve=0.1186331, (0 missing)
##       school05 < 0.5     to the right, improve=0.1154077, (0 missing)
##       movie15  < 9.5     to the right, improve=0.1145823, (0 missing)
##   Surrogate splits:
##       subway15 < 8.5     to the left,  agree=0.982, adj=0.904, (0 split)
##       kid15    < 145.5   to the left,  agree=0.976, adj=0.873, (0 split)
##       subway05 < 1.5     to the left,  agree=0.972, adj=0.853, (0 split)
##       school05 < 1.5     to the right, agree=0.968, adj=0.828, (0 split)
##       movie15  < 9.5     to the right, agree=0.922, adj=0.585, (0 split)
##
## Node number 10: 196 observations
##   mean=502.2512, MSE=4859.875
##
## Node number 11: 1604 observations,    complexity param=0.02250559
##   mean=593.9159, MSE=5805.885
##   left son=22 (1044 obs) right son=23 (560 obs)
##   Primary splits:
##       transaction_month splits as  LLLLLLRRRRR, improve=0.16372090, (0 missing)
##       kid15             < 187.5   to the left,  improve=0.13814160, (0 missing)
##       bigMarket15       < 11.5    to the left,  improve=0.12591020, (0 missing)
```

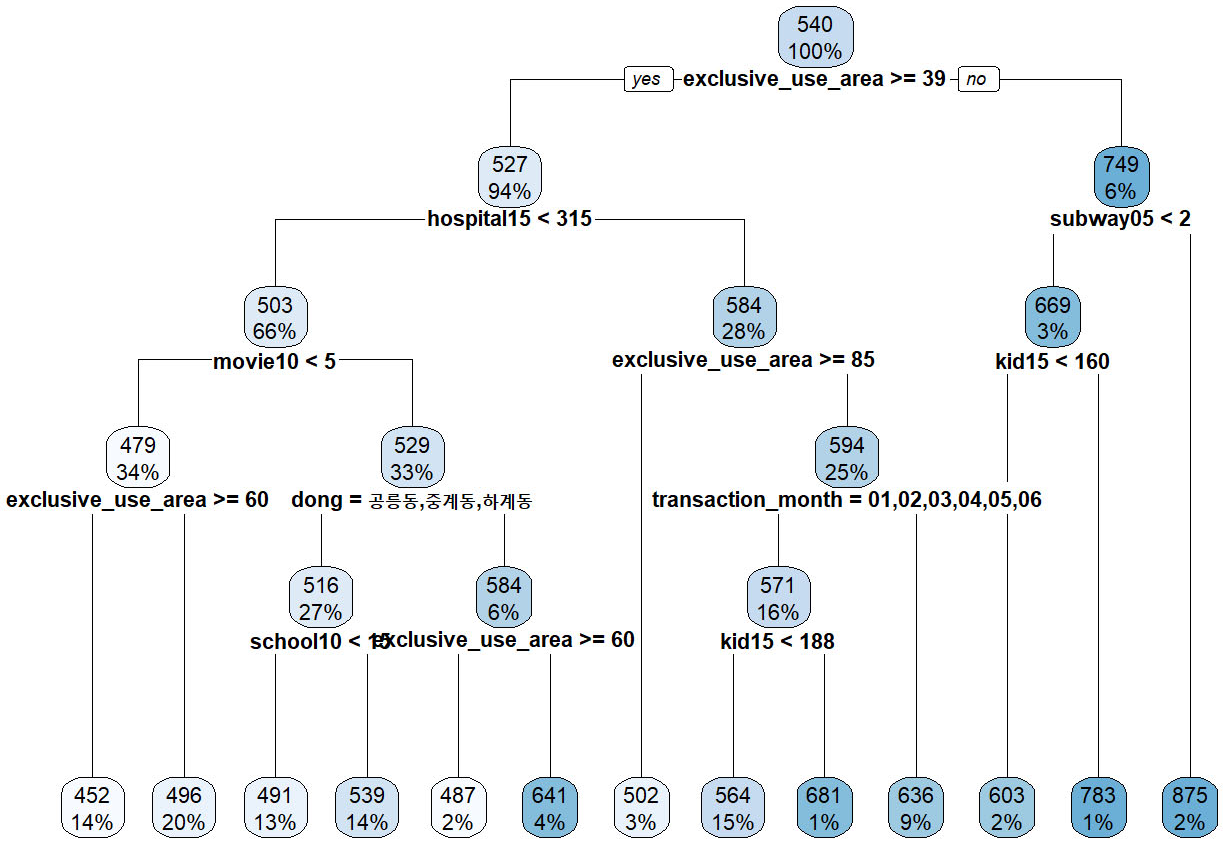
```
##       school10          < 23.5    to the left,  improve=0.11328200, (0 missing)
##       kid05             < 17.5    to the right, improve=0.09603846, (0 missing)
##   Surrogate splits:
##       school10           < 23.5    to the left,  agree=0.658, adj=0.021, (0 split)
##       kid10              < 61.5    to the right, agree=0.658, adj=0.020, (0 split)
##       exclusive_use_area < 84.89   to the left,  agree=0.657, adj=0.018, (0 split)
##       movie15            < 16.5    to the left,  agree=0.657, adj=0.018, (0 split)
##       subway10           < 0.5     to the right, agree=0.655, adj=0.011, (0 split)
##
## Node number 12: 139 observations
##   mean=602.7048, MSE=5121.485
##
## Node number 13: 81 observations
##   mean=783.298, MSE=5180.932
##
## Node number 16: 869 observations
##   mean=452.4992, MSE=4082.516
##
## Node number 17: 1284 observations
##   mean=496.4497, MSE=3460.013
##
## Node number 18: 1706 observations,    complexity param=0.01505464
##   mean=515.8183, MSE=3165.675
##   left son=36 (825 obs) right son=37 (881 obs)
##   Primary splits:
##       school10    < 14.5    to the left,  improve=0.1853096, (0 missing)
##       bigMarket15 < 10.5    to the left,  improve=0.1761978, (0 missing)
##       school15    < 30.5    to the left,  improve=0.1750130, (0 missing)
##       hospital10  < 90.5    to the right, improve=0.1050725, (0 missing)
##       movie05     < 2.5     to the left,  improve=0.1016477, (0 missing)
##   Surrogate splits:
##       school15    < 30.5    to the left,  agree=0.858, adj=0.707, (0 split)
##       bigMarket15 < 9.5     to the left,  agree=0.841, adj=0.670, (0 split)
##       kid10       < 62.5    to the left,  agree=0.817, adj=0.621, (0 split)
##       bigMarket10 < 3.5     to the left,  agree=0.814, adj=0.616, (0 split)
##       dong        splits as  L--RR,       agree=0.805, adj=0.596, (0 split)
##
## Node number 19: 395 observations,    complexity param=0.02769431
##   mean=584.4001, MSE=8363.396
##   left son=38 (145 obs) right son=39 (250 obs)
##   Primary splits:
##       exclusive_use_area < 59.95   to the right, improve=0.6579316, (0 missing)
##       kid05              < 23.5    to the left,  improve=0.3775827, (0 missing)
##       kid10              < 78.5    to the left,  improve=0.3027202, (0 missing)
##       hospital10         < 92      to the right, improve=0.2802189, (0 missing)
##       office05           < 1.5     to the right, improve=0.2750958, (0 missing)
##   Surrogate splits:
##       kid05       < 23.5    to the left,  agree=0.853, adj=0.600, (0 split)
##       bigMarket10 < 3.5     to the left,  agree=0.830, adj=0.538, (0 split)
##       hospital15  < 194.5   to the right, agree=0.830, adj=0.538, (0 split)
##       movie05     < 0.5     to the right, agree=0.830, adj=0.538, (0 split)
##       kid10       < 71      to the left,  agree=0.830, adj=0.538, (0 split)
##
## Node number 22: 1044 observations,    complexity param=0.01236347
##   mean=571.3356, MSE=4478.66
##   left son=44 (980 obs) right son=45 (64 obs)
```

```
##   Primary splits:
##       kid15       < 187.5   to the left,  improve=0.1757781, (0 missing)
##       bigMarket15 < 11.5    to the left,  improve=0.1459677, (0 missing)
##       hospital10  < 237     to the left,  improve=0.1226298, (0 missing)
##       school10    < 17.5    to the right, improve=0.1181325, (0 missing)
##       school15    < 37.5    to the left,  improve=0.1088628, (0 missing)
##   Surrogate splits:
##       hospital10  < 265.5   to the left,  agree=0.967, adj=0.469, (0 split)
##       school10    < 17.5    to the right, agree=0.958, adj=0.312, (0 split)
##       bigMarket05 < 0.5     to the right, agree=0.951, adj=0.203, (0 split)
##
## Node number 23: 560 observations
##   mean=636.012, MSE=5557.581
##
## Node number 36: 825 observations
##   mean=490.7893, MSE=2982.446
##
## Node number 37: 881 observations
##   mean=539.2563, MSE=2201.285
##
## Node number 38: 145 observations
##   mean=486.9981, MSE=4179.327
##
## Node number 39: 250 observations
##   mean=640.8932, MSE=2096.138
##
## Node number 44: 980 observations
##   mean=564.1653, MSE=3747.213
##
## Node number 45: 64 observations
##   mean=681.1298, MSE=2836.908
```

```
install.packages("rpart.plot", repos = "http://cran.us.r-project.org")
```

```
## 패키지 'rpart.plot'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:WUsersWLUISWAppDataWLocalWTempWRtmpEZswNfWdownloaded_packages
```

```
library(rpart.plot)

rpart.plot(tree1, cex = 0.7)
```

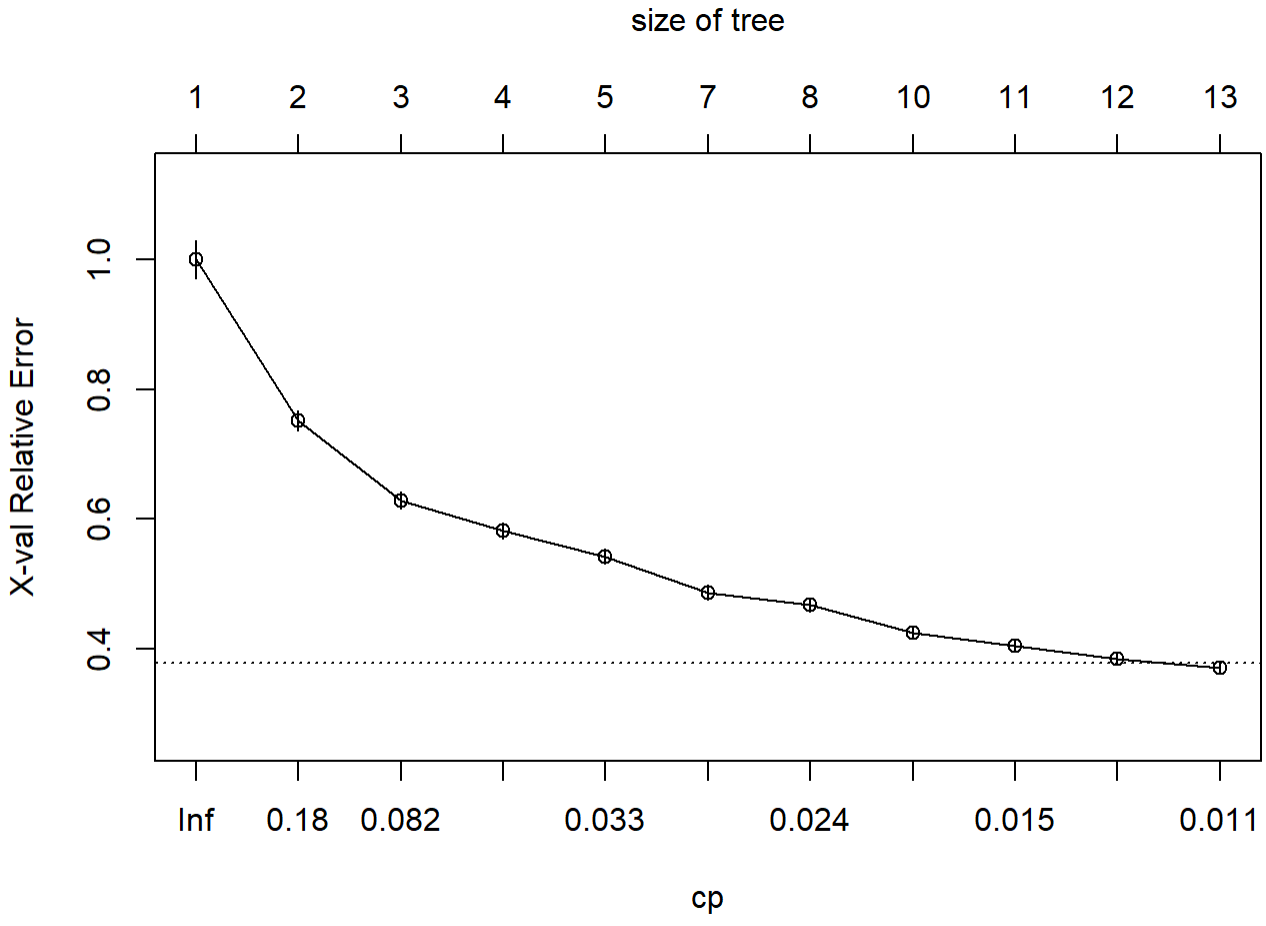


Decision Tree parameter tuning

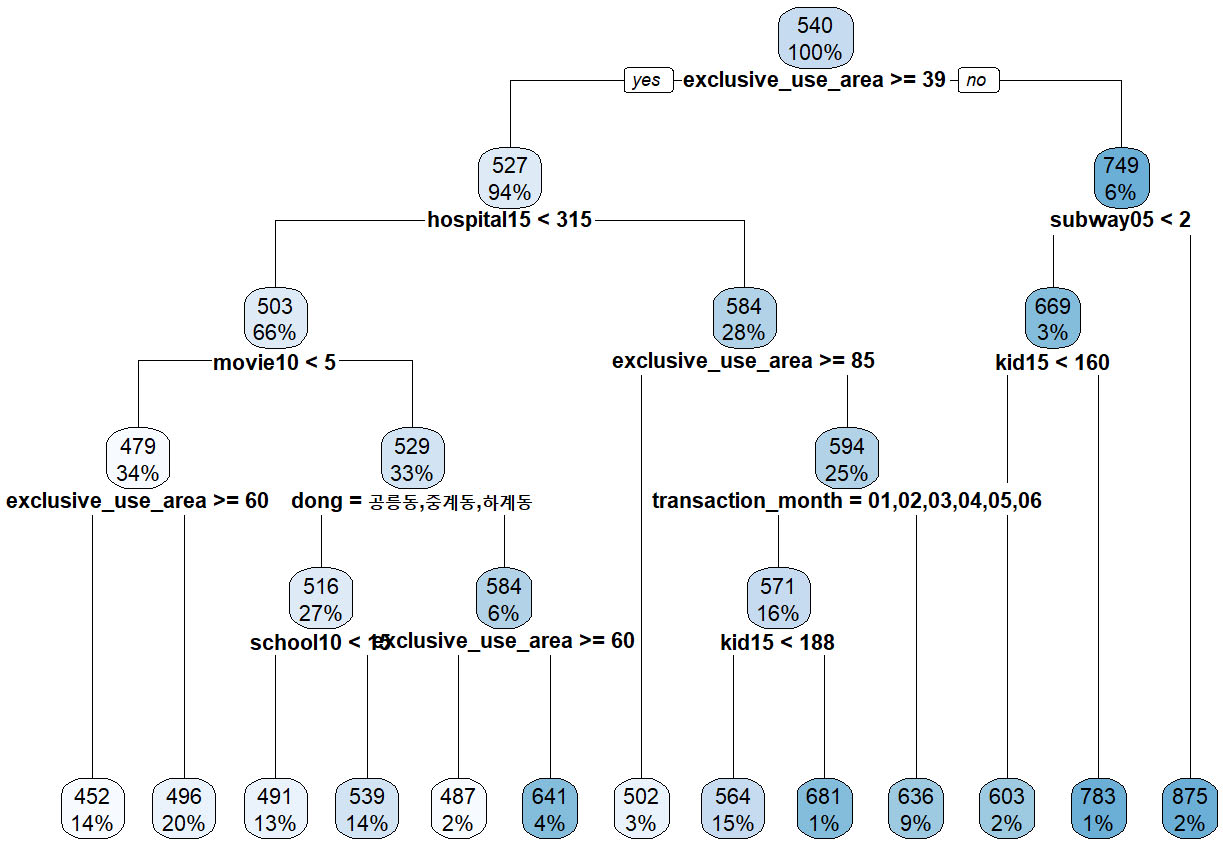
```
printcp(tree1)
```

```
##
## Regression tree:
## rpart(formula = unit_price ~ . - year, data = data_train1, method = "anova",
##     control = rpart.control(minsplit = 50, maxdepth = 5))
##
## Variables actually used in tree construction:
## [1] dong               exclusive_use_area hospital15         kid15
## [5] movie10            school10           subway05           transaction_month
##
## Root node error: 66477236/6412 = 10368
##
## n= 6412
##
##          CP nsplit rel error  xerror      xstd
## 1  0.248787      0   1.00000 1.00023 0.0290602
## 2  0.123384      1   0.75121 0.75223 0.0150142
## 3  0.054060      2   0.62783 0.62891 0.0133603
## 4  0.039992      3   0.57377 0.58232 0.0122489
## 5  0.027694      4   0.53378 0.54247 0.0118086
## 6  0.025108      6   0.47839 0.48726 0.0112679
## 7  0.022506      7   0.45328 0.46818 0.0109434
## 8  0.015059      9   0.40827 0.42487 0.0099900
## 9  0.015055     10   0.39321 0.40479 0.0089851
## 10 0.012363     11   0.37816 0.38520 0.0087919
## 11 0.010000     12   0.36579 0.37080 0.0082922
```

```
plotcp(tree1)
```



```
tree1 <- prune(tree1, cp= tree1$cptable[which.min(tree1$cptable[, "xerror"]), "CP"])

rpart.plot(tree1, cex = 0.7)
```



Decision Tree prediction & RMSE calculation

```
# test data 에 적용

predict_1 <- predict(tree1, data_test1)
summary(predict_1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   452.5   490.8   502.3   537.7   564.2   875.1
```

```
# actual, predicted cbind

databind1 <- cbind(data_test1[,25],predict_1)
databind1 <- as.data.frame(databind1)
summary(databind1)
```

```
##        V1           predict_1
##  Min.   : 242.2   Min.   :452.5
##  1st Qu.: 475.3   1st Qu.:490.8
##  Median : 522.9   Median :502.3
##  Mean   : 536.3   Mean   :537.7
##  3rd Qu.: 582.7   3rd Qu.:564.2
##  Max.   :1094.4   Max.   :875.1
```

```
# RMSE 계산
install.packages("Metrics", repos = "http://cran.us.r-project.org")
```

```
## 패키지 'Metrics'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUISW₩AppData₩Local₩Temp₩RtmpEZswNf₩downloaded_packages
```

```
library(Metrics)
rmse(databind1$V1, databind1$predict_1)
```

```
## [1] 60.02746
```

Linear regression

```
# factor 변수 중 unique value 있는지 찾아보기
str(data_train1)
```

```
## 'data.frame':    6412 obs. of  28 variables:
##  $ dong               : Factor w/ 5 levels "공릉동","상계동",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ exclusive_use_area : num  59.9 60 84.8 114.8 50.1 ...
##  $ floor              : int  9 22 4 4 5 6 10 2 2 6 ...
##  $ bigMarket05        : int  1 0 2 0 2 2 1 1 1 2 ...
##  $ bigMarket10        : int  4 5 3 4 5 5 4 4 3 5 ...
##  $ bigMarket15        : int  9 5 8 6 9 9 8 8 6 9 ...
##  $ school05           : int  5 6 0 2 0 0 3 3 9 0 ...
##  $ school10           : int  15 9 13 8 12 12 15 15 14 12 ...
##  $ school15           : int  28 28 30 28 25 25 28 28 25 25 ...
##  $ subway05           : int  1 0 2 2 2 2 0 0 0 2 ...
##  $ subway10           : int  1 5 2 4 5 5 3 3 1 5 ...
##  $ subway15           : int  5 7 8 9 9 9 5 5 4 9 ...
##  $ hospital05         : int  13 8 11 8 14 14 12 12 7 14 ...
##  $ hospital10         : int  38 86 43 61 101 101 40 40 38 101 ...
##  $ hospital15         : int  113 183 175 193 194 194 122 122 99 194 ...
##  $ movie05            : int  0 2 1 2 0 0 0 0 0 0 ...
##  $ movie10            : int  1 4 5 4 5 5 0 0 0 5 ...
##  $ movie15            : int  7 7 6 8 9 9 5 5 6 9 ...
##  $ kid05              : int  12 18 20 21 24 24 14 14 11 24 ...
##  $ kid10              : int  53 58 75 71 72 72 65 65 43 72 ...
##  $ kid15              : int  139 140 148 152 154 154 137 137 111 154 ...
##  $ office05           : int  3 5 1 2 2 2 2 2 1 2 ...
##  $ office10           : int  5 10 9 10 10 10 7 7 8 10 ...
##  $ office15           : int  15 21 21 20 19 19 18 18 13 19 ...
##  $ unit_price         : num  484 509 401 372 574 ...
##  $ transaction_month  : Factor w/ 11 levels "01","02","03",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ year               : Factor w/ 32 levels "1976","1983",..: 25 19 25 19 5 5 11 11 17 5
...
##  $ year_of_completion_f: Factor w/ 4 levels "1st","2nd","3rd",..: 3 2 3 2 1 1 1 1 2 1 ...
```

```
sapply(lapply(data_train1, unique), length)
```

```
##                dong   exclusive_use_area               floor
##                   5                  423                  30
##         bigMarket05         bigMarket10         bigMarket15
##                   6                   10                  14
##            school05            school10            school15
##                  11                   25                  36
##            subway05            subway10            subway15
##                   4                    8                  11
##          hospital05          hospital10          hospital15
##                  67                  141                 172
##             movie05             movie10             movie15
##                   7                   11                  18
##               kid05               kid10               kid15
##                  32                  71                  112
##            office05            office10            office15
##                   8                   12                  18
##          unit_price    transaction_month                year
##                4315                   11                  32
## year_of_completion_f
##                   4
```

```
# Linear Model  (dong은 제외하고 분석:삭제)
linear1 <- lm(unit_price ~.-year, data = data_train1)
#linear1 <- lm(unit_price ~ dong+exclusive_use_area+floor+bigMarket05+bigMarket10+bigMarket15+school05+school10+school15+subway05+subway10+subway15+hospital05+hospital10+hospital15+movie05+movie10+movie15+kid05+kid10+kid15+office05+office10+office15+transaction_month+year_of_completion_f, data = data_train1)

summary(linear1)
```

```
##
## Call:
## lm(formula = unit_price ~ . - year, data = data_train1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -295.22  -42.86   -5.39   36.21  373.50
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             267.92465   12.57748  21.302  < 2e-16 ***
## dong상계동               87.25876    6.38799  13.660  < 2e-16 ***
## dong월계동               74.43998    5.74248  12.963  < 2e-16 ***
## dong중계동               66.80086    6.81155   9.807  < 2e-16 ***
## dong하계동               -3.69633    6.10889  -0.605 0.545152
## exclusive_use_area       -1.34112    0.04239 -31.637  < 2e-16 ***
## floor                     0.42586    0.17765   2.397 0.016548 *
## bigMarket05              -8.39306    1.05213  -7.977 1.76e-15 ***
## bigMarket10              -0.36912    1.07931  -0.342 0.732363
## bigMarket15               9.65886    0.93830  10.294  < 2e-16 ***
## school05                 -2.36050    0.59088  -3.995 6.54e-05 ***
## school10                  2.59363    0.45444   5.707 1.20e-08 ***
## school15                 -1.14520    0.32923  -3.478 0.000508 ***
## subway05                 28.58813    2.09706  13.632  < 2e-16 ***
## subway10                 -1.53773    1.07752  -1.427 0.153601
## subway15                -10.14500    1.02849  -9.864  < 2e-16 ***
## hospital05                0.37737    0.05042   7.485 8.13e-14 ***
## hospital10               -0.25663    0.03814  -6.728 1.86e-11 ***
## hospital15               -0.40752    0.03135 -12.997  < 2e-16 ***
## movie05                   3.39318    0.93015   3.648 0.000266 ***
## movie10                  14.06352    0.67937  20.701  < 2e-16 ***
## movie15                   8.69721    0.61961  14.037  < 2e-16 ***
## kid05                     1.44848    0.21436   6.757 1.53e-11 ***
## kid10                     0.07240    0.18008   0.402 0.687677
## kid15                     1.95364    0.13301  14.687  < 2e-16 ***
## office05                -15.00090    0.83562 -17.952  < 2e-16 ***
## office10                 -0.29944    0.58914  -0.508 0.611284
## office15                 -5.98075    0.57143 -10.466  < 2e-16 ***
## transaction_month02      -2.86157    5.58146  -0.513 0.608184
## transaction_month03      -8.44675    5.20203  -1.624 0.104481
## transaction_month04      -3.23329    5.13982  -0.629 0.529327
## transaction_month05       5.12644    4.88349   1.050 0.293874
## transaction_month06      24.50749    4.84921   5.054 4.45e-07 ***
## transaction_month07      43.88002    4.76147   9.216  < 2e-16 ***
## transaction_month08      51.09161    5.93484   8.609  < 2e-16 ***
## transaction_month09      54.94882    5.72229   9.603  < 2e-16 ***
## transaction_month10      56.55586    5.83680   9.690  < 2e-16 ***
## transaction_month11      60.81567    5.63293  10.796  < 2e-16 ***
## year_of_completion_f2nd  34.26662    2.90071  11.813  < 2e-16 ***
## year_of_completion_f3rd  35.18841    3.93333   8.946  < 2e-16 ***
## year_of_completion_f4th 163.88689    6.69444  24.481  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.33 on 6371 degrees of freedom
```

```
## Multiple R-squared:  0.5655, Adjusted R-squared:  0.5628
## F-statistic: 207.3 on 40 and 6371 DF,  p-value: < 2.2e-16
```

```
print(linear1)
```

```
##
## Call:
## lm(formula = unit_price ~ . - year, data = data_train1)
##
## Coefficients:
##             (Intercept)               dong상계동               dong월계동
##                267.9246                  87.2588                  74.4400
##              dong중계동               dong하계동       exclusive_use_area
##                 66.8009                  -3.6963                  -1.3411
##                   floor              bigMarket05              bigMarket10
##                  0.4259                  -8.3931                  -0.3691
##             bigMarket15                 school05                 school10
##                  9.6589                  -2.3605                   2.5936
##                school15                 subway05                 subway10
##                 -1.1452                  28.5881                  -1.5377
##                subway15                hospital05                hospital10
##                -10.1450                   0.3774                  -0.2566
##              hospital15                  movie05                  movie10
##                 -0.4075                   3.3932                  14.0635
##                 movie15                    kid05                    kid10
##                  8.6972                   1.4485                   0.0724
##                   kid15                 office05                 office10
##                  1.9536                 -15.0009                  -0.2994
##                office15      transaction_month02      transaction_month03
##                 -5.9807                  -2.8616                  -8.4467
##     transaction_month04      transaction_month05      transaction_month06
##                 -3.2333                   5.1264                  24.5075
##     transaction_month07      transaction_month08      transaction_month09
##                 43.8800                  51.0916                  54.9488
##     transaction_month10      transaction_month11  year_of_completion_f2nd
##                 56.5559                  60.8157                  34.2666
## year_of_completion_f3rd  year_of_completion_f4th
##                 35.1884                 163.8869
```

```
linear1$coefficients
```

```
##             (Intercept)              dong상계동              dong월계동
##            267.92464690             87.25875530             74.43998152
##              dong중계동              dong하계동       exclusive_use_area
##             66.80085546             -3.69632729             -1.34112176
##                   floor             bigMarket05             bigMarket10
##              0.42585903             -8.39305796             -0.36912238
##             bigMarket15                school05                school10
##              9.65885929             -2.36050423              2.59362559
##                school15                subway05                subway10
##             -1.14519898             28.58813362             -1.53772606
##                subway15              hospital05              hospital10
##            -10.14500157              0.37737490             -0.25662697
##              hospital15                 movie05                 movie10
##             -0.40751572              3.39318092             14.06351881
##                 movie15                    kid05                    kid10
##              8.69720765              1.44848315              0.07239747
##                    kid15                office05                office10
##              1.95364199            -15.00090174             -0.29944255
##                office15     transaction_month02     transaction_month03
##             -5.98074544             -2.86157291             -8.44674972
##     transaction_month04     transaction_month05     transaction_month06
##             -3.23329185              5.12643559             24.50749386
##     transaction_month07     transaction_month08     transaction_month09
##             43.88002462             51.09160901             54.94882224
##     transaction_month10     transaction_month11 year_of_completion_f2nd
##             56.55586334             60.81566911             34.26662268
## year_of_completion_f3rd year_of_completion_f4th
##             35.18841301            163.88688661
```

Linear regression parameter tuning

```
step(linear1, direction = "both")
```

```
## Start:  AIC=54024.89
## unit_price ~ (dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket10 + bigMarket15 + school05 + school10 + school15 +
##     subway05 + subway10 + subway15 + hospital05 + hospital10 +
##     hospital15 + movie05 + movie10 + movie15 + kid05 + kid10 +
##     kid15 + office05 + office10 + office15 + transaction_month +
##     year + year_of_completion_f) - year
##
##                        Df Sum of Sq      RSS   AIC
## - bigMarket10           1       530 28882519 54023
## - kid10                 1       733 28882722 54023
## - office10              1      1171 28883160 54023
## <none>                               28881989 54025
## - subway10              1      9233 28891222 54025
## - floor                 1     26052 28908041 54029
## - school15              1     54852 28936841 54035
## - movie05               1     60329 28942318 54036
## - school05              1     72349 28954338 54039
## - school10              1    147663 29029652 54056
## - hospital10            1    205236 29087225 54068
## - kid05                 1    207000 29088989 54069
## - hospital05            1    253972 29135961 54079
## - bigMarket05           1    288484 29170473 54087
## - subway15              1    441085 29323074 54120
## - bigMarket15           1    480387 29362376 54129
## - office15              1    496602 29378591 54132
## - hospital15            1    765775 29647764 54191
## - subway05              1    842499 29724488 54207
## - movie15               1    893183 29775172 54218
## - kid15                 1    977944 29859933 54236
## - office05              1   1460955 30342944 54339
## - dong                  4   1678825 30560814 54379
## - movie10               1   1942672 30824661 54440
## - year_of_completion_f  3   2971001 31852990 54647
## - transaction_month    10   3608986 32490975 54760
## - exclusive_use_area    1   4537313 33419302 54958
##
## Step:  AIC=54023.01
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital10 + hospital15 +
##     movie05 + movie10 + movie15 + kid05 + kid10 + kid15 + office05 +
##     office10 + office15 + transaction_month + year_of_completion_f
##
##                        Df Sum of Sq      RSS   AIC
## - kid10                 1       562 28883082 54021
## - office10              1       988 28883508 54021
## <none>                               28882519 54023
## - subway10              1      9284 28891803 54023
## + bigMarket10           1       530 28881989 54025
## - floor                 1     25902 28908422 54027
## - school15              1     55608 28938128 54033
## - movie05               1     59970 28942489 54034
## - school05              1     72591 28955110 54037
## - school10              1    155392 29037911 54055
```

```
## - hospital10            1    204897 29087416 54066
## - kid05                 1    227565 29110085 54071
## - hospital05            1    262169 29144688 54079
## - bigMarket05           1    289723 29172242 54085
## - subway15              1    444789 29327308 54119
## - bigMarket15           1    480269 29362788 54127
## - office15              1    582655 29465174 54149
## - hospital15            1    773627 29656146 54190
## - subway05              1    847210 29729729 54206
## - movie15               1    892863 29775382 54216
## - kid15                 1    978104 29860623 54235
## - office05              1   1483152 30365671 54342
## - dong                  4   1682139 30564659 54378
## - movie10               1   1960762 30843282 54442
## - year_of_completion_f  3   3052170 31934689 54661
## - transaction_month    10   3610070 32492589 54758
## - exclusive_use_area    1   4538689 33421209 54957
##
## Step:  AIC=54021.13
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital10 + hospital15 +
##     movie05 + movie10 + movie15 + kid05 + kid15 + office05 +
##     office10 + office15 + transaction_month + year_of_completion_f
##
##                        Df Sum of Sq      RSS   AIC
## - office10              1       700 28883781 54019
## <none>                              28883082 54021
## - subway10              1     10120 28893201 54021
## + kid10                 1       562 28882519 54023
## + bigMarket10           1       360 28882722 54023
## - floor                 1     25701 28908782 54025
## - school15              1     55122 28938204 54031
## - movie05               1     59470 28942552 54032
## - school05              1     72086 28955168 54035
## - school10              1    156372 29039454 54054
## - hospital10            1    206126 29089208 54065
## - kid05                 1    258433 29141514 54076
## - hospital05            1    272302 29155384 54079
## - bigMarket05           1    290976 29174058 54083
## - subway15              1    450077 29333158 54118
## - bigMarket15           1    490756 29373838 54127
## - office15              1    587609 29470691 54148
## - hospital15            1    773148 29656229 54189
## - subway05              1    914778 29797860 54219
## - movie15               1    916765 29799846 54219
## - kid15                 1   1438766 30321847 54331
## - office05              1   1501107 30384188 54344
## - dong                  4   1791680 30674761 54399
## - movie10               1   1967752 30850834 54442
## - year_of_completion_f  3   3062938 31946020 54661
## - transaction_month    10   3615320 32498401 54757
## - exclusive_use_area    1   4550053 33433135 54957
##
## Step:  AIC=54019.29
## unit_price ~ dong + exclusive_use_area + floor + bigMarket05 +
```

```
##     bigMarket15 + school05 + school10 + school15 + subway05 +
##     subway10 + subway15 + hospital05 + hospital10 + hospital15 +
##     movie05 + movie10 + movie15 + kid05 + kid15 + office05 +
##     office15 + transaction_month + year_of_completion_f
##
##                         Df Sum of Sq      RSS   AIC
## <none>                                28883781 54019
## - subway10               1     10681 28894462 54020
## + office10               1       700 28883082 54021
## + kid10                  1       274 28883508 54021
## + bigMarket10            1       267 28883514 54021
## - floor                  1     25544 28909325 54023
## - school15               1     55797 28939578 54030
## - movie05                1     62385 28946166 54031
## - school05               1     71405 28955187 54033
## - school10               1    156248 29040029 54052
## - hospital10             1    206297 29090079 54063
## - kid05                  1    259109 29142890 54075
## - hospital05             1    271602 29155384 54077
## - bigMarket05            1    290297 29174078 54081
## - subway15               1    454109 29337890 54117
## - bigMarket15            1    490893 29374675 54125
## - office15               1    606273 29490054 54150
## - hospital15             1    801242 29685023 54193
## - subway05               1    920697 29804478 54218
## - movie15                1    959305 29843086 54227
## - kid15                  1   1439776 30323557 54329
## - office05               1   1543005 30426786 54351
## - dong                   4   1841420 30725201 54408
## - movie10                1   2004939 30888720 54448
## - year_of_completion_f   3   3062406 31946188 54659
## - transaction_month     10   3614789 32498570 54755
## - exclusive_use_area     1   4570269 33454051 54959
```

```
##
## Call:
## lm(formula = unit_price ~ dong + exclusive_use_area + floor +
##     bigMarket05 + bigMarket15 + school05 + school10 + school15 +
##     subway05 + subway10 + subway15 + hospital05 + hospital10 +
##     hospital15 + movie05 + movie10 + movie15 + kid05 + kid15 +
##     office05 + office15 + transaction_month + year_of_completion_f,
##     data = data_train1)
##
## Coefficients:
##             (Intercept)              dong상계동              dong월계동
##                267.4724                 87.4444                 74.2028
##              dong중계동              dong하계동       exclusive_use_area
##                 66.2626                 -3.8361                 -1.3428
##                   floor             bigMarket05             bigMarket15
##                  0.4213                 -8.3369                  9.5983
##               school05                school10                school15
##                 -2.3135                  2.5541                 -1.1412
##               subway05                subway10                subway15
##                 28.7422                 -1.6315                -10.1384
##              hospital05              hospital10              hospital15
##                  0.3745                 -0.2551                 -0.4079
##                movie05                 movie10                 movie15
##                  3.3535                 14.0517                  8.6922
##                  kid05                   kid15                office05
##                  1.4912                  1.9746                -14.9690
##               office15     transaction_month02     transaction_month03
##                 -6.0568                 -2.8530                 -8.4119
##     transaction_month04     transaction_month05     transaction_month06
##                 -3.1819                  5.2115                 24.5491
##     transaction_month07     transaction_month08     transaction_month09
##                 43.9323                 51.1828                 55.0075
##     transaction_month10     transaction_month11 year_of_completion_f2nd
##                 56.6557                 60.8148                 34.2951
## year_of_completion_f3rd year_of_completion_f4th
##                 35.2770                164.0195
```

Linear regression prediction & RMSE calculation

```
linear_best<-lm(formula = unit_price ~ dong + floor + bigMarket05 + bigMarket15 +
    school05 + school10 + school15 + subway05 + subway10 + subway15 +
    hospital05 + hospital15 + movie05 + kid05 + kid10 + office10 +
    transaction_month + year_of_completion_f, data = data_train1)

# test data 에 적용
predict_2 <- predict(linear_best, data_test1[,-25])
summary(predict_2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   328.5   493.7   534.2   538.4   577.3   896.8
```

```
data_test1 %>% select(dong) %>% unique()
```

```
##         dong
## 25339 월계동
## 25402 공릉동
## 25441 하계동
## 25469 상계동
## 25583 중계동
```

```
data_train1 %>% select(dong) %>% unique()
```

```
##         dong
## 25338 월계동
## 25399 공릉동
## 25440 하계동
## 25464 상계동
## 25584 중계동
```

```
# actual, predicted cbind

databind2 <- cbind(data_test1[,25],predict_2)
#databind2 <- cbind(data_test1[,28],predict_2)
databind2 <- as.data.frame(databind2)
summary(databind2)
```

```
##        V1           predict_2
##  Min.   : 242.2   Min.   :328.5
##  1st Qu.: 475.3   1st Qu.:493.7
##  Median : 522.9   Median :534.2
##  Mean   : 536.3   Mean   :538.4
##  3rd Qu.: 582.7   3rd Qu.:577.3
##  Max.   :1094.4   Max.   :896.8
```

```
# RMSE 계산
# install.packages("Metrics")
library(Metrics)
rmse(databind2$V1, databind2$predict_2)
```

```
## [1] 73.26962
```

Random Forest

```
install.packages("randomForest", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'randomForest'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:WUsersWLUISWAppDataWLocalWTempWRtmpEZswNfWdownloaded_packages
```

```
library(randomForest)
```

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

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

```
##
## 다음의 패키지를 부착합니다: 'randomForest'
```

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

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

```
rf.tree1 <- randomForest(unit_price~.-year, data = data_train1,
                         importance = TRUE,
                         ntree = 1000,mtry = 2)

# tree 결과
print(rf.tree1)
```

```
##
## Call:
##  randomForest(formula = unit_price ~ . - year, data = data_train1,      importance = TRUE, n
tree = 1000, mtry = 2)
##                Type of random forest: regression
##                      Number of trees: 1000
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 1661.905
##                     % Var explained: 83.97
```

```
summary(rf.tree1)
```

```
##                Length Class  Mode
## call              6   -none- call
## type              1   -none- character
## predicted      6412   -none- numeric
## mse            1000   -none- numeric
## rsq            1000   -none- numeric
## oob.times      6412   -none- numeric
## importance       52   -none- numeric
## importanceSD     26   -none- numeric
## localImportance   0   -none- NULL
## proximity         0   -none- NULL
## ntree             1   -none- numeric
## mtry              1   -none- numeric
## forest           11   -none- list
## coefs             0   -none- NULL
## y              6412   -none- numeric
## test              0   -none- NULL
## inbag             0   -none- NULL
## terms             3   terms  call
```

```
install.packages("rpart.plot", repos ="http://cran.us.r-project.org")
```

```
## Warning: 패키지 'rpart.plot'가 사용중이므로 설치되지 않을 것입니다
```

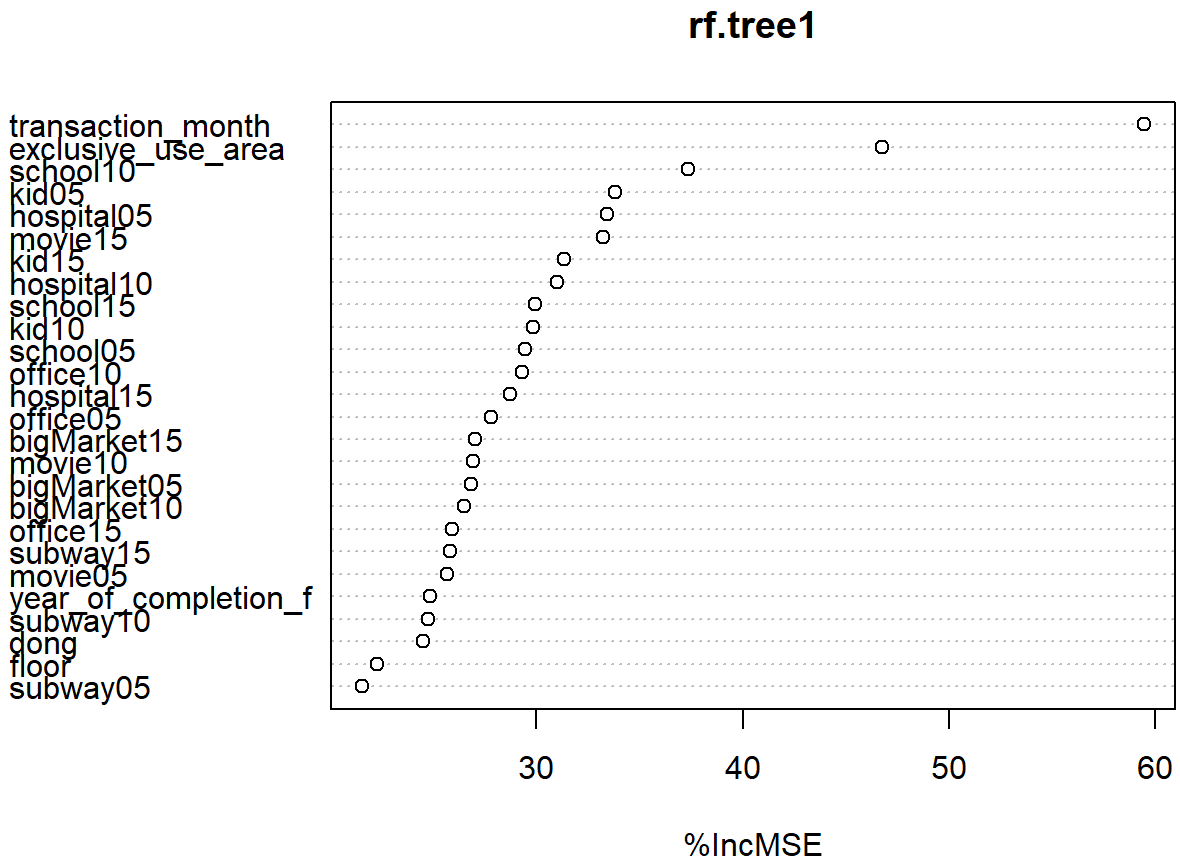
```
library(rpart.plot)

importance(rf.tree1)
```

```
##                       %IncMSE IncNodePurity
## dong                 24.49754      908739.2
## exclusive_use_area   46.76956     6030039.8
## floor                22.28866      778855.0
## bigMarket05          26.83788      952677.8
## bigMarket10          26.48396     1521739.6
## bigMarket15          27.05565     2963691.9
## school05             29.47173     1617299.9
## school10             37.36272     3359813.5
## school15             29.94456     2881900.5
## subway05             21.57542     1362195.4
## subway10             24.74190     1241018.2
## subway15             25.79850     1441278.2
## hospital05           33.42352     2939074.2
## hospital10           31.00597     3132437.2
## hospital15           28.71962     3582309.8
## movie05              25.67522     1615504.6
## movie10              26.91830     1881952.2
## movie15              33.22050     2361349.7
## kid05                33.80069     2071678.8
## kid10                29.86568     1915799.3
## kid15                31.33094     4018828.5
## office05             27.81733     1093384.1
## office10             29.31772     1379995.2
## office15             25.91966     1935098.7
## transaction_month    59.43216     1820304.2
## year_of_completion_f 24.82926      929453.3
```

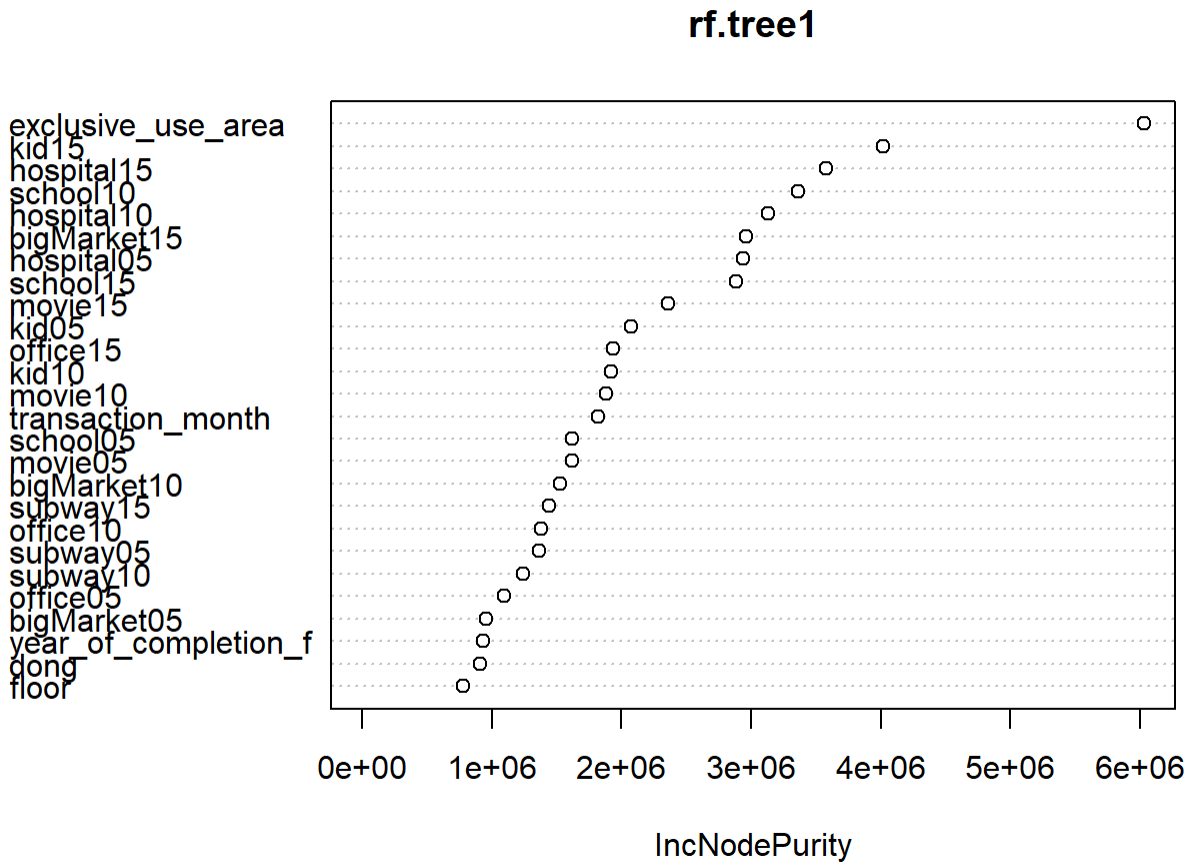
```
importance(rf.tree1, type = 1)
```

```
##                       %IncMSE
## dong                 24.49754
## exclusive_use_area   46.76956
## floor                22.28866
## bigMarket05          26.83788
## bigMarket10          26.48396
## bigMarket15          27.05565
## school05             29.47173
## school10             37.36272
## school15             29.94456
## subway05             21.57542
## subway10             24.74190
## subway15             25.79850
## hospital05           33.42352
## hospital10           31.00597
## hospital15           28.71962
## movie05              25.67522
## movie10              26.91830
## movie15              33.22050
## kid05                33.80069
## kid10                29.86568
## kid15                31.33094
## office05             27.81733
## office10             29.31772
## office15             25.91966
## transaction_month    59.43216
## year_of_completion_f 24.82926
```

```
varImpPlot(rf.tree1, type = 1)
```




```
varImpPlot(rf.tree1, type = 2)
```

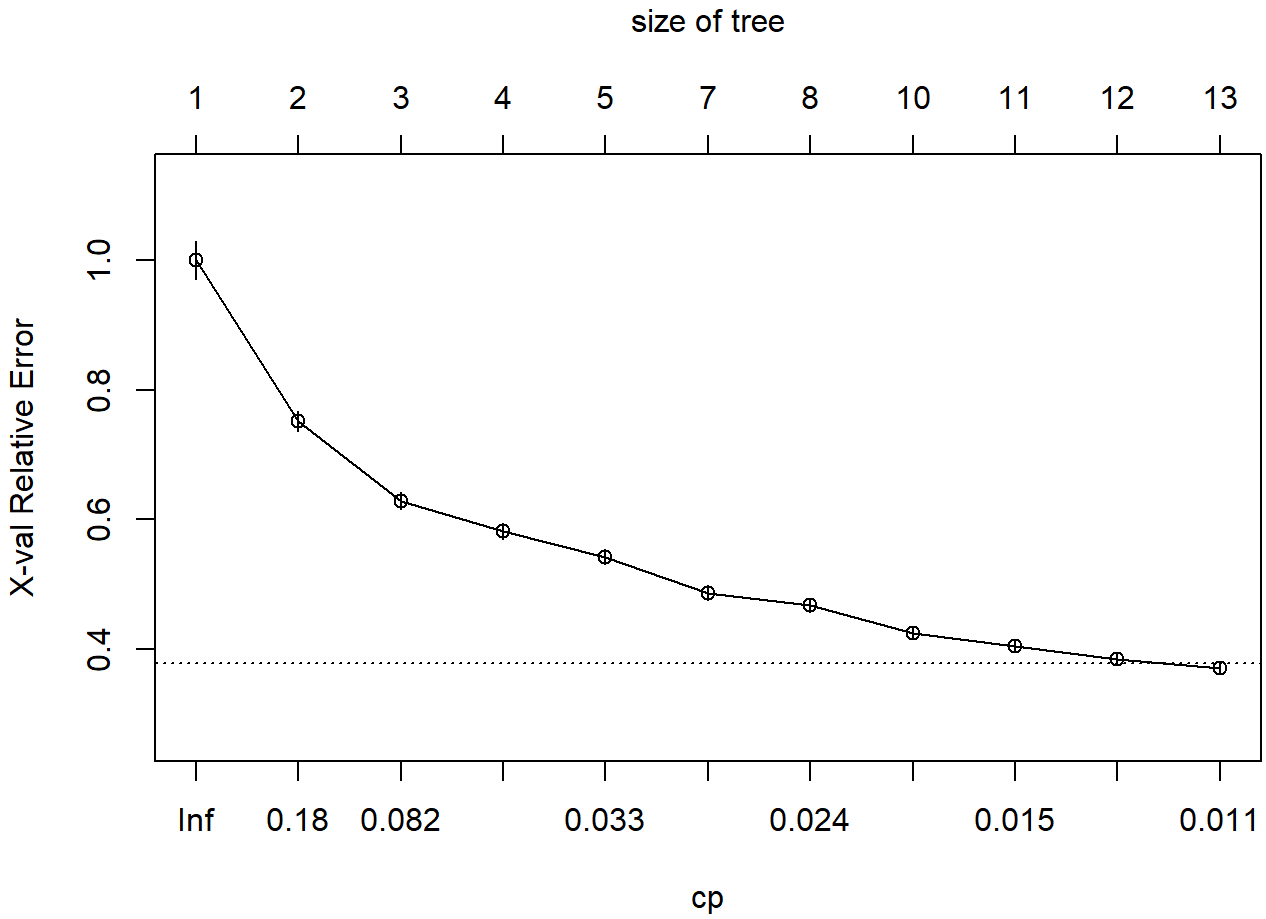


Random Forest parameter tuning

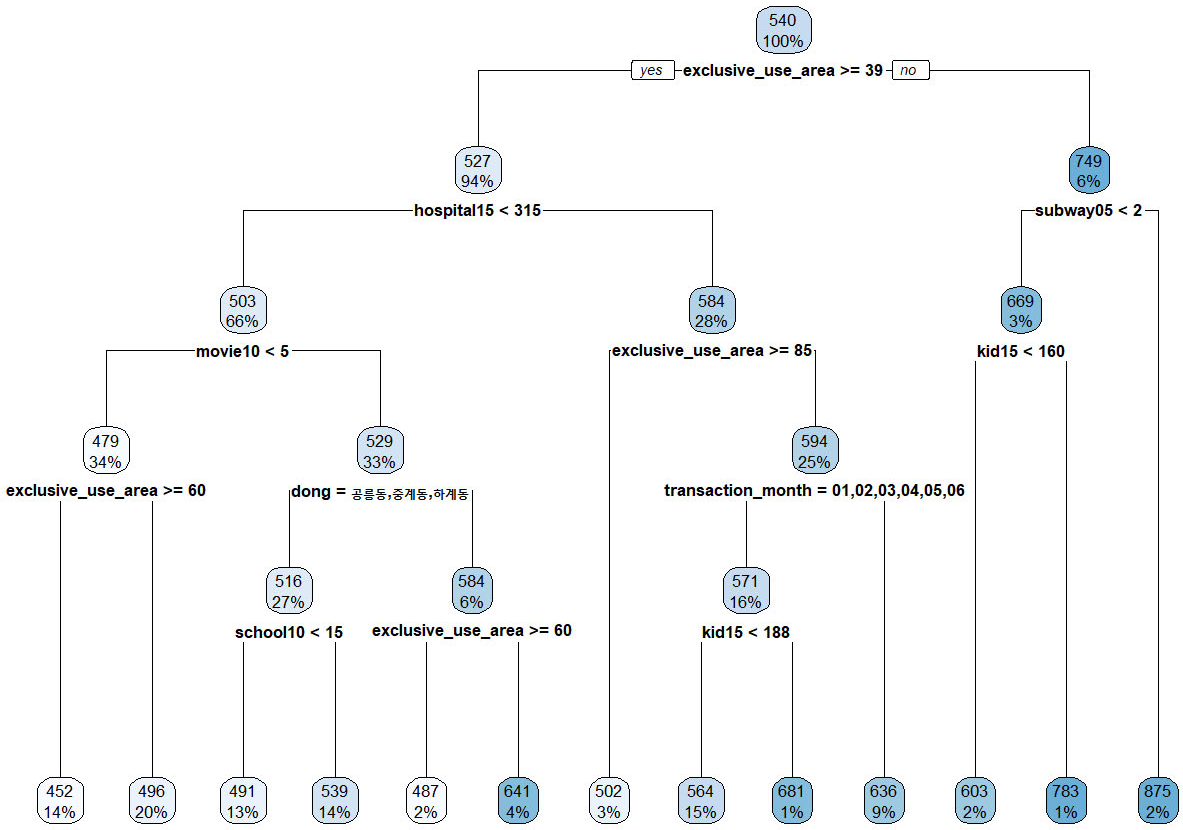
```
printcp(tree1)
```

```
##
## Regression tree:
## rpart(formula = unit_price ~ . - year, data = data_train1, method = "anova",
##     control = rpart.control(minsplit = 50, maxdepth = 5))
##
## Variables actually used in tree construction:
## [1] dong               exclusive_use_area hospital15         kid15
## [5] movie10            school10           subway05           transaction_month
##
## Root node error: 66477236/6412 = 10368
##
## n= 6412
##
##          CP nsplit rel error  xerror      xstd
## 1  0.248787      0   1.00000 1.00023 0.0290602
## 2  0.123384      1   0.75121 0.75223 0.0150142
## 3  0.054060      2   0.62783 0.62891 0.0133603
## 4  0.039992      3   0.57377 0.58232 0.0122489
## 5  0.027694      4   0.53378 0.54247 0.0118086
## 6  0.025108      6   0.47839 0.48726 0.0112679
## 7  0.022506      7   0.45328 0.46818 0.0109434
## 8  0.015059      9   0.40827 0.42487 0.0099900
## 9  0.015055     10   0.39321 0.40479 0.0089851
## 10 0.012363     11   0.37816 0.38520 0.0087919
## 11 0.010000     12   0.36579 0.37080 0.0082922
```

```
plotcp(tree1)
```




```
tree1 <- prune(tree1, cp= tree1$cptable[which.min(tree1$cptable[, "xerror"]), "CP"])

rpart.plot(tree1)
```



Random Forest prediction & RMSE calculation

```
# test data 에 적용

predict_3 <- predict(rf.tree1, data_test1)
summary(predict_3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   293.7   485.6   530.3   538.6   575.1   936.9
```

```
# actual, predicted cbind

databind3 <- cbind(data_test1[,25],predict_3)
databind3 <- as.data.frame(databind3)
summary(databind3)
```

```
##        V1          predict_3
##  Min.   : 242.2   Min.   :293.7
##  1st Qu.: 475.3   1st Qu.:485.6
##  Median : 522.9   Median :530.3
##  Mean   : 536.3   Mean   :538.6
##  3rd Qu.: 582.7   3rd Qu.:575.1
##  Max.   :1094.4   Max.   :936.9
```

```
# RMSE 계산
install.packages("Metrics", repos = "http://cran.us.r-project.org")
```

```
## Warning: 패키지 'Metrics'가 사용중이므로 설치되지 않을 것입니다
```

```
library(Metrics)
rmse(databind3$V1, databind3$predict_3)
```

```
## [1] 39.82788
```

Gradient Boost Model

```
install.packages("gbm", repos ="http://cran.us.r-project.org")
```

```
## 패키지 'gbm'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
##
## 다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
##  C:₩Users₩LUISW₩AppData₩Local₩Temp₩RtmpEZswNf₩downloaded_packages
```

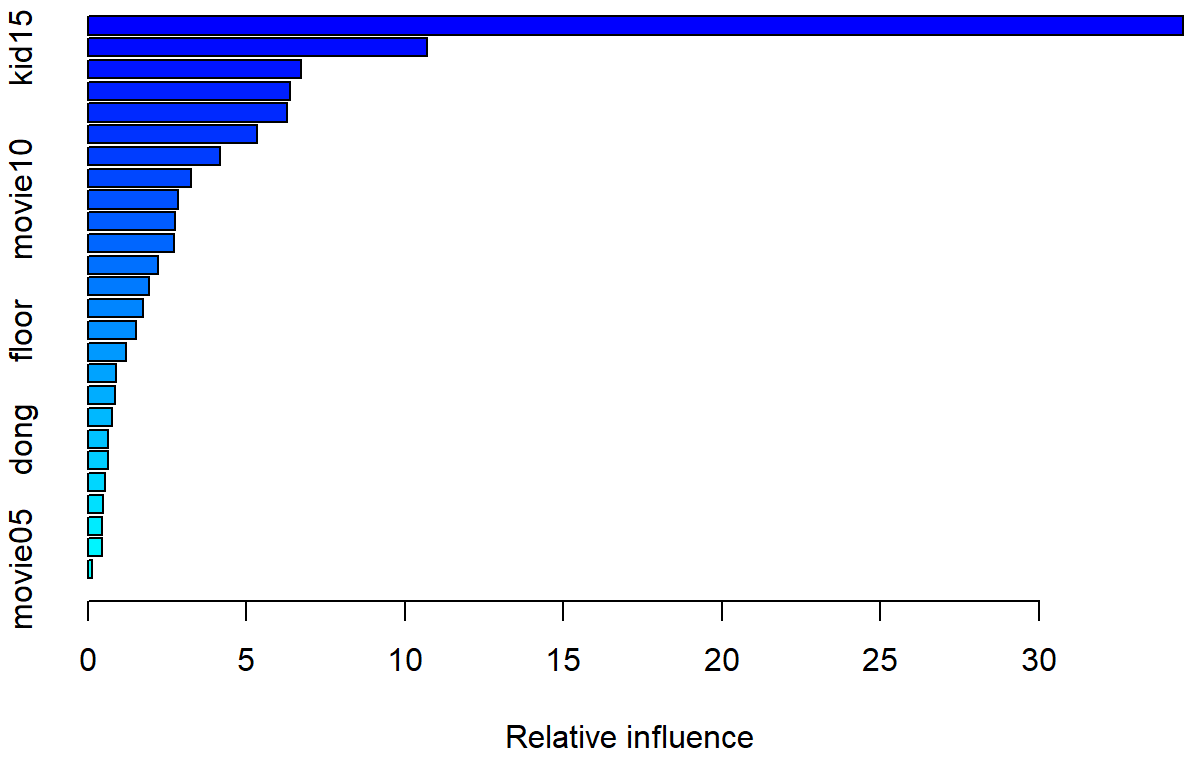
```
library(gbm)
```

```
## Loaded gbm 2.1.8.1
```

```
gbm.tree1 <- gbm(unit_price~.-year, data = data_train1, distribution = "gaussian",
                 n.trees = 1000, shrinkage = 0.01, interaction.depth = 4)

# tree 결과
print(gbm.tree1)
```

```
## gbm(formula = unit_price ~ . - year, distribution = "gaussian",
##     data = data_train1, n.trees = 1000, interaction.depth = 4,
##     shrinkage = 0.01)
## A gradient boosted model with gaussian loss function.
## 1000 iterations were performed.
## There were 26 predictors of which 26 had non-zero influence.
```

```
summary(gbm.tree1)
```




```
##                                    var    rel.inf
## exclusive_use_area     exclusive_use_area 34.5482521
## kid15                             kid15 10.7052232
## transaction_month       transaction_month  6.7197065
## hospital15                   hospital15  6.3859379
## school10                     school10  6.2678641
## subway05                     subway05  5.3212966
## bigMarket15                 bigMarket15  4.1623442
## hospital10                   hospital10  3.2557460
## movie10                       movie10  2.8344654
## hospital05                   hospital05  2.7558402
## year_of_completion_f year_of_completion_f  2.7031452
## school15                     school15  2.2236193
## kid05                           kid05  1.9226744
## kid10                           kid10  1.7323478
## floor                           floor  1.5140958
## movie15                       movie15  1.2075796
## subway15                     subway15  0.8793640
## office10                     office10  0.8533598
## school05                     school05  0.7460993
## dong                             dong  0.6210443
## office15                     office15  0.6190419
## subway10                     subway10  0.5435329
## bigMarket10                 bigMarket10  0.4798855
## bigMarket05                 bigMarket05  0.4423694
## office05                     office05  0.4308640
## movie05                       movie05  0.1243005
```

Gradient Boost Model parameter tuning

```
# printcp(tree1)
# plotcp(tree1)
# tree1 <- prune(tree1, cp= tree1$cptable[which.min(tree1$cptable[, "xerror"]), "CP"])
#
# rpart.plot(tree1)
```

##Gradient Boost Model prediction & RMSE calculation

```
# test data 에 적용
predict_4 <- predict.gbm(object = gbm.tree1,
                             newdata = data_test1,
                             n.trees = 1000,
                             type = "response")
summary(predict_4)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   348.9   484.6   524.2   537.5   572.4   971.3
```

```
# actual, predicted cbind

databind4 <- cbind(data_test1[,25],predict_4)
databind4 <- as.data.frame(databind4)
summary(databind4)
```

```
##        V1           predict_4
##  Min.   : 242.2   Min.   :348.9
##  1st Qu.: 475.3   1st Qu.:484.6
##  Median : 522.9   Median :524.2
##  Mean   : 536.3   Mean   :537.5
##  3rd Qu.: 582.7   3rd Qu.:572.4
##  Max.   :1094.4   Max.   :971.3
```

```
# RMSE 계산
install.packages("Metrics", repos ="http://cran.us.r-project.org")
```

```
## Warning: 패키지 'Metrics'가 사용중이므로 설치되지 않을 것입니다
```

```
library(Metrics)
rmse(databind4$V1, databind4$predict_4)
```

```
## [1] 38.38795
```