

# 대전광역시 부동산 동향

---

집값과 차세대 투자지역



수업: 데이터 시각화

교수님: 박영호 교수님

학번: 20182890

이름: 임성민

## [목차]

1. 발표 주제 및 선정 이유 -----	2
2. 데이터 수집 과정 -----	2
3. 데이터 변수 속성 및 전처리 과정 -----	3
4. 자료 분석 도구와 분석 과정 -----	5
5. 분석 결과 -----	6
6. 결론 -----	9

## 1. 발표 주제 및 선정 이유

### ① 발표 주제

: 대전광역시 부동산 동향 (집값과 차세대 투자지역을 중심으로)

### ② 주제 선정 이유

: 타 교양 수업과제로 지방자치에 대한 조사 보고서를 작성해야 했다. 대전의 부동산 정책에 대한 비평을 짧게 하면서 근거를 찾기 위해 자료를 찾아보니 전혀 생각지도 못한 가격에 거래되는 아파트가 꽤 많은 걸 보았다. 또한 대전광역시 집값이 세종시에 영향을 꽤나 받지 않나 생각했는데, 실제로 자료를 대조해보니 아닌 것 같다는 생각이 들었다. 집값이 오르는 시기가 세종이랑 전혀 관계가 없을 때가 더 많았고 만약이게 사실이라면 어떤 요인이 대전시 집값을 좌지우지하는지 알고 싶었다. 정치 데이터나 텍스트 데이터 마이닝 다루기에는 한계가 있어 내가 가능한 수준을 인지하고 분석에 임했다.

이번 분석에서는 대전 광역시 아파트 값이 지리적 요인, 건설회사의 네임벨류, 투자 방향 등 어떤 움직임에서 가장 영향을 많이 받는지 알아본다. 평균 거래가가 가장 높은 스마트 시티를 기준으로 차세대 스마트 시티가 어디인지 알아보고, 다음 투자 방향성도 제시해본다.

## 2. 데이터 내용 및 출처

데이터 이름	데이터 출처	데이터 내용
아파트 실거래,	공공데이터 포털 대전광역시_주택  <a href="https://data.go.kr/tcs/dss/selectFileDataDetailView.do?publicDataPk=15067893">https://data.go.kr/tcs/dss/selectFileDataDetailView.do?publicDataPk=15067893</a>	
연립 다세대 실거래	공공데이터 포털 대전광역시_주택  <a href="https://data.go.kr/tcs/dss/selectFileDataDetailView.do?publicDataPk=15067893">https://data.go.kr/tcs/dss/selectFileDataDetailView.do?publicDataPk=15067893</a>	

### 3. 데이터 변수 속성 및 전처리 과정

#### ① 변수의 속성

##### ■ 아파트 실거래

열: 기준년월, 기준일, 법정동, 지번, 아파트명, 층, 전용면적, 거래금액. 일부보기

행: 64037개

```
> apt %>% head()
  기준년월 기준일 법정동 지번 아파트명 층 전용면적 거래금액. 일부보기 x
1  201501  1~10      인동 179-1  현대  2    70.96      12600 NA
2  201501  1~10      인동 179-1  현대  9    70.96      15000 NA
3  201501  1~10      인동 179-1  현대  9    70.96      15000 NA
4  201501  1~10      인동 179-1  현대  9    70.96      15000 NA
5  201501  1~10      인동 179-1  현대  9    70.96      15000 NA
6  201501  1~10      인동 179-1  현대  9    70.96      15000 NA

> apt %>% summary()
  기준년월 기준일 법정동 지번 아파트명
Min.   :201401 Length:64037
1st Qu.:201410 Class :character
Median :201507 Mode  :character
Mean   :201576
3rd Qu.:201801
Max.   :202009

  층 전용면적 거래금액. 일부보기 x
Min.   : 1.000 Min.   : 12.01 Min.   : 2000
1st Qu.: 4.000 1st Qu.: 59.65 1st Qu.: 13000
Median : 8.000 Median : 84.36 Median : 19500
Mean   : 8.817 Mean   : 76.78 Mean   : 21046
3rd Qu.:12.000 3rd Qu.: 84.97 3rd Qu.: 26500
Max.   :50.000 Max.   :285.00 Max.   :250000
NA's   : 4      NA's   : 4      NA's   : 4
Mode:logical
NA's:64037
```

##### ■ 연립 다세대 실거래

열: 기준년월, 기준일, 법정동, 지번, 연립주택명, 층, 건축년도, 전용면적, 거래금액. 일부보기

행: 8159개

```
> v111 %>% head()
  기준년월 기준일 법정동 지번 연립주택명 층 건축년도 전용면적 거래금액. 일부보기 x
1  201501 21~31      인동 179-1  현대  2    70.96      12600 NA
2  201501 21~31      인동 179-1  현대  9    70.96      15000 NA
3  201501  1~10      인동 179-1  현대  9    70.96      15000 NA
4  201501 11~20      인동 179-1  현대  9    70.96      15000 NA
5  201501  1~10      인동 179-1  현대  9    70.96      15000 NA
6  201501 21~31      인동 179-1  현대  9    70.96      15000 NA

> v111 %>% summary()
  기준년월 기준일 법정동 지번 연립주택명
Min.   :201401 Length:8159
1st Qu.:201410 Class :character
Median :201507 Mode  :character
Mean   :201581
3rd Qu.:201802
Max.   :202009

  층 건축년도 전용면적 거래금액. 일부보기 x
Min.   : 1.000 Min.   : 12.01 Min.   : 2000
1st Qu.: 4.000 1st Qu.: 59.65 1st Qu.: 13000
Median : 8.000 Median : 84.36 Median : 19500
Mean   : 8.817 Mean   : 76.78 Mean   : 21046
3rd Qu.:12.000 3rd Qu.: 84.97 3rd Qu.: 26500
Max.   :50.000 Max.   :285.00 Max.   :250000
NA's   : 4      NA's   : 4      NA's   : 4
Mode:logical
NA's:8159

> dim(v111)
[1] 8159 10
```

## ② 전처리 과정

## -워드 클라우드

[illegible]

아파트와 빌라 데이터에 포함된 법정동을 기준으로 '구' 정보가 들어간 열을 생성한다.

```
# 결측값 제거
apt1 <- subset(apt1920, !is.na(법정등))
apt2 <- subset(apt1, !is.na(아파트명))
apt3 <- subset(apt2, !is.na(층))
apt4 <- subset(apt3, !is.na(거래금액.일부보기))
apt5 <- subset(apt4, !is.na(시군구))
apt5$층 <- as.numeric(apt5$층)
apt19 <- subset(apt, 기준년월 >= 201901)
vilr19 <- subset(vilr, 기준년월 >= 201901)
```

아파트와 빌라 데이터의 결측값들을 제거하고 2019년 1월 이후의 데이터만 추출해낸다.

### -계층적 군집분석 (덴드로그램)

```

#아파트 거래건수합산 상위 30위 추출
apt1 <- table(apt$아파트명)
apt12 <- sort(apt1, decreasing = T)
dim(apt12)
apt123 <- apt12[c(1:30)]
apt123 <- data.frame(apt123)
apt123 <- rename(apt123, '아파트명' = 'var1')
apt123 <- rename(apt123, '거래건수' = 'Freq')
apt123 %>% head()

# 각 아파트의 거래 전용면적평균, 거래금액평균, 중평균 구하기
apt123$전용면적평균 <- 0
apt123$거래금액 <- 0
apt123$중평균 <- 0
apt123 %>% head()
a <- apt[apt$아파트명 == '현대', ]
apt123[apt123$아파트명 == '현대', '전용면적평균'] = mean(a$전용면적)
apt123[apt123$아파트명 == '현대', '거래금액평균'] = mean(a$거래금액, na.rm = TRUE)
apt123[apt123$아파트명 == '현대', '중평균'] = mean(a$중)

#아파트 거래량 상위 30개
apt123

# 계층적 군집 만들기
apt1234 <- apt123

#행번호를 이름으로 변경
rownames(apt1234) <- paste0(apt1234$아파트명, "_", rownames(apt1234))

# 아파트명 행 삭제
apt1234 <- apt1234[,-c(1)]

```

```

> apt1234
현대_1      82.85044      15604.755      8.013536
주공_2      50.90492      9169.077      7.435631
역삼_3      107.73674      25472.898      9.323077
금강역출투타워_4 113.21688      30038.123      23.947447
주공_5      49.28410      10311.538      2.966258
오류그랜드_6   97.17482      28617.548      12.178187
강성큰마름_7    70.84692      20263.164      8.154905
유성_8       77.72335      15394.682      6.955166
송파리_9      84.95000      27689.556      10.107071
영양마을_10    51.50887      12013.438      11.955975
주공_11      51.47754      10759.358      5.252677
버드내_12     85.39919      22730.719      10.871460
주공_13      44.84930      9492.885      9.909692
맑은아침_14    84.62386      26158.184      11.632287
현대스몰트_15  84.98691      30295.891      9.456057
화성내_16     81.68840      34276.843      9.950860
삼성(175)_17   85.34264      24028.928      7.177057
삼록수_18     57.83343      15782.475      7.845960
송강그린_19    70.50208      16788.602      7.902375
해지_20       75.62722      21333.069      8.201058
모내빌_21     66.70909      13765.070      7.252688
수정타운_22    78.51532      20644.274      7.844086
영양마을_23    54.60288      12333.476      12.130435
주공_24      45.75981      7693.152      8.377717
계룡_25       73.89978      12285.451      6.890710
트리플시티_26  121.36425      59492.740      12.443836
크로바_27     118.07690      53377.465      7.845070
목양마을_28    99.71493      26452.422      10.475783
한가람_29     52.89206      16924.029      7.744928
구룡마을주공8-1_30 55.49904      13374.478      12.450746

```

같은 아파트명 중복행을 이용하여 데이터 프레임을 생성한다.

apt123 데이터 중 아파트명이 동일한 열에 해당하는 아파트 명들이 가진 값의 평균을 계산하여 넣는다.

이를 반복 수행하여 apt123이라는 데이터 프레임을 완성한다.

계층적 군집분석을 위해 행번호에 아파트명을 부여하고 아파트명 열을 삭제한다.

## 4. 자료 분석 도구와 분석 과정

### ① 워드클라우드

: 이번 분석에서는 워드클라우드를 활용하여 데이터의 핵심 단어를 그대로 드러내는 직관적인 분석기법을 사용한다. 워드클라우드는 방대한양의 데이터에서 특징을 도출해내기 위해서 사용한다. 이번 분석에서는 아파트의 이름들을 모두 하나의 범주형 변수로 보기 때문에 워드클라우드를 사용한다.

#### # 워드클라우드 핵심코드

```
# 수업에서 다루지 않은 워드 클라우드로 진행
# 패키지 설치
library(RColorBrewer)
library(wordcloud)
library(ggplot2)

# 거래 빈도가 50 이하인 아파트는 제외한다 최대 거래량은 한계를 두지 않는다.
# 단어의 위치와 색은 빈도순이다.
#
palette <- brewer.pal(8, "Dark2") # 글자색 지정
wordcloud(apt5$아파트명, apt5$거래금액, 일부보기, scale=c(4,0.5),min.freq=50,max.words=Inf,
          rot.per=0.25,random.order=F, random.color=F, colors=palette)

# 거래 금액을 내림차순으로 정리
apt5_1_0 <- apt5[order(apt5$거래금액, 일부보기),]
rownames(apt5_1_0) <- NULL
apt5_1_0 %>% head()
apt5_1_0 %>% tail()
```

\*분석 과정은 코드의 주석에 적어 생략합니다.

### ② 계층적 군집분석

: 가까운 대상끼리 순차적으로 군집을 묶어 나가는 군집 분석 방법이다. 거리를 기반으로 데이터를 축소하며 기존 관측치들 간의 모든 거리를 계산한다. 아파트들이 어떠한 유사성을 가지고 있는지 알아보고자 사용하였다. 전처리를 거친 데이터는 모두 연속형으로 병합적 방법과 응집형 방법을 사용하였다. 거리와 유사성은 다른 개념임을 명심한다.

거리 계산에 사용한 계산법(함수): 유클리드 거리, 맨하탄 거리, Ward 연결법,

#### # 계층적 군집분석 코드

```
#거리 계산 ( 유클리디안 방법 )
dist_apt1234 <- dist(scale(apt1234), method = "euclidean")

#군집분석 ( ward.D : 비슷한 크기의 군집끼리 묶기 )
apt1234_hclust <- hclust(dist_apt1234, method = "ward.D")

#계층적 군집분석
summary(apt1234_hclust)

#군집 : 덴드로그램
plot(apt1234_hclust, hang = -1)
#비유사도(Height) 6 정도에서 각기 군집 형성

#그룹 나누기
groups <- cutree(apt1234_hclust, k=4)
groups

#그림 그리기
hca <- hclust(dist(apt1234))
plot(hca)
rect.hclust(hca, k = 4, border = "red")
rect.hclust(hca, h = 5000, which = c(4,7), border = 3:4)
# 높이(h) 5000에서 cut 수행, 4,7번에 사각형 추가, 테두리 색상(border) 지정

### 아그네스
library(cluster)
(agn1 <- agnes(apt1234, metric="manhattan", stand=TRUE))
plot(agn1)

agn2 <- agnes(daisy(apt1234), diss=TRUE, method="complete")
plot(agn2)

agn3 <- agnes(apt1234, method = "flexible", par.meth=0.6)
plot(agn3)
```

\*분석 과정은 코드의 주석에 적어 생략합니다.

## 5. 분석 결과

### ① 워드 클라우드 - 아파트

2015년부터 2020년 9월 사이에 거래된 아파트의 거래 금액을 사용한 워드 클라우드는 워드 클라우드에서 제외하였으며 단어의 위치와 색은 빈도순으로 나열되어 있다. 거래 금액만을 볼 때 스마트 시티가 압도적 1위지만 주변 거래 빈도를 보면 한라 비발디와 트리폴시트를 찾는 사람이 더 많다는 것을 알 수 있다.

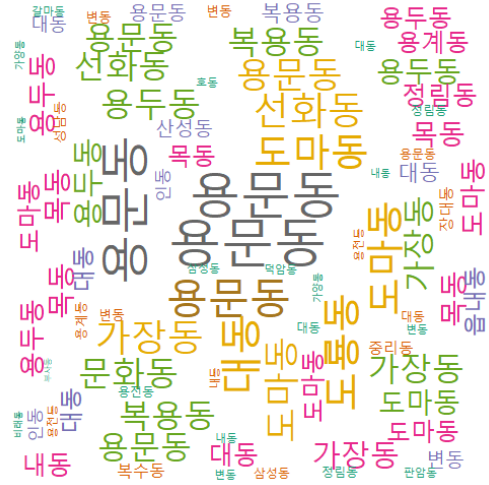


2015년	2016년	2017년
2018년	2019년	2020년

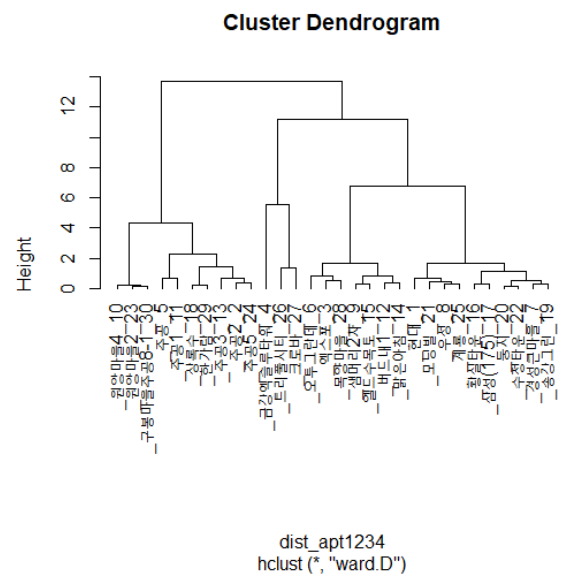
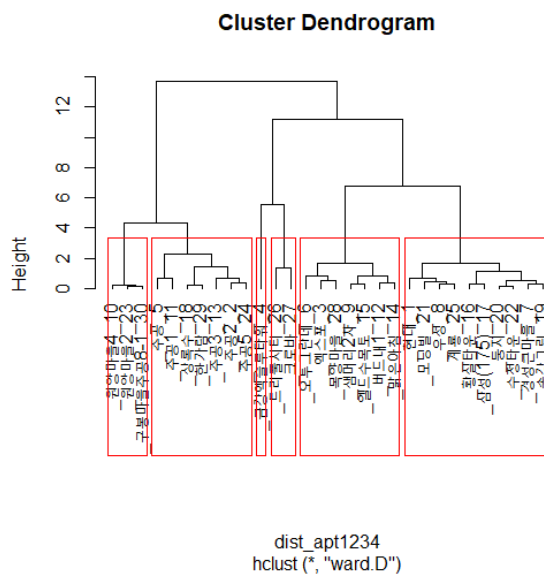
위 표는 2015년부터 2020년까지 같은 조건으로 만들어진 워드 클라우드 자료이다. 2015년에는 크게 두드러지지 않던 트리폴시티와 한라비발디 단지가 2019년 도안 신도시 갑천 3블록 전매 제한이 풀리고 매매 건수와 집값이 크게 상승하며 스마트 시티의 거래량을 앞질렀다. 단일 거래가격으로는 아직 스마트 시티를 앞지르기 힘들지만, 투기 과열지구 선정 이후에도 10억을 가볍게 웃도는 도안 지구의 집값을 생각할 때 차기 스마트 시티는 트리폴시티와 한라 비발디라고 예측할 수 있다. 또한 서구 도안 지구가 위치한 상대동 일대의 부지가격이 상승할 것이라고 예측가능 했다.

## ② 워드클라우드 - 빌라

아파트에 대한 분석을 진행하면서 빌라를 투기에도 사용하는지 궁금증이 생겼다. 빌라를 구매하더라도 투기가 목적이려면 값나가는 부지에 빌라를 구매해야 한다. 이 가설이 사실인지 알아보려고 빌라매매 장소로서 가장 인기인 지역을 워드클라우드 만들어 보았다. 결과는 아파트 매매지역과 정반대인 중구와 서구 구도심 지역으로 신도심에 비해 상대적으로 낙후된 곳이다. 때문에 빌라는 매매일지라도 투기목적보다는 실거주가 목적이라고 짐작할 수 있었다.



## ③ 계층적 군집 분석



1차 덴드로그램에 사용한 응집형 방법의 거리 계산법은 유클리드 거리 계산법이며, 군집 방법은 ward방법이다. 총 6개의 군집으로 분류되었다. 유사성이 아주 높은 군집이 존재하는데 3, 4, 5, 6번째 군집으로 보통 사람들이 '좋은 아파트' 라고 생각하는 아파트들이다. 1, 2번째 군집에는 주공 아파트단지가 모두 포함되었으며, 보통 '나쁘지 않은 아파트' 하고 생각하는 아파트들이었다.

```
> rev(apt1234_hclust)
$dist.method
[1] "euclidean"

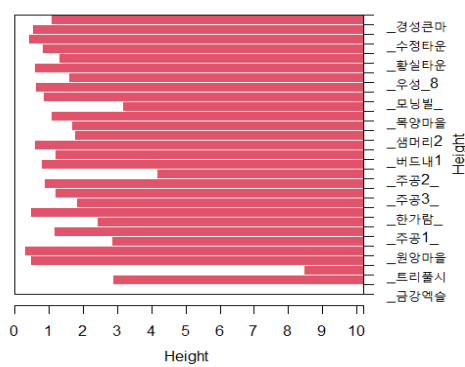
$call
hclust(d = dist_apt1234, method = "ward.D")

$method
[1] "ward.D"

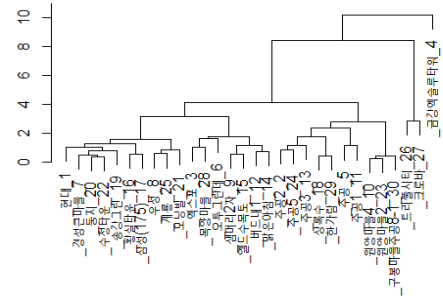
$labels
[1] "_현대_1"          "_주공2_2"
```



Banner of `agnes(x = apt1234, metric = "manhattan", stan`

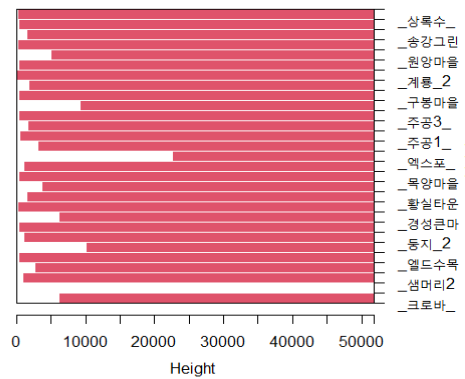


Agglomerative Coefficient = 0.88

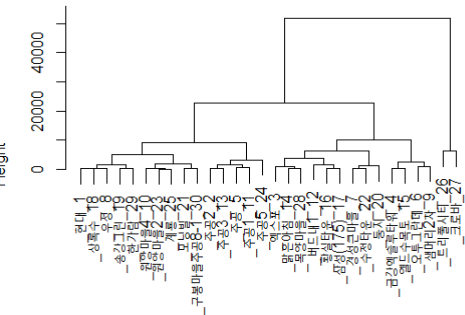


apt1234  
Agglomerative Coefficient = 0.88

Banner of `agnes(x = daisy(apt1234), diss = TRUE,gram of agnes(x = daisy(apt1234), diss = TRUE, method =`

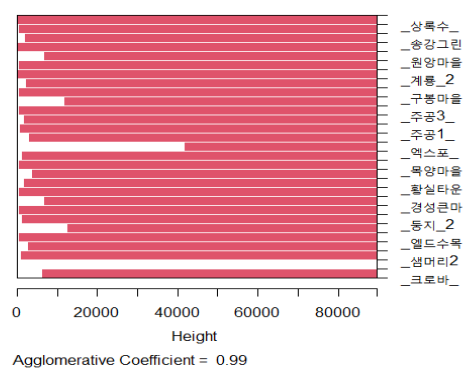


Agglomerative Coefficient = 0.98

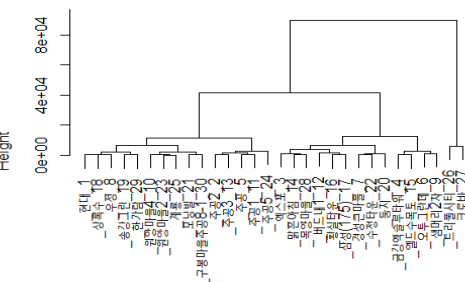


daisy(apt1234)  
Agglomerative Coefficient = 0.98

Banner of `agnes(x = apt1234, method = "flexibleogram of agnes(x = apt1234, method = "flexible", par.me`



Agglomerative Coefficient = 0.99



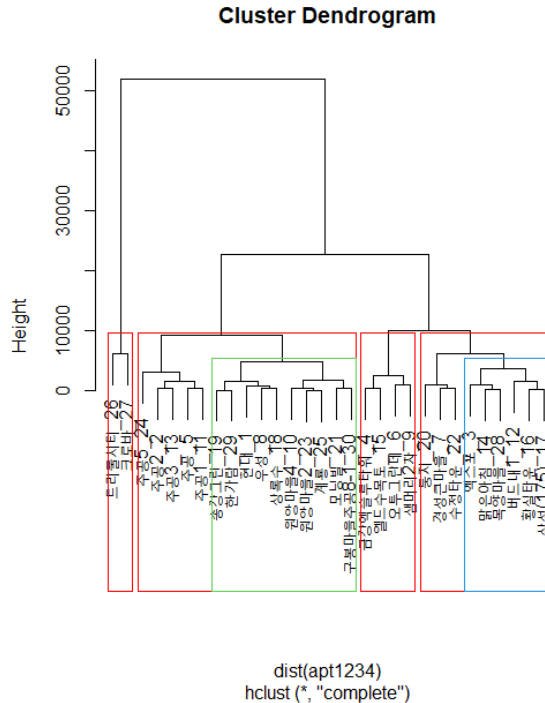
apt1234  
Agglomerative Coefficient = 0.99

2차 덴드로그램은 `agnes` 함수를 활용하였다. 이는 병합적 방법 중 하나로 작은 군집에서 출발하여 거리가 가까운 순으로 병합해 나가는 방법이다. 3가지 방법을 활용하였으며 3가지 모두 각기 결과가 다르다. 하지만 아주 작은 군집을 제외하고 어느정도 크기가 있는 군집을 구성한 아파트명은 4가지 모두 거의 비슷하다.

1차와 응집형 방법과 2차 병합적 방법에서 거의 비슷한 결과를 볼 수 있었다.

## 6. 결론

```
hca <- hclust(dist(apt1234))
plot(hca)
rect.hclust(hca, k = 4, border = "red")
rect.hclust(hca, h = 5000, which = c(4,7), border = 3:4)
# 높이(h) 5000에서 cut 수행, 4,7번에 시각형 추가, 테두리 색상(border) 지정
```



가장 기본적인 방법으로 만든 덴드로그램

: 대전에서 부동산 투자를 하고자 한다면 도안 신도시내의 아파트에 투자해야 한다. 또한 트리폴시티와 한라비발디는 앞으로도 계속 상승세일 것이다. 빌라를 신축하거나 빌라 자체로 돈을 벌고자 한다면 서구 인근 구도심과 중구 한가운데의 빌라를 통째로 매매할 것을 추천한다.

응집형 방식의 덴드로그램을 볼 때 주공아파트와 원앙마을은 사람들에게 비슷한 수준으로 인식되어 거래가 이루어진다. 그러나 병합적 방법과 따로 방법을 설정하지 않고 만든 위의 덴드로그램에서는 다른 군집으로 묶여 나타났다. 전용면적, 거래금액, 층 수의 평균은 사람들의 선호도가 많이 반영되어 있다고 할 수 있다.

거래량 상위 30개에 주공 아파트단지만 6개로 대전에서는 주공아파트의 거래가 빈번하다. 다른 아파트들에 비해 상대적으로 저렴한 임대아파트라는 인식을 생각할 때 대전의 부동산 구매 방향이 투자보다도 실 거주 매매가 많다고 생각 할 수 있었다.

단기 투자를 위해서라면 매매율이 높은 현대, 주공, 엑스포아파트를 매매하는 것을 장기 투자를 위해서라면 매매율과 아파트 거래금액이 모두 높은 트리폴시티를 구매하는 것을 추천한다.