

서울특별시 행복주택 최적입지 선정

1. 배경

□ 주제 선정

‘행복주택’이란 임대주택의 일종으로, 청년계층에게 저렴하면서도 양질의 주거 환경을 제공하자는 취지로 시행된 정부의 보편적 주거복지 정책의 일환이다. 주거복지정책의 사각지대에 있던 청년계층(사회초년생, 신혼부부, 대학생 등)에게 공급되는 공공 임대주택으로 각광받던 행복주택은 기대와는 다르게 현실적인 문제에 봉착해 있다.

□ 분석 필요성(문제점) 및 전략

먼저, 행복주택의 임대료와 관리비는 저렴하지 않다. 행복 주택의 보증금은 보통 4000~8000만 원이고 임대료는 주변 시세의 60~80% 수준으로 비교적 저렴한 편이지만, 관리비는 월세에 비해 터무니없이 비싸다. 즉, 사회 초년생들에게는 여전히 부담스러운 가격이다. 이 때문에 100:1의 치열한 경쟁률을 뚫고 당첨이 됐음에도 청년들의 계약률은 70%대에 그친다. 또한, 많은 행복주택들이 경제성만을 우선시한 나머지 접근성이 좋지 않은 도시 외곽에 건설된 것 역시 청년들이 행복주택을 주저하게 되는 이유이다. 서로 상충되는 가치인 경제성과 접근성을 적절히 고려하지 못하여 상대적으로 학교 및 사업체와 떨어져 있는 도심 외곽에 행복주택을 짓는 것은 수요를 제대로 반영하지 못하는 것이다.

따라서 본 조는 PCA 분석, 클러스터링(군집화), 다목적 최적화 기법을 통해 접근성과 경제성 그리고 주변 인프라를 최대한 만족하는 서울특별시 행복주택의 최적 입지를 분석하여 총 3개의 후보지를 선정하고자 한다.

□ 선행연구

기존 선행연구에서는 대중교통 접근성이 우수하면서 저렴한 가격으로 공급할 수 있는 행복주택의 최적 입지를 제안하고 있다.(진찬우, 이건축, 2015) 따라서 다목적 공간 최적화 모델에서 접근성과 경제적 비용을 고려하였다. 하지만 본 분석에서는 경제성과 접근성에 지역 인프라 상황까지 고려를 한 다목적 최적화 모델을 만들어 차별성을 두고자 한다. 또한 선행연구 당시 2015년의 상황과 현 2022년의 상황은 다를 것이라 판단하였다. 따라서 최근 자료를 바탕으로 다양한 분석 툴(R, python)을 사용하여 분석을 진행한다.

2. 데이터 분석

□ 데이터 선정

총 5가지 영역에 대하여 데이터를 수집하였으며, 각 데이터의 사용 용도는 다음과 같다.

영역	데이터명	출처	사용 용도
경제	① 서울시 자치구별 주택매매가격 (txt)	서울 열린데이터 광장	서울시 자치구별 주택매매가격 파악
	② 서울시 자치구별 공시지가 분포 (csv)	서울 열린데이터 광장	서울시 자치구별 공시지가 분포 파악
교통	③ 서울시 자치구별 통근통학 인구 비율 (txt)	서울 열린데이터 광장	높은 통근통학 인구 비율을 통한 서울시 행복주택의 필요성 강조
	④ 서울시 자치구별 지하철역 총개수 (csv)	서울 열린데이터 광장	서울시 자치구별 지하철역 총개수 파악
	⑤ 서울시 지하철 노선별 평일 운행회수 및 수송인원 (txt)	서울 열린데이터 광장	서울시 지하철 노선별 평일 운행회수 및 수송인원 파악
인프라	⑥ 서울시 자치구별 경찰서, 지구대, 소방서, 119안전센터 개수 (txt)	서울 열린데이터 광장	서울시 자치구별 경찰서, 지구대, 소방서, 119안전센터 개수 파악
	⑦ 서울시 자치구별 대형마트, 백화점 등 개수 (xlsx)	kosis 국가통계포털	서울시 자치구별 대형마트, 백화점 등 개수 파악
	⑧ 서울시 자치구별 문화공간 개수 (xlsx)	서울 열린데이터 광장	서울시 자치구별 문화공간 개수 분포 파악
수요	① 서울시 대학교별 위치, 재학생 수 및 기숙사 수용률 (xlsx)	서울 열린데이터 광장	서울시 내 대학교별 위치, 재학생 수 및 기숙사 수용률 파악
	② 서울시 사업체수 데이터	MDIS(통계데이터센터 공공용 마이크로데이터)	서울시 내 자치구별 사업체수 및 합계종사자 수 파악
기타	③ 서울시 행복주택 입주 후보지 (csv)	서울 열린데이터 광장	서울시 자치구별 기존 행복주택 개수 파악
	④ 서울시 재정비촉진구역 목록	서울시 누리집 공식 사이트	서울시 행복주택 입지 후보지 파악
	⑤ 행복주택 공급계획(csv)	서울주택도시공사	서울 특별시에 공급될 행복주택의 면적 및 세대수 파악

• 모든 데이터는 각각의 기관의 데이터베이스 활용 지침을 준수하여 추출하였으며, 검증된 공식 통계 사이트에서 최신의 데이터를 사용하였다.

□ 데이터 분석

1. 전처리

전처리는 기본 작업을 우선 실행한 뒤 아래의 과정을 순서대로 진행하여 다음 단계인 클러스터링에 용이하도록 하였다. 여기서 기본 작업이라 함은 데이터에서 필요한 변수만을 추출, 서울시 데이터만을 추출, 변수명 변경(가독성), 중복행 제거, 변수 형태 변환을 포함한다.

1) 먼저 추가적인 데이터 조사를 통해 결측치를 제거하거나 보간하고, Boxplot을 이용하여 이상치를 확인하였다. 다음으로 각 대학교별 통학하는 대학생 수의 정도를 수요량= (재학생수×(1-기숙사수용률))이라는 새로운 변수로 정의하였다. 최적 입지의 자치구를 선정하기 위하여 데이터의 단위를 자치구로 변환하였으며, 변수들의 척도의 다양성을 감안해 스케일링 작업을 적용하였다. 다목적 최적화할 때의 용이성을 감안하여 min-max scaling을 전체 변수에 적용하였다.

	변수명	내용	자료형
1	자치구	서울 특별시 내 자치구명	object
2	재학생수	해당 자치구 소재 대학 합계재학생 수	int64
3	수용률	$((\text{기숙사수용인원수})/(\text{재학생수})) \times 100(\%)$	float64
4	수요량	$(\text{재학생수} \times (1 - \text{기숙사수용률}))$	int64
5	공공기관수	해당 자치구 내 도서관, 기타 공공기관 합계	int64
6	경찰서	해당 자치구 내 경찰서 합계	int64
7	쇼핑	해당 자치구 내 쇼핑공간 합계	int64
8	역개수	해당 자치구 내 역개수 합계	int64
9	평일운행횟수	해당 자치구 내 지하철역 평균 평일운행횟수	float64
10	수송인원	해당 자치구 내 지하철역 평균 하루 수송인원	float64
11	공시지가	단위 면적(m^2)당 토지 가격	float64
12	주택매매가격지수	2021.6월 기준 주택매매가격 변동률	float64
13	세대수	해당 자치구 소재 행복주택 세대수	int64
14	주택공급물량	해당 자치구 내 공급될 행복주택 물량	int64
15	사업체수	해당 자치구 소재 사업체 총개수	int64
16	종사자수	해당 자치구 소재 사업체 합계종사자 수	int64

2) PCA

PCA란 차원 축소 기법으로 고차원 데이터를 저차원의 데이터로 변환함으로써 차원의 저주를 문제를 해소해 준다. 클러스터링에 필요한 저차원의 변수를 만들어 내기 위해 데이터에 위 기법을 적용한 결과, 차원축소는 성공적으로 되었으나 파생변수들의 해석이 명확하지 않은 것과 더불어 이후 단계인 클러스터링에서는 지수들만으로도 충분하다 판단하였다. 따라서 PCA를 기각하였고 대신에 클러스터링의 기준 축으로 사용할 경제성 지수, 접근성 지수를 자체적으로 설정하였다.

3) 경제성 지수, 접근성 지수 생성

경제성과 접근성, 그리고 인프라와 관련된 주요 변수들을 선정한 뒤 해당 변수들의 선형 결합을 통해 각 지수식을 설정하였다. 다음으로 각 변수의 가중치를 계산하기 위해 해당 지수의 주요 변수들을 독립변수로, 주택공급물량을 종속변수로 둔 랜덤포레스트 모델을 구축하였다. 랜덤포레스트는 변수 중요도라는 척

도를 통해 어떤 변수가 예측력이 가장 강한지 추정할 수 있는 분석 방법이다. 모델 결과를 반영한 최종 지수 식은 다음과 같다.

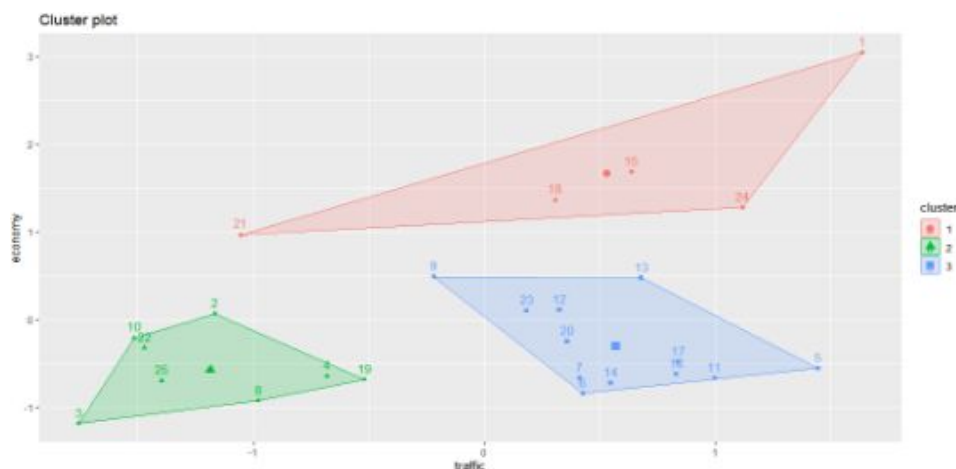
- 경제성 지수 = $0.6 \times (\text{공지시가}) + 0.4 \times (\text{주택매매가격지수})$
- 접근성 지수 = $0.3 \times (\text{수요량}) + 0.3 \times (\text{종사자수}) + 0.2 \times (\text{역개수}) + 0.3 \times (\text{수송인원})$
- 인프라 지수 = $0.3 \times (\text{공공기관수}) + 0.3 \times (\text{경찰서}) + 0.4 \times (\text{쇼핑})$

2. 클러스터링 분석

클러스터링 분석이란 각 데이터의 유사성을 측정하여 집단으로 분류하는 방법이다. 따라서 전체 데이터에서 자치구를 기준으로 하여 특성에 따라 구분하기 위해 클러스터링 분석을 진행하였다. 이때, x축은 접근지수로 y축은 경제지수로 두고 접근지수가 크고 경제지수가 작은 군집을 찾아야 한다. 본 연구에서는 k-means, GMM, k-medoids 방법을 사용하여 군집화를 시행하였다.

1) k-means 클러스터링

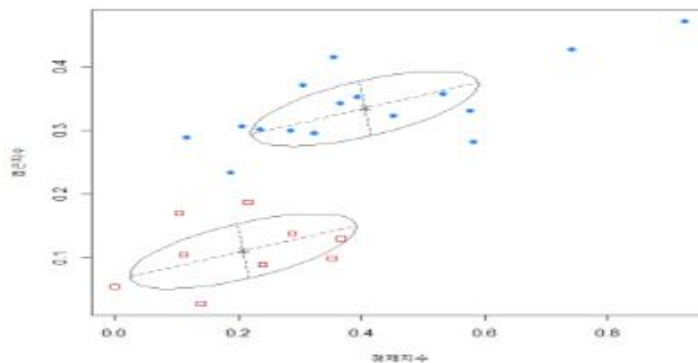
k-means 클러스터링은 k개의 중심점을 정하여 각 데이터들을 가장 가까운 중심점과 같은 그룹으로 할당한다. 그리고 그룹의 중심점을 다시 업데이트하여 다시 개별 데이터의 그룹을 할당하는 과정을 반복하는 방법이다. 최적의 군집 수를 정하기 위해 Elbow Method를 사용하였다. Elbow Method란 군집 수에 따라 SSE의 변동을 비교하여 최적의 군집 수를 정하는 방법이다. k가 3일 때 그래프의 선이 완만하게 꺾여 최적의 군집 수라 판단하여 클러스터링을 진행하였다. 접근지수가 높고 경제지수가 낮은 클러스터(3번 클러스터)가 생성되었으나 각 클러스터에 속한 데이터의 개수가 12개, 5개, 8개로 나누어져 클래스 불균형이 심하다 판단하였다. (<그림 2.1>) 따라서 k-means 방법은 기각한다.



<그림 2.1>

2) GMM 클러스터링

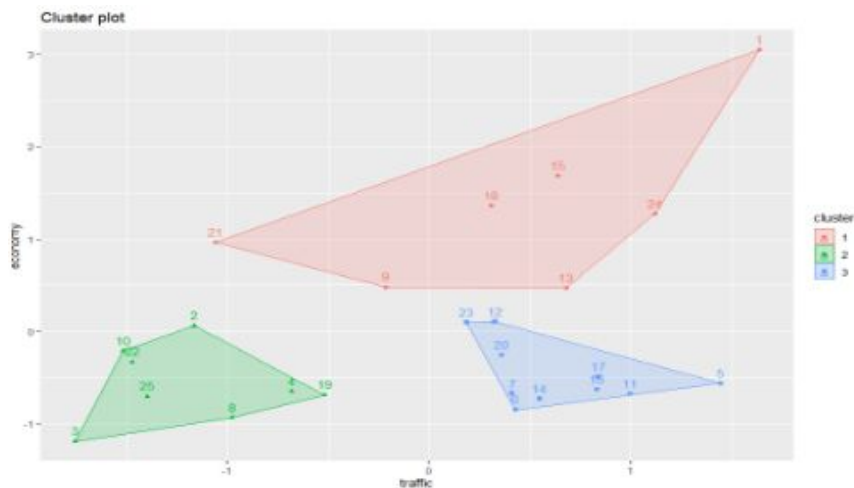
GMM 클러스터링은 클러스터링을 적용하고자 하는 데이터가 여러 개의 가우시안 분포를 가진 데이터 집합들이 섞여서 생성되었다는 가정하에 군집화를 수행하는 방식이다. GMM 클러스터링도 몇 개의 군집으로 나눌 것인지를 초기에 정해주어야 한다. 본 연구에서는 변수들의 분포를 살펴보고 $k=2$ 로 하여 클러스터링을 하였으나 군집이라 할 수 있는 뚜렷한 클러스터가 생성되지 않았다. (<그림 2.2>) 이에 따라 GMM 클러스터링 방법은 기각하였다.



<그림 2.2>

3) k-medoids 클러스터링

k-medoids 방법은 k-means 방법을 변형한 것으로 군집의 중심점을 데이터의 평균이 아닌 중앙값을 사용하여 정하는 방법이다. 데이터의 이상치에 민감한 k-means 방법의 단점을 보완해준다. k-medoids 방법도 최적 군집 수(k)를 Elbow Method를 사용하여 3으로 정하여 진행하였다. 각 군집에 속한 데이터의 개수가 7개, 8개, 10개로 클래스 불균형이 심하지 않고, 접근지수가 높고 경제지수가 낮은 클러스터(3번 클러스터)가 만들어졌다. (<그림 2.3>) 최종적으로 k-medoids 클러스터링 방법을 사용하였다.



<그림 2.3>

최종적으로 k-medoids 클러스터링 방법으로 접근지수가 높고 경제지수가 낮은 3번 군집을 선택하여 관악구, 광진구, 구로구, 동대문구, 동작구, 서대문구, 성동구, 서대문구, 성북구, 영등포구, 종로구로 총 10개의 자치구가 선택되었다.

3. 다목적 최적화

후보 자치구를 세 곳으로 줄이기 위해 뽑힌 자치구들에 대해 접근성, 경제성, 주변 인프라를 기준으로 다목적 최적화를 진행했다. 다목적 최적화는 제약조건과 함께 최대화 또는 최소화해야 할 목적함수가 여러 개인 문제에서 최적해를 찾는 방법이다. 목적함수와 다목적 최적화 식은 다음과 같다.

- $Maximize Z_1 = 0.3 \times (\text{수요량}) + 0.3 \times (\text{종사자수}) + 0.1 \times (\text{역개수}) + 0.3 \times (\text{수송인원})$

- $Minimize Z_2 = 0.6 \times (\text{공지시가}) + 0.4 \times (\text{주택매매가격지수})$

- $Maximize Z_3 = 0.3 \times (\text{공공기관수}) + 0.3 \times (\text{경찰서}) + 0.4 \times (\text{쇼핑})$

$$\Rightarrow Maximize Z = a \times Z_1 + b \times (1 - Z_2) + c \times Z_3, (a + b + c = 1, a, b > c)$$

목적함수 Z_1 , Z_2 , Z_3 는 각각 접근지수와 경제지수 그리고 인프라 지수에 해당한다. 다목적 최적화의 가중치 a , b , c 는 합산해서 1이 되어야 하고, 행복주택의 주요 목적인 접근성과 경제성은 주변 인프라보다 더 중요한 요소이므로 c 는 a 와 b 보다 작아야 한다. 그리고 접근성과 경제성은 최대한 비슷한 정도의 가중치를 가질수록 좋다. 이 조건들을 모두 만족시키는 0.1단위의 가중치를 찾기 위해 랜덤포레스트를 사용했다. 앞서 만든 경제지수, 접근지수, 인프라지수를 독립변수로, 주택공급물량을 종속변수로 두어 랜덤포레스트 모델을 구축하였다. 그 결과로, 접근지수, 경제지수, 인프라지수 순으로 중요도가 높게 나왔고, 가능한 가중치 경우는 $(a=0.5, b=0.3, c=0.2)$ 와 $(a=0.4, b=0.4, c=0.2)$ 였다. 다목적 최적화 결과에서 두 경우 모두 동일하게 동대문구, 관악구, 성북구가 뽑혔지만, 다목적 최적화 결과값이 더 높은 것은 $(a=0.4, b=0.4, c=0.2)$ 이므로 이를 다목적 최적화의 가중치로 정했다.

	동대문구	관악구	성북구
$(a=0.5, b=0.3, c=0.2)$	0.530	0.522	0.512
$(a=0.4, b=0.4, c=0.2)$	0.577	0.561	0.557

최종 자치구 후보지 세 곳으로 동대문구, 관악구, 성북구가 뽑혔고, 주거지형 재정비촉진지구 목록¹⁾에서 후보지 자치구에 해당하는 촉진 지구를 모두 추출했

1) 서울특별시 주택건축 분야별 정보 중 재정비촉진지구 현황(2022년 1월 기준) 목록을 참고하였다.
(<https://news.seoul.go.kr/citybuild/archives/516748>)



다. 그다음, 자치구별로 하나의 재정비촉진지구를 선정하기 위해 각 촉진 지구의 구역별 재개발 진행 단계²⁾와 행복주택이 들어서기 좋은 주변 환경을 가졌는지를 고려하였고 다음과 같이 선정되었다.

자치구	재정비촉진지구 명칭	위치
동대문구	이문·휘경 재정비촉진지구	동대문구 이문, 휘경동 일원
동대문구	전농·답십리 재정비촉진지구	동대문구 전농 1, 4동, 답십리 1, 3, 5동
관악구	신림 재정비촉진지구	관악구 신림동 일원
성북구	길음 재정비촉진지구	성북구 길음동 624, 971 일대
성북구	장위 재정비촉진지구	성북구 장위동 일원

4. 최종 입지 선정

마지막으로 각 재정비촉진지구 내에서 행복주택 건설이 가능한 구역을 선정하기 위해 지역 특성과 주변 환경을 조사하였고, 동대문구의 전농8 구역, 관악구의 신림4 구역, 성북구의 장위15 구역이 서울시 행복주택의 최종 입지로 선정되었다.

(1) 전농·답십리 재정비촉진지구 -> 전농8 구역


전농·답십리 재정비촉진지구	지역 특성
	<ul style="list-style-type: none"> • 2014년에 전농·답십리뉴타운 사업이 진행됨 • 전농7, 답십리16구역은 2014년에, 답십리 18구역은 2019년에 재개발이 완료됨 • [교통] 청량리역(1호선,경의중앙선), 신답역(2호선), 답십리역(5호선) • [환경] 배봉산, 답십리 공원, 근린공원, 청계천 • [학교] 서울시립대, 답십리초, 동대문중 등 • [기타] 전농119 안전센터 <p>⇒ 가능한 구역 : 전농8, 답십리12, 답십리18</p>
전농8 구역	구역 특성
	<ul style="list-style-type: none"> • 답십리12, 18구역에 비해 서울시립대학교와 가깝고 청량리역과도 멀지 않음 • [교통] 서울 동부의 주요 교통 거점인 청량리역과 가까움 • [환경] 인근에 3곳의 공원이 있음 • [학교] 서울시립대학교와 가까움 • [기타] 전농로타리시장이 있고 주변 상권이 편함

2) 서울시 정비사업 정보모핑 사이트의 사업장검색 기능을 이용하여 각 지역의 재개발 현황을 볼 수 있다.
<https://cleanup.seoul.go.kr/cleanup/bsnssttt/lsrMainIndx.do>

(2) 신림 재정비촉진지구 -> 신림4 구역

신림 재정비촉진지구	지역 특성
	<ul style="list-style-type: none"> • 신림1, 2, 3구역은 재개발(주택정비형) 진행 단계에 있음 • [교통] 서울대벤처타운역(신림선) • [환경] 좌측에 진우봉, 하측에 관악산 있고, 근처에 어린이공원이 3곳 있음 • [학교] 원신초, 광신중·고, 신우초, 미림여고 • [기타] 신신림시장, 삼성동시장 등 시장터가 있음 <p>⇒ 가능한 구역 : 신림4</p>
신림4 구역	구역 특성
	<ul style="list-style-type: none"> • 신림4 구역만이 행복주택 건설이 가능함 • [교통] 도보로 10분 거리에 서울대벤처타운역(신림선)이 있음 • [환경] 하측에 돌산이 붙어있고, 근처에 어린이공원이 2곳 있음 • [학교] 미림여고와 신우초 사이에 위치함

(3) 장위 재정비촉진지구 -> 장위15 구역

장위 재정비촉진지구	지역 특성
	<ul style="list-style-type: none"> • 현재 뉴타운사업이 진행 중이고, 장위 1구역에 장위지웰에스테이트(도시형)가 입주예정 • 장위4, 6, 7, 10구역에는 재개발(주택정비형) 진행 단계에 있음 • [교통] 북부간선도로, 상월곡역(6호선), 돌곶이역(6호선), 석계역(1호선) • [환경] 북서울꿈의숲, 월곡산, 영축산, 소곶내 어린이 공원 • [학교] 광운대, 동덕여대 • [기타] 장위전통시장, 장위119 안전센터 <p>⇒ 가능한 구역 : 장위3, 장위14, 장위15</p>
장위15 구역	구역 특성



- 장위3, 14 구역은 장위15 구역에 비해 지하철역과 멀고, 주변 인프라가 좋지 않음
- **[교통]** 북부간선도로가 있고 상월곡역(6호선)과 가까움
- **[환경]** 월곡산이 있고 운동시설이 잘 되어있음
- **[학교]** 광운대, 동덕여대, 한국예술종합학교(석관동 캠퍼스)와 가까움
- **[기타]** 장위전통시장과 가까움

3. 분석 활용 전략

□ 기대효과

기존 행복주택은 경제적인 측면 때문에 본래의 목표였던 대중교통이 편리하며 저렴한 좋은 조건의 주택을 제공하는 것이 어려운 상황이었다. 저렴한 곳은 도시 외곽에 위치해 교통 편이 좋지 않고, 접근성이 좋은 곳은 그만큼 경제적으로 부담이 되는 한계를 보였다. 그러나 본 연구에서 선정된 최적 입지인 세 곳은 가격도 저렴하며 대중교통이 편리하고 주변 인프라까지 고려를 한 곳이다. 따라서 행복주택의 주요 수요자인 대학생과 사회초년생들은 경제적인 부담에서 벗어날 수 있을 것이다. 또한 접근성의 한계에서도 벗어나 출근과 통학에 부담이 없다.

□ 방향제시

본 연구는 기존의 공공 주택 공급 방법에서 벗어나 접근성, 경제성, 주변 인프라를 고려하여 최적화된 입지 대안을 제시하고 있다. 이러한 방법은 공간만 따지는 것이 아닌, 실제 행복주택이 가지고 있던 문제점에 대해 해결할 수 있는 현실적인 대안이다. 따라서 서울 행복주택 건설에 있어 다양한 현실적인 상황을 고려해야 할 때 적절히 활용될 수 있을 것이다.

참고문헌

- 김지현, 강수진(2020), 「행복주택 성과지표 개선방안에 대한 연구 = A Study on the Improvement of Performance Indicators for Happy Housing Project」, 한국주택학회
- 민지현, 박하연, 정하은, 정희진, Zheng Haihua(2020), 「청년 공공임대주택의 문제점과 개선방안에 대한 연구」, 梨花女子大學校 法政大學 行政學科
- 진찬우, 이건학(2015), 「행복주택 최적 입지 선정에 관한 연구: 다목적 공간 최적화 접근」, 한국도시지리학회지 제18권 2호(81~95), 한국도시지리학회