

서울시 빅데이터 캠퍼스 공모전

서울시
시립 산후조리원
'최적입지' 찾기

NEW TURN

김상현 김성준 차다현 하예인

CONTENTS

프로젝트 배경, 목적 및 순서소개_ **프로젝트 소개**

회귀모형을 이용한 소득집단 구하기_ **데이터 분석**

GIS를 활용한 입지분석을 통한 종합분석_ **입 지 선 정**

최종 입지 선정 및 활용방안과 한계점_ **결론 및 제언**

1

프로젝트 소개

프로젝트 배경, 목적 및 순서소개

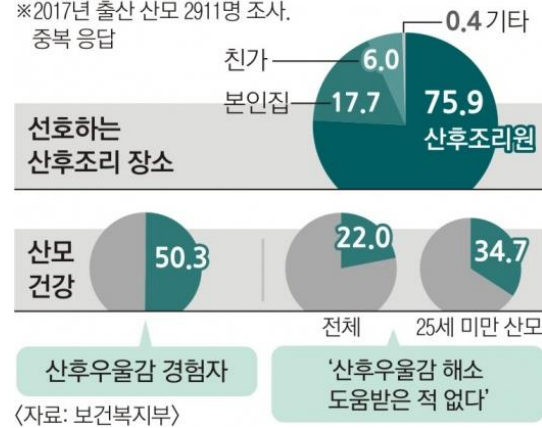
1 프로젝트 소개 _ 프로젝트 배경 및 현황분석

산후조리원이 왜?

출산율 감소의 원인?

2018년 산후조리 실태조사 (단위: %)

※2017년 출산 산모 2911명 조사.
중복 응답



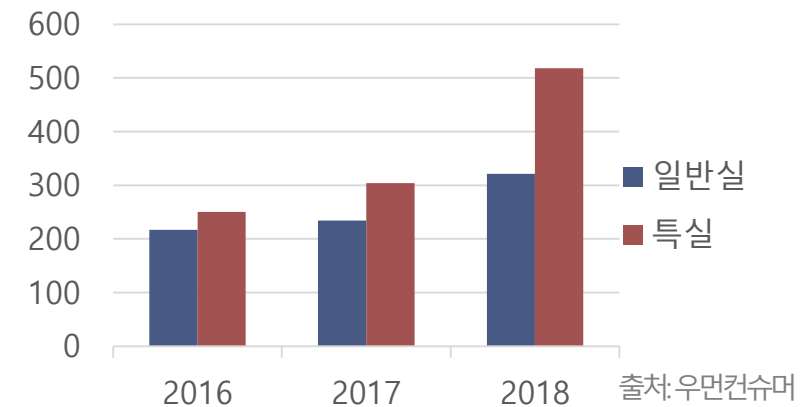
2018 산후조리 실태조사 결과
“산모에 대한 복지 부족”

전국 산후조리원 현황



출생아 대비 산후조리원 수 지속적 증가
“민간 산후조리원은 성장사업”

서울시 민간 산후조리원 가격



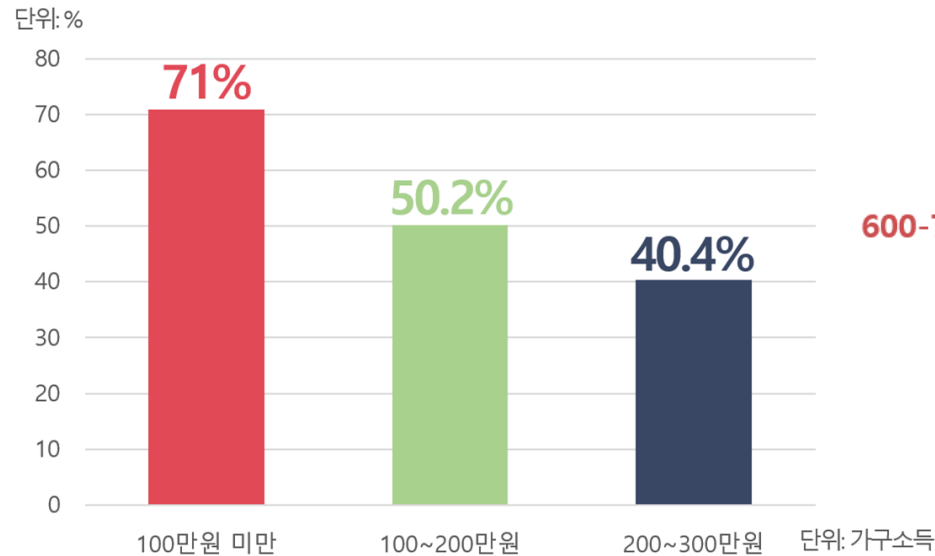
갈수록 비싸지는 민간 산후조리원
일반실 평균 321만원, 특실 평균 518만원
소비자물가상승률(1%)보다 큰 상승폭
“예비부모의 경제적 부담”

산후조리원의 필요성 증가와 민간 산후조리원의 고급화
“산모에 대한 복지를 마련해 출산을 장려”

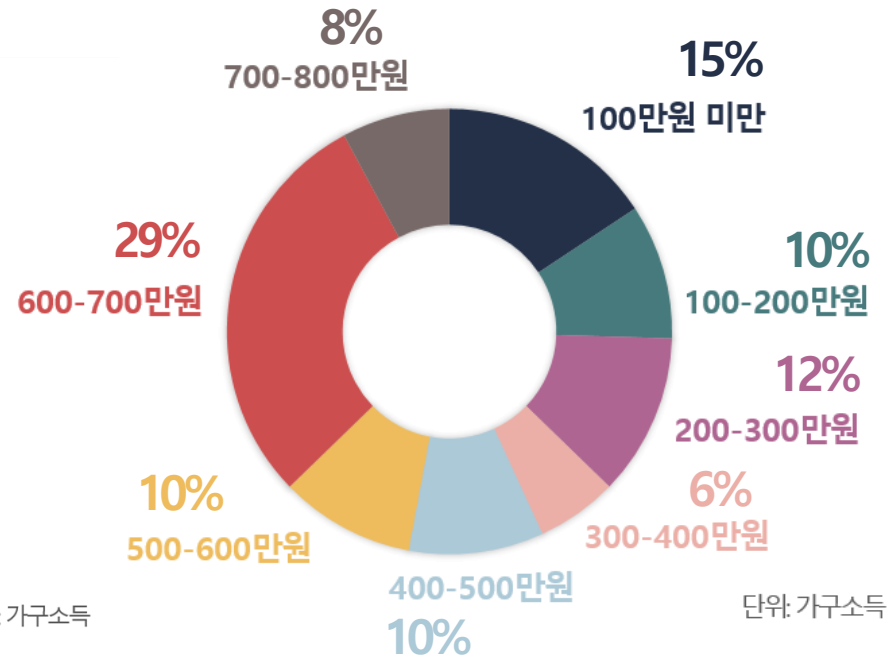
1 프로젝트 소개 _ 프로젝트 배경 및 현황분석

2018 산후조리 실태조사

산후조리원 비이용 이유를 비용부담이라고 뽑은 비율



공공 산후조리원 수혜비율



저소득층의 산후조리원 비이용 이유
비용부담

서울시에서 운영하는
공공산후조리원 1곳

지자체 단위로 운영되는 공공산후조리원
주요 수혜계층은 중산층

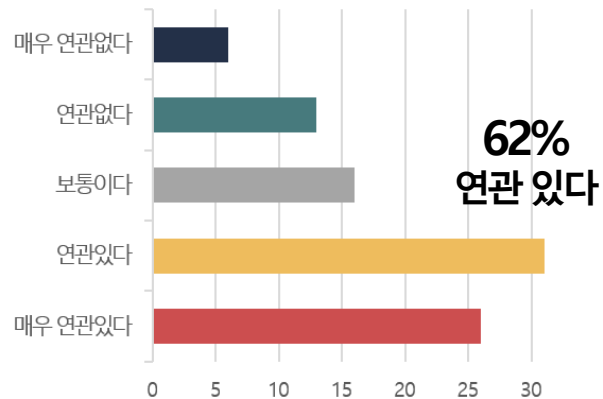
저소득층의 수혜를 최대화할 수 있는, "서울 시립산후조리원 필요"

1 프로젝트 소개 _ 프로젝트 배경 및 현황분석

프로젝트 타당성 검토를 위한 자체 설문조사 실시

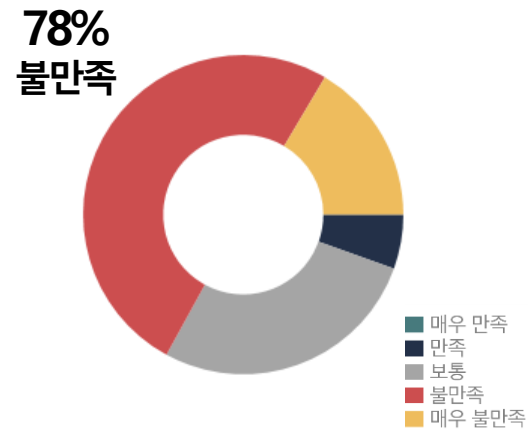
응답자 수: 92명
조사 방법: 구글 설문

산후조리 복지확대와 출산율과의 관계



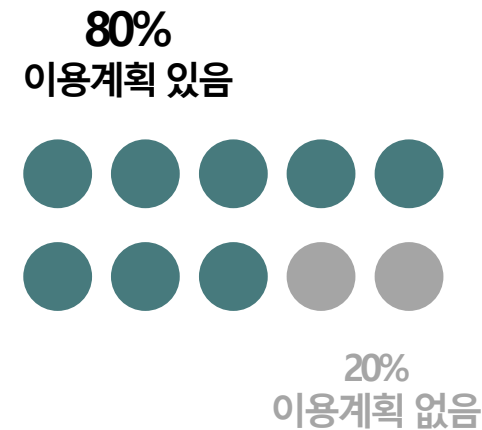
출산율 개선의 KEY 중 하나,
산후조리 복지 확대

산후조리원 비용만족도



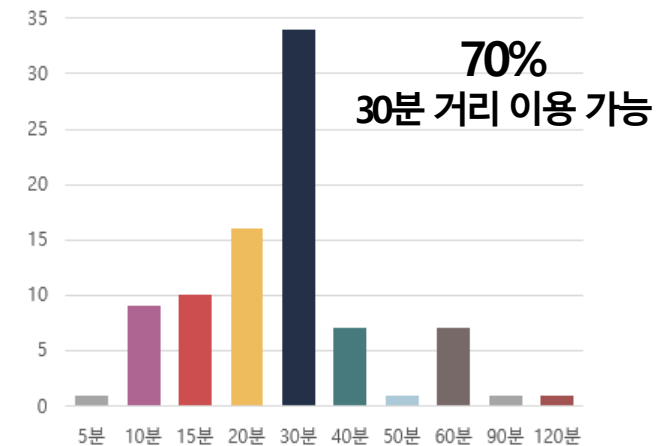
산모들에게 부담되는
과도하게 비싼 산후조리원

산후조리원 이용계획



산후조리원이 필수라고
생각하는 국민들의 확대

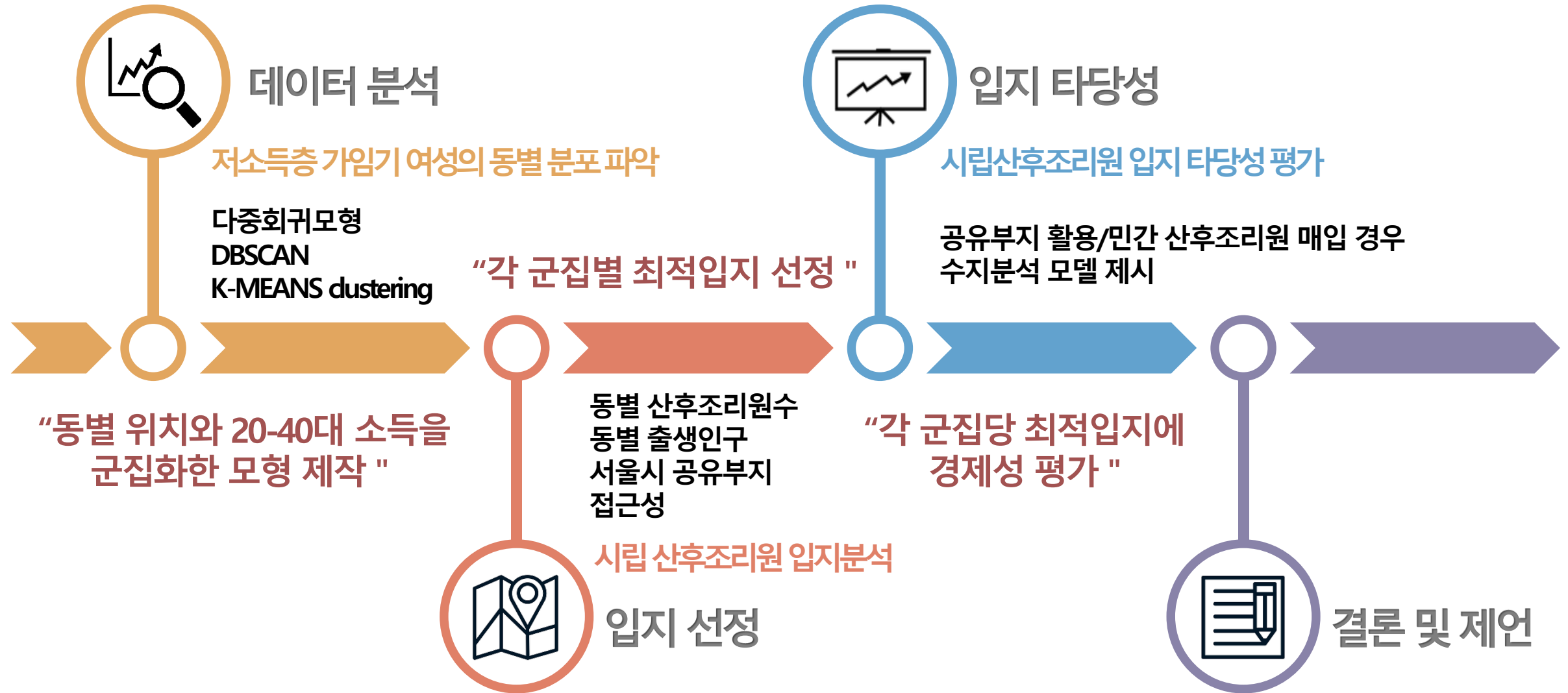
산후조리원까지의 이동 최대시간(분)



70%의 사람들이 30분거리까지
이용 가능

출산율 개선을 위한 저렴한 가격의 '시립 산후조리원' 건립 필요성 확인

1 프로젝트 소개 _ 프로젝트 과정



2

데이터 분석

회귀모형을 이용한 소득집단 구하기

2 데이터 분석_데이터 가공 과정 및 사용방법



데이터 분석

전제조건

분석과정

데이터 전처리

소득수준 데이터

행정동별 20, 30, 40 대의 3천만원 미만 소득자의 비율 순위화

부동산 실 거래가 데이터

단독다가구, 연립다세대, 오피스텔의 평당 월세(월세 * 100 = 전세) 환산, 전세

INPUT DATA

동별 소득순위

구별 GRDP

실거래가(월세)

실거래가(전세)

동별 기초생활수급 가구수

1단계 다중회귀모형

다중회귀모형 제작 후 10만번의 머신러닝

동별 소득과의 비례관계 확인 -> DBSCAN에 들어갈 변수 확정

“각 변수의 가중치 확인”



“위치와 20 - 40 대 저소득층의 소득을 반영한 군집 생성”



2단계 DBSCAN Clustering 밀도기반 클러스터링

각 동별 변수의 패턴을 머신러닝
각 변수의 패턴이 비슷한 동끼리 cluster 형성



“회귀모형에서 구한 가중치를 이용해 각 군집의 소득 위계 지정”

3단계 머신러닝을 통한 “동별 위치와 20 - 40 대 소득을 군집화한 모형 제작”

2 데이터 분석 _ 다중회귀모형

파이썬을 이용한 회귀분석

설명변수가중치 얻기

DBSCAN군집 간 소득수준 비교의 기준으로 가중치 사용

Supervised Learning : Multiple Linear Regression을 이용한 동별 소득 추정

- Multiple Linear Regression 이란?

두 개 이상의 독립변수들과 하나의 종속변수의 관계를 분석하는 기법

‘동별소득 데이터들의 가중치를 구하여 우선순위를 매기고자 사용’

- 다중회귀 모형제작

$$H(x_1, x_2, x_3) = w_1x_1 + w_2x_2 + w_3x_3 + b$$

- cost값으로 회귀모형 적절성 판단

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^m (H(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}) - y^{(i)})^2$$

독립변수

- 1.주택실거래가데이터
_ 면적당 월세를 전세로 반환한 금액
- 2.주택실거래가데이터
_ 면적당 전세 금액
- 3.기초생활수급가구수
4. GRDB (구별)

종속변수

1. 20 - 40 대 소득순위 데이터

2 데이터 분석 _ DBSCAN

파이썬을 이용한 회귀분석

설명변수 가중치 얻기

DBSCAN군집 간 소득수준 비교의 기준으로 가중치 사용

다중선형회귀분석 머신러닝 결과

다중선형회귀분석 결과분석

- 10만번 학습시킨 결과

Cost : 0.049572308 -> 모델의 정당성 확보

- 가중치 값으로 유의미한 변수 추출

[0.3379296] : GRDP

[0.22243083] : 월세 -> 전세환산금액

[0.39320806] : 전세금액

[-1.2396864] : 기초생활수급 가구수

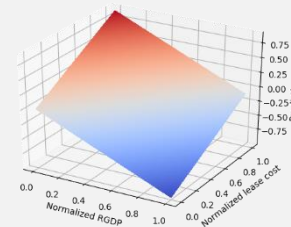
-> 가중치가 음수이므로 변수제거

DBSCAN(density-based spatial clustering of applications with noise)

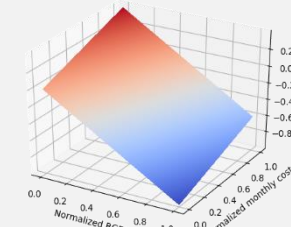
변수: GRDP, 월세->전세환산금액, 전세, 20-40대 동별소득순위

변수의 분포 패턴을 머신러닝하여
비슷한 패턴끼리 군집화하는 '밀도기반 클러스터링' 기법
"다양한 측면을 고려한 최적의 20-40대 동별 소득 추정"

RGDP - 전세



RGDP - 월세



월세 - 전세

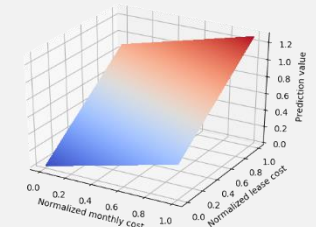


그림1. 선형 회귀분석 시각화

2 데이터 분석 _ DBSCAN

파이썬을 이용한 회귀분석

설명변수 가중치 얻기

DBSCAN군집 간 소득수준 비교의 기준으로 가중치 사용

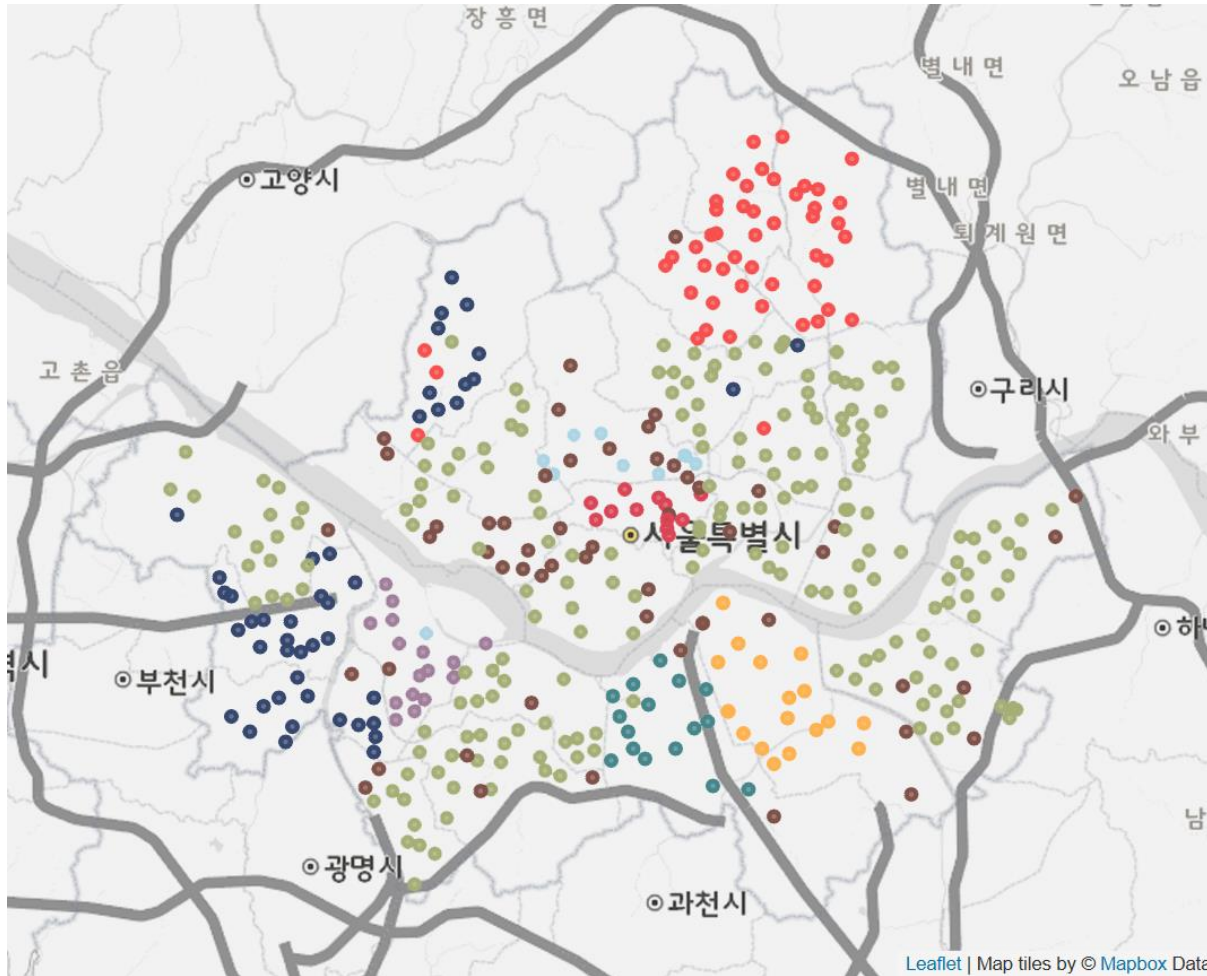


그림 2. DBSCAN 결과

범례

- | | | |
|-----------------|----------------|----------------|
| ● -1: '#E96060' | ● 5: '#9F849B' | ● 2: '#A3AA7C' |
| ● 7: '#F3B15A' | ● 4: '#7C564F' | ● 1: '#CD5362' |
| ● 6: '#4C8184' | ● 3: '#3C4A6B' | ● 0: '#B2D3E1' |

소득순위

“회귀분석의 가중치를 이용해 나눈 소득 순위”

● 4 < ● 3 < ● 0 < ● 2 < ● 5 < ● 1 < ● 6 < ● 7

하위 3개 클러스터 추출

- 4: '#7C564F' ● 3: '#3C4A6B' ● 0: '#B2D3E1'

2 데이터 분석 _ KMeans

K-Means clustering

- 보정소득순위 하위 3개 **cluster의 순위 데이터**와
각 동의 동사무소 주소 기준으로 나온 **x, y 좌표**를 이용해 군집화
동별 동사무소는 python을 이용해 구글 지도에서 crawling

“ 비슷한 보정 소득순위 중 근접한 지역이면 하나의 cluster로 묶어 산후조리원 입지선정을 진행 ”

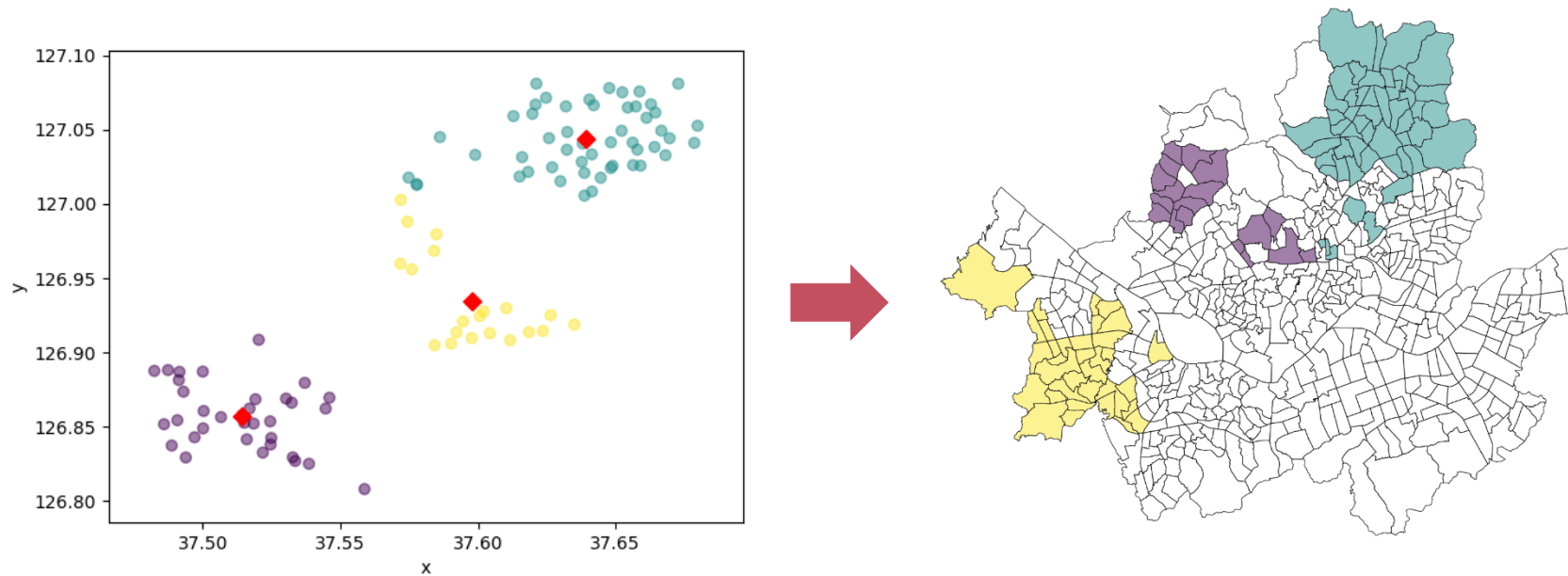
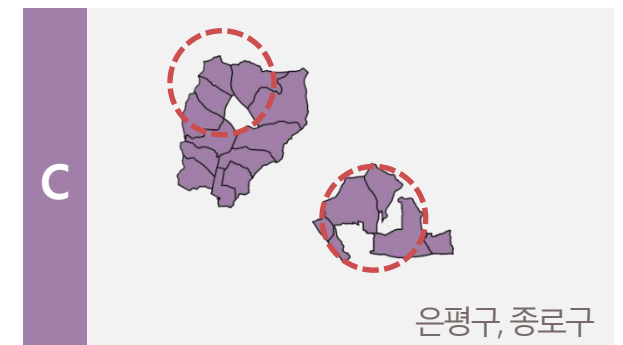
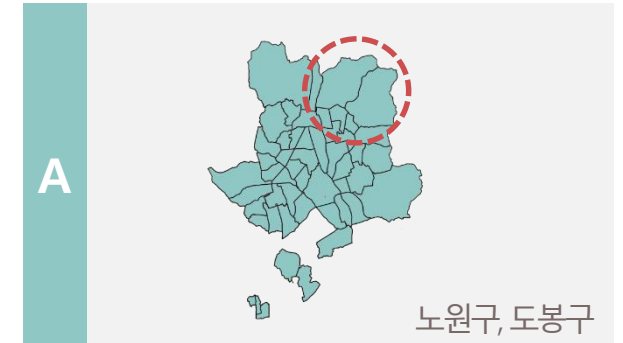


그림 3. K-Means 결과



3

입지 선정

GIS를 활용한 입지분석을 통한 종합분석

3 입지 선정 _ 데이터 가공 과정 및 사용방법



입지 선정

산부인과 위치

대상 군집에 인접한 산부인과 분포 파악_
인접 산부인과 고려

접근성

지하철 반경을 고려한 접근성 분석_
접근성이 좋은 최적입지 선정

전제조건

분석과정

20-40대 소득수준

3단계 머신러닝을 통한 소득모형

소득수준과 동별 위치데이터가 반영된 군집모형

“소득이 낮은 저소득지역 선정”

공유부지 활용

토지매입? 임대료를 최소화

공유부지의 활용으로 빠른 손익분기점 달성

“취약 동과 인접한 공유부지 선정”

산후조리원 수 / 출생인구

동별 산후조리원 수와 출생인구간 상관관계 파악

동별 산후조리원에서 출생인구를 나눈 값이 0에 수렴할수록 취약지역

“출생인구는 많지만, 산후조리원 수가 적은 취약지역 선정”

기존 산후조리원 활용

군집 모두에게? 최대 수혜를 제공

군집의 가운데에 위치한 기존 산후 통해 수혜지역 극대화?

“입지가 적절한 기존 산후조리원 매입”

입지분석을 통한 “시립 산후조리원의 최적입지 선정”

3 입지 선정 _ 취약지역 선정

GIS를 활용해 각 클러스터의 취약지역 선정

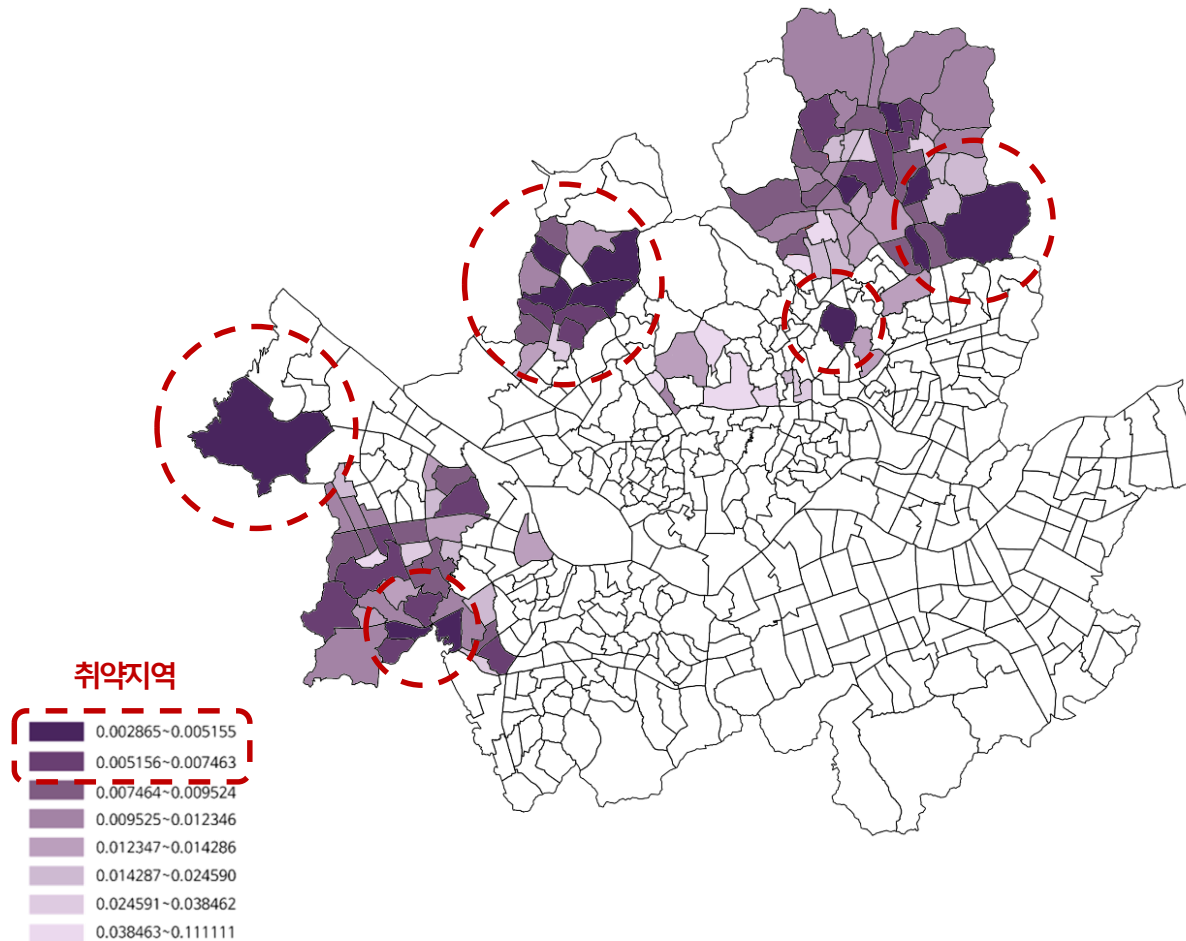


그림4. K-Means 결과

산후조리원 입지선정을 위해 고려해야 할 요소

1. 유입예상인구(출생인구) 2. 주변 산후조리원 수 3. 접근성



산후조리원 수+1
출생인구

산후조리원이 0개인 지역소득에 +1을 합
을 계산해 0에 가까운 지역을 산출함으로써

20-40대 소득이 낮은,
출생인구가 많지만
산후조리원 수가 적은,
산후조리 **취약지역**을 선정

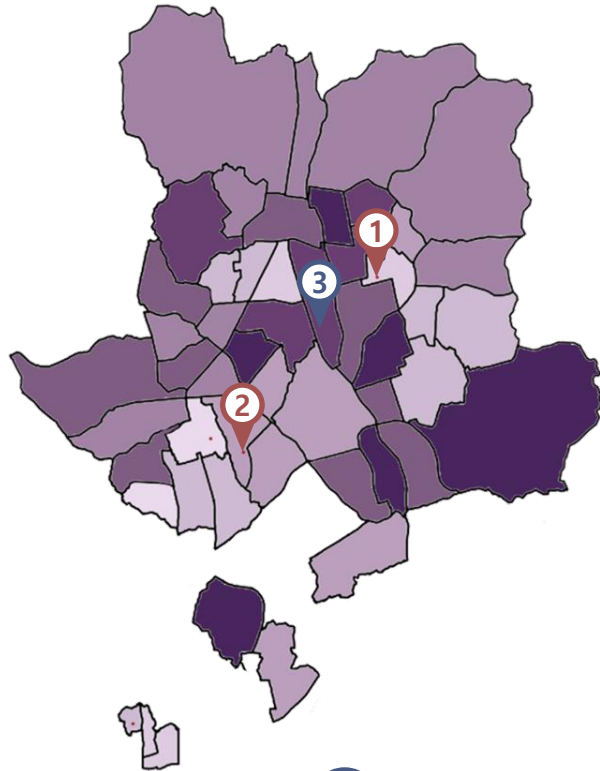


“ 클러스터 내 공유부지 중
취약지역과 접근성을 고려해 **적절부지**를 선정 ”

서울시 소유의 공유재산

3 입지 선정 _ 공유부지 접근성 분석

A 집단



○ 반경 500m

■ 서울시 공유부지

■ 기존 산후조리원

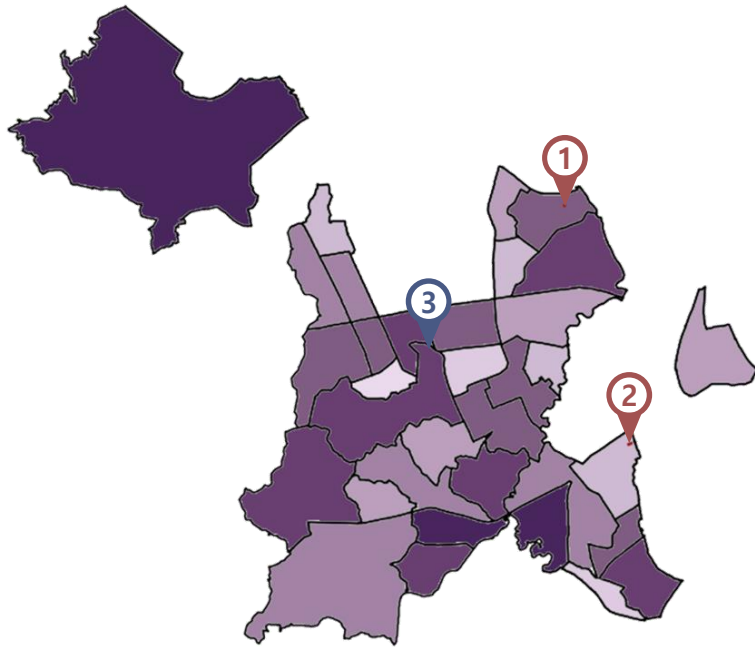


2번은 지하철 반경 500m에 포함되지 않으므로 1번이 더 적합

3번은 군집의 가운데에 위치하며 역 반경 500m내에 입지해있으며, 가장 가까운 산부인과까지의 거리가 820m로 매입 후보지로 적절

3 입지 선정 _ 공유부지 접근성 분석

B 집단



○ 반경 500m

■ 서울시 공유부지

■ 기존 산후조리원

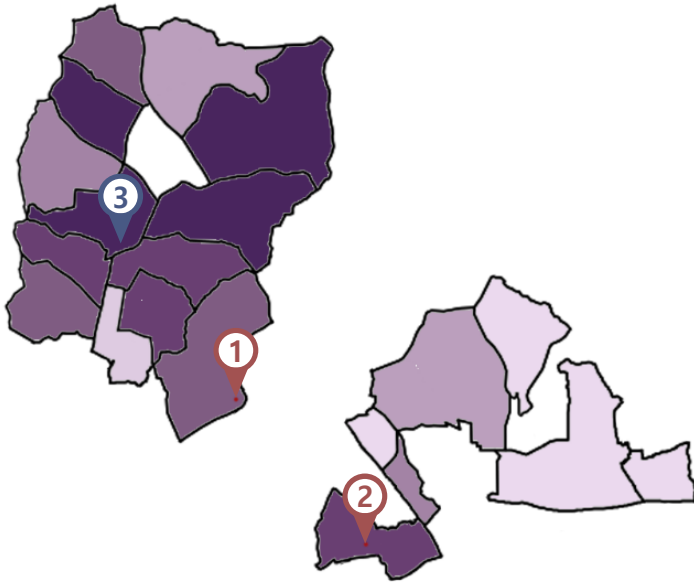


1번, 2번 두 곳 다 지하철 접근성이 좋지만, 취약 동과 인접한 **1번**이 더 적합

3번은 군집의 가운데에 위치하며 두 역의 결절지 내에 입지해있으며, 산부인과와 같은 건물로 **매입 후보지로 적절**

3 입지 선정 _ 공유부지 접근성 분석

c 집단



○ 반경 500m

■ 서울시 공유부지

■ 기존 산후조리원

1. 홍은동 429



2. 충현동 3-438



3. 아이젤 산후조리원



지하철과의 접근성이 더 좋고, 취약 동과 인접한 **2번**이 더 적합

3번은 군집의 가운데에 위치하며 역 반경 500m내에 입지해있으며, 가장 가까운 산부인과까지의 거리가 34m로 **매입 후보지로 적절**

3 입지 선정 _ 타당성 검토

수지모델분석을 통한 타당성 검토

연간유입예상인원 산정모델

$$I = B \times E$$

I: 연간유입예상인원

B: 저소득지역 출생인구

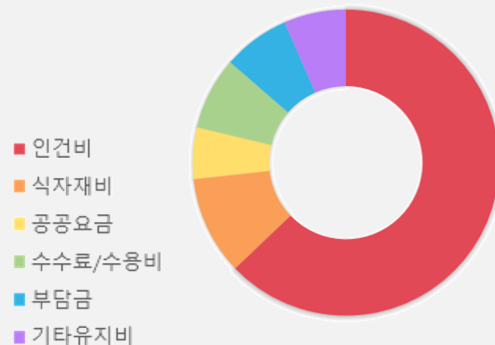
E: 저소득층 중 시립산후조리원 예상이용희망 비율(34.8%)

$$E = (LP_y \times 0.4 + LP_n \times W_{income}) \times 0.8$$

LP_y: 저소득층 산후조리원 이용인원

LP_n: 저소득층 산후조리원 비 이용인원

W_{income}: 비이용 원인이 소득인 비율



산후조리원 운영 지출액

수지분석모델

$$\pi = TR - TC = N(P - \frac{MP}{A})$$

π: 연간 순수익

P: 시립 산후조리원 예상 이용 금액 (120만원)

-수지분석을 바탕으로

TR: 연간 총수입 = $N \times P$

TC: 연간 총비용 = $N \times \frac{MP}{A}$

N: 연간수용인원

-I(연간유입예상인원)를 고려

MP: 산후조리원 운영 비용 (17163.8천 원)

- 산후조리원 보고서 참조

A: 수용 가능 인원 (160명)

- 차용한 보고서에 해당하는 산후조리원의 수용 가능 인원

$$\text{손익분기점} = \frac{\text{초기비용}}{\pi}$$

공유부지 사용 시

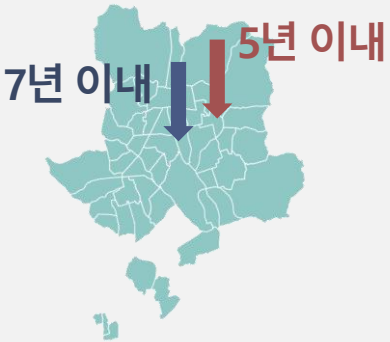
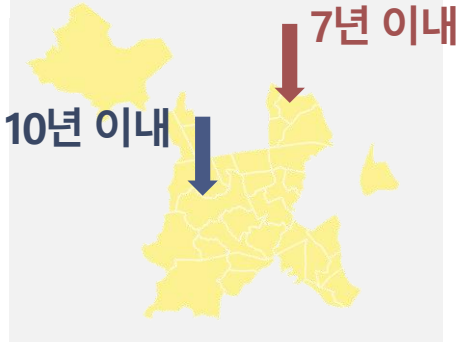
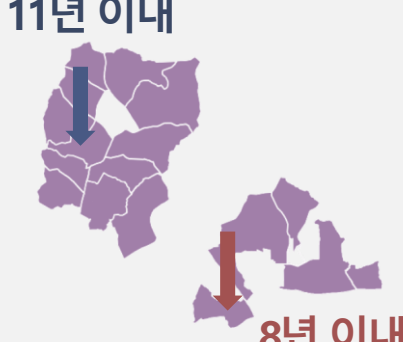
산후조리원 설립 비용
→ 88억(원)

민간 산후조리원 매입 시

산후조리원 설립비용(88억원) + 건물매입가(35억원)
→ 123억(원)

3 입지 선정 _ 타당성 검토

수지모델분석을 통한 타당성 검토

A 지역	B 지역	C 지역
<p>동시수용인원 : 50명 연간 수용인원 1500명</p> <p>예상유입희망인구 = 수용인원 * 147%</p> <p>수익 : 180억원 - 지출 : 160.910억원</p> <p>순수익 : 19.09억원</p> <p>손익분기점</p> 	<p>동시수용인원 : 34명 연간 수용인원 1000명</p> <p>예상유입희망인구 = 수용인원 * 178%</p> <p>수익 : 120억원 - 지출 : 107.274억원</p> <p>순수익 : 12.73억원</p> <p>손익분기점</p> 	<p>동시수용인원 : 30명 연간 수용인원 900명</p> <p>예상유입희망인구 = 수용인원 * 104%</p> <p>수익 : 108억원 - 지출 : 96.546억원</p> <p>순수익 : 11.45억원</p> <p>손익분기점</p> 

연간 예상 산후복지 수혜가구
= 3400가구

산후조리원 설립 비용 88억원
-> 손익분기점은 공유부지 활용할 경우
모두 8년 이내 도달

이후 연간 시립산후조리원 운영을 통한
재정확보액
= 약43.3억원



이를 다시 산모복지에 사용해
<아이 낳기 좋은 서울>을 위한
"선순환" 가능

4

결론 및 제언

결론 및 제언, 참고문헌

4 결론 및 제언 _ 결론 및 제언

☑ <서울시립산후조리원> 공공산후조리원의 한계를 넘다
철저한 복지대상 선정과 입지분석을 통한
저소득층 가임기 인구가 최대수혜 도모

☑ 공공산후조리원의 경우 인건비가 적자의 원인
이를 철저한 입지분석으로 유입인구를 극대화함으로 해결
기존 공공산후조리원보다 더 낮은 가격으로 복지혜택 가능

☑ 저소득층이 평균 이용하는 산후조리원 가격 224만원
시립산후조리원은 53.6% 가격으로 시민에게 혜택제공

☑ 손익분기점 8년 이내 예상
그 후 생기는 순수익으로 산후복지의 선순환
손익분기점 이후 매년 흑자 433억원 상당의 공공재정 확보
이를 다시 서울시 산모 복지를 위해 사용하는 복지 선순환

☑ 산후조리원 일반 서비스업으로 분류되어 35%의 세금징수
이 비율을 낮춘다면 훨씬 빨리 손익분기점에 도달할 것

☑ 기존 구단위로 시행되던 산모사업을 서울시가 자체적으로 관리
보다 평등하고 합리적인 예산 운영 가능

4 결론 및 제언 _참고문헌 및 분석데이터

참고문헌

2018, 산후조리원 실태조사 결과분석', 한국보건사회연구원
 2018, 제6기 지역보건의료계획 2017년 시행결과 및 2018년 시행계획, 송파구청
 이정원, 2013, 공공산후조리원 설치 및 운영 타당성 조사보고서, 디제이케이기획
 2014, 예상매출액 산정서 표준양식에 관한 규정, 공정거래위원회
 2019, 산모 4명 중 3명은 산후조리원 이용...만족도는 '친정' 1위, <http://www.etoday.co.kr/news/view/1711655>
 2018, '복지 사각지대' 산후우울증.. 산모.아기까지 위협한다, <https://news.vdaum.net/v/20180528171818123>
 2019, '여성 60% "임산부 정부 지원정책 잘 모르겠다"', http://news.khan.co.kr/kh_news/khan_art_view.html?art_id=201930417138001

사용도구



사용 데이터

	데이터셋목록	기간	활용목적	출처
1	저소득층비율	2018년	동별20, 30, 40대의저소득층비율이 높은위치산정	열린데이터광장
2	소득수준데이터	2015년		서울특별시빅데이터캠퍼스
3	서울시산후조리원현황및이용요금가격	2019년 8월		서울특별시 - 서울소식 - 고사공고
5	부동산실거래가	2019년 8월		국토교통부
8	구별GRDP	2018년		통계청
4	동별출생인구	2016년 4월 7일	산후조리원위치산정	열린데이터광장
6	서울시부지	2018년	산후조리원위치산정	서울특별시정보소통광장
7	교통데이터	지하철 - 2014	산후조리원위치산정	서울특별시빅데이터캠퍼스