

2022년 데이터 분석 청년 인재양성 데이터 분석 전문교육과정

데이터로 같이, 가치 있게(With Value)!

데이턴십 해커톤 제 5회

# 공동 보육을 위한 어린이집 공동체 분류 모델

분석 결과보고서

참여조: 서울 18조

참여자: 정은아(조장),  
이예원, 이은지, 이제형,  
조유경, 한지혜

씨에스리 컨소시엄


 CSLEE kpc 한국생산성본부

Copyright © CSLEE Consortium

CSLEE Consortium의 사전 승인 없이 본 내용의 전부 또는 일부에 대한 복사, 배포, 사용을 금합니다.

## 목 차

1. 분석 개요 .....	5
1.1. 분석 배경 및 개요 .....	5
1.2. 분석 목적 및 방향 .....	9
1.3. 분석 결과 활용 방안 .....	11
2. 분석 데이터 .....	12
2.1. 분석 데이터 목록 .....	12
2.2. 데이터 상세 설명 .....	13
2.3. 데이터 정제 방안 .....	16
3. 분석 프로세스 .....	19
3.1. 분석 프로세스 .....	19
3.2. 분석 내용 및 방법 .....	21
4. 최적 어린이집 공동체 분석결과 .....	37
4.1. 시범 운영 탐색 .....	37
4.2. 인천광역시 어린이집 분석 .....	38
4.3. 인천광역시 중심 국공립어린이집 선정 .....	40
4.4. 인천형 모아어린이집 공동체 구성 어린이집 추출 .....	41
4.5. 인천형 모아어린이집 위치 시각화 .....	42
5. 활용 방안 .....	44
5.1. 문제점 개선 방안 .....	44
5.2. 업무 활용 방안 .....	44
6. 참고자료(Reference) .....	45
7. 부록 .....	47

## 그 림 목 차

[그림 1-1] 출생아 수 및 합계출산율 추이, 1970~2021년 (통계청) .....	5
[그림 1-2] OECD 38개 회원국 출산율, 2020년 (OECD) .....	5
[그림 1-3] 서울형 모아어린이집 추진 체계 (서울시) .....	9
[그림 1-4] 서울형 모아어린이집 예시 (공동체의 국공립 비율이 높은 경우) .....	10
[그림 1-5] 서울형 모아어린이집 예시 (어린이집 간의 거리가 먼 경우) .....	10
[그림 1-6] 서울형 모아어린이집 예시 (공동체 주변 필요시설이 없는 경우) .....	11
[그림 2-1] 전체 시/도 어린이집 일반현황 .....	13
[그림 2-2] 행정구역(시군구)별/1세별 주민등록인구 .....	14
[그림 2-3] 시도별 아동학대 신고접수 건수 .....	14
[그림 2-4] 향후 늘려야 할 공공시설 설문조사 .....	15
[그림 2-5] 향후 늘려야 할 복지서비스 설문조사 .....	15
[그림 2-6] 인천광역시 어린이집 기본정보 .....	16
[그림 2-7] 지역 채택 데이터셋 .....	17
[그림 2-8] 인천 어린이집 데이터셋 .....	18
[그림 3-1] 분석 프로세스 .....	19
[그림 3-2] 중심 모아어린이집 선정 순서도 .....	20
[그림 3-3] 모아어린이집 그룹 생성 순서도 .....	20
[그림 3-4] 시도별 기준에 대한 데이터셋 .....	21
[그림 3-5] 시도별 기준 시각화 .....	22
[그림 3-6] 시도별 기준1 시각화 .....	22
[그림 3-7] 시도별 기준2 시각화 .....	22
[그림 3-8] 시도별 기준3 시각화 .....	23
[그림 3-9] 시도별 기준4 시각화 .....	23
[그림 3-10] 시도별 기준5 시각화 .....	23
[그림 3-11] 시도별 기준6 시각화 .....	23
[그림 3-12] 시도별 기준 순위 데이터셋 .....	25
[그림 3-13] 인천 어린이집 시각화 .....	25
[그림 3-14] 인천 어린이집 데이터셋 .....	26
[그림 3-15] 차원축소 모델 .....	26
[그림 3-16] umap을 이용하여 2차원 공간에 어린이집 데이터 시각화 .....	29
[그림 3-17], [그림 3-18] 차원축소한 그림 시각화 .....	31
[그림 3-19] condensed_tree_ 예시 .....	32
[그림 3-20] 인천 어린이집 중심 국공립 어린이집 선정 시각화 .....	33
[그림 3-21] 중심 국공립 어린이집 500m 이내 어린이집 데이터셋 .....	34
[그림 3-22] 인천형 모아어린이집 공동체 구성 어린이집 추출 과정 .....	35
[그림 3-23] 최종 인천 모아어린이집 데이터셋 .....	35
[그림 3-24] 인천형 모아어린이집 시각화 .....	36
[그림 4-1] 지역 채택 종합 순위 시각화 .....	37
[그림 4-2] 지역 채택 종합 순위 데이터셋 .....	37
[그림 4-3] 최종 클러스터링 결과 .....	39

[그림 4-4] 그룹별 어린이집 특성 시각화 .....	39
[그림 4-5] '그룹2'와 '그룹3'의 어린이집 특성 시각화 .....	40
[그림 4-6] 어린이집 특징 분류 데이터셋 .....	40
[그림 4-7] 인천 국공립 어린이집 선정 시각화 .....	41
[그림 4-8], [그림 4-9] 인천형 모아어린이집 공동체 어린이집 추출 데이터프레임 .....	42
[그림 4-10] 최종 인천광역시 모아어린이집 데이터셋 (그룹 고유 ID 포함) .....	42
[그림 4-11] 인천형 모아어린이집 시각화 .....	43
[그림 4-12] 두 그룹의 모아어린이집 시각화 .....	44

## 표 목 차

[표2-1] 분석 데이터 .....	12
[표2-2] 데이터 상세 설명 .....	13

## 1. 인천선택아기

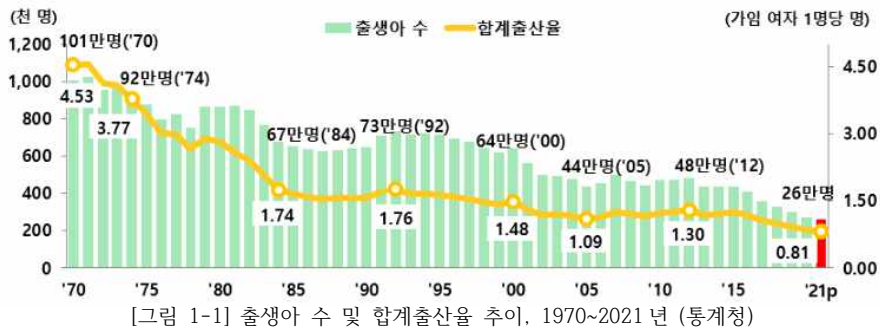
## 1. 분석개요

### 1.1. 분석 배경 및 개요

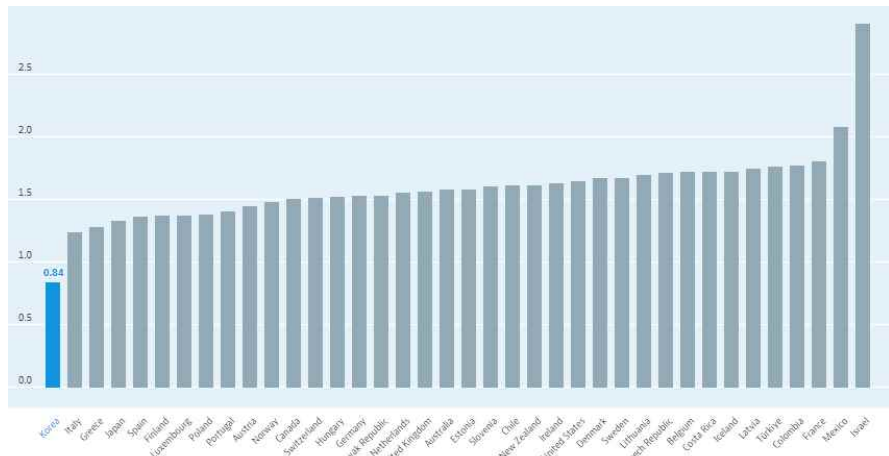
#### 1.1.1. 분석 배경

##### 1) 대한민국 저출산 현황

- 출생아 수는 2015년 12월부터 78개월째 전년 대비 감소세를 이어가고 있으며, 합계출산율이 2015년엔 1.24명이었으나 2018년 0.98명부터 2021년 기준 0.81명을 기록하며 평균 출생아 수가 한 명도 되지 않는다.
- 올해 합계출산율은 0.7명대로 전망되고 있다.



- 경제협력개발기구(OECD) 38개 회원국의 평균 합계출산율은 2019년 기준 1.61명으로 합계출산율이 1명 미만인 나라는 한국이 유일하다.



[그림 1-2] OECD 38 개 회원국 출산율, 2020년 (OECD)

#### 2) 저출산으로 인한 문제점

- (인구절벽 현상 발생) 인구절벽 현상이란 어느 순간을 기점으로 한 국가나 구성원의 인구가 급격히 줄어들어 인구 분포가 마치 절벽이 깎인 것처럼 역삼각형 분포가 된다는 이론으로, 생산가능인구(만 15~64세)가 급격히 줄어들고 고령인구(만 65세이상)가 급속도로 늘어나는 현상을 가리킨다. 현재 대한민국은 저출산 추세로 인구가 감소하는 것은 기정 사실로 인구절벽에 대한 대비가 필요하다.
- (인구소멸 위험지역 발생) 소멸위험지수란 65세 이상 인구 대비 20~39세 가임기 여성인구의 비율로, 한국고용정보원의 보고서는 소멸위험지수가 0.5 이하인 지역을 소멸위험지역으로 파악하고, 소멸위험지수가 낮을수록 소멸위험이 높다고 파악이 된다. 2022년 3월 기준 소멸위험 지역은 전국 228개 시군구의 절반에 가까운 113개(49.6%)에 달한다.
- (사회 시스템 붕괴) 저출산 현상이 가속화될 경우, 생산가능인구가 감소하고 고령화 추세로 인한 노인 부양 부담이 증가할 것이다. 이로 인해 경제성장 속도가 급속히 둔화하여 국가 경제에 위협이 될 수 있으며 국민연금, 의료보험 등 사회보장 시스템의 붕괴 또한 야기할 수 있다.
- [관련 자료] 여론조사 전문기관 한길리서치가 나의미래연구소 의뢰로 7월 30일에서 8월 1일까지 전국 만 18세 이상 성인 1050명을 대상으로 '저출산·고령화 문제에 따른 국가소멸 여부'에 대해 조사한 결과, 국가소멸 가능성에 공감한다는 응답이 85.7%(적극 공감한다 65.4%, 조금 공감한다 20.3%)로 나타났다. 연령별로는 60대 이상 89.4%가 저출산으로 국가소멸할 거 같다고 답해 가장 높았다. 이어 50대(89.2%), 40대(87.3%), 30대(82.8%), 18~29세(76.1%) 순이었다. '저출산 지속 시 국민연금·의료보험 등 사회보장 시스템이 유지될지' 묻는 질문에 응답자 83.8%(유지가 매우 힘들다 55.1%, 유지가 조금 힘들다 28.7%)가 유지하기 어렵다고 바라봤다.

#### 1.1.2. 분석의 필요성

##### 1) 기존 현금성 지원정책의 미비한 성과와 재원 마련의 한계점

- 2006년 정부가 '저출산, 고령사회 기본계획'을 시행한 뒤 2006년부터 2020년까지 15년간 투입한 예산은 380조 2000억원에 달한다.

- ☐ 2021년 정부는 저출산 문제를 해결하기 위해 아동수당, 출산장려금 등 46조 7000억원의 예산을 투입하였으나 출생아 수를 고려하면 신생아 한 명당 약 1억 7658만원 사용한 셈으로 지원 대비 효과 미비하다.
- ☐ 여성단체들은 정제된 출산율에 대한 원인 중 하나로 정부의 잘못된 ‘출산 보상정책’ 때문이라 주장하기도 한다.
- ☐ 이처럼 현금성 지원 정책과 더불어 지속적인 재원 조달은 저출산 문제 해결에 있어서 명백한 한계를 가진다.

## 2) 저출산 정책 새로운 추진전략 “함께 일하고 함께 돌보는 사회 조성”

- ☐ 저출산이란 사회적 문제를 해결하기 위해 대응 방안과 전략, 실행 계획에 대한 새로운 접근 방식이 필수불가결한 시기로 기존 정책에서 개선된 방향을 모색할 필요가 있다.
- ☐ 최근 수립된 제4차 저출산·고령사회 기본계획(2021~2025년)에 따르면, 남성과 여성 모두 ‘노동 중심 생애’를 지향하며 청년층의 노동권 실현을 위해 남녀가 함께 일하고 함께 돌보는 여건을 조성할 필요가 있음을 알 수 있다.
- ☐ 이에 “함께 일하고 함께 돌보는 사회 조성”, “모두의 역량이 고루 발휘되는 사회” 등이 주요 추진전략으로 선정되었다. (대한민국 정부, 2021a: 43)
- ☐ MZ세대가 ‘나’의 행복, 삶의 질 등의 가치를 중시한다는 점에서 제4차 기본계획 저출산 대응 정책 패러다임이 함께 일하고 함께 돌보는 사회로의 전환을 시도하였다는 점은 의의가 있다. (출처\_한국정책연구 제22권 제 1호, 저출산 대응 정책의 요구도 및 우선순위분석: MZ세대 인식을 중심으로)

## 3) 새로운 보육 방식의 필요성

- ☐ 코로나19가 유행한 이후, 어린 자녀를 둔 부모들은 돌봄 공백이 발생하여 새로운 형태의 보육 방식을 시작했다. 이들은 각자의 자녀를 품앗이 개념으로 돌보는 공동육아 방식을 선택했다.
- ☐ 공동육아 방식을 통해 ‘독박육아’에서 벗어날 수 있어 육아 스트레스를 감소시킬 수 있으며, 돌봄 공백이 줄어들어 워킹맘에게도 실질적인 도움이 된다. 특히 맞벌이 부부는 공동육아를 통해 돌봄 공백을 서로 채워주게 돼 자녀를 더 낳을 수 있는 환경이 조성된다.

- ☐ 이처럼 공동육아에 대한 만족도는 어린이집에 비해 크지만, 현재 개원한 어린이집과 공동육아 나눔터의 수는 확연히 차이가 난다.
- ☐ 윤재석 영유아보육학회 부회장은 “공동육아나눔터로 부모들의 양육 부담을 덜어주기엔 그 숫자가 턱없이 적은 것이 현실”이라 말했다.
- ☐ 현재 발생하고 있는 문제점들을 해결하기 위하여 공동육아 방식을 도입한 어린이집이 현실적인 대안이 될 수 있다.

## 1.2. 분석 방향 및 목적

### 1.2.1. 분석 방향

#### 1) 미래지향적 보육 운영 모델 ‘서울형 모아어린이집’

- 공립, 민간, 가정 어린이집 간 보육 수급 격차를 해소하고, 어린이집 간 협력과 상생을 지향하며, 저출산에 대해 선제적 대응하고자 2021년부터 서울특별시 일부 자치구에서 시행한 보육 운영 모델



[그림 1-3] 서울형 모아어린이집 추진 체계 (서울시)

#### 2) ‘서울형 모아어린이집’의 성과

- 2021년 성과를 토대로 2022년 서울시 25개 전 자치구에서 시행되고 있으며 현 정부에서 보육 전략으로 서울형 모아어린이집 채택하였다. 본 사업이 전국으로 확산될 수 있도록 리플릿 및 사업 매뉴얼 제작하여 타 시도에 배포할 예정이다. (2022.3.22\_ 연합뉴스\_서울시, '尹 공약' 서울형 모아어린이집 전국 확대 추진(종합))

#### 3) ‘서울형 모아어린이집’ 보육 모델 전국 확대를 통한 인식 개선

- 2022년 서울시 모든 자치구에서 시행되고 있는 공유형 어린이집 ‘서울형 모아어린이집’에서 착안하여 ‘서울형 모아어린이집’에 대한 데이터 분석을 진행할 것이다.

- 이를 통해 새로운 ‘공유 어린이집’ 분류 모델을 제작하여 ‘서울형 모아어린이집’ 보육 지원 사업의 전국 확대 시 더 효율적으로 적용될 수 있도록 한다. 나아가 국가 책임적 공동보육 인식을 확산시키는데 기여할 수 있도록 한다.

#### 4) 현재 서울시에서 시행되고 있는 모아어린이집의 문제점

- (보육 격차 해소 어려움) 공동체를 구성하는 어린이집들의 유형 중 국공립 유형의 비율이 높다면 민간과 국공립 간의 보육 격차를 해소하려는 모아어린이집의 취지를 실현하기 어렵다. [그림1-4]에서의 모아어린이집 공동체의 경우 4개 구성원 중 3곳이 국공립 어린이집임을 확인할 수 있다.



[그림 1-4] 서울형 모아어린이집 예시 (공동체의 국공립 비율이 높은 경우)

- (도보권 이용의 어려움) 어린이집 이용권역으로 통학 거리 10분, 즉 750m를 넘어서지 않는 것이 권장된다. 그러나 서울형 모아어린이집은 현재 자율적인 참여를 기반으로 하고 있어 [그림 1-5]와 같이 어린이집 간의 거리가 먼 경우가 존재한다.

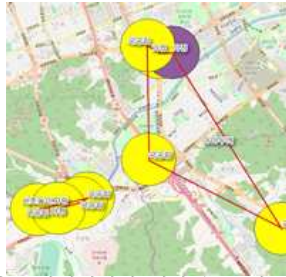


[그림 1-5] 서울형 모아어린이집 예시 (어린이집 간의 거리가 먼 경우)

- (주변 시설 미흡) 모아어린이집의 주목적 중 하나인 생태 친화를 위하여 공동체 주변에 공원, 체육시설과 같은 공간이 필요하다. 마찬가지로 모아어린이집에 대한 자율 참여 방식으로 인해 공동체 주변에 필요 시설이 없더라도 모아어린이집으로 기능할 수 있다.

[그림 1-6]에서 어린이집의 반경 500m내에 공원이 없는 경우를 노란색 버퍼를 통해서 확인 가능

좌측 하단에 위치한 모아어린이집 공동체의 모든 구성원 중 근방에 공원이 존재하는 어린이집은 없다는 점 또한 발견할 수 있다.



[그림 1-6] 서울형 모아어린이집 예시 (공동체 주변 필요시설이 없는 경우)

### 1.2.2. 분석 목적

- 본 분석의 목적은 저출산 해결 방안으로 “아이를 낳으면 사회가 함께 키운다”라는 국가 책임적 공동 보육 강화
- 공동 보육을 통해 변화하는 인구 구조에 대응하며 청년들이 안심하며 일하고 부담 없이 아이를 키울 수 있는 사회 분위기 조성
- 종국엔 저출산으로 인한 국가적 위기 해소 목적

## 1.3. 분석 결과 활용 방안

### 1.3.1. 선정 예측 모델 활용

- '서울형 모아어린이집' 관련 빅데이터 분석을 통해 핵심적인 지리적 특징 정보 정량화
- 모아어린이집 위치 선정 근거 자료로 활용 가능

### 1.3.2. 전국 확대 제안

- 서울형 모아어린이집 분석 내용을 바탕으로 전국 확대에 필요한 지역 및 위치 선정의 실효성 검증 가능
- 지역 예측 모델을 통해 모아어린이집 최적 입지 분석

## 2. 분석 데이터

### 2.1. 분석 데이터 목록

[표 2-1] 분석 데이터

활용 데이터	구분	필수 (Y / N)	생성 주기	활용 방법
도시 공원정보	정형/ 내부	N	연간	관리번호, 공원명, 공원구분, 소재지도로명주소, 소재지번주소, 위도, 경도, 공원면적, 공원보유시설(운동시설,유흥시설,교양시설,기타시설), 지정고시일, 전화번호, 데이터기준일자
유치원알리미 어린이집 일반현황	정형/ 내부	Y	2분기	전국 국립, 공립, 사립 유치원 대상으로 매년 1회이상 주요정보 공시
시도별 아동학대 신고접수 건수	정형/ 외부	Y	연간	시도별, 총계, 응급아동학대 의심사례, 아동학대 의심사례
어린이집 보육현황	정형/ 내부	Y	연간	연령별 인원수, 어린이집 유형
보육아동현황	정형/ 내부	N	연간	시도별 아동인구, 보육상태별 구분, 연령별
서울형 모아어린이집 시행 유치원 정보	정형/ 외부	Y	매년	-자치구, 공동체명, 어린이집명, 주요내용 - 2021년 진행된 '서울형 공유어린이집' 시행 자치구 및 공동체와 어린이집 정보
전국 유아 인구 수	정형/ 내부	Y	연간	-전국 유아 인구 수에 대한 shp 파일
서울시 영유아 인구	정형/ 내부	Y	수시	-연령별, 가구별 -6세미만 영유아 인구수
행정구역(시군구)별 /1세별 주민등록인구	정형/ 외부	Y	수시	-행정구역별, 1세 기준 연령별 총 인구수, 남자 인구수, 여자 인구수
2021년 사회조사보고서	정형/ 내부	Y	연간	-향후 늘려야 할 공공시설 -향후 늘려야 할 복지서비스
어린이집정보공개 포털 인천광역시 어린이집 기본정보	정형/ 내부	Y	수시	-어린이집 기본정보 :어린이집명, 유형, 운영현황, 주소, 경도, 위도, 보육실수, 놀이터수, 등

## 2.2. 데이터 상세 설명

[표 2-2] 데이터 상세 설명

분석데이터	형식	생성주기
도시 공원정보	Excel	연간
유치원알리미	Csv	2분기
어린이집 일반현황	Csv	연간
시도별 아동학대 신고접수 건수	Csv	연간
어린이집 보육현황	Excel	연간
보육아동현황	Excel	연간
서울형 모아어린이집 시행	웹페이지	매년
유치원 정보	shp	연간
전국 유아 인구 수	Excel	수시
서울시 영유아 인구	Csv	수시
행정구역(시군구)별/1세별 주민등록인구	Excel	연간
2021년 사회조사보고서	Csv	수시
어린이집정보공개포털	Excel	수시
인천광역시 어린이집 기본정보		

### 2.2.1. 전체 시/도 어린이집 일반현황[.CSV]

- [유치원 알리미] - [공시자료 다운로드] - [시/도 : 전체 시/도, 공시차수 : 2022년 1차, 공시항목 : 일반현황]
- 교육청명, 교육지원청명, 유치원명, 설립유형, 대표자명, 원장명, 설립일, 개원일, 주소, 전화번호, 홈페이지, 운영시간, 만 3세~5세 학급수, 혼합 학급수, 특수 학급수, 인가총정원, 3세~5세 정원수, 혼합모집정원수, 특수학급모집정원수, 만 3세~5세 유아수, 혼합 유아수, 특수 유아수 데이터 포함

	A	B	C	D	E
1	행정구역(시군구)별	연령별	2022.06	2022.06	2022.06
2	행정구역(시군구)별	연령별	총인구수 (명)	남자인구수 (명)	여자인구수 (명)
3	충청북도	계	1597118	811127	785991
4	충청북도	0세	7594	3860	3734
5	충청북도	1세	8477	4319	4158
6	충청북도	2세	8948	4643	4305
7	충청북도	3세	10033	5204	4829

[그림 2-1] 전체 시/도 어린이집 일반현황

### 2.2.2. 행정구역(시군구)별/1세별 주민등록인구[.CSV]

- [KOSIS 국가통계포털] - [인구] - [주민등록인구현황] - [행정구역(시군구)별/1세별 주민등록인구]
- 행정구역(시군구)별, 연령별, 총인구수(명), 남자 인구수(명), 여자 인구수(명) 데이터 포함

	A	B	C	D	E
1	행정구역(시군구)별	연령별	2022.06	2022.06	2022.06
2	행정구역(시군구)별	연령별	총인구수 (명)	남자인구수 (명)	여자인구수 (명)
3	충청북도	계	1597118	811127	785991
4	충청북도	0세	7594	3860	3734
5	충청북도	1세	8477	4319	4158
6	충청북도	2세	8948	4643	4305
7	충청북도	3세	10033	5204	4829
1835	강원도	97세	414	100	314
1836	강원도	98세	297	86	211
1837	강원도	99세	235	60	175
1838	강원도	100세 이상	354	74	280

[그림 2-2] 행정구역(시군구)별/1 세별 주민등록인구

### 2.2.3. 시도별 아동학대 신고접수 건수[.CSV]

- [KOSIS 국가통계포털] - [범죄·안전] - [학대피해아동보호현황] - [시도별 신고접수 건수]
- 시도별, 총계, 응급아동학대 의심사례, 아동학대 의심사례 데이터 포함

	A	B	C	D
1	시도별(1)	2020	2020	2020
2	시도별(1)	계	응급아동학대 의심사례	아동학대 의심사례
3	강원	1404	51	1353
4	경기	9192	564	8628
5	경남	1743	40	1703
6	경북	1875	96	1779
7	광주	858	11	847
8	대구	1719	66	1653

[그림 2-3] 시도별 아동학대 신고접수 건수

### 2.2.4. 향후 늘려야 할 공공시설 설문조사에서의 국·공립 어린이집 유치원 응답 비율[.EXCEL]

- [2021 사회조사보고서] - [복지 3-1]
- 향후 늘려야 할 공공시설 설문에서의 전체 선택지 :
  - ①국·공립 어린이집 유치원 ②공원, 녹지, 산책로 ③보건 의료 시설
  - ④사회 복지 시설 ⑤공영 주차 시설 ⑥문화 예술 회관 ⑦도서관
  - ⑧생활 체육 시설 ⑨기타



주: 1) 2021년부터 '체육시설 및 경기장' → '생활체육시설'로 변경  
Note: 1) Since 2021 survey item has been changed.  
(Sports facilities or stadiums → Recreational sports facilities)

**2-1. Public Facilities Requiring Expansion (Main response)**

(Unit: %)

	계	국립 어린이 유치원	공립 놀이 시설	보건 의료 시설	사회 복지 시설	공영 주차 시설	문화 예술 회관	도서관	생활 체육 시설	기타
	Total	Public daycare Centers	Park & amusement parks	Health or medical facilities	Social welfare facilities	Public parking facilities	Cultural centers	Libraries	Recreational sports facilities	Others
서울	100.0	14.0	22.2	20.9	14.1	12.3	4.7	4.3	7.4	-
남자	100.0	14.5	21.6	19.6	12.7	14.9	4.6	4.2	7.9	-
여자	100.0	13.7	22.7	22.1	15.3	10.0	4.8	4.4	7.0	-
부산	100.0	9.6	22.5	23.3	13.8	12.0	5.4	4.0	9.4	0.1**
남자	100.0	9.7	19.0	22.6	12.4	15.9	4.2	4.1	12.1	-
여자	100.0	9.5	25.8	23.9	15.1	8.3	6.5	3.9	6.8	0.2**
대구	100.0	10.7	19.1	22.2	15.6	12.1	4.7	4.8	10.1	0.6
남자	100.0	9.9	16.0	21.2	14.8	15.6	4.2	4.5	13.4	0.6
여자	100.0	11.5	22.1	23.2	16.3	8.9	5.3	5.1	7.0	0.6
인천	100.0	12.1	22.6	22.2	15.4	13.5	4.3	1.8	8.1	0.1**
남자	100.0	9.6	21.1	21.2	14.5	17.6	5.4	1.2	9.4	-
여자	100.0	14.5	24.1	23.1	16.3	9.4	3.3	2.4	6.7	0.2**

[그림 2-4] 향후 늘려야 할 공공시설 설문조사

## 2.2.5. 향후 늘려야 할 복지서비스 설문조사에서의 보육서비스 응답 비율[.EXCEL]

□ 향후 늘려야 할 복지서비스 설문에서의 전체 선택지 :

- ①고용(취업) 지원 서비스 ②소득 지원 서비스 ③주거 관련 서비스  
④안전 관련 서비스 ⑤보건 의료·건강 관리 서비스 ⑥보육 관련 서비스  
⑦교육 관련 서비스 ⑧노인 돌봄 서비스 ⑨문화·여가 생활 지원 서비스  
⑩기타

주: 1) 2021년부터 '문화·여가 생활 지원 서비스' → '문화·여가 생활 지원 서비스'로 변경  
Note: 1) Since 2021 survey item has been changed. (Culture and leisure support services → Culture and leisure support services)

**3-1. 향후 늘려야 할 복지서비스 (주된 응답) - 계속**

3-1. Welfare Services Requiring Expansion (Main response) - Continued

	계	고용·취업 지원 서비스	소득 지원 서비스	주거 관련 서비스	안전 관련 서비스	보건·의료· 건강 관리 서비스	보육 관련 서비스	교육 관련 서비스	노인 돌봄 서비스	문화·여가 생활 지원 서비스	기타
	Total	Employment support services	Income support services	Housing related services	Safety related services	Medical and health care services	Childcare related services	Education related services	Elderly care services	Culture and leisure support services	Others
서울	100.0	28.8	13.2	17.7	4.4	18.0	4.6	3.1	9.1	1.0	0.1
남자	100.0	27.7	12.6	17.7	4.3	17.7	4.0	3.0	9.4	1.2	0.1
여자	100.0	29.9	13.8	17.7	4.5	18.3	5.1	3.2	8.8	0.8	0.1
부산	100.0	14.4	17.1	18.9	7.2	22.1	2.0	3.0	9.8	2.6	0.1
남자	100.0	16.2	17.1	19.2	6.8	19.9	3.7	2.7	10.7	2.9	0.1
여자	100.0	12.6	17.1	18.6	7.6	24.3	0.3	3.3	8.9	2.3	0.1
대구	100.0	16.2	15.7	19.7	9.9	15.1	4.1	3.4	9.0	6.3	0.1
남자	100.0	15.9	15.7	19.7	9.9	15.1	4.1	3.4	9.0	6.3	0.1
여자	100.0	16.5	15.7	19.7	9.9	15.1	4.1	3.4	9.0	6.3	0.1
인천	100.0	20.6	15.9	15.1	4.7	19.4	2.4	3.2	8.8	4.5	0.1
남자	100.0	20.6	15.9	15.1	4.7	19.4	2.4	3.2	8.8	4.5	0.1
여자	100.0	20.6	15.9	15.1	4.7	19.4	2.4	3.2	8.8	4.5	0.1

[그림 2-5] 향후 늘려야 할 복지서비스 설문조사

## 2.2.6. 인천광역시 어린이집 기본정보[.EXCEL]

- [어린이집 정보 공개 포털] - [보육정보공개API] - [Open API] - [어린이집 기본정보] - [기준조건 : 비정기(수시), 어린이집 지역 : 인천광역시]
- 시도, 시군구, 어린이집명, 어린이집유형구분, 운영현황, 우편번호, 주소, 어린이집전화번호, 어린이집팩스번호, 보육실수, 보육실면적, 놀이터수, 보육교직원수, 정원수, 현원수, 위도, 경도, 통학차량운영여부, 홈페이지주소, 인가일자, 휴지시작일자, 휴지종료일자, 폐지일자 데이터 포함

주: 1) 2021년부터 '체육시설 및 경기장' → '생활체육시설'로 변경  
Note: 1) Since 2021 survey item has been changed.  
(Sports facilities or stadiums → Recreational sports facilities)

**2-1. Public Facilities Requiring Expansion (Main response)**

(Unit: %)

	계	국립 어린이 유치원	공립 놀이 시설	보건 의료 시설	사회 복지 시설	공영 주차 시설	문화 예술 회관	도서관	생활 체육 시설	기타
	Total	Public daycare Centers	Park & amusement parks	Health or medical facilities	Social welfare facilities	Public parking facilities	Cultural centers	Libraries	Recreational sports facilities	Others
서울	100.0	14.0	22.2	20.9	14.1	12.3	4.7	4.3	7.4	-
남자	100.0	14.5	21.6	19.6	12.7	14.9	4.6	4.2	7.9	-
여자	100.0	13.7	22.7	22.1	15.3	10.0	4.8	4.4	7.0	-
부산	100.0	9.6	22.5	23.3	13.8	12.0	5.4	4.0	9.4	0.1**
남자	100.0	9.7	19.0	22.6	12.4	15.9	4.2	4.1	12.1	-
여자	100.0	9.5	25.8	23.9	15.1	8.3	6.5	3.9	6.8	0.2**
대구	100.0	10.7	19.1	22.2	15.6	12.1	4.7	4.8	10.1	0.6
남자	100.0	9.9	16.0	21.2	14.8	15.6	4.2	4.5	13.4	0.6
여자	100.0	11.5	22.1	23.2	16.3	8.9	5.3	5.1	7.0	0.6
인천	100.0	12.1	22.6	22.2	15.4	13.5	4.3	1.8	8.1	0.1**
남자	100.0	9.6	21.1	21.2	14.5	17.6	5.4	1.2	9.4	-
여자	100.0	14.5	24.1	23.1	16.3	9.4	3.3	2.4	6.7	0.2**

[그림 2-6] 인천광역시 어린이집 기본정보

## 2.3. 데이터 정제 방안

## 2.3.1. 지역 채택을 위한 시도별 통계 데이터

## 1) 전체 시/도 어린이집 일반현황

- 설립유형 국립, 사립에 따른 어린이집 수를 구하기 위해, 유치원명, 주소, 설립유형 컬럼 불러오기
- '설립유형'의 하위 범주(단설, 병설, 사인, 국립)를 제외한 상위 범주만 이용하여 위해 문자열 추출, '설립유형' 컬럼에 값 치환
- 주소에서 행정구역을 가져오기 위해 띄어쓰기를 기준으로 첫번째 문자열 추출, '행정구역(시군구)별' 컬럼 생성
- '행정구역(시군구)별', '설립유형' 컬럼을 이용하여 그룹화
- pivot 함수를 이용하여 컬럼 기준을 '설립유형', 행 기준을 '행정구역(시군구)별', 값 영역을 '유치원명'으로 설정
- 행정구역별 국립 어린이집 수, 사립 어린이집 수, 총 어린이집 수 확인

## 2) 행정구역(시군구)별/1세별 주민등록인구

- '연령별' 컬럼에서 숫자만 추출하기 위해 정규 표현식 이용
- NaN 값 포함한 데이터 및 전국 주민등록인구 데이터 제거
- '연령별' 컬럼에 대해 AND 연산자를 이용한 조건식 적용하여 0세, 1세, 2세, 3세, 4세, 5세 인구 수 데이터 추출
- '행정구역(시군구)별' 컬럼을 이용하여 그룹화
- 행정구역별 0세 이상 5세 이하 인구 총합 확인

## 3) 시도별 아동학대 신고접수 건수

- NaN 값 포함한 데이터 및 전국 통계 데이터 제거
- 행정구역별 읍면동학대 의심사례, 아동학대 의심사례를 합한 값을 아동학대 건수로 설정



## 4) 2021 사회조사보고서

- ☐ 향후 늘려야 할 공공시설로 국공립 어린이집을 늘려야 한다고 응답한 비율 수집
- ☐ 향후 늘려야 할 복지서비스로 보육서비스를 늘려야 한다고 응답한 비율 수집

## 5) 지역 채택을 위한 데이터프레임 생성

- ☐ 1) ~ 4)에서 구한 변수 : 행정구역(시군구)별, 어린이집 수, 아동인구 수, 국립, 사립, 아동학대 건수, 향후 늘려야 할 공공시설(국공립어린이집), 향후 늘려야 할 복지 서비스(보육 관련 서비스)
- ☐ 신규 파생 변수 : 국립/사립, 아동학대건수/아동인구수, 아동인구수/어린이집수

행정구역(시군구)별	어린이집	아동인구 (명)	국립	사립	아동학대	향후 늘려야 할 공공시설(국공립어린이집)	향후 늘려야 할 복지 서비스(보육 관련 서비스)	국립/사립	아동학대건수/아동인구수	아동인구수/어린이집수
0 강원도	350	51693	271	79	1404	9.9	3.9	3.43	0.0272	147.69
1 경기도	2050	150294	1220	830	9192	13.7	3.9	1.47	0.0167	266.44
2 경상남도	649	117284	422	227	1743	11.6	3.5	1.86	0.0149	180.71
3 경상북도	637	88760	432	205	1875	8.2	3.7	2.11	0.0216	136.20
4 광주광역시	269	55356	126	143	858	10.1	3.7	0.88	0.0155	205.78
5 대구광역시	311	81271	111	200	1719	10.7	3.8	0.56	0.0212	261.32
6 대전광역시	245	53472	103	142	1645	11.5	3.6	0.73	0.0308	218.25
7 부산광역시	371	107187	135	236	2112	9.6	2.9	0.57	0.0167	288.91
8 서울특별시	754	253910	292	462	4167	14.0	4.6	0.63	0.0142	389.80
9 세종특별자치시	63	24614	61	2	371	11.7	5.4	30.50	0.0151	390.70
10 울산광역시	180	44167	93	87	1402	9.6	4.0	1.07	0.0317	245.37
11 인천광역시	385	108998	187	198	3099	12.1	2.5	0.94	0.0285	282.85
12 전라남도	475	60701	388	87	2398	8.9	3.3	4.46	0.0390	127.79
13 전라북도	464	55976	341	123	2453	9.8	3.9	2.77	0.0438	120.64
14 제주특별자치도	118	28308	101	17	859	10.0	4.7	5.94	0.0303	239.90
15 충청남도	482	80068	363	119	2374	9.3	3.6	3.05	0.0296	166.12
16 충청북도	308	58191	237	71	1288	9.1	4.0	3.34	0.0221	188.93

[그림 2-7] 지역 채택 데이터셋

## 2.3.2. 인천광역시 어린이집 데이터

- ☐ 2,803개의 데이터를 가진 인천광역시 어린이집 기본정보를 raw data로 이용
- ☐ raw data의 우편번호, 어린이집 전화번호, 어린이집 팩스 번호, CCTV 설치수, 위도, 경도, 홈페이지 주소, 인가일자, 휴지시작일자, 휴지종료일자, 폐지일자 컬럼 삭제
- ☐ 운영 현황이 '폐지'인 어린이집 데이터 및 null 값 포함한 데이터 삭제

☐ boxplot, barplot, histogram 확인 \*부록 7.2. 참고

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	시도	시군구	어린이집명	어린이집유형구분	공영형태	주소	보육장소	보육장명	보육장소재지	보육장소재지	보육장소재지	보육장소재지	보육장소재지
2	인천광역시	중구	한솔어린이집	국공립	재계	인천광역시 중구 한솔대로861번길 47 1층 1015호	3	55	0	1	20	5	미공영
3	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	3	59	0	5	22	6	미공영
4	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	10	261	2	12	12	12	미공영
5	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	3	59	0	5	22	6	미공영
6	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	4	89	1	5	30	7	미공영
7	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	2	208	1	5	35	10	미공영
8	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	6	165	1	12	75	12	미공영
9	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	6	165	1	6	12	12	미공영
10	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	10	402	1	24	152	154	미공영
11	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	4	295	1	8	42	15	미공영
12	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	8	248	2	15	91	60	미공영
13	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	11	448	2	18	122	86	미공영
14	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	3	82	0	5	18	9	미공영
15	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	3	85	0	5	20	10	미공영
16	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	3	84	0	6	19	15	미공영
17	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	3	80	0	3	30	6	미공영
18	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	40	440	2	16	111	77	미공영
19	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	5	86	0	8	20	18	미공영
20	인천광역시	동구	한솔어린이집	민간	재계	인천광역시 동구 한솔로 17 (한솔동)	2	117	0	5	39	15	미공영

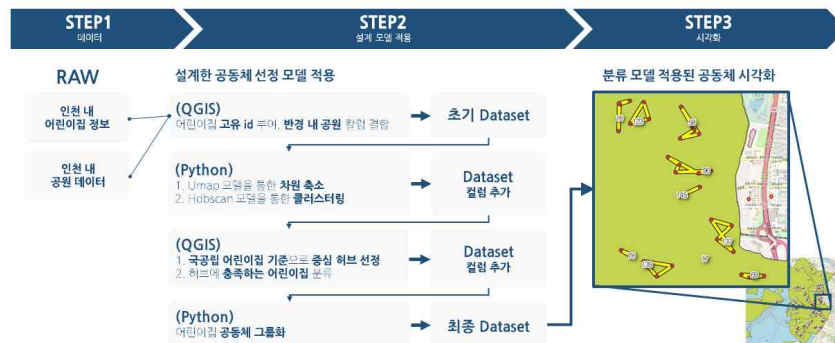
[그림 2-8] 인천 어린이집 데이터셋

### 3. 분석 프로세스

#### 3.1. 분석 프로세스

##### 1) 데이터 분석 및 모델링

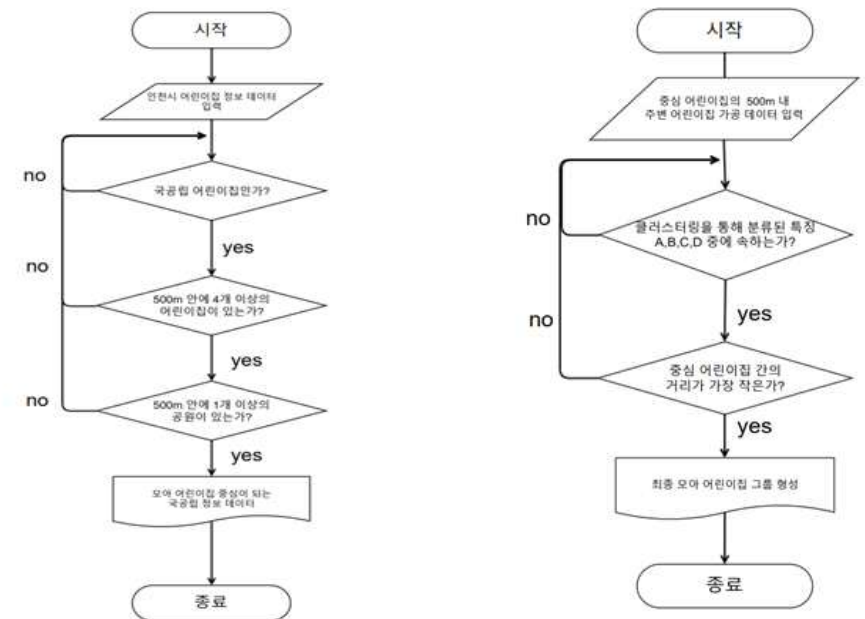
- ☐ 6가지 기준을 적용하여 우선적으로 시범 운영할 지역 탐색 -> 인천 지역 채택
- ☐ 인천광역시의 어린이집 통합 데이터에 대해 umap 분석 방법, hdbscan 분석 방법을 적용해 어린이집 군집화
- ☐ 모아 어린이집 공동체의 중심으로써 안정적으로 기능할 국공립어린이집 확인
- ☐ 소속 군집별로 국공립어린이집으로부터 최소 거리인 어린이집 추출



[그림 3-1] 분석 프로세스

##### 2) 결과 시각화 및 최적 모아 어린이집 공동체 선정

- ☐ 인천광역시 모아 어린이집 공동체 위치 시각화



[그림 3-2] 중심 모아어린이집 선정 순서도

[그림 3-3] 모아어린이집 그룹 생성 순서도

## 3.2. 분석 내용 및 방법

### 3.2.1. 시범 운영 지역 탐색

#### 1) 적용할 기준 6 가지 선정

##### ① 사립 대비 국립 어린이집 수

- \* 모아 어린이집의 시행 목적인 국공립 어린이집과 민간 어린이집 간 불균형 해소를 고려하여, 사립/국립 비율이 낮은 지역에 우선순위 설정

##### ② 향후 늘려야 할 공공시설로 국공립 어린이집을 늘려야 한다는 비율

##### ③ 향후 늘려야 할 복지서비스로 보육 서비스를 늘려야 한다는 비율

##### ④ 아동 인구수 대비 아동학대 피해 건수

- \* 어린이집 간 교류를 통해 아동학대를 예방하기 위한 차원에서 아동학대 피해 비율이 높은 지역에 우선순위 설정

##### ⑤ 아동 인구수

##### ⑥ 어린이집 수 대비 아동 인구수

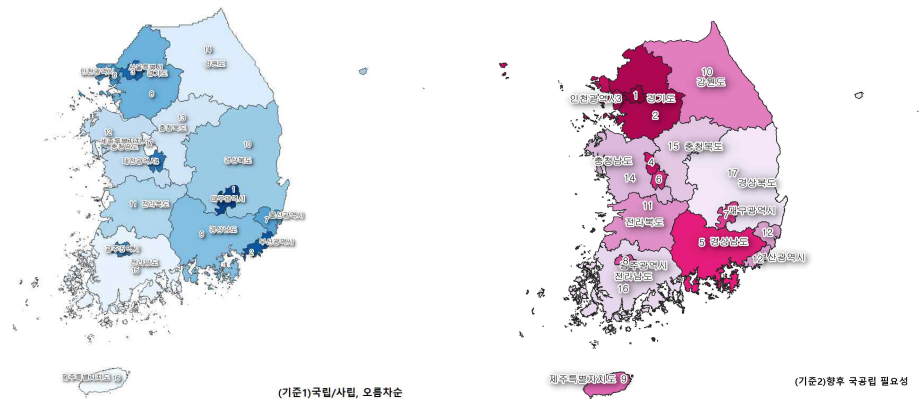
#### 2) 시도별 기준 6 가지에 대한 수치 데이터 확인 및 시각화

구분	A	C	D	E	F	G	H	I	J	K
1	행정구역(시군구)별	아동인구(만)	국립	사립	아동학대	향후 늘려야 할 공공시설(국공립어린이집)	향후 늘려야 할 복지 서비스(보육 관련 서비스)	국립/사립	아동학대건수/아동인구수	아동인구수/어린이집 수
2	서울특별시	293910	292	462	4167	14	6.6	0.61	0.0142	389.8
3	경기도	555094	1220	830	9197	13.7	3.9	1.47	0.0147	268.44
4	대구광역시	252719	111	200	1719	10.7	3.8	0.96	0.0212	209.32
5	인천광역시	508698	407	198	3699	12.1	2.5	0.94	0.0205	252.05
6	대전광역시	53472	103	142	1645	11.5	3.6	0.73	0.0308	218.25
7	울산광역시	44167	91	87	1402	9.6	4	1.07	0.0317	245.37
8	부산광역시	107187	135	236	2112	9.6	2.9	0.57	0.0197	288.91
9	세종특별자치시	24614	65	2	371	11.7	5.4	80.5	0.0518	390.7
10	제주특별자치도	28208	101	17	859	10	4.7	3.94	0.0383	239.9
11	전라북도	55976	341	525	2451	9.9	3.9	2.77	0.0448	120.64
12	전라남도	117284	423	227	1741	11.6	3.5	1.86	0.0149	180.71
13	경주광역시	55356	126	143	958	10.1	3.7	0.88	0.0555	205.08
14	충청북도	58191	217	71	1288	9.1	4	3.14	0.0221	188.93
15	충청남도	80068	163	119	2574	9.3	3.6	1.05	0.0296	166.12
16	강원도	51693	271	79	1804	9.9	3.9	3.41	0.0272	147.69
17	울산광역시	46360	432	205	1871	8.2	3.7	2.15	0.0276	156.2
18	전라남도	40701	188	87	2368	8.9	3.3	4.46	0.0339	127.29

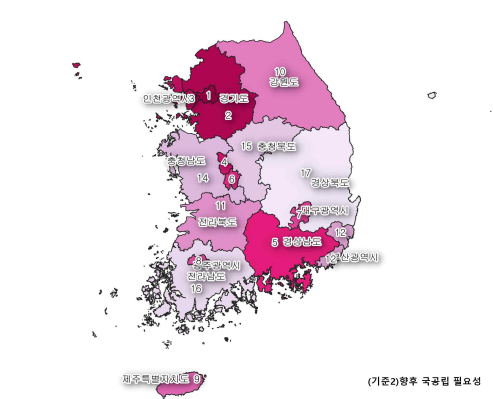
[그림 3-4] 시도별 기준에 대한 데이터셋



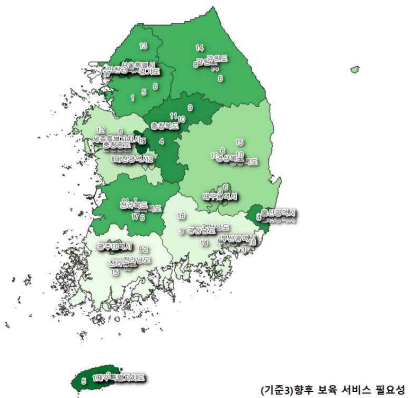
[그림 3-5] 시도별 기준 시각화



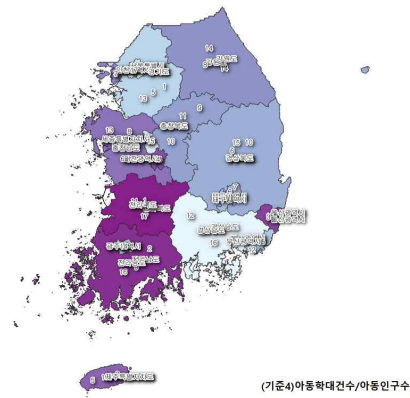
[그림 3-6] 시도별 기준 1 시각화



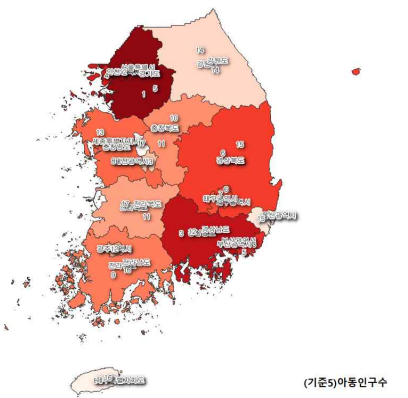
[그림 3-7] 시도별 기준 2 시각화



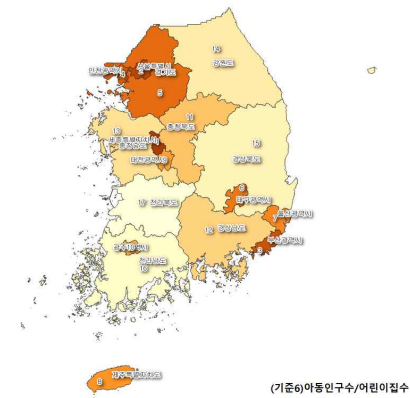
[그림 3-8] 시도별 기준 3 시각화



[그림 3-9] 시도별 기준 4 시각화



[그림 3-10] 시도별 기준 5 시각화



[그림 3-11] 시도별 기준 6 시각화

3) Python rank 함수를 이용해 시도별 기준 6 가지에 대한 수치 데이터로 각 기준에서의 순위를 파악한다. 다음으로 각 기준에서의 순위 점수를 합하여 시도별 종합 점수를 책정한다. 종합 점수를 오름차순 기준으로 rank 함수를 적용해 종합 순위를 도출하여 가장 우선적으로 시범 운영할 지역을 채택한다.

#### 4) 시도별 기준 6 가지에 대한 순위 확인

##### □ (기준1) 사립 대비 국립 어린이집 수

상위 4개 행정구역 : 대구(1) - 부산(2) - 서울(3) - 대전(4)

하위 4개 행정구역 : 세종(17) - 제주(16) - 전남(15) - 강원(14)

##### □ (기준2) 향후 늘려야 할 공공시설로 국립 어린이집을 늘려야 한다는 비율

상위 4개 행정구역 : 서울(1) - 경기(2) - 인천(3) - 세종(4)

하위 4개 행정구역 : 경북(17) - 전남(16) - 충북(15) - 충남(14)

##### □ (기준3) 향후 늘려야 할 복지서비스로 보육서비스를 늘려야 한다는 비율

상위 4개 행정구역 : 세종(1) - 제주(2) - 서울(3) - 울산(4)

하위 4개 행정구역 : 인천(17) - 부산(16) - 전남(15) - 경남(14)

##### □ (기준4) 아동 인구수 대비 아동학대 피해건수

상위 4개 행정구역 : 전북(1) - 전남(2) - 울산(3) - 대전(4)

하위 4개 행정구역 : 서울(17) - 경남(16) - 세종(15) - 광주(14)

##### □ (기준5) 아동 인구수

상위 4개 행정구역 : 경기(1) - 서울(2) - 경남(3) - 인천(4)

하위 4개 행정구역 : 세종(17) - 제주(16) - 울산(15) - 강원(14)

##### □ (기준6) 어린이집 수 대비 아동 인구수

상위 4개 행정구역 : 세종(1) - 서울(2) - 부산(3) - 인천(4)

하위 4개 행정구역 : 전북(17) - 전남(16) - 경북(15) - 강원(14)

## 5) QGIS 를 활용해 시도별 각 기준에서의 순위 시각화

	A	L	M	N	O	P	Q
1	행정구역(시군구)별	(기준1)국립/사립	(기준2)향후 국공립 필요성	(기준3)향후 보육 서비스 필요성	(기준4)아동학대건수/아동인구수	(기준5)아동인구수	(기준6)아동인구수/어린이집수
2	서울특별시	3	1	1	17	2	2
3	경기도	8	2	6	13	1	5
4	대구광역시	1	7	9	18	7	6
5	인천광역시	6	3	17	7	4	4
6	대전광역시	4	6	12	4	13	9
7	울산광역시	7	12	4	3	15	7
8	부산광역시	2	12	16	12	5	3
9	세종특별자치시	17	4	1	15	17	1
10	제주특별자치도	16	9	2	5	16	8
11	전라북도	11	11	6	1	11	17
12	경상남도	9	5	14	16	3	12
13	광주광역시	5	8	10	14	12	10
14	충청북도	13	15	4	9	10	11
15	충청남도	12	14	12	6	8	13
16	강원도	14	10	6	8	14	14
17	경상북도	10	17	10	10	6	15
18	전라남도	15	16	15	2	9	16

[그림 3-12] 시도별 기준 순위 데이터

- 인천광역시 어린이집 중심으로 500m버퍼 형성
- 폴리곤에 포함하는 포인트 개수 계산으로 인천광역시 어린이집 주변 공원 수 도출
- 공원수 데이터를 포함한 포인트 레이어 형성
- 인천광역시 어린이집 데이터에 주변 공원 수 컬럼 추가

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	시도	시군구	어린이집문어린이집문 운영현황	주소	보육실수	보육실면적	보육실면적/인구	보육교직원정원수	현원수	총학자정수	500m내 공원수	id			
2	인천광역시 중구	송현어린이집	재개	인천광역시	3	59	0	5	22	6 미운영	3	1			
3	인천광역시 중구	천송어린이공유집	재개	인천광역시	3	55	0	3	20	5 미운영	0	2			
4	인천광역시 연수구	빛송어린이집	재개	인천광역시	4	89	1	5	30	7 운영	1	3			
5	인천광역시 미추홀구	파란돌어린이집	재개	인천광역시	7	370	2	9	84	32 운영	4	4			
6	인천광역시 남동구	아이꿈어린이집	재개	인천광역시	4	93	0	3	35	10 미운영	4	5			
7	인천광역시 남동구	봄소리어린이집	재개	인천광역시	7	208	1	12	75	52 운영	4	6			
8	인천광역시 남동구	한솔빛어린이집	재개	인천광역시	10	402	1	24	152	134 운영	1	7			
9	인천광역시 남동구	다솜어린이집	재개	인천광역시	6	295	1	8	82	31 운영	6	8			
10	인천광역시 남동구	이삭어린이가정	재개	인천광역시	3	84	0	6	19	15 운영	7	9			
11	인천광역시 남동구	고운어린이가정	재개	인천광역시	3	65	0	5	20	10 미운영	7	10			

[그림 3-14] 인천 어린이집 데이터셋

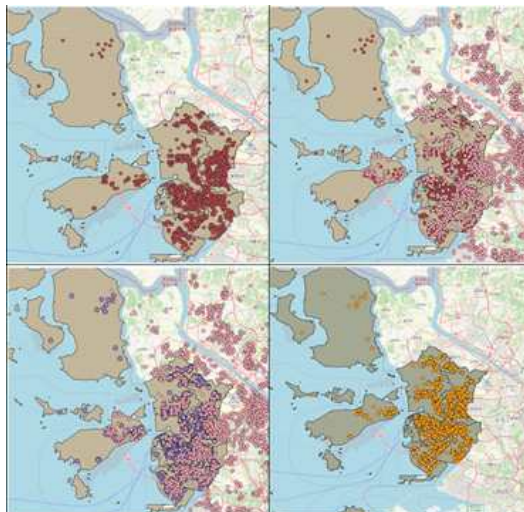
## 2) 차원축소

## (1) 클러스터링 전 차원축소를 진행하는 이유

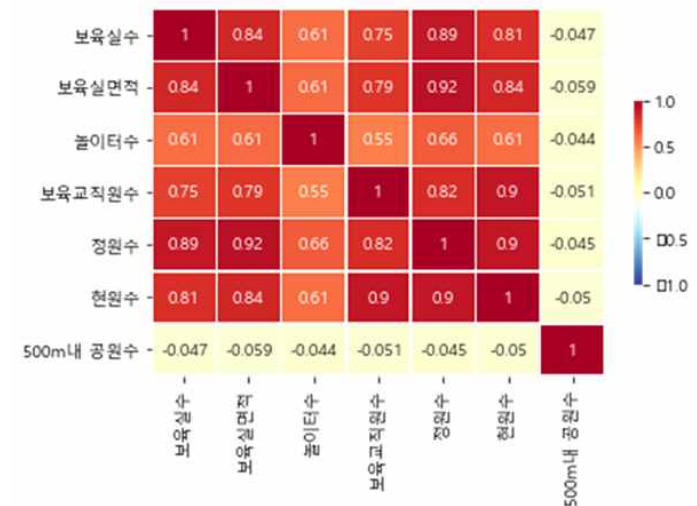
차원이 커지거나 독립변수 간 강한 상관관계가 있는 경우, 모델의 정확도가 떨어질 수 있다. [그림3-15]를 보면 독립변수 간에 강한 상관관계를 가지고 있다는 것을 확인 가능. 따라서 차원축소를 이용하여 여러 변수의 정보를 최대한 유지하며 데이터 변수의 개수를 줄이는 방법이 필요

## 3.2.2. 인천광역시 어린이집 분석

## 1) 인천광역시 어린이집 데이터에 어린이집 고유 id 및 어린이집 주변 공원 수 컬럼 추가



[그림 3-13] 인천 어린이집 시각화



[그림 3-15] 차원축소 모델



## (2) 전처리 과정

- sklearn.preprocessing.LabelEncoder을 이용하여 '어린이집유형구분' 변수 라벨링 변화
- '통학차량운여부' 변수를 더미변수로 바꿔 0과1로 표현
- sklearn.preprocessing.MinMaxScaler을 이용하여 연속형 변수들의 스텔일을 변화시킴
- 범주형 데이터를 라벨링

## (3) UMAP 차원축소 모델

## ■ PCA와 T-SNE 차원축소 모델

- 차원축소로 많이 사용되는 주성분분석(PCA)의 경우 선형 변환으로 데이터를 저차원으로 임베딩하여 가시화할 때 결측치 정보가 손실되는 경우가 생긴다. 또한 선형 가정을 하기 때문에 범주형 변수를 처리할 수 없다.
- T-SNE는 비선형적인 차원축소 방법이다. 기본적인 원리는 고차원의 데이터 간 거리를 저차원으로 축소하였을 때도 똑같이 유지. 점들의 유사도를 기반으로 축소하며, 점들의 유사도는 A를 중심으로 한 정규 분포에서 확률 밀도에 비례하여 이웃을 선택하면 포인트 A가 포인트 B를 이웃으로 선택한다는 조건부 확률로 계산된다.

## ■ UMAP 모델 개념

- Uniform Manifold Approximation and Projection(UMAP)은 t-SNE와 유사하게 시각화에 사용할 수 있으며 일반적인 비선형 차수 감소에도 사용할 수 있는 차원축소 기법이다.
- UMAP의 컨셉 또한 고차원의 데이터 간 거리를 저차원으로 축소하였을 때 동일하게 유지하려는 것이므로 t-SNE와 매우 유사하다.
- UMAP이 사용하는 방법은 t-SNE와 전혀 다른 수학적 원리를 기반으로 한다. 이러한 수학적 기반을 이용하여 t-SNE보다 더 global한 데이터 분포를 학습할 수 있다. 최종적으로 UMAP 모델을 이용하여 차원축소를 진행했다.

## ■ UMAP과 hdbscan

- 임베딩 차원에서도 잘 확장이 된다. 단순히 시각화를 위해서가 아니라 다른 기계 학습 작업에 대한 예비 단계로 범용 축소 기술로 사용할 수 있다. 특히 hdbscan 클러스터링과 잘 협력한다.

## (4) UMAP 파라미터

## ■ N\_neighbor

- 데이터의 다양체 구조를 학습하려고 할 때 UMAP이 살펴볼 로컬 이웃의 크기를 제한하여 데이터의 로컬 구조와 글로벌 구조의 균형을 조정한다.
- N\_neighbor에 낮은 값을 전달한다면 UMAP은 매우 국소적인 구조에 집중하는 반면, 큰 값을 전달할 시 데이터의 다양체 구조를 추정할 때 UMAP으로 하여금 더 많은 수의 이웃을 보게 만든다.
- 넓은 범위를 얻는 대신 세부 구조를 잃게 된다. N\_neighbor의 기본값은 15이며, 2에서 200까지의 범위 내에서 설정한다.

## ■ Min\_dist

- 저차원 공간에서 포인트 간의 최소 거리를 제어한다. Min\_dist의 값이 낮다면 포인트가 촘촘하게 무리짓게 되며, 값이 크면 상대적으로 느슨하게 퍼져 있게 된다.
- 군집화 또는 더 세밀한 위상 구조에 관심이 있는 경우 유용하다. 값을 크게 설정하면 UMAP은 점들을 함께 채우는 것을 방해하고 대신 넓은 위상 구조의 보존에 초점을 맞추기 때문이다. 기본값은 0.1이며 0.0에서 0.99까지의 범위 내에서 설정한다.

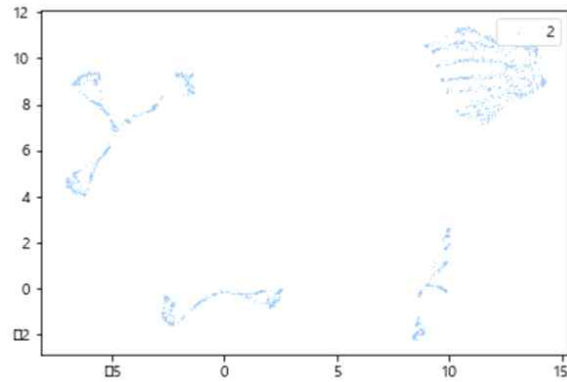
## ■ N\_components

- 사용자가 데이터를 내장하는 축소 차원 공간의 차원성을 설정하는 옵션이다. t-SNE와 같은 일부 다른 시각화 알고리즘과 달리 UMAP은 임베딩 차원에서 잘 확장되므로 2차원 또는 3차원 시각화에만 사용할 수 있다.

- 매우 국소적인 구조에 집중하기보다 많은 수의 이웃을 보게 하여 넓은 범위를 얻기 위해 n\_neighbors=50으로 설정하였고, 포인트 간의 최소 거리는 기본값인 0.1로 설정하여 차원축소를 진행하였다.



## (5) 어린이집 특징을 차원축소한 결과



[그림 3-16] umap 을 이용하여 2 차원 공간에 어린이집 데이터 시각화

## 3) 클러스터링(군집화)

## (1) Hdbscan 군집화 모델 선택 이유

- ☐ HDBSCAN은 탐색적 데이터 분석에 이상적이며, 의미 있는 클러스터(있는 경우)를 반환할 수 있는 빠르고 강력한 알고리즘이며, 앞에서 진행한 umap 차원축소 모델과 잘 협력하는 모델이므로 어린이집 데이터를 특징에 맞게 군집화하기 위한 모델로 선택하였다.

## (2) Hdbscan 군집화 모델

## ■ Hdbscan 모델 개념

- ☐ 비지도 학습모델이며, DBSCAN 알고리즘과 같은 밀도(density) 기반 클러스터링 모델이다.
- ☐ 클러스터별로 다양한 밀도를 가질 수 있으며 Single-linkage 클러스터링을 이용하여 클러스터 간 위계(hierarchy)를 구성, 더 높은 유연성을 제공한다.

## ■ DBSCAN 알고리즘

- ① 밀도/희소에 따라 공간을 변환
- ② 거리 가중 그래프의 최소 스패닝 트리를 만들기
- ③ 연결된 구성 요소의 클러스터 계층을 구성
- ④ 최소 클러스터 크기를 기준으로 클러스터 계층을 압축

## ⑤ 응축된 트리에서 안정적인 클러스터를 추출

## (3) Hdbscan 파라미터

## ■ min\_samples

- ☐ 핵심 포인트에 대한 최소 인접 이웃 수를 설정하는 파라미터이다. 값이 높을수록 더 많은 점이 노이즈/이상치로 삭제됨

## ■ min\_cluster\_size

- ☐ 최종 군집이 될 수 있는 최소 크기이다. 값이 높을수록 군집이 커진다.

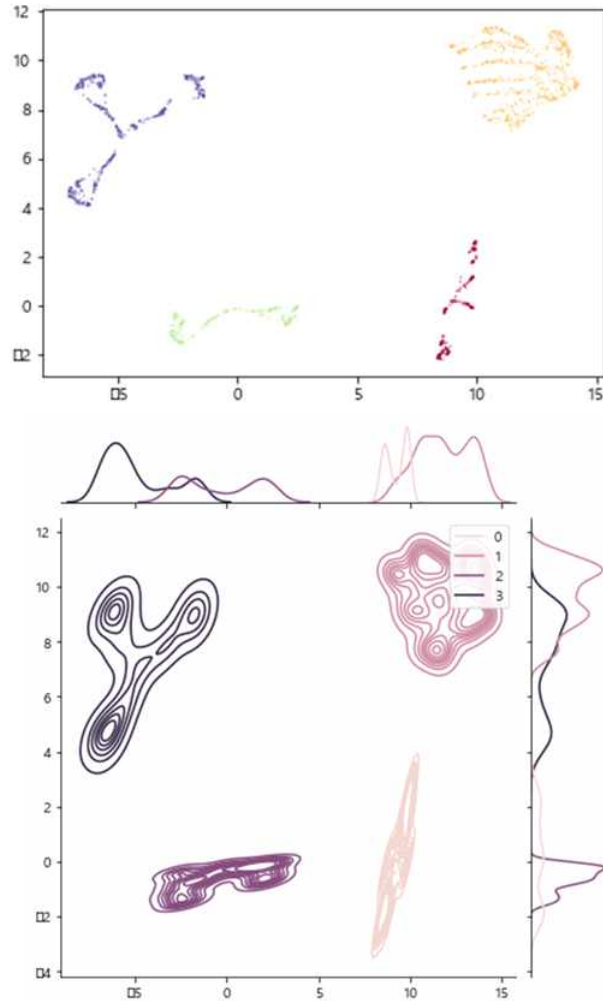
## ■ min\_samples과 min\_cluster\_size 관계

- ☐ 상대적으로 min\_samples를 작게 유지하면서 min\_cluster\_size를 증가시키면 특이치는 유지되지만 대신 모든 클러스터가 min\_cluster\_size보다 클 때까지 가장 유사한 인접 클러스터와 작은 클러스터를 병합시킨다.
- ☐ 특정 클러스터를 많이 사용하려면 작은 min\_sample과 작은 min\_cluster\_size를 사용
- ☐ 일반화된 클러스터를 원하는 동시에 대부분의 세부 정보를 유지하고 싶다면 작은 min\_samples와 큰 min\_cluster\_size를 사용한다. 매우 일반적인 클러스터를 사용하고 클러스터에서 많은 노이즈를 제거하려면 큰 min\_sample과 큰 min\_cluster\_size를 사용하면 된다.

- 최종적으로 일반화된 클러스터를 만들면서 대부분의 세부 정보를 유지하기 위해 min\_samples=20, min\_cluster\_size=100으로 설정하였다.

## (4) 최종 클러스터링 결과

그림과 같이 총 4개의 그룹으로 적절하게 클러스터링된 것을 직관적으로 알 수 있다.



[그림 3-17], [그림 3-18] 차원축소한 그림 시각화

<계층 비지도 클러스터링 문제>

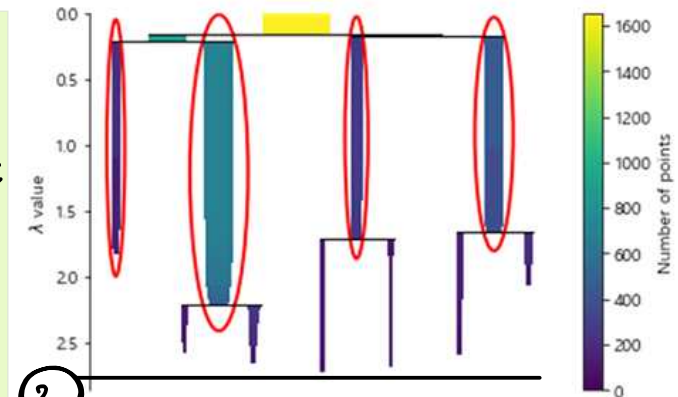
minimum size 이상의 크기를 가진 component들이 나온다. (선택치에 따른 component에 포함된 데이터)

(탄생시기)  
 $\lambda = \frac{1}{\text{distance}}$   
 local density를 더 잘 표현하기 위해

stability  $I(\lambda_p - \lambda_{p-1})$

## ■ condensed\_tree\_ 그려보기

□ 어떤 클러스터가 서로 가까이 있거나 병합될 수 있는지, 어느 클러스터가 멀리 떨어져 있는지 클러스터 객체의 condensed\_tree\_속성을 통해 기본 계층에 액세스할 수 있다. 각 계층 구조를 덴드로그램으로 볼 수 있으며, 각 분기의 폭(및 색상)은 해당 수준에서 클러스터의 점수를 나타낸 그림이다. [그림 3-19]를 보면 빨간색 부분이 선택된 분기를 의미하며 람다값이 크고 길게 나타나는 것을 통해 안정적으로 클러스터를 구성하고 있음을 확인할 수 있다.

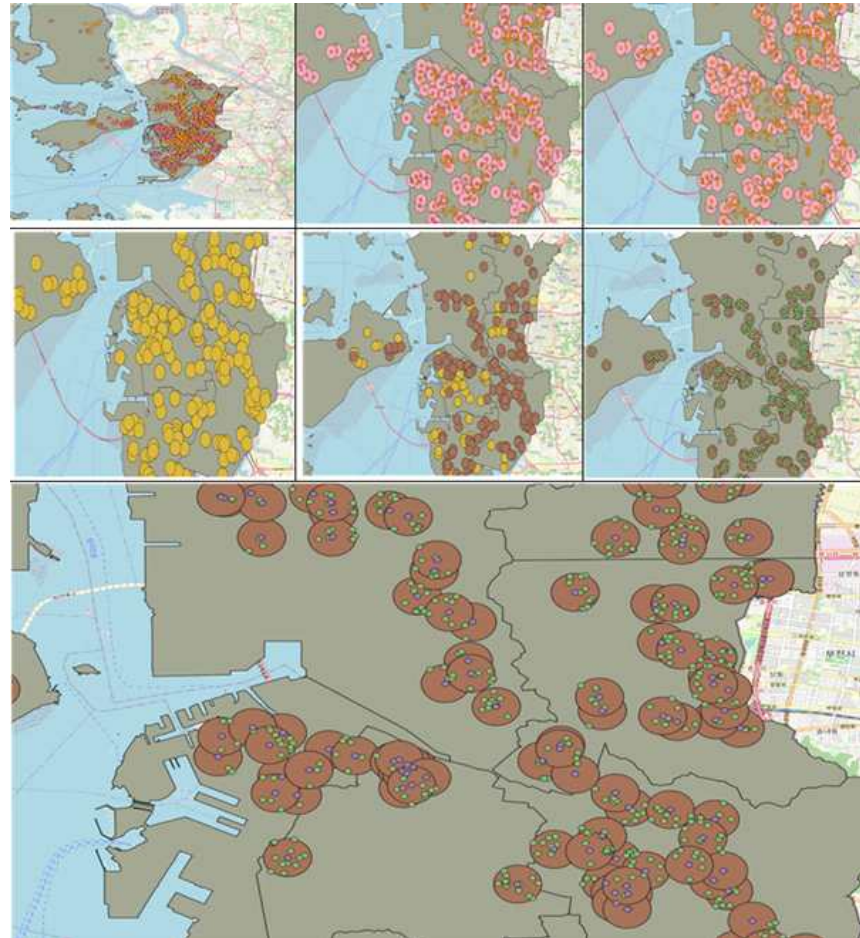


[그림 3-19] condensed\_tree\_ 예시

## 2.3. 인천광역시 중심 국공립 어린이집 선정

## 1) 중심 국공립 어린이집 선정

- 인천의 어린이집 중 설립 유형이 국공립인 어린이집을 선정
- 선택된 국공립 어린이집을 기준으로 반경 500m 버퍼 형성
- 폴리곤에 포함하는 포인트 개수 계산으로 국공립 어린이집 주변 어린이집 수 도출
- 주변 어린이집 수가 도출된 버퍼에서 일정 기준치(주변 어린이집 4개 이상, 주변 공원 1개 이상)를 만족하는 버퍼 선택
- 선택된 버퍼의 중심에 있는 국공립 어린이집을 중심 국공립 어린이집으로 선정



[그림 3-20] 인천 어린이집 중심 국공립 어린이집 선정 시각화

## 2) 중심 국공립 어린이집 반경 500m 이내의 주변 어린이집 확인

- ☐ 선정된 중심 국공립 어린이집으로부터 반경 500m 이내에 들어오는 주변 어린이집들만 추출해 데이터 파일을 만든다.
- ☐ 2개 이상의 중심 국공립 어린이집의 반경 이내에 들어오는 경우, 가장 가까운 중심 국공립 어린이집을 선택한다. 해당 중심 국공립 어린이집의 고유 id 정보를 담은 HubName 컬럼 추가
- ☐ 주변 어린이집과 허브(중심 국공립 어린이집) 간의 거리를 구해 HubDist 컬럼 추가

	A	B	C	D	E	F	G	H	I	AG	AH	AI	AJ	AK	AL	AM
1	시도	시군구	어린이집명	어린이집종류	영양형	우편번호	주소	어린이집종류	법정동명	라벨+유형	구유형명	id	특정분류	HubName	HubDist	
2	인천광역시 중구	갈매기어린이국공립	정상	22327	인천광역시 032-885-8	도원동	0국공립	중구국공립	50	1	50	0				
3	인천광역시 중구	중산어린이국공립	정상	22407	인천광역시 032-751-3	중산동	0국공립	중구국공립	74	1	74	0				
4	인천광역시 중구	운서어린이국공립	정상	22376	인천광역시 032-751-3	운서동	0국공립	중구국공립	84	1	84	0				
5	인천광역시 중구	스위첸어린이국공립	정상	22389	인천광역시 032-751-3	중산동	0국공립	중구국공립	118	1	118	0				
6	인천광역시 중구	오션하일어린이국공립	정상	22394	인천광역시 032-751-8	중산동	0국공립	중구국공립	134	1	134	0				
266	인천광역시 계양구	부평초등학교	정상	21049	인천광역시 032-541-9	계산동	0초등학교	계양구계산동	1239	1	1259	88.55713				
267	인천광역시 남동구	예다인어린이민간	정상	21553	인천광역시 032-433-1	간석동	1민간	남동구간석동	760	1	642	92.31339				
268	인천광역시 남동구	미래어린이가정	정상	21525	인천광역시 032-466-5	만수동	0가정	남동구만수동	732	2	635	93.18242				
269	인천광역시 연수구	아기솔어린이가정	정상	21967	인천광역시 032-816-2	동춘동	0가정	연수구동춘동	489	0	563	97.38247				
270	인천광역시 부평구	소나무어린이민간	정상	21450	인천광역시 032-3330-	십정동	1민간	소나무부평구민간	946	1	1069	99.5223				
271	인천광역시 부평구	매물어린이민간	정상	21312	인천광역시 032-526-4	청천동	1민간	매물부평구민간	1032	1	886	100.3535				
272	인천광역시 남동구	코마사어린이민간	정상	21510	인천광역시 032-422-6	간석동	1민간	코마남동구민간	675	1	796	102.9185				
273	인천광역시 부평구	예교어린이가정	정상	21384	인천광역시 032-521-9	산곡동	2가정	예교부평구민간	954	3	1116	102.9621				
274	인천광역시 연수구	봉림예린이가정	정상	21977	인천광역시 032-815-0	동춘동	2가정	연수구동춘동	410	2	511	104.8396				
275	인천광역시 중구	온유어린이가정	정상	22394	인천광역시 032-268-0	중산동	0가정	온유중구민간	129	0	134	106.8139				
276	인천광역시 남동구	e편한어린이가정	정상	21613	인천광역시 032-467-7	서창동	0가정	e편한남동구민간	857	0	861	110.0562				

[그림 3-21] 중심 국공립 어린이집 500m 이내 어린이집 데이터셋

## 3.2.4. 인천형 모아어린이집 공동체 구성 어린이집 추출

## 1) 허브(중심 국공립 어린이집)의 특성 분류별 최종 주변 어린이집 추출

- ☐ Python을 이용해 가장 가까운 중심 국공립 어린이집, 즉 허브와의 거리 컬럼을 결합한 csv 파일을 불러온다.
- ☐ 국공립 어린이집의 경우, 허브로서의 기능도 수행하고 있다면 동일 어린이집 간 연결된다. 때문에 해당 행들에 대해 어린이집 id와 허브 id의 일치 여부 조건을 이용해 제거
- ☐ 특정 허브로부터 500m 이내에 위치한 어린이집들을 클러스터 집단별로 허브와 최소 거리에 있는 어린이집 1곳 추출
- ☐ 사용할 주요 컬럼명
  - HubName : 한 어린이집이 여러 허브로부터 500m 이내 범위에 포함된 경우 그 중 최단거리 허브의 id



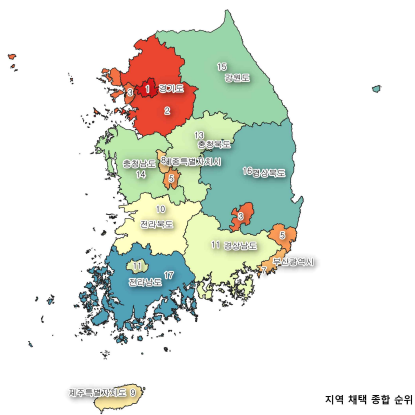


- 중심 국공립어린이집을 허브로 선정
- 허브의 반경 500m 버퍼 범위 내에 속하는 인천 어린이집들을 최근접 허브와 라인으로 연결하고 라인의 거리를 도출, 하나의 허브와 연결된 어린이집들을 하나의 그룹으로 간주
- 그룹 내의 어린이집들을 허브와의 거리와 특징 분류를 고려하여 상호 보완적인 어린이집들을 인천 공동체로 도출
- 도출된 인천 공동체 중 구성 어린이집 수가 2개 이하인 공동체 제외한다.
- 공동체 분포 시각화

## 4. 최적 어린이집 공동체 분석 결과

### 4.1. 시범 운영 지역 탐색

- 각 기준에서의 순위를 합한 값인 종합 점수를 오름차순 기준으로 rank 합수를 적용하여 종합 순위를 도출
- 그 결과, 우선적으로 시범 운영할 지역 1위는 서울특별시, 2위는 경기도, 3위는 공동으로 인천광역시, 대구광역시, 공동 5위는 대전광역시, 울산광역시, 7위는 부산광역시가 선정
- 종합순위에서 낮은 순위를 차지한 시도로는 8위 세종특별자치시, 9위 제주특별자치도, 10위 전라북도, 공동 11위 광주광역시와 경상남도, 13위 충청북도, 14위 충청남도, 15위 강원도, 16위 경상북도, 17위 전라남도



[그림 4-1] 지역 채택 종합 순위 시각화

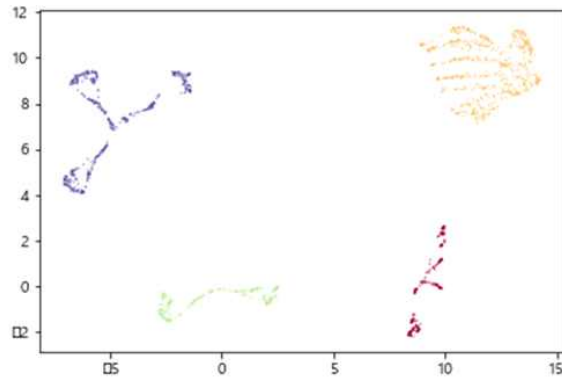
	A	R	S
1	행정구역(시군구)별	종합점수	종합순위
2	서울특별시	28	1
3	경기도	35	2
4	대구광역시	41	3
5	인천광역시	41	3
6	대전광역시	48	5
7	울산광역시	48	5
8	부산광역시	50	7
9	세종특별자치시	55	8
10	제주특별자치도	56	9
11	전라북도	57	10
12	경상남도	59	11
13	광주광역시	59	11
14	충청북도	62	13
15	충청남도	65	14
16	강원도	66	15
17	경상북도	68	16
18	전라남도	73	17

[그림 4-2] 지역 채택 종합 순위 데이터셋

- 공동 보육 목적의 어린이집 공동체 형성을 위한 위치 선정 중요성에 있어서 1위를 차지한 서울특별시의 경우 이미 서울형 모아어린이집이 시행되고 있기에 제외한다.
- 2위 경기도의 경우 상당수의 행정구역을 포괄하고 있어 행정구역 간 어린이집의 분포 편차를 고려했을 때 최우선 시범 운영 지역으로 선정하기에 적합하지 않다고 판단되어 배제한다.
- 인천광역시와 대구광역시가 공동 3위를 차지했기 때문에 두 광역시를 중심으로 분석 기준 6개 측면에서 비교할 필요가 있다.
- (기준1) 사립 대비 국립 어린이집 수를 봤을 때 인천광역시가 0.94, 대구광역시가 0.56의 비율을 보여 대구광역시가 국립보다 사립 어린이집이 많다.
- (기준2) 향후 국공립 어린이집을 늘려야 한다고 응답한 비율은 인천광역시가 12.1%, 대구광역시가 10.7%로 인천광역시가 국공립 추가 설립에 대한 시급성이 높다.
- (기준3) 향후 보육서비스를 늘려야 한다는 응답 비율은 인천광역시가 2.5%, 대구광역시가 3.8%로 복지서비스의 개선 필요성에선 대구광역시가 높다.
- (기준4) 아동 인구수 대비 아동학대 피해 건수 비율을 봤을 때 인천광역시가 0.0285%로 0.0212%인 대구광역시보다 높다.
- (기준5) 아동 인구는 인천이 108,898명, 대구광역시가 81,217명으로 인천이 대구의 1.34배의 아동 인구를 포함하고 있다.
- (기준6) 어린이집 수 대비 아동 인구 비율은 인천이 282.85, 대구가 261.32로 어린이집 인원 수용에 있어서 문제 발생률이 인천이 대구보다 높음을 확인할 수 있다.
- 기준1, 기준3에서 대구의 점수가 높다면, 인천은 기준2, 기준4, 기준5, 기준6에서 점수가 상대적으로 높다. 이를 감안하여, 인천광역시를 모아어린이집 전국 확대를 위한 초석으로 삼을 지역으로 채택했다.

## 4.2. 인천광역시 어린이집 분석

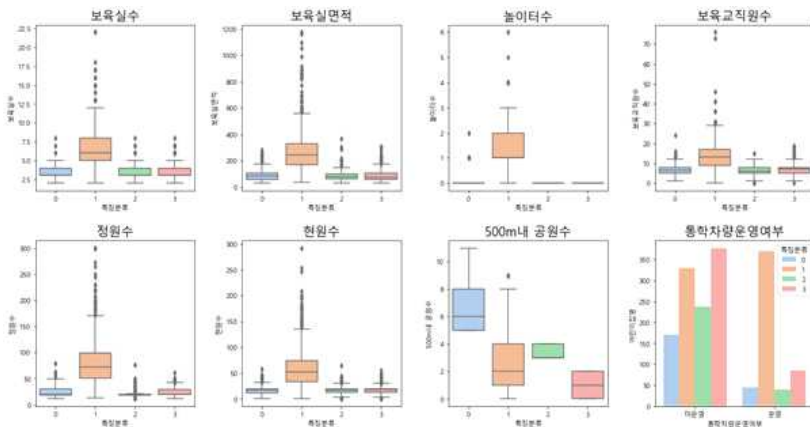
### 1) 최종 클러스팅 결과



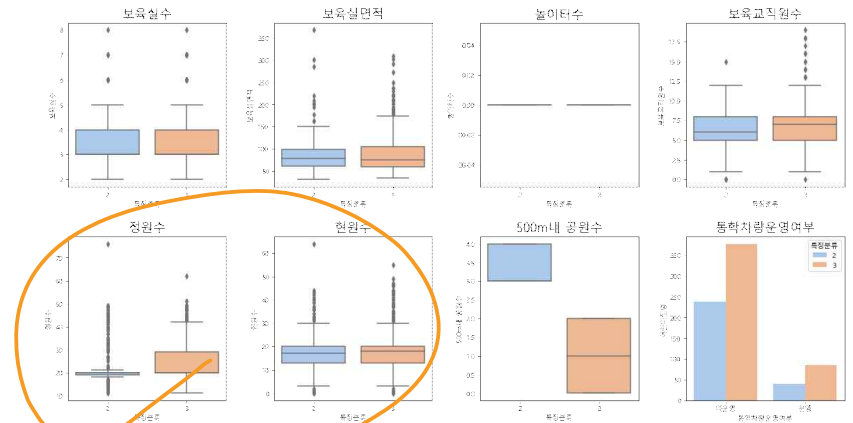
[그림 4-3] 최종 클러스터링 결과

### 2) 군집화된 그룹별 어린이집 특징

- [그림4-4]를 보면 '그룹1'의 경우 보육실수, 보육실면적, 놀이터수, 보육교직원수, 정원수, 현원수 모두 다른 그룹보다 더 큰 값들을 가진다. '그룹0'의 경우 500m내의 공원수가 특히 많다.
- [그림4-5]는 '그룹2'와 '그룹3'의 차이를 자세히 보기 위해 그려보았다. '그룹3'이 500m 내 공원수가 '그룹2'보다 많다. 그리고 '그룹2'의 경우 현원수와 정원수를 비교하면 현재 만원이라는 것을 알 수 있고, '그룹3'의 경우 현원수보다 정원수가 더 많기 때문에 추가인원을 받을 수 있다.



[그림 4-4] 그룹별 어린이집 특성 시각화



[그림 4-5] '그룹2'와 '그룹3'의 어린이집 특성 시각화

아담

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	시도	시군구	어린이집명	어린이집유형	현황	주소	보육실수	보육실면적	놀이터수	보육교직원수	정원수	통학차량운행	500m내 공원수	공id		특성분류
2	인천광역시	동구	송현어린이집	재개	인천광역시	3	59	0	5	22	6 미운영	3	1			2
3	인천광역시	중구	잔송어린이집	재개	인천광역시	3	55	0	3	20	5 미운영	0	2			3
4	인천광역시	연수구	빛솔어린이집	재개	인천광역시	4	89	1	5	30	7 운영	1	3			1
5	인천광역시	마추홀구	파란돌어린이집	재개	인천광역시	7	370	2	9	84	32 운영	4	4			1
6	인천광역시	남동구	아이꿈어린이집	재개	인천광역시	4	93	0	3	35	10 미운영	4	5			2

[그림 4-6] 어린이집 특성 분류 데이터셋



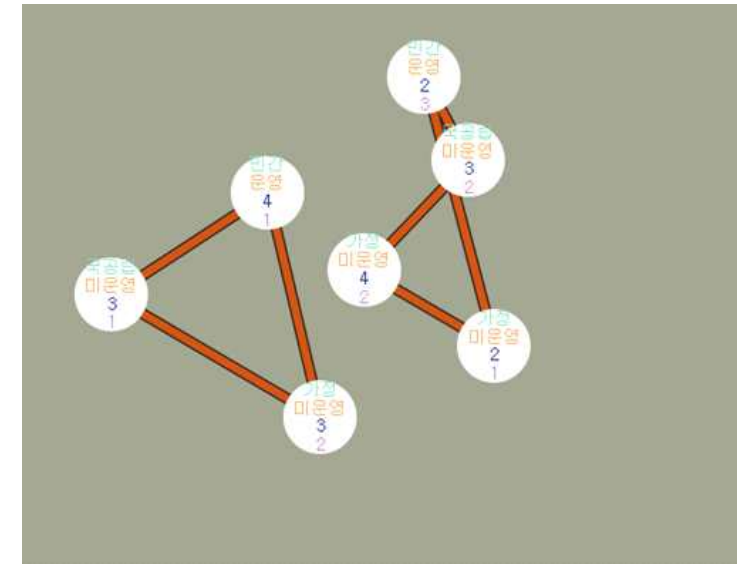


#### 4.5. 인천형 모아어린이집 위치 시각화

- 2곳만 속하는 공동체를 제외한 최종 공동체 수는 108개이다.
- hexagon 그리드를 이용하여 최종 공동체의 분포를 시각화한다.



[그림 4-11] 인천형 모아어린이집 시각화



[그림 4-12] 두 그룹의 모아어린이집 시각화  
(유형, 통학차량 운영여부, 주변공원수, 특징분류 포함)

- 좌측의 모아어린이집은 국공립 어린이집 1개, 가정 어린이집 1개 그리고 민간 어린이집 1개가 공동체도 묶여있다. 국공립 어린이집과 가정 어린이집의 경우 보육실수, 보육실면적, 놀이터수, 보육직원수가 상대적으로 많아 다른 어린이집과 공유하며 서로 이득을 줄 수 있다.
- 우측의 모아어린이집의 경우 국공립 어린이집 1개, 가정 및 민간 어린이집 3개로 묶여있다. 이 그룹에는 3종류의 특징 유형을 가진 어린이집들이 다양하게 묶여있다.
- 모든 어린이집이 공원과 가까운 거리에 있으며 4개 어린이집 간의 거리는 도보로 이용 가능하다. 따라서 인천 모아어린이집이 적절하게 묶여 있음을 확인할 수 있다.

## 5. 활용 방안

### 5.1. 문제점 개선 방안

#### 5.1.1. 문제점

- (관련 자료 부족) 모아 어린이집 사례가 많지 않았다.
- (변수 파악 미흡) 공원 외에도 다른 시설 및 어린이 교육프로그램을 고려하지 못했다.
- (참여의 한계) 선정 조건에 만족하지 못해 공동체 속에 포함되지 못한 어린이집이 있다.

#### 5.1.2. 개선방안

- (변수의 다양화) 생태 친화라는 모아어린이집의 목적에 대한 변수를 공원으로 선정해 분석을 진행했으나 분석 모델을 통해 다양한 변수를 적용해 분석 가능하다.
- (분석 모델 활용) 시간 제약으로 모아어린이집에 대한 분석을 인천광역시로 한정하여 진행했으나 분석 모델을 통해 다른 지역도 분석 가능하다.

### 5.2. 업무 활용 방안

#### 5.2.1. 기존 문제 보완

- (국공립 비율 조정) 국공립, 가정 어린이집 간의 보육 수급 격차를 해소할 수 있다.
- (공동체간의 거리를 최소화) 공동체 간의 어린이집 거리를 최소화하여 거리가 멀어서 발생하는 운영의 불편함을 최소화할 수 있다.
- (운영 환경 개선) 공원과 같은 환경을 고려함으로써 기존의 모델보다 생태 친화 기능 효과를 증대할 수 있다.

#### 5.2.2. 정책 실현 가능성 증대

- (모델의 재활용성) 유연하게 사용 가능한 코드를 이용함으로써 데이터를 추가하여 공동체가 원하는 목적에 맞게 활용 가능하다.
- (전국 시행 가능성) 전국으로 확대 시행 시 선제적으로 제안을 하여 참여율을 증가시킬 수 있다.
- (지자체 참여) 지자체별 테마에 맞는 시설을 추가하여 지자체만의 새로운

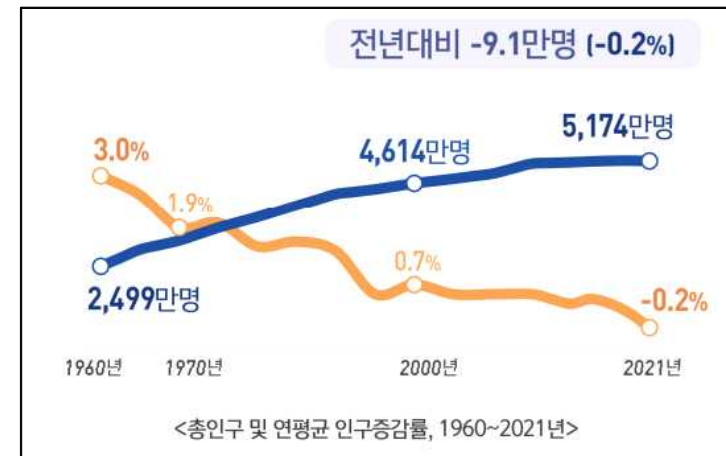
방식으로 서울형 모아어린이집 보육 모델 활용 가능

- (어린이집 참여 독려) 사전에 공동체를 분류하여 자발적인 지원과 더불어 어린이집에 공동체 구성원으로서 참여를 제안할 수 있다.

## 6. 참고자료(Reference)

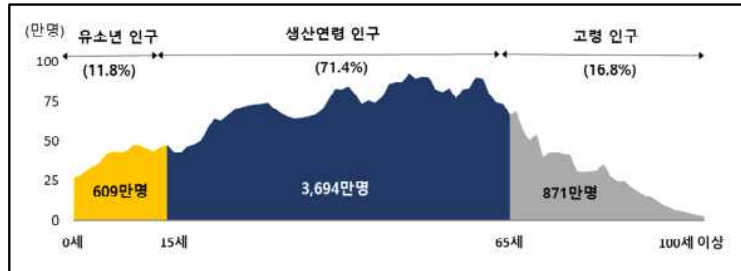
### 6.1. 저출산 원인

- 2022년 7월 28일 통계청이 발표한 '2021년 인구주택총조사 인구 부분 집계 결과'에서 지난해 11월 1일 기준으로 우리나라 총인구는 5173만 8000명으로 1년전보다 9만 1000명(0.2%) 감소한 것으로 나타남.
- 국내 거주 외국인을 포함한 우리나라 총인구가 정부 수립 후 72년 만에 처음으로 감소

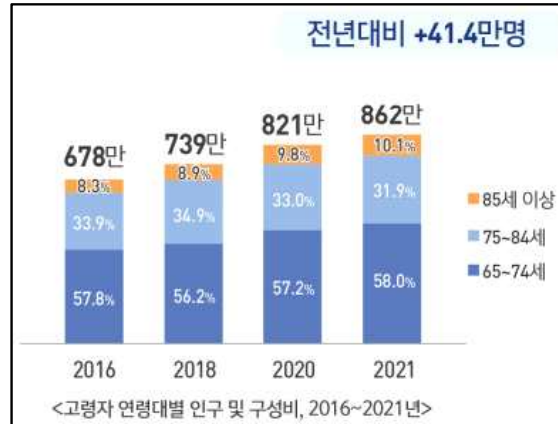


<출처- 통계청-총인구 및 연평균 인구증감률, 1960~2021년>

- 노인 인구(65세 이상)는 한 해동안 41만명이 급증하여 862만 명으로 총인구 중 고령화 인구가 차지하는 비율은 16.8%이며, 전년대비 5.1% 증가



&lt;출처\_통계청\_연령별인구, 2021년&gt;



&lt;출처\_통계청\_고령자 연령대별 인구 및 구성비, 2016~2021년&gt;

< 표 3 > 유소년인구, 생산연령인구, 고령인구, 2005~2021년<sup>R</sup>

(단위 : 천 명, %, 여자 100명당)

구 분		2005년	2010년	2015년 <sup>R</sup>	2016년 <sup>R</sup>	2017년 <sup>R</sup>	2018년 <sup>R</sup>	2019년 <sup>R</sup>	2020년 <sup>R</sup>	2021년 <sup>R</sup>	전년 대비 증감
계		47,279	48,580	51,069	51,270	51,423	51,630	51,779	51,829	51,738	-91
인구	유소년인구 (0~14세)	8,994	7,806	6,954	6,822	6,683	6,543	6,392	6,254	6,087	-167
	생산연령인구 (15~64세)	33,913	35,340	37,490	37,621	37,569	37,632	37,570	37,208	36,944	-344
	고령인구 (65세 이상)	4,372	5,434	6,617	6,827	7,171	7,455	7,817	8,287	8,707	419
구성비	유소년인구 (0~14세)	19.0	16.1	13.6	13.3	13.0	12.7	12.3	12.1	11.8	-0.3
	생산연령인구 (15~64세)	71.7	72.7	73.4	73.4	73.1	72.9	72.6	71.9	71.4	-0.5
	고령인구 (65세 이상)	9.2	11.2	13.0	13.3	13.9	14.4	15.1	16.0	16.8	0.8
유소년부양비 <sup>1)</sup>		26.5	22.1	18.5	18.1	17.8	17.4	17.0	16.8	16.5	-0.3
노년부양비 <sup>2)</sup>		12.9	15.4	17.6	18.1	19.1	19.8	20.8	22.2	23.6	1.3
노령화지수 <sup>3)</sup>		48.6	69.6	95.2	100.1	107.3	113.9	122.3	132.5	143.0	10.5

1) 생산연령인구 1백명당 부양해야 할 유소년인구

2) 생산연령인구 1백명당 부양해야 할 고령인구

3) 유소년인구 1백명당 고령인구

1) 생산연령인구 1백명당 부양해야 할 유소년인구

2) 생산연령인구 1백명당 부양해야 할 고령인구

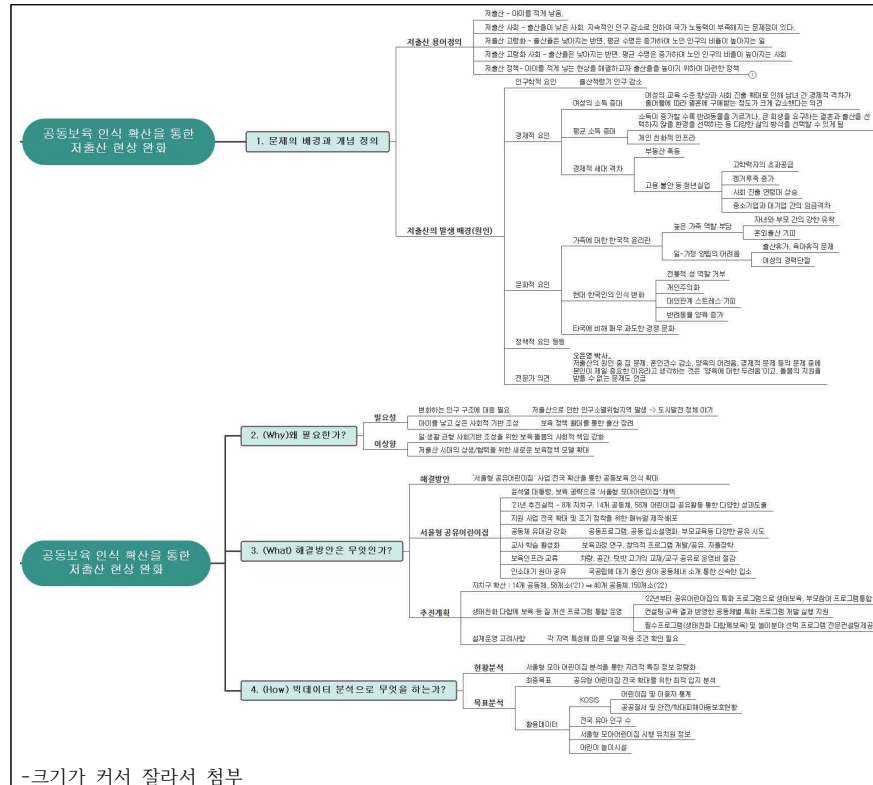
3) 유소년인구 1백명당 고령인구

&lt;출처\_통계청\_유소년인구, 생산연령인구, 고령인구 2005~2021년&gt;



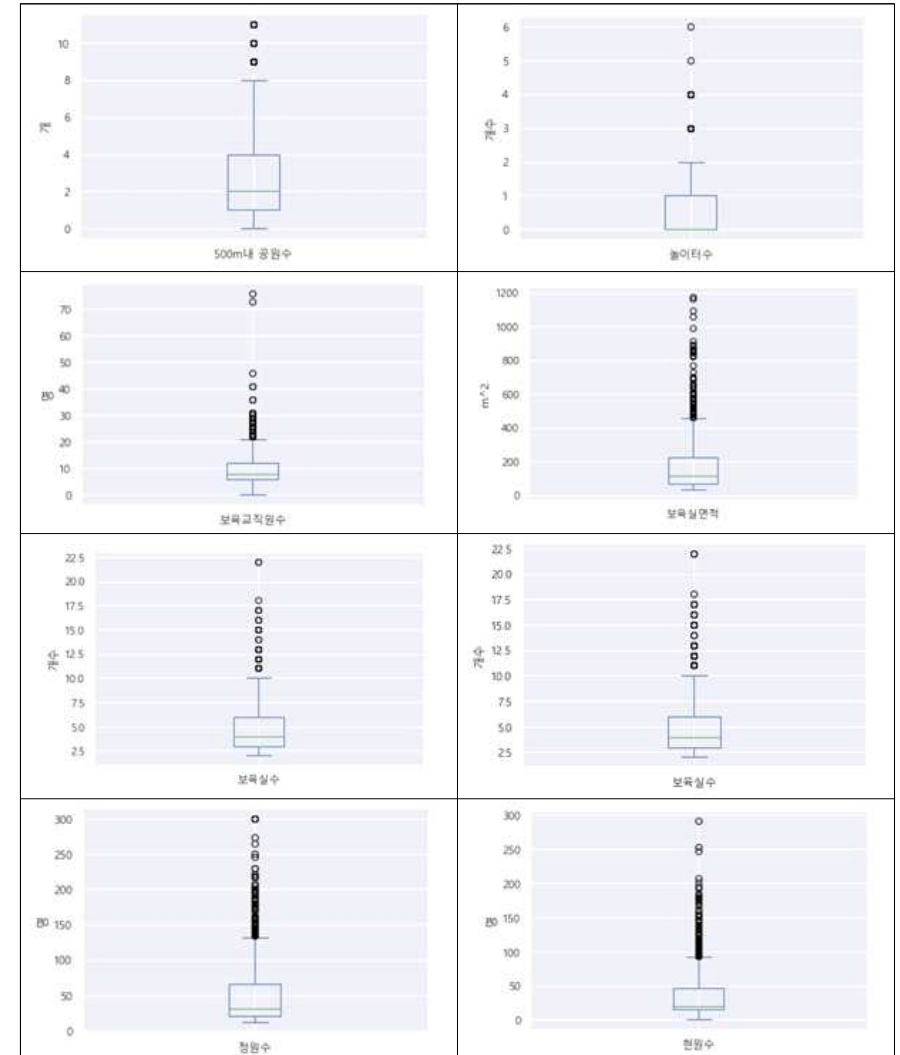
## 7. 부록

## 7.1. 마인드맵

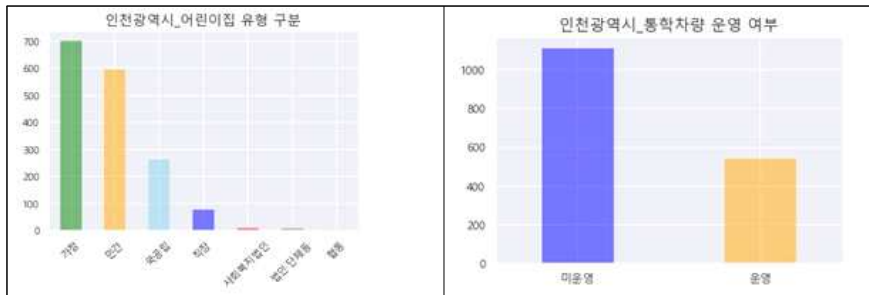


## 7.2. 인천광역시 어린이집 데이터 EDA(box plot, bar plot, histogram, scatter plot)

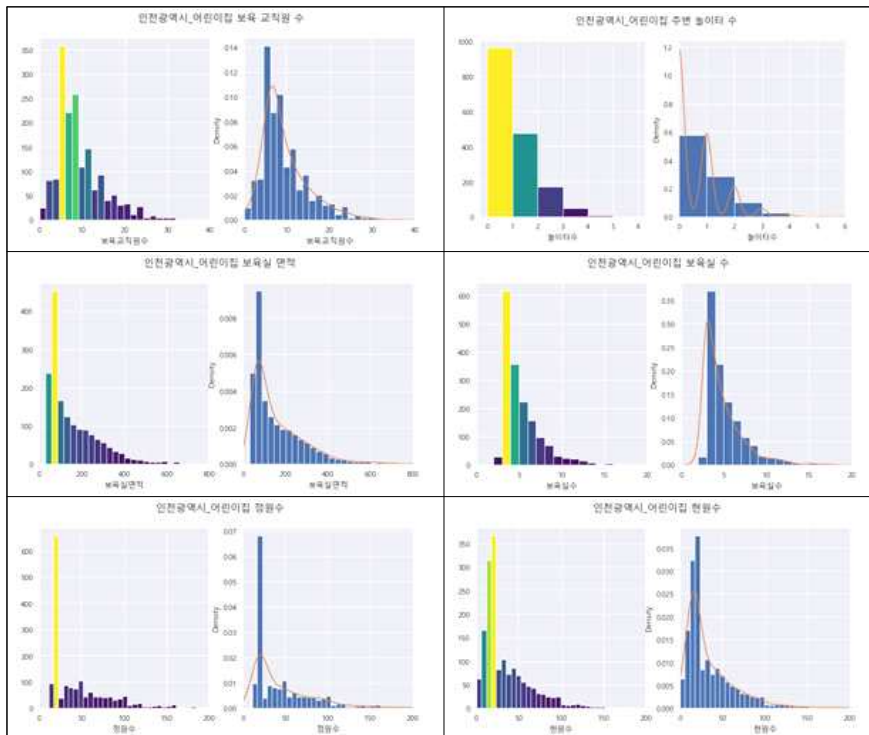
□ box plot



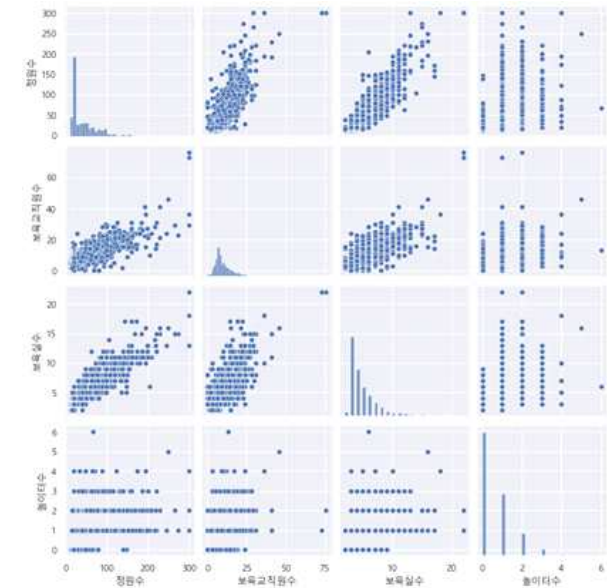
## □ bar plot



## □ histogram



## □ scatter plot



## 7.3. 차원축소와 클러스터링 코드

```
#####
##### 데이터 불러오기
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

# 인천 어린이집 데이터 불러오기
data = pd.read_csv('data/데이터2.csv', encoding='euc-kr')
data.head(3)
```



```

data.shape # 데이터 확인
data.columns # 칼럼 확인
# 사용할 칼럼만 추출
df = data[['id','어린이집명','어린이집유형구분','보육실수','보육실면적','놀이터수','
보육교직원수','정원수','현원수',
'통학차량운영여부','500m내 공원수']]
df.head()
df['어린이집유형구분'].value_counts()
#####
##### 데이터 전처리
# 어린이집유형구분 라벨링
from sklearn.preprocessing import LabelEncoder
label = LabelEncoder()
df['어린이집유형구분2'] = label.fit_transform(df['어린이집유형구분'])
print(label.classes_)
print(label.inverse_transform([0,1,2,3,4,5]))

# 통학차량운영여부 라벨링
df['통학차량운영여부'] = df['통학차량운영여부'].replace({'운영':1, '미운영':0})
# 스케일 변환
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
col = ['보육실수','보육실면적','놀이터수','보육교직원수','정원수','현원수','500m내
공원수']
df[col] = scaler.fit_transform(df[col])
df.head()
df.isnull().sum()
df = df.dropna()
df.head()

#####
##### 변수간 상관관계 파악
# 칼럼간 상관관계 확인
import seaborn as sns
import matplotlib
cor = df[col].corr()

```

```

matplotlib.rc('font', family='Malgun Gothic')
sns.heatmap(cor,
             cmap = 'RdYlBu_r',
             annot = True, # 실제 값을 표시한다
             #mask=mask, # 표시하지 않을 마스크 부분을 지정한다
             linewidths=.5, # 경계면 실선으로 구분하기
             cbar_kws={"shrink": .5},# 컬러바 크기 절반으로 줄이기
             vmin = -1,vmax = 1 # 컬러바 범위 -1 ~ 1
             )
plt.show()

#####
##### 차원축소-umap

from sklearn.datasets import fetch_openml
from sklearn.decomposition import PCA
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

# Dimension reduction and clustering libraries
import umap
import hdbscan
import sklearn.cluster as cluster
from sklearn.metrics import adjusted_rand_score,
adjusted_mutual_info_score
# umap
clusterable_embedding = umap.UMAP(
    n_neighbors=50,
    min_dist=0.1,
    n_components=2, # 2-dim
    random_state=20,
).fit_transform(df[col])
clusterable_embedding
# umap plot
plt.scatter(clusterable_embedding[:, 0], clusterable_embedding[:, 1],

```

```

c=df['어린이집유형구분2'],
s=0.1, cmap='Spectral')
plt.legend(df['어린이집유형구분2'], loc= 'upper right')
plt.show()
#####
##### 클러스터링-hdbscan
# hdbscan
import hdbscan
hdbscan = hdbscan.HDBSCAN(
    min_samples=20,
    min_cluster_size=100,
).fit(clusterable_embedding)
# condensed_tree_
hdbscan.condensed_tree_.plot(select_clusters=True)
import hdbscan
labels = hdbscan.HDBSCAN(
    min_samples=20,
    min_cluster_size=100,
).fit_predict(clusterable_embedding)

# 모두다 클러스터링 되었다.
clustered = (labels >= 0)
plt.scatter(clusterable_embedding[~clustered, 0],
            clusterable_embedding[~clustered, 1],
            color=(0.5, 0.5, 0.5),
            s=0.1,
            alpha=0.5)

# 차원축소한 그래프에 차원축소하여 클러스터링하여 그리기
plt.scatter(clusterable_embedding[clustered, 0],
            clusterable_embedding[clustered, 1],
            c=labels[clustered],
            s=0.1,
            cmap='Spectral');
```

```

sns.jointplot(
    x=clusterable_embedding[clustered, 0],
    y=clusterable_embedding[clustered, 1],
    hue=labels[clustered], kind="kde"
)

np.unique(labels)
unique, counts = np.unique(labels, return_counts=True)
print (np.asarray((unique, counts)).T)

df['특징분류'] = labels
df_group = df.groupby(['특징분류', '어린이집유형구분'])['어린이집유형구분'].count()

df_2 = df[['id', '특징분류']]
df_2.shape

df_2['특징분류'].unique()

df_final = pd.merge(data, df_2, how='outer', on='id')
df_final.columns

df_final['특징분류'].unique()

df_final.shape

df_final.to_csv('data/데이터3.csv')

df_eda = df_final[['어린이집명', '어린이집유형구분', '보육실수', '보육실면적', '놀이터수', '보육교직원수', '정원수', '현원수',
                    '통학차량운영여부', '500m내 공원수', '특징분류']]

# 특징분류별 변수 평균값
df_eda.groupby('특징분류').mean()

# 통학차량운영여부 barplot을 위한 그룹통계
```

```

bus_sum = df_eda.groupby(['특징분류','통학차량운영여부']).어린이집명.count()
bus_sum = pd.DataFrame(bus_sum)
bus_sum = bus_sum.reset_index()
bus_sum

# 특징 그룹별 그래프
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(20,10))
plt.subplots_adjust(wspace = 0.3, hspace = 0.3)
sns.set_palette("pastel")
plt.subplot(2,4,1)
sns.boxplot(x = '특징분류', y = '보육실수',data=df_eda).set_title("보육실수",size=18)
plt.subplot(2,4,2)
sns.boxplot(x = '특징분류', y = '보육실면적',data=df_eda).set_title("보육실면적",size=18)
plt.subplot(2,4,3)
sns.boxplot(x = '특징분류', y = '놀이터수',data=df_eda).set_title("놀이터수",size=18)
plt.subplot(2,4,4)
sns.boxplot(x = '특징분류', y = '보육교직원수',data=df_eda).set_title("보육교직원수",size=18)
plt.subplot(2,4,5)
sns.boxplot(x = '특징분류', y = '정원수',data=df_eda).set_title("정원수",size=18)
plt.subplot(2,4,6)
sns.boxplot(x = '특징분류', y = '현원수',data=df_eda).set_title("현원수",size=18)
plt.subplot(2,4,7)
sns.boxplot(x = '특징분류', y = '500m내 공원수',data=df_eda).set_title("500m내 공원수",size=18)
plt.subplot(2,4,8)
sns.barplot(x = '통학차량운영여부',y='어린이집명', hue = '특징분류',data=bus_sum).set_title("통학차량운영여부",size=18)
plt.show()

df_eda2 = df_eda[(df_eda['특징분류']==2)|(df_eda['특징분류']==3)]

```

```

# 통학차량운영여부 barplot을 위한 그룹통계
bus_sum2 = df_eda2.groupby(['특징분류','통학차량운영여부']).어린이집명.count()
bus_sum2 = pd.DataFrame(bus_sum2)
bus_sum2 = bus_sum2.reset_index()
bus_sum2

# 특징 그룹별 그래프
import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(20,10))
plt.subplots_adjust(wspace = 0.3, hspace = 0.3)
sns.set_palette("pastel")
plt.subplot(2,4,1)
sns.boxplot(x = '특징분류', y = '보육실수',data=df_eda2).set_title("보육실수",size=18)
plt.subplot(2,4,2)
sns.boxplot(x = '특징분류', y = '보육실면적',data=df_eda2).set_title("보육실면적",size=18)
plt.subplot(2,4,3)
sns.boxplot(x = '특징분류', y = '놀이터수',data=df_eda2).set_title("놀이터수",size=18)
plt.subplot(2,4,4)
sns.boxplot(x = '특징분류', y = '보육교직원수',data=df_eda2).set_title("보육교직원수",size=18)
plt.subplot(2,4,5)
sns.boxplot(x = '특징분류', y = '정원수',data=df_eda2).set_title("정원수",size=18)
plt.subplot(2,4,6)
sns.boxplot(x = '특징분류', y = '현원수',data=df_eda2).set_title("현원수",size=18)
plt.subplot(2,4,7)
sns.boxplot(x = '특징분류', y = '500m내 공원수',data=df_eda2).set_title("500m내 공원수",size=18)
plt.subplot(2,4,8)
sns.barplot(x = '통학차량운영여부',y='어린이집명', hue = '특징분류',data=bus_sum2).set_title("통학차량운영여부",size=18)
plt.show()

```

#### 7.4. 인천형 모아어린이집 공동체 구성 어린이집 추출 파이썬 코드

```
#####
#####
### 중심 허브의 주변 어린이집 중 클러스터 집단별 최소거리에 있는 어린이집 추출 ###

### 파일 불러오기
import pandas as pd
import numpy as np

df = pd.read_csv('데이터4.csv', encoding='euc-kr')
df.columns

# 특징분류 >> 민간 어린이집 클러스터 넘버
# HubName >> 한 어린이집이 여러 허브의 범위에 포함된 경우 그중 최단거리 허브(국공립 어린이집)
# HubDist >> 민간 어린이집과 최단 허브간의 거리

df.head()

### 국공립 어린이집 - (동일)국공립어린이집 연결된 행 제거하기
# 국공립 어린이집 - (동일)국공립어린이집 연결된 행들도 있기 때문에 해당 행들 제거 필요
# 'id'랑 'HubName' 일치여부 조건 이용해 행 제거

df.info()

drop_index = df[df['id'] == df['HubName']].index
drop_index
```

```
len(drop_index)

df.drop(drop_index, inplace = True)

df[df['id'] == df['HubName']]

df.head()

# 기존 index 제거 O
df.reset_index(drop=True, inplace = True)

df.head()

### 클러스터 집단별 최소거리에 있는 어린이집 추출하기
df['특징분류'].unique()

df['특징분류'].nunique()

df['HubName'].nunique()

## 방법 : 허브(중심 어린이집)의 특성분류별 최종 주변 어린이집을 추출해 데이터프레임에 담기
# 중심(c1), 주변(c2) 컬럼을 가진 빈 데이터프레임 만들기
final_kindergarten = pd.DataFrame(columns=['중심 국공립 어린이집', '주변 어린이집'])

for i in range(df['HubName'].nunique()): # df['HubName']의 고유값 개수만큼 for loop 진행

    # 중심 국공립 어린이집 별 : 같은 허브와 연결된 주변 어린이집들을 모아 kind_group 변수에 담기
    kind_group = df[df['HubName'] == df['HubName'].unique()[i]]

    # 중심 국공립 어린이집 id를 데이터프레임에 넣기 위함
    c1 = df['HubName'].unique()[i]
```

```

for j in range(df['특징분류'].nunique()): # df['특징분류'] 고유값 개수(클러스터 개수)만큼 for loop 진행

    # 소속 클러스터별 : 같은 허브와 연결된 주변 어린이집들 중 같은 특징으로 분류된 어린이집들을 모아 same_cluster 변수에 담기
    same_cluster = kind_group[kind_group['특징분류'] == j]

    if len(same_cluster['HubDist']) > 0: # same_cluster에 값이 존재할 때

        # 소속 클러스터별(위에서의 for loop에 의해) 거리가 가장 작은 데이터 찾기

        # argmin() 이용하여 Distance 기준으로 인덱스 찾기
        min_index = np.argmin(same_cluster['HubDist'])

        # 사용 방법 : 특징분류 그룹에서 최소거리 어린이집이 뿔힐 때마다 그것을 바로 행으로 추가하는 방법
        # 인덱스에 해당하는 데이터의 어린이집의 id 찾아 변수에 저장
        c2 = same_cluster['id'].iloc[min_index]

        # 데이터프레임 만드는 작업 >> 행 추가하는 방식으로 데이터프레임 구성

        final_kindergarten =
final_kindergarten.append(pd.DataFrame([c1,c2], columns = ['중심 국공립 어린이집','주변 어린이집']))

final_kindergarten

final_kindergarten.reset_index(drop=True, inplace = True)
final_kindergarten

#####
####
##### final_kindergarten 데이터를 가지고 그룹핑해주기 #####

# 중심 국공립 어린이집 id 종류 확인

```

```

final_kindergarten['중심 국공립 어린이집'].unique()

# 기준 어린이집 id 각각에 인덱스 부여
final_kindergarten['중심 국공립 어린이집'].unique().tolist()

# 인덱스가 잘 부여됐나 확인
# 인덱스 번호가 0인 id=535으로 확인
final_kindergarten['중심 국공립 어린이집'].unique().tolist().index(535)

# temp_list에 저장

temp_list = final_kindergarten['중심 국공립 어린이집'].unique().tolist()
temp_list

# 그룹id인 "구분" 칼럼 생성
final_kindergarten.insert(0, '구분',final_kindergarten['중심 국공립 어린이집'].map(lambda x: temp_list.index(x)))
final_kindergarten

#####
#####
##### 그룹핑한 행을 전체 모아 어린이집 데이터에 넣어주기 #####

# 그룹id를 가진 데이터와 특징분류가 추가된 인천 어린이집 데이터 불러오기
df2 = pd.read_csv('데이터3.csv', encoding='euc-kr')

# 인천 어린이집 데이터에 중심 어린이집을 기준으로 한 "구분" 칼럼 추가
df3 = pd.merge(final_kindergarten, df2, left_on='중심 국공립 어린이집', right_on='id',how='right')
df3.columns

# 필요없는 칼럼 삭제
df3.drop(labels=['중심 국공립 어린이집', '주변 어린이집'],axis=1,inplace=True)
df3.head()

```

```
# 인천 어린이집 데이터에 주변 어린이집을 기준으로 한 "구분" 칼럼 추가
df4 = pd.merge(final_kindergarten, df3, left_on='주변 어린이집',
right_on='id',how='right')
# 주변 어린이집에 그룹핑 열 추가

df4.columns

df4.drop(labels=['중심 국공립 어린이집', '주변 어린이집'],axis=1,inplace=True)

# 구분_x = 주변 어린이집 , 구분_y = 중심 어린이집
df4

df4.shape

# 중복 id 제거
df4.drop_duplicates(['id'],inplace=True)
df4.shape

df4['구분_y'].notnull().sum() # 기준
df4['구분_x'].notnull().sum() # 주변

df4.to_csv('데이터5_1.csv')

## 이후 Excel 기능을 이용해 구분_x, 구분_y 통합한 컬럼 '그룹id'를 생성하여 '데이터5' csv 파일로 저장한다.
```