

수요예측 기반 돌봄센터 입지분석 빅데이터 분석결과

돌봄센터 수요예측 빅데이터 분석을 통해 초등인구수요를 예측하고 돌봄센터설치 우선지역을 도출함으로써 지역맞춤형 돌봄 지원계획 수립 지원
※ 관련근거 : 빅데이터 분석[아동청소년과-29819(2023.10.31.)]

I 분석 개요

- 분석명 : 수요예측 기반 돌봄센터 입지분석
- 분석기간 : 2024. 1. 2. ~ 2. 23.
- 분석범위 : 서대문구 전지역
- 분석내용
 - 우리시 전입, 전출, 연령별 정지인구¹⁾ 등을 통한 수요예측
 - 초등학교, 돌봄기관, 공동·단독주택, 수요지수 등을 통한 입지선정
- 수집데이터 목록

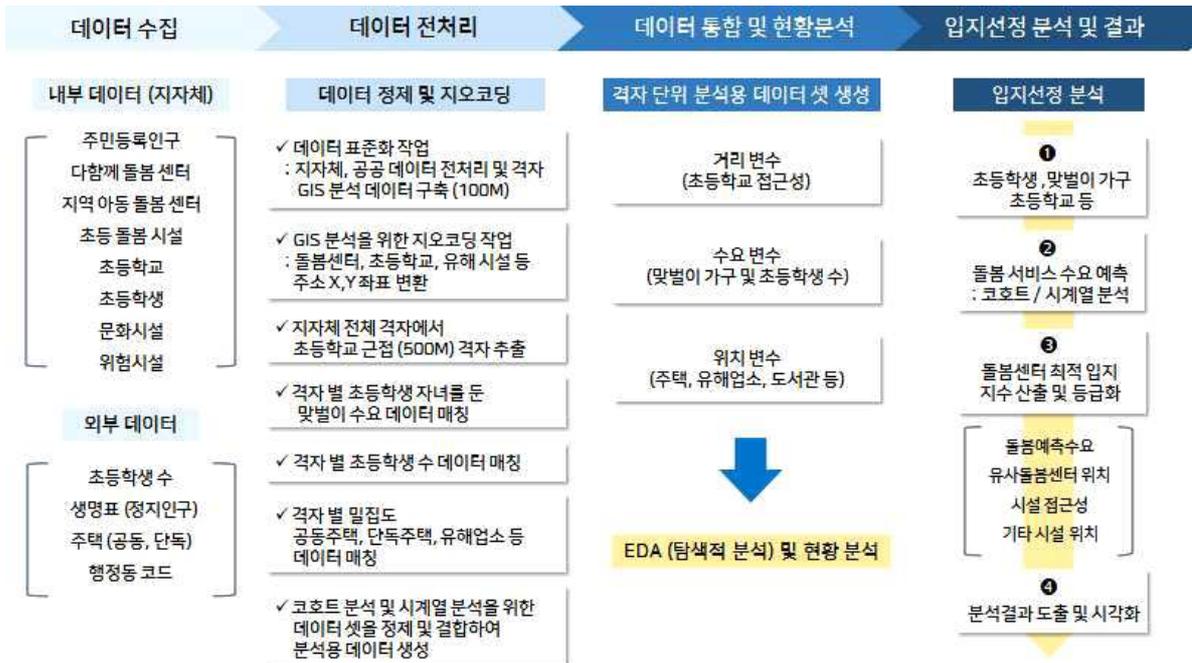
연번	구분	데이터명	내용	출처	기준연월
1	수요예측	인구이동통계	순 이동자수	통계청(공동활용데이터)	2023
2		정지인구	5~12세 정지인구	통계청	2023
3		주민등록 인구통계	연령별(6~12세)인구수	행정안전부	2019~2023
4	입지선정	초등학생 수	격자ID, 초등학생 수	국토지리정보원	2023
5		공동주택 현황	격자ID, 공동주택 수 등	국토지리정보원	2023
6		단독주택 현황	격자ID, 단독주택 수 등	국토지리정보원	2023
7		지역아동센터	시설명, 주소	자체(아동청소년과)	2023
8		다함께돌봄센터	시설명, 주소	자체(아동청소년과)	2023
9		초등학교	학교명, 주소	공공데이터포털	2023
10		유해업소	사업자명, 주소	자체(보건위생과)	2024
11	공통	행정구역 경계	행정동, 법정동 경계	국가공간정보포털	2023

1) 정지인구 : 출생 코호트 내 특정 연령의 생존자들이 다음 연령에 도달하는 동안에 생존해낼 것으로 기대

II

분석 방법

- 분석도구 : 헤안 분석참조모델²⁾, 파이썬(Python)³⁾, QGIS⁴⁾
- 분석프로세스
 - 다함께돌봄센터 이용 대상 만6~12세 아동의 수요를 예측하기 위해 비교적 정확성이 높은 코호트⁵⁾, 시계열(Prophet)⁶⁾ 모델을 활용하여 변동요인별 비교 분석
 - 본 모델에 사용되는 데이터는 SHP 형식과 엑셀 CSV 형식이며 개발 환경은 Python, 지도 시각화는 Q-GIS 툴, 그림은 PNG 파일로 구성됨
 - 본 모델에 포함된 내용 중 공공 데이터를 제외한 데이터는 서대문구 자체 데이터를 활용함



2) 헤안(빅데이터공통기반플랫폼) : 공공·민간의 데이터를 연계·수집·저장 분석하고 분석결과를 공유·활용하는 범정부 빅데이터 분석시스템

3) 파이썬(Python) : 데이터 전처리, 분석 및 시각화를 위한 프로그래밍 언어

4) QGIS(Quantum GIS) : 지리정보 기반의 데이터 시각화, 편집, 분석을 위한 공개소프트웨어

5) 코호트(Chohrt) 분석 : 특정기간 공통된 특성이나 경험을 갖는 집단 간의 행동 패턴을 비교 분석

6) 시계열 분석 : 시계열(시간의 흐름에 따라 기록된 것) 자료를 분석하고 여러 변수들 간 인과관계를 분석

○ 분석 내용 및 방법

① 데이터 전처리를 통한 분석 데이터 생성

- 데이터 정제, 변환, 결합 등 전처리를 통해 분석 데이터 생성
- (데이터 정제) 데이터들은 기준에 맞게 파일명 및 칼럼 명을 변경하여 표준화하고, 행정동 경계 데이터를 서대문구 읍면동에서 추출
- (데이터 변환) 도서관, 도시공원, 초등학교, 돌봄센터 등 위치 정보 좌표를 지오코딩⁷⁾ 후, Shape 파일로 변환
- (데이터 결합) 지자체에서 수집한 [지역아동돌봄센터], [다함께돌봄센터] 데이터와 교육부에서 수집한 [초등돌봄시설] 데이터를 지오코딩하여 X/Y 좌표로 변환한 후, '센터명' 과 'X/Y 좌표' 를 기준으로 결합하여 하나의 [유사돌봄센터] 데이터 생성

유사돌봄센터 데이터 결합 코드
<code>together = pd.read_csv('서대문구_다함께돌봄센터.csv', encoding = 'euc-kr') together</code>
<code>region = pd.read_csv('서대문구_지역아동센터.csv', encoding = 'euc-kr') region</code>
<code>element = pd.read_csv('서대문구_초등돌봄센터.csv', encoding = 'euc-kr') element</code>
<code>dolbom = pd.concat([together, region]) dolbom</code>
<code>dolbom = pd.concat([region, element], ignore_index = True) dolbom</code>
<code>dolbom = dolbom.fillna('-')</code>
<code>dolbom</code>
<code>dolbom.to_csv("서대문구_유사돌봄센터.csv", encoding = 'euc-kr')</code>

- 전체 인구이동 데이터에서 서대문구 전출/전입 데이터를 활용하여 5세~12세 [순 이동자 수] 데이터 생성

7) 지오코딩(Geocoding) : 고유명칭(주소나 산,호수의 이름등)으로 위도와 경도의 좌표 값을 얻는 것

- 서대문구 5세~12세 [순 이동자 수] 데이터와 [주민등록인구 수], [생명표(정지인구)]를 ‘읍면동 이름’ 과 ‘나이’ 를 기준으로 결합하여 코호트 분석용 데이터 생성
- 코호트 수요예측

```

만 5 ~ 12세 전출자 수 코드(일부)

move_out.columns = move_out.columns.str.replace('만나이', '')
move_out.columns = move_out.columns.str.replace('.', '')
move_out.columns = move_out.columns.str.replace('(', '')
move_out.columns = move_out.columns.str.replace(')', '')
move_out.columns = move_out.columns.str.replace('outemd', 'emd')
move_out = move_out[(move_out.전입자1 > 4)&(move_out.전입자1 < 13)|(move_out.전입자2 > 4)&(move_out.전입자2 < 13)|(move_out.전입자3 > 4)
|(move_out.전입자5 > 4)&(move_out.전입자5 < 13)|(move_out.전입자6 > 4)&(move_out.전입자6 < 13)|(move_out.전입자7 > 4)&(move_out.
|(move_out.전입자9 > 4)&(move_out.전입자9 < 13)|(move_out.전입자10 > 4)&(move_out.전입자10 < 13)]
move_out = move_out.fillna(0)
move_out['전입자1'] = np.where((move_out.전입자1 > 12) | (move_out.전입자1 < 5), 0, move_out.전입자1)
move_out['전입자2'] = np.where((move_out.전입자2 > 12) | (move_out.전입자2 < 5), 0, move_out.전입자2)
move_out['전입자3'] = np.where((move_out.전입자3 > 12) | (move_out.전입자3 < 5), 0, move_out.전입자3)
move_out['전입자4'] = np.where((move_out.전입자4 > 12) | (move_out.전입자4 < 5), 0, move_out.전입자4)
move_out['전입자5'] = np.where((move_out.전입자5 > 12) | (move_out.전입자5 < 5), 0, move_out.전입자5)
move_out['전입자6'] = np.where((move_out.전입자6 > 12) | (move_out.전입자6 < 5), 0, move_out.전입자6)
move_out['전입자7'] = np.where((move_out.전입자7 > 12) | (move_out.전입자7 < 5), 0, move_out.전입자7)
move_out['전입자8'] = np.where((move_out.전입자8 > 12) | (move_out.전입자8 < 5), 0, move_out.전입자8)
move_out['전입자9'] = np.where((move_out.전입자9 > 12) | (move_out.전입자9 < 5), 0, move_out.전입자9)
move_out['전입자10'] = np.where((move_out.전입자10 > 12) | (move_out.전입자10 < 5), 0, move_out.전입자10)
move_out = pd.melt(move_out, id_vars = ['emd'], value_vars = ['전입자1', '전입자2', '전입자3', '전입자4', '전입자5', '전입자6', '전입자7', '전입자8', '전입자9', '전입자10'],
var_name = 'out_value', value_name = 'age')
move_out = move_out[move_out['age'] > 0]
move_out = move_out.groupby(by=['emd', 'age'], as_index=False)['out_value'].count()

```

- 2018년도부터 2023년까지의 서대문구 만 6세~12세 [초등학생 수] 데이터를 ‘읍면동 이름’ 을 기준으로 결합하고, 이상치를 필터링하여 시계열 분석용 데이터 생성
- 시계열 수요예측

```

Prophet(시계열) 분석 코드(일부)

child = year_child[year_child['emd_nm'] == emd[0]]
child = child.loc[:, ['ds', 'y']]
prophet = Prophet(yearly_seasonality=True,
weekly_seasonality=False,
daily_seasonality=False,
changepoint_prior_scale=0.05)
prophet.fit(child)
future = prophet.make_future_dataframe(periods=3, freq='y')
forecast = prophet.predict(future)
fore = forecast.iloc[[-3, -2, -1], [0, 15]]
fore.columns = ['ds', 'y']
result = fore.copy(deep=True)
result['emd_nm'] = emd[0]
for i in range(1, len(emd)):
    child = year_child[year_child['emd_nm'] == emd[i]]
    child = child.loc[:, ['ds', 'y']]
    prophet = Prophet(yearly_seasonality=True,
weekly_seasonality=False,
daily_seasonality=False,
changepoint_prior_scale=0.05)
    prophet.fit(child)
    future = prophet.make_future_dataframe(periods=3, freq='y')
    forecast = prophet.predict(future)
    fore = forecast.iloc[[-3, -2, -1], [0, 15]]
    fore.columns = ['ds', 'y']
    result2 = fore.copy(deep=True)
    result2['emd_nm'] = emd[i]
result = pd.concat([result, result2], axis = 0)

```

- 시군구마다 공동주택과 단독주택의 밀집도를 비교하여 도심과 농촌 중 알맞은 입지선정 분석 코드를 사용함
- 공동주택의 밀집도가 단독주택 밀집도보다 크면 도심, 공동주택 밀집도가 단독주택 밀집도보다 작으면 농촌의 입지선정분석 코드를 사용
- 초등학교 접근성 (500M 격자), 유사돌봄센터 접근성 (500M 격자), 격자 단위 맞벌이 데이터 (초등학생 수 * 맞벌이 비율), 격자 단위 공동/단독주택 수 데이터를 격자 ID를 기준으로 결합하여 하나의 데이터 셋을 생성
- 완성된 격자 데이터 셋을 돌봄센터 입지선정분석에 사용
- 돌봄센터 입지선정 분석(도심/농촌 - 특성에 따른 분석 진행)

도심 돌봄센터 입지선정 분석 코드(일부)
<p>맞벌이 비율 입력</p> <pre>work = (0.4694)</pre>
<p>데이터정제</p> <pre>child = gpd.read_file('data/03_돌봄센터입지선정/child.shp', encoding = 'euc-kr') apt = gpd.read_file('data/03_돌봄센터입지선정/apt.shp', encoding = 'euc-kr') dandok = gpd.read_file('data/03_돌봄센터입지선정/dandok.shp', encoding = 'euc-kr')</pre>
<p>맞벌이</p> <pre>child = child.drop(['lbl', 'axis = 1]) child.rename(columns = {'val' : 'child'}, inplace = True) child['child'] = child['child'] + work</pre>
<p>주거밀집(공동,단독)</p> <pre>apt = apt.drop(['lbl', 'geometry'], axis = 1) apt.rename(columns = {'val' : 'apt'}, inplace = True) dandok = dandok.drop(['lbl', 'geometry'], axis = 1) dandok.rename(columns = {'val' : 'dandok'}, inplace = True) temp = pd.merge(child, apt, on = ['gid'], how = 'left') temp = pd.merge(temp, dandok, on = ['gid'], how = 'left') temp = temp[['gid', 'geometry', 'child', 'apt', 'dandok']]</pre>
<p>초등학교 접근성</p> <pre>school = gpd.read_file('data/03_돌봄센터입지선정/school.shp', encoding = 'euc-kr') school = school.loc[:, ['school', 'geometry']] school['geometry'] = school.buffer(500) intersection = gpd.overlay(temp, school, how = 'intersection') intersection['school'] = intersection['geometry'].area intersection['school'] = round(intersection['school'], 2) intersection = intersection[intersection['school'] != 0] intersection = intersection.groupby(by=['gid'], as_index=False)['school'].sum() temp1 = pd.merge(temp, intersection, on = ['gid'], how = 'left')</pre>

② 데이터 분석 도구

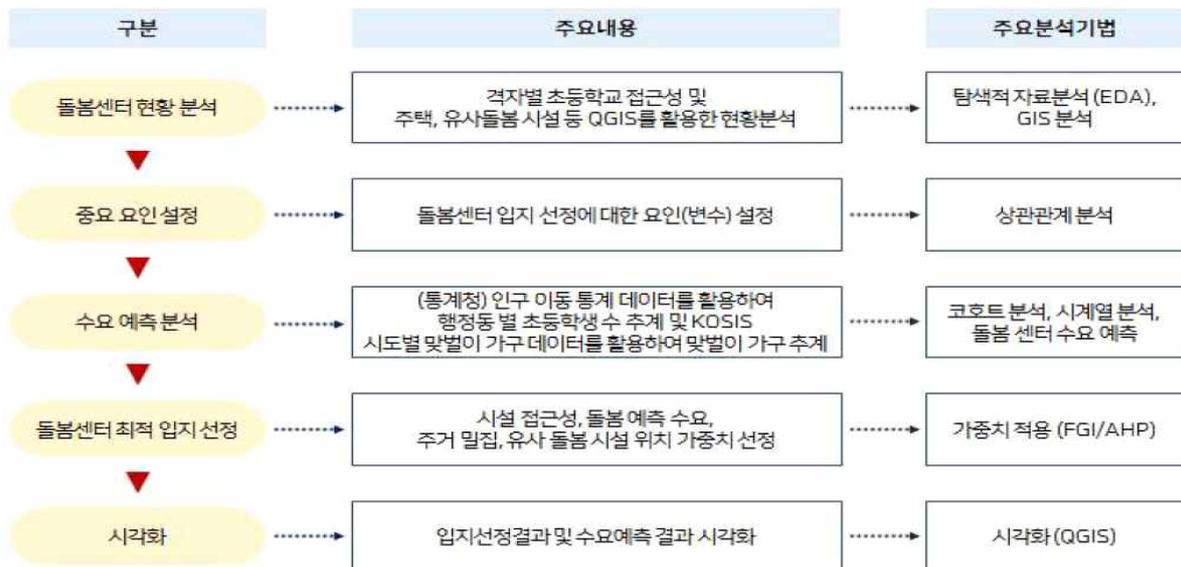
- 중요 요인 도출, 입지분석 등의 분석 및 시각화를 개발도구 Python, Q-GIS를 활용하여 진행함

Anaconda	QGIS
 <p>수학과 과학 분야에서 사용되는 여러 패키지를 묶어 놓은 파이썬 배포판으로서 Numpy, Matplotlib, Pandas 등을 포함. 특히 최근에 데이터 사이언스와 머신러닝 분야에서 파이썬을 사용하기 위해 기본적으로 설치하는 배포판</p>	 <p>데이터 뷰, 편집, 분석을 제공하는 크로스 플랫폼 자유-오픈 소스 데스크톱 지리 정보 체계(GIS) 응용 프로그램으로, 본 연구에서는 입지분석 결과를 지도 시각화 도구로 활용됨</p>

③ 표준분석모델을 활용한 입지분석

- 지역 내 공적 돌봄 서비스 현황 및 각 지원사업 간 연계를 감안하여 다함께 돌봄센터가 균형있게 분포되도록 설치하기 위해 수요가 많고 이용 아동이 학교 혹은 집을 기준으로 이동(도보)하기 편한 곳으로 최적 입지 분석 실시
- FGI(Focus Group Interview, 심층 집단면접), AHP(Analytic Hierarchy Process, 분석적 계층 프로세스)를 통하여 산정된 지수별 가중치를 최적입지점수에 반영

$$\text{최적입지점수} = (\text{접근성}^{\text{초등학교}} \text{지수} * 39.468) + (\text{수요}^{\text{초등학생}} \text{지수} * 26.247) + (\text{주거밀집}^{\text{공동7:단독3}} \text{지수} * 17.352) - (\text{유사돌봄지수} * 16.933)$$



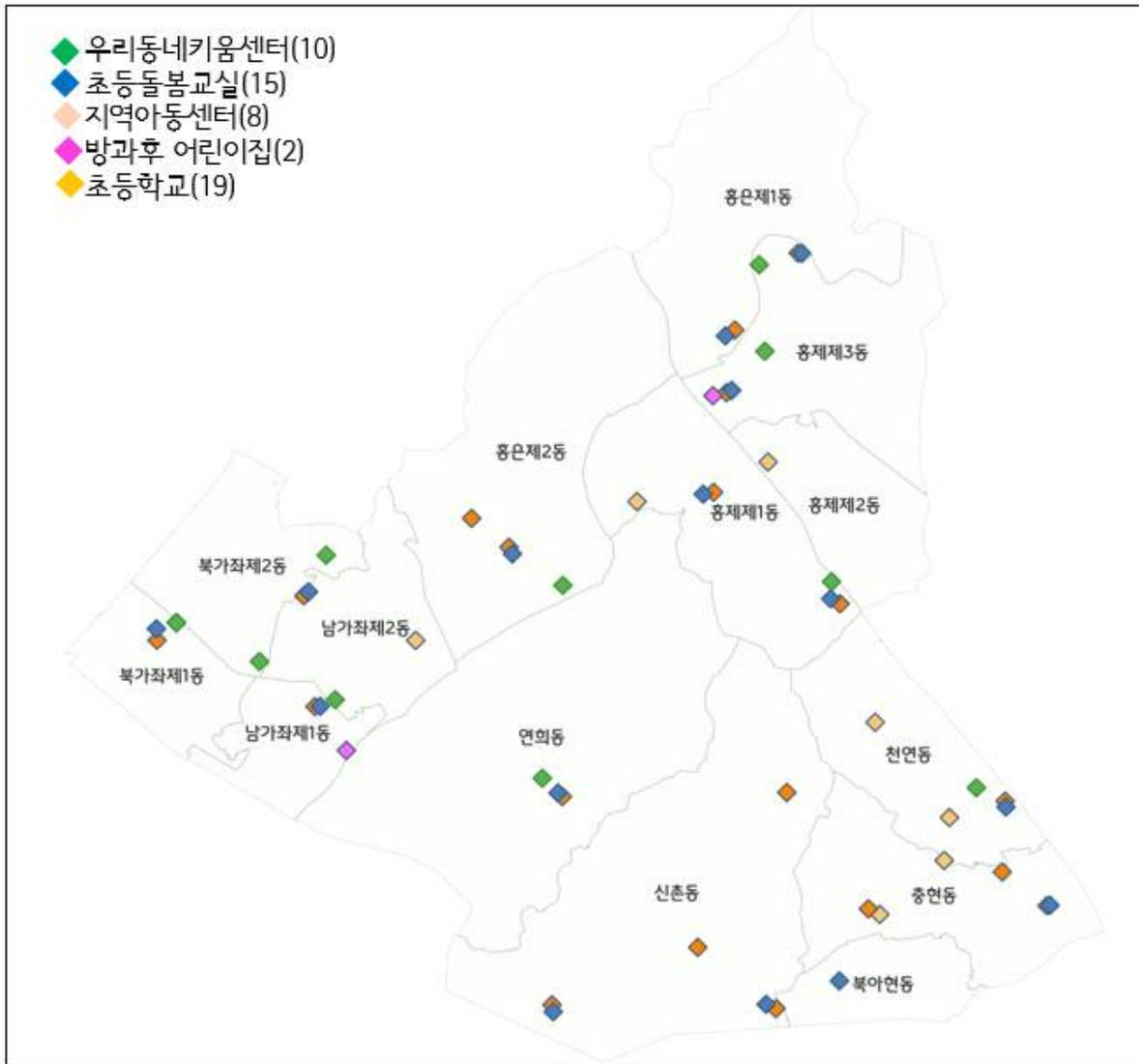
Ⅲ

현황분석

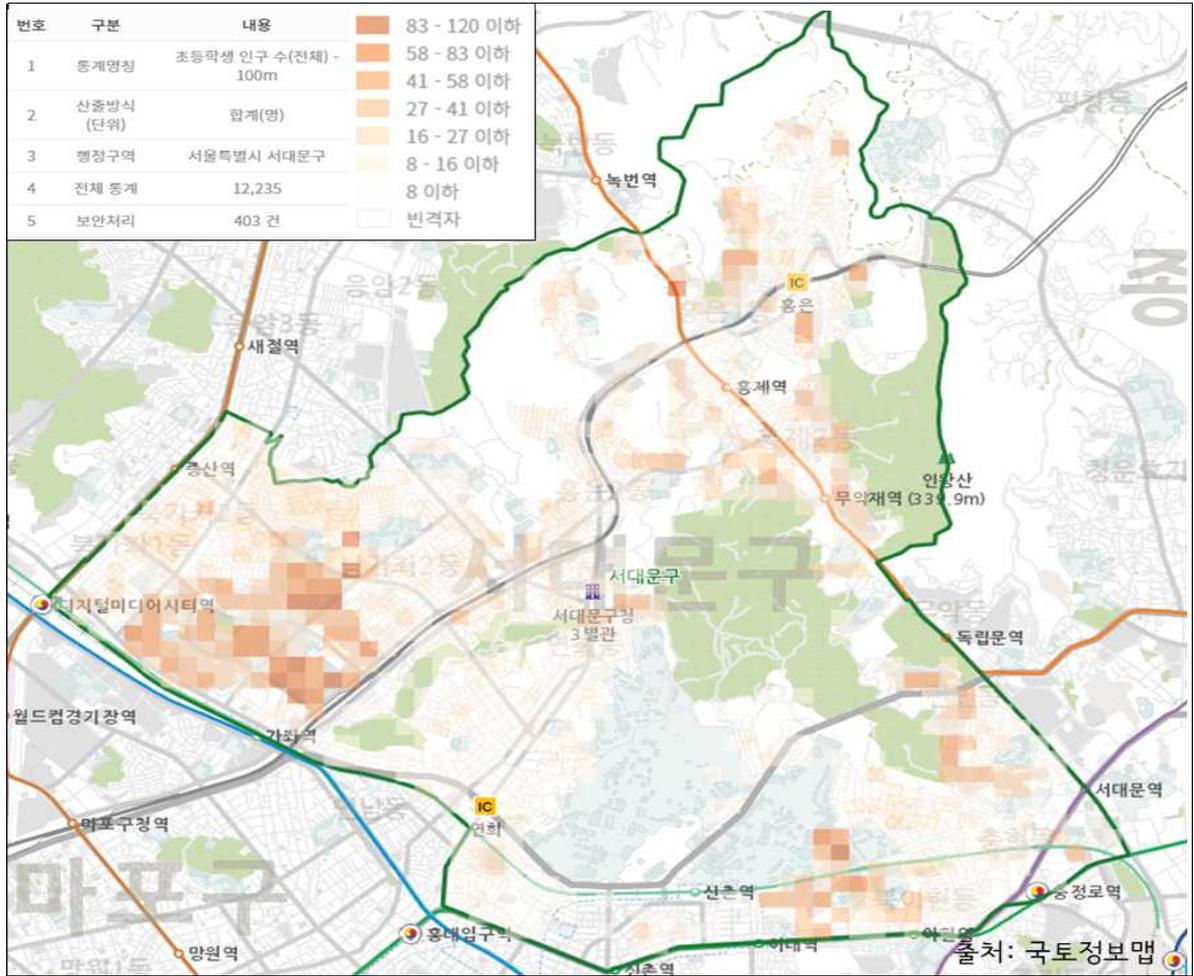
○ 초등학교 및 돌봄시설 현황

- 초등학교 19개소(부속학교포함)
- 유사돌봄시설 총35개소(초등돌봄교실 15, 우리동네키움센터 10, 지역아동(복지)센터 8, 방과후교실어린이집 2)
- 우리동네키움센터는 초등학교와 같이 지역별 고루 분포되어 있음

|| 초등학교 및 유사돌봄센터 현황 ||



|| 서대문구 초등학생 인구 수(2023년10월기준) ||



구분	5~12세 아동	돌봄센터	초등학교 (사립제외)	돌봄센터1개당 아동수	학교1개당 아동수
총현동	1012	3	1	337.3	1012
천연동	856	4	1	214	856
북아현동	1009	1	1	1009	1009
신촌동	424	2	2	212	212
연희동	1277	2	1	638.5	1277
홍제1동	1250	4	2	312.5	625
홍제2동	569	1	-	569	-
홍제3동	815	5	2	163	407.5
홍은1동	1123	1	1	1123	1123
홍은2동	1472	2	1	736	1472
남가좌1동	2021	2	1	1010.5	2021
남가좌2동	2119	3	1	706.3	2119
북가좌1동	1243	2	1	621.5	1243
북가좌2동	1250	3	-	416.6	-

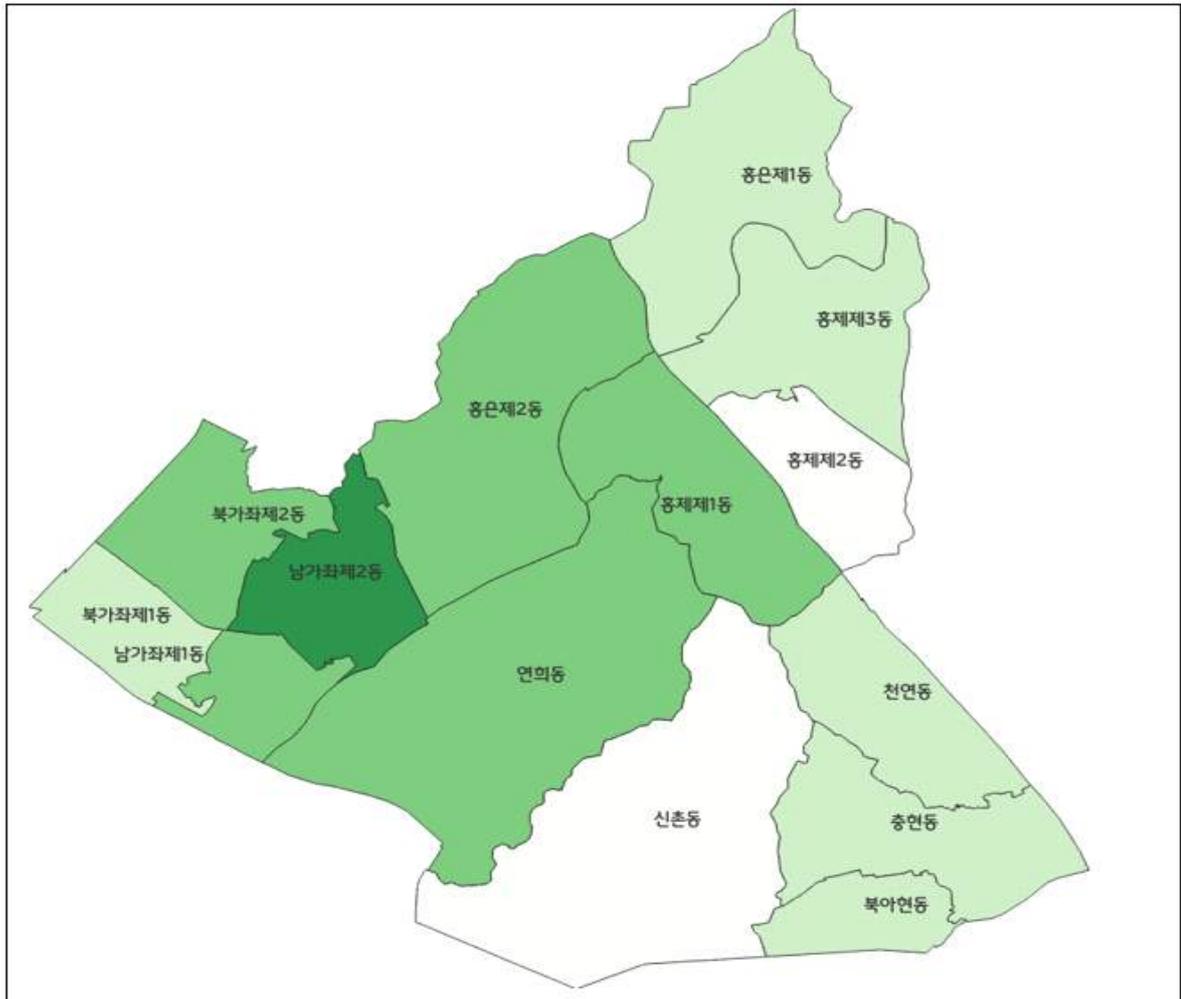
IV

입지분석결과

□ 수요예측분석

- 코호트 및 시계열 수요예측 결과, 향후 돌봄 수요가 많을 것으로 예상되는 지역은 남가좌제2동 > 홍은동 > 연희동 순으로 나타남

|| 코호트 수요예측결과 시각화 ||



※ 색이 진할수록 수요가 높은 지역임

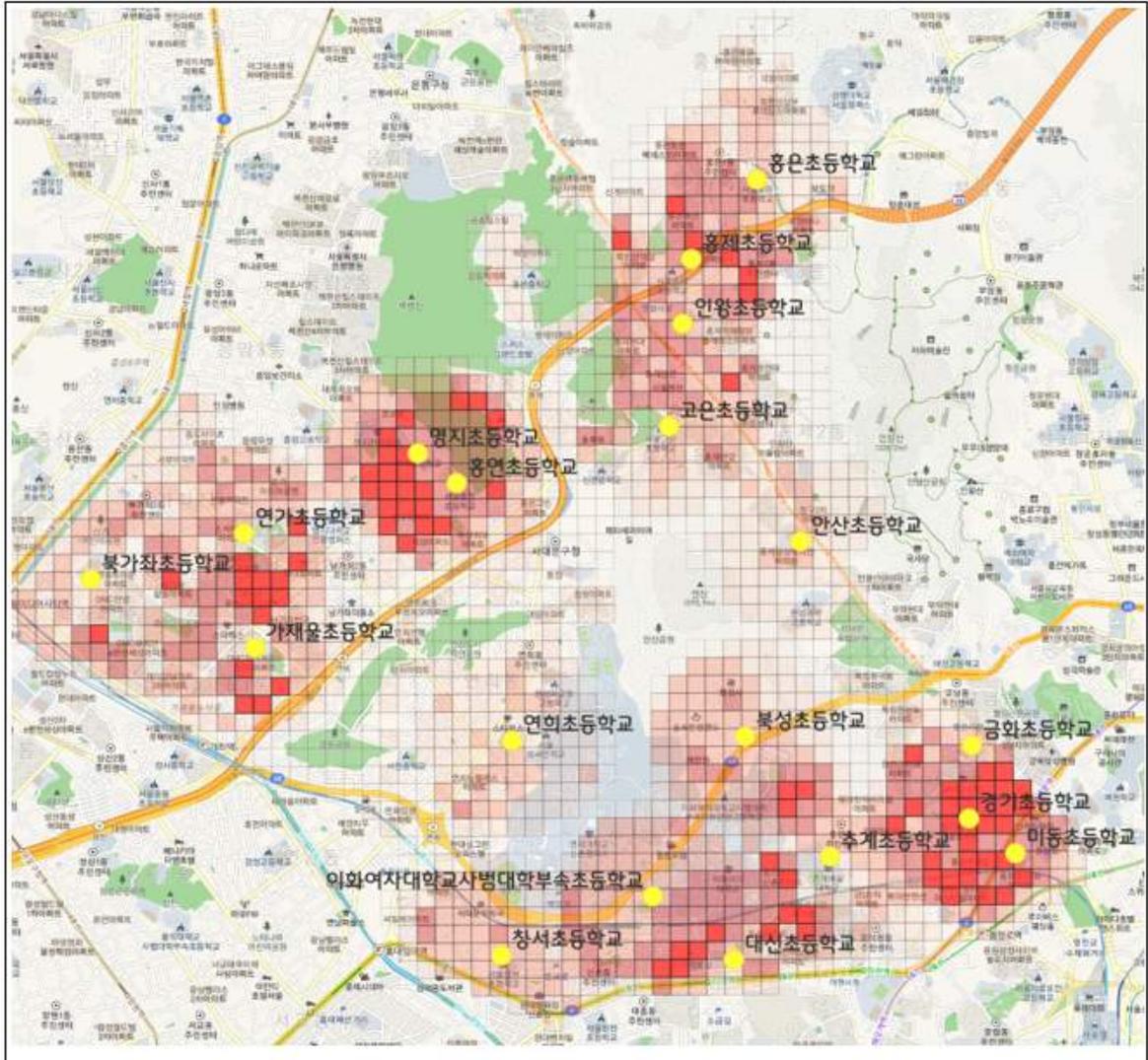
순위	구분	한계 (코호트+ 시계열평균)	코호트 분석결과	시계열(2024) 분석결과	시계열(2025) 분석결과	시계열(2026) 분석결과
1	남가좌제2동	4,712	1,977	2,650	2,730	2,825
2	남가좌제1동	4,155	1,871	2,134	2,283	2,435
3	홍은제2동	2,560	1,305	1,155	1,248	1,361
4	홍제제1동	2,505	1,140	1,276	1,363	1,457
5	연희동	2,423	1,186	1,297	1,217	1,198

□ 입지선정분석

① 분석결과

- 최적입지점수가 높은 지역들이 홍은2동, 충현동, 남가좌2동에 집중되어 있음

|| 입지선정 분석결과(입지선정(1~100위)) ||



※ 색이 진할수록 최적입지점수가 높은 지역임

- 전체 분석결과 중 100위 격자만 선택하여 등급분류
- 대단지아파트 밀집지역 및 초등학교 인접지역에 돌봄수요가 높을 것을 볼 수 있음

|| 입지선정 분석결과(입지선정(1~10위)) ||



순위	구분	최적입지 점수	접근성 지수	수요 지수	주거밀집 지수	유사돌봄지수
1	e편한세상홍제가든 플라즈@	399.06	118.40	262.47	52.06	33.87
2	건우,서호빌라 인근	361.64	394.67	0.00	34.70	67.73
3	신촌푸르지오@	343.90	236.80	157.48	17.35	67.73
4	힐스테이트신촌@	343.71	157.87	236.22	17.35	67.73
5	DMC쌍용스윗닷홈@	328.66	236.80	209.98	17.35	135.47
6	서대문성당 인근	327.77	394.67	0.00	34.70	101.60
7	서대문역인근	326.94	394.67	0.00	0.00	67.73
8	돈의문센트레빌@	320.15	394.67	26.25	34.70	135.47
9	DMC에코자이@	319.35	236.80	183.73	17.35	118.53
10	래미안루센티아@	319.35	236.80	183.73	17.35	118.53

② 최적위치 선정

○(후보지 1) 홍은제동 104번지 인근(e편한세상홍제가든플라츠 일대)

- 1순위는 5세~12세 아동 수요점수가 최고점으로 해당 격자 내에 아동의 수가 많으며 코호트, 시계열 분석을 통한 수요예측결과 미래 돌봄 수요가 가장 높은 지역으로 나타남
- 특히, 해당 지역은 e편한세상홍제가든플라츠아파트, 북한산더샵아파트, 홍은풍림아파트, 홍은더아더프라임아파트 등과 같은 대단지아파트가 밀집되어 있고, 홍제초등학교가 있어 돌봄 수요가 매우 높음
- 다만, 500m 반경 내외로 유흥·단란주점이 있음(북한산더샵@기준)



순위	구분	최적입지 점수	접근성 지수	수요 지수	주거밀집 지수	유사돌봄지수
1	e편한세상홍제가든@	399.06	118.40	262.47	52.06	33.87
18	홍은백산@	300.72	236.80	131.24	17.35	84.67
19	신홍빌라, 해주빌라 인근	300.70	236.80	78.74	86.76	101.60
20	홍은 무명, 대우빌라 인근	298.17	236.80	26.24	52.05	16.93

○ (후보지 2) 충청로2가 인근(건우,서호빌라 인근)

- 해당지역은 돈의문센트레빌아파트와 다수의 빌라 및 연립주택이 밀집되어 있고, 500m 반경 내에 다수의 초등학교가 있어 돌봄 수요가 높음
- 다른 후보지에 비해 수요지수가 다소 낮고 인근에 지역아동센터가 2곳 있지만, 초등학교 접근성지수가 높음
- 다만, 1km 반경 내외로 유흥·단란주점이 있음(돈의문센트레빌@기준)



순위	구분	최적입지 점수	접근성 지수	수요 지수	주거밀집 지수	유사돌봄지수
2	건우,서호빌라 인근	361.64	394.67	0.00	34.70	67.73
6	서대문성당 인근	327.77	394.67	0.00	34.70	101.60
7	서대문역인근	326.94	394.67	0.00	0.00	67.73
8	돈의문센트레빌@	320.15	394.67	26.25	34.70	135.47

- (후보지 3,4) 북아현동 1011번지 일대(신촌푸르지오아파트 인근)
 - 해당 지역은 힐스테이트신촌아파트, 신촌푸르지오아파트, 두산아파트, 럭키대현아파트 등과 같은 대단지 아파트가 밀집되어 있고, 서울북성초등학교가 있어 돌봄 수요가 높음
 - e편한세상신촌아파트는 신촌푸르지오아파트와 500m정도의 거리가 있지만, 인접한 대단지아파트로서 같은 지역으로 고려해볼 수 있음



순위	구분	최적입지 점수	접근성 지수	수요 지수	주거밀집 지수	유사돌봄지수
2	신촌푸르지오@	343.90	236.80	157.48	17.35	67.73
3	힐스테이트신촌@	343.71	157.87	236.22	17.35	67.73

○ (후보지 5) 남가좌동 380번지 일대(DMC쌍용스윗닷홈2@ 인근)

- 남가좌제2동은 코호트, 시계열 분석을 통한 수요예측결과 미래 돌봄 수요가 매우 높은 지역이며, 흥은제2동과 인접한 지역임
- 해당 지역은 DMC에코자이아파트, 래미안루센티아아파트, DMC쌍용스윗닷홈아파트, 남가좌 현대아파트와 같은 아파트가 밀집되어 있으며, 가재울초등학교와 연가초등학교가 있어 돌봄 수요가 높음



순위	구분	최적입지 점수	접근성 지수	수요 지수	주거밀집 지수	유사돌봄지수
5	DMC쌍용스윗닷홈2	361.64	394.67	0.00	34.70	67.73
9	DMC에코자이@	327.77	394.67	0.00	34.70	101.60
10	래미안루센티아@	326.94	394.67	0.00	0.00	67.73

③ 돌봄 사각지대 중 최적 입지 후보지

○ 유사돌봄지수가 0인 격자 중 최적입지점수가 높은 순으로 산출하여 돌봄시설 설치가 필요한 지역 산출 ⇒ 홍은제2동이 적절한 후보지로 도출됨



순위	구분	최적입지 점수	접근성 지수	수요 지수	주거밀집 지수	유사돌봄지수
1	현대아트빌라 인근	214.06	118.40	26.25	69.41	0.00
2	충현한양홈타운인근	209.34	197.34	5.00	7.00	0.00
3	백련사~가좌로2길 인근	195.57	157.87	3.00	34.70	0.00
4	신촌봉원교회 인근	179.35	118.40	26.25	34.70	0.00
5	코오롱아파트 인근	179.35	118.40	26.25	34.70	0.00

V

한계점 및 향후분석방향

- 도서관, 도시공원, 유해시설 등의 다양한 변수들이 입지분석에 함께 활용된다면 보다 안전하고 활용적인 센터 입지를 선정할 수 있을 것으로 기대됨
- 지역적 특성 및 인구유입 등 유동인구에 대한 데이터가 추가적으로 활용되어 특징에 맞게 가중치를 조절하면 더욱 유의미한 결과를 얻을 것으로 기대됨

VI

행정사항

- 분석결과 제공 및 활용방안 검토(아동청소년과)
 - 신규 돌봄시설 설치 및 서대문구 지역맞춤형 돌봄 지원계획 수립시 활용
- 다함께 돌봄센터 위치 선정 최적화
 - 초등학교 근거리, 초등학생 수, 맞벌이 가구 및 유사돌봄시설 분포를 통한 수요예측 가능
 - 유해 및 위험시설, 문화시설 등의 위치를 고려해 신규 우선 설치 지역을 선정하고 돌봄 서비스를 제공하여 합리적 예산 활용
- 아동 관련 정책 개발 시 활용
 - 돌봄센터 외에도 아동 관련 정책 개발 시 기반 데이터로 활용 가능하여 이용자 편의 증대 가능