

|  |
| --- |
| 데이터 분석 기초  5조 보고서  2018 |
| 분석 주제 - 노인복지 거점 선정 |
| 분석 기간: 2025.4.8. ~ 2025.6.17.  팀명: 어벤져스  팀원: 김민혁(조장), 조주성, 최원경, 장윤경 |



전자공학부 인공지능전공

# 데이터 처리 및 분석 설계

### 데이터 전처리와 EDA, 시각화 설계

1. 데이터 가공(전처리)

본 프로젝트에서는 대구광역시 북구 지역을 대상으로 시니어 클럽 거점 후보지를 선정하고자 다양한 공공 데이터를 수집 및 전처리하였다. 주요 데이터는 크게 세 가지 범주로 구성되었다. 첫째, 행정동별 독거노인 수를 포함한 인구통계 데이터로, 고령 인구의 분포를 분석하는 데 사용되었다. 둘째, 노인 요양시설 및 복지기관의 위치 데이터(silver\_center, welfare\_center)로, 복지 시설 간의 거리 및 밀집도를 파악하는 데 활용되었다. 마지막으로 시내버스 정류소, 지하철 출구(sub1\_df, sub2\_df, sub3\_df), 병·의원(med\_df) 위치 데이터로, 각 시설의 물리적 접근성을 분석하기 위한 핵심 자료가 되었다.

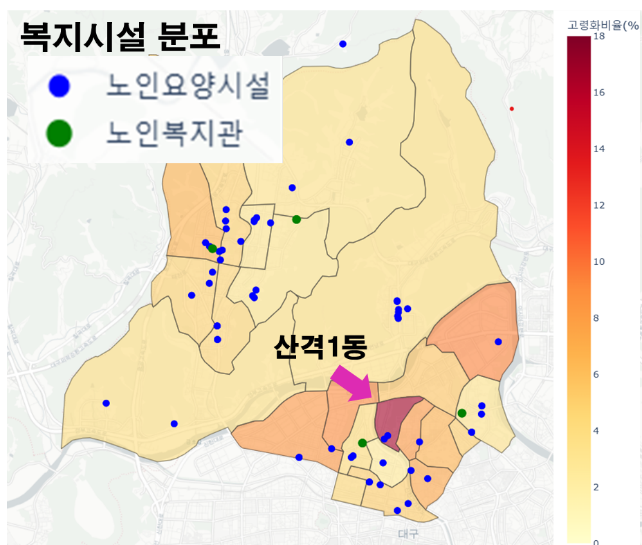
데이터는 모두 2년 이내의 최신 자료만을 선별하여 분석의 현실성과 신뢰성을 확보하고자 하였으며, 위도·경도 좌표 정보를 포함하여 정확한 거리 계산이 가능하도록 하였다. 전처리 과정에서는 결측값을 탐지하여 제거하거나 평균, 중위값 등으로 대체하였다. 좌표 정제 및 지역 매칭 과정도 함께 진행되어 각 시설 데이터를 행정동 단위로 통합 및 정렬하였다. 텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

<데이터 병합 코드>

<데이터 병합 코드>와 같이 요양시설 및 복지시설 데이터를 통합하여 welfare\_combined\_df.csv라는 하나의 통합 복지시설 데이터를 생성하였다. 지하철 데이터(sub1\_df~sub3\_df)는 병합하여 subway\_df로 구성하였으며, 버스(bus\_df) 및 병원(med\_df) 데이터도 병합하여 접근성 평가에 사용할 수 있도록 통일된 형태로 정리하였다.

1. 분석 설계

<북구 복지시설 분포와 고령화 비율 히트맵>과 같이 북구 행정동의 고령화 비율과 북구의 복지시설의 분포를 히트맵으로 나타내어 노인 인구 밀집 지역과 복지시설 분포를 시각적으로 확인하였다.

<북구 복지시설 분포와 고령화 비율 히트맵>

텍스트, 지도, 아틀라스, 폰트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

<버스 정류장과 요양시설 거리기반 클러스터링 결과>

<버스 정류장과 요양시설 거리기반 클러스터링 결과>를 보았을 때 이러한 클러스터링 기반 분석은 물리적 분포는 잘 들어내었다. 하지만<북구 복지시설 분포와 고령화 비율 히트맵>과 <버스 정류장과 요양시설 거리기반 클러스터링 결과>의 결과로는 실제 거점 선정에는 한계가 있다. 접근성, 지역 수요, 정책적 중요도를 반영하지 못했기 때문에 분석의 목적을 충분히 충족시키지 못하는 것으로 판단되었다.

이에 따라, 보다 정량적이고 정책 목적에 부합하는 방식으로 전환하고자 접근성 점수 기반 분석 체계를 도입하였다. 먼저, 각 시설의 위도와 경도 값을 라디안으로 변환하고, 이를 기반으로 3차원 데카르트 좌표로 바꾸어 정확한 거리 계산이 가능하도록 하였다. 이후 KD-Tree 기반의 최근접 이웃 탐색 알고리즘을 통해 각 복지시설이 주변 교통 및 의료 인프라와 얼마나 가까운지를 계산하였다. 계산된 거리값은 지구 반지름(6,371,000m)을 곱하여 유클리드 거리로 환산되었으며, 거리가 가까울수록 높은 점수를 부여하는 방식으로 접근성 점수를 부여하였다.

이 접근성 점수는 버스, 지하철, 병원에 대해 최소-최대 정규화(MinMax Scaling)를 적용하여 0~1 범위로 정규화되었고, 서로 다른 지표 간의 비교가 가능하도록 하였다. 그러나 물리적 거리만으로는 독거노인의 고립 문제와 같은 사회적 문제를 완전히 설명하기 어렵기 때문에, 각 행정동의 독거노인 수를 함께 반영한 다중 요인 점수화 체계를 설계하였다.

이를 위해 복지시설의 주소에서 행정동 이름을 추출하여 독거노인 수 데이터와 병합하고, 정규화 과정을 거쳐 독거노인 점수 항목을 생성하였다. 이후 접근성 점수와 독거노인 점수를 병합하여 통합 점수를 산출하였다. 정책적으로 중요도가 높은 지역인 산격1동의 경우에는 독거노인 점수에 2배의 가중치를 부여하여 그 영향력을 강화하였다.

최종 통합 점수는 거리 기반 접근성과 고령 인구 비율이라는 두 가지 기준을 종합적으로 고려하여 산출된 가중 평균 점수로, 이후 시니어 클럽 후보지를 선정하는 데 직접적으로 활용되었다.

1. EDA 및 시각화 구상

탐색적 데이터 분석(EDA)은 시각적 인사이트 확보와 의사결정 기반 마련을 위해 다양한 형태로 진행되었다. <북구 복지시설 분포와 고령화 비율 히트맵>와 같이 대구 북구 행정동별 고령자 인구 분포를 히트맵으로 시각화하여, 고령 인구가 집중된 지역을 시각적으로 파악하였다. 복지시설과 교통 인프라(지하철, 버스, 병원)는 지도 위에 마커 형태로 시각화를 진행하는 것이 좋다고 판단했다. 이러한 내용을 기반으로 EDA 및 시각화를 진행했다.

# 최종 시각화 기반 거점 비교 분석

## 맵핑, 스코어링 기반 거점 선정 과정 & EDA 정리

1. 최종 스코어링

본 분석의 목적이 노인 복지 활동을 진행할 최적의 거점을 선정하는 것이므로, **독거노인의 분포가 많은 지역에 위치한 시설에는 총점수에 가중치를 추가로 부여**하여 최종 점수를 산출했다. 앞서 말했듯이, 사전 조사를 통해 알아낸 독거노인 분포가 높은 지역에 높은 가중치를 할당 시켰었고, 이를 각 시설에 적용하기 위해 거리기반 점수를 반영한 welfare\_scored\_df 칼럼에서 “행정동”을 필터링하여 각 행정동 별로 할당된 가중치를 각각 매칭하여 다음의 수식을 적용하여 총 점수를 계산했다.

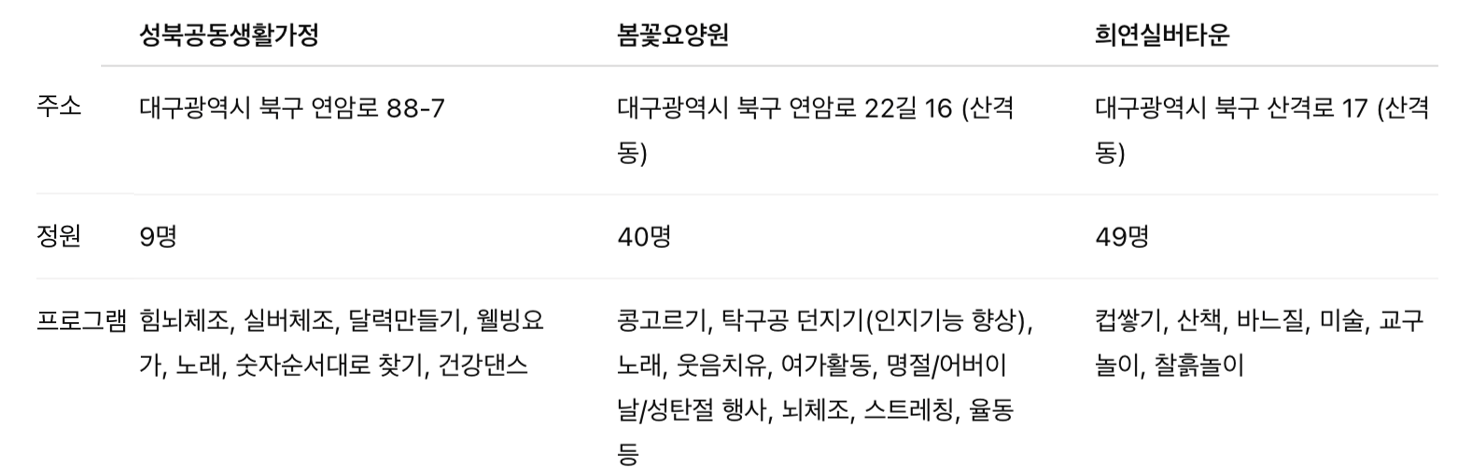
이 총점을 반영한 데이터프레임을 top3\_df 로 저장하고, 총점수를 기준으로 내림차순 정렬하여 상위 3곳을 출력한 결과, 다음과 같이 출력되었다.



<top3\_df 를 ascending=False 로 출력한결과>

분석 결과, 독거노인 분포가 가장 높았던 산격동에 위치한 세 요양시설이 출력되었고, 이 세 시설을 기반으로 거점을 선정하는 것이 타당하다는 결론이 도출되었다. 위 세 시설은 이미 거리기반 점수 및 독거노인 점수가 높은 시설들이므로, 교통 및 의료를 제외하고 거점 선정 방식으로 활동 공간의 유무, 수용가능 인원, 진행중인 활동 등에 중점을 두었다. 먼저, 조사를 통해 수용가능 인원 및 현황, 시설 별 진행중인 프로그램, 활동 공간 등을 찾아보고 간단한 비교를 해보았다.

1. 거점 후보지 비교



<정원 및 프로그램 비교>

먼저, 수용가능 인원 및 프로그램 조사인데, 여기서 성북공동생활가정은 시설 특성 상 어르신들이 출퇴근과 비슷한 형태로 오가며 지내는 간단한 가정 형식의 시설이었기에, 단체 활동에 적합하지 않은 정원을 갖췄다는 것을 확인했다. 또한, 모든 시설에서 진행중인 프로그램이 노인들의 건강과 관련하여 치매예방 및 재활 등에 집중되어 있었기에, 독거노인들의 고립감을 해소하기 위한 프로그램 개선이 필요할 것이라고 판단했다.



<시설 및 주변 환경 비교>

다음은 시설 조사이다. 성북공동생활가정은 바로 이전 조사에서 프로그램 공간이 협소하다는 점에서 거점 후보로 알맞지 않음을 확인했고, 희연실버타운은 타 시설들에 비해 비교적 대중교통 접근성이 약 2~3블럭 가량 떨어졌고, 타 시설들에 비해 점수도 낮았고 프로그램 환경에 큰 플러스 요소로써 고려되는 점들이 별로 없었다.



<봄꽃요양원 지하1층 소강당 공간>

반면에 봄꽃요양원은 지하 1층에 소강당 처럼 큰 공간이 있어 요양시설과 별개로 독립적인 활동공간이 있었고, 복지시설이 밀집되어 있는 지역에 위치하였으며, 바로 근처에 버스정류장과 의료시설이 다수 분포해 있어 다측면에서 접근성이 우수하여 거점 후보로 가장 알맞다고 판단하였다.

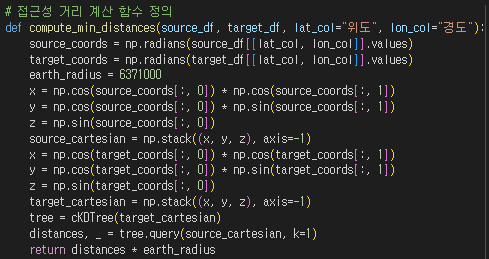


<봄꽃요양원 활동게시판>

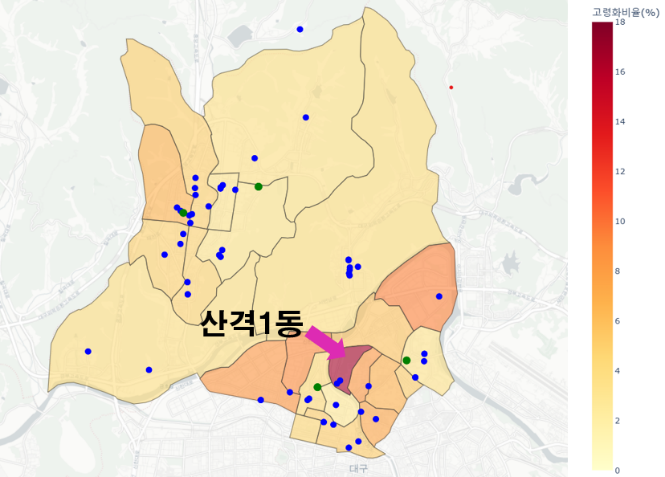
또한, 지하에 있는 넓은 공간 외에도 잘 보이는 곳에 활동 게시판이 있어, 주간, 월간 활동 등에 있어 지속적인 업데이트와 교류, 친목이 가능하고, 점수제나 최근 뉴스 기재 등을 도입하기에 매우 적절해 보였다.

1. 거점 선정과정 정리 - EDA (기술적 - *technological* 측면)

다음은 최종 결과를 도출하기 직전까지의 과정에서 필요하였으며, 필요에 따라 사용한 방식들의 기술적인 측면이다.

옆은 거리 기반 접근성 점수를 매길 때 사용한 거리 계산 함수이다. 최근접 이웃 탐색에 유용한 KDTree를 scipy 로부터 가져와 사용했으며, 실제 척도, 지구 표면상 거리를 고려하여 정확한 거리를 재기 위해 계산한 거리에 지구의 반지름을 곱한 값을 리턴한다.

이 함수를 기반으로 버스, 지하철역, 의료시설과의 접근성 거리를 계산하였고, pandas 라이브러리의 dropna 함수를 통해 결측치를 처리한 후, 거리가 추가된 welfare\_combined\_df 의 각 칼럼에 sklearn의 MinMaxScaler를 적용하여 0과 1사이의 값으로 정규화 하여 이를 점수 산출에 사용하였다. 최종점수는 정규화 된 각 컬럼의 거리의 평균을 1에서 뺀 값으로, 거리기반 최종 점수에 해당하는 점수이다. 이렇게 산출된 점수와 함께 folium 의 지도 생성 함수를 통해 시설들이 마커로 표시된 지도를 생성하였고, geoJSON을 활용한 독거노인 분포 지도와 결합하여 히트맵과 함께 분석하였다.

시설의 좌표값을 토대로 지도 위에 겹치게 표시하여 folium 맵과 비교하였고, 지도에 히트맵을 결합하여 독거노인 비율의 분포를 나타내었고, 그 결과 산격1동의 비율이 가장 높았다. 이를 통해 독거노인 점수를 매긴 것이다.

<독거노인 비율 분포도>

이러한 시각화 과정에서 겪은 어려움도 있었는데, 처음엔 folium 맵과 geoJSON 경계 데이터를 병합하여 매핑하여 클러스터링을 통해 거점을 선정하려고 하였던 것에서 문제가 발생했다.

병합 과정에서 행정동 명칭의 불일치로 인해 일부 정보가 누락되는 문제가 발생하여 지도에 일부 구역이 미출력 되는 현상이 발생하여, 병합 대신 따로 시각화를 하되, 점수를 도입하여 이를 기반으로 거점을 선정하는 방식으로 바꿔 문제를 우회하는 데 성공하였다. 이렇게 하여 독거노인 비율까지 반영한 최종 점수인 ‘총점수’ 칼럼의 값들을 모두 구하였고, 인터넷 서핑을 통해 자료를 조사하여 각 후보군들의 정보를 상호 비교를 하여 봄꽃요양원이라는 최종 후보를 도출해낸 것이다. 다음은 이러한 결론을 기반으로 적용할 프로그램 및 개선방안이다.

