

「2023년 제11회 문화데이터 활용 경진대회」 분석보고서

분석 프로그램

☒ Python

☐ R

☒ Tableau

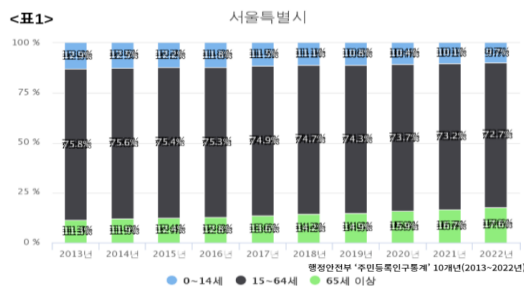
☐ 기타

1) 분석 주제

문화데이터 분석을 통한 노인 인프라 구축을 위한 분석 인사이트 도출

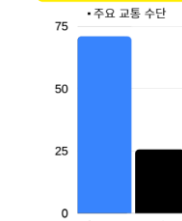
2) 분석의 배경 및 목적

우리가 언제까지 젊음을 유지할 수 있을까? 당신은 늙어서도 품격있고 교양있게 살고 싶다는 생각을 한 적 있는가? 젊었던 그 시절의 자신을 떠올리며 남은 노후를 즐겁게 보내고 싶지 않은가? 노인생활 실태 및 욕구는 국가와 가족이 함께 책임져야 하는 것이다.



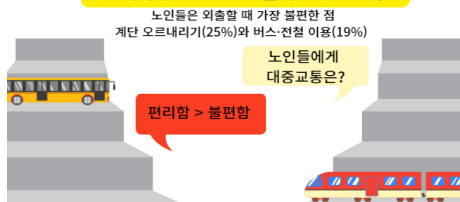
행정안전부에서 제시한 '주민등록인구통계' 10개년(2013~2022년) 자료 <표1>에 따르면 노인 인구(65세 이상)의 비율은 10년 전 대비 약 6% 증가함을 알 수 있다. 그에 반해 0~14세 이상의 비율은 점점 감소하고 있는 추세이다.

<표2> 노인의 외출시 주요 교통수단 현황



<표2>는 보건복지부와 한국보건사회연구원에서 공동 실시한 [2020년도 노인실태조사] 자료집에 근거하여 제작한 인포그래픽이다. 좌측에 그래프를 볼 때 노인의 경우 주요 '교통수단'은 대중교통이 약 71%로 자가용의 약 2.5배가량 높은 비율을 차지한다.

<표3> 보건복지부 '2020 노인실태조사 보고서'



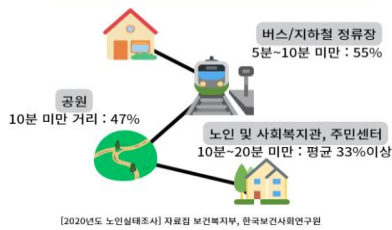
<그림1>



그러나 <표3>과 <그림1>를 통해 '대중(大衆)교통'이라는 이름이 무색할 만치 대중에게 어려운 대상이 됨을 파악할 수 있다. '계단이나 경사로 오르내리기'가 약 38%를, '대중교통 승하차 이용의 어려움'이 약 29%의 비율을 차지했다.

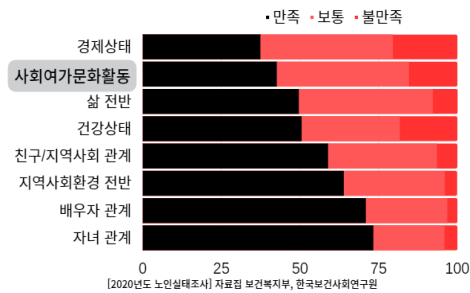
<표4>는 노인이 주로 이용하는 주요기관 및 시설이 '거리'와의 상관관계가 있음을 명확히 보여준다. 도보 5분~10분 미만 거리에 '버스/지하철 정류장'은 약 55%로 전체 인원의 과반수 이상의 비율을 차지했다. '공원'의 경우 10분 미만 거리 이용자가 약 47%의 비율을 차지했다.

<표4> 노인의 주요 기관 및 시설과의 도보 이동시간



이와 반대로 '노인 및 사회복지관, 주민센터'는 이용을 위해 이동시간이 10분~20분 정도일 때가 평균 33%이상의 비율을 차지했다. 이를 통해 노인이 주로 이용하는 주요기관 및 시설이 생각보다 거리가 있어, 대중교통 이용의 필요성이 절실함을 파악할 수 있었다.

<표5> 노인의 생활만족도



<표5> 또한 보건복지부와 한국보건사회연구원에서 공동으로 실시한 [2020년도 노인실태조사_노인의 생활만족도] 그래프를 첨부한 것이다. 노인의 생활만족도를 조사해보았을 때, '사회 여가 문화 활동'이 약 42%의 보통 만족도와 약 15%의 불만을 표했다. 9가지의 항목 중 8번째로 적은 만족도를 확인할 수 있다.

이에 대한 논의는 지금으로부터 약 17년 전인 2006년 대한민국 정책 브리핑 기사 <노인복지 준비없이 '고령사회' 오면...[예산쟁점③]노인보건복지예산 확대는 미래위험 예방투자 (2006.12.07)>에서도 볼 수 있었다. 옛날부터 대두되어 온 주제이나 아직까지도 완전한 사회적 인프라를 구축하지 못한 현실을 파악할 수 있다.

급속한 고령화 사회를 맞고 있는 우리는 노인 보건복지 서비스 인프라를 구축하는데 시간적·사회적 여유가 그리 많지 않다. 노인 보건복지 예산의 선제적 투자를 통해 활기차고 안정된 노후를 보장하는 사회적 인프라를 제대로 구축해야 우리 경제의 지속가능한 성장과 사회 발전을 기대할 수 있다. 고령사회 대응체계 구축, 즉 **노인 복지문화 인프라 확충**이야말로 지속발전 가능성을 확보하는 지름길일 것이다.

3) 활용 데이터 선정

데이터는 '공공데이터포털, 서울열린데이터광장, 문화데이터광장, 한국철도공사, 보건복지부, 산림빅데이터거래소'에서 '승하차교통데이터, 생활이동데이터, 노인복지시설데이터, 문화재데이터, 동식물원휴양림데이터'를 얻었고, 이를 기반으로 최종 Dataset 'seoul_final.csv'을 제작하였다.

노인이 '대중교통'을 이용해 '생활공간'으로 이동한다는 것, 이때 생활공간은 주로 '복지시설', '공원'과 같은 '휴양 공간'을 이용한다는 것, 그리고 다른 연령대에 비해 '역사의식'을 가지고 문화 공간으로 '문화재' 시설을 이용한다는 것을 기반으로 데이터를 확보하였다.

한국관광공사가 발표한 2023년 국내관광 트렌드는 '일상의 모든 순간이 여행'이다. 즉, 일상 속 개인의 관심사나 취향과 관련된 경험을 추구하며, 다양한 테마의 여행을 통해 현재의 행복을 만끽한다는 것이다. 여행의 테마는 크게 6개로 '친환경 여행, 로컬관광, 아웃도어/레저여행, 농촌 여행, 체류형 여행, 취미 여행'로 나뉘었다. 그 중 '친환경 여행'은 소비자 조사 결과에서 6개 여행 테마 중 가장 높은 70.1%의 높은 의향을 보였다.

우리팀은 노인 복지문화 인프라 예시로 '곤지암 화담숲'을 설명하고자 한다. 이는 경기도 곤지암에 위치하여 LG상록재단이 공익사업의 일환으로 설립, 운영하는 수목원이다. 화담숲 안을 운행하는 모노레일과 화담숲 주차장을 운행하는 리프트가 있어 남녀노소 즐기기에 충분하다.

<그림 2> 화담숲 구글 리뷰 4.5점



이미 많은 분들이 알고 계신 곳인지라, 두말 할 필요없이 **경관이 좋고 아기가기 하면서도 깨끗하고 (중략) 하루 휴가 쓰고 방문하시는 것을 추천드립니다.**

카페 등 편의시설도 잘 되어있고 수목관 등 생태체험 장소도 있습니다. 잘 조성된 숲을 다니는 것도 모노레일도 가능하고 도보도 가능합니다. 가격이 조금 비싸다고 느낄 수도 있는데 **충분히 돈을 한다고 느꼈어요**

개인적으로 우리나라에서 다녀본 수목원중에서는 제일 관리가 잘 되고 있고 정말 좋았던 수목원이었던 것 같다.

까다롭기로 유명한 구글 리뷰가 4.5의 평점을 보였다. <그림2>는 화담숲 구글 리뷰 중 상단의 3개를 이미지화한 인포그래픽이다.

화담숲은 '경기도' 문화공간이라는 점에서 '서울특별시'로 지역을 한정지은 우리팀의 방향성과 다르다는 의문이 제기될 수도 있다. 하지만, 서울특별시에 거주하는 노인이, 혹은 노인과 함께 거주하는 가족이 시간과 돈을 들여 경기도 곤지암으로 힐링여행을 떠나는 것만으로도 예시로 설정할 근거가 충분하다.

자택 인근 공원과 노인복지문화시설 방문을 즐겨하는 노인들의 특성을 고려하였을 때, 상대적인 신체적 제약으로 먼 곳을 이동하지 못하는 노인들이 가볍게 산책하고 여행의 기분을 만끽할 수 있는 인프라가 구축되어야 하는 것이다.

하단부터는 선정한 데이터에 대한 간략한 설명과 그림이다.

[1] 교통데이터_승하차데이터

대중교통이란, 국토교통부에 정의된 바 버스와 지하철 데이터를 의미한다. 노인은 경로 우대권을 사용하는 65세 이상의 고령자이다.

연번	일자	역번호	역명	구분	06시~07시	07시~08시	08시~09시	09시~10시
1	2022-01-01	150	서울역	출가 : 승차	38	33	24	41
2	2022-01-01	150	서울역	출가 : 환차	21	101	62	78
3	2022-01-01	151	시정	출가 : 승차	13	20	14	14
4	2022-01-01	151	시정	출가 : 환차	7	26	23	27
5	2022-01-01	152	종각	출가 : 승차	12	29	18	31
6	2022-01-01	152	종각	출가 : 환차	15	49	37	43
7	2022-01-01	153	종로3가	출가 : 승차	25	21	26	88
8	2022-01-01	153	종로3가	출가 : 환차	24	78	44	94
9	2022-01-01	154	종로5가	출가 : 승차	7	10	26	27
10	2022-01-01	154	종로5가	출가 : 환차	4	30	41	38

: 전체 승객유형 중 65세 이상 즉, 우대권을 소지하는 노인 승객 데이터를 선별하며 2번째 데이터인 생활이동데이터 논문자료를 통해 발견한 노인이 보통 대중교통을 이용하는 시간대 13시~16시 데이터로 전처리한다.

[2] 인구데이터_생활이동데이터

A	B	C	D	E	F	G	H	I	J
대상연월	요일	도착시간	출발 시간	도착 지역	출발 지역	이동유형	평균 이동	이동인구	이동인구
2	202303월	13	11010	11010M		20 WH	25	48.67	
3	202303월	13	11010	11010F		40 EW	21	88.26	
4	202303월	13	11010	11010F		40 HW	38	443.57	
5	202303월	13	11010	11010F		40 HH	20	33.34	
6	202303월	13	11010	11010F		40 HW	45	91.78	
7	202303월	13	11010	11010F		40 WE	19	58.33	
8	202303월	13	11010	11010F		40 WH	22	50.84	
9	202303월	13	11010	11010F		40 WW	24	71.95	
10	202303월	13	11010	11010F		45 EE	81	1849.61	
11	202303월	13	11010	11010F		45 EH	41	278.76	
12	202303월	13	11010	11010F		45 EW	23	89.43	
13	202303월	13	11010	11010F		45 HE	38	317.88	
14	시도	시군구	읍면동	name	full_name	45 HH	19	32.39	
15	11000	11010	1101052 사천동	서울특별시 중구구 서각동		45 HW	20	115.04	
16	11000	11010	1101054 상천동	서울특별시 중구구 선릉동		45 WE	25	71.26	
17	11000	11010	1101055 부암동	서울특별시 중구구 부암동		45 WH	27	46.16	
11000	11010	1101056 경안동	서울특별시 중구구 경안동						
11000	11010	1101057 무악동	서울특별시 중구구 무악동						
11000	11010	1101058 고남동	서울특별시 중구구 고남동						
11000	11010	1101060 신원동	서울특별시 중구구 신원동						
11000	11010	1101061 동구1234서울특별시 중구구 동구12347동							
11000	11010	1101063 동부56기 서울특별시 중구구 동부56기동							
11000	11010	1101064 미원동	서울특별시 중구구 미원동						
11000	11010	1101067 청간1동	서울특별시 중구구 청간1동						
11000	11010	1101068 청간2동	서울특별시 중구구 청간2동						
11000	11010	1101069 청간3동	서울특별시 중구구 청간3동						
11000	11010	1101070 용인1동	서울특별시 중구구 용인1동						

출발지/도착지/
나이 선별

행정동코드

: KT에서 제공하는 생활이동데이터이다. 휴대폰을 소지하는 이들의 모든 이동 데이터가 추출된다. KT생활이동매뉴얼에 따라 노인들은 13~16시 사이의 이동량이 제일 활발하기에 13시~16시 시간대 4개의 데이터를 활용했다. 이때 이동유형과 같은 필요 없는 값은 전처리과정을 거친다.

[3] 인프라데이터 (1)_노인요양복지시설데이터

가. 노인주거복지시설 현황(시·군·구)																
구	구명	노인요양시설			노인주거복지시설			노인요양시설			노인주거복지시설			노인요양시설		
		시설명	소재지	연립수	시설명	소재지	연립수	시설명	소재지	연립수	시설명	소재지	연립수	시설명	소재지	연립수
1	강남구	강남노인요양시설	강남구	1	강남노인주거복지시설	강남구	1	강남노인요양시설	강남구	1	강남노인주거복지시설	강남구	1	강남노인요양시설	강남구	1
2	강북구	강북노인요양시설	강북구	1	강북노인주거복지시설	강북구	1	강북노인요양시설	강북구	1	강북노인주거복지시설	강북구	1	강북노인요양시설	강북구	1
3	강서구	강서노인요양시설	강서구	1	강서노인주거복지시설	강서구	1	강서노인요양시설	강서구	1	강서노인주거복지시설	강서구	1	강서노인요양시설	강서구	1
4	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
5	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
6	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
7	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
8	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
9	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
10	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
11	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
12	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
13	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1
14	강원구	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1	강원노인주거복지시설	강원구	1	강원노인요양시설	강원구	1

- 재가노인복지시설.csv
- 노인주거복지시설.csv
- 노인의료복지시설.csv
- 노인원분데이터.xlsx
- 노인여가복지시설.csv

: 노인복지시설 현황(시군구) 데이터이다. 재가노인복지시설, 노인주거복지시설과 같은 노인 인프라 관련 데이터가 많았지만 그중에서도 주거, 의료복지시설 데이터를 중심으로 활용하였다

[3] 인프라데이터 (2)_동식물휴양림데이터, 공원데이터

구	구명	분류1	분류2	분류3	검색어	명칭	행정 시	행정 구	행정 동
1	강남구	BE_IW08-C관광/숙박	테마-보호	동물원	동물원	동물원	서울특별시	강남구	강남1동
2	강남구	BE_IW08-C관광/숙박	테마-보호	동물원	세계파충류세계파충류서울특별시강남구	세계파충류세계파충류서울특별시강남구	서울특별시	강남구	강남1동
3	강남구	BE_IW08-C관광/숙박	테마-보호	동물원	원숭이학교원숭이학교서울특별시강남구	원숭이학교원숭이학교서울특별시강남구	서울특별시	강남구	강남1동
4	강남구	BE_IW08-C관광/숙박	테마-보호	동물원	코끼리동물원코끼리동물원서울특별시강남구	코끼리동물원코끼리동물원서울특별시강남구	서울특별시	강남구	강남1동
5	강남구	BE_IW08-C관광/숙박	테마-보호	식물원	곤충식물원곤충식물원서울특별시강남구	곤충식물원곤충식물원서울특별시강남구	서울특별시	강남구	강남1동
6	강남구	BE_IW08-C관광/숙박	테마-보호	식물원	관악산야생관악산야생서울특별시강남구	관악산야생관악산야생서울특별시강남구	서울특별시	강남구	강남1동
7	강남구	BE_IW08-C관광/숙박	테마-보호	식물원	금강산식물원금강산식물원서울특별시강남구	금강산식물원금강산식물원서울특별시강남구	서울특별시	강남구	강남1동
8	강남구	BE_IW08-C관광/숙박	테마-보호	식물원	금강조경/금강조경서울특별시강남구	금강조경/금강조경서울특별시강남구	서울특별시	강남구	강남1동
9	강남구	BE_IW08-C관광/숙박	테마-보호	식물원	남산야외/남산야외서울특별시강남구	남산야외/남산야외서울특별시강남구	서울특별시	강남구	강남1동
10	강남구	BE_IW08-C관광/숙박	테마-보호	식물원	남산야외/남산야외서울특별시강남구	남산야외/남산야외서울특별시강남구	서울특별시	강남구	강남1동

: 서울특별시 내에 존재하는 동물원, 식물원과 휴양림 데이터이다. 이를 통해 문화재 인근 버스 정류장과 지하철 노선을 확인할 수 있었다. 추후 인프라 구축을 요하는 곳에 활용될 데이터이다.

[3] 인프라데이터 (3)_문화재데이터

순번	종목	지정번호	문화재명	한자명	지정일	수량/면적	시군구
1	시도동독제1호		한강대교	漢江大橋	1969.05.11	교량(구) 서울특별시	강남구
2	시도동독제2호		보신각 앞 지하철 수	보신각 앞 지하철 수	1970.05.11	1/0.1㎡ 서울특별시	강남구
3	시도동독제3호		구 통계국 청사	구 통계국 청사	1970.05.11	건물 1동(서울특별시)	강남구
4	시도동독제4호		경향신문	京鄕新聞	1970.05.11	1건 220㎡ 용산구	강남구
5	시도동독제5호		서울시립대학교 구 본	서울시립대학교 구 본	1970.05.11	건물 2동 서울특별시	강남구
6	시도동독제6호		서울시립대학교 구 2	서울시립대학교 구 2	1970.05.11	건물 1동 서울특별시	강남구
7	시도동독제7호		고려대학교 4.18 기념	고려대학교 4.18 기념	1970.05.11	기념탑 1개 서울특별시	강남구
8	시도동독제8호		독립신문	獨立新聞	1970.05.11	1점 마포구	강남구
9	시도동독제9호		인광노	引光奴	1970.05.11	1건 3㎡ 서울특별시	강남구
10	시도동독제10호		남산도서관 소장 일지	南山道서관 소장 일지	1970.05.11	3종 16책 용산구	강남구
11	시도동독제11호		남산도서관 소장 경신	南山道서관 소장 경신	1970.05.11	16종 156책 용산구	강남구
12	시도동독제12호		남산도서관 소장 대한	南山道서관 소장 대한	1970.05.11	5종 18책 용산구	강남구
13	시도동독제13호		남산도서관 소장 대한	南山道서관 소장 대한	1970.05.11	1종 1책 용산구	강남구

: 문화재청으로부터 추출한 문화재 데이터이다. 이를 통해 문화재 인근 버스 정류장과 지하철 노선을 확인할 수 있었다. 추후 인프라 구축을 요하는 곳에 활용될 데이터이다.

위 데이터들을 통합한 최종 dataset이 'seoul_final.csv'이다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	행정구	동식물원수공원수	승하차	지하철역구노인인구	경로당	노인시설	노인시설	노인시설	노인시설	노인시설	노인시설	노인시설	노인시설	노인시설	노인시설	노인시설	문화재수
2	용산구	2	2	6693008	10	39779	88	25	194	113	0.002841	1.69E-05	0.113355	669300.8	1901.279	0	450
3	마포구	0	5	5881397	14	54002	154	17	284	171	0.003167	2.91E-05	0.157935	420099.8	1330.267	0	20
4	성북구	0	3	6467009	9	73613	175	31	715	206	0.002798	3.19E-05	0.1166	718556.6	2010.822	0	157
5	성동구	1	5	7078565	13	45435	162	20	391	182	0.004006	2.57E-05	0.168689	544505	2181.136	0	29
6	종로구	0	12	22039837	12	28311	63	21	404	84	0.002967	3.81E-06	0	1836653	5449.432	1	455
7	서대문구	0	4	4959137	5	53756	111	26	388	137	0.002549	2.76E-05	0.148356	991827.4	2527.724	1	147
8	서초구	5	6	11696077	10	60072	141	14	322	155	0.00258	1.33E-05	0.033632	1169608	3017.865	1	70
9	중구	0	5	16276987	14	24174	48	8	74	56	0.002317	3.44E-06	0.003011	1162642	2693.305	1	107
10	강남구	1	7	17425049	17	74959	167	32	215	199	0.002655	1.14E-05	0.006936	1025003	2721.162	1	41
11	노원구	0	2	12334867	14	85420	243	29	1008	272	0.003184	2.21E-05	0.048318	881061.9	2805.536	1	55
12	동작구	0	7	10736494	13	64928	140	14	190	154	0.002372	1.43E-05	0.033753	825884.2	1958.88	0	35
13	강북구	0	4	5709043	4	63313	100	26	503	126	0.00199	2.21E-05	0.083692	1427261	2840.41	1	32
14	은평구	0	7	12361204	11	85010	154	40	1095	194	0.002282	1.57E-05	0.019555	1123746	2564.483	1	31
15	중랑구	0	6	5719284	9	69341	127	26	859	153	0.002206	2.68E-05	0.098245	635476	1402.169	0	27
16	관악구	1	2	9236593	4	78430	114	30	435	144	0.001836	1.56E-05	0.025094	2309148	4239.67	1	131
17	동대문구	1	4	10266451	9	61612	134	21	224	155	0.002516	1.51E-05	0.043186	1140772	2869.891	1	31

4) 분석 내용 및 결과

모든 데이터는 다음과 같은 전처리 과정(하단의 이미지는 승하차 데이터의 전처리 과정을 보인 것이다)을 거쳤다. (1) data.info()와 data.describe() 함수를 통해 데이터 기본 정보 확인 후, (2) data.isna()와 data.isna().sum() 함수를 통해 데이터 결측치를 확인한다. 이후 (3)과 같이 data.drop('제거 열', axis=1) 함수를 통해 필요없는 열이나 행을 제거한다. 그리고 heatmap()을 통해 결측치를 최종적으로 확인한다.

1 데이터 기본 정보 확인

```
data.info()
data.describe()
```

5 데이터 결측치 확인

```
data.isna()
data.isnull().sum()
data.isna().sum()
```

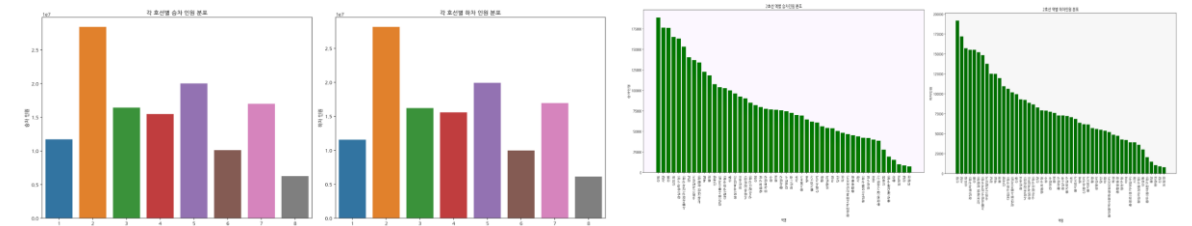
7 데이터 전처리

```
sns.heatmap()
data.dropna(inplace=True)
```

8 sns.heatmap()

다들 알겠네 = 輿의 X = 결측치

[1] 교통데이터_승하차데이터



막대 그래프 형식으로 호선별 승하차 인원분포를 시각화했는데 이때 2호선 > 5호선 > 7호선 > 3호선 > 4호선으로 상위 5개의 값을 얻었고, 주황색 막대 2호선의 승하차 비율이 가장 높았다. Filtered_data로 2호선 열을 추출하고 ascending 값을 false로 도출되는 내림차순 정렬 시각화 실행했다. 상위 5개의 승하차 인원 분포 확인 결과는 다음과 같다. '승차'는 '신림, 사당, 선릉, 신도림, 잠실(송파구청)'이다. '하차'는 '신림, 사당, 신도림, 잠실(송파구청), 선릉'이다.

Q "원하시는 호선의 총 승하차 인원 분포를 보여드리겠습니다. 1-8 사이의 자연수를 입력해주세요."

```
try:
    n = int(input("원하시는 호선의 총 승하차 인원 분포를 보여드리겠습니다. 1-8 사이의 자연수를 입력해주세요: "))
    if n < 1 or n > 8:
        raise ValueError("입력한 호선 번호가 범위를 벗어났습니다.")
    > n호선 데이터 추출
    > 역별 승하차 총 인원 계산
```

Q "원하시는 역의 승하차 총 인원 분포를 보여드리겠습니다. 역명을 입력해주세요."

```
station = input("원하시는 역의 승하차 총 인원 분포를 보여드리겠습니다. 역명을 입력해주세요: ")
if station not in data['역명'].unique():
    print("올바른 역명을 입력해주세요.")
    exit()
> 입력된 역명에 대한 승차 및 하차 분포 데이터 추출
> 왼쪽 서브플롯 - 승차 분포 / 오른쪽 서브플롯 - 하차 분포
```

더 나아가 좌측 이미지처럼 if 함수로 사용자로부터 원하는 호선 숫자를 입력받아 해당 호선의 총인원 분포도를 보여주도록 만들었다. 1에서 8을 제외한 수의 경우 value error값이 도출된다. 또, 우측 이미지처럼 사용자로부터 역명을 input 받고 해당 역 승하차 분포를 시각화하는 것입니다. 입력된 역명에 대한 승하차 분포 데이터를 추출하여 왼쪽에는 승차, 오른쪽에는 하차 플롯을 출력할 수도 있다.

[2] 인구데이터_생활이동데이터

pd.merge()

: 데이터프레임 병합

ex) ss1과 loc 데이터프레임

'출발 행동동 코드' 열 기준 병합

> 새로운 데이터프레임 S1 생성

pd.merge()

: 데이터프레임 병합

ex) ss1과 loc 데이터프레임

'출발 행동동 코드' 열 기준 병합

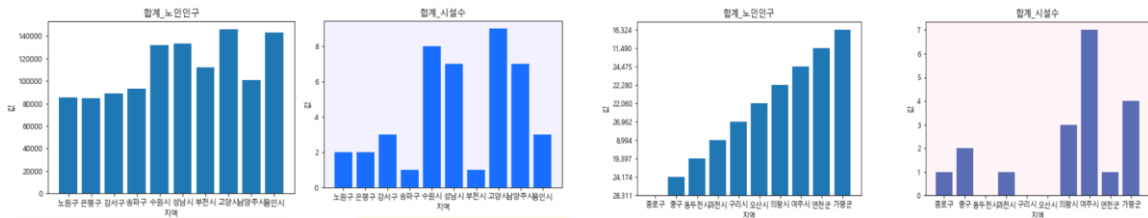
> 새로운 데이터프레임 S1 생성

KT생활이동 매뉴얼을 통해 오후 1시부터 4시 사이의 노인들 이동량이 제일 활발함을 확인했기 때문에 오후 1~4시 시간대 4개의 데이터를 활용했다. 이동유형과 같은 필요없는 값을 가공하고, 데이터 전처리와 결측치 확인을 통해 이상이 없음을 파악했다. 그리고 시간대별/행정별 데이터를 병합했다.

이후 지하철 승하차데이터와 지역별 유동인구 데이터 상관관계를 분석할 때는 데이터 크기 차이로 인해 size를 조절했다.

[3] 인프라데이터 (1)_노인요양복지시설데이터

노인복지시설 현황(시군구) 데이터는 병합처리된 부분이 많아 전처리 과정이 오래걸렸다. 병합 해제 후 결측치를 파악하고 상위권/하위권 10개 지역을 그래프로 시각화했다.



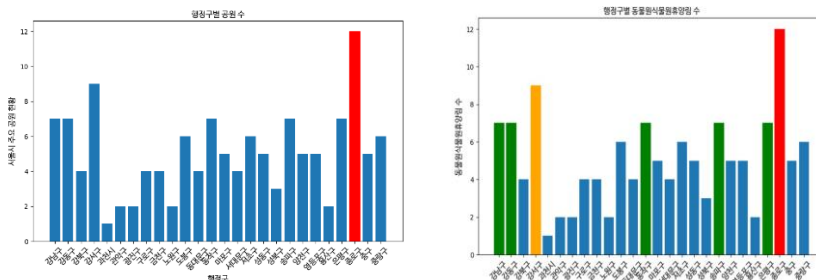
좌측이 노인 인프라가 잘 갖춰진 상위권 10개 지역, 우측이 하위권 10개 지역이다.



좌측은 상위권 10개 지역 그래프이다. 에서 1위를 차지한 고양시의 경우, 상대적으로 노인인구 대비 시설이 비례하게 잘 구축되어 있었고, 용인시의 경우에는 시설 수에 비해 노인인구가 많았다. 양로시설의 경우, 용인시는 과다 투입이 의심될 정도로 현재 시설에 거주하는 인원과 종사자 수에 비해 시설 수가 매우 적었다. 노인복지주택의 경우 성남시가 선두를 달렸다.

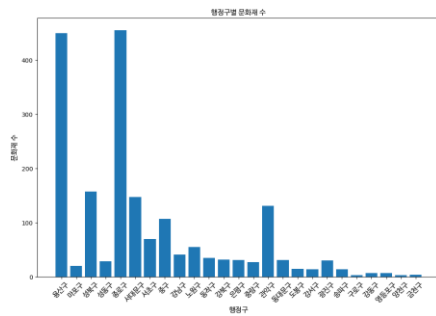
이와 같은 단계로 우측 그래프를 통해 하위권 10곳, 즉 인프라 구축을 요하는 곳도 확인하였다. 이중 과천시와 제일 잘 구축되어 있었고, 동두천은 노인인구가 존재함에도 시설이 아예 마련되어 있지 않았다. 양로시설의 경우, 상위권 데이터와 대비되게 노인인구 대비 노인인프라가 잘 마련되어 있는 곳들이 있었다. 노인복지주택의 경우, 중구와 가평군은 반대의 성향을 보였다.

[3] 인프라데이터 (2)_공원데이터, 동식물휴양림데이터



공원데이터에서 종로구가 서울에서 12곳의 공원을 보유하고 있었고, 동식물휴양림의 경우 또한 종로구가 12곳, 강서구가 9곳으로 뒤를 이었다.

[3] 인프라데이터 (3)_문화재데이터



문화재데이터의 경우, 종로구가 약 450여개, 다음은 용산구가 440여개로 뒤를 이었다. 서울특별시의 중심부에 위치한 자치구로, 중구와 함께 서울의 도심을 이루는 종로구의 지리적 특성을 파악할 수 있었다. 역사적으로는 조선시대 궁궐인 경복궁과 창경궁, 세계문화 유산인 창덕궁과 종묘 등도 이 종로구에 속해 있어 전통과 문화가 잘 보존된 지역이기도 한 것이다.

하단부터는 tabular public을 활용하여 보다 더 이해 용이성을 위해 지도로 시각화한 자료이다.

(https://public.tableau.com/app/profile/woondy.lim/viz/_16880186033340/1_1?publish=yes)

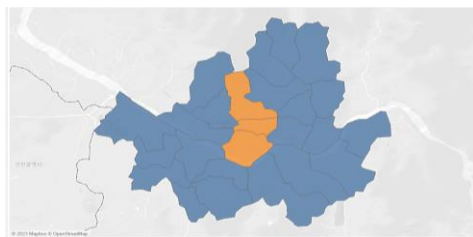


<표1>

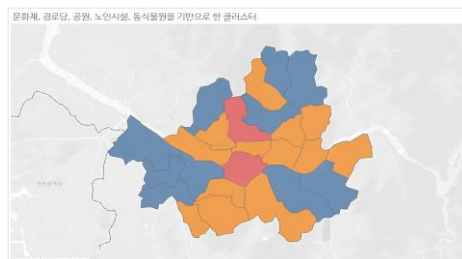


<표2>

<표1>은 '노인인구'와 '승하차' 2개의 데이터를 하나의 표에 나타낸 그림이다. <표2>는 '노인인구', '노인인프라 구축현황' 그래프로 2개의 변수 간의 상관관계를 확인할 수 있다. 노인인구가 많은 곳에는 노인인프라가 더 많이 구축되어 있을 것이라고 예상했다. 그러나, '**종로구**'를 제외하고는 인구수에 비해 인프라가 제대로 구축되어 있지 않았다. 노인을 제대로 수용할 수 없는 시설들이 많이 존재한다는 뜻이었다. 이를 통해 **노인인프라의 열악성**을 확인할 수 있었다.



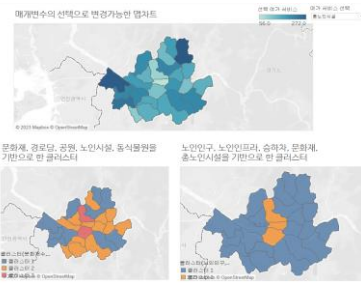
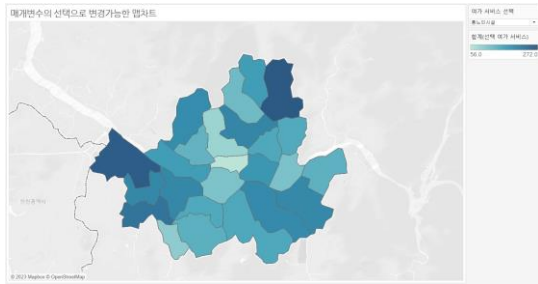
<표3>



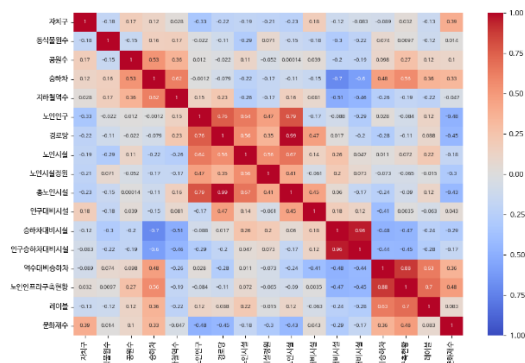
<표4>

<표3>을 통해 '노인인구', '총노인시설(노인인프라)', '승하차', '문화재', '공원 및 휴양시설'을 기반으로 제작한 클러스터에서 '**종로구, 중구, 용산구**'가 상위 3개의 클러스터로 선정됨을 알 수 있다.

<표4>는 노인인프라 데이터를 기반으로 제작한 클러스터 지도이다. <표2>에서 확인했던 것과 마찬가지로 '노인인프라(문화재 공간, 경로당, 노인시설, 공원 및 동식물휴양림)'는 '**종로구**'에 제일 잘 구축되어 있었고, 다음으로는 용산구가 많은 비중을 차지했다.



<그림1>처럼 tabular data는 tabular public 사이트를 통해 매개변수 선택으로 변경이 가능하다. 예측 서비스 제작 시 서비스 이용자가 원하는 데이터를 선택하면 ‘이용자 맞춤 데이터’를 제공할 수 있다는 것이다. <그림2>는 tabular public 사이트에서 확인가능한 그래프들을 첨부한 것이다.



이는 seoul_final.csv파일을 통해 확인한 상관관계 분석 결과이다. +(양의 값)을 중심으로 살펴보면 노인인프라구축현황-승하차의 관계성이 0.56으로 비교적 높은 수치를 가지고 있음을 파악할 수 있다. 또 노인인구-총노인시설 또한 0.76으로 노인인구가 존재하는 곳의 노인시설 분포하고 있음을 알 수 있다. 총노인시설-경로당의 상관관계가 0.99인 것을 볼 때 대표적인 노인시설은 경로당이였다.

앞선 그래프의 시각화를 기반으로 한 분석을 기반으로 예측 모델 10개에 대입하여 알고리즘을 돌렸다. 상기 예측 모델로 인한 결과는 다음과 같다.

[1] 로지스틱 회귀 / [2] 선형회귀 / [3] 랜덤 포레스트

[illegible][illegible]

3. 랜덤 포레스트

```
[ ] from sklearn.ensemble import RandomForestRegressor

# 랜덤 포레스트 회귀 모델 구축
model = RandomForestRegressor(n_estimators=100, random_state=42)

# 모델 학습
model.fit(X_train, y_train)

# 모델 예측
y_pred = model.predict(X_test)

# 모델 평가
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
mse = np.sqrt(mean_squared_error(y_test, y_pred))

print('평균 제곱 오차(MSE):', mse)
print('결정 계수(R^2):', r2)
print('평균 절대 오차(MAE):', mae)
print('평균 제곱근 오차(RMSE):', mse)
```

[4] GBM / [5] Catboost / [6] AdaBoost

[illegible][illegible]

6. AdaBoost

- AdaBoost의 아이디어는 많은 약한 분류기를 학습해서 그들을 다시 조합하여 더 강력한 분류기를 만드는 것. 즉, 각각의 모델은 조금씩 틀릴 수 있지만, 이들을 조합하면 오차를 줄여나갈 수 있다. 이 아이디어는 1990년 데이비드 로버트 스타인바움에 의해 소개된 것이다.
- AdaBoost는 이 아이디어를 구현하기 위해 다음과 같은 과정을 거친다.
 1. 초기 데이터셋을 생성한다. 이 데이터셋은 2개의 클래스로 나뉘어 있다. 이 클래스를 각각 클래스 0과 클래스 1로 표시한다.
 2. 이 데이터셋을 사용하여, 약한 분류기를 학습한다. 이 약한 분류기는 클래스 0과 클래스 1을 분류하는 데 사용된다.
 3. 이 약한 분류기의 성능을 평가한다. 이 성능은 이 분류기가 클래스 0과 클래스 1을 분류하는 데 얼마나 잘하는지를 나타낸다.
 4. 이 약한 분류기의 성능을 기반으로, 이 분류기의 가중치를 결정한다. 이 가중치는 이 분류기가 클래스 0과 클래스 1을 분류하는 데 얼마나 잘하는지를 나타낸다.

```

1 from sklearn.metrics import roc_auc_score
2
3 def AdaBoostClassifier(X_train, X_test, y_train, y_test):
4     # AdaBoost를 사용하여 약한 분류기를 학습한다
5     classifier = AdaBoostClassifier(n_estimators=100)
6     classifier.fit(X_train, y_train)
7
8     # AdaBoost를 사용하여 테스트 데이터를 분류한다
9     y_pred = classifier.predict(X_test)
10
11     # AdaBoost의 성능을 평가한다
12     auc = roc_auc_score(y_test, y_pred)
13
14     # AdaBoost의 성능을 출력한다
15     print("AdaBoost의 성능: {}".format(auc))
16
17 # AdaBoost를 사용하여 테스트 데이터를 분류한다
18 AdaBoostClassifier(X_train, X_test, y_train, y_test)
  
```

AdaBoost는 이 아이디어를 구현하기 위해 다음과 같은 과정을 거친다.

1. 초기 데이터셋을 생성한다. 이 데이터셋은 2개의 클래스로 나뉘어 있다. 이 클래스를 각각 클래스 0과 클래스 1로 표시한다.
2. 이 데이터셋을 사용하여, 약한 분류기를 학습한다. 이 약한 분류기는 클래스 0과 클래스 1을 분류하는 데 사용된다.
3. 이 약한 분류기의 성능을 평가한다. 이 성능은 이 분류기가 클래스 0과 클래스 1을 분류하는 데 얼마나 잘하는지를 나타낸다.
4. 이 약한 분류기의 성능을 기반으로, 이 분류기의 가중치를 결정한다. 이 가중치는 이 분류기가 클래스 0과 클래스 1을 분류하는 데 얼마나 잘하는지를 나타낸다.

7. LightGBM

```

x XGBBoost

1) %load xgbdata
from sklearn.datasets import load_svmlight_file

2) from sklearn.metrics import roc_auc_score
from sklearn.cross_validation import train_test_split

3) #데이터 불러오기
data = load_svmlight_file('data/train', sparse=True)
train_data, test_data, train_label, test_label = train_test_split(
    data.data, data.target, test_size=0.2, random_state=360)

4) #XGB를 이용해 모델 학습
model = xgb.XGBClassifier()
model.fit(train_data, train_label)

5) #테스트 데이터에 대한 예측
pred = model.predict(test_data)

6) #ROC 곡선 그리기
roc_auc_score(test_label, pred)

7)
import xgboost as xgb
model = xgb.XGBClassifier()
model.fit(train_data, train_label)
print model.score(test_data, test_label)

```

```

9. RNN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

[illegible]

[11] LIME

```

LIME(Local Interpretable Model-agnostic Explanations)

1. 파이썬 라이브러리
from sklearn.metrics import pairwise_distances
from sklearn.metrics.pairwise import euclidean_distances
from sklearn.metrics.pairwise import cosine_distances
from sklearn.metrics.pairwise import manhattan_distances
from sklearn.metrics.pairwise import mahalanobis_distances

2. 데이터셋 로드
X = load_data()
y = load_labels()

3. 모델 학습
model = LogisticRegression()
model.fit(X, y)

4. LIME 라이브러리 설치
pip install lime

5. LIME 모델 생성
lime_model = LimeModel(model=model, data=X, labels=y)

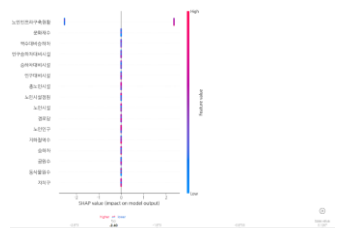
6. LIME 모델 사용
lime_model.predict(X)
lime_model.explain(X)
lime_model.explain(X, y)
lime_model.explain(X, y, distance='euclidean')
lime_model.explain(X, y, distance='cosine')
lime_model.explain(X, y, distance='manhattan')
lime_model.explain(X, y, distance='mahalanobis')

7. LIME 모델 평가
lime_model.score(X, y)
lime_model.score(X, y, distance='euclidean')
lime_model.score(X, y, distance='cosine')
lime_model.score(X, y, distance='manhattan')
lime_model.score(X, y, distance='mahalanobis')

```



[12] SHAP

[illegible]

5) 시사점 및 기대효과

노인들의 관심사, 선호도, 접근성 등을 파악하고, 이를 기반으로 개선사항과 최적화 방안을 제시함으로 상대적으로 체력적 부담이 크지 않게 하면서, 노인여가문화생활의 만족도를 향상시킬 수 있을 것이다. 또한 노인 자택 인근 공원 및 휴양림과 문화인프라가 구축된다면 지역상생의 효과도 기대해볼 수 있다. 혼자 사는 노인의 증가, 가족 구성원 간의 상호 도움의 감소 등의 문제도 해결할 수 있을 것으로 기대된다.