

국민의 소중한 데이터, 국민을 위해

“취약계층 삶의 질 향상을 위한 데이터 분석”

2020.02.20

국민연금공단
청년인턴 나영비

CONTENTS

01 공단 소개 및 인턴 과정

- 1-1 국민연금공단 소개
- 1-2 6개월 인턴 과정

02 주요 분석 과제

- 2-1 기초연금 신청안내 효율화
- 2-2 저임금 근로자 국민연금 가입 확대
- 2-3 장애인 활동지원 신뢰도 제고

03 추가 분석 과제

- 3-1 증명서 발급 취약지역 파악
- 3-2 상담내역 텍스트 분석

04 청년인턴을 마치며

- 4-1 느낀점
- 4-2 향후 발전 방안

01

공단소개 및 인턴 과정

1-1 국민연금공단 소개

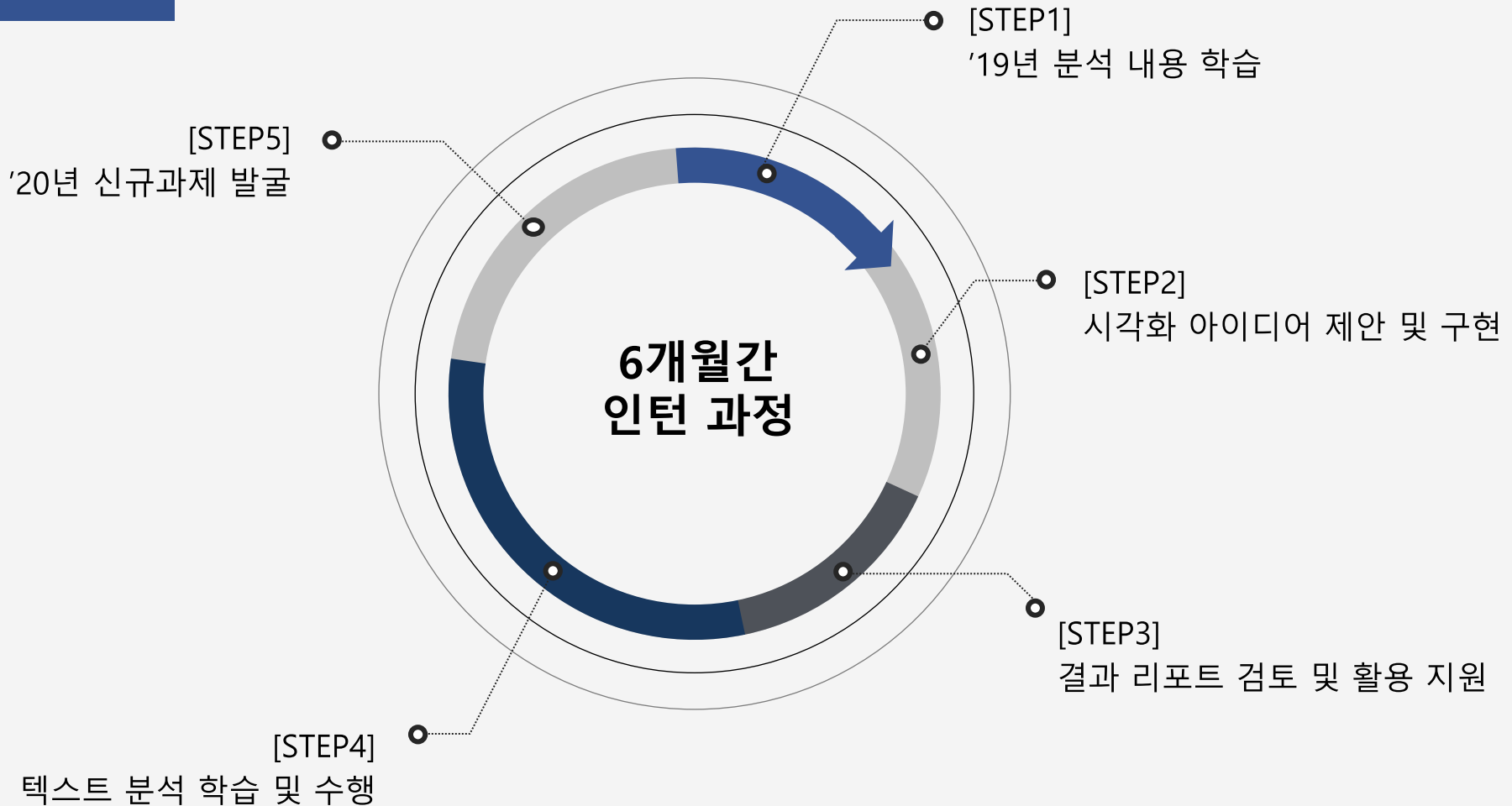
1-2 6개월 인턴 과정

국민연금공단 소개



- 국민연금공단은 연금 업무 외에도 기초연금, 노후 설계, 장애인 지원 등 취약계층을 위한 다양한 업무를 수행하고 있으며, **2,231만 명의 가입자 및 520만 명의 수급자 관리와 930조의 연 기금**을 운용
- 빅데이터부에서는 4,800억 건의 데이터를 활용하여 **사회취약계층 지원과 복지사각지대를 해소**하기 위한 공단 내·외부의 다양한 분석을 수행.

인턴 과정



※ 분석 프로세스 상 신규과제 발굴이 첫번째 단계이지만 인턴 배치 시기로 인하여 마지막 단계부터 체험

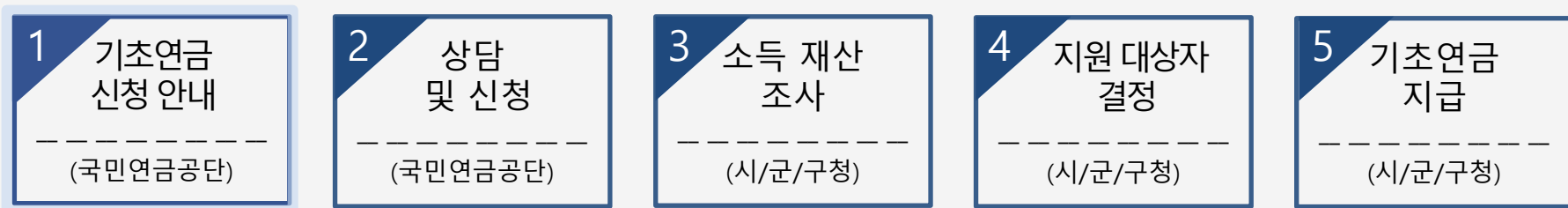
2-1

주요 분석과제

기초연금 신청안내 효율화

□ 기초연금이란?

- 노후를 대비하지 못한 어르신들의 생활 안정에 도움을 주기 위한 제도로
만 65세 이상 **소득 수준 하위 70%**에 해당하는 분들에게 일정 금액을 지원



□ 필요성

- 기초연금 수급 자격이 되어도 신청을 하지 않으면 지원받을 수 없음
- 과거 분석 결과 안내가 잘되면 그에 비례하여 신청률도 높아지는 것을 파악

무엇보다
'안내'가 중요

“기초연금 **안내**에서 발생하는 2가지 한계점”
1) 경계 미수급자 파악의 어려움 2) 한정된 인력과 시간

신청 가능성 높은 안내대상자 타겟팅 모형 개발

□ 분석 범위

- 안내대상자 예측에 대한 영향도가 높은 변수 파악
- 기초연금 수급 가능성이 높은 안내대상자 예측모형 개발
- 일선 현장에서 실제 업무에 활용할 안내대상자 명부 제공

□ 활용 데이터

- 기초연금 대상자, 공시지가, 각종 모바일 및 안내문 발송내역

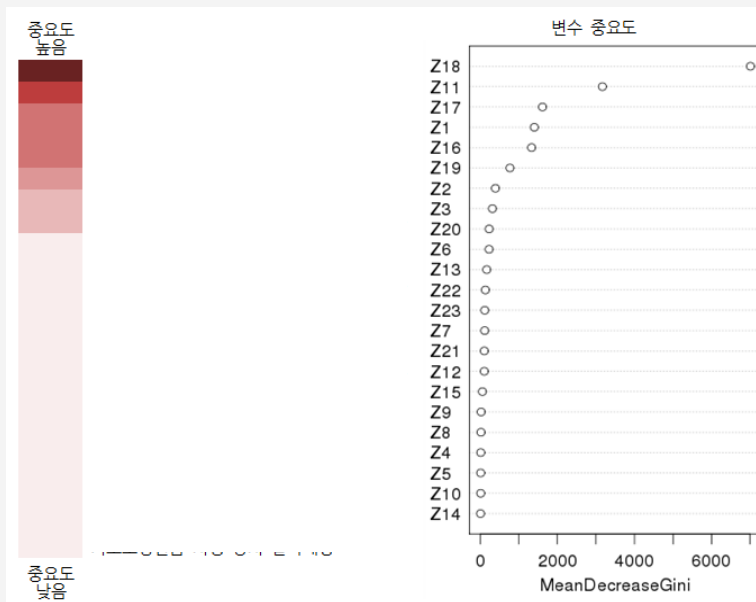
구 분	주요 항목
종속변수	기초연금 수급 여부
독립변수	(인구 통계적 특성) 성별, 연령, 거주지
	(경제적 요인) 공시지가, 소득인정금액
	(사회적 요인) 1인 가구 비율, 고령인구 비율, 노령화 지수
	(기초연금 업무 관점) 신청이력, 안내여부, 안내결과, 미신청 사유, A급여액, 당월 기본연금금액

□ 적용 알고리즘 검토

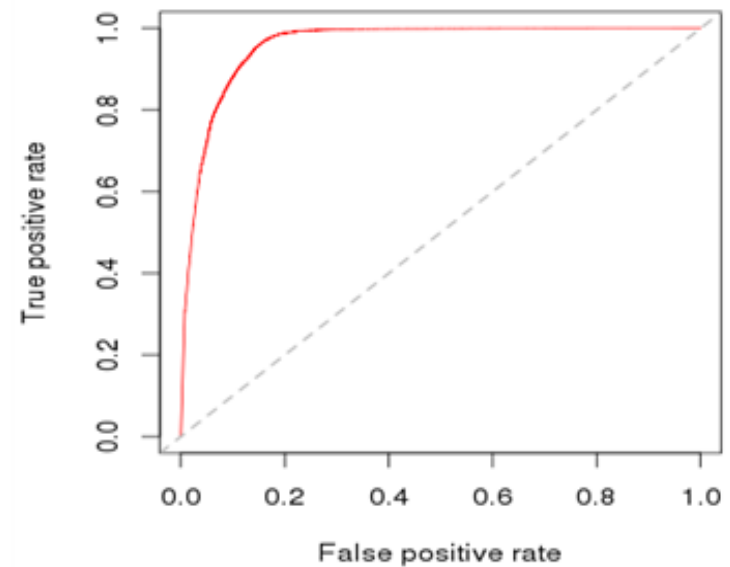
- GLM, SVM, 랜덤포레스트 등 다양한 알고리즘을 검토 후, 공단 환경에 가장 적합한 알고리즘인 랜덤포레스트 선정

<랜덤포레스트(Random Forest) 선정 이유>

- 검토한 알고리즘 중 가장 성능이 우수하고 업무부서에 모형에 대한 설명 용이



주요 영향변수 파악(23개 항목)



모형 성능평가(93.6%)

- 개발한 모형의 업무 활용도를 높이기 위해 안내대상자가 밀집된 집중관리 지역 파악
 - 청년인턴 교육과정 중 배웠던 QGIS를 활용한 시각화 아이디어 제안 및 구현

<데이터>

- 기초연금 우선안내대상자 정보
- 시도, 시군구, 읍면동 공간 데이터

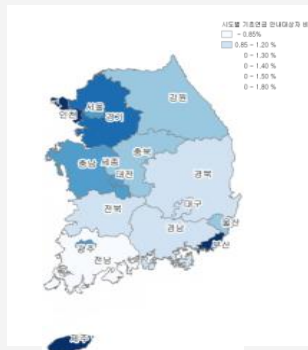
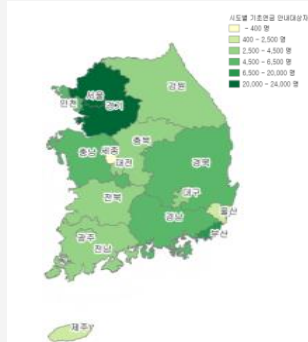
<시각화 과정>

1. 전국 시도, 시군구, 읍면동 데이터와 지역 코드 결합
2. 지역코드를 key로하여 1의 데이터와기초연금 안내대상자 정보를 결합
3. 단계 구분도를 통해 안내대상자 밀집지역시각화
4. 공간시각화 지식공유를 위해 상기 과정을 '공간시각화 가이드'로 작성

안내대상자수

안내대상자비율

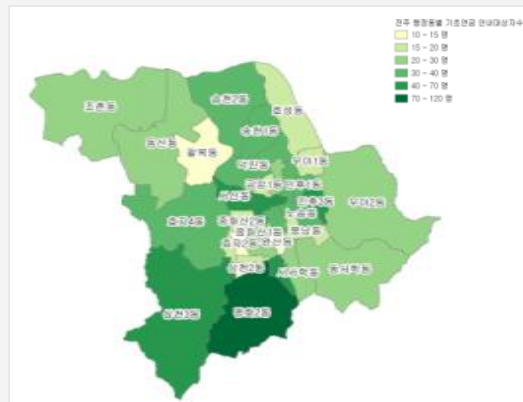
광역시도



□ 활용 방법

- 각 지역별 특성에 맞게 대상자를 분류(10분위)하여 신청안내 활동 전개

분위	수급 가능확률 범위
1	(0.5,0.541]
2	(0.541,0.580]
·	·
·	·
·	·
9	(0.969,0.996]
10	(0.996,1.0]



(예시) 도시지역은 거주인구 및 담당직원 수를 감안하여 수급가능성이 높은 9분위, 농촌지역은 더 많은 안내를 위하여 7분위 이상을 대상으로 함

□ 활용 성과

모형 활용 후 기초연금 신청률 9.8%p 증가 (전년 동기간대비)

□ 느낀점

- 교육 중 배웠던 알고리즘이 실제 업무에 어떻게 활용하는지 체험
- 시각화 구현 시 배웠던 내용을 더욱 명확하게 알게 되었고, 시각화 지식을 팀원들과 공유할 수 있어 보람을 느낌

2-2

주요 분석과제

저임금 근로자 가입 확대

소규모 사업장의 저임금 근로자 가입 확대를 위한 예측 모형 개발

□ 추진 배경

- 10인 미만 사업장 저임금 종사자들의 노후 소득보장을 위한 가입 확대 업무 효율화

□ 분석 범위

- 10인 미만 사업장 및 해당 사업장근로자 특성 분석
- 가입가능성을 점수화한 지수 생성 모형 개발

□ 활용 데이터

- 소규모 사업장 및 근로자, 국세청 및 고용노동부 입수자료 활용

구 분	주요 항목
종속변수	가입 여부
독립변수	(인구 통계적 특성) 성별, 연령
	(경제적 요인) 임금 수준, 사업장 규모
	(가입 업무 관점) 업종, 형태, 최초취득일, 근무일수, 자료 입수기관, 가입종별 가입현황

※ 분석 기법은 기초연금 우선안내대상자 타겟팅 모형 개발과 유사하므로 자세한 설명 생략

□ 정확도 향상

- 초기 예측모형은 정확도가 91.9%였으나 2주 간의 개선 노력을 통하여 1% 이상의 정확도가 높아짐으로써 10만 건 이상의 성과가 더 나타날 것으로 기대

		실제 (Reference)	
		가입	미가입
예측 (Prediction)	가입	Type I	Type II
	미가입	Type III	Type IV

□ 활용 방법

- 가입가능성 지수 및 대상자의 기존 가입 패턴에 대한 정보 생성
- 지사에서 가입 안내에 이용할 수 있도록 정보를 제공하는 화면 개발

□ 활용 성과

모형 활용 후 기존대비 저임금 근로자 국민연금 가입률 향상 (전년 동기간대비)

□ 느낀점

- 실제 회사에서는 단 1%의 정확도 향상을 위하여 각고의 노력을 기울인다는 것을 배움

2-3

주요 분석과제

장애인 활동지원 신뢰도 제고

□ 장애인 활동지원이란?

- 장애로 인해 일상생활이 어려운 분들에게 활동지원 서비스를 제공함으로써 삶의 질 향상

□ 장애 등급제 폐지에 따라 서비스지원 종합조사의 신뢰성 확보가 중요한 시점

개선 전

많은 지원 대상 및 다양한 조사 항목에 의하여
심사결과의 편차 발생



개선 후

동일한 장애유형과 정도에서는
심사결과의 편차 최소화

□ 추진 배경

- 장애인 활동지원 심사결과 신뢰도 제고를 위한 이상치 검출 모형(Outlier Detection Model) 개발

□ 분석 범위

- 장애유형별 점수 분포 분석 및 이상치 검출모형 개발
- '19.7월 장애등급제 개편 이후 일정기간(3개월 내외) 축적된 실 데이터를 활용한 모형 업그레이드

□ 활용 데이터

- 장애인활동지원 심사 결과 : (파일럿 데이터) 588건
- 주요 항목: 장애유형, 종합조사 점수, 장애 정도(경·중), 성인/아동 구분

□ 데이터 증폭

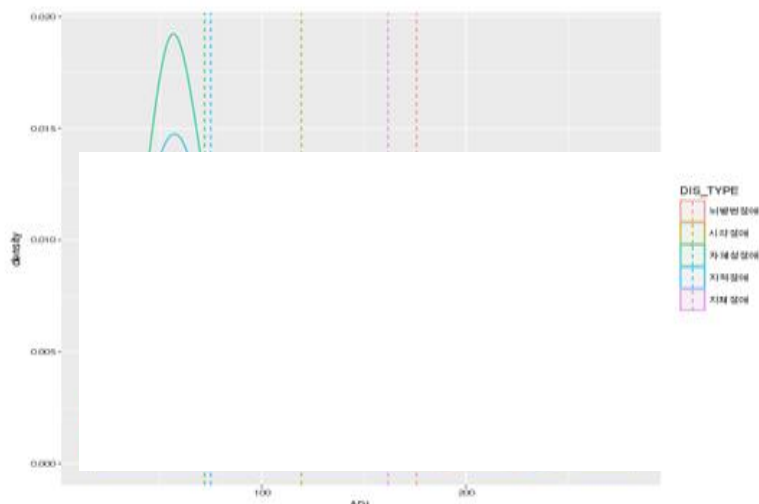
- 소량의 파일럿 데이터로는 모형 개발 시 추정불가 결측구간 다수 발생
- MCMC알고리즘을 사용하여 부족한 데이터를 증폭

※ 데이터 부족 상황 극복을 위한 방안마련 및 구현 과정에 참여

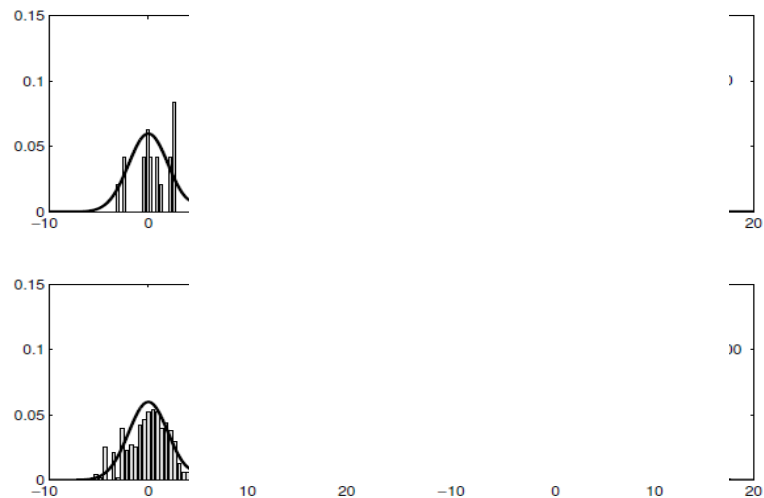
<MCMC(Markov Chain Monte Carlo) 선정이유>

- 데이터 부족으로 인한 모형 개발 불가(최소 20만 건 이상 필요) → 데이터 증폭이 요구됨

□ Outlier Detection Model



[장애유형별 점수 분포]



[파일럿 데이터 증폭]

□ 활용 방법

- 검증화면 구성 및 시범운영('19.11.13~'20.6.30)
- 종합조사 결과관리 화면에서 집단간 신뢰구간(95%)에 따른 점수범위를 확인하여 영역별 평가결과 검증

※ 향후 매년 검증결과의 신뢰도 향상을 위해 모델 고도화 지속 추진

□ 활용 성과

조사결과 입력 시 정상 범주를 벗어나는 경우 가이드를 제시해 조사결과의 신뢰도 향상

□ 느낀점

- 데이터가 많아질수록 분석이 어렵다고 생각했는데 적은 데이터는 증폭이 필요한 경우가 있어 까다로울 수 있다는 것을 알게 됨

3-1

추가 분석 과제

증명서 발급 취약지역 파악

민원 편의성 향상을 위한 무인발급기 위치 선정

※ 공단 증명서 발급 업무 개선과제 중 일부인 무인발급부분 참여

□ 분석 배경

- 전체 민원업무의 증명서 발급이고, 매년 지속적으로 증가
- 비대면채널로의 전환비율이 높아지고 있지만, 여전히 방문 민원건수도 많은 상황(연간 130만 건)
- 발급수요 및 접근성을 고려한 무인발급기 설치 최적지 파악 필요

□ 활용 데이터

- 증명서 발급 이력, 무인발급기 위치 정보, 지역별 거주인구, 버스정류소 및 지하철역, GIS정보

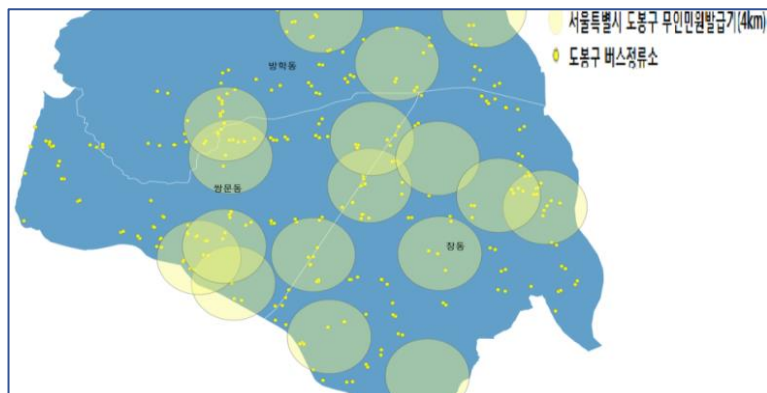
□ 분석 기법

- 무인발급기 위치를 기준으로 주변 대중교통 및 인구수를 고려한 접근성 파악
- 거주지부터 무인발급기 까지 최단거리 거리계산 및 취약지역 시각화 분석
- 취약지수 산출을 위한 모형 개발

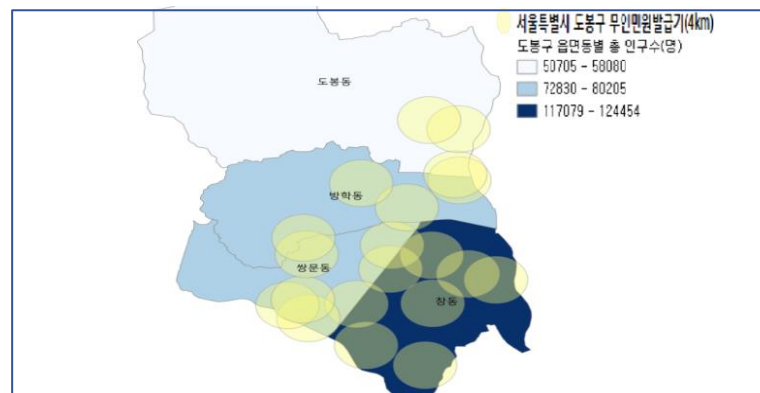
□ 무인발급기의 접근성에 대한 기초현황 분석

- 무인발급기 위치 기준 주변 지역의 버스정류소, 인구 수를 비교하는 접근성을 파악

□ 취약지역 파악



무인민원발급기와 버스정류소의 위치 비교



인구수와의 비교

예시) 힐트리움아파트 ↔ 도봉구청 무인발급기 최단경로는 4.1Km이지만, 도보 및 대중교통을 이용 시 17분 소요

- 인구수, 버스정류소가 많은 지역임에도 불구하고 반경 4Km내에 발급기가 없는 **사각지대 발생**

□ 추진성과

- 무인발급기 확대 설치계획 수립 시 분석결과 활용 예정('21년 예산 요구)

□ 느낀점

- 교육 때 배웠던 표준 분석모델 중 하나인 CCTV설치 분석방법을 직접 활용해볼 수 있는 기회였음
- 일상생활에서 느끼는 불편함을 분석적 시각에서 바라보면 좋은 아이디어가 될 수 있다고 생각함

3-2

추가 분석 과제

상담내역 텍스트 분석

고객 상담 불만원인 파악 및 서비스 품질 개선

※ 공단 상담서비스 개선과제 중 일부인 일시금 지급 개선부분 참여

□ 추진 배경

- 일시금 지급 상담내용에 대한 심층분석으로 불만 발생 주요원인 파악

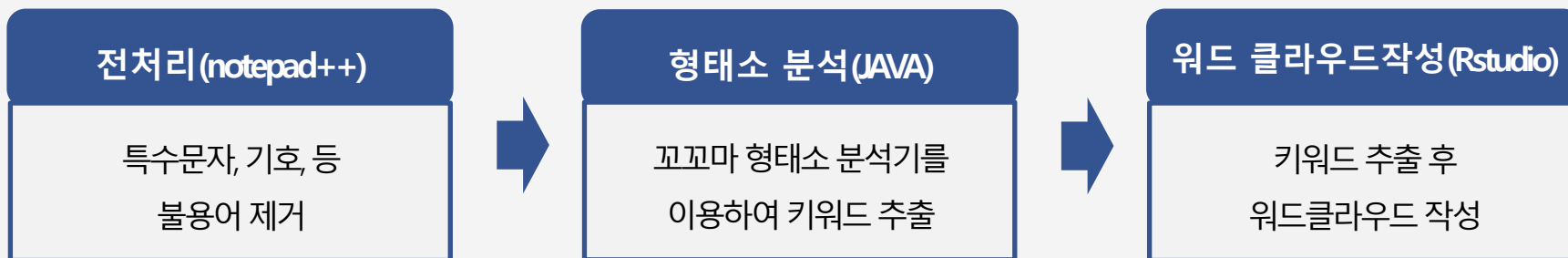
* 일시금 : 연금 수급조건에 해당하지 않아 한꺼번에 금액을 받는 것

□ 활용 데이터

- 고객접촉이력 및 상담내역등록자료

□ 분석기법

- 고객 상담내용 등록데이터를 기초로 텍스트분석 수행

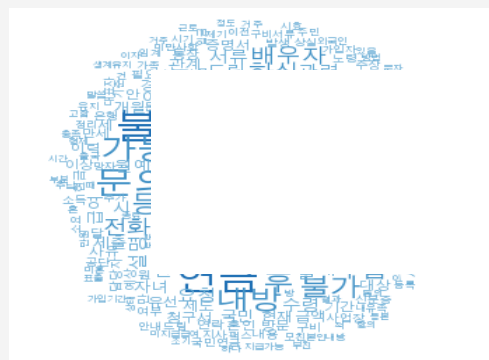


□ 불만원인 파악

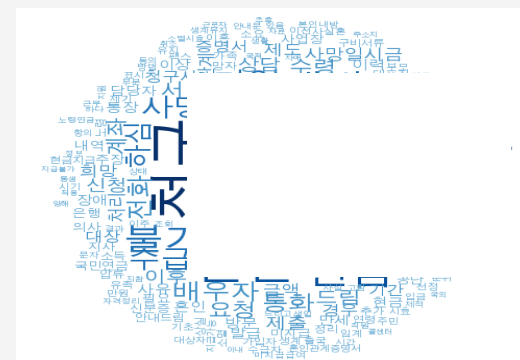
- 분석결과 주요원인은 상담내용 자체의 문제가 아니라 본인여부를 확인하는 절차의 불편함 (공인인증, 신분증 제시 등)



1) 상위 90% 단어 추출



2) '일시금', '지금', '안내' 제외



3) 상위 80% 단어 추출

□ 추진성과

- 신분증 진위여부 확인 자동화 등 본인 확인절차 단축으로 상담 편의성 향상

□ 느낀점

- VOC분석 등 다양한 과제 수행 시 핵심기술이 될 수 있는 텍스트분석에 대해 배울 수 있는 기회였음

04

청년인턴을 마치며

4-1 느낀점

4-2 향후 발전방안

□ 업무적 측면

- 데이터 분석 학습 외에도 사회 취약계층을 위한 공익적 과제에 참여한 것에 보람을 느낌
- 직원 분들이 분석대상업무에 대해 너무나 열심히 사전학습을 하는 모습을 보며, 다양한 분야의 사회적 이슈들에 대한 관심을 가져야겠다고 생각하게 됨
- 처음엔 공단 용어가 낯설어 어려움을 겪었는데, 직원 분들과의 인터뷰, 자료조사 같은 방법을 통해 극복하였고 새로운 분야에 대해 이해하는 과정을 배웠음
- 분석과제를 추진하면서 분석결과의 활용, 성과를 내는 방법에 대해 배웠고 이런 것들이 매우 중요한 부분이라는 것을 알게 되었음

□ 분석적 측면

- 인턴 경험을 통하여 분석가로서 더 공부해야 할 것, 부족한 점들을 파악할 수 있었음
- 이전에는 이론적인 공부를 했다면, 이제는 배웠던 내용이 실제 업무에는 어떻게 적용되었는지 떠올릴 수 있어 보다 구체적인 목표를 가지고 역량을 키워나갈 수 있을 것임
- 데이터 건수가 많을수록 분석이 어려울 것이라고 생각했었는데, 오히려 적은 데이터 분석이 더 어려울 수 있다는 것을 알게 됨
- 제대로 된 데이터 분석을 하기 위해 요구되는 통계학 지식에 대한 중요성을 인식함

□ 참여 과제

- 개발한 모형들에 대한 정확도 개선 등 지속적인 성능 업그레이드 필요
- 업무 처리중심으로 되어있는 데이터 구조를 고객 중심으로 재구성하여 업무 활용도 향상

□ 자기 계발

- 평소 다양한 분석 아이디어를 찾고 구현하는 연습을 통하여, 어떤 문제가 주어졌을 때 해결할 수 있는 능력 배양

□ 청년인턴 제도

- 보다 많은 청년들이 인턴과정에 참여해서 실전 능력을 배양할 수 있는 기회가 확대되길 바램
- 원하는 기관에 가기 위해 연고지와 떨어진 곳에 오다 보니 체류비용이 많이 들었는데, 앞으로는 이러한 부분을 위해 경제적 지원도 조금 더 많이 해주시면 좋겠음

6개월간 청년 인턴에 참여하며 많은 것을 배우고 경험할 수 있었습니다.
소중한 기회를 주셔서 감사합니다.

이외
공단에 남긴 흔적...

[별첨] 상담고객 사례집 첨부 자료(1)

평생상담 한 건당 A4종이 1.1장

→ 총 14,853,651 장 * 8.08원 = 1억 2천만원

→ 120,000,000원 / 치킨 한 마리 가격 = **치킨 706마리**



평생상담 한 건당 A4종이 1.1장

→ 총 14,853,651 장 * 8.08원 = 1억 2천만원

→ 120,000,000원 / nps카페 아아 = 총 6000잔



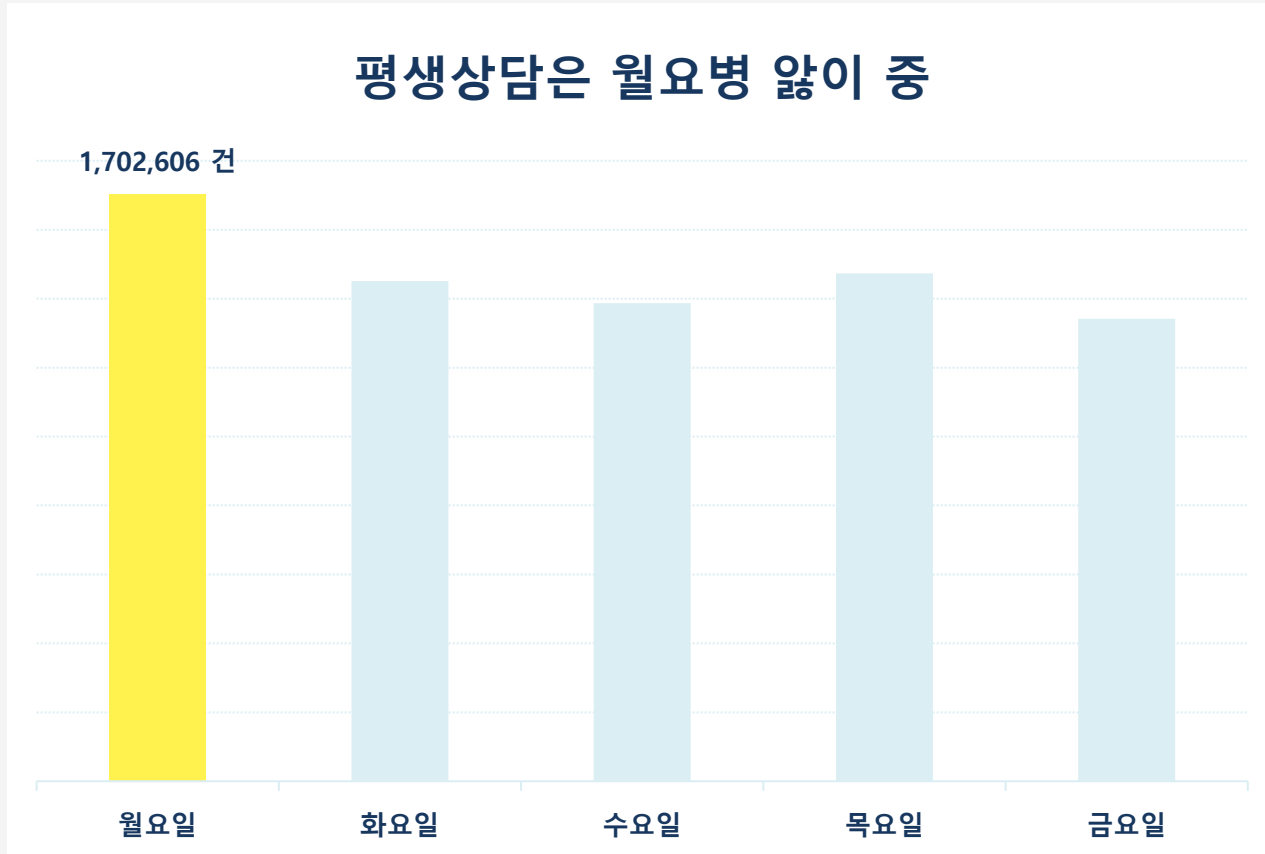
평생상담고객이 일렬로 줄을 선다면...

총 14,853,651 건 * 1인 차지 공간(0.5m)

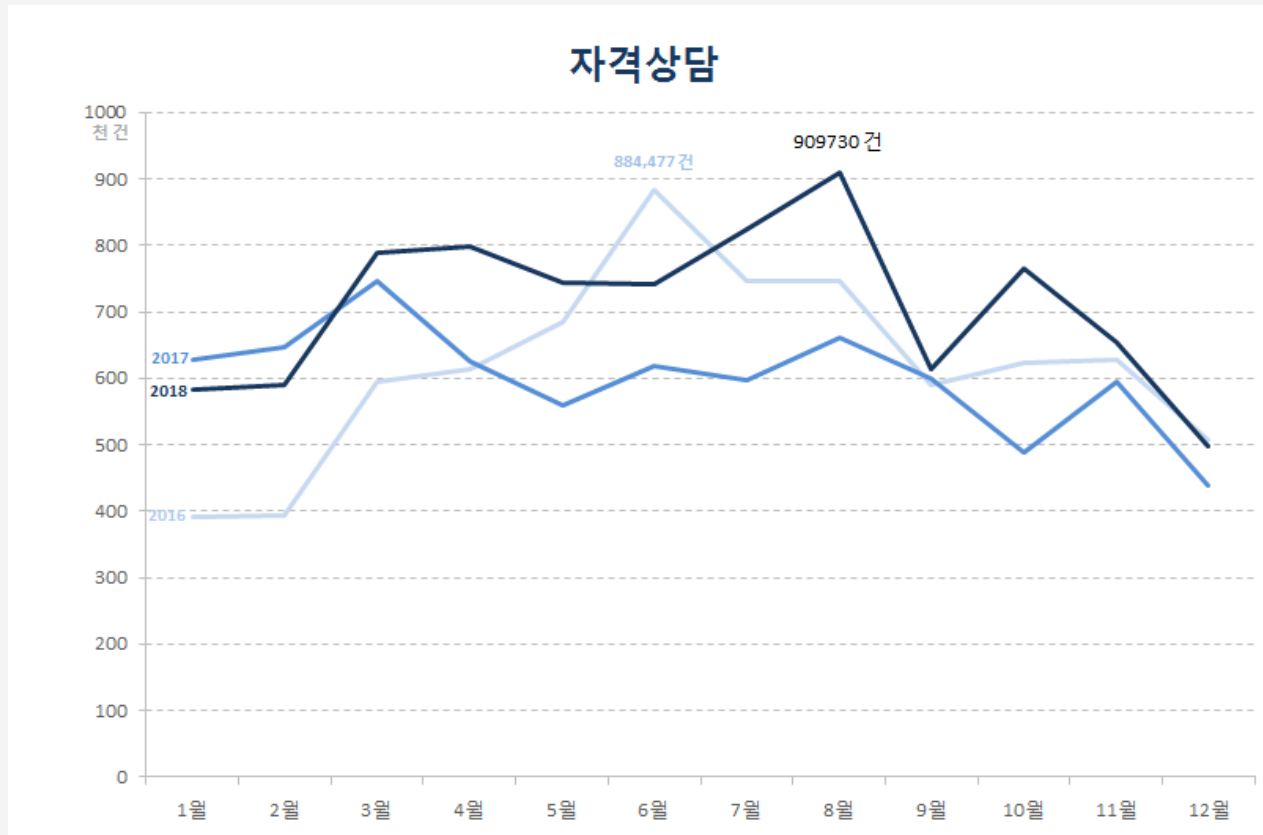
= 약 7426km → 공단에서 전주한옥마을 왕복 345번



[별첨] 상담고객 사례집 첨부 자료(2)

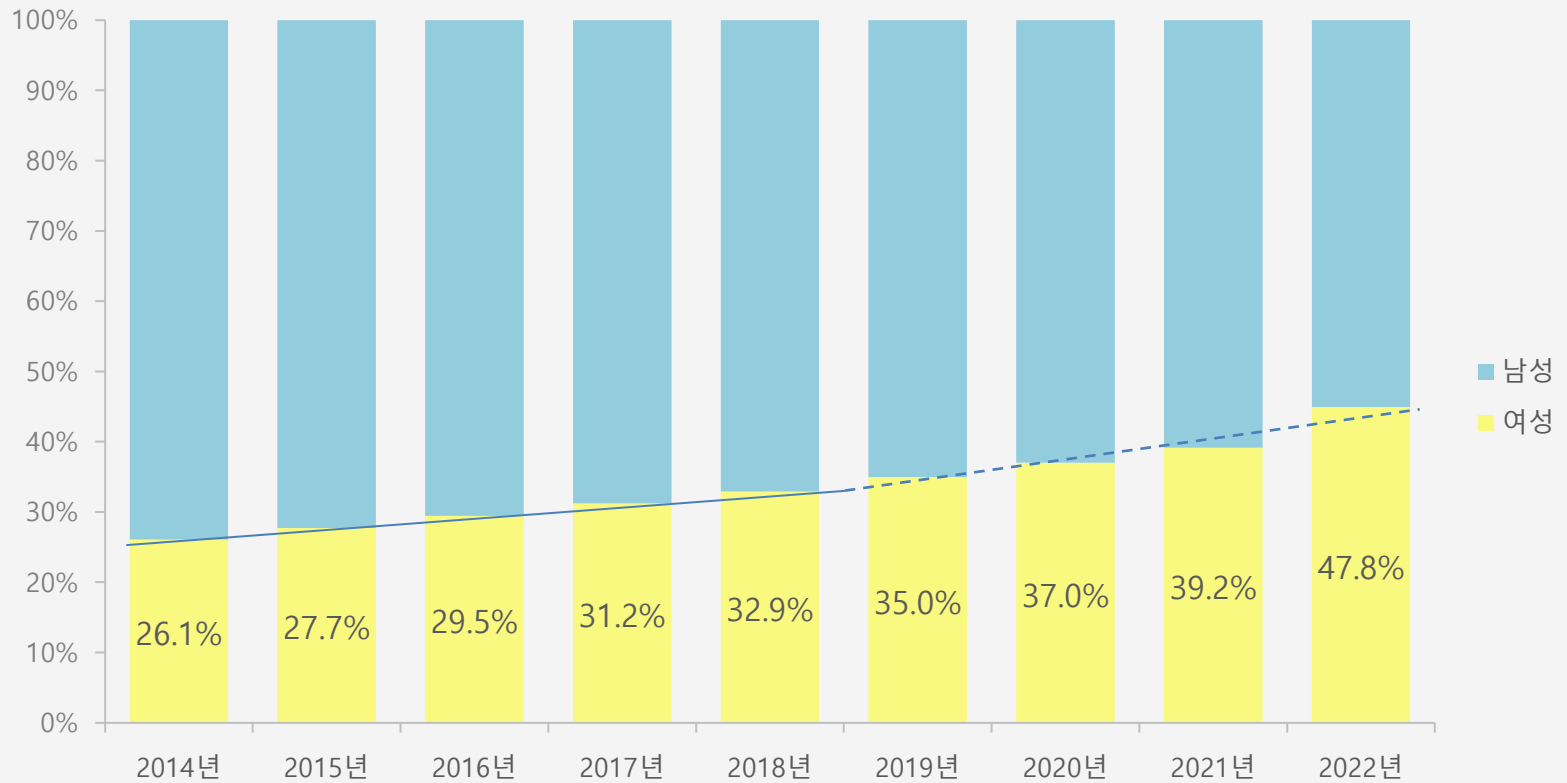


[별첨] 상담고객 사례집 첨부 자료(3)

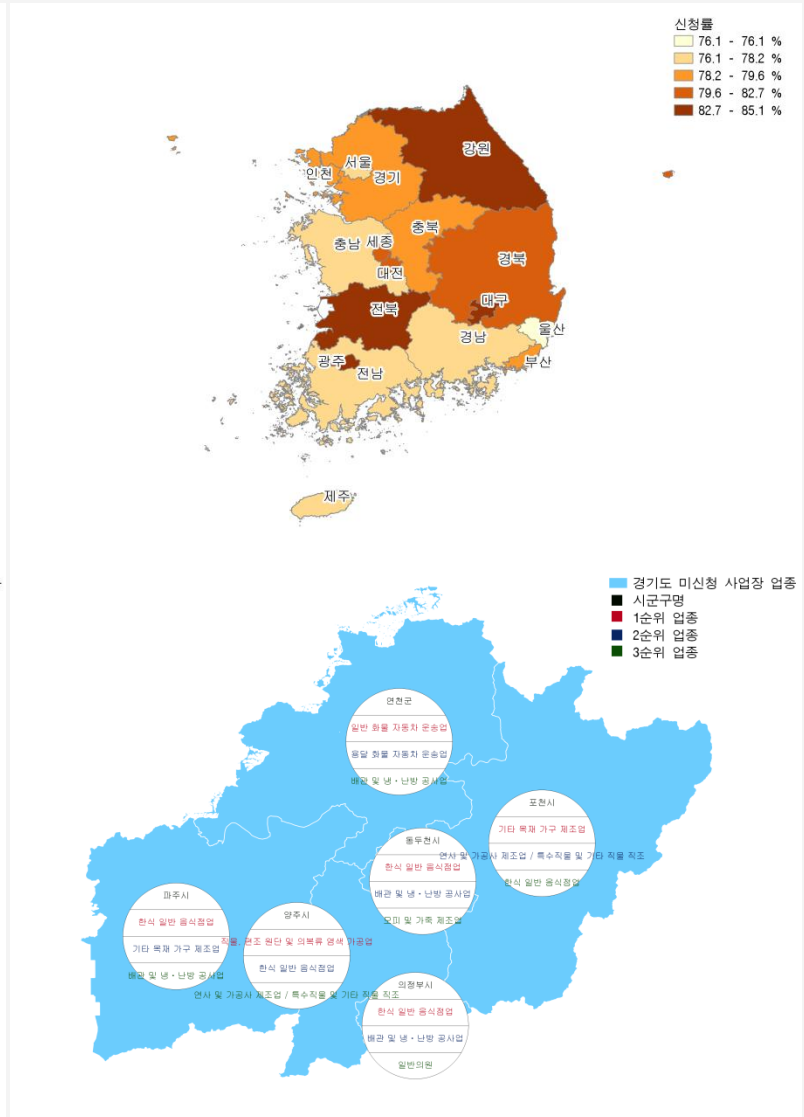
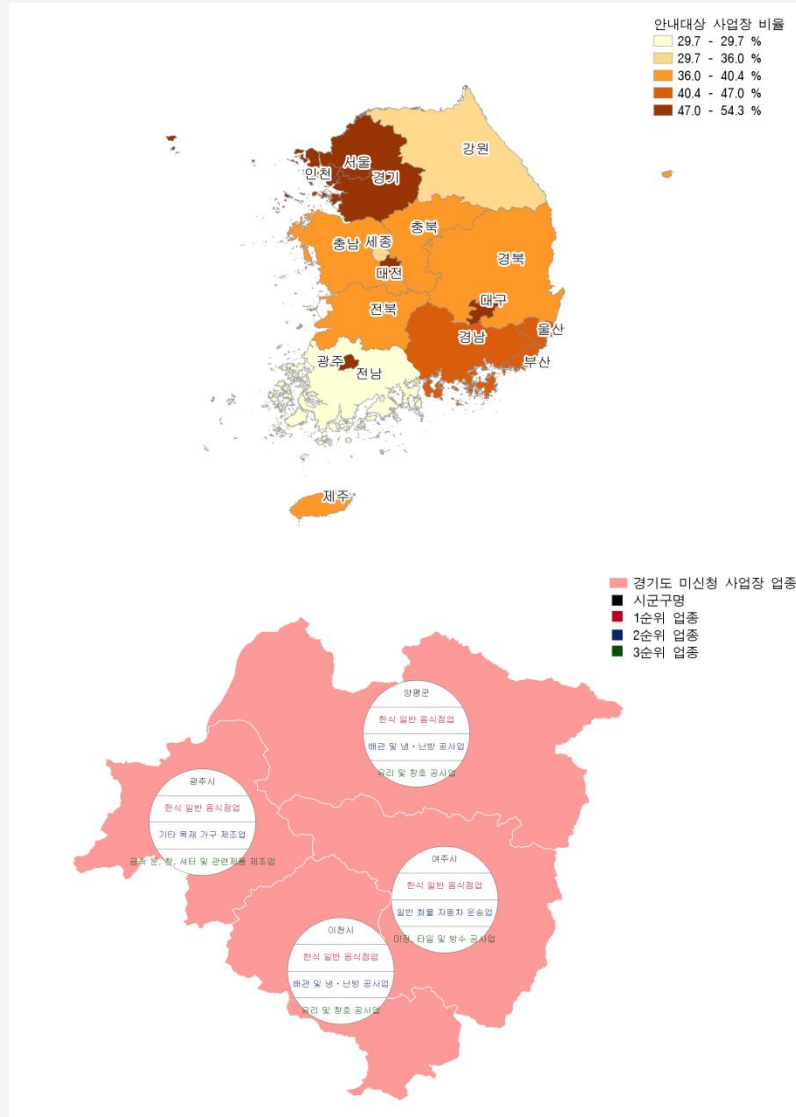


- 평생상담이력 중 자격상담은 9월에 하락하는 경향이 있고, 연초와 연말에는 3개년인 비슷한 추세를 보이지만 나머지 월에서는 모두 다른 경향을 보임

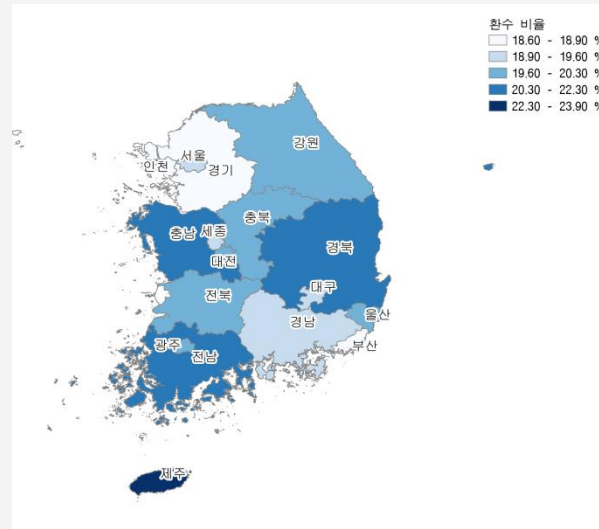
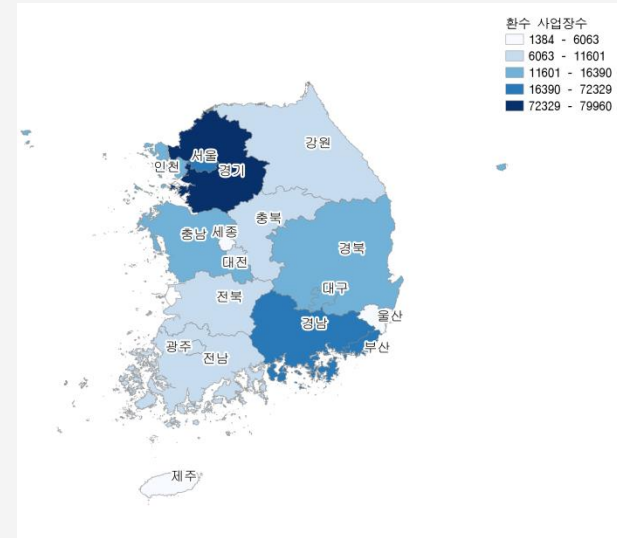
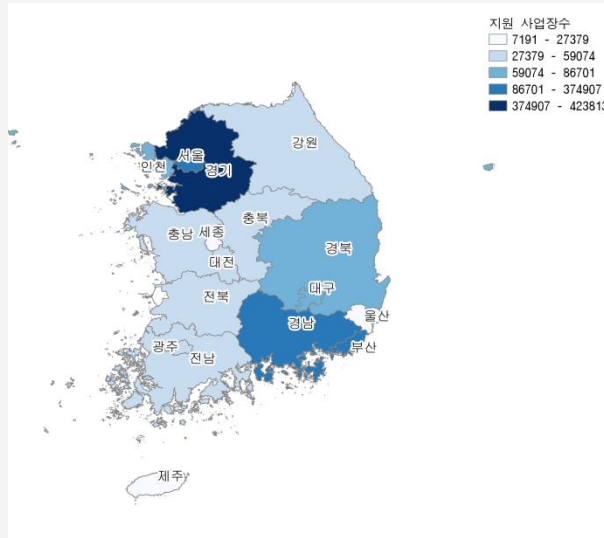
[별첨] 기관평가 지표 분석과제 시각화



[별첨] 국민연금 보험료 지원 사업 관련 시각화



[별첨] 국민연금 보험료 지원 사업 관련 시각화



“ Q & A ”

“ THANK YOU ”