A solid yellow circle is positioned to the left of the word 'PORTFOLIO', partially overlapping the letter 'P'.

PORTFOLIO

About me



안녕하세요 반갑습니다😊
세상에 없던 새로운 컬러,
예비 데이터분석가 정별입니다.
신선한 접근과 맡은 바 임무에 헌
신적인 태도로 다양한 프로젝트를
진행해왔습니다.

Education

한국외국어대학교 AI융합전공, 정치외교학과

(2019~현재 휴학중)

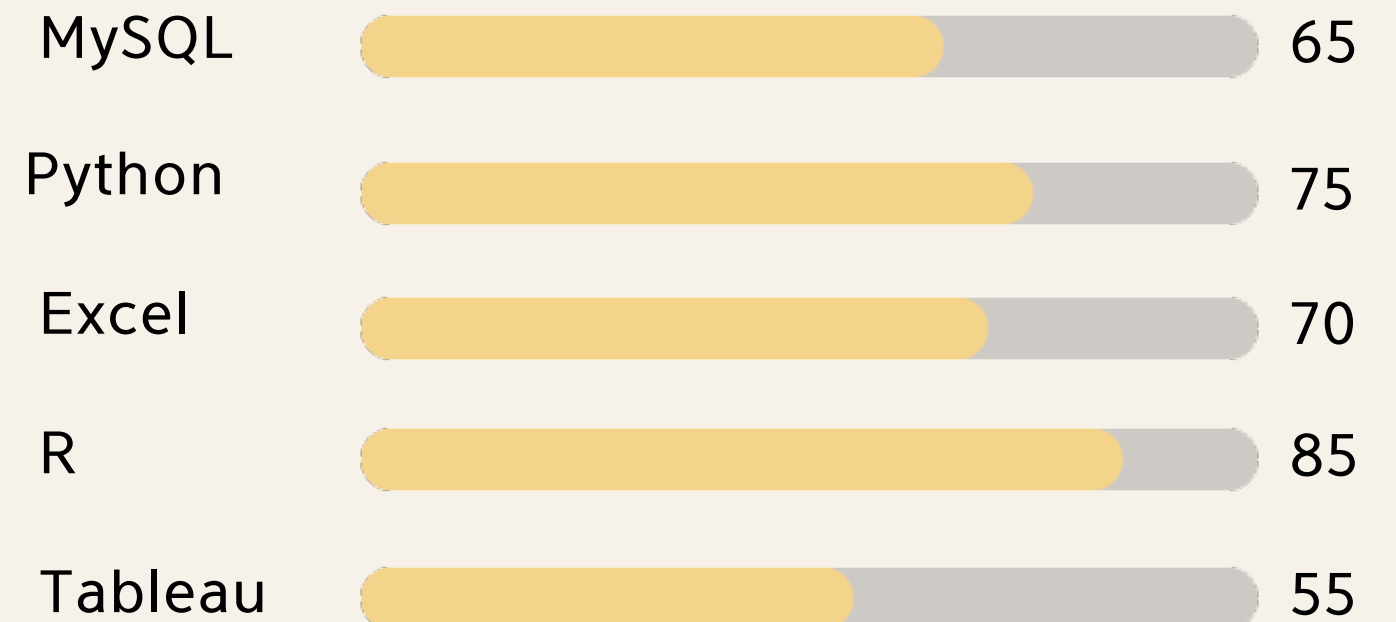
주요 이수과목: 데이터베이스, 데이터사이언스, 빅데이터, 빅데이터
와 정치, 정책학개론 등

패스트캠퍼스 데이터분석 부트캠프 수료

(2023.12~2024.5)

주요 이수내용: 기업연계 프로젝트, EXCEL, SQL, PYTHON,
TABLEAU 강의 및 각 프로젝트

Stacks



Awards

KOSSDA 대학생 공모전 2024 최우수상

멋쟁이사자처럼대학 11기 아이디어톤 9등
(총 354팀)

Project 01

이커머스 기업 G사의 새로운 주력 제품 발굴 프로젝트

프로젝트 개요

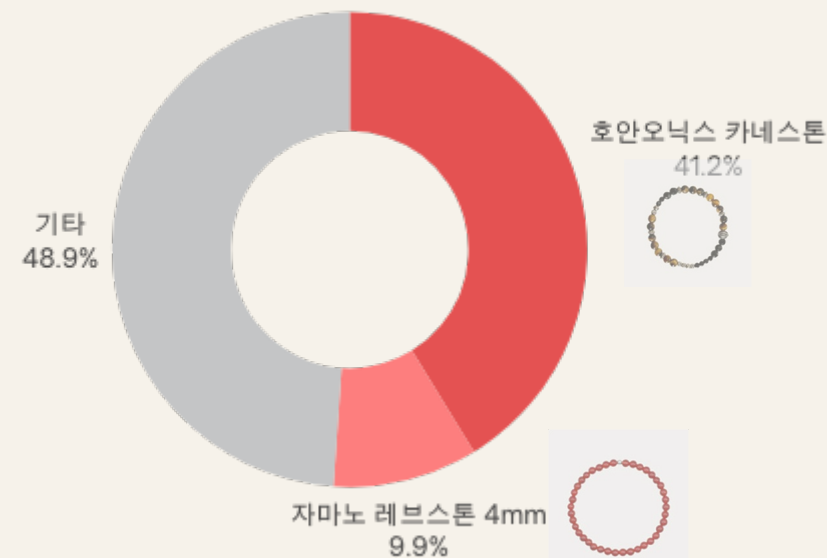
분석 배경(Why)

- **주력 제품에 대한 높은 매출 의존도:** 주력 주얼리 제품 2개의 매출 비중이 51.1%→매출의 불안정성 증대
- **광고 집행의 일관성 부족:** 매출 상위 5개 제품의 매출 대비 광고비에대한 큰 변동폭 (14.2%~86.1%)

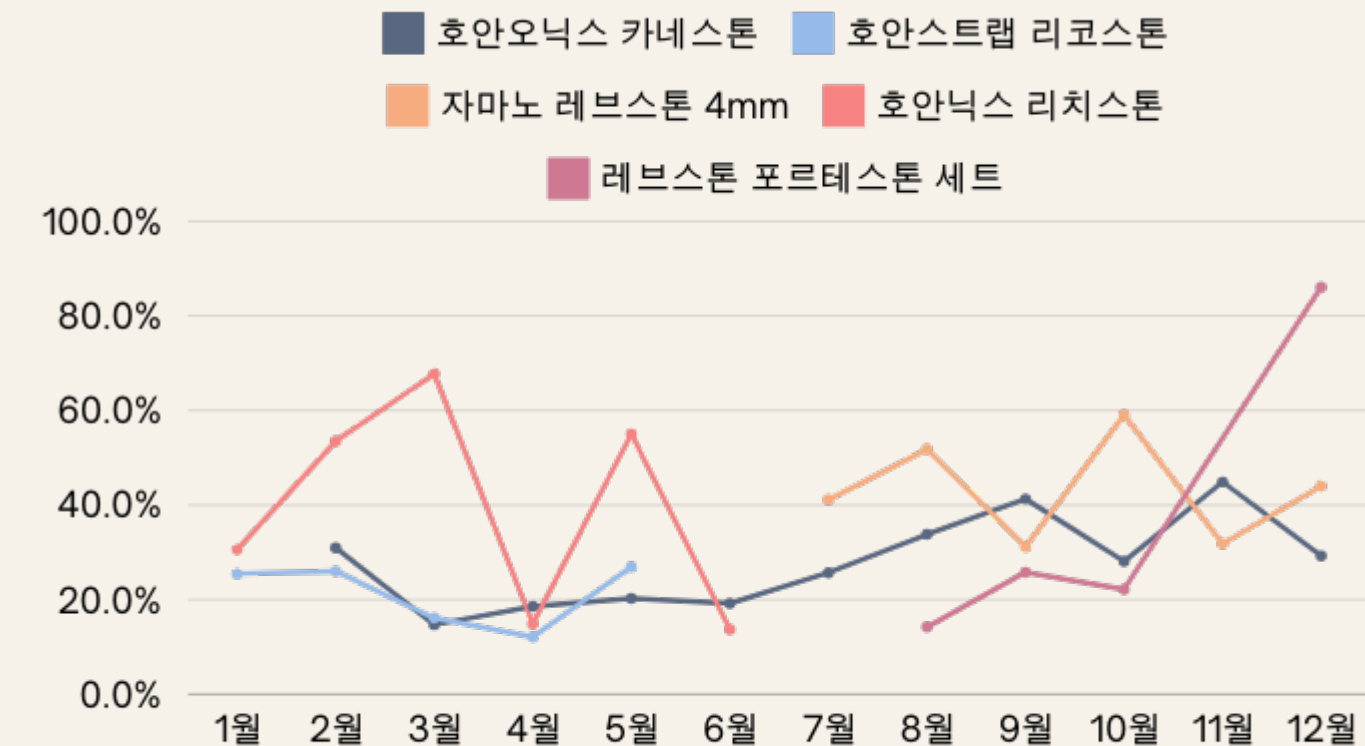
분석 목표(What)

- **G사의 새로운 주력 제품 발굴**→ 주문 데이터, 광고 데이터에 기반한 새로운 지표 생성 및 기준에 부합한 새로운 제품군 선정
- **발굴된 후보 제품에대한 마케팅 전략 제시**
→ 리뷰 데이터, 광고 데이터에 기반한 고객 분류 및 제품별 브랜딩 제시

<주력 제품 매출 비중>



<제품별 매출 대비 광고비 비율>



분석 방법(How)

- 데이터: **G사 내부 데이터 활용**- 주문, 고객 정보, Google Analytics, Looker Studio, 제품 데이터셋 활용
- 사용 tool: Pandas, Excel, MySQL, Tableau, R
- 분석 기법: Cohort 분석, TFIDF 및 빈도분석 등

이커머스 기업 G사의 새로운 주력 제품 발굴 프로젝트

개념 설계 및 분석

Data Preprocess

- 일괄적으로 제품별 매출, 구매 수량, 고객 정보를 추출(시각화)하기 위하여 전처리 진행
- 최종적으로 2개의 통합 데이터 생성(주문 통합, 광고 통합)-> 1번의 주문 정보/ 1개의 광고 링크 column을 primary key로 설정

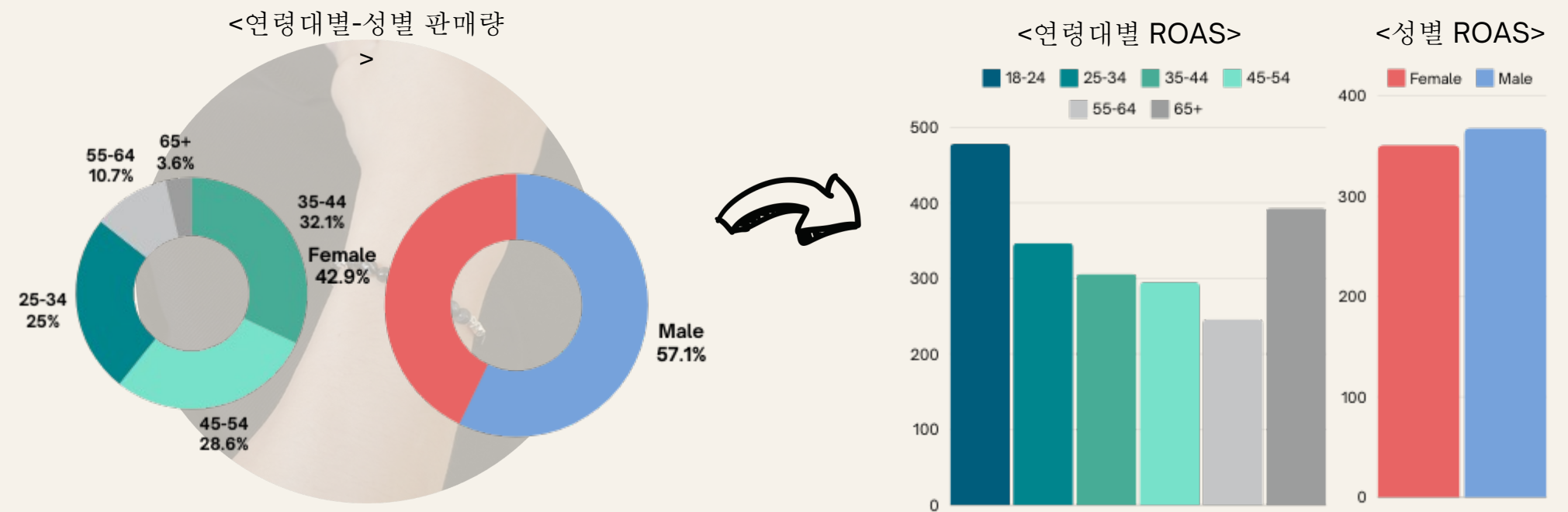


EDA

- 매출액과 광고비의 시계열 분석을 통한 G사의 '기존 주력 제품에 대한 높은 매출 의존도' 및 '광고 집행의 불안정성' **문제 발견**
- 시장조사에 기반한 제품별/월별/계절별 ROAS 분석을 통한 **후보제품 기준 선정**(+고객 주소 정보에 기반한 바이럴 마케팅 분석 진행)

데이터분석 및 시각화

- Cohort Analysis
-선정된 후보 제품군에대한 연령대/성별 판매량, ROAS, 구매 전환율, 고객 유지율 분석 진행



- Text Analysis
G사 전체 리뷰, 부정 리뷰, 후보 제품별 리뷰에 대하여 **TFIDF** 및 **Keyword-Network** 분석 진행

Project 01

이커머스 기업 G사의

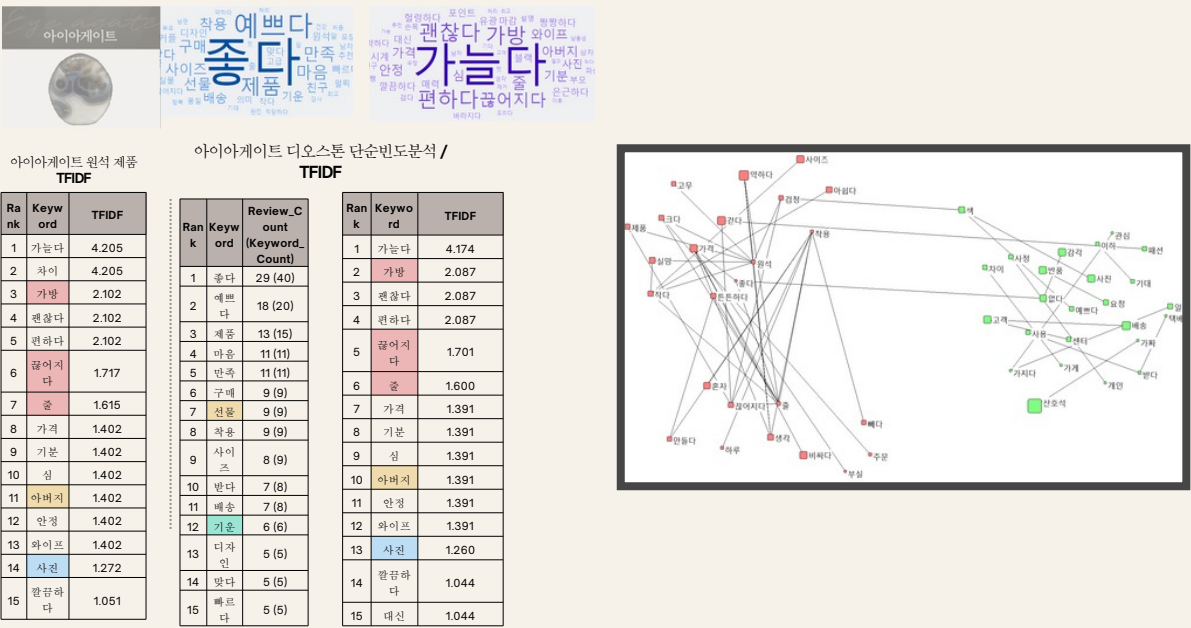
새로운 주력 제품 발굴 프로젝트

개념 설계 및 분석

데이터분석 및 시각화

- Text Analysis
- TFDIDF 및 단순빈도분석: 선정된 후보 제품들에대한 고객들의 소구점 파악

Keyword-Network 분석: 전체/부정 리뷰에서 확인되는 부정적인 특징들이 후보 제품들에서는 나타나지 않거나 오히려 긍정적으로 나타나는 현상 파악



분석 결론

- G사의 새로운 주력 제품 3개를 제시->주력 제품과 함께 새로운 제품 발굴의 지표 제시
- 제시된 3개 제품에 대하여 차별화된 브랜딩 전략 제시->성별/연령대별 ROAS,리뷰에 기반한 브랜딩

기대효과

- 새로운 주력 제품을 제시함으로써 기존 주력 제품의 높은 매출 의존성 해소
- 급진적 변화나 추가적 비용 지출 없이, 기존에 낭비 되던(월별 ROAS 200 미만) 광고 집행비를 새로운 후보 제품에 전용하도록 제안함으로써 효율적 광고 집행 기대
- 간단하지만 확실한 전략을 통해 G사 유관부서에서 바로 시행 가능함

평가

*G사 대표이사 및 현직 데이터분석가들의 평가

- 새롭게 제시해주신 주력 후보군을 바탕으로, 새롭게 광고를 집행해 볼 수 있을 것 같습니다.

- 고객과 마케팅 기법에만 집중한 다른 팀들에 비해 '왜 사람들이 이 제품을, 이 원석을 살까?'에 대한 근본적인 질문을 던지고 그에대한 답을 찾아가는 과정이 논리적이고 특색 있었음.
- 프레젠테이션 과정을 통해 결국 청자를 설득하는 것은 차가운 데이터가 뒷받침된, 피부에 와닿는 감동임을 경험함.

주제에 맞는 분석내용 이었는가? (30)	주장하는 바가 논리적이었는가? (30)	도출한 인사이트 가 도움이 되었는 가?(30)	발표(10)	비즈니스 스에 적용하면 좋은 전략 및 인 사이트	총점	평가
30	30	30	9	새로운 제시해주 신 주력 후보군을 바탕으로, 새롭게 광고를 집행해볼 수 있을 것 같습니 다.	99	시장 현황부터 시작한 것이 흥미로웠습니다. 눈에 잘 들어오는 발표 자료였습니다. 시장 조사와 시장 트렌드를 바탕으로 주력 제품 확장을 제시해주신 부분이 새롭고 흥미로웠습 다. 그 내용을 연결해서 연령대별 마케팅 방안도 제시해주신 것이 좋았습니다. 수고 많으셨습니다.

Project 02

비도심 지역간 노인복지 불균형은 얼마나 심각한가?

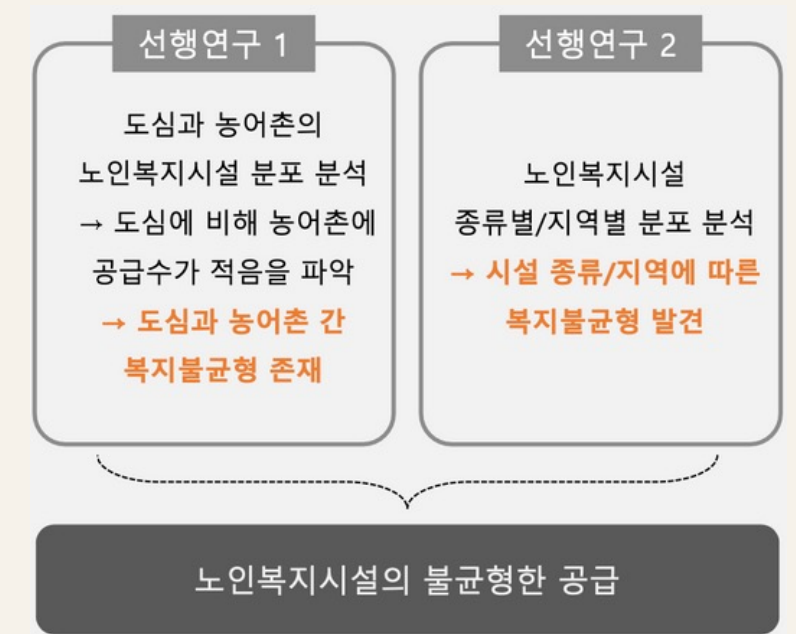
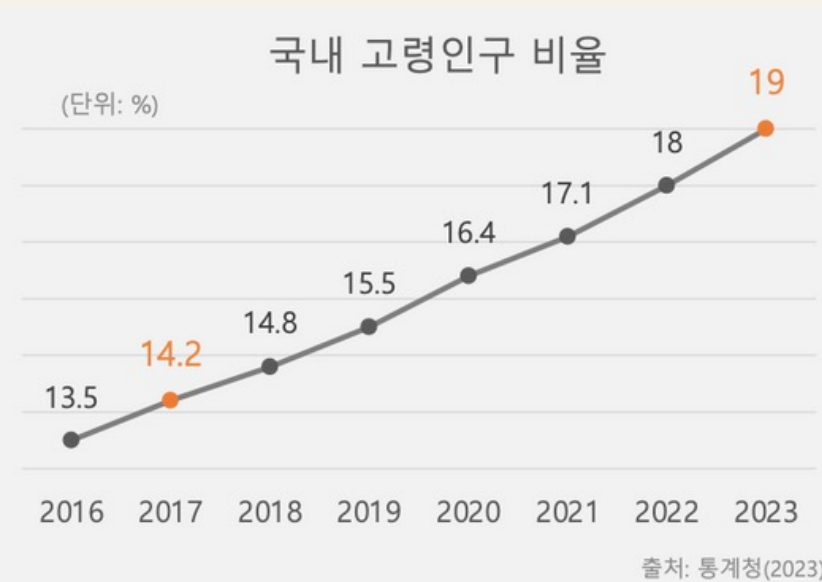
프로젝트 개요

분석 배경(Why)

- **초고령 사회로의 진입을 목전에 둔 대한민국:** 고령 인구 비율이 2017년 14.21%로 고령 사회 진입, 2023년 19%로 초고령 사회로의 진입을 앞둔 상태
- **노인복지의 지역 간 불평등을 지적한 선행연구 동향:** 2000년대부터 도심과 농어촌 간의, 지역 간의 노인복지 불평등을 지적함

분석 목표(Question)

- **비도심 지역 사이에서 노인복지시설의 불균형 문제는 존재하는가?:** 노인복지 불균형을 다룬 선행연구 동향(도심vs농어촌, 수도권vs비수도권)의 이분법적 구도를 벗어남. '진짜' 수요인 비도심 지역 간 차이에 주목
- **노인복지시설이 불균형한 지역 간에 어떤 차이가 나타나는가?:** 노인복지시설이 과대/과소 공급 되는 지역들에대한 분석 진행



분석 방법(How)

- 데이터: 공공기관 2차데이터 활용- 지방재정365, kosis 국가통계포털
- 사용 tool: Excel, R, Tableau, Pandas
- 단순 회귀 분석, 지표분석 활용

Project 02

KOSDA 대학생 공모전 2024 ‘데이터가 보여주는 한국사회’

비도심 지역간 노인복지 불균형은 얼마나 심각한가?

개념 설계 및 분석

Data Preprocess

- 서울 및 6개 광역시 제거 / 행정구역 명칭(시도 / 시군구) 표준화 / 숫자 단위 통일 / 노인장기요양보험 1~3등급 이외 등급 제거 등

```
11 #시군구별 65세이상 전체 건강보험 대상자
12 total_pop<- read_excel('total_pop.xlsx') #preproceesed data
13 #데이터 읽음
14 total_pop<- total_pop%>% filter(!is.na(건강보험_대상자))
15 View(total_pop)
16 #시군구별 노인복지 기관수
17 place <- read_excel('newplace.xlsx') #preproceesed data
18 #데이터 읽음
19 place<- place %>% filter(!is.na(전체))
20 View(place)
21 #시군구별 장기요양보험 대상자(1-3등급) 수
22 in_pop <- read_excel('new_insurance_pop.xlsx')
23 View(in_pop)
24 #등급별 총합 전처리
25 library(tidyr)
26 # NA 값을 이전 행의 도시 이름으로 채우기
27 in_pop_filled <- in_pop %>%
28   fill(시군구)
29 View(in_pop_filled)
30 # 도시별로 그룹화하여 합을 계산
31 in_pop_sum <- in_pop_filled %>%
32   group_by(시군구) %>%
33   summarise(인원수 = sum(인원수))
34 View(in_pop_sum)
35 write.xlsx(in_pop_sum, 'in_pop_sum
(시군구별_장기요양보험_대상자).xlsx') #preproceesed data
36 #데이터 읽음
37 #데이터를 조인
38 total_in_pop_join<- left_join(in_pop_sum, place, by
="시군구")
39
40 #지역별 인구 1000명당 보험 대상자 대비 시설 비율(실재수요)
41 #해당지역에서 모든 노인인구 1000명씩 나뉠때, 각 그룹이
이용할 수 있는 시설의 수 => 해당 시설 n개가 1000명을
커버한다.
```

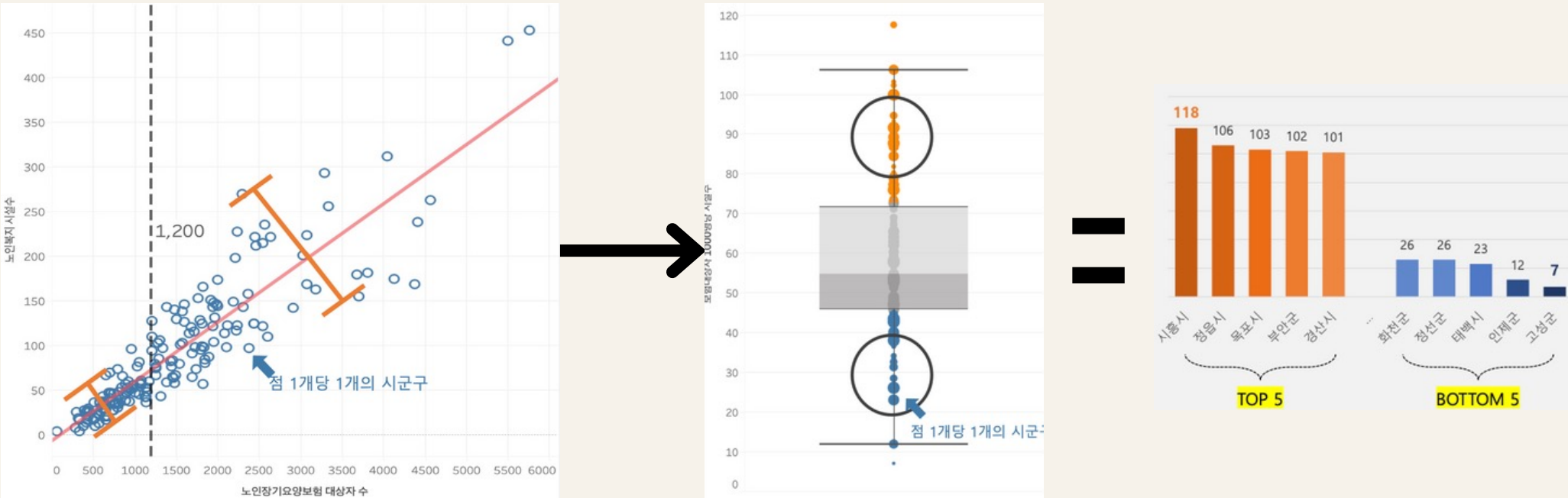
<R코드중 시군구별 노인장기요양보험
대상자 전처리 일부>

EDA

- 각 지표 별 상하위권 지역 추출(정렬 활용)
- 실재수요 파악 위해 각 지역의 전체 노인인구(건강보험대상자) 대비 노인장기요양보험 대상자 수, 시설 수, 정원 등 분석

데이터분석 및 시각화

- 단순 회귀분석(Simple Regression Analysis)
 - 노인장기요양보험 인구(X)와 노인복지시설(Y) 간 상관관계 분석
 - X값이 1,200이상일 때, 회귀선(적정수준)으로부터 각 점들의 거리가 증가



- >’노인장기요양보험 인구 1,000명당 노인복지시설 수’의 지역별 편차가 매우 크다는 사실 시각화
- > **Top5** <> **Bottom5** 지역 사이의 차이가 매우 두드러짐

Project 02

KOSDA 대학생 공모전 2024 ‘데이터가 보여주는 한국사회’

비도심 지역간 노인복지 불균형은 얼마나 심각한가?

개념 설계 및 분석

데이터분석 및 시각화

- 지표 분석
 - 불균형 지역간 노인복지 격차를 파악하기 위한 정량지표 3가지 선정 및 전처리:
‘전체 노인인구 1,000명당 의료기관 수’, ‘노인장기요양보험 인구 대비 노인복지시설 정원 수’, ‘지자체별 노인복지 예산 규모’

분석 결론(QnA)

- 비도심 지역 사이에서 노인복지시설의 불균형 문제는 존재하는가?: 존재함. ‘노인장기요양보험 인구 수’와 ‘노인복지시설 수’ 사이의 관계에서 지역간 실제 수요대비 시설 공급의 불균형한 양상을 발견함
- 노인복지시설이 불균형한 지역 간에 어떤 차이가 나타나는가?: 노인복지자원 지표에서 불균형이 나타남. 3가지 지표 모두 TOP5지역에서 BOTTOM5지역에 비해 높게 나타남

시사점

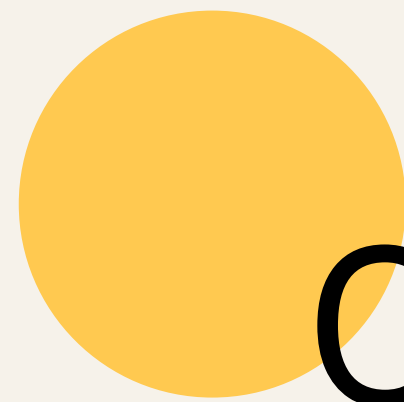
- 공모전의 취지와 같이 대학생의 시선에서 데이터의 눈으로 한국사회를 조망할 수 있는 시간이자, 새로운 질문과 새로운 답변을 데이터의 언어로 문답하는 과정이었음.

평가

*한국사회과학자료원 심사위원 평가

- 노인복지 불균형 문제에대해 선행연구들의 분석틀을 벗어나 비도심 지역으로 대상을 한정한 점이 인상깊었음
- 오늘날 정부와 사회에서 주요하게 다루는 주제를 데이터를 통해 논리적으로 잘 전달하였음





Contact.



+82 010-4354-2784



jb7458@naver.com



<https://github.com/JungByoul>



Seoul, Republic of Korea

Thank you

감사합니다