



PORTFOLIO

About me

안녕하세요 반갑습니다😊
세상에 없던 새로운 컬러,
예비 데이터분석가 정별입니다.
신선한 접근과 맡은바 주인의식을
갖는 태도로 다양한 프로젝트를 진
행해왔습니다.

Education

한국외국어대학교 AI융합전공, 정치외교학과
(2019~졸업예정)

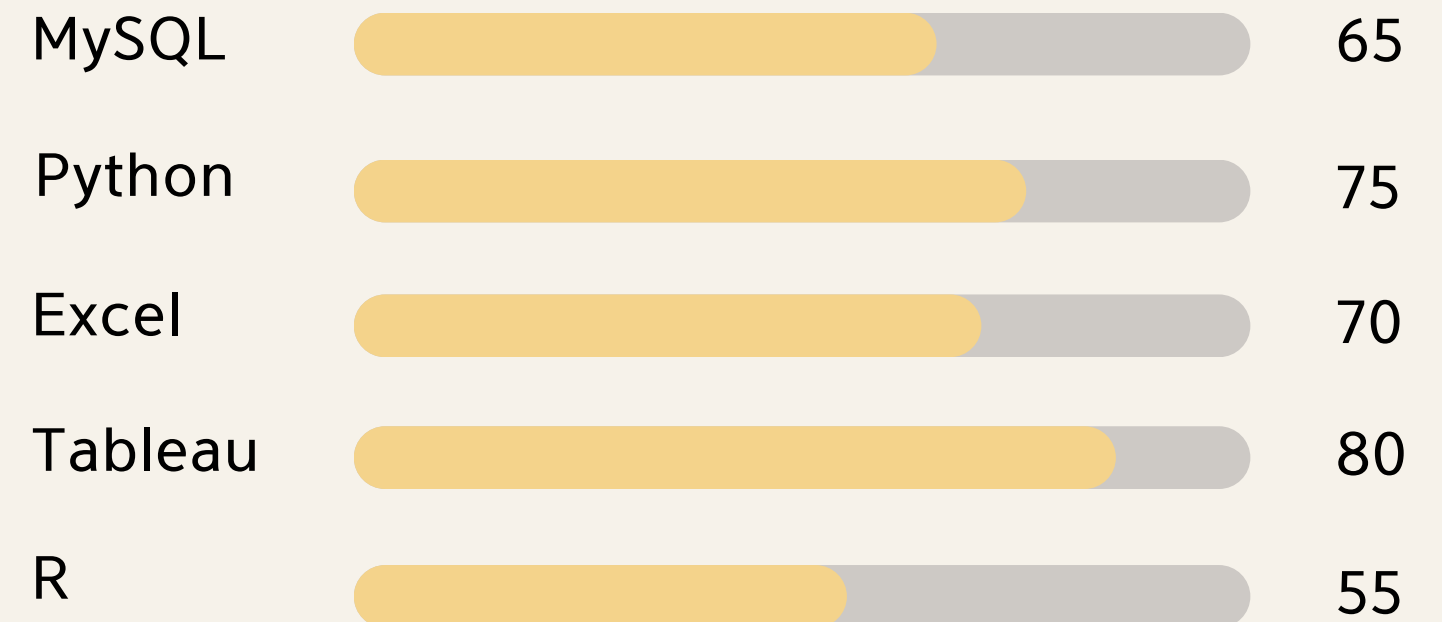
패스트캠퍼스 데이터분석 부트캠프 수료

(2023.12~2024.5)

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

Data.Popcorn n8n 스터디 및 프로젝트 수료
(2025.1~2025.3)

Stacks



Awards

KOSSDA 대학생 공모전 2024 최우수상

2024 위해정보 데이터 분석 보고서 공모전 3위(장려상)

멋쟁이사자처럼대학 11기 아이디어톤 9등
(총 354팀)

Project 01

이커머스 기업 G사의 새로운 주력 제품 발굴 프로젝트

프로젝트 개요

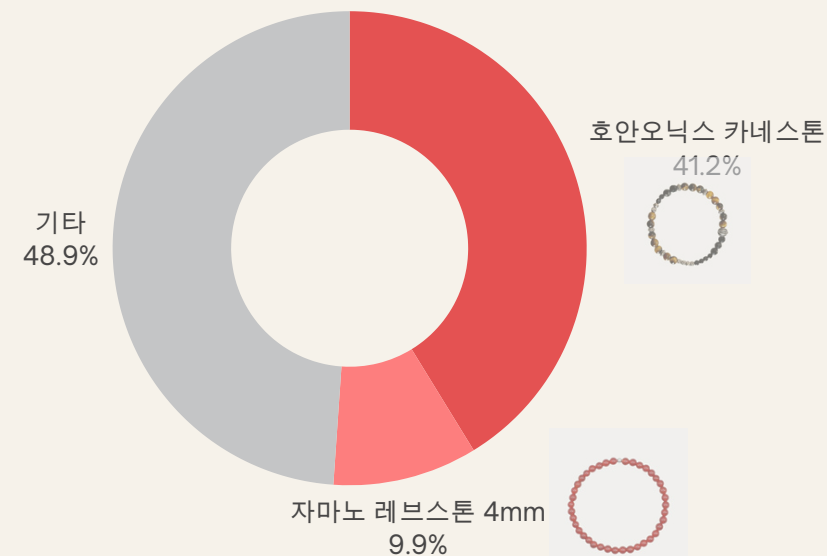
분석 배경(Why)

- **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

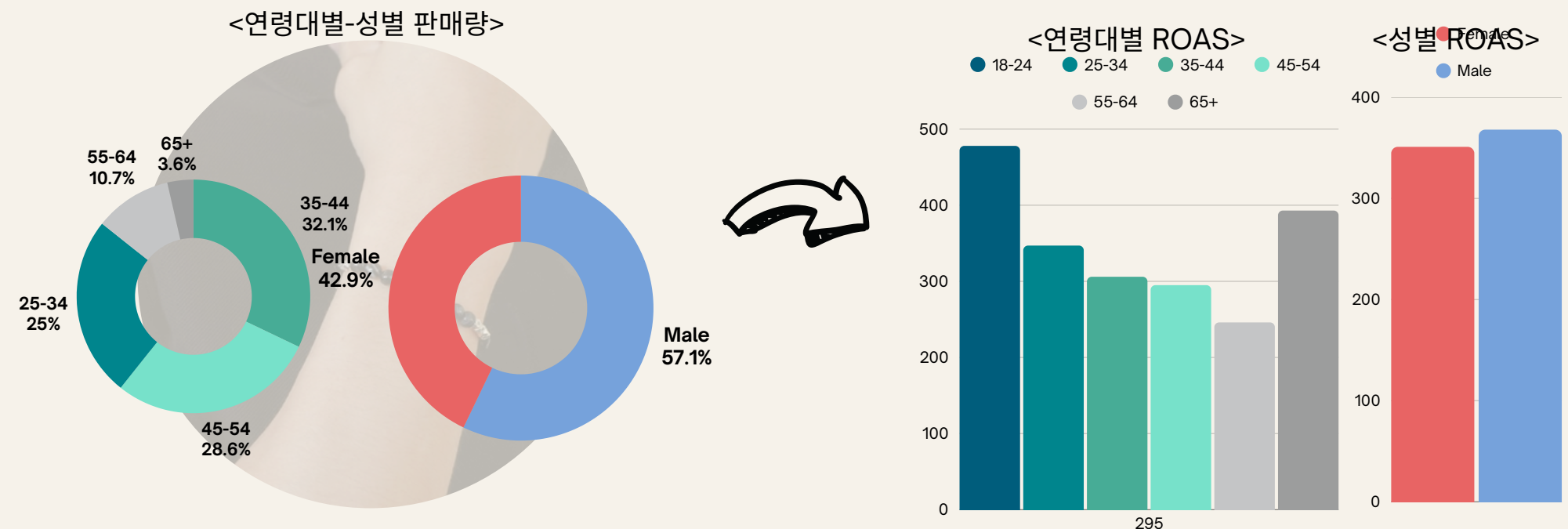


EDA

- 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

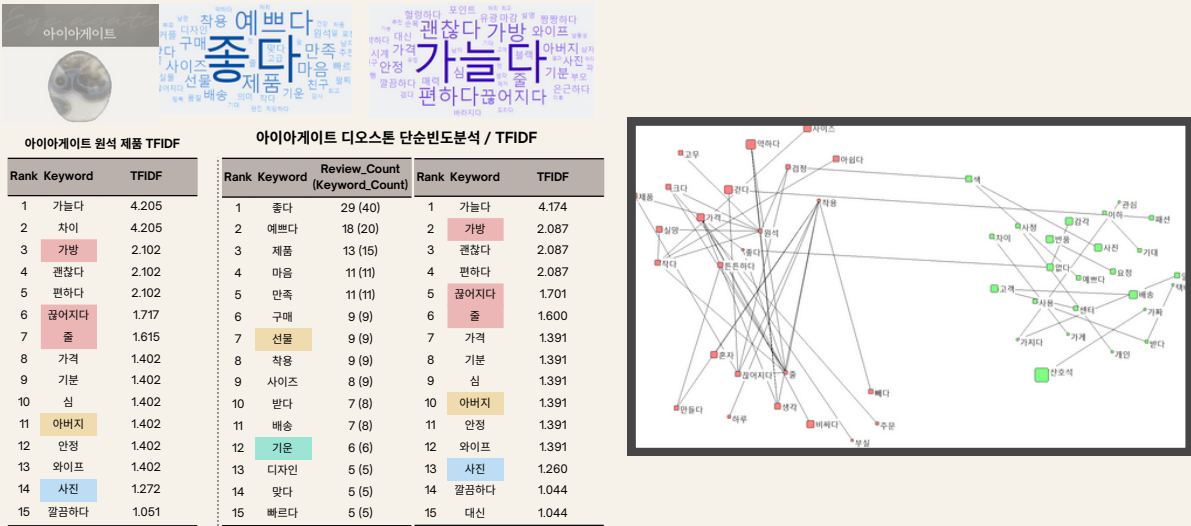
개념 설계 및 분석

데이터분석 및 시각화

- Text Analysis

TFDIDF 및 단순빈도분석: 선정된 후보 제품들
에대한 고객들의 소구점 파악

Keyword-Network 분석: 전체/부정 리뷰에서 확
인되는 부정적인 특징들이 후보 제품들에서는 나
타나지 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 니다. 그 내용을 연결해서 연령대별 마케팅 방안도 제시해주신 것이 좋았습니다. 수고 많으셨습니다.

Project 02

비도심 지역간 노인복지 불균형은 얼마나 심각한가?

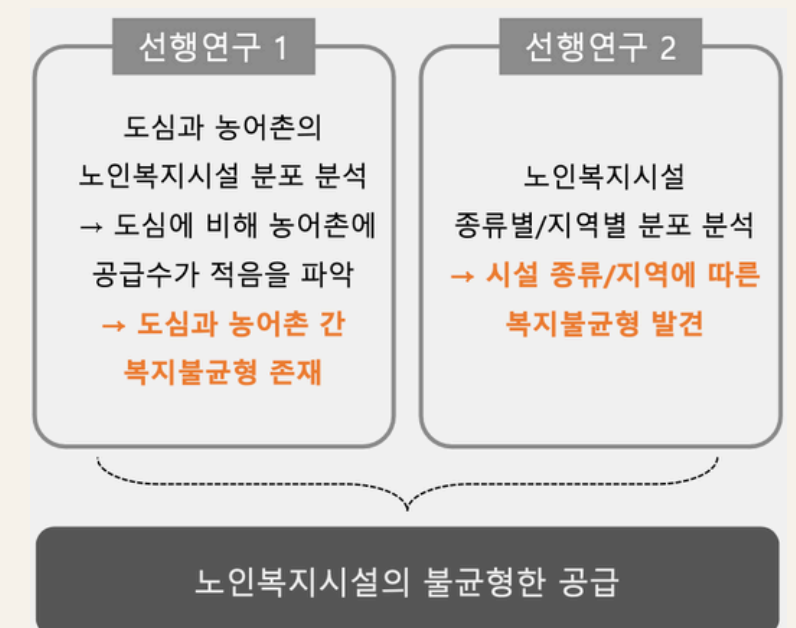
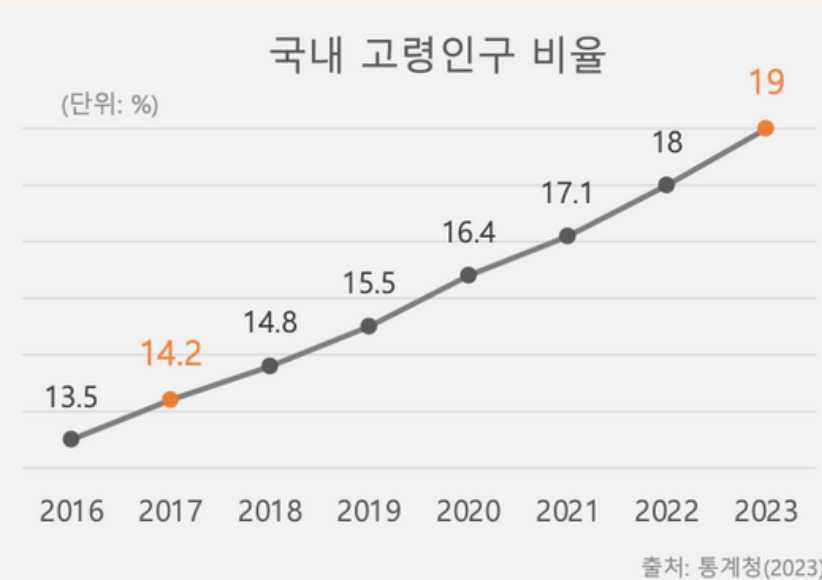
프로젝트 개요

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.
```

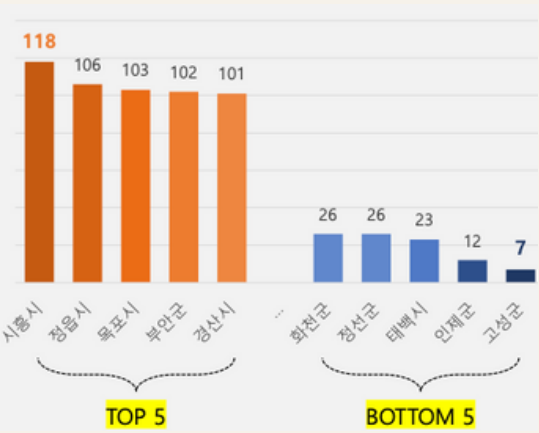
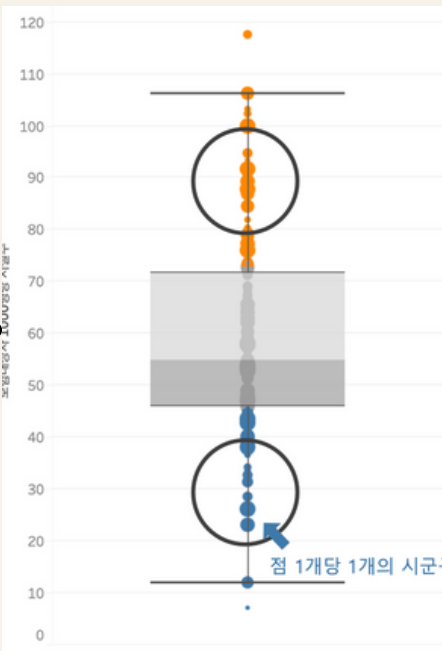
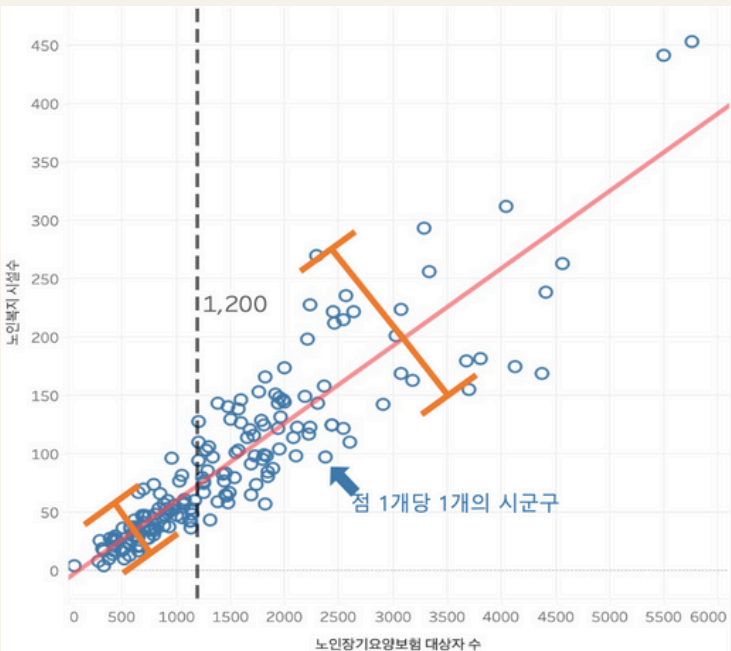
<R코드중 시군구별 노인장기요양보험 대상자 전처리 일부>

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



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 실재 수요대비 시설 공급의 불균형한 양상을 발견함
- 노인복지시설이 불균형한 지역 간에 어떤 차이가 나타나는가?: 노인복지자원 지표에서 불균형이 나타남. 3가지 지표 모두 TOP5지역에서 BOTTOM5 지역에 비해 높게 나타남

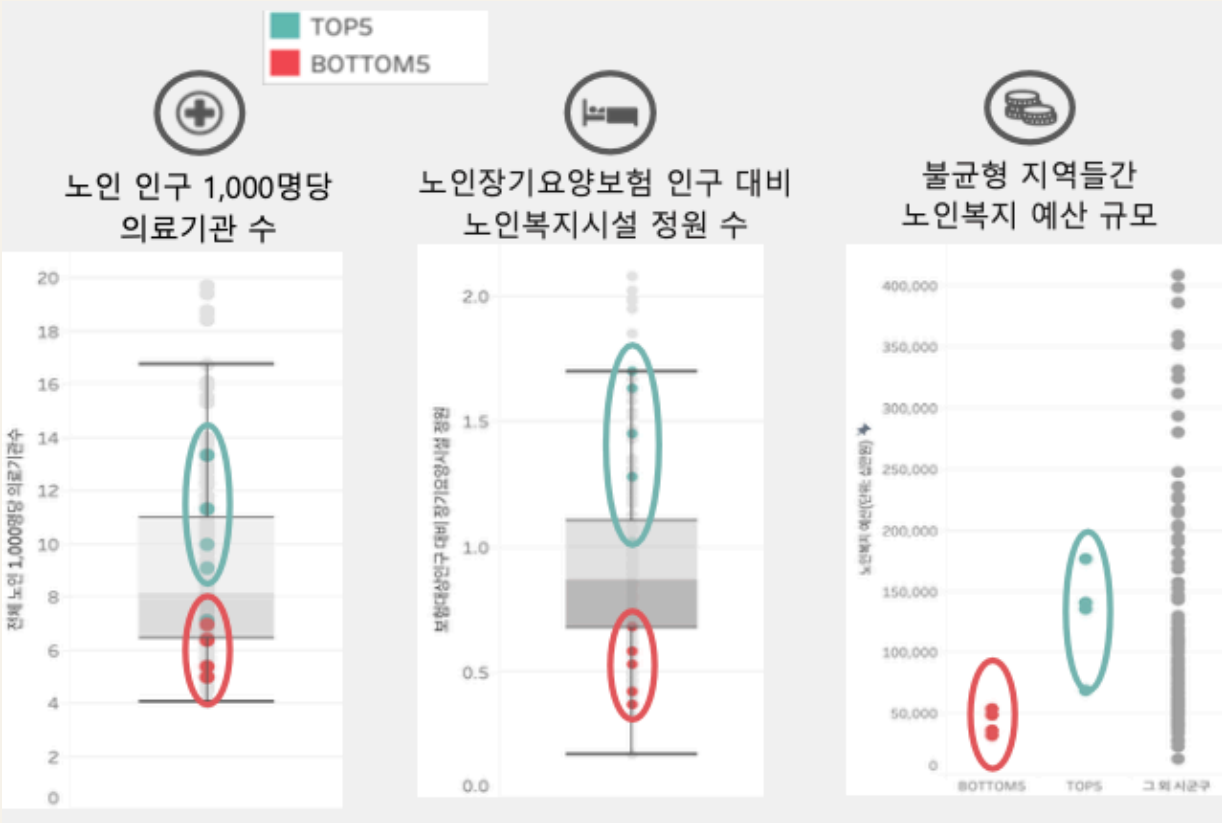
시사점

- 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



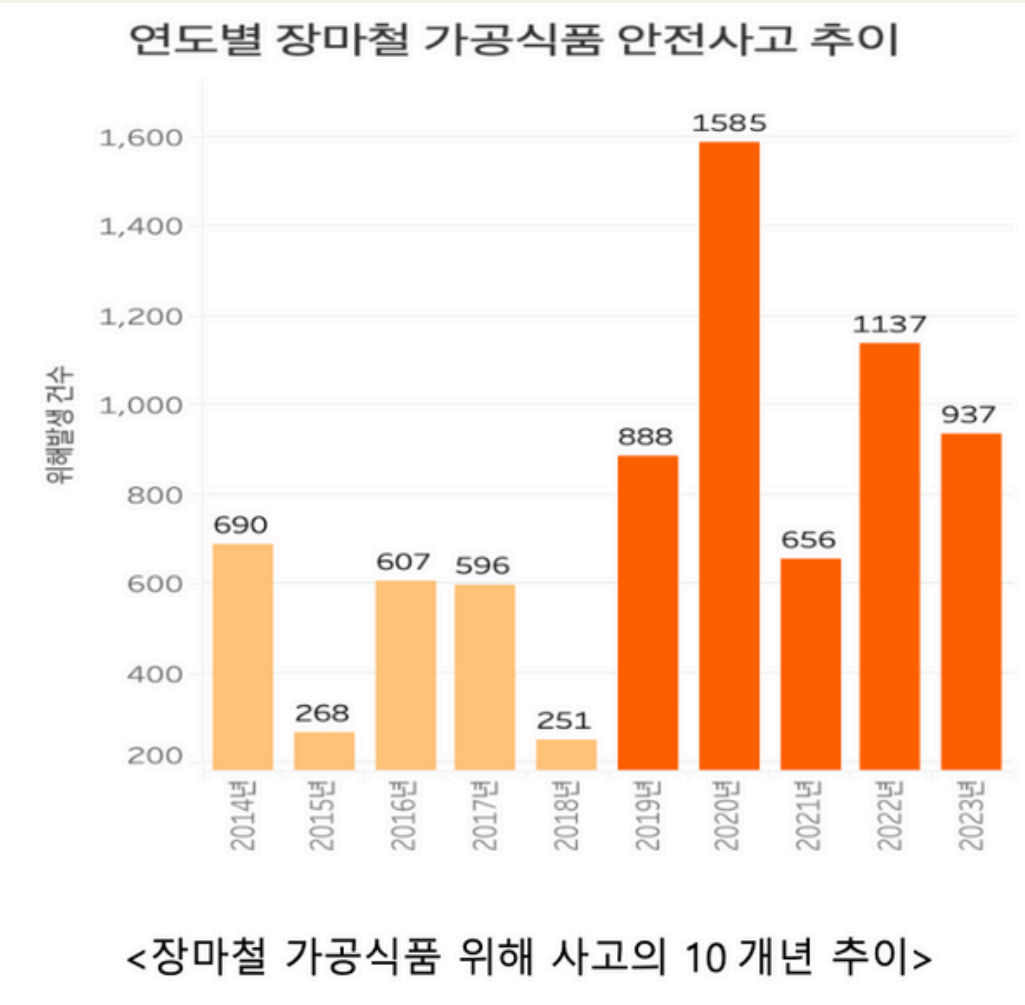
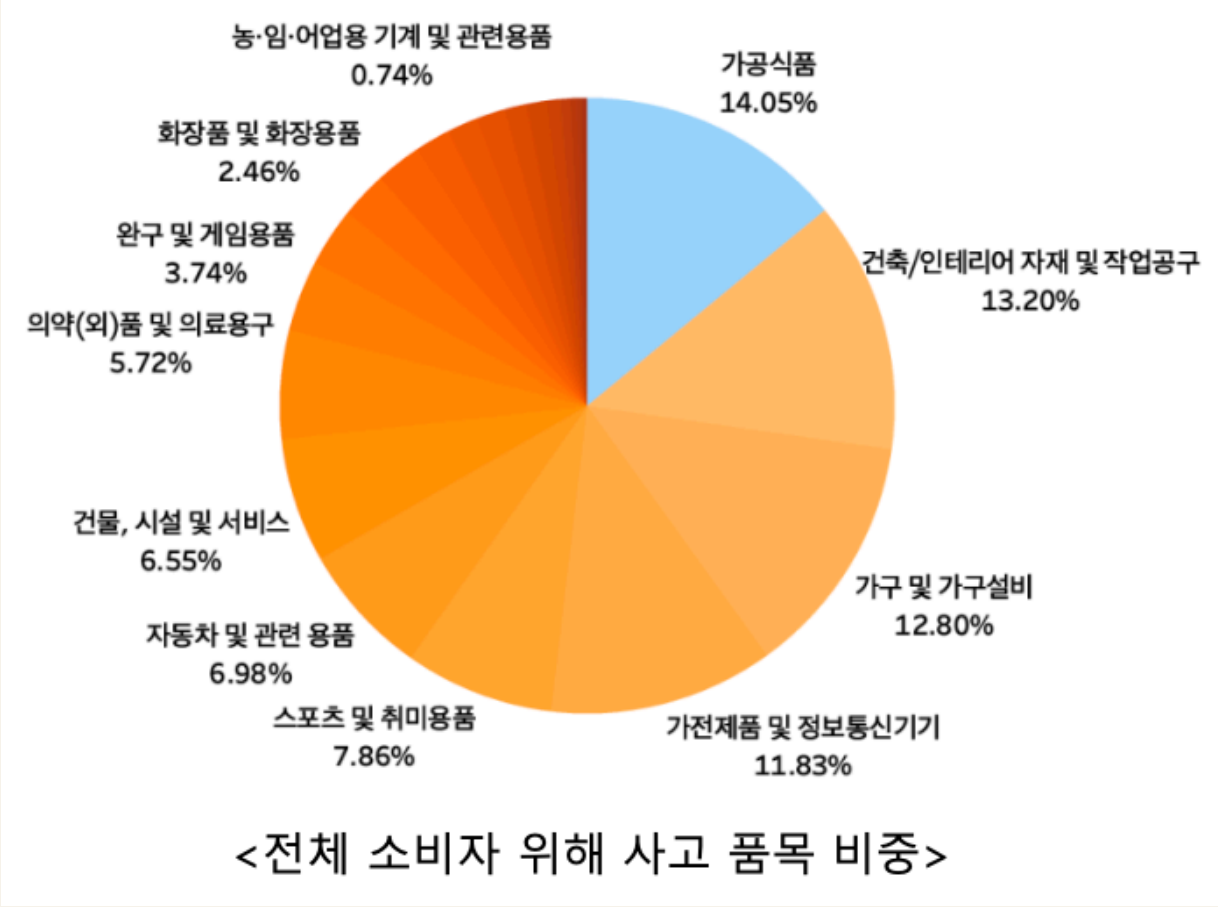
Project 03

장마철 가공식품 안전사고 예방 대책을 위한 종합분석

프로젝트 개요

분석 배경(Why)

- **식품안전사고의 높은 비중:**
2014~2023년 식품안전사고는 전체 사고 대비 가장 많은 비중을 차지(14.05%, 78,400건)
- **장마철 식품 안전사고(식중독) 추이의 급증:**
- 2014~2018 년 251~690 건(중양값 596, 평균 482) ->
2019~2024년656~1,585 건(중양값 937, 평균 1,040)으로 급증



Project 03

2024 위해정보 데이터 분석 보고서 공모전

장마철 가공식품 안전사고 예방 대책을 위한 종합분석

프로젝트 개요

분석 목표(Question)

- ‘장마’가 식품 안전사고에 미치는 영향의 규명 및 대응방안 제시
 - 10개년(2014~2023) ‘가공식품 위해 데이터’를 장마철/비장마철로 비교하여 EDA
 - ‘연관분석’ 기법을 활용-> 연령대/성별 위해 품목 분석
 - 기상 변수와 식중독 사이의 상관관계 분석

분석 방법(How)

- 데이터: 한국소비자원 자체 제공- 소비자 위해감시 시스템(CISS) 수집 데이터(2014~2023), 기상데이터(기상청)
- 사용 tool: Excel, Python, Tableau
- 연관분석, 단순 회귀 분석 활용

대분류	중분류	소분류*	비고
위해사항	신고자 인적사항	성명, 연락처	미제공(개인정보)
	위해자 인적사항	성명 연락처, 주소	미제공(개인정보)
		성별, 나이	제공(통계화 가능)
	위해관련사항	발생일	제공(통계화 가능)
		위험/위해원인	제공(통계화 가능)
		위해증상	제공(통계화 가능)
		위해부위	제공(통계화 가능)
		발생장소	제공(통계화 가능)
위해품목/ 업체	위해품목사항	품목명	제공(통계화 가능)
		업체명, 모델명	미제공(비공개 대상)

Project 03

2024 위해정보 데이터 분석 보고서 공모전

장마철 가공식품 안전사고 예방 대책을 위한 종합분석

개념 설계 및 분석

Data Preprocess

- ‘가공식품’ 소비자 위해사고 접수 67,836건(N/A 제외) 중 326일의 장마 기간 6,413건과 비장마기간 3326일의 61,423건 추출
- 2014~2023년 10개 기상변수 단위 조정 및 통합

N44									
	A	B	C	D	E	F	G	H	I
1	날짜	위해발생 건수	강수량(mm)	평균기온(°C)	최저기온(°C)	최고기온(°C)	평균습도(%rh)	최저습도(%rh)	일조시간(h)
2	2014.1.1	225	0	4.8	-0.3	9.2	51	11	7.7
3	2014.1.2	181	0	2.6	-2.2	9.1	49	10	8.4
4	2014.1.3	126	0	2.6	-3.2	10.1	54	12	4.9
5	2014.1.4	181	1.2	1.5	-2.8	6.9	63	12	6.8
6	2014.1.5	214	0.1	0.7	-3.7	6.6	62	9	7.1
7	2014.1.6	113	0	0.2	-5.1	7.6	59	12	8.4
8	2014.1.7	127	0	1.9	-4.3	7.8	60	16	3.4
9	2014.1.8	157	1.8	2.1	-1.4	5.3	65	25	1.3
10	2014.1.9	151	0	-4.4	-8.1	-0.6	34	9	8.9
11	2014.1.10	154	0	-3.7	-8.6	-0.2	56	10	7.0

EDA

- 장마/비장마 기간 식품 위해 사고 15개 품목 비교
- 장마/비장마 기간 ‘부패 및 변질’ 위해 원인 비교
- 기간별 ‘피부 및 피하조직 손상’ 위해증상 비교

데이터분석 및 시각화

- 연관 분석(Apriori 알고리즘) 결과
 - 장마기간 특정 성별 및 연령대에서 발생하는 식품 위해사고 품목
 - 빈번하게 발생하는 항목 집합 발견 -> 항목들간 연관 규칙 도출

성별	연령	주요 위해품목	주요 위해증상
남자	10대 미만	식육 조리식품, 우유류, 기타 조리식품	두드러기, 구토, 피부염 또는 피부발진
여자	10대 미만	-	두드러기, 구토, 피부염 또는 피부발진
남자	10대	치킨	소화기계통 장기손상 및 통증, 두드러기
여자	10대	-	소화기계통 장기손상 및 통증
남자	20대	회, 고기류, 수산물, 치킨, 초밥, 식육 조리식품	소화기계통 장기손상 및 통증, 두드러기
여자	20대	회, 고기류, 수산물, 치킨, 초밥	소화기계통 장기손상 및 통증
남자	30대	회, 고기류, 수산물, 탕류	소화기계통 장기손상 및 통증, 설사
여자	30대	회, 고기류	소화기계통 장기손상 및 통증, 설사
남자	40대	회, 수산물, 탕류	두드러기
여자	40대	회, 수산물, 탕류	-
남자	50대	회, 수산물, 탕류, 식육 조리식품	소화기계통 장기손상 및 통증, 두드러기, 설사
여자	50대	회, 수산물, 탕류	소화기계통 장기손상 및 통증, 두드러기
남자	60대	기타 영양보충제품, 탕류, 일반 다이어트 식품	-
여자	60대	기타 영양보충제품, 수산물	구토
남자	70대 이상	-	소화기계통 장기손상 및 통증
여자	70대 이상	탕류, 식육 조리식품	소화기계통 장기손상 및 통증, 구토

- >1 연령대가 높아질수록 영양 보충 제품 및 탕류에 의한 위해 증가
- >2 20~30대에서 치킨, 고기류, 초밥이 주요 위해품목
- >3 10대 미만에서 피부 관련 질환에서 민감하게 반응

개념 설계 및 분석

데이터분석 및 시각화

- 선형 상관분석
 - 복합적인 상호작용 및 노이즈가 많다는 기상 데이터의 본질적 특징 고려 = 상관계수가 낮게 나타남
 - 장마 여부에 따른 명확한 경향성 확인
- >1 기온은 모든 변수에서 장마 여부가 중요 요인
- >2 최저습도와 일사량 합에서 양의 상관관계
- >3 일조시간의 낮은 유의성

	평균기온	최저기온	최고기온	강수량	평균습도	최저습도	일조시간합	일사량합
장마	0.2003	0.1937	0.1674	-0.032	0.0677	0.1594	0.0737	0.1349
비장마	0.1511	0.1531	0.1459	0.0437	0.1152	0.1397	0.0003	0.1258

분석 결론(QnA)

- EDA 및 연관분석 기법을 활용한 연령대/성별 위해 품목 분석
 - 1 식품 섭취에 의한 위해는 장마와 상관없이 ‘소화기 계통의 질환’에 주의가 필요
 - 2 **장마 기간 10대 미만의 피부 질환에 각별한 주의**
 - 3 장마 기간 60대의 영양 보충 제품에 의한 위해 증가
- 기상변수와 식중독 사이의 상관관계 분석
 - 1 전 기간 식중독은 기온, 최저습도, 일사량의 영향을 받음
 - 2 장마기간 식중독은 기온 및 최저습도와 양의 상관관계를 보임
 - 3 강수량, 평균습도, 일조시간은 유의미한 상관관계 없음

대응방안

- 1 장마철 직전 10대이하 청소년 및 아동 대상, 알레르기 유발 혹은 피부에 자극적인 식품과 예방 수칙 교육
- 2 장마철 피부질환 중심 진료 및 상담 인프라 구축
- 3 기온 및 최저습도가 평년 대비 높은 여름철 식품 제조, 유통업체 및 외식업체에대한 관리 감독 강화
- 4 장마철 연령대별 주의 위해 식품 정보 제공 통해 경각심 높이기

Thank you