

텍스트마이닝을 이용한 출소자 대상 보호사업과 심리상담 연계 효과 분석

K-SIPI (Korean Social re-Integration Predictive Indicator)

팀장 : 박준수

팀원 : 김지은, 배성은, 서유진, 이인배

목차

01 서론

02 연구방법

03 연구 결과 및 시사점

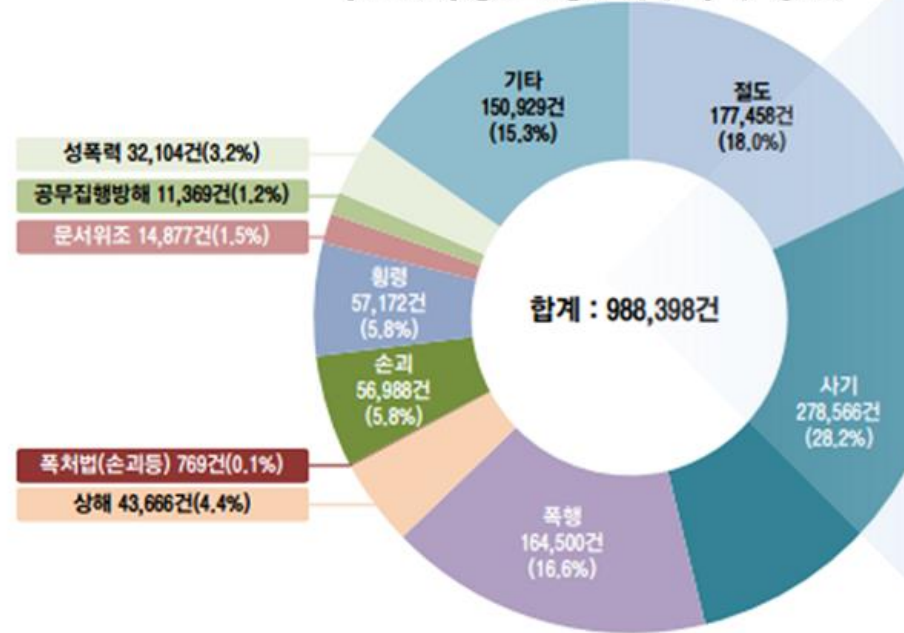
04 연구 결과 확산

01 서론

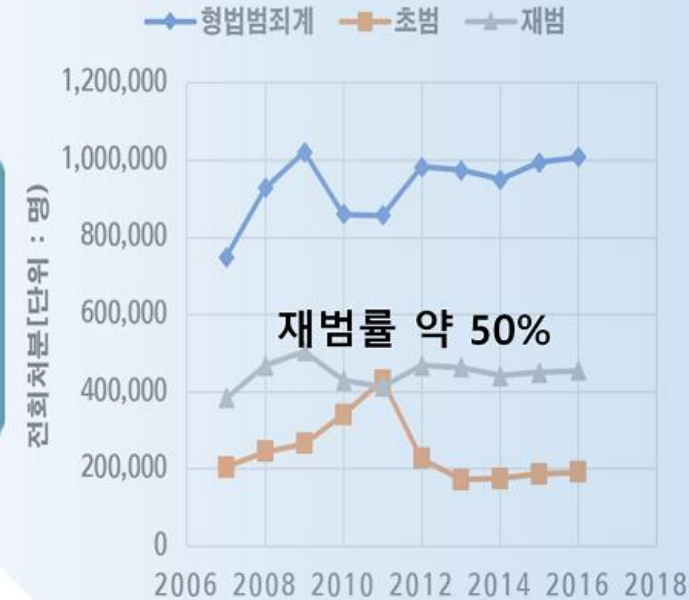
[1] 연구의 필요성

2018년도 전체 범죄발생건수: 1,738,190건 중

주요 죄명별 형법범죄 구성비



형법범죄자 재범 추이



출처: 법무부. (2019). 2018 범죄백서

※ 최근 10년(2008~2018)간 금고 이상의 형을 받은 수형자의 입소 경력별 인원 분석결과 **범죄경력자의 50% 내외는 재범으로 인해 교정시설에 재수용**

“다시 감옥 가고싶다”... 10년 복역 후 강도짓한 40대 남성



[인사이트] 김한솔 기자 = 감옥에 다시 보내달라며 편의점에서 강도짓을 한 남성이 경찰에 붙잡혔다.

15일 MBC는 한 남성이 경상북도 청송군 모처의 편의점에서 흥기를 들고 아르바이트생을 협박하는 모습이 담긴 CCTV 영상을 공개했다.

남성 A(44)씨는 지난 13일 새벽 1시 45분께 이 편의점에 들어와 한참을 서성이다가 아르바이트생이 계산대로 간 틈을 타 모자에서 흥기를 꺼내 금전을 요구했다.

편의점 점장에 의해 붙잡힌 A씨는 곧바로 경찰에 구속됐다.

때마침 편의점에 있던 점장은 A씨가 잠시 방심한 틈을 타 순식간에 A씨의 손을 낚아채 흥기를 떨어트리도록 유도해 제압했다.

그런데 제압 과정에서 A씨는 강력한 저항 의지나 도주 시도가 거의 없었다.

알고보니 그는 청송 출소 후 사회부적응을 겪던 재범한 죄수였다.

청송경찰서에 따르면 조사에서 A씨는 건강 악화와 사회부적응, 생활고 등으로 인해 다시 감옥에 가고 싶다는 취지로 범행 동기를 진술한 것으로 알려졌다.

범죄경력자의 사회부적응 사례 多



사회적응의 필요 : 범죄경력자가 출소 후 사회에 원만히 적응하여 재범하지 않는 것



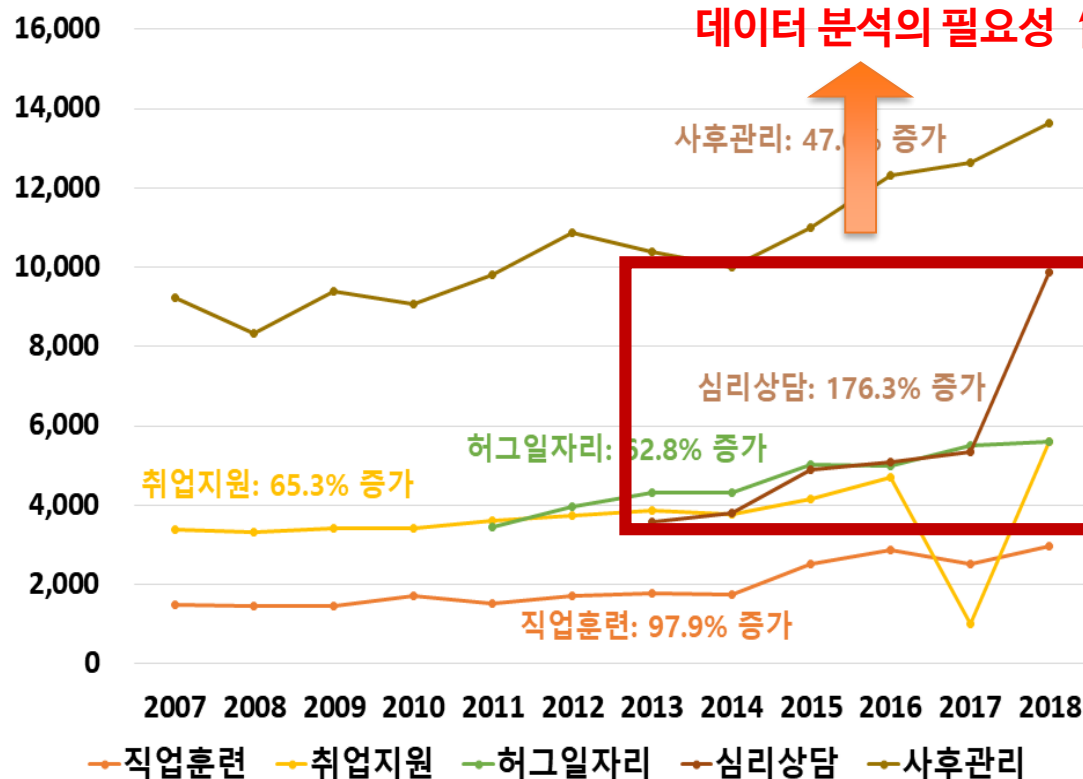
재범의 **심리적 요소와 욕구**를 파악,
재범방지에 도움 줄 수 있는 방안을 모색할
필요가 있음

출소 후 초기정착 및 사회적응을 위한 국가적 지원기관 :



한국법무보호복지공단
KOREA REHABILITATION AGENCY

주요 보호사업별 건수



매년 1만명 이상의 출소자에 대한
심리상담 데이터 축적

1인당 평균 10번 상담 -> 10만건

10년 이상 데이터축적 -> 100만건

빅데이터화

보호기간별 심리상담 분석을 통한
보호사업의 **질적 효과성** 측정 가능

출소 후 초기정착 및 사회적응을 위한 국가적 지원기관:



숙식제공 분석 이유

17.4%

/ 서비스 목적

안정된 숙식제공 서비스로 취업 및 자립을 위한 기반 마련
(생활관자치운영제도를 도입한 자율형생활관 운영, 단, 청소년시설은 자치운영제 유예)

숙식제공

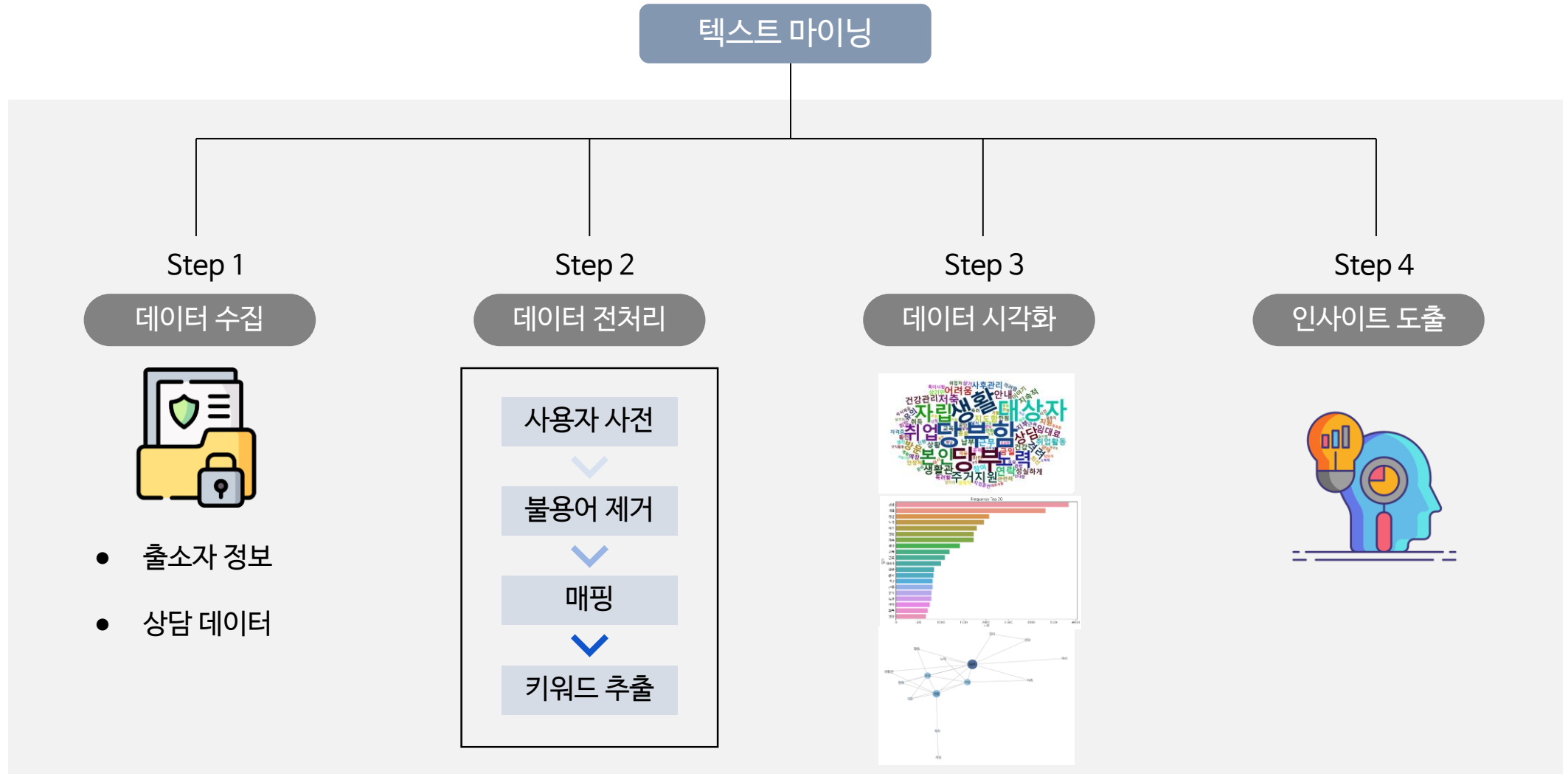
주거지원

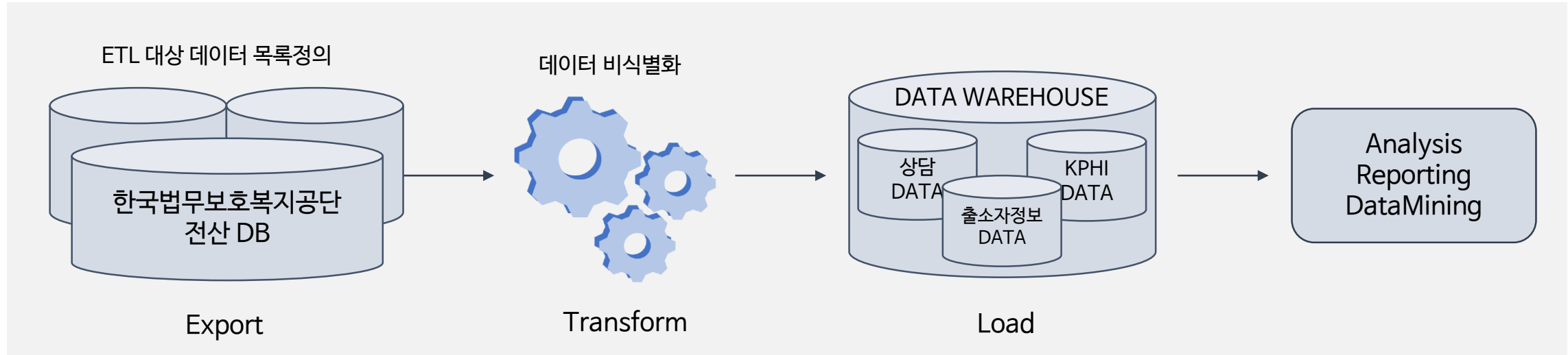
전체

■ 초범 ■ 재범

02 연구방법

- [1] 데이터 수집
- [2] 데이터 전처리 과정
- [3] 데이터 시각화
- [4] 인사이트 도출





1) 출소자 정보 데이터

보호카드번호	개인식별코드	상담날짜	보호사업코	출생년도	연령	학력	혼인	죄명	석방사유	범죄횟수
100020110268	bAZUzQi/863H/CLyucxEYg==	20120619	20	1977	30	1	NULL	80	30	1
100020120175	WD8sGuuBdIRVGVW5Z0coUw	20130404	20	1970	40	1	NULL	990	990	1
100020120191	VMIHtU0pmbElq7cFCvq6uw==	20130217	20	1969	40	1	NULL	10	30	2
100020120191	VMIHtU0pmbElq7cFCvq6uw==	20130307	20	1969	40	1	NULL	10	30	2
100020120191	VMIHtU0pmbElq7cFCvq6uw==	20130311	20	1969	40	1	NULL	10	30	2
100020120224	rb4DYhFwPSEloyy8StxbLg==	20130320	20	1961	50	1	NULL	40	10	3

2) 상담데이터

주민번호(비식별)	보호카드번호	상담일자	보호사업코드	보호사업명	상담구분	상담구분명	상담내용
///tqlgHDzalE9DI/BBYJQ==	12020160195	20160808	40	직업훈련	10	멘토링	허그일자리 프로그램 참여자로 대형운전면허 자격증 취득을 위해 학원에 입교조치하고 조기 자격증 취득하여
///tqlgHDzalE9DI/BBYJQ==	12020160195	20160830	40	직업훈련	10	멘토링	대형면허 자격증을 합격하였으며, 취득 후 솔로몬방재엔지니어링(주),소방설비 및 운전직으로 취업하였는바, 장
///tqlgHDzalE9DI/BBYJQ==	12020160195	20161031	20	취업지원	20	사후관리	다니던 회사를 그만 두었다고 함. 재취업을 위해서 노력해 볼 것을 당부하고 취업담당 직원과 통화해 볼 것을
///tqlgHDzalE9DI/BBYJQ==	12020160195	20161130	20	취업지원	20	사후관리	-상기인은 솔로몬소방방재 엔지니어링에서 근무하였으나 개인적인 사정으로 인하여 퇴사하였으며 사후관리

한글 형태소 분석기 - Python Konlpy

문장 예시

doc = '회사 판단하에 정식으로 채용되지 않아 상심이 큼, 최근애로사항에 대해 상담함.'

Hannanum

[('회사', 'N'), ('판단하', 'N'), ('에', 'J'), ('정식직원', 'N'), ('으로', 'J'), ('최근애로사항', 'N'), ('에', 'J'), ('대', 'P'), ('어', 'E'), ('하', 'P'), ('어', 'E')]

Kkma

[('회사', 'NNG'), ('판', 'NNG'), ('단하', 'NNG'), ('에', 'JKM'), ('정식', 'NNG'), ('조', 'NNG')]

Komoran

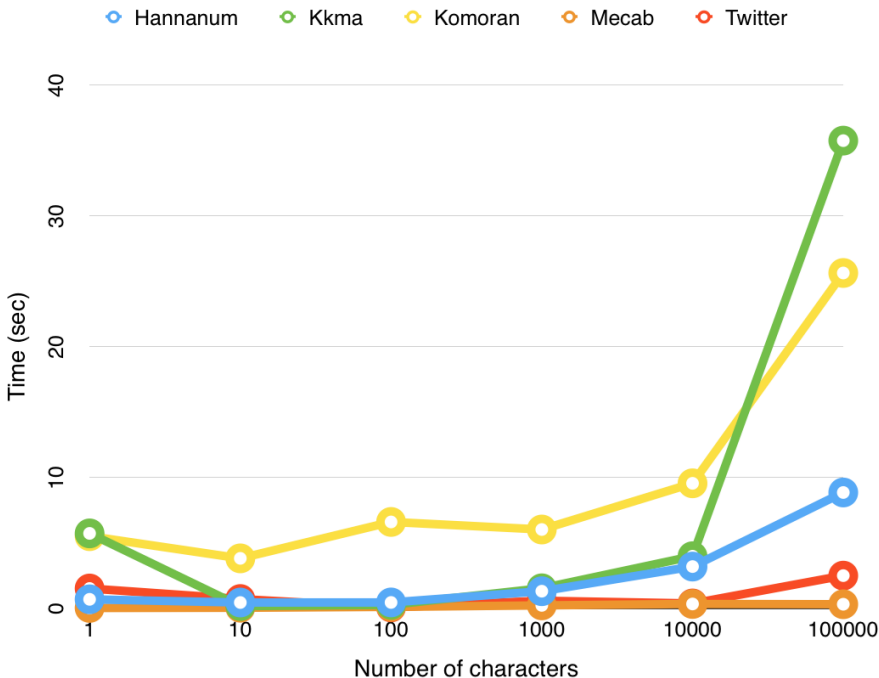
[('회사', 'NNG'), ('판단', 'NNG'), ('하', 'NNG'), ('에', 'JKB'), ('정식', 'NNG'), ('직원', 'NNG')]

Mecab

[('회사', 'NNG'), ('판', 'NNG'), ('단하', 'NNG'), ('에', 'JKB'), ('정식', 'NNG'), ('직원', 'NNG'), ('최근', 'NNG'), ('애', 'NNG'), ('로', 'JKB'), ('사항', 'NNG'), ('에', 'JKB'), ('대해', 'VV+EC')]

Twitter

[('회사', 'Noun'), ('판단', 'Noun'), ('하', 'Suffix'), ('에', 'Josa'), ('정식', 'Noun'), ('직원', 'Noun'), ('최근', 'Noun'), ('애로', 'Noun'), ('사항', 'Noun'), ('에', 'Josa'), ('대해', 'Noun'), ('상담', 'Noun')]



1. 로딩시간 : Hannanum : 0.66 secs

Kkma : 5.70 secs

Komoran : 5.49 secs

Mecab : 0.01 secs

Twitter : 1.49 secs

2. 실행시간 : Hannanum : 8.83 secs

Kkma : 35.72 secs

Komoran : 25.61 secs

Mecab : 0.28 secs

Twitter : 2.47 secs

사용자사전

고유명사 등의 단어를 사용자가 색인어로 추출하고 싶을 때 직접 추가하여 색인할 수 있는 기능이다.

```
from konlpy.tag import Twitter
twitter = Twitter()

text = '허그일자리프로그램'
token = twitter.pos(text)
print(token)
```

[('허그', 'Noun'),
(('일자리', 'Noun'),
(('프로그램', 'Noun'))]

〈적용 전〉



```
from konlpy.tag import Twitter
twitter = Twitter()
twitter.add_dictionary(['허그일자리프로그램'], 'Noun')

text = '허그일자리프로그램'
token = twitter.pos(text)
print(token)
```

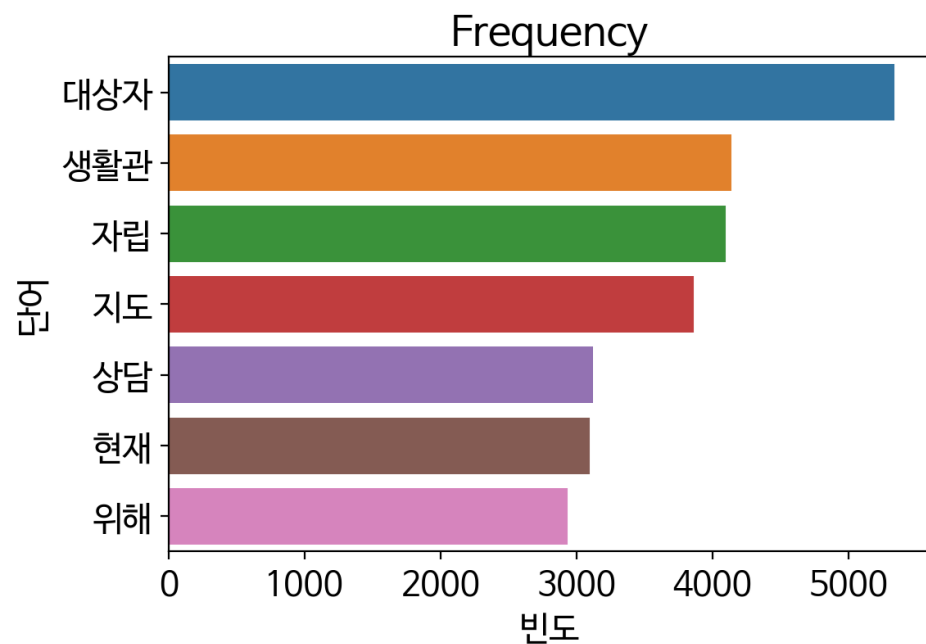
[('허그일자리프로그램', 'Noun')]

〈적용 후〉

불용어 : 자주 등장하지만 데이터를 분석하는데 있어 큰 의미를 갖지 않는 단어

데이터에서 유의미한 단어 토큰만을 선별하기 위해서 큰 의미가 없는 단어 토큰을 제거하는 작업

[총 655개 불용어 지정]



#날짜 계정 등 시기를 나타내는 토큰

'개월', '금일', '기간', '겨울', '년도', '날짜', '금월', '까지', '날씨',
'다음주', '평소', '차후', '지난', '요즘', '아침', '오전', '오후', '예전', '이제'

#지역 위치나타내는 토큰

'경기도', '김해', '서울', '울산', '천안'

#관련없는 토큰

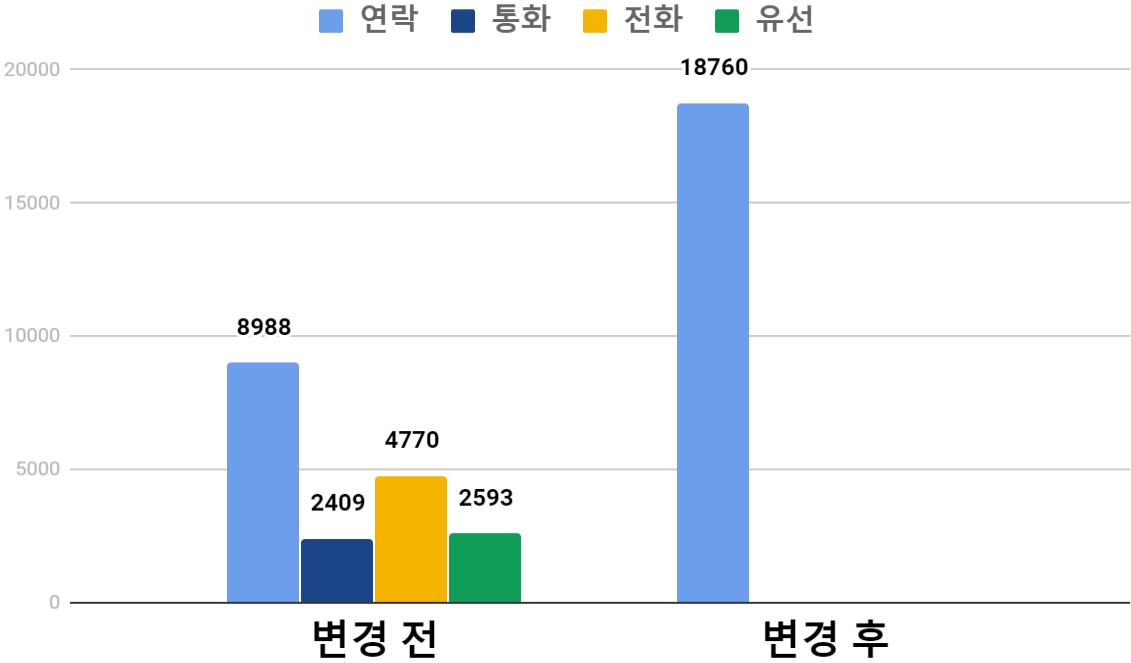
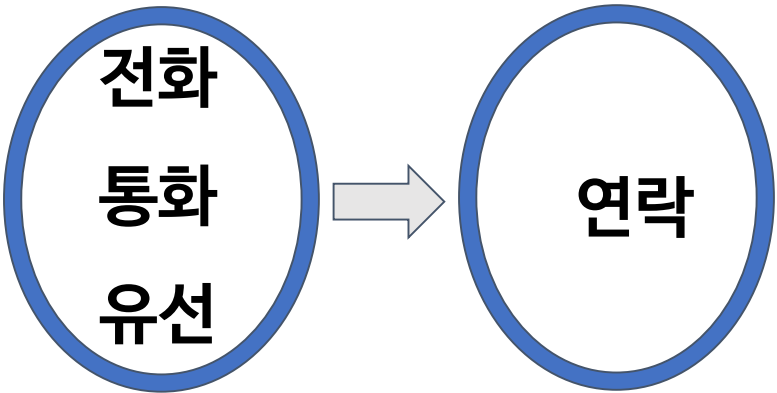
'격려', '결과', '경우', '계속', '계획', '공단', '과정', '관련', '곤
'대상자', '독려', '마련', '멘토링', '문자', '문제', '발생', '방문',
'아직', '여건', '연락', '예정', '외박', '외출', '요청', '유의', '이
'주거지', '주거지원', '준비', '지도', '지원', '지지', '직원', '진행',
'보호종료', '개시', '동일', '파악', '재차', '사용', '주기', '연결', '마무리',
'해당', '해지', '인계', '각종', '번호', '취지', '참고', '보도', '약간', '디

#쓸모없는 토큰

'가지', '고자', '고함', '구들', '금은', '기반', '기하', '동안', '되어',
'만원', '만큼', '면서', '모습', '문하', '바로', '보이', '보임',
'인의', '인하', '있는', '있다고', '있도록', '있어', '있으며', '있음

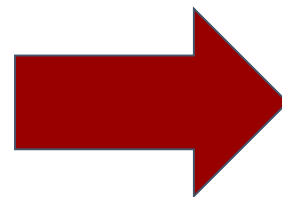
매핑(Mapping)

매핑: 의미가 같거나 비슷한 의미를 가진 단어를 하나의 단어로 통일하는 작업

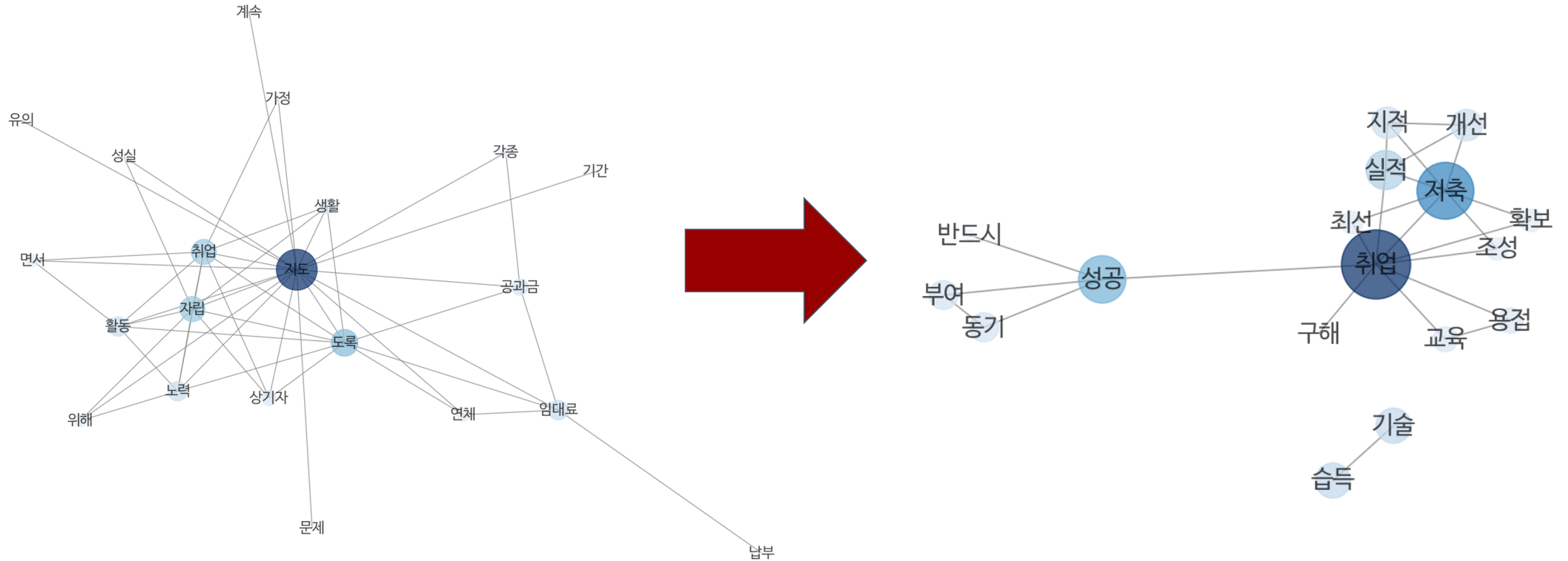


문장	추출된 단어(명사)
전화로 통화하여 연락	before : 전화, 통화, 연락
	after : 연락

-



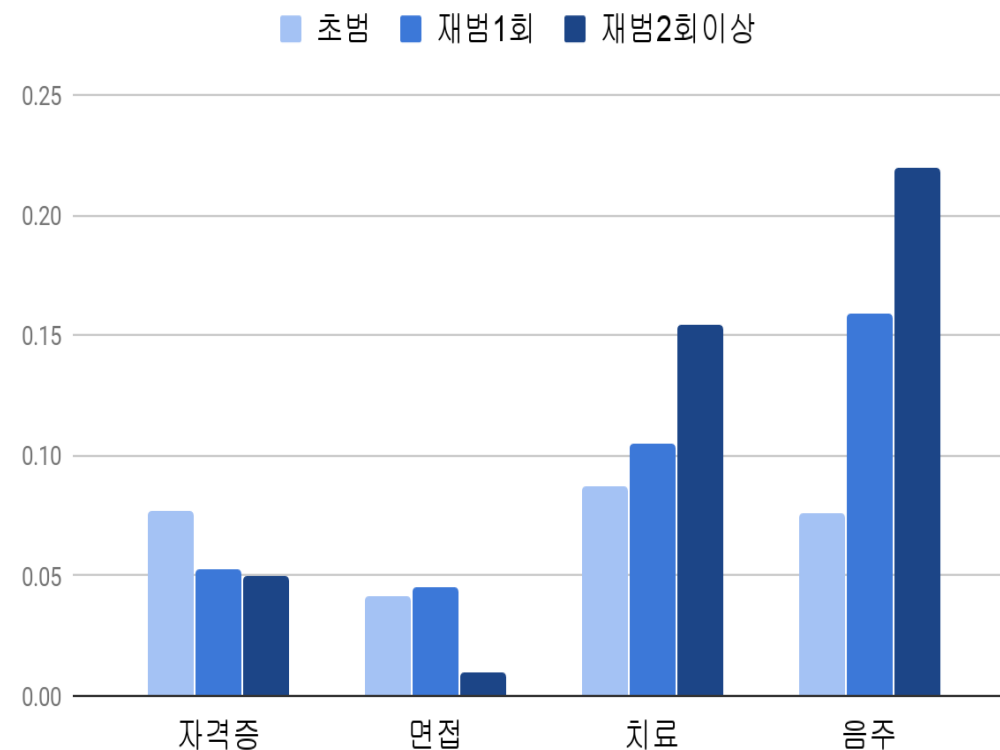
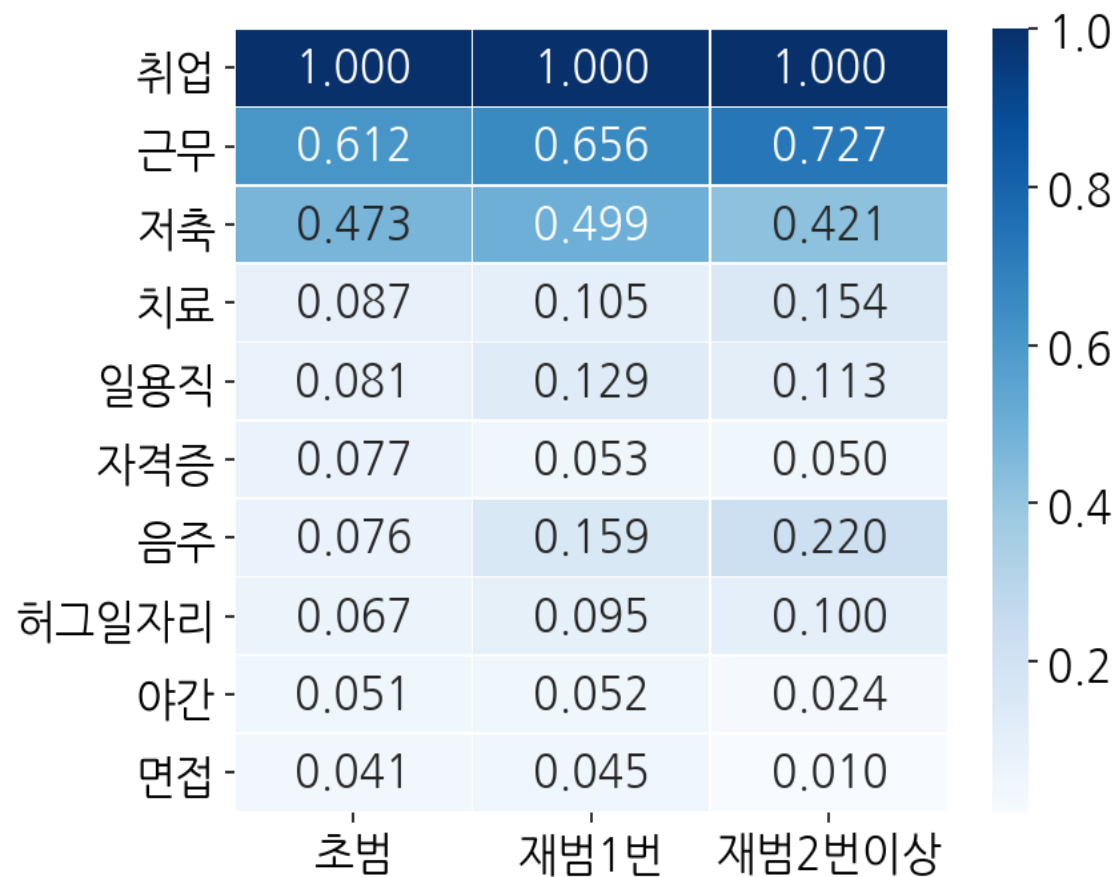
• 전처리 후 숙식 데이터 네트워크 변화



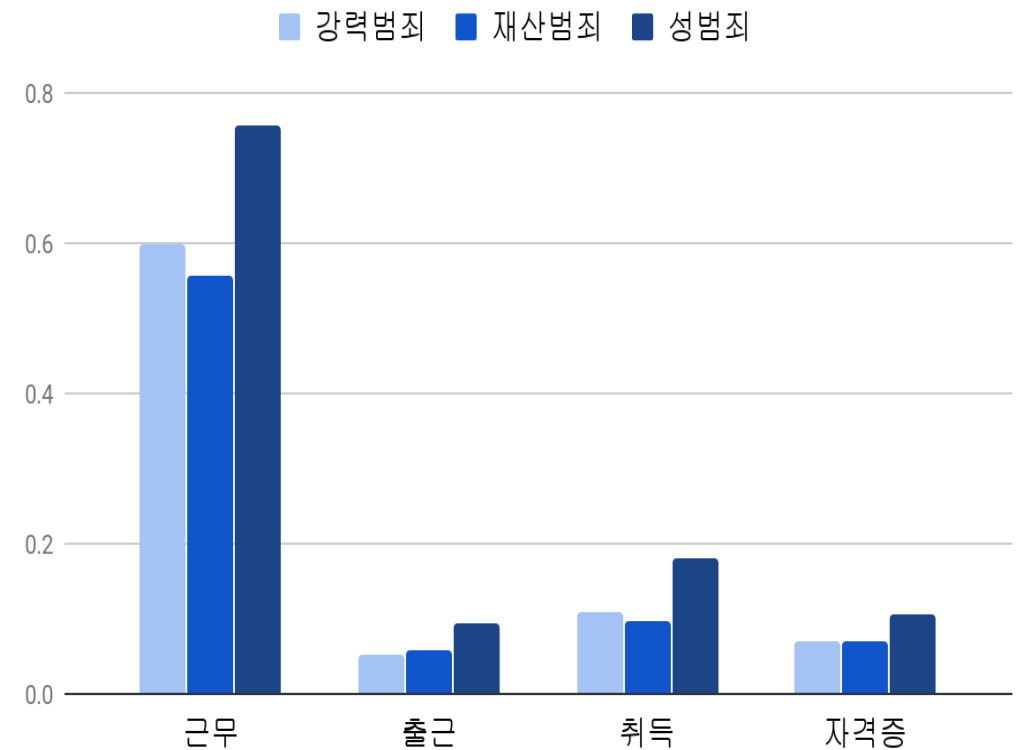
-

활동	인원
자립	2.4
취업	1.7
노력	1.5
저축	1.5
숙식제공	1.1
안내	1.0
취업활동	0.9
어려움	0.9
건강관리	0.9
건강	0.8

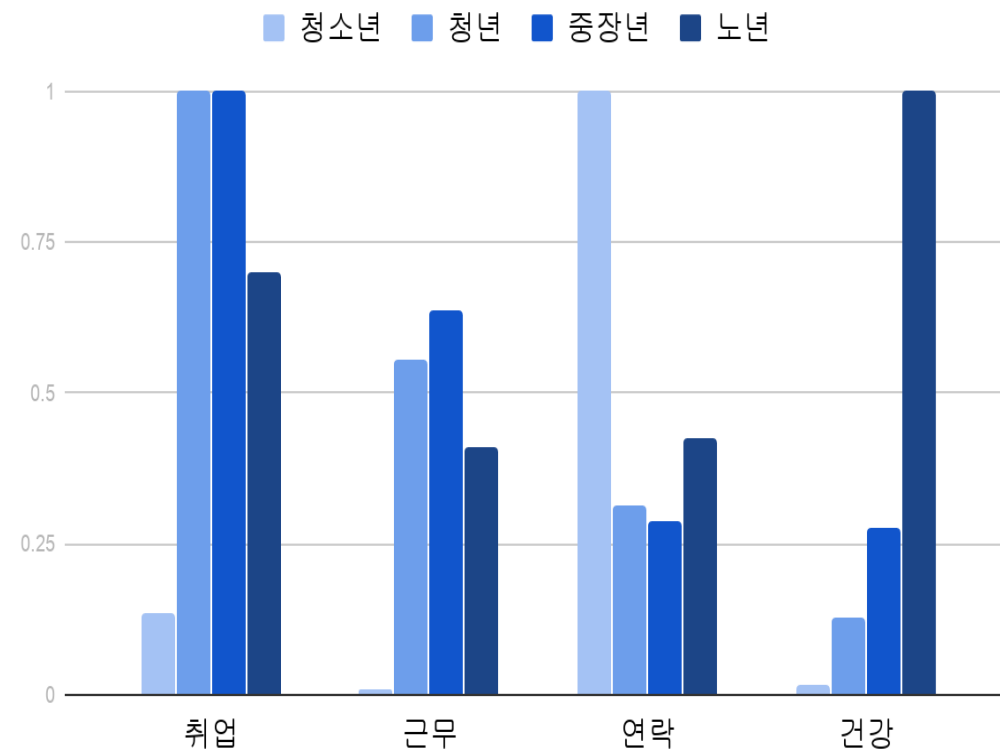
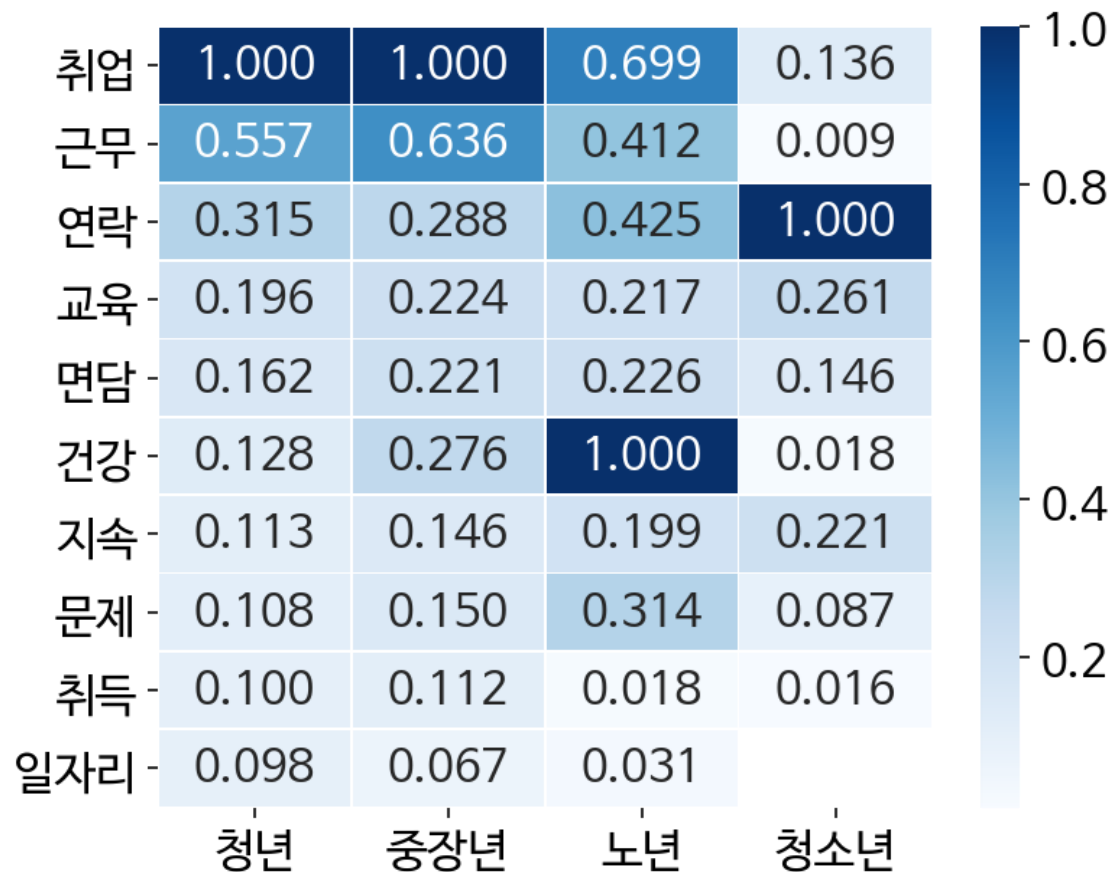
재범 횟수에 따라 키워드 분석



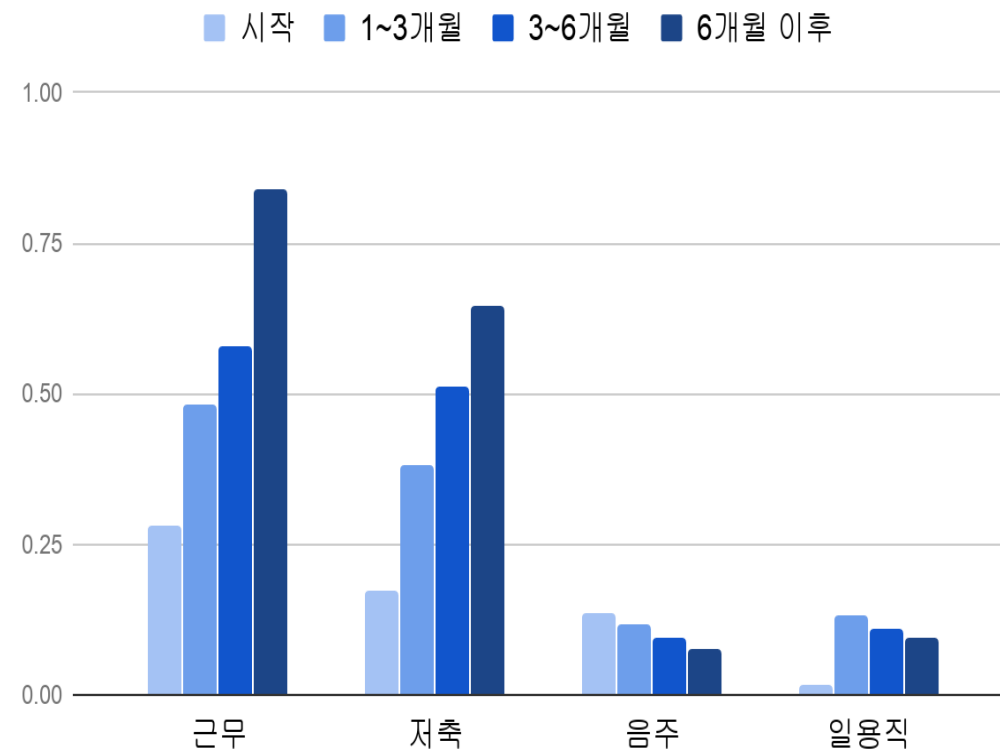
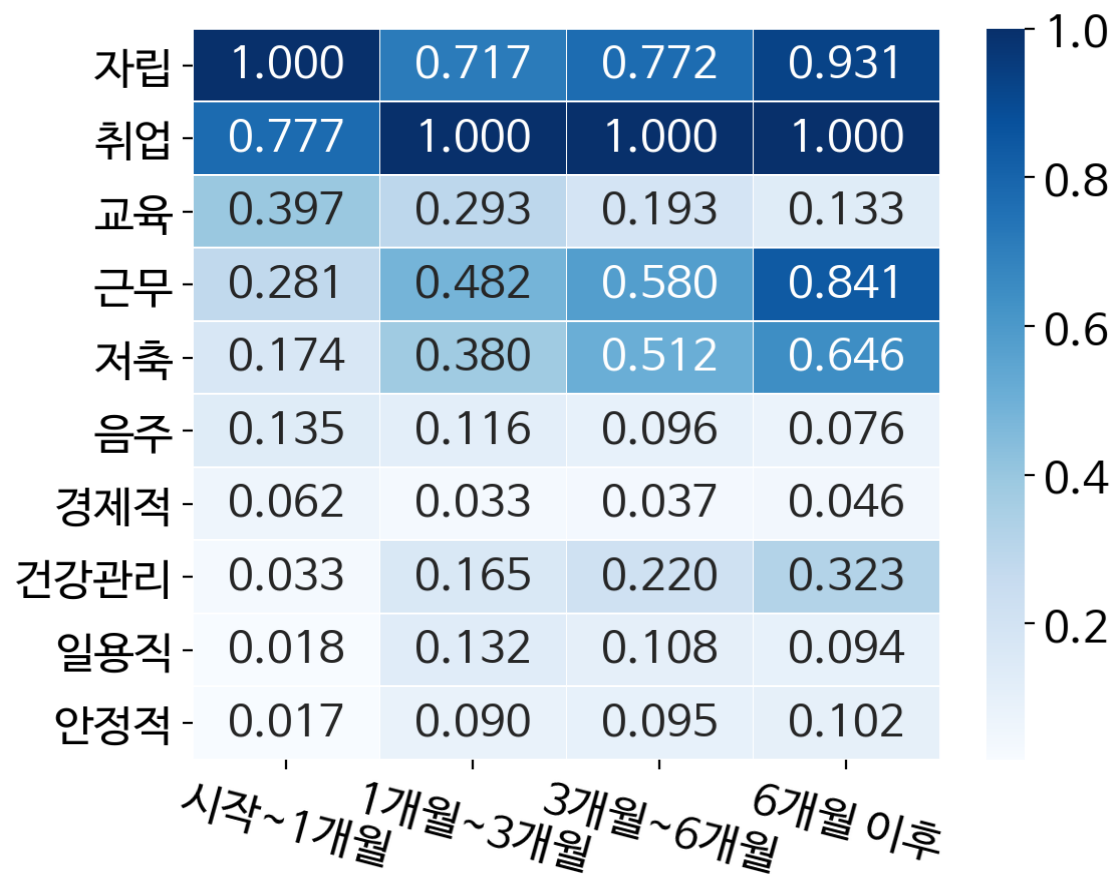
· 죄명에 따라 키워드 분석



• 연령에 따라 키워드 분석

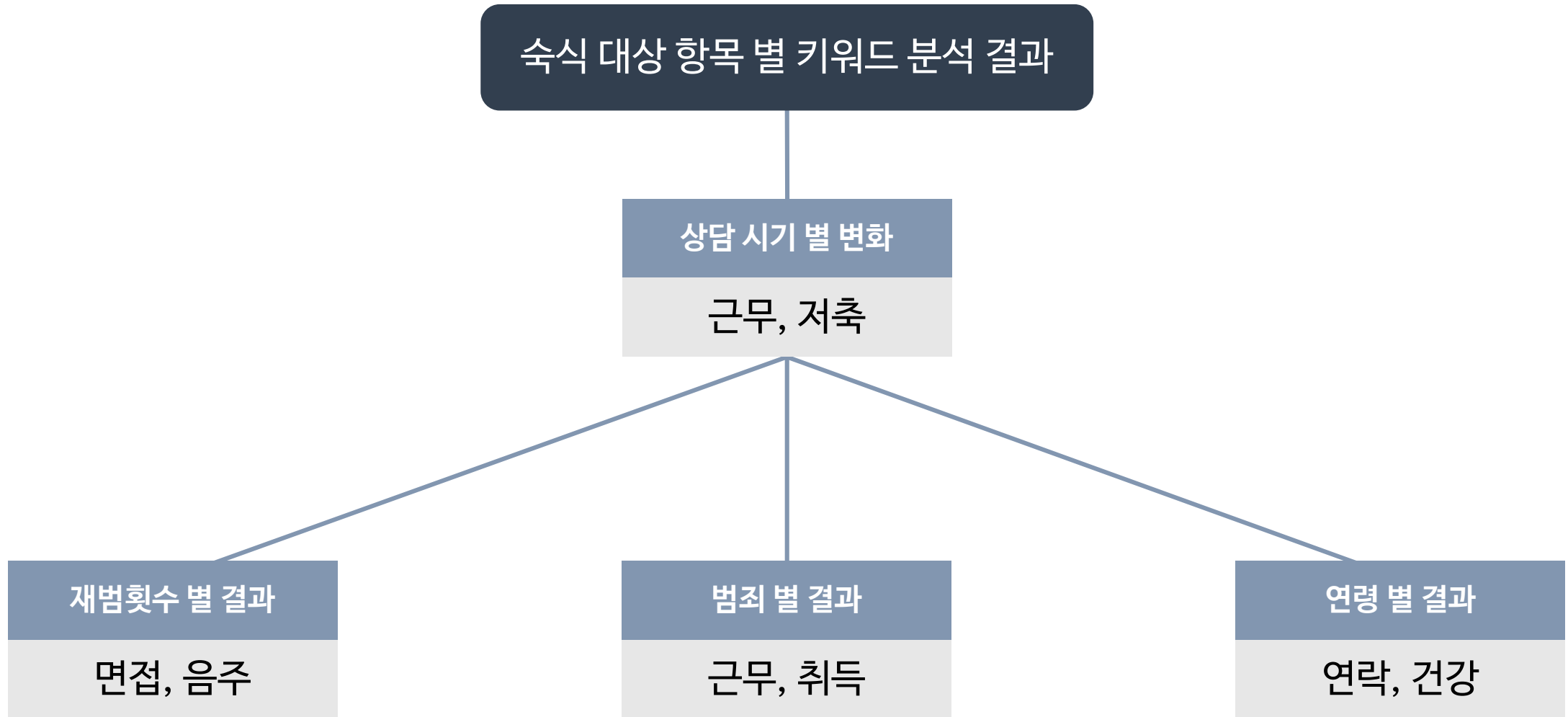


• 숙식제공 시계열 키워드 분석



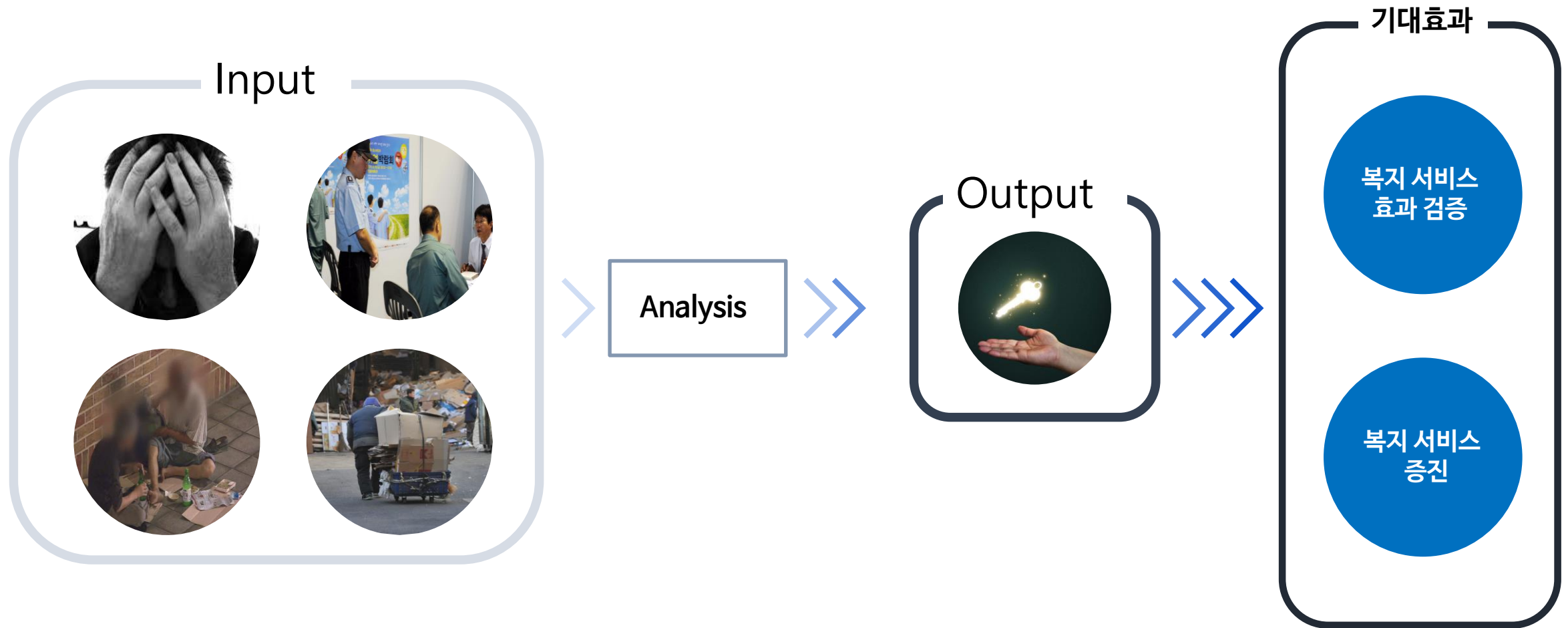
03 연구 결과 및 시사점

[1] 텍스트 마이닝 기법을 활용한 심리상담 분석 결과



04 연구결과 확산

[1] 연구결과 확산



감사합니다.

Appendix

가족접견지원



가족친화지원



기타지원



숙식지원



원호지원



주거지원



직업훈련



창업지원



취업지원



학업지원



허그일자리

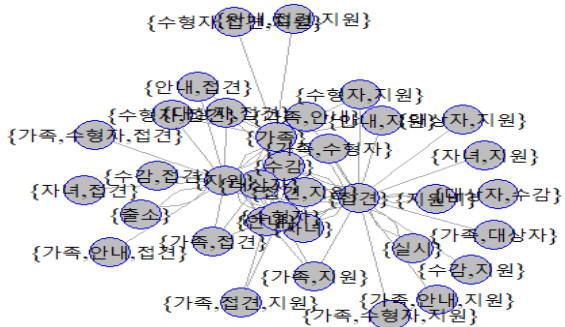


심리상담



Appendix

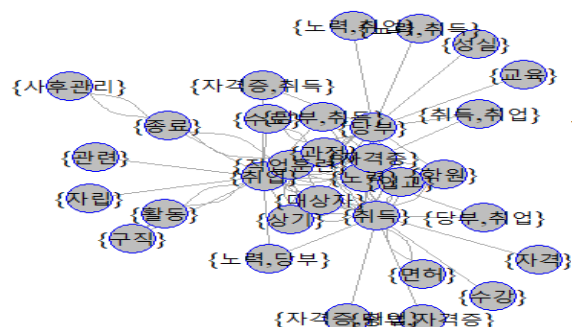
가족접견지원



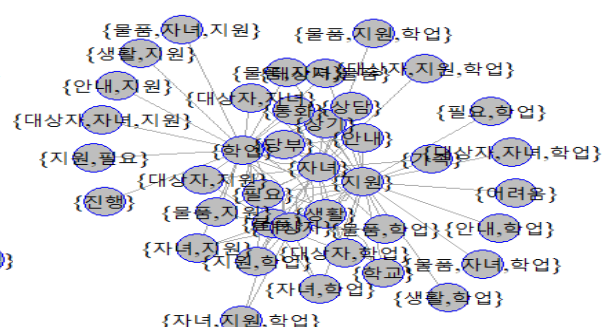
숙식지원



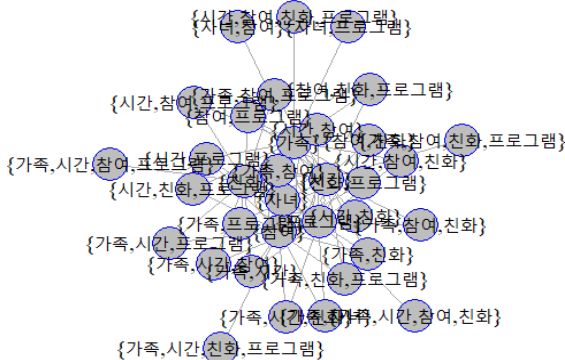
직업훈련



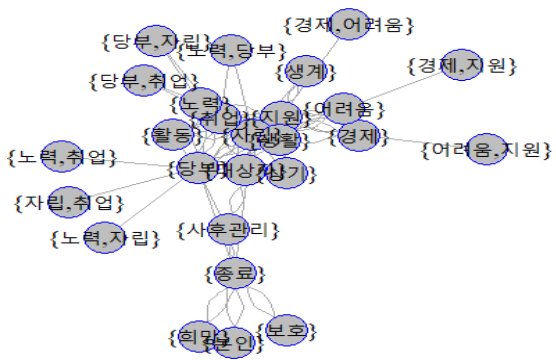
학업지원



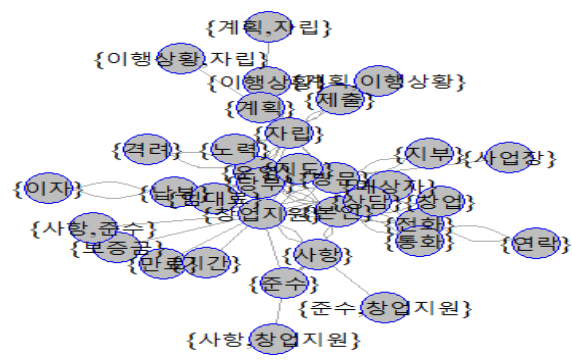
가족친화지원



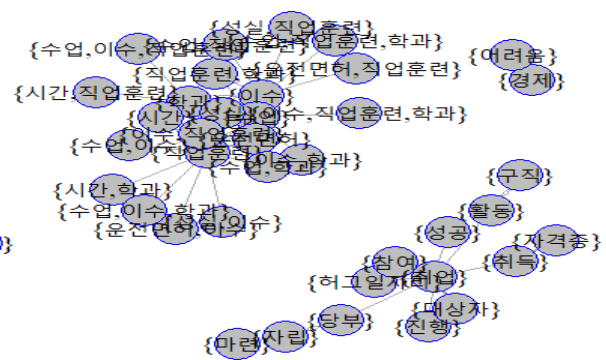
원호지원



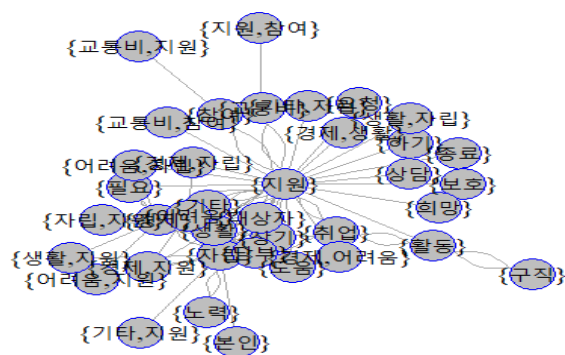
창업지원



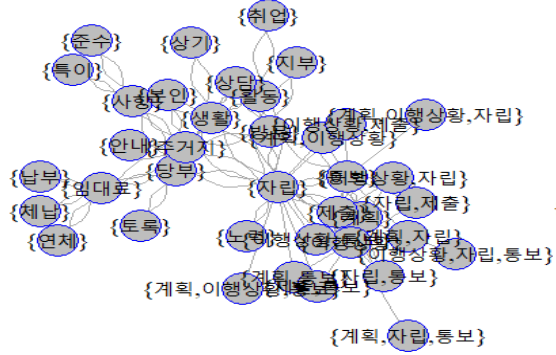
허그일자리



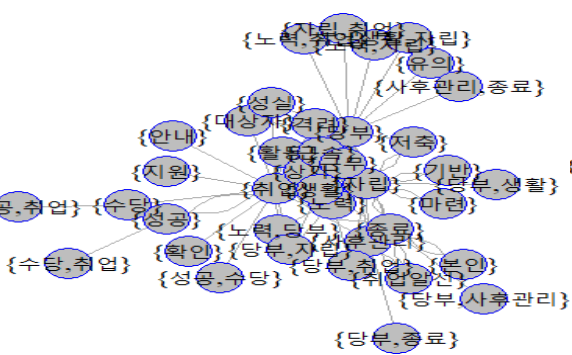
기타지원



주거지원



취업지원



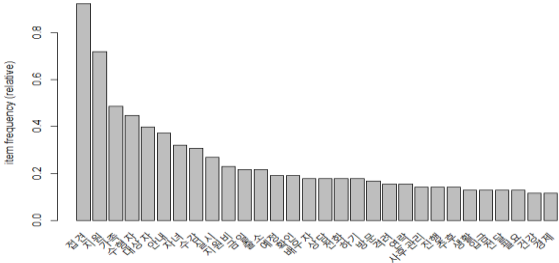
심리상담



Appendix

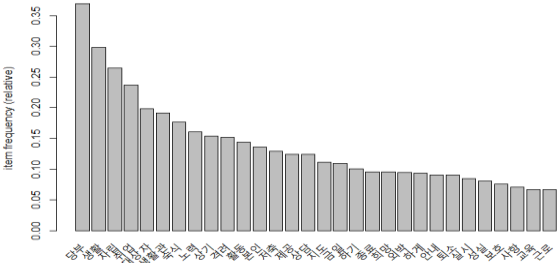
가족접견지원

가족접견지원



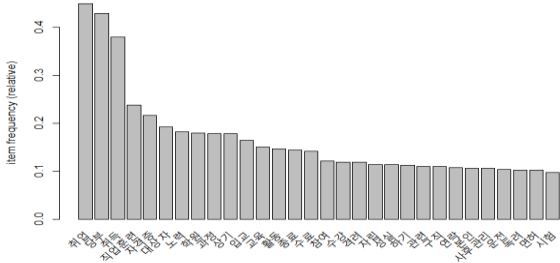
숙식지원

숙식제공



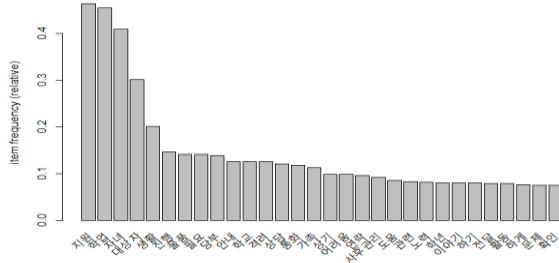
직업훈련

직업훈련



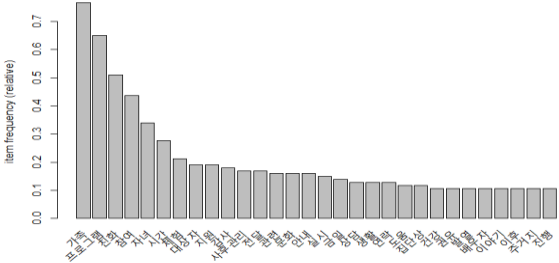
학업지원

학업지원



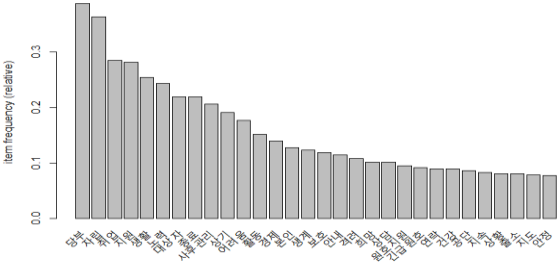
가족친화지원

가족친화지원



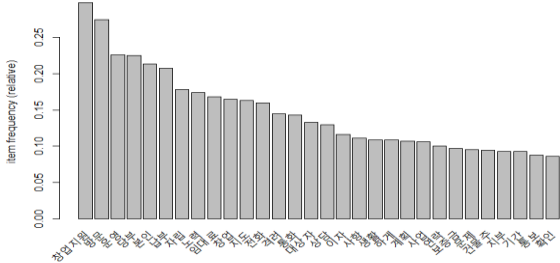
원호지원

원호지원



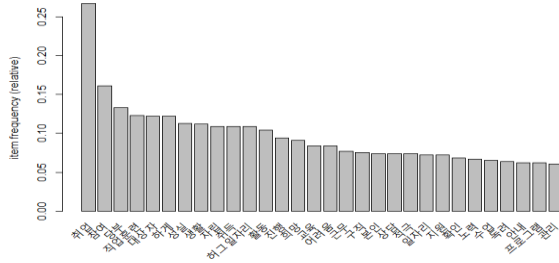
창업지원

창업지원



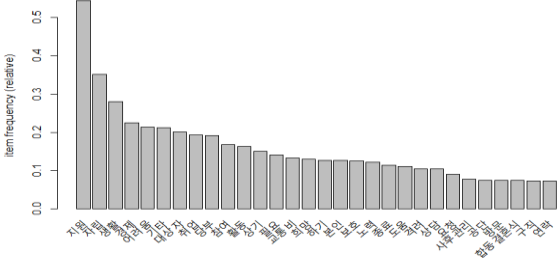
허그일자리

허그일자리



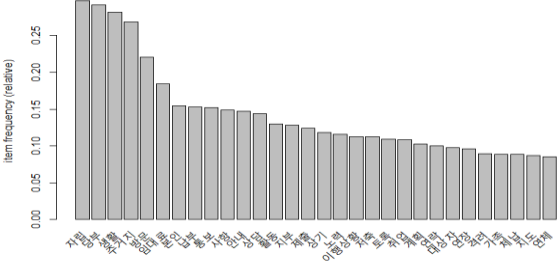
기타지원

기타지원



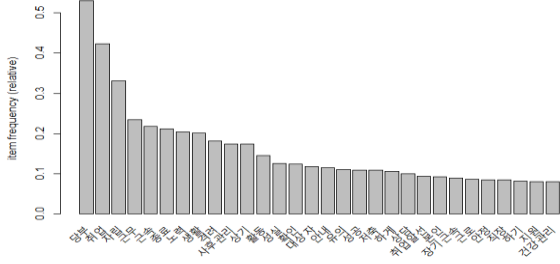
주거지원

주거지원



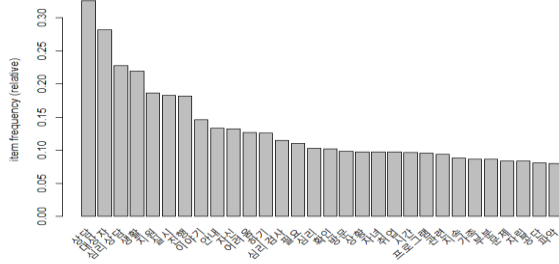
취업지원

취업지원



심리상담

심리상담



Appendix

Data columns

변수 명	의미	타입	변수 명	의미	타입
DSJ_CARD_NO	보호카드보호	Int64	JAEBUM	재범 횟수	Int64
DSJ_JUMIN_NO	개인식별코드	Object	JAEBUM_YN	재범 여부	Object
CONSULT_DATE	상담 일자	Int64	AGE	연령	Int64
CARE_OPEN_CLASS_NM	보호사업코드	Int64	CRIME_GR	범죄 그룹화	Object
CONSULT_CONT	상담 내용	Object	EDU_GR	학력 그룹화	Object
CONSULT_GBN	상담 구분	Int64	CONSULT_N	상담 횟수	Int64
YEAR	출생 년도	Int64	CONSULT_DATE_COUNT	상담 기간	Int64
EDU	학력	Int64	RISK	심리상담 고위험군	Object
CRIME	죄명	Int64	CARE_OPEN_CLASS_NM_GR	보호사업 그룹화	Object
MARRIAGE	혼인	Int64	TF1 ~ 9	심리상담 점수	Int64

4p : <http://www.ioj.go.kr/homepage/index/MainAction.do?method=index> (범죄백서)

5p : <https://www.insight.co.kr/news/160862> (인사이트)

6p : <https://koreha.or.kr/> (한국법무보호복지공단)

8p : <https://koreha.or.kr/> (한국법무보호복지공단)

10p : <https://www.flaticon.com/packs/charity-27> (flaticon)

12p : <https://konlpy-ko.readthedocs.io/ko/v0.4.3/morph/> (konlpy)

26p : <http://m.civilpost.net/2035> (시빌포스트)

<http://www.seoul.co.kr/news/newsView.php?id=20100911010009> (서울신문)

<http://mn.kbs.co.kr/news/view.do?ncd=2946131> (KBS NEWS)

<http://www.womennews.co.kr/news/articleView.html?idxno=71321> (여성신문)

<https://www.forbes.com/sites/falonfatemi/2018/08/27/insight-selling-is-the-new-solution-selling/#57413e537646> (Forbes)