

ESTsoft 와썹 AI 모델 개발 12기

잡플래닛 리뷰 기반 텍스트 마이닝

결과보고서

2025.09.29 | SAPI | 김정철(팀장), 신민경, 정준범, 조아람

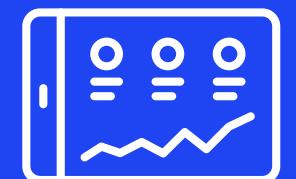
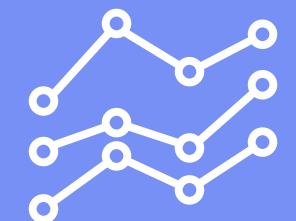


Table of Contents

속보!! 팀 SAPI 두번째 프로젝트 본격 개시!!

01

프로젝트 개요

02

프로젝트 팀 구성 및 역할

03

프로젝트 수행 절차 및 방법

04

프로젝트 수행 경과

01

프로젝트 개요

1-1. 주제 및 선정 배경



EST HR팀의 데이터 기반 문제 해결 노력



이스트소프트 HR팀은 최근 퇴사율 증가와 채용의 어려움을 해결하려고 한다. 채용 플랫폼에 기록된 수많은 회사 리뷰들을 텍스트 마이닝으로 분석을 통해 퇴사 사유와 입사 포기 이유를 파악하고 조직 문화를 개선하여 문제를 처리하고자 한다.

1-2. 기획 의도



데이터 기반 기업 이해와 실무 응용 경험



잡플래닛 리뷰 데이터를 기반으로 이스트소프트의 장점·단점·경영진 의견 등을 분석하고, 텍스트 마이닝 기법을 활용해 설정한 가설들을 검증하였다. 이를 통해 수업에서 배운 내용을 실제 기업 데이터에 적용하며, 데이터 기반 기업 이해와 실무 응용 경험을 쌓는 데 목적이 있다.

1-3. 프로젝트 내용



콘셉트

Jobplanet

1. 기업 평판의 주관성 문제

일반적으로 기업에 대한 인식은 소문, 개인 경험, 홍보 자료에 의존하기 쉽다. 그러나 이는 객관적이지 않고, 실제 재직자들의 목소리를 충분히 반영하지 못한다.

2. 채용사이트 리뷰 데이터의 활용 필요성

채용사이트에는 직원들이 작성한 구체적인 장점·단점·경영진 의견 데이터가 쌓여 있다. 하지만 이 데이터는 방대한 텍스트 형태로 존재하기 때문에, 체계적인 분석 없이는 통찰을 얻기 어렵다.

1-3. 프로젝트 내용



구현내용



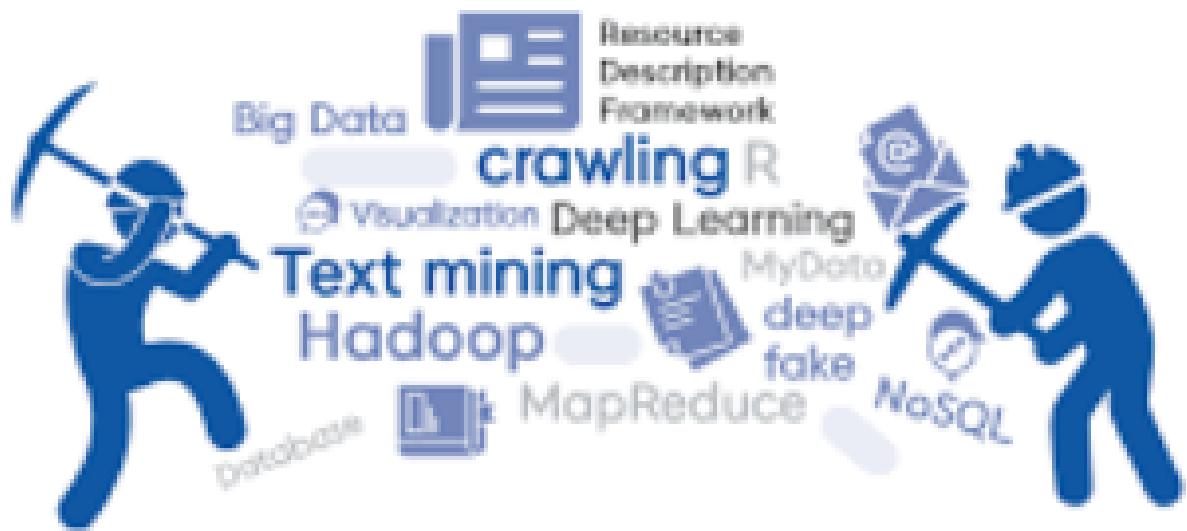
자연어 처리, 시각화 등을 활용한 가설 설정 및 검증을 통해 문제를 해결할 솔루션을 탐색한다.

- 가설 1: "이스트소프트는 단점이 장점보다 많은 회사이다."
- 가설 2: "신입 교육이 지원자 감소의 원인 중 하나이다."
- 가설 3: "개발 부서의 업무 강도가 타 부서보다 높을 것이다."
- 가설 4: "우수 인재 유출의 가장 큰 원인은 성장 가능성의 부재일 것이다."
- 가설 5: "직원들은 운영진에게 칭찬보다는 쓴소리를 더 많이 남길 것이다."

1-3. 프로젝트 내용



훈련 내용과의 연관성



자연어 처리(NLP): KoNLPy를 활용한 한국어 토큰화, 불용어 처리 등 텍스트 전처리 기술의 중요성을 체감하고 실습한다.

분석 모델 적용: 키워드 빈도 분석, 감성 분석, 토픽 모델링 등 수업에서 배운 다양한 텍스트 마이닝 기법을 비즈니스 문제 해결이라는 명확한 목적 아래 적용한다.

시각화 능력: WordCloud, Matplotlib, Seaborn을 사용해 분석 결과를 효과적으로 시각화하고, 이를 통해 데이터로부터 설득력 있는 스토리를 만드는 능력을 기른다.

1-4. 활용 장비 및 재료



파이썬이 베이식과 동급이라면 옵티머스 프라임은 트럭이다



텍스트마이닝에 Selenium

텍스트처리에 KoNLPy

EDA 단순 단어 빈도 분석, 워드클라우드 시각화, tf-idf

분석결과시각화에 WordCloud / Matplotlib

1-5. 프로젝트 구조



프로젝트 구조도

Jobplanet

데이터 탐색



데이터 추출

python™

데이터 분석



인사팀

02

프로젝트 팀 구성 및 역할

2. 프로젝트 팀 구성 및 역할



우리팀을 소개합니다.

훈련생	역할	담당업무
김정철	팀장	데이터 수집, 데이터 전처리, 주제 선정 배경 기획, EDA, 가설 1 & 가설 5 설정 및 검증, 인사이트 후보 도출, 모델링
신민경	팀원	시나리오 작성, EDA, 가설 2 설정 및 검증, 모델 활용 및 응용 방안 구상
정준범	팀원	PPT 자료 제작, EDA, 가설 4 설정 및 검증
조아람	팀원	페르소나 생성, EDA, 가설 3 설정 및 검증, 발표

03

프로젝트 수행 절차 및 방법

3. 프로젝트 수행 절차 및 방법



REIMAGINING YOUR EVERYDAY LIFE WITH AI HUMAN

구분	기간	활동	비고
사전 기획	9/26(금)	프로젝트 기획 및 주제 선정	아이디어 선정
데이터 수집	9/29(월)	필요 데이터 수집 및 정리	선정 사이트 크롤링
데이터 전처리	9/29(월) ~ 10/1(화)	데이터 정제 및 정규화 결측치 처리 텍스트 분리·통합	
데이터 분석	10/1(수) ~ 10/2(목)	EDA / 가설 설정 및 검증	검증 후 결론 도출
모델링	10/13(월) ~ 10/15(수)	모델 구현	팀별 중간보고 실시
고도화	10/16(목) ~ 10/17(금)	활용·응용 방안 및 개선점 도출	최종 수정

04

프로젝트 수행 경과

4-1. 학습 데이터 소개



잡플래닛의 (주)이스트소프트의 리뷰데이터

기획/경영 | 현직원 | 서울 | 2025. 09 작성

3.0

고인물들의 세계가 깨지고 있다. 새로운 변화중

★★★☆☆

승진 기회



복지/급여



워라밸



사내문화



경영진



장점 높은 브랜드 인지도, 자유로운 휴가, 워라밸을 지지하는 문화 등

단점 비전없음, 경영진 능력 부족, 뭘 해야할지 잘 모름, 이것저것 간만 봄

경영진에 바라는 점 뭘 해야할지 모르겠으면 공부하세요. 비전있는 경영진이 필요 합니다.

- 1년 후에도 비슷할 것 같아요

👉 기업을 추천해요

도움이 돼요 ⚡ 1

신고하기

2014. 04 부터 작성된 이스트소프트에 재직했던 모든 직군의 현직자의 리뷰를 사용

<데이터 구성>

제목, 장점, 단점, 경영진에게 하고 싶은 말

4-1. 학습 데이터 소개



잡플래닛의 (주)이스트소프트의 리뷰데이터



제목 : 자율과 평등의 회사 자유로움을 원한다면 강추

장점 : 수평적인 문화에, 자유로운 근무, 주3일 재택근무 최고의 회사

단점 : 자유에는 책임이 따르고, 알아서 '잘' 해야한다는 점, 알아서 어렵게 해야하는지에 대한 끊임없는 고민이 필요하다.

경영진에게 하고싶은 말 : 좋은 회사를 만들기 위해 꾸준히 노력하시는 것 같습니다

제목 : 게임쪽은 죽어가는 분위기 월급 루팡이 많음, 칼퇴들을 잘함 그외는 잘 모름

장점 : 팀에 따라 다르지만 연차, 반차, 반반차 사용이 쉬움 술없는 회식 워크샵은 잘안감

단점 : 인센티브가 없음

경영진에게 하고싶은 말 : 게임은 만들때는 많은 사람들의 의견을 받거나 프로토입부터 사내 테스트로 의견을 좀 받아서 개발했으면

제목 : 개발자 우대가 좋고, 시험을 통한 승진 제도가 있어라 당신의 운명이 천국과 지옥으로 나뉘게됩니다

장점 : 휴가를 상상 눈치보지 않고 자유롭게 다녀올 수 있는 구조입니다.

단점 : 부서 내 변화가 자주 발생하고 있어 그 와중에 회생자들이 발생하는 경우가 있다. 개별 구성원에 대한 배려가 있으면 좋겠음

경영진에게 하고싶은 말 : 적극적인 투자를 하고 있습니다 사내에 적극적으로 공유하려는것도 좋음

제목 : 오래 다닐만한 회사는 아님, 2년에서 3년 정도 다니기에는 좋은 회사입니다.

장점 : 다들 젊어서 확실히 보수적이지는 않음, 술강요는 전혀 없습니다.

단점 : 회사 정원 평균 연령이 30대 초반으로 어림 이듯인 죽은 여기에 오래 있으려 하기 보단 거쳐가는 회사로 인식

경영진에게 하고싶은 말 : 경영진이 모두 창립멤버 위주여서, 다들 회사 생활을 해본 적이 없음, 보이지 않는 부조리가 많음 이부분을 잘 헤아려주길

제목 : 팀이 어느 팀이냐에 따라 당신의 운명이 천국과 지옥으로 나뉘게됩니다

장점 : 연차를 사용하는 것이 자유로운 편이다 절은 분위기에서 일할 수 있다

단점 : 점점 사라지는 복지 불투명한 비전 체계가 잡혀있지않아서 업무하기 힘들 관리자급이 너무 적고 사원들만 많음

경영진에게 하고싶은 말 : 실무자에게 현실적으로 지시를 내렸으면 좋겠습니다 아무래도 어렵겠지만..

제목 : 고인물들의 세계가 깨지고 있다. 새로운 변화중 다니기에는 좋은 회사입니다.

장점 : 높은 브랜드 인지도, 자유로운 휴가, 워라밸을 지지하는 문화 등

단점 : 비전없음, 경영진 능력 부족, 뭘 해야할지 잘 모름, 이것저것 간만 봄

경영진에게 하고싶은 말 : 뭘 해야 할지 모르겠으면 공부하세요. 비전있는 경영진이 필요합니다.

제목	장점	단점	경영진에게 하고싶은 말
고인물들으 높은 브랜드 비전없음, 뭘 해야할지 모르겠으면 공부하			
기업문화가 재택 근무 재택근무가 AI도 좋지만 다른 서비스도 신경			
네임밸류 회사나 제품 인력 관리 핵심 인력들의 장기근속을 위한			
신입들은 그나마 재택근무 되지도 딱히 바라는건 없고 그냥 하던대			
자유롭고 재택근무 자율이란 연구,개발보단 기획,마케팅이 우선			
신입 개발자 신입 초봉 인사발령 직원들이 실제로 원하는게 무엇인			
예술의전당 젊은 IT 회사 추구미에 화이팅 더 좋은 회사로 만드시길.			
개발자들을 위라밸 훌륭한 경영진 수준 이제 스톡옵션 욕심 버리시고 사			
좋은 사람 1층 카페와 인사팀이 회사의 발전을 진심으로 원한다			
젊은 꼰대들 간식이 있고 회장라인, 아직도 이해할 수 없는 일화가 너			
앞으로의 1대기업 수준 수평적인 "대체적으로 경영진이 유한 성격은			
가족같지만 실수를 하면 전문적인 외부 인력을 적절히 배치하여 조			
사공이 많아서마다 사내 정치 직원들이 바라는건 이미 알고 있			
신사업 신사업 유연근무 재택근무 신사업에 대한 끊임없는 연구로			
무난하면서 신입으로 회사가 이직원들 보상은 약하고 경영진들만			
인사부서에 많은 경험 많은 시기 왜직원들이 퇴사를이 높은지 남			
부부부가 장비지원은 그 외에 모 좋은 능력을 가진 사람들의 말을			

셀레니움으로 동적 크롤링 후, 추출한 텍스트를 CSV파일로 변환

4-2. 전처리



데이터셋 불러오기 및 크기 확인

```
# 데이터셋 불러오기  
  
import pandas as pd  
  
df = pd.read_csv("./data/jobplanet_review.csv")  
df.head(5)
```

	제목	장점	단점	경영진에게 하고싶은 말
0	고인물들의 세계가 깨지고 있다. 새로운 변화중	높은 브랜드 인지도, 자유로운 휴가, 워라밸을 지지하는 문화 등	비전없음, 경영진 능력 부족, 뭘 해야할지 잘 모르는 것 같아요.	뭘 해야할지 모르겠으면 공부하세요. 비전 있는 경영진이 필요합니다.
1	기업문화가 좋습니다. 되게 자유롭고 일에 몰두할 수 있었어요	재택근무가 가장 큰 장점입니다. 1층에 카페도 저렴해서 좋았습니다.	재택근무가 아니라면 단점이 될 수도 있는 점이 꽤 많을 것 같습니다.	AI도 좋지만 다른 서비스도 신경써주세요.
2	네임밸류 있고 나름 좋은 기업으로 기억함	회사나 제품이 네임밸류가 좀 있다보니 일하는데 있어서 자부심 비슷한게 있음	인력 관리가 잘 안되는 듯. 그래서 단기간 근무하고 이직하는 직원들이 많음	핵심 인력들의 장기근속을 위한 조치를 마련해주세요.
3	신입들은 들어와서 회사가 이런곳이구나 알고 가면 좋은 회사	그나마 재택근무 있어서 회사에 계속 다녔다. 이거마저 없어지면 남아서 다닐 사람은 ...	잘 되지도 않고 다른 회사랑 차별점도 없는 AI 툴 구름만 보고 끊어감, 수직적인	딱히 바라는 건 없고 그냥 하던대로 하세요
4	자유롭고 자율적이네 실적이 필요하다.	재택근무 3일의 장점이 크다 아주 큰 장점 복지도 꽤나 좋다	자율이란 이름의 방치 혹은 압박 존재 개발자만 그럴수도..	연구, 개발보단 기획, 마케팅이 우선되는 느낌입니다

```
# 데이터셋 기본 구조
```

```
print("데이터 크기 : {}".format(df.shape))  
print("컬럼명 : {}".format(df.columns.tolist()))
```

데이터 크기 : (367, 4)

컬럼명 : ['제목', '장점', '단점', '경영진에게 하고싶은 말']

4-2. 전처리



결측치 확인

```
# 결측치 확인
```

```
print(df.isna().sum())
```

제목

0

장점

0

단점

0

경영진에게 하고싶은 말

2

dtype: int64

```
# 결측치 위치 확인
```

```
df[df['경영진에게 하고싶은 말'].isna()]
```

제목	장점	단점	경영진에게 하고싶은 말
336 여자가 다니기 괜찮은 회사	부서마다 차이가 있겠지만 남녀 차별이 없다. 술 권하지 않는다. 출산 휴가가 자유롭...	어딜가나 있겠지만, 뒷담화 하는 사람이 있음. 부서마다 차이 있겠지만 끼리끼리 어울...	Nan
365 이스트소프트	회사 문화와 사람들의 충성도, 그리고 매니지먼트의 도덕성 장점인 회사	너무 커지다 보니, 창조성이 점점 떨어지고 있음	Nan

4-2. 전처리



결측치 확인

여자가 다니기 괜찮은 회사

장점 부서마다 차이가 있겠지만 남녀 차별이 없다.

술 권하지 않는다.

출산 휴가가 자유롭다. 기본 1년이고 2년 쉬는 사람도 있었던 것 같다.
복귀해도 동일한 대우를 받는다.

단점 어딜가나 있겠지만, 뒷담화 하는 사람이 있음.

부서마다 차이 있겠지만 끼리끼리 어울리는 경향이 있다.

이스트소프트

장점 회사 문화와 사람들의 충성도, 그리고 매니지먼트의 도덕성 장점인 회사

단점 너무 커지다 보니, 창조성이 점점 떨어지고 있음

경영진에 바라는 점

경영진에 바라는 점

4-2. 전처리



결측치 제거 및 확인

결측치 공백으로 빼우기

```
df = df.fillna("")  
print(df.isna().sum())
```

제목	0
장점	0
단점	0
경영진에게 하고싶은 말	0
dtype:	int64

결과 확인

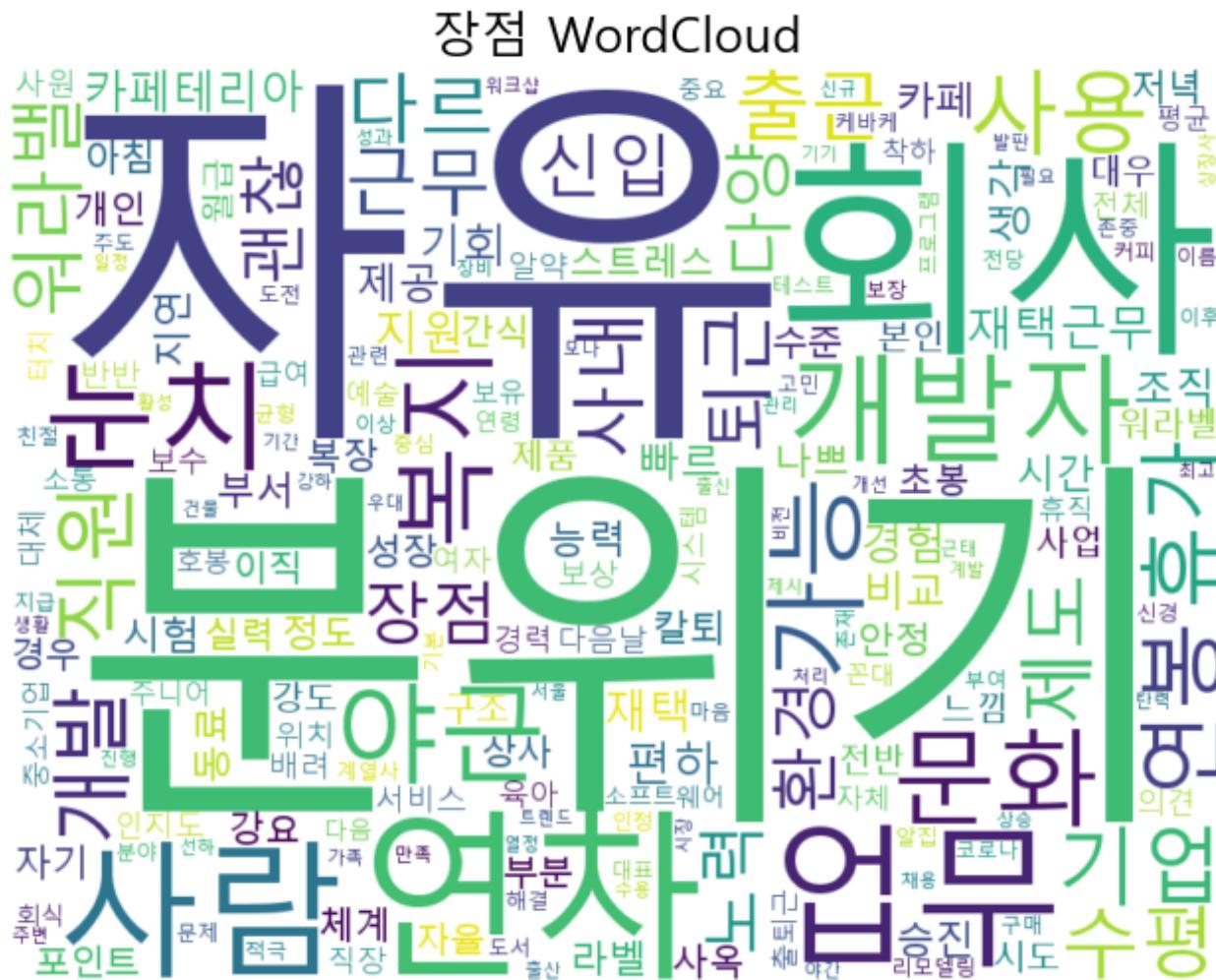
```
df.loc[[336, 365]]
```

제목	장점	단점	경영진에게 하고싶은 말
336 여자가 다니기 괜찮은 회사	부서마다 차이가 있겠지만 남녀 차별이 없다. 술 권하지 않는다. 출산 휴가가 자유롭...	어딜가나 있겠지만, 뒷담화 하는 사람이 있음. 부서마다 차이 있겠지만 끼리끼리 어울...	
365 이스트소프트	회사 문화와 사람들의 충성도, 그리고 매니지먼트의 도덕성 장점인 회사	너무 커지다 보니, 창조성이 점점 떨어지고 있음	

4-3. EDA(탐색적자료분석)



워드클라우드(장점, 단점)



4-3. EDA(탐색적자료분석)



워드클라우드(경영진에게 하고 싶은 말, 제목

경영진에게 하고 싶은 말 WordCloud



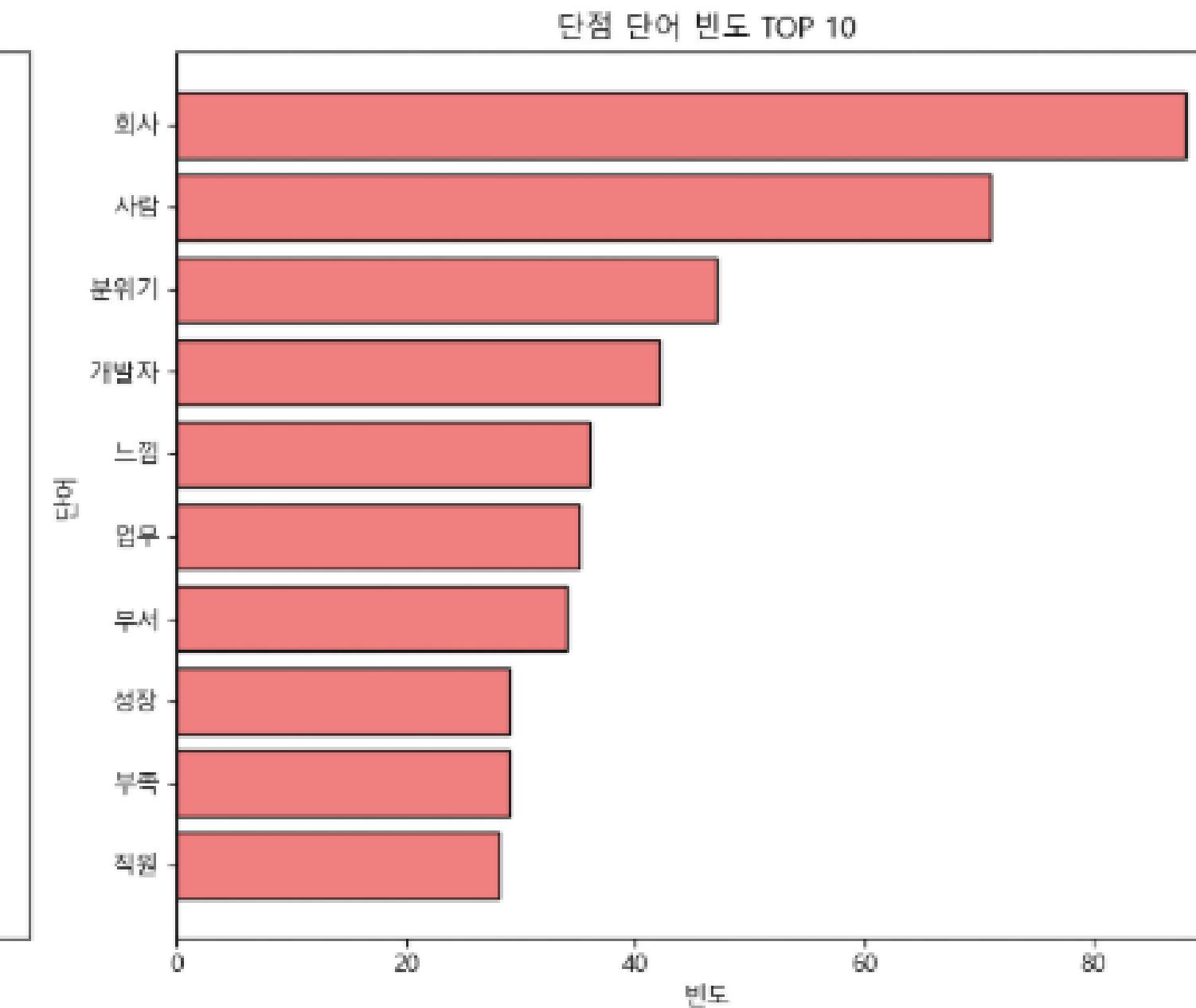
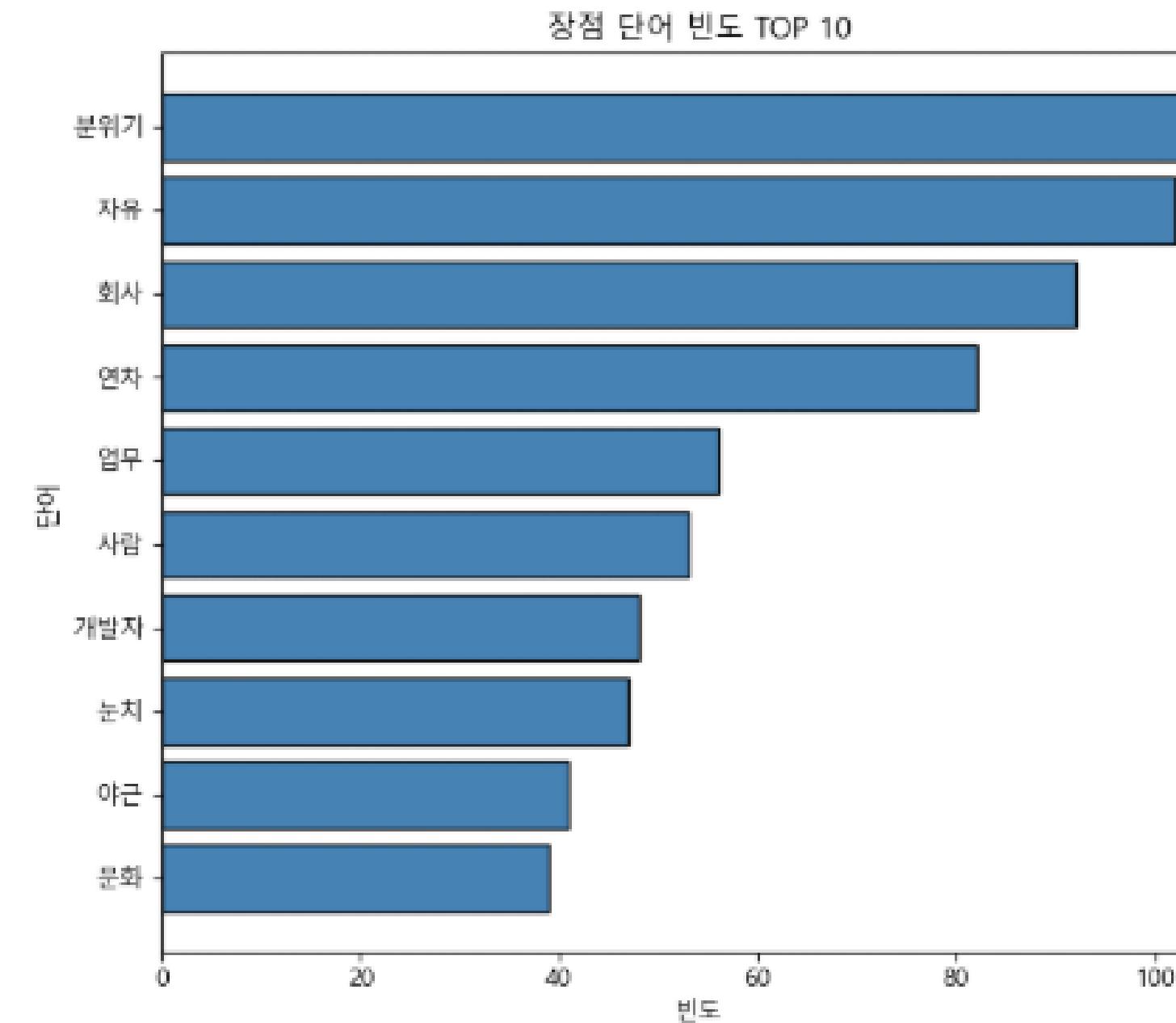
제목 WordCloud



4-3. EDA(탐색적자료분석)



단어빈도수 분석



4-3. EDA(탐색적자료분석)



단어빈도수 분석

장점 키워드 단점 키워드 복합적 키워드

분위기	사람	회사
자유	느낌	개발자
연차	부서	
업무	성장	
눈치	부족	
야근	직원	
문화		

결과:

공통적으로 분위기, 회사, 사람, 업무, 개발자가 많이 언급되었다.
이 중에서 분위기와 업무는 장점 쪽에서 더 많이 등장하므로 장점으로,
사람은 단점에서 더 자주 언급되어 단점으로 분류하였다.

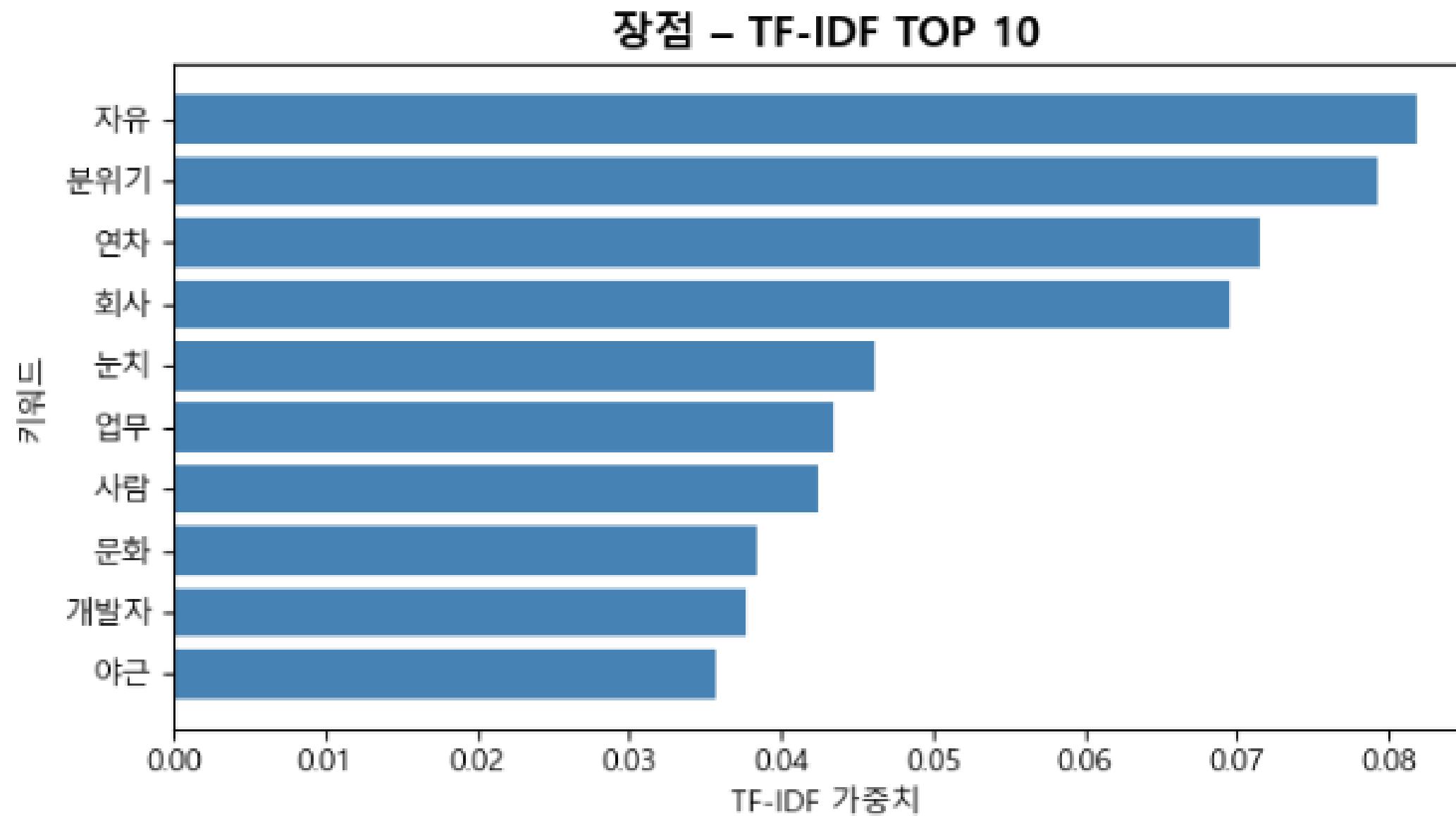
다만, 회사와 개발자는 긍정·부정 양쪽에서 고르게 나타나
복합적인 양상이 관찰되었다.

장점 쪽에서는 분위기, 자유, 연차, 업무, 눈치, 야근, 문화 등이
자주 언급되었고, 단점 쪽에서는 사람, 느낌, 부서, 성장, 부족, 직원 등이
자주 언급되었다.

4-3. EDA(탐색적자료분석)



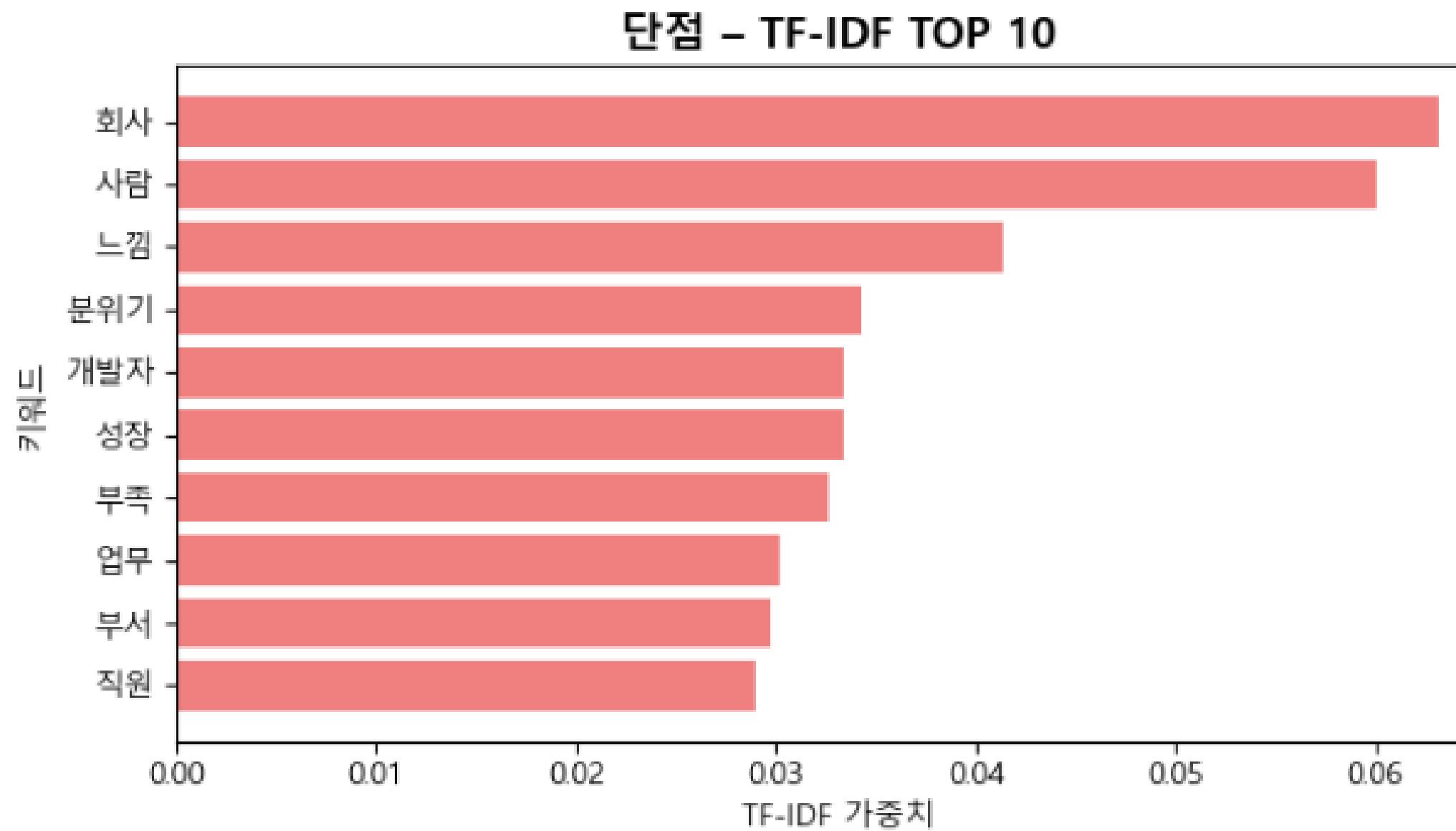
TF-IDF



4-3. EDA(탐색적자료분석)



TF-IDF



4-3. EDA(탐색적자료분석)



TF-IDF

장점 키워드 단점 키워드 복합적 키워드

자유	사람	회사
분위기	느낌	개발자
연차	성장	
눈치	부족	
업무	부서	
문화	직원	
야근		

결과:

빈도 분석과 TF-IDF 분석은 결과적으로 동일한 키워드를 분류해냈다.
그러나 단어들의 순위에는 차이가 나타났는데,
이는 빈도 분석이 단순히 자주 언급되는 키워드를 강조하는 반면,
TF-IDF는 집단 간 특성을 더 잘 구분 지어주는 키워드에
가중치를 두기 때문이다.

4-4. 가설검정

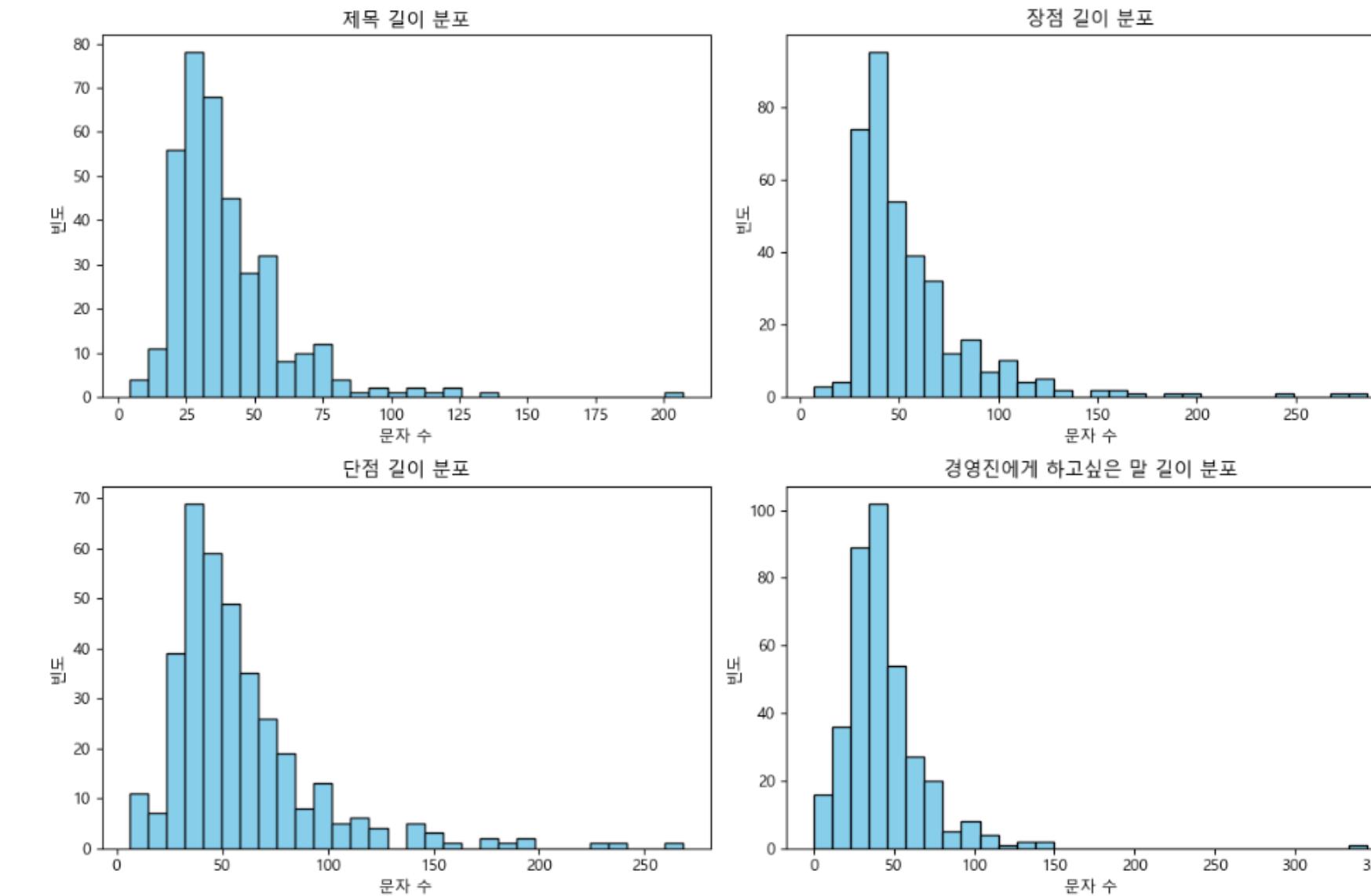


1. 이스트소프트는 단점이 장점보다 많은 회사이다.

```
# 각 컬럼별 텍스트 길이 통계
```

```
stat = length.describe().round(2)  
stat
```

	제목	장점	단점	경영진에게 하고싶은 말
count	367.00	367.00	367.00	367.00
mean	39.55	55.93	58.44	43.55
std	21.42	33.82	35.69	27.45
min	4.00	7.00	6.00	0.00
25%	26.00	35.00	36.00	28.00
50%	34.00	46.00	49.00	39.00
75%	47.00	64.00	69.00	51.00
max	207.00	286.00	268.00	345.00



리뷰에서 장점 문장의 길이가 단점 문장의 길이보다 길면 장점이 더 많은 회사로,
반대의 경우 단점이 더 많은 회사로 간주하였다.

4-4. 가설검정



1. 이스트소프트는 단점이 장점보다 많은 회사이다.

```
## Welch's t-검정

import numpy as np
from scipy.stats import ttest_ind

# 컬럼 선택
pros_col = '장점'
cons_col = '단점'

# 가설 검정
t_stat, p_value = ttest_ind(cons, pros, equal_var = False, alternative = 'greater')

# 결과 출력
print("T 통계량: {:.3f}".format(t_stat))
print("p-value: {:.3f}".format(p_value))

alpha = 0.05 # 유의수준

# 판단
if p_value < alpha:
    print("이스트소프트 기업은 장점보다 단점이 많은 회사일 것이다.")
else:
    print("이스트소프트 기업은 장점보다 단점이 많은 회사라고 보기 어렵다.")
```

T 통계량: 1.047
p-value: 0.148

이스트소프트 기업은 장점보다 단점이 많은 회사라고 보기 어렵다.

이를 통계적으로 검정하기 위해 Welch의 t-검정 (등분산 가정 없음, 단측 검정: 단점 > 장점)을 실시한 결과,

- $t = 1.047$,
- $p = 0.148$ (유의수준 $\alpha = 0.05$)

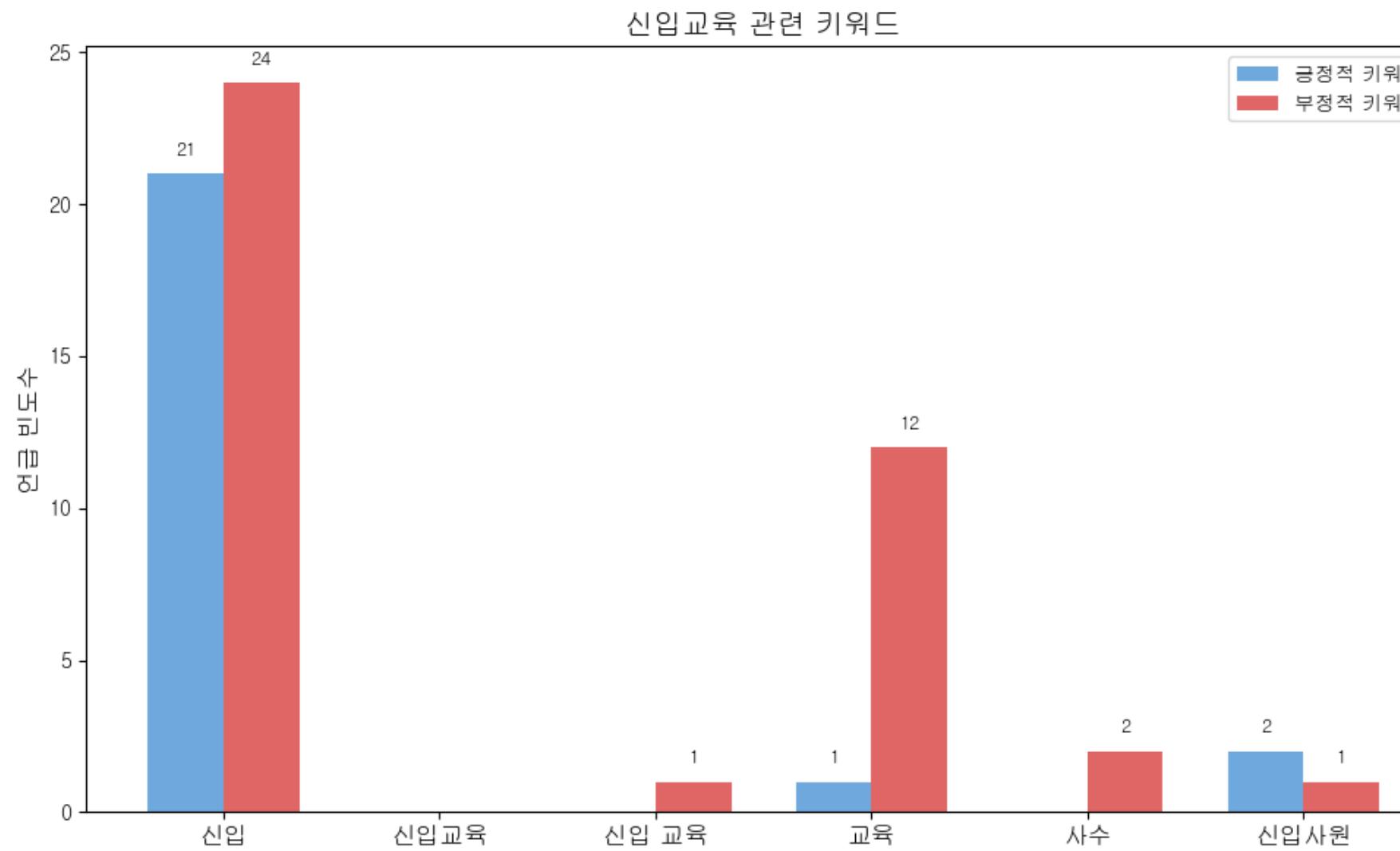
으로 나타났다.

즉, 실제로는 장점과 단점의 문장 길이 차이가 통계적으로 유의하지 않으며, 따라서 이스트소프트 기업을 단점이 장점보다 많은 회사라고 보기 어렵다고 결론지을 수 있다.

4-4. 가설검정



2. 신입 교육이 지원자 감소의 원인 중 하나이다.



결과 :

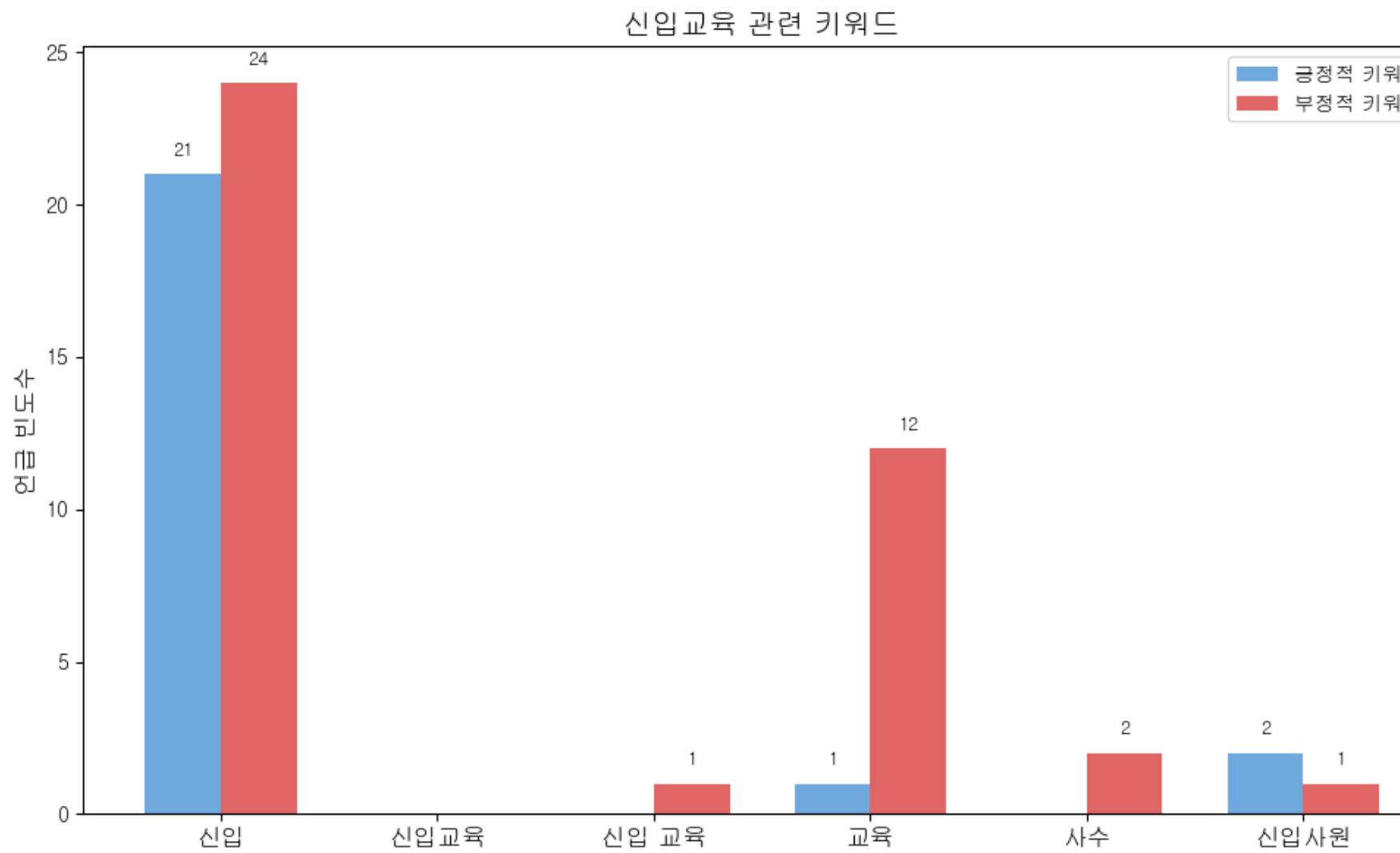
신입은 복합적으로 결과가 도출됐지만 교육 이외에 다른 변수(신입사원의 근태 등)가 많으므로 이 가설에서 확신을 내리기엔 무리가 있다.

교육과 신입교육, 신입 사원의 교육을 담당하는 사수에서 부정적인 이미지(단점)이 많이 추출됐지만 통계적으로 유의미한 양의 자료가 아니라 가설을 뒷받침하기엔 무리가 있다.

4-4. 가설검정



2. 신입 교육이 지원자 감소의 원인 중 하나이다.



결과 :

또한 p-value검정을
진행한 결과 p-value 값이 0.19100이 나와 해당 가설의
값이 긍정적 키워드와 부정적 키워드가 유의미한 차이가
있다고 보기 어렵다.

따라서 데이터의 수를 더 확보하여 텍스트 마이닝을 진행하
지 않는 이상 위의 수치로는 신입 교육이 지원자들이 줄어드
는 원인 중 하나라고 보기엔 무리가 있다.

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

☞

키워드 빈도 시각화 (WordCloud)



<'개발'이 포함된 리뷰를 키워드 빈도(word frequency) 기반으로 워드클라우드 실행한 모습>

개발 부서에 대한 키워드를 간접적으로 보여준다.

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.



<'개발'이 포함된 리뷰를 키워드 빈도(word frequency) 기반으로 워드클라우드 실행한 모습>

개발 부서에 대한 긍정 키워드를 간접적으로 보여준다.

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.



Negative Keyword Word Cloud

A word cloud visualization where the size of each word represents its frequency. The most prominent words are '야근' (Overtime) in large green font and '실적' (Performance) in dark blue font, indicating they are the most frequently mentioned negative keywords associated with development department work.

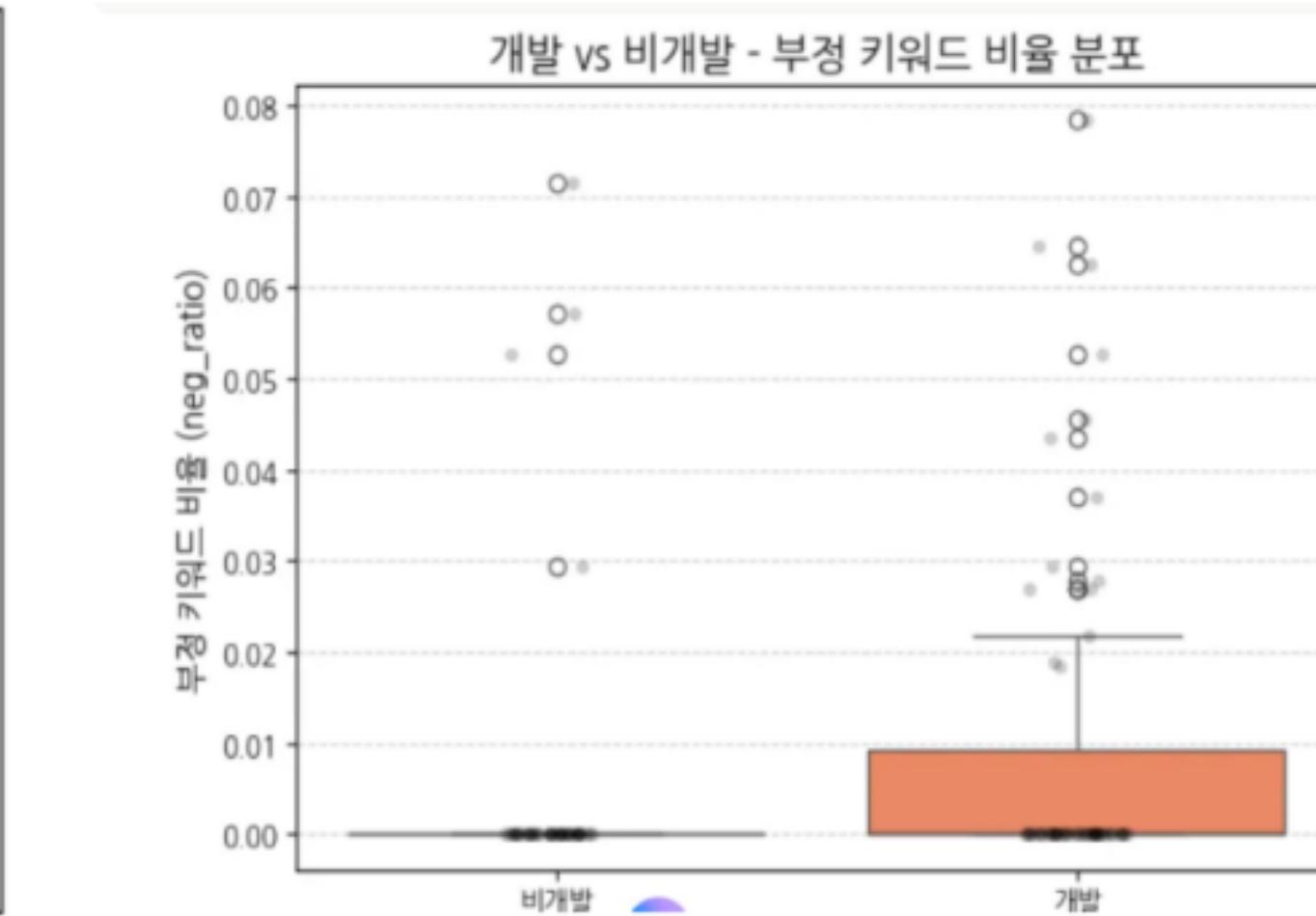
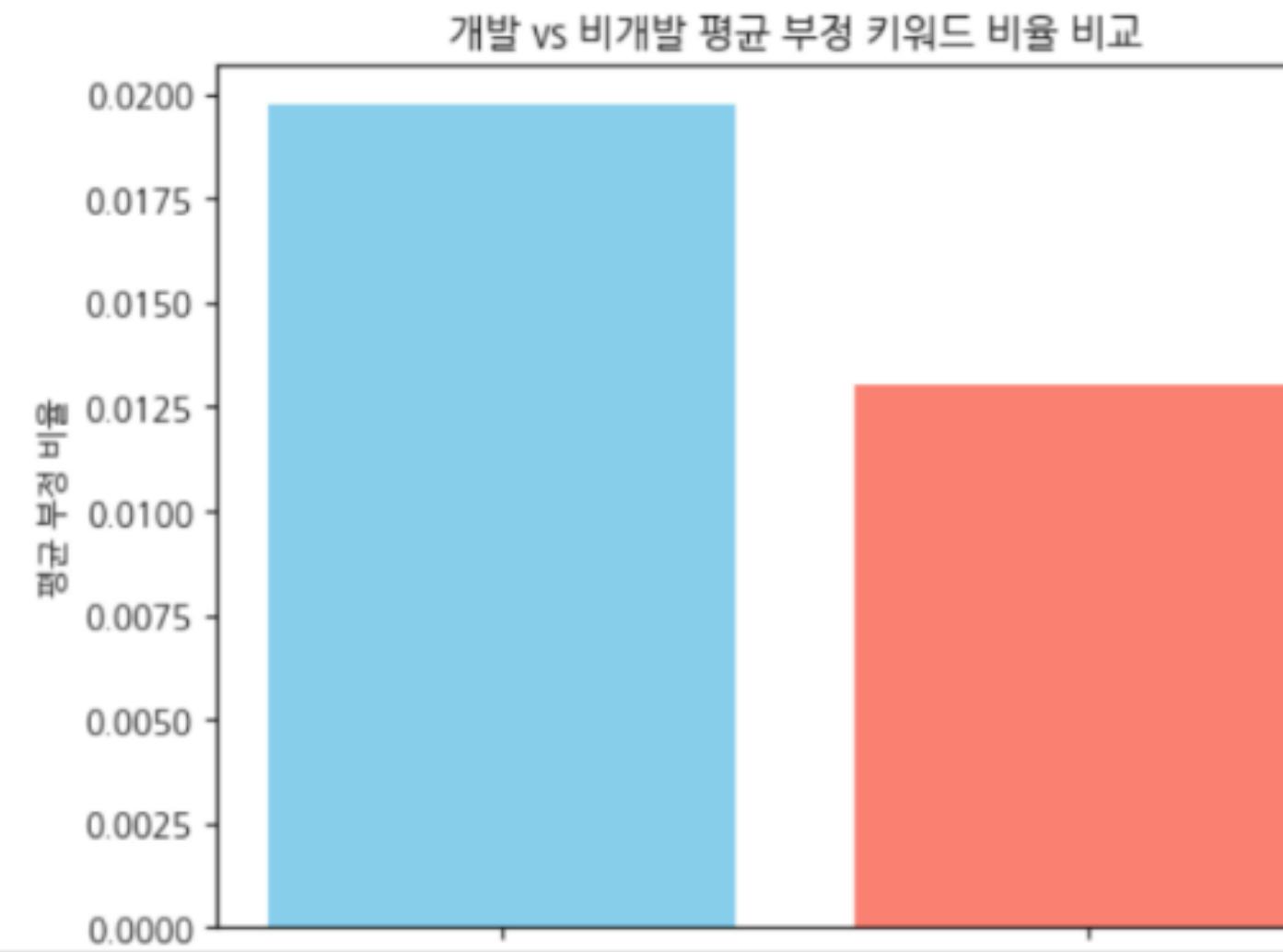
<'개발'이 포함된 리뷰를 키워드 빈도(word frequency) 기반으로 워드클라우드 실행한 모습>
개발 부서에 대한 부정 키워드를 간접적으로 보여준다.

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

- ✓ 개발 평균 부정 비율: 0.02
- ✓ 비개발 평균 부정 비율: 0.013
- ✓ t값: 1.216, p값: 0.227



개발 부서와 비개발 부서의 평균 부정 키워드 비율이 차이가 있는가를 통계적으로 검정

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

```
# =====
# 잡플래닛 리뷰 데이터 기반
# 개발 vs 비개발 리뷰의 '스트레스 관리 단어' 비율 비교
# =====

# 1. 라이브러리 불러오기
import pandas as pd
from kiwipy import Kiwi
from scipy.stats import ttest_ind

# 2. 데이터 불러오기
df = pd.read_csv('/content/jobplanet_review (3).csv')
df = df.fillna('')

# 3. 텍스트 통합 (리뷰 전체 문장 만들기)
df['text'] = df['질적'] + ' ' + df['단점'] + ' ' + df['결연진에게 하고싶은 말']

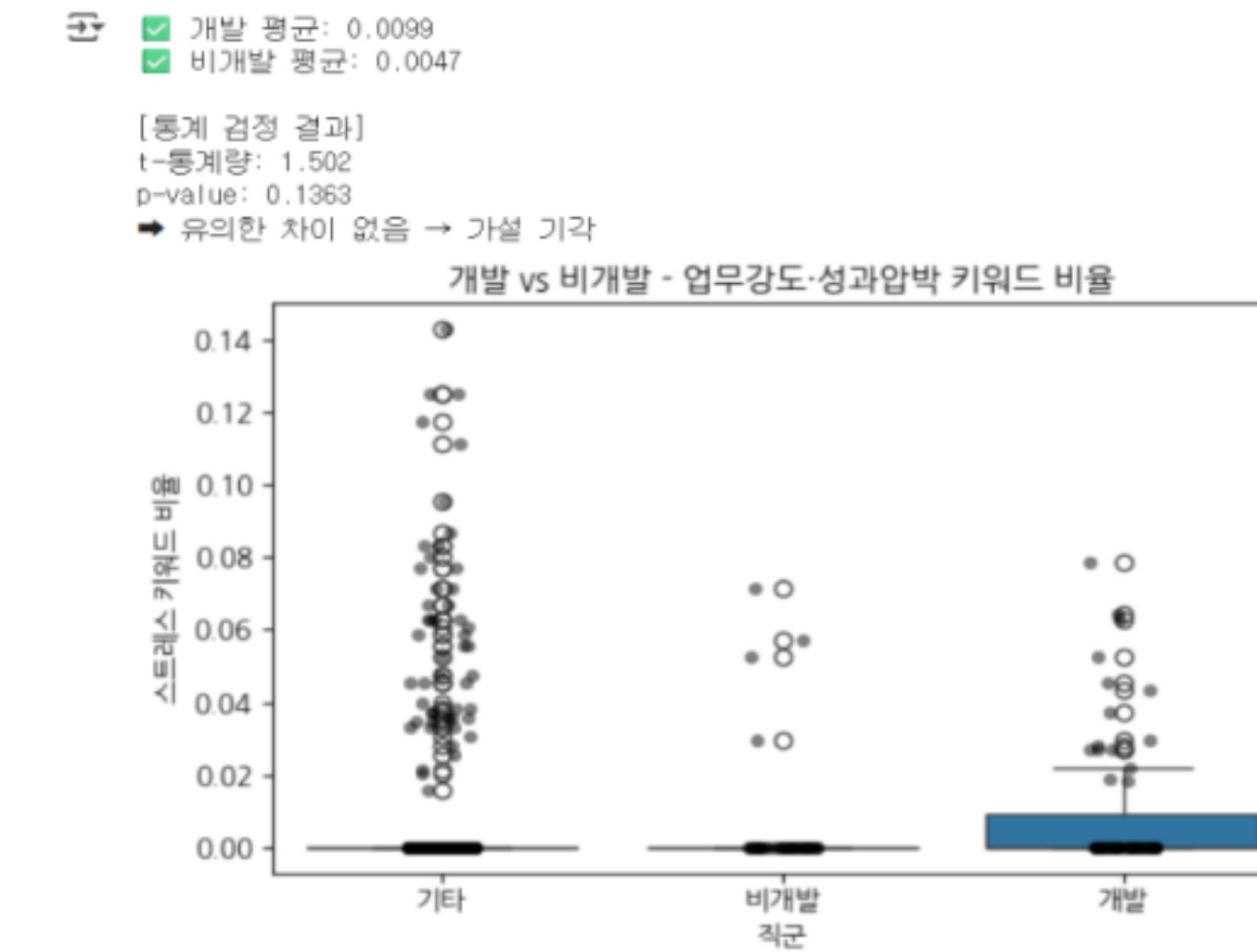
# 4. 형태소 분석기 초기화
kiwi = Kiwi()

# 명사만 추출
def noun_tokenizer(text):
    return [token.form for token in kiwi.tokenize(str(text)) if token.tag.startswith("N")]

df['tokens'] = df['text'].apply(noun_tokenizer)

# 5. 개발 / 비개발 분류 규칙
dev_keywords = {"개발", "엔지니어", "프로그래머", "코드", "IT", "데이터", "서비스개발"}
non_dev_keywords = {"영업", "마케팅", "디자인", "인사", "회계", "문영", "기획", "관리"}

def classify_dept(tokens):
    if any(word in tokens for word in dev_keywords):
        return "개발"
    else:
        return "비개발"
```



개발 부서와 비개발 부서의 평균 부정 키워드 비율이 차이가 있는가를 통계적으로 검정

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

```
# -----
# [Jobplanet 리뷰 데이터 기반
# 개발 vs 비개발 리뷰의 '스트레스 관리 단어' 비율 비교
# -----

# 1. 라이브러리 불러오기
import pandas as pd
from kiwipy import Kiwi
from scipy.stats import ttest_ind

# 2. 데이터 불러오기
df = pd.read_csv('/content/jobplanet_review (3).csv')
df = df.fillna('')

# 3. 텍스트 통합 (리뷰 전체 문장 만들기)
df['text'] = df['장점'] + ' ' + df['단점'] + ' ' + df['경영진에게 하고 싶은 말']

# 4. 형태소 분석기 초기화
kiwi = Kiwi()

# 명사만 추출
def noun_tokenizer(text):
    return [token.form for token in kiwi.tokenize(str(text)) if token.tag.startswith("N")]

df['tokens'] = df['text'].apply(noun_tokenizer)

# 5. 개발 / 비개발 분류 규칙
dev_keywords = {"개발", "엔지니어", "프로그래머", "코드", "IT", "데이터", "서비스개발"}
non_dev_keywords = {"영업", "마케팅", "디자인", "인사", "회계", "운영", "기획", "관리"}

def classify_dept(tokens):
    if any(word in tokens for word in dev_keywords):
        return "개발"
    elif any(word in tokens for word in non_dev_keywords):
        return "비개발"
    else:
        return "기타"

df["dept"] = df["tokens"].apply(classify_dept)

# 6. 리뷰를 토큰 단위로 펼치기
rows = []
for i, row in df.iterrows():
    for token in row["tokens"]:
        rows.append({
            "review_id": i,
```

☞ 개발 평균: 0.0099
 비개발 평균: 0.0047

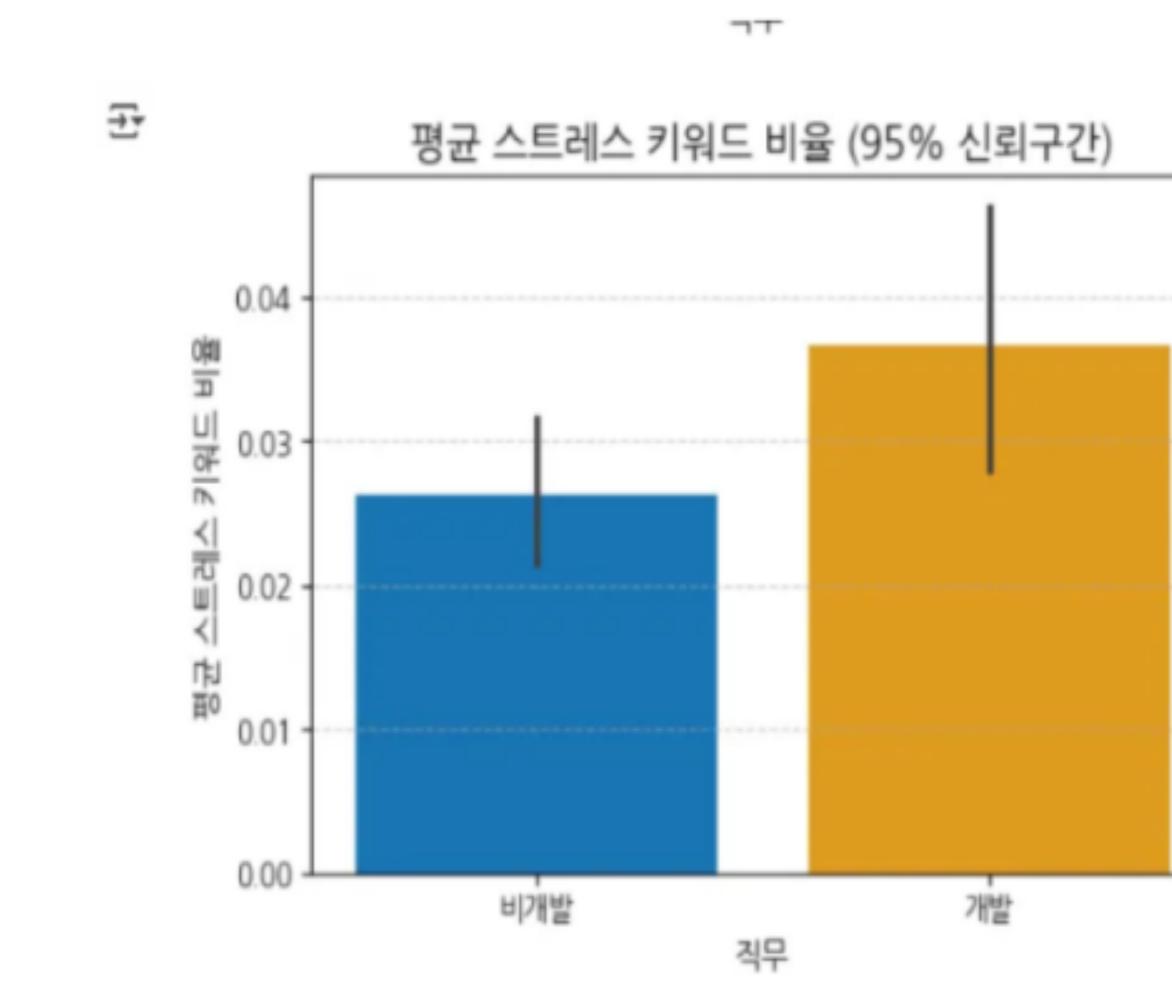
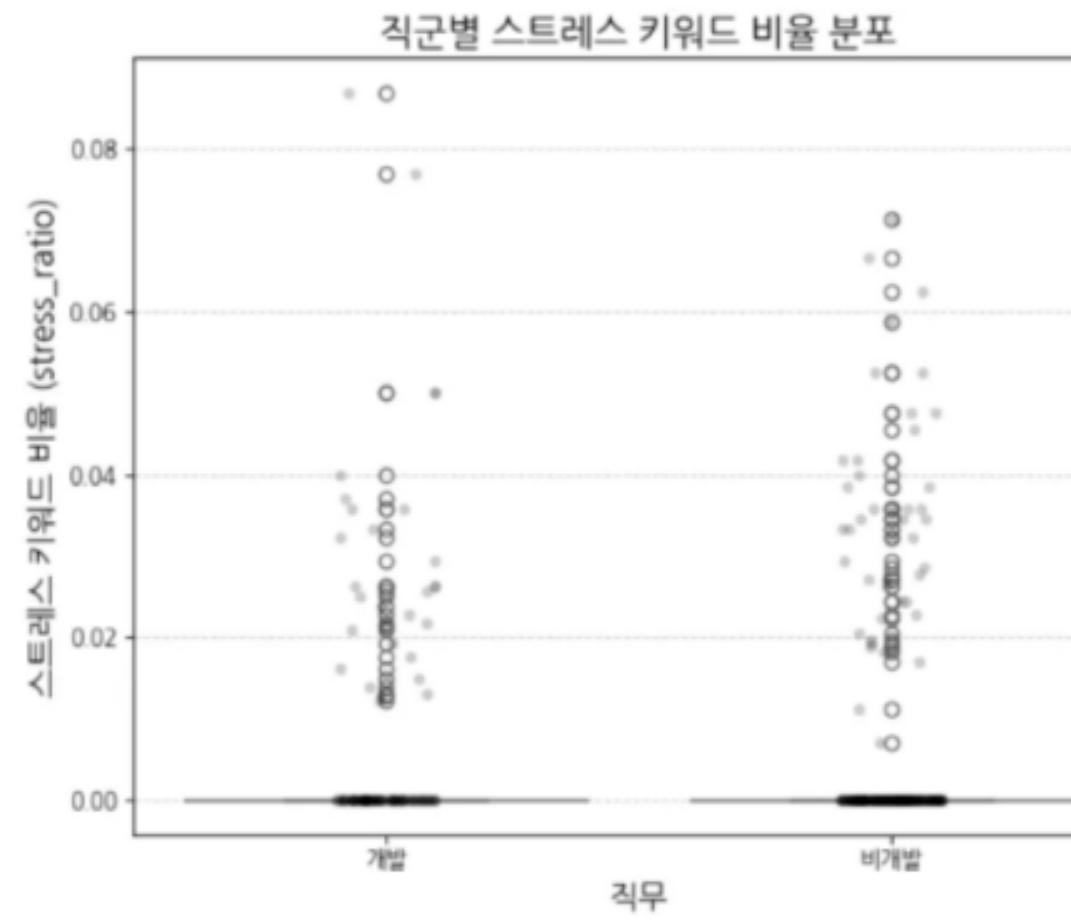
[통계 검정 결과]
t-통계량: 1.502
p-value: 0.1363
➡ 유의한 차이 없음 → 가설 기각

개발 부서와 비개발 부서의 평균 부정 키워드 비율이 차이가 있는가를 통계적으로 검정

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.



개발 부서와 비개발 부서의 평균 부정 키워드 비율이 차이가 있는가를 통계적으로 검정

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

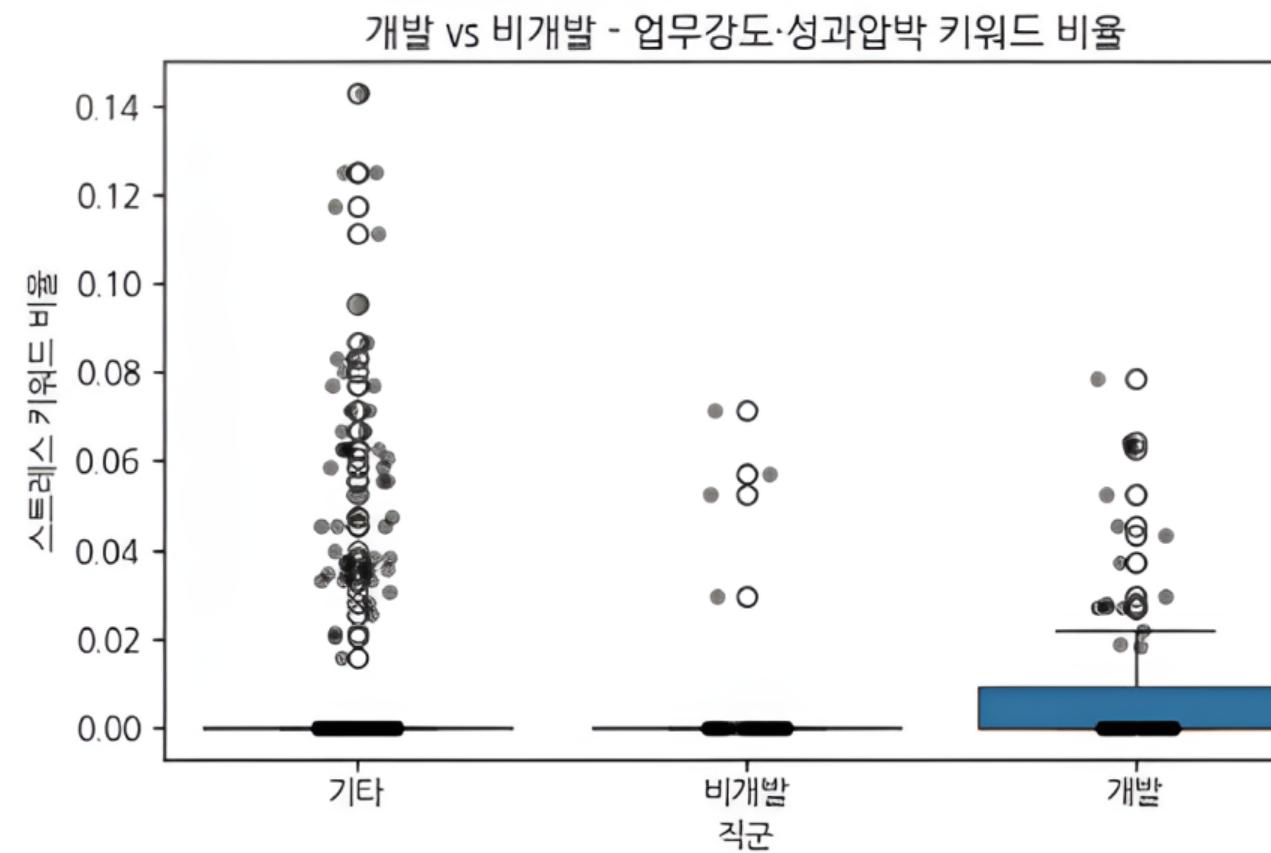
→ 개발 평균: 0.0099
비개발 평균: 0.0047

[통계 검정 결과]

t-통계량: 1.502

p-value: 0.1363

→ 유의한 차이 없음 → 가설 기각



개발 부서 리뷰는 비개발 부서 리뷰에 비해 부정적인 감정이나 불만 관련 키워드를 유의미하게 더 많이 포함하고 있다. 하지만, 단순히 '야근', '성과'와 같은 몇 가지 핵심 스트레스 키워드만으로는 유의미한 차이를 발견하기 어렵거나, 정교한 분류 시 차이가 미미하다.

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

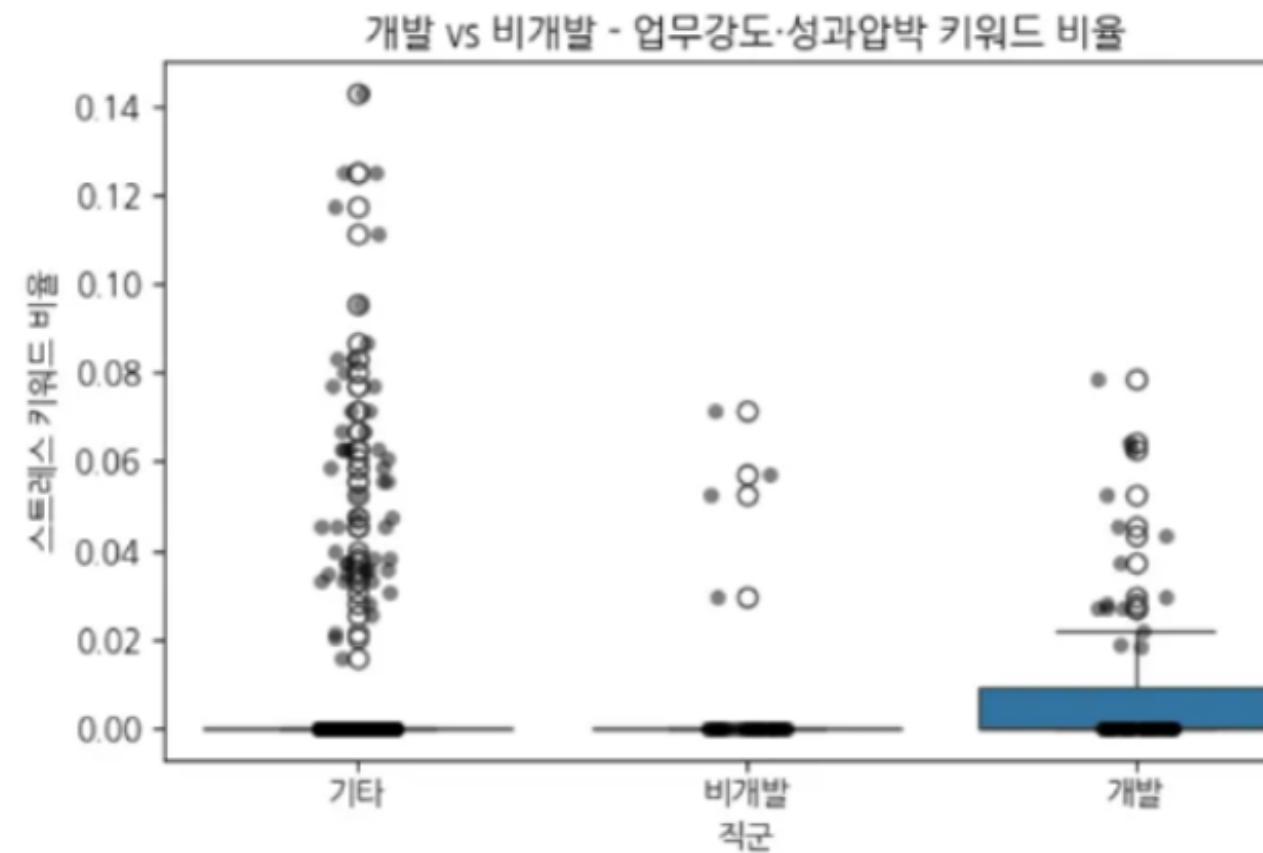
▣ 개발 평균: 0.0099
▣ 비개발 평균: 0.0047

[통계 검정 결과]

t-통계량: 1.502

p-value: 0.1363

→ 유의한 차이 없음 → 가설 기각



<가설검정 >

1. 부정 키워드 비율 비교

개발 평균: 약 0.018

비개발 평균: 약 0.013

t-검정 결과: t = 1.216 / p-value = 0.227

결론:

→ p값이 0.05보다 크므로, 유의미한 차이가 없음 (귀무가설 채택)

→ 즉, 개발직과 비개발직 간의 부정 키워드 비율 차이는 통계적으로 유의하지 않음

4-4. 가설검정



3. 개발 부서의 업무 강도가 타 부서보다 높을 것이다.

<가설검정 >

2. 스트레스 키워드 비율 비교

개발 평균: 약 0.0099

비개발 평균: 약 0.0047

t-검정 결과:t = 1.582 / p-value = 0.1368

결론:

→ 역시 p값이 0.05보다 크므로, 유의미한 차이가 없음 (귀무가설 채택).

→ 개발직이 약간 더 높은 경향은 있지만, 통계적으로 유의한 수준은 아님.

모든 그래프에서 개발직군이 평균적으로 부정/스트레스 키워드 비율이 조금 더 높게 나타남.

하지만 분산이 크고, 표본 내 변동성이 커서 통계적으로 유의하지 않음.

즉, “개발직이 더 스트레스를 받는다”는 직관적 해석은 가능하지만, 통계적 근거는 약함.

개발 부서와 비개발 부서 간의 부정 및 스트레스 관련 키워드 비율은 평균적으로는 개발 부서가 높지만, 통계적으로 유의한 차이는 없다 ($p > 0.05$).

따라서, 현재 데이터로는 두 집단 간 명확한 스트레스·부정 인식 차이를 단정할 수 없다.

4-4. 가설검정



4. 우수 인재 유출의 가장 큰 원인은 성장 가능성의 부재일 것이다.

	미직 사유 유형	언급된 리뷰 수
0	성장 기회 없음	70
1	관리자의 문제	52
2	기대가 없거나 불분명함	41
3	IT에 대한 최상층 지원 부족	38
4	경쟁력 없는 연봉	35
5	번아웃	23
6	업무의 의미 인식 부족	19
7	최첨단 기술/기회 없음	16
8	교류 부족	11
9	불충분한 유연성	10
10	정기적인 상태 확인 없음	7
11	직장 내 외로움	2

우수 인재 유출의 원인을 구글링 하여,
다음과 같은 12가지 원인들로 구분했다.

이직 사유 유형별 키워드 정의
→ 키워드가 언급된 리뷰 수가 많은 순으로 정렬

데이터 크기 : (367, 4)인 df에서 단점컬럼에 핵심키워드가
얼만큼 판독됐는지 수를 셨다.

다음과 같은 결과가 나왔고, 성장 기회 없음 관련 키워드를 언급한
리뷰 수가 70으로 가장 높았고, 그 다음으로 관리자의 문제,
기대가 없거나 불분명함을 뽑았다.

4-4. 가설검정



4. 우수 인재 유출의 가장 큰 원인은 성장 가능성의 부재일 것이다.

1. 경쟁력 없는 연봉

- 핵심 키워드: 연봉, 월급, 급여, 보상, 인상률, 상승률, 연봉테이블, 인센티브, 성과급

7. 관리자의 문제

- 핵심 키워드: 관리자, 팀장, 리더, 꼴대, 수직적, 정치, 불통, 편애, 감정적

2. 교류 부족

- 핵심 키워드: 교류, 소통, 협업, 팀워크, 사회 활동, 동료, 어울리기, 파벌

8. 성장 기회가 없음

- 핵심 키워드: 성장, 커리어, 발전, 배울점, 기회, 승진, 진급, 고인물, 정체

3. 기대가 없거나 불분명함

- 핵심 키워드: 온보딩, 신입, 교육, 체계, 시스템, 목표, 가이드, 방치, 인수인계

9. 최첨단 기술이나 기회가 없음

- 핵심 키워드: 기술, 신기술, 트렌드, 레거시, 코드, 개발문화, 도전, 기회

4. 업무의 의미 인식 부족

- 핵심 키워드: 비전, 방향성, 목표, 의미, 보람, 뜬구름, 목적의식, 동기부여

10. 정기적인 상태 확인(Check-in) 없음

- 핵심 키워드: 피드백, 면담, 원온원(1on1), 관심, 소통 부재, 정기적

5. IT에 대한 최상층 지원 부족

- 핵심 키워드: 경영진, 리더십, 탑다운, 탑다운, 의사결정, 지원, 투자, 구시대적

11. 직장 내 외로움

- 핵심 키워드: 외로움, 소속감, 개인주의, 삭막, 차가운, 겉도는, 비인간적

6. 불충분한 유연성

- 핵심 키워드: 유연, 재택, 원격, 자율, 출퇴근, 워라밸, 마이크로매니징, 통제

12. 번아웃

- 핵심 키워드: 번아웃, 야근, 업무량, 과도, 혹사, 압박, 스트레스, 휴식, 소진

4-4. 가설검정



4. 우수 인재 유출의 가장 큰 원인은 성장 가능성의 부재일 것이다.

카이제곱 적합도 검정 (Chi-squared Goodness of Fit Test)

이 검정은 **하나의 범주형 변수에 대해 관측된 빈도가 기대되는 이론적 빈도와 얼마나 잘 맞는지를 검증하는** 데 사용됩니다.

수식은 다음과 같습니다.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

전체 12개 사유 간의 빈도 차이가 통계적으로 있을까???

-> 카이제곱 적합도 검정 시행

4-4. 가설검정



4. 우수 인재 유출의 가장 큰 원인은 성장 가능성의 부재일 것이다.

```
1 import numpy as np
2 from scipy.stats import chisquare
3
4 observed_frequencies = np.array([70, 52, 41, 38, 35, 23, 19, 16, 11, 10, 7, 2])
5 chi2_statistic, p_value = chisquare(observed_frequencies)
6
7 print("--- 12개 아직 사유 유형 간 빈도 차이 검증 (카이제곱 적합도 검정) ---")
8 print(f"관측된 빈도: {observed_frequencies}")
9 print(f"총합: {np.sum(observed_frequencies)}")
10 print(f"기대 빈도 (각 유형별): {np.sum(observed_frequencies) / len(observed_frequencies)}")
11 print("-" * 50)
12 print(f"카이제곱 통계량 (Chi-squared Statistic): {chi2_statistic:.4f}")
13 print(f"P-value: {p_value:.4f}")
14 print("-" * 50)
15
16 # 4. 결과 해석
17 alpha = 0.05
18 print(f"유의수준(alpha): {alpha}")
19
20 if p_value < alpha:
21     print("결론: P-value가 유의수준보다 작으므로, 귀무가설을 기각합니다.")
22     print("→ 12가지 아직 사유 유형별 언급 빈도에는 통계적으로 유의미한 차이가 있습니다.")
23     print("→ 즉, 특정 아직 사유가 다른 사유에 비해 더 많이 또는 적게 언급되는 경향이 있으며, 이는 우연이라고 보기 어렵습니다.")
24 else:
25     print("결론: P-value가 유의수준보다 크므로, 귀무가설을 기각할 수 없습니다.")
26     print("→ 12가지 아직 사유 유형별 언급 빈도 차이가 통계적으로 유의미하다고 말할 수 없습니다.")

--- 12개 아직 사유 유형 간 빈도 차이 검증 (카이제곱 적합도 검정) ---
관측된 빈도: [70 52 41 38 35 23 19 16 11 10 7 2]
총합: 324
기대 빈도 (각 유형별): 27.0
-----
카이제곱 통계량 (Chi-squared Statistic): 171.3333
P-value: 0.0000
-----
유의수준(alpha): 0.05
결론: P-value가 유의수준보다 작으므로, 귀무가설을 기각합니다.
→ 12가지 아직 사유 유형별 언급 빈도에는 통계적으로 유의미한 차이가 있습니다.
→ 즉, 특정 아직 사유가 다른 사유에 비해 더 많이 또는 적게 언급되는 경향이 있으며, 이는 우연이라고 보기 어렵습니다.
```

결론:

p-value가 0.05보다 작으므로, 이스트소프트 12개 아직 사유 유형별 키워드 포함된 잡플래닛의 리뷰 사이의 언급 빈도에는 통계적으로 유의미한 차이가 있다는 것을 알 수 있다.

즉, 특정 아직 사유가 다른 사유에 비해 더 많이 또는 적게 언급되는 경향이 있으며, '성장 기회 없음, 관리자의 문제, 기대가 없거나 불분명함, IT에 대한 최상층 지원 부족, 경쟁력 없는 연봉' 등이 높은 빈도에 해당된다.

4-4. 가설검정



5. 직원들은 운영진에게 칭찬보다는 쓴소리를 더 많이 남길 것이다.

KNU 한국어 감성사전

작성자 : 온병원, 박상민, 나철원 작성일 : 2018.05.11 소속 : 군산대학교 소프트웨어융합공학과 Data Intelligence Lab 홈페이지
[: http://dilab.kunsan.ac.kr/](http://dilab.kunsan.ac.kr/)

<> 개요

- 음식, 여행지, 영화, 음악, 자동차, 스마트폰, 강의, 컴퓨터 등 특정 도메인에서 사용되는 긍부정어보다는 인간의 보편적인 기본 감정 표현을 나타내는 긍부정어로 구성된다. 보편적인 긍정 표현으로는 '감동받다', '가치 있다', '감사하다'와 보편적인 부정 표현으로는 '그저 그렇다', '도저히 ~수 없다', '열 받다' 등을 들 수 있다.
- 각 도메인의 감성사전을 빠르게 구축하기 위한 기초 자료로 활용하기 위해 개발되었음
- 본 한국어 감성사전은 다음과 같은 소스로부터 통합되어 개발되었음 (1) 국립국어원 표준국어대사전의 뜻풀이(glosses) 분석을 통한 긍부정 추출(이 방법을 통해 대부분의 긍부정어 추출) (2) 김은영(2004)의 긍부정어 목록 (3) SentiWordNet 및 SenticNet-5.0에서 주로 사용되는 긍부정어 번역 (4) 최근 온라인에서 많이 사용되는 축약어 및 긍부정 이모티콘 목록
- 총 14,843개의 1-gram, 2-gram, 관용구, 문형, 축약어, 이모티콘 등에 대한 긍정, 중립, 부정 판별 및 정도(degree)값 계산

<> 특징

- 표준국어대사전을 구성하는 각 단어의 뜻풀이를 분석하여 긍부정어를 추출하였음
- 1-gram, 2-gram, n-gram(어구, 문형), 축약어, 이모티콘 등의 다양한 종류의 긍부정어 포함
- 영화, 음악, 자동차 등 어떤 도메인에도 사용될 수 있는 보편적인 긍부정어로 구성

분석에 활용한 도구 : KNU 한국어 감성사전

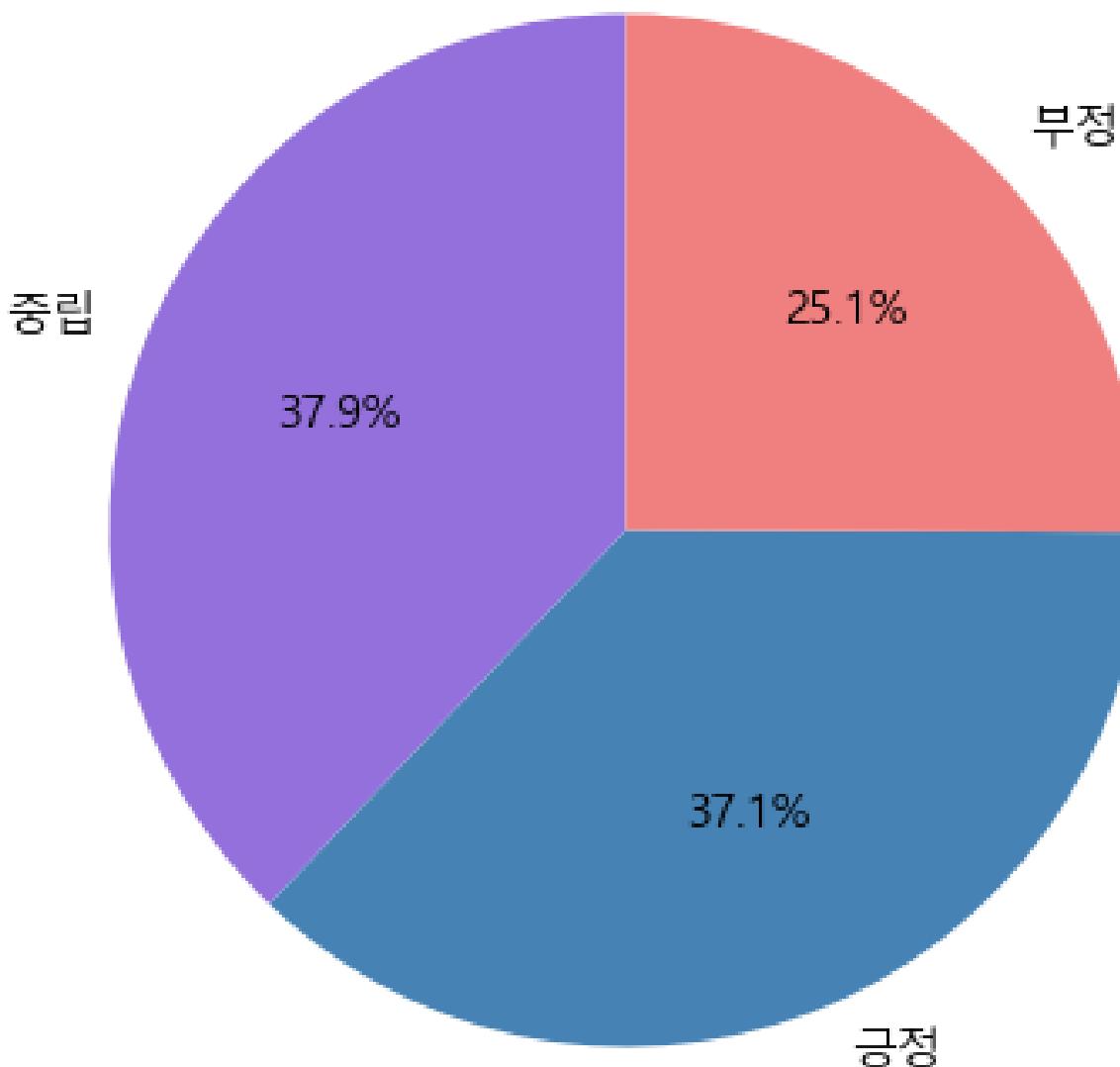
- 긍정/부정/중립 사전을 기반으로 텍스트의 감성을 분류
 - '감동받다, 가치 있다' 등은 긍정으로, '그저 그렇다, 도저히 ~수 없다, 열 받다' 등은 부정으로 분류
- 사전 기반 접근을 통해 리뷰 텍스트를 자동 감성 분류

4-4. 가설검정



5. 직원들은 운영진에게 칭찬보다는 쓴소리를 더 많이 남길 것이다.

경영진에게 하고 싶은 말 감성 분석 비율



“경영진에게 하고 싶은 말” 텍스트만 추출 후, KNU 감성사전을 활용
해 각 리뷰를 긍정 / 중립 / 부정으로 분류

4-4. 가설검정



5. 직원들은 운영진에게 칭찬보다는 쓴소리를 더 많이 남길 것이다.

```
## 이항 검정

from scipy.stats import binomtest

# 긍정 부정만 추출
sent = df[df["경영진_감성_KNU"].isin(["긍정", "부정"])]  
  
n_neg = (sent["경영진_감성_KNU"] == "부정").sum()  
n_pos = (sent["경영진_감성_KNU"] == "긍정").sum()  
n_total = n_neg + n_pos  
  
print("부정:", n_neg, "긍정:", n_pos, "총:", n_total)  
  
# 가설 검정
res = binomtest(n_neg, n_total, p = 0.5, alternative = 'greater')
print("p-value: {:.4f}".format(res.pvalue))

alpha = 0.05 # 유의수준
if res.pvalue < alpha:
    print("직원들은 운영진에게 칭찬보다 쓴소리를 더 많이 남긴다.")
else:
    print("직원들이 운영진에게 칭찬보다 쓴소리를 더 많이 남긴다고 보기 어렵다.")
```

부정: 92 긍정: 136 총: 228

p-value: 0.9986

직원들이 운영진에게 칭찬보다 쓴소리를 더 많이 남긴다고 보기 어렵다.

결과:

분석 결과를 바탕으로 긍정적 발언을 '칭찬', 부정적 발언을 '쓴소리'로 분류하고, 두 비율 간 차이를 통계적으로 검정하기 위해 이항검정(Binomial test, 단측 검정: 부정 > 긍정)을 실시하였다.

검정 결과, $p = 0.9896$ (유의수준 $\alpha = 0.05$) 으로 나타났다.

즉, 실제로는 직원들이 운영진에게 남긴 쓴소리의 비율이 칭찬보다 통계적으로 유의하게 많다고 보기 어렵다.

따라서 본 분석 결과에 따르면, 직원들이 운영진에게 칭찬보다 쓴소리를 더 많이 남긴다는 가설은 통계적으로 지지되지 않는다.

4-5. 모델링



인사이트 후보 도출

1. '야근'이 언급된 리뷰는 전반적으로 부정적인 감정을 띠 것이다.
2. '복지' 관련 리뷰는 다른 주제보다 긍정적인 어휘가 많이 등장할 것이다.
3. '연봉' 관련 리뷰가 부정적인 경향을 보일 경우, 보상 수준이나 성과 대비 급여의 불합리함을 문제로 인식하고 있을 것이다.
4. '소통' 관련 리뷰가 긍정적인 경향을 보일 경우, 협업 분위기나 자유로운 의견 교환을 장점으로 인식하고 있을 것이다.
5. '승진'이 언급된 리뷰는 평가·공정성 관련 단어와 함께 등장하며 부정적인 경향을 보일 것이다.
6. '평가'가 언급된 리뷰는 성과·보상 관련 단어와 함께 등장하며 긍정적인 경향을 보일 것이다.



4-5. 모델링



인사이트 후보 도출



→ 위와 같은 인사이트들에 대해, 특정 키워드를 입력했을 때
의미적으로 유사한 리뷰 집단을 신속히 모아서 그 집단의 감정
경향·공동출현 단어·문맥을 비교 검증하기 위해, KIWI 형태소
분석기와 TF-IDF 벡터화, K-NN 최근접 검색을 결합한 텍스트
검색 모델이 필요하다!

4-5. 모델링



모델 개요



KIWI 형태소 분석기와 TF-IDF 벡터화, K-NN 최근접 검색 알고리즘을 이용해 사용자가 입력한 키워드와 의미적으로 유사한 리뷰를 찾아주는 텍스트 검색 모델

4-5. 모델링



모델 구조

단계	내용
1. 전처리	JobPlanet 리뷰의 제목, 장점, 단점, 경영진에게 하고싶은 말 내용을 한번에 나타내기 위해 하나의 텍스트로 통합
2. 형태소 분석 및 토크나이징	KIWI 분석기로 명사·동사·형용사만 추출하여 핵심 단어 중심 토큰화 수행
3. 불용어 처리	stopwords-iso(최신 공개 불용어 세트로 soynlp보다 성능이 우수)를 활용해 불필요 단어를 제거
4. 벡터화 (TF-IDF)	단어 중요도(TF-IDF)를 이용해 각 리뷰를 수치 벡터로 변환
5. 유사도 계산 (K-NN)	코사인 유사도를 사용해 입력 키워드와 가장 유사한 리뷰를 탐색
6. 하이퍼파라미터 탐색	n-gram, min_df, max_df, max_features 등 256개 조합 중 평균 유사도(mean_sim) 기준으로 최적의 조합 도출

4-5. 모델링



평가 지표



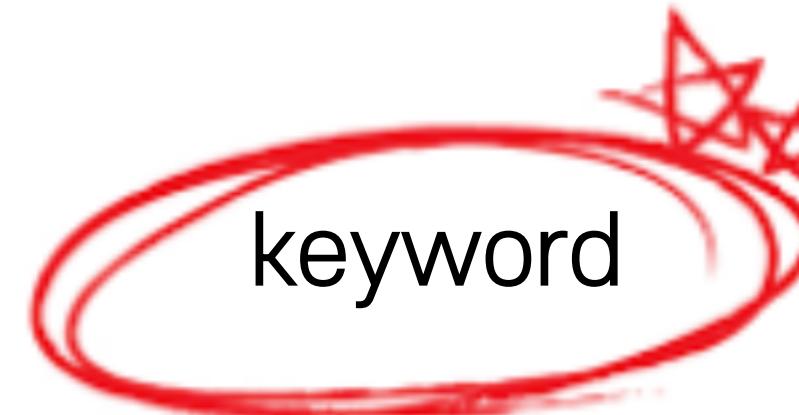
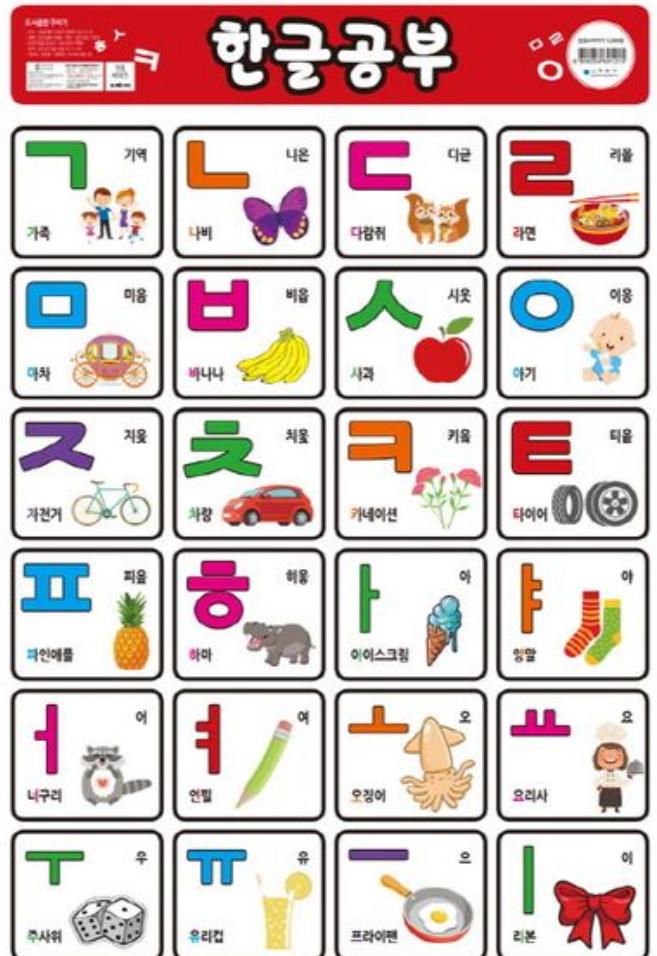
평균 유사도 (Mean Similarity)

- 여러 키워드(‘야근’, ‘복지’, ‘연봉’ 등)에 대해 검색 결과의 코사인 유사도 평균값
- 높을수록 결과들이 일관성 있게 묶였다는 의미

4-5. 모델링



모델 특징 및 장점



- 한국어 특화: KIWI + stopwords-iso로 정제된 텍스트 분석
- 직관적 검색: 키워드 강조([]) 실시간 유사도 탐색 기능
- 확장 가능: BERT 계열 임베딩 모델로 성능 고도화 가능

4-5. 모델링



활용 예시

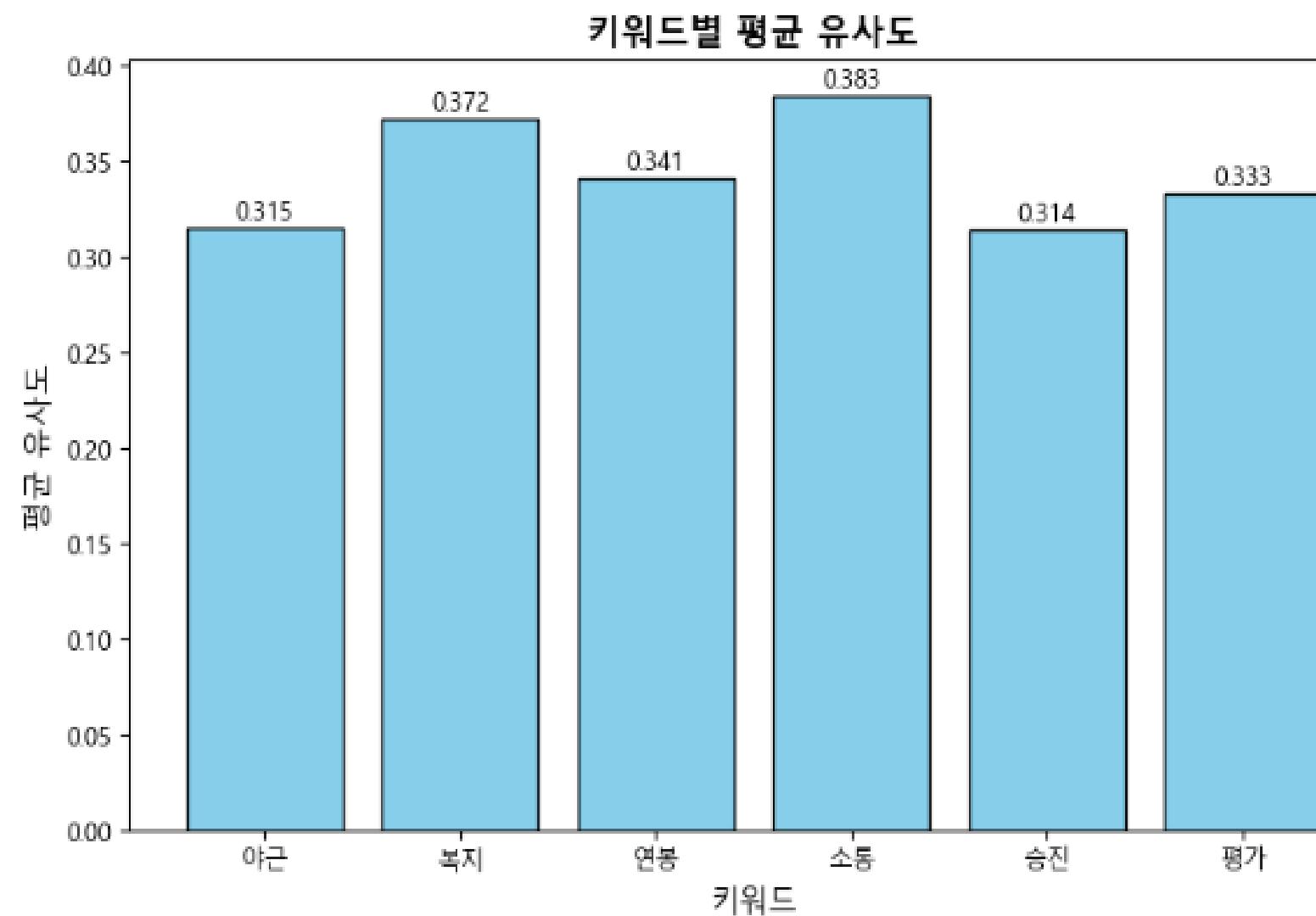
순위	유사도	내용
1	0.406	제목- 개발자 실력들이 좋다. 같이 커갈 수 있는 곳 장점- 사람들이 좋다. 서실 이게 제일 중요하다고 본다. 동료가 좋아야 일할 맛도 나니까 단점- 좀 일이 많다. 요샌 어떤지 모르겠는데 예전엔 [[야근]][[야근]]의 연속이었다. 경영진에게 하고싶은 말- 글쎄 딱히 바라는건 없고 일을 좀 줄였으면 한다.
2	0.348	제목- 어느 개발사가 그렇듯 개발자가 갑인 회사 정치알력싸움이 상당한 회사 장점- [[야근]]하면 다음날 지연출근 제도가 있습니다 [[야근]]을 많이 했었죠 단점- 나이 많아지면 자동 승진인데 뭣하려 이런 걸 만들었는지 알 수 없음 경영진에게 하고싶은 말- 너무 돈 줄일 생각만 하지 말고 인프라에도 투자를 좀 해 주세요
3	0.343	제목- 개발자 중심. 개발자를 잘 대우함. 비람만 잘 타면 크게 될 곳 장점- 실력 우선주의. 코딩 테스트로 연봉 점프 가능. 개발자는 괜찮은 경력 단점- [[야근]]이 좀 많다?? 어쩔 수 없는 상황이지만 [[야근]]이 많았음. 경영진에게 하고싶은 말- 일정은 낙관하게. 다 좋음. 앞으로도 잘 될거 같음
4	0.328	제목- 직원 연령도 어리고 자유로운 분위기의 회사. 하지만 실적의 압박과 [[야근]]이 단점입니다. 장점- 스펙을 정말 안보기때문에 직원 연령도 어리고 비교적 자유로운 분위기입니다. 그리고 사람들도 친절하고 착합니다. 단점- ...실적의 압박이 많습니다. 상사에게 밀보이면 매일 갈궈지고 그때문에 퇴사한 직원도 꽤 있습니다. 그리고 부서에 따라 다르겠지만 [[야근]]이 일상입니다. 경영진에게 하고싶은 말- 구체적인 프로세스가 구축되어 있어서 일의 효율성이 높았으면 좋겠습니다.
5	0.305	제목- 개발자를 높이 쳐주는 곳 실력이 [[야근]]없이 워라밸이 살아있느 회사 장점- 편한 회사 수평적인 조직 개발자가 다른 직군에 비해 연봉이 높음 단점- 공무원같이 경직된 업무환경 경영진에게 하고싶은 말- 신규 사업을 엉뚱한 쪽으로 하는 것 같아서 회사가 점점 더 어려워지는 느낌

사용자가 “야근”을 검색하면 ‘야근 관련 언급이 포함된 리뷰’를 유사도 순으로 정렬해 출력
→ 키워드별 감정 분석이나 주제별 리뷰 탐색 등에 활용 가능

4-5. 모델링



모델 평가



“복지”와 “소통” 키워드의 평균 유사도가 상당히 높은 수치로 나와 리뷰 내에서 관련 문맥을 가장 일관되게 탐색했음을 확인할 수 있다.

“야근”과 “승진” 키워드는 유사도는 상대적으로 낮아 리뷰 간 표현 다양성이 크거나 언급 빈도가 낮음을 시사한다.

전반적으로 평균 유사도 0.31 ~ 0.38 수준으로, TF-IDF 기반 모델로서는 안정적인 의미 탐색 성능을 보인다.

4-5. 모델링



모델링 활용 및 응용 방안



1. 회사 홍보 '워라밸', '좋은', '복지' 등에 키워드를 검색했을 때 보이는 긍정적인 반응을 활용하여 회사의 브랜드 가치를 올리고 새로운 구직자들에게 매력을 어필할 수 있음.

2. 회사 실태 파악 각종 키워드(복지, 워라밸, 퇴사 등등)을 검색하여 회사에 대한 문제점과 개선점을 파악할 수 있음.

4-5. 모델링



모델링 활용 및 응용 방안



3. 유사한 의견을 한 눈에 파악할 수 있음 유사한 리뷰를 그룹으로 묶어 해당 주제에 대한 의견들을 보여주는 시각적인 대시보드를 구축해 사용자가 원하는 정보를 한 눈에 파악 가능하게 만들 수 있음.

4. 이직 및 퇴사 위험 관리 이직 및 퇴사와 관련된 키워드를 검색한 빈도를 측정(가능하다면)하여 이직이나 퇴사에 대한 위험도를 측정할 수 있음. 만약 위험도가 높다면 리뷰를 통해 어떤 점에서 불만이 있는지 파악할 수도 있고 개선점도 도출할 수 있음.



4-5. 모델링



기대효과



1. 분석시간 및 비용 절감 수많은 데이터를 일일이 읽고
분류할 필요가 없어 데이터 분석에 소요되는 시간과
인력을 절감할 수 있음.

2. 정성적 데이터 활용을 통한 의견 파악 가능 단순한
키워드 분석을 넘어 문맥적 유사도까지 검색 가능하니
직원들의 의견을 구체적으로 살펴볼 수 있음. 이를 통해
정성적으로 데이터를 깊게 분석 할 수 있음.

4-5. 모델링



기대효과



3. 데이터 기반을 활용한 유사 사례 방지 특정 이슈가 발생시, 과거에 비슷한 리뷰를 신속하게 찾아 유사 사례를 방지 및 대응할 수 있음.
즉 위험 요소에 대해 즉각적인 의사 결정이 가능해짐.

감사합니다

