

T-SENT

NLP project 곽민수

시나리오

코로나 이후 매출이 줄어 걱정이예요. 새로운 서비스의 제공과 변화가 필요해요.

인간의 감성을 터치하는 건 어떨까요?
코로나라는 위기의 상황이 또 다른 기회가 될 수 있어요.

MK

비밀
영향
상황
확언
기쁨

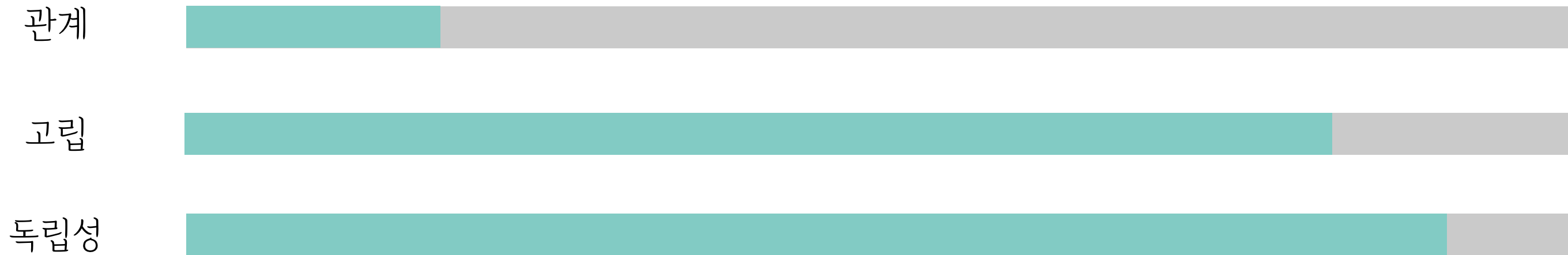
Sentiment

마음을 터놓는

관계의 단절

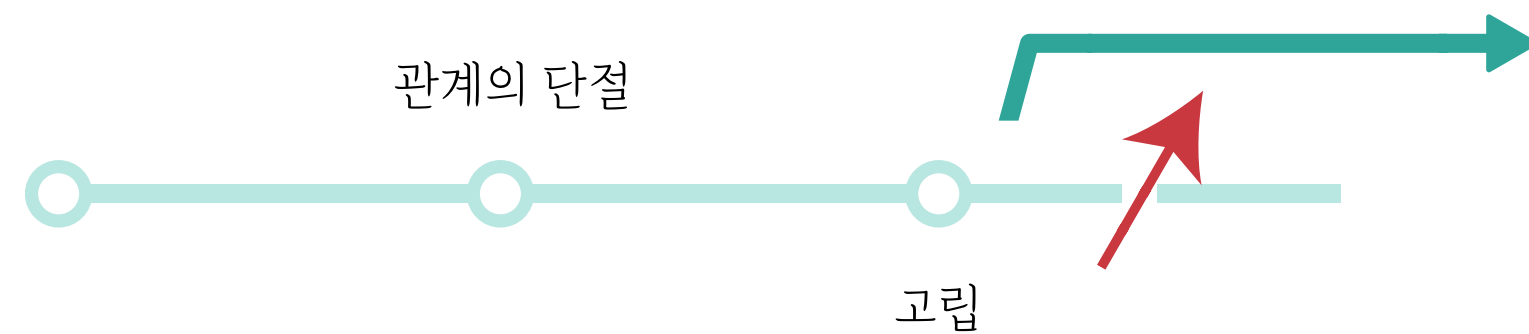
DS

2022년 현황 (Post-Corona)



NOTE

- 마음을 터놓치 못함
- 기계에 익숙해지는 인간

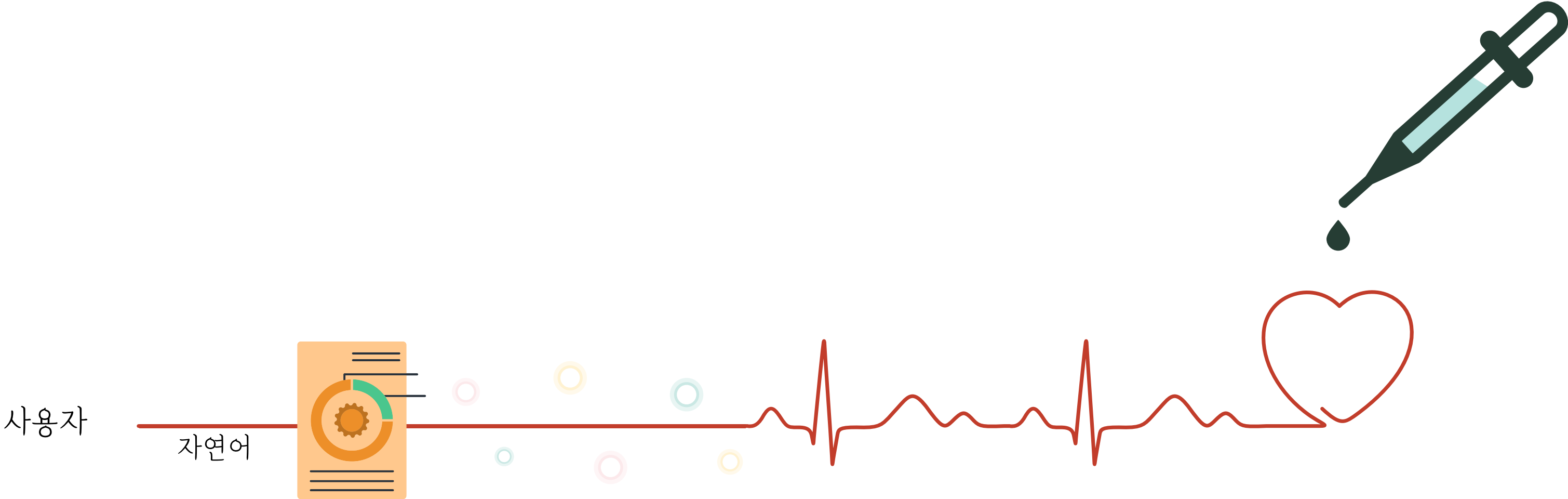




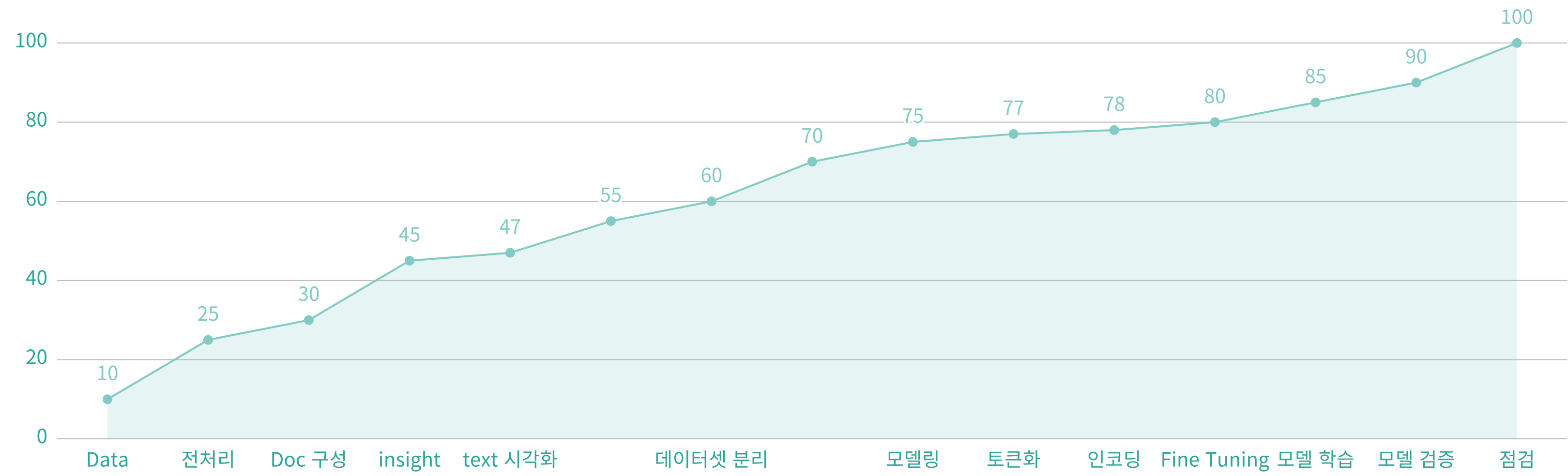
시나리오

New Shift

- 자신의 속 마음을 털어놓을 수 있는 서비스
- 기계를 사용하여 부담감 완화
- 자연어를 통한 감정 분석
- 개인형 맞춤형 티 서비스 제공



PIPE-LINE



DATA AI HUB: 감성대화 말뭉치

	번호	연령	성별	상황키워드	신체질환	감정_대분류	감정_소분류	사람문장1	시스템응답1	사람문장2	시스템응답2	사람문장3	시스템응답3	사람문장4	시스템응답4
0	44164	청년	남성	연애, 결혼, 출산	해당 없음	기쁨	신이 난	아내가 드디어 출산하게 되어서 정말 신이 나.	아내분이 출산을 하시는군요. 정말 축하드려요.	아 지금 정말 신이 나.	잘 된 일이네요.	아기가 점점 클게 벌써 기대가 되네. 내가 많이 높아줘야지.	좋은 아빠가 되실 거 같아요. 진심으로 축하드려요.	NaN	NaN
1	3926	노년	남성	건강, 죽음	만성질환유	불안	스트레스 받는	당뇨랑 합병증 때문에 먹어야 할 약이 열 가지가 넘어가니까 스트레스야.	약 종류가 많아 번거로우시겠어요.	건강할 때 관리 좀 잘할 걸 하는 생각이 들더라고요.	현재 상황에서 변화를 주기 위해 어떻게 하면 될까요?	약을 잘 챙겨 먹고 나을 수 있도록 노력해야지.	약을 잘 챙겨 먹어 건강해지시길 바랄게요.	NaN	NaN
2	50882	청소년	여성	학업 및 진로	해당 없음	당황	당황	고등학교에 올라오니 중학교 때보다 수업이 갑자기 어려워져서 당황스러워.	고등학교 수업이 중학교 때와는 다르게 갑자기 어려워서 당황스러우시군요.	아직 학기 초인데 내가 수업에 잘 따라갈 수 있을지 걱정돼.	이 상황에서 할 수 있는 게 어떤 것이 있을까요?	중학교 수업을 들을 때보다 훨씬 더 열심히 할 거야.	수업을 어떤 방식으로 열심히 듣는게 좋을까요?	선생님이 강조하신 부분을 필기하고 집에서 매일 수업 내용을 복습하려고 해.	열심히 해서 좋은 결과가 있었으면 좋겠어요.

Training Set (40879,15) --> Train_df (40879, 9)

Validation Set (5130,15) --> Valdidation_df(5130, 9)

- 번호
- 인간에 대한 정보 : 연령, 성별, 신체질환
- 상황 키워드
- 감정 데이터 : 대분류, 소분류
- 사람문장: 1, 2, 3, 4
- 시스템 문장 : 1, 2, 3, 4

Hypothesis

가설1. 성별에 따라 우세하는 감정이 다를 것이다.

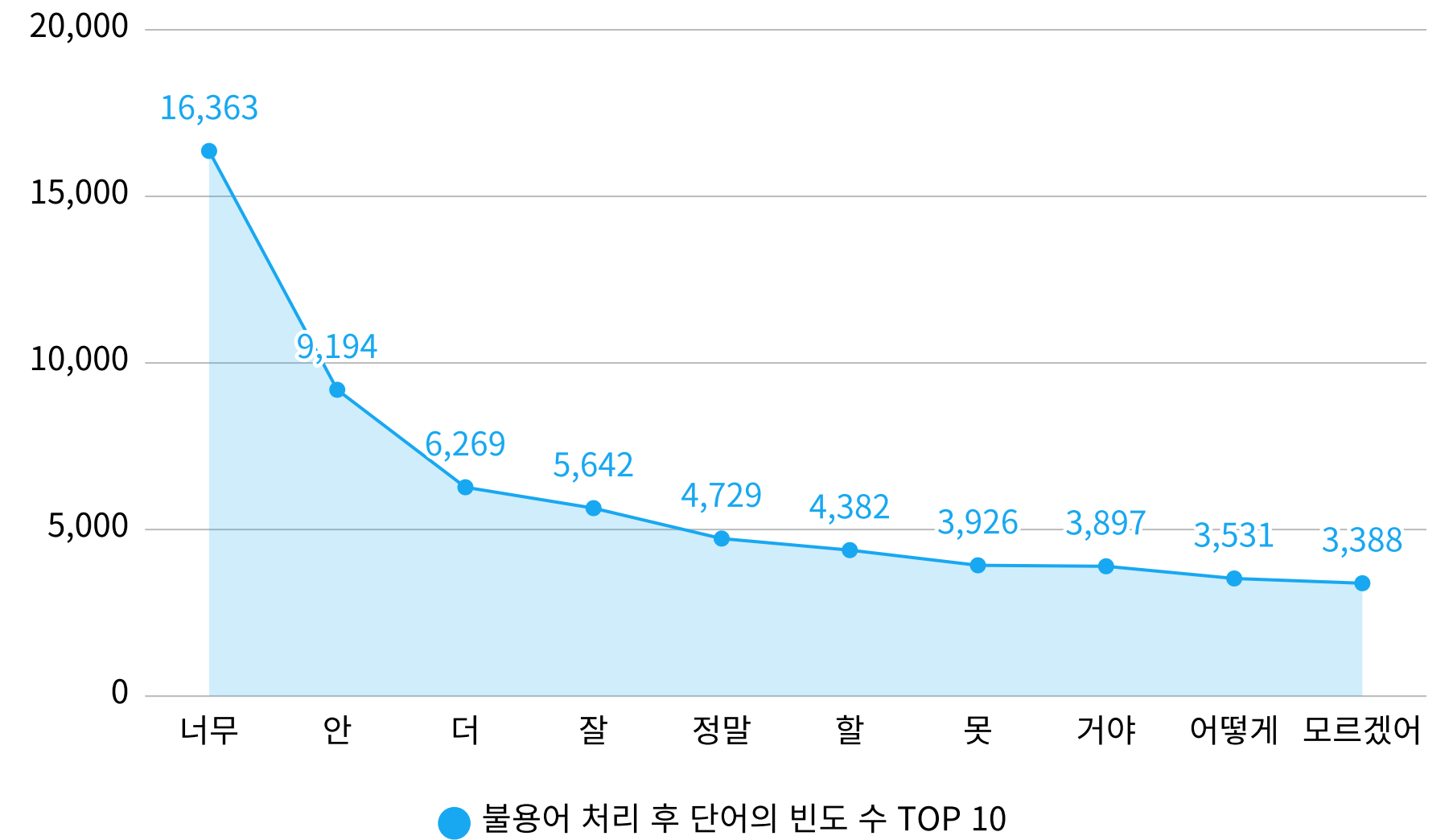
가설2. 연령에 따라 감정의 분포도 다를 것이다.

시각화를 위해 합친 데이터셋 shape: (46009, 9)

	번호	연령	성별	상황키워드	신체질환	감정_대분류	감정_소분류	human_document	system_document
0	44164	청년	남성	연애, 결혼, 출산	해당없음	기쁨	신이 난	아내가 드디어 출산하게 되어서 정말 신이 나. 아 지금 정말 신이 나. 아기가 점점...	아내분이 출산을 하시는군요. 정말 축하드려요. 잘 된 일이네요. 좋은 아빠가 되실 ...
1	3926	노년	남성	건강, 죽음	만성질환 유	불안	스트레스 받는	당뇨랑 합병증 때문에 먹어야 할 약이 열 가지가 넘어가니까 스트레스야. 건강할 때 ...	약 종류가 많아 번거로우시겠어요. 현재 상황에서 변화를 주기 위해 어떻게 하면 될까...
2	50882	청소년	여성	학업 및 진로	해당없음	당황	당황	고등학교에 올라오니 중학교 때보다 수업이 갑자기 어려워져서 당황스러워. 아직 학기 ...	고등학교 수업이 중학교 때와는 다르게 갑자기 어려워서 당황스러우시군요. 이 상황에서...
3	31303	노년	남성	재정	만성질환 무	기쁨	신이 난	재취업이 돼서 받게 된 첫 월급으로 온 가족이 외식을 할 예정이야. 너무 행복해. ...	재취업 후 첫 월급이라 정말 기쁘시겠어요. 이 기분을 유지하려면 어떻게 해야 할까요...
4	47200	노년	여성	재정	만성질환 유	기쁨	안도	빚을 드디어 다 갚게 되어서 이제야 안도감이 들어. 빚도 다 갚았으니 당분간은 아무...	기분 좋으시겠어요. 앞으로는 어떻게 하실 생각이신가요? 빚도 다 갚았으니 당분간은 ...
...
46004	37859	중년	남성	재정, 은퇴, 노후준비	해당없음	기쁨	느긋	부동산 임대 소득으로 현재 여유롭게 살 수 있어서 좋다. 그래서 마음의 안정을 가...	여유롭게 생활할 수 있어 좋으시겠어요. 그렇군요. 계속해서 여유로우려면 어떻게 해야 좋...
46005	40171	노년	여성	건강, 죽음	만성질환 유	불안	회의적인	폐결핵은 이미 완치된 것 같은데 약을 한 달이나 더 먹으라고 하네? 아직 안 나온 ...	치료가 완료되지 않은 게 아닐까 싶으셔서 염려하고 계시는군요. 많이 불안하시군요. ...
46006	13717	청년	남성	연애, 결혼, 출산	해당없음	분노	좌절함	연애하고 싶은데 소개팅만 나가면 광이야. 이러다가 난 결혼 못 하고 늙어 죽을 거야...	소개팅에서 만나는 사람과 계속 잘 안 풀려서 많이 좌절하신 것 같아요. 댄스 동호...
46007	33784	노년	남성	재정	만성질환 유	분노	안달하는	은행 대출이 막혀서 생활비를 구할 수가 없어. 이제 어떻게 살아야 하나 막막해. 당...	은행 대출이 안 돼서 마음이 괴로우시군요. 이 상황을 해결하기 위해서는 어떻게 할 ...
46008	43490	노년	남성	재정	만성질환 유	분노	짜증내는	자식들은 키워 준 은혜도 모르고 내게 오지도 않네. 너무 외롭고 슬퍼. 응 이제 다...	자녀들이 찾아오지 않아서 슬퍼하시는 것 같아요. 특별한 날뿐만 아니라 평소에도 자녀...

46009 rows x 9 columns

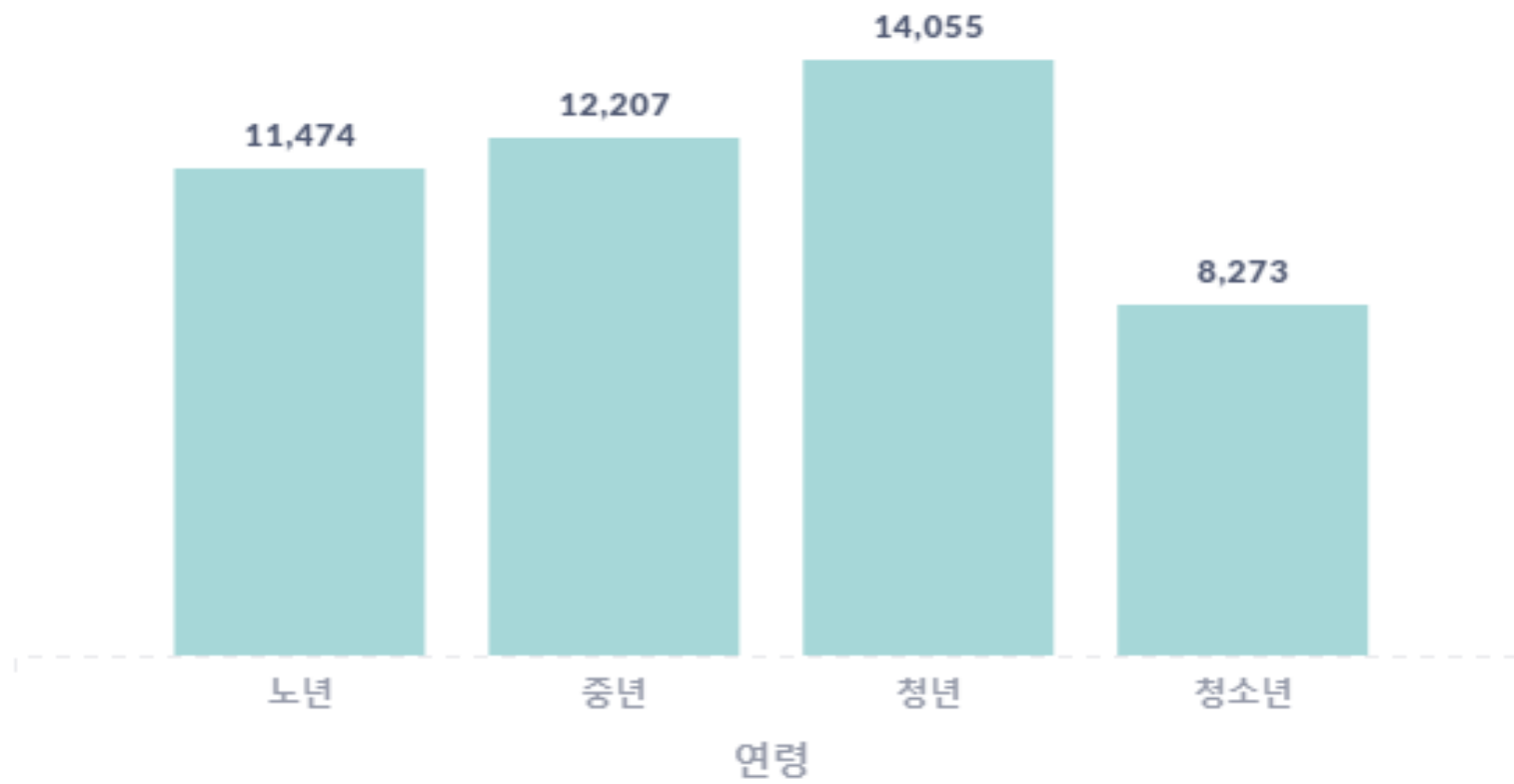
INSIGHT



DATA

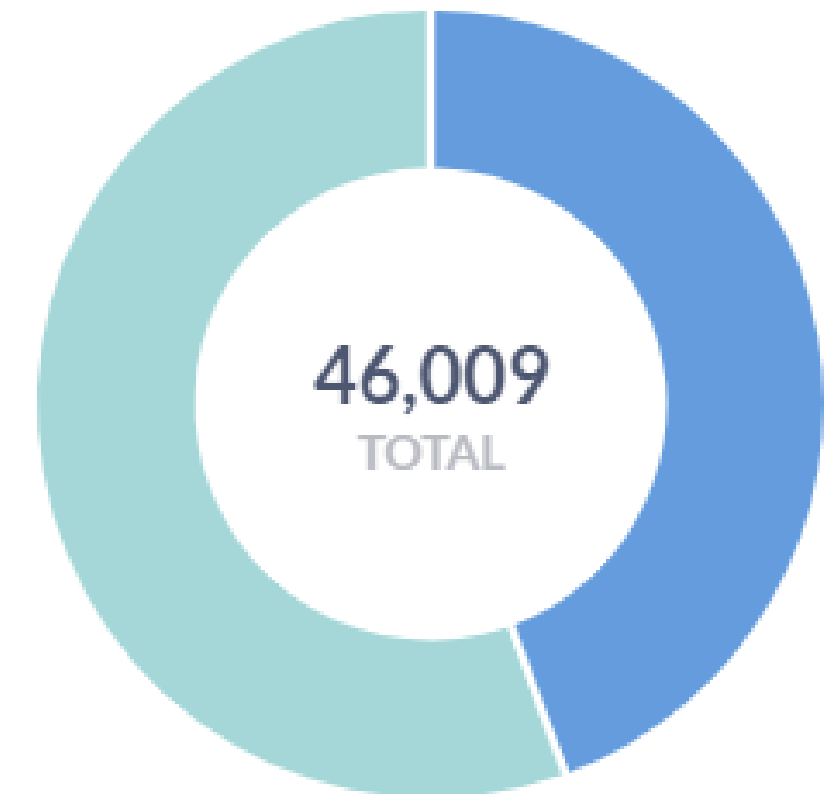
- 청년> 중년> 노년> 청소년 순 데이터 포함
- 비교적 균일하므로 감성분석 진행

연령 별 분포



성별 분포

● 남성
● 여성

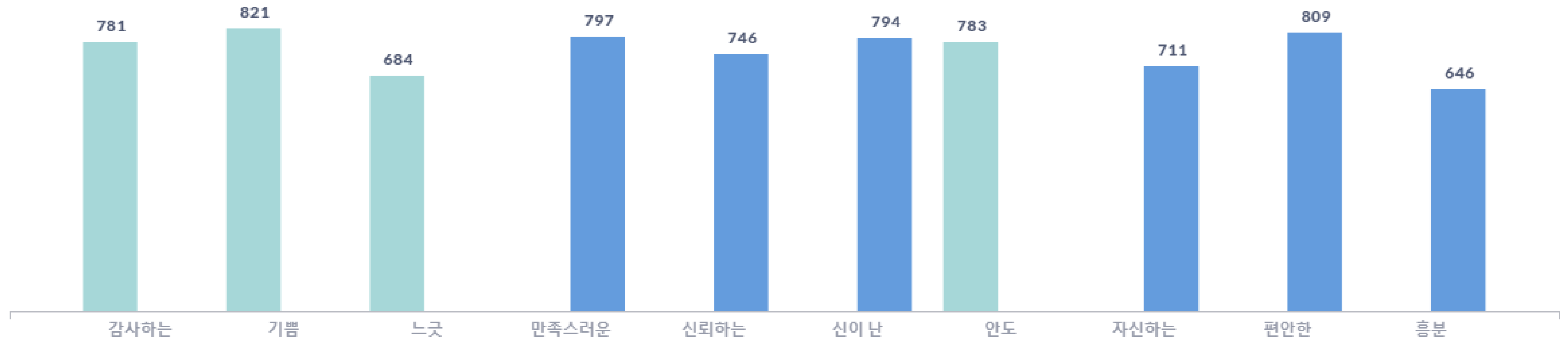


INSIGHT

감정_소분류

성별에 따른 감정 분포

● 여성 ● 남성

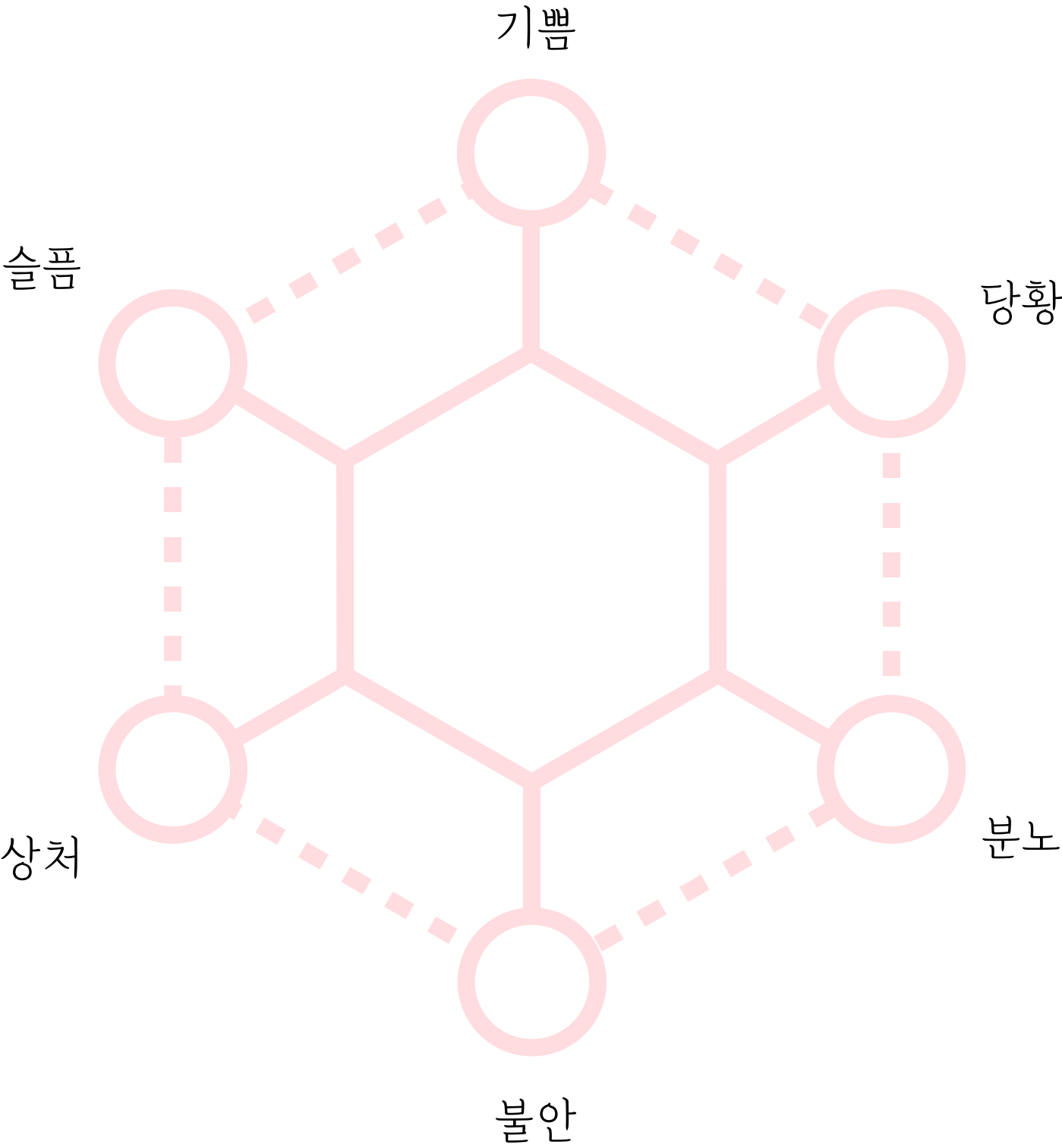


INSIGHT

감정_ 소분류

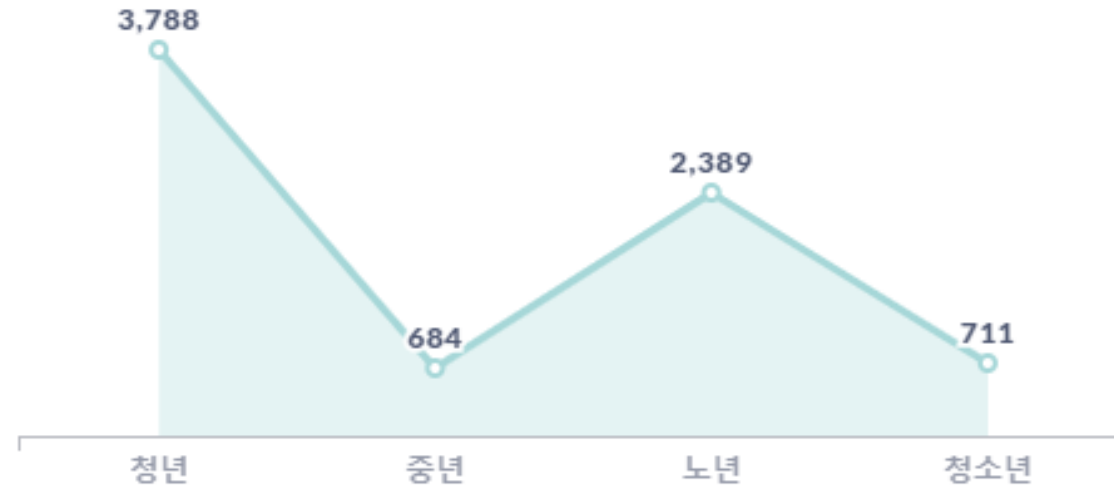
- 약 60개의 카테고리
- 세세한 분석은 가능하나 불필요함
- 감정_대분류에 의한 분류가 더 적합함

'신이 난' '스트레스 받는' '당황' '안도' '취약한' '우울한' '구역질 나는' '좌절한' '눈물이 나는' '배신당한', '남의 시선을 의식하는' '죄책감의' '당혹스러운' '분노' '두려운' '낙담한'
'회의적인' '마비된' '혼란스러운' '흥분', '한심한' '염세적인' '안달하는' '걱정스러운' '조심스러운' '충격 받은' '상처' '불안' '노여워하는' '기쁨' '짜증내는', '환멸을 느끼는' '비통한'
'혐오스러운' '억울한' '만족스러운' '부끄러운' '버려진' '고립된' '괴로워하는' '자신하는' '외로운' '방어적인' '편안한' '질투하는' '감사하는' '초조한' '실망한' '슬픔' '신뢰하는'
'가난한, 불우한' '후회되는' '툭툭대는' '성가신' '열등감' '희생된' '느긋' '악의적인'

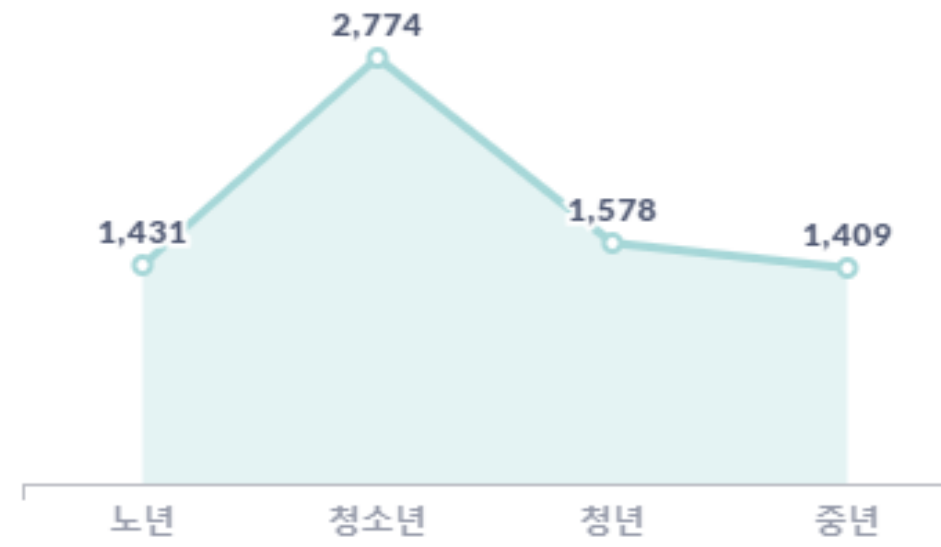


INSIGHT

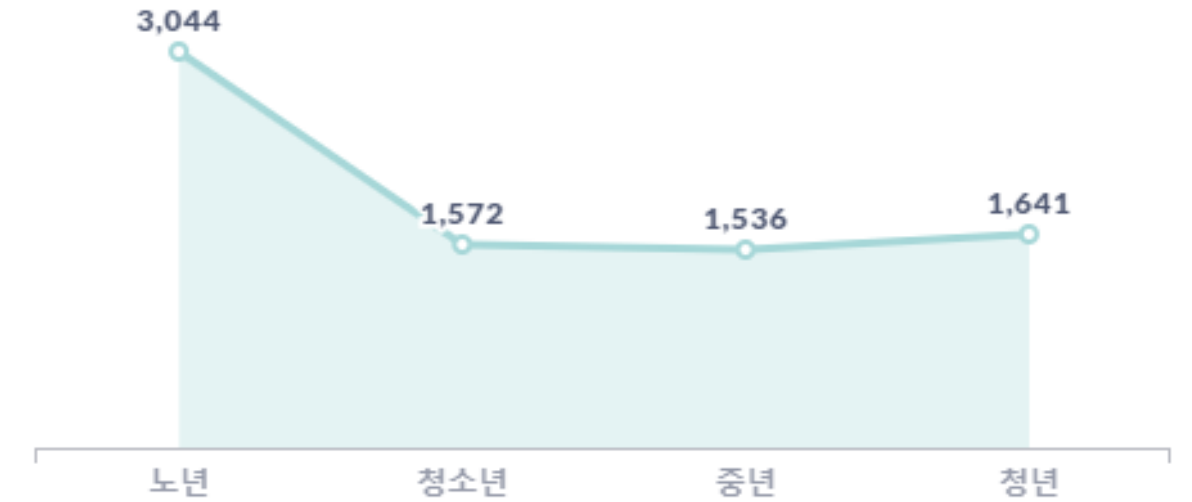
기쁨 지수 (연령)



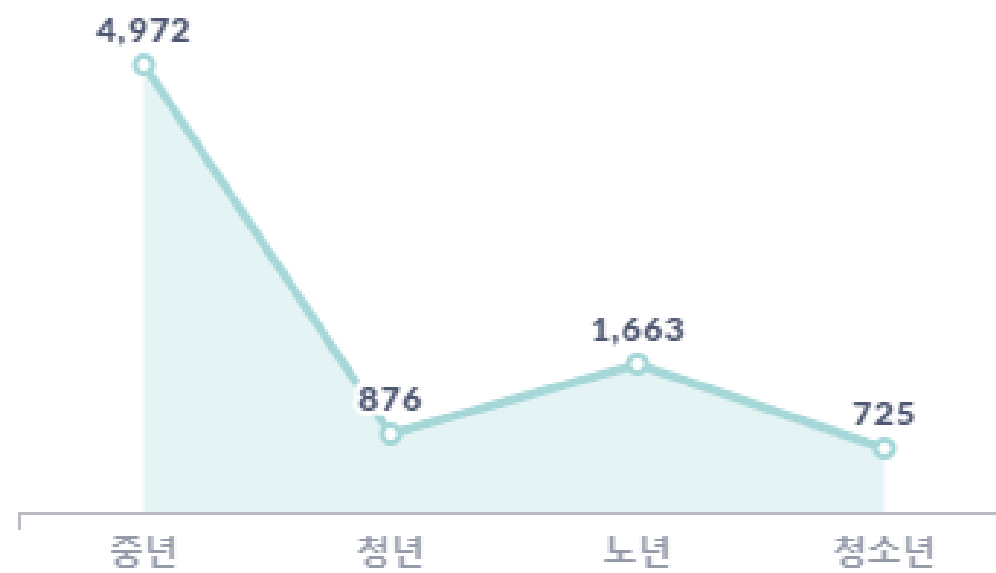
당황 지수 (연령)



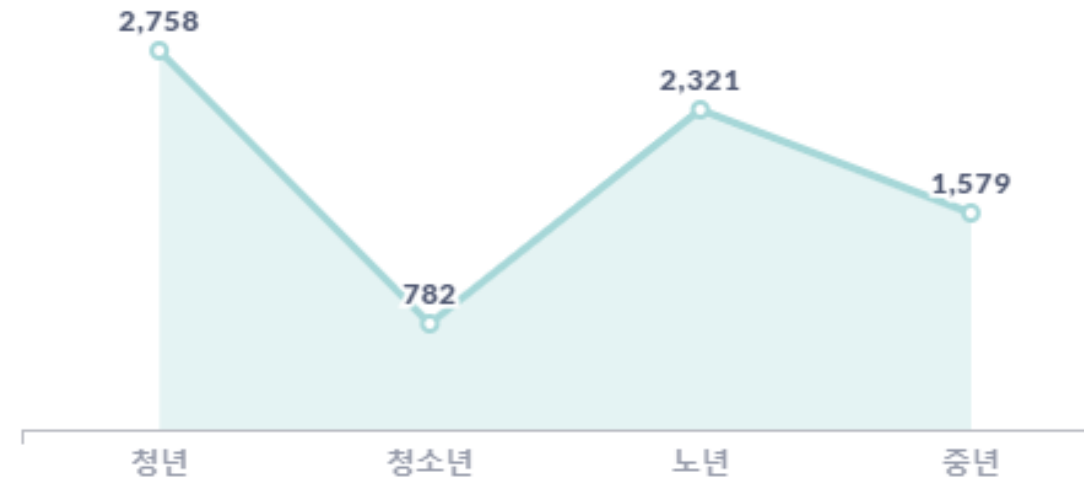
분노 지수 (연령)



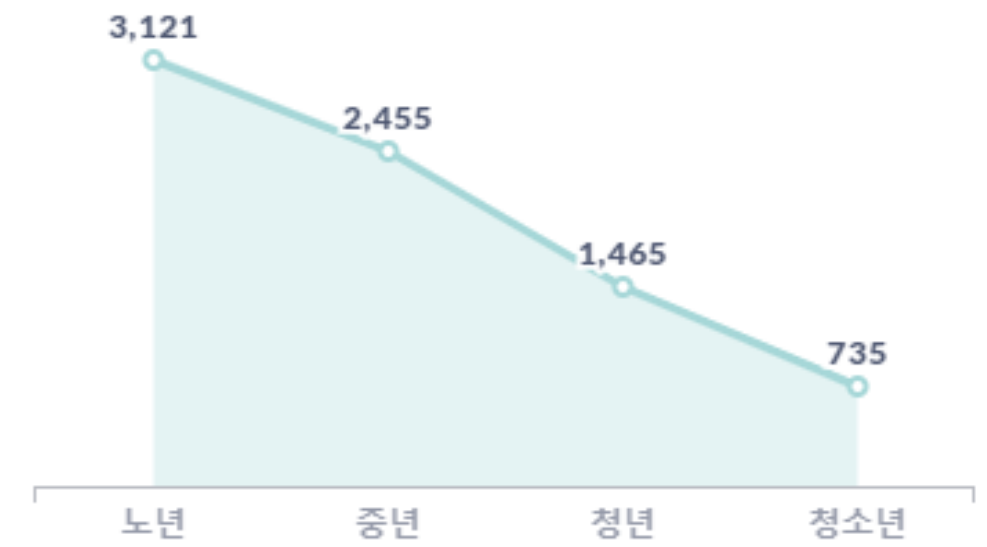
불안 지수 (연령)



상처 지수 (연령)

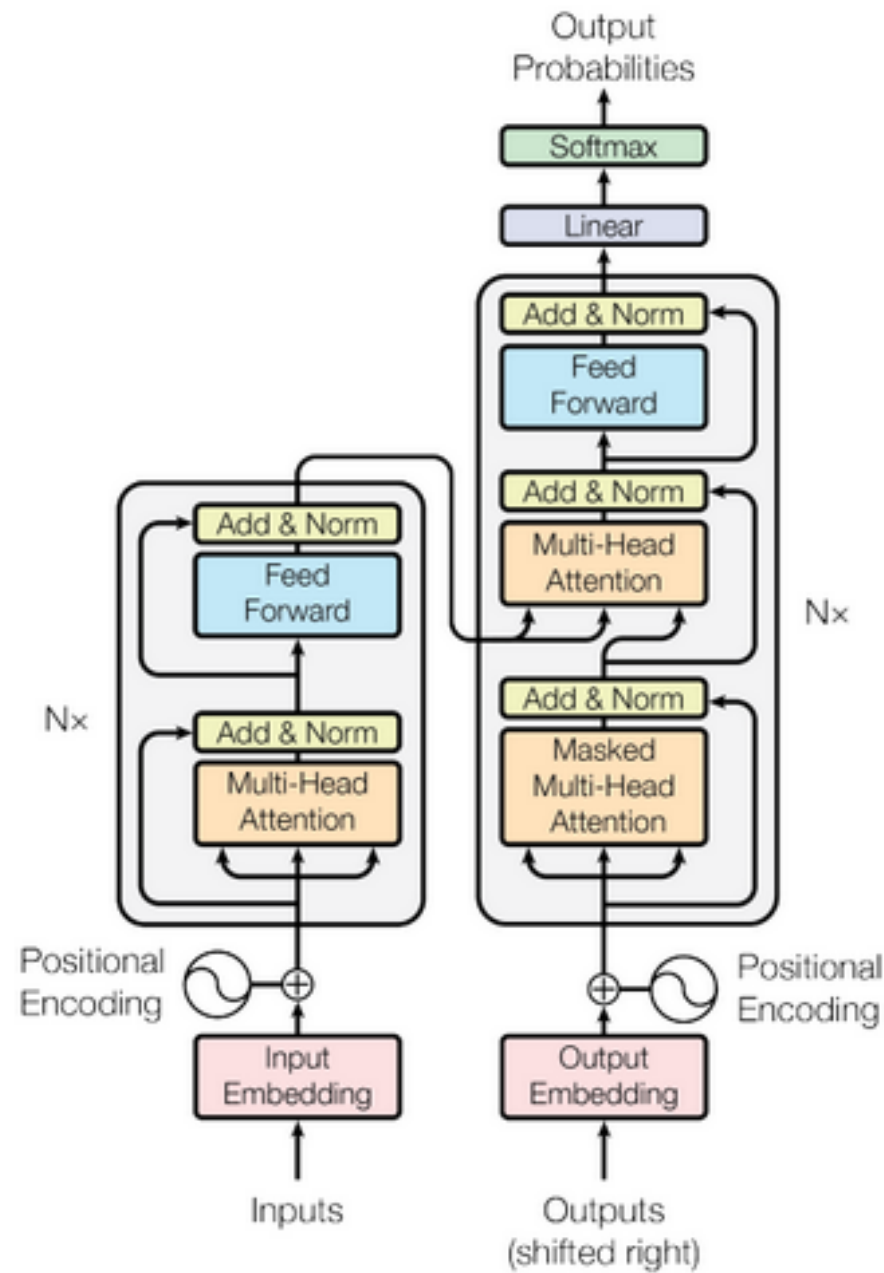


슬픔 지수 (연령)



연령 별 강세한 감정이 상이함을 통해 타겟을 구체화 할 수 있으며, 그에 적합한 티 제조에 참고 가능

MODEL : BERT



TRANSFORMER 구성도

TRANSFORMER를 이용하면서 ENCODING(좌측) 부분만 사용하는 사전학습모델

- 자연어 처리의 대가
- 사전 훈련 언어 모델
- 자연어 처리 부분에서 좋은 성능

- 문맥 반영 (보통의 텍스트 코퍼스만 이용해 교육 없이 양방향 처리)

동작 방식: 사전학습 (BERT 모델) + Fine Tuning (추가적인 모델)을 쌓아 Task 수행



BERT 모델 학습: Fine Tuning

	output
global_step	5110
training_loss	1.7819

metrics	output
train_runtime	2999.6767
train_samples_per_second	54.511
train_steps_per_second	1.704
total_flos	1.0756
train_loss	1.7819
epoch	5.0

BERT 모델 학습 결과 (검증)

===== Epoch 1/5 =====

평균 학습 loss : 1.7749110460281372

검증 정확도 : 0.2598876953125

===== Epoch 2/5 =====

평균 학습 loss : 1.7740951776504517

검증 정확도 : 0.259033203125

===== Epoch 3/5 =====

평균 학습 loss : 1.7744091749191284

검증 정확도 : 0.2584228515625

===== Epoch 4/5 =====

평균 학습 loss : 1.7735663652420044

검증 정확도 : 0.2584228515625

===== Epoch 5/5 =====

평균 학습 loss : 1.7738864421844482

검증 정확도 : 0.2584228515625

- 본 모델은 6개의 감정 카테고리로 감정 분석을 진행한다.
- CHANCE LEVEL : $100 / 6 = 16.66 \rightarrow 0.16$
- 모델 학습의 검증 정확도는 0.25 로
CHANCE LEVEL인 0.16 을 상회한다.

BERT 모델 학습 결과 (테스트셋)

테스트 정확도 : 0.25846273291925465

- CHANCE LEVEL : $100 / 6 = 16.66$ --> 0.16
- 테스트 정확도 역시 검증 데이터셋과 유사하게 0.258로 나타난다.
- CHANCE LEVEL 인 0.16을 넘어선다.

REFLECTION

- 불용어 처리 : BERT 모델 Tokenizer 아닌 직접 Tokenizing 할 때 불용어 처리가 쉽지 않았음.
 - 불용어를 일일이 처리해야 한다는 점
- 감정 분류에 대하여: 감정 카테고리
- 모델 과정에 대하여: 오류와의 전쟁
- Bert 모델을 처음 접한 것의 한계
 - 진행 과정 중에도 잘 하고 있는지 확실하지 않아 어려움을 겪음
 - 그러나 새로운 것을 시도하며 배우는게 많았음
 - 가장 문제였던 것은 거듭되는 error에 Trouble shooting 하는 것이었지만,
오히려 Trouble shooting이 과정을 이해하는 것을 도왔다 생각됨.
- Bert 모델이 문맥을 반영하기 때문에 doc 부분을 결합하는데 있어 고민이 됐음.
 - 추후 수정이 필요하다면 데이터셋을 수정할 것임

Tea에 마음을 더하다

T-SENT