



# 2팀 대회 회고

팀원 김희숙 이수연  
한예송

# 목차

## #01 쇼핑몰 리뷰 평점 분석 시각화 경진대회

- 대회 소개
- 분석 및 인사이트
- 1등 분석 리뷰

## #02 쇼핑몰 지점별 매출액 예측 경진대회

- 대회 소개
- 분석 및 인사이트
- 1등 분석 리뷰

## #03 쇼핑몰 지점별 매출액 분석 시각화 경진대회

- 대회 소개
- 분석 및 인사이트
- 1등 분석 리뷰



# 쇼핑몰 리뷰 평점 분석 시각화 경진대회



# # 대회 소개

## 대회 개요

- 쇼핑몰 리뷰 데이터 셋을 이용하여 인사이트를 발견하고, 쇼핑몰 상품 리뷰 및 평점을 효과적으로 분석 시각화

## 심사 기준

구분	영역	평가항목	배점
기술력 (35)	분석	데이터 특성에 따른 적합한 분석 기법을 활용하였는가?	5
	시각화	데이터 분석 결과를 적절한 시각화 방법으로 효과적으로 시각화했는가?	20
	논리	시각화 콘텐츠의 기승전결이 잘 구성되어 있는가?	5
	코드	코드의 가독성을 고려하며, 주석이 성실하게 작성되었는가?	5
데이터 (30)	활용	다양한 데이터를 활용하였는가?	20
	연계	데이터 간의 융합 및 <b>feature engineering</b> 을 적절하게 시도했는가?	10
인사이트 (35)	창의	다른 사람들이 생각하지 못한 새로운 관점을 제시하였는가?	20
	가독성	전문지식이 없어도 이해하기 쉬운가?	15

## Data

- id : 샘플 아이디
- reviews : 쇼핑몰 리뷰 텍스트
- target : 상품 평점

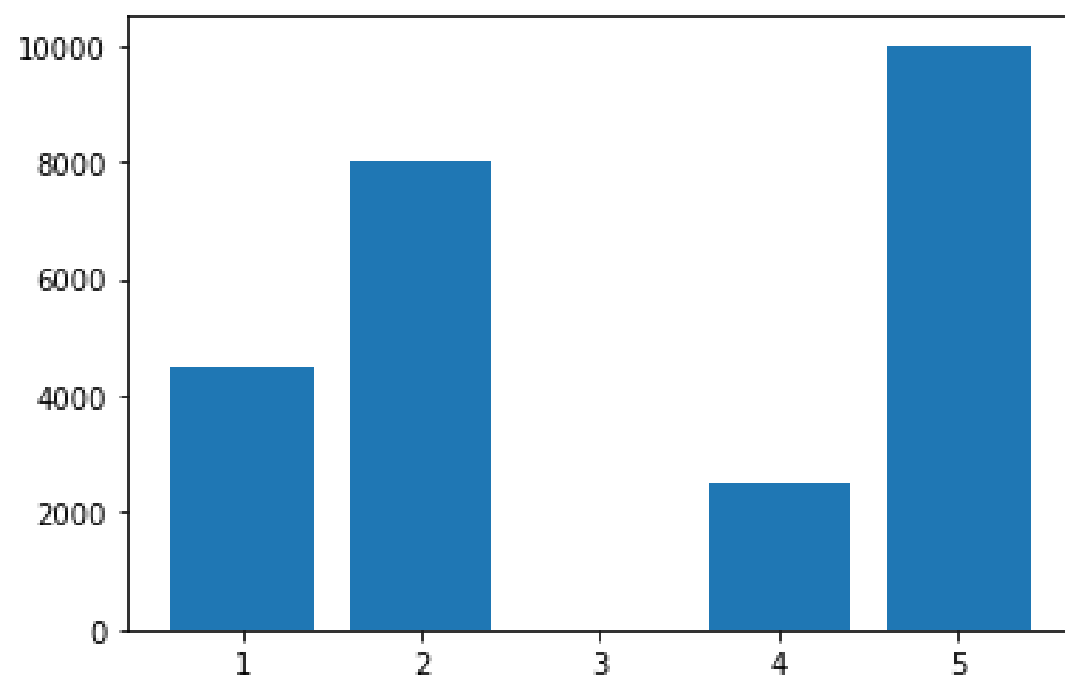
```
data.head()
```

	id	reviews	target
0	0	조아요 처음구입 싸게했어요	2
1	1	생각보다 잘 안돼요 매지 바른지 하루밖에 안됐는데ㅠㅠ 25천원가량 주고 사기 너무 ...	1
2	2	디자인은괜찮은데 상품이 금이가서 교환했는데 두번째받은상품도 까져있고 안쪽에 금이가져...	2
3	3	기전에 이 제품말고 이마트 트레이더스에서만 팔던 프리미엄 제품을 사용했었습니다. 샘...	2
4	4	튼튼하고 손목을 잘 받쳐주네요~	5

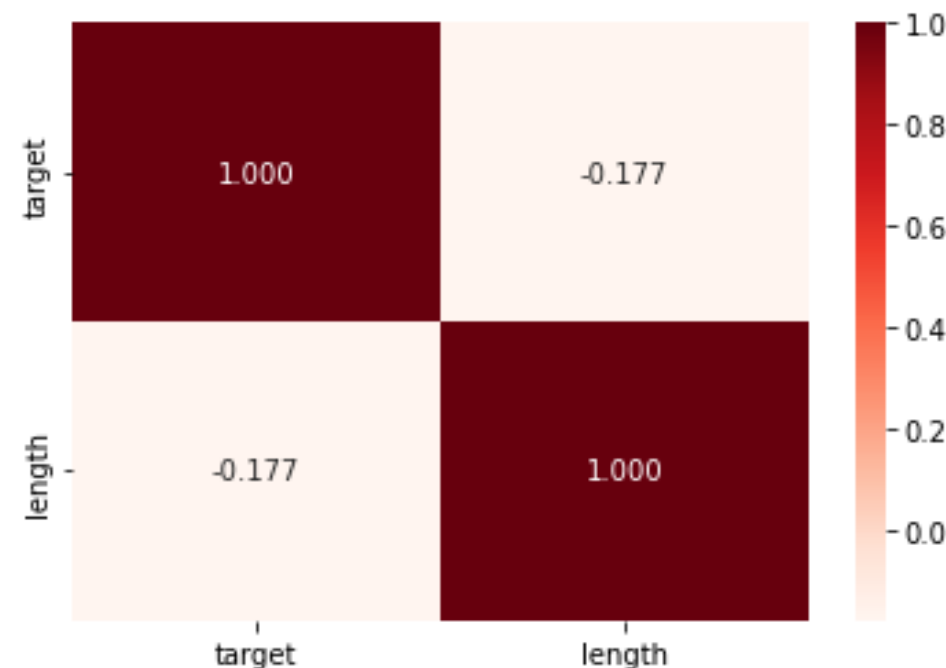
- (25000, 3)

# # 분석 및 인사이트

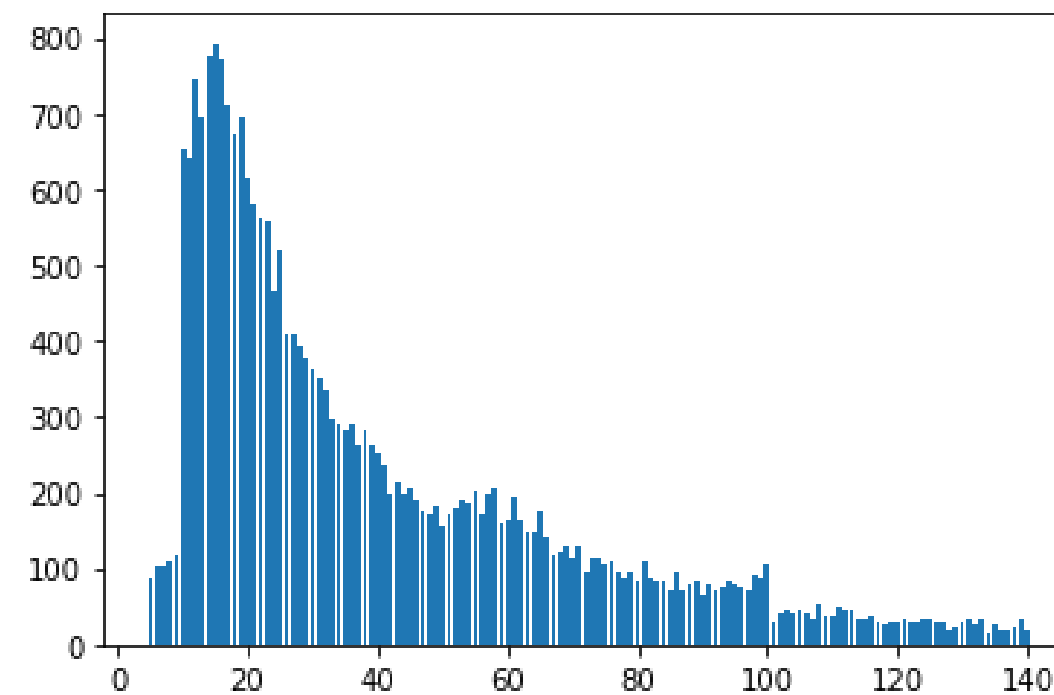
## EDA



[평점별 응답자 수]



[리뷰 글자수와 평점 상관관계]



[리뷰 글자수 분포]

## 전처리

- 워드클라우드를 만들기 위한 전처리 진행
- 특수문자 제거(\*정규표현식, re 라이브러리 이용) 후 명사/동사/형용사에 해당하는 단어의 어간(\*토큰화 Okt 라이브러리 이용)만 추출하여 사용

```
re.sub('[^a-zA-Z0-9가-힣]', ' ', text1)
```

```
from konlpy.tag import Okt
okt = Okt()

def filtering(text):
    word_list = okt.pos(text, stem = True) # 단어 토큰화
    pos_list = ['Noun', 'Verb', 'Adjective']
    filtered_word = []

    for word, pos in word_list:
        if pos in pos_list:
            filtered_word.append(word) # 명사, 동사, 형용사에 해당하는 단어 추출
    return filtered_word
```

# # 분석 및 인사이트

## 평점별 리뷰 데이터 워드클라우드 시각화



[평점 =  
1]



[평점 =  
2]



[평점 =  
4]



[평점 =  
5]

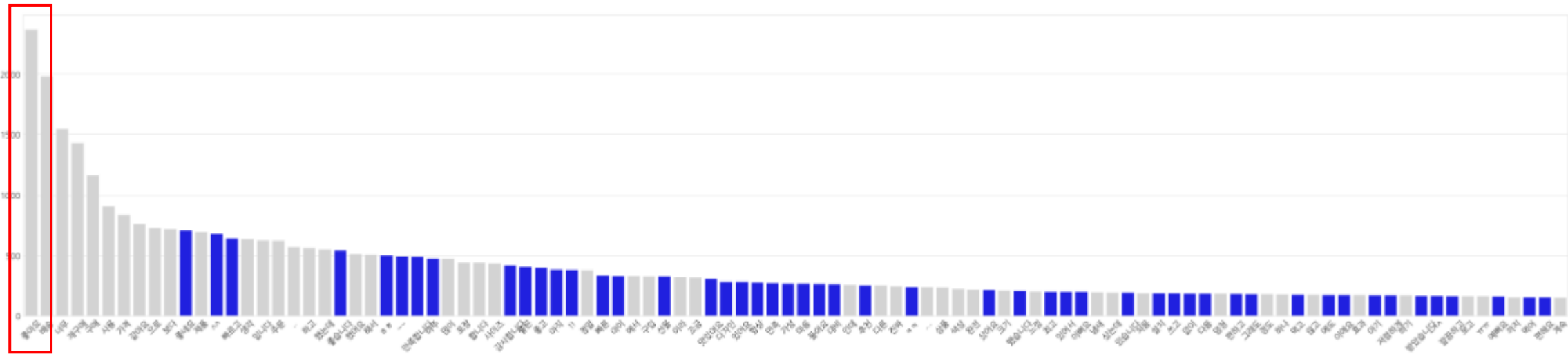
## 결론

- 평점이 낮을수록 부정적인 단어 많이 언급, 평점이 높을수록 긍정적인 단어 많이 언급  
(ex. '좋다'의 언급량: 평점 2점 << 평점 5점)
- '배송'의 경우 평점에 관계없이 모든 워드클라우드에서 나타남 -> 이는 배송 기간이나 배송 상태가 평점에 크게 영향을 미치는 것으로 보임
- 워드클라우드의 인사이트를 구체화하려면 '하다', '있다'와 같은 단어는 제거하는 것이 필요해 보임

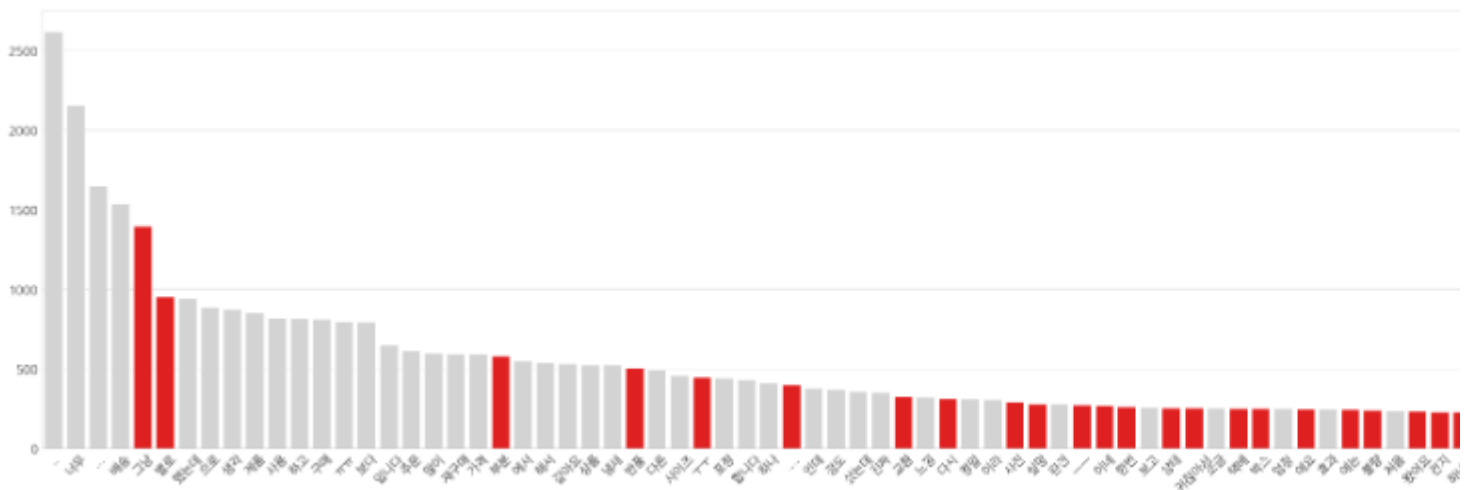


# # 1등 분석 리뷰

Data



[Positive - Negative]



[Negative - Positive]

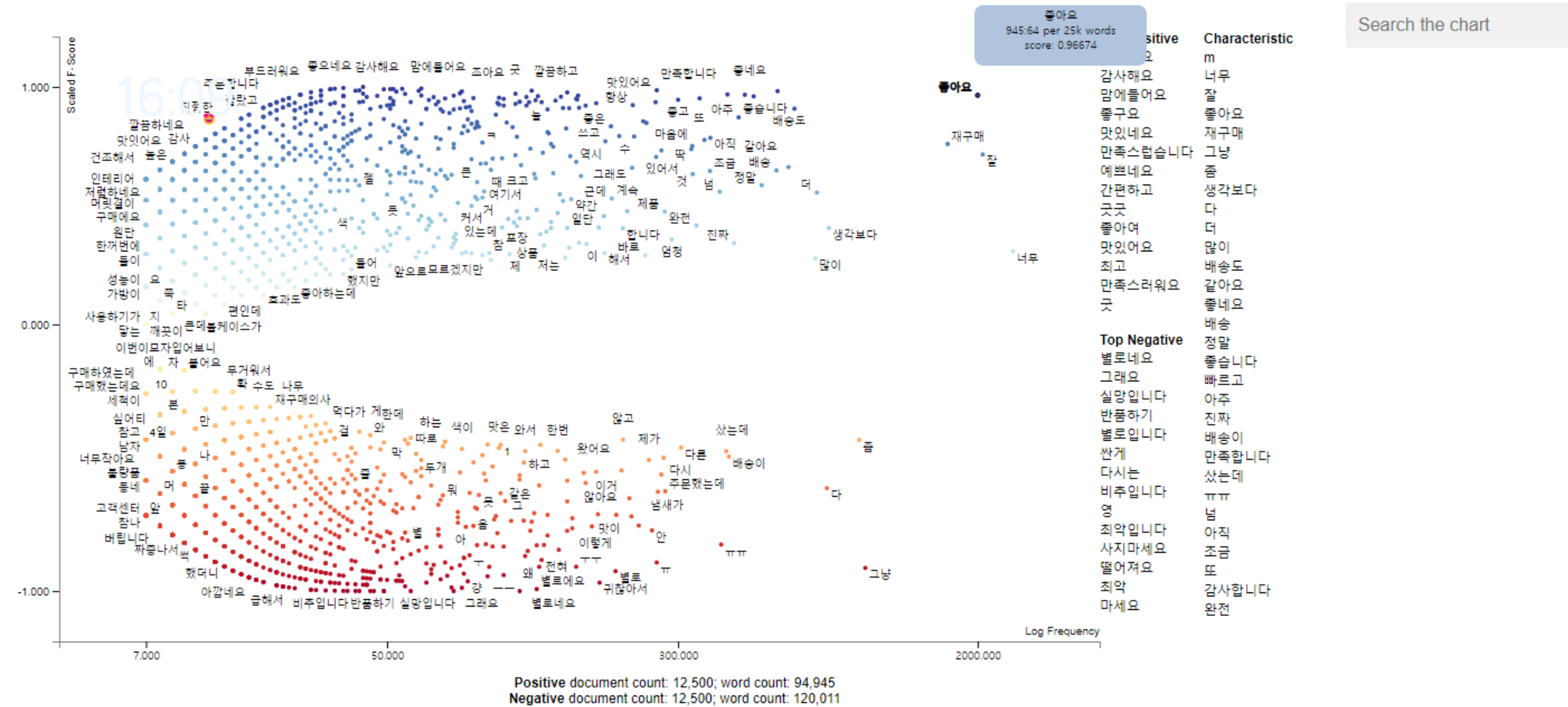
평점 4, 5점(긍정) 리뷰 형태소 상위 100개

- 파란색으로 하이라이트된 부분은 평점 1, 2점(부정) 리뷰의 상위 100개 형태소에 존재하지 않는 형태소
- 즉, 긍정 리뷰를 대표할 수 있는 특징
- 특이한 것은 ‘좋아요’가 부정 리뷰의 상위 100개 형태소에 포함된 것

id		reviews	target	category
214	214	좋아요 조아요 조아요	2	Negative
239	239	너무 싱싱한 꽃을 저렴한 가격에 받을수 있다니~~너무좋아요	2	Negative
381	381	상태안좋아요	2	Negative
422	422	두대 거치하실봉은 싱글 두개 사세요 이유는 간섭이너무심합니다 만듦새는 너무 좋아요	2	Negative
452	452	한달동안 마셔봤는데 얼굴에 별루 효과가 없어요 더마셔봐야하나 생각중~먹기엔맛은괜찮고...	2	Negative
...	...	...	...	...
24391	24391	상품 자체는 이쁘고 좋아요 ..다만 박스 열자마자 은행동냄새 같은 동냄새가ㅠ .....	2	Negative
24410	24410	배송은 빨라서 좋아요~~ 둘이 먹긴 좋은데...예열을 오래 해놨는데도 너무 오~~~	2	Negative
24566	24566	배송은 빠르고 좋았으나..너무 얇아서 서핑한번 탔더니 구멍이... ㅜㅜㅜ 그냥 서핑...	1	Negative
24765	24765	배송빠르고 좋아요 ㅎㅎ	1	Negative
24878	24878	지지도가 있어서 보관하기좋아요 화학 냄새가 빠지지않아서 걱정이죠	2	Negative

# # 1등 분석 리뷰

## Scattertext



Term: 좋아요

**Positive frequency:**  
945 per 25,000 terms  
180 per 1,000 docs  
**Some of the 1906 mentions:**

10 재구매 항상 여기서만 주문해서 잘 쓰고있어요 후딱쓰기좋아요 크기도 좋구

17 튼튼하고,좋아요 깔끔하게 정리되어서좋아요

22 빠른 배송에 친절한 기사님은 기본! 생각보다 고급스럽고 깔끔하게 주방에 두기 딱 좋아요~ 수납은 적힌 그대로 넉넉합니다^^

**Negative frequency:**  
64 per 25,000 terms  
19 per 1,000 docs  
**Some of the 133 mentions:**

214 좋아요 조아요 조아요

239 너무 싱싱한 꽃을 저렴한 가격에 받을수 있다니~~너무좋아요

381 상태안좋아요