

# 최종 발표

클러스터링 기법을 활용한  
e커머스 사용자 리뷰에 따른 시장 세분화 연구

A Study on Market Segmentation Based on E-Commerce User Reviews  
Using Clustering Algorithm

# 목차

01 연구 개요

02 클러스터링

03 군집 분석

04 연구 결론

05 참고 문헌



### 연구 배경 및 목적

- 코로나로 인한 **e커머스 시장 확대** 및 인터넷 쇼핑물 이용률 증가
- **고객 리뷰**가 소비자 구매 의사결정에 미치는 영향력 증대
- 다양한 형태의 고객 리뷰어들에 대한 **고객 분석 및 관리** 필요



### 연구 단계



## 연구 단계 &lt;데이터 수집&gt;

The screenshot shows a product review interface for a brand named 'eun'. At the top, there's a header with the brand name, a 'TOP500' badge, and statistics: '랭킹 180', '도움 수 211', and '작성한 후기 281'. Below this is a grid of product images. The main review section features a large image of a white garment with a red arrow pointing to the '수집 요소' box. The review text includes a 5-star rating, the date '2021.09.11', and a detailed description of the product's quality and fit. At the bottom, there are buttons for '도움 됨' (Helpful) and '도움 안 됨' (Not helpful).

## 수집 요소

- 사용자 이름
- 랭킹
- 전체 도움이 돼요 수
- 작성 리뷰 수
- 리뷰 상품명
- 리뷰 평점
- 리뷰 작성 날짜
- 이미지 여부
- 리뷰 내용
- 각 리뷰의 도움이 돼요 수

출처 : 위메프

## 연구 단계 &lt;전처리&gt;

기준	변수명		
감성	긍정 리뷰 비율	분기	2020년 1월 - 2020년 3월 작성 비율
	부정 리뷰 비율		2020년 4월 - 2020년 6월 작성 비율
	중립 리뷰 비율		2020년 7월 - 2020년 9월 작성 비율
	전체 리뷰 긍정 비율		2020년 10월 - 2020년 12월 작성 비율
	전체 리뷰 부정 비율		2021년 1월 - 2021년 3월 작성 비율
	전체 리뷰 중립 비율	계절	3월 - 5월 작성 비율
	내용 없는 리뷰 비율		6월 - 8월 작성 비율
	평점 - 댓글 불일치 비율		9월 - 12월 작성 비율
			12월 - 2월 작성 비율
별점	별점 1 비율	작성주기	중복 날짜 제거 작성 비율
	별점 2 비율		작성 주기 평균
	별점 3 비율		리뷰 작성하지 않은 기간
	별점 4 비율	기타 사용자 정보	리뷰어 랭킹 점수
	별점 5 비율		전체 리뷰 개수
	별점 평균		'도움이 돼요' 받은 리뷰 비율
리뷰 내용	이미지 첨부 리뷰 비율		총 '도움이 돼요' 개수
	리뷰 길이 평균		

32개 입력변수 전처리 및 정규화

## 연구 단계 &lt;전처리&gt;

## 텍스트 마이닝(감성분석)

lee\*\*\*\*\*

★★★★★ 5.0 | 2021.04.30 | 신고

반려견타입 중형견

강아지이름 (기본사이즈) 반려동물인식표, 강아지인식표(기본사이즈) 푸들(블랙), 형광다홍

눈에 확 띄네요  
 고급재 이라고 하셨는데 가격에 비해 무직포재질을 사용하신 느낌이랄까요? 더  
 좋았을 것 같아요.  
 아담한 디자인 포장이 좋지도 보았으니 너무 감동!  
 항상 산책 다닐때 기본에 쓰고있는데 인식표가 가슴줄 사이드에 있다보니 이름을 물어보셨는  
 데 등에 딱 매달아주면 확실히 선명해서 너무 좋을것같아요  
 아무래도 검둥이라ㅋㅋㅋ 밝은 나시 하나 입혀야 할것같네요ㅋㅋㅋ 재구매 의사있어요~~~

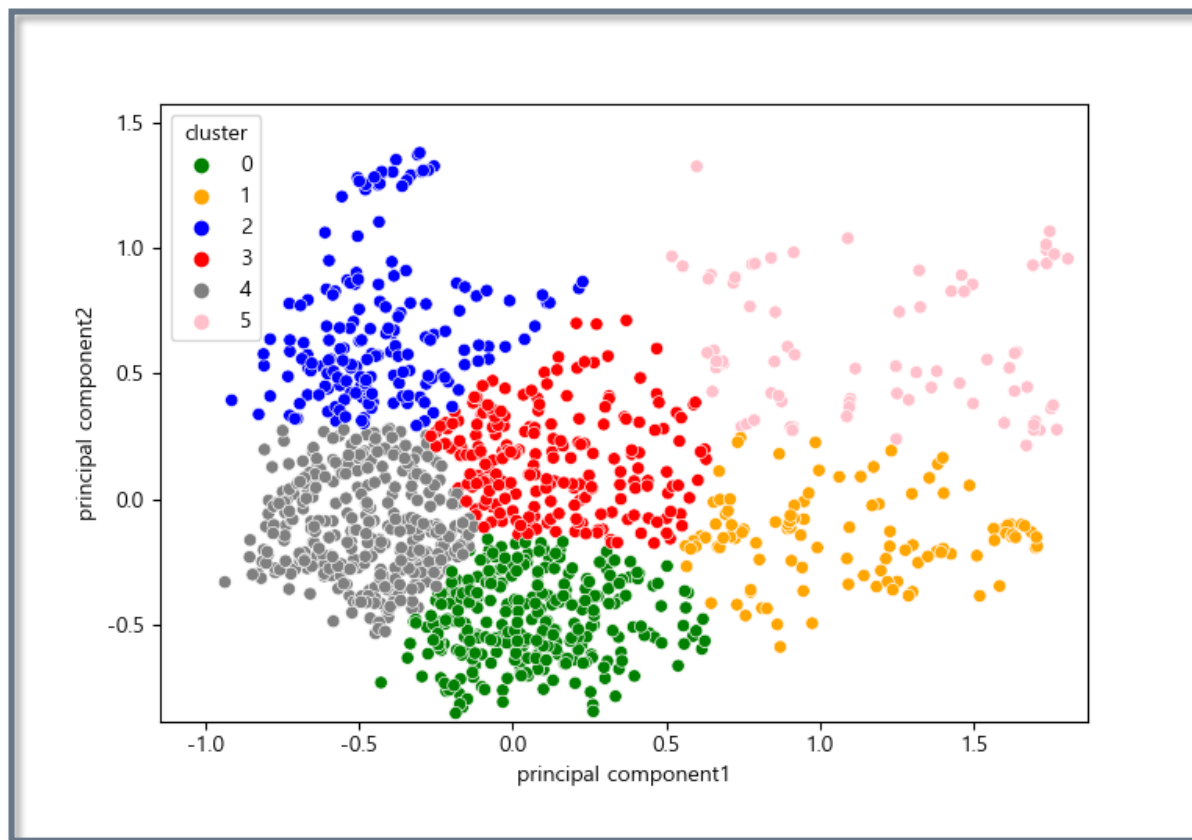


## [감성분석 과정]

1. '리뷰내용' 칼럼 사용
2. 데이터 정제
  - 특수문자 및 불용어 제거, 형태소 분리 등
3. KOSAC 감성사전 변형 및 구축
4. 감정 사전을 통한 각 리뷰의 긍/부정 여부 판단

**K-means** 클러스터링 기법을 이용한 군집화 진행

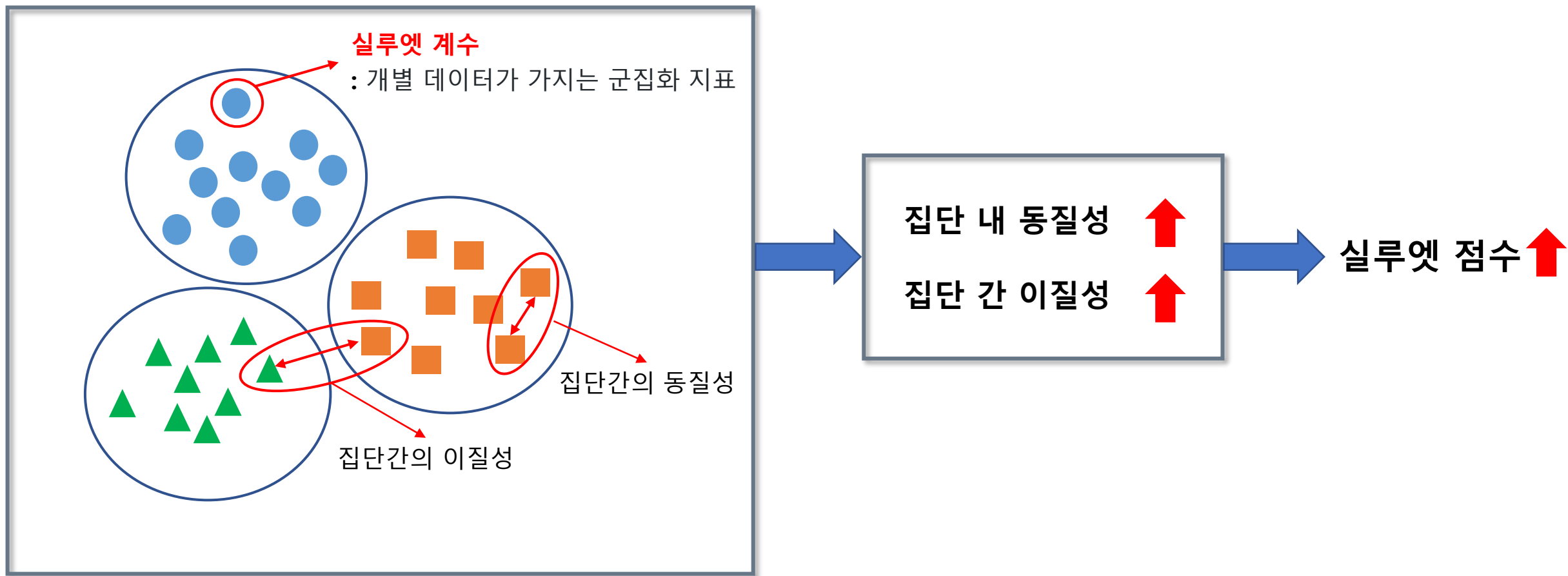
모델 시각화 및 성능향상을 위해 **PCA (차원 축소)**, 입력변수 간 **상관계수 분석**



클러스터링 2차원 시각화 그래프

## 실루엣 점수\*

: 클러스터링 성능 평가 지표로 각 군집 간의 거리가 얼마나 효율적인지를 나타내는 점수 (-1.0 ~ 1.0)



\* 참고문헌 : 김성수, 백준영, 강범수, (2017), "실루엣을 적용한 그룹탐색 최적화 데이터클러스터링", 한국경영과학회지, pp.25-34



## 모델 성능 향상을 위한 입력변수 조정

분기과 계절에  
대한 입력변수

기준	변수명		
감성	긍정 리뷰 비율	분기	2020년 1월 - 2020년 3월 작성 비율
	부정 리뷰 비율		2020년 4월 - 2020년 6월 작성 비율
	중립 리뷰 비율		2020년 7월 - 2020년 9월 작성 비율
	전체 리뷰 긍정 비율		2020년 10월 - 2020년 12월 작성 비율
	전체 리뷰 부정 비율		2021년 1월 - 2021년 3월 작성 비율
	전체 리뷰 중립 비율	계절	3월 - 5월 작성 비율
	내용 없는 리뷰 비율		6월 - 8월 작성 비율
	평점 - 댓글 불일치 비율		9월 - 12월 작성 비율
			12월 - 2월 작성 비율
별점	별점 1 비율	작성주기	중복 날짜 제거 작성 비율
	별점 2 비율		작성 주기 평균
	별점 3 비율		리뷰 작성하지 않은 기간
	별점 4 비율	기타 사용자 정보	리뷰어 랭킹 점수
	별점 5 비율		전체 리뷰 개수
	별점 평균		'도움이 돼요' 받은 리뷰 비율
리뷰 내용	이미지 첨부 리뷰 비율		총 '도움이 돼요' 개수
	리뷰 길이 평균		

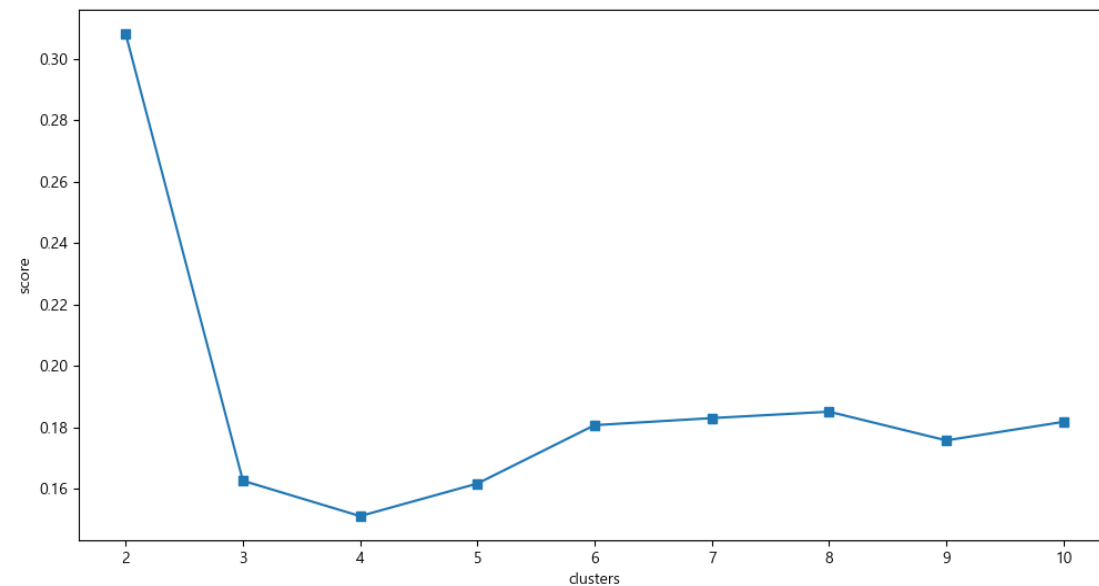
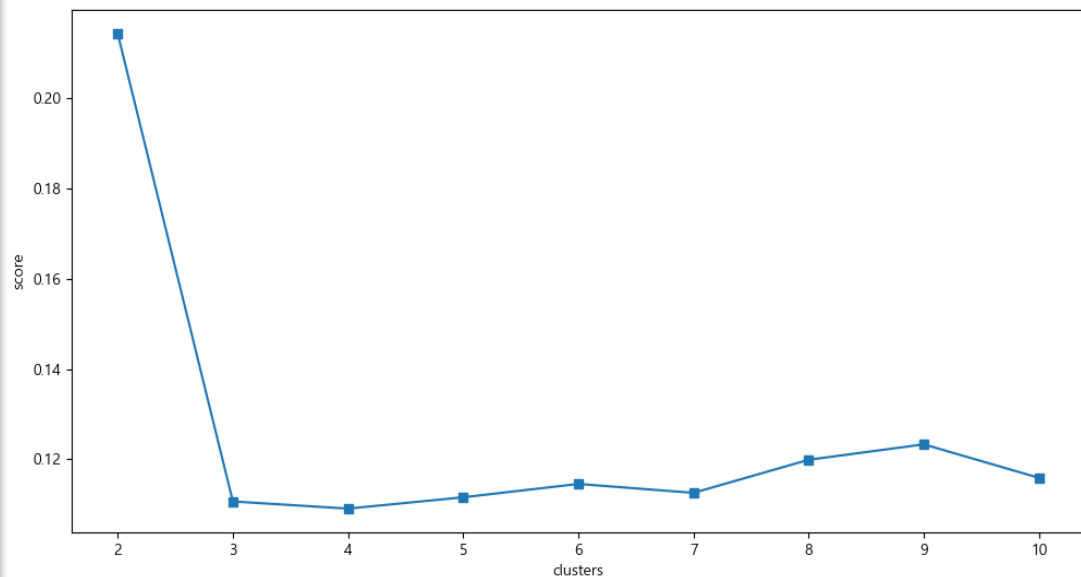
→ 제거

## 모델 성능 향상을 위한 입력변수 조정

32개의 변수 모델

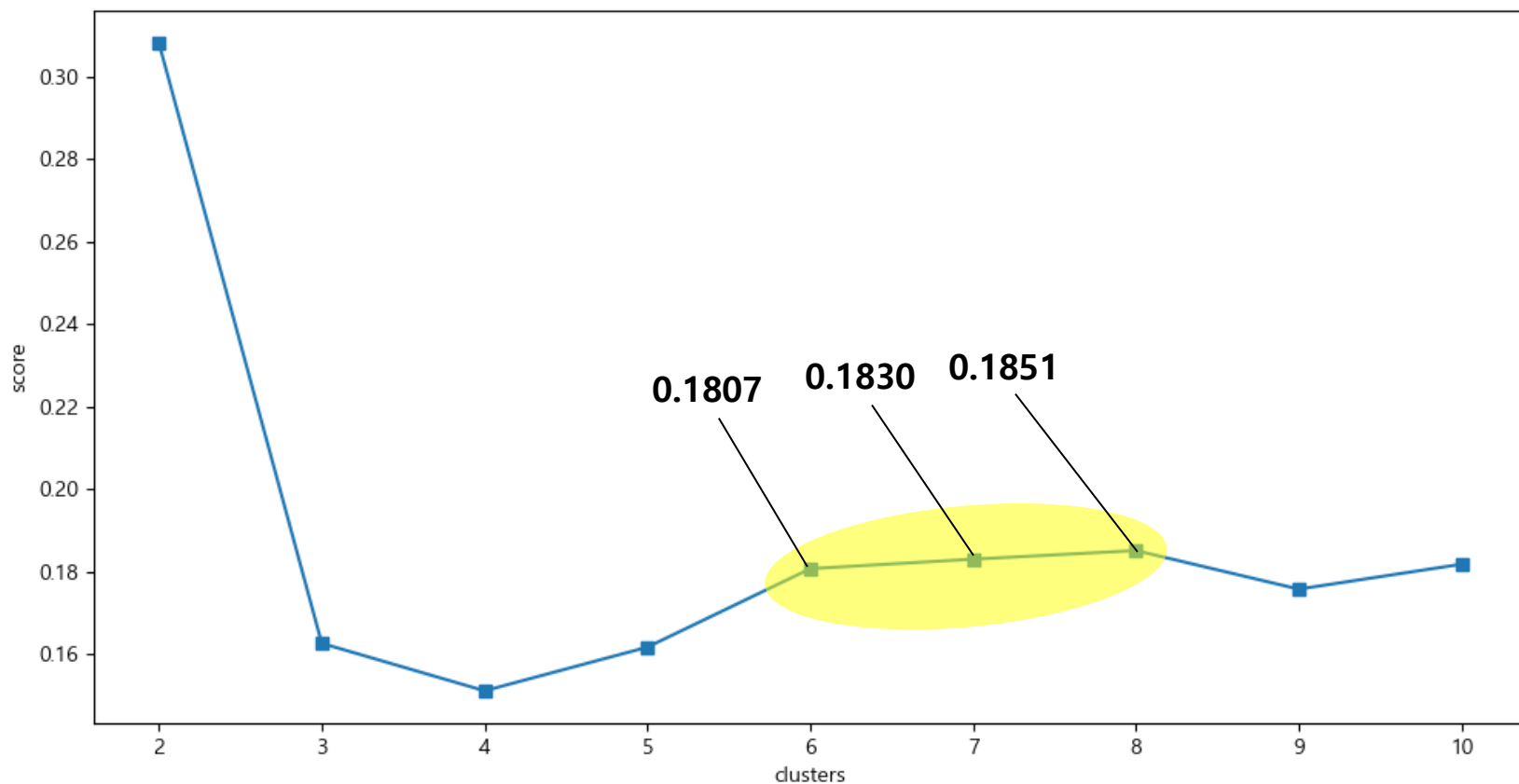
실루엣 점수  
모두 0.5이상 증가

23개의 변수 모델  
( 분기와 계절 입력변수 제거 )



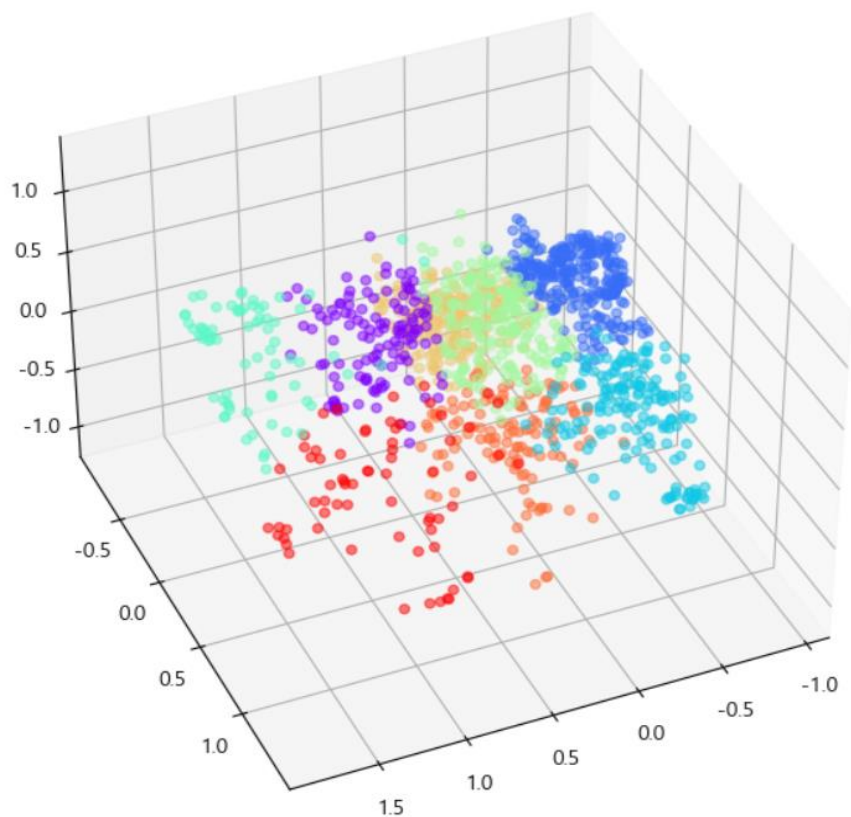
## 최종모델 선정: 실루엣 점수 비교 분석

23개의 변수 모델( 분기와 계절 입력변수 제거 )



## 최종모델 선정: 실루엣 점수 비교 분석

PCA = 3, Clustering = 8



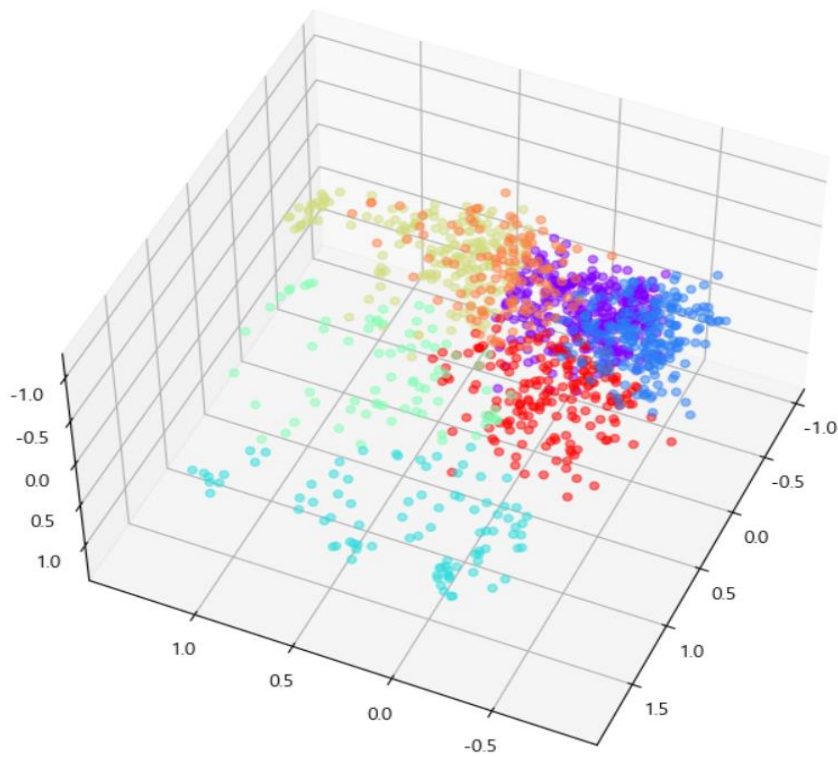
[ 모델 내 군집 별 실루엣 점수 ]

1	0.265
2	0.404
3	0.392
4	0.445
5	0.188
6	0.382
7	0.307
8	0.150

군집5과 군집8의 실루엣 점수가 현저히 낮음

## 최종모델 선정: 실루엣 점수 비교 분석

PCA = 3, Clustering = 7



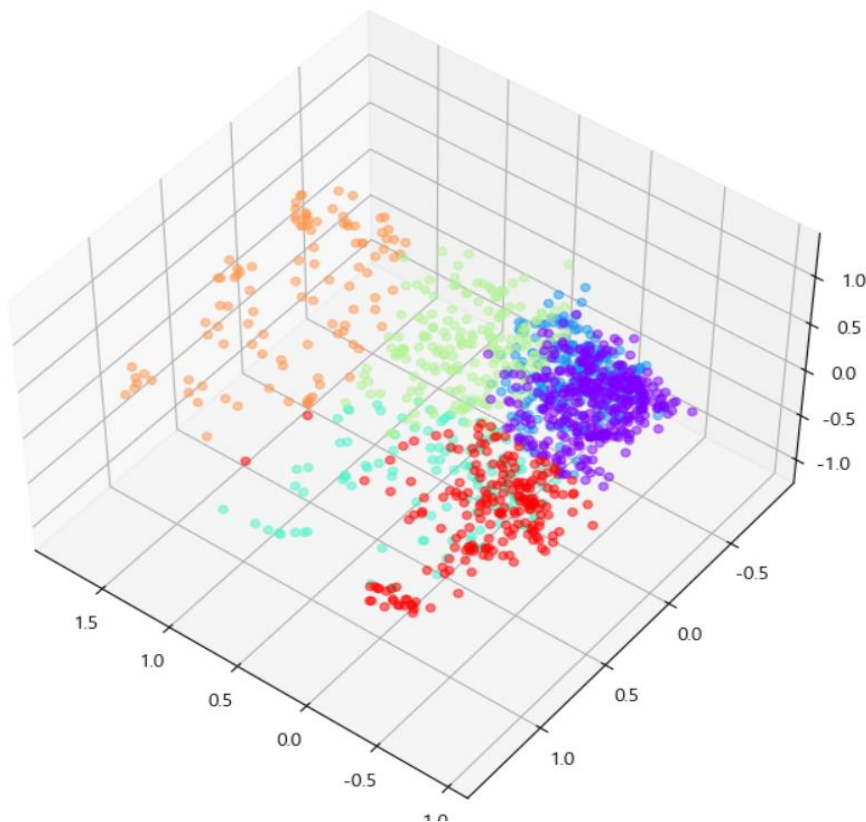
[ 모델 내 군집 별 실루엣 점수 ]

1	0.405
2	0.387
3	0.400
4	0.189
5	0.356
6	0.314
7	0.213

군집4과 다른 군집들 간의 실루엣 점수 차이가 크며  
현저히 낮은 점수를 가짐

## 최종모델 선정: 실루엣 점수 비교 분석

PCA = 3, Clustering = 6



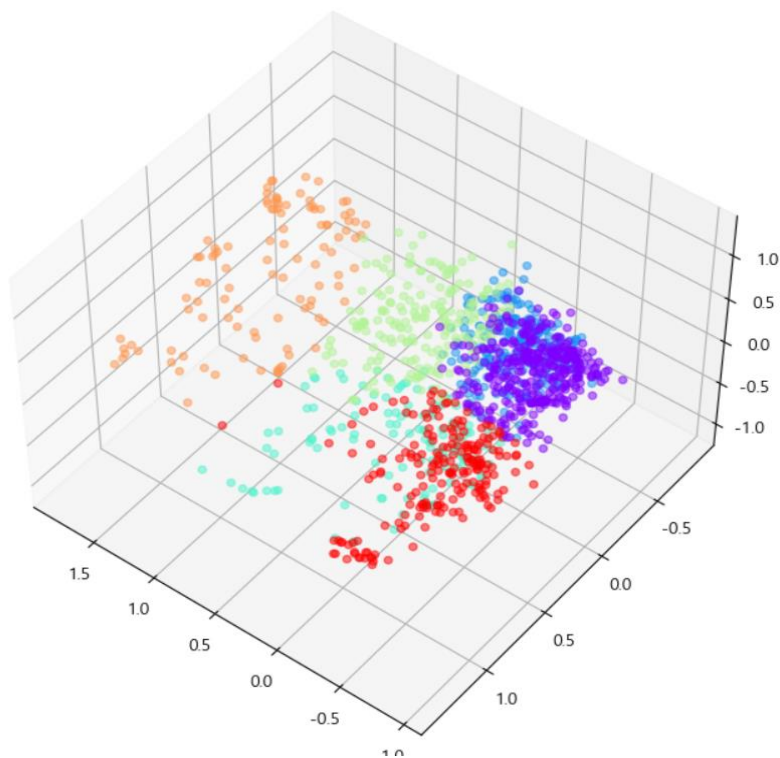
[ 모델 내 군집 별 실루엣 점수 ]

1	0.407
2	0.394
3	0.235
4	0.202
5	0.392
6	0.340

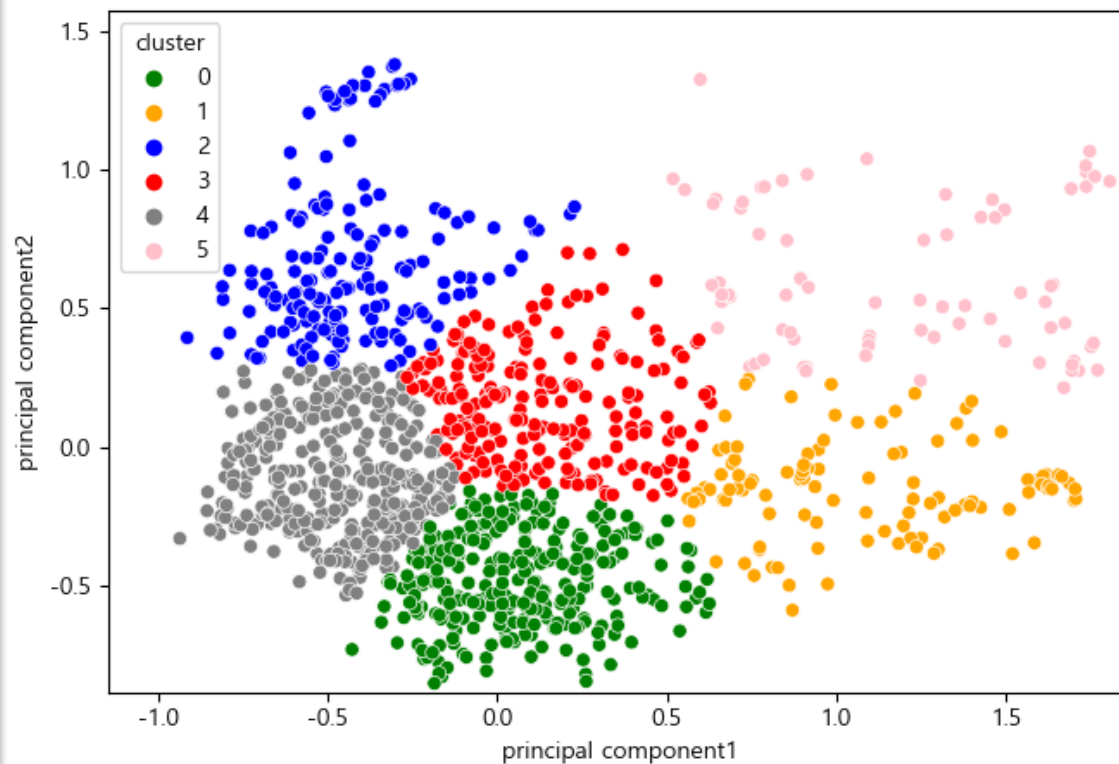
군집 label 별 비교적 비슷한 실루엣 점수  
군집 개수 6,7,8개 중 가장 높은 실루엣 점수 가짐

## 최종 모델

PCA = 3, Clustering = 6  
실루엣 점수: 0.18



PCA = 2, Clustering = 6



## 군집 분석: 입력변수 별 각 군집 기술통계 비교분석

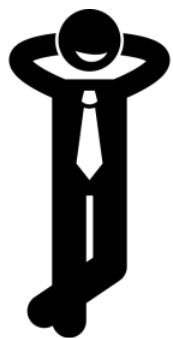
	군집 1	군집 2	군집 3	군집 4	군집 5	군집 6
리뷰어 랭킹 점수				값이 0	값이 0	가장 높음
도움이 돼요 수			0 값이 많음		가장 낮음	가장 높음
별점 1 비율	가장 낮음			값 골고루 분포	가장 높음 (격차 큼)	
별점 2 비율				값 골고루 분포	가장 높음	가장 낮음
별점 3 비율		가장 높음 (골고루)			평균 대비 1 많음	가장 낮음
별점 4 비율		높은 편	가장 높음		가장 낮음	
별점 5 비율	높은 편				가장 낮음	가장 높음
별점 평균	높은 편	높은 편	높은 편	높은 편	가장 낮음	가장 높음
중복 날짜 제거 작성 기간		가장 낮음	높은 편	높은 편	가장 높음 (격차 큼)	높은 편
작성 주기 평균				가장 높음	가장 낮음	
리뷰 작성하지 않은 기간	가장 낮음				가장 높음 (골고루)	
이미지 댓글 비율	낮은 편	가장 낮음	높은 편	낮은 편		가장 높음 (격차 큼)
리뷰 길이 평균	가장 낮음	낮은 편				가장 높음 (비슷함)
리뷰어의 전체 댓글 수		가장 높음 (비슷함)			가장 낮음	
도움이 돼요 수 받은 댓글		낮은 편	높은 편	낮은 편	높은 편	가장 높음
None 리뷰 비율	높은 편	높은 편	낮은 편	높은 편	값이 0	낮은 편
긍정 댓글 비율	높은 편	높은 편			가장 낮음	가장 높음 (격차 큼)
부정 댓글 비율	가장 낮음		낮은 편	높은 편	가장 높음	낮은 편
중립 댓글 비율				가장 높음 (비슷함)		가장 낮음 (비슷함)
전체 긍정 비율	높은 편	높은 편	높은 편		가장 낮음	가장 높음
전체 부정 비율	낮은 편	낮은 편			가장 높음 (격차 큼)	낮은 편
전체 중립 비율						가장 낮음 (격차 큼)
평점 댓글 불일치	낮은 편		높은 편	가장 높음		낮은 편



## 시장세분화



## 군집 1



무던한 귀차니즘 고객

대충대충 좋게 별점만 !

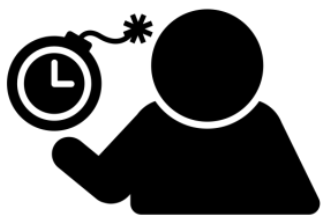
## 리뷰 특징

- 별점 5 점만을 많이 부여
- 리뷰 길이 짧음
- None 리뷰 비율 높음
- 도움이 돼요 수 낮음
- 반복되고 성의 없는 리뷰 많음

## 마케팅 방안

- 리뷰 자동완성 기능 제공
- 리뷰 10글자 이상 글자 수 제한

## 군집 2



마감임박 베틀치기 고객

빨리빨리 리뷰 복붙 !

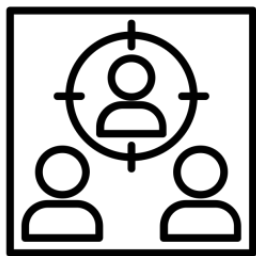
## 리뷰 특징

- **별점 2,3,4** 점을 다양하게 부여함
- 이미지 댓글 비율 낮음
- **리뷰를 몰아서 작성함**

## 마케팅 방안

- 리뷰 작성 기간 세분화 및 **기간에 따른 포인트 차등 지급**
- 리뷰 **이미지 유/무**에 따른 포인트 차등 지급

## 군집 3



미리 잡자 잠재 고객

좋은데 별점 5점은  
안줄래 !

## 리뷰 특징

- 별점 3,4점을 많이 부여함
- 별점 5점에 인식함
- 도움이 돼요 받은 댓글 비율 높은 편
- 이미지 리뷰 많은 편

## 마케팅 방안

- AI 기반 상품 추천 시스템을 통한 1인 맞춤형 마케팅

## 군집 4



논리정연 스마트 고객

내 주관대로 꼼꼼히 열심히 !

## 리뷰 특징

- 금/부정 댓글 골고루 작성
- 상품 구매에 대한 주관 뚜렷
- 별점 1~5점 다양
- 작성주기 높음

## 마케팅 방안

- 제품 체험단으로 선정하여 충성고객으로의 유도
- 우수 체험단 선정 후 추가 보상 및 혜택 제공

## 군집 5



주의경보 이탈 고객

상품 구매가 실망스러워 !

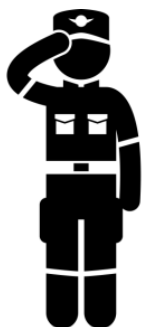
## 리뷰 특징

- 별점 1,2점을 많이 부여함
- 부정 댓글 작성 비율 높음
- 리뷰작성 자주 안함
- 중복날짜 가장 높음
- 전체 리뷰 수 적음
- 리뷰를 1개만 작성한 사람이 많음

## 마케팅 방안

- VOC(Voice of Customer)를 통한 실시간 피드백 제공
- 최근 1년 미구매 고객 대상 할인쿠폰 및 포인트 지급

## 군집 6



한결같은 충성 고객

상품도 서비스도 좋아 ! 충성 !

## 리뷰 특징

- 별점 5점을 많이 부여함
- 긍정 댓글 작성 비율 높음
- 리뷰어 랭킹 달성 고객 많음
- 전체리뷰 수 많음
- 도움이 돼요 수 많음
- 이미지 댓글 비율 높음

## 마케팅 방안

- 멤버십 서비스 혜택 제공
- 고객 등급 관리제 도입 및  
매월 등급에 따른 혜택 제공

### ✓ 연구 결과

- 이커머스 시장의 리뷰고객을 총 6개의 군집으로 세분화할 수 있음
- 시장세분화 결과를 통해 고객 관리 방안을 제시함

### ✓ 기대효과 및 시장성

- e커머스 시장의 고객 유형 파악 및 고객 관리 용이
- 잠재 고객 발굴을 통한 맞춤형 고객 관리 서비스 제공
- 고객 군집별 타겟 마케팅을 통한 기업의 마케팅 비용 절감 및 수익 창출