Feb 2023. eMFORCE Report

2019-2022 대한민국 쇼핑트렌드 2탄

온라인 쇼핑 유발 마케팅의 New Paradigm

진짜 데이터 드리븐 마케팅을 하고싶다면

eMFORCE Digital Marketing Group 데이터랩

쇼핑트렌드 보고서 1탄

쇼핑 유발 마케팅의 NEW PARADIGM

지난 2023년 1월 20일에
발간한 보고서에서는

네이버에서 제공하는
쇼핑트렌드의 중분류 상품
카테고리 데이터를 활용해

동시기에 많이 클릭되는
상품들을 그룹화시키는
작업을 진행했다.

쇼핑트렌드 보고서 2탄

이제 우리는, 무엇을 광고해야 하는가

2번째로 발간한 보고서는

동일한 작업을 '소분류'
데이터를 기준으로 진행했다

대상만 다르게 진행했으니
결과가 드라마틱하지는 않다

하지만 알 수 있을 것이다
이제는 이게 가능하다는 것을.

검색 광고 운영 데이터에 대한 데이터 분석 접근 가설 및 관점

검색광고는 검색을 하는 상품에 대한 정보를 노출시켜서 클릭을 유도하는 과정이다.

이 간단한 과정에도 수많은 전략과 전술이 존재한다.
사람들이 어떤 키워드로 상품을 주로 검색하는지도 알아야 하고 (소파, 쇼파, 1인용소파, 거실소파,..)
그게 성, 연령별로 어떻게 다른지도 알아야 하고, 검색 시기별로 키워드가 어떻게 달라지는지도 알아야 한다.
심지어 소파 상품 하나를 팔아도 사이트로 유입되는 키워드가 비단 소파 종류에만 국한된 것도 아니다.

예를 들면, 소파 상품을 클릭하고 구매했지만, 유입 키워드는 그릇, 테이블, 인테리어 등 다양할 수 있다는 얘기다.
그리고 이렇게 유입되는 키워드들은 100개가 될지, 1,000개가 될지, 10,000개가 될지 알 수 없다.

개인이 이 모든 키워드들을 매일 일일이 관리할 수 없기 때문에
주로 유입/구매가 일어나는 상위 키워드들을 집중하여 관리하는 전략을 택한다.

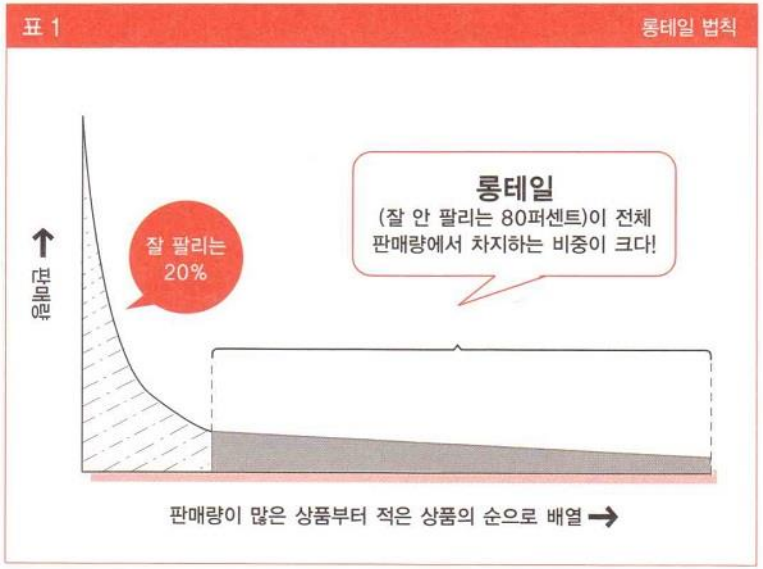
만약 한 상품에 100개의 유입 키워드가 존재하는데 상위 20개 키워드가 전체 매출의 80%를 담당한다면?
옛날 어르신의 말처럼 '선택'과 '집중'을 하는 전략이 효율적이고 효과적일 것이다.

이런 계산법이 19세기 말에 발표된 <**파레토 법칙** (Pareto's Law)>이다.
실제 전 세계 인구의 20%가 나머지 80%를 먹여 살린다는(20대 80의 사회), 괴상한 이론도 있다.

그런데 이와 반대로, 하위 80% 상품이 상위 20% 상품의 매출을 상회한다는 <**롱테일 법칙** (Long Tail)>도 있다.

온라인 광고는 어떨까? 상위 20%의 검색 키워드가 전체 매출의 80%를 차지하는가?
아니 하위 80%에 해당하는 키워드들을 제대로 들여다보고 관리한 적은 있을까? ~~(아.. 너무.. 도발적인가..?)~~

전체 검색 키워드 뽀개기! 검색 광고 운영 데이터 분석 기법 개발! 이게 우리가 이 작업을 시작한 이유다.



출처 : 스가야 오시히로 <롱테일 법칙>
출처 : 소비자평가 (언론 기사)
http://www.iconsumer.or.kr/news/articleView.html?idxno=4706

시계열 클러스터링(Time Series Clustering) 데이터 분석 기법에 대한 이해

시계열 클러스터링 분석 기법은 통계상에서의 클러스터링(군집) 분석 기법의 일종으로,
풀어서 얘기하면, '시간'이라는 조건을 가진 '계열'의 데이터들을, '시간'이라는 기준을 통해 몇 개의 '군집'으로 묶는 작업이다.

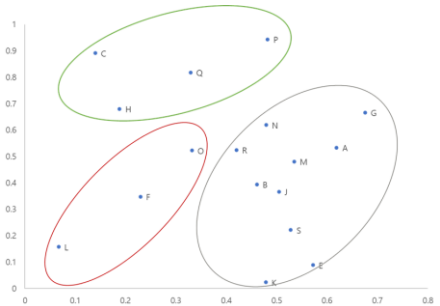
여름에 반팔, 원피스, 여행용 가방, 수영복 등이 많이 팔린다는 사실은 누구나 알고 있다.
알고 있지 않더라도 조금만 생각해 보면 누구나 예상할 수 있다.
상식 선에서 상품들 간의 연관성을 쉽게 연결 지을 수 있고 실제 매출 데이터로도 증명되기 때문이다.

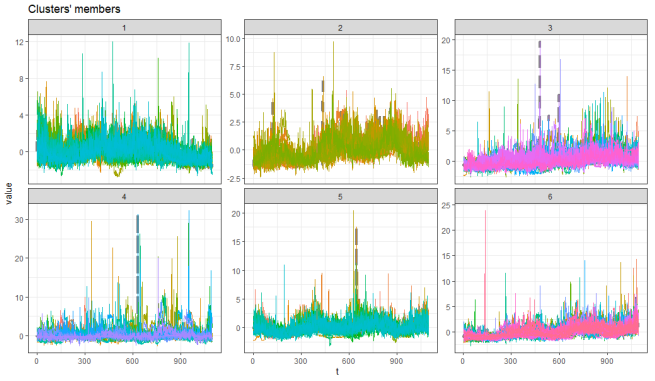
그래서 오프라인이나 온라인 할 것 없이 여름 시즌이 되면 이런 상품들을 돋보이게 진열하고
각종 명목으로 할인 행사를 펼치면서 충분히 지갑을 열 준비가 된 소비자들의 구매 욕구를 자극한다.

그래서 이런 특정 시즌에 판매되는 상품들을 가리켜 시즌성(Seasonality)이 있다고 한다.

그런데 1년은 365일이고 온라인이라는 환경에서는 매일 셀 수도 없이 많은 상품들이 판매된다.
어제 판매된 상품의 종류와 오늘 판매된 상품의 종류가 동일하지 않을 것이다.
오프라인에서는 진열이 곧 판매를 결정하지만, 소비자의 자발적 검색과 탐색이 두드러지는
온라인에서는 전면에 내세운 상품군 외에도 다양한 상품들이 매일 다채롭게 판매될 것이다.

period	3ce	ahc아이크림	ts샴푸	고운발크림	나스쿠션
2018-01-01	38.83079	32.04419	56.65523	3.73675	39.54943
2018-01-02	37.5	34.56108	69.74271	4.18293	41.2599
2018-01-03	38.49809	31.65541	61.26929	4.2387	46.59991
2018-01-04	38.54562	35.37957	65.31732	5.24261	38.63162
⋮	⋮	⋮	⋮	⋮	⋮
2020-12-28	10.59885	56.7833	31.6638	35.02509	16.43721
2020-12-29	11.31178	49.13034	31.54373	37.6464	10.97204
2020-12-30	15.58935	44.58768	26.1578	24.65142	10.17939
2020-12-31	12.83269	35.72744	23.5506	21.08198	8.96954





만약 동시기에 함께 판매가 증가하는 상품들을 알게 된다면 어떨까?

반팔과 원피스와 여행용 가방이 동시기에 판매될 확률이 높다는, 누구나 예상 가능한 것 외에
반팔과 후라이팬이 동시기에 판매되고 있다는 현상을 발견하게 된다면 어떨까?

반팔과 후라이팬이 동시기에 판매되는 이유를 반드시 규명할 필요는 없다.
서로 아무 연관이 없고, 구매하는 층이 겹치지 않을지도 모른다.
하지만 여기는 온라인이지 않은가. 함께 팔린다는 것만으로도 마케팅을 할 충분한 이유가 된다.

데이터 분석 접근 방향

- 월별 클릭지수를 활용, 통계적 분석 기법인 **시계열 클러스터링(Serial Clustering)**을 적용하기로 결정 후
  일부 중분류 항목 중 몇몇 기간에 데이터 값이 누락, 혹은 존재하지 않는 경우를 발견하고
  이에 해당하는 데이터(이상치) 항목들을 소거하여 살아남은 데이터만으로 분석

STep 01 *Data Crawling*

쇼핑 카테고리별 월간 클릭 지수 수집

네이버에서 제공하는
전체 쇼핑 카테고리에 대해
'중분류'를 기준으로
최근 4년 동안의(48개월)
월별 클릭 지수 데이터 추출

STep 02 *Cleansing*

분석하기에 데이터 값이 충분하지 않은 정보 소거

추출된 상품 카테고리 중
일부 기간의 데이터 값이
존재하지 않는 경우,
즉 48개월 기간이 모두
채워지지 않아서
분석이 어려운 카테고리 소거

STep 03 *Hypothesis*

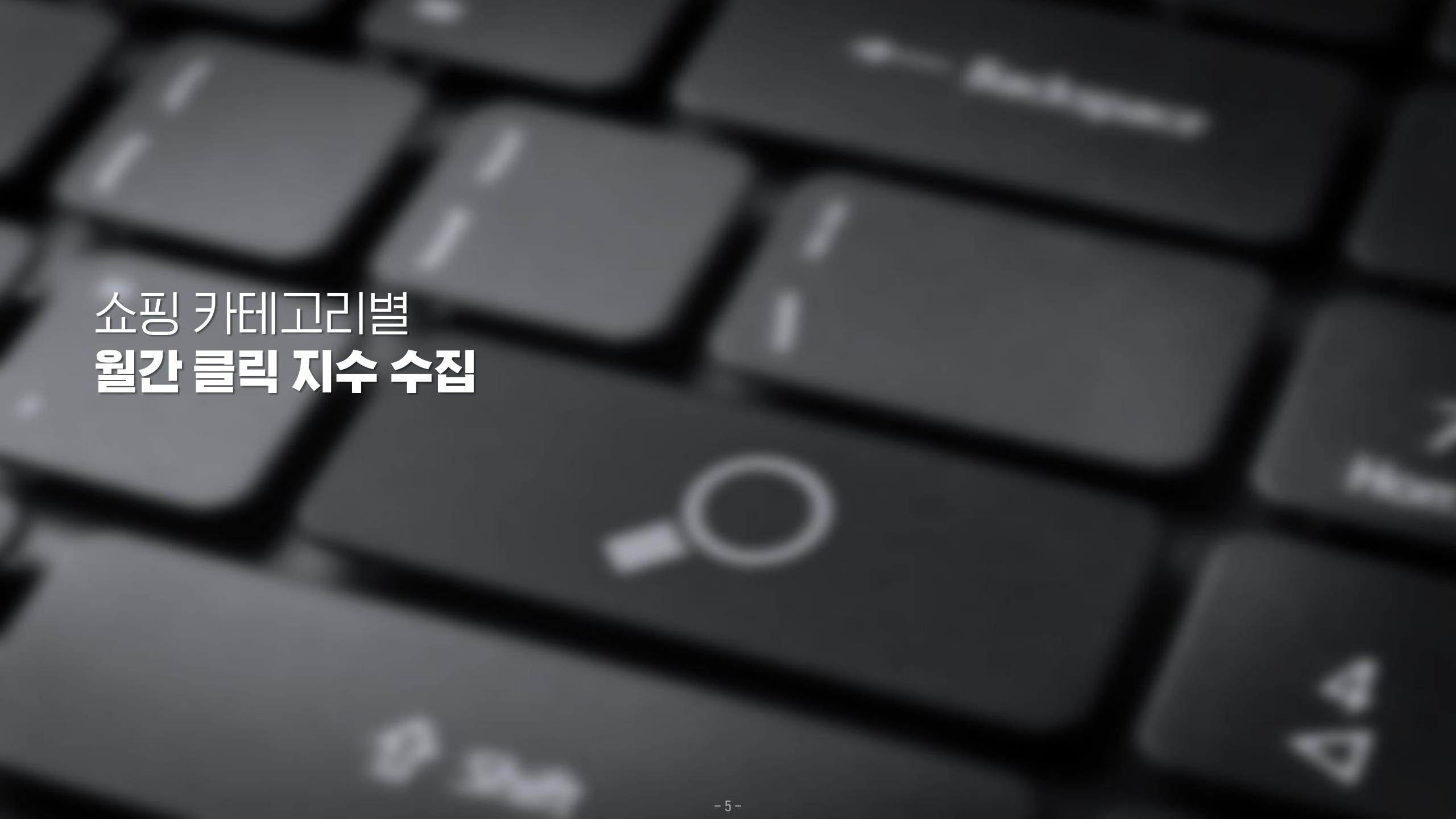
상관성 분석을 통해 그룹화 가능성 진단

살아남은 카테고리 간의
상관성(Correlation) 분석을 통해
통계적으로 그룹화될
가능성이 있는지
유의미한 해석이 가능할지를
사전에 판단하는 단계

STep 04 *Analysis*

시계열 클러스터링 (Serial Clustering) 기법을 통해 본격적 분석

통계적 분석 기법인
시계열 클러스터링 기법을
활용하여
몇 개의 그룹으로 묶이는지,
묶인 그룹 안에 어떤 상품군이
응집해 있는지 분석



쇼핑 카테고리별
월간 클릭 지수 수집

데이터 설계

- 네이버 데이터랩 사이트에서 제공하는, 전체 상품 카테고리에 대한 <쇼핑 클릭 지수> 데이터를 API로 호출해서 활용
- 네이버에서 자체적으로 분류한, 대분류 / 중분류 / 소분류 중 본 보고서는, **소분류**를 기준으로 분석 진행
  소분류 데이터를 추출함에 있어 특이점은 뒷장에 설명
  - 대분류는 11개, 중분류는 269개, 소분류는 2,301개...

항목	내용
데이터 종류	네이버 쇼핑클릭 데이터 - URL : https://datalab.naver.com/shoppingInsight/sCategory.naver - 네이버 포털 사이트에서 제공되는 쇼핑 상품 항목을 소비자가 "클릭"한 결과를 제공해주며 다만, 실제 클릭 규모가 아닌 지수화된 데이터를 제공
데이터 특성	1. 접근 가능한 온라인 데이터 중, 실제 소비 행동에 가장 근접한 데이터 2. 개별 상품군을 넘어 전체 상품 카테고리 조망 가능
데이터 한계	1. 소비자의 실제 구매 여부 확인 불가 2. 트렌드, 유행에 따라 카테고리 분류 정의 가변적 3. 사이트에서 검색되는 데이터 추이와 API로 추출하는 데이터 추이 상이
분석 범위	2019년 1월 ~ 2022년 12월까지의 월간 클릭 추세
추출 규모	소분류 기준 2,301개 상품군 분석 활용

대분류 (11개)	중분류 (269개)	소분류 (2,301개)
가구/인테리어	여성의류	니트
디지털/가전	남성의류	롱부츠
면세점	스킨케어	스킨
생활/건강	베이스메이크업	토너
스포츠/레저	생활가전	아이라이너
식품	계절가전	워크스테이션
여가/생활편의	PC부품	라디오
출산/육아	침실가구	CPU
패션의류	거실가구	RAM
패션잡화	인테리어소품	소파
	전통주	막걸리
	조미료	올리고당
	...	...

*출처 : 나무위키

데이터 추출

- 추출에 앞서, 네이버 쇼핑 인사이트 API 호출 스크립트 확인 시 **상위 분류 체계의 코드 필요**
- 예를 들어, 여성의류(중분류)에만 있는 '원피스' 키워드의 경우 여성의류(중분류) 코드와 패션의류(대분류) 코드 둘 다 활용 가능하지만, '스웨터' 키워드의 경우 패션의류(대분류) 코드를 활용하면 여성의류(중분류)의 '스웨터'와 남성의류(중분류)의 '스웨터' 구별 불가능
- 따라서 보다 정확한 데이터 산출을 위해 **대분류 코드와 중분류 코드 중 어떤 것을 활용**하는 것이 적합한지 확인 필요

```
{
  "startDate": "2019-01-01",
  "endDate": "2019-12-31",
  "timeUnit": "month",
  "category": "50000000",
  "keyword": [
        {"name": "패션의류/원피스", "param": [ "원피스" ]},
        {"name": "패션의류/스웨터", "param": [ "스웨터" ]} ],
  "device": "",
  "gender": "",
  "ages": [ ]
}
EOF
```

대분류	대분류 코드	중분류	중분류 코드	소분류
패션의류	50000000	여성의류	50000167	원피스
패션의류	50000000	여성의류	50000167	스웨터
패션의류	50000000	남성의류	50000169	스웨터
…	…	…	…	…

*출처 : 네이버 오픈 API, https://developers.naver.com/docs/serviceapi/datalab/shopping/shopping.md

데이터 추출

- 대분류 코드와 중분류 코드를 각각 활용하여 데이터 추출 후 **단순 개수 비교**
  ① 투입하였으나 추출이 되지 않은 키워드 개수 = NA 키워드 개수　　② 데이터 값이 0으로 나오는 키워드의 개수 = ratio 0 키워드 개수
- 전체 투입 키워드 중 (①+②)의 비율을 계산하여 일명 **"불량률"**로 명명, 불량률이 낮은 쪽의 코드를 선택하고자 하였으나 **데이터 신뢰성에 대한 의문**

		패션의류	패션잡화	화장품/미용	디지털/가전	가구/인테리어	출산/육아	식품	스포츠/레저	생활/건강	여가/생활편의	면세점
투입한 카테고리 전체 키워드		64	150	134	365	166	352	277	243	431	58	61
대분류 코드 사용	추출된 키워드 개수	39	137	125	359	156	341	275	237	413	55	61
	NA 키워드 개수	25	13	9	6	10	11	2	6	18	3	0
	ratio 0 키워드 개수	0	0	0	3	0	6	8	1	2	8	22
	(NA + ratio 0)/전체	39.1%	8.7%	6.7%	2.5%	6.0%	4.8%	3.6%	2.9%	4.6%	19.0%	36.1%
중분류 코드 사용	추출된 키워드 개수	64	150	134	365	166	352	277	243	431	58	61
	NA 키워드 개수	0	0	0	0	0	0	0	0	0	0	0
	ratio 0 키워드 개수	0	0	0	3	0	9	12	4	13	14	23
	(NA + ratio 0)/전체	0.0%	0.0%	0.0%	0.8%	0.0%	2.6%	4.3%	1.6%	3.0%	24.1%	37.7%
선택 분류 코드		중분류	중분류	중분류	중분류	중분류	중분류	대분류	중분류	중분류	대분류	대분류

데이터 추출

- 기간 개수가 동일한 키워드의 평균 오차와 피어슨(Pearson) 상관계수 확인
- 전체 카테고리의 **평균 오차는 1.98**, **상관계수는 0.96**으로 산출, 이는 곧 **중분류 코드 데이터**와 **대분류 코드 데이터**의 **유사도가 높다**는 뜻
- 이 둘을 혼용하여 사용해도 무방하다고 판단, 따라서 **카테고리별 불량률((NA + ratio 0)/전체)이 낮은 쪽의 코드를 사용**하기로 결정

0 ≤ **평균 오차** ≤ 100
0 에 가까울수록 유사

-1 ≤ **피어슨 상관 계수** ≤ 1
1 에 가까울수록 유사

		패션의류	패션잡화	화장품/미용	디지털/가전	가구/인테리어	출산/육아	식품	스포츠/레저	생활/건강	여가/생활편의	면세점
투입한 카테고리 전체 키워드		64	150	134	365	166	352	277	243	431	58	61
대분류 코드 사용	추출된 키워드 개수	39	137	125	359	156	341	275	237	413	55	61
	NA 키워드 개수	25	13	9	6	10	11	2	6	18	3	0
	ratio 0 키워드 개수	0	0	0	3	0	6	8	1	2	8	22
	(NA + ratio 0)/전체	39.1%	8.7%	6.7%	2.5%	6.0%	4.8%	3.6%	2.9%	4.6%	19.0%	36.1%
중분류 코드 사용	추출된 키워드 개수	64	150	134	365	166	352	277	243	431	58	61
	NA 키워드 개수	0	0	0	0	0	0	0	0	0	0	0
	ratio 0 키워드 개수	0	0	0	3	0	9	12	4	13	14	23
	(NA + ratio 0)/전체	0.0%	0.0%	0.0%	0.8%	0.0%	2.6%	4.3%	1.6%	3.0%	24.1%	37.7%
기간 개수 동일 키워드		37	130	116	309	147	275	146	196	347	12	52
기간 개수 동일 키워드의 평균 오차		4.94	1.66	0.50	0.85	1.60	0.75	3.08	0.33	0.69	6.57	0.75
기간 개수 동일 키워드의 Pearson		0.89	0.97	0.99	0.98	0.96	0.98	0.90	0.99	0.98	0.91	0.99
선택 분류 코드		중분류	중분류	중분류	중분류	중분류	중분류	대분류	중분류	중분류	대분류	대분류

자, 이제부터
본격적인 분석

Step 04 / 심화 분석 접근 방향

이상치 **제거 기준을 이원화**하여 합리적인 분석 기법 탐색

- 앞서의 상관분석에서 바통을 이어받아, 통계적 기법을 적용하여 군집화(Clustering) 시도
- 이상치 소거 역시 **① 통계적 분석 기법을 적용한 이상치 검출 결과**와 **②연구자 판단에 따라 임의로 소거 기준을 설정한 결과** 두 개를 모두 적용하여 최종 분석 결과에 기반, 유의미하다고 판단되는 결과를 선택할 예정 -> **두가지 기준 모두 같은 분석 기준으로 수렴**
- <참고> 시계열 클러스터링 (Serial Clustering) : 48개월 동안의 항목별 클릭지수 추이들을 통해 유사한 그룹이 몇 개로 나뉘는지를 분석한 결과
  자사의 엠포스 데이터랩 게시물을 참조 (http://bigdata.emforce.co.kr/index.php/2021021901)

STep 04

시계열 클러스터링
(Serial Clustering)
기법을 통해 본격적 분석

통계적 분석 기법인
시계열 클러스터링 기법을
활용하여
몇 개의 그룹으로 묶이는지,
묶인 그룹 안에 어떤 상품군이
응집해 있는지 분석

Analysis ver. 01
- 분석 기법에 의해 이상치로 분류된 항목 제거

2019~2022년까지의 48개월 기간 중 월간 클릭량이 누락된 데이터를 소거하기 위해 통계적 분석 기법을 활용

Analysis ver. 02
- 월간 클릭량이 하나라도 없는 데이터 모두 소거

통계적 분석 기법을 적용하지 않고
48개의 월간 데이터 중 하나라도
누락 값이 있으면 해당 항목 전체 소거

유의미한 Cluster 개수 도출

Cluster 별 포함 항목 추출

포함 항목 기반 데이터 해석

두 가지 기준에 대해
하나의 결론으로 수렴

Final Analysis
- 월간 클릭량이 하나라도 없는 이상치로 분류된 데이터 제거
- 최종 분석 개수 : 1,597개

2019~2022년까지의 48개월 기간 중 월간 클릭량을 기준으로 통계적 기법을 이용하여 이상치로 판단된 상품군은,
곧 월간 데이터 중 하나라도 누락 값이 있는 상품군과 일치한다는 결과가 산출

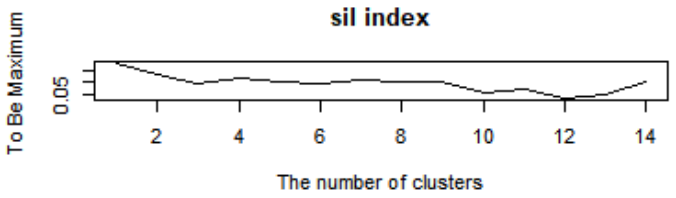
최적의 Cluster 개수 선정 방법

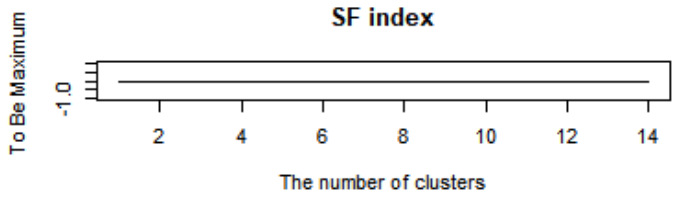
7가지의 다양한 Clustering(군집화) 평가 지수를 적용해 최적의 개수 추출

- 이상치를 제거한 220개의 쇼핑 항목이, 월간 쇼핑 클릭 추이에 따라 몇 개의 그룹으로 나뉘는지를 통계적으로 검증하기 위한 작업으로 7가지의 다양한 군집화 평가 지수 분석을 통해 공통된 Cluster 수를 산정
- (각 군집화 평가 지수에 대한 설명이 난해한 것은 어쩔 수 없음.. 기술적 용어에 대한 사전 이해가 필요하므로.)

Silhouette index

각 데이터 별로 그 데이터가 속한 **클러스터 내의 유사도**와 **인접한 클러스터와의 유사도**를 비교하는 지표



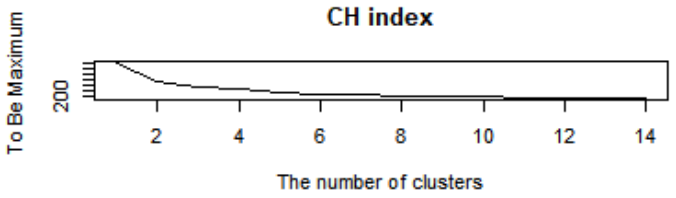


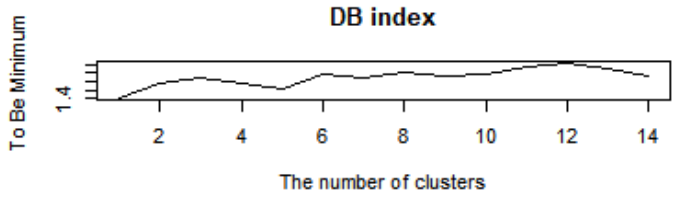
Score Function index

일반적인 **인공지능 연구에서 사용**하는 성능 지표. 클러스터간 거리를 score화 하여 거리를 함수로 계산하는 지표

Calinski-Harabasz index

데이터셋과 특정 클러스터간 중심거리를 **클러스터 간 중심거리의 분산 비율로 계산**하는 지표



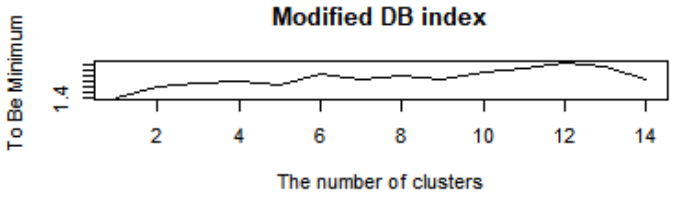


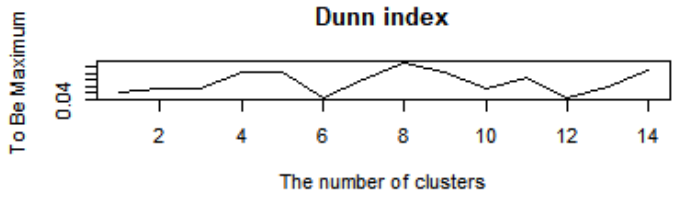
Davies-Bouldin index

특정 클러스터와 그것과 가장 유사한 클러스터 사이의 평균 거리로 계산하는 지표

Modified Davies-Bouldin index

수정된 Davies-Bouldin index라고 하며, **클러스터간 중심 거리의 최소값**을 구하는 수식이 포함된 지표



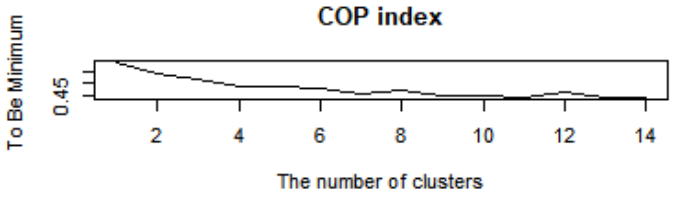


Dunn index

가장 유명한 평가 지표이며, Sil index와는 다르게 해당 데이터가 속한 클러스터 내의 유사도와 **모든 클러스터의 유사도를 비교**하는 지표

COP index

특정 클러스터의 중심과 해당 클러스터에 속해 있는 모든 데이터셋의 평균거리를 **가장 멀리 떨어진 클러스터의 거리 비율로 계산**하는 지표



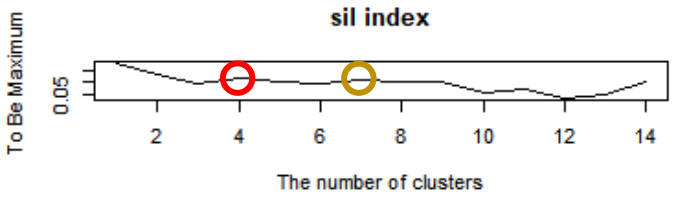
최적의 Cluster 개수 선정 방법

7개 중 5개의 평가 지표에서 유의미한 평가 분석 도출, 최종 7개의 Cluster 선정

- Silhouette, Davies-Bouldin, Modified Davies-Bouldin, Dunn, COP 지표 분석 상에서 공통적으로, Cluster가 5개와 7개로 나뉘었을 때 유의미함을 확인
- 보다 다양한 클러스터 선정으로 세분화된 상품군의 해석을 위해 7개의 Cluster로 선정
- Score Function, Calinski-Harabasz는 값의 변화가 없이 나타나거나, 우하향하여 Cluster 수를 도출하기 어려움

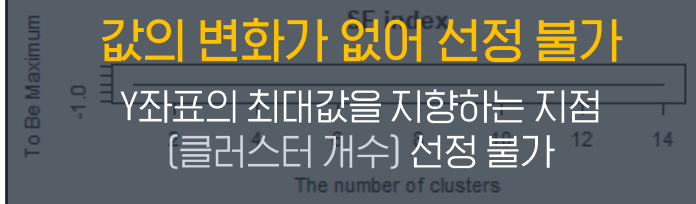
Silhouette index

각 데이터 별로 그 데이터가 속한 **클러스터 내의 유사도**와 **인접한 클러스터와의 유사도**를 비교하는 지표



값의 변화가 없어 선정 불가

Y좌표의 최대값을 지향하는 지점
(클러스터 개수) 선정 불가



Score Function index

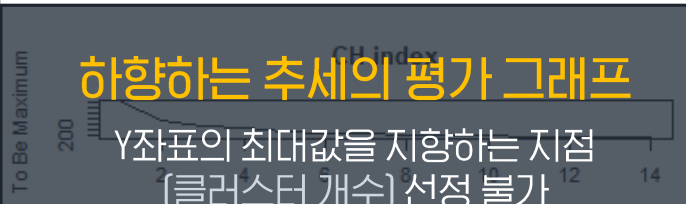
일반적인 **인공지능 연구에서 사용**하는 성능 지표. 클러스터간 거리를 score화 하여 거리를 함수로 계산하는 지표

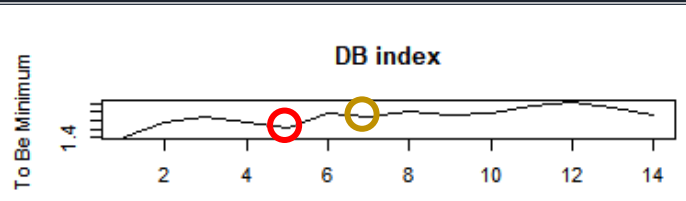
Calinski-Harabasz index

데이터셋과 특정 클러스터간 중심거리를 **클러스터 간 중심거리의 분산 비율로 계산**하는 지표

하향하는 추세의 평가 그래프

Y좌표의 최대값을 지향하는 지점
(클러스터 개수) 선정 불가



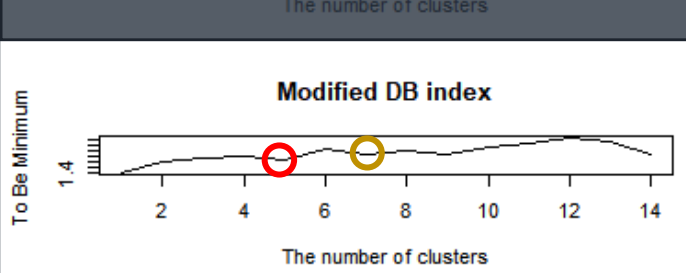


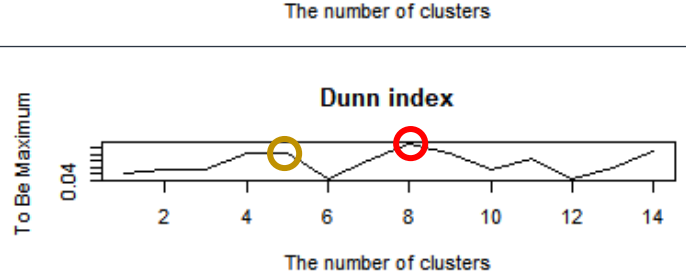
Davies-Bouldin index

특정 클러스터와 그것과 가장 유사한 클러스터 사이의 평균 거리로 계산하는 지표

Modified Davies-Bouldin index

수정된 Davies-Bouldin index라고 하며, **클러스터간 중심 거리의 최소값**을 구하는 수식이 포함된 지표



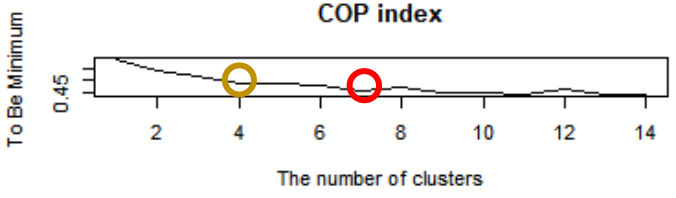


Dunn index

가장 유명한 평가 지표이며,
Sil index와는 다르게 해당 데이터가 속한 클러스터 내의 유사도와 **모든 클러스터의 유사도를 비교**하는 지표

COP index

특정 클러스터의 중심과 해당 클러스터에 속해 있는 모든 데이터셋의 평균거리를 **가장 멀리 떨어진 클러스터의 거리 비율로 계산**하는 지표

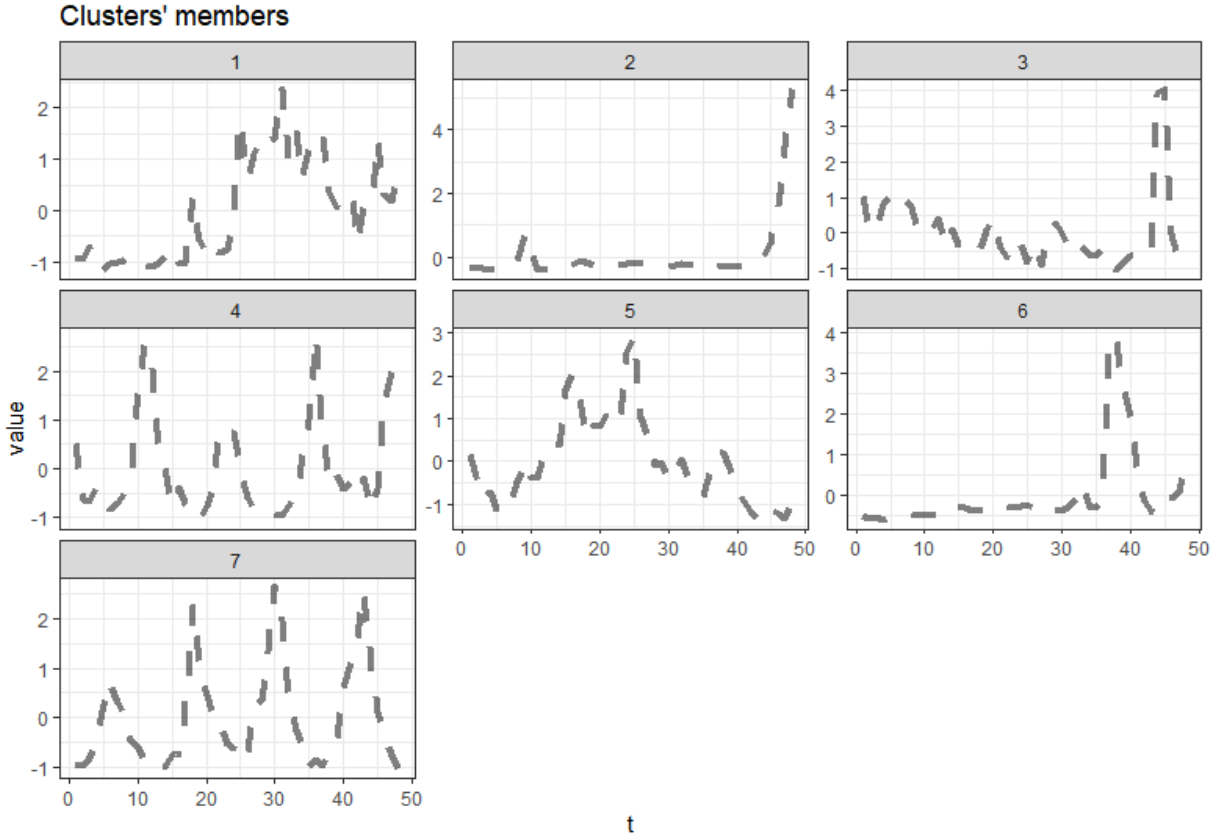


7가지 Cluster의 클릭지수 흐름

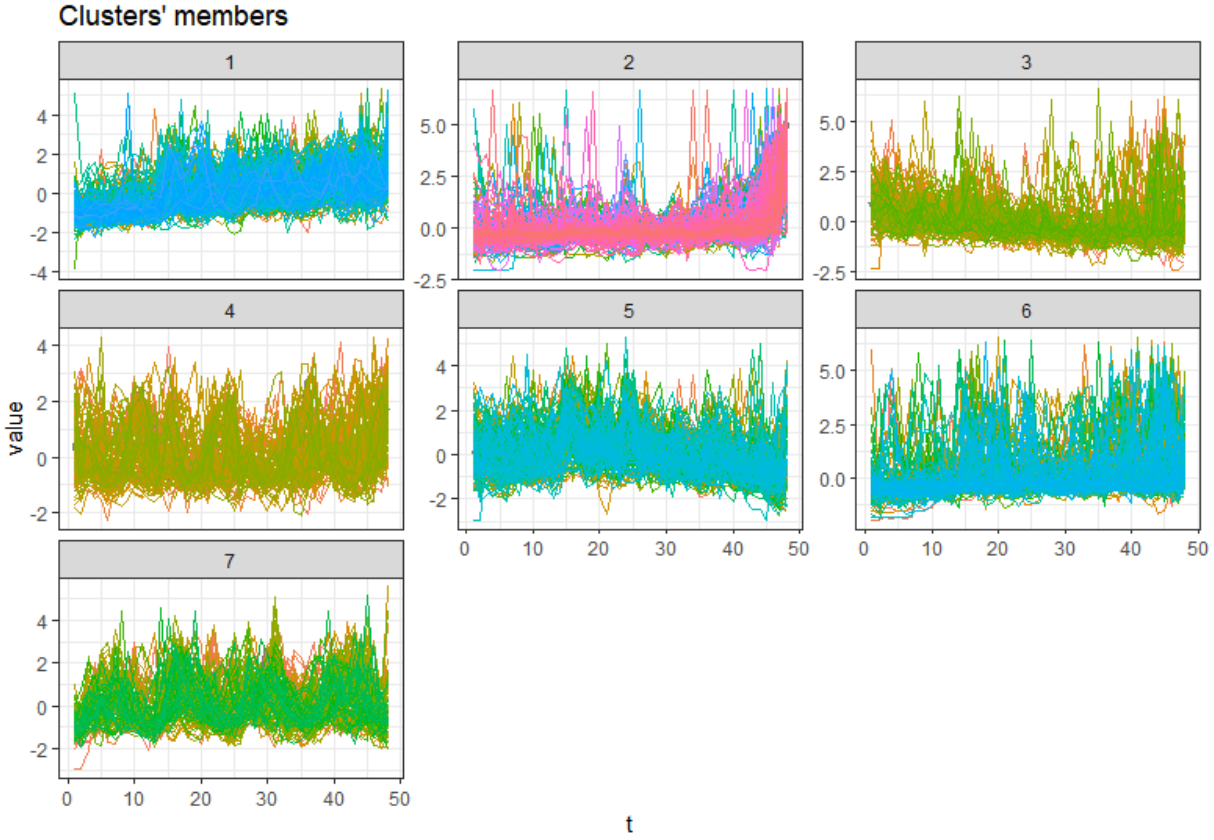
7개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

- 시즌성을 보이는 유형, 최근 시점에 증가하는 유형, 기간에 따라 급상승하는 유형 등으로 분류

[대표 추세 시각화]



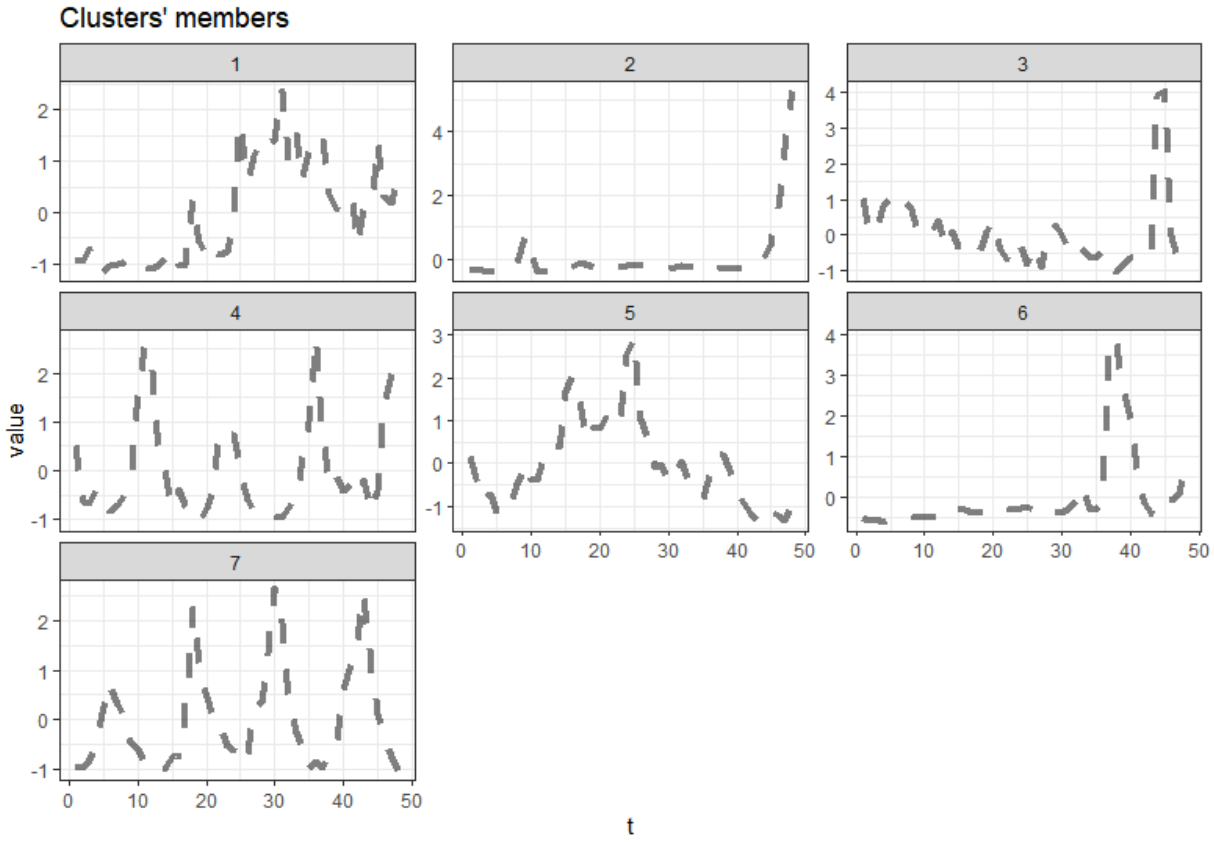
[Cluster별로 포함된 개별 키워드들의 추이를 중첩한 시각화]



7가지 Cluster의 클릭지수 흐름

7개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

- 시즌성을 보이는 유형, 최근 시점에 증가하는 유형, 가파른 상승 후 하락하는 유형 등으로 분류



Centeroids (각 클러스터의 중심점)

각각의 클러스터는 거리 측정 기준값이 있어야 하는데
해당 기준값이 되는 중심점을 Centeroids라고 일컬음.

Centeroids 선정 방법은 임의로 선정하며,
임의로 선정된 각 Centeroids로부터
가까운 거리의 개체들끼리 그룹을 형성

즉, 시계열 클러스터의 Centeroids는
"각 클러스터의 대표 분포 개체"이며,
해당 분포와 "비슷한 분포"를 가진 개체끼리 분류

7가지 Cluster의 클릭지수 흐름

7개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

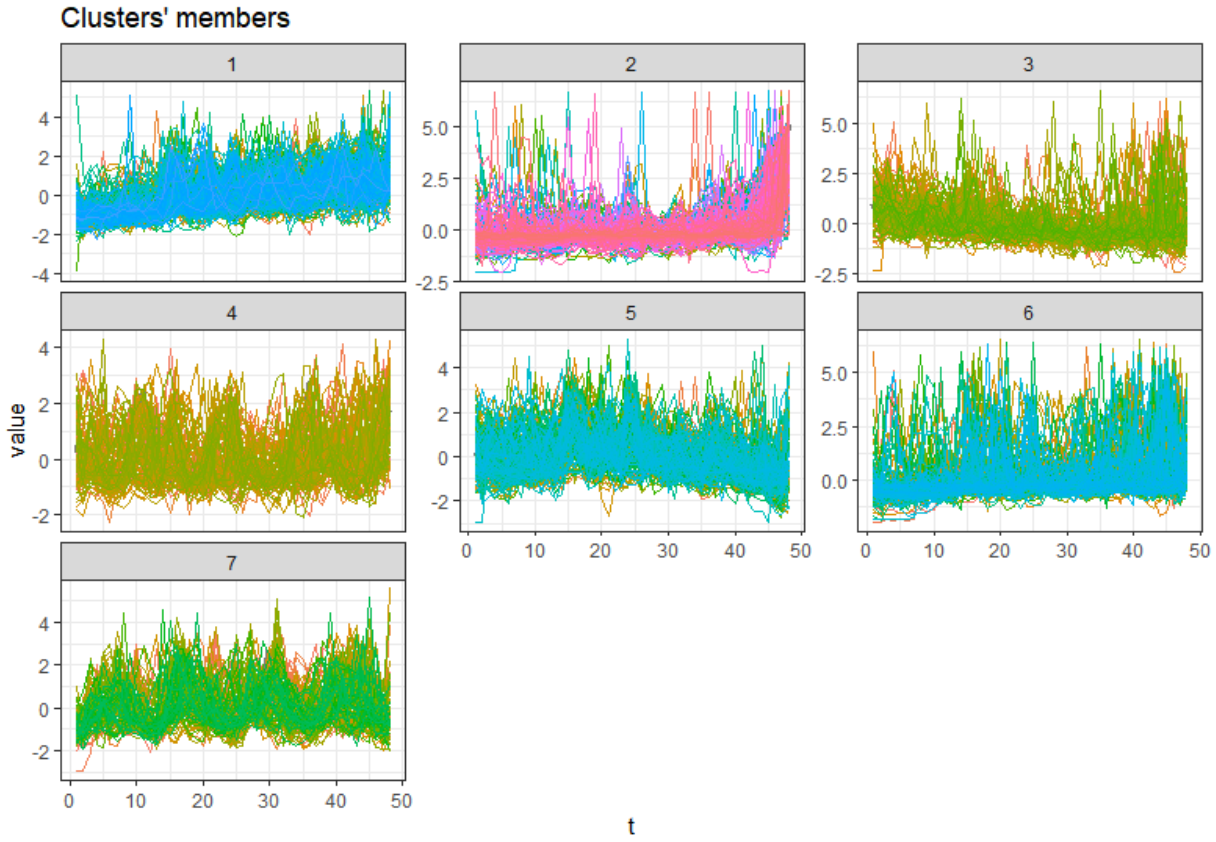
- 시즌성을 보이는 유형, 최근 시점에 증가하는 유형, 가파른 상승 후 하락하는 유형 등으로 분류

시계열 클러스터링 결과 그래프

시계열을 각 시간 인덱스에 대한 열이 있는 테이블로 평면화하고
클러스터링 알고리즘을 적용

Centeroids로부터 거리가 가까운 개체들끼리 묶여서
동일 클러스터에 형성

시계열에 따라 클러스터로 묶인 개체들의 데이터를
그래프로 확인 가능



이제부터는 도출된 7개 각각의 Cluster 안에
어떤 키워드들이 묶였는지 나열할 것이다.

깊게 설명하거나 해석하지 않고 단순히 '나열'만 할 것이다.

왜냐면, 각 Cluster는 조합된 키워드의 연관성 내지는 상관성에 따라
어떤 의미를 보일 수도 있고, 전혀 그렇지 않을 수도 있는데

키워드 간의 상호 연관성을 파악하려면,
즉, 왜 함께 판매가 증가하거나 감소하는지를 파악하려면

일일이 **개별 관계 분석**이 이뤄져야 하며
본 클릭 데이터만으로는 소비자 심리를 이해하는 데 한계가 있어
SNS 데이터 등을 통한 트렌드 반영 여부 등 검증이 필요한데,

이 부분은 아쉽게도 본 결과물에는 담기지 않았다.

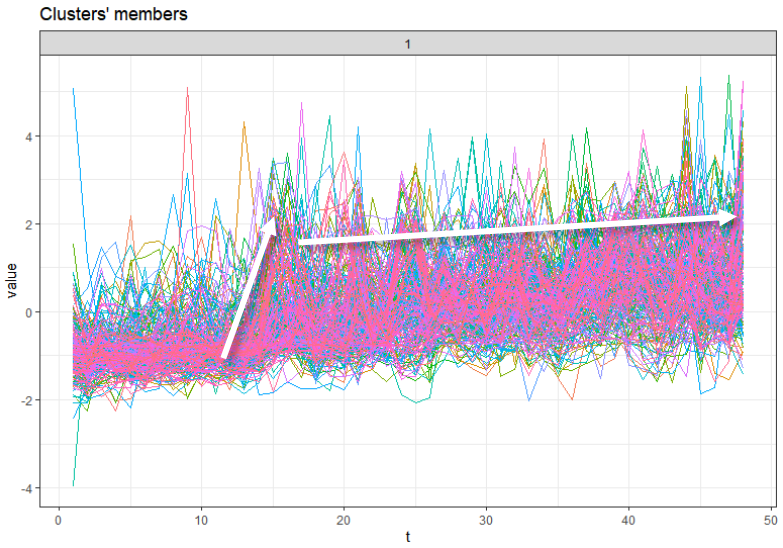
< Cluster 분석 결과 해석 예시 >

	그룹 유형	분석 결과 설명
Case 01	Seasonality Keyword Group	특정 시즌에 많이 판매되는 상품들이 묶인 경우로, 당연히 의미 있지만, 아마 마케팅을 하는 담당자라면 익히 알고 있고 예상 가능한 결과일 것이라 새롭지는 않을 것
Case 02	New Trend Catching Group	기존에 연관성을 예상하기는 어려웠지만 새로운 트렌드에 따라 함께 판매되는 경우인데 분석 결과만 가지고 새로운 트렌드인지 여부를 판단할 수 없어 일일이 추가 분석이 필요한 영역
Case 03	Disrelation Group	연관성이 전혀 없지만 같은 시기에 판매가 증가/감소하는 경우로 당장 어떤 액션을 취할 필요는 없지만 나중을 위해 적당한 모니터링은 필요할 듯

Cluster 01 – 총 282개의 키워드 중 상위 50개 키워드

New Trend Catching Group

- 코로나19로 인해 집 밥에 대한 수요가 늘면서 수요 급증한 영역



돈가스소스	햄스터용품	드레스	메이크업 브러시	우유
나비장	고가구	나박김치	마요네즈	안경소품
빵가루	옷커버	스킨	건과류	보행보조용품
배드민턴라켓	순금돌반지	물엿	봉투	운동화
건어물	업소용 자외선소독기	쿠키	떡	젓갈류
작동완구	우표	소품걸이	이유식재료	토트백
diy아기용품	수의	도어벨	캐릭터인형	수첩
원예공구	화분영양제	국	정장세트	짜장
굴소스	틀니관리용품	필기도구	불교용품	솥
유치보관함	푸딩	유아변기커버	테니스가방	미트

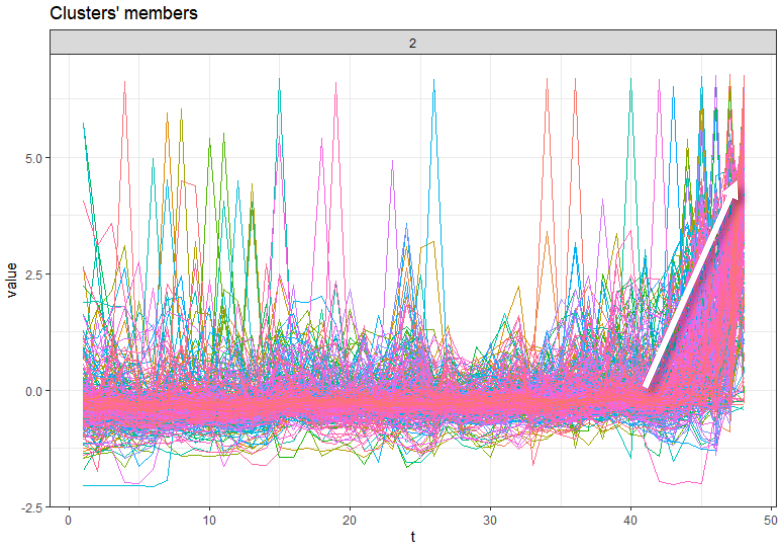
*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

식량 비축 & 요리 취미

Cluster 02 – 총 431개의 키워드 중 상위 50개 키워드

New Trend Catching Group

- 뷰티용품과 운동용품 및 야외활동과 관련된 상품군 형성
  식품 관련 상품군(1번 클러스터와는 다른 추세)
- 2022년 중순부터 크게 상승하는 추세



즉석밥	로스트볼	안전문	하이라이트	베개솜
건강팔찌	머플러	스파게티	침낭	생선
헤어글레이즈	머스타드소스	시서스	속눈썹영양제	머리보호대
핸드드라이어	잡곡	김	셀카봉	전기그릴
즉석카메라	tv거실장	음식물처리기	패션시계	유아마스크
네일케어세트	치석제거기	이발기	브라	선크림
천연감미료	쌀	욕실매트	수련화	공기정화식물
진공포장기	축구공	아령	콧물흡입기	렌즈세정액
스티커	신발장	조화	피규어	칫솔
휴대폰케이블	스쿼시라켓	속싸개	타격대	콤비블라인드

*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

엔데믹의 도래로
본격적인
외출 준비 시작

Cluster 03 – 총 125개의 키워드 중 상위 50개 키워드

Disrelation Group

- 유아용품관련 상품군이 상위에 다수 포진해있기는 하지만
  특정 연관성으로 묶였다기 보다 대부분 점진적으로 하락하는 추세의 상품들

장롱	유아면봉	유아욕조	수액시트	볼링의류
디지털도어록	보닛	아기띠패드	보낭	유아포크
면	임산부 바디필로우	휴대폰렌즈	제수용품	발패드
한복	쌍안경	패널	심박계	머니클립
산모방석	헤드유닛	하이패스	구둣주걱	환자식
바운서	점퍼루	수영가방	유아변기	앞치마
노리개젖꼭지	노트북 도난방지	여권지갑	호각	카메라렌즈
글러브	비치웨어	레몬밤	바디용품	국악기
볼링화	얼룩제거제	야구보호대	쉐이딩	기저귀정리함
재봉틀	어학학습기	붓	낚싯줄	잠금벨트

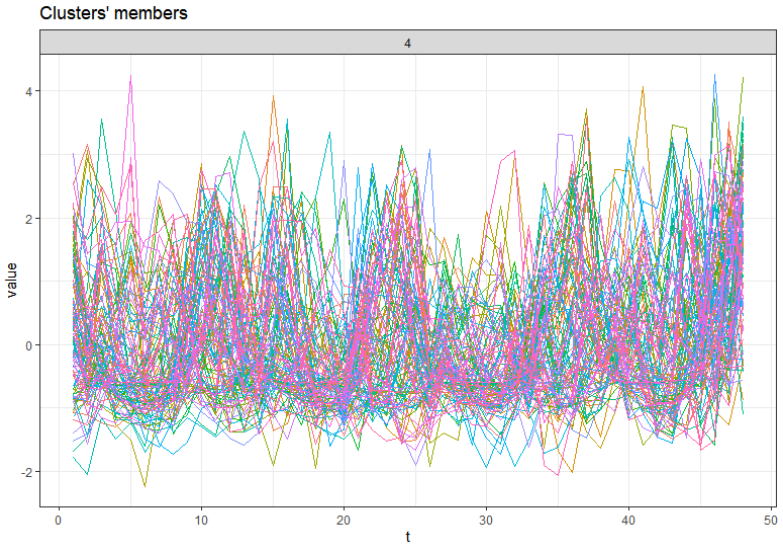
*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

유아용품 & 출산선물

Cluster 04 – 총 106개의 키워드 중 상위 50개 키워드

Seasonality Keyword Group

- 연말(겨울)에 튀는 시즌성을 보임
  (겨울과 관련된 상품군들 다수 형성)



매트솜	축구보호대	수납장	통장지갑	업소용 음식물처리기
하이라이터	우주복	업소용반죽기	부츠키퍼	클렌징로션
아동시계	티켓	슬립	천체망원경	이불커버
헌팅캡	스키복	수중총	파우더	다이어리
전기쿠커	니트	홍삼제조기	스키용품	스웨터
스프	족구화	구강청정기	주얼리소품	털신
키재기	타이즈	보틀워머	남방	클렌징티슈
남성수영복	와펜	세탁볼	김치냉장고	클렌징세트
과실주	유아세탁비누	pc스피커	마사지크림	스포츠 선글라스
제습기	타투	코트	붕대	좌훈기

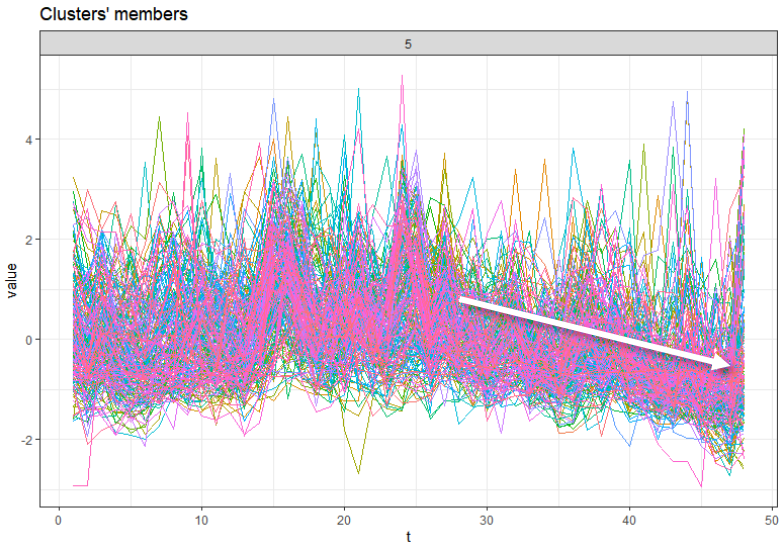
*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

와~ 겨울이다~~

Cluster 05 – 총 241개의 키워드 중 상위 50개 키워드

New Trend Catching Group

- 2020년 말, 정점을 찍고 하락 추세



잡지꽂이	유아두유	수족관	공간박스	자동차모니터
서버	프린터공유기	카메라그립	캔들용품	낚싯대
민물낚시	수면팩	폼보드	악기놀이	승용완구
가검	헤어핀	절삭공구	반지	여성지갑
몰딩	미니화분	일립티컬	데코스티커	키홀더
쉐이빙폼	보행기	임산부보호대	헤어롤	햄
쏘서	찹쌀가루	칠리	모뎀	가정용게임기
목검	수유복	핫플레이트	숯부작	녹즙기
수예용품	버터	목재	젤리	안경줄
유아손톱가위	캠코더	미술작품	코너장	올리브

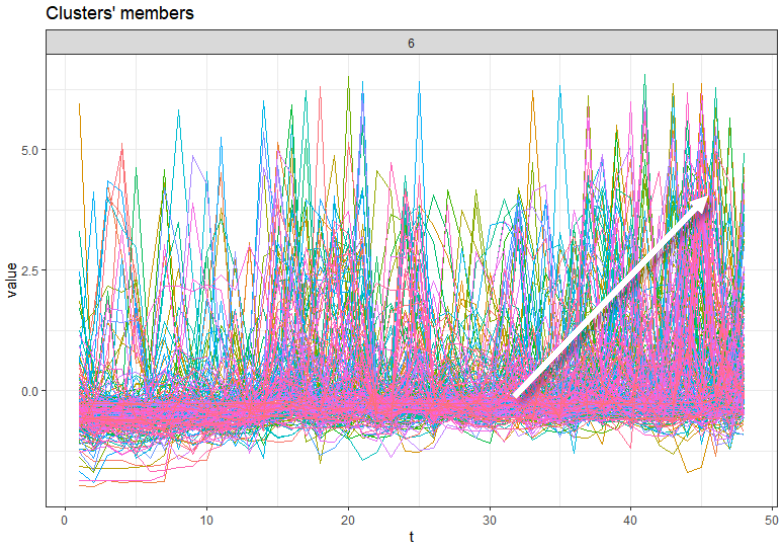
*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

in the house~
Interior

Cluster 06 – 총 254개의 키워드 중 상위 50개 키워드

Disrelation Group

- 일부 상품군들에 한해 2021년 말 큰 상승



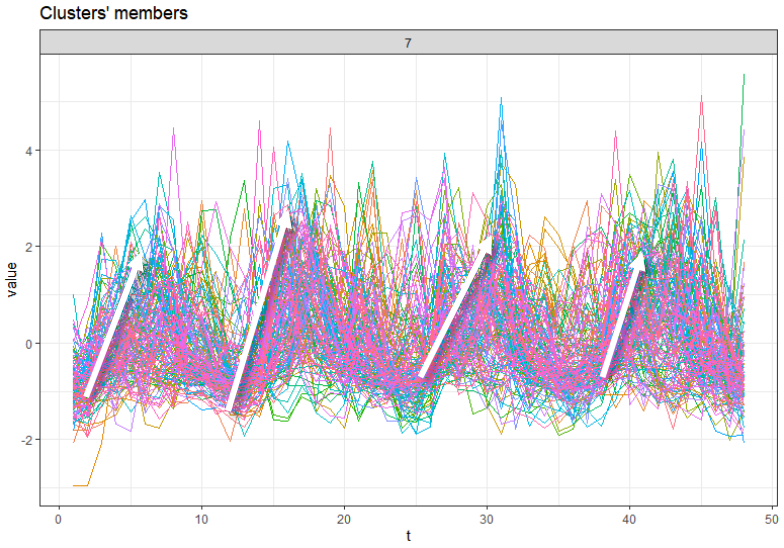
운반용품	인덕션	디지털액자	건강환	캐리어소품
팔보호대	웹캠	생활선물세트	쿠션안마기	귀달이모자
부부테이블	헤어왁스	발냄새제거제	스테이크소스	유아유연제
토스터기	죽	아이케어	유아입욕제	코클리너
즉석탕	해초	nas	캐릭터카드	모카신
퍼즐	몸통보호대	된장	린스	pc케이블
호떡믹스	남성화장품 세트	응급용품	저주파패드	스크럽
장난감소독기	소금	커튼액세서리	안전잠금장치	욕조
잔	코팩	아로마테라피	밀가루	비료
키보드	헤어미스트	유아 이유식의자	cd플레이어	냉풍기

whatever~

*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

Cluster 07 – 총 158개의 키워드 중 상위 50개 키워드
Seasonality Keyword Group

- 봄, 여름에 튀는 시즌성을 보임
  (여름과 관련된 상품군들 다수 형성)



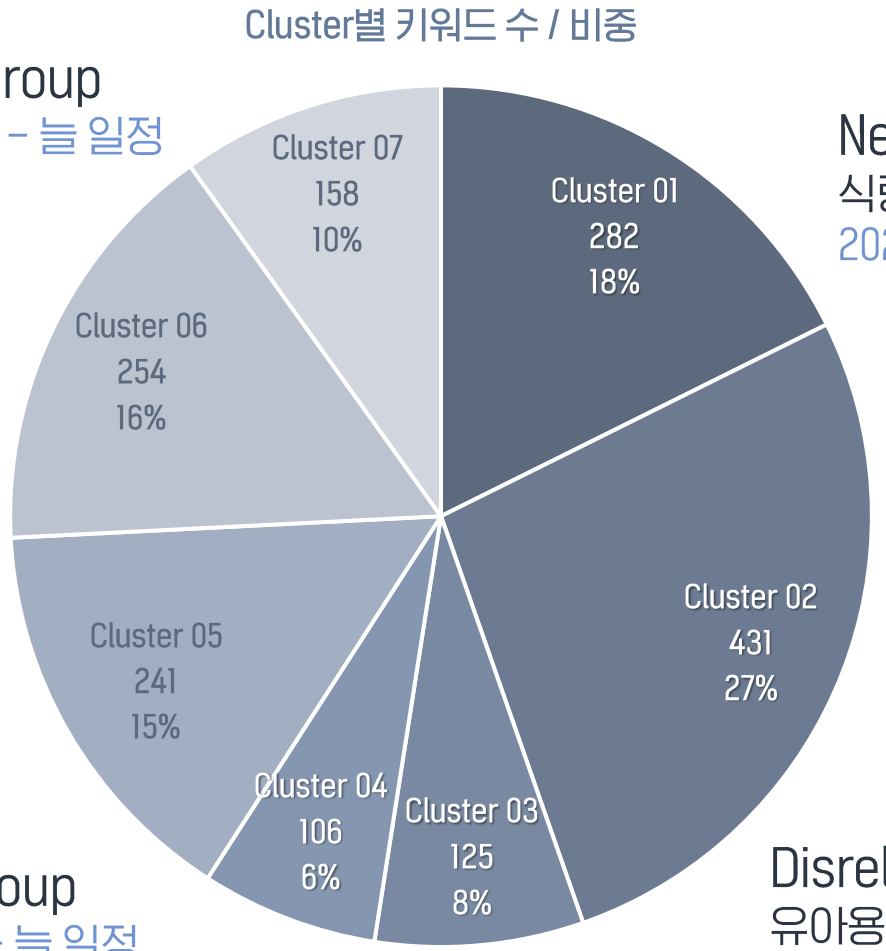
햄버거	구두	흙	선글라스 케이스	모종
죽도	카시트용품	팔찌	야구가방	허브식물
스커트	캐미솔	포수장비	간장	낚시용품
딱지	커피용품	아이스머플러	테니스의류	설탕
농구공	샌들	업소용빙수기	에어컨	자전거의류
아이스크림	부채	태닝	빙수기	공간놀이기구
야구양말	홑이불	페도라	선캡	화분
잔티젠	속바지	골뱅이	원피스	물조리개
재활운동기구	야외벤치	냉동고	정원그네	손수건
스네이크보드	화폐	매니큐어	타프	오토바이의류

*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

와~ 여름이다~~

요약 : Cluster 별 키워드 분포 등 유형 정리

만약 우리 사이트에 유입된 키워드들이
아래와 같은 유형으로 구분이 되었다면 어떨까?



Seasonality Group
와~ 여름이다~~ – 늘 일정

New Trend Catching Group
식량 비축 & 요리 취미
2020 코로나로 급등 이후 꾸준히 유지

Disrelation Group
Whatever – 최근 상승

New Trend Catching Group
in the house~ – 요즘 인기 시들

New Trend Catching Group
본격적인 외출 준비 시작
2022년 중순부터 급등

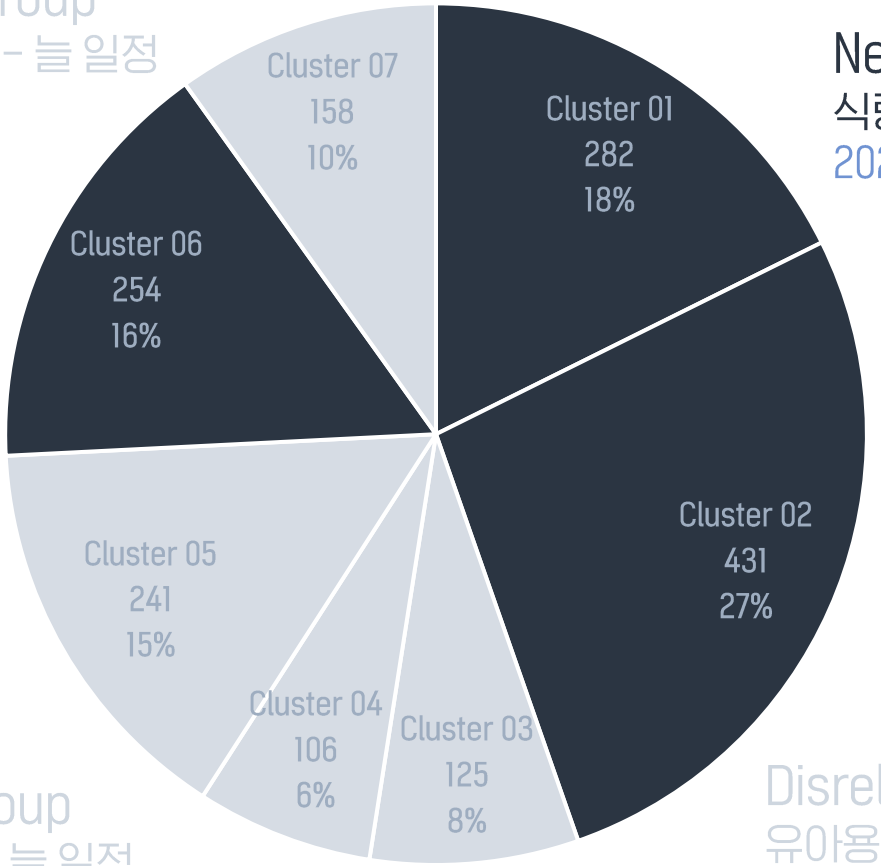
Seasonality Group
와~ 겨울이다~~ – 늘 일정

Disrelation Group
유아용품 & 출산 선물 – 점진적 하락

요약 : Cluster 별 키워드 분포 등 유형 정리

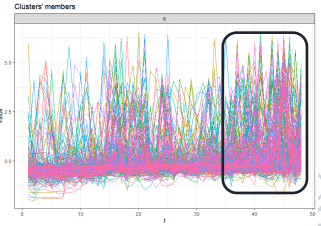
키워드 조합 유형, 키워드 종류, 최근 상승 여부 등을 기준삼아 다음의 마케팅, 브랜딩 활동을 기획해 볼 수 있을 것이다.

Cluster별 키워드 수 / 비중

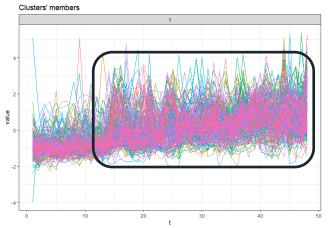


Seasonality Group
와~ 여름이다~~ – 늘 일정

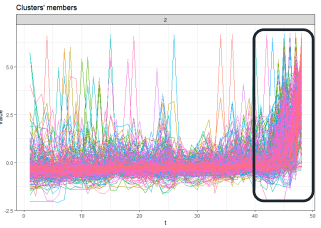
Disrelation Group
Whatever – 최근 상승



New Trend Catching Group
식량 비축 & 요리 취미
2020 코로나로 급등 이후 꾸준히 유지



New Trend Catching Group
본격적인 외출 준비 시작
2022년 중순부터 급등



Disrelation Group
유아용품 & 출산 선물

Seasonality Group
와~ 겨울이다~~ – 늘 일정

이렇게 군집 분석을 통해 키워드가 묶인 이유를 살펴보면

실제 소비 품목의 유사성이나 트렌드 연관성에 따라 묶이는 경우도 있지만
단순히 해당 품목들의 소비가 최근 증가, 감소함에 따라 묶이는 경우도 존재한다

수많은 키워드들의 소비 패턴을 알고 싶어 시작한 일이 본의 아니게,
최근 달라지고 있는 소비 행동을 보여주기도 한다는 의미다.

그게 시계열 클러스터링의 묘미다.

본 보고서에서는 **쇼핑클릭** 데이터를 기반으로 이 같은 현상을 설명했지만
온라인 광고 데이터에서는 이러한 패턴이 더욱 분명하게 드러날 수 있다.

마케팅/광고 시장이 오프라인에서 온라인으로 넘어오면서
전과는 비교도 안 될 정도로 수많은 고객 행동 데이터를 볼 수 있게 되었는데

활용 데이터 종류는 더 단출해졌고 / 분석 과정은 말도 안 되게 단순해졌다.

활용 데이터의 종류가 단출해지고 분석 과정이 단순해지려면
훨씬 복잡한 분석과 난해한 해석 과정이 반복적으로 **선행**되어야 한다.

지금은 데이터를 복잡하게 봐야 할 때이다.
설사 그 끝에 결국 ROAS 밖에 없더라도.

이제 우리는
진짜 광고 운영 데이터를
분석하러 간다

End Of Document

Contact Us

Website URL http://bigdata.emforce.co.kr

T 02. 6177. 1871 eMAIL khbak@emforce.co.kr