

Jan 2023. eMFORCE Report

2019-2022 대한민국 쇼핑트렌드 1탄  
2023년, 쇼핑 유발 마케팅의  
**NEW PARADIGM**



**eMFORCE 데이터랩**  
Digital Marketing Group



혹시, 이런 얘기  
들어 보셨어요?

“기저귀 옆에 맥주를 두면 잘 팔린다”



# “나이키의 상대는 닌텐도다”



A man in a grey suit, white shirt, and light blue tie is looking upwards and to the right with a confused expression. His hands are outstretched in a questioning gesture. The background is white with a large, faint grey circle around him.

**What the..**

“기저귀 옆에 맥주”는 기저귀 구매자들이 육아 스트레스로 맥주를 함께 구매하는 경우가 많다는 얘기고

“나이키의 상대가 된 닌텐도”는 나이키와 리복이 아닌, 지출 비용 관점에서 나이키와 닌텐도 게임기를 고민한다는 얘기인데

둘의 공통점은, 서로 다른 카테고리 간의 **관계성**이라는데 있다. 그것이 동반 구매로 상호 작용하는 것이든, 경쟁 관계든 말이다.

...

영업에서 **마케팅**으로, 마케팅에서 **브랜딩**으로 판매전략이 진화함에 따라 소비자들에게 던지는 기업의 질문도 “무엇을” → “왜” → “어떻게”로 바뀌었으며  
[뭘 샀니? 왜 샀니? 어떻게 할거니? 어떻게 생각하니? 어떻게? 어떻게?]

질문을 해소하는 데이터의 역할도 ‘무엇’과 ‘왜’를 묻는 **여론조사**에서 소비자 자신도 모르는 ‘어떻게’를 밝혀내는 **빅데이터**로 진화했다.

온라인 환경에서 생성된 쇼핑 데이터(빅데이터)는, 부지불식간에 실시간으로 벌어지는 자발적 행동을 담고 있으므로 ‘무엇’과 ‘왜’보다 **어떻게**를 설명하는데 더 적합하고 가깝다.

...

그래서 지금부터는, <**온라인 쇼핑 클릭 지수**>라는 빅데이터를 활용해 어떤 상품들이 서로 함께 관심을 받는지를 세밀하게 살펴 볼 것이다.

물론 누가봐도 연관성 높아 보이는 상품 간 관계는 관심이 없을 것이다. 우리의 관심사는 **기저귀 옆에 맥주**, **나이키의 상대가 된 닌텐도**일 테니까.

분석을 얘기하기에 앞서, 한 가지 더 미리 밝혀둘 것이 있다.

우리는 <쇼핑 클릭 지수> 데이터를 활용한 분석 과정에서  
몇가지 검증하기 어려운 **오류**를 발견했다.

간단히 말하면, 일부 쇼핑 항목에서 API로 호출한 데이터 값들과  
네이버 데이터랩 사이트에서 확인되는 데이터 값들이 일치하지 않는 것인데

아마도 [그저 막연한 추측이지만..]

네이버 자체적으로 **유행에 따라 쇼핑 카테고리를 추가하고 이동하는 과정에서**  
일종의 '간극'이 발생한 것 같은데, 일단은 “밝혀내기”보다 **“무시하기”**를 택했다.

본 보고서에서 보여주고 싶었던 건 선명한 트렌드가 아니라  
**트렌드에 이르기까지의 접근 방식과 분석 방법론**이니까.

...

그래서, 여러분도 본 분석 결과에 대한 감론을박 보다는  
분석 과정 자체에 대한 흥미가 생기셨으면 좋겠다.

우리가 하고 있는 시도는 사실, 트렌드 설파가 아닌  
**온라인 광고 운영 데이터**를 향하는 길목이니까.

# 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, **중분류**를 기준으로 분석되었으며 **소분류 데이터를 활용한 분석은, 이후 최종 분석 보고서를 통해 공개 예정**  
 - 대분류는 11개, 중분류는 269개, 소분류는, ... 몇 천 개..

항목	내용
데이터 종류	네이버 쇼핑클릭 데이터 - URL : <a href="https://datalab.naver.com/shoppingInsight/sCategory.naver">https://datalab.naver.com/shoppingInsight/sCategory.naver</a> - 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 행동을 가장 근접하게 보여주는 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	중분류 기준 269개 상품군 분석 활용

대분류 [11개]	중분류 [269개]
가구/인테리어	diy용품
디지털/가전	diy자재
면세점	dvd
생활/건강	pc
스포츠/레저	pc부품
식품	pc액세서리
여가/생활편의	가루
출산/육아	가루류
패션의류	간편조리식품
패션잡화	거실가구
	건강
	건강관리용품
	...

\*출처: 나무위키



# 데이터 분석 접근 방향

- 월별 클릭지수를 활용, 통계적 분석 기법인 **시계열 클러스터링(Serial Clustering)**을 적용하기로 결정 후 일부 중분류 항목 중 **몇몇 기간에 데이터값이 누락, 혹은 존재하지 않는 경우를 발견**하고 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 데이터값이  
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 그룹화 될  
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

본격적인 통계적 분석에 앞서  
**분석 가능성을 진단**

Step 02 ~ Step 03



# 소거과정에서 살아 남은 232개의 데이터들을 기반으로 상관관계(Correlation) 분석을 통해 항목별 상관성 측정

- 상관관계 분석은 48개월 동안의 추세를 기반으로 측정되므로 항목 간 상관성이 높다는 것은, 48개월 동안 유사한 흐름으로 상품이 클릭되고 있음을 추정 가능

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12	Column 13	Column 14	Column 15	Column 16	Column 17	Column 18	Column 19	Column 20	Column 21	Column 22	Column 23	Column 24	Column 25		
2	Column 1	1																									
3	Column 2	0.149044	1																								
4	Column 3	0.14338	0.147849	1																							
5	Column 4	-0.20629	-0.18862	-0.16806	1																						
6	Column 5	-0.07525	-0.15491	-0.07495	-0.16359	1																					
7	Column 6	-0.07599	-0.22984	-0.11031	0.050277	0.023686	1																				
8	Column 7	-0.04639	0.127886	-0.13181	-0.14213	0.053515	-0.06947	1																			
9	Column 8	-0.05304	0.046736	-0.02186	-0.18159	-0.2339	-0.22634	0.093605	1																		
10	Column 9	-0.00369	-0.02019	-0.1289	0.036003	0.342006	-0.03362	-0.07893	-0.14327	1																	
11	Column 10	0.46201	-0.39607	-0.1043	0.093925	0.475238	0.155269	-0.15476	-0.60382	0.360348	1																
12	Column 11	0.04795	0.197372	-0.0256	-0.09051	-0.04297	-0.14636	-0.09741	0.353566	0.00162	-0.32391	1															
13	Column 12	0.23353	0.173524	0.149803	-0.01113	0.027301	-0.02768	-0.0821	-0.12738	-0.02142	0.012615	-0.05897	1														
14	Column 13	-0.03396	0.267595	-0.0235	0.028107	-0.07633	-0.06146	-0.01398	-0.13091	0.681959	-0.06584	-0.0741	-0.01552	1													
15	Column 14	0.54974	0.588178	0.164625	-0.17095	-0.04511	-0.23661	0.35871	0.512524	-0.05911	-0.45446	0.400273	-0.05444	0.06353	1												
16	Column 15	0.13113	0.403816	-0.08684	-0.16783	-0.01353	-0.05689	0.17626	0.151488	-0.02181	-0.17318	0.462603	-0.00821	0.0547	0.461307	1											
17	Column 16	-0.01212	-0.01256	-0.20298	0.264221	-0.17061	0.274084	-0.0637	-0.54749	-0.06693	0.212356	-0.40828	-0.06088	0.26455	-0.35321	0.29978	1										
18	Column 17	-0.01798	-0.0376	0.000729	-0.00409	0.742928	0.062059	0.068315	-0.50348	0.221183	0.533087	-0.29978	-0.00537	0.0353	-0.22426	0.39978	0.29978	1									
19	Column 18	-0.05121	-0.30042	-0.12886	0.211631	0.40505	0.061111	-0.09697	-0.32043	0.470621	0.663651	-0.14556	0.091029	-0.1599	-0.22426	0.39978	0.29978	-0.12644	1								
20	Column 19	0.57149	-0.01019	0.06819	0.134398	0.254278	0.011003	-0.13794	-0.47009	-0.04051	0.522637	-0.37978	-0.10199	-0.0304	-0.12644	-0.18272	0.39978	0.29978	-0.12644	1							
21	Column 20	0.0564	-0.13863	-0.16466	0.123541	-0.00971	-0.04044	0.028487	-0.26642	-0.16268	0.134486	-0.19249	-0.02938	-0.0528	-0.18544	0.010111	0.39978	0.29978	-0.12644	-0.18544	1						
22	Column 21	0.10776	0.087617	-0.02755	0.210198	0.025575	0.016161	-0.03973	-0.26285	0.369101	0.254659	-0.21986	-0.00077	-0.0499	-0.16761	0.010111	0.39978	0.29978	-0.12644	-0.18544	0.010111	1					
23	Column 22	-0.08385	-0.03309	-0.00515	-0.13142	-0.0672	-0.04252	-0.0656	0.081954	-0.02514	-0.11065	0.500572	-0.04904	-0.0511	0.007434	0.010111	0.39978	0.29978	-0.12644	-0.18544	0.010111	-0.07889	1				
24	Column 23	0.53526	-0.01534	0.074608	-0.03007	0.225643	0.046028	0.091908	-0.24147	0.111363	0.425944	-0.09263	0.152631	-0.0222	-0.08388	0.010111	0.39978	0.29978	-0.12644	-0.18544	0.010111	-0.07889	0.010111	1			
25	Column 24	0.02201	-0.19651	-0.25407	-0.0121	-0.2047	0.004319	0.26959	0.363535	0.012518	-0.08725	0.369061	-0.18503	0.07984	0.146731	-0.04329	-0.08413	-0.33114	-0.23889	-0.26812	-0.12767	-0.3157	0.050913	-0.0754	0.035734	1	
26	Column 25	-0.03381	-0.05261	-0.07973	0.414277	-0.08907	-0.09097	0.008177	-0.18685	-0.09192	-0.05762	-0.185	-0.00374	0.1532	-0.23731	-0.09331	0.231361	-0.02846	-0.14692	0.04481	0.01349	-0.09339	-0.0864	0.00389	0.035734	0.035734	1
27	Column 26	-0.03623	-0.07309	-0.03773	0.068921	-0.01669	-0.00911	-0.05492	-0.16488	0.041336	0.438808	-0.06097	0.014938	0.12513	-0.04446	-0.03528	0.289558	0.099462	-0.05775	0.383697	-0.03349	-0.00346	-0.01538	-0.0138	0.312756	0.007025	0.007025
28	Column 27	-0.06596	-0.36817	-0.20413	-0.01126	0.788092	0.102063	-0.08675	-0.24872	0.556484	0.631364	-0.09348	-0.03962	-0.0934	-0.28623	-0.08608	-0.194	0.492384	0.676445	0.117235	-0.04753	0.325587	-0.07138	0.13997	-0.2389	-0.16129	-0.16129
29	Column 28	-0.0218	-0.36025	-0.19237	0.024305	0.785308	0.168584	-0.11442	-0.41504	0.47618	0.653504	-0.1854	-0.0086	-0.0841	-0.36895	-0.13157	0.005526	0.576554	0.59014	0.171265	0.065501	0.315795	-0.07674	0.06369	-0.303	-0.07724	-0.07724
30	Column 29	-0.02585	0.224154	-0.10509	0.068682	-0.01424	-0.05989	0.060427	-0.13346	0.71849	-0.04078	-0.02333	-0.02866	0.94517	-0.22415	-0.01424	-0.05989	0.060427	-0.13346	0.71849	-0.04078	-0.02333	-0.02866	0.94517	0.177455	0.149317	0.149317

48개월 동안의 클릭 추세를 기반으로 카테고리 간 '상관관계(Correlation)'를 분석  
 동일 항목 간 상관성은 "1"  
 상관성이 높을수록 양(+)'의 값을 가지며  
 상관성이 낮을수록 음(-)'을 값을 가짐

# 항목별 상관성이 높게 측정된 상품 카테고리들을 정렬한 결과

- 상관성이 낮게 측정된 음(-)의 값들은 제외한 결과
- Ex. 「간편조리식품」과 상관성이 높게 측정된 항목은, 주니어가구, 수납용품, 아동가구, 신생아의류... 순이며

이를 활용해 해석해 보면 ① 간편조리식품 구매층 특징(Profiling)을 타 품목 클릭 행태로 유추 가능 → “간편식 구매자들은 어린 자녀를 둔 부모층”  
 ② 간편조리식품 구매 증가시기에 함께 수요가 증가할 수 있는 상품군 유추 가능 → “간편식과 주니어가구를 함께 마케팅한다면?”

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	간편조리식품	거실가구	게임기	과자	구강청결용품	기저귀	김치	남시	냉동	농산물	당뇨관리용품	댄스	등산	러그	모니터	모니터주변기기	목욕용품	물티슈	밀키트	바디용품	반려동물	반찬
2	주니어가구	침구단품	정리용품	세탁용품	반려동물	러그	러그	등산	주방가전	냉동	교구	카페트	축구	김치	김치	장갑	캠핑	러그	전통주	카메라	구강청결용품	커피
3	수납용품	수예	태블릿pc	블루레이	악기	김치	기저귀	축구	정리용품	캔	세탁용품	네트워크장비	남시	양말	러그	밀키트	서재	기저귀	김치	주니어가구	면류	재활운동용품
4	아동가구	뷰티소품	수납	물티슈	면류	물티슈	벨트	수예	보드	주방가전	배드민턴	커피	수예	잠옷	유아세제	전통주	스킨케어	양말	커피	신생아의류	악기	종교
5	신생아의류	축구	수납용품	유아동잡화	생활용품	양말	잠옷	뷰티소품	농산물	소독	아기간식	pc부품	뷰티소품	커피	커피	커피	음료	김치	이유식	수납용품	생활용품	좌욕용품
6	여성언더웨어	아웃도어가구	주방용품	당뇨관리용품	홀데코	커피	블라인드	검도	보드	과자	이유식	검도	블라인드	네트워크장비	김치	블라인드	물티슈	아동가구	축산물	홀데코	러그	이유식
7	헤어케어	침실가구	주니어가구	축산물	잠옷	양말	침실가구	주변기기	정리용품	수집품	러그	침실가구	기저귀	반찬	재활운동용품			반찬	pc부품	간편조리식품	홀데코	이유식
8	카메라	수영	냉동	사무용가구	순금	블라인드	물티슈	수영	캔	요가	사무용가구	종교	아웃도어가구	물티슈	장류	종교		사무용가구	종교	야구	순금	양말
9	바디용품	등산	휴대폰	기저귀	과자	반찬	커피	아웃도어가구	수납	수납	유아동잡화	좌욕용품	수영	반찬	잠옷	이유식		블루레이	러그	스케이트	과자	네트워크장비
10	마스크	남시	카메라	건강관리용품		벨트	밀키트	베이스메이크업	게임기	시험	블라인드	재활운동용품	베이스메이크업	벨트	기저귀	반찬		잠옷	재활운동용품	여성언더웨어		장류
11	게임기	검도	신생아의류	사무용품		밀키트	유아세제	테니스	요가	주방용품		반찬	테니스	네트워크장비	재활운동용품	러그		밀키트	기저귀	게임기		pc부품
12		축구	제과	카시트		유아세제	네트워크장비	거실가구	자전거	자전거		장류	거실가구	카페트	이유식	좌욕용품		커피	반찬	태블릿pc		물티슈
13		베이스메이크업	보드	블라인드		이유식	반찬	수납가구	주방용품	제과		잠옷	침구단품	유아세제	양말	기저귀		과자	잠옷			기저귀
14		모니터주변기기	아동가구	양말		네트워크장비	이유식	침구단품	원예	베이커리		카시트	청소용품	사무용가구	물티슈	카페트		카시트	좌욕용품			카페트
15		장갑	소독	축산물		pc부품	전통주	청소용품	영상가전	제빵재료		김치	수납가구	이유식	종교	축구		유아세제	장갑			잠옷
16		테니스	주방가전	침구단품		사무용가구	카페트	주방가구	물리치료	휴대폰		유아세제	주방가구	pc부품	카페트	이유식용품		벨트	블라인드			김치
17		생활용품	diy자재	김치		카페트	pc부품	육실용품	제과	물리치료		벨트	축구	밀키트	벨트	양말		pc부품	인형			밀키트
18		기저귀	냉동식품	러그		종교	모니터	축구	제빵재료	소스		출산	인테리어소품	종교	pc부품	pc부품		전통주	네트워크장비			전통주
19		양말	주방가구	라면		재활운동용품	사무용가구	요가	계절가전	수산물		양말	육실용품	재활운동용품	모니터주변기기	잠옷		이유식	이유식용품			블라인드
20		벨트	주변기기	여행용가방		전통주	카시트	인테리어소품	홀데코	점		기저귀	자동차기기	좌욕용품	좌욕용품	인형		재활운동용품	공구			모니터
21		물티슈	수납가구	배드민턴		여행용가방	종교	자동차기기	휴대폰	주변기기		블라인드	요가	모니터	블라인드	장류		네트워크장비	모니터주변기기			이유식용품
22			캔	조미료		이유식용품	재활운동용품	생활용품	태블릿pc	스쿠터		전통주	생활용품	전통주	밀키트	네트워크장비		카페트	양말			사무용가구
23			바디용품	반찬		카시트	모니터주변기기		수산물	홀데코		밀키트		카시트	전통주	모니터		종교	카페트			모니터주변기기
24			제빵재료	구강청결용품		모니터	여행용가방		가루	매트		공구		장류	사무용가구	물티슈		여행용가방	장류			장갑
25			요가	밀키트		좌욕용품	좌욕용품					장갑		장갑	장갑	벨트		모니터	벨트			인형
26			간편조리식품	반려동물		블루레이	장류					물티슈		모니터주변기기	이유식용품	음반		축산물	실버용품			카시트
27			원예			모니터주변기기	장갑					모니터		댄스	카시트	블라인드		좌욕용품	음반			유아세제
28			마스크			과자	이유식용품					소품		블루레이	소품	스키		건강관리용품	축산물			축산물
29						축산물	건강관리용품					이유식용품		이유식용품	댄스	공구		사무용품	카시트			벨트
30						장류	댄스					음반		여행용가방	유모차	거실가구		이유식용품	사무용품			댄스
31						장갑	블루레이							축산물		건강관리용품		모니터주변기기	임부복			공구
32						건강관리용품	사무용품							과자		실버용품		장류	건강관리용품			문구
33						댄스	남성화장품							인형		문구		장갑	유아세제			족구
34						노트북	과자							공구		카시트		세탁용품	여행용가방			선케어

상관성 높은  
순으로 정렬



What the  
못 믿겠다고?

# 불신의 시대니 (- -+), 하나만 더 예를 들어 보자 코로나19 이후 관심이 증가했던 등산 카테고리

- 등산 쇼핑 카테고리는 상관계수 0.7 이상인 쇼핑 항목들이 13개로 다양한 쇼핑 항목들과 관심 흐름이 유사하다
- 상관성이 높은 순으로 보면, 축구, 낚시, 수영, 뷰티소품, 검도, 침실가구, 아웃도어 가구 항목들이 0.9 이상으로 아주 높게 측정되었는데 이 중 "뷰티소품" 과의 높은 상관성이 아주 생똥맞다. 그리고 유사 카테고리인 베이스메이크업, 욕실용품 등과의 상관성도 꽤 높다 물론 Seasonality의 영향, 그러니까 해당 상품군을 주로 구매하는 시기가 봄, 가을이어서 일 수는 있겠지만, 설사 백번 양보해서 그렇다고 쳐도 발견해야 할 꺼리들은 넘치고 넘친다.

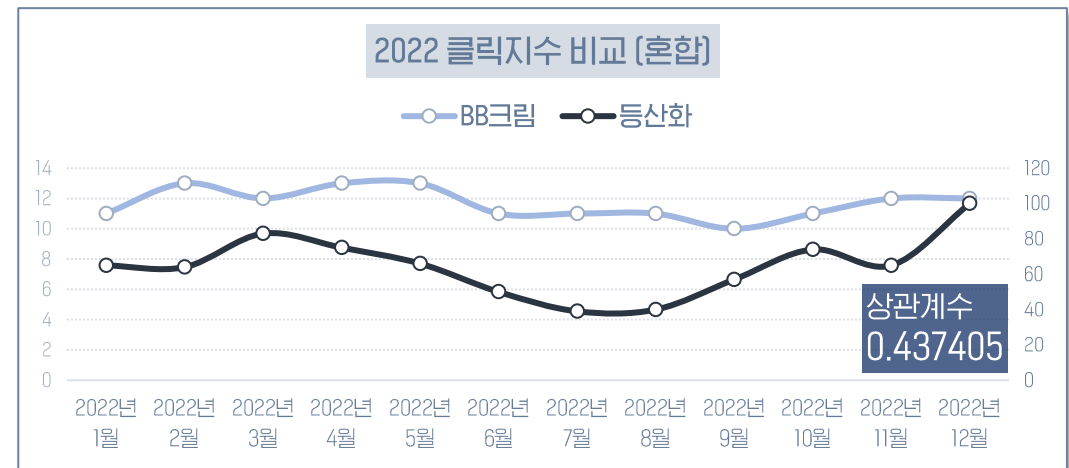
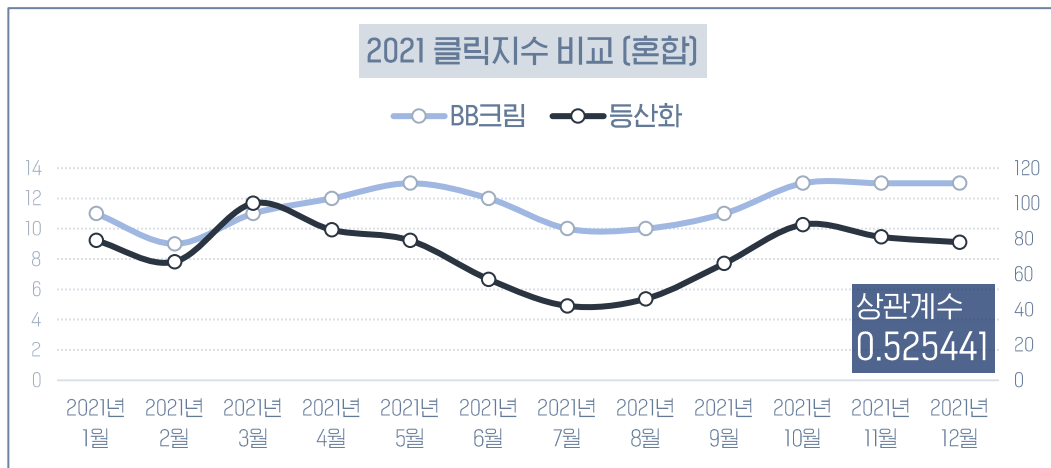
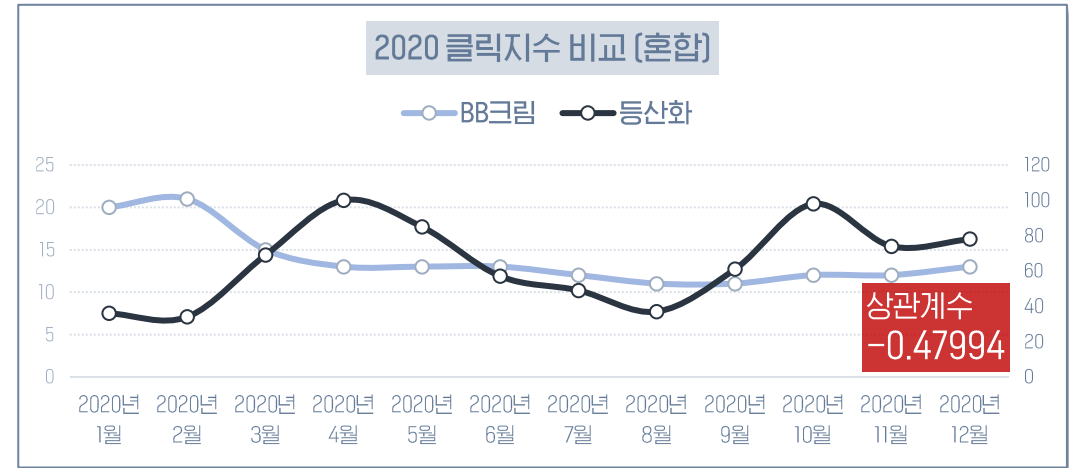
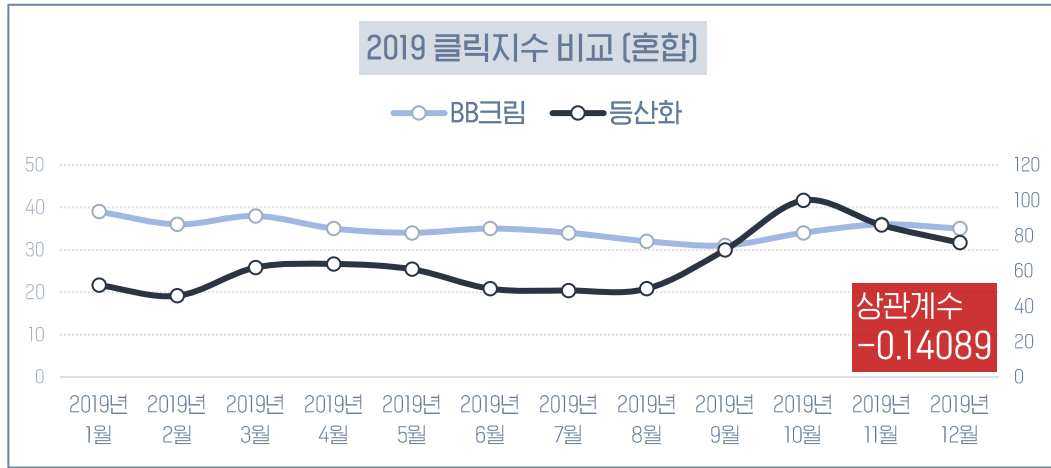
A	B	C	D	E	F	G	M	N	O	P	Q	R	S	T	U	V	
1	간편조리식품	거실가구	게임기	과자	구강청결용품	기저귀	김치	등산	리그	모니터	모니터주변기기	욕실용품	물티슈	밀키트	바다용품	반려동물	반찬
2	주니어가구	침구단품	정리용품	세탁용품	반려동물	러그	러그	축구	김치	김치	장갑	캠핑	러그	전통주	카메라	구강청결용품	커피
3	수납용품	수예	태블릿pc	블루레이	약기	김치	기저귀	낚시	양말	러그	밀키트	서재	기저귀	김치	주니어가구	면류	재활운동용품
4	아동가구	뷰티소품	수납	물티슈	면류	물티슈	벨트	수예	장옷	유아세제	전통주	스킨케어	양말	커피	신생아의류	약기	종교
5	신생아의류	축구	수납용품	유아동장비	생활용품	양말	장옷	뷰티소품	커피	커피	커피	음료	김치	이유식	수납용품	생활용품	좌욕용품
6	여성언더웨어	아웃도어	침실가구	기저귀	양말	양말	장옷	검도	플라인드	네트워크장비	김치	헬스	플라인드	물티슈	아동가구	축산물	러그
7	헤어케어	침실가구	기저귀	양말	양말	양말	장옷	침실가구	기저귀	반찬	재활운동용품		반찬	pc부품	간편조리식품	홀데코	이유식
8	카메라	수영	냉동	사무용가구	손글	플라인드	물티슈	아웃도어가구	물티슈	장류	종교		사무용가구	종교	야구	손글	양말
9	바다용품	등산	휴대폰	커피	아	아	아	수영	반찬	장옷	이유식		블루레이	러그	스케이트	과자	네트워크장비
10	마스크	낚시	카메라	베이스메이크업	베이스메이크업	베이스메이크업	베이스메이크업	테니스	벨트	기저귀	반찬		장옷	재활운동용품	여성언더웨어		장류
11	게임기	아세제	아세제	아세제	아세제	아세제	아세제	테니스	네트워크장비	재활운동용품	러그		밀키트	기저귀	게임기		pc부품
12		아세제	아세제	아세제	아세제	아세제	아세제	거실가구	카페트	이유식	좌욕용품		커피	반찬	태블릿pc		물티슈
13		아세제	아세제	아세제	아세제	아세제	아세제	침구단품	유아세제	양말	기저귀		과자	장옷			기저귀
14		아세제	아세제	아세제	아세제	아세제	아세제	청소용품	사무용가구	물티슈	카페트		카세트	좌욕용품			카페트
15		아세제	아세제	아세제	아세제	아세제	아세제	수납가구	이유식	종교	축구		유아세제	장갑			장옷
16		아세제	아세제	아세제	아세제	아세제	아세제	주방가구	pc부품	카페트	이유식용품		벨트	플라인드			김치
17		아세제	아세제	아세제	아세제	아세제	아세제	족구	밀키트	벨트	양말		pc부품	인형			밀키트
18		아세제	아세제	아세제	아세제	아세제	아세제	인테리어소품	종교	pc부품	pc부품		전통주	네트워크장비			전통주
19		아세제	아세제	아세제	아세제	아세제	아세제	욕실용품	재활운동용품	모니터주변기기	장옷		이유식	이유식용품			플라인드
20		아세제	아세제	아세제	아세제	아세제	아세제	자동차기기	좌욕용품	좌욕용품	인형		재활운동용품	공구			모니터
21		아세제	아세제	아세제	아세제	아세제	아세제	요가	모니터	플라인드	장류		네트워크장비	모니터주변기기			이유식용품
22		아세제	아세제	아세제	아세제	아세제	아세제	생활용품	전통주	밀키트	네트워크장비		카페트	양말			사무용가구
23		아세제	아세제	아세제	아세제	아세제	아세제		카세트	전통주	모니터		종교	카페트			모니터주변기기
24		아세제	아세제	아세제	아세제	아세제	아세제		장류	사무용가구	물티슈		여행용가방	장류			장갑
25		아세제	아세제	아세제	아세제	아세제	아세제		장갑	장갑	벨트		모니터	벨트			인형
26		아세제	아세제	아세제	아세제	아세제	아세제		모니터주변기기	이유식용품	음반		축산물	실버용품			카세트
27		아세제	아세제	아세제	아세제	아세제	아세제		댄스	카세트	플라인드		좌욕용품	음반			유아세제
28		아세제	아세제	아세제	아세제	아세제	아세제		블루레이	소품	스키		건강관리용품	축산물			축산물
29		아세제	아세제	아세제	아세제	아세제	아세제		이유식용품	댄스	공구		사무용품	카세트			벨트
30		아세제	아세제	아세제	아세제	아세제	아세제		여행용가방	유모차	거실가구		이유식용품	사무용품			댄스
31		아세제	아세제	아세제	아세제	아세제	아세제		축산물		건강관리용품		모니터주변기기	일부복			공구
32		아세제	아세제	아세제	아세제	아세제	아세제		과자		실버용품		장류	건강관리용품			돈구
33		아세제	아세제	아세제	아세제	아세제	아세제		인형		공구		장갑	유아세제			족구
34		아세제	아세제	아세제	아세제	아세제	아세제		공구		카세트		세탁용품	여행용가방			선케어

등산 카테고리를 예로들면,  
상관계수 0.7 이상인  
카테고리 13개로 우수한 수준  
축구, 낚시 등 아웃도어 활동과  
뷰티/케어 관련 상품 구입,  
인테리어 가구 분야와 밀접

상관계수 0.70이상 개수	등산
등산	1.00
축구	0.98
낚시	0.98
수예	0.97
뷰티소품	0.97
검도	0.95
침실가구	0.92
아웃도어가구	0.91
수영	0.89
베이스메이크업	0.86
테니스	0.80
거실가구	0.73
침구단품	0.72
청소용품	0.66
수납가구	0.65
주방가구	0.61
족구	0.59
인테리어소품	0.59
욕실용품	0.55
자동차기기	0.54
요가	0.54
생활용품	0.53

# 일부 소분류 상품 간의 최근 4년간 실제 클릭 추이를 보면 2020년까지 무관하던 흐름이, 2021년 이후 묘하게 비례하는 추세로..

- 연간 시계열 차트 흐름이 시각적으로도 눈에 띄는 변화를 보이지만, 통계적으로 확인해 보면 2020년까지 상관계수가 음(-)의 값을 가지던 것이, 2021년 이후에는 양(+)<sup>의 값을 가지며 확연히 전환</sup>
- 상관성이 높다는 것이, 반드시 인과성으로 연결되지는 않을 것이나 연간 흐름으로 끊어 확인한 결과, 두 데이터 간의 관계를 단순히 Seasonality로 해석하기에는 무리가 있을 것





# 실제 소비자 언어를 통해 확인해 볼까?

## 전체 언급 규모가 많은 것은 아니나, 꾸준히 언급 증가 추세

- 등산 등 아웃도어 활동이 최근 코로나19로 인해 2030대 등 젊은 층으로 확대되면서 실제 화장을 고민하는 여성층이 증가하고 있는 것으로 나타남
- 특히 이러한 경향은 2020년 등 코로나19 초기 대비로도 오히려 증가

등산모임 화장하고 가는거 개예바죠??

ㅠㅠㅠㅠ 쌓얼에 자신없는 1인이라... 비비크림은 포기못하겠는데...

등산모임가면 훈남 오빠도 있고 그래가지고.. 하.. 다들 그냥 선크림만 바르고 가시나요??

초가을, 지금처럼 쌀쌀해지기 전이라 동봉까지 올라가며 땀 많이 났지만 화장은 이렇게 그대로 보송하게!

코로나도 끝나고 날도 좋아서 정말 오랜만에 등산 다녀 왔는데 저보다 어르신 분들도 많이 와서 놀랐어요  
요새 등산이 유행한다더니 ㅋㅋ 새삼 실감한 느낌인데ㅠㅠ 진짜 화장 하고 온건 뭐 그러려니 하는데  
남자 여자 구분없이 다들 향수도 엄청 뿌리고

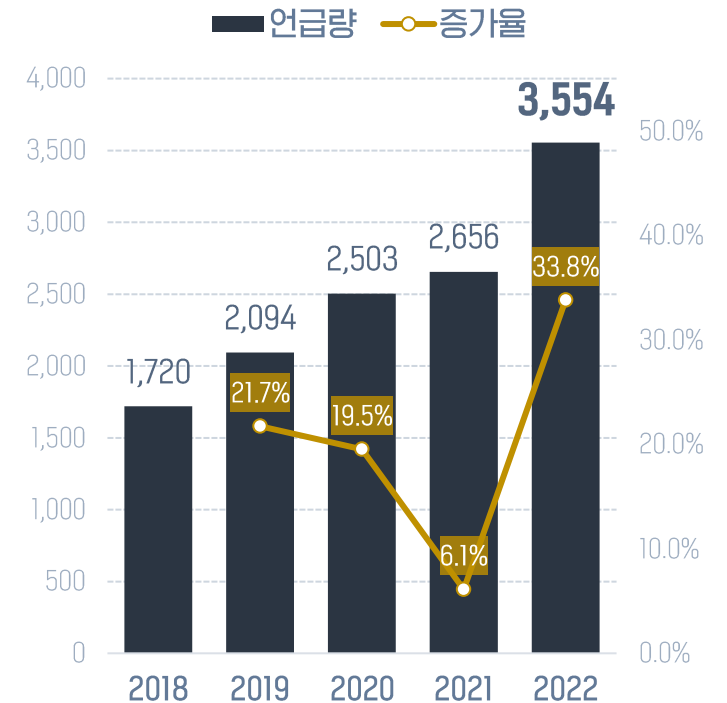
등산을하며 체력을 키워가니 참 좋다 2년전 등산은 완전히 다와서 찍는 컨셉이었는데 가면서 찍는것도 괜찮은거같다  
스킨. 로션. 수분크림. 선크림까지만 바르고 립스틱 살짝 헷는데 꼭 비비크림도 바른거같네

등산가는데 누가 화장을해!!! 했는데 대부분 나도 화장하고 탕구나... 그렇구나...

최근에 등산 다녀온 덕상사 말로는, 요즘 젊은 새렘들~ 이러면서 설명해주는ㅋㅋ  
여자 화장이 산꼭대기에서도 안 지워진데ㅋㅋㅋㅋㅋ

등산하고싶어서 동호회같은걸 가입했는데 제가 화장한거랑 안한거랑 차이가커서 화장하고가면 좀 그래보일까요?..  
안하면 넘 초체한데 굳이 운동인데 화장을.. 이렇게 볼까봐 ㅠㅠ

SNS 상에서의 “화장+”등산”  
동반 언급량 추이 (단위:건)



# statistics

간단한 **상관관계**(correlation) 분석을 통해 중분류 항목 간  
“어떤 **알 수 없는**” 관계가 존재 할 수 있다는 확신이 들었다.  
그래서 이를 기반으로 **한 단계 더** 앞으로 나가 보기로 했다.

자, 이제부터  
**본격적인 분석**

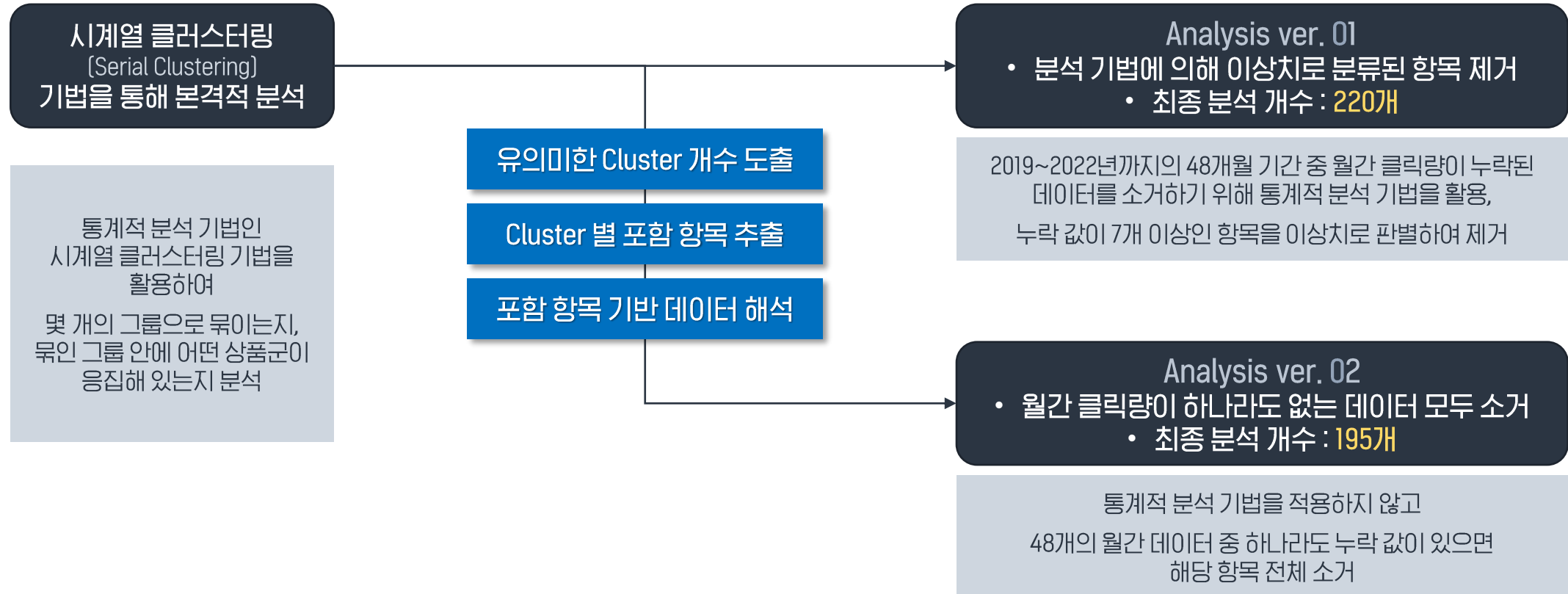
Step 04

## Step 04 / 심화 분석 접근 방향

### 이상치 제거 기준을 이원화하여 합리적인 분석 기법 탐색

- 앞서의 상관분석에서 바통을 이어받아, 통계적 기법을 적용하여 군집화(Clustering) 시도
- 이상치 소거 역시 ① 통계적 분석 기법을 적용한 이상치 검출 결과와 ② 연구자 판단에 따라 임의로 소거 기준을 설정한 결과 두개를 모두 적용하여 최종 분석 결과에 기반, 유의미하다고 판단되는 결과를 선택할 예정
- <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



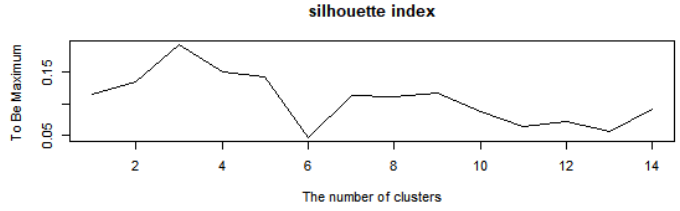

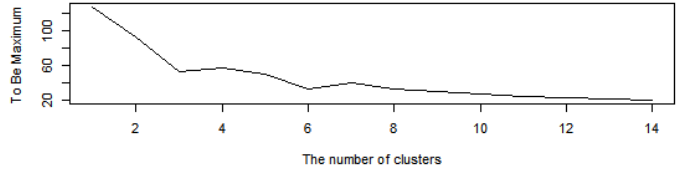
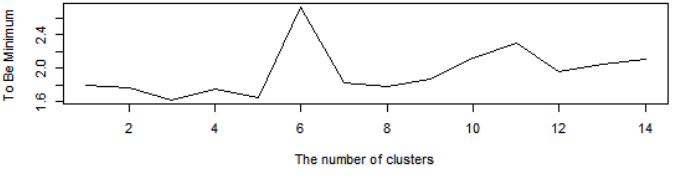
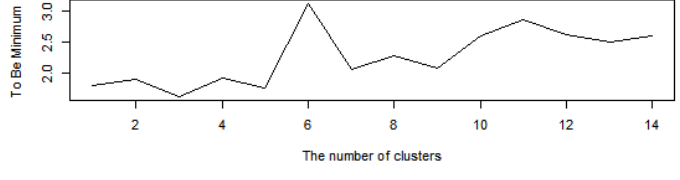
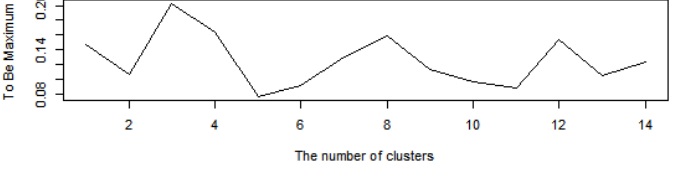
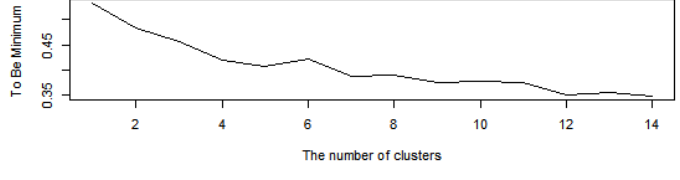
# Version 01

- 분석 기법에 의해  
이상치로 분류된 항목 제거
- 최종 분석 개수 : 220개

# 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 선정
- [각 군집화 평가 지수에 대한 설명이 난해한 것은 어쩔 수 없음.. 기술적 용어에 대한 사전 이해가 필요하므로.]

<p><b>Silhouette index</b></p> <p>각 데이터 별로 그 데이터가 속한 클러스터 내의 유사도와 인접한 클러스터와의 유사도를 비교하는 지표</p>	 <p>silhouette index</p> <p>The number of clusters</p>	 <p>Score Function index</p> <p>The number of clusters</p>	<p><b>Score Function index</b></p> <p>일반적인 인공지능 연구에서 사용하는 성능 지표. 클러스터간 거리를 score화 하여 거리를 함수로 계산하는 지표</p>
<p><b>Calinski-Harabasz index</b></p> <p>데이터셋과 특정 클러스터간 중심거리를 클러스터간 중심거리의 분산 비율로 계산하는 지표</p>	 <p>Calinski-Harabasz index</p> <p>The number of clusters</p>	 <p>Dabies-Bouldin index</p> <p>The number of clusters</p>	<p><b>Davies-Bouldin index</b></p> <p>특정 클러스터와 그것과 가장 유사한 클러스터 사이의 평균 거리로 계산하는 지표</p>
<p><b>Modified Davies-Bouldin index</b></p> <p>수정된 Davies-Bouldin index라고 하며, 클러스터간 중심 거리의 최소값을 구하는 수식이 포함된 지표</p>	 <p>Modified Dabies-Bouldin index</p> <p>The number of clusters</p>	 <p>Dunn index</p> <p>The number of clusters</p>	<p><b>Dunn index</b></p> <p>가장 유명한 평가 지표이며, Sil index와는 다르게 해당 데이터가 속한 클러스터 내의 유사도와 모든 클러스터의 유사도를 비교하는 지표</p>
<p><b>COP index</b></p> <p>특정 클러스터의 중심과 해당 클러스터에 속해 있는 모든 데이터셋의 평균거리를 가장 멀리 떨어진 클러스터의 거리 비율로 계산하는 지표</p>	 <p>COP index</p> <p>The number of clusters</p>		

# 최적의 Cluster 개수 선정 방법

## 7개 중 4개의 평가 지표에서 유의미한 평가 분석 도출, 최종 3개의 Cluster 선정

- Silhouette, Davies-Bouldin, Modified Davies-Bouldin, Dunn 지표 분석 상에서 공통적으로, Cluster가 3개로 나뉘었을 때 유의미함을 확인
- Score Function, Calinski-Harabasz, COP index는 값의 변화가 없이 나타나거나, 우하향하여 Cluster 수를 도출하기 어려움

<p><b>Silhouette index</b></p> <p>각 데이터 별로 그 데이터가 속한 클러스터 내의 유사도와 인접한 클러스터와의 유사도를 비교하는 지표</p>	 <p>The number of clusters</p>	<p><b>Score Function index</b></p> <p>값의 변화가 없어 선정 불가</p> <p>Y좌표의 최대값을 지향하는 지점 (클러스터 개수) 선정 불가</p>  <p>The number of clusters</p>	<p><b>Score Function index</b></p> <p>일반적인 인공지능 연구에서 사용하는 성능 지표. 클러스터간 거리를 score화 하여 거리를 함수로 계산하는 지표</p>
<p><b>Calinski-Harabasz index</b></p> <p>데이터셋과 특정 클러스터간 중심거리를 클러스터 간 중심거리의 분산 비율로 계산하는 지표</p>	<p><b>Calinski-Harabasz index</b></p> <p>하향하는 추세의 평가 그래프</p> <p>Y좌표의 최대값을 지향하는 지점 (클러스터 개수) 선정 불가</p>  <p>The number of clusters</p>	<p><b>Dabies-Bouldin index</b></p>  <p>The number of clusters</p>	<p><b>Davies-Bouldin index</b></p> <p>특정 클러스터와 그것과 가장 유사한 클러스터 사이의 평균 거리로 계산하는 지표</p>
<p><b>Modified Davies-Bouldin index</b></p> <p>수정된 Davies-Bouldin index라고 하며, 클러스터간 중심 거리의 최소값을 구하는 수식이 포함된 지표</p>	<p><b>Modified Dabies-Bouldin index</b></p>  <p>The number of clusters</p>	<p><b>Dunn index</b></p>  <p>The number of clusters</p>	<p><b>Dunn index</b></p> <p>가장 유명한 평가 지표이며, Sil index와는 다르게 해당 데이터가 속한 클러스터 내의 유사도와 모든 클러스터의 유사도를 비교하는 지표</p>
<p><b>COP index</b></p> <p>특정 클러스터의 중심과 해당 클러스터에 속해 있는 모든 데이터셋의 평균거리를 가장 멀리 떨어진 클러스터의 거리 비율로 계산하는 지표</p>	<p><b>COP index</b></p> <p>하향하는 추세의 평가 그래프</p> <p>Y좌표의 최소값을 지향하는 지점 (클러스터 개수) 선정 불가</p>  <p>The number of clusters</p>		

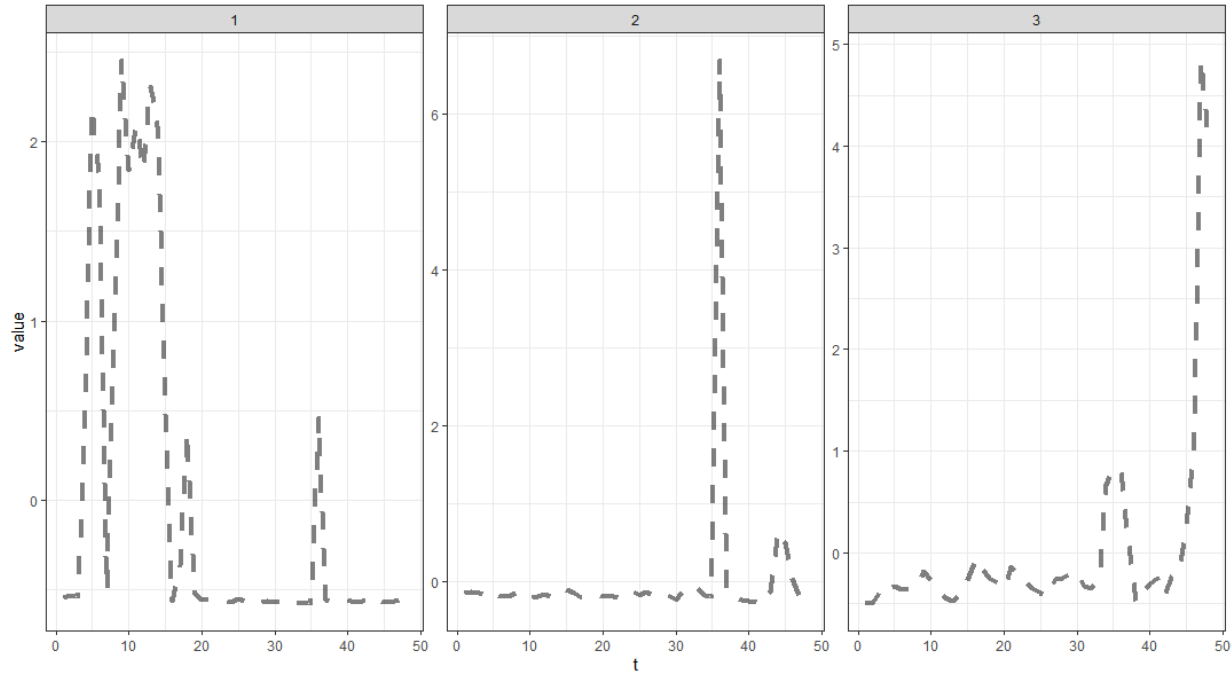
# 3가지 Cluster의 클릭지수 흐름

## 3개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

- 시점 초기에 클릭률이 높았던 유형, 후반부에 높아지는 유형, 최근 시점에 증가하는 유형 등으로 분류

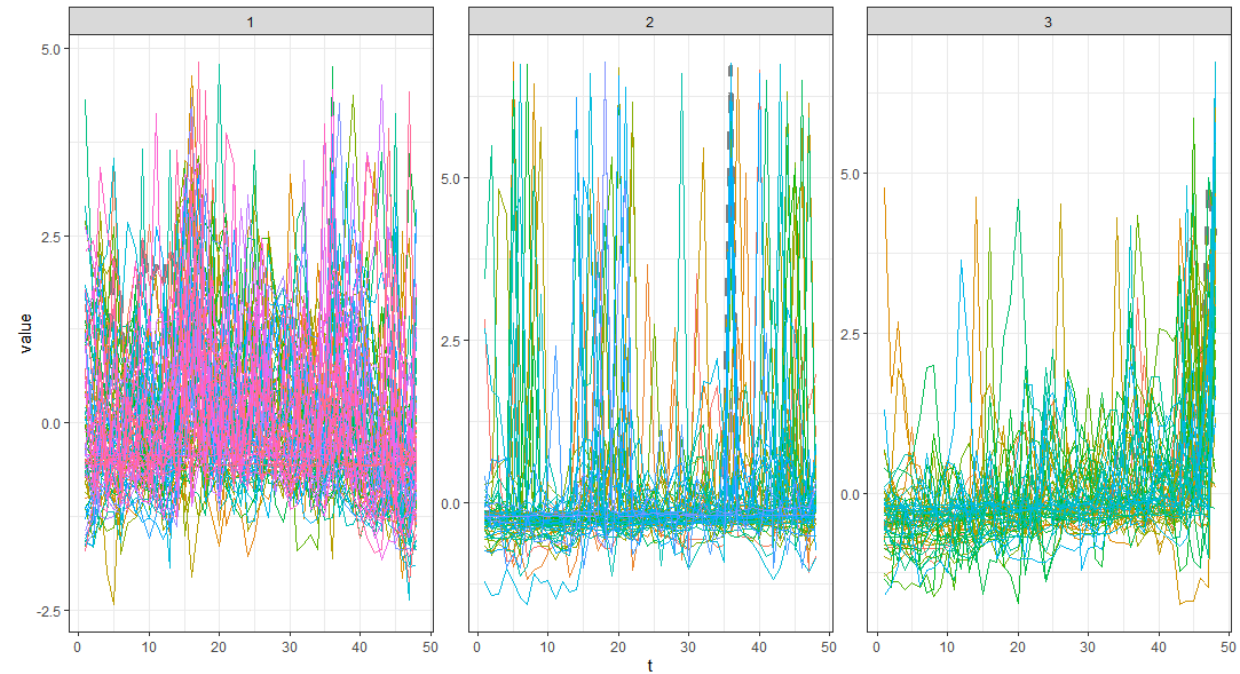
[ 대표 추세 시각화 ]

Clusters' members



[ Cluster별로 포함된 개별 키워드들의 추이를 중첩한 시각화 ]

Clusters' members

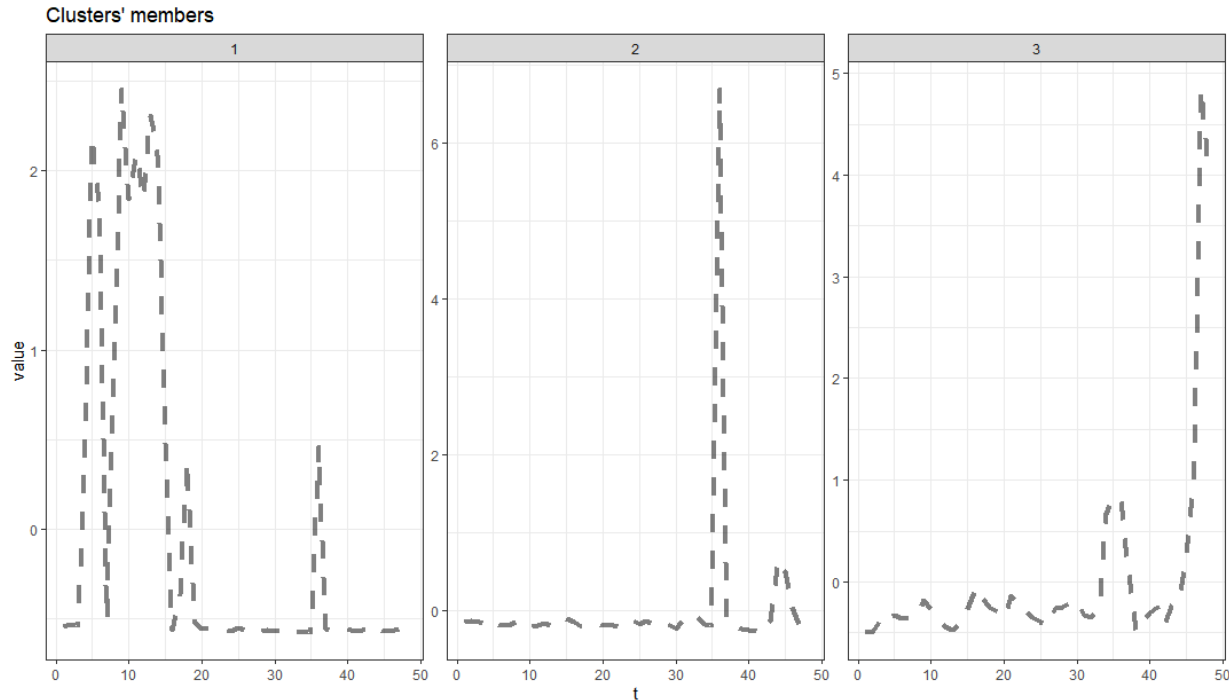




## 3가지 Cluster의 클릭지수 흐름

### 3개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

- 시점 초기에 클릭률이 높았던 유형, 후반부에 높아지는 유형, 최근 시점에 증가하는 유형 등으로 분류



### Centroids (각 클러스터의 중심점)

각각의 클러스터는 거리 측정 기준값이 있어야 하는데 해당 기준값이 되는 중심점을 Centroids라고 일컫음.

Centroids 선정 방법은 임의로 선정하며,  
임의로 선정된 각 Centroids로부터  
가까운 거리의 개체들끼리 그룹을 형성

즉, 시계열 클러스터의 Centroids는  
“각 클러스터의 대표 분포 개체”이며,  
해당 분포와 “비슷한 분포”를 가진 개체끼리 분류

## 3가지 Cluster의 클릭지수 흐름

### 3개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

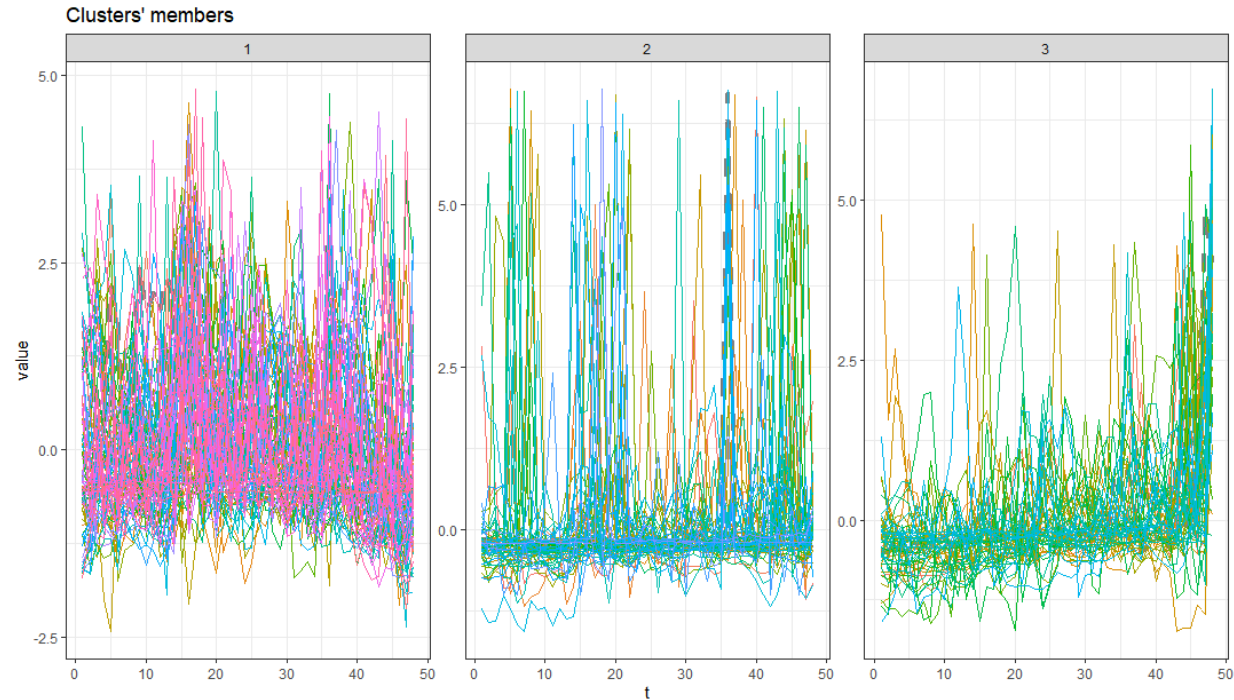
- 시점 초기에 클릭률이 높았던 유형, 후반부에 높아지는 유형, 최근 시점에 증가하는 유형 등으로 분류

#### 시계열 클러스터링 결과 그래프

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

Centeroids로부터 거리가 가까운 개체들끼리 묶여서 동일 클러스터에 형성

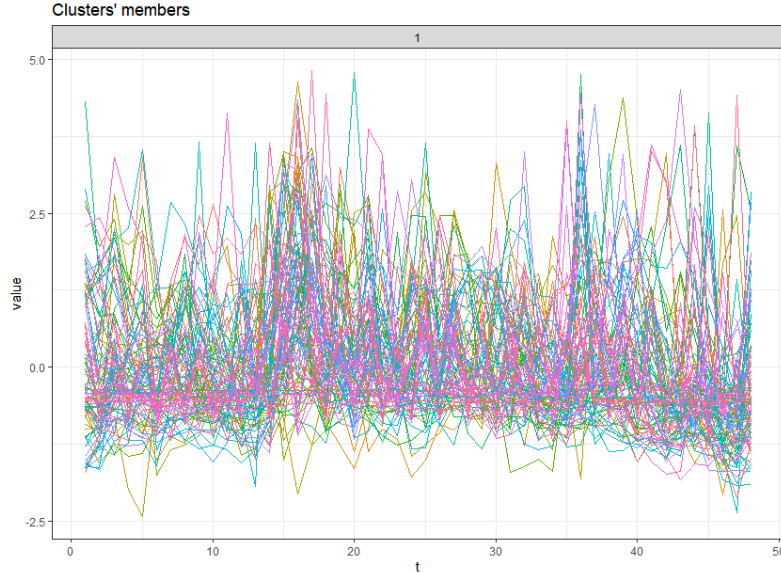
시계열에 따라 클러스터로 묶인 개체들의 데이터를 그래프로 확인 가능



# Cluster 01

## 포함된 상품군 96개

- 육과 관련 상품과 건강용품이 상위에 위치해 있으나, 이것만으로는 패턴을 발견하기 어려울 정도로 많은 상품군이 포함



출산기념품	보드	게임기	당구용품	모자	순금	정원	수산물	신생아인류
여성가방	수납가구	클렌징	발건강용품	인테리어소품	간편조리식품	냉동식품	색조메이크업	인라인스케이트
물리치료용품	주방가전	서재용가구	선글라스	네일케어	오토바이	드레싱	홀데코	조미료
태블릿pc	스쿼시	골프	볼링	스킨케어	향수	원예	안경테	아동가구
좌훈용품	소스	주방용품	야구	헤어케어	타이틀	탁구	통조림	남성언더웨어
시계	롤러	제과재료	마스크	시럽	수납용품	냉동	정원용품	자동차용품
전자제품	헤어액세서리	농산물	가루	눈건강용품	수영복	휴대폰	유아동잠옷	여성의류
팩	찜질용품	욕실용품	좌욕	건강식품	요가	제빵재료	식물	카메라
유아동의류	정리용품	수납	스케이트	소품	주니어가구	남성신발	diy자재	
스쿠터	청소용품	국내여행	베이커리	지갑	주얼리	자전거	렌터카	
완구	돌기념품	바디용품	원예용품	침구세트	남성의류	캠	여성언더웨어	

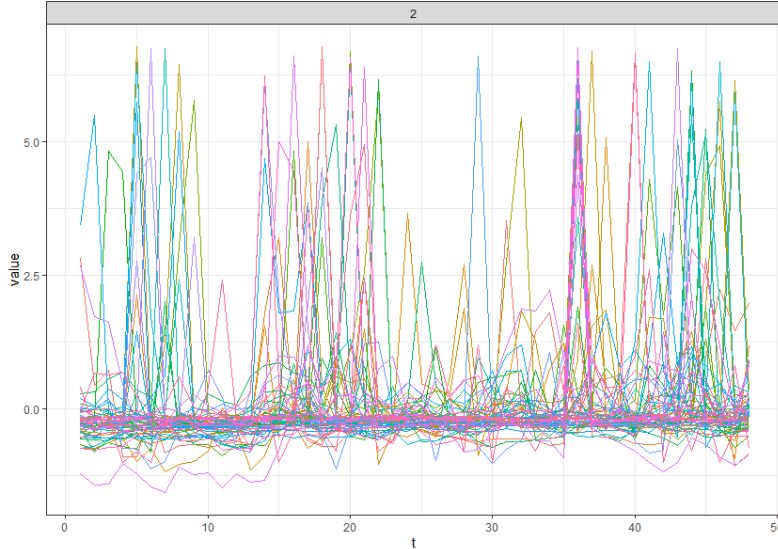
\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순

# Cluster 02

## 포함된 상품군 66개

- 다수의 스포츠, 아웃도어 활동 포함

Clusters' members



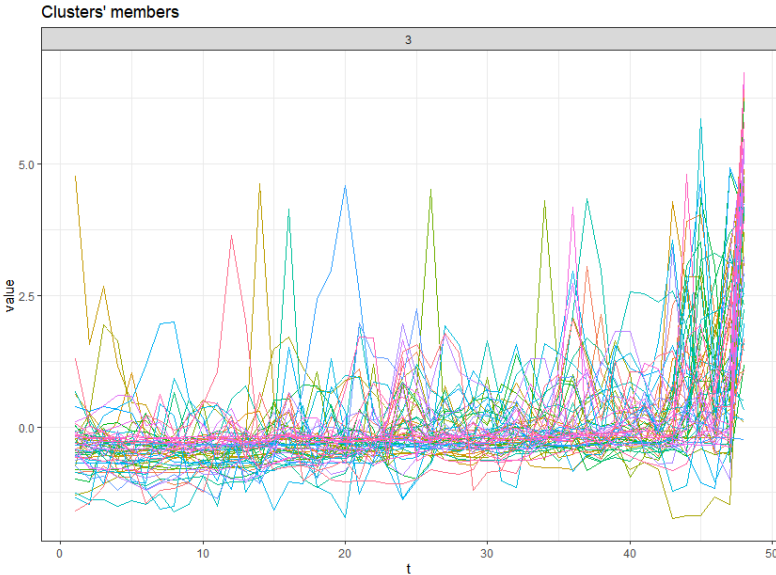
관상어용품	캔	소프트웨어	캠핑	주변기기	생활용품
거실가구	광학기기	노트북액세서리	유아가구	주방가구	남성가방
수유용품	음향가전	문구용품	반려동물	외출용품	계절가전
분유	권투	pc	수예	안마용품	농구
수집품	생활가전	출산	마라톤용품	뷰티소품	식용유
의료용품	낚시	남성화장품	서재	패션소품	저장장치
체험	pc액세서리	축구	헬스	테니스	물리치료
배구	검도	당뇨관리용품	아웃도어가구	침구단품	학습기기
스킨스쿠버	교구	등산	휴대폰액세서리	필라테스	구강위생용품
수영	영상가전	베이스메이크업	유아침구	바디케어	이미용가전
임산부용품	헤어스타일링	침실가구	목욕용품	세탁용품	약기

\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순

# Cluster 03

## 포함된 상품군 57개

- 주로 상위에 사무용품 등의 키워드 상위 차지



음반	이유식	사무용가구	선케어	블라인드	아기간식
실버용품	건강관리용품	종교	화방용품	벨트	매트
카세트	잠옷	라면	pc부품	축산물	
원데이클래스	밀키트	장갑	공구	유모차	
과자	유아동잡화	인형	신발용품	블루레이	
모니터	양말	반찬	임부복	해외여행	
여성신발	다이어트식품	배드민턴	장류	카페트	
무술용품	음료	네트워크장비	스키	이유식용품	
문구	러그	족구	전통주	물티슈	
김치	노트북	커튼	여행용가방	기저귀	
사무용품	재활운동용품	유아세제	모니터주변기기	좌욕용품	

\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순

통계적인 기법에 따라 이상치를 검출한 Version 1의 경우  
가장 큰 문제는 **Cluster 수가 3개로 적게 나왔다**는 것이었다.

Cluster 수가 적게 나오면, 그에 포함되는 세부 중분류 항목들이  
즉, 상당수 이질적인 항목들이 고작 3개 분류로 나뉘어진다는 것이며

이는 [낮선 분석 결과를 거의 처음 접하는 상황에서]  
개별 Cluster의 특징을 간파하기 어렵다는 것을 의미한다.

그래서 우리에게서 Version 2가 필요했다.



# Version 02

- 월간 클릭량이 하나라도 없는 데이터, 모두 소거
- 최종 분석 개수 : 195개

# 최적의 Cluster 개수 선정 방법

## 7가지의 Clustering(군집화) 평가 지수 분석 과정

- 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 선정

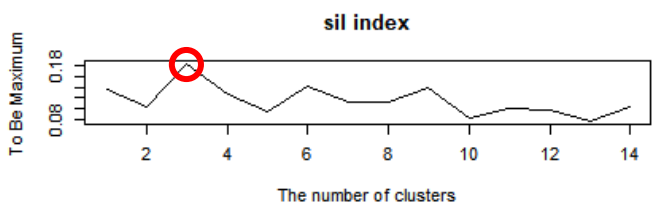
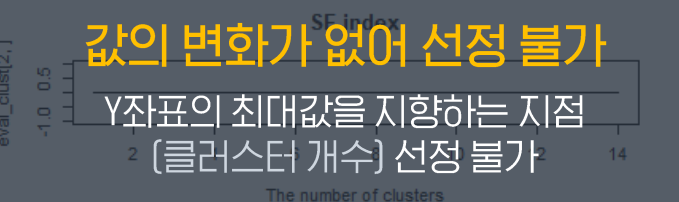
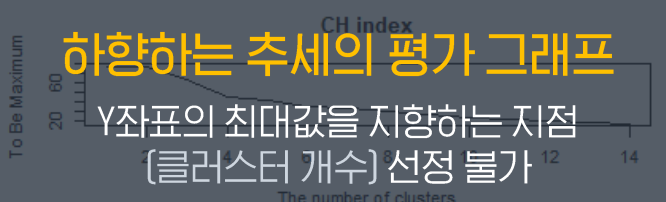
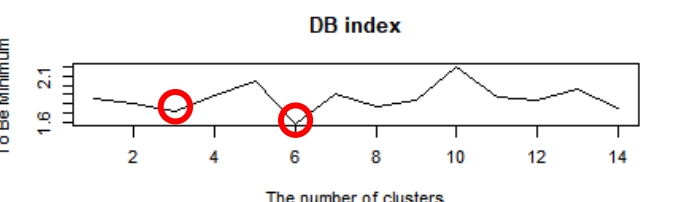
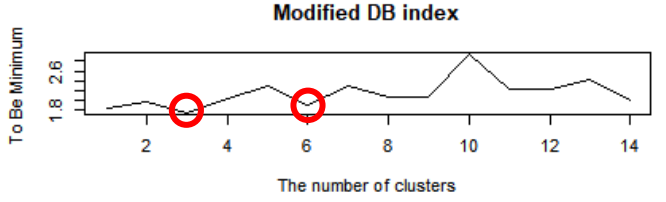
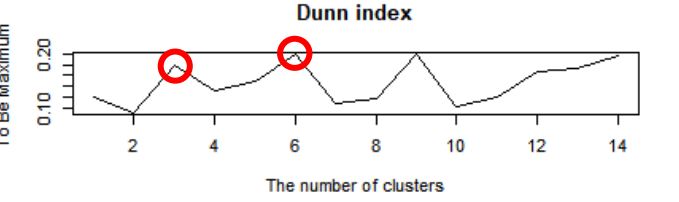
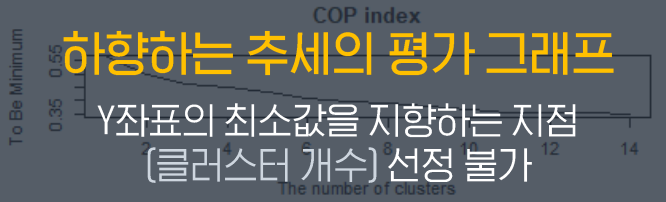
<p><b>Silhouette index</b></p> <p>각 데이터 별로 그 데이터가 속한 클러스터 내의 유사도와 인접한 클러스터와의 유사도를 비교하는 지표</p>	<p><b>sil index</b></p>	<p><b>SF index</b></p>	<p><b>Score Function index</b></p> <p>일반적인 인공지능 연구에서 사용하는 성능 지표. 클러스터간 거리를 score화 하여 거리를 함수로 계산하는 지표</p>
<p><b>Calinski-Harabasz index</b></p> <p>데이터셋과 특정 클러스터간 중심거리를 클러스터간 중심거리의 분산 비율로 계산하는 지표</p>	<p><b>CH index</b></p>	<p><b>DB index</b></p>	<p><b>Davies-Bouldin index</b></p> <p>특정 클러스터와 그것과 가장 유사한 클러스터 사이의 평균 거리로 계산하는 지표</p>
<p><b>Modified Davies-Bouldin index</b></p> <p>수정된 Davies-Bouldin index라고 하며, 클러스터간 중심 거리의 최소값을 구하는 수식이 포함된 지표</p>	<p><b>Modified DB index</b></p>	<p><b>Dunn index</b></p>	<p><b>Dunn index</b></p> <p>가장 유명한 평가 지표이며, Sil index와는 다르게 해당 데이터가 속한 클러스터 내의 유사도와 모든 클러스터의 유사도를 비교하는 지표</p>
<p><b>COP index</b></p> <p>특정 클러스터의 중심과 해당 클러스터에 속해 있는 모든 데이터셋의 평균거리를 가장 멀리 떨어진 클러스터의 거리 비율로 계산하는 지표</p>	<p><b>COP index</b></p>		



# 최적의 Cluster 개수 선정 방법

## 7개 중 4개의 평가 지표에서 유의미한 평가 분석 도출, 최종 6개의 Cluster 선정

- Silhouette, Davies-Bouldin, Modified Davies-Bouldin, Dunn 지표 분석 상에서 공통적으로, Cluster가 6개로 나뉘었을 때 유의미함을 확인
- Score Function, Calinski-Harabasz, COP index는 값의 변화가 없이 나타나거나, 우하향하여 Cluster 수를 도출하기 어려움

<p><b>Silhouette index</b></p> <p>각 데이터 별로 그 데이터가 속한 클러스터 내의 유사도와 인접한 클러스터와의 유사도를 비교하는 지표</p>	 <p>The number of clusters</p>	<p><b>값의 변화가 없어 선정 불가</b></p> <p>Y좌표의 최대값을 지향하는 지점 2 [클러스터 개수] 선정 불가</p>  <p>The number of clusters</p>	<p><b>Score Function index</b></p> <p>일반적인 인공지능 연구에서 사용하는 성능 지표. 클러스터간 거리를 score화 하여 거리를 함수로 계산하는 지표</p>
<p><b>Calinski-Harabasz index</b></p> <p>데이터셋과 특정 클러스터간 중심거리를 클러스터 간 중심거리의 분산 비율로 계산하는 지표</p>	<p><b>하향하는 추세의 평가 그래프</b></p> <p>Y좌표의 최대값을 지향하는 지점 12 [클러스터 개수] 선정 불가</p>  <p>The number of clusters</p>	<p><b>DB index</b></p>  <p>The number of clusters</p>	<p><b>Davies-Bouldin index</b></p> <p>특정 클러스터와 그것과 가장 유사한 클러스터 사이의 평균 거리로 계산하는 지표</p>
<p><b>Modified Davies-Bouldin index</b></p> <p>수정된 Davies-Bouldin index라고 하며, 클러스터간 중심 거리의 최소값을 구하는 수식이 포함된 지표</p>	<p><b>Modified DB index</b></p>  <p>The number of clusters</p>	<p><b>Dunn index</b></p>  <p>The number of clusters</p>	<p><b>Dunn index</b></p> <p>가장 유명한 평가 지표이며, Sil index와는 다르게 해당 데이터가 속한 클러스터 내의 유사도와 모든 클러스터의 유사도를 비교하는 지표</p>
<p><b>COP index</b></p> <p>특정 클러스터의 중심과 해당 클러스터에 속해 있는 모든 데이터셋의 평균거리를 가장 멀리 떨어진 클러스터의 거리 비율로 계산하는 지표</p>	<p><b>하향하는 추세의 평가 그래프</b></p> <p>Y좌표의 최소값을 지향하는 지점 12 [클러스터 개수] 선정 불가</p>  <p>The number of clusters</p>		

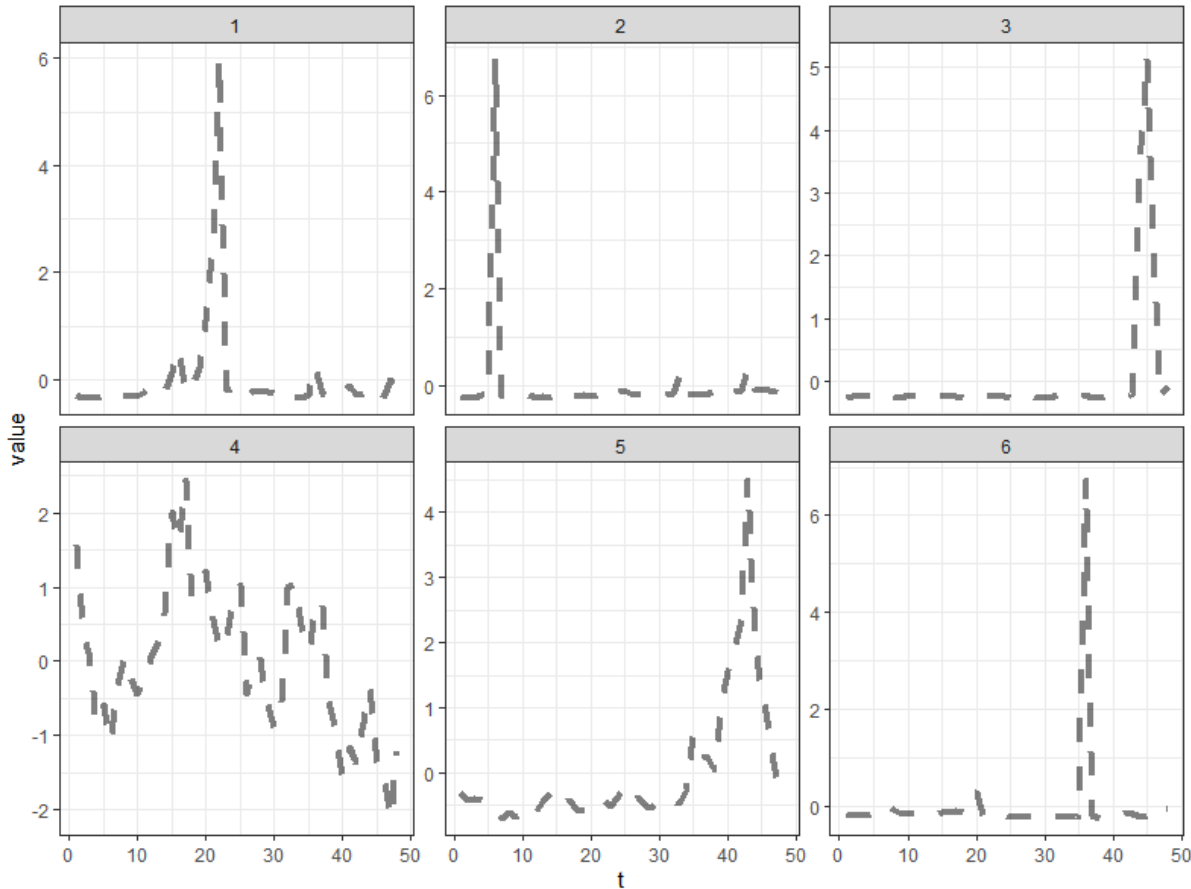
# 6가지 Cluster의 클릭지수 흐름

## 6개의 클러스터에 모인 쇼핑 상품들의 클릭지수 흐름 분석 결과

- 클릭률이 높았던 시기가 각각 다른 유형, 일부 반등은 있지만 대체적으로 감소하는 유형, 미세하게 증가하는 유형 등 총 6가지 클러스터로 분류

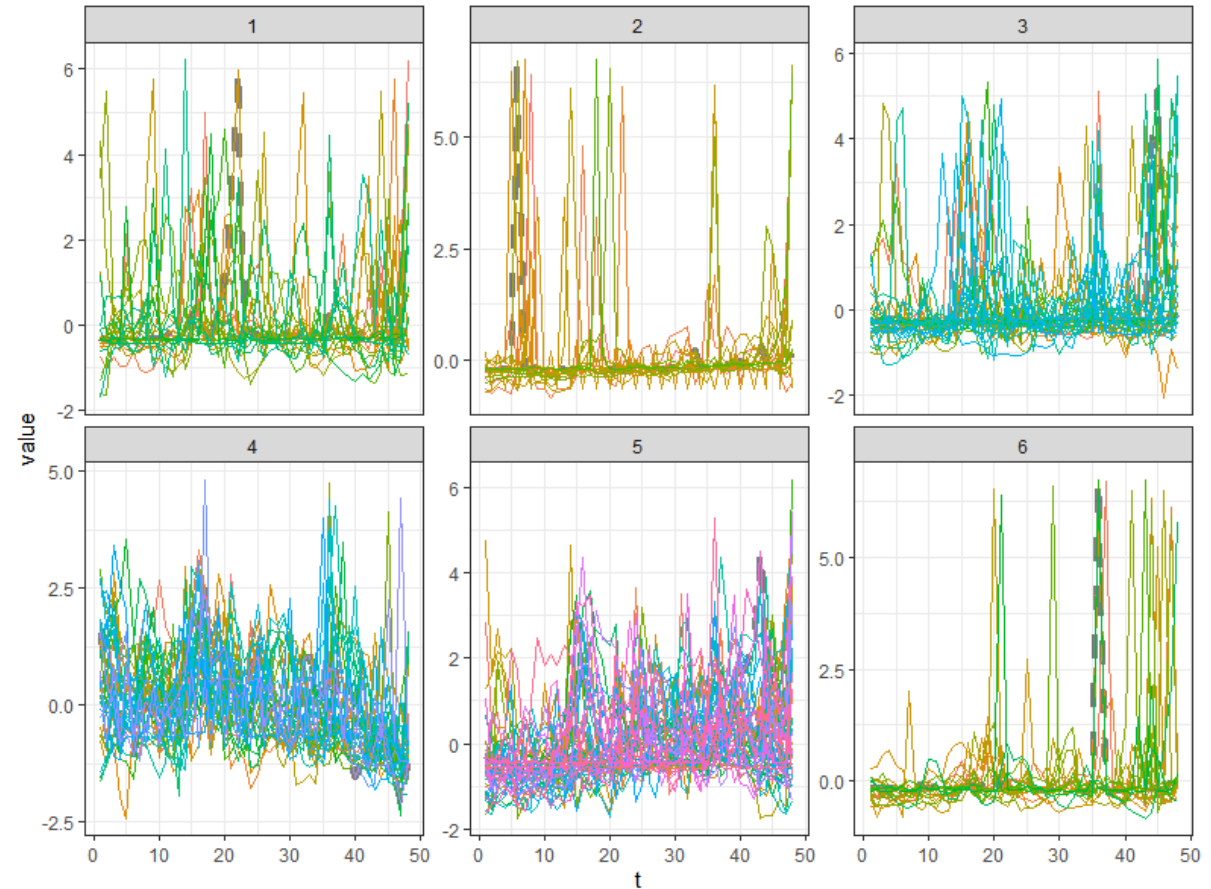
[ 대표 추세 시각화 ]

Clusters' members



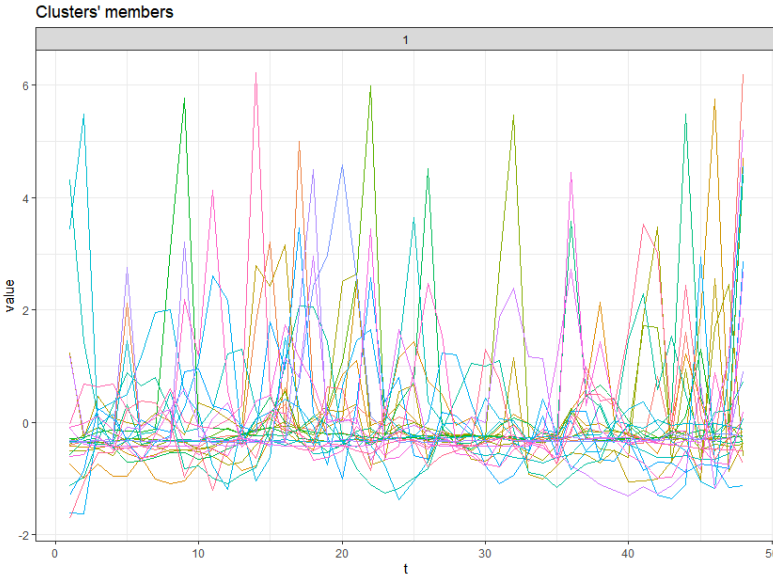
[ Cluster별로 포함된 개별 키워드들의 추이를 중첩한 시각화 ]

Clusters' members



# Cluster 01 - 25개 키워드 포함 육아 과정에서 함께 관심이 증가하는 품목들

- 유아동의류, 헤어액세서리, 청소용품 항목이 해당 Cluster에 적합도가 가장 높게 산출
- 모든 항목의 공통점을 유추할 수는 없지만 대략 그려본 그림으로는, 육아 과정에서 요하게되는 품목들이 아닌가 싶기도..



유아동의류	침구세트	이미용가전
헤어액세서리	공구	유아동잠옷
청소용품	pc부품	자동차용품
문구	마라톤용품	
음료	냉동식품	
선글라스	안마용품	
네일케어	필라테스	
스케이트	생활용품	
장갑	계절가전	
교구	유모차	
노트북액세서리	농구	

\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순

## Parenting

유아동의류, 유모차, 유아동잠옷 등  
육아 관련 품목들과 함께

청소용품, 침구세트, 냉동식품,  
안마용품, 생활용품, 계절가전 등이 다수 포함

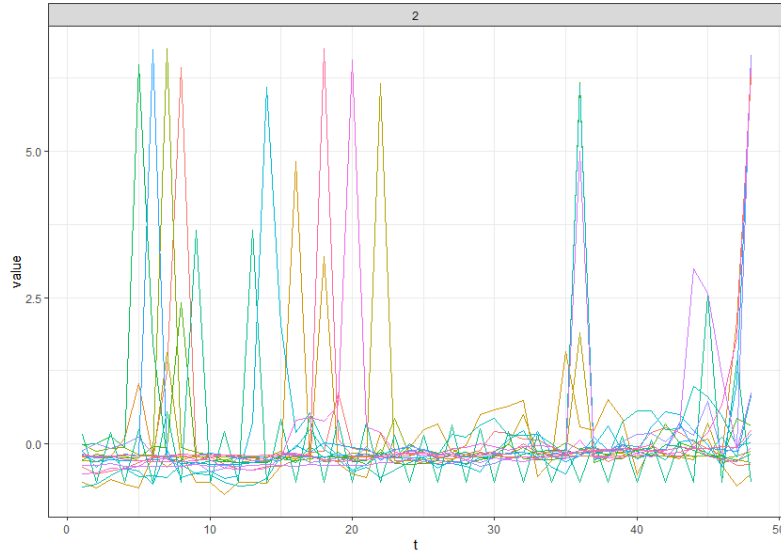
살짝, 주관적 해석을 곁들여서 유추해 보면  
유아 자녀를 둔 부모의 소비행동과 닮아있음

“기저귀 옆에 맥주” .. 기억하는가?

# Cluster 02 - 17개 키워드 포함 캠핑 수요와 비례하는 인테리어 품목

- 패션 아이템, 육아, 캠핑, 인테리어 아이템 등이 조금씩 섞여 있기는 하지만 우선 주목하고 싶은 것은 '캠핑'
- 코로나19로 인해 캠핑에 대한 수요가 증가하면서, 캠핑용품, 캠핑가구, 간편식(밀키트) 등 관련 시장이 호황이었지만 인테리어 아이템과 연계성이 높아진 것도 적지 않은 변화 (물론 인테리어 시장은 캠핑 시장과 별개 트렌드로 성장했지만..)

Clusters' members



시계	캠핑
수유용품	아웃도어가구
이유식	휴대폰액세서리
의료용품	침구단품
배구	남성가방
수영	물리치료
임산부용품	
재활운동용품	
네트워크장비	
소프트웨어	
커튼	

## Outdoor-Camp

캠핑 등 아웃도어 활동으로 인해  
아웃도어 가구에도 관심을 갖게 된 소비자는

인테리어 목적으로 아웃도어 가구를 고르기도 했고

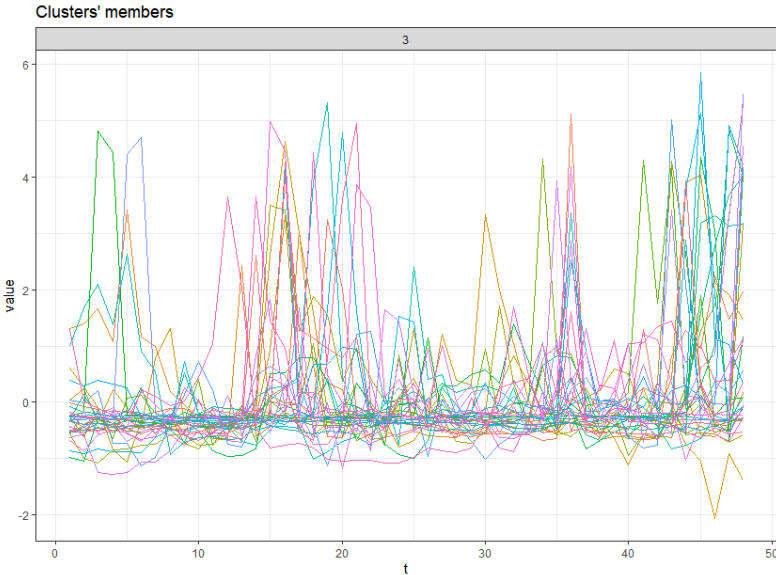
인테리어 목적에서 아웃도어 가구는  
좁은 공간을 언제나 손쉽게 변화시키는 기능적 용도도 컸지만  
그 자체만으로 심미성을 배가시키는 역할을 하기도.

\*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

# Cluster 03 - 35개 키워드 포함

## 집을 꾸미는 소비에 진심

- 코로나19 이후 집은 더 이상 bed place가 아니라 Life 그 자체가 되었고 과거에 소품 등을 활용해 소소하게 집을 꾸미는 것을 넘어 통째로 리모델링하거나 나만의 Identity를 담은 스타일을 구현하는 등 활동의 범위 증대



과자	사무용가구	블라인드	통조림
거실가구	생활가전	바디케어	남성언더웨어
무술용품	배드민턴	탁구	
제과재료	문구용품	벨트	
농산물	스킨케어	세탁용품	
잠옷	건강식품	냉동	
러그	족구	해외여행	
노트북	선케어	색조메이크업	
캔	유아세제	이유식용품	
가루	헬스	홈데코	
음향가전	유아침구	아기간식	

\*왼쪽 위에부터 아래쪽으로 Cluster 적합도 높은 순

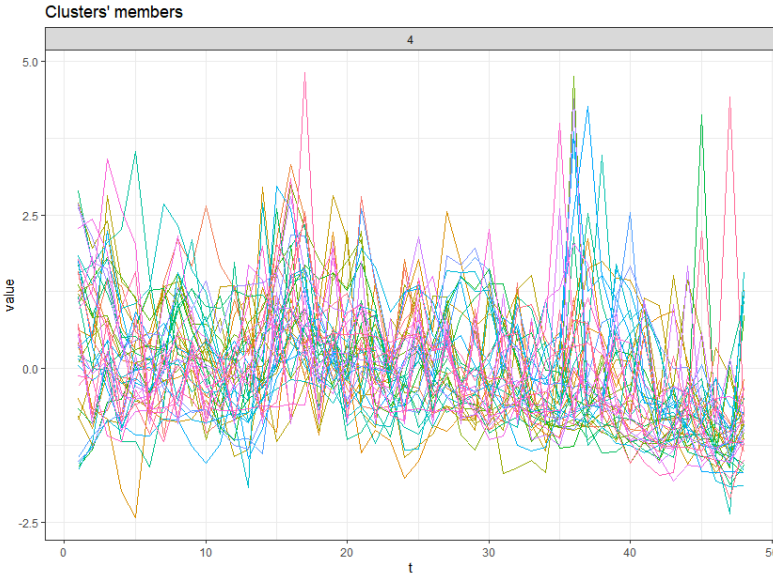
# Sweet Home

집을 꾸미거나, 집에서 하는 활동을 위한 다양한 상품군이 동시에 수요 집중

# Cluster 04 - 41개 키워드 포함

## 돈 쓰고 싶다~ Flex~

- 하이라이트 된 상품군들을 보면 전반적으로 취미에 가깝게 느껴지다가도 소위 돈을 좀 써야하는 상품들로 구성된 듯



여성가방	주방용품	오토바이	정원용품
태블릿pc	욕실용품	향수	diy자재
전자제품	국내여행	타이틀	렌터카
팩	당구용품	주니어가구	여성언더웨어
수납가구	바디용품	주얼리	신생아의류
스쿠터	볼링	정원	아동가구
스쿼시	야구	드레싱	여성의류
롤러	인테리어소품	원예	카메라
수납	지갑	주방가구	
게임기	수납용품	휴대폰	
클렌징	간편조리식품	안경테	

\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순

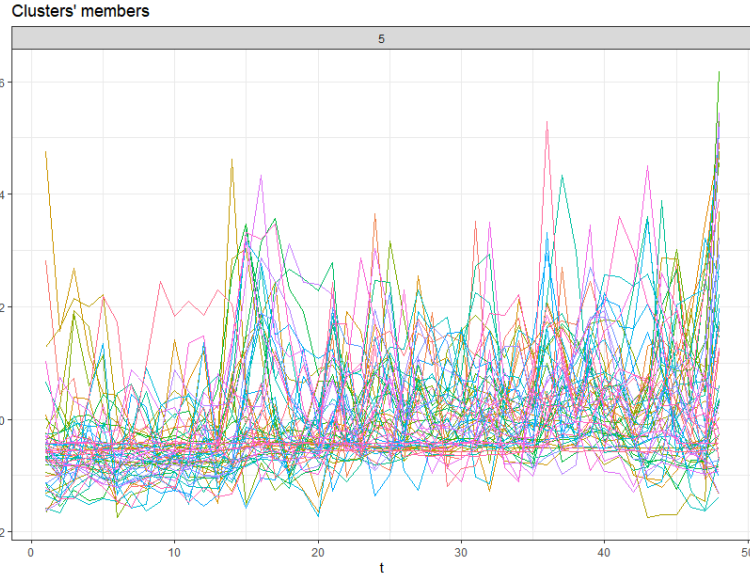
## New Flex~

새로운 취미를 위한 상품이거나

억눌린 소비가 폭발할 수 있는 상품군, 혹은 “지르는~” 행태를 대변

# Cluster 05 - 55개 키워드 포함 다양한 실버용품에 대한 수요 형성

- 물리치료용품, 좌훈, 찜질, 좌욕 등은 상대적으로 고연령대에서 나타날 수 있는 건강용품으로 판단되며 관상어, 원예, 제빵 등 역시 (최근에는 저연령층으로까지 확대되었으나) 여전히 고연령층이 주도하는 시장으로 판단



물리치료용품	찜질용품	베이커리	순금	모니터주변기기
음반	정리용품	마스크	수영복	제빵재료
실버용품	모니터	좌욕	요가	남성신발
좌훈용품	여성신발	라면	남성의류	자전거
카세트	김치	원예용품	화방용품	잼
원데이클래스	밀키트	인형	신발용품	기저귀
보드	사무용품	반찬	임부복	수산물
완구	골프	모자	스키	물티슈
관상어용품	체험	헤어케어	전통주	식물
주방가전	양말	시럽	테니스	매트
소스	다이어트식품	pc	여행용가방	인라인스케이트

## Silver

상위 품목에 아예 “실버용품”이 별도로 명시되어 있지만

그 외 상품들의 공통점을 상상해 보면

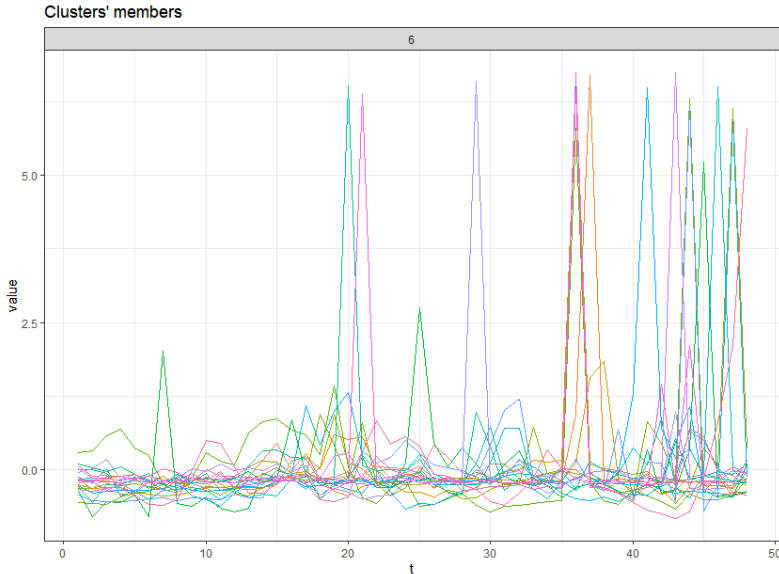
건강 관련, 취미 관련 등인데

고연령층의 수요에 부합하는 점이 있기도..

\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순

# Cluster 06 - 21개 키워드 포함 격렬한 스포츠 활동에도 끄떡없는 뷰티용품 추구

- 6개의 Cluster 중 비교적 적은 키워드들을 포함하고 있음에도, 상대적으로 높은 설명력을 가지는 Cluster로 판단



분류	침실가구
수집품	유아가구
스킨스쿠버	수예
권투	서재
낚시	반려동물
검도	뷰티소품
등산	식용유
남성화장품	카페트
출산	저장장치
축구	악기
베이스메이크업	

## Sports Makeup

스킨스쿠버, 권투, 낚시, 등산 등의 스포츠/아웃도어 활동과

남성화장품, 베이스메이크업, 뷰티소품 등 뷰티 관련 상품군이 같은 Cluster로 묶인 사례

앞서 상관관계 분석에서도 예시를 들었으나

야외 활동에서도 메이크업에 대한 관심이 증가중이며

이러한 활동에서 요구되는 메이크업의 경우

끝까지 유지되는 기능성도 중요하지만 한 듯 안 한 듯 '꾸안꾸' 니즈가 높게 형성

\*왼쪽 위에서부터 아래쪽으로 Cluster 적합도 높은 순



# 이제 결론

# Conclusion

반복해서 말하지만, 본 보고서의 결론은  
어떤 트렌드가 있다.. 보다, “이렇게도 온라인 데이터를 분석할 수 있다!”는 것이다.

빅데이터를 대중적으로 알린, **세스 스티븐스 데이비도위츠(Seth Stephens-Davidowitz)**는  
구글에서 검색되는 단 하나의 키워드를 발견함으로써 미국 트럼프 대통령의 당선을 예측했다.

물론 그러한 시도가 또 다른 여론조사 기법의 주류로 안착되지는 못했지만

우리가 의식조차 못한채로 빈번히 사용하고 있는 “**검색**”과 “**클릭**”이라는 행동은  
빅데이터가 본격적으로 활용되면 어떠한 결과를 가져다 줄 수 있는지, 대대적으로 증명했다.

그러한 실천적 사례가 존재함에도  
**여전히 온라인에서 생성되는 데이터가 제한적으로 활용되고** 있다는 것은 안타까운 일이다.

그래서 우리는 연구된 자료들을 무한히 공유하는 일을 하고 있다.

누군가 우리가 무한히 공유하는 자료들을 보고 천재적인 해법으로  
우리가 따라할 수 있는 방법론들을 만들어주기를 바라면서.

# Next Step

우리의 다음 스텝은 중분류를 넘어 **몇 천 개의 소분류를 대상으로** 지금까지 보여준 작업들을 되풀이하는 것이다.

소분류는 개별 품목에 가깝다. 그만큼 세부적으로 잘 쪼개져 있다는 것이다.

그 상품들을 대상으로 상호 연관성을 조사하고 Cluster로 묶어내는 작업들을 할 것이다.

상품이 세부 단위로 갈수록 꽤 의미있는 결과들을 뱉어내 줄거라 기대한다.  
[일부는 흑여 인과성을 보여줄지도 모르겠다.]

To be Continued..

# End Of Document

## Contact Us

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

T 02. 6177. 1871      eMAIL [khbak@emforce.co.kr](mailto:khbak@emforce.co.kr)