

제5회 L.POINT Big Data Competition

# DIGITAL TREND Analyzer

#Team 빅토리 : 이호창, 전도현, 정채영

# CONTENTS

1. 분석 데이터 & 프로그램
2. 데이터 전처리
3. 온라인 선호지수
4. 수요 트렌드 및 인사이트
5. 신규서비스 제안

# 1 분석 데이터 & 프로그램

데이터	내용	출처
엘포인트	Product	
	Search1/Search2	
	Session	
	Master	
	Custom	
공공데이터	서울시 일별 평균 대기 오염도 정보	서울 열린데이터광장
네이버 데이터랩	화장품/미용, 식품, 생활/건강	쇼핑인사이트 분야통계
소비자동향지수	2018.04-2018.09사이 지수	KOSIS (한국은행, 소비자동향조사)

# 1 분석 데이터 & 프로그램

---

프로그램	사용	
	ORACLE SQL DEVELOPER	데이터 정제
	R Studio	데이터 전처리 및 통계분석
	Excel	데이터 전처리

---

## 2 데이터 전처리

상품군 - Class 2(중분류)에서 선정



화장품



퍼스널케어



냉동식품



건강식품



애완용품

## 2 데이터 전처리

### SQL DEVELOER 를 사용하여 필요한 상품군 데이터 추출

① 모든 상품에 대한 정보가 있는 MASTER 파일에서 필요한 상품군 필터 적용 하여 분류

(EX) 상품군 중 건강식품

D_C	PD_NM	CLACT_NK	CLACT_Z_NK	CLACT3_NM
64382	언더아머 !스프츠패션	남성	일반	남성스프츠티셔츠
62282	여자 플라!속옷/양말	여성	양말	여성일반양말
61729	88A29332! 유아들의	유아	유아	영유아티셔츠/탑
61537	닥터마틴 (패션잡화	남성	화	남성부츠
58820	여성 그레!남성의류	남성	의류	남성남방셔츠
58280	선업스트로!여성의류	여성	의류	여성원피스
58183	[175/5] 플!남성의류	남성	의류	남성캐주얼재킷
58098	V넥 고방!여성의류	여성	의류	여성원피스
58057	백색포인트!여성의류	여성	의류	여성원피스
56306	선업 체크!속옷/양말	여성	속옷	여성잠옷
55076	정장형!스포츠의류	여성	의류	여성바지
318122	H1280YR! 패션잡화	여성	가방	여성토트백
318011	가법고 편!스포츠패션	여성	스포츠	여성현닝/트레이닝화
317956	체크패턴!여성의류	여성	의류	여성스커트
127096	(KM)코찌!여성의류	여성	의류	여성청바지
125372	88-TS018	여성	의류	여성티셔츠/탑
124084	여성백백!패션잡화	여성	가방	여성백팩
123395	네이비 하!속옷/양말	여성	속옷	여성잠옷
123233	빅사이즈!여성의류	여성	의류	여성스커트
122845	(2XL-5XL)!여성의류	여성	의류	여성원피스
210546	컴포트 레!패션잡화	패션	액세서리	기타패션잡화
209216	코튼 롱방!유아들의	유아	의류	남아티셔츠/탑
216029	[보리보리] 유아들의	유아	의류	여아가디건
214635	280D 사틴 속옷/양말	여성	속옷	슬리퍼
213856	한미 연초!건강식품	영양제	오메가3/기타추출물	
213204	남여 솔리!스포츠패션	남성	등산	남성등산바지
213185	스킬렛 16!식기/조리	조리기구	냄비	
209786	쿠키커터!주방잡화	조리도구	제빵용품	
206498	아일렛 A!여성의류	여성	의류	여성남방셔츠
202585	[스타일론!여성의류	여성	의류	여성티셔츠/탑
201823	(GN)프릴!속옷/양말	유아	의류	유아동타이츠
266521	그레이 남!스포츠패션	남성	등산	남성등산티셔츠



197837	6400453	224194	건강보조	미용보조	1
54073	352351	777526	영양제	기타영양제	1
564990	6663007	764050	건강진액	채소즙	2
2815692	10677305	799725	영양제	기타영양제	1
5929151	9525175	764050	건강진액	채소즙	50
4109507	8023889	764050	건강진액	채소즙	10
5445133	6973961	764050	건강진액	채소즙	20
1174878	720939	764050	건강진액	채소즙	3
1623511	8481013	799725	영양제	기타영양제	1
3740538	3738045	259916	건강보조	미용보조	2
440002	6084688	242001	영양제	일반비타민	1
5879631	1115364	454249	영양제	일반비타민	7
127968	2719223	764050	건강진액	채소즙	5
2792811	4288137	764050	건강진액	채소즙	21
5445133	6973961	764050	건강진액	채소즙	20
3348073	4049151	764050	건강진액	채소즙	60
5879631	1741166	454249	영양제	일반비타민	2
2401276	7020928	679375	건강진액	채소즙	1
703414	5403473	259916	건강보조	미용보조	1
1157604	3339017	454249	영양제	일반비타민	5
2121065	8899088	354532	건강진액	과일즙	1
5211193	4537062	185648	건강보조	다이어트보조	1
6567096	7274897	224194	건강보조	미용보조	1
3791556	6078192	679375	건강진액	채소즙	1
682051	7390037	472973	영양제	기타영양제	1
6401558	10658050	472973	영양제	기타영양제	5
5972389	391303	777526	영양제	기타영양제	5
3262480	10849065	472973	영양제	기타영양제	1
6382580	10565142	679375	건강진액	채소즙	1
728915	10673585	679375	건강진액	채소즙	1
5701192	9192399	472973	영양제	기타영양제	1
5981091	1910085	66350	영양제	기타영양제	1
1195939	1629609	66350	영양제	기타영양제	1

처음 MASTER파일의 데이터는 약80만개의행 => 필요한건강식품에해당되는 데이터는약8천개로 정제완료

## 2 데이터 전처리

### SQL DEVELOER 를 사용하여 필요한 상품군 데이터 추출

② PRODUCT와 HEALTHFOOD(MASTER에서 건강식품만 담겨있는 테이블)를 조인하여 구매목록 생성

(EX) 상품군 중 건강식품

PRODUCT table structure:

CLNT_ID	SESS_ID	PD_C	CLAC2_NM	CLAC3_NM	PD_BUY_CT	PD_BUY_C
1	1049501	761674	건강간식	채소즙	1	1049501
2	1111162	728089	건강보조식품	다이어트보조식품	1	1111162
3	1129695	564436	건강간식	혼합즙	1	1129695
4	1135532	336541	영양제	기타영양제	1	1135532
5	1246132	496728	홍삼/인삼가공식품	머리나눔	1	1246132
6	1259546	740296	건강간식	혼합즙	1	1259546
7	1540317	199305	영양제	영양제세트	1	1540317
8	1569021	819778	영양제	일반비타민	1	1569021
9	1572739	810343	건강보조식품	다이어트보조식품	1	1572739
10	158233	834286	영양제	기타영양제	1	158233
11	162411	820999	영양제	자연유래영양제	1	162411
12	1641351	829863	영양제	종합영양제	1	1641351
13	1645305	243114	건강보조식품	미용보조식품	1	1645305
14	1668719	626119	홍삼/인삼가공식품	머리나눔	1	1668719
15	1674853	627651	영양제	유산균/프로바이오틱스	1	1674853
16	1723101	768892	건강간식	과일즙	1	1723101
17	1856100	315	영양제	종합영양제	1	1856100
18	1864797	238039	건강보조식품	다이어트보조식품	1	1864797
19	1866698	754549	홍삼/인삼가공식품	홍삼정/분말/환	1	1866698
20	1866721	754549	홍삼/인삼가공식품	홍삼정/분말/환	1	1866721

PRODUCT

```
SELECT COUNT_ID,  
PSESS_ID,  
PPD_BUY_CT,  
HPD_C,  
HCLAC2_NM,  
HCLAC3_NM
```

```
FROM PRODUCT P  
INNER JOIN
```

```
HEALTHFOOD H
```

```
ON PPD_C=HPPD_C;
```

HEALTHFOOD table structure:

CLNT_ID	SESS_ID	PD_C	CLAC2_NM	CLAC3_NM	PD_BUY_CT	PD_BUY_C
1	1049501	761674	건강간식	채소즙	1	1049501
2	1111162	728089	건강보조식품	다이어트보조식품	1	1111162
3	1129695	564436	건강간식	혼합즙	1	1129695
4	1135532	336541	영양제	기타영양제	1	1135532
5	1246132	496728	홍삼/인삼가공식품	머리나눔	1	1246132
6	1259546	740296	건강간식	혼합즙	1	1259546
7	1540317	199305	영양제	영양제세트	1	1540317
8	1569021	819778	영양제	일반비타민	1	1569021
9	1572739	810343	건강보조식품	다이어트보조식품	1	1572739
10	158233	834286	영양제	기타영양제	1	158233
11	162411	820999	영양제	자연유래영양제	1	162411
12	1641351	829863	영양제	종합영양제	1	1641351
13	1645305	243114	건강보조식품	미용보조식품	1	1645305
14	1668719	626119	홍삼/인삼가공식품	머리나눔	1	1668719
15	1674853	627651	영양제	유산균/프로바이오틱스	1	1674853
16	1723101	768892	건강간식	과일즙	1	1723101
17	1856100	315	영양제	종합영양제	1	1856100
18	1864797	238039	건강보조식품	다이어트보조식품	1	1864797
19	1866698	754549	홍삼/인삼가공식품	홍삼정/분말/환	1	1866698
20	1866721	754549	홍삼/인삼가공식품	홍삼정/분말/환	1	1866721

HEALTH FOOD

HEALTH\_BUY table structure:

CLNT_ID	SESS_ID	PD_C	CLAC2_NM	CLAC3_NM	PD_BUY_CT	PD_BUY_C
1	1049501	761674	건강간식	채소즙	1	1049501
2	1111162	728089	건강보조식품	다이어트보조식품	1	1111162
3	1129695	564436	건강간식	혼합즙	1	1129695
4	1135532	336541	영양제	기타영양제	1	1135532
5	1246132	496728	홍삼/인삼가공식품	머리나눔	1	1246132
6	1259546	740296	건강간식	혼합즙	1	1259546
7	1540317	199305	영양제	영양제세트	1	1540317
8	1569021	819778	영양제	일반비타민	1	1569021
9	1572739	810343	건강보조식품	다이어트보조식품	1	1572739
10	158233	834286	영양제	기타영양제	1	158233
11	162411	820999	영양제	자연유래영양제	1	162411
12	1641351	829863	영양제	종합영양제	1	1641351
13	1645305	243114	건강보조식품	미용보조식품	1	1645305
14	1668719	626119	홍삼/인삼가공식품	머리나눔	1	1668719
15	1674853	627651	영양제	유산균/프로바이오틱스	1	1674853
16	1723101	768892	건강간식	과일즙	1	1723101
17	1856100	315	영양제	종합영양제	1	1856100
18	1864797	238039	건강보조식품	다이어트보조식품	1	1864797
19	1866698	754549	홍삼/인삼가공식품	홍삼정/분말/환	1	1866698
20	1866721	754549	홍삼/인삼가공식품	홍삼정/분말/환	1	1866721

HEALTH\_BUY

PRODUCT의PD\_C (상품코드)를 통해 건강식품테이블 과이너조인함으로써건강식품을구입한정보에대한테이블생성  
INNER JOIN을 한 이유: 먼저 정제한 HEALTHFOOD에상품코드와 구매한상품의코드의교집합데이터를 추출

## SQL DEVELOER 를 사용하여 필요한 상품군 데이터 추출

(EX) 상품군 中 건강식품

HEALTH\_BUY

## INNER JOIN

## SESSION

HEALTH\_SESSION

HEALTH\_BUY와SESSION테이블을 INNER JOIN을하여구매한사람들중에서의세션날짜 & 총페이지조회수추출  
HEALTH\_BUY테이블의SESS\_ID와SESSION의SESS\_ID를통해 INNER JOIN(+똑같은 논리로HEALTH\_BUY테이블 기준으로RIGHT OUTER JOIN 결과도 같게 나옴)



RIGHT OUTER JOIN으로 JOIN하면, 정제해왔던 데이터들을 갖고있는 HEALTH\_SESSION을 기준으로 데이터의 교차점을 불러오기 때문에 CUSTOM 데이터의 CLIENT ID들은 거스름품을 산 사람도 있고 사지 않은 사람도 존재

## 2 데이터 전처리

### 일별 평균 서울의 미세먼지 데이터 - Excel SUMIF 함수, 측정소 39개 평균

Ex) SUMIF(\$A\$1:\$A\$5,E1,\$C\$1:\$C\$5)

일자	측정소명	PM10	PM2.5
20180409	강남구	35	26
20180409	강남대로	59	36
20180409	강동구	33	20
20180409	강변북로	48	31
20180409	강북구	32	18
	.		
	.		
	.		



일자	PM10	PM2.5	평균PM10	평균PM2.5
20180409	1624	1034	41.64103	26.51282
20180410	1935	803	49.61538	20.58974
20180411	2608	996	66.87179	25.53846
20180412	1904	608	48.82051	15.58974
20180413	1480	575	37.94872	14.74359
		.		
		.		
		.		

### 3 온라인 선호지수

---

#### 선호지수 개발 - U,P,M Score

##### U-Score (Usage Score)

: 고객의 아이템 이용 현황을  
바탕으로 고객의  
컨텐츠에 대한 이용 정도

##### P-Score (Preference Score)

: 다른 컨텐츠 대비  
해당 컨텐츠의  
상대적인 선호 정도

##### M-Score (Mania Score)

: 타 고객 그룹에 대비하여  
많이 이용할 수록  
높게 도출

### 3 온라인 선호지수

#### 선호지수 개발 - U,P,M Score

##### 화장품

	남성케어	메이크업	미용소품	선케어	스킨케어	향수
U-score	16901	212029	24736	20611	174184	17831
P-score	0.217	2.728	0.318	0.264	2.241	0.229
M-score	2.134	2.380	2.224	1.793	2.462	2.084

##### 퍼스널케어

	구강케어	바디케어	여성위생용품	핸드/풋케어	헤어케어
U-score	27821	41150	28043	13406	42041
P-score	0.912	1.350	0.920	0.440	1.379
M-score	2.678	1.990	3.375	2.270	2.165

### 3 온라인 선호지수

#### 선호지수 개발 - U,P,M Score

##### 냉동식품

	냉동만두	냉동밥	냉동국탕류	냉동핫도그	냉동튀김	기타냉동간편식
U-score	2445	2907	2391	1256	565	12834
P-score	0.6549692	0.7787302	0.6405036	0.3364586	0.1513528	3.4379855
M-score	2.171403	5.423507	3.783228	1.679144	2.100372	4.629870

##### 건강식품

	건강보조식품	건강진액	영양제	홍삼/인삼가공식품
U-score	19090	15931	48570	42610
P-score	0.605	0.505	1.539	1.350
M-score	3.041	2.816	2.180	2.84

### 3 온라인 선호지수

#### 선호지수 개발 - U,P,M Score

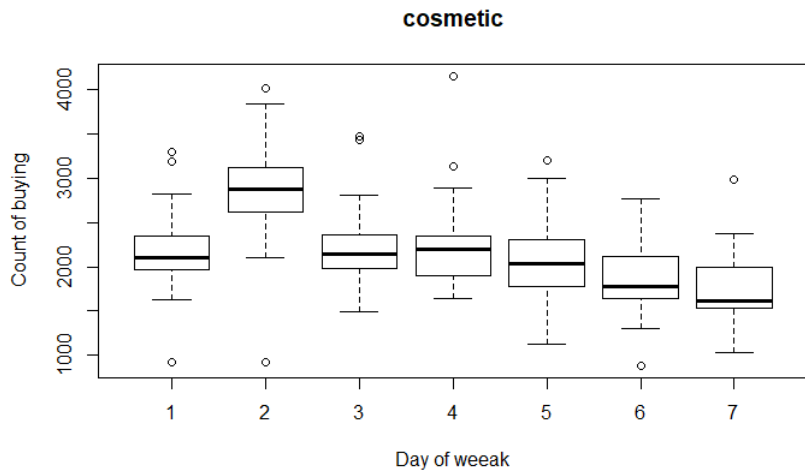
##### 애완용품

	고양이용품	애견용품
U-score	51611	82385
P-score	0.770	1.230
M-score	3.076	2.413

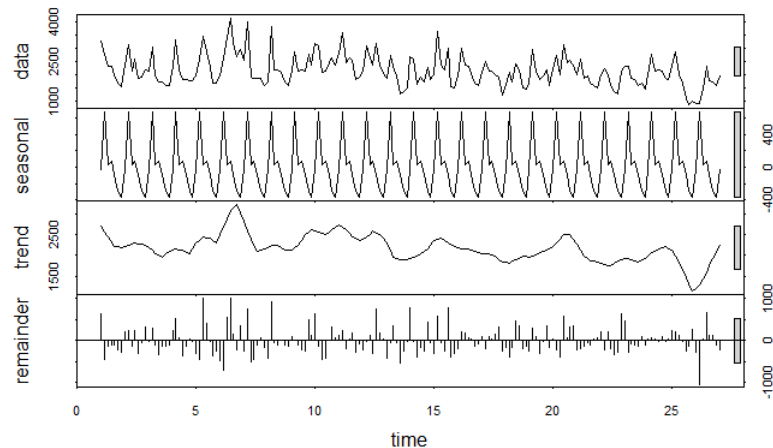
그렇다면, 고객들의 이용정도,선호정도(구매량)는 시간에 따른 패턴이 있을까?

## 4 수요 트렌드 및 인사이트

### 시계열 분석 - 화장품



1: 일요일, 2:월요일, ..., 7: 토요일



시계열 데이터: seasonal 데이터, 일주일 간격으로 매번 같은 모양

구매건수는 주중 **월요일**이 가장 **높고** 수요일 또한 높다  
월별은 **4월 중순**에 가장 **높고**, **9월 추석** 기간에 가장 **낮다**

## 4 수요 트렌드 및 인사이트

### Pearson 상관 - 화장품 구매건수와 소비자심리지수

Pearson's product-moment correlation

```
data: custom and buy.cosmetic
t = 2.3269, df = 4, p-value = 0.08052
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1383268  0.9718146
sample estimates:
      cor
0.7583696
```

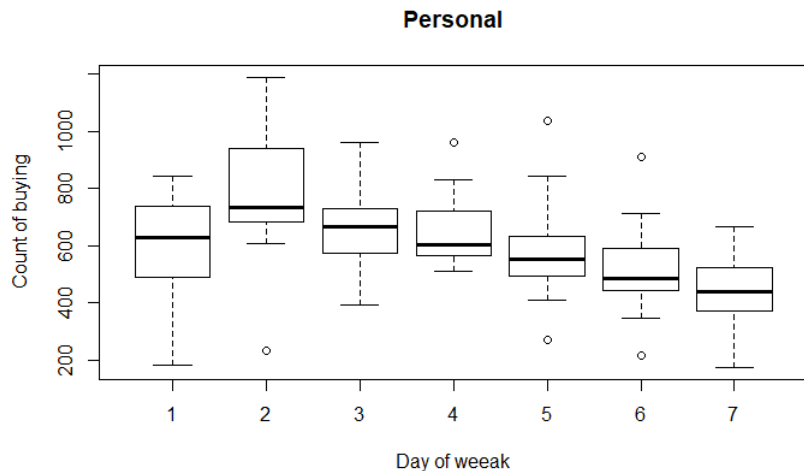
\*P-value가 0.0805로 그렇게 낮지 않기 때문에 어느 정도 상관성을 보인다

다른 상품군에 비해 획기적으로 저렴한 값으로 **화장품**은 **소비자심리지수**와 **관련성**이 있다

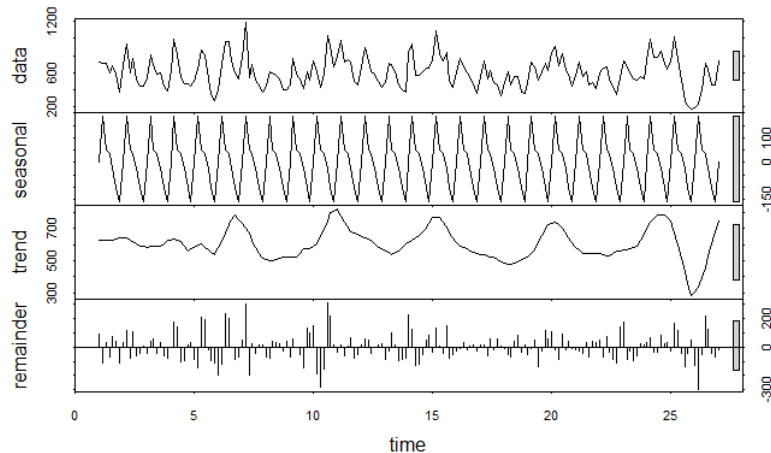


## 4 수요 트렌드 및 인사이트

### 시계열 분석 - 퍼스널케어



1: 일요일, 2: 월요일, ..., 7: 토요일

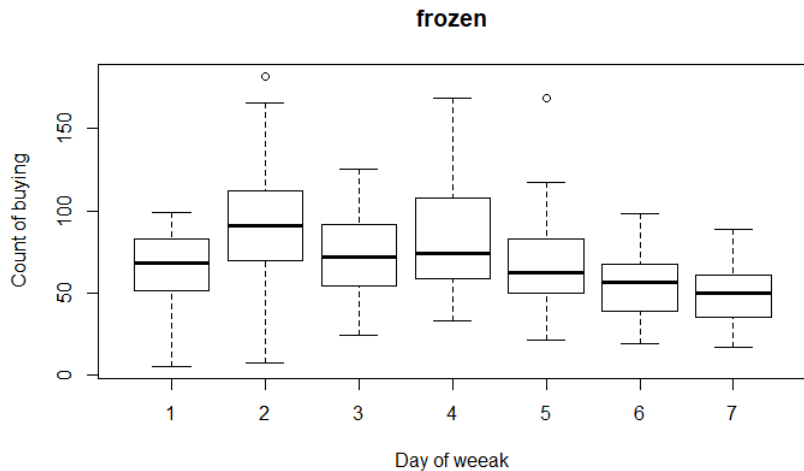


시계열 데이터: seasonal 데이터, 일주일 간격으로 매년 같은 모양

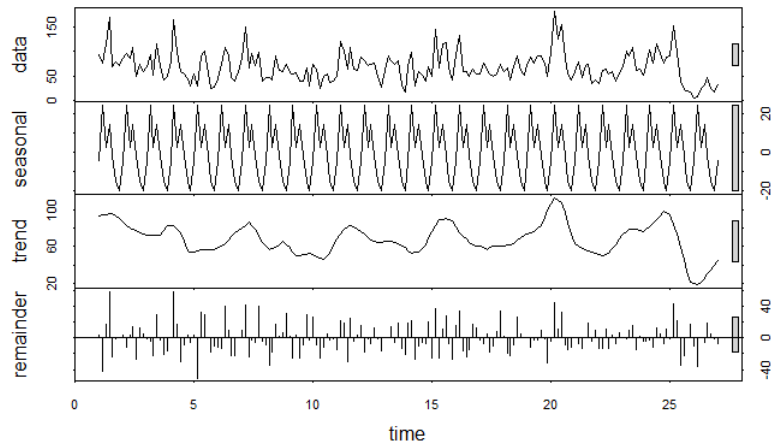
다른 제품과 마찬가지로 **월요일**에 구매건수가 가장 **높고**, **토요일&일요일**에 가장 **낮다**

## 4 수요 트렌드 및 인사이트

### 시계열 분석 - 냉동식품



1: 일요일, 2: 월요일, ..., 7: 토요일

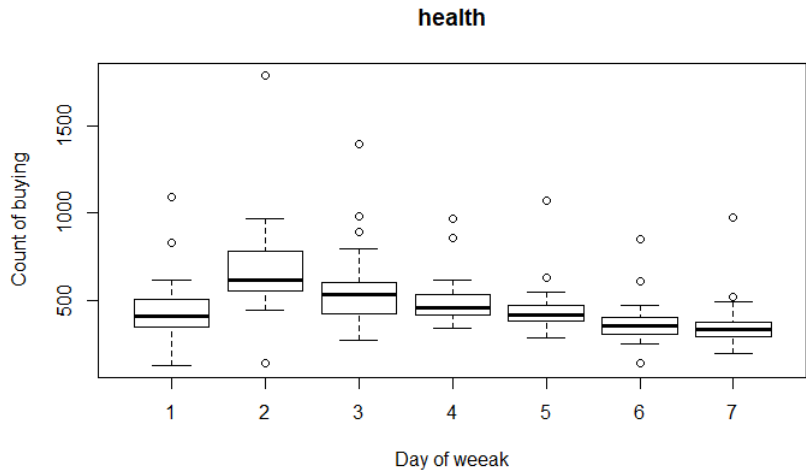


시계열 데이터: seasonal 데이터, 일주일 간격으로 매번 같은 모양

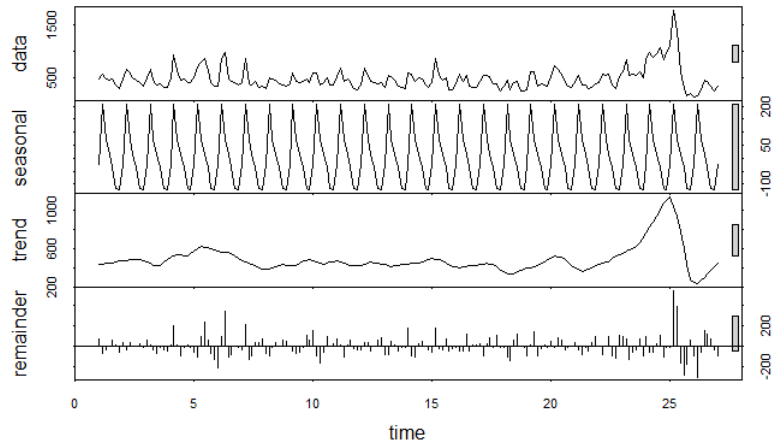
요일에 따라 그렇게 큰 차이는 없지만, **월요일**과 **수요일**에는 **편차**가 **크다**  
**추석**이 되면 구매건수가 **최하**로 떨어지는 것은 동일하다

## 4 수요 트렌드 및 인사이트

### 시계열 분석 - 건강보조식품



1: 일요일, 2:월요일, ..., 7: 토요일



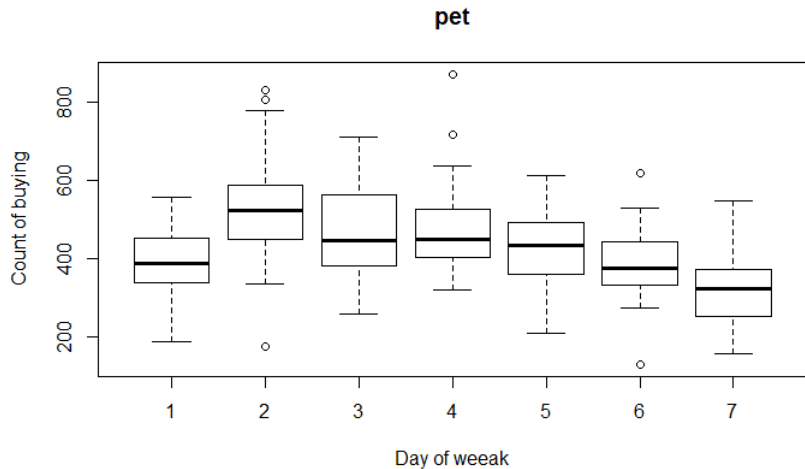
시계열 데이터: seasonal 데이터, 일주일 간격으로 매번 같은 모양

\*요일 별 편차가 크지 않지만 outlier들이 존재! 9월에 구매건수가 갑자기 증가하고 M-score가 높기 때문

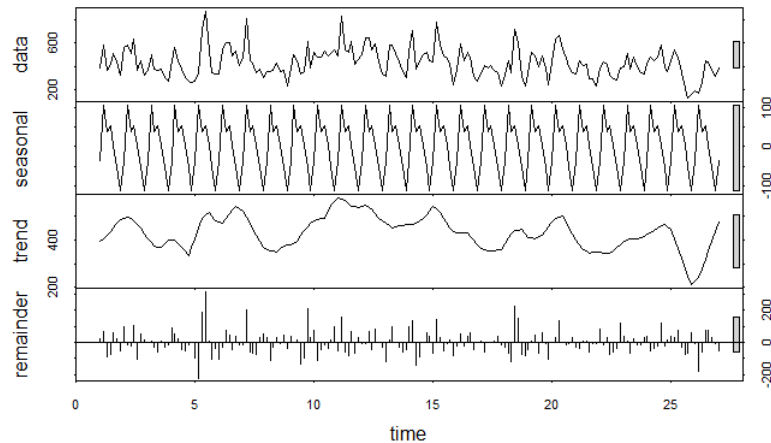
추석 당일에는 다른 제품과 마찬가지로 구매건수가 떨어지지만, 9월에 추석 선물을 위한 구매가 증가한다  
기존에 구매하는 사람들만 구매하는 상품이다

## 4 수요 트렌드 및 인사이트

### 시계열 분석 - 애완동물



1: 일요일, 2: 월요일, ..., 7: 토요일



시계열 데이터: seasonal 데이터, 일주일 간격으로 매년 같은 모양

마찬가지로 **월요일**에 구매건수가 가장 **높고**, **토요일&일요일**에 가장 **낮다**  
**5월**은 **가정의 달**이고, **외출**을 하기 **좋은 날**이 많기 때문에 **수요**가 **늘어난다**

## 4 수요 트렌드 및 인사이트

일간 서울 미세먼지 수준은 네이버 검색량 & 상품 구매건수에 영향을 미칠까?

미세먼지 수준 표

수준	미세먼지	초미세먼지
좋음	0~30 $\mu\text{m}$	0~15 $\mu\text{m}$
보통	31~80 $\mu\text{m}$	16~35 $\mu\text{m}$
나쁨	81~150 $\mu\text{m}$	36~75 $\mu\text{m}$
매우나쁨	151~ $\mu\text{m}$	76~ $\mu\text{m}$

?



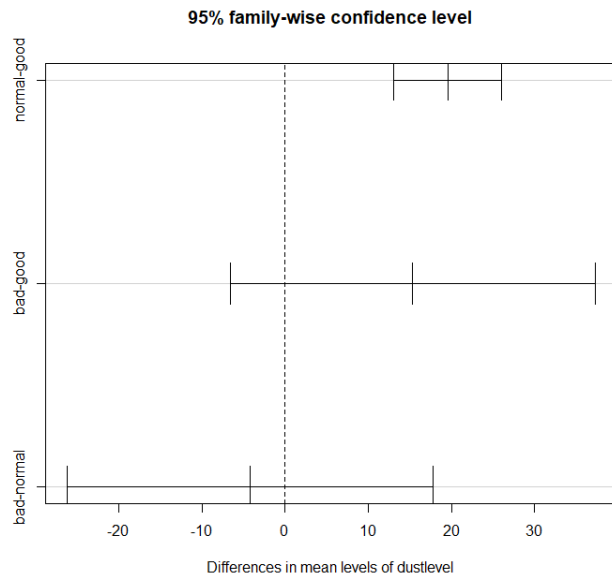
관련 있을 것 같은 상품군

클렌징  
스킨케어  
선크어  
헤어케어  
마스크/팩

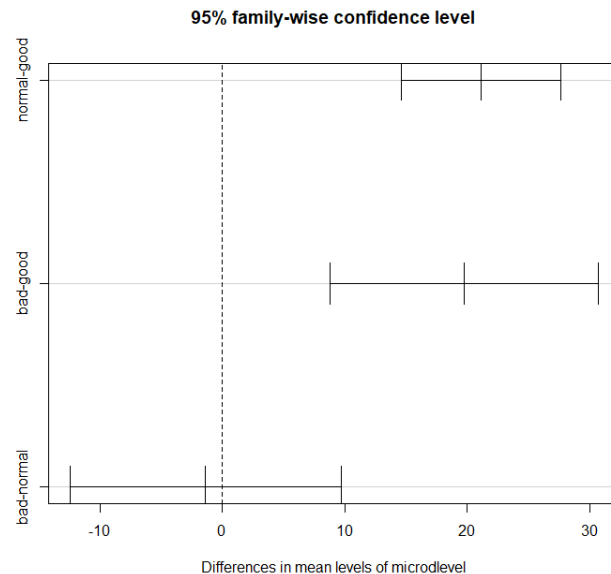
출처: '바로 알면 보인다. 미세먼지, 도대체 뭘까?', 환경부, 2016

## 4 수요 트렌드 및 인사이트

### 분산분석(One-way ANOVA) – 미세먼지와 선행어 검색량



미세먼지: 정도가 보통-좋음 일 때 차이 발생

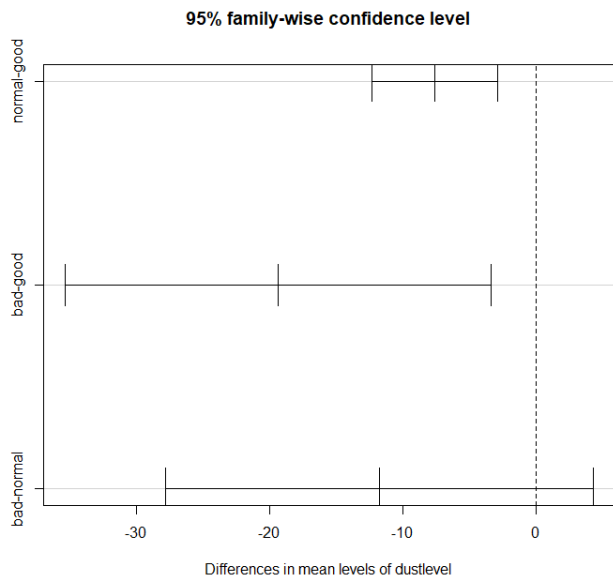


초미세먼지: 정도가 보통-좋음, 나쁨-좋음 일 때 차이 발생

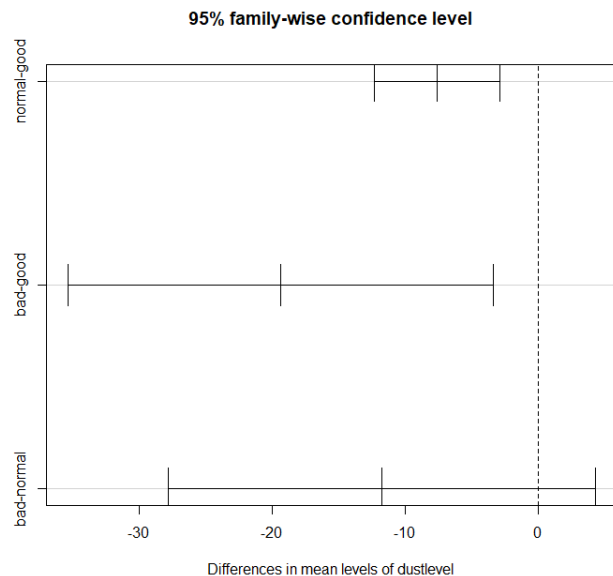
일간 서울 **미세&초미세먼지** 수준에 따라 **선행어와 헤어케어** 상품의 **검색량**이 다르다

## 4 수요 트렌드 및 인사이트

### 분산분석(One-way ANOVA) – 미세먼지와 헤어케어 검색량



미세먼지: 정도가 보통-좋음, 나쁨-좋음 일 때 차이 발생



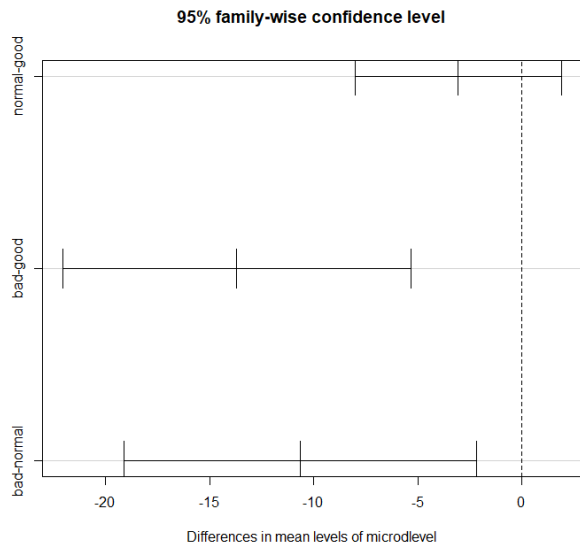
초미세먼지: 정도가 보통-좋음, 나쁨-좋음 일 때 차이 발생

일간 서울 **미세&초미세먼지** 수준에 따라 **선케어와 헤어케어** 상품의 **검색량**이 다르다

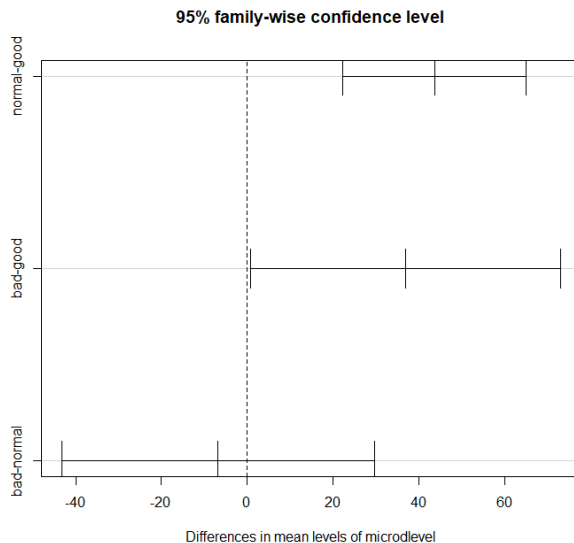
## 4 수요 트렌드 및 인사이트

### 분산분석(One-way ANOVA) – 미세먼지와 선크어 구매건수

; 미세먼지와 검색량 분산분석에서 의미 있었던 상품군만 가지고 분석



미세먼지: 정도가 나쁨-좋음, 나쁨-보통 일 때 차이 발생



초미세먼지: 정도가 보통-좋음, 나쁨-좋음 일 때 차이 발생

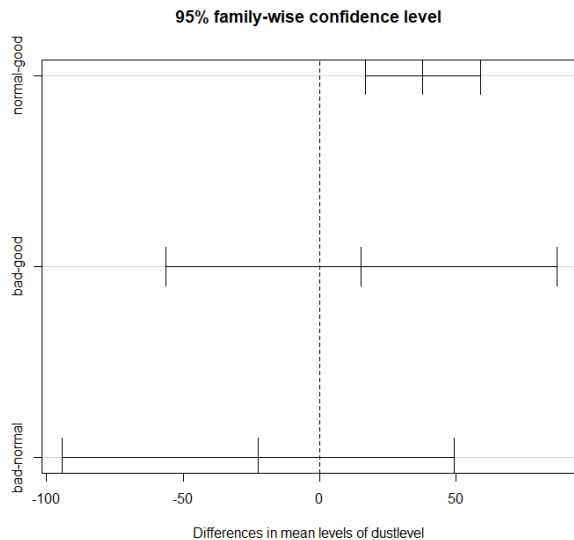
일간 서울 미세&초미세먼지 수준에 따라 선크어와 헤어케어 상품의 구매건수가 다르다



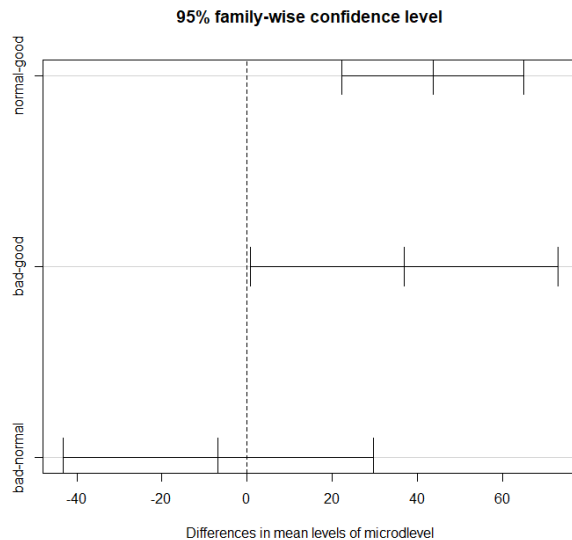
## 4 수요 트렌드 및 인사이트

### 분산분석(One-way ANOVA) – 미세먼지와 헤어케어 구매건수

; 미세먼지와 검색량 분산분석에서 의미 있었던 상품군만 가지고 분석



미세먼지: 정도가 보통-좋은 일 때 차이 발생



초미세먼지: 정도가 보통-좋은, 나쁨-좋은 일 때 차이 발생

일간 서울 미세&초미세먼지 수준에 따라 선풍기와 헤어케어 상품의 구매건수가 다르다

## 4 수요 트렌드 및 인사이트

### 선케어 & 헤어케어 제품의 수요 트렌드



나뻘-좋음, 보통-좋음과 같이 그날의 **미세먼지 수준**에 따라  
**선 케어** 관련 제품과 **헤어 케어** 관련 제품의 관심과 **수요**가 **달라질** 것이다.

## 5 신규 서비스 제안

MONDAY

### 월요일 이벤트 공략

온라인 선호지수 & 요일별  
구매수량 확인 결과  
월요일이 가장 중요



U-Score가 가장 높은 제품군 광고  
U-score가 낮은 제품과의  
결합 상품 추천 서비스 공략

U, M, P SCORE

### 선호지수 고려

M-Score가 높은 상품일수록  
지속적인 소비층 다  
이들을 잡기 위한 서비스!



고객의 구매데이터를 분석하여  
같은 물품을 x번 구매 시,  
1번의 무료 이벤트  
비슷한 상품군 할인권 발매 서비스

WEATHER

### 미세먼지 정보 연동

미세먼지 농도와  
뷰티 케어 제품과의 관련성



L.Point 앱에 대기정보 연동해서  
광고, 맞춤형 쿠폰 제공하기  
Ex) 미세먼지 심한 날  
헤어 케어 제품 광고, 할인쿠폰 제공

## 5 신규 서비스 제안

### 소비자동향지수

#### 소비자동향지수(CSI)활용

소비자동향지수를  
활용하여 화장품 판매전략 수립



화장품과 CSI는 양의 상관관계  
CSI가 낮을 때는 저가 화장품,  
높을 때는 고가 화장품을 광고

### 상품 맞춤형 광고

#### 상품군 별 특색 반영

화장품 - 선물용으로 5월에 판매 증가

냉동식품 - 요리하기 힘든 여름과 추석 전에  
가장 판매량이 높음

건강식품 - 구매하는 사람만 구매  
- 추석전에만 급격히 판매됨

애완용품 - 외출하기 좋은 날 & 가정의 달 5월에  
잘 판매되는 추세



상품군에 맞춰 월별로  
해당 상품에 대한  
맞춤형 광고 필요

감사합니다.