

# NS SHOP+ 판매실적 예측을 통한 편성 최적화 방안(모형) 도출

...

ESC 2조

팀장 신예진 yuree0910@naver.com  
팀원 김정현 coby.kim27@gmail.com  
팀원 이규민 minzzang68@naver.com  
팀원 최정욱 koo802@naver.com  
팀원 한인욱 gorany131@gmail.com

## Contents

1. 문제정의	01
2. 데이터 분석 및 변수 생성	00
3. 데이터 전처리	00
4. 모델링	00
5. 편성 최적화 방안 제안	00

## 1. 문제 정의 ...

NS Shop+편성데이터(NS홈쇼핑) 를 활용하여  
방송편성표에 따른 판매실적을 예측하고,  
최적 수익을 고려한 요일별/ 시간대별 / 카테고리별 편성 최적화 방안(모형) 제시

## 2. 데이터 분석 및 변수 생성 ...

**분석1** test set(제공 데이터)와 train set(평가 데이터)의 기간 상이

test

2019년 1월 1일 ~ 12월 31일 >> 1년, 총 35379번의 방송편성 (무형, 취급액 NA 상품 제외)

train

2020년 6월 1일 ~ 6월 30일 >> 한 달, 총 2716번의 방송편성 (무형, 취급액 NA 상품 제외)

## 2. 데이터 분석 및 변수 생성 ...

### 변수1 분기

분기 별 취급액의 차이를 설명하기 위해 해당 변수를 추가하였다.

### 변수2 온도

기온에 따라 특정 상품군의 매출에 유의미한 변동이 있을 것이라 예측하여 외부 데이터를 이용한 변수를 추가하였다.

모두 기상청의 기상자료개방포털의 데이터를 사용했다.  
(<https://data.kma.go.kr/stcs/grnd/grndTaList.do?pgmNo=70>)

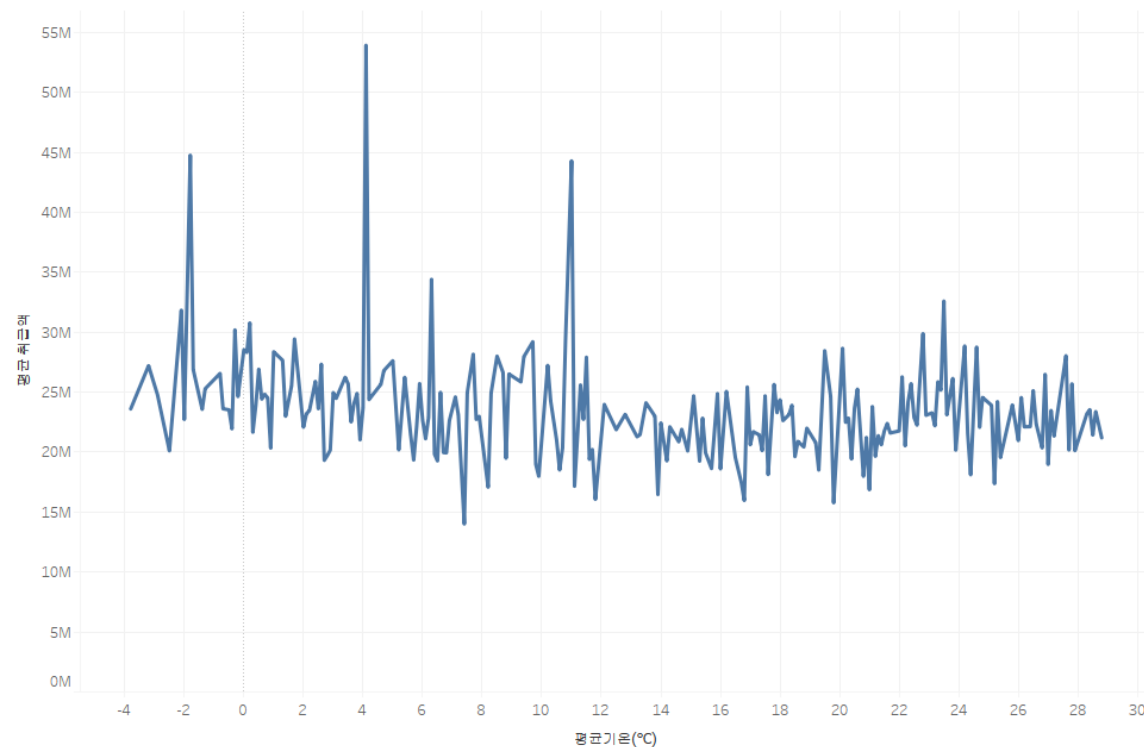
전국 평균은 기후통계분석에 활용되는 45개 지점의 평균으로 산출된다.

"평균기온(°C)": 일 별 전국 평균기온의 평균치에 대한 변수.

"최저기온(°C)": 일 별 전국 최저기온의 평균치에 대한 변수.

"최고기온(°C)": 일 별 전국 최고기온의 평균치에 대한 변수.

3. 기온



평균 기온과 평균취급액의 그래프

## 2. 데이터 분석 및 변수 생성 . . .

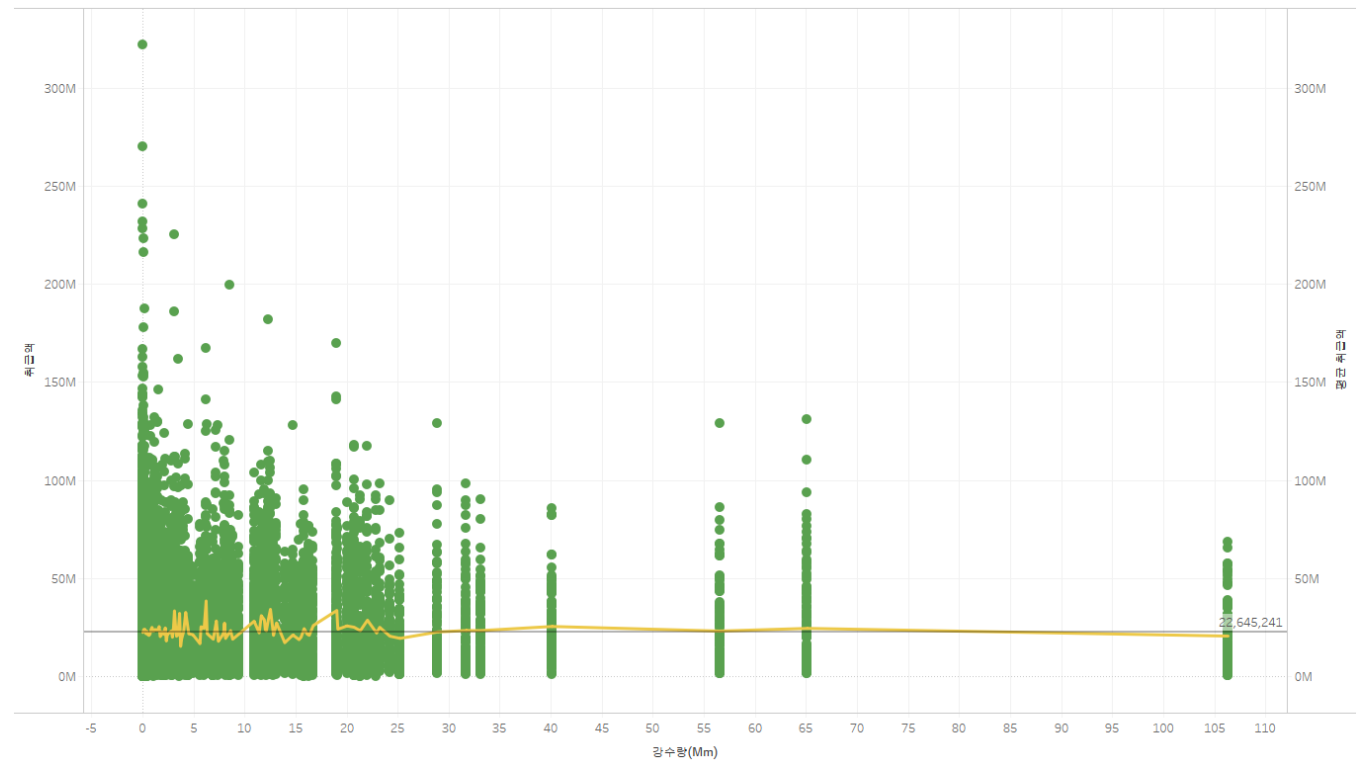
### 변수3 강수량

강수량에 따라 홈쇼핑 방송 시청 인구 및 취급액에 유의미한 변동이 있을 것이라 예측하여 외부 데이터를 이용한 변수를 추가하였다.

기상청의 기상자료개방포털의 데이터를 사용했다.  
(<https://data.kma.go.kr/stcs/grnd/grndRnList.do?pgmNo=69>)

전국 평균은 기후통계분석에 활용되는 45개 지점의 평균으로 산출된다.

"강수량(mm)": 일 별 전국 강수량 평균치에 대한 변수.



강수량과 총 취급액(초록색), 평균 취급액(노란색 선은 비가 올 때, 회색 선은 비가 오지 않을 때)의 그래프

## 2. 데이터 분석 및 변수 생성 ...

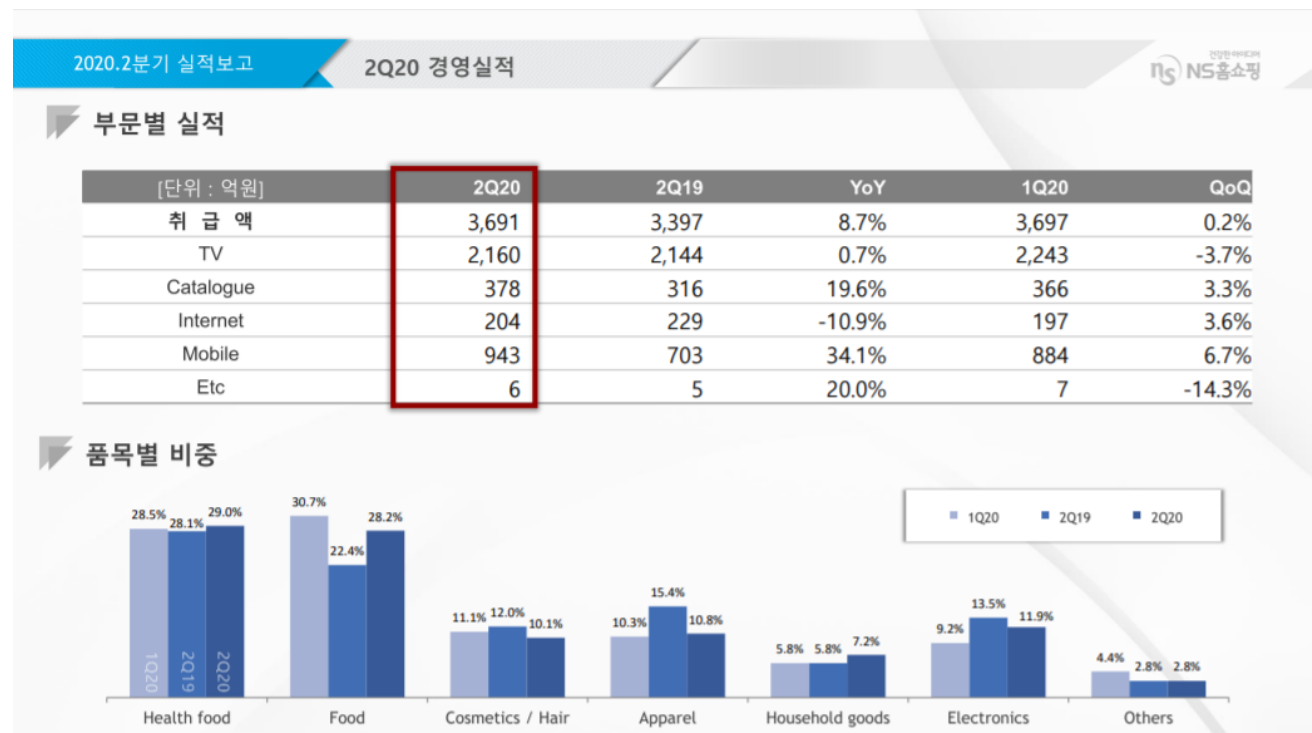
### 변수4 코로나로 인한 변화 반영 필요

#### 식품 매출 따라 희비교차한 홈쇼핑업계

NS홈쇼핑은 2020년 1분기 매출이 1406억원으로 2019년 1분기 대비 19.5% 신장했다. 2분기도 성장이 이어졌다는 분석이 지배적이다. NS홈쇼핑에 따르면 특히 20대의 신선식품 구매율이 눈에 띄게 증가했다. 코로나19 이슈로 소비자들 집에 머무는 시간이 늘며 건강에 좋은 신선식품을 더 많이 찾은 덕을 봤다. 이를 방증하듯 NS홈쇼핑의 대표 신선식품인 '완도활전복'의 상반기 주문건수는 이미 2019년 전체 주문량의 84% 이상을 달성했다.

2020.08.03 10:01 더바이어

<https://www.withbuyer.com/news/articleView.html?idxno=20526>



출처: <http://pr.nsmall.com/invest/ir.do>

2020년 6월 총 취급액 예측치

$$= 2020년\ 2분기\ 실적발표자료의\ 취급액 \times \frac{\text{제공데이터 2019년 2분기 취급액}}{\text{엔에스 쇼핑 2019년 2분기 실적발표자료의 취급액}} \times \frac{\text{제공데이터 2019년 6월 취급액}}{\text{제공데이터 2019년 2분기 취급액}}$$

## 2. 데이터 분석 및 변수 생성 ...

### 분석2 상품군, 마더코드, 상품코드

- 총 11개 상품군: 가구, 가전, 건강기능, 농수축, 생활용품, 속옷, 의류, 이미용, 잡화, 주방, 침구
  - train set 마더코드 688개
  - test set 마더코드 201개 (test set과 같은 것 85개, test set에 없는 것 116개)

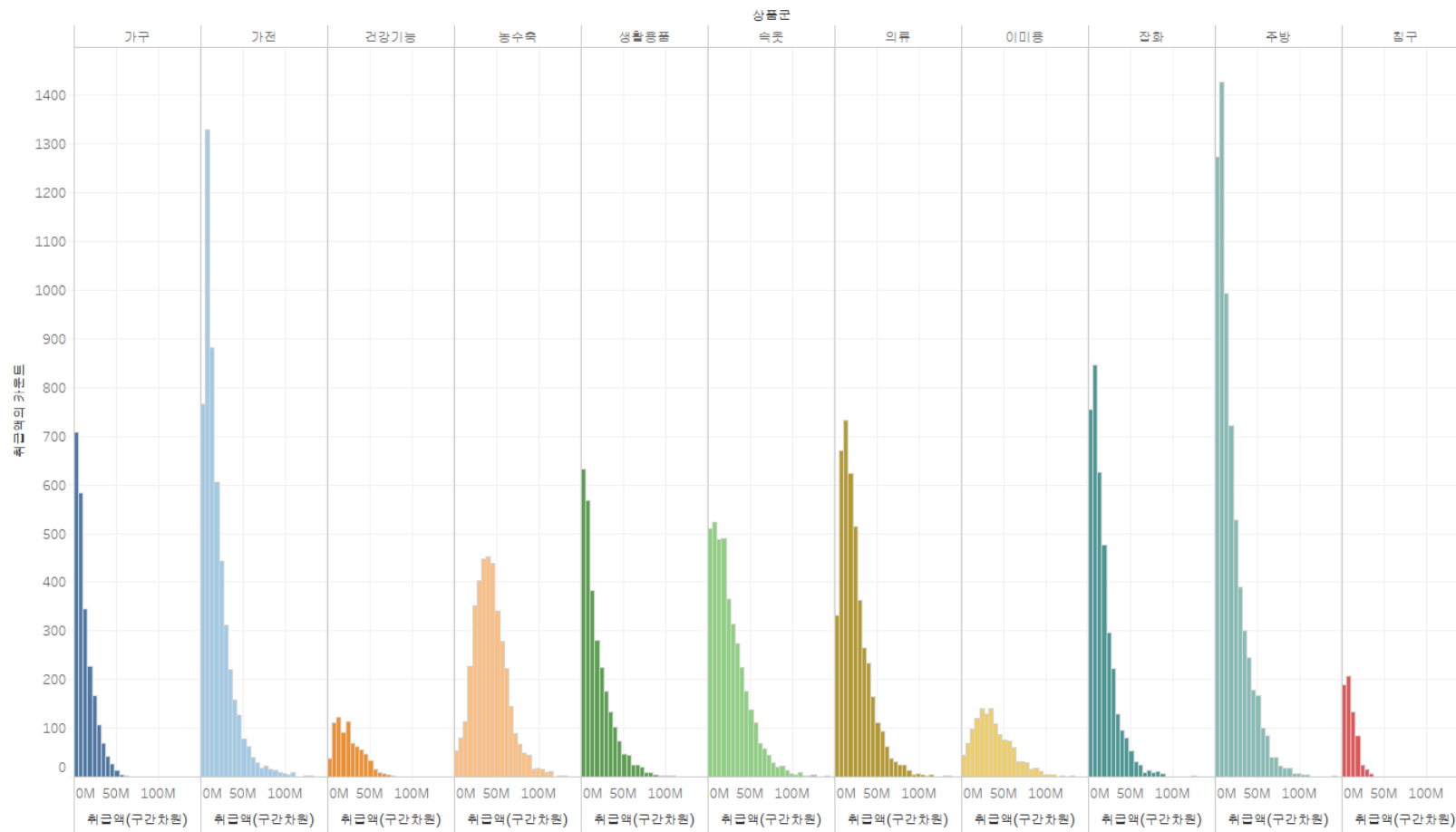


## 2. 데이터 분석 및 변수 생성 ...

### 변수5 상품군별 판매도 반영 필요 -> 상품군 power

각 상품군의 방송편성 빈도와 판매개수 정도가 다르기 때문에 이를 구별해줄 필요가 있다. 아래와 같은 공식으로 상품군별 분당 판매개수를 구해주고, 이를 상품군의 소비 정도로 봐 '상품군 power' 변수로 만들어준다.

$$\sum \frac{\text{취급액/판매단가}}{\text{평균방송분}}$$

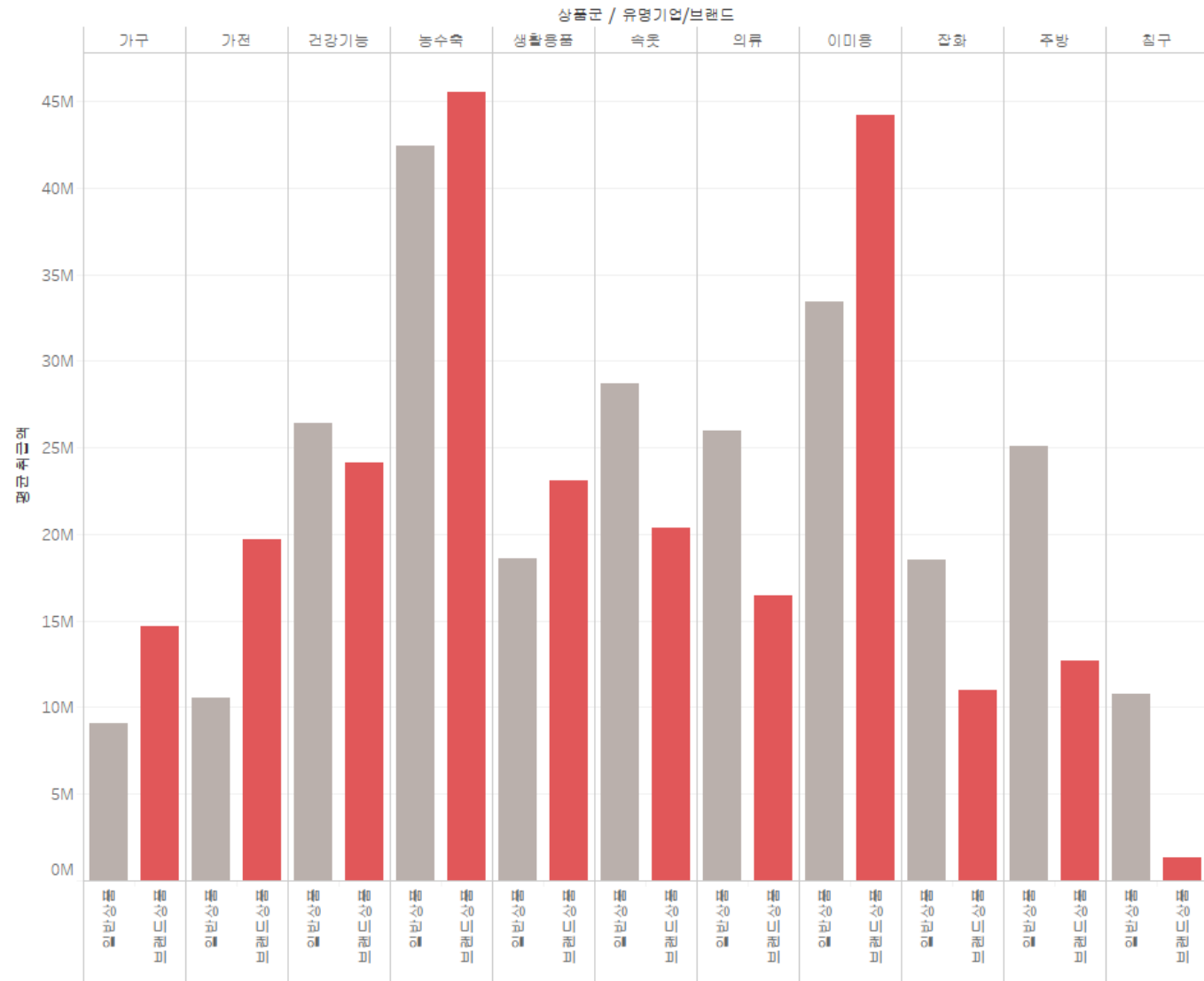


## 2. 데이터 분석 및 변수 생성 ...

### 변수6 마더코드에서 브랜드 추출

같은 브랜드의 비슷한 상품들은 하나의 마더코드로 묶인다.

그러나 각 브랜드를 변수로 만들기에는 하나의 마더코드 내의 상품의 개수가 전체 상품에 비해 작으며, 따라서 모델에 잘 반영되지 않을 것이라는 문제가 있다. 또한 train set과 test set의 마더코드가 일치하지 않으므로, 예측에 사용하기 곤란하다. 따라서 유명 브랜드/기업의 상품만을 추출해 이들을 표시하는 binary 데이터로 만들어주었다.



## 2. 데이터 분석 및 변수 생성 ...

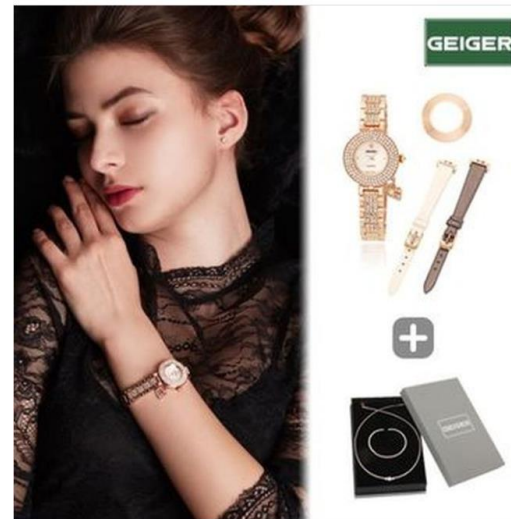
### 변수7 상품코드에서 성별 추출

상품코드가 상이한 이유: 상품 자체가 다름, 성별의 상이함, 결제 방법의 차이

- 의류, 속옷, 잡화, 이미용 상품군에 속한 상품 중 남성/여성 상품 분류
- 상품명 내에 "남성", "여성"이 명시되어 있는 경우 각각 "남성상품", "여성상품"으로 분류하였다.
- 성별이 직접적으로 명시되어 있지 않은 상품도 있었다. 시계, 네일, 파운데이션 등의 경우 직접 상품명을 구글에 검색 후 광고 모델이 특정 성별일 경우 해당 성별 상품으로 분류하였다.

### 변수8 상품코드에서 결제 방법 추출

일시불, 무이자 등의 결제 방법에 따라 상품코드가 달라지므로 이를 각각 binary data로 만들어 3개의 변수 생성(일시불 열, 무이자 열, 없음 열)



NS Shop+ | 2019년 03월 22일 02시 00분 방송종료

[NS Shop+](일시불)가이거 제니스시계 주얼리세트

~~128,000원~~  
128,000원

쇼핑사 NS홈쇼핑 샵플러스  
방송시간 2019년 03월 22일 01:03 ~ 02:00

바로구매

## 2. 데이터 분석 및 변수 생성 ...

### 분석3 방송 노출(분)

- 14976개의 노출(분) 결측치
- 같은 상품의 방송이 연속적으로 편성된 경우 존재
- 한 방송에 여러 개의 상품 동시에 판매하는 경우 존재

## 2. 데이터 분석 및 변수 생성 . . .

### 변수9 결측치 처리

노출(분) 결측치는 같은 방송 시간에 여러 개의 상품이 동시에 방송되는 경우 발생 -> 방송 시간이 같을 때 최초 노출(분)으로 이후의 공백 처리

방송일시	노출(분)	상품명
2019-01-01 6:00	20	테이트 남성 쉐런니트3종
2019-01-01 6:00		테이트 여성 쉐런니트3종

### 변수10 평균방송분

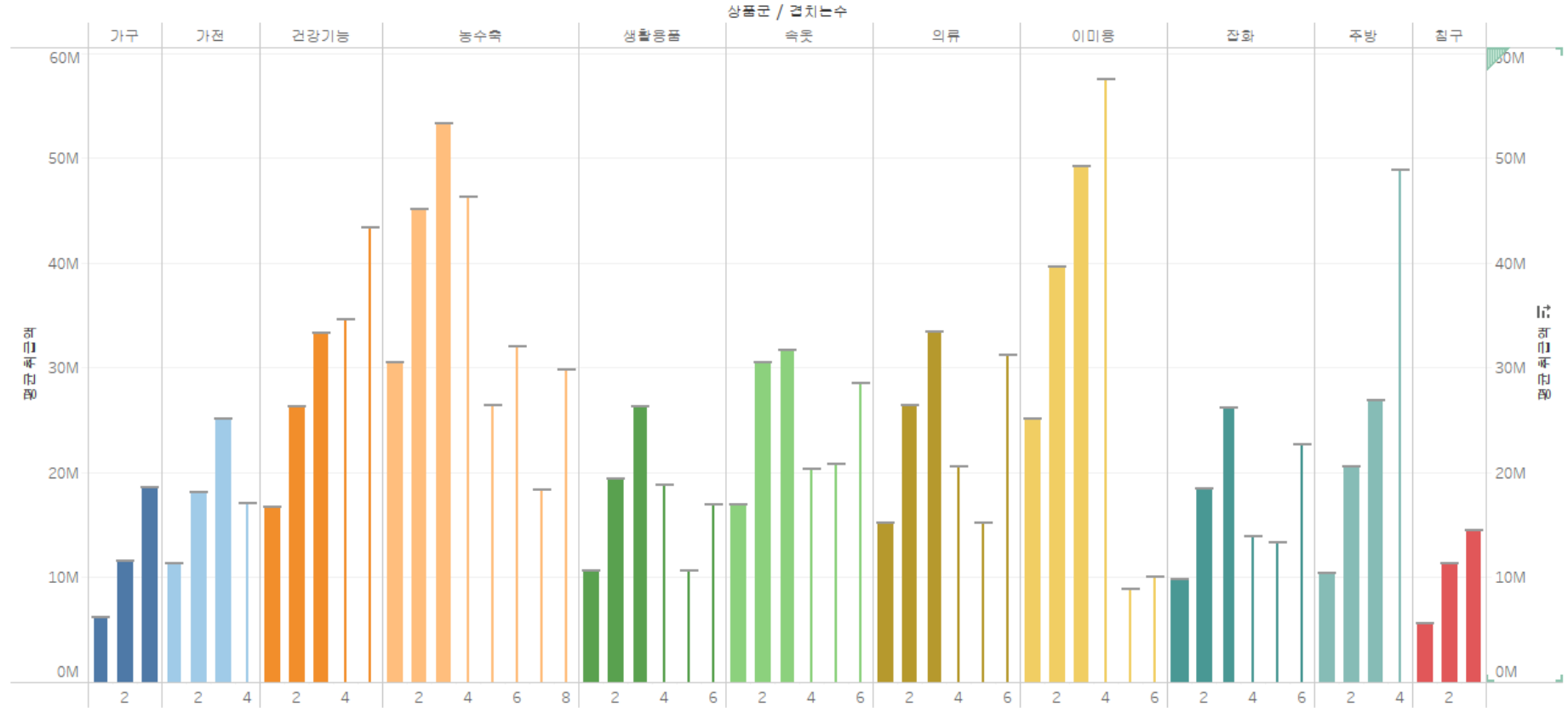
동시에 방영되는 상품들의 경우(같은 마더코드) 각 상품에 대해 평균적인 노출(분)을 추가.

방송일시	노출(분)	상품명	
2019-01-01 6:40	20	테이트 남성 쉐런니트3종	→10
2019-01-01 6:40	20	테이트 여성 쉐런니트3종	→10
2019-01-01 7:00	20	오모떼 레이스 파운데이션 브라	→20
2019-01-01 7:20	20	오모떼 레이스 파운데이션 브라	→20

## 2. 데이터 분석 및 변수 생성 ...

### 변수11 마더코드 겹치는 수

- 동일 상품을 연속 방송하는 경우 이후 편성 시간으로 갈수록 취급액이 유의미하게 높음을 상품군별 시각화를 통해 확인하였다.
- 따라서 마더코드를 기준으로 연속 방송의 횟수를 헤아려서 새로운 변수로 추가



## 2. 데이터 분석 및 변수 생성 ...

### 분석4 시청률

- train set에 대해 2019년 1분 단위로 시청률 데이터 제공
- test set에 대해 2020년 시청률 1분 단위로 존재하지 않음

## 2. 데이터 분석 및 변수 생성 ...

### 변수12 2019년 시청률 가공

해당 방송 시청률은 취급액에 직접적인 영향을 미칠 것이라 생각하여, 제공된 시청률 데이터를 이용한 홈쇼핑 방송의 시청률 데이터 변수를 추가하였다.

“시청률 평균”: 우선, 주말(금요일 포함)과 공휴일의 각 시간(분)에 대한 평균을 생성하였다. 그리고 평일(금요일 제외)의 경우도 각 시간(분)에 대한 시청률 평균을 계산하였다.

이후 위의 시청률 평균을 활용하여 방송 별 노출(분)에 따른 시청률 평균값을 구하여 변수로 추가하였다. 방송지속시간 동안의 계산된 시청률 평균을 모두 더한 후, 방송 지속시간으로 평균을 낸 것이다.

NS SHOP+ 일자별, 시간대별 시청률 (2019년)				
시간대	2019-01-01	2019-01-02	2019-01-03	2019-01-04
02:00	0	0	0	0
02:01	0	0	0	0
02:02	0	0	0	0
02:03	0	0	0.014	0
02:04	0	0	0	0
02:05	0	0	0	0
02:06	0	0	0	0
02:07	0	0	0	0
02:08	0	0	0	0
02:09	0.014	0	0	0

```
weekday.index = weekday['시각']
weekday
```

시간	시청률 평균	시각
600	0.001407	600
601	0.001648	601
602	0.001492	602
603	0.001658	603
604	0.001849	604

방송일시	시청률평균	요일/시간
2019-01-01 06:00:00	0.002631956	화06:00
2019-01-01 06:00:00	0.002631956	화06:00
2019-01-01 06:20:00	0.002434022	화06:20
2019-01-01 06:20:00	0.002434022	화06:20
2019-01-01 06:40:00	0.003161747	화06:40
2019-01-01 06:40:00	0.003161747	화06:40
2019-01-01 07:00:00	0.00297074	화07:00
2019-01-01 07:20:00	0.00285829	화07:20
2019-01-01 07:40:00	0.003773494	화07:40

Ex) 비버리힐스폴로클럽 남성기초세트

방송일시 2019-01-01 17:00, 노출 20분의 시청률

= (2019년 휴일 17:00의 시청률 평균 + 2019년 휴일 17:01의 시청률 평균 + ... + 2019년 휴일 17:18의 시청률 평균 + 2019년 휴일 17:19의 시청률 평균)의 평균



## 2. 데이터 분석 및 변수 생성 . . .

### 변수13 타 채널 시청자 수

재핑 효과(zapping effect)와 같이 다른 방송 시청 중 유입되는 시청 인구가 매출에 미치는 영향을 분석하기 위해 해당 방송일의 타 채널 시청자 수에 대한 변수를 추가하였다.

'닐슨코리아' 홈페이지의 시청자 수 데이터를 사용했다.  
([http://www.nielsenkorea.co.kr/tv\\_terrestrial\\_day.asp?menu=Tit\\_1&sub\\_menu=1\\_1&area=00](http://www.nielsenkorea.co.kr/tv_terrestrial_day.asp?menu=Tit_1&sub_menu=1_1&area=00))'

"타 채널 시청자 수 평균": 각 방송일시에 해당하는 날의 전국 시청자 수 상위 20위 방송의 시청자 수의 평균치에 대한 변수.

TV Ranking				지상파 일일 - TOP 20 LIST FOR TV PROGRAMS			
<ul style="list-style-type: none"> <li>지상파                             <ul style="list-style-type: none"> <li>일일순위</li> <li>주간순위</li> </ul> </li> <li>종합편성</li> <li>케이블</li> </ul>				Data Search 2019년 01월 01일			
				전국 수도권			
				2019.01.01			
가가시청률 TOP 20 (분석기준: 전국13개지역, 가구, 단위:%)				시청자수 TOP 20 (분석기준: 전국13개지역, 개인, 단위:천 명)			
순위	채널	프로그램	시청률	순위	채널	프로그램	시청자수
1	KBS1	일일연속극(비켜라운명야)	15.6	1	KBS1	일일연속극(비켜라운명야)	2,446
2	KBS1	KBS9시뉴스	11.8	2	KBS2	신년특선영화(관상1부)	2,167
2	KBS2	신년특선영화(관상1부)	11.8	3	KBS2	신년특선영화(관상2부)	1,976
4	KBS1	KBS뉴스7	10.8	4	KBS1	KBS9시뉴스	1,852
5	KBS2	신년특선영화(관상2부)	10.3	5	KBS1	KBS뉴스7	1,712
6	KBS1	KBS뉴스(09:30)	9.9	6	SBS	본격연예한밤	1,630

## 2. 데이터 분석 및 변수 생성 ...

### 분석5 시간

- 일 단위: 소비자들이 방송을 시청하는 시간대는 평일과 주말(요일), 공휴일에 따라 달라질 것이다
  - 시간 단위: 시간대에 따라 특정 상품군에 대한 구매 의욕 등이 달라질 것이다.

## 2. 데이터 분석 및 변수 생성 . . .

### 변수14 시간변수 세분화

방송 시간대에 따라 생기는 홈쇼핑 상품 매출의 차이를 반영하기 위해 방송시간을 2시간으로 세분화한 시간 변수를 추가하였다.

'06:00~08:00', '08:00~10:00', '10:00~12:00', '12:00~14:00', '14:00~16:00', '16:00~18:00', '18:00~20:00', '20:00~22:00',  
'22:00~00:00', '00:00~02:00', '02:00~04:00' (각 방송일시에 해당하는 시간대이면 1, 그렇지 않으면 0을 갖는 이진변수)

### 변수15 프라임 타임

상품이 잘 판매되는 시간대인 프라임 타임을 변수로 추가하였다.

과학기술정보통신부에서 발표한 프라임 타임은 오전 9시~11와 오후 8시~11시이다. 이에 더해 농수축 상품을 판매 비중이 높은 NS shop+만의 프라임 타임인 오후 4~6시를 추가해주었다. 이렇게 프라임타임을 세분화해서 오전/식품오후/그외오후/프라임아님 4가지 카테고리로 구분하였다. 출처: <https://www.etnews.com/20180718000049?m=1>

### 변수16 요일별 변수

요일 별로 생기는 홈쇼핑 방송 시청 인구 및 매출의 차이를 반영하기 위해 추가하였다.

"요일\_금", "요일\_목", "요일\_수", "요일\_월", "요일\_일", "요일\_토", "요일\_화"

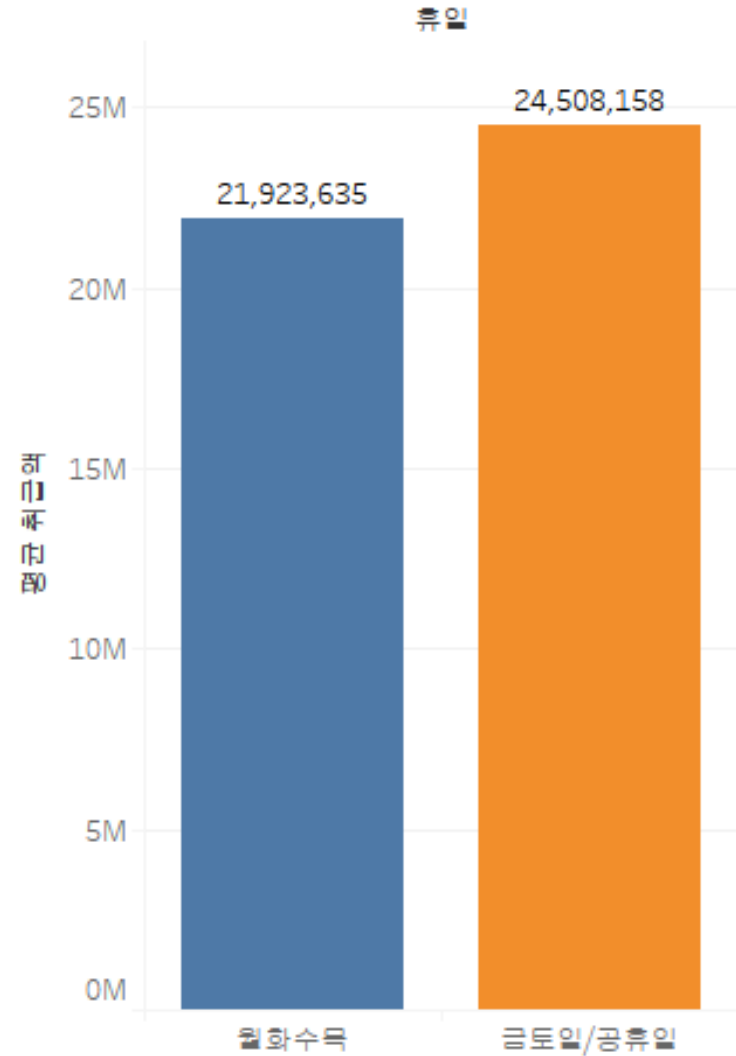
각 방송일시에 해당하는 요일이면 1, 그렇지 않으면 0을 갖는 이진변수

## 2. 데이터 분석 및 변수 생성 ...

### 변수17 금요일·토요일 및 공휴일

휴일: 상품 판매 요일 중 금요일, 토요일 및 공휴일. 휴일이면 1, 휴일이 아니면 0을 갖는 이진변수.

- 직장인이 집에서 보내는 시간이 많은 요일을 금요일, 토요일 및 공휴일로 선정하여 "휴일" 변수로 정의하였다.
- 오른쪽 그래프를 보면 휴일로 처리된 상품의 평균 취급액이 휴일로 처리되지 않은 상품의 평균 취급액 보다 높은 것을 볼 수 있다.



## 2. 데이터 분석 및 변수 생성 ...

### 분석6 상품명

- 상품명 자체에 연예인 이름이 포함되는 등의 경우가 존재한다.

## 2. 데이터 분석 및 변수 생성 ...

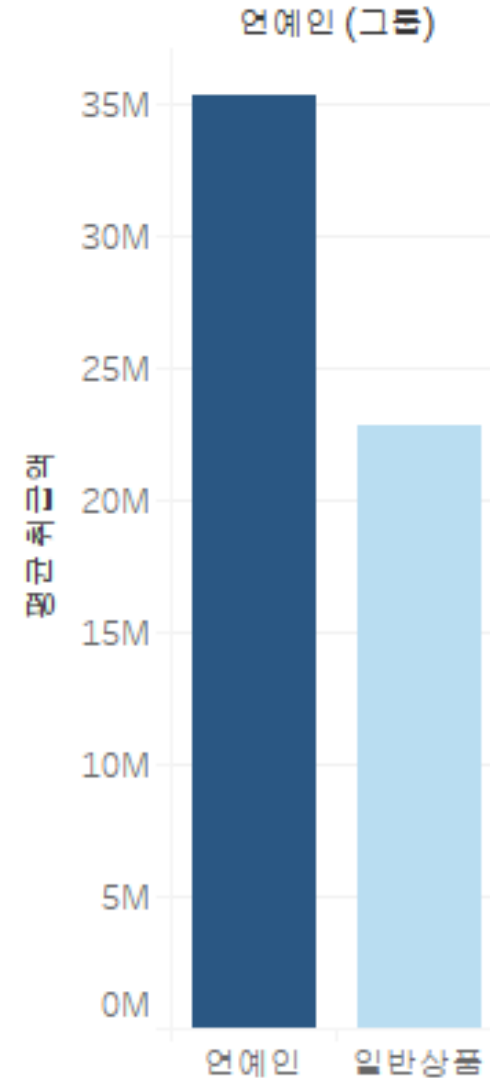
### 변수18 연예인

상품명에 연예인 이름 포함 여부. 포함하면 1, 포함하지 않으면 0을 갖는 이진변수.

- 연예인의 평판으로부터 발생하는 광고효과를 잡아내기 위해 추가하였다. 연예인으로 전지현, 팽현숙, 김병만, 서장훈 등이 있었다.
- 우측 그래프를 보면 연예인 이름을 포함하는 상품이 연예인 이름을 포함하지 않은 상품보다 취급액이 10천만원 이상 차이 나는 것을 볼 수 있다.

### 변수19 손질

농수축 소분류의 어류 상품 중, '손질'이라는 키워드가 들어간 상품명과 그렇지 않은 상품명 이 존재했다. 손질이 된 어류가 더 잘 판매될 것이라 예상해 이를 구분해주는 변수를 만들어 주었다.



## 2. 데이터 분석 및 변수 생성 ...

### 분석7 상품군 내 소분류

- 상품군 내에도 다양한 소분류가 존재한다.
- 그 중에서도 NS홈쇼핑의 주력상품인 농수축 상품군과,

6월에 판매량 증가가 예상되는 선풍기, 에어컨 등의 제품이 포함된 가전제품을 세분화하였다.

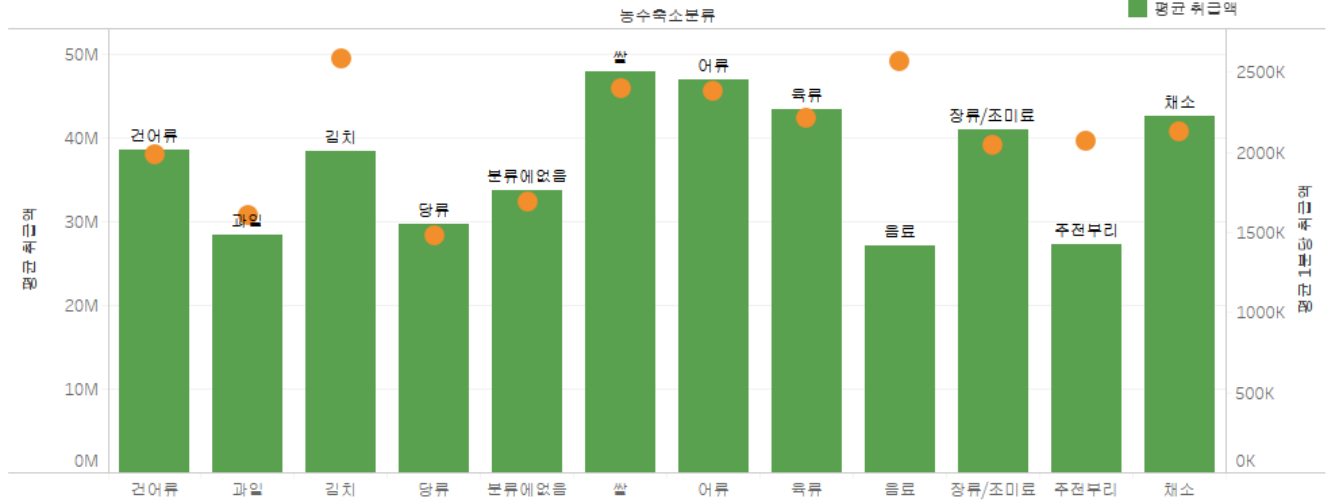
## 2. 데이터 분석 및 변수 생성

### 변수20 농수축 소분류

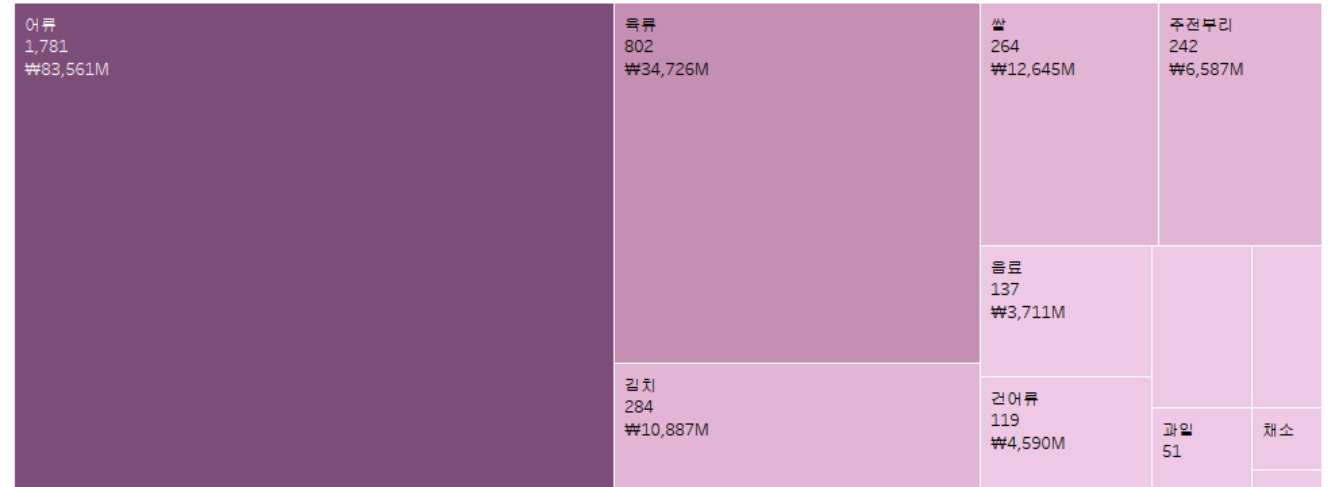
농수축 상품이 많은 빈도를 차지하고 있으므로, 농수축 상품군 내에서도 소분류가 필요하다고 생각되었다.

따라서 건어류, 과일, 김치, 당류, 쌀, 어류, 육류, 음료, 장류/조미료, 주전부리, 채소로 분류했다. 이후 각 소분류의 개수가 너무 작은 카테고리인 당류, 채소는 제거했다.

평균 취급액과 분당 취급액



방송 횟수와 총 취급액



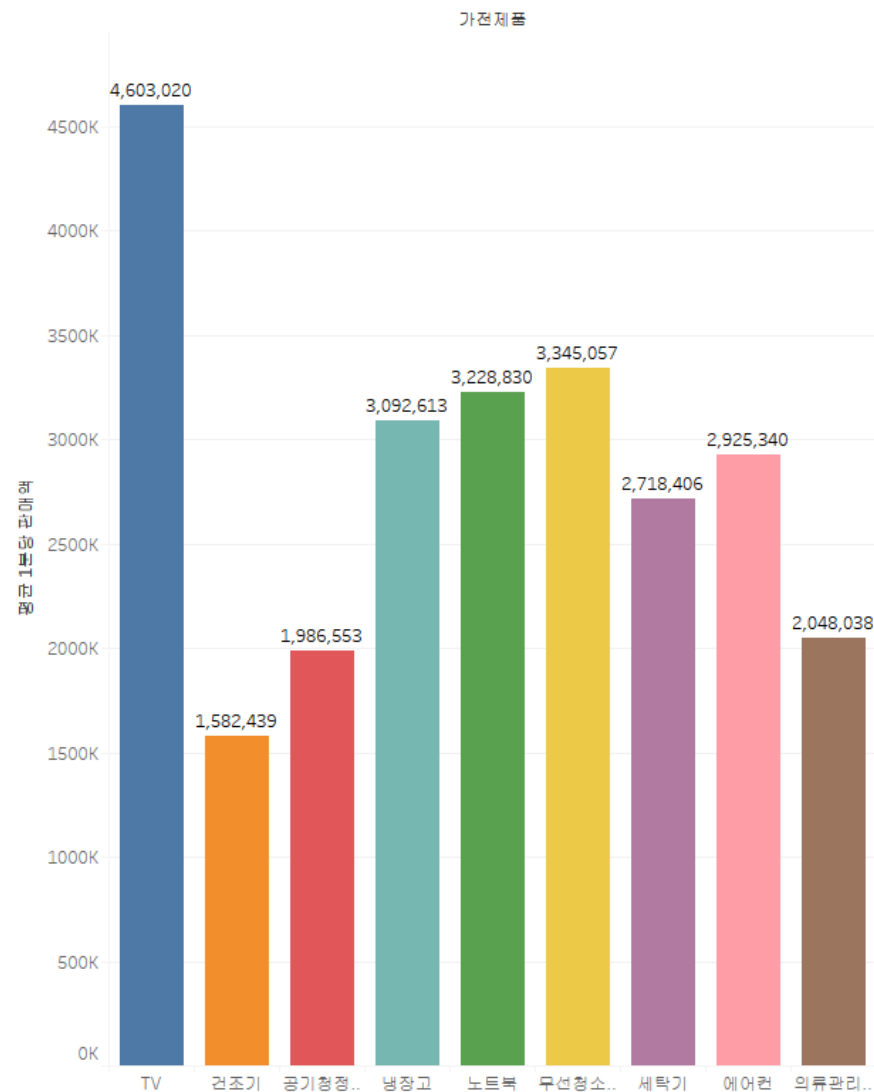


## 2. 데이터 분석 및 변수 생성 ...

### 변수21 가전제품 소분류

가전제품 소분류: 여러 가전제품 종류 간에 취급액이 상이함에 따라 “가전제품 소분류”인 범주형 변수를 생성하였다. 가전제품 종류를 총 9가지로 분류하였으며 prediction 단계에서는 각 종류마다 dummy화 하여 이진변수를 생성하였다.

- Ex1) 가전제품\_TV: 가전제품 중 TV 상품. TV 상품이면 1, TV 상품이 아니면 0을 갖는 이진변수.
- Ex2) 가전제품\_건조기: 가전제품 중 건조기 상품. 건조기 상품이면 1, 건조기 상품이 아니면 0을 갖는 이진변수.



## 2. 데이터 분석 및 변수 생성 ...

### 분석8 경제상황에 따른 소비 변화

- 경제 상황에 따른 소비행태 변화가 존재한다.

## 2. 데이터 분석 및 변수 생성 ...

### 변수22 월급수령일

국민은행 공식 블로그를 통해 월급날을 10일, 20일, 25일로 지정했다. 월급 수령으로 인한 홈쇼핑 상품 소비 증가 수령 당일 이후 약 3일동안 지속될 것을 고려하였다. 월급날: 상품 판매 날짜 중 월급날 (10~13일, 20~23일, 25~28일). 월급 날이면 1, 월급날이 아니면 0을 갖는 이진변수

출처: <https://m.blog.naver.com/youngkbblog/221714986107>

### 변수23 소비자물가지수

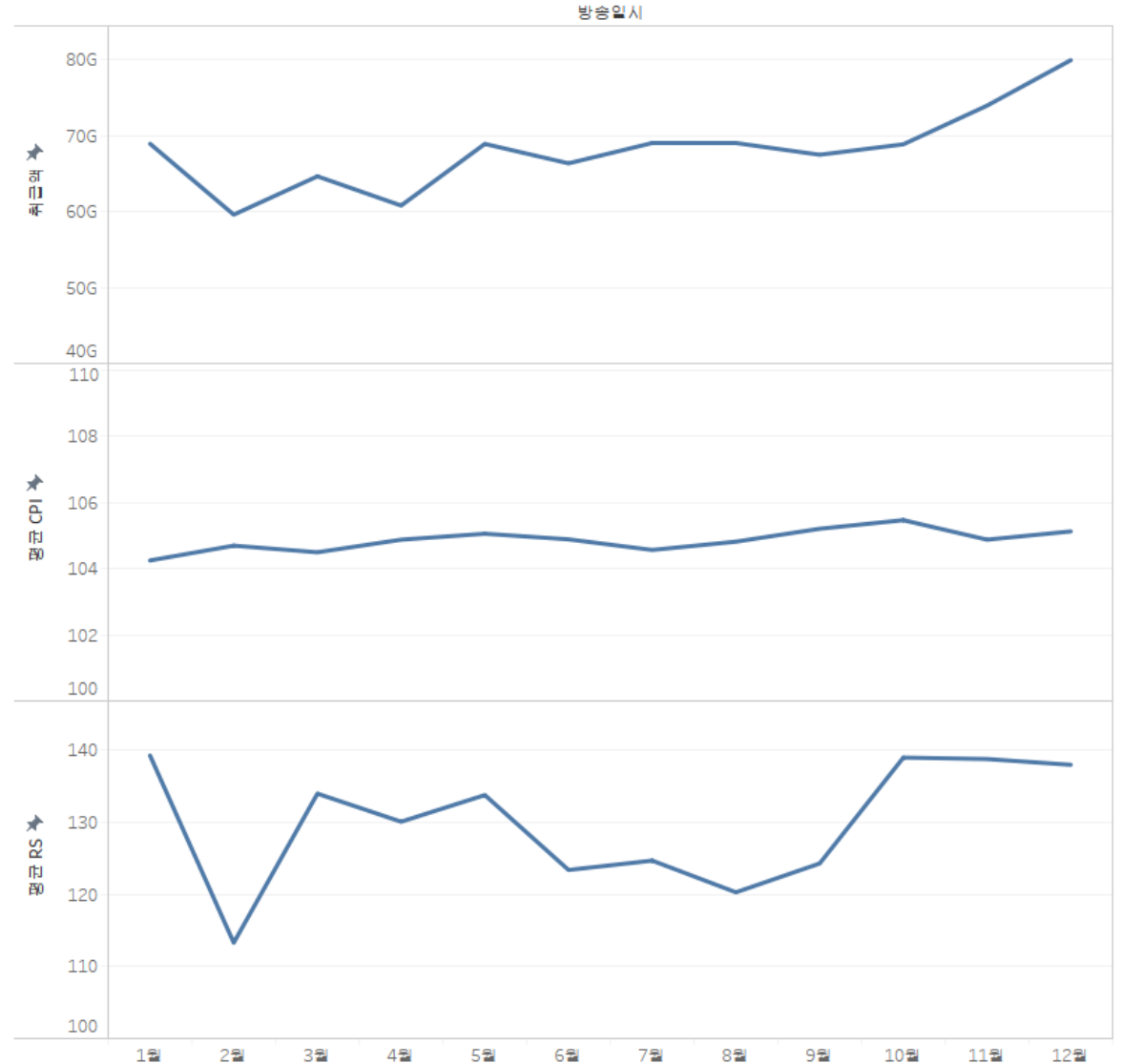
### 변수24 소매업태별(홈쇼핑) 판매액지수(불변지수)

각 가정이 생활을 위해 구입하는 상품과 서비스의 가격 변동을 반영하기 위해 통계청의 월별 소비자물가지수를 참고해 해당 수치를 추가해주었다.

[http://kosis.kr/statisticsList/statisticsListIndex.do?menuId=M\\_01\\_01&wcd=MT\\_ZTITLE&parmTabId=M\\_01\\_01&parentId=P.1;P2.2;P2\\_6.3;#SelectStatsBoxDiv](http://kosis.kr/statisticsList/statisticsListIndex.do?menuId=M_01_01&wcd=MT_ZTITLE&parmTabId=M_01_01&parentId=P.1;P2.2;P2_6.3;#SelectStatsBoxDiv)

통계청이 유통 및 소비동향을 파악하기 위해 작성하는 지표인 소매업태별 판매액지수 중 홈쇼핑 항목을 추가해주었다.

[http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\\_1K31013&vw\\_cd=MT\\_ZTITLE&list\\_id=JG2&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE](http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1K31013&vw_cd=MT_ZTITLE&list_id=JG2&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)

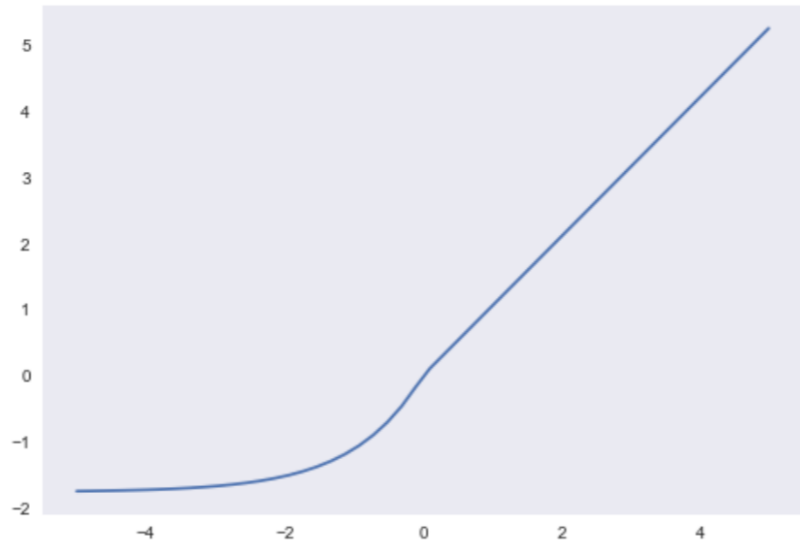


▲ 소비자 물가지수(CPI)와 소매업태별 판매액 불변지수(RS) 중 홈쇼핑의 지수, 취급액의 시간에 따른 변화

### 3. 데이터 전처리 ...

#### standard scaling

모델링에 아래와 같은 SELU activation function을 사용



출처: <https://medium.com/@damoncivin/self-normalising-neural-networks-snn-2a972c1d421>

- 이를 사용하기 위해서는 input 이 평균 0 , 분산 1로 표준화되어야 하기 때문에 모든 feature 에 대해 standard scaling을 적용하였다.
- 모든 변수에 대해서 scaling 하지 않은 경우(즉, 수치형 데이터에만 scaling을 적용하고 범주형에는 적용하지 않은 경우)는 DNN의 수렴속도가 현저히 느렸다.

#### skewness 조정

- 데이터의 빈도가 특정 방향으로 치우쳐 있다기 보다, 아예 몇몇 개의 특정 값에 국한된 변수가 더 많았기 때문에, skewness 조정을 위해서 transform을 하여도 그 효과가 크지 않을 것이라 예상하여 따로 조정하지 않았다.

## 4. 모델링 ...

train set, test set, validation set의 분할

mape값으로 성능 판단

## 4. 모델링 . . .

### 모델 선택 기준 validation set 이용

- data set을 train / valid / test 로 분할
- train set 으로 모델을 훈련 → valid set 으로 모델의 적합성 등 평가 → test set 의 결과로 최종 모델 선택

```
# dataset train/test set 으로 나누기  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=0)
```

```
# dataset 의 train 을 valid / train 으로 다시 나누기  
X_train_, X_valid_, y_train_, y_valid_ = train_test_split(X_train, y_train, test_size=0.20, random_state=0)
```

### 과정

- 기존의 머신러닝 (XGBoost, lightGBM, Random Forest ... ) 방법을 적용한 결과 test set 에 대한 mape 가 44~50 정도의 값을 보임
- relu 층을 5~6개 쌓은 간단한 딥러닝 모형의 mape 가 42 정도의 성능을 보임

→ 딥러닝 모델 선택

- ✓ 다양한 딥러닝 모델 탐색 후, 가장 성능이 좋았던 3개의 모델을 앙상블 하여 예측하기로 하였다.
- ✓ 앙상블을 할 때에는 최대한 서로 다른 모델을 앙상블하는 게 효과가 좋으므로, 서로 구조가 다른 3개의 모델을 선택하였다.
- ✓ 마지막 예측을 할 때에는 위에서 정한 모델로 전체 데이터를 train / valid set 로만 나눈 뒤 train으로 훈련시키고, 예측하였다.

## 4. 모델링 ●●●

```
model1 = Sequential([
    Dense(128, kernel_initializer='normal', activation = "relu", input_shape=X_train.shape[1:]),
    Dropout(0.2),
    Dense(256, kernel_initializer='normal', activation = "relu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='normal', activation = "relu"),
    Dropout(0.2),
    Dense(1024, kernel_initializer='normal', activation = "relu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='normal', activation = "relu"),
    Dropout(0.2),
    Dense(256, kernel_initializer='normal', activation = "relu"),
    Dropout(0.2),
    Dense(32, kernel_initializer='normal', activation = "relu"),
    Dense(1, kernel_initializer='normal'), ])
```

```
Epoch 111/160
885/885 [=====] - 13s 15ms/step - loss: 35.0464 - val_loss: 39.9530
Epoch 112/160
885/885 [=====] - 14s 16ms/step - loss: 34.8940 - val_loss: 39.8072
Epoch 113/160
885/885 [=====] - 14s 16ms/step - loss: 34.9452 - val_loss: 39.8525
222/222 [=====] - 1s 3ms/step - loss: 39.7790
```

- 7층의 hidden layer 로 이루어진 모형을 사용
- activation 은 relu , initializer 는 normal 을 사용
- 중간 dropout 층을 0.2 로 사용하여 overfitting을 방지
- 결과: 총 훈련은 113 epoch 까지 훈련하였고 , test set 에 대한 mape는 39.7790 이 나왔다.

## 4. 모델링 . . .

```

model2 = keras.models.Sequential()
model2.add(keras.layers.Dense(200, activation="selu", kernel_initializer="lecun_normal", input_shape=X_train.shape[1:]))
model2.add(Dropout(0.2))
for layer in range(5):
    model2.add(keras.layers.Dense(200, activation="selu", kernel_initializer="lecun_normal"))
    model2.add(Dropout(0.2))
model2.add(keras.layers.Dense(100, activation="selu", kernel_initializer="lecun_normal"))
model2.add(Dropout(0.2))
model2.add(keras.layers.Dense(10))
model2.add(keras.layers.Dense(1))

```

```

Epoch 146/160
885/885 [=====] - 5s 5ms/step - loss: 36.2665 - val_loss: 39.8117
Epoch 147/160
885/885 [=====] - 5s 5ms/step - loss: 36.4544 - val_loss: 39.8935
Epoch 148/160
885/885 [=====] - 5s 5ms/step - loss: 36.3156 - val_loss: 40.1366
222/222 [=====] - 0s 2ms/step - loss: 39.4675

```

- 7층의 hidden layer 로 이루어진 모형을 사용
- activation 은 selu, initializer 는 lecun\_normal 을 사용
- 중간 dropout 층을 0.2 로 사용하여 overfitting 방지
- 결과: 총 훈련은 148 epoch 까지 훈련하였고, test set 에 대한 mape 는 39.4675 이 나왔다.



## 4. 모델링 . . .

```
model3 = Sequential([
    Dense(128, kernel_initializer='he_normal', activation = "elu", input_shape=X_train.shape[1:]),
    Dropout(0.2),
    Dense(256, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(512, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(256, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(32, kernel_initializer='he_normal', activation = "elu"),
    Dropout(0.2),
    Dense(1, kernel_initializer='he_normal'),
])
```

```
Epoch 85/160
885/885 [=====] - 14s 16ms/step - loss: 36.9328 - val_loss: 39.7265
Epoch 86/160
885/885 [=====] - 14s 16ms/step - loss: 36.9559 - val_loss: 39.7013
Epoch 87/160
885/885 [=====] - 14s 16ms/step - loss: 36.7640 - val_loss: 39.6186
222/222 [=====] - 1s 3ms/step - loss: 39.3965
```

- 9층의 hidden layer 로 이루어진 모형을 사용
- activation 은 elu, initializer 는 he\_normal 을 사용
- 중간 dropout 층을 0.2 로 사용하여 overfitting 방지
- 결과: 총 훈련은 86 epoch 까지 훈련하였고 , test set 에 대한 mape는 39.3965 가 나왔다.

## 4. 모델링 . . .

```
y_pred = 0.3 * model1.predict(X_test) + 0.35 * model2.predict(X_test) + 0.35 * model3.predict(X_test)
MAPE(y_test, y_pred.reshape(-1))
```

38.68272049152054

	방송일시	노출(분)	마더코드	상품코드	상품명	상품군	판매단가	취급액
0	2020-06-01 06:20:00	20	100650	201971	잭필드 남의류		59800	11694454
1	2020-06-01 06:40:00	20	100650	201971	잭필드 남의류		59800	18411246
2	2020-06-01 07:00:00	20	100650	201971	잭필드 남의류		59800	22643726
3	2020-06-01 07:20:00	20	100445	202278	쿠미투니카속옷		69900	7672606
4	2020-06-01 07:40:00	20	100445	202278	쿠미투니카속옷		69900	14808185
5	2020-06-01 08:00:00	20	100445	202278	쿠미투니카속옷		69900	25948110
6	2020-06-01 08:20:00	20	100381	201247	바비리스 아이미용		59000	10279856
7	2020-06-01 08:40:00	20	100381	201247	바비리스 아이미용		59000	16954316
8	2020-06-01 09:00:00	20	100381	201247	바비리스 아이미용		59000	29109222
9	2020-06-01 09:20:00	20	100638	201956	램프쿡 자 주방		109000	15917805
10	2020-06-01 09:40:00	20	100638	201956	램프쿡 자 주방		109000	29477026
11	2020-06-01 10:00:00	20	100638	201956	램프쿡 자 주방		109000	42661628
12	2020-06-01 10:20:00	20	100348	201091	벨레즈온 속옷		59900	14850100
13	2020-06-01 10:40:00	20	100348	201091	벨레즈온 속옷		59900	30497824
14	2020-06-01 11:00:00	20	100348	201091	벨레즈온 속옷		59900	32085572
15	2020-06-01 11:20:00	20	100012	200016	AAC 삼채 농수축		40900	21835160
16	2020-06-01 11:40:00	20	100012	200016	AAC 삼채 농수축		40900	31157364
17	2020-06-01 12:00:00	20	100012	200016	AAC 삼채 농수축		40900	33698708
18	2020-06-01 12:20:00	20	100080	200217	아키 라이 속옷		99900	11053399
19	2020-06-01 12:40:00	20	100080	200217	아키 라이 속옷		99900	19821480
20	2020-06-01 13:00:00	20	100080	200217	아키 라이 속옷		99900	29413250
21	2020-06-01 13:20:00	20	100570	201673	KT휴대폰_무형		0	
22	2020-06-01 13:20:00		100570	201671	(특)KT휴대폰_무형		0	
23	2020-06-01 13:40:00	20	100570	201673	KT휴대폰_무형		0	

- 모델1, 모델2, 모델3을 앙상블 하였다.
- 가중치는 세 계수의 합이 1이 되도록 0.05 단위로 미세조정하면서, test set 에 대해 제일 낮은 mape 가 나오는 가중치로 결정
- 그 결과 test set 에 대하여 38.68 의 mape 가 나와 단일 모델보다 더 나은 성능을 보였다
- 결정된 3개의 앙상블 모델로 예측한 예측치를 형성하였고, 그 예측치를 엑셀로 저장

## 5. 편성 최적화 방안 제안 ...

LIME함수를 이용하여 각 상품의 취급액에 영향을 끼친 변수들을 살펴보고,

이를 바탕으로 시간별/요일별/카테고리별 편성방안을 제안하였다.

LIME 그래프에 녹색으로 표시된 변수는 취급액에 긍정적인 영향을, 빨간색으로 표시된 변수는 취급액에 부정적인 영향을 끼친다.

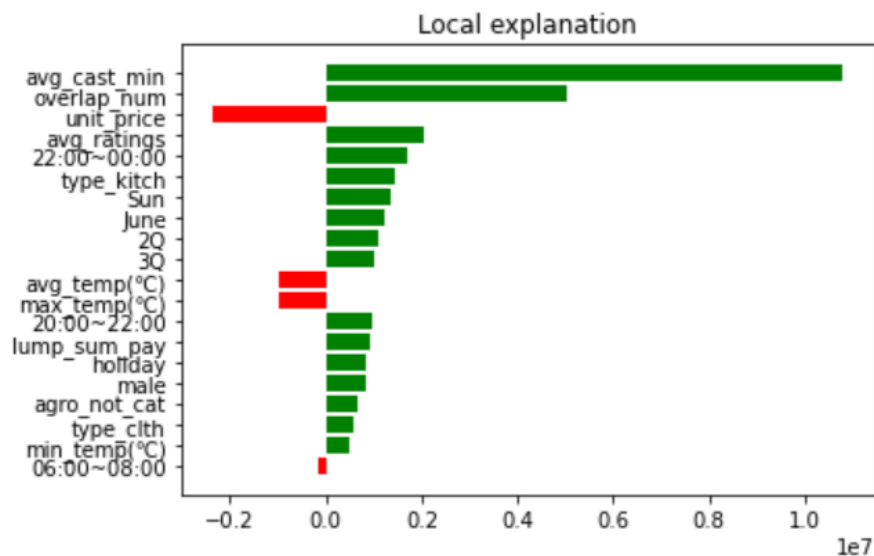
## 5. 편성 최적화 방안 제안 ●●●

시간별 최적 방안

## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 1. 평균 방송분

- 평균방송분은 한 방송에서 하나의 상품이 방송되는 시간을 뜻한다.
  - 모든 변수들 중 LIME 모델 가중치가 압도적으로 높았다.
- 동일 방송에 많은 상품을 소개하면, 방송 효과가 분산되어 매출 감소로 이어진다는 것을 알 수 있다. 따라서 **하나의 방송에 하나의 상품을 배정** 하는 것이 바람직하다.



◀ 2020-06-14 07:00:00에 방영된 '코몽트 남성 반팔 티셔츠 8종(시즌2)' 상품의 취급액에 영향을 끼치는 변수 상위 20개이다.  
해당 상품은 6:00~8:00 시간대의 상품 중 취급액 1위를 기록하였다.  
평균방송분(avg\_cas\_min)변수가 양의 영향을 끼치는 것을 확인할 수 있다.

	avg_cast_min
	평균방송분
상품군 가구 상위	
644	6.666666667
645	6.666666667
646	6.666666667
상품군 가전 상위	
493	20
579	20
1321	20
상품군 건강기능 상위	
2166	20
2138	10
2114	10
상품군 농수축 상위	
1306	20
2543	20
1933	20
상품군 생활용품 상위	
1285	20
1364	20
634	20
상품군 속옷 상위	
2192	20
1664	20
1165	20

	avg_cast_min
상품군_의류 상위	
1891	20
1156	20
1784	20
상품군_이미용 상위	
2189	20
1754	20
741	20
상품군_잡화 상위	
1813	20
248	20
1448	20
상품군_주방 상위	
1195	30
1787	20
2399	20
상품군_침구 상위	
809	2.5
808	2.5
807	2.5
요일 : 월	
2601	20
2598	20
2604	20

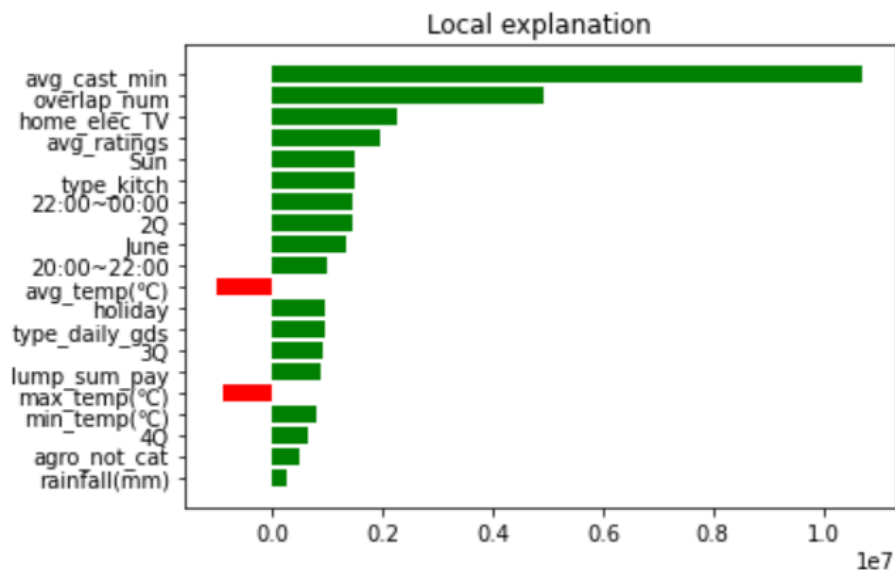
▲ 상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, '평균방송분'은 전부 칠해졌다.

## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 2. 겹치는 수

- 겹치는 수는 해당 상품이 연속으로 방송된 횟수를 뜻한다.
- 거의 모든 경우에서 LIME 분석 시 높은 가중치를 보였다.

→ 연속 방송의 영향이 유의미하므로, 기존의 단일 방송 상품도 **최대한 연속해서 방송**하는 방향으로 편성하는 것이 **매출 증가**에 효과적일 것이다.



◀ 2020-06-14 23:00:00에 방영된 '신일 이동식에어컨' 상품의 취급액에 영향을 끼치는 변수 상위 20개이다. 해당 상품은 생활용품 상품군의 상품 중 취급액 2위를 기록하였다. 겹치는 수(overlap\_num) 변수가 양의 영향을 끼치는 것을 확인할 수 있다.

	overlap_num
상품군 가구 상위	겹치는수
644	3
645	3
646	3
상품군 가전 상위	
493	3
579	3
1321	3
상품군 건강기능 상위	
2166	1
2138	2
2114	2
상품군 농수축 상위	
1306	3
2543	3
1933	3
상품군 생활용품 상위	
1285	3
1364	3
634	3
상품군 속옷 상위	
2192	3
1664	3
1165	3

	overlap_num
상품군 의류 상위	
1891	3
1156	3
1784	3
상품군 이미용 상위	
2189	3
1754	3
741	3
상품군 잡화 상위	
1813	3
248	3
1448	3
상품군 주방 상위	
1195	2
1787	3
2399	3
상품군 침구 상위	
809	6
808	6
807	6
요일 : 월	
2601	3
2598	3
2604	3

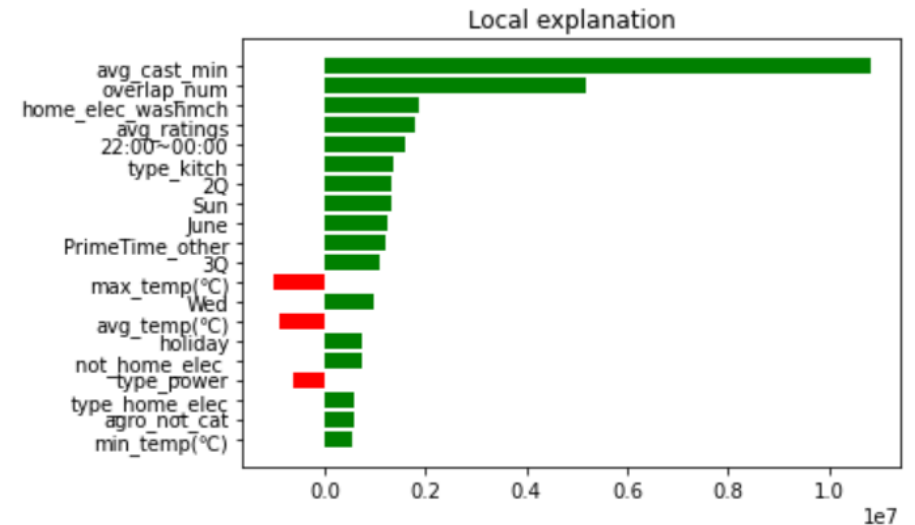
▲ 상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, '겹치는 수'은 대부분 칠해졌다.

## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 3. 시간대(2시간 단위)

- 시간대별로 TOP3 제품들을 LIME으로 분석한 결과, 6~8시 이른 오전과 00~04시 새벽 시간대에는 매출 성과에 유의미한 영향을 끼치지 못한다.
- 반면 20~00시는 매출 성과에 긍정적인 영향을 끼친 것으로 분석되었다.

	06:00~08:00	08:00~10:00	10:00~12:00	12:00~14:00	14:00~16:00	16:00~18:00	18:00~20:00	20:00~22:00	22:00~00:00	00:00~02:00	02:00~04:00
요일 : 월											
2601	0	0	0	0	0	0	1	0	0	0	0
2598	0	0	0	0	0	0	1	0	0	0	0
2604	0	0	0	0	0	0	0	1	0	0	0
요일 : 화											
1502	0	0	0	0	0	0	0	1	0	0	0
2685	0	0	0	0	0	0	0	1	0	0	0
103	0	0	0	0	0	0	1	0	0	0	0
요일 : 수											
223	0	0	0	0	0	0	0	1	0	0	0
929	0	0	0	0	0	0	0	0	1	0	0
2189	0	0	0	0	0	0	0	1	0	0	0
요일 : 목											
2288	0	0	0	0	0	0	0	0	1	0	0
973	0	0	1	0	0	0	0	0	0	0	0
2261	0	0	0	0	0	0	1	0	0	0	0
요일 : 금											
2366	0	0	0	0	0	0	0	1	0	0	0
1754	0	0	0	0	0	0	0	1	0	0	0
1751	0	0	0	0	0	0	1	0	0	0	0
요일 : 토											
493	0	0	0	0	0	0	0	1	0	0	0
1195	0	0	0	0	1	0	0	0	0	0	0
1787	0	1	0	0	0	0	0	0	0	0	0
요일 : 일											
1306	0	0	0	0	0	1	0	0	0	0	0
2543	0	0	0	0	0	0	0	1	0	0	0
579	0	0	1	0	0	0	0	0	0	0	0



▲ 2020-06-10 22:00:00에 방영된 'LG전자 통돌이 세탁기 TR14WK1(화이트)' 상품의 취급액에 영향을 끼치는 변수 상위 20개이다. 해당 상품은 생활용품 상품군의 상품 중 취급액 2위를 기록하였다.

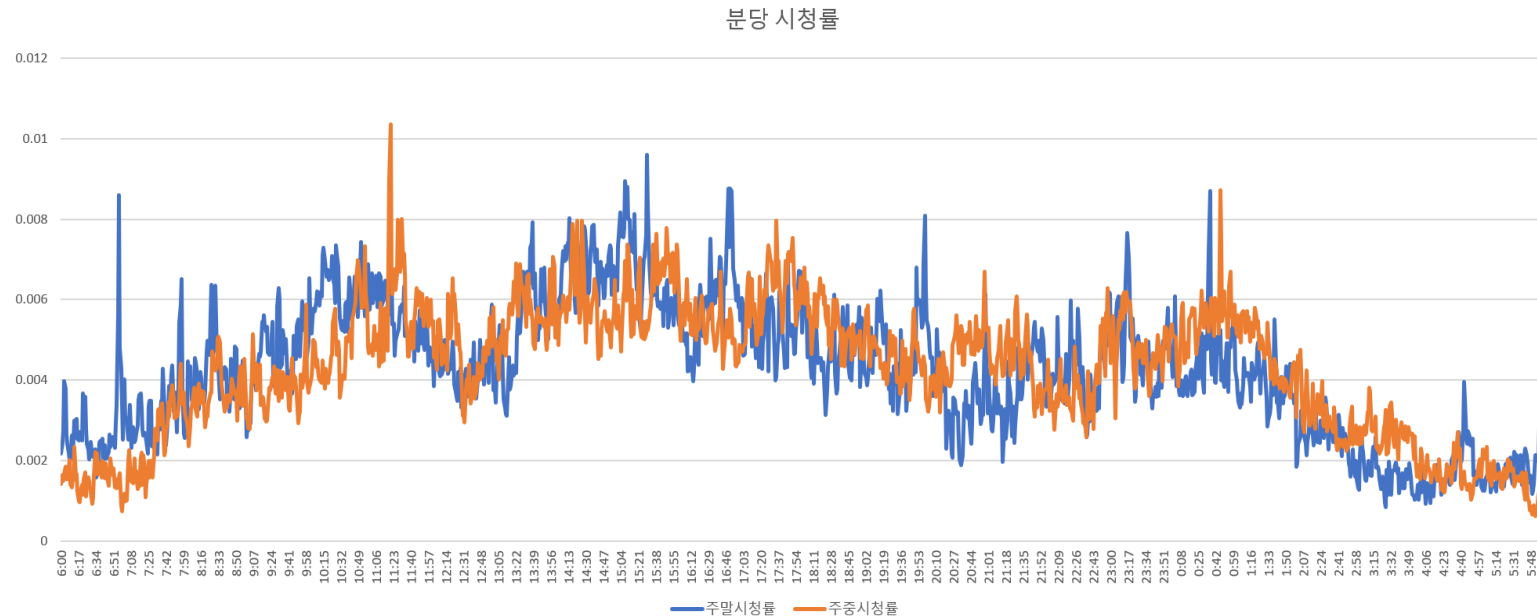
22:00~00:00 시간대 변수가 양의 영향을 끼치는 것을 확인할 수 있다.

◀ 요일별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, 18:00~00:00 시간대 항목이 색칠된 경우가 많았다.

## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 4. 시청률

- LIME 분석 결과, 시청률은 매출액에 중요한 변수 기능을 했다.
- 반면 타 채널 시청자 수 평균은 매출액에 중요한 변수로 작용하지 않았다고 분석되었다. 홈쇼핑 채널의 특성상 뉴스, 오락 등 다른 채널과의 차별성이 있기 때문으로 풀이된다.



	시청률		시청률
6:00~8:00 TOP3		18:00~20:00 TOP3	
1246	0.00299	1321	0.00491
73	0.00344	2601	0.00484
1245	0.00316	103	0.0059
8:00~10:00 TOP3		20:00~22:00 TOP3	
1787	0.00416	493	0.00446
1249	0.00378	2543	0.00446
2396	0.00416	1502	0.00459
10:00~12:00 TOP3		22:00~00:00 TOP3	
579	0.00615	1863	0.00558
2399	0.00615	2288	0.00536
973	0.00591	929	0.00355
12:00~14:00 TOP3		00:00~02:00 TOP3	
1177	0.0041	2478	0.00401
2581	0.00499	1060	0.00409
594	0.0041	160	0.005
14:00 ~ 16:00 TOP3		02:00~04:00 TOP3	
1195	0.00679	1604	0.00352
1305	0.00592	547	0.00264
1285	0.0067	2144	0.00352
16:00~18:00 TOP3			
1306	0.00487		
1933	0.00487		
1201	0.00529		

시간별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. (▲) 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, 시청률은 대부분 칠해졌다.



## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 5. 프라임 타임

- 오전 9~11시, 오후 8~11시(이상 출처 : 과학기술정보통신부)와 ns홈쇼핑의 식품 프라임 타임인 오후 4~6시를 프라임 타임으로 상정하여 분석한 것이다. 오전/식품/그 외/프라임 아님 4가지로 구분하였다.
- 프라임 타임이 아닐 때와 프라임 타임일 때를 비교하여 매출액에 끼친 차이를 볼 수 있다. 특히 식품 프라임 타임은 농수축 상품군의 매출에 유의미한 영향을 보인다.

PrimeTime_그외	ime_식품	ime_오전	ime_프라임아님	PrimeTime_그외	ime_식품	ime_오전	ime_프라임아님	PrimeTime_그외	ime_식품	ime_오전	ime_프라임아님
상품군 가구 TOP3				상품군 농수축 TOP3				상품군 의류 TOP3			
644	1	0	0	1306	0	1	0	1891	0	0	1
645	1	0	0	2543	0	0	0	1156	0	0	0
646	1	0	0	1933	0	1	0	1784	0	0	0
상품군 가전 TOP3				상품군 생활용품 TOP3				상품군 이미용 TOP3			
493	1	0	0	1285	0	0	0	2189	1	0	0
579	0	0	1	1364	0	0	0	1754	1	0	0
1321	0	0	0	634	0	0	0	741	1	0	0
상품군 건강기능 TOP3				상품군 속옷 TOP3				상품군 잡화 TOP3			
2166	0	0	0	2192	1	0	0	1813	0	0	0
2138	0	0	0	1664	1	0	0	248	0	0	1
2114	0	0	0	1165	0	0	1	1448	0	0	1
상품군 주방 TOP3				상품군 침구 TOP3							
1195	0	0	0	809	0	0	0				
1787	0	0	1	808	0	0	0				
2399	0	0	1	807	0	0	0				

▲ 상품군별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, 농수축 상품군에서 식품프라임은 대부분 녹색이다.

## 5. 편성 최적화 방안 제안 ●●●

요일별 최적 방안

## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 1. 월화수목금 vs (토)일

- 주말에 편성된 상품이 평일에 편성된 상품보다 더 잘 판매되며, 주말 중 특히 일요일에 더 잘 판매된다.
- 따라서 인기가 있거나 단가가 높은 상품을 주말에 편성하면 총 매출을 올릴 수 있을 것이다.

상품군 가구 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일	상품군 의류 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일
644	0	0	0	0	0	0	1	1891	0	0	0	0	0	0	1
645	0	0	0	0	0	0	1	1156	0	0	0	0	0	1	0
646	0	0	0	0	0	0	1	1784	0	0	0	0	0	1	0
상품군 가전 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일	상품군 이미지 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일
493	0	0	0	0	0	1	0	2189	0	0	1	0	0	0	0
579	0	0	0	0	0	0	1	1754	0	0	0	0	1	0	0
1321	0	0	0	0	0	0	1	741	1	0	0	0	0	0	0
상품군 건강기능 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일	상품군 잡화 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일
2166	0	0	1	0	0	0	0	1813	0	0	0	0	0	1	0
2138	0	1	0	0	0	0	0	248	0	0	0	1	0	0	0
2114	0	1	0	0	0	0	0	1448	0	1	0	0	0	0	0
상품군 농수축 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일	상품군 주방 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일
1306	0	0	0	0	0	0	1	1195	0	0	0	0	0	1	0
2543	0	0	0	0	0	0	1	1787	0	0	0	0	0	1	0
1933	0	0	0	0	0	0	1	2399	0	0	0	0	0	1	0
상품군 생활용품 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일	상품군 침구 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일
1285	0	0	0	0	0	0	1	809	0	1	0	0	0	0	0
1364	0	0	0	0	0	0	1	808	0	1	0	0	0	0	0
634	0	0	0	0	0	0	1	807	0	1	0	0	0	0	0
상품군 속옷 TOP3	요일_월	요일_화	요일_수	요일_목	요일_금	요일_토	요일_일								
2192	0	0	1	0	0	0	0								
1664	0	0	0	1	0	0	0								
1165	0	0	0	0	0	1	0								

▲ 상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, '일요일'이 대부분 칠해졌다.

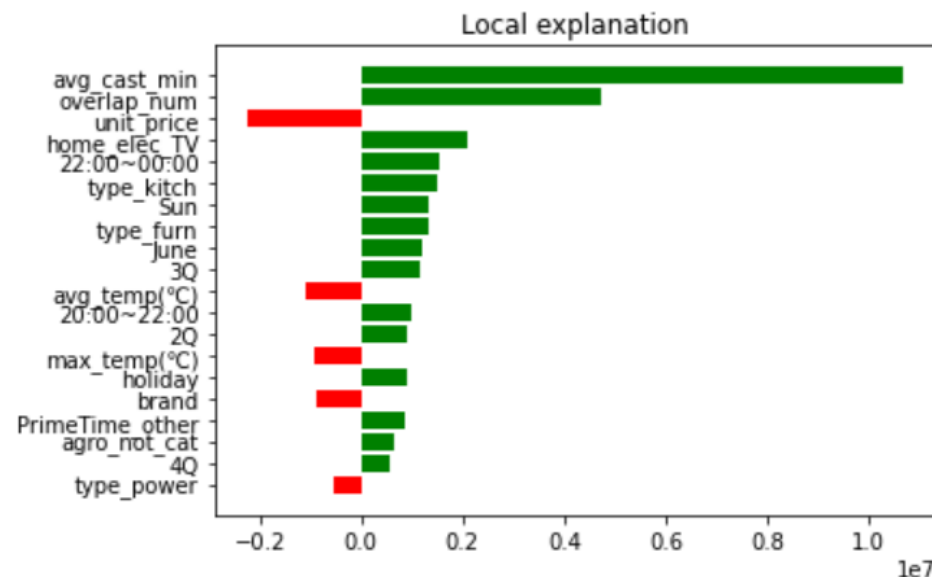
## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 2. 평일 vs 휴일(금, 토, 공휴일)

- 휴일(금요일 포함)에 편성된 상품이 월요일~목요일에 편성된 상품보다 더 잘 판매된다. 앞선 결과와 마찬가지로 휴일에 인기가 있거나 단가가 높은 상품을 편성하면 총 매출을 증가시킬 수 있을 것이다.

상품군 가구 TOP3	휴일	상품군 의류 TOP3	휴일	상품군 생활용품 TOP3	휴일
644	1	1891	1	1285	1
645	1	1156	1	1364	1
646	1	1784	1	634	1
상품군 가전 TOP3		상품군 이미용 TOP3		상품군 속옷 TOP3	
493	1	2189	0	2192	0
579	1	1754	1	1664	0
1321	1	741	0	1165	1
상품군 건강기능 TOP3		상품군 잡화 TOP3		상품군 침구 TOP3	
2166	0	1813	1	809	0
2138	0	248	0	808	0
2114	0	1448	0	807	0
상품군 농수축 TOP3		상품군 주방 TOP3			
1306	1	1195	1		
2543	1	1787	1		
1933	1	2399	1		

- ▲ 상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 녹색으로 칠했는데, 휴일이 대부분 칠해졌다.



- ▲ 2020-06-07 20:00:00에 방영된 '보루네오 델루나 유로탑 슬라이딩 LED침대 킹' 상품의 취급액에 영향을 끼치는 변수 상위 20개이다.  
해당 상품은 가구 상품군의 상품 중 취급액 3위를 기록하였다.  
휴일(holiday)변수가 양의 영향을 끼치는 것을 확인할 수 있다.

## 5. 편성 최적화 방안 제안 . . .

### 핵심변수 3. 평균기온, 최고기온

- 평균기온 및 최고기온이 높을수록 상품 판매에 부정적인 영향을 준다.

상품군 가구 TOP3	평균기온(°C)	최고기온(°C)	상품군 의류 TOP3	평균기온(°C)	최고기온(°C)	상품군 생활용품 TOP3	평균기온(°C)	최고기온(°C)
644	22	28.1	1891	23.1	29.1	1285	23.9	28.2
645	22	28.1	1156	24	28.8	1364	23.9	28.2
646	22	28.1	1784	22.4	27.5	634	22	28.1
상품군 가전 TOP3			상품군 이미용 TOP3			상품군 속옷 TOP3		
493	21.9	27.7	2189	22	24.6	2192	22	24.6
579	22	28.1	1754	21.9	27.4	1664	19.9	22.4
1321	23.9	28.2	741	24.1	31.4	1165	24	28.8
상품군 건강기능 TOP3			상품군 잡화 TOP3			상품군 침구 TOP3		
2166	22	24.6	1813	22.4	27.5	809	25.6	32.9
2138	24.4	30.3	248	24	29.9	808	25.6	32.9
2114	24.4	30.3	1448	23	28.9	807	25.6	32.9
상품군 농수축 TOP3			상품군 주방 TOP3					
1306	23.9	28.2	1195	24	28.8			
2543	23.8	29.2	1787	22.4	27.5			
1933	23.1	29.1	2399	23.6	29.1			

▲ 상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 음의 가중치를 가지는 중요 변수를 빨간색으로 칠했는데, 평균기온과 최고기온이 대부분 칠해졌다.

## 5. 편성 최적화 방안 제안 ●●●

카테고리별 최적 방안



## 5. 편성 최적화 방안 제안 . . .

## 핵심 상품군 1. 가전

- 여러 가전제품 종류 간의 취급액 차이를 반영하기 위해 세분화하였다.
  - LIME 분석 결과 소분류 항목 중 세탁기의 가중치가 가장 긍정적으로 예측되었다.
  - 다른 시간대 별, 요일 별 LIME 분석 결과 TV 역시 매출에 긍정적인 영향을 주는 것으로 예상되었다.
- 가전제품 중 **세탁기**와 **TV** 상품의 방송 비중을 늘리는 것이 **매출 증가**에 좋은 영향을 미칠 것이다.

	home_elec_TV	not_home_elec	home_elec_drie	home_elec_airclean	home_elec_fridge	home_elec_notebk	home_elec_wrlss_vcr	home_elec_washmch	home_elec_aircon	home_elec_clothmcl
	가전제품_TV	가전제품_가전제품 아님	가전제품_건조기	가전제품_공기청정기	가전제품_냉장고	가전제품_노트북	가전제품_무선청소기	가전제품_세탁기	가전제품_에어컨	가전제품_의류관리기
상품군 가전										
493	0	0	0	0	0	0	0	1	0	0
579	0	0	0	0	0	0	0	1	0	0
1321	0	0	0	0	0	0	0	1	0	0

상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 초록색으로 칠했는데, '가전' 상품군의 상위 3개 품목을 분석한 결과, 세탁기의 영향이 긍정적으로 작용하는 것을 알 수 있다.

## 핵심 상품군 2. 농수축

- 상품군 중 가장 비중이 높은 농수축 상품을 면밀히 분석하기 위해 상품군 내에서 소분류를 진행하였다.
  - LIME 분석 결과 농수축 소분류 중 '어류'의 영향이 눈에 띄게 크게 작용하였다.
- 농수축 상품군의 상품 중 **어류**의 비중을 크게 유지한다면, **매출 증가**에 좋은 영향이 있을 것이다.

	agro_driedfish	agro_fruit	agro_kimchi	agro_not_cat	agro_rice	agro_fish	agro_meat	agro_drink	agro_sauce/spice	agro_snacks
	농수축소분류_건어류	농수축소분류_과일	농수축소분류_김치	농수축소분류_분류없음	농수축소분류_쌀	농수축소분류_어류	농수축소분류_육류	농수축소분류_음료	농수축소분류_장류/조미료	농수축소분류_주전부리
상품군 농수축										
1306	0	0	0	0	0	1	0	0	0	0
2543	0	0	0	0	0	1	0	0	0	0
1933	0	0	0	0	0	1	0	0	0	0



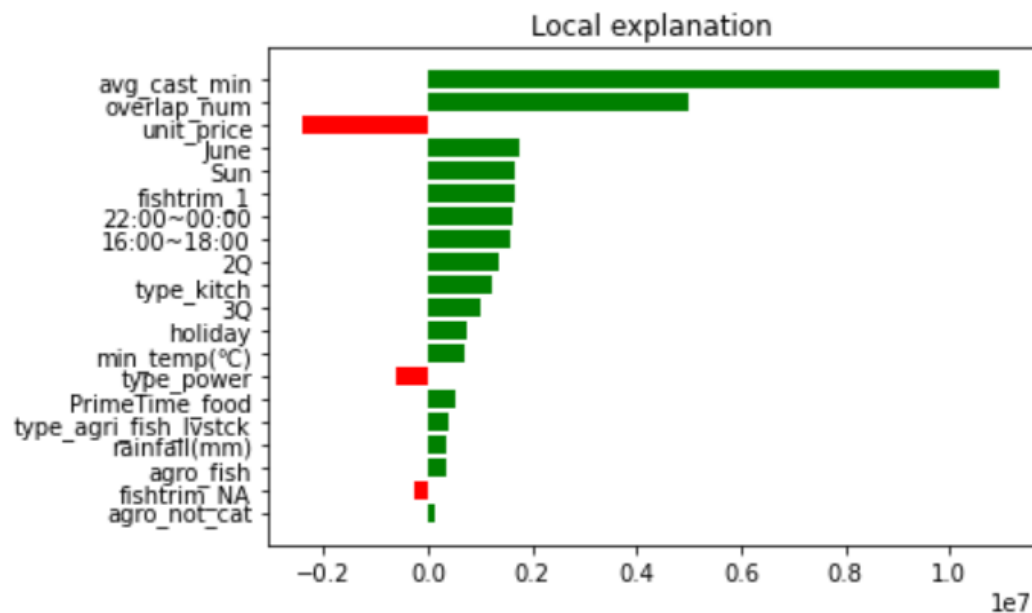
## 5. 편성 최적화 방안 제안 . . .

## 핵심 상품군 2. 농수축 - 손질

- 어류의 경우 손질 여부에 따라 다시 분류하였는데, 손질된 어류 상품의 영향이 유의미하였다.  
→ 특히 **손질된 어류 상품**의 배송 및 판매 비중을 늘린다면, **매출 증가**에 긍정적으로 작용할 것이다.

	fishtrim_0	fishtrim_1	fishtrim_NA	agro_fish
	어류손질여부_0	어류손질여부_1	어류손질여부_해당없음	농수축소분류_어류
12:00~14:00 상위 3개				
1177	0	0	1	0
2581	0	1	0	1
594	0	0	1	0
14:00 ~ 16:00 상위 3개				
1195	0	0	1	0
1305	0	1	0	1
1285	0	0	1	0
16:00~18:00 상위				
1306	0	1	0	1
1933	0	1	0	1
1201	0	0	1	0
18 ~ 20시				
1321	0	0	1	0
2601	0	1	0	1
103	0	1	0	1
20 ~ 22시				
493	0	0	1	0
2543	0	1	0	1
1502	0	0	1	0

▲ 상품군 별 상위 3개 데이터를 추출하여 LIME 분석한 자료이다. 양의 가중치를 가지는 중요 변수를 초록색으로 칠했는데, '어류'에 긍정적인 영향을 받는 데이터의 경우 손질여부가 같이 긍정적인 영향을 미쳤다는 것을 알 수 있다.



- ▲ 2020-06-14 16:00:00에 방영된 '현대어찬 손질갈치 14팩' 상품의 취급액에 영향을 끼치는 변수 상위 20개이다.  
해당 상품은 농수축 상품군의 상품 중 취급액 1위를 기록하였다.  
손질(fishtrim\_1) 변수가 양의 영향을 끼치는 것을 확인할 수 있다.

...

감사합니다