



**X**

# **판매실적 예측 & 편성 최적화 방안**

시로앤마로

이정민 (jungmin9195@naver.com)

임경룡 (kjiky7777@naver.com)

주은혁 (wndmsgur96@naver.com)

황호진 (hojin1223@naver.com)

# INDEX

---



홈쇼핑,  
코로나를 만나다.



EDA



Modeling  
&  
Algorithm



SUMMARY

## 홈쇼핑의 인식 변화 (검색어 : NS 홈쇼핑)



http://find.mk.co.kr/new/search.php?pageNum=1&cat=&cat1=&media\_eco=&pageSize=10&sub=all&dispFlag=OFF&page=news&s\_kwd=NS+%C8%A8%BC%EE%C7%CE&s\_page=new  
s&go\_page=&ord=1&ord1=1&ord2=0&s\_keyord=NS+%C8%A8%BC%EE%C7%CE&s\_j\_k  
eyword=NS+%C8%A8%BC%EE%C7%CE&s\_author=&y1=2020&m1=01&d1=01&y2=2020  
&m2=08&d2=25&media%5B%1%5D=00001&ord=1&area=ttbd

# 2020

## '고객, 소비자' 중심의 인식이 높은 편



# 2019

## 홈쇼핑의 인식 변화 (검색어 : NS 홈쇼핑)



## '고객, 소비자' 중심의 인식이 높은 편



“내가 원하는 상품을 사자”

# 홈쇼핑, 코로나를 만나다.

## 홈쇼핑의 인식 변화 (검색어 : NS 홈쇼핑)

**‘브랜드, 대표, 제품’ 등 특정 상품의 인식과 더불어  
‘기업, 사업’에 대한 인식이 새롭게 늘어남**

## '고객, 소비자' 중심의 인식 약화



지난달 규모 서비스 **정도계획** 운영 인구 시설방송 시작

# 2020

# 홈쇼핑, 코로나를 만나다.

## 홈쇼핑의 인식 변화 (검색어 : NS 홈쇼핑)

**‘브랜드, 대표, 제품’ 등 특정 상품의 인식과 더불어  
‘기업, 사업’에 대한 인식이 새롭게 늘어남**

## '고객, 소비자' 중심의 인식 약화



## 코로나 19 이후, 소비자의 불안감으로 인한 특정 상품 + 기업의 신뢰 증대

**“내가 원하는 상품보다 내가 믿을 수 있는 상품을 사자”**



지난달 규모 서비스 정도 계획 운영 인구 시설 방송 시작

# 2020

# 홈쇼핑, 코로나를 만난다.

홈쇼핑의 인식 변화 (검색어 : NS 홈쇼핑)

‘브랜드, 대표, 제품’ 등 특정 상품의 인식과 더불어  
‘기업, 사업’에 대한 인식이 새롭게 늘어남

‘고객, 소비자’ 중심의 인식 약화



코로나 19 이후, 소비자의 불안감으로 인한  
특정 상품 + 기업의 신뢰 증대

“내가 원하는 상품보다 내가 믿을 수 있는 상품을 사자”

“브랜드 파워의 인식 증대”

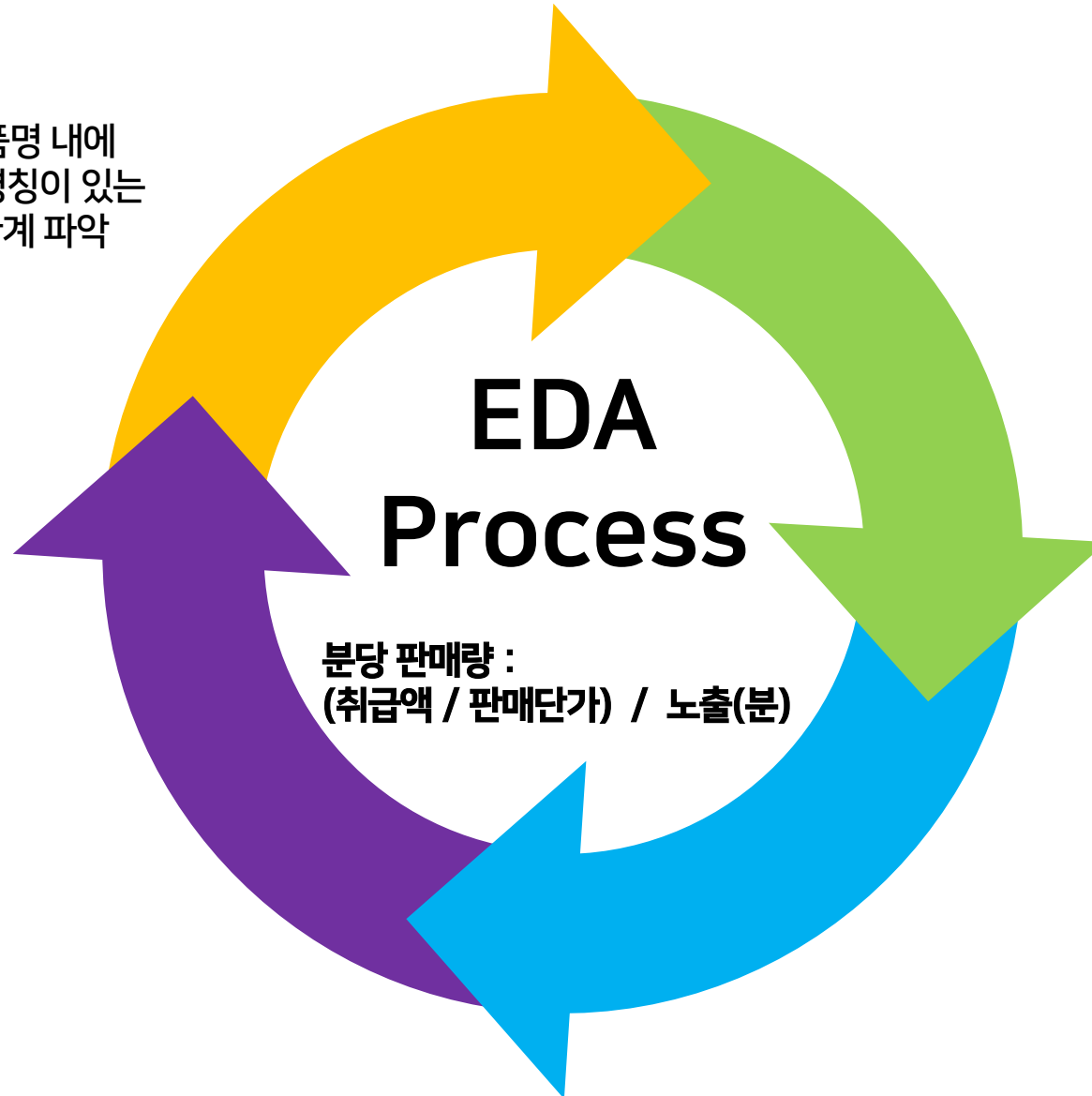


지난달 규모 서비스 정도 계획 운영 인구 시설방송 시작

2020



상품의 광고 효과를 확인하기 위해 상품명 내에  
브랜드명 및 사람 이름, 특정 장소의 명칭이 있는  
경우와 표준화된 척도 (분당 판매량) 관계 파악



# EDA : 브랜드파워

## CASE1



개그우먼 출신 **팽현숙** 씨가 여름철 보양식에 적합한 신제품 ‘팽현숙의 옛날오리부추한판’ 을 출시했다고 밝혔다.

(중략)

팽현숙 옛날 순대국 관계자는 “각 채널 예능에서 많은 사랑을 받고 있는데, 이와 함께 제품들도 사랑을 받고 있는 것 같다”며 “현재 순대국에 이어 홈쇼핑에서 대박 행진을 이어가고 있다”고 덧붙였다.

출처 : [아이티비즈\(http://www.it-b.co.kr\)](http://www.it-b.co.kr)

## CASE2



예산군은 9일 **예산군 대표 쌀 브랜드 ‘미황’**이 지난해 말부터 진행된 8차례 TV홈쇼핑에서 모두 완판을 기록했다고 밝혔다.

예산군에 따르면 홈쇼핑 전문 판매채널인 ‘GS홈쇼핑’과 ‘홈앤쇼핑’에서 소개된 예산의 ‘미황’ 쌀은 3만3787세트(1세트 10kg×2포)를 판매해 총 16억원의 매출을 올리는 성과를 거뒀다.

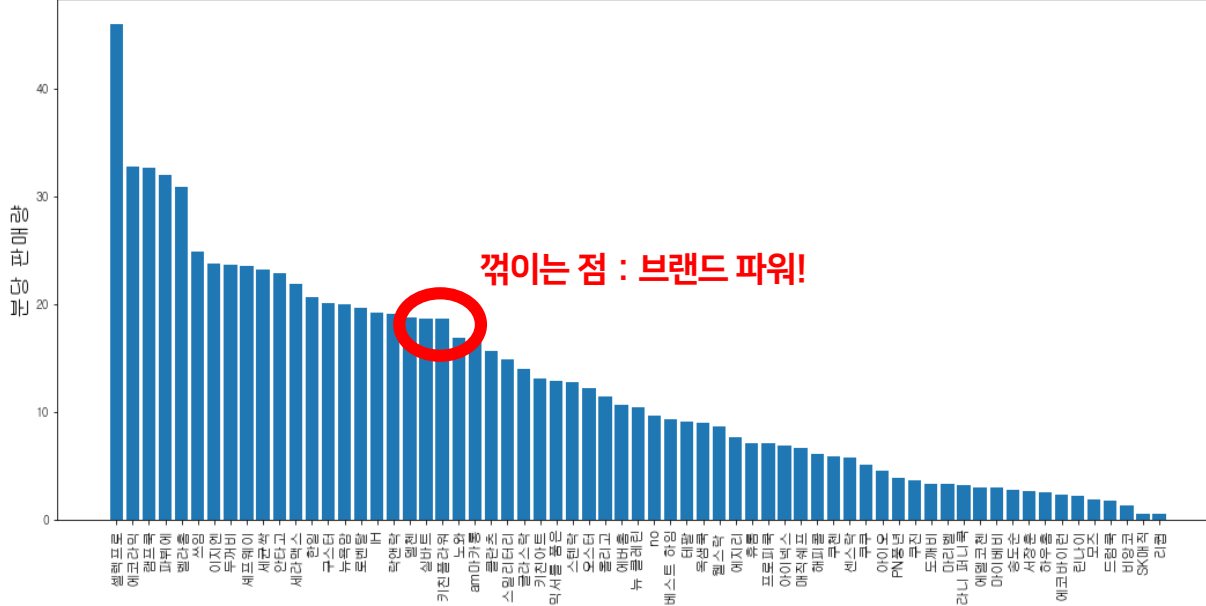
특히 예산 ‘미황’은 홈쇼핑 소비자 구매 만족도 조사에서 GS홈쇼핑 5점 만점에 4.7점, 홈앤쇼핑 100점 만점에 97점의 좋은 평가를 받았다.

출처 : [디트뉴스24\(http://www.dtnews24.com\)](http://www.dtnews24.com)

# 유명인 및 특정 브랜드에 대한 선호 현상 : 브랜드 파워 확인!

# EDA : 브랜드파워

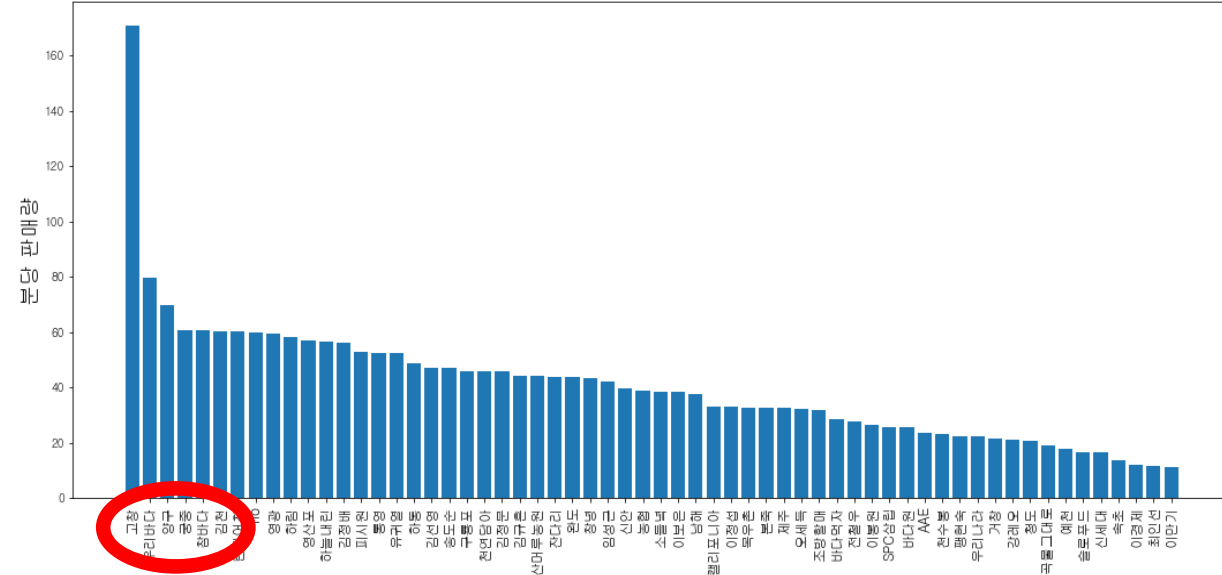
주방 브랜드별 분당 판매량



각 상품군에 있어, 특정 브랜드 상품에 대한 분당 판매량의 차이가 유의미 했음을 알 수 있다.

각 차트의 변곡점까지의 상품에 대해 브랜드 파워가 있다고 판단한다.

농수축 브랜드별 분당 판매량

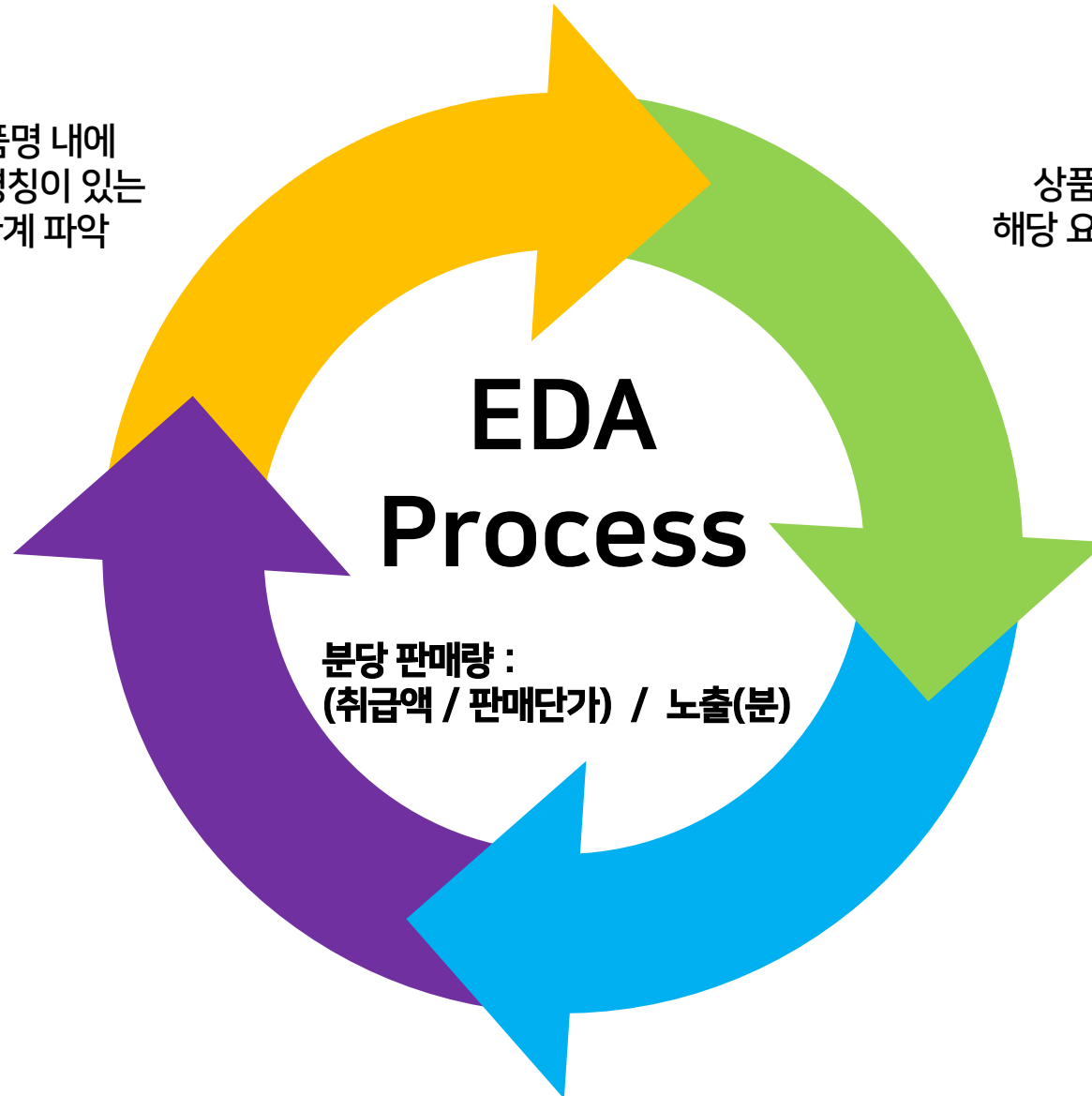


지역명 (고창, 양구, 김천) 을 가진 상품에 대한 선호도 확인

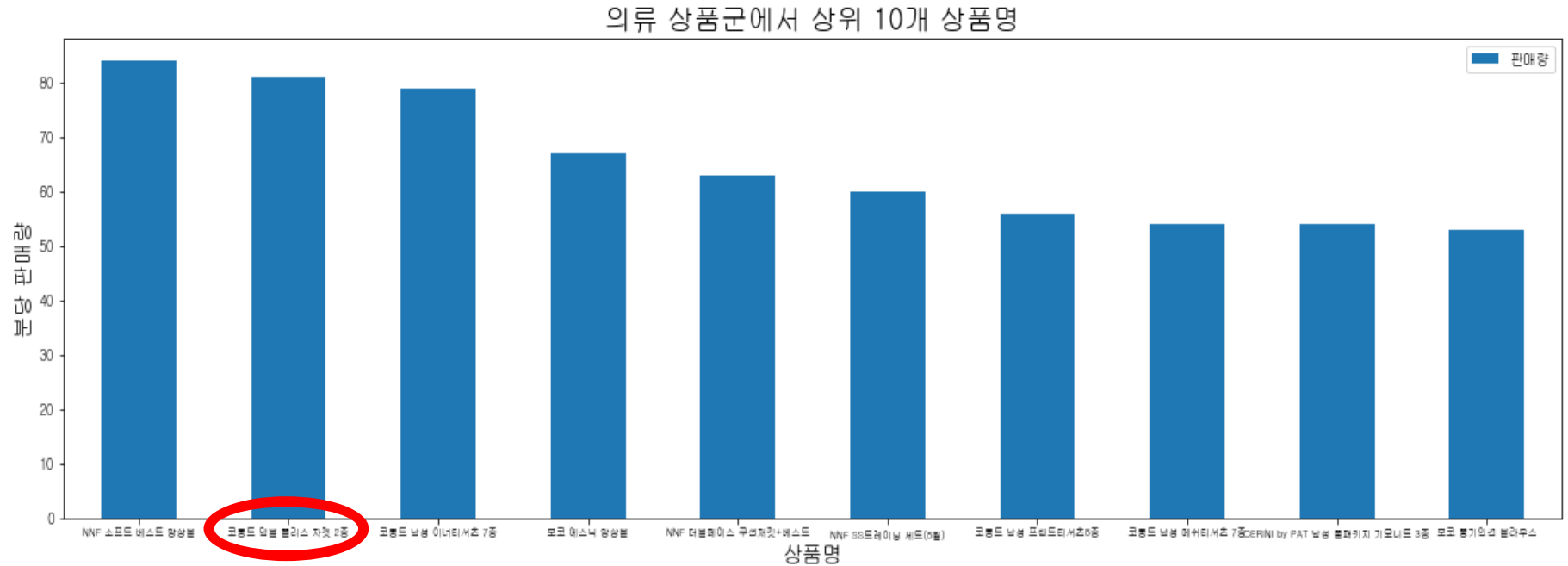
다른 상품군과 달리 농수축에서는 지역 혹은 사람 이름의 상품명을 갖는 경우, 분당 판매량이 높은 것으로 나타났다.

상품의 광고 효과를 확인하기 위해 상품명 내에  
브랜드명 및 사람 이름, 특정 장소의 명칭이 있는  
경우와 표준화된 척도 (분당 판매량) 관계 파악

동일한 상품의 단가가 다른 경우,  
상품에 대한 할인이 있는 것으로 간주하여  
해당 요인이 분당 판매량에 미치는 영향 파악



# EDA : 가격경쟁



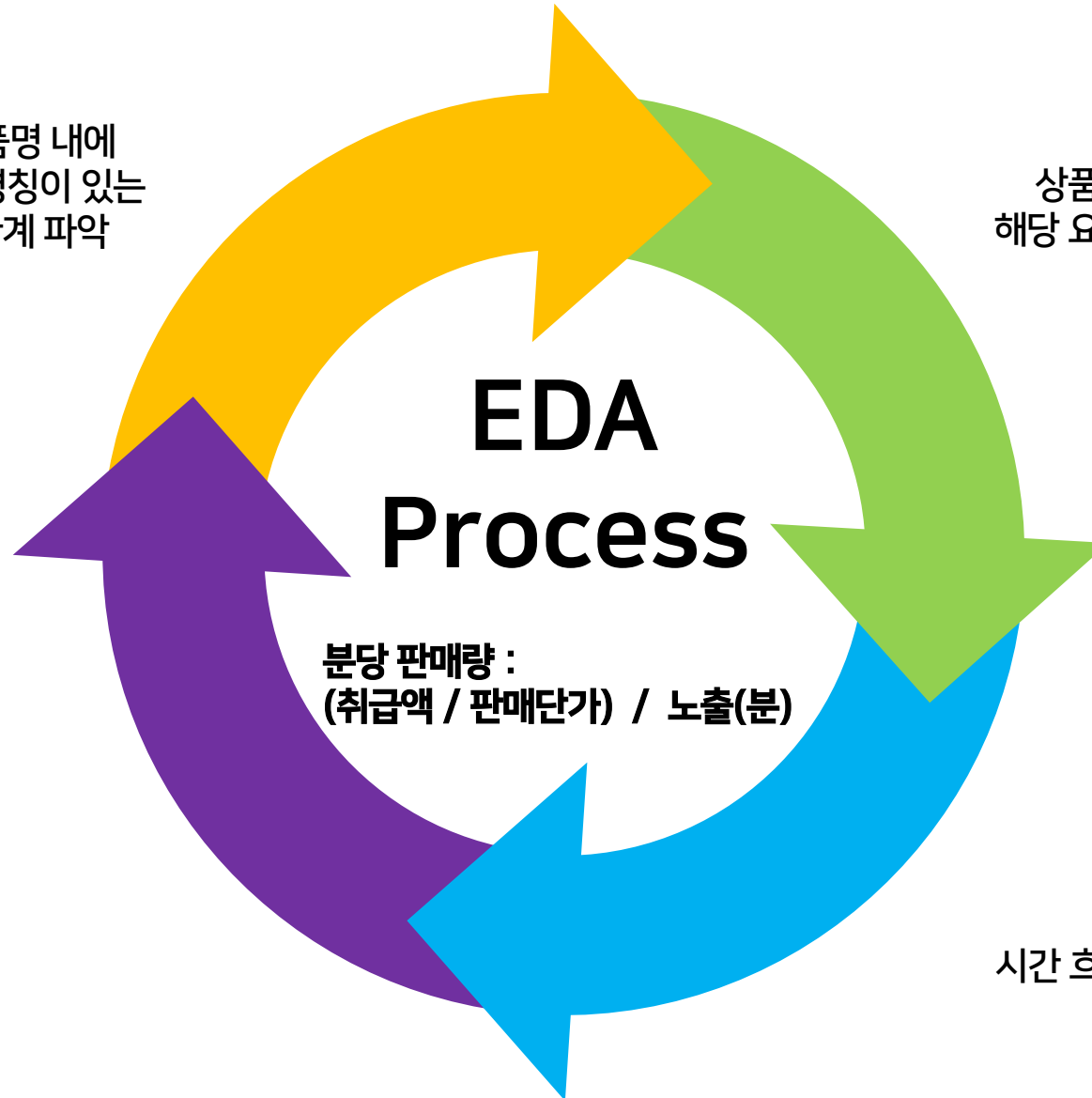
할인율 변수 책정

추성훈을 모델로 한 '코몽트 덤블 플리스 자켓'의 경우, 판매 단가를 낮췄을 때 (49900 -> 39900) 두번째로 높은 분당 판매량을 기록하였다.

이를 통해, 브랜드 파워와 더불어 할인 (가격 경쟁)에 대한 판매 효과가 존재함을 확인할 수 있었다.

상품의 광고 효과를 확인하기 위해 상품명 내에  
브랜드명 및 사람 이름, 특정 장소의 명칭이 있는  
경우와 표준화된 척도 (분당 판매량) 관계 파악

동일한 상품의 단가가 다른 경우,  
상품에 대한 할인이 있는 것으로 간주하여  
해당 요인이 분당 판매량에 미치는 영향 파악



계절, 날짜, 방송순서 등  
시간 흐름의 요인과 분당 판매량의 관계 파악

# EDA : 시간 (프라임 타임)

---

## 모든 경우에서 프라임 타임이 동일한 효과를 나타낼까?

NS홈쇼핑 관계자는 "보통 오후 7시부터 11시까지는 패션 뷰티를 주력하는 다른 홈쇼핑 업체의 프라임 타임인 것은 맞지만 우리의 경우 식품을 주력하고 있기 때문에 미래부에 인정하는 프라임 타임도 오전9시부터 11시, 오후 4시부터 6시다.  
“오전 프라임 타임에는 45.9%,  
오후 프라임 타임에는 98.9% 식품을 판매하고 있다”고 말했다.

출처 : 컨슈머치(<http://www.consumuch.com>)

**프라임 타임이 모든 상품군에 동일하게 적용되지 않는다!**

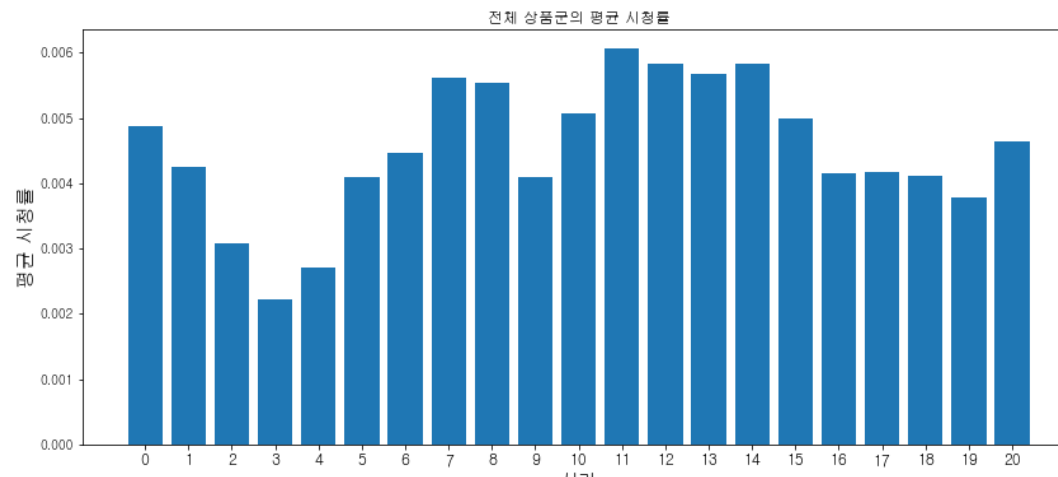
# EDA : 시간 (프라임 타임)

## 모든 경우에서 프라임 타임이 동일한 효과를 나타낼까?

NS홈쇼핑 관계자는 "보통 오후 7시부터 11시까지는 패션 뷰티를 주력하는 다른 홈쇼핑 업체의 프라임 타임인 것은 맞지만 우리의 경우 식품을 주력하고 있기 때문에 미래부에 인정하는 프라임 타임도 오전9시부터 11시, 오후 4시부터 6시다.

“오전 프라임 타임에는 45.9%,  
오후 프라임 타임에는 98.9% 식품을 판매하고 있다”고 말했다.

출처 : 컨슈머치(<http://www.consumuch.com>)



**프라임 타임이 모든 상품군에 동일하게 적용되지 않는다!**

**기존 프라임 타임의 시청률이 높지 않은 것으로 나타났다!**



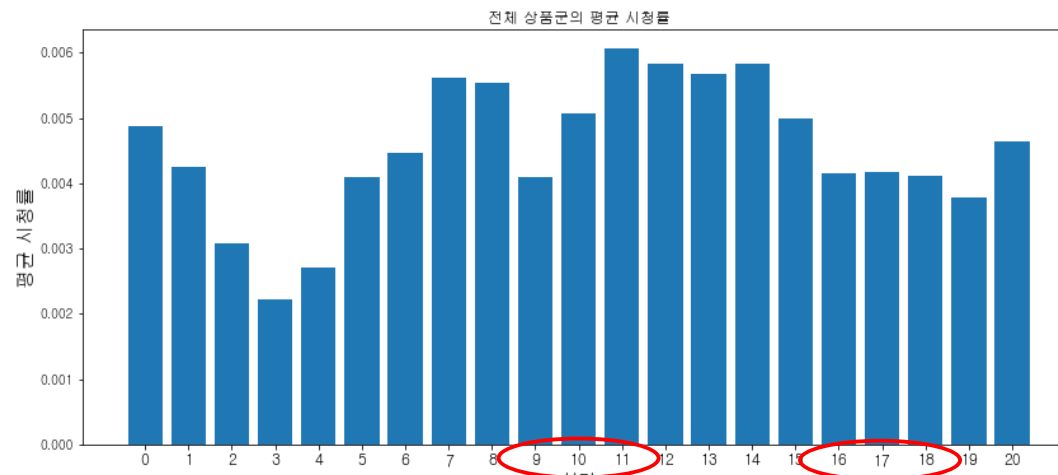
# EDA : 시간 (프라임 타임)

## 모든 경우에서 프라임 타임이 동일한 효과를 나타낼까?

NS홈쇼핑 관계자는 "보통 오후 7시부터 11시까지는 패션 뷰티를 주력하는 다른 홈쇼핑 업체의 프라임 타임인 것은 맞지만 우리의 경우 식품을 주력하고 있기 때문에 미래부에 인정하는 프라임 타임도 오전9시부터 11시, 오후 4시부터 6시다.

“오전 프라임 타임에는 45.9%,  
오후 프라임 타임에는 98.9% 식품을 판매하고 있다”고 말했다.

출처 : 컨슈머치(<http://www.consumuch.com>)



**프라임 타임이 모든 상품군에 동일하게 적용되지 않는다!**

**기존 프라임 타임의 시청률이 높지 않은 것으로 나타났다!**

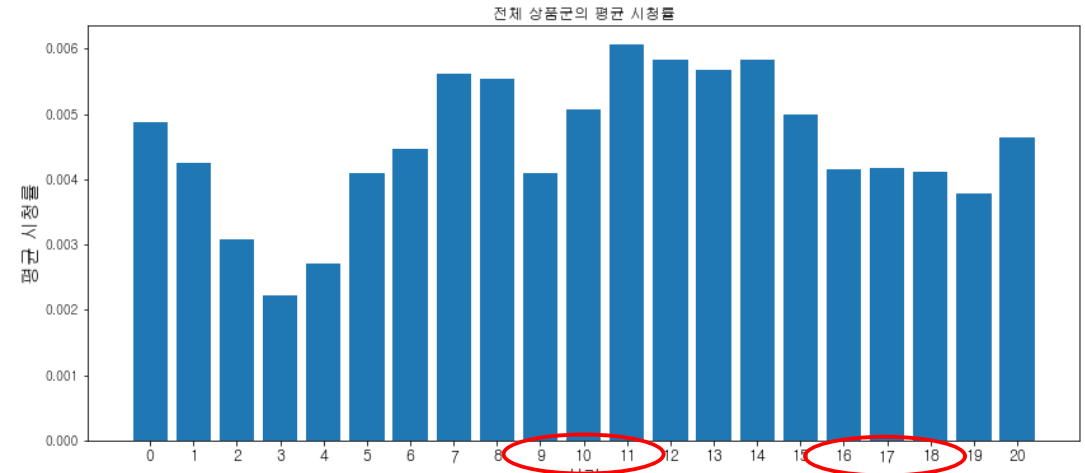
# EDA : 시간 (프라임 타임)

## 모든 경우에서 프라임 타임이 동일한 효과를 나타낼까?

NS홈쇼핑 관계자는 "보통 오후 7시부터 11시까지는 패션 뷰티를 주력하는 다른 홈쇼핑 업체의 프라임 타임인 것은 맞지만 우리의 경우 식품을 주력하고 있기 때문에 미래부에 인정하는 프라임 타임도 오전9시에서 11시, 오후 4시부터 6시다.

“오전 프라임 타임에는 45.9%,  
오후 프라임 타임에는 98.9% 식품을 판매하고 있다”고 말했다.

출처 : 컨슈머치(<http://www.consumuch.com>)

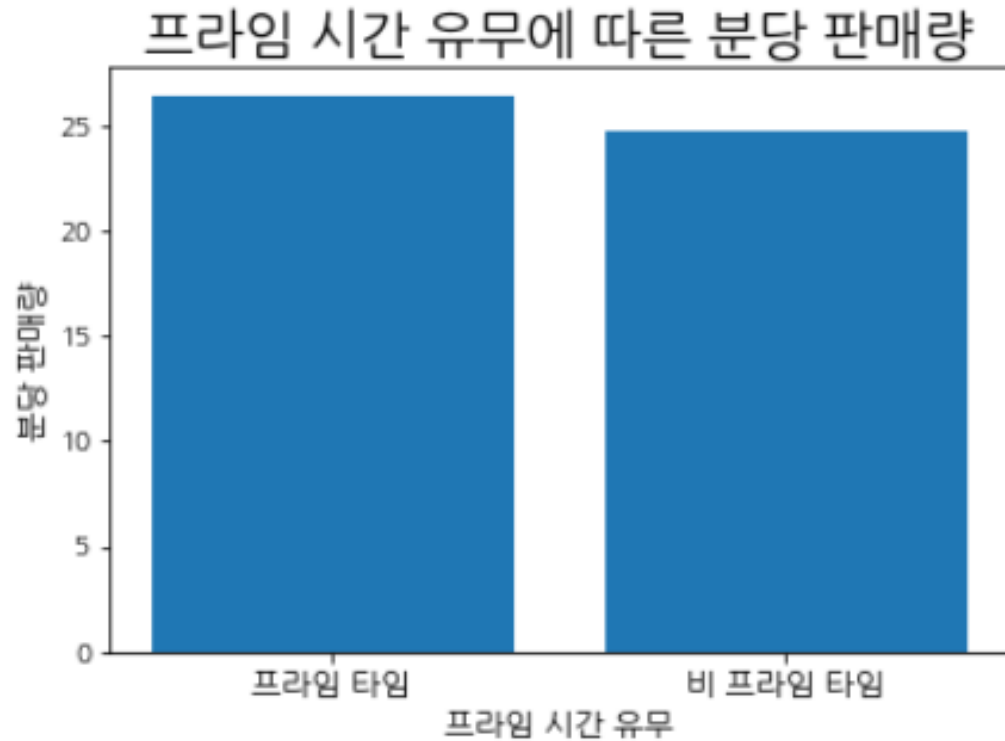


프라임 타임이 모든 상품군에 동일하게 적용되지 않는다!

기존 프라임 타임의 시청률이 높지 않은 것으로 나타났다!

## 프라임 타임의 재설정 필요

# EDA : 시간 (프라임 타임)



새롭게 정의한 프라임 타임에서의 분당 판매량으로 보아, 프라임 타임의 실효성이 존재함을 알 수 있다.

새롭게 정의한 프라임 타임은 다음과 같다. (1시간 단위)

Ex) 2시 : 02시 00분 ~ 02시 59분

상품군	주중	주말
의류	02시, 10시, 18시	14시, 17시, 18시
속옷	02시, 13시, 14시	10시, 14시, 17시
주방	14시, 17시, 23시	06시, 15시, 23시
농수축	00시, 01시, 11시	10시, 15시, 23시
이미용	00시, 10시, 15시	11시, 18시, 23시
가전	14시, 15시, 16시	00시, 07시, 15시
생활용품	16시, 22시, 23시	08시, 16시, 17시
건강기능	00시, 11시, 14시	14시, 17시, 22시
잡화	02시, 10시, 16시	11시, 14시, 15시
가구	01시, 10시, 19시	06시, 08시, 23시
침구	15시, 22시, 23시	12시, 13시, 21시

‘NS홈쇼핑 프라임타임은 주중·주말 오전 9~11시, 오후 4~6시다.’

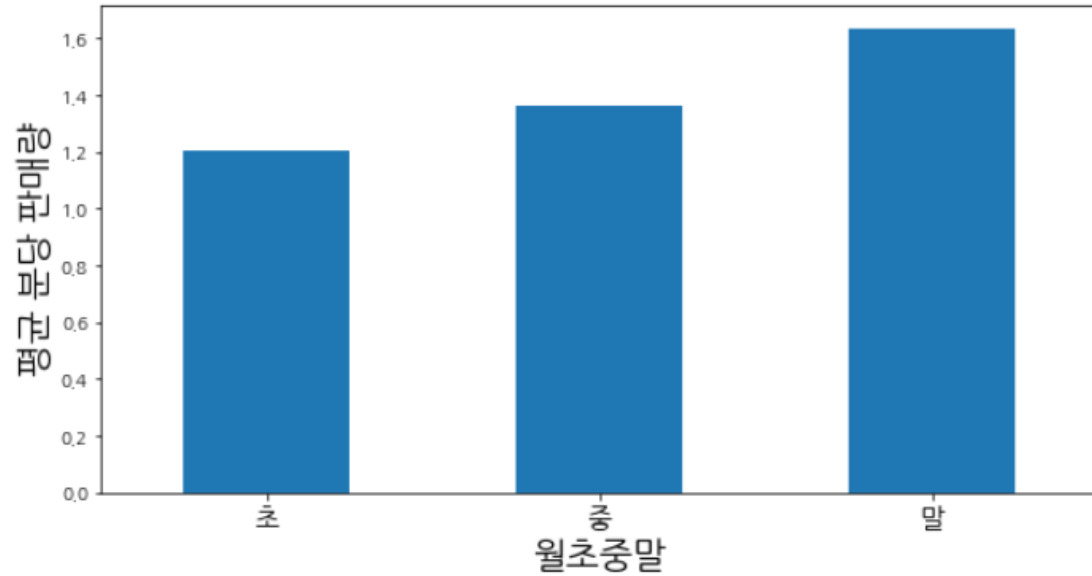
출처 : <https://www.etnews.com/20180718000049?m=1>

# EDA : 시간 (월 구분 , 주중/주말 구분)

월초 : 1~10일

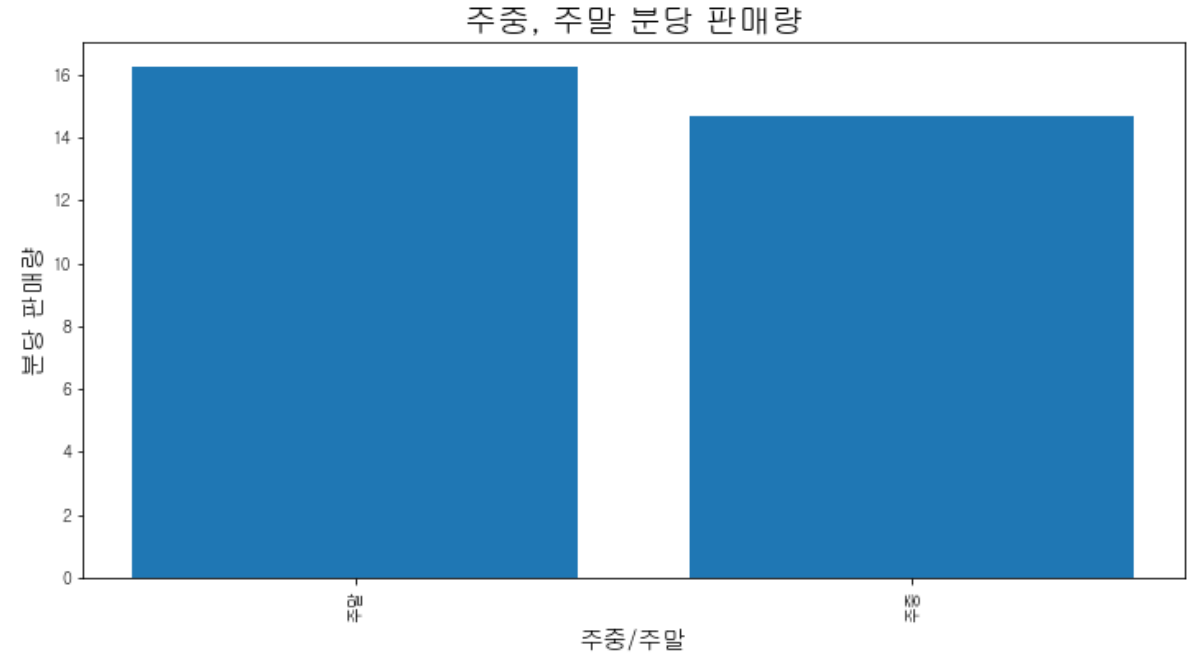
월중 : 11~20일

월말 : 21~30(31)일



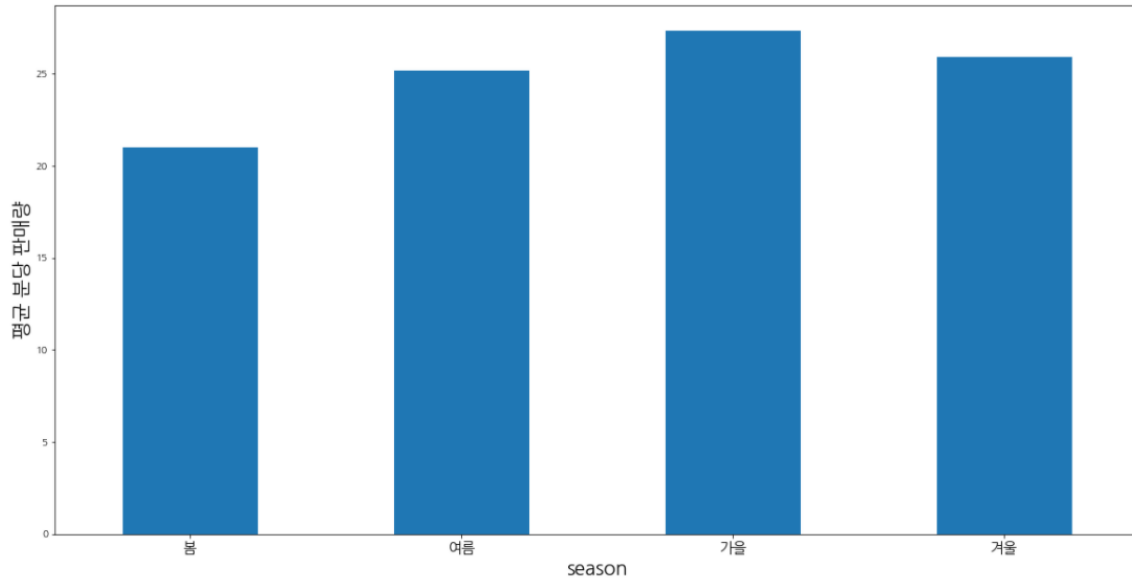
일부 상품군의 경우, 월초 ~ 월말 사이의 평균 분당 판매량에서 유의한 차이가 있었음을 알 수 있다.

특히, 월말의 분당 판매량이 가장 높았음을 알 수 있다.

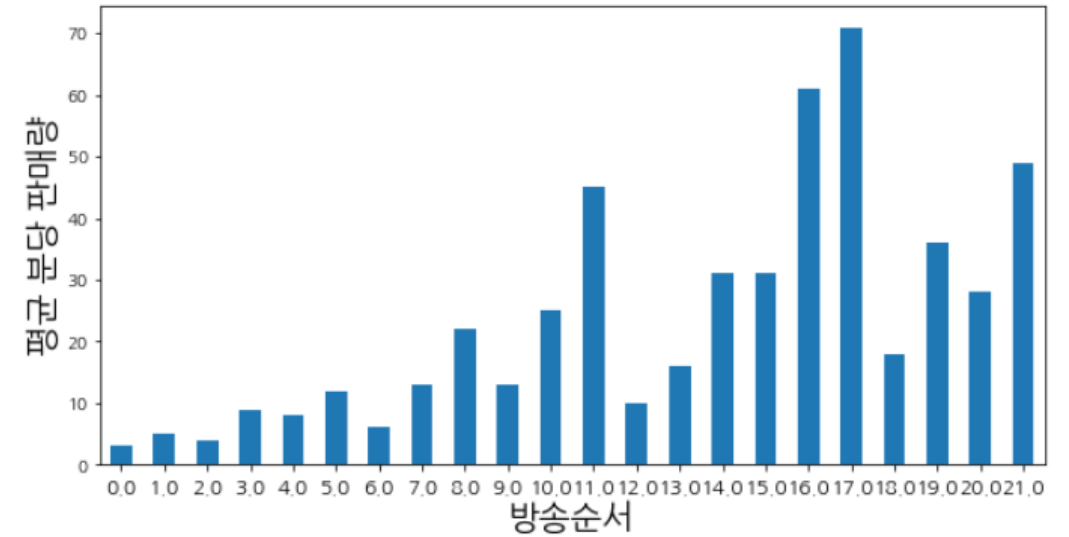


주중, 주말의 분당 판매량에는 차이가 있었으며 주말의 경우 판매량이 낮았음을 알 수 있다.

# EDA : 시간 (계절 / 방송 순서)



계절에 따라 전체 평균 분당 판매량의 차이가 존재함을 알 수 있다.  
특히, 가을의 분당 판매량이 가장 높은 것으로 나타났다.  
따라서, 계절 요인은 분당 판매량에 영향을 미치는 것으로 보인다.



동일한 상품이 일정 시각에 따라 연속으로 방영되는 경우를  
방송 순서로 각각 지정한 결과, 방송 순서가 후에 이어질수록  
분당 판매량이 증가하는 경향을 보이고 있다.

상품의 광고 효과를 확인하기 위해 상품명 내에  
브랜드명 및 사람 이름, 특정 장소의 명칭이 있는  
경우와 표준화된 척도 (분당 판매량) 관계 파악

동일한 상품의 단가가 다른 경우,  
상품에 대한 할인이 있는 것으로 간주하여  
해당 요인이 분당 판매량에 미치는 영향 파악

## EDA Process

분당 판매량 :  
 $(\text{취급액} / \text{판매단가}) / \text{노출(분)}$

기온, 미세먼지를 비롯한  
2020년 6월에 일어난 기상 요인을 파악하여  
분당 판매량과의 관계 파악

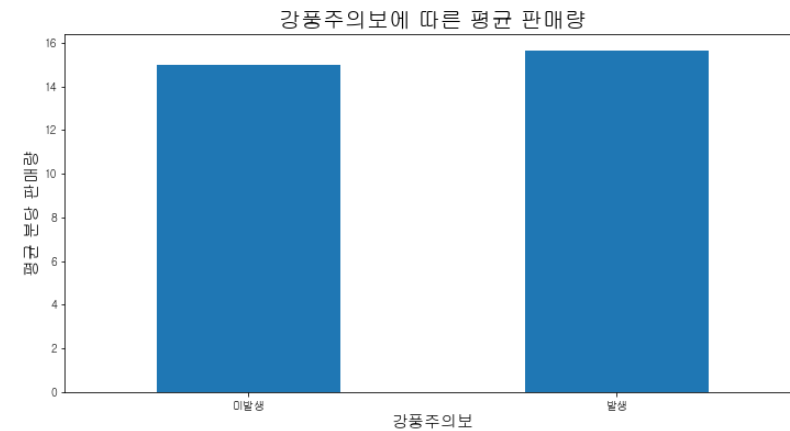
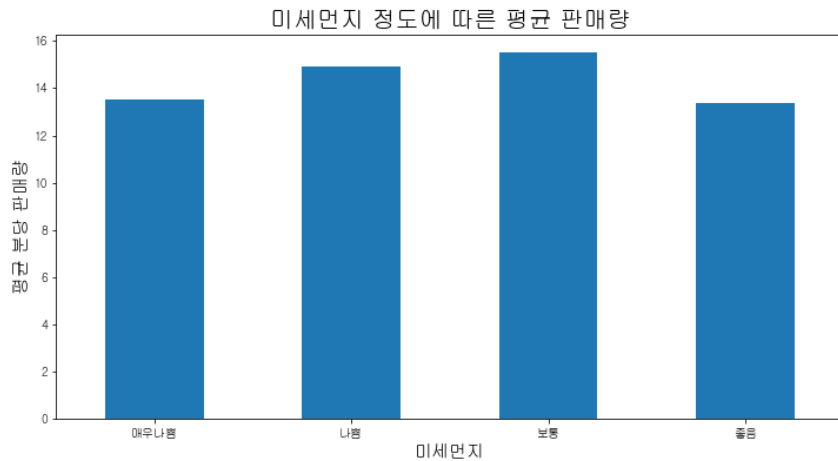
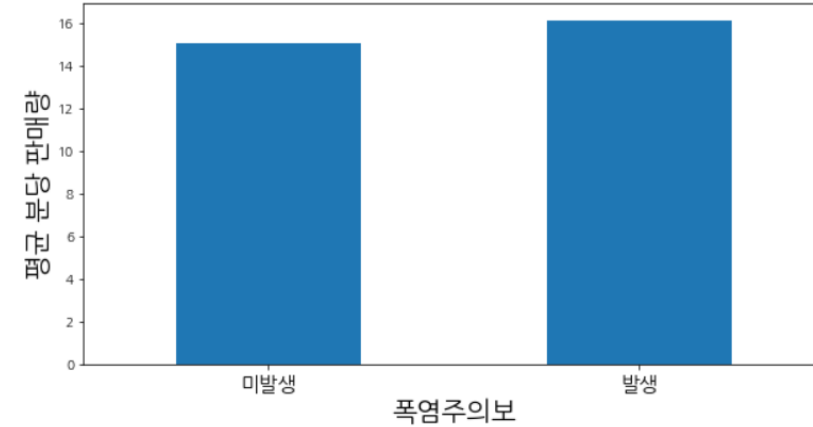
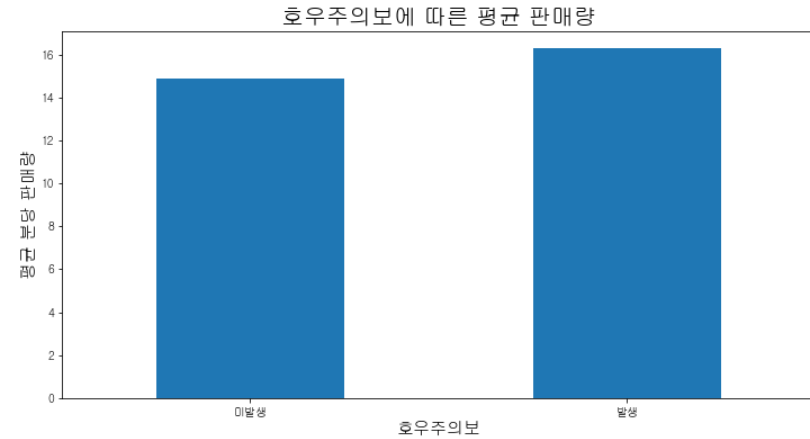
계절, 날짜, 방송순서 등  
시간 흐름의 요인과 분당 판매량의 관계 파악

# EDA : 기상 상황

2020년 6월에 일어난 기상특보(주의보) 상황은 다음과 같다.  
(강풍, 풍랑, 호우, 폭염, 폭풍해일)

※ 경보의 경우, 주의보 발표 내에 포함되어 있음  
※ 풍랑의 경우, 발생 조건이 강풍보다 높은 조건이므로 강풍 내에 포함되어 있음

자료출처 : <https://data.kma.go.kr/cmmn/main.do> (기상자료 개방 포털)



모든 기상 특보 및 미세먼지 상황에서 집단 간 분당 판매량의 차이가 있는 것으로 나타났다.  
기상 상황은 홈쇼핑 판매에 영향을 미칠 수 있는 것으로 보인다.

# EDA : 기상 상황 (기온)

## OLS Regression Results

```
=====
Dep. Variable:      취급액  R-squared (uncentered):      6.626
Model:              OLS    Adj. R-squared (uncentered):      6.626
Method:             Least Squares  F-statistic:          -4.401e+04
Date:               Sat, 26 Sep 2020  Prob (F-statistic):      1.00
Time:               08:54:46  Log-Likelihood:       -6.8787e+05
No. Observations:   37372    AIC:                  1.376e+06
Df Residuals:       37371    BIC:                  1.376e+06
Df Model:           1
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
temp	1.078e+06	7461.993	144.421	0.000	1.06e+06	1.09e+06

```
=====
Omnibus:            11110.890  Durbin-Watson:          0.754
Prob(Omnibus):      0.000     Jarque-Bera (JB):       50193.835
Skew:               1.390     Prob(JB):              0.00
Kurtosis:           7.950     Cond. No.              1.00
=====
```

단순 회귀 적합 결과, 기온과 취급액은 양의 상관관계를 보였으며  
이때, 해당 변수의 계수는 유의한 값을 가지고 있는 것으로 나타났다.

따라서, 기온은 홈쇼핑 판매에 어느정도 영향을 미친다고 볼 수 있다.



# EDA : 기상 상황 (기온)

OLS Regression Results						
Dep. Variable:	취급액	R-squared (uncentered):	6.626			
Model:	OLS	Adj. R-squared (uncentered):	6.626			
Method:	Least Squares	F-statistic:	-4.401e+04			
Date:	Sat, 26 Sep 2020	Prob (F-statistic):	1.00			
Time:	08:54:46	Log-Likelihood:	-6.8787e+05			
No. Observations:	37372	AIC:	1.376e+06			
Df Residuals:	37371	BIC:	1.376e+06			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
temp	1.078e+06	7461.993	144.421	0.000	1.06e+06	1.09e+06
Omnibus:	11110.890	Durbin-Watson:	0.754			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	50193.835			
Skew:	1.390	Prob(JB):	0.00			
Kurtosis:	7.950	Cond. No.	1.00			

**모든 기상상황은 홈쇼핑의 판매에  
영향을 미치는 것으로 판단된다!**

단순 회귀 적합 결과, 기온과 취급액은 양의 상관관계를 보였으며  
이때, 해당 변수의 계수는 유의한 값을 가지고 있는 것으로 나타났다.

따라서, 기온은 홈쇼핑 판매에 어느정도 영향을 미친다고 볼 수 있다.

# 변수 정의

## 연속형 변수

변수명	내용	특징
노출 (분)	1회 방송 당 노출 시간 (분)	수치 Min, Max 표준화
판매단가	한 상품당 가격 (원)	
기온	전국 평균 기온 (℃)	

종속 변수 : 취급액

기상데이터 : 기상자료개방포털  
(<https://data.kma.go.kr/cmmn/main.do>)

2019.01.01 ~ 2019.12.31  
2020.06.01 ~ 2020.06.30 자료 활용

## 범주형 변수

변수명	내용	특징
hour	방송 일자 (시간)	시간대별 범주화
weekday	방송 일자의 주중/주말 여부	1 : 주말 , 0 : 주중
month	방송 일자 (월)	월별 범주화
season	계절 구분	0~3 : 봄, 여름, 가을, 겨울
discount	할인 여부	1 : 할인 , 0 : 미할인
방송순서	연속 방송 편성시, 순서 정의	
호우주의보 강풍주의보 폭염주의보 폭풍해일	해당 방송 시간대 기상특보 상황 발표 여부	1 : 특보 상황 0 : 특보 해제
미세먼지 프라임 타임 초중말	미세먼지 농도 수준 상품군별 프라임 타임 해당 여부 월초,월중,월말 해당 여부	1 : 해당, 0 : 해당하지 않음 0~2 : 초,중,말 순서

# Modeling

---

## Linear Model

직선 모형을 통해 변수의 효과 (계수)와  
유의성 등 변수에 대한 직관적인 판단이 가능하다.

변수 간 다중공선성 발생 가능성이 높아  
모델 성능 저하가 일어날 수 있다.

데이터가 모두 선형으로 설명된다고 볼 수 없다.



# Modeling : Linear Model

---

	kind	lm_mape
0	속옷	1.690931e+10
1	잡화	9.025694e+01
2	건강기능	2.572205e+12
3	농수축	3.904540e+12
4	침구	5.475640e+01
5	생활용품	1.467500e+02
6	주방	1.063720e+02
7	이미용	3.772992e+01
8	가전	1.029648e+02
9	가구	1.072389e+02
10	의류	7.543015e+01

회귀 모형의 경우, 모든 상품군의 MAPE의 값이 다른 모형에 비해 매우 높은 수치를 기록하며, 4개의 모형 중 가장 낮은 성능을 보였다.

그러나 모든 상품군에서 모든 변수가 판매에 영향을 미친다는 모순적인 결과를 나타냈다.

이는 고차원의 데이터를 선형으로만 설명할 수 없다는 점에서 비롯된 오류임을 알 수 있다.

# Modeling

---

## Linear Model

직선 모형을 통해 변수의 효과 (계수)와  
유의성 등 변수에 대한 직관적인 판단이 가능하다.

변수 간 다중공선성 발생 가능성이 높아  
모델 성능 저하가 일어날 수 있다.

데이터가 모두 선형으로 설명된다고 볼 수 없다.

## Random Forest

다수의 의사결정나무를 이용하는 학습법  
여러 변수를 일괄적으로 다룰 수 있다.

모델 중 메모리 소모가 가장 많으며,  
Training data 양이 증가해도 성능이 급격히 향상하지 않는다.



# Modeling : Random Forest

---

	kind	rf_mape
0	속옷	95.983932
1	잡화	62.790722
2	건강기능	53.102343
3	농수축	50.972370
4	침구	31.360894
5	생활용품	78.643059
6	주방	73.959839
7	이미용	28.532725
8	가전	68.165376
9	가구	73.093836
10	의류	55.248364

n\_estimators : 생성할 tree의 개수

max\_features : 최대 선택할 특성의 수

max\_depth : 각 의사결정나무의 깊이를 설정. 트리가 깊어질수록 더 잘게 분류를 시키므로 일반적으로 정확도가 높아진다.

Random Forest의 경우,

특정 상품군에서의 MAPE는 30에 가까운 수치를 보이며 절반 이상의 데이터를 예측할 수 있는 좋은 성능을 보이고 있었으나,

생활용품, 속옷류를 비롯한 일부 상품군에서 100에 가까운 수치를 보이며 데이터를 거의 예측하지 못하는 모델 성능을 보였다.

따라서, 해당 모델을 일부 상품군의 예측에 적용하기에는 많은 어려움이 있을 것으로 보인다.

# Modeling

---

## Linear Model

직선 모형을 통해 변수의 효과 (계수)와  
유의성 등 변수에 대한 직관적인 판단이 가능하다.

변수 간 다중공선성 발생 가능성이 높아  
모델 성능 저하가 일어날 수 있다.

데이터가 모두 선형으로 설명된다고 볼 수 없다.

결정력이 약한 분류기를 다수 사용하여 정확도를 높이는 기법  
Gradient Boosting 과정에서 과적합을 방지할 수 있다.

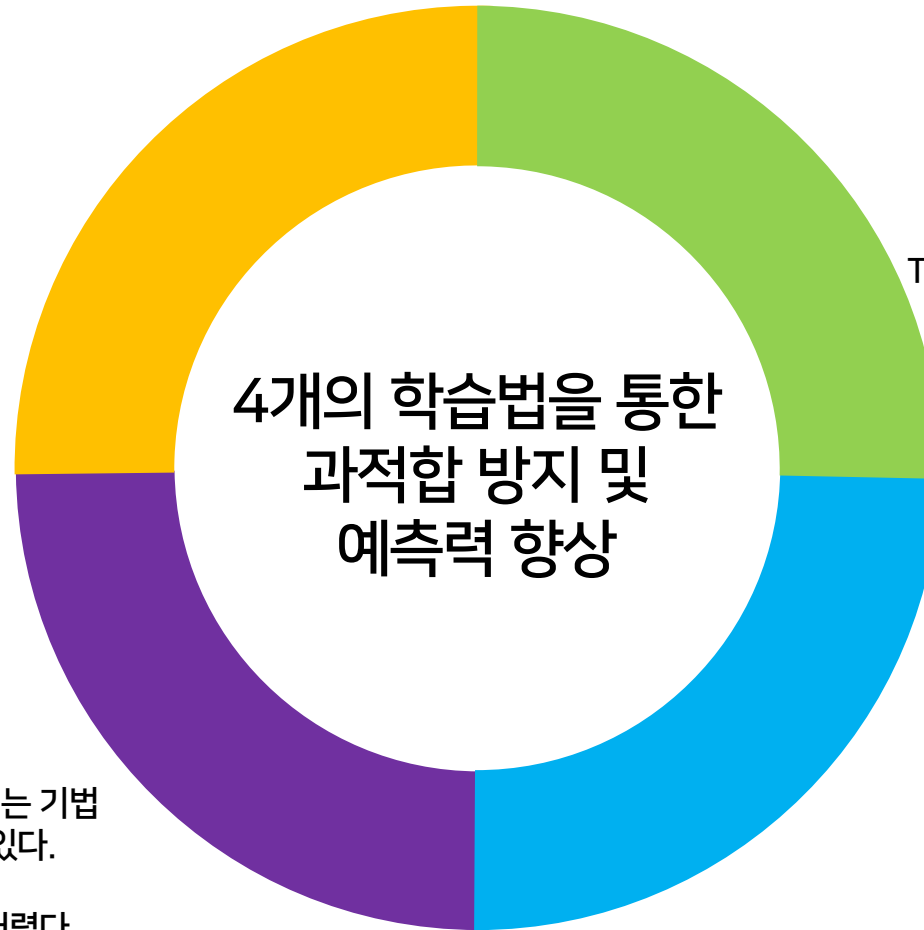
계산 과정 추적이 어려워 변수의 중요도를 파악하기 어렵다.

## XGBoost

## Random Forest

다수의 의사결정나무를 이용하는 학습법  
여러 변수를 일괄적으로 다룰 수 있다.

모델 중 메모리 소모가 가장 많으며,  
Training data 양이 증가해도 성능이 급격히 향상하지 않는다.



# Modeling : XGBoost

	kind	xgb_mape
0	속옷	123.049550
1	잡화	238.515339
2	건강기능	60.507165
3	능수축	57.530119
4	침구	169.139073
5	생활용품	317.600210
6	주방	224.886706
7	이미용	40.235363
8	가전	274.650390
9	가구	385.974469
10	의류	49.109365

eval\_metric : 검증데이터에 적용되는 모델 선택 기준

n\_estimators : 생성할 tree의 개수

learning\_rate : GBM은 초기의 추정값에서 시작하여 각각의Tree 결과를 사용하여 추정값을 업데이트 하는 과정에서 발생하는 변화의 크기

max\_depth : 각 의사결정나무의 깊이를 설정. 트리가 깊어질수록 더 잘게 분류를 시키므로 일반적으로 정확도가 높아진다.

XGBoost의 경우,

의류와 이미용을 제외한 나머지 상품군의 MAPE는 50을 넘겼으며, 그 중 다수가 100을 넘는 수치를 기록하며 좋지 못한 성능을 보였다.

따라서, 해당 모델을 일부 상품군의 예측 과정에 적용하기에는 많은 어려움이 있을 것으로 보인다.



# Modeling

## Linear Model

직선 모형을 통해 변수의 효과 (계수)와  
유의성 등 변수에 대한 직관적인 판단이 가능하다.

변수 간 다중공선성 발생 가능성이 높아  
모델 성능 저하가 일어날 수 있다.

데이터가 모두 선형으로 설명된다고 볼 수 없다.

결정력이 약한 분류기를 다수 사용하여 정확도를 높이는 기법  
Gradient Boosting 과정에서 과적합을 방지할 수 있다.

계산 과정 추적이 어려워 변수의 중요도를 파악하기 어렵다.

## XGBoost

## Random Forest

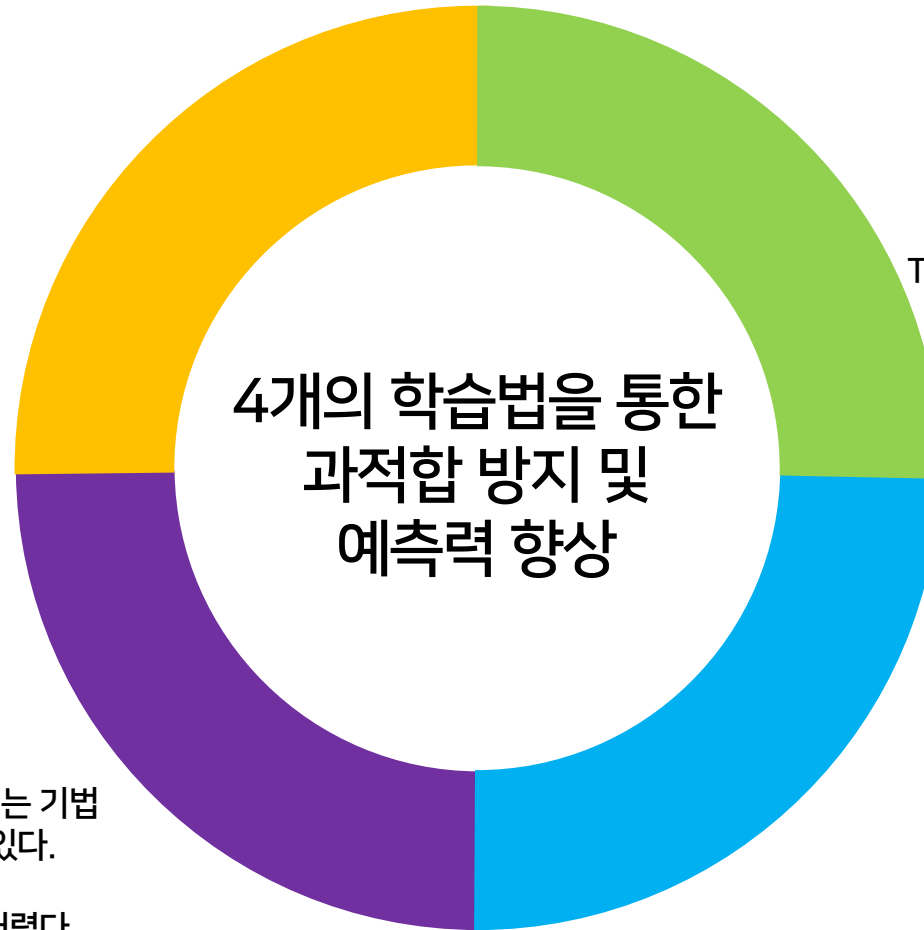
다수의 의사결정나무를 이용하는 학습법  
여러 변수를 일괄적으로 다룰 수 있다.

모델 중 메모리 소모가 가장 많으며,  
Training data 양이 증가해도 성능이 급격히 향상하지 않는다.

기존 Gradient Boosting보다 속도가 빠르며,  
과적합을 방지할 수 있어 좋은 예측 결과를 만들어낼 수 있다.

적은 데이터 세트에 적합하는 경우 과적합이 발생하기 쉽다.

## Light Gradient Boost



# Modeling : Light Gradient Boosting Model

---

	kind	lgb_mape
0	속옷	86.608507
1	잡화	52.903913
2	건강기능	44.603927
3	농수축	41.998009
4	침구	31.859083
5	생활용품	60.482734
6	주방	53.131283
7	이미용	28.949799
8	가전	62.995166
9	가구	65.480531
10	의류	49.109365

learning\_rate : GBM은 초기의 추정값에서 시작하여 각각의Tree 결과를 사용하여 추정값을 업데이트 하는 과정에서 발생하는 변화의 크기

max\_depth : 각 의사결정나무의 깊이를 설정. 트리가 깊어질수록 더 잘게 분류를 시키므로 일반적으로 정확도가 높아진다.

num\_leaves : 전체 Tree의 leave 수

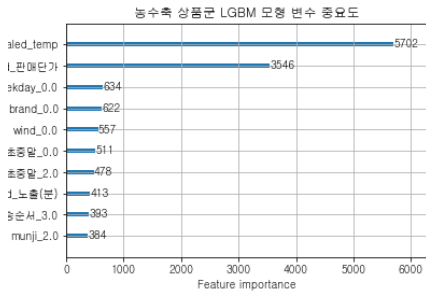
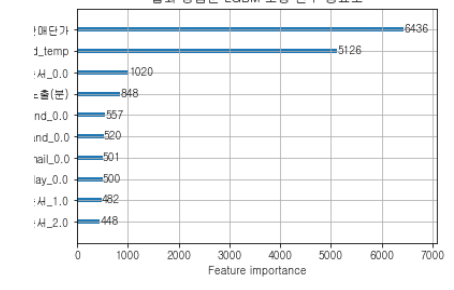
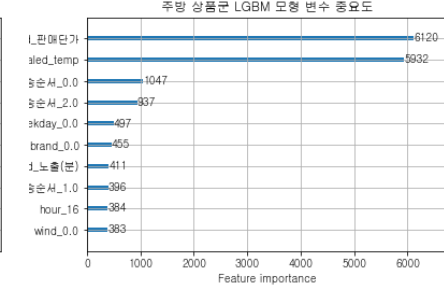
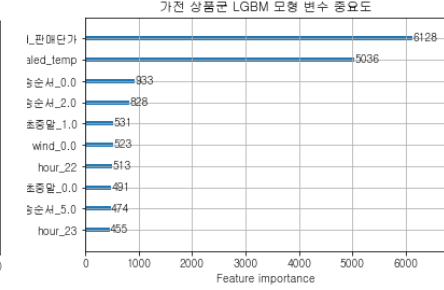
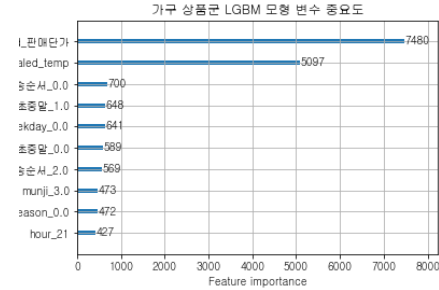
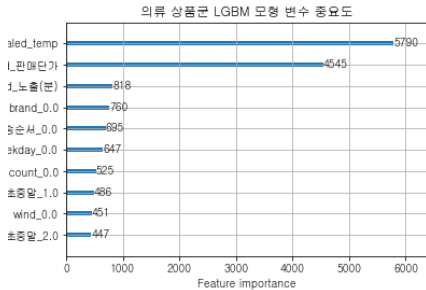
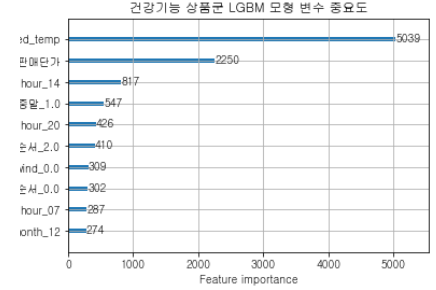
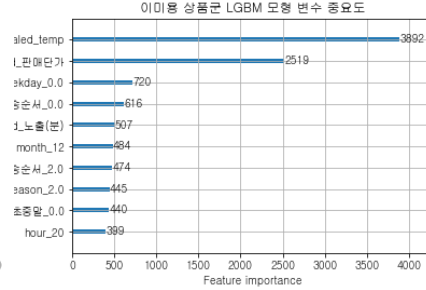
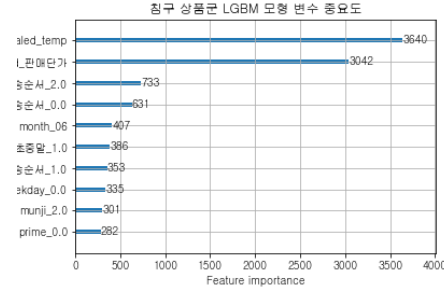
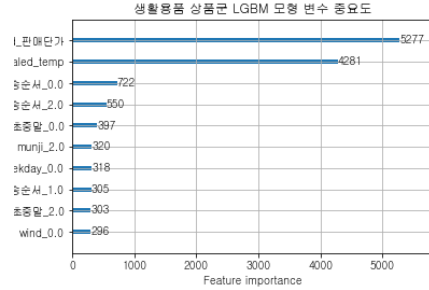
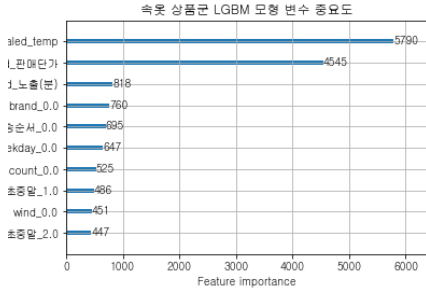
LGBM의 경우,

다수의 상품군이 50에 가깝거나 그 이하의 수치를 기록하며, 해당 상품군의 절반 이상을 예측할 수 있는 가장 좋은 성능을 보여주고 있다.

하지만 여전히 60 이상의 수치를 기록하는 상품군으로 인해, 여전히 전체적인 예측에서는 많은 어려움을 보이고 있다.

# Modeling : Final

## 각 모델별 상위 10% 중요 변수 추출



판매단가, temp(기온), 방송 순서가 모든 상품군에서 가장 중요한 변수로 나타났다.

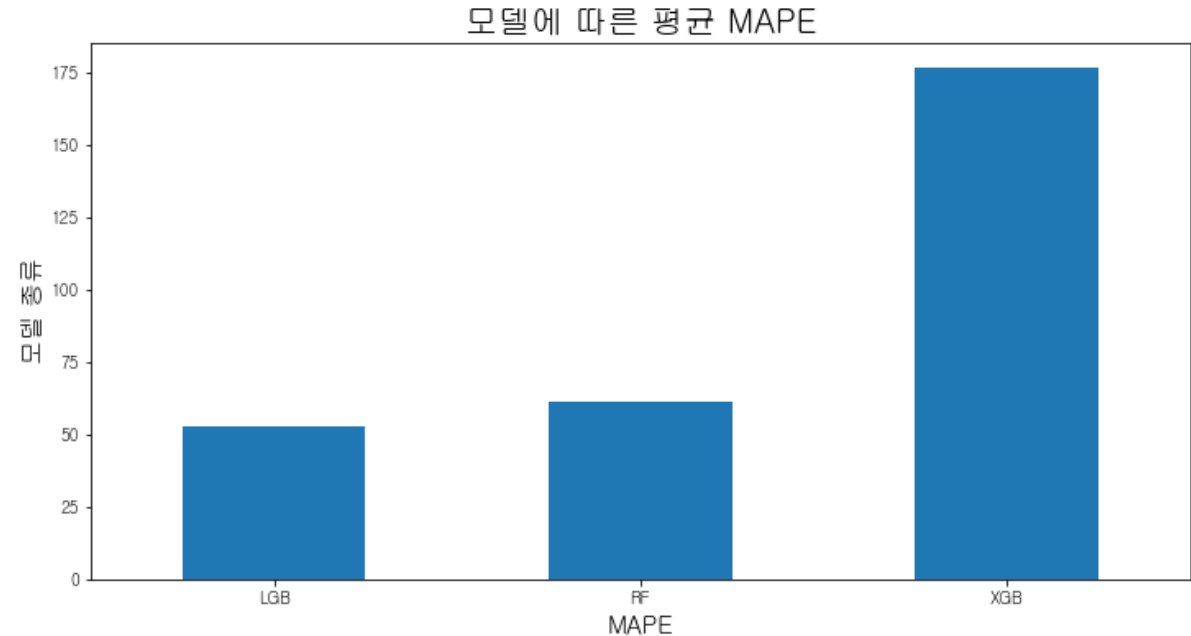
상품군별로 중요 변수의 정도가 상이했기 때문에 변수 제거 없이 모델 훈련을 진행하였다.

다수의 상품군에서 브랜드, 월초~월말 및 기상특보 관련변수 중요도 또한 높은 것으로 나타났다.

# Modeling : Final

	kind	lgb_mape	xgb_mape	rf_mape
0	속옷	86.608507	123.049550	95.983932
1	잡화	52.903913	238.515339	62.790722
2	건강기능	44.603927	60.507165	53.102343
3	농수축	41.998009	57.530119	50.972370
4	침구	31.859083	169.139073	31.360894
5	생활용품	60.482734	317.600210	78.643059
6	주방	53.131283	224.886706	73.959839
7	이미용	28.949799	40.235363	28.532725
8	가전	62.995166	274.650390	68.165376
9	가구	65.480531	385.974469	73.093836
10	의류	49.109365	49.109365	55.248364

: MAPE 최솟값



회귀모형을 제외한 나머지 3개의 모형을 사용한 결과, LGBM이 상대적으로 가장 낮은 평균 MAPE를 기록하여 좋은 성능을 가지고 있는 것으로 판단된다.

따라서, LGBM을 최종 예측 모델로 채택한다.

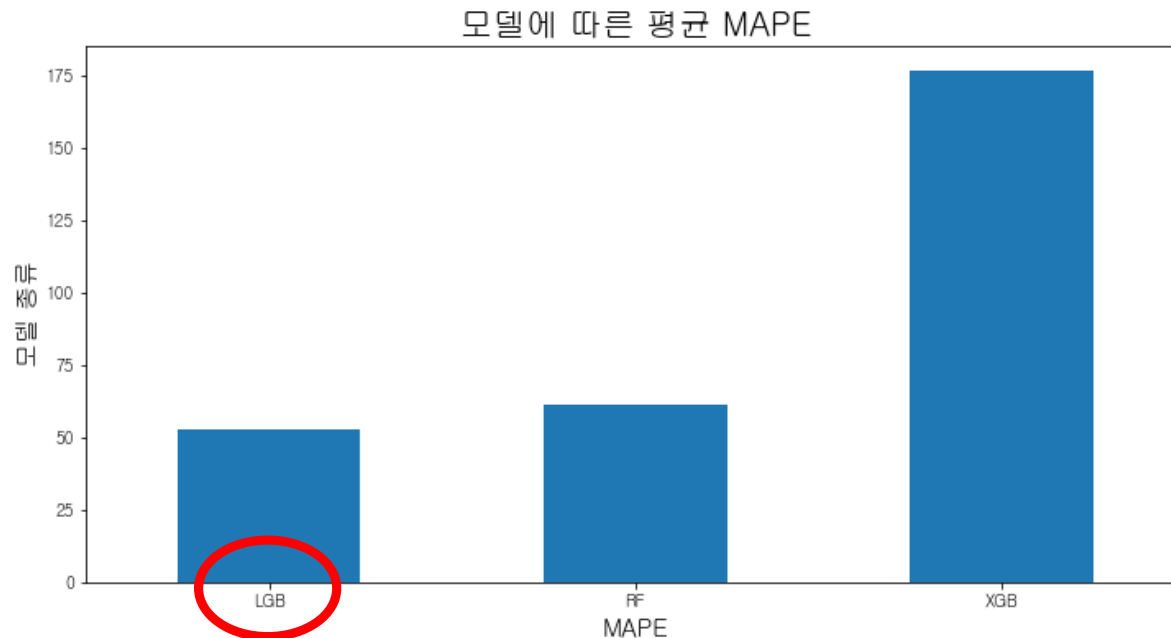
그러나 속옷, 생활용품, 가전, 가구와 같은 상품군에서는 여전히 높은 오류 수치를 기록하고 있었다.

이는 해당 상품군의 판매량 예측을 위한 추가적인 변수 생성의 필요성을 보여주고 있다.

# Modeling : Final

	kind	lgb_mape	xgb_mape	rf_mape
0	속옷	86.608507	123.049550	95.983932
1	잡화	52.903913	238.515339	62.790722
2	건강기능	44.603927	60.507165	53.102343
3	농수축	41.998009	57.530119	50.972370
4	침구	31.859083	169.139073	31.360894
5	생활용품	60.482734	317.600210	78.643059
6	주방	53.131283	224.886706	73.959839
7	이미용	28.949799	40.235363	28.532725
8	가전	62.995166	274.650390	68.165376
9	가구	65.480531	385.974469	73.093836
10	의류	49.109365	49.109365	55.248364

: 가장 낮은 수치



회귀모형을 제외한 나머지 3개의 모형을 사용한 결과, LGBM이 상대적으로 가장 낮은 평균 MAPE를 기록하여 좋은 성능을 가지고 있는 것으로 판단된다.

따라서, LGBM을 최종 예측 모델로 채택한다.

그러나 속옷, 생활용품, 가전, 가구와 같은 상품군에서는 여전히 높은 오류 수치를 기록하고 있었다.

이는 해당 상품군의 판매량 예측을 위한 추가적인 변수 생성의 필요성을 보여주고 있다.

# Algorithm

---

## 데이터 수집

### 시청률 데이터

연도별 시청률 데이터  
1분당 시청률

### 판매 데이터

상품 판매 데이터  
상품명, 판매량, 방송일자, 상품군

### 기상 데이터

기상예보 데이터  
일기예보, 특보, 미세먼지 등

변수 생성



# Algorithm

## 데이터 수집

### 시청률 데이터

연도별 시청률 데이터  
1분당 시청률

### 판매 데이터

상품 판매 데이터  
상품명, 판매량, 방송일자, 상품군

### 기상 데이터

기상예보 데이터  
일기예보, 특보, 미세먼지 등

## 변수 생성

### 시청률 변수

### 판매 변수

### 기상 변수

## 상품군에 알맞는 모델 학습

XGBoost  
RandomForest  
LGBM  
LM

시간대, 상품 및  
기상 상황에 맞춰 최적화된  
매출 예측  
+  
코로나 19 관련 소비자 심리  
(브랜드 파워)

# Algorithm

## 데이터 수집

### 시청률 데이터

연도별 시청률 데이터  
1분당 시청률

### 판매 데이터

상품 판매 데이터  
상품명, 판매량, 방송일자, 상품군

### 기상 데이터

기상예보 데이터  
일기예보, 특보, 미세먼지 등

변수 생성

시청률 변수

판매 변수

기상 변수

판매 상품

ITEM1

ITEM2

ITEM3

ITEM4

ITEM5

ITEM6

ITEM N

⋮

상품군에 알맞는  
모델 학습

XGBoost  
RandomForest  
LGBM  
LM

방송시간 추출

방송시간1

방송시간2

방송시간3

방송시간N

⋮

기상별 / 일자별 / 상품별  
예상 매출생성

시간대, 상품 및  
기상 상황에 맞춰 최적화된  
매출 예측  
+  
코로나 19 관련 소비자 심리  
(브랜드 파워)



# Algorithm

## 데이터 수집

### 시청률 데이터

연도별 시청률 데이터  
1분당 시청률

### 판매 데이터

상품 판매 데이터  
상품명, 판매량, 방송일자, 상품군

### 기상 데이터

기상예보 데이터  
일기예보, 특보, 미세먼지 등

변수 생성

시청률 변수

판매 변수

기상 변수

판매 상품

ITEM1

ITEM2

ITEM3

ITEM4

ITEM5

ITEM6

ITEM N

⋮

상품군에 알맞는  
모델 학습

XGBoost  
RandomForest  
LGBM  
LM

기상별 / 일자별 / 상품별  
예상 매출생성

시간대, 상품 및  
기상 상황에 맞춰 최적화된  
매출 예측  
+  
코로나 19 관련 소비자 심리  
(브랜드 파워)

방송시간 추출

방송시간1

방송시간2

방송시간3

방송시간N

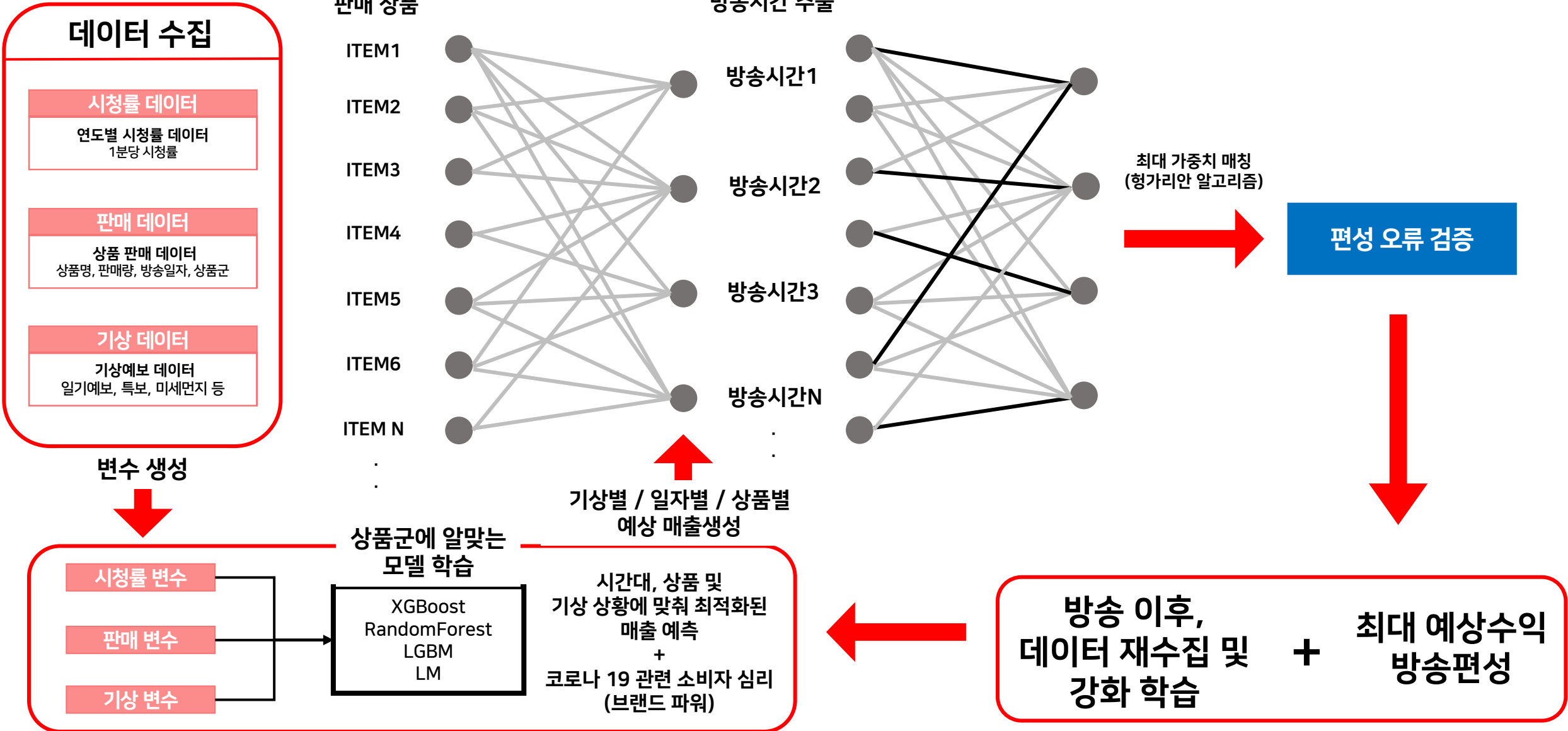
⋮

최대 가중치 매칭  
(헝가리안 알고리즘)

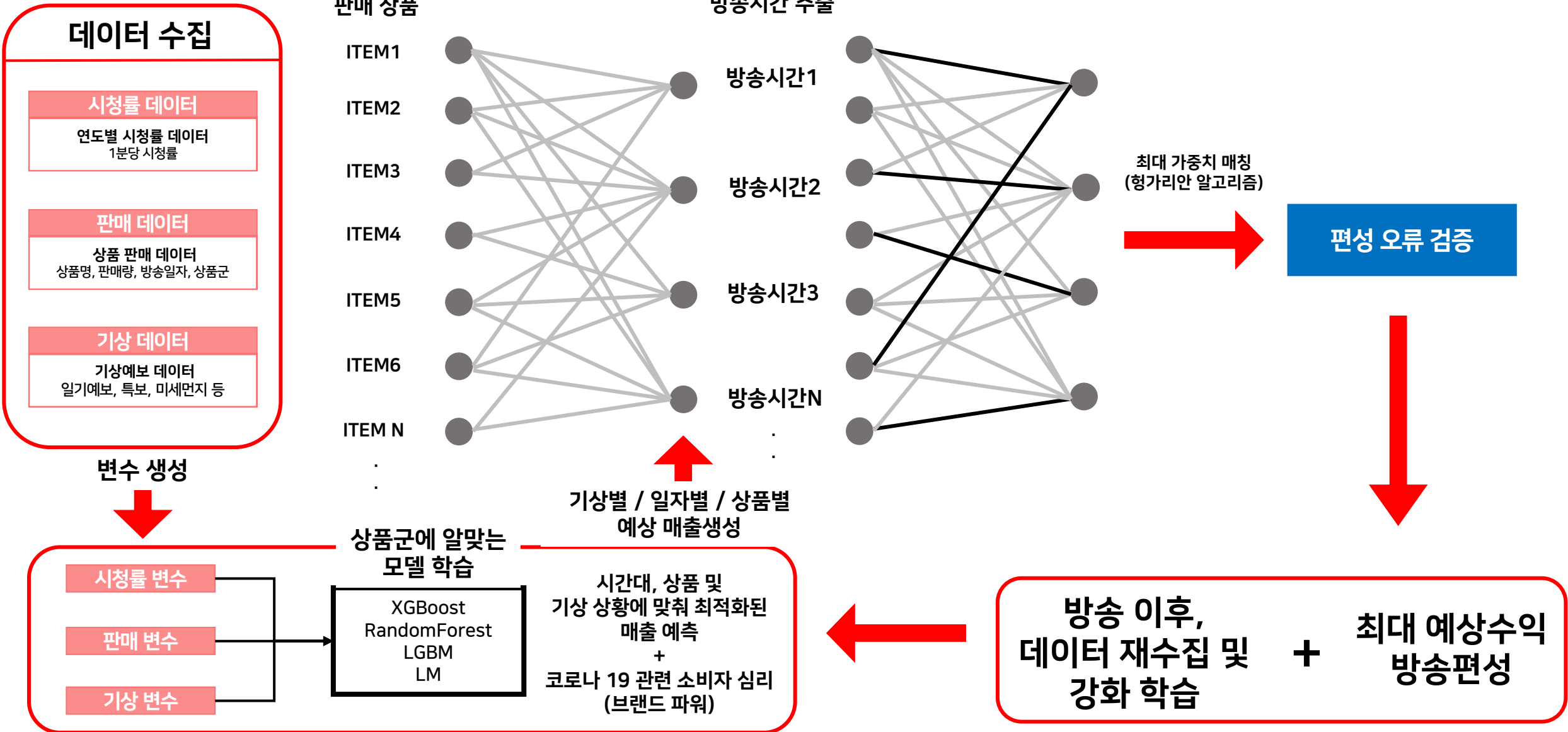
편성 오류 검증

방송 이후,  
데이터 재수집 및  
강화 학습 + 최대 예상수익  
방송편성

# Algorithm



# Algorithm



# SUMMARY

---

## 브랜드 파워의 중요성

코로나 19 이후, 홈쇼핑에서 브랜드에 대한 소비자의 인식이 더욱 강화되었음을 확인

## 외부 기상 정보의 활용

기상 변화에 따라 소비패턴이 변화할 수 있음을 확인

## 한계점

2019년의 환경으로 현재의 재난상황을 예측하는 것에 대해 제한사항 존재

고객 정보의 부재로 인한 고객 구매 행동 패턴 파악의 한계

# THANK YOU

시로앤마로