

# NS Shop+ 판매실적 예측을 통한 편성 최적화 방안 도출

2020. 09. 28

팀명: B526

신승은 ([tmddms1025@naver.com](mailto:tmddms1025@naver.com))

유채연 ([hiily96@naver.com](mailto:hiily96@naver.com))

김수현 ([tngusl95@naver.com](mailto:tngusl95@naver.com))

고유정 ([zzangu919@naver.com](mailto:zzangu919@naver.com))



# INDEX

1

Data Preprocessing

2

Modeling

3

최적화 편성 방안

4

결론

# Data Preprocessing

# I. Data Preprocessing

주제 정의 및 데이터 소개

## NS Shop+ 판매실적 예측을 통한 편성 최적화 방안 도출

- 1) 2019년 편성 데이터 활용하여 2020년 6월 프로그램 상품판매실적을 예측
- 2) 최적 수익을 고려한 요일별 / 시간대별 / 카테고리별 편성 최적화 방안 제시

# I. Data Preprocessing

## 2. 변수 생성 - 1) 범주형

변수	설명	변수	설명
month	월 (1~12)	Group0	상위카테고리 (1~12)
day	일 (1~31)	Group1	중간카테고리 (1~10)
Time	시간(6~25) 자정은 24시, 새벽 1,2시는 25시	Group2	하위카테고리 (1~10)
Week	주차(1~53)	ilsibul	상품의 일시불 여부 (1: 일시불 / 2: 무이자/ 0: otherwise)
Weekday	요일(1~7) / 월요일이 1	Holiday	휴일 여부 (휴일X: 0 / 휴일O: 1)
Cold_wave	한파주의보:1 / 없음:0	Heat_wave	폭염 주의보:1 / 폭염 경보:2 / 없음:0

# I. Data Preprocessing

## 2. 변수 생성 - 1) 범주형

- 상세한 분류를 위해서 자회사인 NS 홈쇼핑+ 카테고리 참조하여 새로운 고유코드를 부여함
- 총 코드 개수는 134개로 기존에 주어진 13개보다 확장함

예시 : 가공식품



고유 코드 부여

가공식품	간편/즉석요리	간편조리	1	1	1
가공식품	간편/즉석요리	냉장/냉동식품	1	1	2
가공식품	과자/간식	간식	1	2	1
가공식품	과자/간식	과자	1	2	2
가공식품	과자/간식	떡/빵	1	2	3
가공식품	김치/반찬	김치	1	3	1
가공식품	김치/반찬	젓갈	1	3	2
가공식품	라면/면/스프	라면/면	1	4	1
가공식품	라면/면/스프	죽/스프	1	4	2
가공식품	조미료/소스/장류	장	1	5	1
가공식품	조미료/소스/장류	조미료/양념	1	5	2
가공식품	커피/생수/음료	두유/우유	1	6	1
가공식품	커피/생수/음료	차/티백	1	6	2

# I. Data Preprocessing

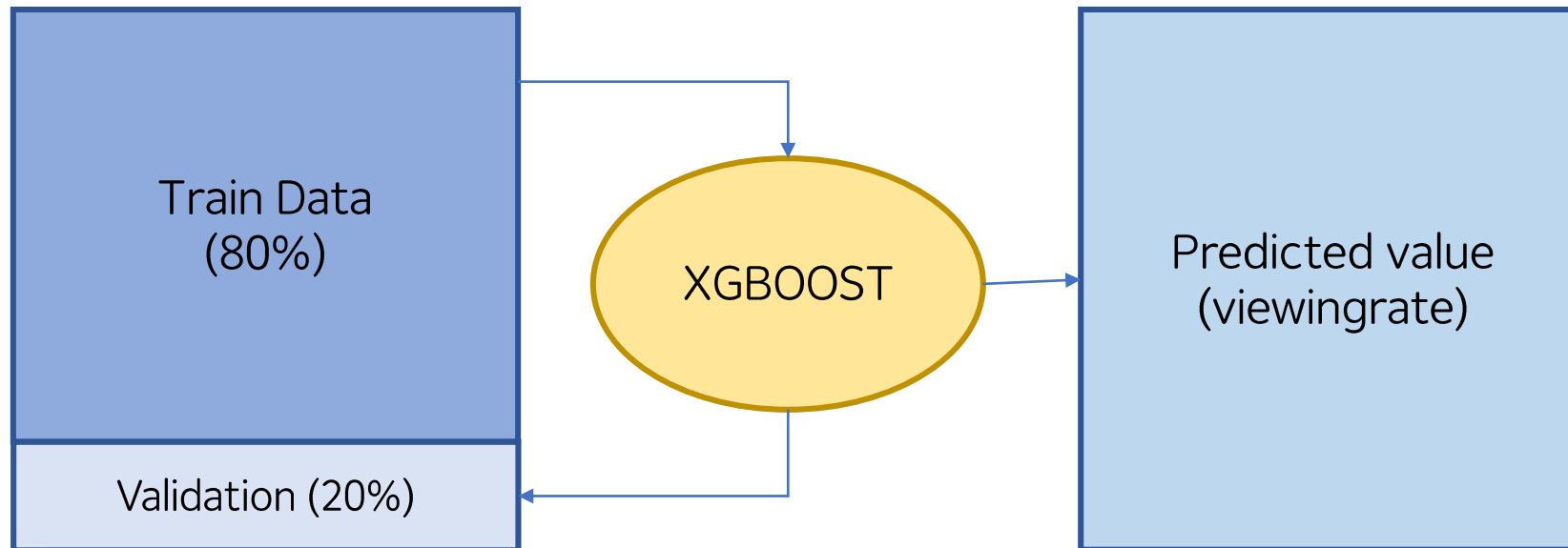
## 2. 변수 생성 - 2) 연속형

변수	설명
min	상품을 방송 시작한 시간의 분 단위
cum_expoT	각 시간, 각 상품의 노출 분수의 누적 합
unit_price	판매 단가
temp	시간당 기온
precipitation	시간당 강수량
viewingrate	홈쇼핑 채널의 분당 시청률을 편성된 시간(분)동안에 평균을 내어 기록한 값
package	2019년 전국 우체국 택배 접수 건수를 하루 단위로 합산 한 값

# I. Data Preprocessing

## 2. 변수 생성 - 2) 연속형

- 2019년 train data에는 viewingrate 변수가 존재함 (주어진 데이터)
- 2020년 test data에는 viewingrate 변수가 주어지지 않아 예측이 필요함
- 여러 모델들 중 XGBoost의 score가 제일 높은 값으로 나타나, 2019년의 시청률을 y값으로 두고 validation을 한 후, y값을 예측해 2020년 시청률을 생성함





# I. Data Preprocessing

## 2. 변수 생성 - 3) 종속변수

- 같은 마더코드를 가진 상품들을 대부분 한 시간 동안 판매함 (대부분 20분 단위로 기록되어있음)
- 노출 분수가 누적될 수록 상품의 총 매출액이 증가하는 것을 확인함
- 따라서, 상품이 한 시간 동안 판매되는 동안의 총 매출액의 누적합을 구한 것을 목표 변수로 설정함

방송일시	상품명	취급액	cumsum
2019-01-01 6:00	테이트 남성 셀린니트3종	2099000	2099000
2019-01-01 6:00	테이트 여성 셀린니트3종	4371000	4371000
2019-01-01 6:20	테이트 남성 셀린니트3종	3262000	5361000
2019-01-01 6:20	테이트 여성 셀린니트3종	6955000	11326000
2019-01-01 6:40	테이트 남성 셀린니트3종	6672000	12033000
2019-01-01 6:40	테이트 여성 셀린니트3종	9337000	20663000

# I. Data Preprocessing

## 3. 변수 정제 - 1) 범주형

- 범주형 변수를 아래의 예시처럼 one-hot encoding 방법으로 바꿈

ex) 왼쪽의 표에서 month 변수에 대한 값을 one-hot-encoding 방법으로 바꿀 시 오른쪽 표와 같음

index	month
1	1
2	2
3	3
⋮	⋮



index	1월	2월	3월	...
1	1	0	0	...
2	0	1	0	...
3	0	0	1	...
⋮	⋮	⋮	⋮	⋮

# I. Data Preprocessing

## 3. 변수 정제 - 2) 연속형

- 연속형 변수들마다 값의 범위가 다르기 때문에, 단위를 통일할 필요가 있음
- 모든 값을 0과 1사이의 값으로 만들어주는 MinMaxScaler 방법을 사용하여 정제함.

cumsum	cumsum (변환 후)
2099000	0.00471356
4371000	0.01022808
5361000	0.01263098

viewingrate	viewingrate (변환 후)
0.00055	0.00437898
0.00415	0.0330414
0.00615	0.04896497

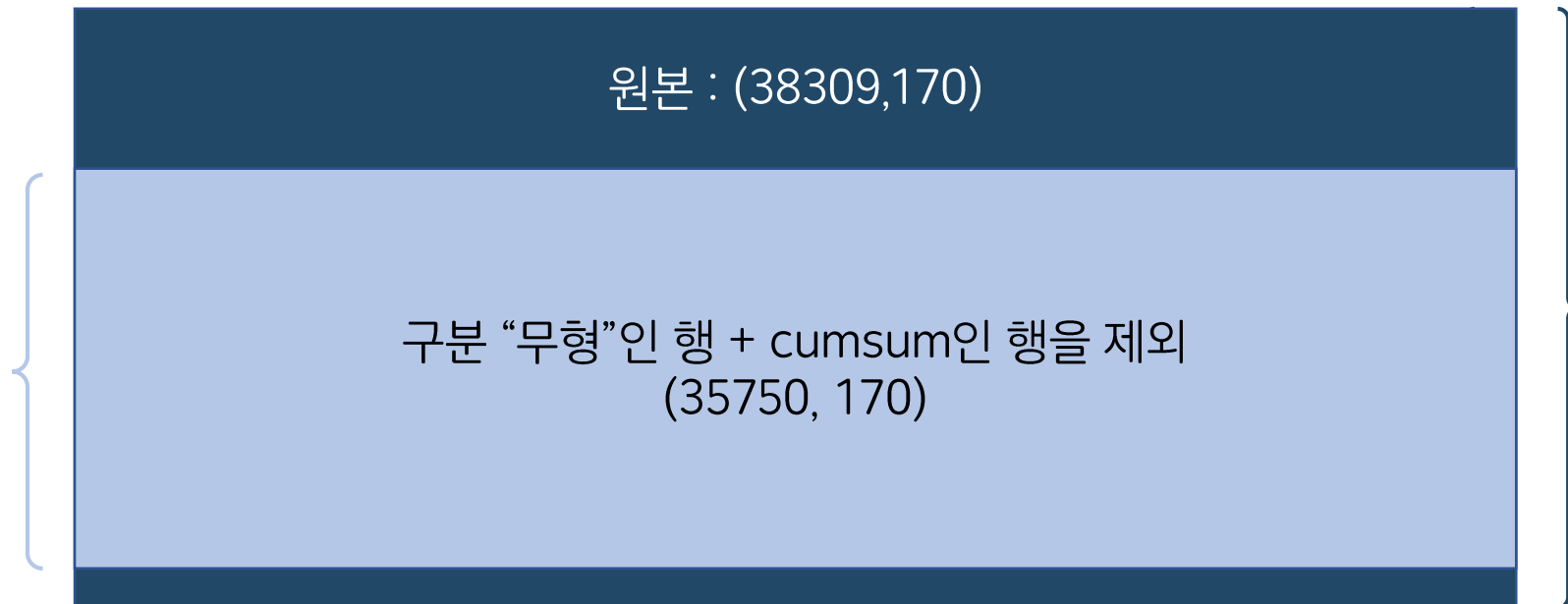
[ MinMaxScaler 계산식 ]

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

# I. Data Preprocessing

## 4. 결측값 제거

- Group0의 값이 NA인 데이터를 제거함
- MAPE 계산 시, 종속변수인 cumsum이 0이 되면 Inf 값을 나타내므로 제거함

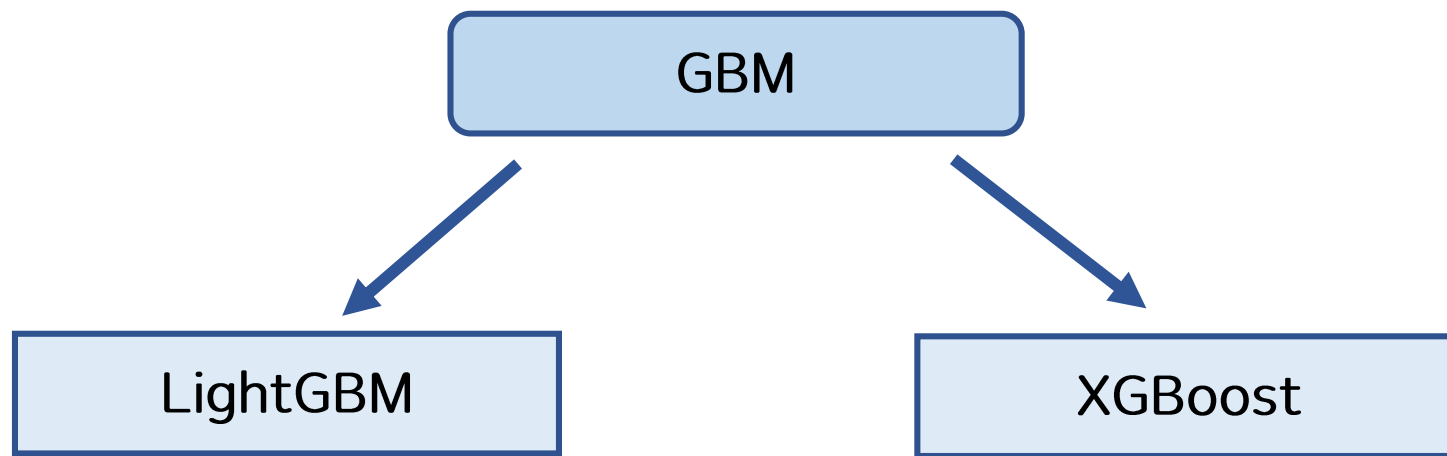


# Modeling

## 2. Modeling

### 1. 모델 선택—Gradient Boosting Model

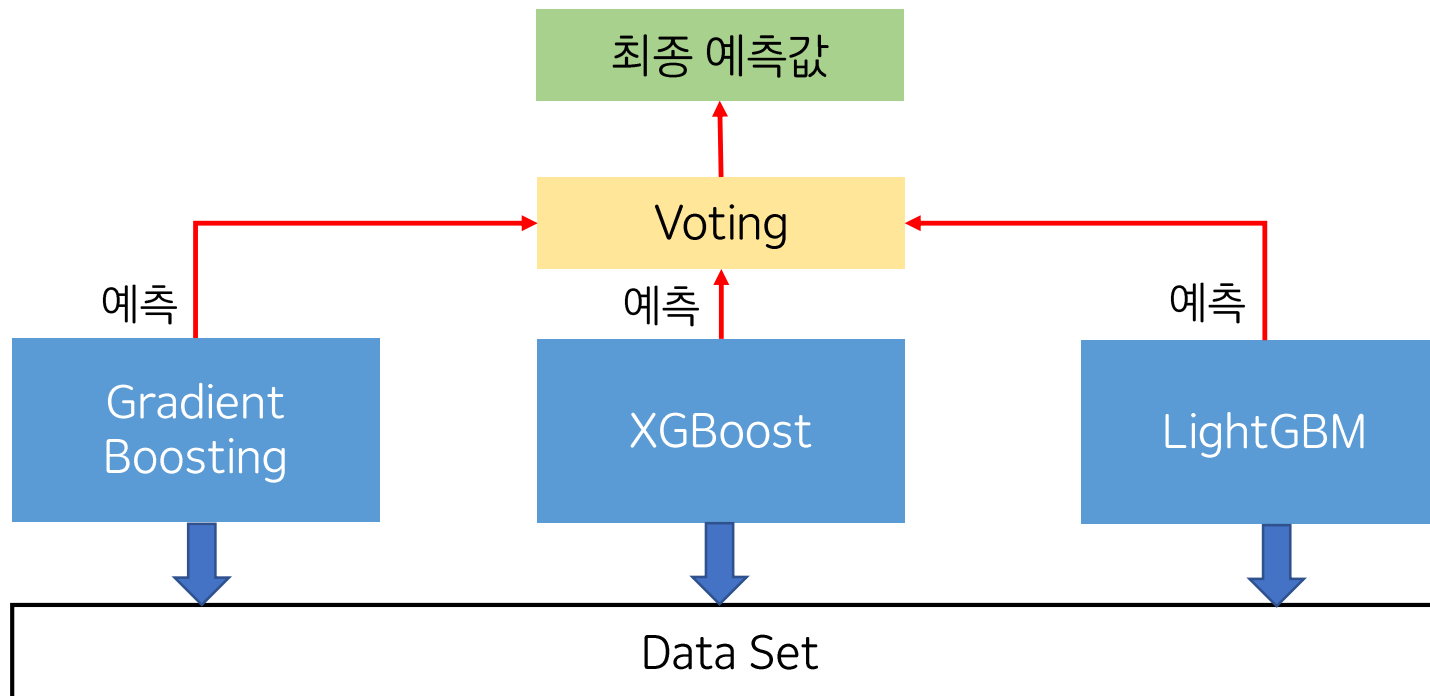
- 부스팅 계열에 속하는 알고리즘으로써 독립변수·종속변수가 있는 데이터에 대한 예측에서 좋은 성능을 보여줌
- 머신러닝 알고리즘 중에서도 가장 예측 성능이 높다고 알려져 있음
- Gradient Boosting Model(이하 GB)은 계산이 많이 필요한 알고리즘이기 때문에 효율적으로 구현하기 위한 LightGBM, XGBoost와 같은 패키지가 있음



# 2. Modeling

## 2. 모델 훈련

- GB, XGBoost, LightGBM 모델을 모두 이용하여 훈련시키고자 함
- 2개, 3개의 모델을 조합한 앙상블 모델도 함께 사용. 소프트 보팅(Soft Voting) 방식으로 앙상블함



## 2. Modeling

### 2. 모델 훈련 – 적절한 하이퍼 파라미터 찾기

- GB, XGBoost, LightGBM 각각의 모델과, 앙상블 한 모델을 최적화 하는 하이퍼 파라미터를 찾고자 함
- GridSearchCV 패키지를 사용

Model	N_estimators	Max_depth		score
GB	600	10	-	0.7481
XGBoost	500	7	Min_child_weight: 2	0.7597
LightGBM	1300	12	Num_leaves: 26	0.7751



# 2. Modeling

## 2. 모델 훈련

– 앞서 계산한 하이퍼 파라미터를 적용한 뒤 모델 훈련을 진행함

Model	모델 개수	MAPE	Score( $R^2$ )
GB	1	0.5922	0.8173
XGB	1	0.6568	0.8203
LGB	1	0.6847	0.8272
GB & XGB	2	0.5873	0.8292
GB & LGB	2	0.6030	0.8349
XGB & LGB	2	0.6332	0.8350
XGB & LGB & GB	3	0.5949	0.8368

## 2. Modeling

### 3. 스택킹

- 앙상블 모델 3개 (GB&LGB, GB&XGB, XGB&LGB)를 이용하여 Kfold기반 스택킹 모델 사용함
- Stacking된 데이터를 MAPE가 가장 낮았던 Gradient Boosting 모델을 이용하여 훈련을 진행함

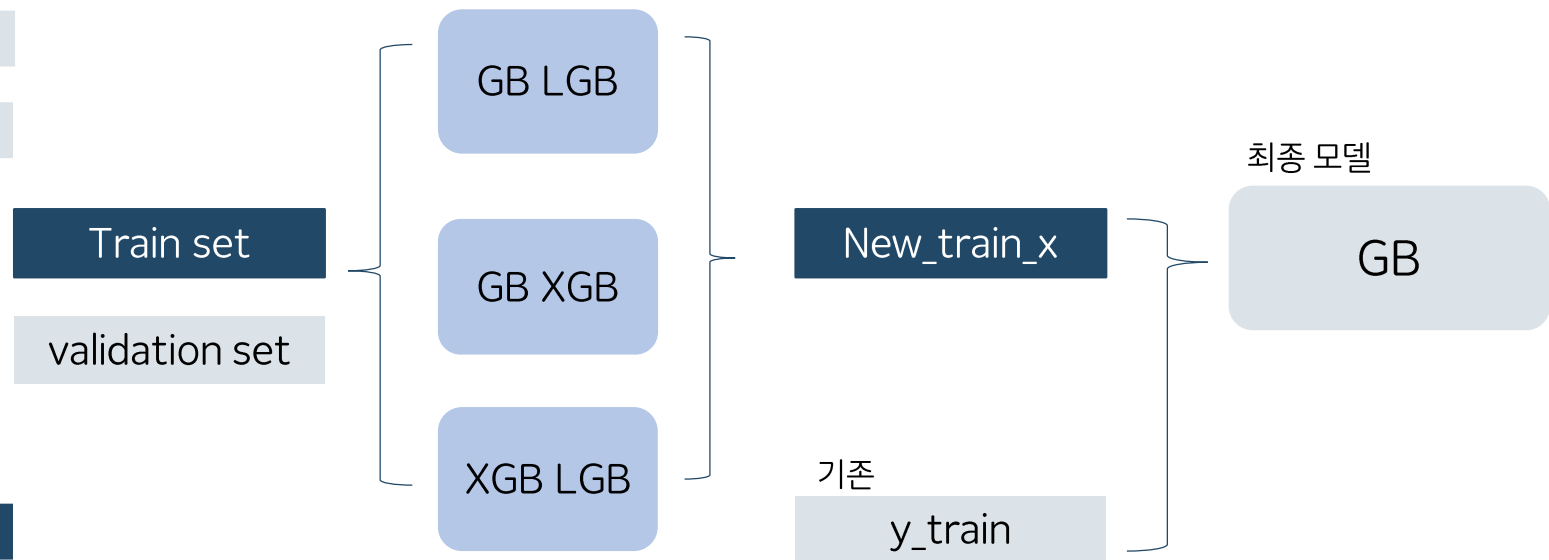
N folds = 10

1	2	3	4	5	...	10
---	---	---	---	---	-----	----

1	2	3	4	5	...	10
---	---	---	---	---	-----	----

⋮

1	2	3	4	5	...	10
---	---	---	---	---	-----	----



## 2. Modeling

### 3. 스택킹

- 앙상블 모델과 스택킹 모델의 스코어 비교

Model	Score( $R^2$ )	Stacking
GB & XGB	0.8292	0.8427
GB & LGB	0.8349	
XGB & LGB	0.8350	

- 앙상블 모델만 돌렸을 때보다 앙상블 모델로 예측한 값을 x값으로 다시 넣고 스택킹 했을 때 score가 향상됨.

# 매출 최적화 방안

# 3. 매출 최적화 방안

## 요일별 & 시간별 최적화 편성 방안

– 요일별 및 시간별 최대 매출을 고려한 최적 편성 방안을 만들기 위해 새로운 데이터 생성함

1) 요일 별로 총 매출액의 중앙값을 구하고, 그 값을 가진 가장 가까운 날짜를 구함

2) 요일 별로 모든 시간대를 고려함

3) 단가는 각 그룹 코드별 상품가격의 중앙값을 부여함

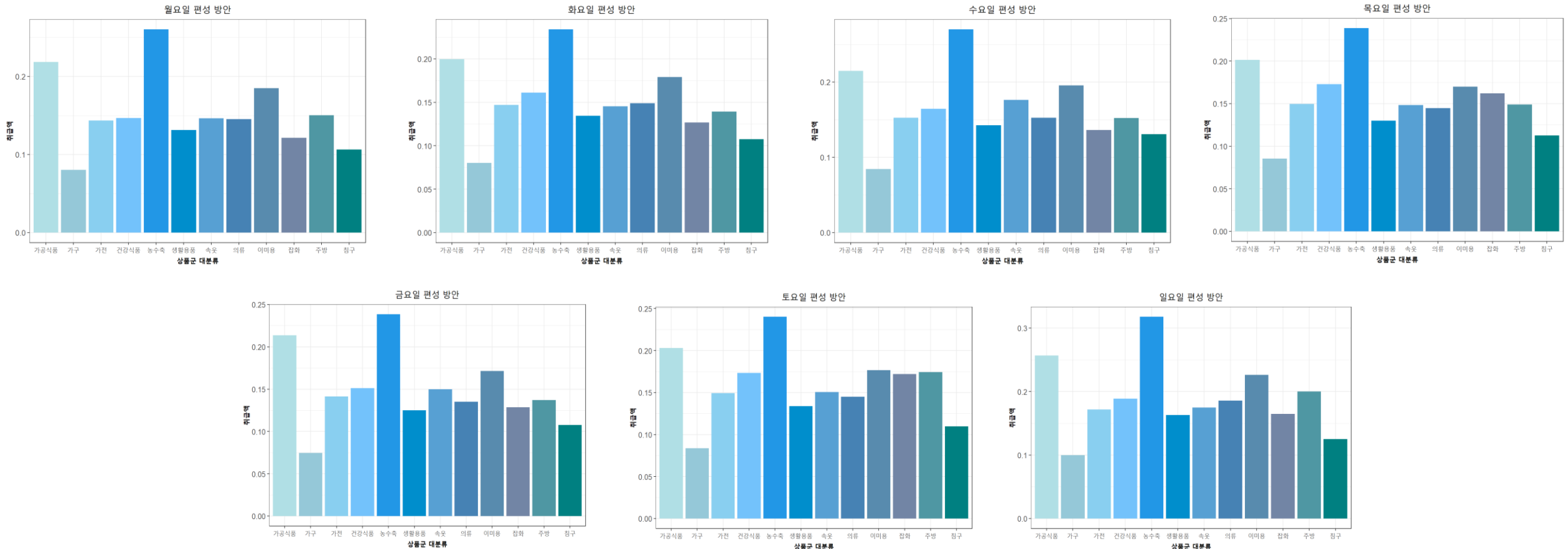
day	month	time	Week	weekday	Holiday	Cum_expoT	group0	group1	group2	...	가격	
29	4	6	18	1	0	60	1	1	1	...		
29	4	6	18	1	0	60	1	1	2	...		
29	4	6	18	1	0	60	1	2	1	...		
⋮	⋮	⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮		
13	1	25	7	7	1	60	12	2	1	...		
13	1	25	7	7	1	60	12	3	1	...		

(18760, 170)

# 3. 매출 최적화 방안

## 1. 요일별

- 요일별로 그래프를 그렸을때, 그 차이가 크게 다르지 않음
- 가공식품과 농수축의 취급액이 제일 높고, 속옷과 침구가 낮았음

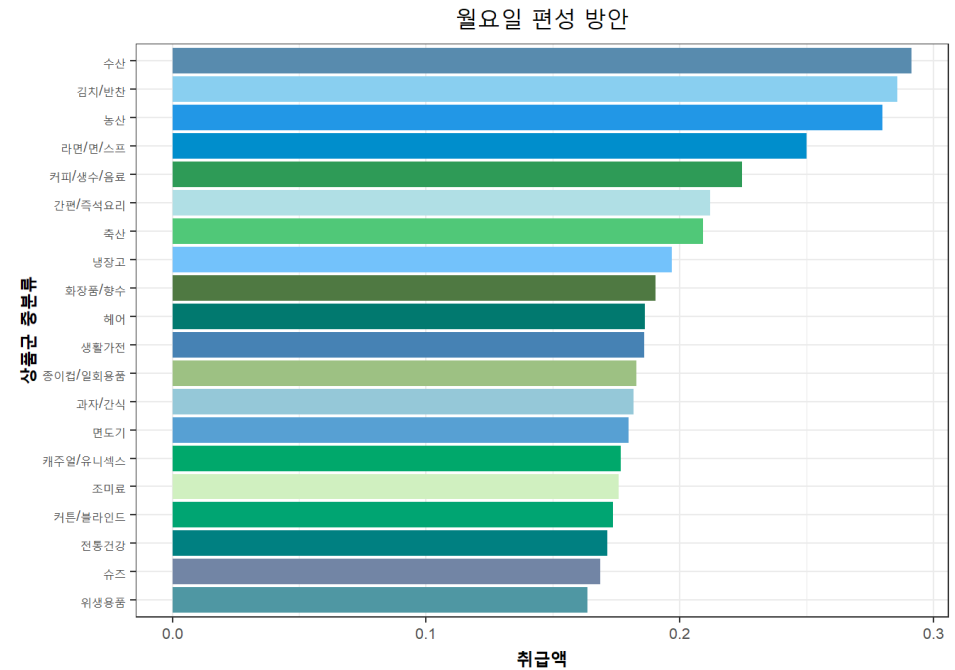


### 3. 매출 최적화 방안

#### 1. 요일별 - 월요일

매출 순위			
1	농수축	7	의류
2	가공식품	8	가전
3	이미용	9	생활용품
4	주방	10	잡화
5	건강식품	11	침구
6	속옷	12	가구

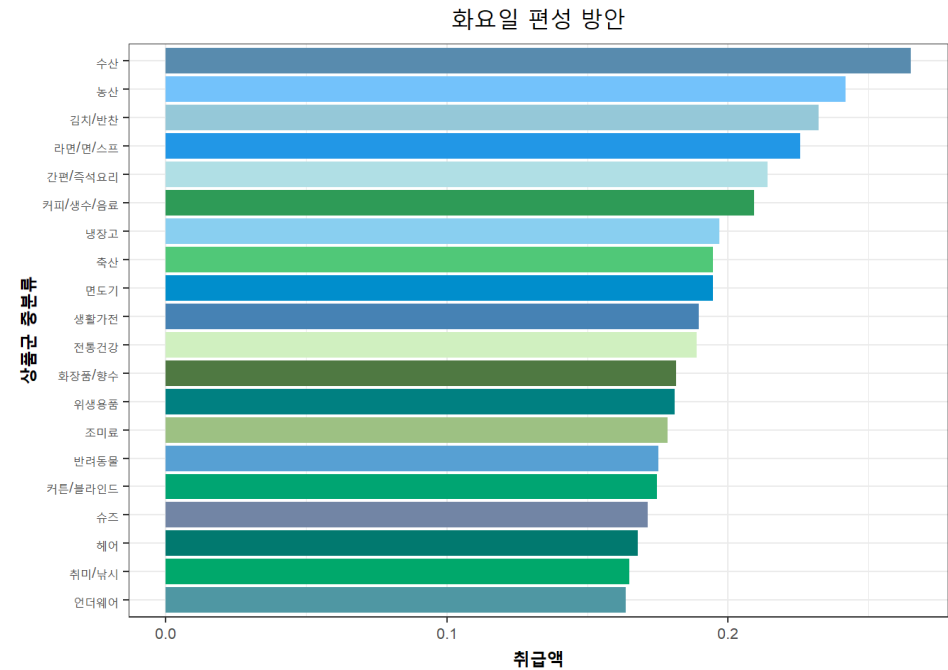
- 농수축과 가공식품의 취급액 순위는 꾸준히 높음
- 타 요일에 비해 냉장고의 취급액 순위가 높음



# 3. 매출 최적화 방안

## 1. 요일별 - 화요일

매출 순위			
1	농수축	7	속옷
2	가공식품	8	주방
3	이미용	9	생활용품
4	건강식품	10	잡화
5	의류	11	침구
6	가전	12	가구



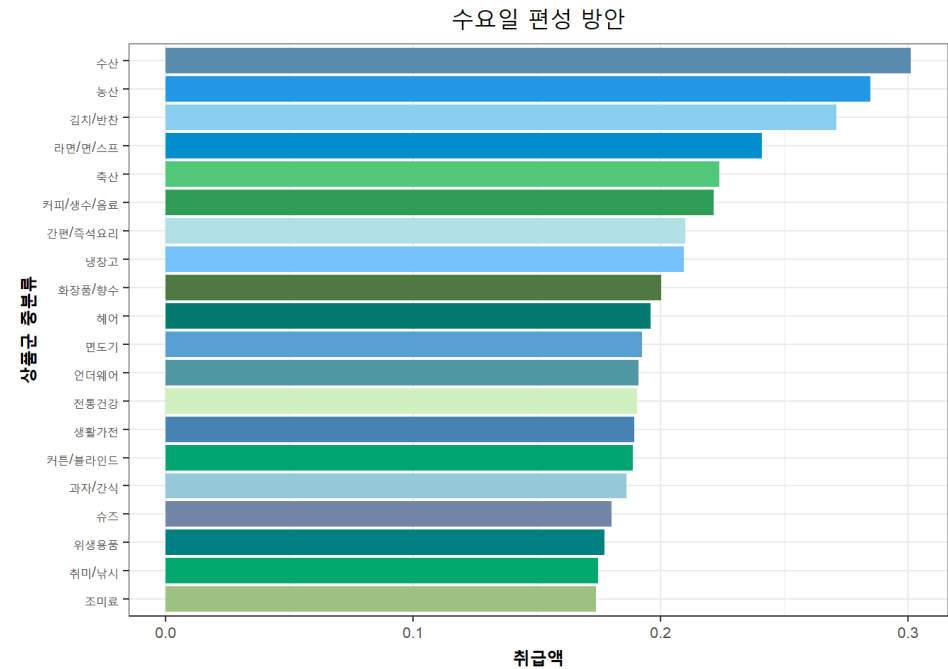
- 전통건강식품(선식, 꿀, 홍삼 등)의 취급액 순위가 타 요일에 비해 높음
- 월요일에 비해 위생용품(바디케어, 샴푸/트리트먼트)의 취급액 순위가 높음



### 3. 매출 최적화 방안

#### 1. 요일별 - 수요일

매출 순위			
1	농수축	7	가전
2	가공식품	8	주방
3	이미용	9	생활용품
4	속옷	10	잡화
5	건강식품	11	침구
6	의류	12	가구

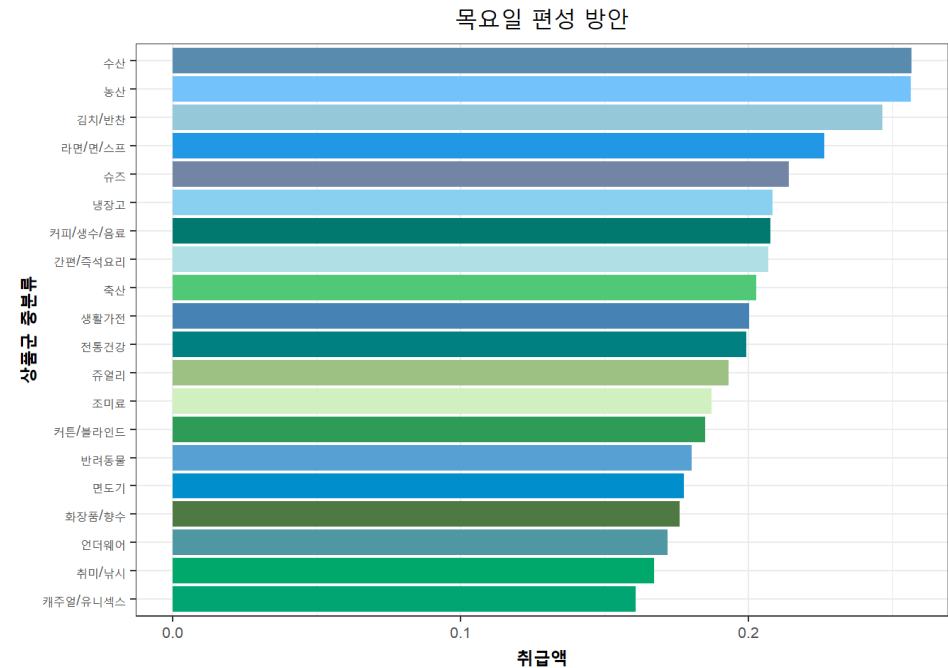


— 농수축 상품 중 축산이 취급액이 다른 요일에 비해 취급액 순위가 높음

# 3. 매출 최적화 방안

## 1. 요일별 - 목요일

매출 순위			
1	농수축	7	가전
2	가공식품	8	속옷
3	건강식품	9	의류
4	이미용	10	생활용품
5	잡화	11	침구
6	주방	12	가구

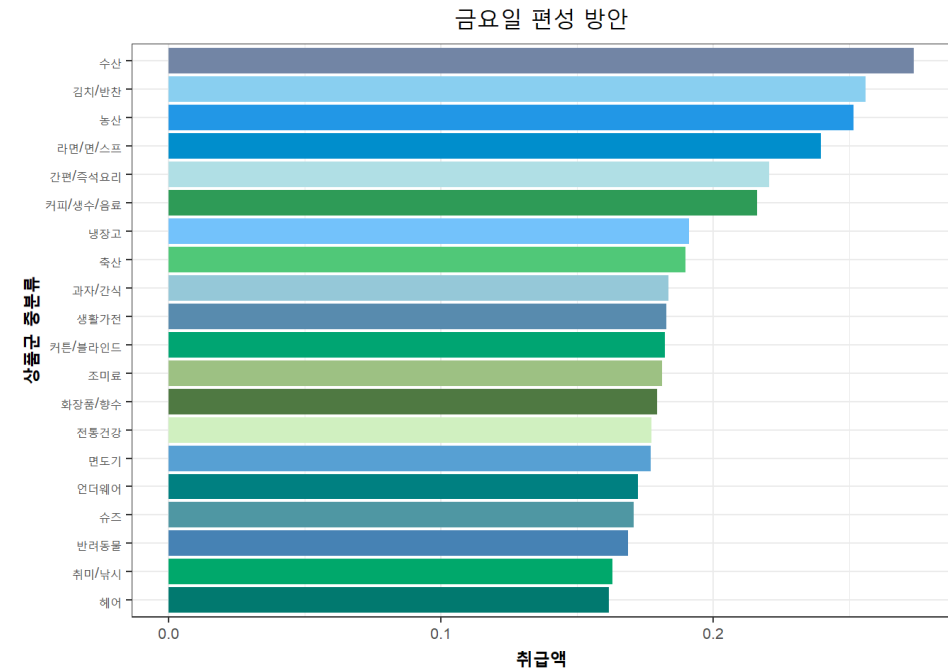


- 냉장고와 신발에 대한 취급액 순위가 다른 요일에 비해 높음
- 다른 요일에 비해 의류 매출이 감소하고, 그 중 캐주얼/유니섹스 상품의 취급액 순위가 가장 크게 하락함
- 다른 요일에 비해 반려동물 관련 상품과 주얼리 상품의 취급액 순위가 상승함

# 3. 매출 최적화 방안

## 1. 요일별 - 금요일

매출 순위			
1	농수축	7	주방
2	가공식품	8	의류
3	이미용	9	잡화
4	건강식품	10	생활용품
5	속옷	11	침구
6	가전	12	가구

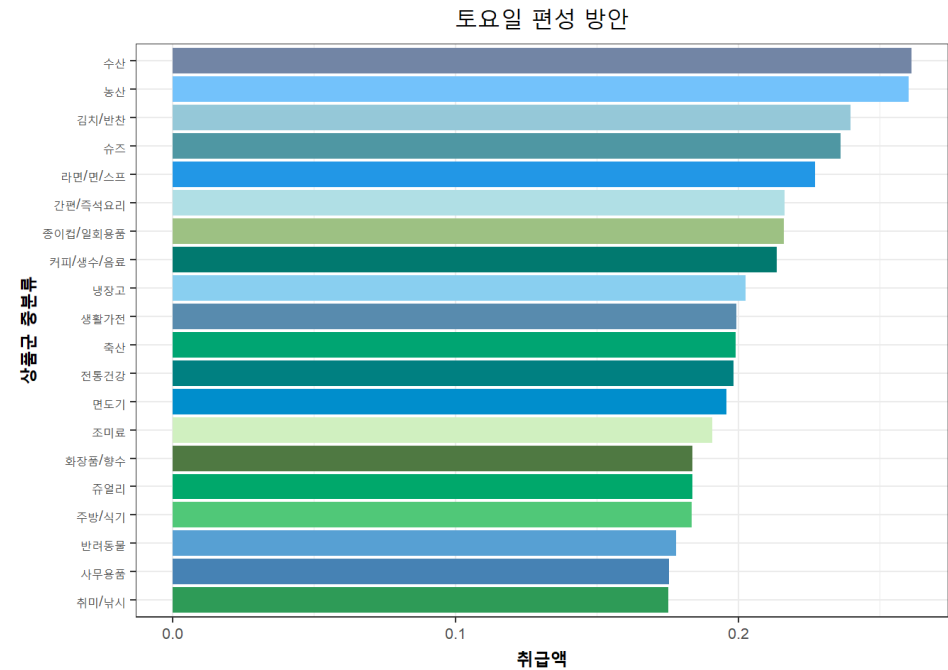


- 신발 상품의 취급액 순위가 다른 요일에 비해 낮음

### 3. 매출 최적화 방안

#### 1. 요일별 - 토요일

매출 순위			
1	농수축	7	속옷
2	가공식품	8	가전
3	이미용	9	의류
4	주방	10	생활용품
5	건강식품	11	침구
6	잡화	12	가구

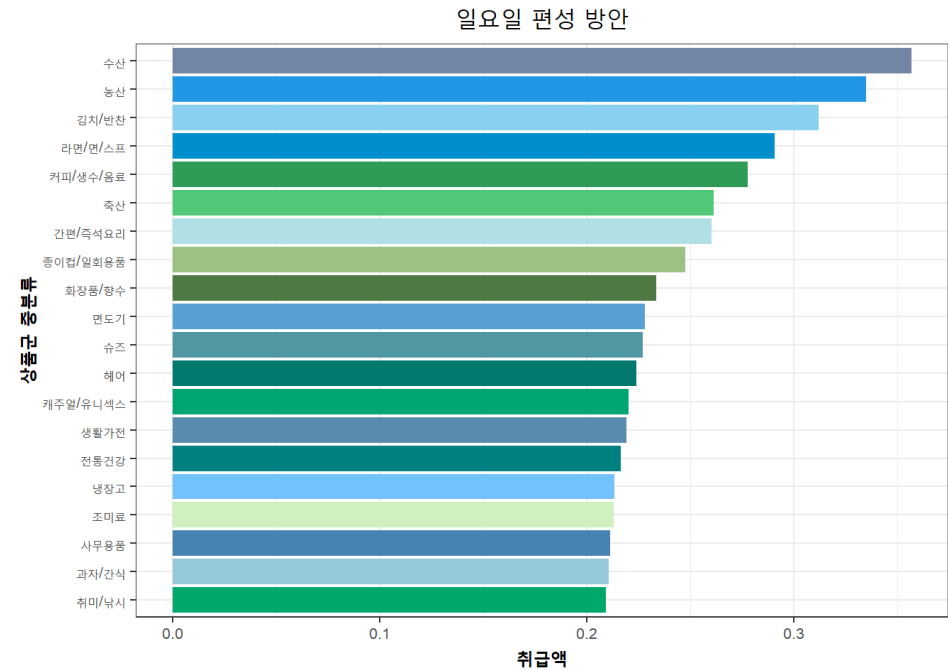


- 다른 요일에 비해 신발 취급액의 순위가 높게 나타남
- 종이컵/일회용품(ex. 행주티슈)의 취급액 순위가 다른 요일에 비해 높게 나타남

# 3. 매출 최적화 방안

## 1. 요일별 - 일요일

매출 순위			
1	농수축	7	속옷
2	가공식품	8	가전
3	이미용	9	생활용품
4	주방	10	잡화
5	건강식품	11	침구
6	의류	12	가구



- 전반적인 취급액 흐름과 크게 다르지 않음
- 토요일 만큼은 아니지만 종이컵/일회용품 취급액 순위가 높은 편임

# 3. 매출 최적화 방안

## 2. 시간별 최적화 편성 방안

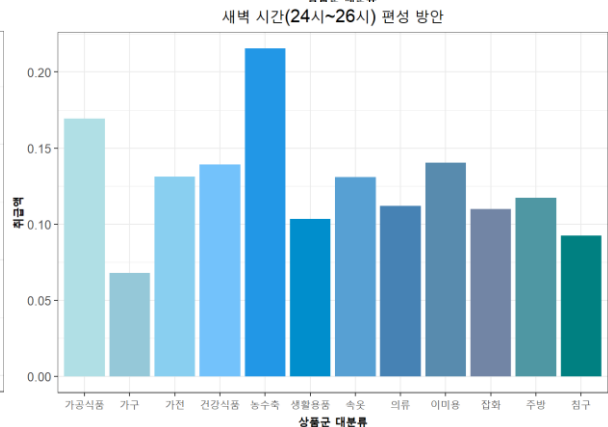
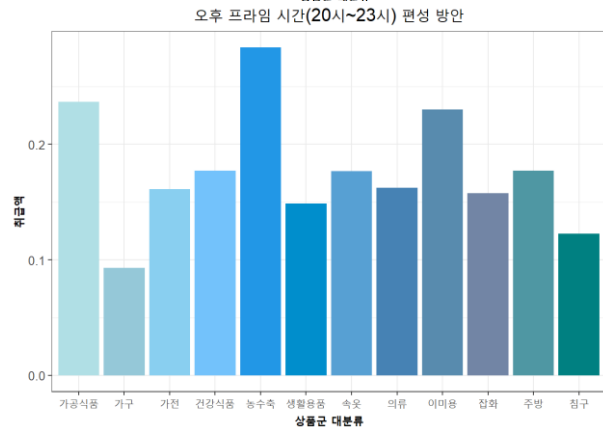
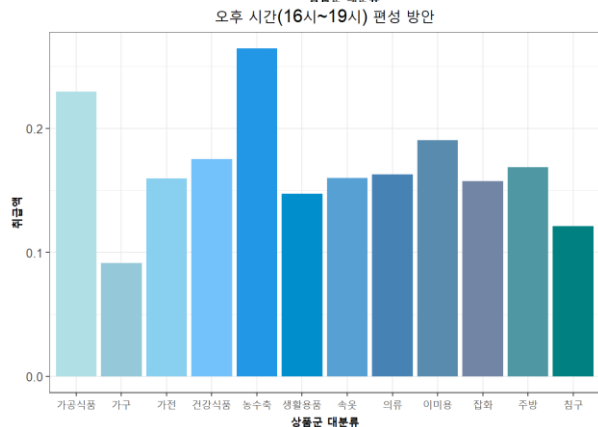
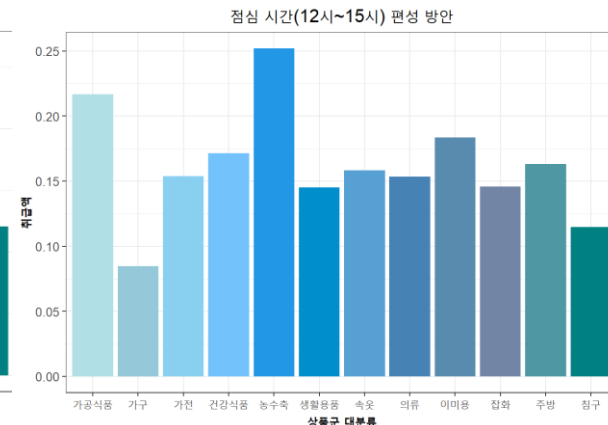
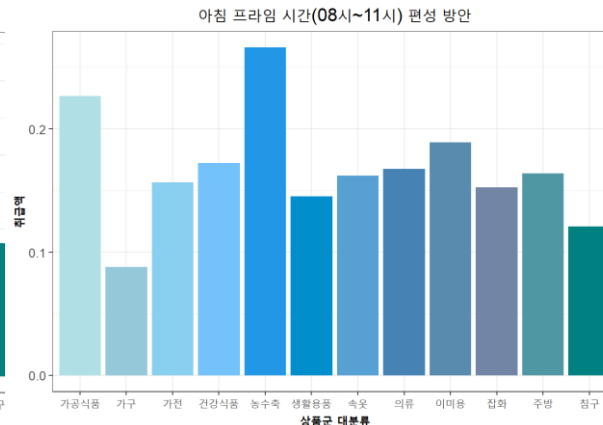
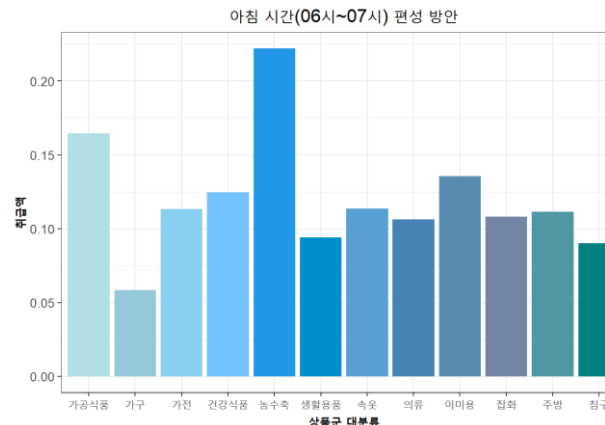
- 시간대별 최적화 편성 방안은 앞의 요일별 최적화 편성 방안과 동일하게 진행함
- 시간대별 최적화 편성 방안은 아래와 같이 시간대를 나누어 분석함

time	시간대
6~8	아침
8~11	아침 PRIME
12~15	점심
16~19	오후
20~23	오후 PRIME
24~25	새벽

# 3. 매출 최적화 방안

## 2. 시간별 - 시간대별 취급액 차이

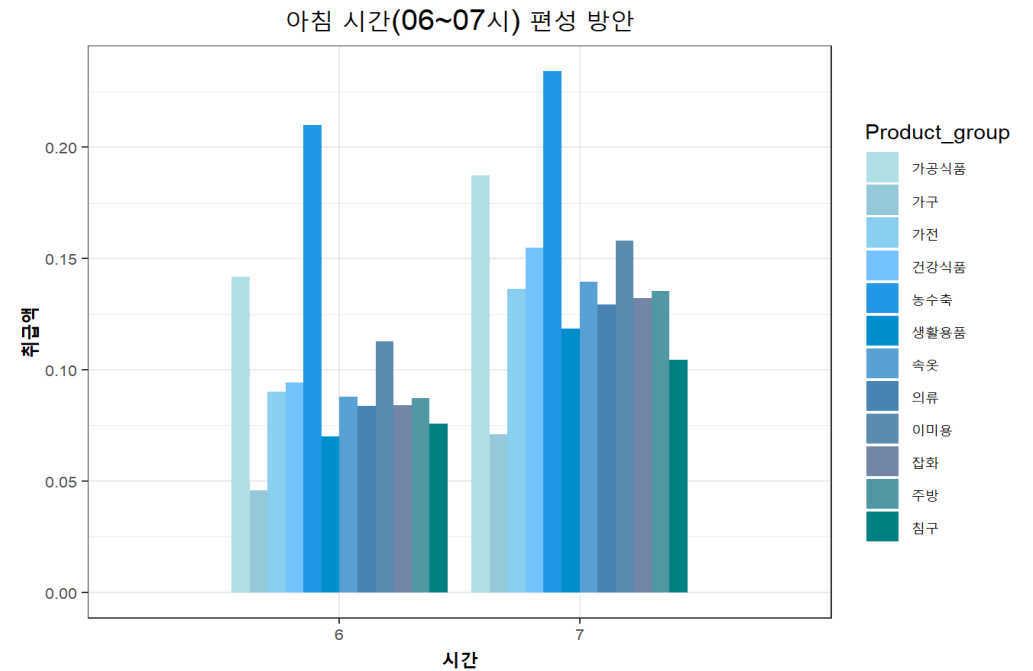
- 시간대별 취급액의 차이는 크게 다르지 않음
- 농수축이 제일 높았고 가구가 가장 낮았음



### 3. 매출 최적화 방안

#### 2. 시간별 - 아침

매출 순위			
1	농수축	7	주방
2	가공식품	8	잡화
3	이미용	9	의류
4	건강식품	10	생활용품
5	가전	11	침구
6	속옷	12	가구



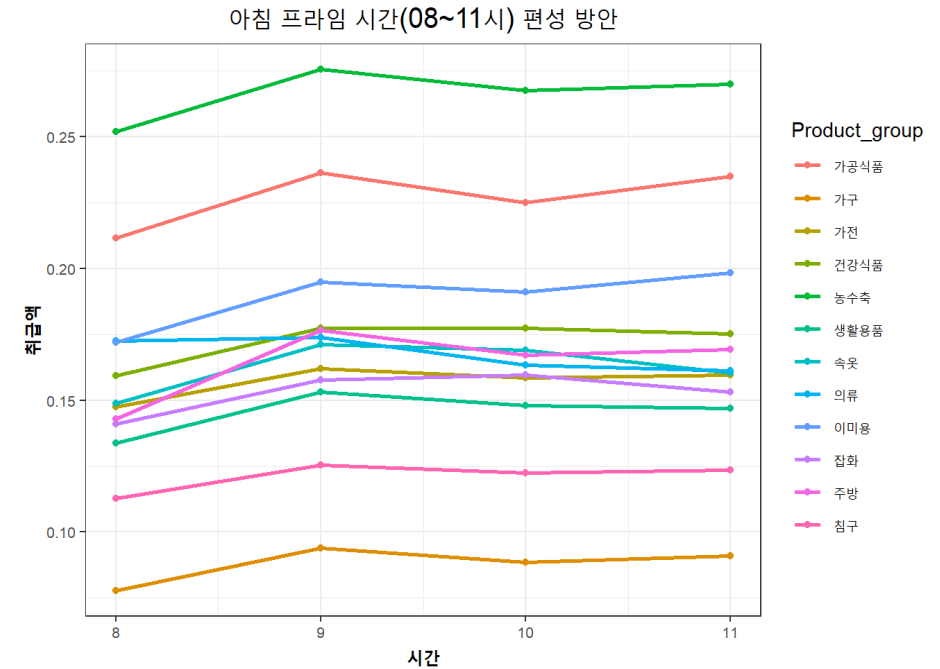
- 시간이 지남에 따라 모든 상품군의 취급액이 뚜렷하게 증가하는 추세를 보임



# 3. 매출 최적화 방안

## 2. 시간별 - 아침 PRIME

매출 순위			
1	농수축	7	속옷
2	가공식품	8	가전
3	이미용	9	잡화
4	건강식품	10	생활용품
5	가전	11	침구
6	속옷	12	가구

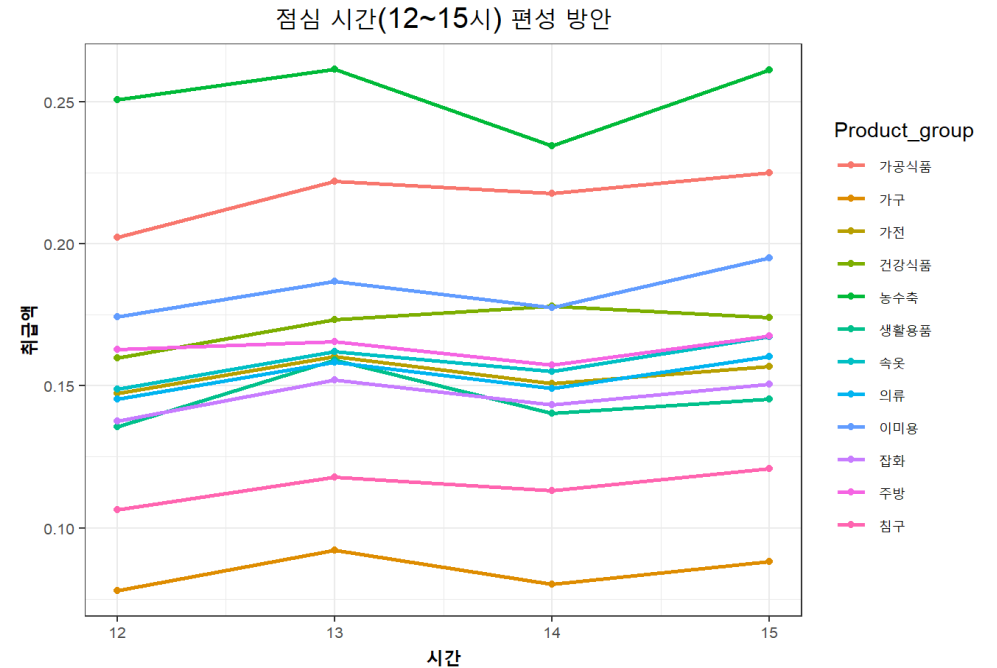


- 대부분의 상품군이 증가와 감소를 반복하는 추세를 보임
- 건강식품, 의류, 생활용품, 속옷은 9시 이후 감소 추세를 보임

# 3. 매출 최적화 방안

## 2. 시간별 - 점심

매출 순위			
1	농수축	7	의류
2	가공식품	8	가전
3	이미용	9	잡화
4	건강식품	10	생활용품
5	주방	11	침구
6	속옷	12	가구

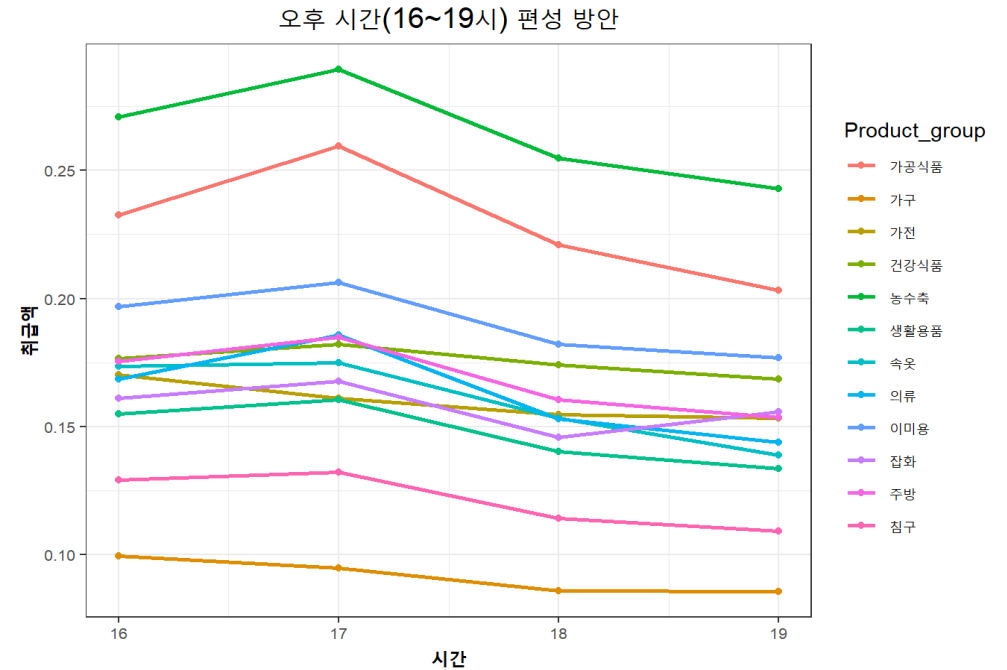


- 대부분의 상품군이 증가와 감소를 반복하는 추세를 보임
- 건강식품은 14시까지 증가했다가 그 이후부터 감소하는 추세를 보임

# 3. 매출 최적화 방안

## 2. 시간별 - 오후

매출 순위			
1	농수축	7	속옷
2	가공식품	8	가전
3	이미용	9	잡화
4	건강식품	10	생활용품
5	주방	11	침구
6	의류	12	가구

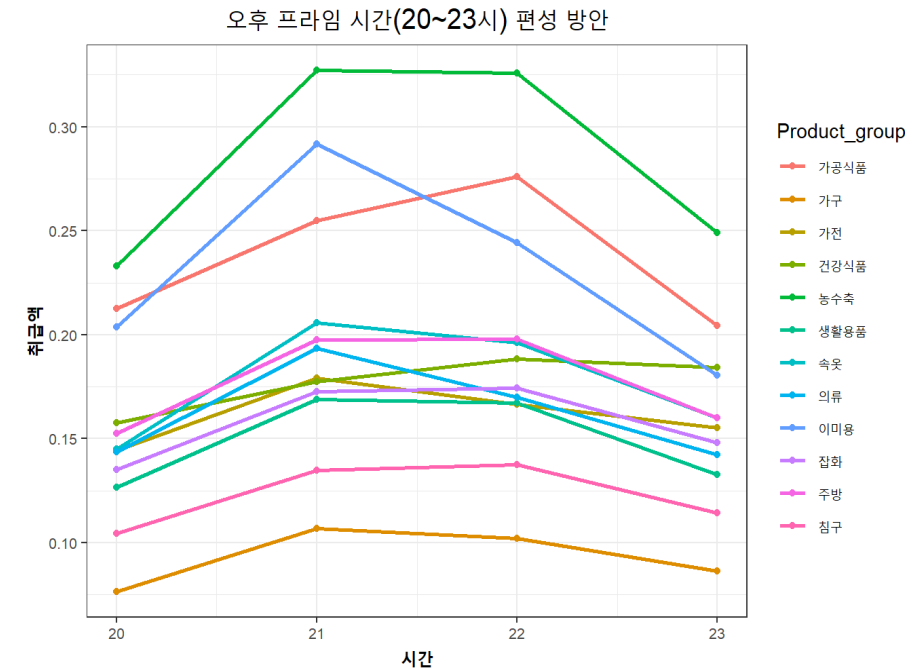


- 17시에서 대부분의 상품군의 취급액이 가장 높고 그 이후 부터는 감소하는 추세를 보임
- 다만, 잡화는 18시 이후에도 증가하는 추세를 보임
- 가구와 가전은 16시 이후로 계속해서 감소하는 추세를 보임

# 3. 매출 최적화 방안

## 2. 시간별 - 오후 PRIME

매출 순위			
1	농수축	7	의류
2	가공식품	8	가전
3	이미용	9	잡화
4	주방	10	생활용품
5	건강식품	11	침구
6	속옷	12	가구

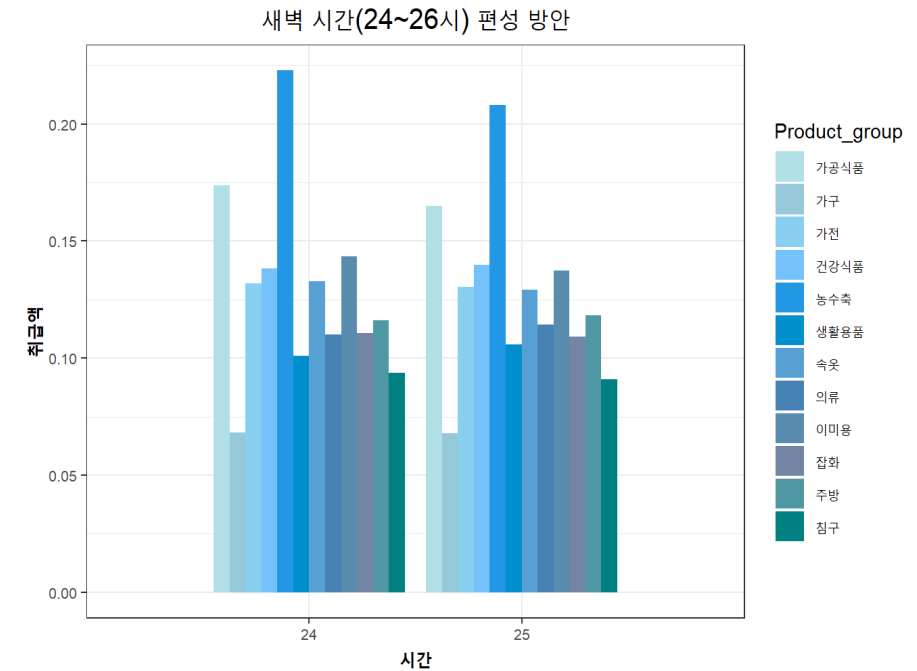


- 21시와 22시가 모든 상품군에서 높은 취급액을 가짐
- 이 시간대에 특히 가공식품의 증가세와 이미용의 감소세가 제일 뚜렷하게 나타남

### 3. 매출 최적화 방안

#### 2. 시간별 - 새벽

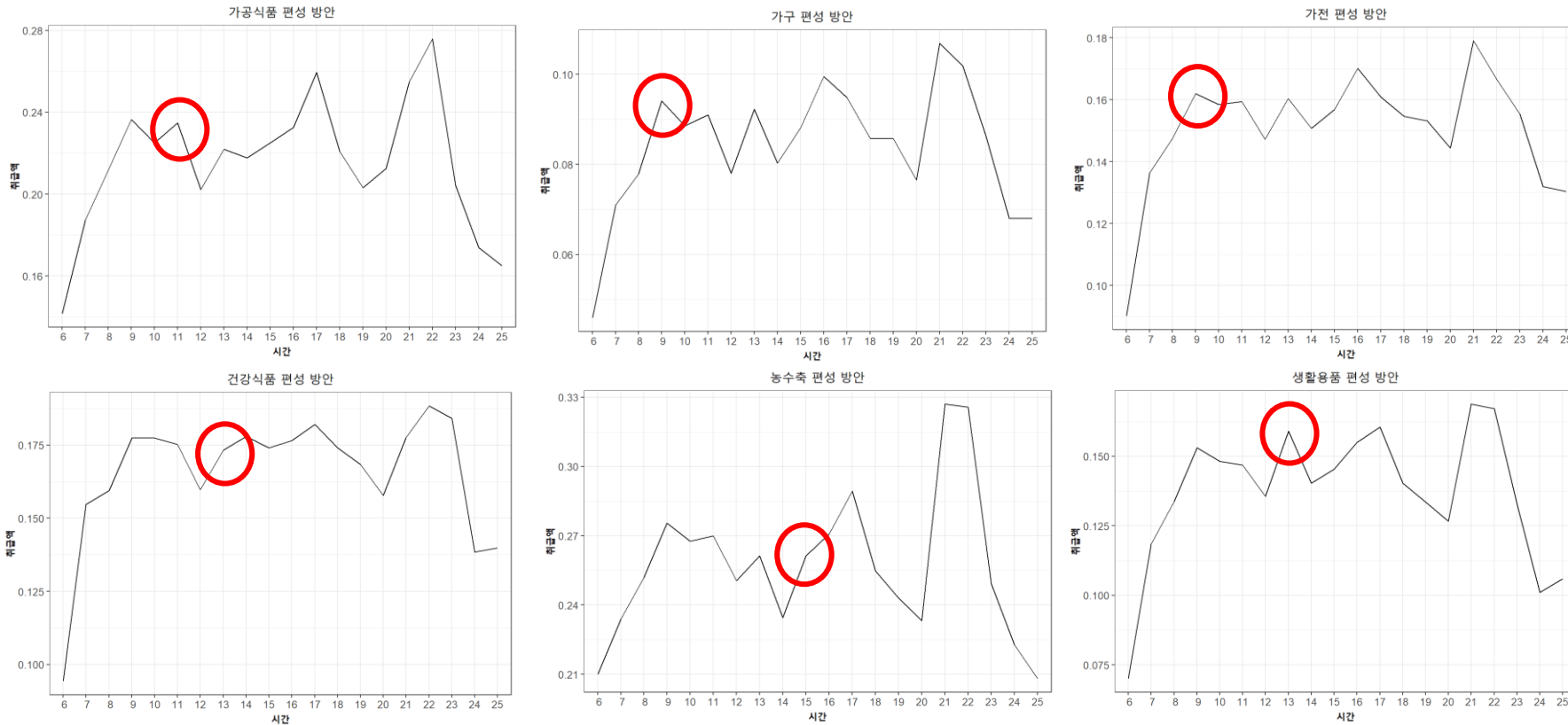
매출 순위			
1	농수축	7	주방
2	가공식품	8	의류
3	이미용	9	잡화
4	건강식품	10	생활용품
5	가전	11	침구
6	속옷	12	가구



– 대부분 상품군이 감소하거나 소폭 증가하였지만, 농수축 상품의 취급액은 감소가 더 뚜렷함

# 3. 매출 최적화 방안

## 3. 카테고리별



[프라임 시간대를 제외한 상품군별 최고 매출 증가 시간대]

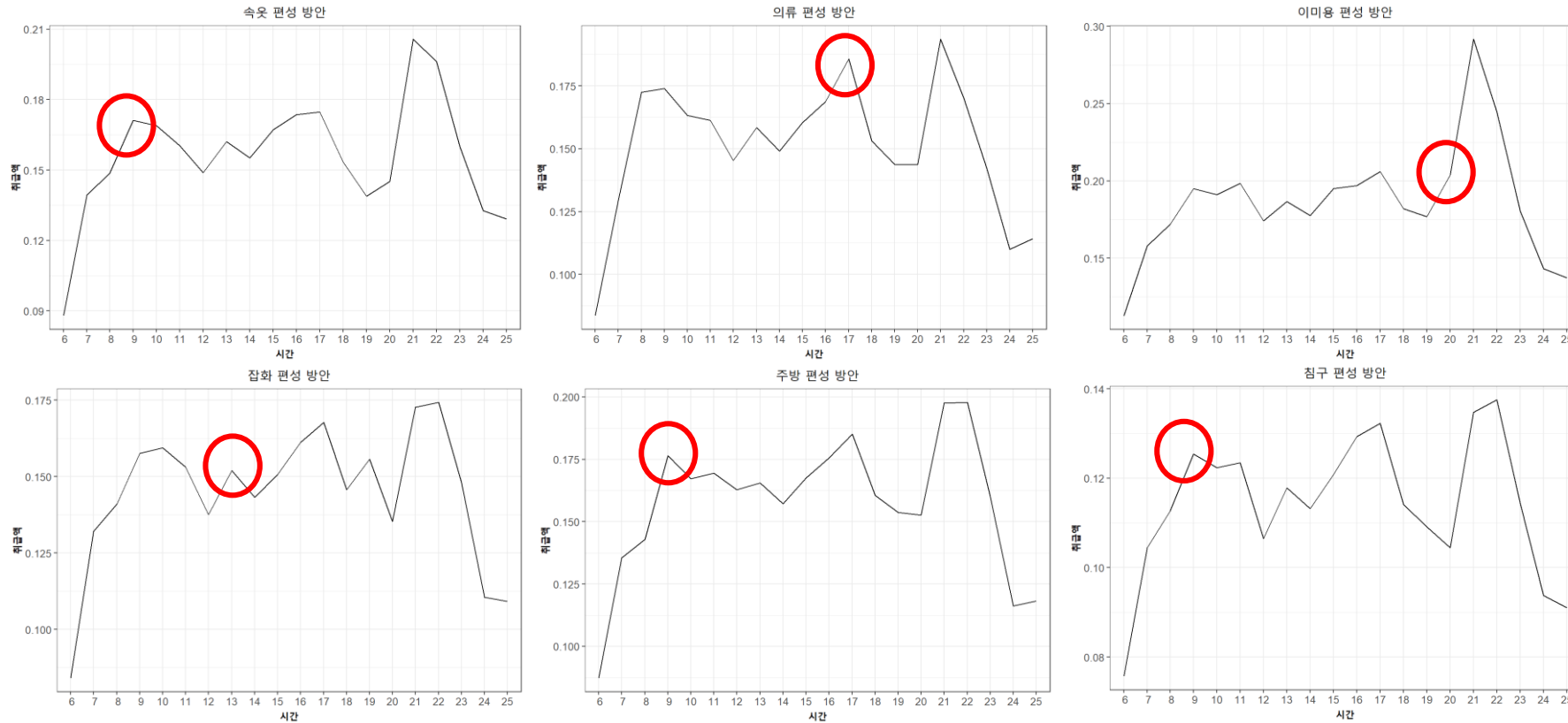
상품군	시간대
가공식품	11시
가구	09시
가전	09시
건강식품	13시
농수축	15시
생활용품	13시

- 카테고리별로 시간에 따른 그래프를 그려 어느 시간대에 특정 카테고리의 매출이 높은지 확인함
- 프라임 시간대에는 모든 상품군이 높은 매출액을 그리고 있음

# 3. 매출 최적화 방안

## 3. 카테고리별

[프라임 시간대를 제외한 상품군별 최고 매출 증가 시간대]



상품군	시간대
속옷	09시
의류	17시
이미용	20시
잡화	13시
주방	09시
침구	09시

- 카테고리별로 시간에 따른 그래프를 그려 어느 시간대에 특정 카테고리의 매출이 높은지 확인함
- 프라임 시간대에는 모든 상품군이 높은 매출액을 그리고 있음

# Summary & Review



# 4. Summary

## 4. Summary – 요일별 최적 편성 방안

요일	최적 편성 제안
월	농수축과 가공식품의 취급액 순위는 꾸준히 높고, 타 요일에 비해 냉장고의 취급액 순위가 높음
화	전통 건강식품(선식, 꿀, 홍삼 등)과 위생용품(바디케어, 샴푸/트리트먼트)의 취급액 순위가 높음
수	농수축 중 축산의 취급액 순위가 높음
목	냉장고, 신발, 반려동물 관련 상품, 주얼리 상품의 취급액 순위가 높음
금	농수축과 가공식품의 취급액 순위는 꾸준히 높고 신발의 취급액 순위가 하락함
토	신발, 종이컵/일회용품의 취급액 순위가 상승함
일	농수축과 가공식품의 취급액 순위는 꾸준히 높고 일회용품의 취급액 순위가 상승함

# 4. Summary

## 4. Summary – 시간별 최적 편성 방안

요일	최적 편성 제안
아침	모든 상품군에서 취급액이 전체적으로 증가하는 추세를 보임
아침 PRIME	건강식품, 의류, 생활용품, 속옷은 9시 이후 감소함
점심	건강식품이 14시까지 증가했다가 그 이후부터 감소함
오후	잡화를 제외한 나머지 상품군이 17시 이후로 감소함 / 잡화는 18시 이후 증가
오후 PRIME	가공식품은 22시까지 증가하고 이미용은 21시 이후로 감소함
새벽	농수축 상품의 취급액의 감소가 뚜렷함

# 4. Summary

## 4. Summary – 카테고리별 최적 편성 방안

상품군	시간대	상품군	시간대
가공식품	11시	속옷	09시
가구	09시	의류	17시
가전	09시	이미용	20시
건강식품	13시	잡화	13시
농수축	15시	주방	09시
생활용품	13시	침구	09시

- 위 표는 프라임 시간대를 제외한 상품군별 최고 매출 증가 시간대를 나타냄
- 주어진 시간대에 상품을 편성할 시 큰 폭의 매출 증가를 기대할 수 있음

# 4. Summary

## 4. Review

### — 한계점

- 1) 1년치의 데이터로 정확한 예측을 기대하기는 어려움
- 2) 더욱더 정확한 예측을 위해서는 상세한 상품 분류가 필요함
- 3) 2019년에는 존재하지 않았던 코로나 관련 이슈를 2020년 test data에 적용하기 어려움. 이를 보완하고자 2019년 데이터에서 총 매출액과의 0.34의 상관계수를 보이고 코로나 이슈로 인해 증가 추세를 보인 우체국 택배 접수량 변수인 'package'를 사용하였으나 2020년 데이터에 유의미한 영향을 주었을지는 판단하기가 어려운 면이 있음

**Thank You!**