

포스코 AI·Big Data 아카데미 26기

포유마켓의 유기농 시장 경쟁력 재고와 고객맞춤형 서비스 도입을 통한 매출 증대

A4조 손정우, 조현서, 길배섭, 정재원, 노성훈





포유 마켓



5년 만에 용인 수지 지역에서 4개의 점포를 늘릴만큼 급성장한
"친환경 식품, 물품 판매 중형마켓"

위 치



용인시 수지구 지역

영업 시간



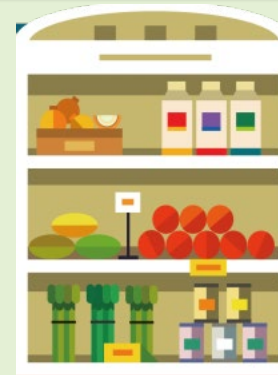
연중무휴
오전 8시 - 오후 8시

서비스



모바일 배달 서비스

운 영



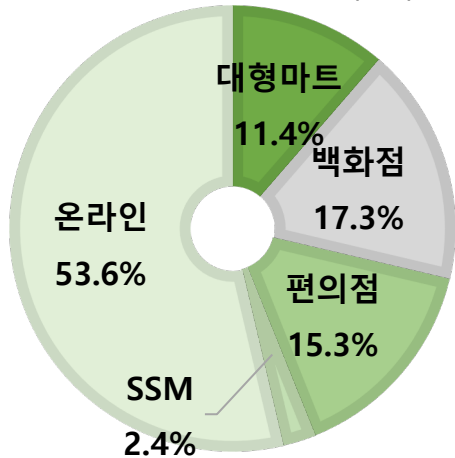
오프라인 매장
4곳 운영

추진 배경

“22년도 매출액 감소로 인해 매출 증대 전략 필요”

시장 과열로 인한 경쟁우위 하락

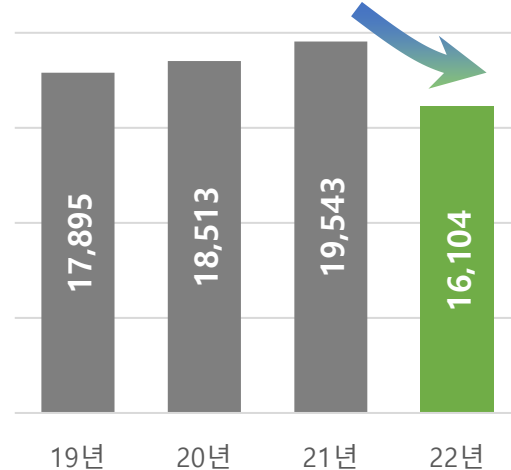
- 산업통상자원부 '23년 11월



< 주요 유통업체 매출 동향 >
대형마트로 인한 경쟁 과열로
자사 시장 경쟁력 저하

당사의 매출액 감소

단위(백만원)

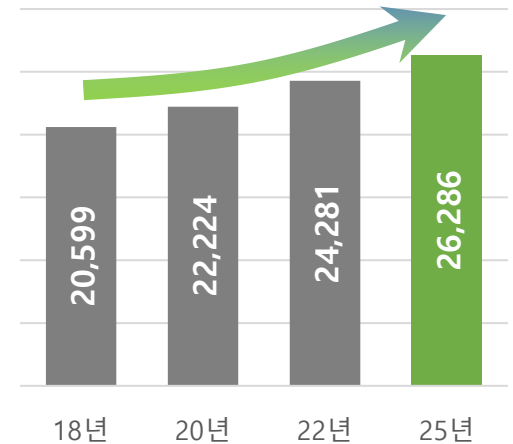


- 자사 매출 현황 자료

소비자 심리지수, 물가 지수 변화에 따른
당사 매출 10% 하락

친환경 유기농 시장 성장

단위(억원)



- 한국농촌경제연구원, 친환경 농산물 시장 규모

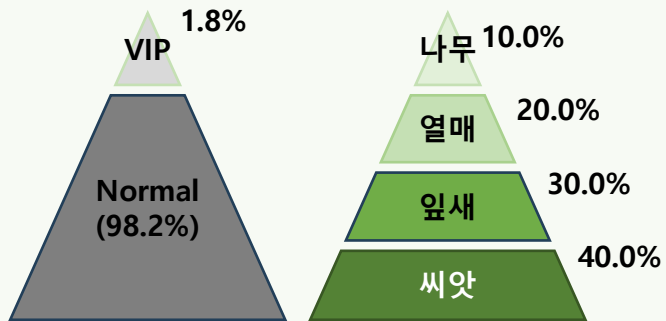
미세먼지와 환경오염 문제로
관심 및 수요 증가

고객 needs와 소비 트렌드 분석을 통해 경쟁우위를 확보 및 매출 증대

현상 및 개선 기회

“고객 needs에 적합한 서비스의 부족으로 매장 경쟁력 저하”

[효과 없는 멤버십 제도]



기존의 멤버십 제도는 CRM에 필요한 고객 분류를 충분히 나타내지 못함

→ RFM 분석 및 고객 특성에 기반한 VIP고객 매출 향상

[친환경, 유기농 제품 경쟁력 부족]



유기농 주요 상품의 믿음과 소비자들의 친환경적 소비가 증가함에 따라 경쟁력 부족을 극복

→ 친환경, 유기농 관련 제품의 라인

[고객 구매패턴을 활용한 판매 부진]



시즌 세일 및 추천 알고리즘을 통한 구매율 향상

→ 연관 상품 진열을 통한 구매촉진
→ 제품 추천 통한 매출액 증대

개선목표 : 매출액 30% 향상 ('22년도 총 매출액: 16,104백만원 → 20,935백만원)

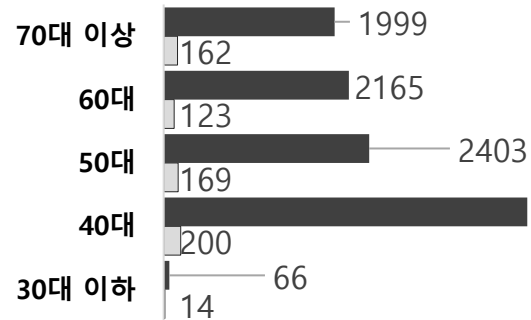
VIP 고객매출	유기농 제품 판매 매출	제품추천 구매율
10% 향상	30%증가	70% 향상

분석계획

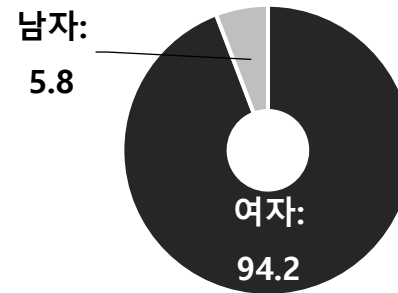
“아래와 같이 고객 데이터와 판매데이터를 활용하여 분석에 활용”

고객데이터 (11555 * 8)

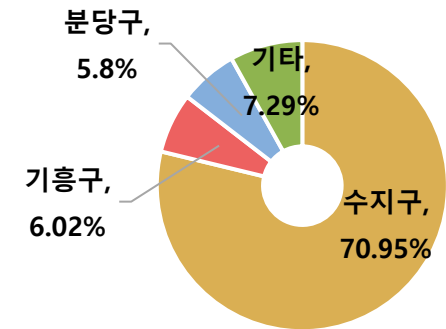
연령대 및 성별 분포



전체 남녀 분포

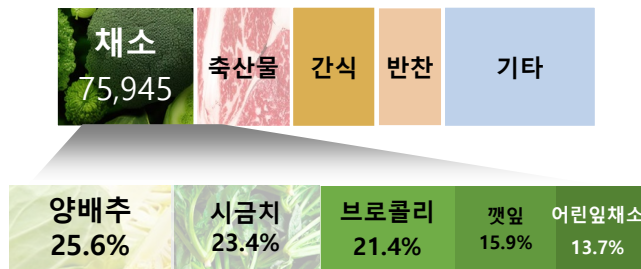


지역별 분포

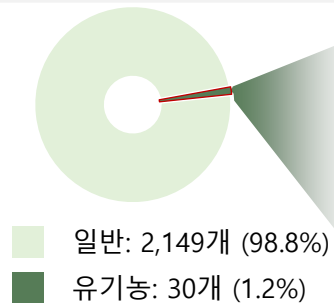


판매데이터 (294645 * 14)

주 판매 품목

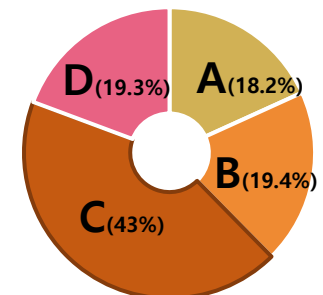


전체 품목 대비 유기농 제품 개수



판매순	품목
1	유정란
2	맵쌀
3	우유/산양유
4	논지엠오유정란
5	유기한우

매장 별 매출



* 22년도 총매출 : 16,104 백만원

* 총 4개 매장 (A, B, C, D)

데이터 분석 계획 수립

“아래와 같이 분석방법을 수립함”

데이터셋: member_data.csv (11555 *8), product_data_001~004.csv (294645 *14)

목적	분석방법	분석 내용	담당자	일정
RFM으로 도출된 VIP 매출증대 방안 수립	K-means 군집분석	각 군집의 가중치의 합을 1로 환산하여 각 군집의 최종 가중치를 확정	정재원	~ 6/3
	막대그래프	'나무', '열매' 등 고객들의 주 구입 품목 분석, 주 구매 시간 파악	길배섭	
	연관분석	RFM 등급별 고객의 주 구매패턴 파악 및 유의미한 추천을 통한 매출 증대 (ex. 두부를 산 후 콩나물을 사는 패턴)	정재원 조현서	
	상관분석	연관분석에서 도출된 confidence를 이용하여 주요 변수를 추출하고 모델에 적합할 변수를 지정	손정우	~6/4
고객 구매 특성 분석을 통한 프로모션 기획	막대그래프	시간대 및 요일별 고객의 구매 상품 및 매출 시각화	조현서	~6/4
	연관분석	성별과 연령대를 기준으로 집단별 연관분석을 실시하여 유의미한 추천을 통한 매출 증대	길배섭 조현서	~6/4
	상관분석	연관분석에서 도출된 confidence를 이용하여 주요 변수를 추출하고 모델에 적합할 변수를 지정	손정우	~6/4
유기농 친환경 상품의 매출 증진을 위한 프로모션 기획	막대그래프	전체 판매 제품 중 유기농 관련 상품의 비율 파악 대략적인 증감 정도 확인	조현서 노성훈	~6/4
	시계열분석 ARIMA PROPHET	시계열분석을 통해 유기농 관련 상품의 추후 매출액 증감 경향을 파악할 수 있음	조현서 길배섭	~6/4

데이터 분석 _ ① 멤버십 제도

“고객 멤버십 등급을 분류하여 ‘나무’등급에 해당하는 고객층의 매출 증가 ”

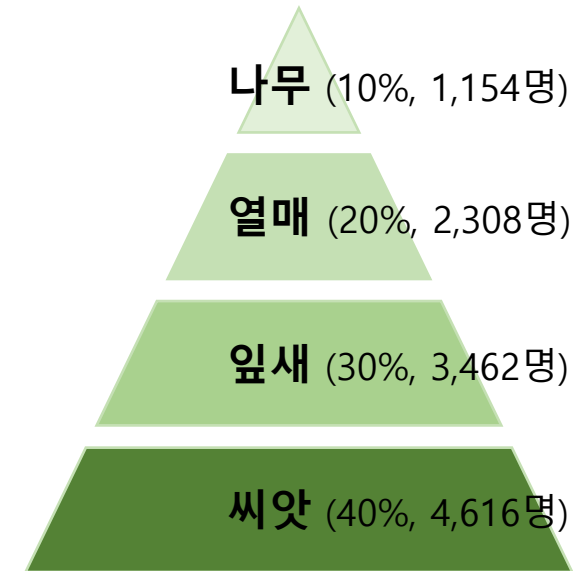
Recency: 얼마나 최근에 구매했는가 | **Frequency:** 얼마나 자주 구매했는가 | **Monetary:** 얼마나 많은 금액을 지불했는가

구분	R	F	M
5	14일 이하	5일 미만	37,071원 이하
4	14일 - 30일	5일 - 13일	37,071원 - 96,487원
3	30일 - 60일	13일 - 26일	96,487원 - 241,249원
2	60일 - 90일	26일 - 279일	241,249원 - 548,101원
1	90일 이상	279일 이상	548,101원 - 14,207,898원

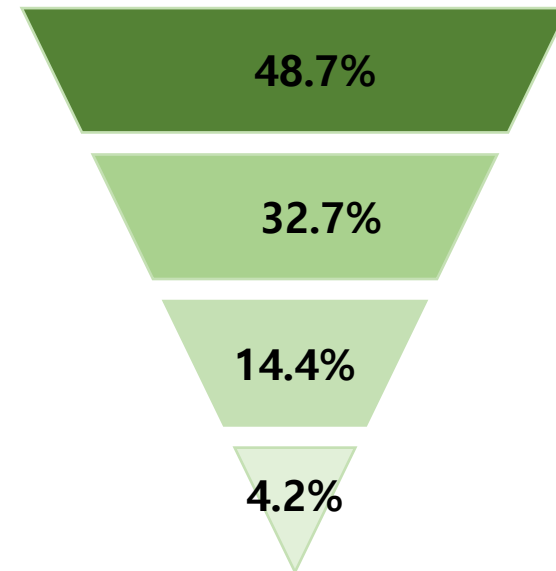
< RFM score >

K-means Clustering 분석을 활용하여 가중치 산출

$$0.268 * R + 0.358 * F + 0.373 * M$$



[멤버십 구성 비율]



[매출액 비중]

데이터 분석 _ ① 멤버십 제도

“멤버십 등급 중 ‘나무’ 고객의 시간대별 구매 특성 및 주요 구매 품목은 아래와 같이 나타남”

나무

총 매출액의 **48.7%** 기여

열매

충성 고객 만족도 평가
의 핵심 표본

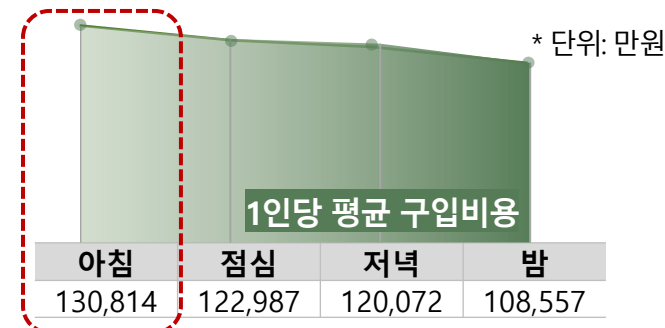
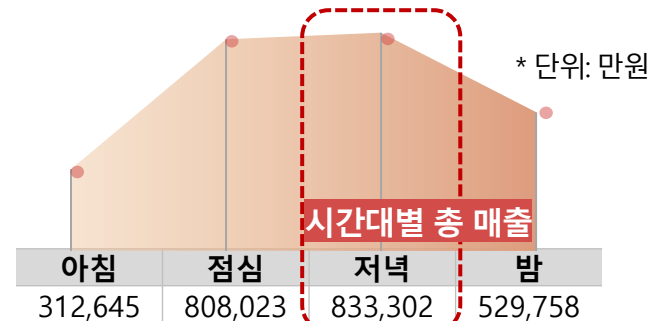
잎새

이탈이 쉬움. 경쟁사와의 우위를
판별할 수 있는 지표로 활용

씨앗

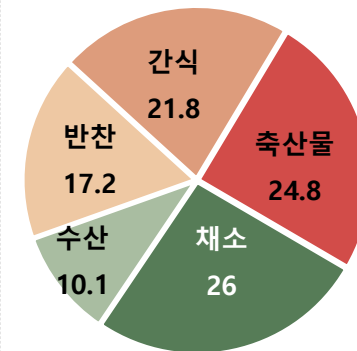
초기 구입 할인 이벤트 등을
통해 진입 유도

구매 시간대 분석

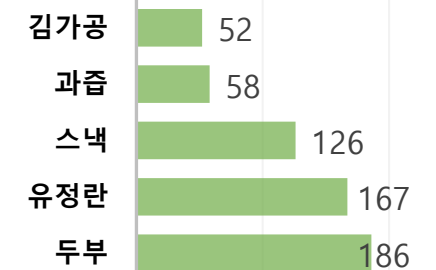


구입 품목

< 물품대분류 별 구매비율 >



< 물품소분류 별 구매비율 >



< 연관분석 결과 >

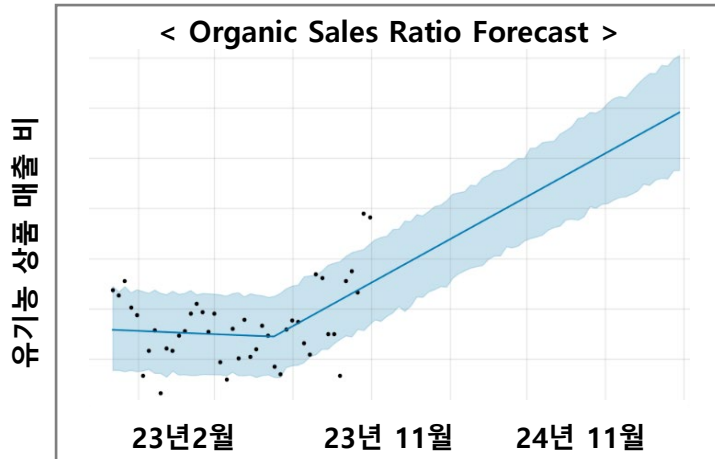
상품A	애호박	양파	라면	
상품B	오이	당근	스낵	...

→ “나무” 고객의 주 구매 시간대와 주 구입 품목을 분석하여 맞춤형 서비스 제공

데이터 분석 _ ② 유기농 제품

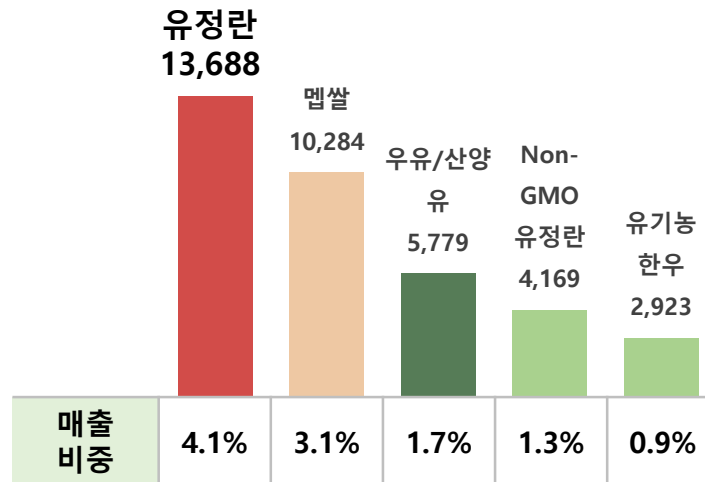
“유기농 제품은 판매량이 급격하게 증가하는 추세를 보이며, 판매품목을 분석해 매출을 증대함”

1/ 유기농 제품의 판매량 증가



시계열분석을 통해 '23년 11월 이후 유기농 상품의 매출이 꾸준히 성장할 것으로 예측

2/ 매출 기여도가 높은 유기농 제품



유기농 상품 중 Best Seller Top5 및 전체 매출 대비 기여도

3/ 유기농 제품 연관분석

상품A	상품B	향상도(Lift)
오이	애호박	2.0
<u>양파</u>	<u>당근</u>	<u>1.7</u>
산양유	요구르트	1.59
...		
산양유	<u>식빵</u>	<u>1.27</u>

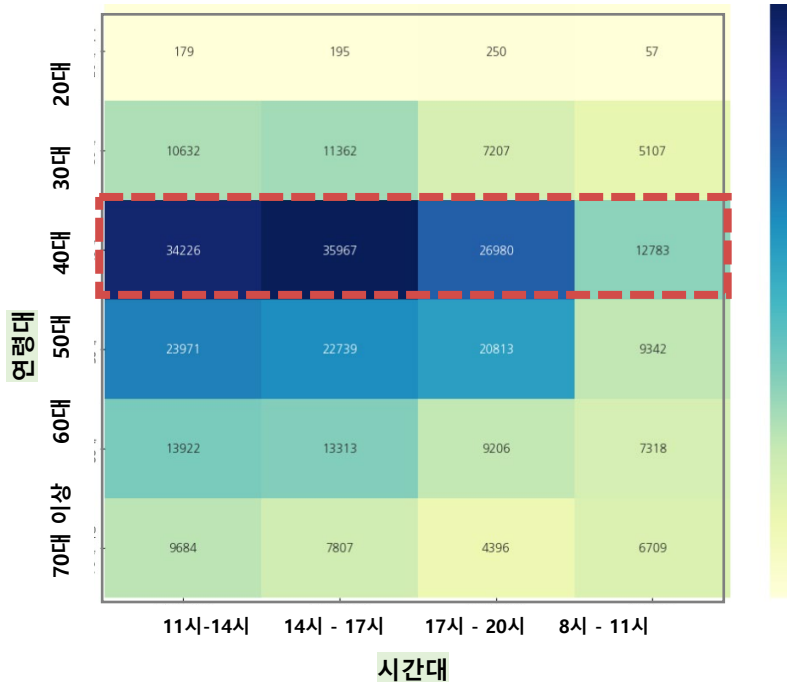
유기농 제품 간 향상도 1.25이상인 상품 식별

→ 유기농 상품판매의 증가를 예측하여, 유기농 상품의 개수 증가 및 연관분석을 통해 매대 진열, 묶음 상품, 제품추천에 적용

데이터 분석 _ ③ 고객

“주 고객의 구매 시간대와 구매 상품을 식별하여 타임세일 등 적용 프로모션 기획”

시간대별 구매 품목



주 고객층 40대 여성의 구매시간
11시 - 14시, 14시 - 17시

주요 구매 품목

< 11시 - 14시 >

물품명	판매횟수
두부	1,078
콩나물	698
유정란	622
찌개용두부	425
오이	374

< 14시 - 17시 >

물품명	판매횟수
두부	1,273
유정란	737
콩나물	694
찌개용두부	519
양파	361

주 구입 시간대 11시 ~ 17시

시즌별 구매 품목 분석: 여름



- 나이대: 40 대
- 성별: 여성
- 직업: 주부
- 성함: 홍○○

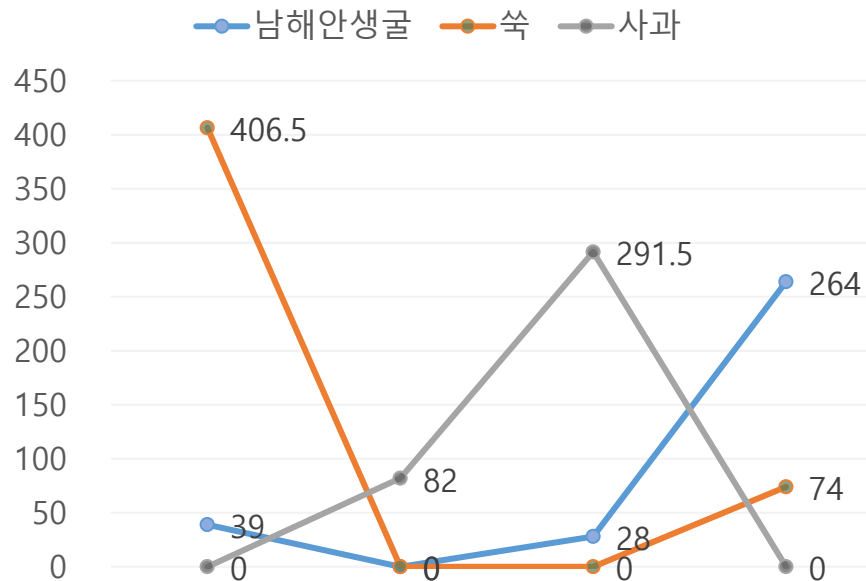


주 구매시간에
한시적 타임세일을 하여
구매 촉진

데이터 분석 _ ④ 계절에 따른 프로모션 진행

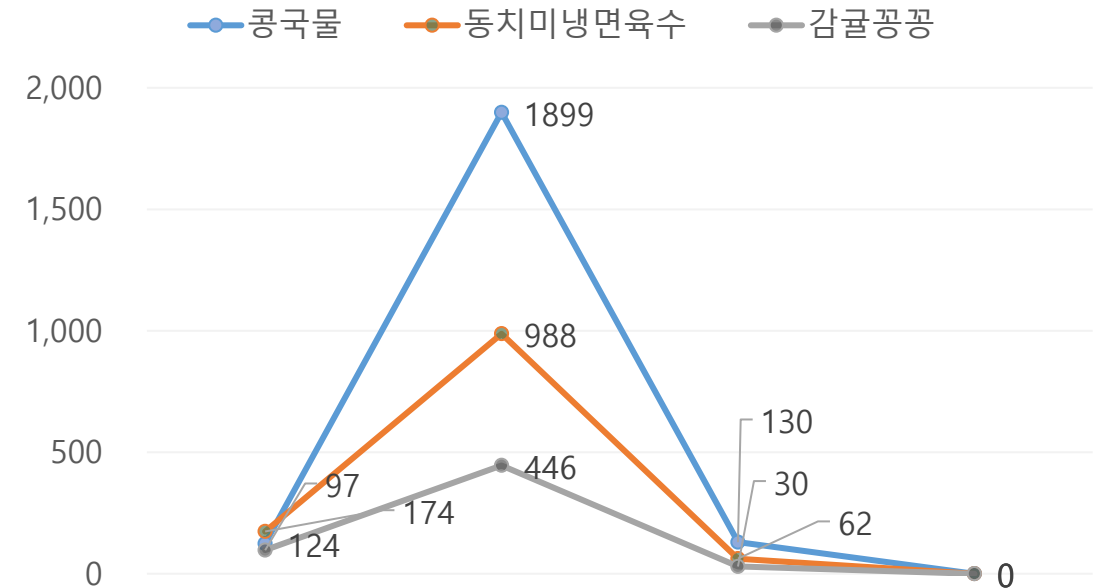
“계절에 따른 구매 매출 차이가 큰 제품을 분석하여 시즌세일 적용”

시즌별 구매 품목 분석 : 봄, 가을, 겨울



상품	봄	여름	가을	겨울
쭈욱	406.5	0.0	0.0	74.0
사과	0.0	82.0	291.5	0.0
남해안생굴	39.0	0.0	28.0	264.0

시즌별 구매 품목 분석: 여름



상품	봄	여름	가을	겨울
콩국물	124	1899	130	0
동치미냉면육수	174	988	62	0
감귤꿀꽂꽂	97	446	30	0

개선안 및 적용방안

“아래와 같이 분석결과를 종합하여 개선안을 도출함”

고객 등급

- 가장 높은 비중을 차지하는 **“나무”**는 신선식품을 구입함
- 저녁 시간대의 총 매출이 높음 (구입하는 사람이 많음)
- 아침에는 적은 인원 사람들의 구매금액 가장 높음

유기농 식품

- '23년 11월 이후 **판매량이 증가**할 것으로 예측함
- 연관분석 결과 연관성이 있는 상품들이 있음

고객 분석

- 주 소비자층은 **40대 여성**으로 아침 시간대에 구입을 함
- 두부, 콩나물, 유정란 등의 식품을 주로 구매
- 계절별 선호가 뚜렷한 쑥, 콩국물, 굴 등 있음

1 매대 진열

- 연관분석을 통해 도출된 품목을 함께 배치
- 계산대 근처에 잘 팔리는 상품을 배치하여 구매 유도

2 모바일 서비스

- 세일 관련 알림 서비스
- 제품 정보 조회 서비스 : 연관성이 있는 제품들에 한해 레시피 제공
- 연관성 높은 제품 추천
- “나무” 등급 관련 중요도 조사

3 각종 프로모션

- 고객 구매 시간대를 이용한 타임세일 도입
- 묶음 상품 판매를 통한 구매 촉진
- 유기농 제품 샘플 제공을 통해 구입 유도

개선안 및 적용방안 _ ① 매장

“매대진열 방안 및 프로모션을 통해 매장매출 증진”

저가 상품 진열

저가 인기상품 구매 촉진을 위한
Last minute 계산대 인근 진열



- 1 두부(총 판매 1위 상품), 가격 2,200원
- 2 스낵 (총 판매 3위), 가격 3,000원 내외
- 3 과일즙(총 판매 3위), 가격 10,000원 내외

핵심 제품의 배치

매장의 가장 안쪽에 핵심 제품을 진열하여,
고객의 동선을 통해 추가 구매 유도

<연관분석>



상품A	상품B
오이	애호박
양파	달근
산양유	요구르트
산양유	식빵



<매장 가장 안>

계절별 특별상품 프로모션



매장 중 가장 핵심적인
진열대에 계절별 주요
판매 상품 진열



연관상품 진열

<연관분석>

상품A	상품B
오이	애호박
양파	달근
산양유	요구르트
산양유	식빵



달근, 양파



식빵, 산양유



...

개선안 및 적용방안 _ ② 모바일

“모바일 서비스를 통해 VIP ‘나무’ 고객층을 차별화하고 다양한 기능을 통해 매출을 증진 ”

VIP “나무” 서비스

1. 신선식품 입고 알림

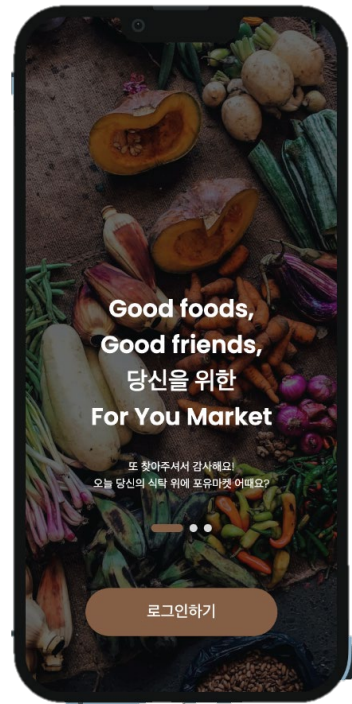
: 신선식품을 주로 구매하는 고객을 대상으로 “배려 마케팅” 수행

배강국 “나무”등급 고객님!
[사과 1KG] 제품이 입고되었습니다.
• 수확일자: 24년도 10월 13일
• 입고일자: 24년도 10월 15일

2. 계절상품 사전예약

: 계절별 주요 상품인 썩, 콩국물, 사과, 굴 등의 상품을 우선확보할 수 있도록 사전예약제도 운영

VIP 혜택



배송서비스

알림서비스

1. 매장별 실시간 재고 확인
2. 쿠폰 발급 및 알림
3. 프로모션 알림

추천서비스

1. 유기농 제품 추천
2. 가격행사 제품 추천
3. 연관 상품 추천

회원 커뮤니티

주민 소통창구, 정보공유, 건강 노하우 공유

레시피 추천 서비스

연관분석 상품 결과를 ‘레시피’로 함께 제시



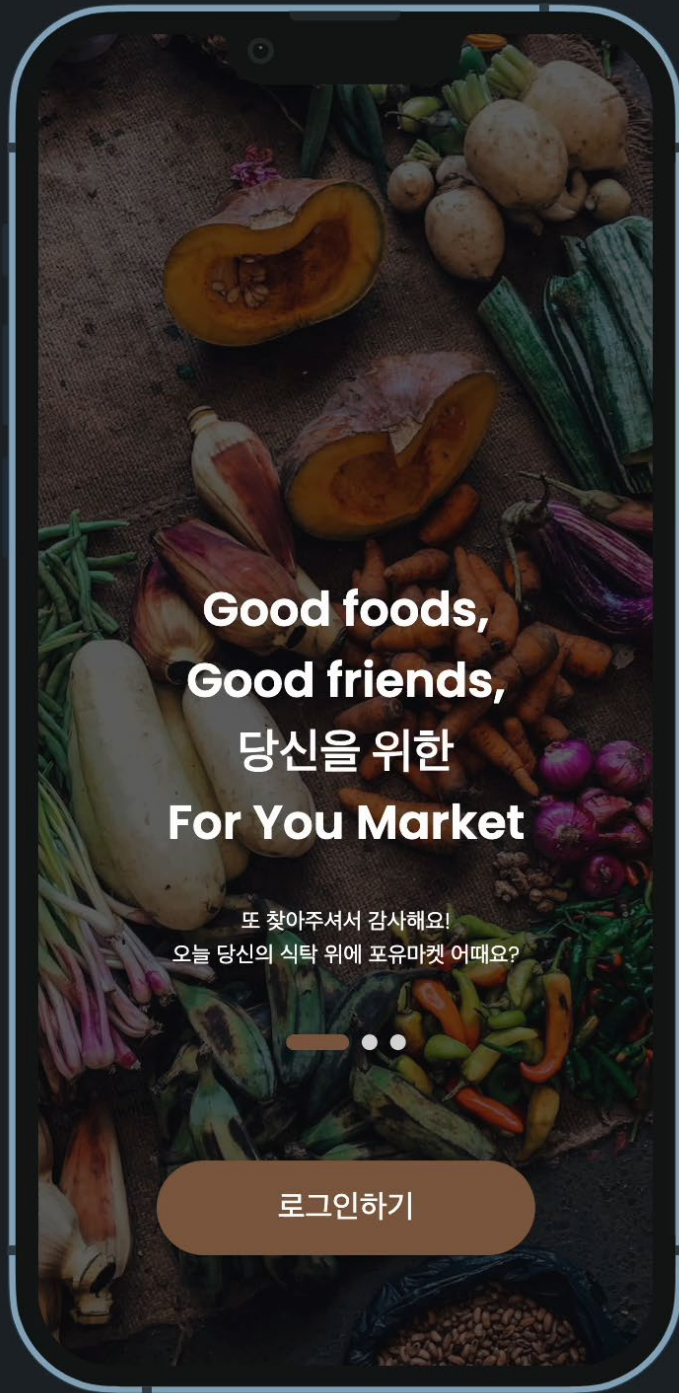
두부 구입



콩나물국 레시피

- 콩나물 200g,
- 두부 1/2모
- 대파 1대
- 멸치된장 1.5스푼

:



시계열 분석

Q1. 시계열 분석할 때 이상하게 튀었던 값은?

A1. 2023년 10월 마지막 주와 11월 첫 주에 특이한 값이 있었으나, 이전 4주간의 증가 추세를 고려해 제거하지 않음.

Q2. 변수들 보면서 중요하다고 생각한 부분 중 누락된 부분은?

A2. AR모델의 P값이 중요하지만, 데이터를 대략적으로 보고 결정해 정확도가 떨어질 가능성이 있음.

Q3. 시도해보고 싶었지만 못했던 분석은?

A3. Seasonal ARIMA를 적용해 보고 싶었으나, 해당 데이터에 계절 요인이 없어 적용하지 못해 아쉬움.

변수선언

Q1. 정보를 어떻게 활용했는지?

A1. 유기농 제품의 비율과 매출을 분석하고, 매장의 판매 전략을 최적화하는 데 사용.

Q2. 분석을 통해 얻은 주요 인사이트는?

A2. 분석을 통해 매장에서 유기농 제품이 차지하는 비율과 제품의 매출 기여도를 파악함. 이를 통해 친환경 제품에 대한 소비자 선호도를 이해하고, 향후 제품 구성과 마케팅 전략에 반영.

RFM score

Q1. 어떤 원리를 활용하여 RFM score를 산출하였는가?

A1. SDM과 K-means 클러스터링을 활용하여 R, F, M에 가중치를 산출했다. RFM 가중치 산정은 로지스틱 회귀분석, 정성적 평가, 팔레토 법칙을 참고.

Q2. R / F / M 변수는 각각 어떻게 설정하였는가?

A2. R : 방문 빈도를 5등분하여 설정, M : 11개월간의 구매 금액을 기준으로, F : 최근 방문일을 90일로 설정. 논문 검색 결과, RFM 산출 기준이 다양하며, 분류 목적에 맞게 정성적으로 설정.

부록

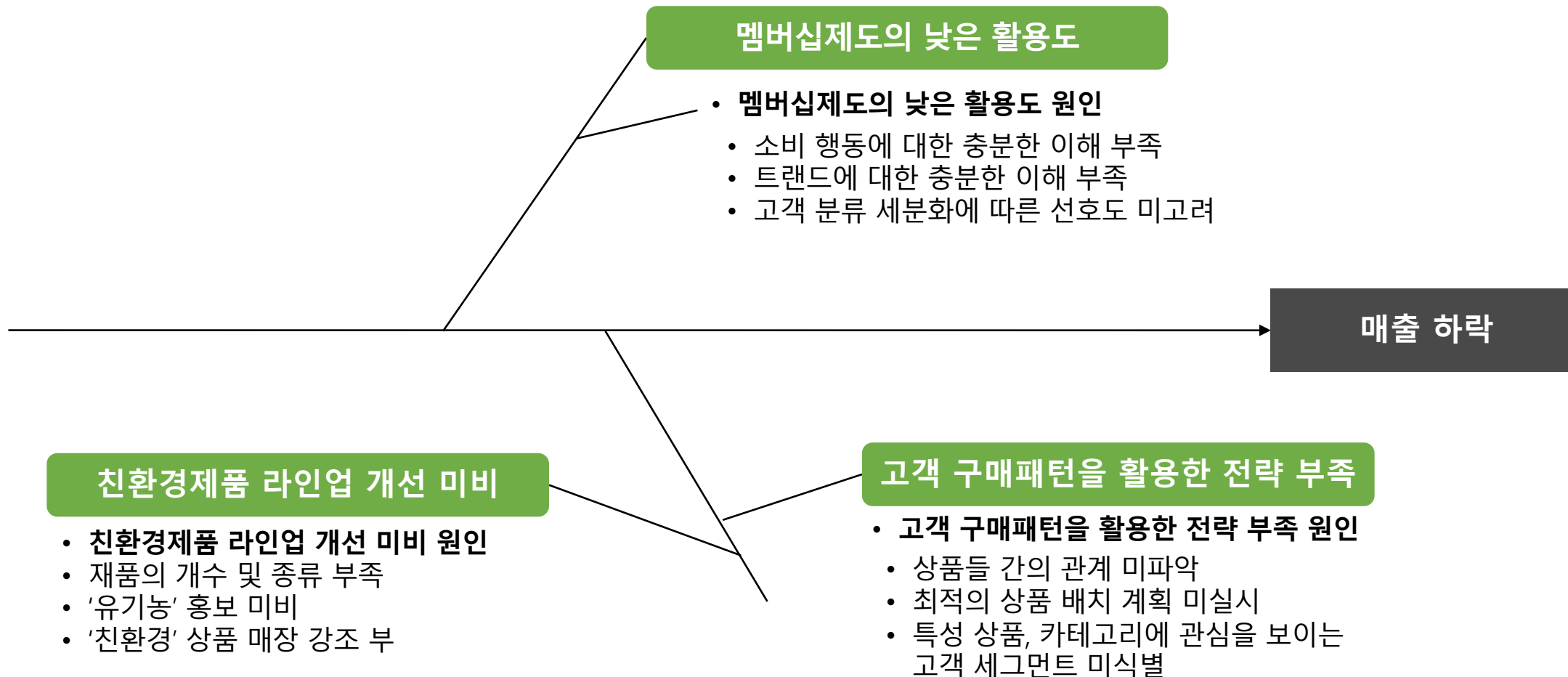
빅데이터분석을 통한 대형마켓 매출 증대

1. 잠재원인도출
2. 역할분담
3. 전처리
4. RFM 분석 방법론
5. ARIMA 모델 활용



잠재원인 도출

“KPI달성을 위해 특성요인도를 통한 잠재원인 분석 수행”



잠재원인 도출

“KPI와 특성요인도를 활용하여 잠재 원인 6건 도출”

KPI	잠재원인	중요도	분석가능성	Score	선정
멤버십제도의 낮은 활용도	소비 행동에 대한 충분한 이해 부족	9	3	12	O
멤버십제도의 낮은 활용도	트랜드에 대한 충분한 이해 부족	9	9	18	O
멤버십제도의 낮은 활용도	고객 분류 세분화에 따른 선호도 미고려	9	9	18	O
친환경 제품 라인업 개선 미비	제품의 개수 및 종류 부족	3	9	12	O
친환경 제품 라인업 개선 미비	‘유기농’ 홍보 미비	9	9	18	O
고객 구매패턴을 활용한 전략 부족	상품들 간의 관계 미파악	1	9	10	
고객 구매패턴을 활용한 전략 부족	최적의 상품 배치 계획 미실시	9	9	18	O
고객 구매패턴을 활용한 전략 부족	특성 상품, 카테고리에 관심을 보이는 고객 세그먼트 미식별	1	9	10	

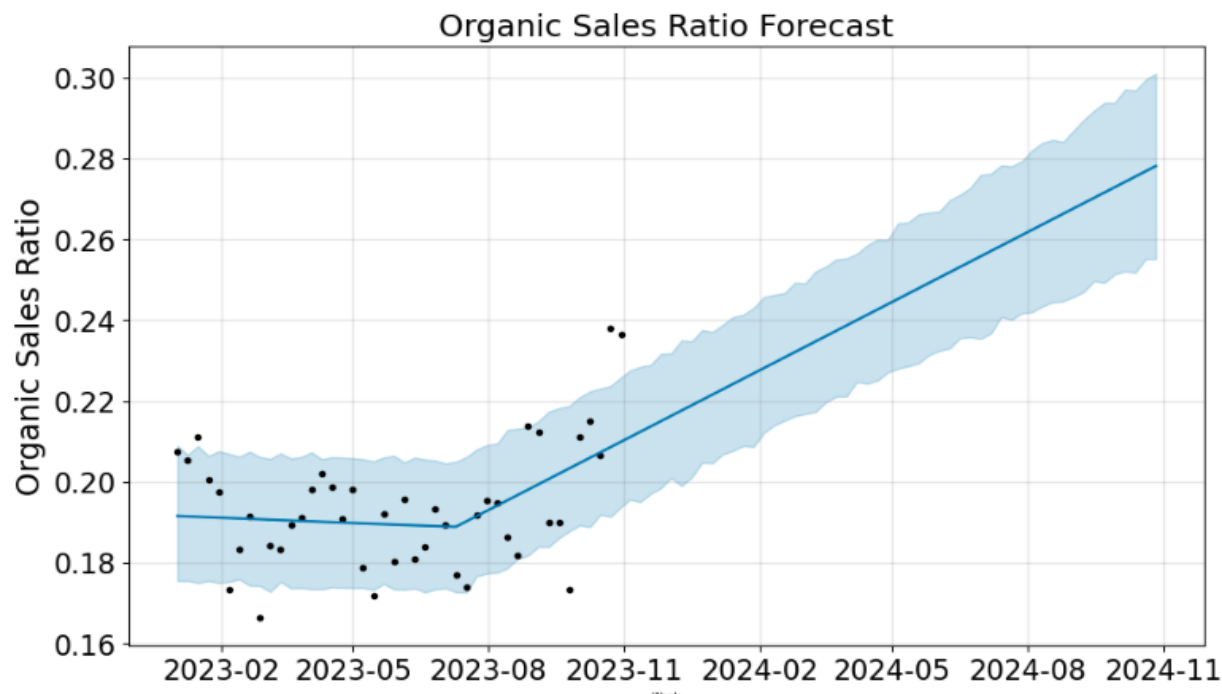
이름	역할	비고
손정우 (조장)	데이터 분석 계획 수립, 분석 프로세스 검토, EDA, 시계열 분석	
조현서	자료조사, 데이터 전처리, EDA, 시연용 앱 개발, EDA, 연관 분석	
노성훈	자료조사, EDA, 연관 분석, PPT	
길배섭	데이터 전처리, EDA, 연관분석, 시계열 분석	
정재원	자료조사, 데이터 전처리, EDA, 시계열 분석, PPT, PM	

데이터셋: member_data.csv (11555 *8), product_data_001~004 (294645 *14).csv

자료명	변수 명	변수 타입	이상치	결측치	확인결과	정제방안
고객정보데이터	회원번호	ID	-	-	-	-
고객정보데이터	조합원상태	범주형	-	-	-	-
고객정보데이터	주소	범주형	-	-	-	-
고객정보데이터	성별	범주형	-	2	회원 성별 'nan'값	행 제거
고객정보데이터	연령	연속형	-	2	회원 연령 'nan'값	행 제거
고객정보데이터	배송서비스신청여부	범주형	-	-	-	-
고객정보데이터	모바일알람신청여부	범주형	-	-	-	-
고객정보데이터	Gold_member	범주형	-	-	-	-
판매 정보 데이터	공급일자	datetime	-	-	-	-
판매 정보 데이터	회원번호	ID	-	-	-	-
판매 정보 데이터	조합원상태	범주형	-	-	-	-
판매 정보 데이터	물품대분류	범주형	-	-	-	-

데이터셋: member_data.csv (11555 *8), product_data_001~004 (294645 *14).csv

자료명	변수 명	변수 타입	이상치	결측치	확인결과	정제방안
판매 정보 데이터	물품중분류	범주형	-	-	-	-
판매 정보 데이터	물품소분류	범주형	-	-	-	-
판매 정보 데이터	물품명	범주형	-	-	-	-
판매 정보 데이터	구매수량	연속형	-	-	-	-
판매 정보 데이터	구매금액	연속형	-	-	-	-
판매 정보 데이터	구매매장	범주형	-	-	-	-
판매 정보 데이터	반품일자	datetime	-	424	'구매수량 < 0'인 데이터 중 반품일자가 누락	제거
판매 정보 데이터	구매시각	time	-	-	-	-

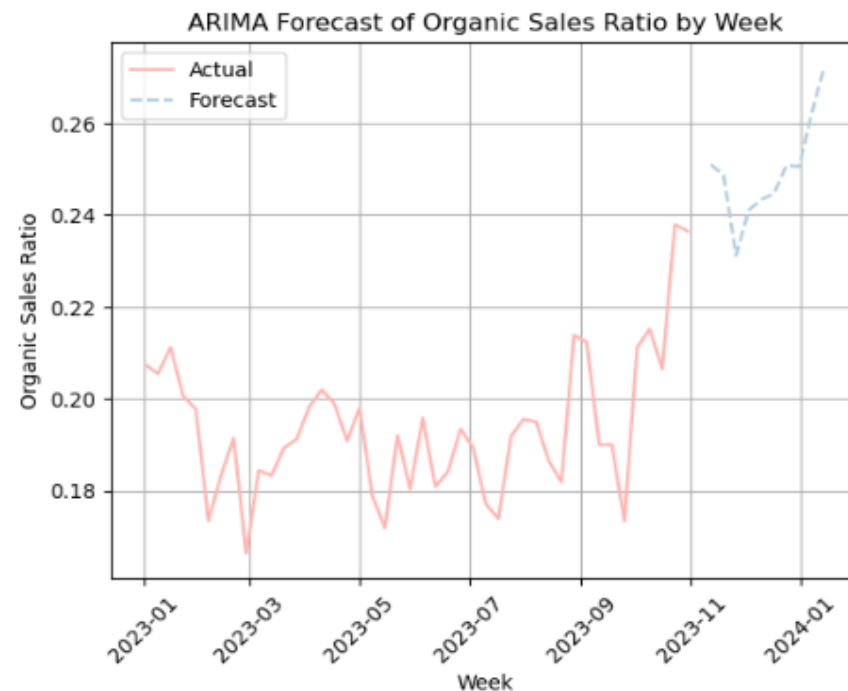


1) 이상치 판단

23년 10월 마지막주, 11월 첫 주의 값이 이전의 전반적인 경향과는 차이가 있음. 그러나 이전 12주의 경향이 증가하는 추세였기에 이상치로 판단하지 않고, 그대로 분석을 진행

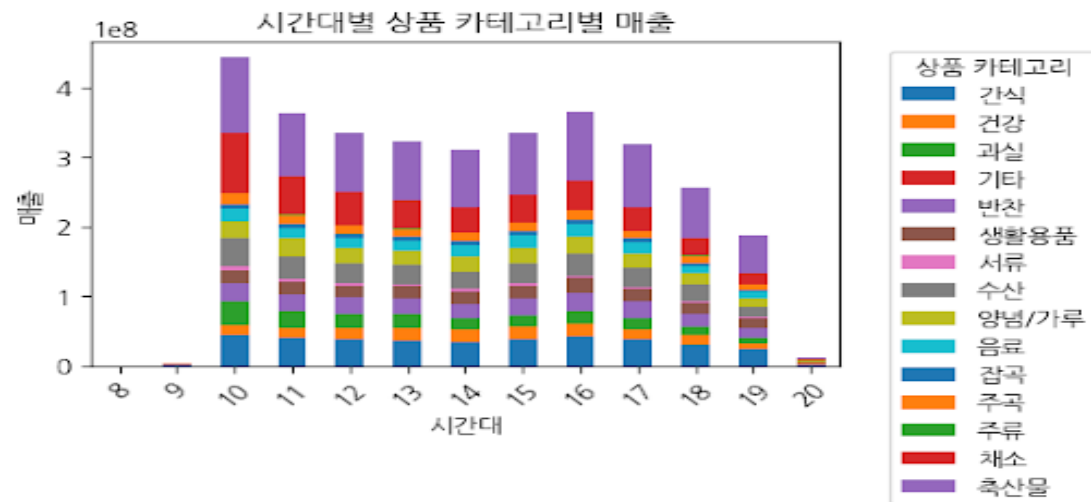
2) ARIMA 분석

Arima 분석에서는 자기회귀 정도에 사용되는 p 값이 가장 중요한데, 데이터의 크기를 확인하고 가장 적정하다 생각한 값인 12를 지정

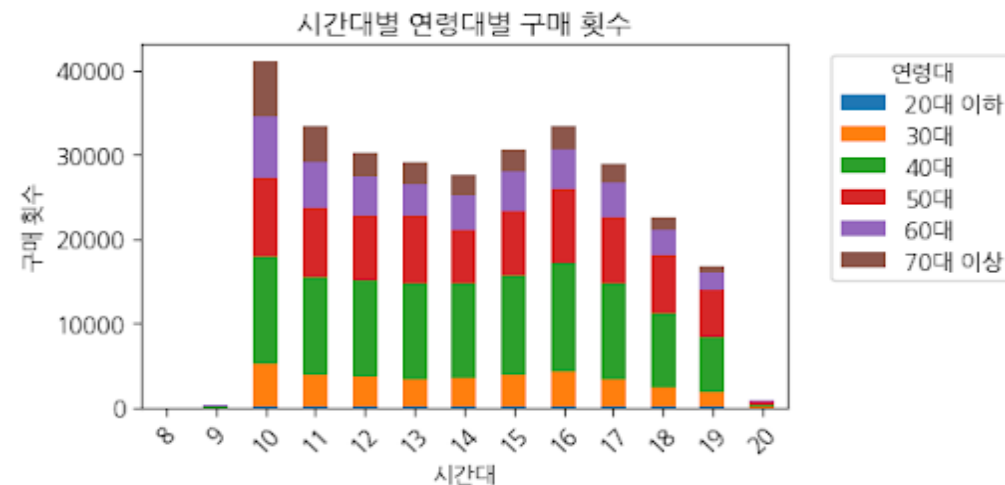


3) 추가로 실시해보고 싶은 분석

원 자료에서 특정 모양이 반복되는 계절요인이 있다면 S.arima를 적용해야하는데 이번의 데이터 셋에서는 발견되지 않았음.

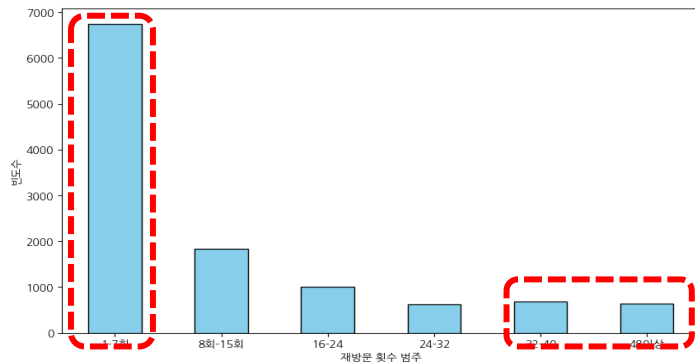


- 전체 매출: 10시에 가장 높고 시간에 따라 점차 감소하다가 15시 이후에 증가추세로 전환 16시에도 비교적 높은 매출을 보였습니다.
- 8~9시와 19시 이후 매출이 상대적으로 낮은 것을 확인할 수 있습니다.
- 카테고리별 매출 패턴: '축산물'이 가장 많이 판매된 영역을 차지하고 이어 채소가 뒤잇습니다.
- 시간대별 특이사항: 카테고리별 시간의 흐름에 따른 증감경향을 보려했지만 보이지 않았습니다. 따라서 모든 시간대에서 일정한 카테고리별 매출이 기록된 것을 확인했습니다.



- 전체 경향: 시간대별 상품 카테고리별 매출과 거의 동일한 분포를 보입니다. 이는 상품 카테고리별 평균 가격이 크게 다르지 않음을 의미합니다.
- 다양한 연령대 중 40대가 가장 많은 비중을 차지했고 50대가 뒤잇습니다.
- 8~9시와 19시 이후 구매 횟수가 상대적으로 낮은 것을 확인할 수 있습니다.

[재구매 횟수에 따른 고객 수 분포]



재구매횟수 1~7회

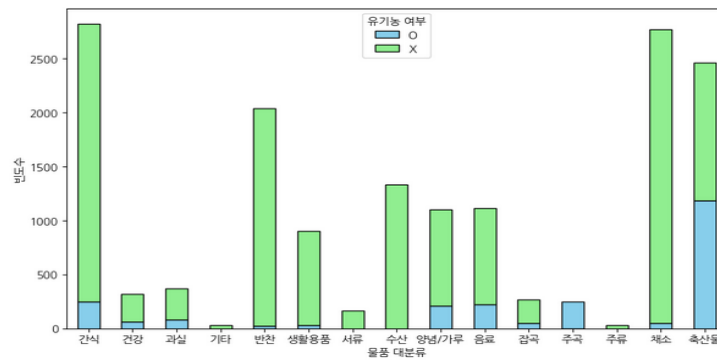
재구매횟수 32회 이상

시간대	방문자 수
14시~17시	5,000
11시~14시	4,500
17시~20시	3,000
8시~11시	1,000

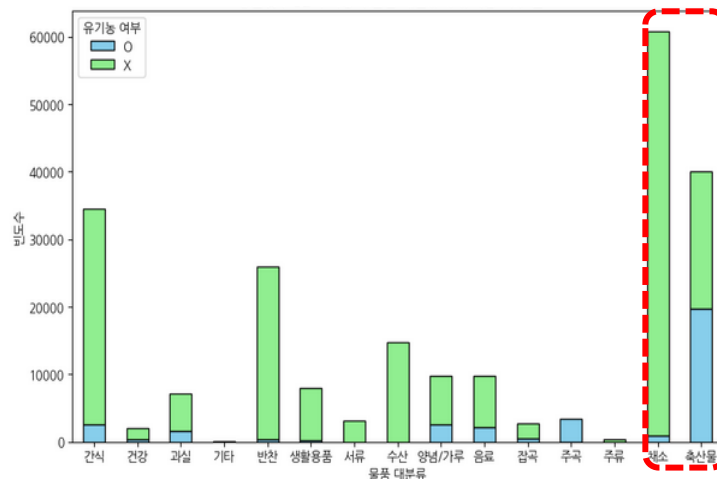
시간대	방문자 수
11시~14시	70,000
14시~17시	50,000
17시~20시	40,000
8시~11시	20,000

재구매횟수는 1~7회가 가장 많으며 횟수가 증가할수록 고객 수는 감소하는 경향을 보입니다. 재구매횟수 1~7회의 경우에는 14~17시 사이에 가장 많이 방문합니다. 또한, 재구매횟수 32회 이상의 경우에는 11~14시 사이에 가장 많은 고객님이 방문하시는 것을 확인 할 수 있습니다.

[재구매 횟수 1-7회]



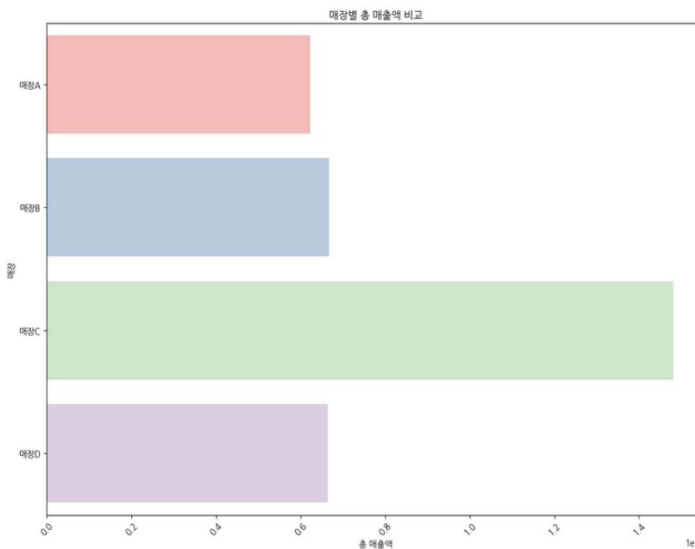
[재구매 횟수 32회 이상]



- 방문 횟수가 적은 고객의 경우, **간식**을 구입하는 경우가 가장 많았습니다. 또한, 방문횟수와 관계없이 채소와 축산물은 꾸준히 높은 수치를 보입니다.
- 이 집단에 대해 연관 분석을 시행한 후 묶음 상품 구성 및 번들의 형식을 통해 매출을 증진할 수 있습니다. 또한 매장 앞에 간식을 노출시켜 고객님들의 방문을 유도 할 수도 있습니다.

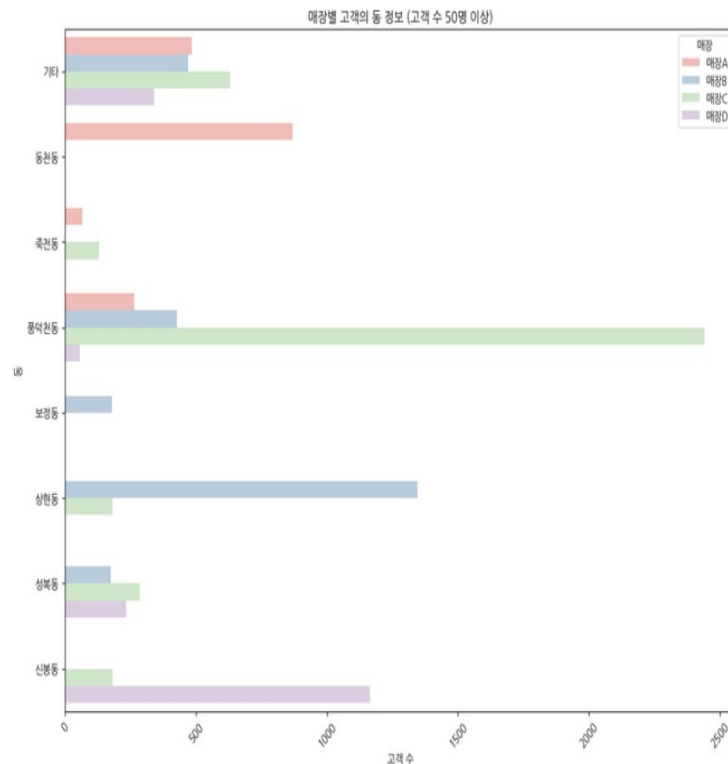
- 방문 횟수가 많은 고객의 경우, 식자재를구입하는 경우가 가장 많았습니다. 또한 채소와 축산물은 방문횟수와 관계없이 꾸준히 높은 수치를 보이지만 재구매 횟수가 적은 집단에 비해서도 높은 수치를 보입니다. 다만, 유기농 구입에 대한 비율은 차이가 없음을 확인할 수 있습니다.
- 채소 및 축산품 관련 재고의 신선도 유지 방안 개선을 통해 매출의 향상을 도모할 수 있으며, 모바일 어플리케이션의 상품 입고날짜 확인 시스템을 통해 재구매 횟수가 많은 고객님에게 더욱 편리한 매장이용을 추구할 수 있습니다.

- 매장별 총 매출액 비교



매장별 총 매출액 비교 그래프를 보면 c매장을 제외한 3개의 매장의 총 매출액이 동일한 것을 확인할 수 있고, c매장은 다른 매장의 2배 이상의 매출액을 올리고 있기 때문에 c매장만의 지리적 특성, 판매 품목 등에 대해서 추가적 조사가 필요할 것이라 생각합니다.

- 매장별 고객 거주지 합계



매장 A의 경우 동천동에 가장 많은 고객이 있는것을 확인할 수 있습니다. 또한 동천동 고객들은 오직 A매장만 이용하는 것으로 미루어보아 매장 A는 동천동에 있는 것을 확인할 수 있습니다.

매장 B의 경우 5개 이상의 동에 고객이 분포되어 있으며, 상현동에 대부분의 고객이 있는 것으로 추측해볼 때, 매장 B는 상현동에 위치해 있을 것으로 예상됩니다.

매장 c의 경우 수지구 대부분의 동네에 고객이 퍼져있는 것을 확인할 수 있습니다. 또한 풍덕천동에 가장 많은 고객이 있으며 수지구에서 가장 규모가 큰 동이 풍덕천동이기때 다른 매장에 비해 많은 매출을 올린 것이라 추측해 볼 수 있습니다.

매장 D의 경우 신봉동에 가장 많은 고객이 있으며, 다른 동에서의 회원이 적기 때문에 신봉동에 있을 것으로 예상 됩니다.